Ph.D. Thesis

Graph Matching Using Position Coordinates and Local Features for Image Analysis

Gerard Sanromà Güell

Departament d'Enginyeria Informàtica i Matemàtiques

Universitat Rovira i Virgili

This thesis has been written in El Pont d'Armentera (Tarragona),
December 2011

Gerard Sanromà Güell

Graph Matching Using Position Coordinates and Local Features for Image Analysis

Ph.D. Thesis

Advisors: Francesc Serratosa i Casanelles[a] and
René Alquézar Mancho[b]

[a] Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili.
[b] Institut de Robòtica i Informàtica Industrial, CSIC - Universitat Politècnica de Catalunya.

Universitat
Rovira i Virgili

This thesis has been written in El Pont d'Armentera (Tarragona),
December 2011

UNIVERSITAT
ROVIRA I VIRGILI

DEPARTAMENT D'ENGINYERIA INFORMÀTICA
I MATEMÀTIQUES

Av. Països Catalans, 26
43007 Tarragona
Tel. 34 977 559 703
Fax 34 977 559 710
secdeim@urv.cat

FEM CONSTAR que aquest treball titulat "Graph Matching Using Position Coordinates and Local Features for Image Analysis" que presenta en Gerard Sanromà Güell per a l'obtenció del títol de Doctor, ha estat realitzat sota la direcció del Dr. Francesc Serratosa i Casanelles del Departament d'Enginyeria Informàtica i Matemàtiques d'aquesta universitat i sota la direcció del Dr. René Alquézar Mancho de l'Insitut de Robòtica i Informàtica Industrial de la Universitat Politècnica de Catalunya.

Tarragona, 3 de desembre de 2011

El director de la tesi doctoral

Dr. Francesc Serratosa i Casanelles
Universitat Rovira i Virgili
Catalunya

El co-director de la tesi doctoral

Dr. René Alquézar Mancho
Universitat Politècnica de

Acknowledgements

First of all I wish to express my profound gratitude to Dr. René Alquézar and to Dr. Francesc Serratosa for their commitment with this project and for their friendship. Without their support this thesis would have never been written. I also want to thank Dr. Blas Herrera for the helpful discussions about geometry. It has been a pleasure to collaborate with him. I wish to thank Universitat Rovira i Virgili and the Departament of Computer Science and Mathematics for the economic sustenance through a pre-doctoral scholarship at first, and a position as assistant professor at this time. I am also grateful to Insititute of Robotics and Industrial Informatics (CSIC-UPC) for furnishing the infrastructure where most of our meetings did take place and for granting me access to their computational grid.

I wish to thank both to my family and to my girlfriend's family for their unconditional support and confidence. The last words are of gratitude to my girlfriend Mercè, my most important support.

Agraïments

Primer de tot desitjo expressar la meva profunda gratitud al Dr. René Alquézar i al Dr. Francesc Serratosa pel seu compromís amb aquest projecte i per la seva amistat. Sense el seu suport aquesta tesi mai no s'hauria escrit. També vull agraïr al Dr. Blas Herrera per les útils discussions sobre geometria. Ha sigut un plaer colaborar amb ell. Desitjo agrair a la Universitat Rovira i Virgili i al Departament d'Enginyeria Informàtica i Matemàtiques el sustent econòmic per mitjà d'una beca predoctoral al principi, i un lloc com a professor associat en aquests moments. També estic agraït a l'Institut de Robòtica i Informàtica Industrial (CSIC-UPC) per proveir la infraestructura on hem realitzat la majoria de les reunions i per facilitar-me l'accés al seu grid computacional.

Vull agrair tant a la meva família com a la de la meva parella el seu suport incondicional i la seva confiança. Les últimes paraules son de gratitud cap a la meva parella Mercè, el meu suport més important.

Abstract

Finding the correspondences between two images is a crucial problem in the computer vision & pattern recognition field. It is relevant to a broad range of purposes going from object recognition applications in the areas of biometry, document analysis and shape analysis to applications involving multiple view geometry such as pose recovery, structure from motion and localization & mapping.

Many existing techniques approach this problem either using local image features or point-set registration methods (or a mixture of both). In the former ones, a sparse set of features is first extracted from the images and then characterized in the form of descriptor-vectors using the local image evidence. Features are associated according to the similarity between their descriptors. In the second ones, feature-sets are regarded as point-sets which are associated using non-linear optimization techniques. These are iterative procedures that estimate correspondence and alignment parameters in alternate steps.

Graphs are representations that allow for binary relations between the features. Accounting for binary relations in the correspondence problem often leads to the so-called graph matching problem. There exists a number of methods in the literature aimed at finding approximate solutions to different instances of the graph matching problem, which in most cases is known to be NP-hard.

Regardless of the type of representation used, part of our work is devoted to the comparison of local image features. Specifically, we investigate the benefits of using cross-bin measurements such as the Earth Movers' Distance to that end. The rest of our work is dedicated to formulating both the image features association and point-set registration problems as instances of the graph matching problem. In all the cases, we propose approximate algorithms to solve these problems and compare to a number of existing methods from different areas, namely, outlier rejectors, point-set registration methods and other graph matching methods.

Experiments show that in most cases the proposed methods outperform the rest. Occasionally the proposed methods either share the best performances with some competing method or they get slightly worse results. In these cases, the proposed methods usually present lower computational times.

Resum

Trobar les correspondències entre dues imatges és un problema crucial en el camp de la visió per ordinador i el reconeixement de patrons. És rellevant per un ampli ventall de propòsits des d'aplicacions de reconeixement d'objectes en les àrees de biometria, anàlisi de documents i anàlisi de formes fins aplicacions relacionades amb geometria des de múltiples punts de vista tals com recuperació de pose, estructura des del moviment i localització i mapeig.

Moltes tècniques existents enfoquen aquest problema o bé usant característiques locals a la imatge o bé mètodes de registre de conjunts de punts (o bé una mescla d'ambdós). En les primeres, un conjunt dispers de característiques és primerament extret de les imatges i després caracteritzat en la forma de vectors descriptors usant evidències locals de la imatge. Les característiques son associades segons la similitud entre els seus descriptors. En les segones, els conjunts de característiques son considerats com conjunts de punts els quals son associats usant tècniques d'optimització no lineal. Aquests son procediments iteratius que estimen els paràmetres de correspondència i d'alineament en passos alternats.

Els grafs son representacions que contemplen relacions binaries entre les característiques. Introduir les relacions binàries al problema de la correspondència sovint porta a l'anomenat problema de l'emparellament de grafs. Existeix una gran quantitat de mètodes a la literatura destinats a trobar solucions aproximades a diferents instàncies del problema d'emparellament de grafs, el qual en la majoria de casos és del tipus "NP-hard".

Una part del nostre treball està dedicada a investigar els beneficis de les mesures de "bins" creuats per a la comparació de característiques locals de les imatges. La resta està dedicada a formular ambdós problemes d'associació de característiques d'imatge i registre de conjunt de punts com a instàncies del problema d'emparellament de grafs. En tots els casos proposem algoritmes aproximats per solucionar aquests problemes i ens comparem amb un nombre de mètodes existents pertanyents a diferents àrees com eliminadors d'"outliers", mètodes de registre de conjunts de punts i altres mètodes d'emparellament de grafs.

Els experiments mostren que en la majoria de casos els mètodes proposats superen a la resta. En ocasions els mètodes proposats o bé comparteixen el millor rendiment amb algun mètode competidor o bé obtenen resultats lleugerament pitjors. En aquests casos, els mètodes proposats normalment presenten temps computacionals inferiors.

Contents

Introduction

The correspondence problem is of pivotal importance in the Computer Vision & Pattern Recognition field. It plays an important role in many object recognition tasks in diverse areas such as biometry, shape analysis or document analysis. It is also crucial in many applications involving multiple view geometry such as pose recovery, structure from motion or localization & mapping.

The correspondence problem consists in identifying a set of objects by assigning labels from a label-set. We will consider the features extracted from an origin and destination images, the object-set and label-set, respectively.

We will focus on the following types of features: local image features and coordinate positions.

Local image features are built upon the local image evidence at salient points (chapter 1). Main research in local image features is focused on developing highly discriminative image features invariant to a number of image deformations such as viewpoint or illumination changes.

Matching of local features is usually posed as a linear assignment problem in which the sum of similarities (or distances) between the matched features must be maximized (or minimized).

While the matching of local image features aims to solve the correspondence problem, point-set registration aims to simultaneously solve the correspondence and alignment problems. From an optimization point of view, association of non-discriminant features such as coordinate positions is more complex than the case of local image features.

Once the underlying pose (alignment) parameters are known, the correspondence problem can be reduced to an instance of the linear assignment problem in which points in the origin set are associated with points in the destination set so that the sum of distances between the associated points is minimized. On the other hand, once the correspondences are known, estimation of the alignment parameters reduces to a least-squares estimation problem with known closed-form solutions for various types of geometric transformations (chapter 2). The case when neither the correspondence nor the alignment parameters are known leads to the point-set registration problem, a kind of chicken-and-egg problem for which efficient non-linear optimization techniques have been devised (chapter 3).

Most of the existing point-set registration methods rely on some initial correspondence (or alignment) estimates that might be located with the aid of feature-based approaches.

From a graph-theoretic point of view, approaches to the association of features can be divided according to whether they rely exclusively on unary measurements or they allow for relational information. Graph matching literature deals with the problem of locating the correspondences by taking into account binary relations while, in some cases, still being able to incorporate unary measurements as well (chapter 4).

Part I of this thesis covers aspects related to the association of local image features and position coordinates using either unary or binary measurements, or both.

Contributions

A widely followed approach to solve the correspondence problem between two images is to extract two sparse sets of features and associate each one from one set to its closest neighbour in the other set. Key to this approach is the choice of the distance measure used.

Local image features are commonly represented by means of distribution-based descriptors such as histograms. Distance measures between histograms are computed either in a bin-to-bin basis or in a cross-bin basis. The former ones assume that correspondences between bins are known in advance while the latter ones decide such correspondences during the process.

Since local image descriptors are often compared using bin-to-bin measures, our aim is to investigate whether the use of cross-bin distances such as the Earth Mover's Distance (EMD) is more appropriate for comparing local image feature descriptors.

In chapter 5 we propose an efficient algorithm to compute the EMD that is based on an heuristic that favours movements involving locations in the boundaries of the histograms.

The proposed algorithm presents a theoretical cost lower than similar approaches. We present image retrieval experiments and point-set registration experiments with *Shape Contexts* aimed at evaluating both the discrimination ability of the EMD-based metric and the effectiveness of the proposed algorithm to compute it. We also present an empirical study of the time complexity in order to assess the efficiency of the algorithm in real situations.

Correspondences between local features are usually decided by solving a linear assignment problem or a weighted bipartite matching problem based on the distances between unary local descriptors. However, allowing for binary relations usually leads to NP-hard optimization problems.

In chapter 6 we devise approximate graph matching methods aimed at investigating the benefits of allowing for binary relations between neighboring features to the problem of local image features matching.

We evaluate the matching performance of the proposed methods in a series of SIFT matching experiments with synthetic images. We compare to outlier rejectors as well as to point-set registration and graph matching methods.

The correspondence problem must be usually solved at some stage of an object recognition application. The recognition of handwritten characters is not an exception, and graphs provide natural means for representing such objects.

Graph representations from handwritten characters are usually derived from their skeletal descriptions. Such graphs are often sparse, and may induce to ambiguities to many structural models used for graph matching, even to the most distinctive ones such as the model by Wilson & Hancock (1997). This model gauges the structural consistency induced by a match by averaging the Hamming distances between the matching realizations of the cliques in a data-graph and the cliques in a model-graph.

In chapter 7 we propose an extension to the model by Wilson & Hancock (1997) that aims to improve the matching accuracy of the aforementioned types of graphs by using position coordinates as nodes' attributes. In order to be invariant to the specific pose of the point-sets we introduce the estimation of the similarity alignment parameters into the problem. The proposed formulation leads to a new measure of consistency based on Hamming and Procrustes distances. We investigate two optimization methods: discrete relaxation and genetic search.

We evaluate the matching ability of the proposed method using graphs extracted from handwritten letters.

There exist graph matching approaches to point-set registration that take into account binary relations. We will refer to this problem as simultaneous (or joint) structural graph matching and point-set registration. This problem is usually posed as an iterative process of alternate estimation of correspondence and alignment parameters.

From the standpoint of this thesis, the most relevant approaches to solve this problem use either the statistical estimation apparatus of the EM algorithm or deterministic annealing procedures in conjunction with Softassign. The former ones have the advantage of offering statistical insights of such decoupled estimation processes while the latter ones benefit from the well-known robustness and convergence properties of the Softassign embedded within deterministic annealing processes.

In chapters 8 and 9 we try to bridge the gap between these two families of approaches by proposing methods that formulate these processes within a principled statistical framework at the same time that they exhibit the desirable properties of the Softassign and deterministic annealing ensemble.

We evaluate the registration and recognition abilities of the proposed methods in a series of image and shape registration experiments as well as shape retrieval experiments with both real and synthetic data.

Part II is devoted to exposing our contributions to the development of graph matching methods aimed at exploiting evidence coming from coordinate positions and local image features.

All the methodologies developed in this thesis tolerate a certain disagreement between the corresponding unary measurements as well as the binary relations. The sources of errors are explained through the use of probability distributions.

Outliers are features in one set with no correspondence in the other set explainable under the actual error-assumptions. In fact, they are noisy measurements that do not necessarily follow any probability distribution and that deserve a special attention. Since outliers may dramatically affect to the matching performance, it is a constant concern in all our contributions to achieve a certain robustness against this type of noise.

Before we finalize the introduction, we wish to mention a few usual ways of extracting graph representations from images as well as the ones that we have adopted.

When extracting a graph-representation from an image, the nodes usually represent a sparse set of regions obtained by some kind of segmentation process (e.g., local features extraction). Nodes' attributes convey information relevant to the matching process such as position or local image information.

Edges represent some kind of relationship between nodes such as proximity or spatial adjacency (in the case of region adjacency graphs). They may include some attribute of this relationship such as distance, orientation or relative position. Alternatively, edges they may be unattributed ($\{0, 1\}$-valued).

Graphs used in this thesis have been extracted by one of the following approaches:

- Nodes representing feature-points and edges following either a Delaunay triangulation on the point-set or a K-nearest-neighbour approach (it can be considered that the Delaunay triangulation approach leads to a region adjacency graph over the Voronoi tessellations seeded at the feature-points).
- In the case of black and white shape images, graphs are derived from the medial axis representation of the shape.

As nodes' attributes we have used either position coordinates or feature descriptor vectors. We have used unattributed edges in all the cases.

All these approaches lead to sparse graphs in which binary relations are limited to the neighborhood of each node. This sparsity can be exploited in order to produce efficient implementations.

The terms match, correspondence and assignment are used interchangeably throughout this thesis.

Notation

	Indices are in lowercase
a, b	Latin letters are used to index the origin object-set
α, β	Greek letters are used to index the destination object-set
	Vectors are in lowercase bold. Vectors are column by default
$\mathbf{x} = (x^{V}, x^{H})$	Column-vector with the vertical and horizontal coordinates of a planar point.
$\mathbf{y} = (y^{V}, y^{H})$	Idem.
$\mathbf{x}_a, \mathbf{y}_\alpha$	Column vectors of the a-th and α-th points from the origin and destination sets, respectively.
$\mathbf{h}, \mathbf{k}$	Descriptor vectors (column).
$\mathbf{h}_a(k), \mathbf{k}_\alpha(k)$	k-th elements from the descriptor vectors of the a-th and α-th objects from the origin and destination sets, respectively.
	Some special functions are in capital caligraphic letters
$\mathcal{F}$	Objective function (optimization).
$\mathcal{T}(\cdot, \Phi)$	Spatial transformation according to parameters Φ.
	More vectors (in bold)
$\mathbf{t} = (t^{V}, t^{H})$	Translation vector with vertical and horizontal coordinates.
$\tilde{\mathbf{x}} = \left(x^1, x^2, x^3\right)$	Point in homogeneous coordinates.
$\mathbf{x}' = \mathcal{T}(\mathbf{x}, \Phi)$	Transformed point according to transformation $\mathcal{T}$ with parameters Φ.
$\mathbf{w}_a$	Virtual observation (planar point)
	Some scalars (in greek letters)
τ	Ratio value.
λ_i	i-th eigenvalue (in decreasing order).
η^{V}, η^{H}	Vertical and horizontal scaling parameters.
	Spatial transformation matrices are in typewriter

$\mathtt{R}$	2×2 orthogonal rotation matrix.
$\mathtt{A}$	2×2 non-singular matrix of affine transformation parameters.
$\mathtt{H}$	3×3 homography matrix.
	Sets are in capital caligraphic letters
$\mathcal{I} = 1 \ldots n$	Origin index-set.
$\mathcal{J} = 1 \ldots m$	Destination index-set.
$\mathcal{X} = \{\mathbf{x}_a \vert a \in \mathcal{I}\}$	Origin point-set.
$\mathcal{Y} = \{\mathbf{y}_\alpha \vert \alpha \in \mathcal{J}\}$	Destination point-set.
$\mathcal{H} = \{\mathbf{h}_a \vert a \in \mathcal{I}\}$	Origin descriptor vector-set.
$\mathcal{K} = \{\mathbf{k}_\alpha \vert \alpha \in \mathcal{J}\}$	Destination descriptor vector-set.
$\mathcal{U} = \{\mathbf{u}_a \vert a \in \mathcal{I}\}$	Origin node-set.
$\mathcal{V} = \{\mathbf{v}_\alpha \vert \alpha \in \mathcal{J}\}$	Destination node-set.
	Functions are in italic
I	Image
$I\left(x^{V}, x^{H}\right)$	Value of the image function at position (x^{V}, x^{H}).
$f : \mathcal{I} \to \{\mathcal{J} \cup \emptyset\}$	Assignment function.
$\emptyset$	Symbol of the null-assignment.
	Optimization
Θ	Symbol used to denote some generical parameters to be optimized.
$\Theta^{\star}$	A star indicates the optimal value for the parameters.
$\omega_{a\alpha}^{(n)}$	Missing-data estimates for the match $a \to \alpha$ given the parameters at iteration n (EM algorithm).
	Matrices are in capital italic and elements are indexed with subscripts
$C_{a\alpha}$	(a, α)-th element of the cost matrix C.
B	Benefit or similarity matrix.
W	Weights matrix.
D, M	Adjacency matrices of the data (origin) and model (destination) graphs, respectively.
S	Assignments (or correspondences) matrix.
	Graphs are in capital bold letters
$\mathbf{G} = (\mathcal{U}, D)$	Data-graph.
$\mathbf{H} = (\mathcal{V}, M)$	Model-graph.

Part I

Background

Chapter 1

Image Matching using Local Features

1.1 Introduction

During the last decade there has been an increasing interest in feature-based approaches to image matching. These are approaches aimed at solving the correspondence problem that rely on a set of discriminant features built upon local image evidence around some interest points. They have proven to be successful in a wide variety of applications such as object recognition (Lowe, 2004), robot localization (Frank-Bolton *et al.*, 2008; Ila *et al.*, 2010), texture recognition (Lazebnik *et al.*, 2005), object registration (Belongie *et al.*, 2002), clustering (Xia & Hancock, 2009) and building panoramas (Brown & Lowe, 2003).

Since correspondences are decided solely on the basis of local information, it is usual to apply a further refinement process aimed at removing spurious matches by enforcing some global consistency. In the next chapters we will review approaches to enforce global consistency by imposing geometric or structural constraints. In the present chapter we will focus on solving the correspondence problem using local discriminant features.

There are three different steps involved in this process, namely, interest point detection, feature description and matching.

First, a set of salient points are detected and appropriate neighborhood regions are determined. Saliency is determined by their stability under certain changes in the imaging conditions (e.g. illumination or viewpoint). Invariance to viewpoint changes is attained by determining the orientation, scale and/or shape of the neighboring regions.

Next, descriptions of the detected regions are encapsulated in the form of feature-vectors. At this step, the orientation, scale and/or shape information is used to normalize the regions and thereby build invariant descriptors. Robustness to affine illumination changes may be introduced at this step by normalizing the local intensity values.

Finally, a matching strategy is utilized in order to establish the correspondences between the feature-descriptors extracted from two images. This is usually done by finding the correspondences that minimize the sum of distances between the matched descriptors.

1.2 Point and Region Detection

Interest points in an image can be characterized as follows:

- have a mathematically well-founded definition
- have a well-defined position in image space
- their local image information is rich
- can be stably located under global perturbations such as changes in illumination or viewpoint.

Figure 1.1 illustrates this concept.

In the following we introduce the main point and region detectors used throughout this thesis.

We use the notation $\mathbf{x} = (x^V, x^H)$ to represent a point by its vertical and horizontal coordinates. The variable I denotes an image function, whereas $I(x^V, x^H)$ denotes the image's intensity value at point $\mathbf{x}$.

1.2.1 Harris Corners

Harris corner detector (Harris & Stephens, 1988) is a popular point detector because of its ability to stably detect points across changes in rotation, scale, illumination and image noise. It is based on the local auto-correlation function of a signal, where the local auto-correlation function measures the local changes of the image within a patch by shifting the patch along the image. The Harris corner detector is a continuous extension to the discrete one presented by Moravec (1981).

Given a point $\mathbf{x} = (x^V, x^H)$ and a shift $\Delta = (\Delta^V, \Delta^H)$, the auto-correlation function is defined as

$$autocorr(\mathbf{x}) = \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left[I\left(y_i^V, y_i^H\right) - I\left(y_i^V + \Delta^V, y_i^H + \Delta^H\right)\right]^2 \tag{1.1}$$

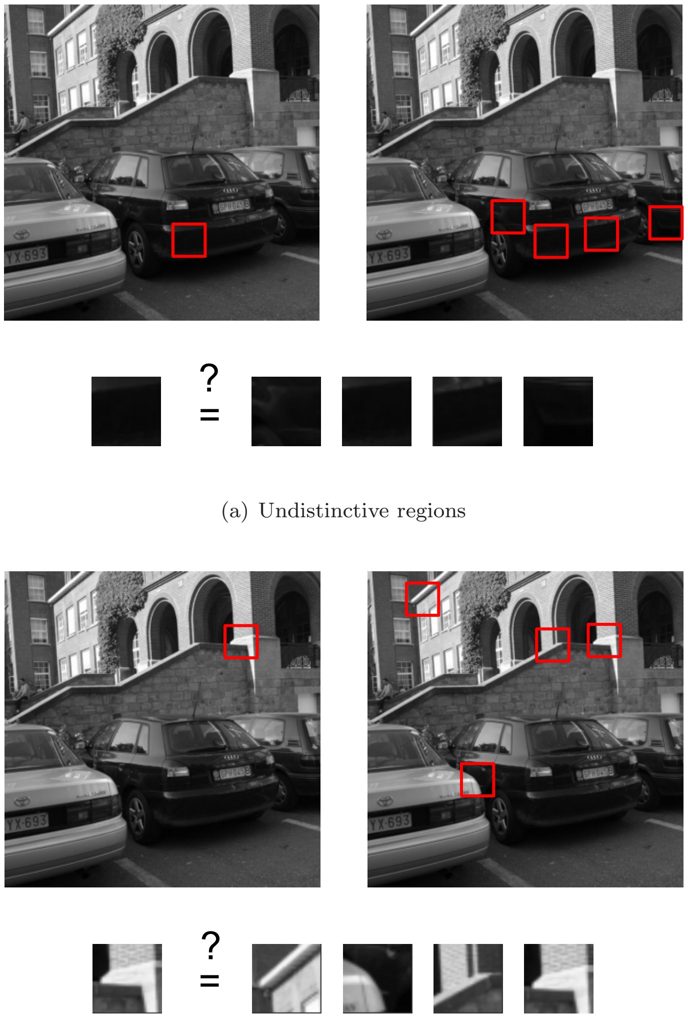
where $\mathcal{W}^{\mathbf{x}}$ is the set of points inside a window centered on $\mathbf{x}$.

One of the differences between the Harris & Stephens (1988) corner detector and the one by Moravec (1981) is that the former uses a Gaussian weighting factor $e^{-\left({x^V}^2 + {x^H}^2\right)/2\sigma^2}$ to define the window while the latter uses a discrete patch.

The shifted image is approximated by a Taylor series expansion truncated to the first order terms,

$$I\left(y_i^V + \Delta^V, y_i^H + \Delta^H\right) \approx I\left(y_i^V, y_i^H\right) + \left[I^V\left(y_i^V, y_i^H\right) I^H\left(y_i^V, y_i^H\right)\right] \begin{bmatrix}\Delta^V \\ \Delta^H\end{bmatrix} \tag{1.2}$$

where I^V and I^H are the partial derivatives of the image function in the vertical and horizontal directions, respectively.



(a) Undistinctive regions

(b) Distinctive regions

Figure 1.1: (a) shows an example of ambiguous point/region detection where the local image information is not distinctive enough in order to establish a correspondence. On the contrary, regions in (b) contain enough information in order to select a distinctive matching candidate.

Substituting approximation (1.2) into (1.1) yields,

$$\begin{aligned} \mathit{autocorr}\left(\mathbf{x}\right) &= \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left[I\left(y_i^V, y_i^H\right) - I\left(y_i^V + \Delta^V, y_i^H + \Delta^H\right)\right]^2 \\ &= \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left(I\left(y_i^V, y_i^H\right) - I\left(y_i^V, y_i^H\right) - \left[I^V\left(y_i^V, y_i^H\right) I^H\left(y_i^V, y_i^H\right)\right] \begin{bmatrix} \Delta^V \\ \Delta^H \end{bmatrix}\right)^2 \end{aligned}$$

$$= \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left(- \left[I^V \left(y_i^V, y_i^H \right) I^H \left(y_i^V, y_i^H \right) \right] \begin{bmatrix} \Delta^V \\ \Delta^H \end{bmatrix} \right)^2$$

$$= \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left(\left[I^V \left(y_i^V, y_i^H \right) I^H \left(y_i^V, y_i^H \right) \right] \begin{bmatrix} \Delta^V \\ \Delta^H \end{bmatrix} \right)^2$$

$$= \left[\Delta^V \Delta^H \right] \begin{bmatrix} \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left[I^V \left(y_i^V, y_i^H \right) \right]^2 & \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} I^V \left(y_i^V, y_i^H \right) I^H \left(y_i^V, y_i^H \right) \\ \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} I^V \left(y_i^V, y_i^H \right) I^H \left(y_i^V, y_i^H \right) & \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} \left[I^H \left(y_i^V, y_i^H \right) \right]^2 \end{bmatrix} \begin{bmatrix} \Delta^V \\ \Delta^H \end{bmatrix}$$

$$= \left[\Delta^V \Delta^H \right] \mathtt{A}^{\mathbf{x}} \begin{bmatrix} \Delta^V \\ \Delta^H \end{bmatrix} \quad (1.3)$$

where the 2×2 matrix $\mathtt{A}^{\mathbf{x}}$ captures the intensity structure of the local neighbourhood centered at $\mathbf{x} = (x^V, x^H)$.

In order to give an idea of the information provided by matrix $\mathtt{A}^{\mathbf{x}}$, figure 1.2 presents three example images along with the scatter plots of the values of the partial derivatives I^V, I^H for each window $\mathcal{W}^{\mathbf{x}}$ in the images.

The eigenvalues λ_1, λ_2 of matrix $\mathtt{A}^{\mathbf{x}}$ form a rotationally invariant descriptor describing the direction of the intensity changes of the windowed region. There are three cases to be considered:

1. If both λ_1, λ_2 are small (i.e. little change in $\mathtt{A}^{\mathbf{x}}$ in any direction), the windowed image region is approximately flat.

2. If one eigenvalue is high and the other is low (i.e. changes in $\mathtt{A}^{\mathbf{x}}$ are in one direction), then the windowed region contains an edge.

3. If both eigenvalues are high (i.e. changes in $\mathtt{A}^{\mathbf{x}}$ are in any direction), then the windowed region contains a peak.

Figure 1.3 shows an illustrative scheme of the classification space via eigenvalues λ_1, λ_2.

The Harris corner detector does not provide any scale or shape information aimed at determining an appropriate neighborhood region around each interest point. In the following we introduce an interest point detector that determines both orientation and scale information.

1.2.2 Difference of Gaussians

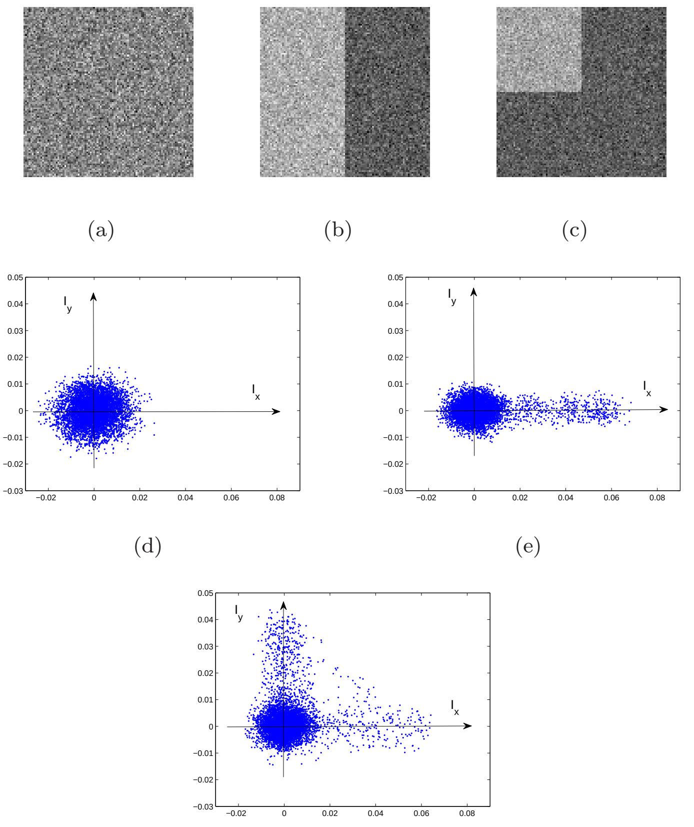
Lowe (2004) developed this detector aimed at selecting interest regions in a scale and rotation invariant way.

The scale space of an image is defined as a function $L(x^V, x^H, \sigma)$ obtained by the convolution of an input image $I(x^V, x^H)$ with a variable-scale Gaussian $G(x^V, x^H, \sigma)$:

$$L(x^V, x^H, \sigma) = G(x^V, x^H, \sigma) * I(x^V, x^H) \quad (1.4)$$

where $*$ is the convolution operation and

$$G(x^V, x^H, \sigma) = \frac{1}{2\pi\sigma^2} e^{\left(x^{V2} + x^{H2}\right)/2\sigma^2} \quad (1.5)$$



(a) (b) (c)

(d) (e)

(f)

Figure 1.2: Figures (a), (b) and (c) are example images showing a flat region, an edge and a corner, respectively. Figures (d), (e) and (f) are scatter plots showing the distributions of the sum of partial derivatives $\left(\sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} I^V(\mathbf{y}_i), \sum_{\mathbf{y}_i \in \mathcal{W}^{\mathbf{x}}} I^H(\mathbf{y}_i)\right)$ for each window $W^{\mathbf{x}}$ across the respective images.

In order to detect stable keypoint locations, the scale-space extrema of the Difference-of-Gaussian function convolved with the image $D(x^V, x^H, \sigma)$ is used. This function can be computed from two nearby scales separated by a constant multiplicative factor k in the following way:

$$\begin{aligned} D(x^V, x^H, \sigma) &= (G(x^V, x^H, k\sigma) - G(x^V, x^H, \sigma)) * I(x^V, x^H) \\ &= L(x^V, x^H, k\sigma) - L(x^V, x^H, \sigma) \end{aligned} \tag{1.6}$$

This function is particularly appropriate since it provides a close approximation to the scale-normalized Laplacian of Gaussian (the one producing the most stable image features compared to other ones, Mikolajczyk & Schmid (2002)) and it can be efficiently computed by simply subtracting the smoothed images L (see figure 1.4).

Difference-of-Gaussian images are computed for all scales as shown in figure

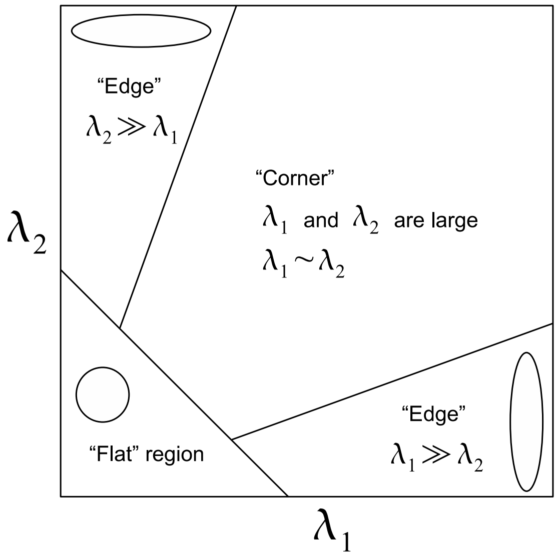

"Edge"
$\lambda_2 \gg \lambda_1$
"Corner"
λ_1 and λ_2 are large
$\lambda_1 \sim \lambda_2$
λ_2
"Edge"
$\lambda_1 \gg \lambda_2$
"Flat" region
λ_1

Figure 1.3: Different classification areas for a point (x^V, x^H) as function of the eigenvalues λ_1, λ_2 of matrix $\mathtt{A}^\mathtt{x}$.

1.5.

Local extrema are detected by selecting the points in scale space that are higher or lower than all its eight neighbors in the current scale and nine neighbors in the scale above and below (see figure 1.6).

Accurate keypoint localization is performed from the candidate locations by fitting a 3D quadratic function to the local sample points. This allows to reject low contrast points as well as those that are poorly localized along an edge since they are considered unstable.

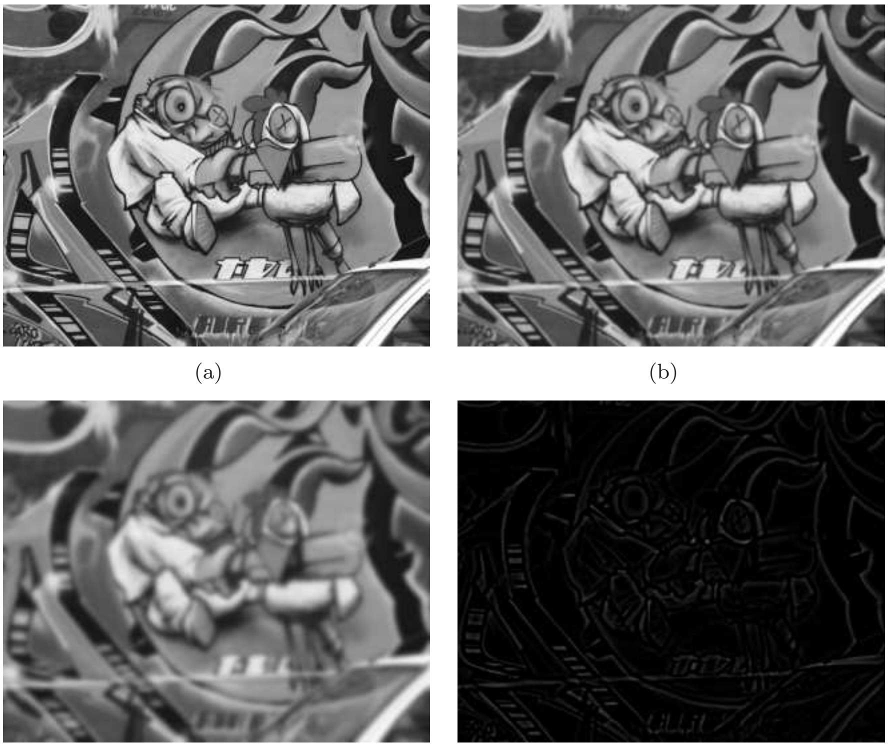
An orientation is assigned to each keypoint so that descriptors can be represented relative to this orientation and therefore achieve invariance to image rotation.

The following procedure is used. The scale of the keypoint is used to select the Gaussian smoothed image, L, so that all computations are preformed in a scale-invariant way.

An orientation histogram is built from the gradient orientations of sample points within a region around the keypoint. Gradient orientation $\theta(x^V, x^H)$ of a point from the smoothed image $L(x^V, x^H)$ at the selected scale is computed using pixel differences in the following way

$$\theta(x^V, x^H) = \tan^{-1}\{[L(x^V, x^H+1) - L(x^V, x^H-1)] / [L(x^V+1, x^H) - L(x^V-1, x^H)]\} \tag{1.7}$$

The orientation histogram has 36 bins covering the 360 degree range of orientations. Each sample added to the histogram is weighted by its gradient


(a)

(b)

(c)

(d)

Figure 1.4: (a) Original image. (b) Gaussian with scale parameter σ. (c) Gaussian with a higher scale parameter $k\sigma$. (d) Difference-of-Gaussians (i.e. subtraction of two Gaussians)

magnitude and by a Gaussian term depending on its distance to the center (i.e. the keypoint). The gradient magnitude $w\left(x^{V}, x^{H}\right)$ is computed in the following way

$$w\left(x^{V}, x^{H}\right) = \sqrt{\left[L\left(x^{V}, x^{H}+1\right) - L\left(x^{V}, x^{H}-1\right)\right]^2 + \left[L\left(x^{V}+1, x^{H}\right) - L\left(x^{V}-1, x^{H}\right)\right]^2} \tag{1.8}$$

Peaks in the orientation histogram correspond to dominant orientations in the local gradients. The keypoint orientation is assigned to the one corresponding to the highest peak in the orientation histogram. Any other peak within 80% of the highest one is used to create another keypoint with the same location and a different orientation. Therefore, keypoints with multiple peaks of similar magnitude will generate multiple keypoints with same location and different orientations. This has demonstrated to contribute to a better stability in the matching.

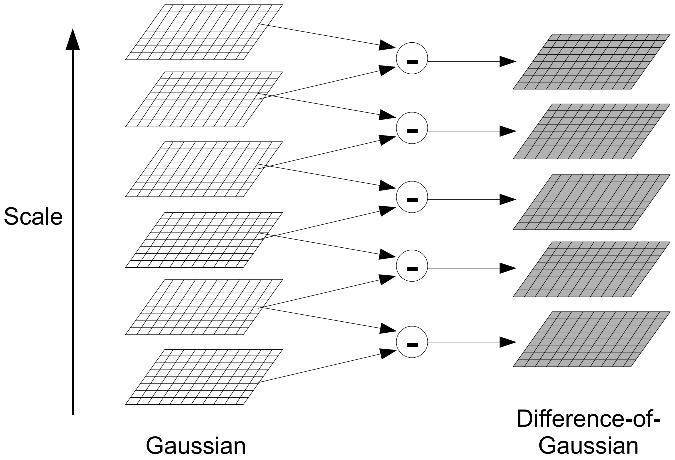


Figure 1.5: The initial image is successively convolved with Gaussians (i.e. blurred) to produce the set of scale space images shown on the left. Adjacent images are subtracted to produce the Difference-of-Gaussian images on the right.

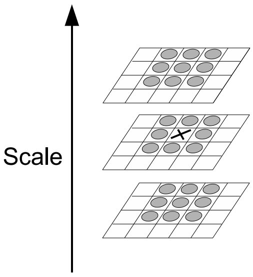


Figure 1.6: A point is selected as candidate only if it is larger or smaller than all its neighbours in the same and adjacent scales.

1.2.3 Other Detectors

Although not directly used in this thesis, other detectors worth mentioning are the following.

Harris-Affine. Mikolajczyk & Schmid (2004) developed an affine invariant detector. In previous approaches, scale invariance was attained by locating points at the extrema of the three-dimensional scale-space of an image. Then, the radius of the appropriate circular neighborhood region was determined by the scale at which the point was detected. However, in the case of large viewpoint changes, circular neighborhoods may no longer fit corresponding regions between two images. Mikolajczyk & Schmid (2004) addressed this problem by selecting the appropriate *affine shape* (elliptical region) of the scale-space extrema.This was attained with an iterative algorithm based on the second-moment matrix. The characteristic scale and the affine shape of a point determined the affine-invariant region.

Maximally Stable Extremal Regions. Matas *et al.* (2002) presented an approach to detect connected regions of pixels that were either brighter or darker than all the pixels in its outer boundary. These connected components have some desirable properties such as their detection is stable against monotonic illumination changes and that geometric transformations (affine, homography or non-rigid) preserve connectedness of the regions.

As seen, point detectors usually provide circular or elliptic regions with a given orientation and scale. Before the description stage, the scale or shape information is used to map all the regions to the same fixed-radius circular form. The estimated orientation is used in order to normalize for rotation as well. This way, descriptors are scale, rotation and (in some cases) affine invariant.

1.3 Feature Descriptors

Perhaps the simplest way of describing a region is to use a raw vector of pixel intensities. Despite its simplicity, this leads to high-dimensional representations that may render inefficient in practical situations. Distribution-based descriptors using histograms have become popular due to its high efficiency and descriptive facilities. In the following we introduce the descriptors more relevant to our work.

1.3.1 Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT) descriptors (Lowe, 2004) are based upon a model of biological vision by Edelman *et al.* (1997). They observed that neurons in the visual cortex respond to gradients at a particular orientation and rough location instead of absolute intensity values and precise locations.

Lowe (2004) build a 3D histogram of 8 × 8 location bins and 8 gradient orientations bins for each region. The regions are previously normalized for scale and rotation using the parameters estimated during the detection phase. For each location bin, each one of its 8 corresponding orientation planes are assigned the sum of occurrences responding at that particular gradient orientation in the image patch falling inside that location bin. Each occurrence is weighted by its gradient magnitude and by a Gaussian weighting factor giving less emphasis to gradients that are far from the center of the descriptor. This is illustrated in figure 1.7.

1.3.2 Shape Contexts

Belongie *et al.* (2002) present *Shape Contexts*, a descriptor that accounts for the spatial distribution and frequency of occurrences of the remaining points around a keypoint.

The original descriptor consists of a 2D log-polar histogram where one dimension encodes the distance from the center in a logarithmic scale and the other dimension encodes the angle. Using a logarithmic scale, a finer resolution

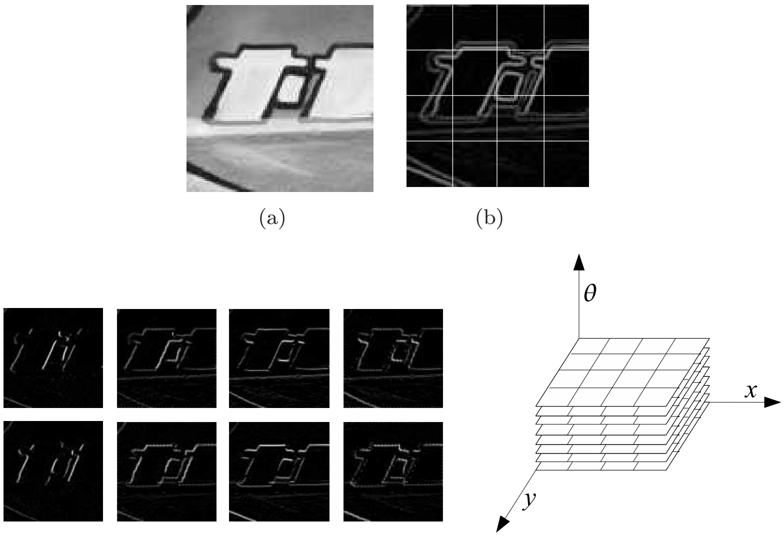

(a) (b) (c) (d)

Figure 1.7: (a) Image patch corresponding to an interest region. (b) Gradient image with the 4 × 4 location grid superimposed. (c) Gradient images corresponding to the 8 orientation planes. (d) 3D histogram of location (2D) + orientation (1D).

is attained at nearby points, thus providing robustness to global deformations as far as they preserve the local shape topology. Figure 1.8 illustrates this idea.

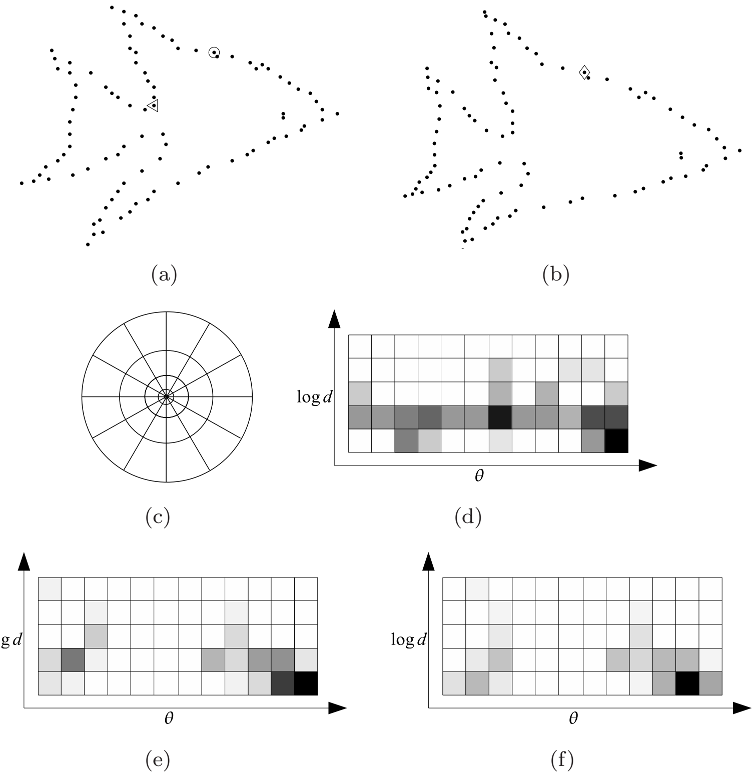
Mikolajczyk & Schmid (2005) used an adapted version of this descriptor aimed at the matching of natural images in the context of a performance evaluation of local descriptors. They used a log-polar histogram for describing the edge distribution inside an interest region, as opposed to the global description used in the original approach. Edges were extracted with the Canny edge detector (Canny, 1986). Additionally, they added one more dimension to the log-polar histogram to account for the gradient orientations.

1.3.3 Other Descriptors

Although not used in this thesis, another local image descriptor worth mentioning is the one presented by (Lazebnik *et al.*, 2005) based on the *spin images* (Johnson & Hebert, 1999). This is a two-dimensional histogram encoding the distribution of brightness values in an affine normalized patch. The two dimensions correspond to the distance from the center of the patch and the intensity value. Since these measures are invariant to rotations, spin images offer the right degree of invariance for representing scale or affine-normalized regions.

An exhaustive evaluation of different types of descriptors by Mikolajczyk & Schmid (2005) in terms of matching accuracy concluded that SIFT descriptors perform the best. A few approaches using these descriptors are the following.

Ila *et al.* (2010) use SIFT for *Simultaneous Localization And Mapping* (SLAM)



(a)

(b)

(c)

(d)

(e)

(f)

Figure 1.8: (a) and (b) show two example point-sets of a fish template related by a nonrigid deformation. (c) illustrates the 5×12 log-polar grid where each cell corresponds to a bin of the descriptor's histogram. (d), (f) and (e) show the Shape Contexts of the points marked with $\triangleleft, \circ, \diamond$ in the template shapes. Horizontal and vertical axis correspond to 12 angle and 5 log-distance bins, respectively. Although the deformation in the templates, Shape Contexts allow to distinguish the corresponding points from the non-corresponding one, thus providing an effective framework for feature-matching.

of a mobile robot. Frank-Bolton *et al.* (2008) use SIFT for localization of a mobile robot from multiple views. Lowe (2001) present an approach for learning robust descriptors from a set of training images. These local descriptor models, which are based on SIFT, are aimed to be robust to a wide range of transformations. Brown & Lowe (2003) present an approach for building panoramas from sets of images using SIFT descriptors. Forssén & Lowe (2007) build SIFT descriptors from *Maximally Stable Extremal Regions* (MSER) (Matas *et al.*, 2002). These adapted SIFT descriptors introduce affine-invariance by incorporating the shape information from the MSER detector.

1.4 Matching Criteria

At this point, correspondences are decided solely on the basis of the distance between descriptor-vectors, disregarding any coordinate information about the interest points.

While there exist different approaches to determine the distance measure between two descriptors, selection of the matches is usually done so as to minimize the sum of distances between the matched features. This is an instance of the linear assignment problem which can be solved within polynomial time with low exponent.

Consider two sets of interest points $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$, where $\mathcal{I} = 1 \ldots |\mathcal{X}|$ and $\mathcal{J} = 1 \ldots |\mathcal{Y}|$ are the index-sets, extracted from two images.

Consider also the descriptor-vector sets $\mathcal{H} = \{\mathbf{h}_a\}$ and $\mathcal{K} = \{\mathbf{k}_\alpha\}$ such that $\mathbf{h}_a(i), \mathbf{k}_\alpha(i)$ denote the i-th elements of the descriptors-vectors from the regions associated with the points $\mathbf{x}_a$ and $\mathbf{y}_\alpha$. Without loss of generality, we use the same notation regardless the descriptors consist of multi-dimensional histograms or raw vectors of pixel intensities (note that, in any case, all the descriptor-vectors have the same length).

We denote the cost matrix as C where the (a, α)-th element $C_{a\alpha}$ contains the cost of matching a-th feature in the first image to α-th feature in the second image. The matching heuristic tries to estimate the matching function $f : \mathcal{I} \to \{\mathcal{J} \cup \emptyset\}$ that optimize some objective function $\mathcal{F}$ defined over the elements of matrix C.

Many features from the first image may have no correct match in the other image because they arise from background clutter or they were not detected in the second image. The symbol $\emptyset$ represents the index of the *null-assignments*, so that any feature a matched to $\emptyset$ is considered an *outlier* (i.e., it has no corresponding feature in the other set).

In the following we present some commonly adopted combinations of distance measures and match heuristics, however, many other combinations are possible.

1.4.1 Cross-Correlation & Hungarian

This is a commonly adopted approach when descriptors consist of raw vectors of pixel intensities.

Cross-Correlation estimates the degree to which two image patches $\mathbf{h}_a$ and $\mathbf{k}_\alpha$ are correlated.

This is a benefit (or similarity) measure rather than a cost one. The benefit matrix is composed by the benefits of the match from the a-th feature in the first image I_1 to the α-th feature in the second image I_2. Hence, it is defined as

$$B_{a\alpha} = \frac{\sum_i \left(\mathbf{h}_a\left(k\right) - \bar{I}_1\right)\left(\mathbf{k}_\alpha\left(i\right) - \bar{I}_2\right)}{\sqrt{\sum_i \left(\mathbf{h}_a\left(i\right) - \bar{I}_1\right)^2}\sqrt{\sum_i \left(\mathbf{k}_\alpha\left(i\right) - \bar{I}_2\right)^2}} \tag{1.9}$$

where $\bar{I}_1$ and $\bar{I}_2$ are the mean values of the images.

Such a benefit matrix B can be easily turned into a cost matrix C in the following way

$$C = \gamma \mathbf{1}\mathbf{1}^\top - B \tag{1.10}$$

where $\gamma = \max\left(B_{a\alpha} \in B\right)$ is the maximum benefit value in matrix B and $\mathbf{1}$ is a column vector of ones of appropriate size.

The final correspondences are decided by solving a linear assignment problem. This is, find the matching function f that minimizes the following expression.

$$f^\star = \arg\min_f \sum_{a \in \mathcal{I} \cup \emptyset} C_{af(a)} \tag{1.11}$$

subject to $\nexists a, b$ s.t. $f\left(a\right) = f\left(b\right) \neq \emptyset$. This problem can be efficiently solved with the Hungarian method (Munkres, 1957).

An extra column of costs $\forall a, C_{a\emptyset} = \epsilon$ representing the assignments to $\emptyset$ may be introduced in order to manage outliers so that no matches $f\left(a\right) = \alpha$ with a cost $C_{a\alpha} \geq \epsilon$ are selected by the Hungarian method.

1.4.2 Euclidean Distance & Ratio Test

Lowe (2004) proposed to use the Euclidean distance between the SIFT feature descriptors as a cost measure. This is,

$$C_{a\alpha} = \sqrt{\sum_i \left[\mathbf{h}_a\left(i\right) - \mathbf{k}_\alpha\left(i\right)\right]^2} \tag{1.12}$$

With regards to the outlier management, they noticed that in the case of SIFT descriptors a global threshold on distance did not perform well since some descriptors were more discriminative than others. A more effective constraint is obtained by comparing the distance of the closest neighbor to that of the second-closest neighbor. Accordingly, feature a is matched to its closest candidate α (according to the Euclidean distance), if and only if the ratio of the distance $C_{a\alpha}$ to that of the second closest candidate $C_{a\beta}$ is not higher than a certain threshold. This is,

$$f\left(a\right) = \begin{cases} \alpha & \text{if } \; C_{a\alpha}/C_{a\beta} \leq \tau \\ \emptyset & \text{otherwise} \end{cases} \tag{1.13}$$

where $\tau \in \left[0 \ldots 1\right]$ is a predefined threshold and β is the index corresponding to the second closest feature to a in Euclidean distance.

1.4.3 Chi-Square Test & Hungarian

Belongie *et al.* (2002) used the χ^2 test statistic as a natural way to denote the dissimilarity between two shape-context histograms. This is,

$$C_{a\alpha} = \frac{1}{2} \sum_{i} \frac{\left[\mathbf{h}_a\left(i\right) - \mathbf{k}_\alpha\left(i\right)\right]^2}{\mathbf{h}_a\left(i\right) + \mathbf{k}_\alpha\left(i\right)} \tag{1.14}$$

They used the Hungarian method to seek for the set of assignments f imposing a maximum cost threshold ϵ in order for a feature to be considered an outlier.

Chapter 2

Geometric Transformation Models for Point-Set Alignment

2.1 Introduction

Alignment of point-sets is frequently used in *pattern recognition* when objects are represented by a set of coordinate-data. The idea behind is to be able to compare two point-sets regardless the effects of a given family of spatial transformations. This is at the core of many object recognition applications e.g., medical image analysis, shape retrieval, learning shape models (Cootes *et al.*, 1995; Dryden & Mardia, 1998) or reconstructing a scene from various views (Hartley & Zisserman, 2000).

Point-set alignment requires that the correspondences between both point-sets are known a priori. The techniques related to feature descriptors described in the previous chapter may be used in order to establish the correspondences. The techniques described in the present chapter allow to align two point-sets in a way tolerant to a certain amount of noise in their position coordinates. In the more realistic case of further sources of noise as, for example, erroneous correspondences, the use of robust techniques such as RANSAC (section 2.4) or iterative point-set registration methods (chapter 3) may be applied. In either way, the methodology described in this chapter is used as an intermediate step of these robust estimation techniques.

Consider two corresponding sets of points $\mathcal{X} = \{\mathbf{x}_i, i = 1 \ldots n\}$ and $\mathcal{Y} = \{\mathbf{y}_i, i = 1 \ldots n\}$ where $\mathbf{x}_i = (x_i^V, x_i^H)$ and $\mathbf{y}_i = (y_i^V, y_i^H)$ are column-vectors containing the 2D (vertical and horizontal) position coordinates of each point. It is known a priori that points $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^2$ are in correspondence.

The alignment problem may be posed as the one of finding the optimal transformation parameters $\Phi^\star$ that minimize the following sum of squared norms

$$\Phi^\star = \arg\min_{\Phi} \sum_{i=1}^{n} \left\| \mathbf{x}_i - \mathcal{T}\left(\mathbf{y}_i; \Phi\right) \right\|^2 \qquad (2.1)$$

where $\mathcal{T}(\mathbf{y};\Phi)$ is the result of transforming point $\mathbf{y}$ according to a transformation model with parameters Φ and, $\|\cdot\|$ is the Euclidean norm.

Following Hartley & Zisserman (2000), we present a hierarchy of geometric transformation models starting from the most specialized, the isometries, until the projective transformations are reached.

2.2 A Hierarchy of Transformations

2.2.1 Isometries

Isometries are transformations that preserve Euclidean distance (*iso* = same, *metric* = measure). An isometry is represented as

$$\begin{pmatrix} y^{V\prime} \\ y^{H\prime} \end{pmatrix} = \begin{bmatrix} \gamma\cos\theta & -\sin\theta \\ \gamma\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} y^{V} \\ y^{H} \end{pmatrix} + \begin{pmatrix} t^{V} \\ t^{H} \end{pmatrix} \tag{2.2}$$

where $\gamma = \pm 1$. If $\gamma = 1$ then the isometry is *orientation-preserving* and is a *Euclidean* transformation (a composition of a translation and a rotation). If $\gamma = -1$ then the isometry reverses orientation (i.e., includes a reflection).

Euclidean transformations model the motion of a rigid object. A planar Euclidean transformation can then be written as

$$\mathbf{y}' = \mathtt{R}\mathbf{y} + \mathbf{t} \tag{2.3}$$

where $\mathtt{R}$ is a 2×2 rotation matrix (an orthogonal matrix such that $\mathtt{R}^\top\mathtt{R} = \mathtt{R}\mathtt{R}^\top = \mathtt{I}$, where $\mathtt{I}$ is the identity matrix), and $\mathbf{t}$ is a translation 2-vector. Special cases are a pure rotation (when $\mathbf{t} = (0,0)$) and a pure translation (when $\mathtt{R} = \mathtt{I}$).

A planar Euclidean transformation has three degrees of freedom, one for the rotation and two for the translation.

The transformation can be computed from two point correspondences.

2.2.2 Similarity Transformations

A similarity transformation is an isometry composed with an isotropic scaling. In the case of a Euclidean transformation composed with a scaling (i.e. no reflection) the similarity has the representation

$$\begin{pmatrix} y^{V\prime} \\ y^{H\prime} \end{pmatrix} = \begin{bmatrix} \eta\cos\theta & -\eta\sin\theta \\ \eta\sin\theta & \eta\cos\theta \end{bmatrix} \begin{pmatrix} y^{V} \\ y^{H} \end{pmatrix} + \begin{pmatrix} t^{V} \\ t^{H} \end{pmatrix} \tag{2.4}$$

This can also be written as

$$\mathbf{y}' = \eta\mathtt{R}\mathbf{y} + \mathbf{t} \tag{2.5}$$

where scalar η represents the isotropic scaling. A similarity transformation preserves the ”shape” (form). A planar similarity transformation has four degrees of freedom, the scaling accounting for one more degree of those in an Euclidean transformation.

A similarity can be computed from two point correspondences.

2.2.3 Affine Transformations

An affine transformation (or an *affinity*) is a non-singular linear transformation followed by a translation. It has the matrix representation

$$\begin{pmatrix} y^{V\prime} \\ y^{H\prime} \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} y^{V} \\ y^{H} \end{pmatrix} + \begin{pmatrix} t^{V} \\ t^{H} \end{pmatrix} \tag{2.6}$$

This can also be written as

$$\mathbf{y}' = \mathtt{A}\mathbf{y} + \mathbf{t} \tag{2.7}$$

with $\mathtt{A}$ a 2×2 non-singular matrix. A planar affine transformation has six degrees of freedom corresponding to the four matrix elements plus the two translation elements.

The transformation can be computed from three point correspondences.

To understand the effects of the linear component $\mathtt{A}$ it is useful to consider it as the composition of two transformations, namely rotations and non-isotropic scalings. The affine matrix $\mathtt{A}$ can always be decomposed as

$$\mathtt{A} = \mathtt{R}(\theta)\,\mathtt{R}(-\phi)\,\mathtt{D}\mathtt{R}(\phi) \tag{2.8}$$

where $\mathtt{R}(\theta)$ and $\mathtt{R}(\phi)$ are rotations by θ and ϕ respectively, and $\mathtt{D}$ is a diagonal matrix:

$$\mathtt{D} = \begin{bmatrix} \eta^{V} & 0 \\ 0 & \eta^{H} \end{bmatrix} \tag{2.9}$$

The affine matrix $\mathtt{A}$ is seen as a concatenation of a rotation (by ϕ); a scaling by η^V and η^H respectively in the (rotated) vertical and horizontal directions; a rotation back (by $-\phi$); and finally another rotation (by θ). The two additional degrees-of-freedom of the affinity over the similarity are due to the angle ϕ specifying the scaling direction, and the ratio of the scaling parameters $\eta^V : \eta^H$. The essence of an affinity is this scaling in orthogonal directions, oriented at a particular angle (see figure 2.1).

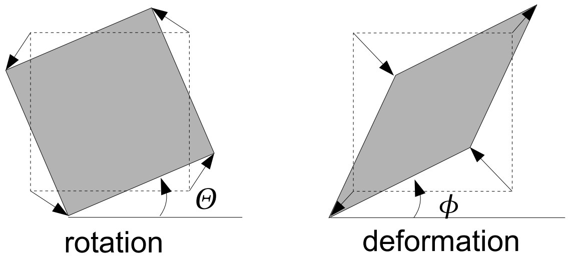


Figure 2.1: Distortions arising from a planar affine transformation. A rotation by $\mathtt{R}(\theta)$ and a deformation $\mathtt{R}(-\phi)\,\mathtt{D}\mathtt{R}(\phi)$. Note that the scaling directions in the deformation are orthogonal.

An affinity is orientation-preserving or -reversing according to whether $\det \mathtt{A}$ is positive or negative respectively. Since $\det \mathtt{A} = \eta^V \eta^H$ the property depends only on the sign of the scalings.

2.2.4 Projective Transformations

A projective transformation (i.e. *projectivity*) is a mapping from points in the *projective plane* $\mathbb{P}^2$ to points in $\mathbb{P}^2$.

The projective plane can be thought of a set of rays through the origin in $\mathbb{R}^3$. Hence, a point in $\mathbb{P}^2$ is the set of all vectors $k\left(y^1, y^2, y^3\right), k = [0 \ldots \infty]$. As k varies, it forms a ray through the origin which can be thought of as representing a single point in $\mathbb{P}^2$. See figure 2.2 for an illustration.

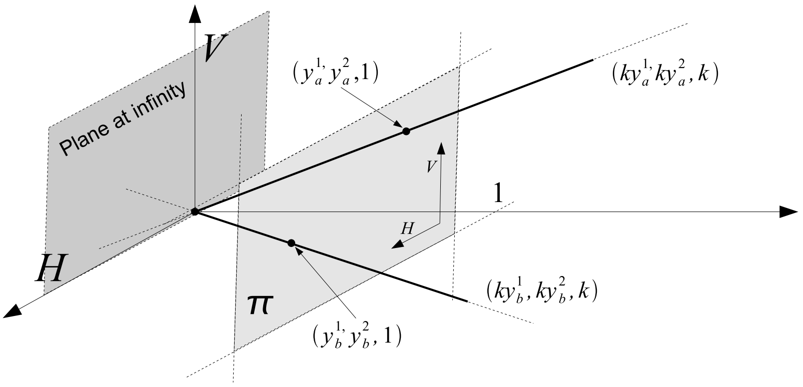


Figure 2.2: Each point in the projective space corresponds to a ray through the origin in $\mathbb{R}^3$. The representative of a ray in homogeneous coordinates is taken as the point of intersection of the ray with plane π. A ray lying in the plane at infinity does not intersect with plane π and hence has no homogeneous representation (it involves a division by zero).

A projectivity is also called a *collineation* (since it maps lines to lines), or also a *homography.*

We use *homogeneous* coordinates in order to represent planar points in the projective plane. We represent a planar point $\mathbf{y} = (y^V, y^H)$ in homogeneous coordinates by adding an extra 1, i.e., $\tilde{\mathbf{y}} = (y^V, y^H, 1)$. In homogeneous form, points are represented by *equivalence classes* of coordinate triples, where two triples are equivalent when they differ by a common multiple. Therefore, points $(y^V, y^H, 1)$ and $(2y^V, 2y^H, 2)$ represent the same point (y^V, y^H). Given a coordinate triple (ky^V, ky^H, k), we can get the original planar coordinates by dividing by k to get (y^V, y^H).

A planar projective transformation is of the form

$$\begin{pmatrix} y^{1'} \\ y^{2'} \\ y^{3'} \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t^V \\ a_{21} & a_{22} & t^H \\ w_1 & w_2 & 1 \end{bmatrix} \begin{pmatrix} y^V \\ y^H \\ 1 \end{pmatrix} \tag{2.10}$$

or more briefly, $\tilde{\mathbf{y}}' = \mathtt{H}\tilde{\mathbf{y}}$, where $\tilde{\mathbf{y}}$ denote the homogeneous representation of point $\mathbf{y}$, and $\mathbf{w} = (w_1, w_2)$ are the elation parameters which are responsible for the non-linear effects of the projectivity.

As we said, a ray through the origin in $\mathbb{R}^3$ (i.e., $k\left(y^1, y^2, y^3\right), \forall k$) corresponds to a point in $\mathbb{P}^2$ and thus, only the ratio of the elements is significant.

As consequence, the $\mathtt{H}$ matrix occurring in equation (2.10) may be changed by multiplication by an arbitrary non-zero scale factor without altering the projective transformation (since $k\tilde{\mathbf{y}}' = kH\tilde{\mathbf{y}}$).

We can standardise matrix $\mathtt{H}$ for scale by multiplying it so that one of its elements (e.g., the last-row last-column element) is set to one. In the cases involving rays lying on the plane at infinity (see figure 2.2) it is not possible to perform this scaling. We will assume that the transformation is carefully built so that no rays lying on the plane at infinity are present.

We define a function $g : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ that maps a point $\left(y^1, y^2, y^3\right)$ to point $\left(y^1/y^3, y^2/y^3\right)$, aimed at obtaining the planar representation of an homogeneous vector. Therefore, the result of a projectivity $\mathtt{H}$ is mapped back to the plane according to the following expression.

$$g\left(\mathtt{H}\tilde{\mathbf{y}}\right) = \left(\frac{a_{11}y^V + a_{12}y^H + t^V}{w_1 y^V + w_2 y^H + 1}, \frac{a_{21}y^V + a_{22}y^H + t^H}{w_1 y^V + w_2 y^H + 1}\right) \tag{2.11}$$

(notice the non-linear effects of the elation parameters w_1, w_2).

A projective transformation has eight degrees of freedom (one for each element of matrix $\mathtt{H}$) and can be computed from four point correspondences.

There are other types of geometric transformations such as non-rigid (e.g. thin-plate splines by Bookstein (1989)), but these are out of the scope of this thesis.

2.3 Recovery of the Transformation Parameters

There is an extensive work done towards the goal of finding the alignment parameters that minimize a similar measure than that of equation (2.1). To cite a few, Dryden & Mardia (1998); Kendall (1984) deal with isometries and similarity transformations; Berge (2006); Umeyama (1991) deal with Euclidean transformations (i.e. excluding reflections from isometries); Haralick *et al.* (1989) deal with similarity and projective transformations; and Hartley & Zisserman (2000) deal exclusively with projective transformations.

In the following we describe how to compute the optimal transformation parameters that minimize equation (2.1) for the type of geometric transformations described above, namely, similarities, affinities and projectivities (isometries and Euclidean transformations are particular cases of similarities). Similarity transformations use techniques from *Procrustes* analysis aimed at dealing with the cases when the matrix component is orthogonal. They are explained in section 2.3.1. Neither affinities or projective transformations impose orthonormality constraints and they are explained in sections 2.3.2 and 2.3.3, respectively.

2.3.1 Similarity Transformations

Consider two sets of corresponding planar points $\mathcal{X} = \{\mathbf{x}_i,\ i = 1, \ldots, n\}$ and $\mathcal{Y} = \{\mathbf{y}_i,\ i = 1, \ldots, n\}$ such that for all i, point $\mathbf{x}_i = (x_i^V, x_i^H)$ is supposed to be in correspondence with point $\mathbf{y}_i = (y_i^V, y_i^H)$.

Procrustes analysis (Dryden & Mardia, 1998) deals with finding the similarity transformation that aligns the two point-sets. This is, locate the optimal rotation matrix $\mathtt{R}^\star$, scaling parameter $\eta^\star$ and translation 2-vector $\mathbf{t}^\star$ that minimize

$$d_P^2\left(\mathcal{X},\mathcal{Y}\right) = \min_{\mathtt{R},\eta,\mathbf{t}} \sum_i^n \left\|\mathbf{x}_i - \left(\eta \mathtt{R}\mathbf{y}_i + \mathbf{t}\right)\right\|^2 \tag{2.12}$$

subject to $\det \mathtt{R} = \pm 1$, where $\left\|\cdot\right\|^2$ is the squared Euclidean norm.

The above quantity $d_P^2\left(\mathcal{X},\mathcal{Y}\right)$ corresponds to the squared *Procrustes distance* between point-sets $\mathcal{X}$ and $\mathcal{Y}$ (Dryden & Mardia, 1998).

In the following, we detail the solution to this problem.

Consider the following quantities

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_i^n \mathbf{x}_i \tag{2.13}$$

$$\bar{\mathbf{y}} = \frac{1}{n}\sum_i^n \mathbf{y}_i \tag{2.14}$$

the mean vectors of $\mathcal{X}$ and $\mathcal{Y}$,

$$\sigma_x^2 = \frac{1}{n}\sum_i^n \left\|\mathbf{x}_i - \bar{\mathbf{x}}\right\|^2 \tag{2.15}$$

$$\sigma_y^2 = \frac{1}{n}\sum_i^n \left\|\mathbf{y}_i - \bar{\mathbf{y}}\right\|^2 \tag{2.16}$$

the variances around the mean vectors of $\mathcal{X}$ and $\mathcal{Y}$,

$$\Sigma_{xy} = \frac{1}{n}\sum_i^n \left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\left(\mathbf{y}_i - \bar{\mathbf{y}}\right)^\top \tag{2.17}$$

a covariance matrix of $\mathcal{X}$ and $\mathcal{Y}$ whose *singular value decomposition* is

$$\Sigma_{xy} = \mathtt{U}\Lambda\mathtt{V}^\top \tag{2.18}$$

where $\mathtt{U}$ and $\mathtt{V}$ are square orthonormal matrices and $\Lambda = \left[\begin{smallmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{smallmatrix}\right]$ is a diagonal matrix of eigenvalues.

The optimal transformation parameters are determined uniquely as follows (Umeyama, 1991)

$$\mathtt{R}^\star = \mathtt{U}\mathtt{V}^\top \tag{2.19}$$

$$\eta^\star = \frac{1}{\sigma_y^2} Tr\left(\Lambda\right) \tag{2.20}$$

$$\mathbf{t}^\star = \bar{\mathbf{x}} - \eta^\star \mathtt{R}^\star \bar{\mathbf{y}} \tag{2.21}$$

where *Tr* denotes the trace of a matrix.

In the case of isometries and Euclidean transformations (i.e., no scalings are allowed), we set the scaling parameter η to 1.

The determinant of $\mathtt{R}$ is either 1 or -1, the former case representing a rotation without reflection and the latter a rotation with reflection. In most applications, this distinction may play no role at all. However, in certain cases reflections are not permitted (i.e. Euclidean transformations). In these cases, orthogonality of $\mathtt{R}$ is not enough, and the additional constraint that $\det \mathtt{R} = 1$ must be imposed.

Umeyama (1991) provides a solution for this case. For the cases when $\det \mathtt{R} = -1$ the optimal rotation without reflection is given by

$$\mathtt{R}^\star = \mathtt{UEV}^\top \tag{2.22}$$

where $\mathtt{E} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$.

2.3.2 Affine Transformations

Let $\mathcal{X} = \{\mathbf{x}_i,\ i = 1, \ldots, n\}$ and $\mathcal{Y} = \{\mathbf{y}_i,\ i = 1, \ldots, n\}$ be corresponding point-sets in $\mathbb{R}^2$. The optimal affine transformation parameters $\mathtt{A}^\star$ and $\mathbf{t}^\star$ are those satisfying

$$\min_{\mathtt{A},\mathbf{t}} \sum_i^n \|\mathbf{x}_i - (\mathtt{A}\mathbf{y}_i + \mathbf{t})\|^2 \tag{2.23}$$

where $\mathtt{A} = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$ is a non-singular matrix, $\mathbf{t}$ is a translation 2-vector and $\mathbf{x}_i = (x_i^V, x_i^H), \mathbf{y}_i = (y_i^V, y_i^H)$ are corresponding points in $\mathbb{R}^2$.

This unconstrained least-squares problem consists on minimizing the following sum of squared residuals

$$\mathcal{F}_a = \sum_i^n (r_i^V)^2 + (r_i^H)^2 \tag{2.24}$$

where

$$\begin{aligned} r_i^V &= x_i^V - (a_{11}y_i^V + a_{12}y_i^H + t^V) \\ r_i^H &= x_i^H - (a_{21}y_i^V + a_{22}y_i^H + t^H) \end{aligned} \tag{2.25}$$

We take derivatives of the above expression with respect to the parameters. We write

$$\begin{aligned} \frac{\delta \mathcal{F}_a}{\delta a_{11}} &= \sum_i^n -2r_i^V y_i^V\,(x_i^V - a_{11}y_i^V - a_{12}y_i^H - t^V) \\ \frac{\delta \mathcal{F}_a}{\delta a_{12}} &= \sum_i^n -2r_i^V y_i^H\,(x_i^V - a_{11}y_i^V - a_{12}y_i^H - t^V) \\ \frac{\delta \mathcal{F}_a}{\delta a_{21}} &= \sum_i^n -2r_i^H y_i^V\,(x_i^H - a_{21}y_i^V - a_{22}y_i^H - t^H) \\ \frac{\delta \mathcal{F}_a}{\delta a_{22}} &= \sum_i^n -2r_i^H y_i^H\,(x_i^H - a_{21}y_i^V - a_{22}y_i^H - t^H) \\ \frac{\delta \mathcal{F}_a}{\delta t^V} &= \sum_i^n -2r_i^V\,(x_i^V - a_{11}y_i^V - a_{12}y_i^H - t^V) \\ \frac{\delta \mathcal{F}_a}{\delta t^H} &= \sum_i^n -2r_i^H\,(x_i^H - a_{21}y_i^V - a_{22}y_i^H - t^H) \end{aligned} \tag{2.26}$$

Optimal affine alignment parameters $a_{11}^{\star}, a_{12}^{\star}, a_{21}^{\star}, a_{22}^{\star}, t^{V\star}, t^{H\star}$ are obtained by simultaneously solving the set of linear equations

$$\frac{\delta \mathcal{F}_a}{\delta a_{11}} = 0 \ , \quad \ldots \quad , \ \frac{\delta \mathcal{F}_a}{\delta t^{H}} = 0 \tag{2.27}$$

which yields a matrix equation of the form $M\mathbf{a} = \mathbf{b}$, where M is a 6×6 matrix and $\mathbf{a} = (a_{11}, a_{12}, a_{21}, a_{22}, t^{V}, t^{H})$ and $\mathbf{b}$ are 6-column-vectors. This can be solved by matrix inversion (i.e., $\mathbf{a} = M^{-1}\mathbf{b}$).

2.3.3 Projective Transformations

Let $\mathcal{X} = \{\mathbf{x}_i,\ i = 1, \ldots, n\}$ and $\mathcal{Y} = \{\mathbf{y}_i,\ i = 1, \ldots, n\}$ be corresponding point-sets in $\mathbb{R}^2$. Let $\tilde{\mathcal{Y}} = \{\tilde{\mathbf{y}}_i,\ i = 1, \ldots, n\}$ be the corresponding homogeneous representation of the points in $\mathcal{Y}$. Let $g : \mathbb{R}^3 \to \mathbb{R}^2$ be a function that maps a point in homogeneous form to its corresponding point in the plane. We seek the optimal homography $\mathtt{H}^{\star}$ such that

$$\min_{\mathtt{H}} \sum_{i}^{n} \left\| \mathbf{x}_i - g\left(\mathtt{H}\tilde{\mathbf{y}}_i\right) \right\|^2 \tag{2.28}$$

where

$$\mathtt{H} = \begin{bmatrix} a_{11} & a_{12} & t^{V} \\ a_{21} & a_{22} & t^{H} \\ w_1 & w_2 & 1 \end{bmatrix} \tag{2.29}$$

is an homography matrix and, $\mathbf{x}_i = (x_i^{V}, x_i^{H})$ and $\tilde{\mathbf{y}}_i = (y_i^{V}, y_i^{H}, 1)$ are corresponding points in inhomogeneous and homogeneous coordinates, respectively.

Again, this unconstrained least-squares problem consists on minimizing a sum of squared residuals of the following form

$$\mathcal{F}_h = \sum_{i}^{n} (r_i^{V})^2 + (r_i^{H})^2 \tag{2.30}$$

This time, the residuals have the following expressions

$$\begin{aligned} r_i^{V} &= x_i^{V} - \left(\frac{a_{11} y_i^{V} + a_{12} y_i^{H} + t^{V}}{w_1 y^{V} + w_2 y^{H} + 1} \right) \\ r_i^{H} &= x_i^{H} - \left(\frac{a_{21} y^{V} + a_{22} y^{H} + t^{H}}{w_1 y^{V} + w_2 y^{H} + 1} \right) \end{aligned} \tag{2.31}$$

The system of equations obtained by equating to zero the partial derivatives has no longer solution in closed form due to the non-linearity in the residuals. Instead, a non-linear iterative method such as Gauss-Newton or Levenberg-Marquardt should be applied. Levenberg-Marquardt is a nonlinear optimization technique that offers a compromise between the steepest gradient and inverse Hessian methods. The former is used when close to the optimum while the latter is used far from it.

Other approaches that state this problem in terms of cross-products do have closed-form solutions (Hartley & Zisserman, 2000).

2.4 Outlier Rejection with RANSAC

Up to this point we have assumed that we have been presented with two sets of corresponding points $\mathcal{X} = \{\mathbf{x}_i,\ i = 1, \ldots, n\}$ and $\mathcal{Y} = \{\mathbf{y}_i,\ i = 1, \ldots, n\}$ where the only source of errors is the measurement of the points' position. In many practical situations this is not true since some points may be mismatched, this is, some correspondences may be erroneous. These are *outliers* that can severely disturb the estimated transformation, and therefore should be identified. The goal is to select the set of *inliers* from the presented correspondences so that the geometric transformation can be computed from the set of inliers in an optimal way using the algorithms described in the previous sections. This is *robust estimation* since the estimation is robust (i.e., tolerant) to outliers.

Following Hartley & Zisserman (2000), we start with a simple example that can be easily visualized: fitting a straight line to a set of 2-dimensional points. This can be thought of as estimating a 1-dimensional affine transformation, $x' = ax + b$, between corresponding 1-dimensional points. This way, each correspondence between a pair of 1-dimensional points is defined by each coordinate-pair of the 2-dimensional input point-set.

The problem, which is illustrated in figure 2.3, is the following: given a set of 2D data points, find the line which minimizes the sum of squared perpendicular distances, subject to the condition that none of the valid points deviates from this line by more than t units. This is actually two problems: a line fit to the data and a classification of the data into inliers (valid points) and outliers.

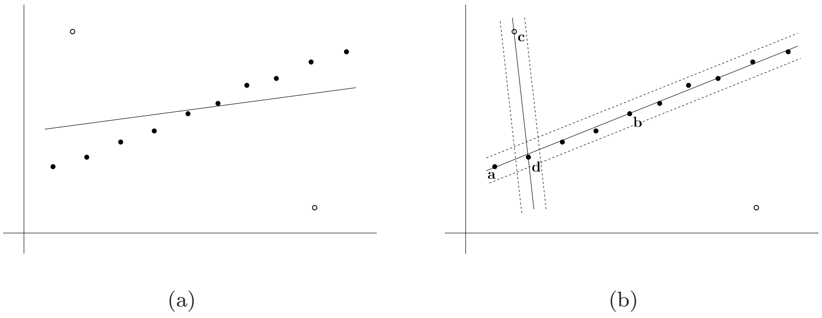


Figure 2.3: The solid points are inliers, the open points outliers. (a) A least-squares fit to the point data is severely affected by the outliers. (b) In the RANSAC algorithm the support for lines through randomly selected point pairs is measured by the number of points within a threshold distance of the lines. The dotted lines indicate the threshold distance. For the lines shown the support is 10 for line $\overline{\mathbf{ab}}$ (where both of the points $\mathbf{a}$ and $\mathbf{b}$ are inliers); and 2 for line $\overline{\mathbf{cd}}$ where the point $\mathbf{c}$ is an outlier.

The RANdom SAmple Consensus (RANSAC) algorighm by Fischler & Bolles (1981) is a very successful robust estimator which is able to deal with a large number of outliers. The idea is very simple: two of the points are selected randomly; these points define a line. The *support* for this line is measured by the number of points that lie within a distance threshold. This random selection is repeated a number of times and the line with most support is considered the robust fit. The points within the distance threshold are the inliers (constituting the *consensus* set). The main intuition behind is that if a point is an outlier it will gain not much support (see figure 2.3.(b)). Furthermore, scoring a line by its support has the advantage of favouring better fits. For example, the line $\overline{\mathbf{ab}}$ in figure 2.3.(b) has a support of 10, whereas the line $\overline{\mathbf{ad}}$, has a support of only 4. Consequently, even though both samples contain no outliers, the line $\overline{\mathbf{ab}}$ will be selected.

More generally, we wish to fit a model, in this case a line, to data, and the *random sample* consists of a minimal subset of the data, in this case two points, sufficient to determine the model. If the model is one of the transformation models described above, namely, a similarity, affinity or projectivity, and the data a set of 2D point correspondences, then the minimal subsets correspond to two, three and four correspondences, respectively.

As stated by Fischler & Bolles (1981) "The RANSAC procedure is opposite to that of conventional smoothing techniques: Rather than using as much of the data as possible to obtain an initial solution and then attempting to eliminate the invalid data points, RANSAC uses as small an initial data-set as feasible and enlarges this set with consistent data when possible".

The RANSAC algorithm is summarized as follows.

1. Randomly select a sample of m corresponding data point pairs from $\mathcal{X}$ and $\mathcal{Y}$, where m is the size of the minimal subset according to the type of transformation, and compute the transformation parameters.

2. Transform one point-set, e.g. using the methods depicted in section 2.3, and determine the subset of corresponding point pairs $\mathbf{x}_i, \mathbf{y}_i$ which are within a distance threshold t. This subset is the consensus set of the sample and defines the inliers of $\mathcal{X}$ and $\mathcal{Y}$.

3. If the number of inliers is greater than some threshold T, re-compute the transformation parameters using all the inliers and terminate.

4. If the number of inliers is less than T, select a new subset and repeat the above.

5. After N trials the largest consensus set is selected, and the transformation parameters are re-estimated using all the points in the consensus set.

Chapter 3

Point-Set Registration

3.1 Introduction

In the previous chapter we characterized several different geometrical transformation models and showed ways to estimate their parameters so as to minimize the point-to-point distance between two corresponding point-sets. However, the accuracy of these techniques is greatly affected by the presence of spurious correspondences or large errors in the location of the features. In practical situations therefore, it is more convenient to use robust techniques aimed at revising both the correspondence and alignment parameters. In this chapter, we describe methods to address the *point-set registration* problem, this is, the joint estimation of the correspondence and alignment parameters.

Although non-iterative algorithms exist for specific types of transformation models and limited amounts of noise (Ho & Yang, 2011), the point-set registration problem is usually solved by means of non-linear iterative methods that at each iteration estimate correspondence and alignment parameters. These methods do not guarantee to find the optimal solution as well as they usually depend on some rough initial estimate that may be located using, for example, the techniques described in chapter 1. However, they are fairly general in the sense of being easily extrapolated to a number of different geometric models as well as they usually present an acceptable robustness against noise.

We distinguish between two families of approaches for solving the registration problem.

In the first family, each point in one point set is influenced only by its nearest point in the other point set. This is the case of the popular *Iterative Closest Point* (ICP) algorithm (Besl & McKay, 1992), and variants described in section 3.1.1.

In the second family of approaches, each point is influenced by all the other points by means of a *multiply-linked* utility measure. We divide the approaches falling within this family into two categories: the ones that pose the problem in a statistical estimation framework using the *Expectation-Maximization* algorithm (EM) by Dempster *et al.* (1977), and the ones that use a technique known as *Softassign* in conjunction with deterministic annealing processes (Gold *et al.*, 1998; Rangarajan *et al.*, 1997), which are described in sections 3.1.2 and 3.1.3, respectively. The first ones have the advantage of offering statistical insights to the point-set registration problems while the others show desirable convergence properties.

In section 3.1.4 we stress some methods in the literature aimed at complementing either of the aforementioned approaches to point-set registration by the use of local discriminant features.

The remaining part of the chapter is divided as follows. In section 3.2, we introduce the theoretical background of the EM algorithm focused on the registration (i.e., correspondence and alignment) problem.

This algorithm, aimed at probabilistic estimation, is not restricted to a specific type of representations as far as one can define the probability distribution explaining a class of objects, no matter if they are point-sets or graphs or whatever. It is for this reason that we develop all the formalism by making no assumptions about neither the domain of the elements to be associated nor the probability distributions involved.

In section 3.3, we derive the EM algorithm for the case of point-set registration using Gaussian Mixture Models (GMM).

In section 3.4, the RPM algorithm is presented (Gold *et al.*, 1998; Rangarajan *et al.*, 1997), a multiply-linked approach to robust point matching that uses Softassign.

3.1.1 Point-Set Registration using ICP and Variants

Essentially, the steps of the ICP algorithm are the following.

1. Associate points by the nearest neighbor criterion

2. Estimate transformation parameters using a mean square cost function (e.g. as explained in section 2.3 for different types of transformation models).

3. Transform the points using the estimated parameters

4. Iterate (re-associate and so on).

Many variants of this algorithm have been proposed performing different choices on the following stages.

> *Selection* of some sets of points in one or both meshes. To cite some examples, Besl & McKay (1992) selects all available points or Chetverikov *et al.* (2005) selects trimmed subsets of points.
>
> *Matching* these points to points in the other point-set. Perhaps the simplest approach is to associate each point to its nearest neighbor (Besl & McKay, 1992).
>
> *Weighting* the pairs of corresponding points found by the previous two steps. One can assign a constant weight to each corresponding pair or, assign lower weights to pairs with higher distances (Godin *et al.*, 1994).
>
> *Rejection* of some corresponding pairs. This is a discontinuous version of the above weighting approach in which certain point correspondences are rejected to be considered outliers.

Error Metric. Although other error metrics may be used, the most common one is the sum-of-squared differences. Estimation of the transformation parameters within each iteration using this metric has usually closed form solutions for several types of geometric models such as the ones proposed in chapter 2.

Although ICP is attractive for its efficiency, it can be easily trapped in local minima due to the strict selection of the best point-to-point assignments. This makes ICP to be particularly sensitive both to initialization and the choice of a threshold needed to accept or to reject a match.

3.1.2 Point-Set Registration using the EM Algorithm

In the following, we cite some approaches to point-set registration using the EM algorithm.

Horaud *et al.* (2011) casted the problem of point-set registration as one of Gaussian mixture modelling. In this probabilistic framework, correspondences were decided in a Maximum A Posteriori (MAP) fashion, while alignment parameters were recovered by means of Maximum Likelihood (ML) estimation. They noted the advantages of using multiple conditional maximization steps within the EM algorithm in the typical case of maximizing over multiple parameters. They proposed methods for estimating optimal Euclidean transformation parameters both in the case of isotropic and anisotropic covariances in the mixture components. Euclidean registration was extended to articulated registration. Robustness against outliers was ensured by introducing a uniform component to the Gaussian mixture.

Myronenko & Song (2010) posed the rigid and non-rigid point-set registration as a Gaussian mixture modelling problem within the EM algorithm. The similarity case was defined for arbitrary dimensions. Robustness against outliers was achieved.

Jian & Vemuri (2005, 2011) posed the point-set registration problem of one of aligning two Gaussian mixtures such that a statistical measure of discrepancy between two Gaussians was minimized. This contrasts with most of the approaches that align a Gaussian mixture to a point-set. They presented solutions for the rigid and non-rigid registration cases in the presence of significant amounts of noise and outliers.

Although not using directly the EM algorithm, Tsin & Kanade (2004) presented an approach that extended the correlation technique typically used for aligning intensity images to the case of point-sets. According to this approach, point-sets are interpreted as intensity images where the presence or absence of a given point, defines the intensity value at that location. By means of a benefit measure that they called kernel correlation, registration problem was posed as one of finding the maximum kernel correlation (minimum entropy) configuration of the the two point-sets to be registered. A solution was proposed for the Euclidean transformations. This method presented robustness against outliers.

3.1.3 Point-Set Registration using Softassign and Deterministic Annealing

With regards to the point-set registration approaches using Softassign and deterministic annealing, we stress the following ones.

Gold *et al.* (1998); Rangarajan *et al.* (1997) developed Robust Point Matching (RPM), an approach for jointly estimating alignment and correspondence parameters. At each iteration continuous correspondences were estimated with Softassign and an annealing procedure was used to gradually push from continuous to $\{0, 1\}$ solutions.

Chui & Rangarajan (2000, 2003) presented TPS-RPM, a new algorithm that generalized RPM to non-rigid spatial mappings which were parameterized with thin-plate splines (TPS). Both approaches RPM and TPS-RPM have the ability of detecting and rejecting outliers.

Lee & Won (2011); Zheng & Doermann (2006) outperformed the TPS-RPM algorithm in a series of synthetic experiments by incorporating information about local neighborhood relations. They implemented an ICP-like algorithm with an heuristic lift which combined the refinement of a set of tentative matches through probabilistic relaxation and the non-rigid deformation of the point-sets with TPS.

3.1.4 Point-Set Registration using Local Discriminant Features

One possibility to improve the convergence of the point-set registration methods is by incorporating discriminant features. Approaches to attributed point-set registration have been presented that combine geometric information from salient points with feature vectors built from local evidence.

Belongie *et al.* (2002) use Shape Contexts for shape matching and object recognition in an ICP-like procedure. First, correspondences are estimated as described in section 1.4.3. Last, estimated correspondences are used to non-rigidly register the shapes using thin plate splines (Bookstein, 1989). This procedure is repeated until convergence.

Dungan & Potter (2010) presented an approach for attributed point-set registration that minimized a Mahalanobis distance between the concatenation of both spatial coordinates and image feature vectors, along spatial transformations. An appropriately chosen error covariance matrix facilitated to compare the various dimensions at appropriate scales. Robustness against outliers was achieved by the use of a least trimmed squares Hausdorff distance, which discards the worst measurements from the computations to be considered outliers.

Yang *et al.* (2011) presented an iterative approach for alternate alignment and correspondence using Softassign within a deterministic annealing procedure. The benefit (or similarity) matrix was computed as a product of point-location and feature-descriptor similarities. This benefit matrix was used to infer a continuous correspondence matrix which in turn, was used to compute the optimal nonrigid transformation parameters. Images were warped according to that transformation. This alternating correspondence and transformation procedure was repeated within an annealing scheme that gradually turned the continuous correspondences into $\{0, 1\}$ ones. Outliers were handled by means of the slack variables of Softassign.

Silletti *et al.* (2011) presented a general, non-iterative method for point-set matching. They combined position coordinates, raw image information and structural information in order to find the matches by means of the extremum principle reported by Scott & Longuet-Higgins (1991).

3.2 Theoretical Background of the EM Algorithm

Probabilistic registration assumes that the elements from the first set (the data) are random observations drawn from a probability distribution parameterized by the second set (the model). We will make no assumptions on the domain of the elements that we will denote with the sets $\mathcal{U} = \{u_a\, ,\, a \in \mathcal{I})$ (the data) and $\mathcal{V} = \{v_\alpha\, ,\, \alpha \in \mathcal{J})$ (the model), where $\mathcal{I} = 1 \ldots |\mathcal{U}|$ and $\mathcal{J} = 1 \ldots |\mathcal{V}|$ are the index-sets.

Therefore, the element-to-element assignment problem can be recast into one of estimating the parameters of the distribution that maximize the likelihood of observation of the first set. This is essentially a registration problem since we are seeking the hypothesis that maximize the overlap between the two sets.

Formally, we seek the parameters Θ (acting on the model-set $\mathcal{V}$) that maximize the observed-data likelihood

$$\Theta^\star = \arg\max_\Theta P\left(\mathcal{U}|\Theta\right) = \prod_{a\in\mathcal{I}} P\left(u_a|\Theta\right) \tag{3.1}$$

where we assume independence between the data observations.

This is known as the Maximum Likelihood (ML) estimate for Θ.

Direct maximization of (3.1) over the parameters is intractable due to the presence of *hidden variables*, namely, the unknown corresponding model elements $v_\alpha \in \mathcal{V}$.

This is the "chicken and egg" problem. We cannot take the two sets into overlap if we ignore the correspondences between their elements and, we cannot estimate the correspondences if we ignore how their elements overlap.

Since $\mathcal{V} = \{v_\alpha,\ \alpha \in \mathcal{J}\}$ is a partition of the event space, the marginal distribution of an observation can be expressed

$$P\left(u_a\right) = \sum_{\alpha\in\mathcal{J}} P\left(u_a, v_\alpha\right) \tag{3.2}$$

It is typical to introduce the *log likelihood function* in the ML estimation problems. Then, the *observed-datalog-likelihood*, i.e., $lnP\left(\mathcal{U}|\Theta\right) = \mathcal{L}\left(\Theta\right)$ becomes

$$\mathcal{L}\left(\Theta\right) = \sum_{a\in\mathcal{I}} ln\left(\sum_{\alpha\in\mathcal{J}} P\left(u_a, v_\alpha|\Theta\right)\right) \tag{3.3}$$

which is an expression involving the logarithm of a sum.

Since $ln\left(x\right)$ is a strictly increasing function, the value of Θ which maximizes $P\left(\mathcal{U}|\Theta\right)$ also maximizes $\mathcal{L}\left(\Theta\right)$.

Following the lines by Borman (2004), the EM algorithm is an iterative procedure for maximizing $\mathcal{L}\left(\Theta\right)$. Assume that after the n-th iteration the current estimate for Θ is given by $\Theta^{(n)}$. Since the objective is to maximize $\mathcal{L}(\Theta)$, we wish to compute an updated estimate Θ such that,

$$\mathcal{L}(\Theta) > \mathcal{L}(\Theta^{(n)}) \tag{3.4}$$

Equivalently, we want to maximize the difference

$$\mathcal{L}(\Theta) - \mathcal{L}(\Theta^{(n)}) = \sum_{a\in\mathcal{I}} \left\{ ln \left(\sum_{\alpha\in\mathcal{J}} P(u_a, v_\alpha|\Theta) \right) - lnP(u_a|\Theta^{(n)}) \right\} \tag{3.5}$$

where Θ are the parameters we are maximizing over and, $\Theta^{(n)}$ are the parameters from the previous iteration.

A result known as the *Jensen's inequality* states that

$$ln \sum_{i=1}^{m} \omega_i x_i \geq \sum_{i=1}^{m} \omega_i ln x_i \tag{3.6}$$

for constants $\omega_i \geq 0$ with $\sum_{i=1}^{m} \omega_i = 1$. This result may be applied to equation (3.5) provided that the constants ω_i can be identified. Consider letting the constants of the form $P(v_\alpha|u_a, \Theta^{(n)})$. Since $P(v_\alpha|u_a, \Theta^{(n)})$ is a probability measure, we have that $P(v_\alpha|u_a, \Theta^{(n)}) \geq 0$ and $\sum_{\alpha\in\mathcal{J}} P(v_\alpha|u_a, \Theta^{(n)}) = 1$ as required.

Then starting with equation (3.5) the constants $P(v_\alpha|u_a, \Theta^{(n)})$ are introduced as,

$$\begin{aligned}
\mathcal{L}(\Theta) - \mathcal{L}(\Theta^{(n)}) &= \sum_{a\in\mathcal{I}} \left\{ ln \left(\sum_{\alpha\in\mathcal{J}} P(u_a, v_\alpha|\Theta) \right) - lnP(u_a|\Theta^{(n)}) \right\} \\
&= \sum_{a\in\mathcal{I}} \left\{ ln \left(\sum_{\alpha\in\mathcal{J}} P(u_a, v_\alpha|\Theta) \frac{P(v_\alpha|u_a, \Theta^{(n)})}{P(v_\alpha|u_a, \Theta^{(n)})} \right) - lnP(u_a|\Theta^{(n)}) \right\} \\
&= \sum_{a\in\mathcal{I}} \left\{ ln \left(\sum_{\alpha\in\mathcal{J}} P(v_\alpha|u_a, \Theta^{(n)}) \frac{P(u_a, v_\alpha|\Theta)}{P(v_\alpha|u_a, \Theta^{(n)})} \right) - lnP(u_a|\Theta^{(n)}) \right\} \\
&\geq \sum_{a\in\mathcal{I}} \left\{ \sum_{\alpha\in\mathcal{J}} P(v_\alpha|u_a, \Theta^{(n)}) \, ln \left(\frac{P(u_a, v_\alpha|\Theta)}{P(v_\alpha|u_a, \Theta^{(n)})} \right) - lnP(u_a|\Theta^{(n)}) \right\} &(3.7) \\
&\geq \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} P(v_\alpha|u_a, \Theta^{(n)}) \, ln \left(\frac{P(u_a, v_\alpha|\Theta)}{P(v_\alpha|u_a, \Theta^{(n)}) \, P(u_a|\Theta^{(n)})} \right) \overset{def}{=} \Delta(\Theta|\Theta^{(n)}) &(3.8)
\end{aligned}$$

In going from equation (3.7) to (3.8) we make use of the fact that

$$\sum_{\alpha\in\mathcal{J}} P(v_\alpha|u_a, \Theta^{(n)}) = 1$$

so that

$$lnP(u_a|\Theta^{(n)}) = \sum_{\alpha\in\mathcal{J}} P(v_\alpha|u_a, \Theta^{(n)}) \, lnP(u_a|\Theta^{(n)})$$

which allows the term $lnP(u_a|\Theta^{(n)})$ to be brought into the summation.

Accordingly, we can define a new function $l\left(\Theta|\Theta^{(n)}\right)$ which is bounded above by the original function $\mathcal{L}\left(\Theta\right)$ in the following way

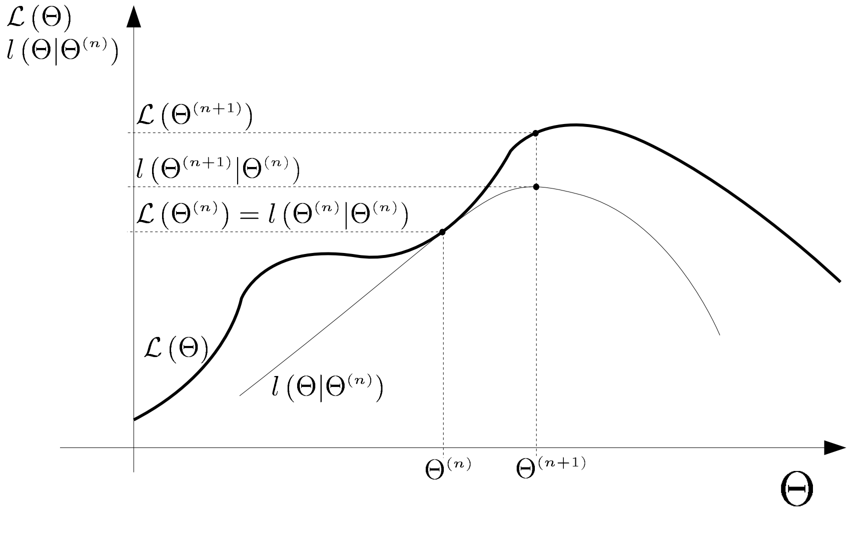
$$l\left(\Theta|\Theta^{(n)}\right) \overset{def}{=} \mathcal{L}\left(\Theta^{(n)}\right) + \Delta\left(\Theta|\Theta^{(n)}\right) \leq \mathcal{L}\left(\Theta\right) \tag{3.9}$$

Additionally, observe that

$$\begin{aligned} &l\left(\Theta^{(n)}|\Theta^{(n)}\right) = \mathcal{L}\left(\Theta^{(n)}\right) + \Delta\left(\Theta^{(n)}|\Theta^{(n)}\right) \\ &= \mathcal{L}\left(\Theta^{(n)}\right) + \sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}} P\left(v_\alpha|u_a,\Theta^{(n)}\right) ln\left(\frac{P\left(u_a,v_\alpha|\Theta^{(n)}\right)}{P\left(v_\alpha|u_a,\Theta^{(n)}\right)P\left(u_a|\Theta^{(n)}\right)}\right) \\ &\qquad = \mathcal{L}\left(\Theta^{(n)}\right) + \sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}} P\left(v_\alpha|u_a,\Theta^{(n)}\right) ln\left(\frac{P\left(u_a,v_\alpha|\Theta^{(n)}\right)}{P\left(u_a,v_\alpha|\Theta^{(n)}\right)}\right) \\ &\qquad = \mathcal{L}\left(\Theta^{(n)}\right) + \sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}} P\left(v_\alpha|u_a,\Theta^{(n)}\right) ln1 \\ &\qquad = \mathcal{L}\left(\Theta^{(n)}\right) \end{aligned} \tag{3.10}$$

so for $\Theta = \Theta^{(n)}$ the functions $l\left(\Theta|\Theta^{(n)}\right)$ and $\mathcal{L}\left(\Theta\right)$ have the same value.

Our objective is to find Θ so that $\mathcal{L}\left(\Theta\right)$ is maximum. We have shown that the function $l\left(\Theta|\Theta^{(n)}\right)$ is bounded above by the likelihood function $\mathcal{L}\left(\Theta\right)$ and that the value of the functions $l\left(\Theta|\Theta^{(n)}\right)$ and $\mathcal{L}\left(\Theta\right)$ is the same at the current estimate $\Theta = \Theta^{(n)}$. Therefore, any Θ which increases $l\left(\Theta|\Theta^{(n)}\right)$ in turn increases $\mathcal{L}\left(\Theta\right)$. In order to achieve the greatest possible increase in the value of $\mathcal{L}\left(\Theta\right)$ it is preferable to select Θ such that $l\left(\Theta|\Theta^{(n)}\right)$ is maximized. We denote this updated value as $\Theta^{(n+1)}$. This process is illustrated in Figure 3.1.



$\mathcal{L}\left(\Theta\right)$
$l\left(\Theta|\Theta^{(n)}\right)$
$\mathcal{L}\left(\Theta^{(n+1)}\right)$
$l\left(\Theta^{(n+1)}|\Theta^{(n)}\right)$
$\mathcal{L}\left(\Theta^{(n)}\right) = l\left(\Theta^{(n)}|\Theta^{(n)}\right)$
$\mathcal{L}\left(\Theta\right)$
$l\left(\Theta|\Theta^{(n)}\right)$
$\Theta^{(n)}$
$\Theta^{(n+1)}$
Θ

Figure 3.1: Graphical interpretation of an EM iteration. The function $l\left(\Theta|\Theta^{(n)}\right)$ is upper-bounded by the function $\mathcal{L}\left(\Theta\right)$. The functions have the same value at $\Theta = \Theta^{(n)}$. The EM algorithm chooses $\Theta^{(n+1)}$ so that maximizes $l\left(\Theta|\Theta^{(n)}\right)$. Since $\mathcal{L}\left(\Theta\right) \geq l\left(\Theta|\Theta^{(n)}\right)$ and $\mathcal{L}\left(\Theta^{(n)}\right) = l\left(\Theta^{(n)}|\Theta^{(n)}\right)$, increasing $l\left(\Theta|\Theta^{(n)}\right)$ ensures that the value of the likelihood function $\mathcal{L}\left(\Theta\right)$ is increased at each step.

Formally we have,

$$\begin{aligned}
\Theta^{(n+1)} &= \arg\max_{\Theta} l\left(\Theta|\Theta^{(n)}\right) \\
&= \arg\max_{\Theta} \Bigg\{ \mathcal{L}\left(\Theta^{(n)}\right) + \\
&+ \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} P\left(v_\alpha|u_a,\Theta^{(n)}\right) ln \left(\frac{P\left(u_a,v_\alpha|\Theta\right)}{P\left(v_\alpha|u_a,\Theta^{(n)}\right) P\left(u_a|\Theta^{(n)}\right)} \right) \Bigg\} \\
&\text{Now drop the terms which are constant w.r.t } \Theta \\
&= \arg\max_{\Theta} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} P\left(v_\alpha|u_a,\Theta^{(n)}\right) lnP\left(u_a,v_\alpha|\Theta\right) \\
&= \arg\max_{\Theta} \sum_{a\in\mathcal{I}} E_{\mathcal{V}|u_a,\Theta^{(n)}}\left[lnP\left(u_a,v_\alpha|\Theta\right)\right] \qquad (3.11)
\end{aligned}$$

There are two quantities involved in the maximization of equation (3.11), namely, the posterior probability terms $P\left(v_\alpha|u_a,\Theta^{(n)}\right)$ and the *complete-data log-likelihood* terms $lnP\left(u_a,v_\alpha|\Theta\right)$.

The EM algorithm consists in two steps: the E-step and the M-step. In the E-step, the conditional expectation function $E_{\mathcal{V}|u_a,\Theta^{(n)}}\left[lnP\left(u_a,v_\alpha|\Theta\right)\right]$ is determined. This is, a weighted sum of complete-data log-likelihoods, i.e., $\sum_\alpha \omega_{a\alpha}^{(n)} lnP\left(u_a,v_\alpha|\Theta\right)$, the weights being the missing data estimates under the current parameters. These missing data estimates $\omega_{a\alpha}^{(n)}$ play the role of selecting which is the appropriate mixture component $P\left(u_a,v_\alpha|\Theta\right)$ for evaluating the element u_a. In the M-step, the new parameters Θ are estimated that maximize the conditional expectation.

The main advantage of substituting the original observed-data log-likelihood term $lnP\left(u_a|\Theta\right)$ by a weighted sum of complete-data terms is that the former is much harder to model because no assumption on the correspondences is made.

3.2.1 Convergence of the EM Algorithm

The convergence properties of the EM algorithm are discussed in detail by McLachlan & Krishnan (1997). We discuss general convergence properties of the algorithm. Starting with the current estimate for Θ, which is $\Theta^{(n)}$, we had that $\Delta\left(\Theta^{(n)}|\Theta^{(n)}\right) = 0$. Since $\Theta^{(n+1)}$ is chosen to maximize $\Delta\left(\Theta|\Theta^{(n)}\right)$, we have that $\Delta\left(\Theta^{(n+1)}|\Theta^{(n)}\right) \geq \Delta\left(\Theta^{(n)}|\Theta^{(n)}\right) = 0$, so for each iteration $\mathcal{L}\left(\Theta\right)$ is non-decreasing.

3.2.2 The Generalized EM Algorithm

In the formulation used above, $\Theta^{(n+1)}$ was chosen so as to maximize the function $\Delta\left(\Theta^{(n+1)}|\Theta^{(n)}\right)$. While this ensures the maximum possible increase in $\mathcal{L}\left(\Theta\right)$, it is also possible to relax the maximization requirement to one of simply choosing Θ so that $\Delta\left(\Theta|\Theta^{(n)}\right) \geq \Delta\left(\Theta^{(n)}|\Theta^{(n)}\right)$. This approach, to simply increase and not necessarily maximize $\Delta\left(\Theta|\Theta^{(n)}\right)$ is known as the Generalized Expectation Maximization (GEM) algorithm and is often used in cases where the maximization is difficult. The convergence of the GEM algorithm can be argued as above.

3.3 Point-Set Registration with the EM Algorithm

In the standard approach for point-set registration with the EM algorithm, this problem is solved as an estimation of the parameters of a mixture of Gaussians.

Consider two point-sets $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ (the data) and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$ (the model), where $\mathcal{I} = 1 \ldots |\mathcal{X}|$ and $\mathcal{J} = 1 \ldots |\mathcal{Y}|$ are the index-sets.. A Gaussian Mixture Model (GMM) is fitted to the data-set $\mathcal{X}$ such that the centers of the Gaussian densities are constrained to coincide with the transformed model points $\mathbf{y}'_\alpha = \mathcal{T}(\mathbf{y}_\alpha; \Phi),\ \alpha \in \mathcal{J}$. Therefore, each density in the mixture is characterized by a mean vector $\mathbf{y}'_\alpha$ and a covariance matrix Σ_α. In the point-set registration approach the means are constrained to vary according to the family of geometric transformations $\mathcal{T}(\cdot; \Phi)$ which enforce prior knowledge about the transformation that exists between the two sets of points.

The operation of the EM algorithm is as follows:

1. Provide initial values for the model parameters (i.e., $\Phi^{(0)}$ and $\Sigma^{(0)} = \{\Sigma_\alpha^{(0)},\ \alpha \in \mathcal{J}\}$).

2. *E-step.* Compute the posterior probabilities given the current estimates of the alignment parameters $\Phi^{(n)}$ and the covariance matrices $\Sigma^{(n)}$.

$$\omega_{a\alpha}^{(n)} = P\left(\mathbf{y}_\alpha | \mathbf{x}_a, \Phi^{(n)}, \Sigma_\alpha^{(n)}\right) \tag{3.12}$$

3. *M-step.* Maximize the conditional expectation with respect to the parameters.

$$\{\Phi^{(n+1)}, \Sigma^{(n+1)}\} = \arg\max_{\Phi,\Sigma} \left\{ \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} ln\left(P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi, \Sigma_\alpha\right)\right) \right\} \tag{3.13}$$

4. Check for convergence

3.3.1 Expectation

We make explicit the posterior probabilities by using the Bayes' rule. Hence, we write,

$$\begin{aligned} P\left(\mathbf{y}_\alpha | \mathbf{x}_a, \Phi^{(n)}, \Sigma_\alpha^{(n)}\right) &= \frac{P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi^{(n)}, \Sigma_\alpha^{(n)}\right)}{P\left(\mathbf{x}_a | \Phi^{(n)}, \Sigma_\alpha^{(n)}\right)} \\ &= \frac{P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi^{(n)}, \Sigma_\alpha^{(n)}\right)}{\sum_{\alpha'} P\left(\mathbf{x}_a, \mathbf{y}_{\alpha'} | \Phi^{(n)}, \Sigma_{\alpha'}^{(n)}\right)} \end{aligned} \tag{3.14}$$

We use a Gaussian distribution with mean $\mathcal{T}(\mathbf{y}_\alpha; \Phi)$ and covariance Σ_α to model the complete-data likelihood term. This is, the likelihood of an observation given its assignment to a certain model point $\mathbf{y}_\alpha$. Note the convenience of modeling the complete-data likelihood as opposed to modeling the incomplete-data likelihood $P\left(\mathbf{x}_a | \Phi^{(n)}, \Sigma_\alpha^{(n)}\right)$ which do not assumes any assignment. Hence, we write,

$$P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi^{(n)}, \Sigma_\alpha^{(n)}\right) = \frac{1}{|2\pi\Sigma_\alpha^{(n)}|^{1/2}} \exp\left[-\tfrac{1}{2}\left\|\mathbf{x}_a - \mathbf{y}_\alpha^{(n)}\right\|^2_{\Sigma_\alpha^{(n)}}\right] \tag{3.15}$$

where $\mathbf{y}_\alpha^{(n)} = \mathcal{T}(\mathbf{y}_\alpha; \Phi^{(n)})$ is the transformed model point $\mathbf{y}_\alpha$ according to transformation parameters $\Phi^{(n)}$ and $\|\mathbf{x}\|_\Sigma^2 = \mathbf{x}^\top \Sigma^{-1} \mathbf{x}$ is the squared Mahalanobis distance with covariance matrix Σ.

The final expression for the posterior probabilities of equation (3.14) is

$$\omega_{a\alpha}^{(n)} = \frac{\frac{1}{|\Sigma_\alpha^{(n)}|^{1/2}} \exp\left[-\frac{1}{2}\left\|\mathbf{x}_a - \mathbf{y}_\alpha^{(n)}\right\|_{\Sigma_\alpha^{(n)}}^2\right]}{\sum_{\alpha'} \frac{1}{|\Sigma_{\alpha'}^{(n)}|^{1/2}} \exp\left[-\frac{1}{2}\left\|\mathbf{x}_a - \mathbf{y}_{\alpha'}^{(n)}\right\|_{\Sigma_{\alpha'}^{(n)}}^2\right]} \tag{3.16}$$

3.3.2 Maximization

In the maximization step we have to find the parameters Φ and Σ that maximize the conditional expectation of equation (3.13). Since the complete-data likelihoods are modeled with Gaussian distributions, maximization of equation (3.13) aims at estimating the parameters of a GMM where the missing data estimates $\omega_{a\alpha}^{(n)}$ are the membership variables. By replacing conditional probabilities with the Gaussian distributions of equation (3.15) and by neglecting constant terms not depending on neither Φ or Σ, this can be expressed in the following minimization form

$$\{\Phi^{(n+1)}, \Sigma^{(n+1)}\} = \arg\min_{\Phi,\Sigma} \left\{ \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \left\|\mathbf{x}_a - \mathcal{T}(\mathbf{y}_\alpha; \Phi)\right\|_{\Sigma_\alpha}^2 + ln|\Sigma_\alpha| \right\} \tag{3.17}$$

In the standard GMM estimation with the EM algorithm, estimation of the means and covariances is fairly straightforward. In the case of point registration, however, the means are constrained by the alignment parameters, and moreover, the functions $\mathcal{T}(\mathbf{y}_\alpha; \Phi)$ are complicated by the presence of rotation matrices. The simultaneous estimation of all the model parameters within the M-step would lead to a difficult nonlinear minimization problem. It is a well-established strategy to replace this maximization with a sequence of *conditional maximization* steps. Then, it turns into an instance of the *expectation conditional maximization* (ECM) algorithm, which has proven to be more broadly applicable than EM, while it shares its desirable convergence properties (Meng & Rubin, 1993).

According to ECM, minimization of equation (3.17) can be decomposed into two steps. First, minimize (3.17) over Φ while keeping the covariance matrices constant and next, estimate empirical covariances using the newly estimated alignment parameters. This is,

$$\Phi^{(n+1)} = \arg\min_{\Phi} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \left\|\mathbf{x}_a - \mathcal{T}(\mathbf{y}_\alpha; \Phi)\right\|_{\Sigma_\alpha^{(n)}}^2 \tag{3.18}$$

and, for all $a \in \mathcal{I}$

$$\Sigma_a^{(n+1)} = \frac{\sum_\alpha \omega_{a\alpha}^{(n)} (\mathbf{x}_a - \mathcal{T}(\mathbf{y}_\alpha; \Phi^{(n+1)}))(\mathbf{x}_a - \mathcal{T}(\mathbf{y}_\alpha; \Phi^{(n+1)}))^\top}{\sum_\alpha \omega_{a\alpha}^{(n)}} \tag{3.19}$$

As stated by Bishop (2006); Ingrassia & Rocci (2007), it may happen that when one of the components of the GMM collapses into a data point while the others are infinitely away, the elements of the covariance matrix tend to zero. In order to avoid possible degeneracies, one may model all the components of the mixture with a common covariance matrix:

$$\Sigma^{(n+1)} = \frac{\sum_a \sum_\alpha \omega_{a\alpha}^{(n)} \left(\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi^{(n+1)}\right)\right) \left(\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi^{(n+1)}\right)\right)^\top}{\sum_a \sum_\alpha \omega_{a\alpha}^{(n)}} \tag{3.20}$$

In section 3.5 we describe how to estimate the optimal transformation parameters of equation (3.18).

3.4 Point-Set Registration with Softassign

Gold *et al.* (1998); Rangarajan *et al.* (1997) developed robust point matching (RPM), a robust algorithm for recovering the alignment and correspondence parameters relating two sets of points. They noted that once one of these parameters is fixed, it is an easy problem to estimate the other. Hence, inspired in earlier works on neural networks and statistical physics, they devised an iterative framework consisting on alternative updates of correspondence and alignment parameters. The behavior of the algorithm is governed by Softassign, a procedure that starting from an ambiguous solution, gradually disambiguate it by means of an annealing procedure. This feature gives the algorithm the ability of avoiding poor local optima. Another novelty is the two-way constraints satisfaction in the correspondences matrix such that one point in one point set can only be assigned to one point in the other point-set, and vice-versa. This is achieved due to a result by Sinkhorn (1964) consisting on alternative row and column normalization.

Having already stressed the most important features of the method, let us introduce some notation. Let $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$ be two sets of points, where $\mathcal{I} = 1, \ldots, |\mathcal{X}|$ and $\mathcal{J} = 1, \ldots, |\mathcal{Y}|$ are the index-sets. Let S be the matrix of correspondences such that its (a, α)-th element $s_{a\alpha} \in [0, 1]$ denote the probability of point $\mathbf{x}_a$ being in correspondence with point $\mathbf{y}_\alpha$.

The aim is to locate the correspondence and alignment parameters, $S^\star, \Phi^\star$, that minimize the following energy function

$$\{S^\star, \Phi^\star\} = \arg\min_{S,\Phi} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} s_{a\alpha} \left(\left\| \mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi\right) \right\|^2 - \rho \right) \tag{3.21}$$

subject to

$$\forall a, \sum_{\alpha\in\mathcal{J}} s_{a\alpha} \leq 1,\ \forall \alpha, \sum_{a\in\mathcal{I}} s_{a\alpha} \leq 1,\ \forall a, \alpha\ s_{a\alpha} \in \{0, 1\} \tag{3.22}$$

Some approaches (Gold *et al.*, 1998) may include a regularization term (which we have omitted) in order to penalize the less likely transformations.

This minimization is essentially a weighted sum of squared alignment errors, where the correspondence indicators gate the contributions of the alignment errors between each possible pair of points. Accordingly, optimal alignment parameters tend to minimize the distances between corresponding points and, conversely, optimal correspondence parameters tend to match points which are closer to each other. The term ρ acts as a threshold error distance indicating how far apart two points must be before the points must be treated as outliers.

This issue is addressed by the *slack* variables that will be explained later.

In order to simplify the formulation, the inequality constraints of equation (3.22) were ignored, turning them into equality constraints. Accordingly, correspondence matrix S must be a *permutation matrix.*

Next, *deterministic annealing* methods were used to turn the discrete problem into a continuous one. Deterministic annealing methods are methods indexed by a control parameter that minimize a series of objective functions. As the parameter is increased the solution to the objective function approach to that of the discrete problem. Alternative row and column normalization of the correspondence matrix S were used in order to enforce two-way constraints. This way, matrix S is relaxed from a permutation matrix to a *doubly stochastic* matrix, i.e., the continuous analog of a permutation matrix.

As said, they devised an iterative algorithm in which correspondence and alignment parameters were updated in decoupled steps. Supposing that alignment parameters are fixed, the aim is to find the matrix S according to the following expression

$$\arg\max_S \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} s_{a\alpha} B_{a\alpha} \tag{3.23}$$

subject to

$$\forall a \sum_{\alpha} s_{a\alpha} = 1,\ \forall\alpha \sum_{a} s_{a\alpha} = 1,\ \forall a, \alpha\ s_{a\alpha} \in \{0, 1\}$$

where, in the problem at hand,

$$B_{a\alpha} = -\left(\left\|\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi\right)\right\|^2 - \rho\right) \tag{3.24}$$

where the sign is reversed because the assignment problem is a maximization one instead of a minimization one.

This is known as an assignment problem, a classic problem in combinatorial optimization where $B_{a\alpha}$ is the benefit coefficient for the assignment $a \to \alpha$.

This discrete problem may be turned into a continuous problem by introducing a control parameter $\mu > 0$ and setting S as

$$s_{a\alpha} = \frac{\exp\left[\mu B_{a\alpha}\right]}{\sum\limits_{\alpha'\in\mathcal{J}} \exp\left[\mu B_{a\alpha'}\right]} \tag{3.25}$$

This is known as the *softmax.* Note that the exponentiation ensures that all the elements of S are positive. It is easy to see that as the control parameter $\mu \to \infty$ the elements of S tend to $\{0, 1\}$ values.

In order to enforce two-way constraints they replaced the one-way normalization of equation (3.25) by a Sinkhorn normalization. Since the method returns a doubly stochastic matrix instead of a permutation one, the ensemble of softmax and Sinkhorn normalization is known as the *Softassign.* This process is illustrated in figure 3.2.

They devised the solution to the correspondence and alignment problem of equation (3.21) as a succession of correspondence and alignment problems embodied within an annealing procedure. Discarding the quantities constant with respect to Φ from equation (3.21), recovery of the alignment parameters

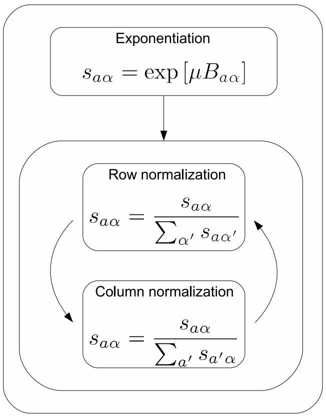

Exponentiation
$s_{a\alpha} = \exp\left[\mu B_{a\alpha}\right]$
Row normalization
$s_{a\alpha} = \frac{s_{a\alpha}}{\sum_{\alpha'} s_{a\alpha'}}$
Column normalization
$s_{a\alpha} = \frac{s_{a\alpha}}{\sum_{a'} s_{a'\alpha}}$

Figure 3.2: Overview of the Softassign procedure.

was performed according to the following expression.

$$\arg\min_{\Phi} \sum_{\alpha\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} s_{a\alpha} \left\| \mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha;\Phi\right) \right\|^2 \tag{3.26}$$

whose solution for different types of spatial transformations is explained in the next section.

With these ingredients, the final algorithm for solving the minimization of equation (3.21) is the following.

1. Initialize $S^{(0)}$ to some valid value.
2. Update alignment parameters from equation (3.26) as explained in section 3.5.
3. Update correspondence parameters by Softassign.
4. Increase the control parameter μ and repeat until convergence or until μ reaches a predefined threshold.

This process is illustrated in figure 3.3.

In order to be able to handle with outliers in a statistically robust way, constraints on S must be inequality constraints, not equality constraints. This is done by introducing slack variables, a standard technique from linear programming. This is,

$$\forall a \sum_{\alpha\in\mathcal{J}} s_{a\alpha} \leq 1 \rightarrow \forall a \sum_{\alpha\in\{\mathcal{J}\cup\emptyset\}} s_{a\alpha} = 1 \tag{3.27}$$

and likewise for column constraints. An augmented matrix of assignments $\tilde{S}$ is created in order to hold the slack variables by adding an extra row and column to the matrix S.

Recall from equation (3.21) that the constant ρ established the thresholding error in order to consider an outlier correspondence. Accordingly, a thresholding correspondence is associated with a zero benefit value of equation (3.24). This way, a point $\mathbf{x}_a \in \mathcal{X}$ is considered an outlier (and therefore, matched to the

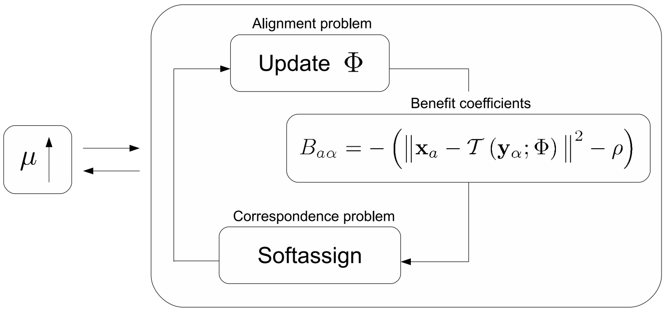

Alignment problem
Update Φ
Benefit coefficients
$B_{a\alpha} = -\left(\left\| \mathbf{x}_a - \mathcal{T}(\mathbf{y}_\alpha; \Phi) \right\|^2 - \rho \right)$
μ
Correspondence problem
Softassign

Figure 3.3: Overview of the Robust Point Matching algorithm. The procedure iterates until convergence or until μ reaches a predefined threshold.

null-node) if $\forall \alpha, B_{a\alpha} < 0$. Since the exponential of the zero-benefit equals to unity, this is the value that must hold the slack variables at each iteration before the Sinkhorn normalization.

This process gradually pushes from ambiguous to unambiguous states as the control parameter increases.

3.5 Estimation of the Transformation Parameters

In the present section we describe the details of the recovery of the transformation parameters in the case of the aforementioned methods.

Although they may look similar, there are substantial differences between the recovery of the transformation parameters in a strict correspondences setting and in a multiply-linked setting. In the former case, which is the case of the ICP-like approaches, recovery of the transformation parameters reduces at solving an optimization problem of the form

$$\Phi^\star = \arg\min_{\Phi} \sum_i \left\| \mathbf{x}_i - \mathcal{T}(\mathbf{y}_i; \Phi) \right\|^2 \tag{3.28}$$

which is discussed in section 2.3.

In the latter case, and specially in the case of approaches with a statistical motivation, the squared distance above usually becomes a weighted sum of squared Mahalanobis distances of the form

$$\Phi^\star = \arg\min_{\Phi} \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} \omega_{a\alpha} \left\| \mathbf{x}_a - \mathcal{T}(\mathbf{y}_\alpha; \Phi) \right\|^2_{\Sigma_\alpha} \tag{3.29}$$

Recovery of the optimal transformation parameters at each iteration is more complex in the multiply-linked approaches than in the strict correspondences ones due to the weights $\omega_{a\alpha}$ and the covariance matrices Σ_α.

The cases without a statistical motivation, such as the RPM algorithm in section 3.4, can be seen as special cases of equation (3.29) where the covariances are equal to the identity matrix.

In the following, we describe approaches to recover the optimal transformation parameters in the case of similarity, affine and projective transformations in a multiply-linked setting.

3.5.1 Similarity Transformations

From equation (3.18) we seek the optimal rotation matrix $\mathtt{R}^\star$, scaling parameter $\eta^\star$ and translation 2-vector $\mathbf{t}^\star$ that minimize the following quantity

$$\min_{\mathtt{R},\eta,\mathbf{t}} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} \big\| \mathbf{x}_a - (\eta \mathtt{R} \mathbf{y}_\alpha + \mathbf{t}) \big\|^2_{\Sigma_\alpha} \tag{3.30}$$

subject to $\det \mathtt{R} = \pm 1$.

As noticed by Horaud *et al.* (2011), weights $\omega_{a\alpha}$ define a spatial mapping of points in $\mathcal{X}$. This way, equation (3.30) can be simplified by introducing the virtual observation $\mathbf{w}_\alpha$ and its weight φ_α that are assigned to a model point $\mathbf{y}_\alpha$

$$\mathbf{w}_\alpha = \frac{1}{\varphi_\alpha} \sum_{a\in\mathcal{I}} \omega_{a\alpha} \mathbf{x}_a \tag{3.31}$$

$$\varphi_\alpha = \sum_{a\in\mathcal{I}} \omega_{a\alpha} \tag{3.32}$$

By introducing the virtual observation of equations (3.31) and (3.32) minimization of equation (3.30) can be expressed in the simpler form

$$\min_{\mathtt{R},\eta,\mathbf{t}} \sum_{\alpha\in\mathcal{J}} \varphi_\alpha^{(n)} \big\| \mathbf{w}_\alpha - (\eta \mathtt{R} \mathbf{y}_\alpha + \mathbf{t}) \big\|^2_{\Sigma_\alpha} \tag{3.33}$$

In order to simplify the problem we will assume that the covariances are isotropic, namely, $\Sigma_\alpha = \sigma_\alpha^2 \mathtt{I}_2$ where $\mathtt{I}_2$ is the 2×2 identity matrix. This way, the Mahalanobis distance reduces to the Euclidean distance. We obtain

$$\min_{\mathtt{R},\eta,\mathbf{t}} \sum_{\alpha\in\mathcal{J}} \varphi_\alpha^{(n)} \sigma_\alpha^{-2} \big\| \mathbf{w}_\alpha - (\eta \mathtt{R} \mathbf{y}_\alpha + \mathbf{t}) \big\|^2 \tag{3.34}$$

Weights φ_α prevent us to find the parameters in the same way as the unweighted case (section 2.3.1). The reason is that we cannot estimate the means and variances because different importance is given to each point $\mathbf{w}_\alpha, \mathbf{y}_\alpha$ according to weights φ_α. We follow the Ansatz by Rangarajan *et al.* (1997) and compute the weighted mean and variances of the point-sets. This is,

$$\bar{\mathbf{w}} = \frac{\sum_\alpha \varphi_\alpha \mathbf{w}_\alpha}{\sum_\alpha \varphi_\alpha} \tag{3.35}$$

$$\bar{\mathbf{y}} = \frac{\sum_\alpha \varphi_\alpha \mathbf{y}_\alpha}{\sum_\alpha \varphi_\alpha} \tag{3.36}$$

$$\sigma_w^2 = \frac{\sum_\alpha \varphi_\alpha \left\| \mathbf{w}_\alpha - \bar{\mathbf{w}} \right\|^2}{\sum_\alpha \varphi_\alpha} \tag{3.37}$$

$$\sigma_y^2 = \frac{\sum_\alpha \varphi_\alpha \left\| \mathbf{y}_\alpha - \bar{\mathbf{y}} \right\|^2}{\sum_\alpha \varphi_\alpha} \tag{3.38}$$

(note that the variances σ_α^{-2} in the quotient cancel out).

Optimal parameters $\mathtt{R}^\star$, $\eta^\star$ and $\mathbf{t}^\star$ are then found following the same approach as in the unweighted case which is explained in section 2.3.1, equations (2.19), (2.20) and (2.21).

3.5.2 Affine Transformations

In the case of affinities we seek the optimal non-singular matrix $\mathtt{A}^\star$ and translation 2-vector $\mathbf{t}^\star$ that minimize the following quantity

$$\min_{\mathtt{A},\mathbf{t}} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} \left\| \mathbf{x}_a - (\mathtt{A}\mathbf{y}_\alpha + \mathbf{t}) \right\|^2_{\Sigma_\alpha} \tag{3.39}$$

where $\mathtt{A} = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$ and $\mathbf{t} = (t^{V}, t^{H})$.

We denote the elements of the inverse covariance matrix as $\Sigma_\alpha^{-1} = \left[\begin{smallmatrix} \sigma_{\alpha 11}^{-2} & \sigma_{\alpha 12}^{-2} \\ \sigma_{\alpha 21}^{-2} & \sigma_{\alpha 22}^{-2} \end{smallmatrix}\right]$.

Consider the following residuals from the alignment of points $\mathbf{x}_a$ and $\mathbf{y}_\alpha$.

$$\begin{aligned} r^{V}_{a\alpha} &= x^{V}_{a} - (a_{11} y^{V}_{\alpha} + a_{12} y^{H}_{\alpha} + t^{V}) \\ r^{H}_{a\alpha} &= x^{H}_{a} - (a_{21} y^{V}_{\alpha} + a_{22} y^{H}_{\alpha} + t^{H}) \end{aligned} \tag{3.40}$$

Then, optimization problem of equation (3.39) is equivalent at minimizing the following quantity

$$\mathcal{F}_{wa} = \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} \left(\sigma_{\alpha 11}^{-2} \left(r^{V}_{a\alpha}\right)^2 + \sigma_{\alpha 22}^{-2} \left(r^{H}_{a\alpha}\right)^2 + \sigma_{\alpha 12}^{-2} r^{V}_{a\alpha} r^{H}_{a\alpha} + \sigma_{\alpha 21}^{-2} r^{V}_{a\alpha} r^{H}_{a\alpha} \right) \tag{3.41}$$

Taking derivatives of $\mathcal{F}_{wa}$ with respect to the parameters we obtain the following expressions.

$$\begin{aligned}
\frac{\delta\mathcal{F}_{wa}}{\delta a_{11}} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} y^{V}_{\alpha} \left(\sigma_{\alpha 11}^{-2} 2 r^{V}_{a\alpha} + r^{H}_{a\alpha} \left(\sigma_{\alpha 12}^{-2} + \sigma_{\alpha 21}^{-2} \right) \right) \\
\frac{\delta\mathcal{F}_{wa}}{\delta a_{12}} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} y^{H}_{\alpha} \left(\sigma_{\alpha 11}^{-2} 2 r^{V}_{a\alpha} + r^{H}_{a\alpha} \left(\sigma_{\alpha 12}^{-2} + \sigma_{\alpha 21}^{-2} \right) \right) \\
\frac{\delta\mathcal{F}_{wa}}{\delta a_{21}} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} y^{V}_{\alpha} \left(\sigma_{\alpha 22}^{-2} 2 r^{H}_{a\alpha} + r^{V}_{a\alpha} \left(\sigma_{\alpha 12}^{-2} + \sigma_{\alpha 21}^{-2} \right) \right) \\
\frac{\delta\mathcal{F}_{wa}}{\delta a_{22}} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} y^{H}_{\alpha} \left(\sigma_{\alpha 22}^{-2} 2 r^{H}_{a\alpha} + r^{V}_{a\alpha} \left(\sigma_{\alpha 12}^{-2} + \sigma_{\alpha 21}^{-2} \right) \right) \\
\frac{\delta\mathcal{F}_{wa}}{\delta t^{V}} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} \left(\sigma_{\alpha 11}^{-2} 2 r^{V}_{a\alpha} + r^{H}_{a\alpha} \left(\sigma_{\alpha 12}^{-2} + \sigma_{\alpha 21}^{-2} \right) \right) \\
\frac{\delta\mathcal{F}_{wa}}{\delta t^{H}} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha} \left(\sigma_{\alpha 11}^{-2} 2 r^{H}_{a\alpha} + r^{V}_{a\alpha} \left(\sigma_{\alpha 12}^{-2} + \sigma_{\alpha 21}^{-2} \right) \right)
\end{aligned} \tag{3.42}$$

Optimal parameters $\mathtt{A}^\star$ and $\mathbf{t}^\star$ are found by solving the set of equations

$$\frac{\delta\mathcal{F}_{wa}}{\delta a_{11}} = 0, \ldots, \frac{\delta\mathcal{F}_{wa}}{\delta t^{H}} = 0$$

with respect to the parameters.

This linear system can be expressed in matrix form $M\mathbf{a} = \mathbf{b}$, where M is a 6×6 matrix and $\mathbf{a} = (a_{11}, a_{12}, a_{21}, a_{22}, t^{V}, t^{H})$ and $\mathbf{b}$ are 6-column-vectors. This can be solved by matrix inversion (i.e., $\mathbf{a} = M^{-1}\mathbf{b}$).

3.5.3 Projective Transformations

Recall that a planar point $\mathbf{y} = (y^V, y^H)$ in homogeneous coordinates is represented by the 3-vector $\tilde{\mathbf{y}} = (y^V, y^H, 1)$, and that a planar projective transformation is a transformation involving homogeneous vectors. Let $g : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ be a function that maps a point in homogeneous form to its corresponding point in the plane by mapping point $\left(y^1, y^2, y^3\right)$ to point $\left(y^1/y^3, y^2/y^3\right)$.

We seek the optimal homography matrix $\mathtt{H}^\star$ that minimize the following expression

$$\min_{\mathtt{H}} \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} \omega_{a\alpha} \left\| \mathbf{x}_a - g\left(\mathtt{H}\tilde{\mathbf{y}}_\alpha\right) \right\|^2_{\Sigma_\alpha} \tag{3.43}$$

where

$$\mathtt{H} = \begin{bmatrix} a_{11} & a_{12} & t^V \\ a_{21} & a_{22} & t^H \\ w_1 & w_2 & 1 \end{bmatrix} \tag{3.44}$$

We denote the elements of the inverse covariance matrix as $\Sigma_\alpha^{-1} = \left[\begin{smallmatrix} \sigma_{\alpha}{}_{11}^{-2} & \sigma_{\alpha}{}_{12}^{-2} \\ \sigma_{\alpha}{}_{21}^{-2} & \sigma_{\alpha}{}_{22}^{-2} \end{smallmatrix}\right]$. Consider the following residuals from the alignment of points $\mathbf{x}_a$ and $\mathbf{y}_\alpha$.

$$\begin{aligned} r^V_{a\alpha} &= x^V_a - \left(\frac{a_{11}y^V + a_{12}y^H + t^V}{w_1 y^V + w_2 y^H + 1}\right) \\ r^H_{a\alpha} &= x^H_a - \left(\frac{a_{21}y^V + a_{22}y^H + t^H}{w_1 y^V + w_2 y^H + 1}\right) \end{aligned} \tag{3.45}$$

Then, optimization problem of equation (3.43) is equivalent at minimizing the following quantity

$$\mathcal{F}_{wh} = \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} \omega_{a\alpha} \left(\sigma_{\alpha}{}_{11}^{-2} \left(r^V_{a\alpha}\right)^2 + \sigma_{\alpha}{}_{22}^{-2} \left(r^H_{a\alpha}\right)^2 + \sigma_{\alpha}{}_{12}^{-2} r^V_{a\alpha} r^H_{a\alpha} + \sigma_{\alpha}{}_{21}^{-2} r^V_{a\alpha} r^H_{a\alpha} \right) \tag{3.46}$$

The system of equations obtained by equating to zero the partial derivatives has no solution in closed form due to the non-linearity in the residuals. Instead, a non-linear iterative method such as Gauss-Newton or Levenberg-Marquardt should be applied.

Chapter 4

Graph Matching

4.1 Introduction

In the previous chapter we have reviewed point-set registration methods aimed at enforcing global geometric consistency by means of unary coordinate measurements. The aims of graph matching is to enforce structural consistency allowing for relational measurements. In the graph matching literature the features to be associated are called *nodes* and the binary relations are represented by a set of links or *edges* between the nodes. Such a linkage conveys the structure of the graph, and is usually referred to as *structural* information. The task of graph matching is then to find the correspondences between the nodes in a way that the constraints imposed by the structure of the graphs are fulfilled.

Conte *et al.* (2004) present a review of the different approaches to graph matching during the last thirty years.

In the case of *unweighted* binary relations (i.e., with values in $\{0, 1\}$), this is the task of determining a structure-preserving matching. This is, if two nodes are linked by an edge, they have to be matched to two nodes linked by an edge as well. In its most stringent form, *graph isomorphism*, this condition must hold in both directions and the matching must be *bijective*. Graph isomorphism requires that the two graphs are identical in order for a feasible matching to exist. A less restrictive type of matching is *subgraph isomorphism* which requires a graph isomorphism between one graph and a node-induced subgraph of the other. Subgraph isomorphism is usually referred to a weaker form which only requires that structure is preserved in one direction.

Early attempts tried to solve the aforementioned problems in an *exact* way, this is, they either returned a feasible solution or halted. Most of them relied on some sort of tree search with backtracking such as the ones by Ghahraman *et al.* (1980); Ullmann (1976).

Graph or subgraph isomorphism are rarely found in practical situations because of the noise inherent to the graph extraction processes from real data. *Inexact* graph matching tries to overcome these difficulties by posing the graph matching problem as an optimization problem. Associated with inexact approaches there is a cost function which evaluates each configuration of matches. Then, the objective turns to finding the configuration of matches with the minimum cost, even when an exact graph or subgraph isomorphism does not exist.

An interesting property is that attribute information can be used to quantify this cost.

A well-known approach to inexact graph matching tries to determine the minimum-cost sequence of graph-edit operations needed to transform one graph into the other (Sanfeliu & Fu, 1983). The most commonly operations involved are node and edge insertion, deletion and substitution. Correspondences between the nodes are deduced from the edit-operations. Associated with each operation there is an edit-cost that reflects our prior knowledge about the likelihood of occurring that deformation (Bunke, 1999). Most of the algorithms to graph-edit distance computation rely on some sort of tree search with some heuristics in order to prune search space.

Approaches to inexact graph matching are divided into optimal and suboptimal.

Optimal approaches guarantee to find the best solution according to the cost function used. This is at the expenses of a high computational time, since computational times in graph matching often grow exponentially with the size of the graphs (the subgraph isomorphism problem is NP-complete). They can be seen as a generalization of the exact approaches since a graph or subgraph isomorphism will be found, if it exists. Some algorithms used to find the optimal sequence of edit operations are for example those by Sanfeliu & Fu (1983); Shapiro & Haralick (1981); Ullmann (1976).

Suboptimal approaches do not guarantee to find the optimal solution but they often find a reasonable solution in acceptable time. Most approaches to find a suboptimal sequence of edit operations are based on tree search as, for example, those by Eshera & Fu (1984); Serratosa *et al.* (2000).

Another approach to suboptimal inexact matching takes this inherently discrete combinatorial problem to the continuous domain. This way, nonlinear continuous optimization techniques can be applied, which normally have polynomial cost with low exponent. They start from an initial estimate which is improved at each iteration until a local optimum of the objective function is attained or a predefined number of iterations is done. There are usually a number of parameters that can be tuned in order to obtain more accurate solutions for each specific problem at hand.

We differentiate the existing approaches according to whether

- they incorporate unary measurements or not.
- binary measurements between nodes are *weighted* (values in $\mathbb{R}$) or *unweighted* (values in $\{0, 1\}$).
- correspondence variables are discretized at each iteration of the algorithm or, on the contrary, they remain continuous throughout the process and are discretized in a last step.

Probabilistic relaxation was among the first approaches to cast the graph matching problem as a continuous optimization problem. The objective is to find the correspondences such that the sum of supports received by the nodes is maximum. The support of each node is defined as the sum of compatibilities from the matched edges incident upon it. The influence from one node to its neighbours is represented by means of the compatibility coefficients which quantify the compatibility of each pair of simultaneous assignments. This influence can be interpreted as a weighted binary relation. Another feature of probabilistic relaxation is that it discretizes the correspondence variables in the last step of the algorithm, thus maintaining the uncertainty in the solutions during all the process.

Rosenfeld *et al.* (1976) developed a model to relax Waltz's discrete labels (Waltz, 1975) by means of probabilistic assignments. They introduced the notion of compatibility coefficients and laid the bases of probabilistic relaxation in graph matching. Hummel & Zucker (1983) firmly positioned the probabilistic relaxation into the continuous optimization domain by demonstrating that finding consistent labelings was equivalent at maximizing a local average consistency functional. Thus, the problem could be solved with standard continuous optimization techniques such as gradient ascent. Later on, Pelillo (1997) showed that in the case of positive compatibilities holding certain symmetry conditions, the heuristic update rule proposed by Rosenfeld *et al.* (1976), constitutes a *growth transformation* for the average local consistency defined by Hummel & Zucker (1983). Furthermore, they showed that these local maxima are local attractors under the action of the update rule. This means that if started near a local consistent labeling the process tends to go towards it.

Gold & Rangarajan (1996) developed an optimization technique, *Graduated Assignment*, specifically designed to the type of objective functions used in probabilistic relaxataion. They casted the problem in the more specific setting of matching two graphs instead of assigning a set of labels to a set of objects as in probabilistic relaxation. Accordingly, they introduced two-way constraints in the assignments function such that a node in one graph can only be assigned to a node in the other graph and vice-versa. They used a Taylor series expansion to approximate the solution of a quadratic assignment problem as a succession of easier linear assignment problems. They used *Softassign* (Chui & Rangarajan, 2003; Rangarajan *et al.*, 1996; Sinkhorn, 1964) to solve the linear assignment problems in the continuous domain. The main novelties were two-way constraints satisfaction and a continuation method to avoid poor local minima.

The main drawbacks of the approaches based on probabilistic relaxation are that they do not contemplate unary measurements in the nodes.

Along the same lines than probabilistic relaxation, Christmas *et al.* (1995) used statistical estimation to solve the problem. They derived the relaxation process in a Bayesian framework and showed how to bring together the unary measurements, binary relations and prior knowledge.

Also with a statistical motivation and the possibility to incorporate unary node measurements Cross & Hancock (1998); Wilson & Hancock (1997) performed graph matching using *cliques*, a kind of graph sub-entities. By using cliques the support for a node is converted into a more sophisticated measure based on a dictionary of feasible structural mappings. Cross & Hancock (1998); Wilson & Hancock (1997); Wilson *et al.* (1998) used cliques in order to detect outliers by measuring the net effects of a node deletion in the reconfigured graph. Accordingly, an outlier is a node that lead to an improvement in the consistency of the affected cliques after its removal. Nodes are regularly tested for deletion or reinsertion following this criterion. The main drawback is that this process of outlier detection is very time consuming since each node must be tested twice (for deletion and reinsertion), each time involving a graph reconfiguration. Although cliques provide a sophisticated measure of structural consistency they are limited to unweighted binary relations.

Luo & Hancock (2001) formulated the problem of graph matching as one of probability mixture modeling with *missing data*. This way, the correspondence variables are the parameters of the distribution and the corresponding nodes in the model-graph are the hidden variables. They used the Expectation-Maximization (EM) Algorithm (Dempster *et al.*, 1977) to find the Maximum Likelihood (ML) estimate of the correspondence indicators. Cross & Hancock (1998) used position coordinates as unary node measurements together with unweighted binary relations in order to recover the correspondences and alignment parameters in dual-steps of an EM algorithm. Bin & R. (1999); Cross & Hancock (1998); Luo & Hancock (2003, 2002) presented approaches to jointly solve the correspondence and alignment problems by exploiting both the unary position coordinates of the nodes and their binary relations. All these approaches discretize the correspondence variables at each iteration, hence disregarding the uncertainty in the correspondences during the matching process.

In the following section, we introduce some definitions and notation.

4.2 Definitions and Notation

We denote the attributed graphs (i.e., graphs) to be matched as the 3-tuples $\mathbf{G} = (\mathcal{U}, D, \mathcal{X})$ and $\mathbf{H} = (\mathcal{V}, M, \mathcal{Y})$.

The node-sets are denoted by $\mathcal{U} = \{u_a,\ a \in \mathcal{I}\}$ and $\mathcal{V} = \{v_\alpha,\ \alpha \in \mathcal{J}\}$, where $\mathcal{I} = 1, \ldots, |\mathcal{U}|$ and $\mathcal{J} = 1, \ldots, |\mathcal{V}|$ are the index-sets.

The adjacency matrices D and M account for the binary relations between the pairs of nodes in the graph. In the *weighted* case, adjacency matrices usually convey information about the similarities between pairs of nodes. In the *unweighted* case they convey information about the links between the nodes.

The advantage of the weighted case is that it allows for continuous adjacency relations in the cases that they are useful for the application. In some cases, the unweighted case can be seen as the discrete counterpart of the weighted case in which the nodes are linked by an edge following some criterion based on their similarity. Other types of relations, however, are $\{0, 1\}$-valued in nature such as those represented by the *region adjacency graphs*, the *k-nearest-neighbor graphs* or the graphs from *skeletal representations*. Either way, the unweighted case benefits from the computational advantages of dealing with sparse matrices, which usually leads to speeded up implementations.

We will consider adjacency matrices of the form

$$D_{ab} = \begin{cases} e\left(u_a, u_b\right) & \text{if } u_a \text{ and } u_b \text{ are linked by an edge} \\ 0 & \text{otherwise} \end{cases}$$

where $e\left(u_a, u_b\right) \in (0, 1]$ is the value of the strength of the edge (the same applies for $M_{\alpha\beta}$). We will consider the case of *undirected* graphs where adjacency matrices are symmetric.

The attribute-sets $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$ contain the column-vectors with the unary measurements on the nodes.

The assignment function $f : \mathcal{I} \to \{\mathcal{J} \cup \emptyset\}$ encodes a *one-to-one* assignment from the (indices of) nodes in $\mathbf{G}$ to the (indices of) nodes in $\mathbf{H}$; and a *many-to-one* assignment from the (indices of) nodes in $\mathbf{G}$ to *null*. The task of graph matching is to optimize some objective function over the set of assignments f.

We introduce the assignment matrix S which is the matrix analog of the assignment function. In the case of discrete assignments, the elements $s_{a\alpha} \in S,\ a \in \mathcal{I},\ \alpha \in \mathcal{J}$ of this matrix are defined as $s_{a\alpha} = \begin{cases} 1 & \text{if } f(a) = \alpha \\ 0 & \text{otherwise} \end{cases}$

Eventually, this matrix representation can be relaxed in order to allow for *ambiguous* assignments. In this case, $s_{a\alpha} \in [0,1]$ stands for the probability of assigning node u_a to node v_α. This matrix is subject to $\forall a \sum_{\alpha \in \mathcal{J}} s_{a\alpha} \leq 1$ being the remaining quantity up to 1 the probability of assigning node u_a to null.

It is common within the statistical estimation context to consider one of the graphs as the data-graph drawn from a distribution parameterized by the other graph which is considered as the model-graph. In most cases it is unimportant which one of the two graphs is considered as the model and which one as the data since both are supposed to belong to the same class of objects.

In the next few sections we review the continuous optimization approaches to graph matching which are most relevant to our thesis. We differentiate among three main approaches. Firstly, the structural graph matching approach that uses exclusively binary relations between pairs of nodes either in a weighted or unweighted form. This excludes the unary measurements $\mathcal{X}$ and $\mathcal{Y}$ from the graph representations. Secondly, the attributed graph matching approach that contemplates both binary and unary measurements. Finally, a special case of attributed graph matching that uses position coordinates as unary measurements on the nodes, namely, the structural graph matching and point-set registration approach. Position coordinates are special cases of unary attributes that deserve special attention since they incorporate the recovery of the alignment parameters (in addition to the correspondence ones) to the graph matching problem.

4.3 Structural Graph Matching

The use of structural relations is at the core of any graph matching method. Purely structural graph matching is the fundamental form of graph matching. It makes use exclusively of the binary relations among the nodes in order to find the matches. There are many continuous optimization approaches to structural graph matching in the literature. In the following we review probabilistic relaxation (Hummel & Zucker, 1983; Pelillo, 1997; Rosenfeld *et al.*, 1976), Graduated Assignment by Gold & Rangarajan (1996), Luo & Hancock (2001)'s structural graph matching approach with the EM algorithm, and the approach by Aguilar *et al.* (2009) to structural graph matching based on a graph transformation.

In the structural graph matching literature, graphs to be matched are represented by the 2-tuples $\mathbf{G} = (\mathcal{U}, D)$ and $\mathbf{H} = (\mathcal{V}, M)$ where $\mathcal{U}, \mathcal{V}$ are the node-sets and D, M the adjacency matrices conveying either the weighted or unweighted binary relations.

4.3.1 Probabilistic Relaxation

The primal motivation of probabilistic relaxation, as proposed by Rosenfeld *et al.* (1976), was to introduce ambiguity, in the form of probabilistic labelings, to the discrete labelings proposed by Waltz (1975). They refer to *labelings* instead of assignments (or correspondences) because they casted the problem as one of assigning *labels* to objects. This way, objects and labels can be seen as the nodes from a data and a model graph, respectively, at the same time that structural constraints are imposed by means of the *compatibility coefficients*. Hummel & Zucker (1983) noticed that in the case of symmetric compatibilities, local maxima of the following function correspond to consistent labelings.

$$\mathcal{F}_{pr} = \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{a\alpha} s_{b\beta} Q_{a\alpha b\beta} \tag{4.1}$$

where $s_{a\alpha} \in [0, 1]$, $\forall a \sum_{\alpha\in\mathcal{J}} s_{a\alpha} = 1$ and $Q_{a\alpha b\beta}$ stand for the compatibility of the simultaneous matches $u_a \to v_\alpha$ and $u_b \to v_\beta$.

In the case of positive compatibility coefficients $Q_{a\alpha b\beta}$, Pelillo (1997) showed that the iterative update of the object-label assignments according to the following equation, inevitably leaded to the consistent labelings of Hummel & Zucker (1983)

$$s_{a\alpha}^{(n+1)} = s_{a\alpha}^{(n)} \frac{\delta\mathcal{F}_{pr}}{\delta s_{a\alpha}} \Bigg/ \sum_{\alpha'\in\mathcal{J}} s_{a\alpha'}^{(n)} \frac{\delta\mathcal{F}_{pr}}{\delta s_{a\alpha'}} \tag{4.2}$$

where, in the case of symmetric compatibilities,

$$\frac{\delta\mathcal{F}_{pr}}{\delta s_{a\alpha}} = 2 \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta} Q_{a\alpha b\beta} \tag{4.3}$$

is usually referred to as the *support function* for the match $u_a \to v_\alpha$.

Specifically, Pelillo (1997) showed that, in the case of nonnegative symmetric compatibilities, the negative of (4.1) is a *strict Liapunov* function for the nonlinear operator defined in (4.2). This fact leads to the important result that strictly consistent labelings of Hummel & Zucker (1983) are local attractors for the update rule of (4.2). This means that, starting from an initial labeling $S^{(0)}$, iterative application of (4.2) lead to a consistent labeling in the vicinity of $S^{(0)}$.

4.3.2 Graduated Assignment

Gold & Rangarajan (1996) proposed a weighted graph matching method aimed at optimizing the type of objective functions used in the probabilistic relaxation approaches. Specifically they aim to minimize the following quantity,

$$\mathcal{F}_{ga} = -\frac{1}{2} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{a\alpha} s_{b\beta} Q_{a\alpha b\beta} \tag{4.4}$$

subject to

$$\forall a \sum_{\alpha\in\mathcal{J}} s_{a\alpha} \leq 1,\ \forall \alpha \sum_{a\in\mathcal{I}} s_{a\alpha} \leq 1,\ \forall a, \alpha\ s_{a\alpha} \in \{0, 1\}$$

where $Q_{a\alpha b\beta}$ is the similarity between the binary relations D_{ab} and $M_{\alpha\beta}$ (i.e., the compatibility coefficients) that conveys the compatibility of the edge-match $(u_a, u_b) \to (v_\alpha, v_\beta)$.

Typically, this consistency is of the form

$$Q_{a\alpha b\beta} = \begin{cases} 0 & \text{if either } D_{ab} = 0 \text{ or } M_{\alpha\beta} = 0 \\ sim\,(D_{ab}, M_{\alpha\beta}) & \text{otherwise} \end{cases} \tag{4.5}$$

where *sim* $(D_{ab}, M_{\alpha\beta})$ is some similarity measure between the edges.

Although it does not contemplate unary measurements, it can handle binary relations in the form of weighted adjacency matrices.

Similarly to the probabilistic relaxation approaches it maintains uncertainty in the form of a continuous assignment variable throughout the matching process. This is done by means of a continuation method, the *softmax* which, indexed by a control parameter, gradually pushes from continuous to discrete solutions, thus reducing the chances of getting trapped in local minima. The update of the assignment variable effected by softmax consists on the exponentiation of the assignment benefits multiplied by a control parameter. The higher the control parameter is, the higher the difference between the best assignment and the rest.

For the moment we will assume that inequality constraints of equation (4.4) are equality constraints. Such constraints enforce the discrete assignments matrix to be a permutation matrix, this is, one node in the first graph can only be assigned to one node in the second graph and vice-versa. They used a result due to Sinkhorn (1964) in order to enforce the continuous assignment matrix to be *doubly stochastic*, i.e., the continuous analog of a permutation matrix. This a result states that any square matrix with positive elements will converge to a doubly stochastic matrix just by the iterative process of alternatively normalizing the rows and columns. This is known as the *Sinkhorn* normalization.

The ensemble consisting of continuous maximization (i.e., softmax) and two-way constraints satisfaction (i.e., Sinkhorn normalization) is known as *Softassign*. See figure 3.2 in section 3.4 for an illustration.

They approximated the objective in (4.4) via Taylor series expansion about an initial condition $S^{(0)}$. This is,

$$-\frac{1}{2}\sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}}\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}} s_{a\alpha}s_{b\beta}Q_{a\alpha b\beta} \approx \\ -\frac{1}{2}\sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}}\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}} s^{(0)}_{a\alpha}s^{(0)}_{b\beta}Q_{a\alpha b\beta} - \sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}} B_{a\alpha}\left(s_{a\alpha} - s^{(0)}_{a\alpha}\right) \tag{4.6}$$

where

$$B_{a\alpha} = \left.\frac{\delta\mathcal{F}_{ga}}{\delta s_{a\alpha}}\right|_{S=S^{(0)}} = +\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}} s^{(0)}_{b\beta}Q_{a\alpha b\beta} \tag{4.7}$$

is the support for the match $u_a \to v_\alpha$ obtained from the neighboring nodes.

Then, minimizing the Taylor series expansion is equivalent to maximizing

$$\sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}} s_{a\alpha}B_{a\alpha} \tag{4.8}$$

which is an assignment problem where $B_{a\alpha}$ represents the benefit value.

The general procedure is the following.

1. Start with some valid initial value for $S^{(0)}$

2. Do a first order Taylor expansion, taking the partial derivative (i.e., compute matrix B).

3. Find the Softassign corresponding to the current assignment estimate.
4. Substitute the resulting S back to step 1, increase the control parameter μ and repeat.

This is illustrated in figure 4.1.

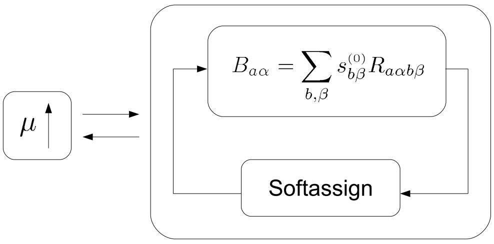

μ
$B_{a\alpha} = \sum_{b,\beta} s^{(0)}_{b\beta} R_{a\alpha b\beta}$
Softassign

Figure 4.1: Overview of the graduated assignment procedure. The procedure iterates until convergence of the correspondence matrix S or until μ has reached a predefined value.

At the beginning of the procedure, for low values of μ, all the assignments $s_{a\alpha}$ have roughly the same scale. As the value of μ increases, the differences between the consistent and inconsistent assignments also increase, while the assignment matrix S gradually tends to a permutation matrix after the Sinkhorn normalization.

So far, we have supposed equality constraints in equation (4.4). Inequality constraints are transformed into equality ones by introducing the *slack* variables. This is,

$$\forall a \sum_{\alpha \in \mathcal{J}} s_{a\alpha} \leq 1 \rightarrow \forall a \sum_{\alpha \in \{\mathcal{J} \cup \emptyset\}} s_{a\alpha} = 1 \tag{4.9}$$

and likewise for column constraints. An augmented matrix of assignments $\tilde{S}$ is created in order to hold the slack variables by adding an extra row and column to the matrix S. By incorporating the slack variables the graph matching algorithm can deal with graphs of different sizes at the same time that they handle outliers in a statistically robust way.

4.3.3 Structural Graph Matching Using the EM Algorithm and Singular Value Decomposition

Luo & Hancock (2001) posed the matching of graphs as a statistical estimation problem. They considered that a data-graph is generated from a model-graph through a noisy process following a Bernoulli distribution with a (low) probability of error P_e. This is, a data-graph is seen as a corrupted instance of a model-graph where the probability of an edge in the model graph being preserved in the data-graph is $1 - P_e$. Correspondences between nodes are the parameters of the Bernoulli distribution. They sought the correspondence parameters that lead to the maximum likelihood of the data-graph being drawn from the model-graph following a Bernoulli distribution.

They posed the problem as a ML estimation of the optimal correspondence parameters S^* in the presence of missing data where the corresponding model-graph nodes are the hidden variables. This is,

$$\begin{aligned} S^* &= \arg\max_S P\left(\mathbf{G}|S\right) \\ &= \arg\max_S \prod_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} P\left(u_a, v_\alpha|S\right) \end{aligned} \tag{4.10}$$

where u_a are the data-graph nodes and v_α are the corresponding model-graph nodes.

They showed how the complete-data likelihood in the right-hand side of equation (4.10) can be factorized into terms of individual correspondence indicators using the Bayes rule in the following way

$$P\left(u_a, v_\alpha|S\right) = K_a \prod_{b\in\mathcal{I}} \prod_{\beta\in\mathcal{J}} P\left(u_a, v_\alpha|s_{b\beta}\right) \tag{4.11}$$

where

$$K_a = \left[\frac{1}{P\left(u_a\right)}\right]^{|\mathcal{I}|\times|\mathcal{J}|-1}$$

is a quantity only depending on the identity of node u_a.

As said, they assumed that edges of the model-graph are preserved in the data-graph with a probability $1 - P_e$. In other words, given that nodes u_a, u_b correspond to nodes v_α, v_β there is edge-consistence with a probability $1 - P_e$ and edge-inconsistency with a probability P_e. This can be expressed in terms of the densities at the right-hand side of equation (4.11) by using a Bernoulli distribution in the following way.

$$\begin{aligned} P\left(u_a, v_\alpha|s_{b\beta}\right) &= \begin{cases} (1-P_e) & \text{if } D_{ab} = 1 \wedge M_{\alpha\beta} = 1 \wedge s_{b\beta} = 1 \\ P_e & \text{otherwise} \end{cases} \\ &= (1-P_e)^{D_{ab}M_{\alpha\beta}s_{b\beta}}\, P_e^{1-D_{ab}M_{\alpha\beta}s_{b\beta}} \end{aligned} \tag{4.12}$$

Hence, the final expression for the complete-data likelihood of equation (4.11) becomes

$$P\left(u_a, v_\alpha|S\right) = Z_a \exp\left[ln\left(\tfrac{1-P_e}{P_e}\right) \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} D_{ab}M_{\alpha\beta}s_{b\beta}\right] \tag{4.13}$$

where $Z_a = P_e^{|\mathcal{I}|\times|\mathcal{J}|}K_a$.

In practice, it is often more convenient to work with the logarithm of the likelihood function. From equations (4.10) and (4.13), we seek the correspondence indicators that satisfy

$$S^\star = \arg\max_S \sum_{a\in\mathcal{I}} ln\left\{\sum_{\alpha\in\mathcal{J}} Z_a \exp\left[ln\left(\tfrac{1-P_e}{P_e}\right)\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}} D_{ab}M_{\alpha\beta}s_{b\beta}\right]\right\} \tag{4.14}$$

which is intractable in closed form due to the mixture structure.

As stated in section 3.2, the EM algorithm addresses this type of problems by iteratively maximizing the expected complete-data log-likelihood conditioned by the observed data. By discarding the terms constant with respect to the correspondence parameters, the EM update rule according to this model becomes

$$S^{(n+1)} = \arg\max_{S} \sum_{a\in\mathcal{I}}\sum_{b\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}}\sum_{\beta\in\mathcal{J}} \omega_{a\alpha}^{(n)} D_{ab} M_{\alpha\beta} s_{b\beta} \tag{4.15}$$

where $\omega_{a\alpha}^{(n)}$ are the missing-data estimates according to the most recent correspondence parameters $S^{(n)}$.

Expectation

In the expectation step the posterior probabilities of the missing data are estimated using the most recent parameters available. This corresponds to the following expression

$$P\left(v_\alpha | u_a, S^{(n)}\right) = \frac{\exp\left[ln\left(\frac{1-P_e}{P_e}\right)\sum_{b,\beta} D_{ab} M_{\alpha\beta} s_{b\beta}^{(n)}\right]}{\sum_{\alpha'} \exp\left[ln\left(\frac{1-P_e}{P_e}\right)\sum_{b,\beta} D_{ab} M_{\alpha'\beta} s_{b\beta}^{(n)}\right]} \overset{def}{=} \omega_{a\alpha}^{(n)} \tag{4.16}$$

Maximization

Luo & Hancock (2001) noted that maximization of equation (4.15) could be equivalently stated in the following matrix form.

$$S^{(n+1)} = \arg\max_{S} Tr\left[D^\top W^{(n)} M S^\top\right] \tag{4.17}$$

where Tr denotes the trace of a matrix, D, M are the adjacency matrices of the data and model graph, respectively; and $W^{(n)}$ are the missing data estimates $\omega_{a\alpha}^{(n)}$ in matrix form.

This is a measure of correlation between the edge sets under the action of the weighted permutation of the posterior probabilities. They use the extremum principle by Scott & Longuet-Higgins (1991) that states that the orthogonal matrix $R^\top$ that maximizes $Tr\left[GR^\top\right]$ can be found by performing the singular value decomposition $G = \mathtt{U}\Lambda\mathtt{V}^\top$. By making the substitution $G = D^\top W^{(n)} M$, the optimal matrix $R^\star$ maximizing $Tr\left[GR^\top\right]$ is

$$R^\star = \mathtt{UEV}^\top \tag{4.18}$$

where $\mathtt{E}$ is the matrix obtained by making the diagonal elements of Λ unity.

Matrix R cannot be interpreted as a probability matrix since its elements are not guaranteed to be either normalized nor positive. It neither can be interpreted as an assignment variable since its elements are not binary. To obtain the discrete assignment variable they follow Scott & Longuet-Higgins (1991) and set the assignments for those elements $R_{a\alpha}$ which are the maximum value for their row and column. Therefore, the updated set of assignment indicators is obtained in the following way.

$$s_{a\alpha}^{(n+1)} = \begin{cases} 1 & \text{if } R_{a\alpha} = \max_{\beta} R_{a\beta} \wedge R_{a\alpha} = \max_{b} R_{b\alpha} \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

As stated in section 3.2, the EM algorithm alternates between the expectation and maximization steps. In the expectation step, the missing data estimates are computed that best correspond nodes in the data and model graph given the recent available parameters. In the maximization step the correspondence parameters that maximize the correlation between the adjacency matrices are computed by using the weighted permutation induced by the missing data estimates.

4.3.4 Outlier Rejection with Graph Transformation Matching

Aguilar *et al.* (2009) presented Graph Transformation Matching (GTM), a graph matching approach based on node-deletion operations. It is similar to RANSAC in the sense that correspondences can only be included or discarded but not modified like in the classical graph matching approaches. However, it is an appealing approach due to its reduced computational cost. It relies on unweighted binary relations local to the nodes instead of on a geometrical model assumption. Hence, it is able to accommodate to both rigid and non-rigid transformations as far as they preserve the local neighbourhood structures.

The algorithm starts with two graphs $\mathbf{G} = (\mathcal{U}, D)$ and $\mathbf{H} = (\mathcal{V}, M)$ with n nodes each, and a bijective mapping between them $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$. It proceeds by deleting nodes in both graphs in a greedy fashion until two isomorphic substructures are found.

Let denote as S the $n \times n$ *permutation matrix* such that

$$s_{a\alpha} = \begin{cases} 1 & \text{if } f(a) = \alpha \\ 0 & \text{otherwise} \end{cases}$$

The algorithm proceeds as follows.

1. Compute the residual matrix $R = |D - SMS^\top|$.

2. *if* $\sum\limits_{a=1}^{n} \sum\limits_{\alpha=1}^{n} R_{a\alpha} = 0$ *then* terminate and return f as the resulting isomorphism.

3. Select the column of R that yields the maximum disparity. This is, $j^{\max} = \arg\max\limits_{j=1,\ldots,n} \sum_{i=1}^{n} R(i,j)$.

4. Delete all references to nodes $u_{j^{\max}}$ and $v_{f(j^{\max})}$ from the node-sets $\mathcal{U}, \mathcal{V}$ as well as the adjacency matrices D, M. Update appropriately the correspondences function f and matrix S. Re-compute the structure of the graphs after the node deletions. Update $n \leftarrow n - 1$.

5. *if* the number of remaining nodes n is less than a predefined threshold, *then* exit without any result, *else* return to step 1.

Aguilar *et al.* (2009) suggested the use of *median K-nearest-neighbor* graphs with $k = 5$ as graph representations.

4.4 Attributed Graph Matching

As opposed to pure structural methods, attributed graph matching methods contemplate the unary measurements in the nodes in order to compute the matches. Attrubuted graph matching methods are often more application-dependant than structural ones. However, this approach is decisive when not all the relevant information to a problem can be meaningfully represented in the form of binary relations.

In the following we review the attributed graph matching method by Wilson & Hancock (1997).

4.4.1 Structural Matching by Discrete Relaxation

Wilson & Hancock (1997) abstracted the matching process in terms of attributed relational graphs. Their aim was to match the graphs denoted by the triples $\mathbf{G} = (\mathcal{U}, D, \mathcal{X})$ and $\mathbf{H} = (\mathcal{V}, M, \mathcal{Y})$, where D and M are *unweighted* adjacency matrices and $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$ convey the unary measurements of each node. Their aim was to find the optimum configuration of matches $f^\star$ with Maximum A Posteriori (MAP) probability with respect to the unary measurements available. This is,

$$f^\star = \arg\max_f P\left(f|\mathcal{X}, \mathcal{Y}\right) = \arg\max_f \frac{P\left(\mathcal{X}, \mathcal{Y}|f\right) P\left(f\right)}{P\left(\mathcal{X}, \mathcal{Y}\right)} \tag{4.20}$$

where $P(f)$ is the joint prior for the matching configuration and $P(\mathcal{X}, \mathcal{Y}|f)$ and $P(\mathcal{X}, \mathcal{Y})$ are, respectively, the conditional measurements density and the probability density function for the sets of unary measurements.

Assuming that the different pairs of unary measurements are conditionally independent of one another given the configuration of matches, the MAP rule writes

$$f^\star = \arg\max_f \prod_{a,\alpha|f(a)=\alpha} P\left(\mathbf{x}_a, \mathbf{y}_\alpha | f\left(a\right) = \alpha\right) P\left(f\right) \tag{4.21}$$

where $\mathbf{x}_a \in \mathcal{X}$ and $\mathbf{y}_\alpha \in \mathcal{Y}$ are the graphs' attributes, and the unconditional densities $P(\mathcal{X}, \mathcal{Y})$ have been discarded since they are a static property of the data.

This optimization is carried out by performing iterative local updates in the matching configuration of the form

$$f\left(a\right) = \arg\max_\alpha P\left(\mathbf{x}_a, \mathbf{y}_\alpha | f\left(a\right) = \alpha\right) P\left(f\right) \tag{4.22}$$

Starting from a tentative matching configuration $f^{(0)}$, the intuition is that global consistency can be attained by virtue of the net effect induced by the series of local updates.

The key issue addressed in that paper was the development of a probability model for the joint prior $P(f)$.

In measuring the probability of the matches from the data graph $\mathbf{G}$ they were interested in exploiting the structural constraints provided by the model graph $\mathbf{H}$. To that end, they devised a new structural subunit associated with each node, namely, the *clique*. A clique $\mathfrak{C}_a$ associated with a node u_a consists of an ordered string of indices from its adjacent nodes u_b (similarily for a clique $\mathfrak{R}_\alpha$ in the model-graph). This is,

$$\mathfrak{C}_a = \left(b_1, \ldots, b_p\right), \text{ s.t. } D_{ab_1} = 1, \ldots, D_{ab_p} = 1 \tag{4.23}$$

$$\mathfrak{R}_\alpha = \left(\beta_1, \ldots, \beta_q\right), \text{ s.t. } M_{\alpha\beta_1} = 1, \ldots, M_{\alpha\beta_q} = 1 \tag{4.24}$$

In the case of planar graphs the order is given by the relative orientations of the adjacent nodes with respect to the central one.

Note that the adjacency matrices have to be unweighted in order to build the clique descriptors.

The matching realization of a data-graph clique $\mathfrak{C}_a$ onto the model graph is denoted by the string

$$\Gamma_a = \left(f\left(b_1\right), \ldots, f\left(b_p\right)\right) \tag{4.25}$$

Structural constraints are imposed in the form of a dictionary of *structural-preserving mappings* (SPM) onto which each data-clique can be mapped. This dictionary encode the structural constraints present on the model-graph in the form of a set of feasible matching realizations. Robustness against rotations and outliers is conferred by generating a bunch of instances from each model-graph clique $\mathfrak{R}_\alpha$ obtained by applying circular shifts and dummy node insertions. This process is illustrated in figure 4.2. The union of the SPM of all the model-graph cliques constitutes the available dictionary Ω.

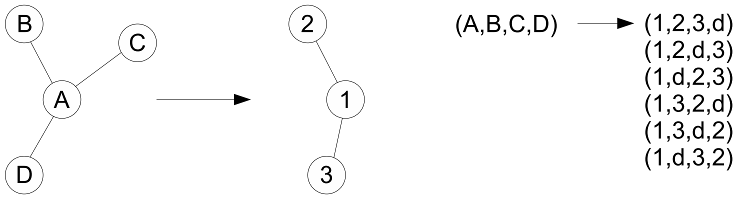


Figure 4.2: Shown in the left-hand side, there are the representations of a clique in the data-graph and a clique in the model-graph. Shown in the right-hand side, there are the string corresponding to the data-graph clique (starting with the symbol of the central node), and the bunch of SPM generated from the model-graph clique by circularly shifting and padding with dummy nodes so as to match the length of the string of the data-clique. Notice that the relative orientations of the adjacent nodes are preserved in the strings.

The idea underpinning their work is that a configuration of matches f is consistent as far as the data-graph cliques map to valid structure-preserving mappings in the model graph.

As we said, their major concern was the model for the joint prior. This is approximated as the average of the consistencies of each data-graph clique. This is,

$$P\left(f\right) = \frac{1}{|\mathcal{I}|}\sum_{a\in\mathcal{I}} P\left(\Gamma_a\right) \tag{4.26}$$

where the probability of a matching realization is expressed in terms of its marginal distribution. This is,

$$P\left(\Gamma_a\right) = \sum_{i=1}^{|\Omega|} P\left(\Gamma_a|\Omega_i\right) P\left(\Omega_i\right) \tag{4.27}$$

where Ω_i is the i-th entry in the dictionary of SPM, Ω.

The conditional probabilities in the right-hand side of equation (4.27) account for the consistence of the mapping Γ_a with regards to each individual

SPM in the dictionary. This consistence is gauged by independently comparing the corresponding symbols in the strings. This is,

$$P\left(\Gamma_a|\Omega_i\right)=\prod_{k=1}^{|\mathfrak{C}_a|}P\left(f\left(b_k\right)|\beta_k\right) \tag{4.28}$$

where b_k is the k-th symbol in the data-graph clique $\mathfrak{C}_a$ and β_k is the k-th symbol in the i-th entry of the dictionary, Ω_i. Note that strings $\mathfrak{C}_a$ and Ω_i are of the same size since possible length differences have been padded with dummy nodes.

Their discrete relaxation scheme was aimed at recovering the correct matching configuration in the presence of two types of perturbations. On one hand, errors in the initial matching configuration were anticipated to occur with a certain probability of error P_e. On the other hand, it was a major concern to recover the true underlying graph structure from point contamination. To that end they introduced the probability $P_\emptyset$ of a given node being an outlier and hence, having to be matched to a dummy node. At that point outlier detection is modeled as an assignment to the null node. In a later step this quantity takes part of a graph editing strategy aimed at actually removing outlier nodes.

With these ingredients, the probability of the match from a data-graph node given a feasible node in the model-graph was

$$P\left(f\left(b_k\right)|\beta_k\right)=\begin{cases} P_\emptyset & \text{if } f\left(b_k\right)=\text{dummy or } \beta_k=\text{dummy} \\ \left(1-P_\emptyset\right)\left(1-P_e\right) & \text{if } f\left(b_k\right)=\beta_k \\ \left(1-P_\emptyset\right)P_e & \text{if } f\left(b_k\right)\neq\beta_k \end{cases} \tag{4.29}$$

By substituting equation (4.29) into equation (4.28), the probability of the match of a data-clique given a SPM is

$$\begin{aligned} P\left(\Gamma_a|\Omega_i\right)=P_\emptyset^{\Psi(\Gamma_a,\Omega_i)} \\ \Big[\left(1-P_\emptyset\right)P_e\Big]^{H(\Gamma_a,\Omega_i)}\Big[\left(1-P_\emptyset\right)\left(1-P_e\right)\Big]^{|\mathfrak{C}_a|-H(\Gamma_a,\Omega_i)-\Psi(\Gamma_a,\Omega_i)} \end{aligned} \tag{4.30}$$

There are two quantities to be considered.

$\Psi\left(\Gamma_a,\Omega_i\right)$ accounts for the number of nodes in the data-graph clique matched to a dummy node plus the number of dummy nodes added to the SPM Ω_i in order to match the length of $\mathfrak{C}_a$. This quantity conveys information about both the structural corruption due to outliers in the data-graph clique and the intrinsic length differences between the cliques.

$H\left(\Gamma_a,\Omega_i\right)$ is the Hamming distance between the matching realization of the data-clique and the SPM. It conveys information about the consistence of the data-clique match. The lower the Hamming distance is, the more feasible the match is.

Assuming equiprobable priors for each model-clique $P\left(\Omega_i\right)=1/|\Omega|$, the final expression for the model of the matching joint prior of equation (4.26) expressed in the exponential form is

$$P\left(f\right)=\frac{1}{|\mathcal{I}|}\sum_{a\in\mathcal{I}}\frac{K_{\mathfrak{C}_a}}{|\Omega|}\sum_{i=1}^{|\Omega|}\exp\Big[-\Big(k_e H\left(\Gamma_a,\Omega_i\right)+k_\emptyset\Psi\left(\Gamma_a,\Omega_i\right)\Big)\Big] \tag{4.31}$$

where $k_e = ln\frac{(1-P_e)}{P_e}$, $k_\emptyset = ln\frac{(1-P_\emptyset)(1-P_e)}{P_\emptyset}$, and $K_{\mathfrak{C}_a} = \Big[(1-P_e)(1-P_\emptyset) \Big]^{|\mathfrak{C}_a|}$.

As we said, a major concern was to develop effective means for rejecting spurious measurements during the matching process. In the next subsection we review their outlier rejection mechanism.

A Principled Criterion for Outlier Rejection

Wilson & Hancock (1997) devised a principled criterion by which nodes were tested for inclusion or exclusion at each iteration of the discrete relaxation algorithm. Graph-edit operations involve a recomputation of the Delaunay triangulation conveying the structure of the graph. The idea is to gauge the contribution to the consistence measure of the affected cliques after a node deletion or reinsertion. Such a graph-edit operation is preserved if it leads to an improvement in the consistency measure. To that end, the nodes involved in the clique of a given node u_a are identified, i.e., $\mathfrak{C}_a - \{a\}$. On one hand, the set of cliques for all these nodes is denoted as χ_a^+. On the other hand, the set of cliques for all these nodes after deletion of node u_a and graph retriangulation is denoted as χ_a^-. Then, the change in the consistency functional $P(f)$ caused by the deletion of node u_a is proportional to

$$\Delta_a^- = P_\emptyset \sum_{b \in \chi_a^-} \frac{K_{\mathfrak{C}_b}}{|\Omega|} \sum_{i=1}^{|\Omega|} \exp\Big[-k_e H(\Gamma_b, \Omega_i) \Big] \tag{4.32}$$

Conversely, when considering a node reinsertion, it is to the clique-set χ_a^+ to which we turn our attention. The change in the MAP criterion caused by reinsertion of node u_a is proportional to

$$\Delta_a^+ = P(\mathbf{x}_a, \mathbf{y}_\alpha | f(a) = \alpha) \sum_{b \in \chi_a^+} \frac{K_{\mathfrak{C}_b}}{|\Omega|} \sum_{i=1}^{|\Omega|} \exp\Big[-k_e H(\Gamma_b, \Omega_i) \Big] \tag{4.33}$$

Following this criterion, if $\Delta_a^+ < \Delta_a^-$ then node u_a is decided to be an outlier (i.e., $f(a) = \emptyset$) and it is removed from the graph followed by the corresponding retriangulation. Conversely, if a previously deleted node u_a leaded to an improvement in the consistency (i.e., $\Delta_a^+ > \Delta_a^-$) then node is reinserted and the graph is retriangulated. Nodes are tested for deletion and reinsertion at each iteration.

4.5 Structural Graph Matching and Point-Set Registration

Here we review two graph matching methods that incorporate position coordinates as unary measurements in the nodes. They are casted within the statistical estimation framework. Position coordinates are a special case of attributes that deserve special attention since they require to include the estimation of the alignment parameters within the statistical apparatus. Therefore, the problem turns one of joint structural graph matching and point-set registration. We review two methods, namely, the dual-step method by Cross & Hancock (1998) and the unified approach by Luo & Hancock (2003).

In these approaches graphs are denoted by the 3-tuples $\mathbf{G} = (\mathcal{U}, D, \mathcal{X})$ and $\mathbf{H} = (\mathcal{V}, M, \mathcal{Y})$ where $\mathcal{U}, \mathcal{V}$ are the node-sets, D, M are either weighted or unweighted adjacency matrices, and $\mathcal{X}, \mathcal{Y}$ convey the information on the position coordinates of each node in the graphs.

4.5.1 Graph Matching with a Dual-Step EM Algorithm

Cross & Hancock (1998) presented an approach to perform graph matching and point-set registration using the EM algorithm. Recovery of the correspondence and alignment parameters is performed in dual maximization steps of an EM algorithm. They sought the optimal correspondence and alignment parameters $f^\star$ and $\Phi^\star$ that maximize the incomplete-data likelihood of an observed graph $\mathbf{G}$. By supposing independence among the observed-graph nodes, the estimation turns into a more tractable form by introducing the corresponding model-graph nodes as hidden variables. This is,

$$\begin{aligned} \{f^\star, \Phi^\star\} &= \arg\max_{f,\Phi} P\left(\mathbf{G}|f, \Phi\right) \\ &= \arg\max_{f,\Phi} \prod_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} P\left(u_a, v_\alpha | f, \Phi\right) \end{aligned} \tag{4.34}$$

where u_a are the data-graph nodes and v_α are the corresponding model-graph nodes.

They devised a model in which the putative match $f^{(n+1)}(a) = \alpha$ is evaluated under the assumption of the matching realization of a data-graph clique centered at node u_a. At a given iteration n, this is effected by the expression

$$P\left(f^{(n+1)}(a) = \alpha | \Gamma_a^{(n)}\right) = \frac{P\left(f^{(n+1)}(a) = \alpha, \Gamma_a^{(n)}\right)}{\sum\limits_{\alpha' \in \mathcal{J}} P\left(f^{(n+1)}(a) = \alpha', \Gamma_a^{(n)}\right)} \tag{4.35}$$

where $\Gamma_a^{(n)}$ is the matching realization of clique centered at node u_a of equations (4.23) and (4.25), at iteration n.

To simplify the development, they used the notation

$$\Gamma_{a\alpha} = \{f^{(n+1)}(a) = \alpha, f^{(n)}(b), \forall b \in \mathfrak{C}_a - \{a\}\}$$

to represent the configuration of matched nodes on the clique $\mathfrak{C}_a$ with the putative update $f^{(n+1)}(a) = \alpha$ at the center node.

Similarly as in the discrete relaxation approach in section 4.4.1, they evaluated the matching realization $\Gamma_{a\alpha}$ by means of a dictionary of structural-preserving mappings (SPM). Using the Bayes rule, they expanded the probability of the matching realization $P(\Gamma_{a\alpha})$ over the set of SPM generated by the model-graph clique $\mathfrak{R}_\alpha$ centered at node v_α. This is,

$$P\left(\Gamma_{a\alpha}\right) = \sum_{i=1}^{|\Omega^\alpha|} P\left(\Gamma_{a\alpha}|\Omega_i^\alpha\right) P\left(\Omega_i^\alpha\right) \tag{4.36}$$

where Ω_i^α is the i-th entry of the dictionary of SPM Ω^α generated from model-clique $\mathfrak{R}_\alpha$ centered at node v_α (see figure 4.2).

The conditional probabilities in the right-hand side of equation (4.36) assess the consistence of the matching realization $\Gamma_{a\alpha}$ given each individual entry in the dictionary, Ω_i^α. As in the discrete relaxation approach of section 4.4.1, this probability measure accounts for different sources of errors independently at each symbol of the mapping $\Gamma_{a\alpha}$. Drawing on the model of errors reported in equations (4.28) and (4.29), the measure of consistency of a match introduced in equation (4.35) has the following expression.

$$P\left(f^{(n+1)}\left(a\right)=\alpha|\Gamma_a^{(n)}\right)=$$
$$\frac{\frac{K_{\mathfrak{C}_a}}{|\Omega^\alpha|}\sum_{i=1}^{|\Omega^\alpha|}\exp\Big[-\Big(k_e H\left(\Gamma_{a\alpha},\Omega_i^\alpha\right)+k_\emptyset\Psi\left(\Gamma_{a\alpha},\Omega_i^\alpha\right)\Big)\Big]}{\sum_{\alpha'\in\mathcal{J}}\frac{K_{\mathfrak{C}_a}}{|\Omega^{\alpha'}|}\sum_{i=1}^{|\Omega^{\alpha'}|}\exp\Big[-\Big(k_e H\left(\Gamma_{a\alpha'},\Omega_i^{\alpha'}\right)+k_\emptyset\Psi\left(\Gamma_{a\alpha'},\Omega_i^{\alpha'}\right)\Big)\Big]}\overset{def}{=}\zeta_{a\alpha}\qquad(4.37)$$

where equiprobable priors $P\left(\Omega_i^\alpha\right)=1/|\Omega^\alpha|$ have been supposed.

The key modeling ingredient in their model was to exploit as exponential indicators the consistency of a match of equation (4.37) in developing a measurement density for the correspondence matches. They considered that it was a measurement density on the point position errors $P\left(\mathbf{x}_a,\mathbf{y}_\alpha|\Phi\right)$ which was appropriate for gauging the probability of a match in the case of structural consistency. Otherwise, they assigned a uniform measurement density ρ which was independent of the position coordinates.

With these ingredients, the expression for the observed-data likelihood of equation (4.34) to be maximized is

$$P\left(\mathbf{G}|f,\Phi\right)=\prod_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}}P\left(\mathbf{x}_a,\mathbf{y}_\alpha|\Phi\right)^{\zeta_{a\alpha}}\rho^{1-\zeta_{a\alpha}}\qquad(4.38)$$

where

$$P\left(\mathbf{x}_a,\mathbf{y}_\alpha|\Phi\right)=\frac{1}{|2\pi\Sigma|^{1/2}}\exp\left[-\tfrac{1}{2}\left\|\mathbf{x}_a-\mathcal{T}\left(\mathbf{y}_\alpha;\Phi\right)\right\|_\Sigma^2\right]\qquad(4.39)$$

is a Gaussian measurement of the point position errors between the data-point $\mathbf{x}_a$ and the transformed model point $\mathcal{T}\left(\mathbf{y}_\alpha;\Phi\right)$ according to transformation parameters Φ; and $\|\mathbf{x}\|_\Sigma^2=\mathbf{x}^\top\Sigma^{-1}\mathbf{x}$ is the squared Mahalanobis distance with covariance matrix Σ.

Cross & Hancock (1998) sought the alignment and correspondence parameters in dual maximization steps. While alignment parameters are recovered in a Maximum Likelihood (ML) way, correspondence parameters are updated following a Maximum A Posteriori (MAP) criterion. According to the well-known development of the EM algorithm (see section 3.2), ML alignment parameters of equation (4.38) are computed by iterative maximization of the expected complete-data log-likelihood conditioned by the observed data. By discarding the constant terms not depending on Φ, this leads to the following expression

$$\Phi^{(n+1)}=\arg\min_{\Phi}\sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}}\omega_{a\alpha}^{(n)}\zeta_{a\alpha}^{(n)}\left\|\mathbf{x}_a-\mathcal{T}\left(\mathbf{y}_\alpha;\Phi\right)\right\|_{\Sigma^{(n)}}^2\qquad(4.40)$$

where $\omega_{a\alpha}^{(n)}$ are the missing-data estimates computed in the expectation step, $\zeta_{a\alpha}^{(n)}$ are the match consistency estimates of equation (4.37) using the current correspondence parameters $f^{(n)}$, and

$$\Sigma^{(n)}=\frac{\sum_a\sum_\alpha\omega_{a\alpha}^{(n)}\zeta_{a\alpha}^{(n)}\left(\mathbf{x}_a-\mathcal{T}\left(\mathbf{y}_\alpha;\Phi^{(n+1)}\right)\right)\left(\mathbf{x}_a-\mathcal{T}\left(\mathbf{y}_\alpha;\Phi^{(n+1)}\right)\right)^\top}{\sum_a\sum_\alpha\omega_{a\alpha}^{(n)}}\qquad(4.41)$$

is the expected value for the covariance matrix.

Expectation

According to their development, the missing-data estimates are updated by substituting the revised alignment parameters into the Gaussian density of equation (4.39). Using the Bayes rule, they expressed the posterior measurement probabilities in terms of the conditional densities in the following way

$$\omega_{a\alpha}^{(n)} = \frac{\exp\left[-\frac{1}{2}\left\|\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi^{(n)}\right)\right\|_{\Sigma^{(n)}}^2\right]}{\sum_{\alpha'} \exp\left[-\frac{1}{2}\left\|\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_{\alpha'}; \Phi^{(n)}\right)\right\|_{\Sigma^{(n)}}^2\right]} \tag{4.42}$$

Maximization

As pointed out earlier, the maximization step is based on dual update processes. The first of these aimed to locate MAP probability correspondence matches. The second update operation is concerned with locating ML spatial transformation parameters.

MAP Correspondence Parameters The update formula for the correspondence parameters is

$$f^{(n+1)}(a) = \arg\max_{\alpha \in \mathcal{J}} \omega_{a\alpha}^{(n)} \zeta_{a\alpha}^{(n)} \tag{4.43}$$

Once the update is applied, each node u_a is checked for deletion and insertion in order to overcome structural corruption in the same way as explained in section 4.4.1. The graph-edit operation is preserved if it leads to an improvement in the consistency measure. Let denote as χ_a^+ and χ_a^- the set of cliques containing the node u_a before and after the graph-edit operation preceding the structure re-computation. Then, the change in the MAP criterion caused by the deletion of node u_a is proportional to

$$\Delta_a^- = P_\emptyset \sum_{b \in \chi_a^-} \frac{K_{\mathfrak{C}_b}}{|\Omega^\alpha|} \sum_{i=1}^{|\Omega^\alpha|} \exp\Big[- k_e H\left(\Gamma_b, \Omega_i^\alpha\right)\Big] \tag{4.44}$$

Conversely, the change in the MAP criterion caused by reinsertion of node u_a is proportional to

$$\Delta_a^+ = \omega_{a\alpha}^{(n)} \sum_{b \in \chi_a^+} \frac{K_{\mathfrak{C}_b}}{|\Omega^\alpha|} \sum_{i=1}^{|\Omega^\alpha|} \exp\Big[- k_e H\left(\Gamma_b, \Omega_i^\alpha\right)\Big] \tag{4.45}$$

Following this criterion, node u_a is removed from the graph provided that $\Delta_a^+ < \Delta_a^-$. Conversely, a previously deleted node u_a is reinserted provided that $\Delta_a^+ > \Delta_a^-$. Nodes are tested for deletion and reinsertion at each iteration.

ML Spatial Transformation Parameters Cross & Hancock (1998) offered solutions to the optimization problem of equation (4.40) for two types of spatial transformations, namely, affinities and projectivities. Solutions for these types of transformations are provided in sections 3.5.2 and 3.5.3, respectively. Note that, in the problem of equation (4.40), the weights are given by the product $\omega_{a\alpha}^{(n)} \zeta_{a\alpha}^{(n)}$.

4.5.2 Unified Framework for Alignment and Correspondence

Luo & Hancock (2003) characterized the alignment and correspondence problems in terms of separate distributions. On one hand, alignment parameters Φ are recovered according to a Gaussian assumption on the alignment errors between the data points $\forall a, \mathbf{x}_a \in \mathcal{X}$ and model points $\forall \alpha, \mathbf{y}_\alpha \in \mathcal{Y}$, i.e., $P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi\right)$. On the other hand, the structure of the graphs is exploited in order to recover the correspondences S between the data-graph nodes $\forall a, u_a \in \mathcal{U}$ and the model-graph nodes $\forall \alpha, v_\alpha \in \mathcal{V}$, by following a Bernoulli distribution on the structural errors, $P\left(u_a, v_\alpha | S\right)$.

Let denote the posterior probabilities of the alignment and correspondence distributions as $\omega_{a\alpha}^{(\Phi)} \overset{def}{=} P\left(\mathbf{y}_\alpha | \mathbf{x}_a, \Phi\right)$ and $\omega_{a\alpha}^{(S)} \overset{def}{=} P\left(v_\alpha | u_a, S\right)$, respectively.

They devised a process in which the two distributions interact via a cross-entropy measure. Specifically, they sought the alignment and correspondence parameters, $\Phi^\star$ and $S^\star$ that maximize the following quantity

$$\left\{\Phi^\star, S^\star\right\} = \arg\max_{\Phi, S} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(S)} P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi\right) + \omega_{a\alpha}^{(\Phi)} P\left(u_a, v_\alpha | S\right) \tag{4.46}$$

This is, correspondence probabilities weight contributions to the log-likelihood function for the alignment errors, and vice-versa.

Alignment errors are assumed to follow a Gaussian distribution. This is,

$$P\left(\mathbf{x}_a, \mathbf{y}_\alpha | \Phi\right) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left[-\tfrac{1}{2}\left\|\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi\right)\right\|_\Sigma^2\right] \tag{4.47}$$

where $\mathcal{T}\left(\mathbf{y}_\alpha; \Phi\right)$ is the transformed model point $\mathbf{y}_\alpha$ according to transformation parameters Φ; and $\|\mathbf{x}\|_\Sigma^2 = \mathbf{x}^\top \Sigma^{-1} \mathbf{x}$ is the squared Mahalanobis distance with covariance matrix Σ.

Correspondences are estimated according to a Bernoulli assumption on the errors in the structure of the graphs, as reported in equation (4.13) in section 4.3.3. This is,

$$P\left(u_a, v_\alpha | S\right) = Z_a \exp\left[ln\left(\tfrac{1-P_e}{P_e}\right) \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} D_{ab} M_{\alpha\beta} s_{b\beta}\right] \tag{4.48}$$

where P_e is the probability of error of the Bernoulli distribution, D, M are the adjacency matrices for the data and model graph, respectively, and Z_a is a constant quantity only depending on the identity of the data-graph node u_a.

The utility measure of equation (4.46) leads to a decoupled process of alignment and correspondence parameters update.

From equations (4.46) and (4.47), alignment parameters are updated according to the following expression.

$$\Phi^{(n+1)} = \arg\min_{\Phi} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(S)} \left\|\mathbf{x}_a - \mathcal{T}\left(\mathbf{y}_\alpha; \Phi\right)\right\|_\Sigma^2 \tag{4.49}$$

Similarity transformation parameters are sought from equation (4.49) as reported in section 3.5.1.

From equations (4.46) and (4.48), correspondence parameters are updated according to

$$S^{(n+1)} = \arg\max_{S} \sum_{a\in\mathcal{I}} \sum_{b\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \sum_{\beta\in\mathcal{J}} \omega_{a\alpha}^{(\Phi)} D_{ab} M_{\alpha\beta} s_{b\beta} \tag{4.50}$$

Correspondence parameters are recovered following the extremum principle by Scott & Longuet-Higgins (1991) reported in section 4.3.3, equations (4.17), (4.18) and (4.19).

4.6 Graph Matching with Genetic Algorithms

Genetic Algorithms (GA) (Goldberg, 1989) are another optimization technique that has been successfully applied to inexact graph matching. Search is performed by means of stochastic genetic operators, thus providing effective means of locating global optimal solutions.

The procedure starts from an initial population. Each individual, called a *chromosome*, encodes a possible solution to the graph matching problem. The cost function that we want to optimize sets the *fitness* value of each chromosome. New solutions are generated by applying *crossover* and *mutation* operators to the individuals in the population. In hybrid GA, a further gradient ascent stage may be introduced at some point. The next generation is decided by means of a *selection* operator. This process is repeated until convergence or a predefined number of times.

Cross (1997) proposed a hybrid GA combining classical genetic operators with a hill-climbing procedure. They sought the correspondences that maximize equation (4.31) (section 4.4.1). In a later paper Cross *et al.* (2000) provided theoretical justification of its convergence.

Wang *et al.* (1997) implemented a GA seeking the permutation that best correlates the adjacency matrices of two graphs.

In the following we briefly describe the different aspects involved in GA.

Encoding Chromosomes are preferably encoded as strings for convenience of the application of genetic operators. Each chromosome constitutes a possible solution to the graph matching problem. In our case, a match $f : \mathcal{I} \to \mathcal{J}$ from a data-graph $\mathbf{G}$ to a model-graph $\mathbf{H}$ is represented by a string of length $|\mathcal{I}|$ of labels drawn from $\mathcal{J}$. Figure 4.3 shows an example.

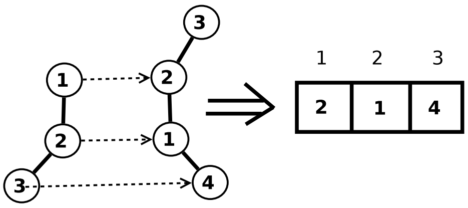


Figure 4.3: A possible solution in the population encoded in the form of a chromosome.

Initialization There is a number of ways of generating the initial population. Perhaps the simplest option is to initialize at random.

Cross (1997) initialized the population so as to uniformly sample the search space. They argued that this way the algorithm is less prone to local optima.

Wang *et al.* (1997) utilized an initialization suggested by the unary measurements on the nodes. They argued that good initial candidates may improve the convergence towards the global optimum.

The size of the population determines the rate of convergence of the GA. The higher the population, the higher the chances of finding the optimal solution at the same time that the computational overhead increases.

Crossover The idea of the crossover operator is to be able to combine partially correct solutions from different individuals in the population in order to generate globally consistent solutions. This operator takes a pair of individuals from the population and interchanges a corresponding portion of their encodings with each other. Figure 4.4 shows an example of this concept.

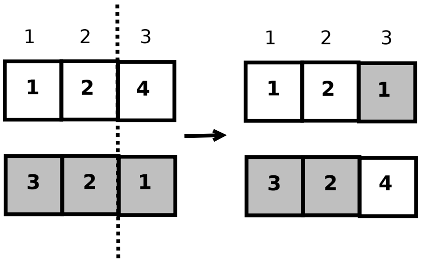


Figure 4.4: Crossover between two individuals at a random point.

Wang *et al.* (1997) randomly chose the initial and ending points of the portions to be interchanged.

Cross (1997) proposed a more meaningful approach in which instead of selecting a contiguous portion from an arbitrarily ordered encoding, they selected a geometrically contiguous portion of the graphs.

Mutation Mutation consists on randomly reassigning the labels at individual sites with a uniform probability. This must be done carefully since a GA with a high mutation probability may degenerate to a random search. Reassignment is done from random labels in $\mathcal{J}$. Figure 4.5 illustrates this concept.

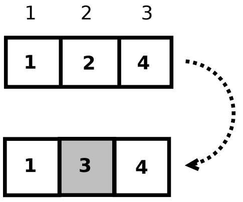


Figure 4.5: Mutation of one of the labels in the chromosome.

Wang *et al.* (1997) applied a low-rated mutation to the best-fitted individuals only. They argued that this strategy provides a compromise between hill-climbing and classical GAs since only solutions in the vicinity of the best fitted individual are explored.

Gradient Ascent Mutation and crossover operators may leave the matches into inconsistent states. In hybrid GA, it is possible to apply a deterministic update procedure to each individual after the crossover and mutation operators in order to remove possible inconsistencies as well as to push them towards their closest local optimum before the selection operation. See figures 4.6 and 4.7 for an illustration.

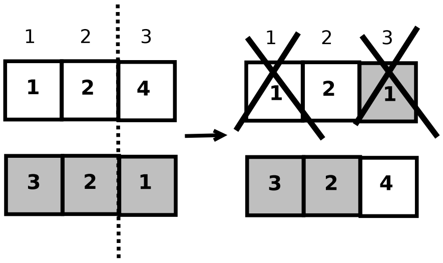


Figure 4.6: Two different data-graph nodes assigned to the same model-graph node due to a crossover operation.

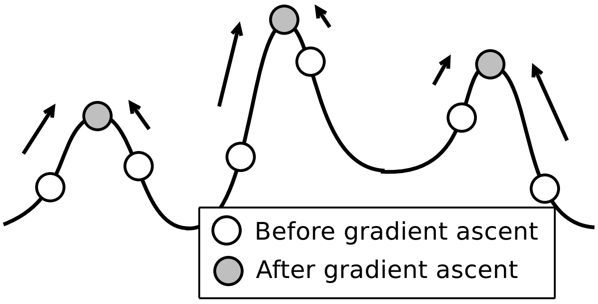


Figure 4.7: Solid curve represents the values of the function along the domain. Empty circles represent the values of the individuals of the population. Shaded circles represent the values of the individuals after the gradient ascent step.

Cross (1997) performed several discrete relaxation iterations to each individual in the population before the selection operator at each iteration of the GA.

Selection The individuals that will constitute the next generation are decided on the basis of the selection operator. It is commonly adopted a roulette wheel approach in which the probability of selecting a given individual is proportional to its fitness value. Note that using this approach, there may be several copies of some individuals while there may be no copies of some others in the resulting population. A too strict selection towards the fittest individuals may narrow the search towards local optima, being unable to find the global optimum.

4.7 Spectral Graph Matching

A relatively new form of graph theory, known as spectral graph theory (Chung, 1997), exploits the spectral properties of the matrices associated with the graph, such as its adjacency matrix or Laplacian matrix. The eigenvalues and eigenvectors of the graph matrices have the useful properties of retaining the global structure of the graph while being invariant to node permutations. These properties have been used for graph matching, graph characterization and graph embedding, among others. One outstanding feature of spectral graph theory is the elegance of relying entirely on algebraic operations on the graph matrices.

One of the first attempts to spectral graph matching was by Umeyama (1988) who presented an eigendecomposition approach to match two weighted graphs. Among the main limitations of this early approach were that the graphs had to be the same size and it was not very robust to structural corruption.

Although not directly aimed at graph matching, Scott & Longuet-Higgins (1991) presented a method for deciding the correspondences from a benefit matrix using its eigendecomposition.

More in the aims of graph matching, Shapiro & Brady (1992) proposed an extension of the work by Scott & Longuet-Higgins (1991) that exploited the eigenvectors of the intra-point distance matrix. Correspondences were decided on the basis of the implicit embedding performed by considering the rows of the eigenvectors matrix as feature vectors. This approach demonstrated to retain the shape information of the point-sets much better than the one reported by Scott & Longuet-Higgins (1991).

Luo & Hancock (2003, 2001) benefited from the work by Scott & Longuet-Higgins (1991) in order to decide the correspondences within iterative graph matching approaches. They were able to match graphs with different sizes and to accommodate structural corruptions.

From a different point of view, Silletti *et al.* (2011) have recently presented a hybrid, non iterative method for point-set matching. They decide the matches from a benefit matrix by using the the extremum principle reported by Scott & Longuet-Higgins (1991). This benefit matrix is built upon a combination of different possible metrics that include measures exploiting the position coordinates of points, intra-point distance matrices (Shapiro & Brady, 1992), structural information, and raw image information.

Another approach consists on extracting a serialized representation of the graph by exploiting the properties of the leading eigenvector of the graph's matrices.

Using the theory of Markov chains, Robles-Kelly & Hancock (2002, 2003, 2005) used the leading eigenvector of the transition probability matrix (i.e., the normalized adjacency matrix) to convert graphs into strings. This is based on the observation that the leading eigenvector corresponds to the steady state random walk on the graph. Graph matching was then reduced to a string-edit distance computation on the seriated graphs.

Along the same lines, Yu & Hancock (2005) presented an approximation to the problem of graph seriation using semidefinite programming.

Another approach consists on perform graph matching by taking advantage of the spectral properties of the association graph or similars.

Wang & Hancock (2006, 2008) presented an alternative formulation for the probabilistic relaxation processes in terms of diffusion processes on an association graph (Barrow & Burstall, 1976). A diffusion process can be regarded as a continuous-time random walk on the state-space defined by the nodes of the graph. Similarly to probabilistic relaxation, diffusion processes propagate confidence in the object labeling globally via local computation. This approach leaded to an iterative procedure of recomputation of the association graph using the newly estimated state-probability vectors.

Emms *et al.* (2009) analyzed the pattern of interferences of a quantum walk on a specially constructed auxiliary graph in order to perform graph matching. Quantum walks are characterized by the eigendecomposition of the graph's Laplacian and have the advantages of avoiding the problem of cospectrality sometimes found with classical random walk approaches. Tentative matches were computed with the Hungarian method and refined with a Maximum Clique procedure (Pelillo, 1996).

Other approaches attach unary measurements on the nodes that characterize the diffusion processes originating on that nodes in order to improve the matching performance.

Lozano & Escolano (2004) used the diffusion processes in order to address the ambiguity sometimes found when using pure structural graph matching methods.

Later on, Lozano & Escolano (2009) extended their previous work by experimenting with regularization kernels. Matching performance was assessed with an adaptation of the Graduated Assignment algorithm aimed at handling unary measurements (Lozano & Escolano, 2004, 2009), as well as an analogous adaptation of the Motzkin-Strauss algorithm (Lozano & Escolano, 2009).

Along the same lines, Lozano & Escolano (2005) proposed a futher improvement on their unary measurements based on local entropic graphs. Following the idea that diffusion kernels are the discrete equivalents of the Gaussian kernels, they approximated the distance between nodes on a locally-Euclidean manifold on the basis of their kernel values. By considering the nodes as realizations of a probability density, they computed their Rényi entropy. This characterization, which is built upon the minimum spanning tree (MST), demonstrated to be more stable against structural corruption than the previously reported ones.

Another approach is to pose the graph matching problem as a point-set registration problem using the embedded node coordinates.

Escolano *et al.* (2011) have recently posed the graph matching problem in terms of non-rigid manifold alignment. They use a metric based on commute times in order to embed the graphs in a low-dimensional manifold. Non-rigid registration of the embedded node coordinates is performed with the EM algorithm and Coherent Point Drift (Myronenko & Song, 2010). Additionally, they propose an information theoretical-based measure in order to gauge the similarity of the two registered graphs.

4.8 Graph Matching Applications

Graph matching methods have been used in a variety of applications. In the following we present some of them.

4.8.1 Local Image Features Matching

Next we present some relevant approaches, to the best of our knowledge, aimed at exploiting some kind of structural relation in order to match local image features.

Shin & Tjahjadi (2010) propose a new multi-descriptor composed by adjacent local features in each clique of a graph. Local feature-descriptors are based on SIFT and neighboring relations are estimated by means of Delaunay triangulations. Clique-based descriptors are matched on the basis of their Hausdorff distance.

Shokoufandeh *et al.* (1998) presented an approach to match hierarchical representations from images. They extracted the salient regions of an image using the wavelet transform and built a *Saliency Map Graph* (SMG), a type of Directed Acyclic Graph with a similar aim than the transitive closure of a tree (Torsello & Hancock, 2003). The SMG captures the containment relations between the salient regions. Hence, a directed edge between two nodes represents that the salient region of the destiny node is contained in the salient region of the origin node. They proposed two different algorithms for matching SMGs, one based on the structural relations and the other accounting with geometrical information.

Todorovic & Ahuja (2008) have recently presented an approach to find corresponding regions between two images with maximum area. They segment the images into regions at different segmentation levels, thus obtaining a natural tree representation encoding the containment relations between the regions across the different levels. They apply the divide-and-conquer strategy reported by Torsello & Hancock (2003) in order to find the maximum common subtree by solving the Maximum Weighted Clique Problem (MWCP) at the different levels of the trees. They devise a series of region similarity measures based on the photometric properties of the regions in order to assign the weights for each match.

4.8.2 Shape Matching

Graphs have been widely used to match shape-representations extracted from binary images. In these cases, graphs are commonly built from the Blum's medial axis or *skeleton* (Blum, 1967; Goh, 2008).

Some approaches to Chinese character recognition represent the strokes in the nodes Chan & Cheung (1992); Suganthan & Yan (1998). However, it is more usual to represent the skeletal end-and-intersection-points in the nodes, and their links in the edges.

Siddiqi *et al.* (1998) distinguished among four types of singularities, called *shocks*, during the skeleton formation process. They introduced *shock-graphs*, a directed attributed graph that retains the shock type along with its formation time. They used a *shock graph grammar* in order to characterize the space of shock-graphs and a tree-matching algorithm to find the correspondences.

Sebastian *et al.* (2004) applied the idea of graph-edit distance (Bunke, 1999; Sanfeliu & Fu, 1983) to the matching of shock-graphs.

Di Ruberto (2004) introduced *Attributed Skeletal Graphs*, another type of attributed graphs closely related to the medial axis representation.

Bai & Latecki (2008) presented a method for matching skeletal graphs based on the similarity of the skeletal paths among the end nodes.

Despite their effectiveness, the applicability of all these methods is restricted to the matching of skeleton-like structures.

Part II

Contributions

Chapter 5

A Fast Approximation of the Earth-Mover's Distance between Multi-Dimensional Histograms

5.1 Introduction

In computer vision, histograms represent the frequency of global or local values in an image. Hence, we can compute the distance of two distributions of such values by comparing their histograms. Comparison of histograms is of interest in a variety of fields such as image retrieval or indexing, as well as local image features matching.

Due to the need of comparing histograms, a number of measures of similarity between histograms have been proposed and used in related applications.

Most of the distance measures presented in the literature only consider the intersection between two histograms as a function of the distance value and they do not take into account the distance between the bins of the histogram. This is an important drawback in sparse histograms, since the distance value between histograms does not reflect the similarity between images.

Figure 5.1 shows images from objects 78, 98 and 88 from the COIL database together with their hue channels and their histograms.

The first two images are green and the last one is blue. Concerning the colour of the images, a human would say that the two first images are the most similar since both images are green; therefore, the distance between their histograms would have to be smaller. Nevertheless, the intersection of the three histograms becomes almost empty and so, if we do not take into account the distance between bins, the distance between the three histograms is similar and almost zero.

We call ground distance as the distance between the elements of the set that the histogram represents. Some of the distances and algorithms presented in the literature need the ground distance to be the L_1 distance. In some applications, this distance is not useful, and other distances are needed, such as

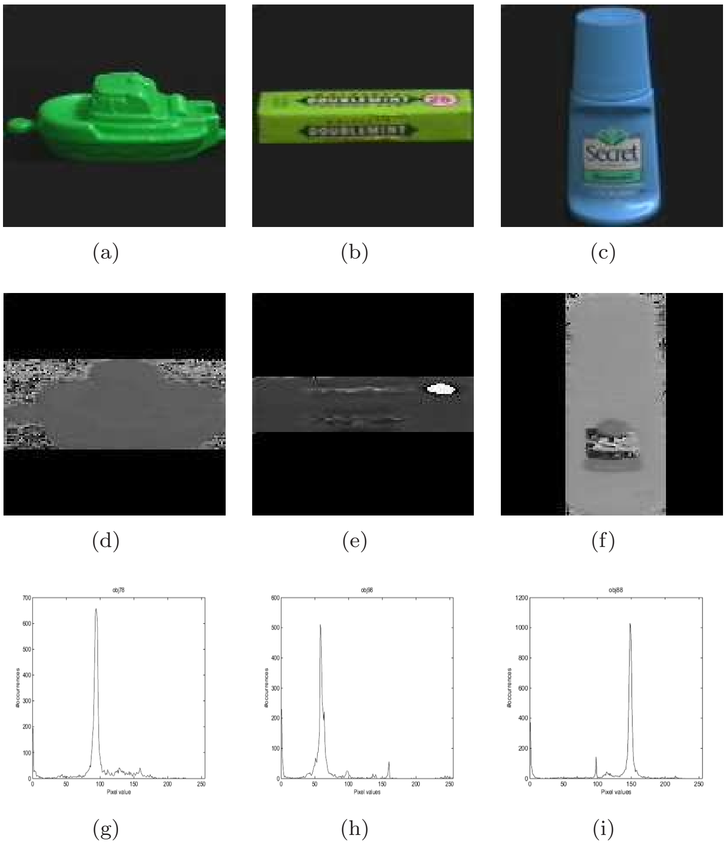


Figure 5.1: (a), (b), (c) Objects 78, 98 and 88 from the COIL database. (d), (e), (f) Their hue channels and (g), (h), (i) their histogram of the hue channels.

the L_2 distance.

Although allowing for the distance between the bins is more accurate when comparing histograms, it is also more computational demanding than using bin-to-bin distances. These requirements are even higher when dealing with histograms with multiple dimensions.

In this chapter, we define a distance between nD-histograms that is able to use any kind of ground distance. It is inspired in the well-known Earth Mover's Distance (EMD) but with a specific *flow* function between bins. With the properties extracted from this new definition, we define an algorithm that computes a sub-optimal solution of EMD with a worst-case complexity of $\mathcal{O}\left(p^2\right)$, being p the total number of bins. Nevertheless, the experiments show a real average cost near to $\mathcal{O}\left(m^2\right)$, being m the number of bins per dimension. Part of the theory and experiments were presented in (Serratosa & Sanromà, 2006; Serratosa *et al.*, 2007).

The structure of this chapter is as follows. In the next section, we discuss the related work. In section 5.3, we define the sets and the histograms. In sections 5.4 and 5.5, we define the distances between histograms and sets and we show the algorithm that computes this distance. In sections 5.6, 5.7 and 5.8

we show, respectively, an empirical study of the time complexity, some image retrieval experiments and some registration experiments using local features. Conclusions are presented in section 5.9.

Appendices A and B are added with the demonstration that the proposed distance between histograms has the same value than the distance between sets and other theoretical properties.

5.2 Related Work

Local descriptors based on histograms are usual tools in many computer vision tasks and pattern recognition. For comparing these descriptors, a distance between histograms and an algorithm to compute it are needed. In this section, we first describe the most relevant distances and algorithms presented elsewhere. Then, we comment some approaches in the fields of image retrieval and indexing, and local image features matching that use histograms. Finally, we comment some structural pattern recognition methods in which histograms are the local descriptors, thus, comparing histograms is needed to compare these graphs. In fact, that was the ground for the research of an efficient method to compare histograms.

5.2.1 Distances and Algorithms

We distinguish between two kind of distances. Firstly, the bin-to-bin functions, including the L_p distances, the χ^2 statistics or the KL divergence, which assume predefined correspondences in the domains of the two histograms to be compared. Secondly, there are cross-bin distances which do not assume any predefined correspondence and, therefore, it has to be comptued. Cha (2002); Rubner *et al.* (2001) present an interesting summary of bin-to-bin distances as well as the cross-bin functions addressing this correspondence problem. Peleg *et al.* (1989); Shen & Wong (1983); Werman *et al.* (1985) presented some early works using cross-bin functions. Adapted from previous work, Rubner *et al.* (2000a) presented a new definition of the distance measure between nD-histograms which was called the *Earth Movers Distance* (EMD). It is defined as the minimum amount of work that must be performed to transform one histogram into the other one by moving the distribution of masses.

Moreover, we can distinguish between the algorithms that compute the distance between 1D-histograms (Cha, 2002; Jou *et al.*, 2004; Serratosa & Sanfeliu, 2006) or nD-histograms (Ling & Okada, 2007; Rubner *et al.*, 2000a; Serratosa & Sanromà, 2006; Serratosa *et al.*, 2007); and also between the algorithms that force the ground distance or the type of measurements to be a specific one (Cha, 2002; Ling & Okada, 2007) and the ones where the ground distance is a parameter of the algorithm (Rubner *et al.*, 2000a; Serratosa & Sanfeliu, 2006).

Often, for specific set measurements, only a small fraction of the bins in a histogram contain significant information, that is, most of the bins are empty. This is more frequent when the dimensions of the histograms increase. In these cases, the methods that use histograms as fixed-sized structures obtain poor efficiency. To overcome this problem, signatures are proposed (Rubner *et al.*, 2000a; Serratosa & Sanfeliu, 2006). They are a compact representation of histograms in which only the non-empty bins are explicitly represented.

The computational cost of the bin-to-bin distances is linear, $\mathcal{O}(p)$. In these algorithms, an operation is performed sequentially on each pair of bins. The computational cost of the cross-bin distances depend on the dimension of the histograms. In the 1D case, Cha (2002) presented three algorithms to obtain the EMD between 1D-histograms when the type of measurements where *nominal*, *ordinal* and *modulo* in $\mathcal{O}(p)$, $\mathcal{O}(p)$ and $\mathcal{O}\left(p^2\right)$, respectively. Latter on, Serratosa & Sanfeliu (2006) reduced the computational cost to $\mathcal{O}(p')$, $\mathcal{O}(p')$ and $\mathcal{O}(p'^2)$, respectively, being p' the number of non-empty bins. They used signatures instead of histograms.

In the nD case ($n > 1$), there is not any known algorithm that computes the EMD in polynomial time. Rubner *et al.* (2000a) presented a sub-optimal method to compute the EMD. They used the simplex algorithm (Nash, 2000) to compute the distance measure and the method by Russell (1969) to search a good initialisation. The computational cost of the simplex iteration is $\mathcal{O}(p'^2)$, where p' is the number of non-empty bins. The main drawback is that the number of iterations is not bounded and that this method needs a good initial solution. The method by Russell (1969) is the most common method used to seek a good initial solution with a computational cost of $\mathcal{O}(p'^3)$.

Recently, Ling & Okada (2007) presented a new algorithm to compare nD-histograms in an average time complexity of $\mathcal{O}\left(p^2\right)$. The main drawback of this method is that the ground distance has to be the L_1 distance. Along the same lines, Kamarainen *et al.* (2003) introduced a subspace projection of the data, called the neighbour-bank projection, where the data (1D-histogram) is projected to a subspace which reduces the dimension of the data by combining adjacent bins, and also represents sparse data in a more tight, smoothed subspace. Jou *et al.* (2004) presented an algorithm to compute the distance between a histogram (obtained from a database image) and a retrieval image (whose histogram was not computed). The cost of the algorithm was $\mathcal{O}(p+n)$, being p and n the number of bins and the number of the pixels of the image, respectively. They used the similarity measurement functions based on the bin-to-bin functions χ^2 and the L_1 and L_2 norms.

5.2.2 Image Indexing and Retrieval

In the image recognition or segmentation domain, the distance between colour histograms has been used for object recognition and image indexing and retrieval.

One of the earliest papers was written by Swain & Ballard (1991). They used the colour histograms to identify an object in a known location and locating a known object. 3D objects were described by few histograms due to the variations on the luminance. The system had a high performance since no segmentation process was needed and a bin-to-bin distance was used.

Hafner *et al.* (1995) presented an approach for filtering images by computing a sub-optimal colour-histogram distance with a linear computational cost.

Kolesnik & Fexa (2005) employed Support Vector Machines as a classifier for automatic segmentation based on histograms. They used this technique to extract chronic wound regions from an image. They showed that colour histograms of higher dimensions provide a better cue for robust separation of classes in the feature space. To do so, they defined an automatic histogram sampling process that gives a denser bin distribution for those histogram parts with larger number of elements. Chapelle *et al.* (1999) applied support vector machines on kernel functions based on the RGB and HSV histograms. In that paper, the number of bins per each dimension of the histograms had to be reduced to 16 due to run time and space requirements.

To further enhance the retrieval effectiveness, recent approaches attempt to evolve more features into histograms.

Ennesser & Medioni (1995) developed the local histogram method to locate an object in a colour image, in which the co-occurrence histogram is employed to improve the discrimination power of colour histogram. Since the colour histogram lacks spatial information, Pass *et al.* (1996) provided the colour coherent vector method for images that integrates colour histogram and spatial relationships among features of the image to relieve this problem.

Finally, Morovic *et al.* (2002) presented an algorithm for transforming an image so as to give it exactly a given target histogram. This is achieved for any original and target 1D-histogram combinations. They developed this algorithm to study the impact of image histograms on image reproduction. That is, having a pair of image sets, they want to show whether the removal or variation in terms of the chosen characteristic also removes variation in the performance of different colour reproduction strategies.

5.2.3 Local Image Features Matching

During the last decade there has been an increasing interest in image matching using local image features. These approaches encapsulate in the form of descriptor vectors the image information local to a set of interest regions.

Many outstanding descriptors represent some kind of distributional information using multidimensional histograms. Belongie *et al.* (2002) used a 2D log-polar histogram of edge occurrences around each keypoint in order to characterize the regions. This way, the dimensions correspond to the angle of the occurrence and the distance from the center (in logarithmic scale). Lowe (2004) built 3D histograms accounting for spatial location and gradient orientations of the gradient locations within the interest regions. Lazebnik *et al.* (2005) built a 2D histogram of brightness values and distance from the center of the interest region. This type of descriptor is specially suited to texture characterization.

Finally, correspondences are established on the basis of the pairwise distance between descriptors. Most approaches use bin-to-bin approaches in order to compute the distance between multidimensional histograms. For example, Belongie *et al.* (2002) use the χ^2 test, Lowe (2004) use the Euclidean distance and Lazebnik *et al.* (2005) use the squared Euclidean distance.

5.2.4 Structural Pattern Recognition

In the structural pattern recognition domain, some graph matching approaches have been presented that use histograms as unary attributes in the nodes.

On one hand, Sanromà *et al.* (2010a,b) presented approaches to match SIFT features using attributed relational graphs.

On the other hand, Sanfeliu *et al.* (2004); Serratosa *et al.* (2002) presented *Function-Described Graphs* (FDG) and *Second-Order Random Graphs* (SORG), both structures aimed at representing a set of attributed graphs.

In all the aforementioned approaches it is necessary to compute the distance between histograms at some stage. Therefore, it is necessary to devise algorithms aimed at comparing histograms in both an accurate and efficient way.

5.3 Sets and Histograms

In this section, we give some definitions and properties related with histograms, which are independent on the dimensionality of the histograms, and the type of measurements that the histograms are composed of. The properties obtained from the definition of the histograms are useful in the definitions of the distances given in the next section. Moreover, we define the distance between the two most used types of measurements; the ordinal and the modulo. Nevertheless, it is important to emphasize that all the definitions and formulation throughout the chapter do not depend on the type of the measurement. At the end of this section, we give an example of a 2D-histogram.

5.3.1 Histogram Definition

Let $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$ be a set of p possible values that can take a measurement. Each value can be represented in a T-dimensional vector as $\mathbf{z}_j = \left(z_j^1, \ldots, z_j^T\right)$. Consider a set of n elements $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ such that $\mathbf{a}_i \in \mathcal{Z}$, $\forall \mathbf{a}_i \in \mathcal{A}$, being $\mathbf{a}_i = \left(a_i^1, \ldots, a_i^T\right)$.

The histogram of the set $\mathcal{A}$ along measurement $\mathcal{Z}$, $H\left(\mathcal{Z}, \mathcal{A}\right)$, is an ordered list of the number of occurrences of the discrete values of $\mathcal{Z}$ among the $\mathbf{a}_i$. We will denote as $\mathbf{h}$ the vectorized representation of the histogram $H\left(\mathcal{Z}, \mathcal{A}\right)$. If $\mathbf{h}\left(j\right)$, $1 \leq j \leq p$, denotes the number of elements of $\mathcal{A}$ that have value $\mathbf{z}_j$, then $\mathbf{h} = \left[\mathbf{h}\left(1\right), \ldots, \mathbf{h}\left(p\right)\right]$, where

$$\mathbf{h}\left(j\right) = \sum_{i=1}^{n} O_{ji}^{\mathcal{A}} \tag{5.1}$$

and the individual occurrences are defined as

$$O_{ji}^{\mathcal{A}} = \begin{cases} 1 & \text{if } \mathbf{a}_i = \mathbf{z}_j \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

The elements $\mathbf{h}\left(j\right)$ are usually called *bins* of the histogram and p is the number of bins of the histogram. In a T-dimensional histogram with m values per each dimension, the number of bins is $p = m^T$. Therefore, $1 \leq j \leq m^T$.

The i-th element of the set $\mathcal{A}$, $\mathbf{a}_i$, has only one value. Therefore, there is only one value of j such that $O_{ji}^{\mathcal{A}} = 1$ (when $\mathbf{a}_i = \mathbf{z}_j$) and for all the other values of j, $O_{ji}^{\mathcal{A}} = 0$ (i.e., $\mathbf{a}_i \neq \mathbf{z}_j$). As consequence, the following equation holds,

$$\sum_{j=1}^{p} O_{ji}^{\mathcal{A}} = 1 \tag{5.3}$$

This result is needed for the demonstration of some properties of the flow in section 5.4.3.

5.3.2 Type of Measurements and Distance between them

Local image descriptors account for the spatial location of the image occurrences together with some other information such as gradient orientation (SIFT) or brightness value (Spin Images). Spatial location may be represented in polar coordinates using angles (Shape Contexts). On the other hand, the most used colour representations are based on the R,G,B or H,S,I descriptors. Both angles and hue measurements are modulo-type (measurement values are ordered but form a ring due to the arithmetic modulo operation). Such measurements may coexist with ordinal-type measurements in other dimensions of a histogram.

Corresponding to these types of measurements mentioned before, we define a measure of difference between two measurement levels $\mathbf{a} = \left(a^1, a^2, \ldots, a^T\right) \in \mathcal{Z}$ and $\mathbf{b} = \left(b^1, \ldots, b^T\right) \in \mathcal{Z}$, where $a^j, b^j \in \mathbb{N}$, as follows:

$$d\left(\mathbf{a}, \mathbf{b}\right) = \sqrt{\sum_{j=1}^{T} r_j^2} \tag{5.4}$$

where the residual

$$r_j = \begin{cases} m - \left|a^j - b^j\right| & \text{if } \left|a^j - b^j\right| > m/2 \text{ and } a^j, b^j \text{ are Modulo type} \\ \left|a^j - b^j\right| & \text{otherwise} \end{cases} \tag{5.5}$$

This measure satisfies the necessary properties of a metric. Since they are straightforward facts, we omit the proofs. The proof of the triangle inequality for the modulo distance is depicted in (Cha, 2002) for the 1D case ($T = 1$).

In the applications that the type of measurements is nominal, $d\left(\mathbf{a}, \mathbf{b}\right)$ can be defined as $d\left(\mathbf{a}, \mathbf{b}\right) = 0$ if $a = b$ and $d\left(\mathbf{a}, \mathbf{b}\right) = 1$ otherwise, without loss of generality for all the other formulation.

5.3.3 Example of the Histogram Definition and Properties

Suppose that we have the domain composed by 9 values ($p = 9$),

$$\mathcal{Z} = \left\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5, \mathbf{z}_6, \mathbf{z}_7, \mathbf{z}_8, \mathbf{z}_9\right\}$$

where $\mathbf{z}_1 = (1,1)$, $\mathbf{z}_2 = (1,2)$, $\mathbf{z}_3 = (1,3)$, $\mathbf{z}_4 = (2,1)$, $\mathbf{z}_5 = (2,2)$, $\mathbf{z}_6 = (2,3)$, $\mathbf{z}_7 = (3,1)$, $\mathbf{z}_8 = (3,2)$, $\mathbf{z}_9 = (3,3)$. The dimensionality of the domain is 2 ($T = 2$) and the domain per each dimension is 3 ($m = 3$). Moreover, we have the set $\mathcal{A} = \left\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6\right\}$ composed by the 6 elements ($n = 6$) $\mathbf{a}_1 = (1,3)$, $\mathbf{a}_2 = (2,3)$, $\mathbf{a}_3 = (1,3)$, $\mathbf{a}_4 = (3,2)$, $\mathbf{a}_5 = (3,2)$, $\mathbf{a}_6 = (3,2)$.

In that case, the individual occurrences $O_{ji}^{\mathcal{A}}$ and the histogram bins $\mathbf{h}\left(j\right)$ take the following values (table 5.1),

Note that, given a column of the table, the addition of the cells is the value of the histogram. Moreover, given a row, the addition is always 1, as it is defined in equation (5.3).

If $\mathbf{a}_i$ are ordinal measurements, $d\left(\mathbf{a}_1, \mathbf{a}_4\right) = \sqrt{2^2 + 1^2} = \sqrt{5}$ and if $\mathbf{a}_i$ are modulo measurements, $d\left(\mathbf{a}_1, \mathbf{a}_4\right) = \sqrt{1^2 + 1^2} = \sqrt{2}$

5.4 Definition of the New Distance

The aim of this section is to define EMD-g_f. EMD-g_f is an EMD but with a specific definition of the flow between bins, g_f, to obtain the two following

$O^{\mathcal{A}}_{ji}$	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8	j=9
i=1			1						
i=2						1			
i=3			1						
i=4								1	
i=5								1	
i=6								1	

	$\mathbf{h}(1)$	$\mathbf{h}(2)$	$\mathbf{h}(3)$	$\mathbf{h}(4)$	$\mathbf{h}(5)$	$\mathbf{h}(6)$	$\mathbf{h}(7)$	$\mathbf{h}(8)$	$\mathbf{h}(9)$
$\mathbf{h}$	0	0	2	0	0	1	0	3	0

Table 5.1: The individual occurrences $O^{\mathcal{A}}_{ji}$ and histogram $\mathbf{h}$ of the set $\mathcal{A}$.

advantages. First, we can deduct that the value of the distance between sets defined later is exactly the same as the value of the EMD-g_f. Second, it allows to define in section 5.5 an approximate algorithm to obtain a sub-optimal value of the EMD-g_f. Obtaining the distance between sets is a NP-problem but there are some good approximations such as the Hungarian method in $\mathcal{O}\left(n^3\right)$ (Ahuja *et al.*, 1993). Our algorithm obtains a sub-optimal solution in an average real cost of $\mathcal{O}\left(m^2\right)$. Therefore, and considering that in some real applications $m < n$, our method might be useful to compare not only histograms but also sets of vector elements.

5.4.1 Distance between Sets

Given two sets of n elements, $\mathcal{A}$ and $\mathcal{B}$, the computation of the distance measure is considered as the problem of finding the minimum difference of pair assignments between both sets. That is, to determine the best one-to-one assignment f (bijective function) between the sets such that the sum of all the differences between two individual elements in a pair $\mathbf{a}_i \in \mathcal{A}$ and $\mathbf{b}_{f(i)} \in \mathcal{B}$ is minimised.

$$D_{\text{set}}\left(\mathcal{A}, \mathcal{B}\right) = \min_{f:\mathcal{A}\to\mathcal{B}} \sum_{i=1}^{n} d\left(\mathbf{a}_i, \mathbf{b}_{f(i)}\right) \tag{5.6}$$

5.4.2 Distance between Histograms: Earth Mover's Distance

Rubner *et al.* (2000b) presented the EMD, a cross-bin histogram distance. Intuitively, given two T-dimensional histograms, one can be seen as a mass of earth properly spread in space, the other as a collection of holes in that same space. Then, the distance measure is the least amount of work needed to fill the holes with earth. More formally, given two histograms $\mathbf{h}$ and $\mathbf{k}$ of two sets $\mathcal{A}$ and $\mathcal{B}$, respectively, where measurements can have one of p values contained in the set $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$, the distance between the histograms D_{EMD} is defined as follows,

$$D_{\text{EMD}}\left(\mathbf{h}, \mathbf{k}\right) = \min_{f:\mathcal{A}\to\mathcal{B}} \sum_{j,j'=1}^{p} d\left(\mathbf{z}_j, \mathbf{z}_{j'}\right) g_f\left(j, j'\right) \tag{5.7}$$

The flow between the bins of both histograms is represented by $g_f(j, j')$, that is, the mass of earth that is moved from the bin j to the bin j', which is subject to the following constraints,

$$g_f(j, j') \geq 0, \ 1 \leq j, j' \leq p \tag{5.8}$$

$$\sum_{j'=1}^{p} g_f(j, j') = \mathbf{h}(j), \ 1 \leq j \leq p \tag{5.9}$$

$$\sum_{j=1}^{p} g_f(j, j') = \mathbf{k}(j), \ 1 \leq j' \leq p \tag{5.10}$$

$$\tag{5.11}$$

The distance between bins is represented by $d(\mathbf{z}_j, \mathbf{z}_{j'})$ (equations (5.4) and (5.5)). The product $d(\mathbf{z}_j, \mathbf{z}_{j'})\, g_f(j, j')$ represents the work needed to transport this mass of earth.

5.4.3 A Specific Flow Function g_f between Bins

In the EMD, an arbitrary value of $g_f(j, j')$ is considered with some constraints (equations (5.8) to (5.10)) to relate the algorithm used to compute this distance and the transportation problem (Rubner *et al.*, 2000b). In EMD-g_f, the flow between bins, $g_f(j, j')$, is defined as a function of the one-to-one assignment f between the sets $\mathcal{A}$ and $\mathcal{B}$ used to compute the distance D_{set} (equation (5.6)) as follows,

$$g_f(j, j') = \sum_{i=1}^{n} O^{\mathcal{A}}_{ji} O^{\mathcal{B}}_{j' f(i)}, \ 1 \leq j, j' \leq p \tag{5.12}$$

where the occurrences matrix O is given in equation (5.2).

With this new definition, we obtain two advantages. First, there is a logical relation between sets and histograms that make possible to demonstrate that $D_{set} = D_{\text{EMD-}g_f}$ (demonstration in appendix A). Second, we transform the imposed constraints (equations (5.8) to (5.10)) into deducted properties of the flow that are crucial to assure that the approximate algorithm presented in section 5.5 converges to a sub-optimal solution of the EMD-g_f (demonstrations in appendix B).

5.4.4 Example of the Distance between Sets and Histograms

Figure 5.2 shows the sets $\mathcal{A}$ and $\mathcal{B}$ and the optimal labelling between them. The type of measurements are ordinal composed by 2D elements. Given this optimal labelling and the distance between sets of equation (5.6), the distance value is:

$$\begin{aligned} &D_{set}(\mathcal{A}, \mathcal{B}) = \\ &\quad d((1,3),(1,2)) + d((2,3),(2,3)) + d((1,3),(1,2)) + d((3,2),(2,1)) + \\ &\quad\ d((3,2),(3,1)) + d((3,2),(2,1)) = 1 + 0 + 1 + \sqrt{2} + 1 + \sqrt{2} = 3 + 2\sqrt{2} \end{aligned}$$

Figure 5.3 shows the histograms of the sets $\mathcal{A}$ and $\mathcal{B}$ plotted in a 2D table and the flow between them.

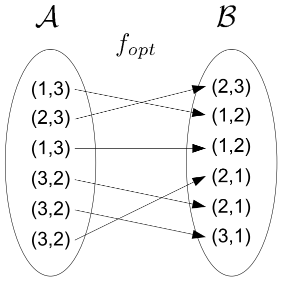


Figure 5.2: Set $\mathcal{A}$ and set $\mathcal{B}$ and the optimal labelling between their elements.

h	**1**	**2**	**3**
1	*0*	*0*	*2*
2	*0*	*0*	*1*
3	*0*	*3*	*0*

k	**1**	**2**	**3**
1	*0*	*2*	*0*
2	*2*	*0*	*1*
3	*1*	*0*	*0*

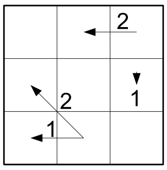


Figure 5.3: Histograms **h** and **k**, and graphical representation of the flow.

The flow, $g_f(j, j')$, is represented by the arrows and its value is drawn above the arrows. The distance between bins, $d(\mathbf{z}_j, \mathbf{z}_{j'})$, is the length of the arrows.

$$D_{\text{EMD-}g_f}(\mathbf{h}, \mathbf{k}) = 1 \cdot 2 + \sqrt{2} \cdot 2 + 1 \cdot 1 + 0 \cdot 1 = 3 + 2 \cdot \sqrt{2}$$

5.5 Approximate Algorithm for the EMD-g_f

Our algorithm that computes the EMD-g_f is inspired on the solution by Russell (1969) of the transportation problem but the cost has been reduced from $\mathcal{O}(p^3)$ to $\mathcal{O}(p^2)$ due to the fact that the cost of transporting a single unit of goods is know a priori. This cost is the distance between bins, $d(\mathbf{z}_j, \mathbf{z}_{j'})$, given the dimensionality and the number of bins of the histograms.

In our algorithm (figure 5.4), n is the number of elements and so, the maximum flow that can be transported. The functions *first* and *next* (commented in the next section) return a pair of bins from both histograms, j and j', at each iteration. Then, the flow is computed, $g_f(j, j')$, and extracted from the histograms and the maximum flow, n. Finally, the cost of this transportation, $g_f(j, j') \cdot d(\mathbf{z}_j, \mathbf{z}_{j'})$ is added to the final distance value. Equations (5.8) to (5.10) assure the algorithm finishes when all the goods have been transported since $\mathbf{h}(j) > 0$ and $\mathbf{k}(j') > 0$ throughout the iterations.

5.5.1 The *first* and *next* Functions

Given a pair (j, j') (supplier and a consumer, respectively), the *first* and *next* functions return the first and next pairs of supplier and consumer to be explored,

```
$D_{EMD-g_f} = 0$
$n = |\mathcal{A}|$    # or equivalently $n = |\mathcal{B}|$ since $|\mathcal{A}| \equiv |\mathcal{B}|$
$(j, j') \leftarrow \mathit{first}\,()$
while $n > 0$ do
    $g_f\,(j, j') \leftarrow \min\left(\mathbf{h}\,(j)\,, \mathbf{k}\,(j')\right)$
    $\mathbf{h}\,(j) \leftarrow \mathbf{h}\,(j) - g_f\,(j, j')$
    $\mathbf{k}\,(j') \leftarrow \mathbf{k}\,(j') - g_f\,(j, j')$
    $n = n - g_f\,(j, j')$
    $D_{EMD-g_f} = D_{EMD-g_f} + g_f\,(j, j') \cdot d\,(\mathbf{z}_j, \mathbf{z}_{j'})$
    $j, j' \leftarrow \mathit{next}\,(j, j')$
end
```

Figure 5.4: Approximate algorithm that computes the EMD-g_f.

respectively. The first pair of supplier-consumer and the order generated by the next function only depends on the dimensionality of histogram and the number of bins but not on the values of the histograms. It is for this reason that the *first* and *next* functions can be computed a priori.

The order of the pairs (j, j') is set by decrementing an energy function E as follows,

$$\begin{aligned}
(i, i') = \mathit{next}\,(j, j') \ \text{ iff } \ & E\,(i, i') \leq E\,(j, j') \\
& \text{and there is no } \ (l, l')\,, (l, l') \neq (i, i')\,, (l, l') \neq (j, j') \ \text{ such that} \\
& \qquad E\,(l, l') \leq E\,(j, j') \ \text{ and } E\,(i, i') > E\,(i, i')
\end{aligned} \tag{5.13}$$

where E is defined as

$$E\,(j, j') = \mathit{path\_deviation}_{j'}\,(j) + \mathit{path\_deviation}_{j}\,(j') \tag{5.14}$$

The $\mathit{path\_deviation}_{j'}\,(j)$ is the difference between the maximum cost from the bin j to any bin of the histogram and the real cost from this bin to the bin j',

$$\mathit{path\_deviation}_{j'}\,(j) = \mathit{max\_d}\,(\mathbf{z}_j) - d\,(\mathbf{z}_j, \mathbf{z}_{j'}) \tag{5.15}$$

This heuristic aims to give preference to the bins located at the extrema of the histograms in order to avoid as much as possible large distance transportations.

Note that several pairs (j, j') can obtain the same energy value. In those cases, the order between them is set arbitrarily.

Figure 5.5.(a) shows an image that represents the energy function E of a 1D-histogram with 25 bins. Figure 5.5.(b) shows an image that represents the order obtained by the *next* function. In both images, dark pixels represent low values. That is, high energy (brighter pixels in the left image) gives the first positions in the order (darker pixels in the right image). Consumer j (j-th bin of $\mathbf{h}$) is represented in the j-th row. Supplier j' (j'-th bin of $\mathbf{k}$) is represented by the j'-th column.

Similarly to figure 5.5, figure 5.6 shows the energy function and order of a 2D-histogram with 5 bins per dimension. The Consumer j (j-th bin of $\mathbf{h}$) which has the 2D value (p, q)) is represented in the image as the row $q \cdot 5 + p$. The Supplier j' (j'-th bin of $\mathbf{k}$ which has the 2D value (t, s)) is represented in the image as the column $t \cdot 5 + s$.

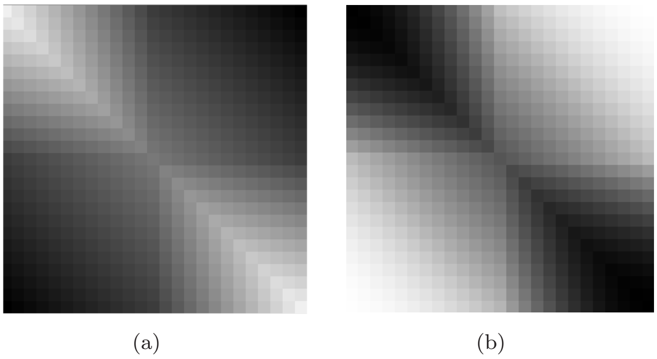

(a) (b)

Figure 5.5: (a) Energy function and, (b) order obtained by the *next* function in a 1D-histogram with 25 bins.

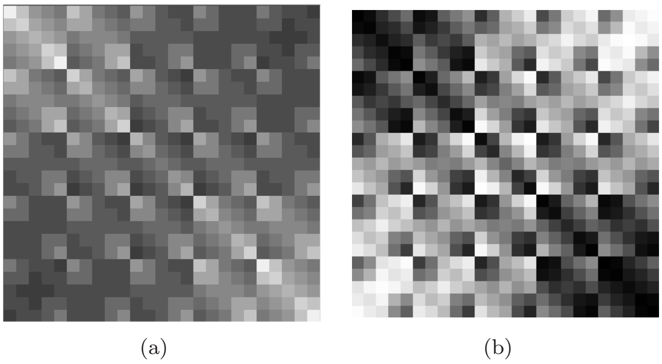

(a) (b)

Figure 5.6: (a) Energy function and, (b) order obtained by the *next* function in a 2D-histogram with 5 bins per dimension. The total number of bins is 25.

5.5.2 The Worst Computational Cost

Each step of the loop of the algorithm has a constant computational cost. The next function is implement as an array that for each pair (j, j'), returns the next pair (i, i'). For this reason, the worst computational cost of our algorithm only depends on the number of iterations. The algorithm finishes when all the goods, n, have been transported and so, the worst case would be in the situation that this is achieved at the last transportation from (j, j'), to (i, i'). The number of possible transportations is p^2.

5.6 Empirical Study of Time Complexity

We have demonstrated that the worst-case complexity of the proposed algorithm is $\mathcal{O}\left(p^2\right)$. In this experiment, we evaluate the real time complexity using real data. To that aim, we used the coil image database (Jou *et al.*, 2004) (figure 5.7 shows 20 objects). Figure 5.8 shows the average number of iterations while comparing the histograms of these images. In the left plot, images are represented by 3D-histograms (RGB, HSV and CIE-lab). In the central plot, images are represented by 2D-histograms (HS and HL) and in the right plot, images are represented by 1D-histograms (H and L). In the three plots, there is the worst computational cost $\mathcal{O}\left(p^2\right)$ and also the $\mathcal{O}\left(m^2\right)$ function in 2D and 3D-histogram plots, and the $\mathcal{O}\left(m\right)$ function in the 1D-histogram plot. The Euclidean-modulo distance was used to compare the Hue channels and the Euclidean distance was used in the other cases.

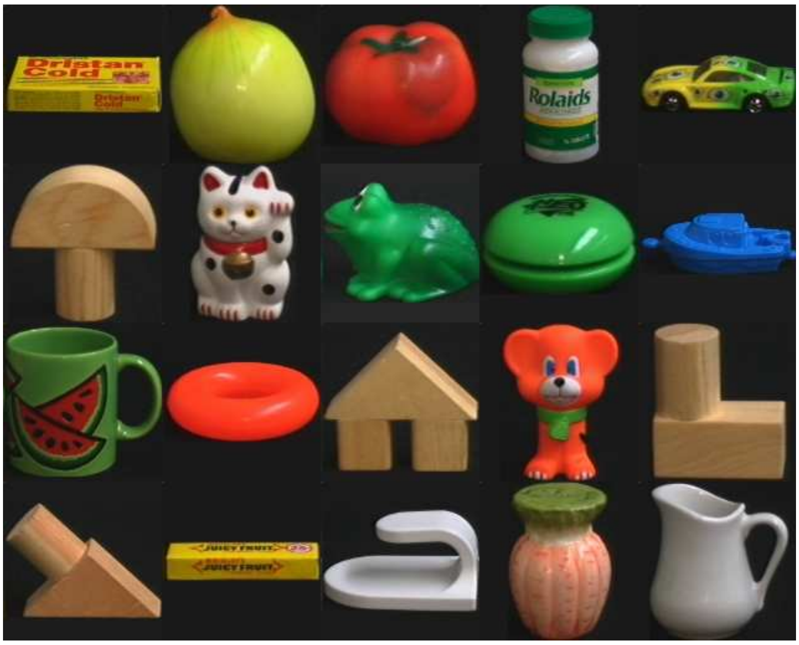
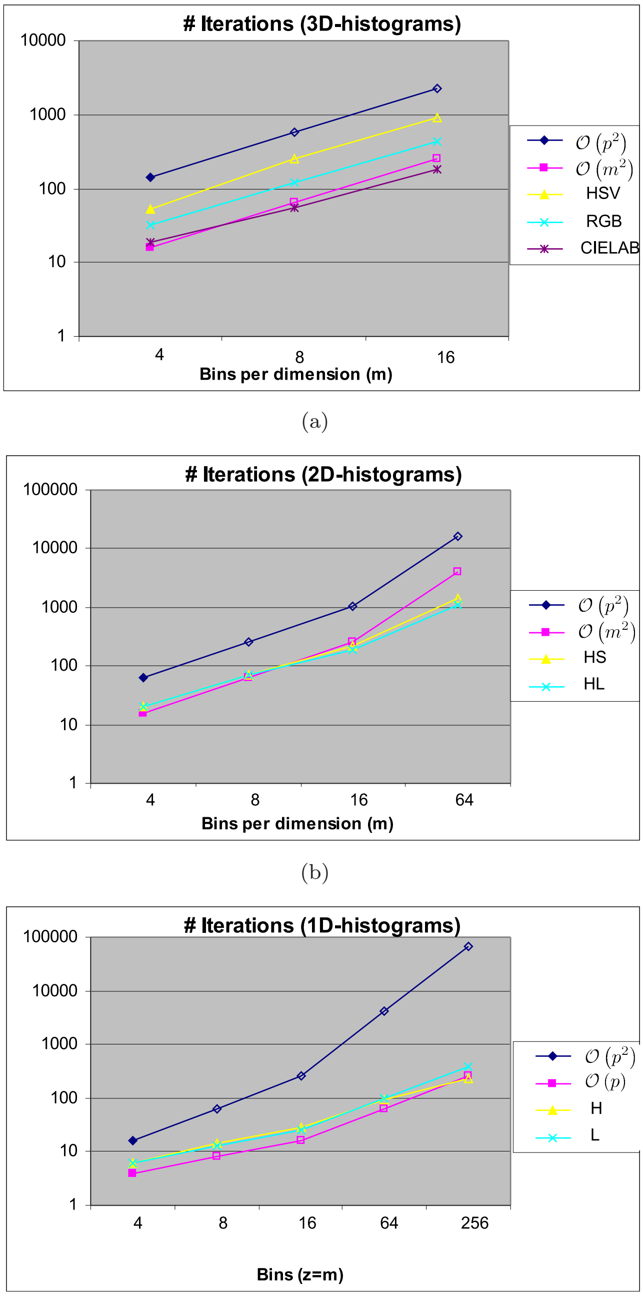


Figure 5.7: Images taken at angle 5 of the 20 objects.

We conclude that the average real time complexity is clearly lower than $\mathcal{O}\left(p^2\right)$ in all the cases. In the CIE-lab, HS and HL cases, the time complexity is lower than $\mathcal{O}\left(m^2\right)$. The experiments reported by Ling & Okada (2007) with randomly generated 2D-histograms show a time complexity of $\mathcal{O}\left(p^2\right)$. Our algorithm is faster and does not need the ground distance to be L_1 as it is needed in their algorithm. Moreover, in the 1D-histograms the time complexity is almost linear, $\mathcal{O}\left(p\right)$. The demos and experiments using 1D-histograms reported by Cha (2002) show a computational cost of $\mathcal{O}\left(p\right)$ for the nominal and ordinal ground distances and $\mathcal{O}\left(p^2\right)$ for the modulo distance. Our algorithm is clearly faster in the modulo case and it has similar results in the other cases.

Finally, the EMD (Rubner *et al.*, 2000a) is the only one that proposes an algorithm to compute a distance between nD-histograms in which any ground distance can be applied. It has a $\mathcal{O}\left(p^3\right)$ computational cost. We have shown in this section that our algorithm has a lower computational cost than the EMD algorithm. We present in the next section a retrieval comparison between both algorithms.



(a)

(b)

(c)

Figure 5.8: Number of iterations of (a) 3D, (b) 2D and (c) 1D-histograms.

5.7 Retrieval Rate Study

The WANG database (figure 5.9) was used for the retrieval rate study. It is a subset of the Corel database of 1000 images which are subdivided into 10 classes (e.g. Africa, beach, ruins, and food). The query set was composed by 2 images for each class and the database set was composed by the other images. We used the 5-NN criterion.

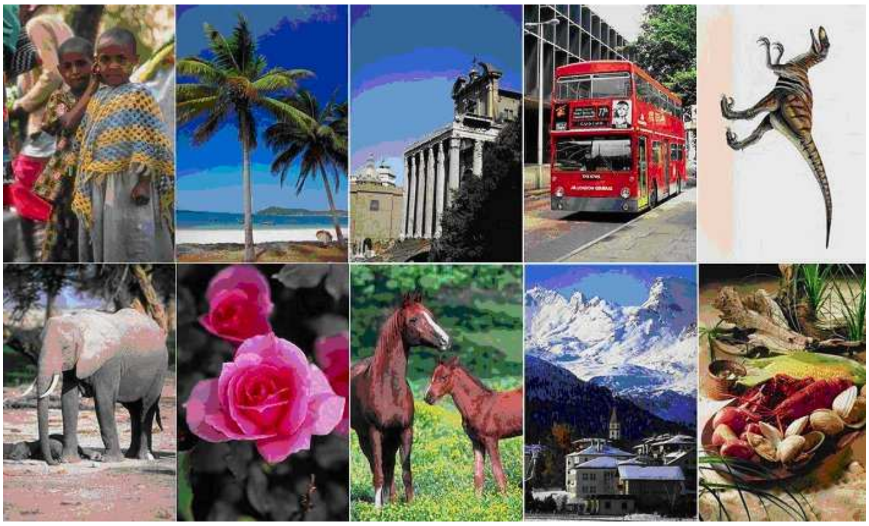


Figure 5.9: Ten images of the WANG database. One from each class.

Figure 5.10 shows the recognition ratio with respect to the number of colours in 6 different cases using the EMD-g_f (top) and EMD (bottom). In the 1D and 2D-histograms cases, the recognition ratio in both plots is almost the same. In the 3D-histogram cases the EMD obtains slightly higher rates.

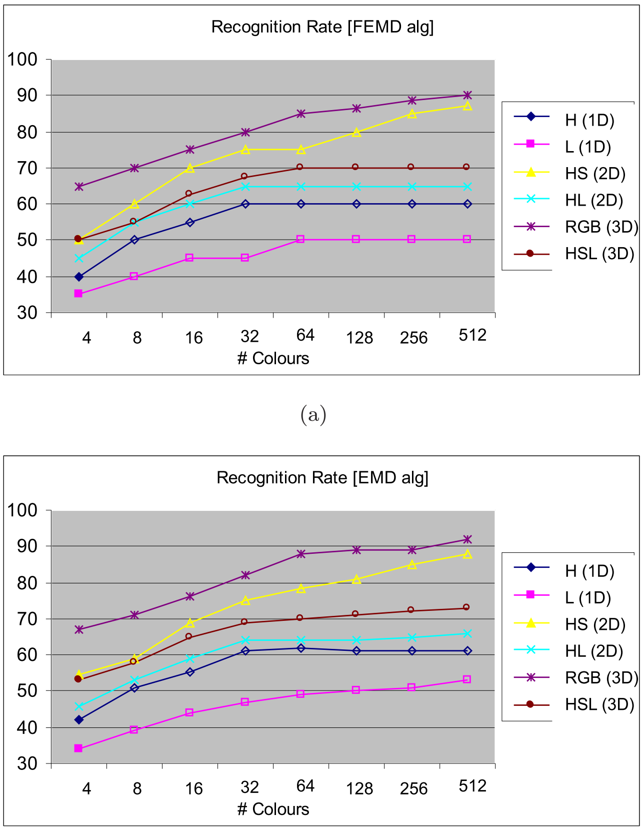
5.8 Non-Rigid Shape Registration with Shape Contexts

We have performed non-rigid registration experiments with the shapes dataset by Chui & Rangarajan (2000) using the *Shape Contexts* feature descriptors (Belongie *et al.*, 2002). This dataset contains perturbed instances of a fish and a Chinese character templates, consisting of 98 and 105 points, respectively. Perturbation levels range from mild to severe, with 100 different instances for each level.

In the present experiments we have used the perturbed instances consisting on non-rigid deformations based on Gaussian radial basis functions (RBF) (Yuille & Grzywacz, 1989), and independent random noise applied to each point independently. A certain amount of ground-level non-rigid deformation is maintained in the random noise perturbations.

Figure 5.11 show examples of the model templates and moderately perturbed instances of each case.

Shape Contexts are highly discriminant features that encode the spatial distribution and frequency of the rest of the points with respect to a given point. The Shape Context for a point consists of a 2D histogram of the occurrences of the remaining points in polar coordinates. The histogram dimension spawning



(a)

(b)

Figure 5.10: Retrieval Rate with respect to the number of colours.

the length magnitude is scaled in a logarithmic fashion in order to achieve a higher resolution in occurrences at nearby distances.

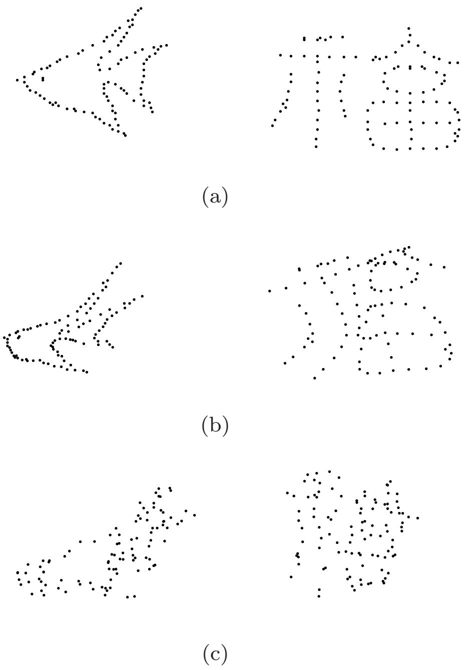
An illustrative example of the Shape Contexts is found in section 1.3.2.

Point-set registration consists on iteratively solving the correspondence and alignment problems until convergence. Correspondences are decided on the basis of the pairwise distances between their Shape Contexts. Given the correspondences, alignment is performed using *Thin-Plate Splines* (Bookstein, 1989) as reported by Belongie *et al.* (2002).

Consider the point-sets from a deformed instance $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and the model template $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$, where $\mathcal{I} = 1 \ldots |\mathcal{X}|$ and $\mathcal{J} = 1 \ldots |\mathcal{Y}|$ are the index-sets.

Let $\mathcal{X}^a = \{g\left(\mathbf{x}_a - \mathbf{x}_b\right),\ \forall b \neq a\}$ be the set of differences between the point $\mathbf{x}_a$ and the rest of the points (the same applies for $\mathcal{Y}^\alpha$). Note that $\mathcal{X}^a$ and $\mathcal{Y}^\alpha$ contain one less element than the original sets $\mathcal{X}$ and $\mathcal{Y}$ since the reference points $\mathbf{x}_a$ and $\mathbf{y}_\alpha$ are not subtracted from themselves.

$g : \mathbb{R}^2 \rightarrow \mathcal{Z}$ is a function that maps Cartesian coordinates to the domain of possible histogram values. $\mathcal{Z}$ divides the domain of possible locations into 12 and 5 bins for the angle and length magnitudes of the polar representation,



(a)

(b)

(c)

Figure 5.11: Fish (left) and Chinese character (right) (a) model templates, (b) non-rigid deformation examples, and (c) independent Gaussian noise examples.

respectively. Hence, the domain $\mathcal{Z}$ contains $12 \times 5 = 60$ values (i.e., $p = 60$ bins).

The g function, apart from converting the differences between points from Cartesian to polar coordinates, also maps these differences to their closest representatives in $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_{60}\}$. In other words, it assigns each value to the bin within it falls into. As said earlier, the length magnitude is discretized into 5 values in logarithmic scale.

For brevity, we denote as $\mathbf{h}_a \equiv H(\mathcal{Z}, \mathcal{X}^a)$ the Shape Context for the point $\mathbf{x}_a$ (the same for $\mathbf{k}_\alpha \equiv H(\mathcal{Z}, \mathcal{Y}^\alpha)$).

The cost $C_{a\alpha}$ of the matching $\mathbf{x}_a \to \mathbf{y}_\alpha$ is computed as the distance between the Shape Contexts $\mathbf{h}_a$ and $\mathbf{k}_\alpha$. In the original approach, Belongie *et al.* (2002) used the χ^2 test to that end. This is,

$$C_{a\alpha}^{\chi^2} = \frac{1}{2} \sum_j \frac{\left[\mathbf{h}_a(j) - \mathbf{k}_\alpha(j)\right]^2}{\mathbf{h}_a(j) + \mathbf{k}_\alpha(j)} \tag{5.16}$$

On the other hand, we apply the algorithm of figure 5.4 in order to compute the distance between the Shape Contexts $\mathbf{h}_a$ and $\mathbf{k}_\alpha$ and thus, obtain the cost $C_{a\alpha}^{\text{EMD}}$ of the matching $\mathbf{x}_a \to \mathbf{y}_\alpha$. We have into account that the angle measurements are *modulo* type.

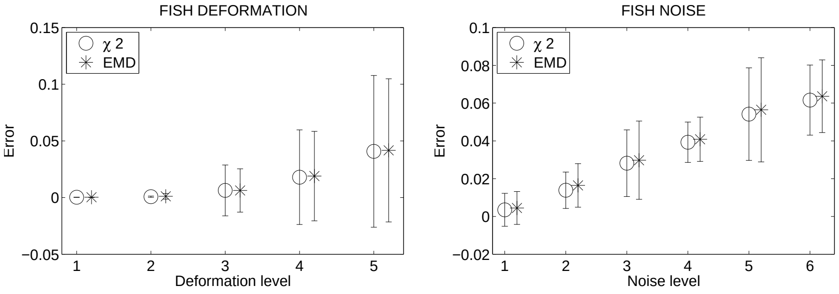
Once the cost matrices C^{χ^2} and C^{EMD} are computed, correspondences are decided in both cases with the Hungarian method (Munkres, 1957).

Thin-Plate Splines (Bookstein, 1989) are used to align the deformed point-set $\mathcal{X}$ with the model template $\mathcal{Y}$ according to the recovered correspondences.

The regularization parameter of the Thin-Plate Splines is set as reported by Belongie *et al.* (2002).

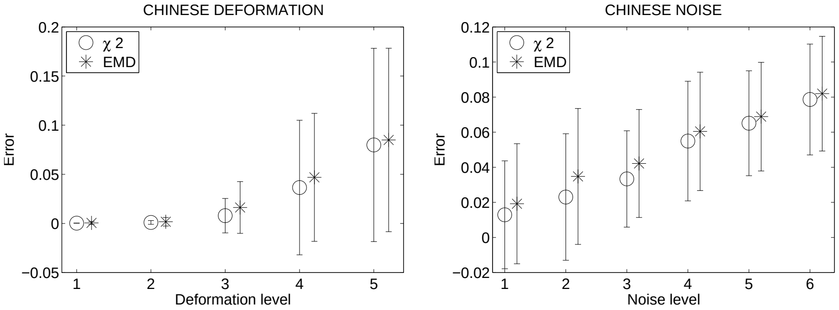
This ICP-like process of alternate correspondence and alignment updates is run a predefined number of times or until convergence of the point-sets.

The *mean registration error* is the mean distance between the model template points and their corresponding counterparts in the deformed template. At the end of the process a perfect registration occurs if all the points in the deformed template coincide with the locations of their corresponding model counterparts. Figures 5.12 and 5.13 show the mean registration errors obtained by using the matching costs computed by both methods.



(a) (b)

Figure 5.12: Mean registration and standard deviations obtained by using our approach and the χ^2 test in order to compute the matching costs in the (a) fish deformation and (b) fish noise datasets. Horizontal axis represents the perturbation's degree.



(a) (b)

Figure 5.13: Mean registration and standard deviations obtained by both approaches in the Chinese character datasets. Horizontal axis represents the perturbation's degree.

The registration performance does not get improved by using the proposed EMD-based distance measure between the Shape Contexts. Apparently, the bin-to-bin correspondence assumption of the χ^2 test measure holds in the case of these descriptors.

5.9 Conclusions and Future Work

We have presented a new definition of the distance between nD-histograms called EMD-g_f and an efficient algorithm to compute an approximation of this distance. Similarly to the EMD, any dimension of histograms or ground distance can be used. Nevertheless, the EMD-g_f outperforms the EMD in two aspects. From the computational point of view, results show that the real computational cost has decreased from $O\left(p^3\right)$ to $\mathcal{O}\left(m^2\right)$ which makes our new algorithm useful to a greater amount of applications. From the theoretical point of view, we have defined a specific flow function between bins, g_f, such that $D_{set} = D_{\text{EMD-}g_f}$. Thus, EMD-$g_f$ can be used as a fast and approximate method to compare sets of elements described by vectors.

In the retrieval rate study the proposed algorithm has shown a similar performance than the orignal EMD algorithm for different choices of colour spaces and dimensions of the histograms.

With regards to the non-rigid registration experiments with the Shape Contexts we have not found any advantage of using our approach instead of the originally proposed χ^2 test as distance measure between these descriptors. This is because the correspondence assumption between bins done by the bin-to-bin measures is fulfilled in the case of Shape Contexts. With regards to other descriptors, the invariance to rotation or affine shape introduced by many methods during the feature description stage suggests that the bin-to-bin correspondence assumption will also hold in these cases. It is for this reason that we do not have a particular interest in extending these experiments to other types of descriptors.

Chapter 6

Attributed Graph Matching for SIFT-Features Association

6.1 Introduction

Image-features matching based on Local Invariant Features Extraction (LIFE) methods has become a topic of increasing interest over the last decade. LIFE methods extract stable representations from a selected set of characteristic regions (features) of the image. These local representations are aimed to be invariant at a certain extent to image deformations as, for example, changes in illumination and viewpoint. Mikolajczyk & Schmid (2005) identified SIFT descriptors (Lowe, 2004) as the most stable representations among a number of approaches.

SIFT features are located at the salient points of the scale-space. Each SIFT feature retains the magnitudes and orientations of the image gradient at the neighboring pixels. This information is represented in a 128-length vector. In chapter 1 we describe several Local Image Feature Extractors.

Despite its efficiency, correspondence matches based on local image information may still present some errors. Outlier rejectors are approaches aimed at fixing these errors by locating and removing the spurious matches that compromise global consistency. To cite some examples, RANSAC (Fischler & Bolles, 1981) and *Graph Transformation Matching* (Aguilar *et al.*, 2009) select a subset of geometrically / structurally consistent matches. In sections 2.4 and 4.3.4 we give more details about these two approaches.

Other approaches that enforce global consistency are, for example, *Robust Point Matching* by Gold *et al.* (1998); Rangarajan *et al.* (1997) or unified alignment and correspondence by Luo & Hancock (2003). Unlike outlier rejectors, these approaches are able to modify the initial correspondence-set during their optimization processes. Check sections 3.4 and 4.5.2 for more details about these methods.

In this chapter we present two graph matching approaches aimed at finding a set of structurally consistent matches. The main novelty is that we use SIFT descriptor-vectors as attribute information during the optimization process. In this way, constraints are imposed both in terms of structural relations and local image information.

The first approach poses the graph matching problem as a *Maximum A Posteriori* (MAP) estimation within a discrete labelling framework (Wilson & Hancock, 1997). The second approach uses the continuous relaxation of the graph matching problem proposed in *Graduated Assignment* (Gold & Rangarajan, 1996).

In section 6.2 we introduce some definitions and notation. The discrete and continuous graph matching approaches are presented respectively in sections 6.3 and 6.4. In section 6.5 we compare them to outlier rejectors as well as to point-set registration and graph matching methods in a series of SIFT-matching experiments with synthetic images. Finally, in section 6.6 some conclusions are given.

6.2 Definitions and Notation

Consider a model image I_M showing a certain scene. Consider a data image I_D showing the same scene as I_M but with some random variations such as viewpoint change, illumination variation, nonrigid deformations in the objects of the scene, etc ... Consider the locations of two sets of SIFT features or *keypoints* $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{x}_\alpha,\ \alpha \in \mathcal{J}\}$, from the data and model images, respectively, where $\mathcal{I} = 1, \ldots, |\mathcal{X}|$ and $\mathcal{J} = 1, \ldots, |\mathcal{Y}|$ are the index-sets. Consider also two sets of SIFT feature descriptors (Lowe, 2004) $\mathcal{H} = \{\mathbf{h}_a\}$ and $\mathcal{K} = \{\mathbf{k}_\alpha\}$ so that each keypoint $\mathbf{x}_a$ and $\mathbf{y}_\alpha$ is respectively associated to a SIFT feature descriptor $\mathbf{h}_a$ and $\mathbf{k}_\alpha$.

Definition 6.1 *We define a graph representing a set of SIFT keypoints from the data-image I_D as a three tuple $\mathbf{G} = (\mathcal{U}, D, \mathcal{H})$ where $u_a \in \mathcal{U}$ is a node representing the SIFT keypoint with position $\mathbf{x}_a \in \mathcal{X}$ and SIFT descriptor-vector $\mathbf{h}_a \in \mathcal{H}$, and D is the adjacency matrix such that*

$$D_{ab} = \begin{cases} 1 & \textit{if } u_a \textit{ and } u_b \textit{ are linked by an edge} \\ 0 & \textit{otherwise} \end{cases} \tag{6.1}$$

Consider also the graph $\mathbf{H} = (\mathcal{V}, M, \mathcal{K})$ that represent a set of keypoints from the model image I_M.

Definition 6.2 *Consider the function $\gamma : \mathcal{I} \to \mathcal{J}$ that maps each descriptor $\mathbf{h}_a$ in the data image I_D to the descriptor $\mathbf{k}_{\gamma(a)}$ in the model image I_M such that $\left\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\right\|$ is the second smallest Euclidean distance $\forall \mathbf{k}_\alpha \in \mathcal{K}$.*

Similarly, the function $g : \mathcal{J} \to \mathcal{I}$ maps descriptors in the model image to descriptors in the data image such that $\left\|\mathbf{k}_\alpha - \mathbf{h}_{g(\alpha)}\right\|$ is the second smallest distance $\forall \mathbf{h}_a \in \mathcal{H}$.

Definition 6.3 *According the SIFT ratio test (Lowe, 2004), a keypoint $\mathbf{x}_a$ with descriptor vector $\mathbf{h}_a$ is matched to a keypoint $\mathbf{y}_\alpha$ with descriptor vector $\mathbf{k}_\alpha$ iff:*

$$\frac{\|\mathbf{h}_a - \mathbf{k}_\alpha\|}{\left\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\right\|} \leq \tau \tag{6.2}$$

where $0 < \tau \leq 1$ is a ratio defining the tolerance to false positives.

This is, a keypoint $\mathbf{x}_a$ from the data-image I_D is matched to the closest (in the descriptor-vector space) keypoint $\mathbf{y}_\alpha$ from the model-image I_M iff the ratio of the distance between $\mathbf{h}_a$ and $\mathbf{k}_\alpha$ to the second smallest distance from $\mathbf{h}_a$ is below a certain value τ. If this condition is not met, then keypoint $\mathbf{x}_a$ is leaved unmatched.

Definition 6.4 *We define the assignment function $f : \mathcal{I} \to \mathcal{J} \cup \emptyset$ that maps keypoints (or nodes) in image I_D to keypoints (or nodes) in image I_M or to* null. *Accordingly, $f(a) = \alpha$ means that node $u_a \in \mathcal{U}$ is matched to node $v_\alpha \in \mathcal{V}$, and $f(a) = \emptyset$ means that it is not matched to any node.*

Analogously, we define the assignment variable $s_{a\alpha} \in S$ such that

$$s_{a\alpha} = \begin{cases} 1 & \textit{if } f(a) = \alpha \\ 0 & \textit{otherwise} \end{cases} \tag{6.3}$$

subject to the constraints $\forall a, \sum_\alpha s_{a\alpha} = \{0, 1\}$ and $\forall \alpha, \sum_a s_{a\alpha} = \{0, 1\}$. This is, each node $u_a \in \mathcal{U}$ can be assigned only to one node $v_\alpha \in \mathcal{V}$.

6.3 A Discrete Labeling Approach

We follow a similar approach than Wilson & Hancock (1997) in which discrete labelling updates are done according to a MAP rule.

The idea of discrete labelling by Wilson & Hancock (1997) is to visit each node and update the assignment variable S in order to gain the maximum improvement in our matching criterion. Unlike probabilistic-relaxation-based approaches (Gold & Rangarajan, 1996; Hummel & Zucker, 1983; Rosenfeld *et al.*, 1976), discrete labelling do not allow for soft assignments.

At iteration $(n+1)$, we assign each data-graph node u_a to the model-graph node v_α with the maximum posterior probability $P\left(v_\alpha | u_a, S^{(n)}\right)$ given the current assignment variable $S^{(n)}$.

We use the following expression for the posterior probabilities

$$\begin{aligned} P\left(v_\alpha | u_a, S^{(n)}\right) &= \frac{P\left(u_a, v_\alpha | S^{(n)}\right) P\left(S^{(n)}\right)}{P\left(u_a, S^{(n)}\right)} \\ &= \frac{P\left(u_a, v_\alpha | S^{(n)}\right) P\left(S^{(n)}\right)}{P\left(u_a | S^{(n)}\right) P\left(S^{(n)}\right)} \\ &= \frac{P\left(u_a, v_\alpha | S^{(n)}\right)}{\sum_{\alpha'} P\left(u_a, v_{\alpha'} | S^{(n)}\right)} \end{aligned} \tag{6.4}$$

where $P\left(u_a, v_\alpha | S^{(n)}\right)$ is the conditional density measurement for the match between nodes u_a and v_α given the current assignment variable $S^{(n)}$.

Since the posterior probability $P\left(v_\alpha | u_a, S^{(n)}\right) \propto P\left(u_a, v_\alpha | S^{(n)}\right)$ is proportional to the conditional density measurement, we define the following updating equation which is equivalent to the MAP rule.

$$s_{a\alpha}^{(n+1)} = \begin{cases} 1 & \text{if } \alpha = \arg\max\limits_{\alpha'} P\left(u_a, v_{\alpha'} | S^{(n)}\right) \wedge \nexists\, b \text{ s.t.} \\ & \quad \Big[\alpha = \arg\max\limits_{\alpha'} P\left(u_b, v_{\alpha'} | S^{(n)}\right) \wedge \\ & \quad P\left(u_b, v_\alpha | S^{(n)}\right) > P\left(u_a, v_\alpha | S^{(n)}\right)\Big] \\ 0 & \text{otherwise} \end{cases} \tag{6.5}$$

where $s^{(n+1)}$ is the assignment variable at iteration $(n+1)$.

In the case of ambiguity due to $P\left(u_b, v_\alpha | S^{(n)}\right) = P\left(u_a, v_\alpha | S^{(n)}\right)$ in equation (6.5) (and the two former conditions hold as well), the assignment is arbitrarily chosen. In practice, this is case is of limited interest since the probability that it happens is negligible.

Starting from an initial estimate $S^{(0)}$, this process is repeated until convergence or a fixed number of times.

6.3.1 A Quality Measure for an Individual Match

Before we proceed to derive a density measurement for a match given the contextual evidence, $P\left(u_a, v_\alpha | S^{(n)}\right)$, we concentrate on the individual probability of match between two keypoints regarding their local image contents, $P_{a\alpha}$, and the threshold probability for the null-match, $P_{a\emptyset}$.

We define

$$P_{a\alpha} = \frac{\frac{1}{\|\mathbf{h}_a - \mathbf{k}_\alpha\| + \epsilon}}{\sum_{\alpha'} \frac{1}{\|\mathbf{h}_a - \mathbf{k}_{\alpha'}\| + \epsilon}} = \frac{1}{\left(\|\mathbf{h}_a - \mathbf{k}_\alpha\| + \epsilon\right) \sum_{\alpha'} \frac{1}{\|\mathbf{h}_a - \mathbf{k}_{\alpha'}\| + \epsilon}} \tag{6.6}$$

which is a quantity proportional to the inverse of the distance between their descriptors (normalized to sum up to one). We have added the small positive scalar ϵ in order to prevent a division by zero in the improbable case of identical descriptors.

Similarly, we define

$$P_{a\emptyset} = \frac{\frac{1}{\tau \left\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\right\| + \epsilon}}{\sum_{\alpha'} \frac{1}{\|\mathbf{h}_a - \mathbf{k}_{\alpha'}\| + \epsilon}} = \frac{1}{\left(\tau \left\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\right\| + \epsilon\right) \sum_{\alpha'} \frac{1}{\|\mathbf{h}_a - \mathbf{k}_{\alpha'}\| + \epsilon}} \tag{6.7}$$

which sets the threshold distance for the null-assignment to $\tau \left\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\right\| + \epsilon$.

Note that the probabilities of equations (6.6) and (6.7) define the same matching criterion as the SIFT ratio test of definition 6.3.

6.3.2 A Density Measurement Incorporating Contextual Evidence

It is now turn to derive a density measurement for the match $u_a \rightarrow v_\alpha$ given the evidence provided by the rest of the matches.

Following a similar development than Luo & Hancock (2001) we factorize the quantity $P\left(u_a, v_\alpha | S\right)$ as follows

$$P\left(u_a, v_\alpha | S\right) = \left[\frac{1}{P(u_a, v_\alpha)}\right]^{|\mathcal{I}| \times |\mathcal{J}| - 1} \prod_{b \in \mathcal{I}} \prod_{\beta \in \mathcal{J}} P\left(u_a, v_\alpha | s_{b\beta}\right) \tag{6.8}$$

where $P\left(u_a, v_\alpha | s_{b\beta}\right)$ is the conditional density of match between nodes u_a and v_α conditioned by the assignment variable $s_{b\beta}$, and $P\left(u_a, v_\alpha\right)$ is the probability of match $u_a \rightarrow v_\alpha$ regarding the nodes' attributes, i.e., $P\left(u_a, v_\alpha\right) = P_{a\alpha}$.

The key modelling ingredient in our model is the conditional density in the right hand side of equation (6.8). This quantity evaluates a matching hypothesis $u_a \rightarrow v_\alpha$ given the evidence provided by the match between nodes $u_b \in \mathcal{U}$ and $v_\beta \in \mathcal{V}$. We distinguish two different cases when modelling this quantity.

- Node u_b is assigned to node v_β, and both pairs (u_a, u_b) and (v_α, v_β) are joined by an edge (i.e., $D_{ab} = 1$ and $M_{\alpha\beta} = 1$). This means that the hypothesis $u_a \to v_\alpha$ is structurally consistent with the match $u_b \to v_\beta$. We assume that this case will be given with a probability proportional to the individual probabilities of match $P_{a\alpha}, P_{b\beta}$ according to the local image contents.

- Otherwise, node u_b does not provide contextual support and hence its contribution is reduced to the probability of null match $P_{b\emptyset}$.

More formally,

$$P\left(u_a, v_\alpha | s_{b\beta}\right) = \left\{ \begin{array}{ll} P_{a\alpha} P_{b\beta} & \text{if } \left(D_{ab} = 1 \wedge M_{\alpha\beta} = 1\right) \wedge s_{b\beta} = 1 \\ P_{a\alpha} P_{b\emptyset} & \text{otherwise} \end{array} \right. \tag{6.9}$$

Note that the density measurement defined in equation (6.9) can be equivalently expressed as

$$P\left(u_a, v_\alpha | s_{b\beta}\right) = \left(P_{a\alpha} P_{b\beta}\right)^{D_{ab} M_{\alpha\beta} s_{b\beta}} \left(P_{a\alpha} P_{b\emptyset}\right)^{1 - D_{ab} M_{\alpha\beta} s_{b\beta}} \tag{6.10}$$

Using the expression for the density measurement of equation (6.10), the conditional density of equation (6.8) expressed in exponential form has the following expression

$$P\left(u_a, v_\alpha | S\right) =$$

$$= \left[\frac{1}{P_{a\alpha}}\right]^{|\mathcal{I}| \times |\mathcal{J}| - 1} \exp\left\{ \sum_{b \in \mathcal{I}} \sum_{\beta \in \mathcal{J}} \left[D_{ab} M_{\alpha\beta} s_{b\beta} ln\left(\frac{P_{b\beta}}{P_{b\emptyset}}\right) + ln\left(P_{a\alpha} P_{b\emptyset}\right) \right] \right\}$$

$$= \left[\frac{1}{P_{a\alpha}}\right]^{|\mathcal{I}| \times |\mathcal{J}| - 1} P_{a\alpha}^{|\mathcal{I}| \times |\mathcal{J}|} \prod_{b \in \mathcal{I}} P_{b\emptyset}^{|\mathcal{J}|} \exp\left[\sum_{b,\beta} ln\left(\frac{P_{b\beta}}{P_{b\emptyset}}\right) D_{ab} M_{\alpha\beta} s_{b\beta} \right] \tag{6.11}$$

$$\propto P_{a\alpha} \exp\left[\sum_{b,\beta} ln\left(\frac{P_{b\beta}}{P_{b\emptyset}}\right) D_{ab} M_{\alpha\beta} s_{b\beta} \right] \tag{6.12}$$

In going from equation (6.11) to (6.12) we have eliminated the term $\prod_b {P_{b\emptyset}}^{|\mathcal{J}|}$ which is constant with respect to u_a, v_α and S.

Note that there are two terms involved in equation (6.12). One accounts for the quality of the match that we are evaluating, $P_{a\alpha}$, and the other accounts for the quality of the matches $u_b \to v_\beta$ that are structurally consistent with $u_a \to v_\alpha$.

Finally, we consider that the *null* node is not joined with and edge to any other node. Therefore, from equation (6.12), we define the conditional measurement for matching node u_a to *null* as

$$P\left(u_a, \emptyset | S\right) = P_{a\emptyset} \exp\left(0\right) = P_{a\emptyset} \tag{6.13}$$

where the term $\exp\left(0\right)$ originates from considering that the null node has no edges at all.

The algorithm iteratively updates the assignment variable S as stated in equation (6.5). This is, at each iteration, each node $u_a \in \mathcal{U}$ is assigned to the node $v_\alpha \in \mathcal{V}$ with the highest conditional density according to equation (6.12).

According to equation (6.13), if $\max\limits_{\alpha} P\left(u_a, v_\alpha | S\right) < P_{a\emptyset}$, then node u_a is leaved unmatched.

6.4 A Continuous Labeling Approach

Graduated Assignment (Gold & Rangarajan, 1996) is a well-known optimization algorithm that has been widely used to solve the graph matching problem. It estimates the assignment variable S that minimizes the following function originated from the relaxation labeling processes (Hummel & Zucker, 1983; Rosenfeld *et al.*, 1976).

$$\mathcal{F}_{ga} = -\frac{1}{2} \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} \sum_{b \in \mathcal{I}} \sum_{\beta \in \mathcal{J}} s_{a\alpha} s_{b\beta} Q_{a\alpha b\beta} \tag{6.14}$$

subject to

$$\forall a \sum_{\alpha \in \mathcal{J}} s_{a\alpha} \leq 1, \ \forall \alpha \sum_{a \in \mathcal{I}} s_{a\alpha} \leq 1, \ \forall a, \alpha \ s_{a\alpha} \in \{0, 1\}$$

where $Q_{a\alpha b\beta}$ is the *compatibility coefficient* conveying the compatibility of the edge-match $(u_a, u_b) \rightarrow (v_\alpha, v_\beta)$.

Gold & Rangarajan (1996) have turned this minimization into an iterative assignment problem where matrix S is relaxed to a double stochastic matrix which is updated according to the following expression

$$S^{(n+1)} = \arg\max_{S} \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} B_{a\alpha} s_{a\alpha} \tag{6.15}$$

where $B_{a\alpha}$ is the support for the match $u_a \rightarrow v_\alpha$ received from the rest of the matches, which is equal to

$$B_{a\alpha} = \sum_{b \in \mathcal{I}} \sum_{\beta \in \mathcal{J}} s_{b\beta} Q_{a\alpha b\beta} \tag{6.16}$$

The assignment problem of equation (6.15) is solved in a continuous (soft) way using a continuation method controlled by a parameter to gradually push from continuous to discrete solutions.

In section 4.3.2 we present more details about the algorithm.

6.4.1 A Quality Measure for an Individual Match

Before we proceed to develop an expression for the compatibility coefficients $Q_{a\alpha b\beta}$, we start by defining a quality measure for an individual match $u_a \rightarrow v_\alpha$ regarding the local image contents encoded by the SIFT descriptors. To that end, we draw on the quantity $P_{a\alpha}/P_{a\emptyset}$ resulting from the previous section. This quantity has the useful property of being > 1 if the match is more likely than the null-match (i.e., $P_{a\alpha} > P_{a\emptyset}$), and ≤ 1 otherwise.

Instead of using the normalized inverse distance of equations (6.6) and (6.7), we will use the exponential of the negative distance. This is,

$$P'_{a\alpha} = \exp\Big[-\|\mathbf{h}_a - \mathbf{k}_\alpha\| \Big] \tag{6.17}$$

$$P'_{a\emptyset} = \exp\Big[-\tau \left\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\right\| \Big] \tag{6.18}$$

Graduated Assignment imposes two-way constraints through Sinkhorn normalization (Sinkhorn, 1964). This means that matches are evaluated in both directions, this is, from **G** to **H** and from **H** to **G**. In order to take advantage from this fact, we propose the following *bidirectional* measure of the quality of an individual match

$$P''_{a\alpha} = \frac{P'_{a\alpha}}{P'_{a\emptyset}} \cdot \frac{P'_{\alpha a}}{P'_{\alpha\emptyset}} \tag{6.19}$$

$$= \frac{\exp\Big[-2\,\|\mathbf{h}_a - \mathbf{k}_\alpha\|\Big]}{\exp\Big[-\tau\,\big\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\big\|\Big]\exp\Big[-\tau\,\big\|\mathbf{h}_{g(\alpha)} - \mathbf{k}_\alpha\big\|\Big]} \tag{6.20}$$

$$= \exp\Big[\tau\Big(\big\|\mathbf{h}_a - \mathbf{k}_{\gamma(a)}\big\| + \big\|\mathbf{h}_{g(\alpha)} - \mathbf{k}_\alpha\big\|\Big) - 2\,\|\mathbf{h}_a - \mathbf{k}_\alpha\|\Big] \tag{6.21}$$

where functions $\gamma(a) : \mathcal{I} \to \mathcal{J}$ and $g(\alpha) : \mathcal{J} \to \mathcal{I}$ are explained in definition 6.2 and we have substituted the matching functionals by their expressions of equations (6.17) and (6.18) in going from equation (6.19) to (6.20).

The quantity defined in equation (6.21) holds the useful property of being $P''_{a\alpha} > 1$ if a match is consistent in both directions (i.e., $P'_{a\alpha} > P'_{a\emptyset}$ and $P'_{\alpha a} > P'_{\alpha\emptyset}$).

6.4.2 The Support for a Match Regarding the Context

It is now turn to propose a support measure for a match that incorporates contextual evidence. We consider that a candidate association $u_a \to v_\alpha$ with a high functional regarding the local information (i.e., $P''_{a\alpha} > 1$) but with low support from the context is likely to be an outlier. On the other hand, a candidate association with a not-enough-high local probability (i.e., $P''_{a\alpha} < 1$) but with high support from the surrounding matches, is likely to be a valid match.

After trying various expressions, we propose the following support function which reflects the desired behaviour

$$E_{a\alpha} = P''_{a\alpha} + P''_{a\alpha}\left[\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}} ln\left(P''_{b\beta}\right) D_{ab} M_{\alpha\beta} s_{b\beta}\right] \tag{6.22}$$

This measure is composed by a sum of two quantities. The first quantity reflects the quality of the match regarding the local image information, $P''_{a\alpha}$. The second quantity reflects the support received from the rest of the matches properly scaled by the probability of match regarding the nodes' attributes.

Consider the case of a candidate association with a high local and a low contextual consistency. Despite of the high quantity of the first part of $E_{a\alpha}$, the negative contribution of the second part would smooth the overall measure. In the case of a candidate association with a not-enough-high local and a high contextual consistency, the positive contribution of the second part would boost the overall measure.

In order to embed the proposed support function $E_{a\alpha}$ within the framework provided by Graduated Assignment, we have to set an expression for the compatibility coefficients $Q_{a\alpha b\beta}$ of equation (6.16) so that $B_{a\alpha} \equiv E_{a\alpha}$.

Finally, we rearrange our support function in the following way

$$E_{a\alpha} = \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta} \left[P''_{a\alpha} \left(ln \left(P''_{b\beta} \right) D_{ab} M_{\alpha\beta} + \frac{1}{N} \right) \right] \tag{6.23}$$

where the expression within square brackets corresponds to the compatibility coefficients $Q_{a\alpha b\beta}$ of equation (6.16), and $N = \sum_{b,\beta} s_{b\beta}$ is a constant (for a fixed S) depending on the number of nodes in the graphs.

We address the null-matching process by means of the *slack* variables of the Graduated Assignment algorithm. We set the value of the slack variables to *unity* which is the thresholding value for the support function $E_{a\alpha}$ in order to consider the match $u_a \rightarrow v_\alpha$ as an outlier. When performing Sinkhorn normalization we have into account that the slack variables are special cases that allow for multiple assignments (i.e., multiple nodes in both graphs can be assigned to null).

6.5 Experiments and Results

We have compared both the discrete and continuous labelling graph matching approaches of the present chapter (*DLGM* and *CLGM*) to the following approaches: the original *SIFT* method; the outlier rejectors Graph Transformation Matching (*GTM*) by Aguilar *et al.* (2009) and *RANSAC* by Fischler & Bolles (1981) explained in sections 4.3.4 and 2.4, respectively; the unified approach to graph matching (*Unified*) by Luo & Hancock (2003) explained in section 4.5.2; and finally the point-set registration method *Robust Point Matching* (*RPM*) by Rangarajan *et al.* (1997) explained in section 3.4.

We have evaluated the matching Precision and Recall scores of each method under the following types of perturbations: photometric distortions, geometrical noise and clutter (point contamination). We have used the *F-measure* to plot the results. The F-measure is defined as the weighted harmonic mean of Precision and Recall and has the following expression

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6.24}$$

The graphs used in the proposed methods (*DLGM* and *CLGM*) follow the representation of definition 6.1. The structures of the graphs for all the methods using them (i.e., *DLGM*, *CLGM*, *GTM* and *Unified*) have been generated using a *mutual K-nearest-neighbours* approach with $K = 5$ (i.e., two keypoints are joined with an edge if both of them are within the 5 nearest neighbours of each other).

The mean size of the keypoint-sets in the experiments without point contamination has been around 150 points. In the point contamination experiments the size of the point-sets increases according to the fraction of added points with respect to the original ~ 150 points.

Our continuous labelling approach (*CLGM*) has been initialized with an equiprobable assignment matrix. The rest of the methods have been initialized with the matches returned by a classical SIFT matching using a ratio $\tau_0 = 0.8$ which is the value suggested by Lowe (2004).

For a fixed value of τ, we have found that the proposed approaches are more prone to send keypoints to null than the classical SIFT method. This is because the methods *DLGM* and *CLGM* introduce the contextual constraint additionally to the individual matching constraint of the original SIFT matching. Therefore, we have relaxed the ratio value for the proposed methods (*DLGM* and *CLGM*) to $\tau = 1$ in the present experiments.

For each experiment we have arbitrarly chosen a grayscale image I_M from the Camera Movements and Deformable Objects' databases used by Aguilar *et al.* (2009).

In the photometric distortion experiments, we generate the data-image I_D by simultaneously applying the following types of perturbations to the model-image I_M: image resizing (to simulate changes in the distance from the objects in the image), image rotation (to simulate changes in viewpoint), image intensity adjustment (to simulate illumination changes), and Gaussian white noise addition to pixel intensity values (to simulate deterioration in the viewing conditions).

We extract the SIFT keypoints from images I_D and I_M, obtaining coordinate vector-sets $\mathcal{X}$ and $\mathcal{Y}$, and SIFT descriptor-sets $\mathcal{H}$ and $\mathcal{K}$, respectively. We define $\widetilde{\mathcal{X}}$ as the result of the mapping from points in $\mathcal{X}$ back to the reference of I_M. We compute $\widetilde{\mathcal{X}}$ by applying to $\mathcal{X}$ the inverse resizing and rotation from the perturbation. We set the ground truth assignments on the basis of the proximity between the points in $\mathcal{Y}$ and $\widetilde{\mathcal{X}}$. Then, for a given $\mathbf{y}_\alpha \in \mathcal{Y}$, we select as its ground truth assignment the most salient $\widetilde{\mathbf{x}}_a \in \widetilde{\mathcal{X}}$ among the ones falling inside a certain radius r from $\mathbf{y}_\alpha$. Saliency is decided according to the gradient magnitude of the SIFT features (Lowe, 2004). The proximity radius has been set to $r = 0.03 \times l$, where l is the diagonal-length of the image. The keypoints which are not involved in any ground truth assignment are discarded. So, at the end of this step we end up with keypoint-sets $\mathcal{Y}' = \{\mathbf{y}'_1, \ldots, \mathbf{y}'_N\}$ and $\mathcal{X}' = \{\mathbf{x}'_1, \ldots, \mathbf{x}'_N\}$, and a bijective mapping $f_{gtr} : \mathcal{X}' \to \mathcal{Y}'$ of ground truth assignments.

Once the N ground truth assignments have been established, we implement the clutter by adding a certain amount of the remaining points in both $\mathcal{X}$ and $\mathcal{Y}$ to $\mathcal{X}'$ and $\mathcal{Y}'$. Clutter points are carefully selected not to fall inside the radius of proximity r of any pre-existent point. Thus, we can safely assume that they have no correspondence in the other point-set.

Finally, geometrical noise consists on adding random Gaussian noise with zero mean and a certain standard deviation σ to the point positions $\mathbf{x}_i = (x^V, x^H)$. This type of noise simulates nonrigid deformations in the position of the features.

We average the experiments over 10 images. Due to the random nature of the noise, we have run 10 experiments for each image. So, each location in the plot is the average of 100 experiments.

Figure 6.1 shows the F-measure plots for each method for an increasing amount of photometric distortions. Both geometrical noise and clutter have been set to zero.

Figure 6.2 shows the results for an increasing number of clutter points. The amount of point contamination has ranged from 0% to 80% of the total N points.

Figure 6.3 shows the results for geometrical noise with σ ranging from 0% to 15% of μ (where μ is the mean of the pairwise distances between the keypoints).

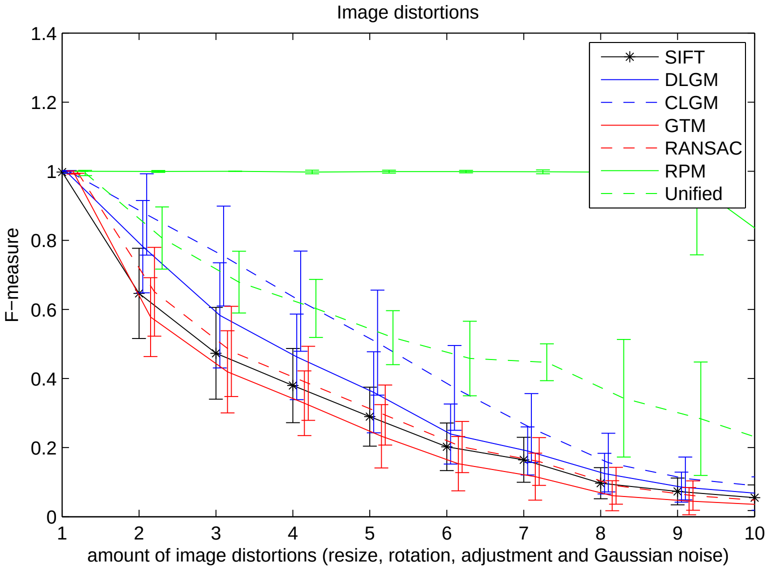


Figure 6.1: Photometric distortions.

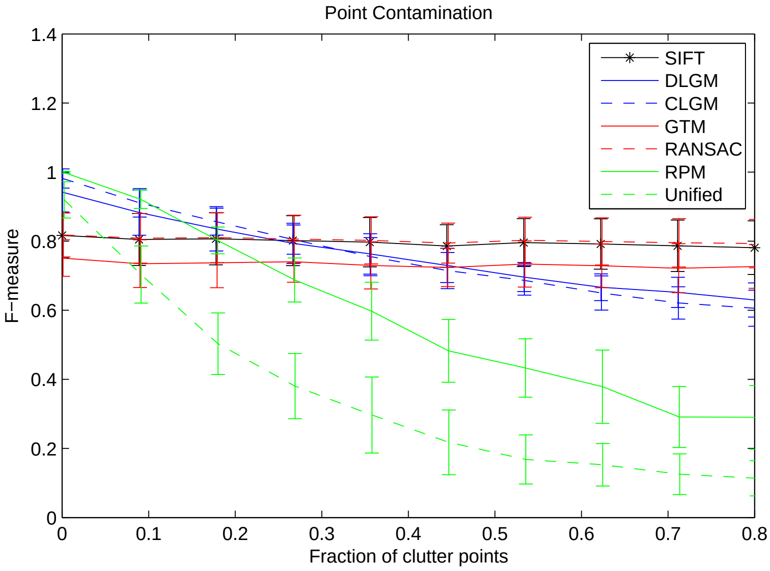


Figure 6.2: Point contamination.

A ground-level amount of photometric distortions have been introduced during the clutter and geometrical noise experiments.

In the photometric distortion experiments, the methods *RPM* and *Unified* have shown the best performance. Both methods use photometric information only to compute the initial matches. *RPM* has shown the best ability to re-

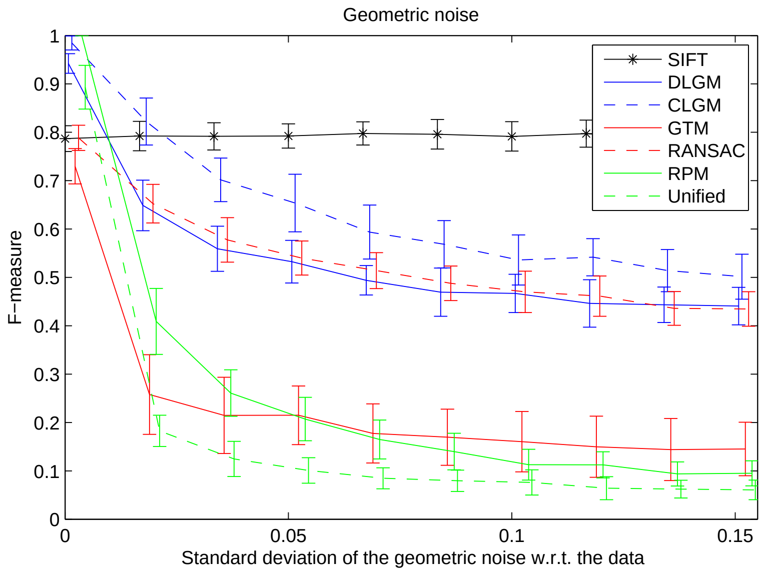


Figure 6.3: Geometric noise.

cover from poor initializations (i.e., high photometric distortion). Although outlier rejectors do not use use photometric information during their optimization processes, they have shown a poor response to photometric noise. This makes evident the inability of the outlier rejectors to recover from bad initial conditions. The proposed methods *DLGM* and *CLGM* present a performance below *RPM* but significantly above *SIFT* matching and the outlier rejectors. The comparison between *SIFT* and the proposed methods elucidates the benefits of incorporating contextual evidence in the case of pure photometric distortions.

In the geometric noise experiments, methods not using photometric evidence have shown the worst performance. Specially, the pure geometric / structural methods *Unified*, *RPM* and *GTM*. The *SIFT* matching method has shown the best performance since it relies solely on the photometric information which has only been disturbed by a low ground-level amount. The proposed methods *DLGM* and *CLGM* present significantly better robustness to geometric distortions than the rest of the methods relying on geometric / structural information. Specifically, the continuous approach performs better than the discrete one, which performs similarly to *RANSAC*.

In the point contamination experiments, both the *SIFT* matching and the outlier rejectors have shown the best performance. Methods relying solely on structural / geometric information (i.e., *RPM* and *Unified*) are the most affected by this type of noise. The proposed methods have shown an intermediate performance.

6.6 Conclusions

We have presented a continuous and a discrete graph matching approach for association of SIFT features. Among their main features we stress that both local and contextual evidence are used during the optimization process. We have presented synthetic experiments evaluating their robustness to different types of deformations.

The proposed methods have presented an intermediate performance in all the experiments, thus representing a compromise between the methods exclusively driven by the local image information and the ones relying on structural / geometrical information. Specifically, the continuous labelling approach using a bidirectional matching functional has overcome the discrete labelling approach in most of the experiments presented.

Chapter 7

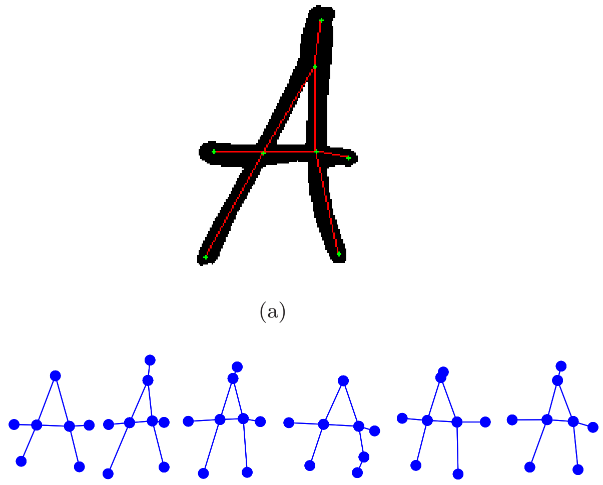
Improving the Matching of Sparse Graphs by Introducing Procrustes Distances into a Dictionary-based Structural Model

7.1 Introduction

Graph matching aims to associate the nodes between two graphs so that the structure is preserved. Early attempts tried to match graphs in an *exact* way, this is, if two nodes of one graph were linked by an edge they had to be matched to two nodes linked by an edge in the other graph, and vice-versa (Ghahraman *et al.*, 1980; Ullmann, 1976).

Later on, graph matching was faced as an energy minimization problem were a cost function evaluated the plausibility of each configuration of matches. One advantage of this approach is that it is possible to match graphs in an *inexact* way by accommodating some deformations. Moreover, we can introduce prior knowledge about these deformations by assigning higher costs to the less likely deformations.

This has many advantages in recognition applications where some variability may be found in the data either due to errors in the feature extraction process or due to natural variations between objects from the same class (see figure 7.1 for an example). In the case of handwritten character recognition, it is a natural approach to extract the graph representations from the medial axis or *skeleton* (Blum, 1967). In this way, nodes can be placed at the end, intersection and high-curvature points, while edges may represent the body of the skeleton. See figure 7.1 for an example of graph extraction from an image of a handwritten character and several other instances from the same class.



(a)

(b)

Figure 7.1: (a) example of a graph extraction from a handwritten character and, (b) several graphs from the same class.

It is a usual strategy among the probabilistic-relaxation-based approaches (Gold & Rangarajan, 1996; Hummel & Zucker, 1983; Luo & Hancock, 2003, 2001; Rosenfeld *et al.*, 1976) to evaluate the likelihood of a given correspondence on the basis of the support received from its context. This support may be interpreted as the amount of corresponding edges incident upon the nodes involved in that correspondence. This is associated to a well-founded measure of structural consistency (Hummel & Zucker, 1983). Nevertheless, in the case of sparse graphs such as the ones extracted from skeletal representations this measure may lead to ambiguities due to the lack of structural evidence.

In this direction, Wilson & Hancock (1997) proposed a more descriptive support accounting for the orientation ordering of the edges incident upon each node. They used a graph sub-entity called *clique* which is a string composed of a central node and all its adjacent nodes disposed in the order established by their relative orientations. Then, structural consistency was measured by means of the *Hamming distance* between cliques in the data and the model graphs. They used a discrete relaxation scheme in order to update the matches following the Maximum-A-Posteriori (MAP) rule. Later on, Cross (1997) used hybrid genetic search to update the matches according to the same model.

Despite the structural representativeness of the model proposed by Wilson & Hancock (1997), it may still present some ambiguities. Consider for example the three different mappings in figure 7.2 between two structures which may be found in the types of graphs addressed in this chapter. Since the orientation ordering around the central nodes is preserved by the mapping, these three matching configurations are equally probable according to the model by Wilson & Hancock (1997). Clearly, the option represented in figure 7.2.(c) is the most appropriate since this is the one that most preserves the spatial arrangement of the nodes.

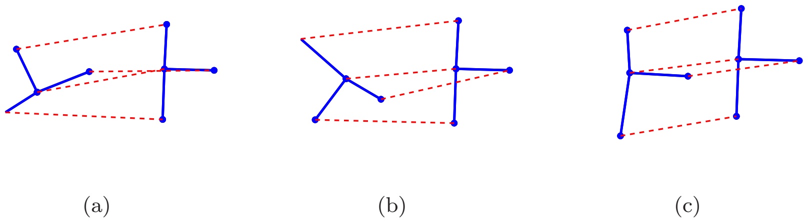


Figure 7.2: Three equally probable mappings from the data-graph clique (on the left) to the model graph-clique (on the right). Cliques are induced by the central nodes.

In order to overcome this ambiguity, we present an approach that accounts for the alignment errors at the clique level by introducing the *Procrustes distance* (Dryden & Mardia, 1998). Our approach leads to a unified model that combines, in a principled way, structural and geometric evidence through the use of Hamming and Procrustes distances, respectively.

The overview of this chapter is the following. In section 7.2 we review the structural model based on cliques by Wilson & Hancock (1997). In section 7.3 we present our unified approach. In sections 7.4 and 7.5 we introduce optimization strategies based on discrete relaxation and hybrid genetic search, respectively. In section 7.6 we provide experimental evidence that our combined approach outperforms the original cliques model when matching the aforementioned types of graphs. We embed the proposed model into discrete relaxation and hybrid genetic search schemes and compare to the original methods.

7.2 A Model of Structural Consistency

Let us denote two graphs with the tuples $\mathbf{G} = (\mathcal{U}, D)$ and $\mathbf{H} = (\mathcal{V}, M)$, where $\mathcal{U} = \{u_a,\ a \in \mathcal{I}\}$ and $\mathcal{V} = \{v_\alpha,\ \alpha \in \mathcal{J}\}$ are the node sets being $\mathcal{I} = 1, \ldots, |\mathcal{U}|$ and $\mathcal{J} = 1, \ldots, |\mathcal{V}|$ their index-sets, and D, M are the $|\mathcal{I}| \times |\mathcal{I}|$ and $|\mathcal{J}| \times |\mathcal{J}|$ adjacency matrices accounting for the binary measurements between each pair of nodes, respectively.

We will consider *undirected unweighted* adjacency matrices of the form

$$D_{ab} = \begin{cases} 1 & \text{if } u_a \text{ and } u_b \text{ are linked by an edge} \\ 0 & \text{otherwise} \end{cases}$$

(the same applies for $M_{\alpha\beta}$).

In some applications, structural information can be inferred from some similarity measurement between nodes. In other applications however, it arises naturally from the data representations such as in the case of the medial axis representations.

Traditionally, probabilistic-relaxation-related approaches (Gold & Rangarajan, 1996; Hummel & Zucker, 1983; Rosenfeld *et al.*, 1976) consider supports for a node proportional to the sum of consistently matched edges incident upon it. In other words, a correspondence $u_a \to v_\alpha$ is more likely to occur as more nodes u_b adjacent to u_a (i.e., $D_{ab} = 1$) correspond to nodes v_β adjacent to v_α.

This approach has demonstrated to be successful in many applications. Nevertheless, in the case of graphs derived from skeletal representations where the adjacency matrices are highly sparse this approach may lead to ambiguities.

Wilson & Hancock (1997) devised an approach aimed at exploiting the ordering of the spatial orientations of the incident edges. Accordingly, in order for a match to be feasible the orientation ordering of the matched incident edges must be preserved.

To that end, they devised the *clique* ($\mathfrak{C}_a$), a new structural subunit associated with each node u_a consisting of an ordered string of indices from its adjacent nodes u_b (similarily for a clique $\mathfrak{R}_\alpha$ in the graph $\mathbf{H}$). This is,

$$\mathfrak{C}_a = (b_1, \ldots, b_p)\,, \text{ s.t. } D_{ab_1} = 1, \ldots, D_{ab_p} = 1 \tag{7.1}$$

$$\mathfrak{R}_\alpha = (\beta_1, \ldots, \beta_q)\,, \text{ s.t. } M_{\alpha\beta_1} = 1, \ldots, M_{\alpha\beta_q} = 1 \tag{7.2}$$

Consider the matching function $f : \mathcal{I} \to \mathcal{J}$ that assigns nodes in $\mathcal{U}$ from graph $\mathbf{G}$ to nodes in $\mathcal{V}$ from graph $\mathbf{H}$. When evaluating a matching f, constraints in the data-graph $\mathbf{G}$ are evaluated in the model graph $\mathbf{H}$. This means that cliques in the data-graph must be matched to valid cliques in the model-graph.

The matching realization of a data-graph clique $\mathfrak{C}_a$ onto the model graph is denoted as $\Gamma_a = (f(b_1), \ldots, f(b_p))$.

The key modelling ingredient in the approach by Wilson & Hancock (1997) was developing a model for the matching prior $P(f)$. This was approximated by the average of the consistencies of each data-graph clique. At its turn, each data-graph clique was evaluated over a dictionary of *structure-preserving mappings* (SPM) Ω, consisting on all the orientation-ordering-preserving mappings on the model-graph onto which each data-graph clique could be mapped. Finally, the symbols in the matching realization of the data-graph clique were independently compared to the symbols in the SPM. This is,

$$P(\Gamma_a|\Omega_i) = \prod_{k=1}^{|\mathfrak{C}_a|} P(f(b_k)|\beta_k) \tag{7.3}$$

where b_k is the k-th symbol in the data-graph clique $\mathfrak{C}_a$ and β_k is the k-th symbol in the i-th entry of the dictionary of SPMs, Ω_i.

See section 4.4.1 for a detailed development of this model.

In their symbolic model, Wilson & Hancock (1997) assumed that structural errors occur with a certain low probability of error P_e. They also introduced a uniform probability $P_\emptyset$ associated with the node-deletion hypothesis. Since it is not our intention to allow for graph-edit operations in our approach we will circumvent this quantity.

With these ingredients, the probability of a mapping from a symbol in a data-clique given its corresponding symbol in a SPM is

$$P(f(b_k)|\beta_k) = \begin{cases} (1 - P_e) & \text{if } f(b_k) = \beta_k \\ P_e & \text{if } f(b_k) \neq \beta_k \end{cases} \tag{7.4}$$

By using the above conditional density, the final expression for the joint prior $P(f)$ according to Wilson & Hancock (1997) expressed in the exponential form is the following

$$P\left(f\right) = \frac{1}{|\mathcal{I}|} \sum_{a \in \mathcal{I}} \frac{K_{\mathfrak{C}_a}}{|\Omega|} \sum_{i=1}^{|\Omega|} \exp\left[- ln\left(\frac{1-P_e}{P_e}\right) d_H\left(\Gamma_a, \Omega_i\right)\right] \tag{7.5}$$

where equiprobable priors have been assumed for each SPM (i.e., $P\left(\Omega_i\right) = \frac{1}{|\Omega|}$), $K_{\mathfrak{C}_a} = \left(1 - P_e\right)^{|\mathfrak{C}_a|}$, and $d_H\left(\Gamma_a, \Omega_i\right)$ is the Hamming distance between the matching realization of the data-clique Γ_a and the SPM Ω_i. This latter quantity conveys information about the consistence of the matching realization of a data-clique given a SPM. The lower the Hamming distance is, the more feasible the mapping is.

Despite its representativeness, this model may lead to some ambiguities as seen in figure 7.2.

In order to overcome this problem, in the next section we present an alternative measurement that replaces the fixed-probability density function of equation (7.4) by a more fine-grained measure accounting for the alignment errors of the consistently matched nodes. Our approach leads to a unified model that gauges structural and spatial consistency through the use of Hamming and Procrustes distances, respectively.

7.3 A Unified Model of Structural and Geometric Consistency

As said, we want to extend the structural model reported by Wilson & Hancock (1997) in order to assess for the spatial consistency of the matched points at the level of each clique. We develop an inexact model allowing for certain alignment errors through the use of Procrustes distances within a probabilistic setting.

Let us augment the graphs $\mathbf{G} = (\mathcal{U}, D)$ and $\mathbf{H} = (\mathcal{V}, M)$ with the point-sets $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$, where $\mathbf{x}_a = \{x_a^{V}, x_a^{H}\}$ and $\mathbf{y}_\alpha = \{y_\alpha^{V}, y_\alpha^{H}\}$ correspond to the 2D position coordinates of nodes u_a and v_α, respectively.

According to the aims of Procrustes analysis, all the geometric measurements are done in a similarity-invariant fashion. To that end, the conditional probability of the matching realization of a data-graph clique Γ_a given a SPM Ω_i of equation (7.3) is taken over the similarity alignment parameters Φ_{ai} in the following way,

$$P\left(\Gamma_a | \Omega_i\right) = \max_{\Phi_{ai}} \prod_{k=1}^{|\mathfrak{C}_a|} P\left(f\left(b_k\right) | \beta_k, \Phi_{ai}\right) \tag{7.6}$$

Our key idea is to replace the fixed probability $(1 - P_e)$ of equation (7.4) with a more fine-grained measure that accounts for the alignment accuracy as well as allowing for certain mis-alignment errors. Hence, we write,

$$P\left(f\left(b_k\right) | \beta_k, \Phi_{ai}\right) = \begin{cases} P_{b_k,\beta_k}^{(\Phi_{ai})} & \text{if } f\left(b_k\right) = \beta_k \\ \rho & \text{if } f\left(b_k\right) \neq \beta_k \end{cases} \tag{7.7}$$

where $P_{b_k,\beta_k}^{(\Phi_{ai})}$ is a probability measurement on the alignment errors according to parameters Φ_{ai}, and ρ is the threshold alignment probability for a match $b_k \rightarrow \beta_k$ to be considered feasible.

We model the alignment probabilities as an exponential of the negative alignment errors properly scaled by a variance parameter. This is,

$$P^{(\Phi_{ai})}_{b_k,\beta_k} = \exp\left[-\frac{\left\|\mathbf{x}_{b_k} - \mathcal{T}\left(\mathbf{y}_{\beta_k};\Phi_{ai}\right)\right\|^2}{2\sigma^2}\right] \tag{7.8}$$

where σ^2 is the expected variance of the alignment errors between corresponding points within the cliques, $\mathcal{T}\left(\mathbf{y}_{\beta_k};\Phi_{ai}\right)$ represents the geometric transformation of point $\mathbf{y}_{\beta_k}$ according to similarity-alignment parameters Φ_{ai}, and $\left\|\cdot\right\|^2$ is the squared Euclidean norm.

The constant ρ of equation (7.7) can be modeled as the exponential of a negative thresholding error $\rho = \exp\left[-\frac{1}{2\sigma^2}\left\|\mathbf{d}\right\|^2\right]$, where $\mathbf{d} = \left(d^{V}, d^{H}\right)$ are the thresholding vertical and horizontal errors, respectively.

With these ingredients, the conditional probability of equation (7.6) can be expressed as

$$P\left(\Gamma_a|\Omega_i\right) = \max_{\Phi_{ai}}\left\{\rho^{d_H\left(\Gamma_a,\Omega_i\right)}\prod_{k|f(b_k)=\beta_k}\exp\left[-\frac{\left\|\mathbf{x}_{b_k}-\mathcal{T}\left(\mathbf{y}_{\beta_k};\Phi_{ai}\right)\right\|^2}{2\sigma^2}\right]\right\} \tag{7.9}$$

$$= \max_{\Phi_{ai}}\exp\left[-ln\left(\tfrac{1}{\rho}\right)d_H\left(\Gamma_a,\Omega_i\right) - \sum_{k|f(b_k)=\beta_k}\frac{\left\|\mathbf{x}_{b_k}-\mathcal{T}\left(\mathbf{y}_{\beta_k};\Phi_{ai}\right)\right\|^2}{2\sigma^2}\right] \tag{7.10}$$

$$= \exp\left[-ln\left(\tfrac{1}{\rho}\right)d_H\left(\Gamma_a,\Omega_i\right) - \frac{1}{2\sigma^2}\left(\min_{\mathtt{R},\eta,\mathbf{t}}\sum_{k|f(b_k)=\beta_k}\left\|\mathbf{x}_{b_k} - \left(\eta\mathtt{R}\mathbf{y}_{\beta_k}+\mathbf{t}\right)\right\|^2\right)\right] \tag{7.11}$$

$$= \exp\left[-ln\left(\tfrac{1}{\rho}\right)d_H\left(\Gamma_a,\Omega_i\right) - \frac{1}{2\sigma^2}d_P^2\left(\hat{\mathcal{X}},\hat{\mathcal{Y}}\right)\right] \tag{7.12}$$

Equation (7.9) merges equations (7.6) and (7.7) into a single expression accounting for the number of structural inconsistencies (through the Hamming distance) and the alignment errors of the structurally consistent mappings between Γ_a and Ω_i. Equation (7.10) uses the exponential form to express equation (7.9). In going from equation (7.10) to equation (7.11) we have focused on the part of the expression affected by the optimization and we have particularized to the case of similarity transformations by introducing the parameters $\mathtt{R},\eta,\mathbf{t}$. In going from equation (7.11) to equation (7.12) we have used the definition of Procrustes distance (section 2.3.1).

Equation (7.12) reveals that structural and spatial consistency are gauged through two well-known distance measures, namely, the Hamming distance $d_H\left(\Gamma_a,\Omega_i\right)$ and the squared Procrustes distance $d_P^2\left(\hat{\mathcal{X}},\hat{\mathcal{Y}}\right)$ between the two corresponding point-sets $\hat{\mathcal{X}} = \left\{\mathbf{x}_{b_k}|f\left(b_k\right)=\beta_k\right\}$ and $\hat{\mathcal{Y}} = \left\{\mathbf{y}_{\beta_k}|f\left(b_k\right)=\beta_k\right\}$ composed by the structurally consistent mapped points from the data-clique Γ_a and the SPM Ω_i.

The final expression for the joint prior by Wilson & Hancock (1997) of equation (7.5) according to the conditional probability proposed in equation (7.12) is the following

$$P\left(f\right) = \frac{1}{|\mathcal{I}|}\sum_{a\in\mathcal{I}}\frac{1}{|\Omega|}\sum_{i=1}^{|\Omega|}\exp\left[-ln\left(\tfrac{1}{\rho}\right)d_H\left(\Gamma_a,\Omega_i\right) - \frac{1}{2\sigma^2}d_P^2\left(\hat{\mathcal{X}},\hat{\mathcal{Y}}\right)\right] \tag{7.13}$$

7.4 Graph Matching by Discrete Relaxation

Wilson & Hancock (1997) devised a discrete labelling approach to update the matches according to the MAP rule (check section 4.4.1 for more details).

In short, this summarizes to the following update rule

$$f(a) = \arg\max_{\alpha} P(\mathbf{x}_a, \mathbf{y}_\alpha | f(a) = \alpha) P(f) \tag{7.14}$$

where $P(\mathbf{x}_a, \mathbf{y}_\alpha | f(a) = \alpha)$ is the probability of the different pairs of unary measurements assuming that they are conditionally independent of one-another given the current state of match.

This independence assumption does not hold for our model since we assume dependence among the position coordinate measurements of the nodes constituting a clique. Moreover, the original approach (Wilson & Hancock, 1997) does not address the details of the computation of this quantity. It is for these reasons that we will not use the quantity $P(\mathbf{x}_a, \mathbf{y}_\alpha | f(a) = \alpha)$ and we will rely solely on the matching priors $P(f)$ in order to perform the iterative updates in both the case of the original approach of equation (7.5) and the case of the proposed method of equation (7.13).

7.5 Graph Matching by Hybrid Genetic Search

Objective functions related to the graph matching problem usually present very irregular landscapes. This makes gradient ascent approaches such as the one presented in the previous section potentially able of getting trapped in local optima.

Genetic Algorithms (GA) are optimization processes effectively capable of locating global optimal solutions (Fogel, 1994). A population of individuals is evolved in GA by the means of *mutation* and *crossover* operations in a manner similar than the natural evolution does. New populations of solutions are decided according to a *fitness function* that measures the adaptation of each individual.

Hybrid GA distinguishes from GA in that it integrates other optimization methods within the evolutionary procedure. It may consist in applying a gradient ascent step after the mutation and crossover operators, before the selection of the next generation of individuals.

See section 4.6 for a more detailed explanation of this process.

Cross (1997); Cross *et al.* (2000) implement a hybrid GA by using the matching prior of equation (7.5) as fitness function and the discrete relaxation scheme reported by Wilson & Hancock (1997) (section 7.4) as gradient ascent step.

We propose a similar approach by using the matching prior of equation (7.13) as fitness function and our own version of the discrete relaxation scheme of the previous section as gradient ascent step.

7.6 Experiments and Results

We have performed matching experiments with a database of handwritten capital letters. This database is composed of 84 graphs from different letters. The mean number of nodes and arcs in the graphs are 5.8 ± 2.1 and 4.9 ± 2.2, respectively. Figure 7.3 shows a set of sample graphs from different classes in this database.

The data graphs of each class are matched against a prototypical graph of that class. Both graph prototypes and ground truth matches have been manually set.

We evaluate the matching accuracy of our model presented in section 7.3 using both optimization procedures described in sections 7.4 (discrete relaxation) and 7.5 (hybrid genetic search).

7.6.1 Discrete Relaxation

In this section we evaluate the ability of our model, optimized through a discrete relaxation scheme, to recover from initial matching corruption and positional noise.

We compare our method (cliques+procrustes) to the structural matching by discrete relaxation method by Wilson & Hancock (1997) (cliques) (sections 7.2 and 7.4), *Graduated Assignment* by Gold & Rangarajan (1996) (grad assig) (section 4.3.2), and structural graph matching using the *Expectation-Maximization* (EM) algorithm by Luo & Hancock (2001) (section 4.3.3). The last two methods use support functions of the types used in probabilistic relaxation approaches.

In the first experiment we have tested the ability of recovering from initial matching corruption. The degree of corruption ranges from zero (initial matching 100% correct) to 1 (initial matching completely erroneous). We have used Graduated Assignment without initialization, so we have plotted the mean correct matching rate. Each location in the plot is the mean of $84 \times 5 = 420$ experiments, so that each one of the 84 graphs in the database is matched to its prototype using 5 different randomly corrupted initial configurations. Figure 7.4 shows the results.

The second experiment evaluates how noise in the position coordinates affects to our method. We have applied Gaussian white noise to the position coordinates of each node. The variance of the noise ranges from zero to the total variance of the data. So, in the extreme case the variance due to noise is the same than due to data. We have run three trials with different fractions of initial corruption in the matching configuration, namely, 0.5, 0.7 and 0.9. Since our method is the only one sensitive to this kind of noise, we plot mean results obtained by the original cliques method (Wilson & Hancock, 1997) under the same levels of corruption. Each location in the plot is the mean of $84 \times 5 = 420$ experiments, this is, 5 random perturbations for each one of the 84 matching experiments. Figure 7.5 shows the results.

With regards to the ability of recovering from initially corrupted matching configurations, our method has shown the best performance followed by the original cliques approach (Wilson & Hancock, 1997). Results of the clique-based approaches show a performance decrease for initial matching corruptions above 70%, specially in the original cliques approach. Methods using probabilistic-relaxation-based support functions are unable to deal with the types of graphs addressed here. This contrasts with the well-known effectiveness of these methods when applied to other types of graphs.

With regards to the positional noise, results show that our method improves cliques method while noise fraction is under $\sim 17\%$ for both initial matching

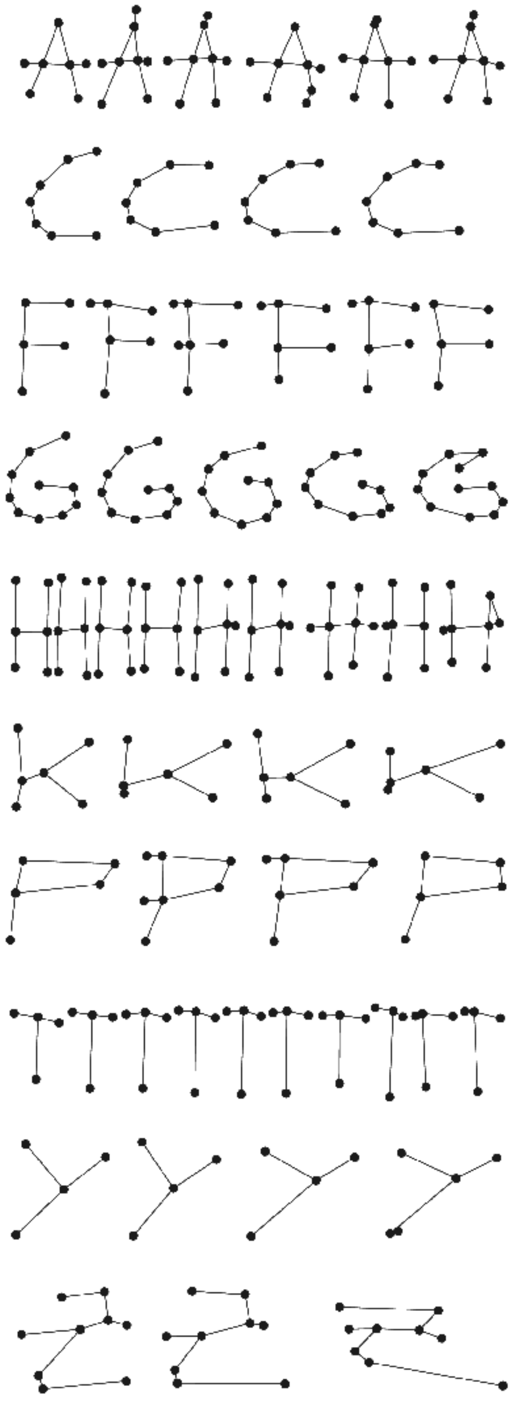


Figure 7.3: A moderate amount of intra-class variations are present in the form of structural and positional deformations.

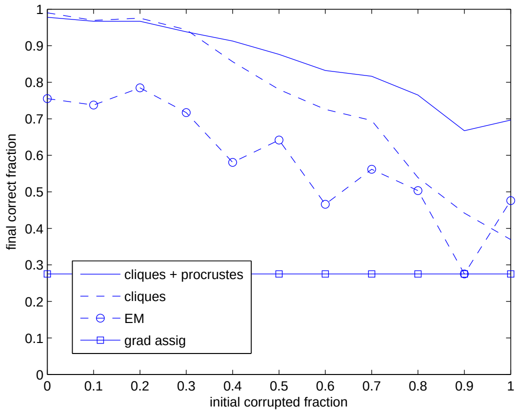


Figure 7.4: Final correct fraction versus initial corrupted fraction of matches.

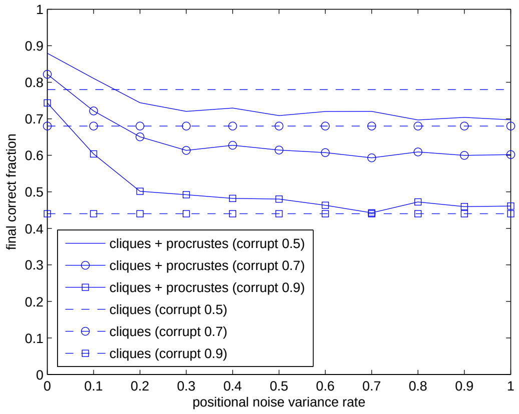


Figure 7.5: Final correct fraction of matches versus positional noise in the node coordinates.

corruptions of 50% and 70%. For an initial corrupted matching fraction of 90%, our method shows the best performance along all the positional noise range used in the experiments. Interestingly, positional noise degrades the performance of our method up to a threshold ($\sim$ 25% of total variance), after which it stabilizes. This may be due to the effect of the structural component of our model which is not affected by this type of noise.

7.6.2 Hybrid Genetic Search

In this section we evaluate the performance of our model optimized using hybrid genetic search as described in section 7.5. We present two types of experiments that evaluate either the matching efficiency or the convergence rate.

Comparative results are presented between our method (HGA cliques + procrustes) and the same hybrid approach but with the original cliques model

(HGA cliques) as reported by Cross (1997).

We have experimentally set the mutation and crossover probabilities to 0.4 and 0.5, respectively. Population is randomly initialized in all the experiments.

As we have seen in the results of the previous section, gradient ascent methods are capable of correcting $\sim 70\%$ of the initialization errors in the type of graphs used here. Therefore, the choice of population size is made so that at least 30% of the loci are correctly assigned in a random initialization. As stated by Myers & Hancock (2001), the probability that at least some fraction s of the nodes will have at least one correct assignment can be expressed as

$$P_s = \sum_{s|\mathcal{I}| \leq k \leq |\mathcal{J}|} \binom{|\mathcal{I}|}{k} P_c^k \left(1 - P_c\right)^{|\mathcal{I}|-k} \tag{7.15}$$

where $P_c = 1 - \left(1 - \frac{1+r}{|\mathcal{J}|+1}\right)^n$ is the probability of at least one correct assignment appearing in the initial population of size n with a fraction r of nodes corrupt.

Figure 7.6 shows the correct assignments fraction at each iteration. The correct fraction is computed at each iteration by averaging the fraction of correct assignments among the best fit individuals. Results at the first iteration correspond to initial population with neither gradient ascent nor genetic operator steps.

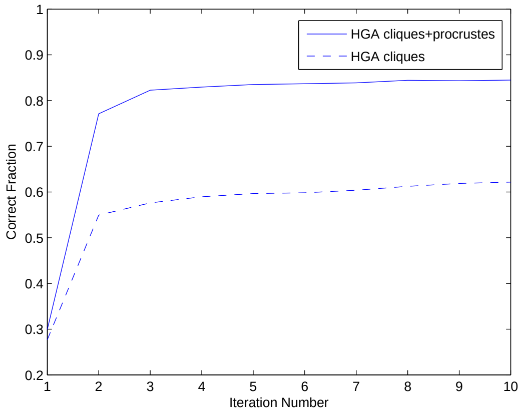


Figure 7.6: Correct assignments fraction of the best fit individuals at each iteration of the genetic algorithm.

The next experiment evaluates the tolerance of our method to severe noise in the position coordinates of the nodes. We have applied Gaussian white noise to the node coordinates with variance ranging from zero to the total variance of the data. Figure 7.7 shows the correct assignments fraction at the end of the algorithm for each noise level. The genetic algorithm iterates either until convergence or after 20 iterations. Since our method is the only sensitive to this kind of noise, we have plotted the mean final correct fraction of the *HGA cliques* method as a base-line.

Convergence rate experiments evaluate the number of iterations needed to converge to the ground-truth solution, under mutation probability and popula-

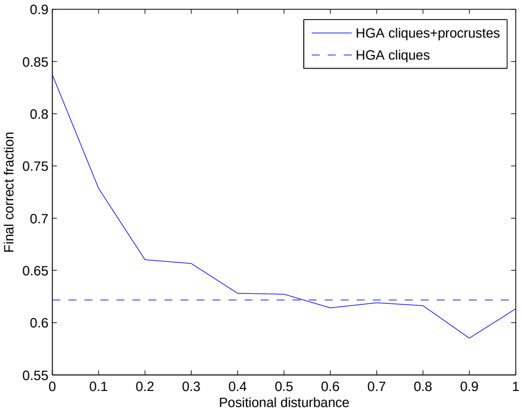


Figure 7.7: Correct assignments fraction with respect to the positional noise.

tion size variations. In the case of no convergence, the algorithm is stopped at iteration 20.

Figure 7.8 shows the number of iterations until convergence with respect to either the mutation probability or population size. The size of the population is controlled by the correct fraction required in a random initialization. In the case that they are not varied, mutation probability and required initial correct fraction are set to 0.4 and 0.3, respectively.

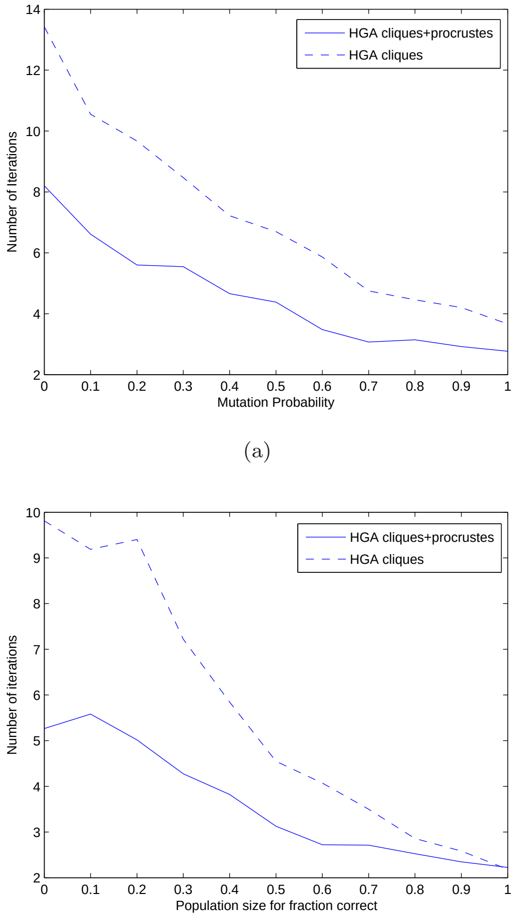
With regards to the matching accuracy through iterations of figure 7.6, results show that our method evolves significantly better than the compared one. A correct fraction of 85% is reached at iteration 10 of our method while, *HGA cliques* is slightly over 60%. This might be due to the ambiguities present when using the purely structural criterion. This suggests that the proposed model agrees with the ground truth more than the pure structural cliques model does. Interestingly, the correct fraction in the randomly initialized population ($\sim 30\%$) is consistent with the choice of the population size according to equation (7.15).

In the positional noise experiments of figure 7.7, results show that our model outperforms *HGA cliques* while positional noise is under $\sim 40\%$ and it maintains a similar performance while this noise is under $\sim 75\%$. Similarly to the results with the discrete relaxation scheme in the previous section, positional noise degrades the performance of our hybrid genetic algorithm until a threshold of $\sim 20\%$, after which performance stabilizes.

As seen in figure 7.8, the number of iterations required until convergence to the correct solution tend to decrease as both the mutation probability and the population size increase. Moreover, our method tends to converge faster than the compared one.

7.7 Conclusions

We have presented a model aimed at improving the matching of sparse graphs such as those obtained from skeletal representations of handwritten letters. Although we have evaluated our model with handwritten characters it is applicable



(a)

(b)

Figure 7.8: (a) Iterations until convergence versus mutation probability. (b) Iterations until convergence versus correct fraction required in the choice of population size.

to any rigid object from which relevant geometric information can be extracted. Our model combines structural and geometric evidence through the use of Hamming and Procrustes distances in a dictionary-based formulation. The model presented integrates smoothly into the cliques framework reported by Wilson & Hancock (1997) at the expenses of low extra computational cost (for the graphs used in our experiments). We have evaluated our model with two different optimization strategies, namely, discrete relaxation and hybrid genetic search. Results show a significant improvement in the ability of recovering from corrupted matching initializations, specially under severe corruption conditions where the improvement is $\sim 100\%$ with respect to the original model by Wilson & Hancock (1997). Geometric noise does not degrade too much the performance of our model, obtaining results comparable to those obtained by the original cliques model.

Chapter 8

A New Graph Matching Method for Point-Set Correspondence using the EM Algorithm and Softassign

8.1 Introduction

The correspondence problem in computer vision tries to determine which parts of one image correspond to which parts of another image. This problem often arises at the early stages of many computer vision applications such as 3D scene reconstruction, object recognition, pose recovery and image retrieval, among others. So, it is of basic importance to develop effective methods that are both robust -in the sense of being able to deal with noisy measurements- and general -in the sense of having a wide field of application-.

The typical steps involved in the solution of the correspondence problem are the following. First, a set of tentative feature matches is computed. These tentative matches can be further refined by a process of outlier rejection that eliminates the spurious correspondences or alternatively, they can be used as starting point of some optimization scheme to find a different, more consistent set.

Tentative correspondences may be computed either on the basis of correlation measures or feature-descriptor distances.

Correlation-based strategies compute the matches by means of the similarity between the image patches around some interest points. Interest points (that play the role of the images' parts to be matched) are image locations that can be robustly detected among different instances of the same scene with varying imaging conditions. Interest points can be corners (intersection of two edges) (Harris & Stephens, 1988; Shi & Tomasi, 1994; Tomasi & Kanade, 1991), maximum curvature points (Han & Brady, 1995; Kitchen & Rosenfeld, 1982;

Koenderink & Richards, 1988) or isolated points of maximum or minimum local intensity (Rosten & Drummond, 2006).

On the other hand, approaches based on feature-descriptors use the information local at the interest points to compute descriptor-vectors. Those descriptor-vectors are meant to be invariant to geometric and photometric transformations. So that, corresponding areas in different images present low distances between their feature-descriptors. A recent paper by Mikolajczyk and Schmid (Mikolajczyk & Schmid, 2005) evaluate some of the most competent approaches.

Another interesting descriptor is Shape Contexts (Belongie *et al.*, 2002; Mori *et al.*, 2005). Given a set of contour points, it consists of a bi-dimensional histogram capturing the spatial distribution of the rest of the points with respect to a given point (see section 1.3.2 for more details). Tentative correspondences are computed on the basis of the similarity between their histograms.

Despite the invariance introduced during the detection/description and the matching phases, the use of local image contents may not suffice to get a reliable result under certain circumstances (e.g., regular textures, multiple instances of a given feature across the images or, large rigid/non-rigid deformations). Figure 8.1 shows an example of a matching by correlation of a scene under rotation and zoom.

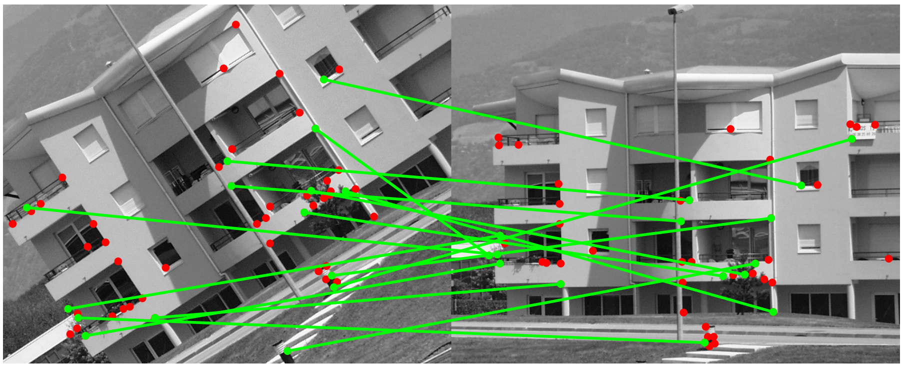


Figure 8.1: Two sample images belonging to the class *Resid* from *http://www.featurespace.org/* with superposed Harris corners (Harris & Stephens, 1988). The green lines represent the tentative correspondences computed by matching by correlation. The red dots are unmatched points. There are several misplaced correspondences.

It is a standard procedure to exploit the underlying geometry of the problem to enforce the global consistency of the correspondence-set. This is the case of the model fitting paradigm RANSAC (Fischler & Bolles, 1981) which is extensively used in computer vision to reject outliers. It selects random samples of correspondences from a tentative set and use them to fit a geometric model to the data. The largest consensus obtained after a number of trials is selected as the inlier class (see section 2.4 for more details). Another effective outlier rejector is based on a Graph Transformation (Aguilar *et al.*, 2009). This is an iterative process that discards one outlying correspondence at a time, according to a graph-similarity measure. After each iteration, the graphs are reconfigured in order to reflect the new state of the remaining correspondences. The process ends up with two isomorphic graphs and the surviving correspondences constitute the inlier class (see section 4.3.4 for more details).

The main drawback of these methods is that their ability to obtain a dense correspondence-set strongly depends on the reliability of the tentative correspondences. Since they are unable either to generate new correspondences or to modify the existing ones, an initial correspondence-set with few successes may result in a sparse estimate. This is illustrated in figure 8.2.

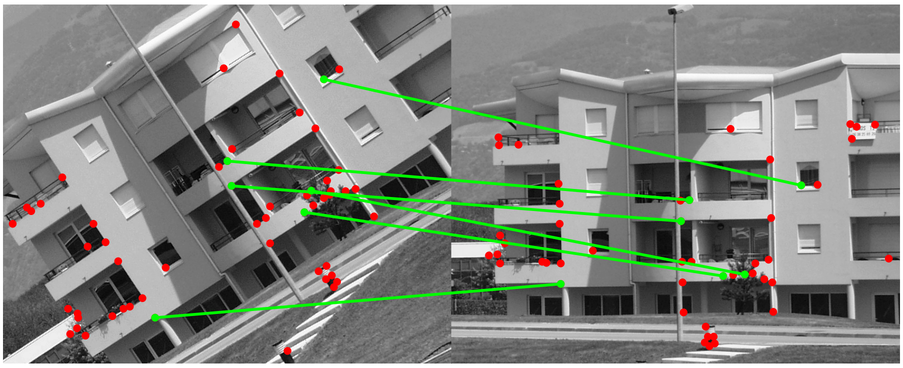


Figure 8.2: The green lines represent the resulting RANSAC inliers from the initial correspondence-set from figure 8.1. Only a few inliers are found by RANSAC. This may not be suitable in the cases when a more dense correspondence-set is needed.

Other approaches such as Iterative Closest Point (ICP) (Besl & McKay, 1992) that fall into the optimization field, attempt to simultaneously solve the correspondence and the alignment problem. Despite they are able to modify the correspondences at each iteration, simple nearest neighbour association is prone to local minima, specially under bad initial alignment estimates.

Attributed Relational Graphs (more generally, graphs) are representational entities allowing for attributes in the nodes and relations among them in the edges. Attributed Graph Matching methods are optimization techniques that contemplate these two types of information to compute the matches and therefore, do not rely on simple nearest neighbour association. In the following section, we review the process of solving the correspondence problem in computer vision using graph techniques.

8.1.1 The Correspondence Problem in Computer Vision using Graphs

The first step at solving the correspondences between two images is to extract their graph representations.

In the case of general images, a commonly adopted representation is to associate feature points to nodes and generate the edge relations following either a Delaunay triangulation (Wilson *et al.*, 1998) or a k-nearest-neighbor strategy (Aguilar *et al.*, 2009).

In the case of binary shape images, it is common to extract the graphs using the shapes' medial axis or *skeleton* (Blum, 1967; Goh, 2008). Some approaches to chinese character recognition represent the strokes in the nodes (Chan & Cheung, 1992; Suganthan & Yan, 1998). However, it is more usual to represent the skeletal end-and-intersection-points in the nodes, and their links in the edges. Some approaches use *Shock-graphs* (Sebastian *et al.*, 2004; Siddiqi *et al.*, 1998; Torsello & Hancock, 2004) or *Attributed Skeletal Graphs* (Di Ruberto, 2004). These are types of graphs which are closely related to the skeletal representations and therefore, cannot be applied to more general computer vision problems.

Another approach uses the similarity of the skeletal paths between the end nodes to establish the shape correspondences (Bai & Latecki, 2008). However, as in the previous case, its applicability is restricted to skeletal representations.

Labeling objects of a scene using their relational constraints is at the core of all general-purpose graph-matching algorithms. An early attempt to discrete labeling was by Waltz (1975). Rosenfeld *et al.* (1976) developed a model to relax the Waltz's discrete labels by means of probabilistic assignments. They introduced the notion of compatibility coefficients and laid the bases of *probabilistic relaxation* in graph matching (section 4.3.1). Hummel & Zucker (1983) firmly positioned the probabilistic relaxation into the continuous optimization domain by demonstrating that finding consistent labellings was equivalent at maximizing a local average consistency functional. Thus, the problem could be solved with standard continuous optimization techniques such as gradient ascent. Gold & Rangarajan (1996) developed an optimization technique, *Graduated Assignment*, specifically designed to the type of objective functions used in graph matching (section 4.3.2). They used a Taylor series expansion to approximate the solution of a quadratic assignment problem by a succession of easier linear assignment problems. They used Softassign (Chui & Rangarajan, 2003; Rangarajan *et al.*, 1996; Sinkhorn, 1964) to solve the linear assignment problems in the continuous domain. The key ingredients of their approach were two-way constraints satisfaction and a continuation method to avoid poor local minima.

Another family of approaches, also in the continuous optimization domain, uses statistical estimation to solve the problem. Christmas *et al.* (1995) derived the complete relaxation algorithm, including the calculation of the compatibility coefficients, following the Maximum A Posteriori (MAP) rule. Cross & Hancock (1998); Wilson & Hancock (1997) used *cliques*, a kind of graph sub-entities, for graph matching. Furthermore, they proposed a new principled way of detecting outliers that consists in measuring the net effects of a node deletion in the reconfigured graph. Accordingly, an outlier is a node that leads to an improvement in the consistency of the affected cliques after its removal. Nodes are regularly tested for deletion or reinsertion following this criterion. In section 4.4.1 we give more details about this process. The main drawbacks are that this process of outlier detection is very time consuming since each node must be tested twice (for deletion and reinsertion), each time involving a graph reconfiguration.

Cross & Hancock (1998); Luo & Hancock (2001) formulated the problem of graph matching as one of probability mixture modeling. This can be thought of as a *missing data* problem where the correspondence indicators are the parameters of the distribution and the corresponding nodes in the model-graph are the hidden variables. They used the Expectation-Maximization (EM) Algorithm (Dempster *et al.*, 1977) to find the Maximum Likelihood (ML) estimate of the correspondence indicators. (Cross & Hancock, 1998; Luo & Hancock,

2003) presented approaches to jointly solve the correspondence and alignment problems. They did so by exploiting both the geometrical arrangement of the points and their structural relations.

The advantages of posing graph matching as a joint correspondence and alignment problem, are twofold. On one hand, structural information may contribute to disambiguate the recovery of the alignment (unlike purely geometric approaches). On the other hand, geometrical information may aid to clarify the recovery of the correspondences in the case of structural corruption (unlike structural graph matching approaches).

We present a new graph matching approach aimed at finding the correspondences between two sets of coordinate points. The main novelties of our approach are:

- Instead of individual measurements, our approach uses relational information of two types: structural and geometrical. This contrasts with other approaches that use absolute geometrical positions (Cross & Hancock, 1998; Luo & Hancock, 2003).
- It maintains a true continuous underlying correspondence variable throughout all the process. Although there are approaches that relax the discrete assignment constraints through the use of statistical measurements, their underlying assignment variable remains discrete (Cross & Hancock, 1998; Luo & Hancock, 2003, 2001; Wilson & Hancock, 1997).
- We face the graph matching problem as one of mixture modelling. To that end, we derive the EM algorithm for our model and approximate the solution as a succession of assignment problems which are solved using Softassign.
- We develop effective mechanisms to detect and remove outliers. This is a useful technique in order to improve the matching results.

Figure 8.3 shows the results of applying our method to the previous matching example.

Although they are more effective, Graph Matching algorithms are also more computationally demanding than other approaches such as the robust estimator RANSAC. Suboptimal Graph Matching algorithms, such as the ones treated in this thesis, often present an $\mathcal{O}\left(N^4\right)$ complexity. However, Graph Matching algorithms can be very useful at specific moments during a real-time operation, e.g. when the tentative correspondence-sets are insufficient for further refinement or when drastic discontinuities appear in the video flow that cause the tracking algorithms to fail. When these circumstances are met, it may be advisable to take a couple of seconds in order to conveniently redirect the process. We present computational time results that demonstrate that our algorithm can match a considerable amount of points in an admissible time using a C implementation.

The outline of this chapter is as follows. In section 8.2, we formalize some concepts such as graphs representations and correspondence indicators. The mixture model is presented in section 8.3. In section 8.4, we give the details on the optimization procedure using the EM algorithm. The mechanisms for outlier detection are presented in section 8.5. We provide experimental validation in section 8.6. Last, discussion about the results and concluding remarks are given in sections 8.7 and 8.8.

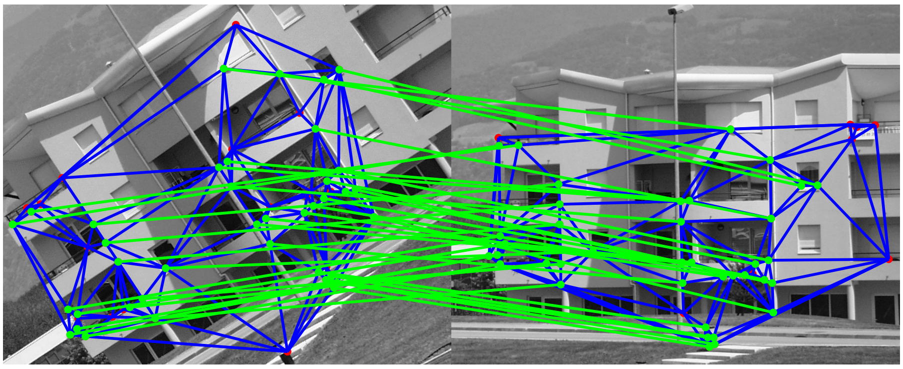


Figure 8.3: Superposed on the images there are the extracted graphs. Blue lines within each image represent the edges, generated by means of a Delaunay triangulation on the nodes. The nodes correspond to the Harris corners. The green lines represent the resulting correspondences of applying our method, using as starting point the correspondence-set of figure 8.1. Our approach arrives at a correct dense correspondence-state, while still leaving a few unmatched outliers in both images.

8.2 Graphs and Correspondences

Consider two graph representations $\mathbf{G} = (\mathcal{U}, D, \mathcal{X})$ and $\mathbf{H} = (\mathcal{V}, M, \mathcal{Y})$, extracted from two images (e.g., figure 8.3).

The node-sets $\mathcal{U} = \{u_a, \forall a \in \mathcal{I}\}$ and $\mathcal{V} = \{v_\alpha, \forall \alpha \in \mathcal{J}\}$ contain the symbolic representations of the nodes, where $\mathcal{I} = 1 \ldots |\mathcal{U}|$ and $\mathcal{J} = 1 \ldots |\mathcal{V}|$ are their index-sets.

The vector-sets $\mathcal{X} = \{\mathbf{x}_a = (x_a^V, x_a^H), \forall a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha = (y_\alpha^V, y_\alpha^H), \forall \alpha \in \mathcal{J}\}$, contain the column vectors of the two-dimensional coordinates (horizontal and vertical) of each node.

The adjacency matrices D and M contain the edge-sets, representing some kind of structural relation between pairs of nodes (e.g., connectivity or spatial proximity).

Hence, $D_{ab} = \begin{cases} 1 & \text{if } u_a \text{ and } u_b \text{ are linked by an edge} \\ 0 & \text{otherwise} \end{cases}$ (the same applies for $M_{\alpha\beta}$).

We deal with *undirected unweighted* graphs. This means that the adjacency matrices are symmetric ($D_{ab} = D_{ba}$, $\forall a,b \in \mathcal{I}$) and its elements can only take the $\{0, 1\}$ values. However, our model is also applicable to the *directed weighted* case.

The variable S represents the state of the correspondences between the node-sets $\mathcal{U}$ and $\mathcal{V}$. Therefore, we denote the probability that a node $u_a \in \mathcal{U}$ corresponds to a node $v_\alpha \in \mathcal{V}$ as $s_{a\alpha} \in S$.

It is satisfied that

$$\sum_{\alpha \in \mathcal{J}} s_{a\alpha} \leq 1 \; , \; \forall a \in \mathcal{I} \tag{8.1}$$

where the probability of node u_a being an outlier equals to

$$1 - \sum_{\alpha \in \mathcal{J}} s_{a\alpha} \tag{8.2}$$

8.2.1 Geometrical Relations

Similarly as it is done with the structural relations, instead of its individual measurements, our aim is to consider the geometrical relations between pairs of nodes. To that end, we define the new coordinate vectors $\mathbf{x}_{ab} = (\mathbf{x}_b - \mathbf{x}_a)\,,\ \forall_{a,b \in \mathcal{I}}$ and $\mathbf{y}_{\alpha\beta} = (\mathbf{y}_\beta - \mathbf{y}_\alpha)\,,\ \forall_{\alpha,\beta \in \mathcal{J}}$, that represent the coordinates of the points $\mathbf{x}_b$ and $\mathbf{y}_\beta$ relative to $\mathbf{x}_a$ and $\mathbf{y}_\alpha$, respectively. Accordingly, we define a new descriptor $\mathcal{X}^a$ for node u_a, as the translated positions of the remaining points so that their new origin is at point $\mathbf{x}_a$, i.e., $\mathcal{X}^a = \{\mathbf{x}_{ai}, {}_{i \in \mathcal{I}}\}$. Similarly for graph $\mathbf{H}$, $\mathcal{Y}^\alpha = \{\mathbf{y}_{\alpha j}, {}_{j \in \mathcal{J}}\}$. This is illustrated in figure 8.4.

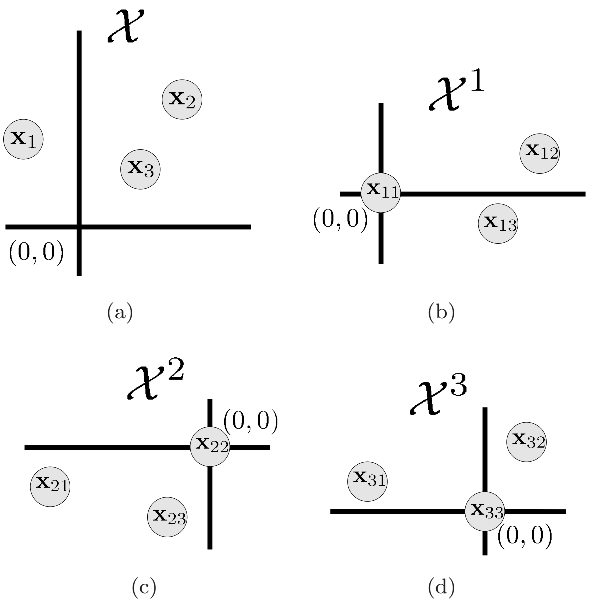


Figure 8.4: The entire point-set $\mathcal{X}$(a) and, the descriptors $\mathcal{X}^1$(b), $\mathcal{X}^2$(c) and $\mathcal{X}^3$(d), that represent the spatial distribution of the point-sets around their new origins $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, respectively.

Affine invariance is introduced at the level of node descriptors so, we consider different affine registration parameters $\Phi_{a\alpha}$ for each possible correspondence $u_a \to v_\alpha$. Since geometrical information is used in a relational way, affine registration does not depend on any translation parameter. Affine registration parameters $\Phi_{a\alpha}$ are then defined by the 2×2 matrix $\Phi_{a\alpha} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$.

We denote the whole set of affine registration parameters as $\Phi = \{\Phi_{a\alpha}\,, \forall_{a,\alpha}\}$.

8.3 A Mixture Model

Our aim is to recover the set of correspondence indicators S that maximize the incomplete likelihood of the relations in the observed graph $\mathbf{G}$. Since the geometrical relations are compared in an affine invariant way, we contemplate the affine registration parameters Φ. Ideally, we seek the optimal correspondence indicators $S^\star$ that satisfy

$$S^\star = \arg\max_S \left\{ \max_\Phi P\left(\mathbf{G}|S,\Phi\right) \right\} \tag{8.3}$$

The mixture model reflects the possibility that any single node can be in correspondence with any of the reference nodes. The standard procedure to build likelihood functions for mixture distributions consists in factorizing over the observed data (i.e., observed graph nodes) and summing over the hidden variables (i.e., their corresponding reference nodes). We write,

$$P\left(\mathbf{G}|S,\Phi\right) = \prod_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} P\left(u_a, v_\alpha|S,\Phi_{a\alpha}\right) \tag{8.4}$$

where $P\left(u_a, v_\alpha|S,\Phi_{a\alpha}\right)$ represents the probability that node u_a corresponds to node v_α given the correspondence indicators S and the registration parameters $\Phi_{a\alpha}$. We are assuming conditional independence between the observed nodes.

Following a similar development than Luo & Hancock (2001) we factorize, using the Bayes rules, the conditional likelihood in the right hand side of equation (8.4) into terms of individual correspondence indicators, in the following way.

$$P\left(u_a, v_\alpha|S,\Phi_{a\alpha}\right) = K_{a\alpha} \prod_{b\in\mathcal{I}} \prod_{\beta\in\mathcal{J}} P\left(u_a, v_\alpha|s_{b\beta},\Phi_{a\alpha}\right) \tag{8.5}$$

where

$$K_{a\alpha} = \left[\frac{1}{P\left(u_a|v_\alpha,\Phi_{a\alpha}\right)}\right]^{|\mathcal{I}|\times|\mathcal{J}|-1} \tag{8.6}$$

If we assume that the observed node u_a is conditionally dependant on the reference node v_α and the registration parameters $\Phi_{a\alpha}$ only in the presence of the correspondence matches S, then $P\left(u_a|v_\alpha,\Phi_{a\alpha}\right) = P\left(u_a\right)$.

If we assume equiprobable priors $P\left(u_a\right)$, then we can safely discard these quantities in the maximization of equation (8.3), since they do not depend neither on S or Φ.

The main aim of equation (8.5) is to measure the likelihood of the correspondence between nodes $u_a \in \mathcal{U}$ and $v_\alpha \in \mathcal{V}$, by evaluating the compatibility of the pairwise relations emanating from them, by means of the correspondence indicators $s_{b\beta}$.

In order to illustrate this process, suppose that we want to measure the likelihood of the correspondence $u_1 \to v_1$ under the situation depicted in figure 8.5.

The likelihood of the correspondence between nodes u_1 and v_1 depends on how compatible are the relations emanating from them to rest of the nodes given the current correspondence indicators. This is illustrated in figure 8.6.

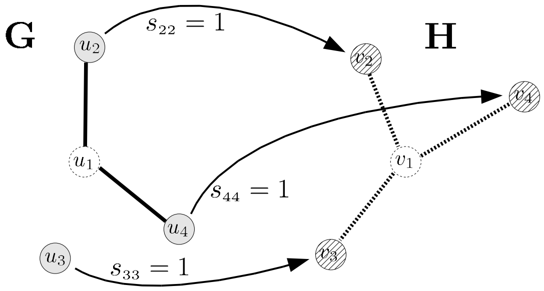


Figure 8.5: We want to measure the likelihood of the correspondence $u_1 \to v_1$ given that nodes $u_2, u_3, u_4 \in \mathcal{U}$ correspond to nodes $v_2, v_3, v_4 \in \mathcal{V}$, respectively.

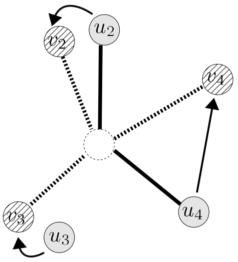


Figure 8.6: Supposing that node u_1 corresponds to v_1 then, nodes u_2 and v_2 have good geometrical and structural compatibility, nodes u_3 and v_3 have good geometrical but not good structural compatibility and, nodes u_4 and v_4 have good structural but poor geometrical compatibility.

8.3.1 A Probability Density Function

In the following, we propose a density function for measuring the conditional likelihood of the individual relations in the right hand side of equation (8.5).

For the sake of clarity, we will define our density function in different stages. First, we will propose separate structural and geometrical models in the case of binary correspondence indicators, i.e., $s_{b\beta} = \{0, 1\}\,, \forall_{b \in \mathcal{I}},\ \forall_{\beta \in \mathcal{J}}$. Next, we will fuse these separate relational models into a combined one and, last we will extrapolate to the case of continuous correspondence indicators.

Regarding the structural relations, we draw on the model by Luo & Hancock (2003, 2001). It considers that structural errors occur with a constant probability P_e. This is, given two corresponding pairs of nodes $u_a \to v_\alpha$, $u_b \to v_\beta$, we assume that there will be lack of edge-support (i.e., $D_{ab} = 0 \vee M_{\alpha\beta} = 0$) with a constant probability P_e. Accordingly, we define the following likelihood function

$$P\left(D_{ab}, M_{\alpha\beta} | s_{b\beta}\right) = \begin{cases} (1 - P_e) & \text{if } s_{b\beta} = 1 \wedge D_{ab} = 1 \wedge M_{\alpha\beta} = 1 \\ P_e & \text{otherwise} \end{cases} \tag{8.7}$$

With regards to the geometrical relations we consider that, in the case of correspondence between nodes u_b and v_β, an affine-invariant measurement of the relative point errors $P\left(\mathbf{x}_{ab}, \mathbf{y}_{\alpha\beta} | \Phi_{a\alpha}\right)$ (for brevity $P^{(\Phi_{a\alpha})}_{a\alpha b\beta}$) is appropriate in gauging the likelihood of the relation $\mathbf{x}_{ab}$. We use a multivariate Gaussian distribution to model this process. We write

$$P^{(\Phi_{a\alpha})}_{a\alpha b\beta} = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\left\|\mathbf{x}_{ab} - \mathcal{T}\left(\mathbf{y}_{\alpha\beta}; \Phi_{a\alpha}\right)\right\|^2_{\Sigma}\right] \tag{8.8}$$

where Σ is a diagonal variance matrix and, $\mathcal{T}\left(\mathbf{y}_{\alpha\beta}; \Phi_{a\alpha}\right) = \Phi_{a\alpha}\mathbf{y}_{\alpha\beta}$ are the transformed coordinates $\mathbf{y}_{\alpha\beta}$ according to the affine registration parameters $\Phi_{a\alpha}$, a 2×2 matrix of affine scale and rotation parameters. Note that $\mathbf{x}_{ab}$ and $\mathbf{y}_{\alpha\beta}$ are already invariant to translation (figure 8.4).

In the case of no correspondence between nodes u_b and v_β, we assign a constant probability ρ that controls the outlier process (see section 8.5). Therefore, the conditional likelihood becomes

$$P\left(\mathbf{x}_{ab}, \mathbf{y}_{\alpha\beta} | s_{b\beta}, \Phi_{a\alpha}\right) = \begin{cases} P^{(\Phi_{a\alpha})}_{a\alpha b\beta} & \text{if } s_{b\beta} = 1 \\ \rho & \text{if } s_{b\beta} = 0 \end{cases} \tag{8.9}$$

Now it is turn to define a combined measurement for the structural and geometrical likelihoods. To this end, we fuse the densities of equations (8.7) and (8.9) into the following expression

$$P\left(u_a, v_\alpha | s_{b\beta}, \Phi_{a\alpha}\right) = \begin{cases} \left(1 - P_e\right) P^{(\Phi_{a\alpha})}_{a\alpha b\beta} & \text{if } s_{b\beta} = 1 \wedge \left(D_{ab} = 1 \wedge M_{\alpha\beta} = 1\right) \\ P_e P^{(\Phi_{a\alpha})}_{a\alpha b\beta} & \text{if } s_{b\beta} = 1 \wedge \left(D_{ab} = 0 \vee M_{\alpha\beta} = 0\right) \\ P_e \rho & \text{if } s_{b\beta} = 0 \end{cases} \tag{8.10}$$

The above density function is defined only in the case of binary correspondence indicators $s_{b\beta}$. We extrapolate it to the continuous case by exploiting, as exponential indicators, the conditional expressions of equation (8.10) in the following way,

$$\begin{gathered} P\left(u_a, v_\alpha | s_{b\beta}, \Phi_{a\alpha}\right) = \\ \left[\left(1 - P_e\right) P^{(\Phi_{a\alpha})}_{a\alpha b\beta}\right]^{D_{ab} M_{\alpha\beta} s_{b\beta}} \left[P_e P^{(\Phi_{a\alpha})}_{a\alpha b\beta}\right]^{(1 - D_{ab} M_{\alpha\beta}) s_{b\beta}} \left[P_e \rho\right]^{(1 - s_{b\beta})} \end{gathered} \tag{8.11}$$

Figure 8.7 shows an illustrative plot of the density function of equation (8.11).

Figure 8.8 illustrates the case when the measurement likelihood $P^{(\Phi_{a\alpha})}_{a\alpha b\beta}$ is lower than the quantity ρ.

Substituting equation (8.11) into (8.5) (and discarding the observed node priors $P\left(u_a\right)$), the final expression for the likelihood of the correspondence between nodes u_a and v_α, expressed in the exponential form, is

$$\begin{gathered} P\left(u_a, v_\alpha | S, \Phi_{a\alpha}\right) = \\ \exp\left\{\sum_{b \in \mathcal{I}} \sum_{\beta \in \mathcal{J}} \left(s_{b\beta}\left[D_{ab} M_{\alpha\beta} \ln\left(\frac{1 - P_e}{P_e}\right) + \ln\left(\frac{P^{(\Phi_{a\alpha})}_{a\alpha b\beta}}{\rho}\right)\right] + \ln\left(P_e \rho\right)\right)\right\} \end{gathered} \tag{8.12}$$

This is, the exponential of a weighted sum of structural and geometrical compatibilities between the pairwise relations emanating from nodes $u_a \in \mathcal{U}$ and $v_\alpha \in \mathcal{V}$. The weights $s_{b\beta}$ play the role of selecting the proper reference

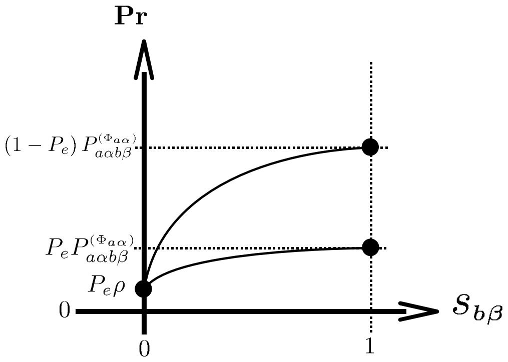
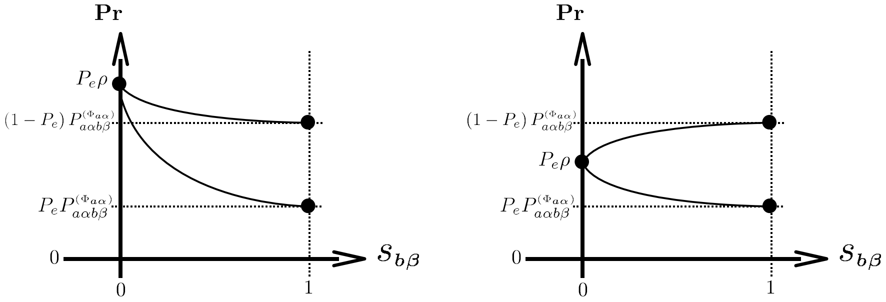


Figure 8.7: Density function of equation (8.11), an extension of the function of equation (8.10) to continuous correspondence indicators. Each solid curve represent either the case of edge-support (i.e., $D_{ab} = 1 \wedge M_{\alpha\beta} = 1$) or lack of it (i.e., $D_{ab} = 0 \vee M_{\alpha\beta} = 0$). At the extrema of each curve (i.e., $s_{b\beta} = \{0, 1\}$), represented with black dots (•), we find the three cases of equation (8.10).



(a) The likelihood value $P^{(\Phi_{a\alpha})}_{a\alpha b\beta}$ is lower than ρ. The density function of equation (8.11) decreases below the outlying threshold ($P_e\rho$) as the correspondence between nodes b and β increases. This means that (a, b) is not a plausible relation under the assumption that node a corresponds to α and b to β.

(b) The density function of equation (8.11) only increases as the correspondence indicator increases in the case of structural consistence (i.e., $D_{ab} = 1 \wedge M_{\alpha\beta} = 1$). In this case, structural consistence makes the difference between considering $(a, b) \rightarrow (\alpha, \beta)$ a plausible mapping or not.

Figure 8.8: In both cases (a) (b) the geometrical likelihood $P^{(\Phi_{a\alpha})}_{a\alpha b\beta}$ is lower than the outlying threshold ρ, but in the case of (b) the structural consistence makes the difference between considering a plausible mapping between relations (a, b) and (α, β) or not.

relation (v_α, v_β) that it is appropriate in gauging the likelihood of each observed relation (u_a, u_b).

These structural and geometrical coefficients (i.e., $D_{ab} M_{\alpha\beta} \ln(\frac{1-P_e}{P_e})$ and $\ln\left(\frac{P^{(\Phi_{a\alpha})}_{a\alpha b\beta}}{\rho}\right)$) are equivalent to the compatibility coefficients of the probabilistic relaxation approaches (Christmas *et al.*, 1995; Hummel & Zucker, 1983; Rosenfeld *et al.*, 1976). In this way, the structural and geometrical compatibilities are posed in a principled, balanced footing.

8.4 Expectation Maximization

The EM algorithm has been previously used by other authors to solve the Graph Matching problem (Cross & Hancock, 1998; Luo & Hancock, 2001). It is useful to find the parameters that maximize the expected log-likelihood for a mixture distribution (check section 3.2 for more details). In our case, we use it to find the correspondence indicators that maximize the expected log-likelihood of the observed relations, given the optimal alignments. From equations (8.3) and (8.4), we write,

$$S^{\star} = \arg\max_{S} \left\{ \max_{\Phi_{a\alpha}} \left\{ \sum_{a \in \mathcal{I}} \ln \left[\sum_{\alpha \in \mathcal{J}} P\left(u_a, v_\alpha | S, \Phi_{a\alpha}\right) \right] \right\} \right\} \tag{8.13}$$

Dempster *et al.* (1977) showed that maximizing the log-likelihood for a mixture distribution is equivalent at maximizing a weighted sum of log-likelihoods, where the weights are the missing data estimates. This is posed as an iterative estimation problem where the new parameters $S^{(n+1)}$ are updated so as to maximize an objective function depending on the previous parameters $S^{(n)}$. Then, the most recent available parameters $S^{(n)}$ are used to update the missing data estimates that, in turn, weigh the contributions of the log-likelihood functions. Accordingly, this utility measure is denoted

$$\Lambda\left(S^{(n+1)} | S^{(n)}\right) = \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} P\left(v_\alpha | u_a, S^{(n)}, \Phi_{a\alpha}\right) \ln P\left(u_a, v_\alpha | S^{(n+1)}, \Phi_{a\alpha}\right) \tag{8.14}$$

where the posterior probabilities of the missing data given the most recent available parameters $P\left(v_\alpha | u_a, S^{(n)}, \Phi_{a\alpha}\right)$ weigh the contributions of the conditional log-likelihood terms.

The basic idea is to alternate between Expectation and Maximization steps until convergence is reached. The expectation step involves computing the posterior probabilities of the missing data using the most recent available parameters. In the maximization phase, the parameters are updated in order to maximize the expected log-likelihood of the incomplete data.

8.4.1 Expectation

In the expectation step, the posterior probabilities of the missing data (i.e., the reference graph measurements v_α) are computed using the current parameter estimates $S^{(n)}$.

The posterior probabilities can be expressed in terms of conditional likelihoods, using the Bayes rule, in the following way

$$P\left(v_\alpha | u_a, S^{(n)}, \Phi_{a\alpha}\right) = \frac{P\left(u_a, v_\alpha | S^{(n)}, \Phi_{a\alpha}\right)}{\sum_{\alpha'} P\left(u_a, v_{\alpha'} | S^{(n)}, \Phi_{a\alpha'}\right)} \equiv \omega_{a\alpha}^{(n)} \tag{8.15}$$

Substituting our expression of the conditional likelihood of equation (8.12) into equation (8.15), the final expression for the posterior probabilities becomes,

$$\omega^{(n)}_{a\alpha} = \frac{\exp\left\{\sum\limits_{b\in\mathcal{I}}\sum\limits_{\beta\in\mathcal{J}}\left[s^{(n)}_{b\beta}D_{ab}M_{\alpha\beta}\ln\left(\frac{1-P_e}{P_e}\right)+s^{(n)}_{b\beta}\ln\left(\frac{P^{(\Phi_{a\alpha})}_{a\alpha b\beta}}{\rho}\right)+\ln\left(P_e\rho\right)\right]\right\}}{\sum\limits_{\alpha'\in\mathcal{J}}\exp\left\{\sum\limits_{b\in\mathcal{I}}\sum\limits_{\beta\in\mathcal{J}}\left[s^{(n)}_{b\beta}D_{ab}M_{\alpha'\beta}\ln\left(\frac{1-P_e}{P_e}\right)+s^{(n)}_{b\beta}\ln\left(\frac{P^{(\Phi_{a\alpha})}_{a\alpha' b\beta}}{\rho}\right)+\ln\left(P_e\rho\right)\right]\right\}} \tag{8.16}$$

8.4.2 Maximization

Maximization is done in two steps. First, optimal registration parameters $\Phi_{a\alpha}$ are computed for each $P\left(u_a, v_\alpha|S, \Phi_{a\alpha}\right)$. Last, global correspondence indicators are updated using the optimal $\Phi_{a\alpha}$'s.

Maximum Likelihood Affine Registration Parameters

We are interested in the registration parameters that lead to the maximum likelihood, given the current state of the correspondences $S^{(n)}$. In other words, the node descriptors $\mathcal{X}^a$ and $\mathcal{Y}^\alpha$ must be optimally registered before we can estimate the next correspondence indicators $S^{(n+1)}$. It is important that the registration do not modify the origins of the node descriptors, since these are the locations of the evaluated nodes u_a and v_α. As consequence, the registration parameters $\Phi_{a\alpha}$ are a 2 × 2 matrix of affine rotation and scaling parameters (without translation).

Therefore, we recover the Maximum Likelihood (ML) registration parameters $\Phi^\star_{a\alpha}$, directly from equation (8.12). This is,

$$\begin{gathered}\Phi^\star_{a\alpha} = \arg\max_{\Phi_{a\alpha}}\left\{\ln P\left(u_a, v_\alpha|S^{(n)}, \Phi_{a\alpha}\right)\right\} = \\ \arg\max_{\Phi_{a\alpha}}\left\{\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}}\left[s^{(n)}_{b\beta}\ln\left(\frac{P^{(\Phi_{a\alpha})}_{a\alpha b\beta}}{\rho}\right)+D_{ab}M_{\alpha\beta}s^{(n)}_{b\beta}\ln\left(\frac{1-P_e}{P_e}\right)+\ln\left(P_e\rho\right)\right]\right\}\end{gathered} \tag{8.17}$$

We discard all the terms constant w.r.t the registration parameters and obtain the following equation

$$\Phi^\star_{a\alpha} = \arg\max_{\Phi_{a\alpha}}\left\{\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}}s^{(n)}_{b\beta}\ln\left(P^{(\Phi_{a\alpha})}_{a\alpha b\beta}\right)\right\} \tag{8.18}$$

Now, we substitute the geometrical likelihood term by its expression of equation (8.8). We discard the constant terms of the multivariate Gaussian function and cancel the exponential and the logarithm functions, thus turning the maximization problem into a minimization one, by removing the minus sign of the exponential. We get the following expression

$$\Phi^\star_{a\alpha} = \arg\min_{\Phi_{a\alpha}}\left\{\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}}s^{(n)}_{b\beta}\left\|\mathbf{x}_{ab}-\mathcal{T}\left(\mathbf{y}_{\alpha\beta};\Phi_{a\alpha}\right)\right\|^2_\Sigma\right\} \tag{8.19}$$

We seek the matrix of affine parameters $\Phi_{a\alpha} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$ that minimize the weighted sum of squared Mahalanobis distances between the relative points $\mathbf{x}_{ab}$ and the transformed relative points $\mathcal{T}\left(\mathbf{y}_{\alpha\beta}; \Phi_{a\alpha}\right) = \Phi_{a\alpha}\mathbf{y}_{\alpha\beta}$. The coefficients $s_{b\beta}$ weigh the contribution of each pairwise distance in a way that the resulting registration will tend to minimize the distances between the relative positions of those u_b and v_β with the larger correspondence indicators.

Minimization of equation (8.19) is explained in section 3.5.2 where the translation parameters must be set to zero.

Maximum Likelihood Correspondence Indicators

One of the key points in our work is to approximate the solution of the graph matching problem by means of a succession of easier assignment problems. Following the dynamics of the EM algorithm, each one of these problems is posed using the most recent parameter estimates. As it is done in Graduated Assignment (Gold & Rangarajan, 1996), we use the *Softassign* (Chui & Rangarajan, 2003; Rangarajan *et al.*, 1996; Sinkhorn, 1964) to solve the assignment problems in a continuous way. The two main features of the Softassign are that, it allows to adjust the level of discretization of the solution by means of a control parameter and, it enforces two-way constraints by incorporating a method discovered by Sinkhorn (1964) (in section 3.4 we give a more detailed explanation of Softassign). The two-way constraints guarantee that one node of the observed graph can only be assigned to one node of the reference graph, and vice versa. In the case of continuous assignments, this is accomplished by applying alternative row and column normalizations (considering the correspondence variable S as a matrix). Moreover, Softassign allows us to smoothly detect outliers in both sides of the assignment (see section 8.5).

According to the EM development, we compute the correspondence indicators $S^{(n+1)}$ that maximize the utility measure of equation (8.14). In our case, this equals to

$$S^{(n+1)} = \arg\max_{S} \left\{\Lambda\left(S|S^{(n)}\right)\right\} =$$

$$\arg\max_{S} \left\{ \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} \left(s_{b\beta} \left[D_{ab} M_{\alpha\beta} \ln\left(\tfrac{1-P_e}{P_e}\right) + \ln\left(\frac{P_{a\alpha b\beta}^{(\Phi_{a\alpha})}}{\rho}\right) \right] + \ln\left(P_e \rho\right) \right) \right\} \tag{8.20}$$

where $\omega_{a\alpha}^{(n)}$ are the missing data estimates.

Rearranging, and dropping the terms constant w.r.t the correspondence indicators, we obtain

$$S^{(n+1)} =$$

$$\arg\max_{S} \left\{ \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \left[D_{ab} M_{\alpha\beta} \ln\left(\tfrac{1-P_e}{P_e}\right) + \ln\left(\frac{P_{a\alpha b\beta}^{(\Phi_{a\alpha})}}{\rho}\right) \right] \right\} \tag{8.21}$$

which, as it can be seen in the following expression, presents the same form as an assignment problem (Gold & Rangarajan, 1996)

$$S^{(n+1)} = \arg\max_{S} \left\{ \sum_{b \in \mathcal{I}} \sum_{\beta \in \mathcal{J}} s_{b\beta} B_{b\beta}^{(n)} \right\} \tag{8.22}$$

where the $B_{b\beta}^{(n)}$ are the benefit coefficients for each assignment.

Softassign computes the correspondence indicators in two steps. First, the correspondence indicators are updated with the exponentials of the benefit coefficients

$$s_{b\beta} = \exp\left(\mu \, B_{b\beta}\right) \tag{8.23}$$

where μ is a control parameter. Second, two-way constraints are imposed by alternatively normalizing across rows and columns the matrix of exponentiated benefits. This is known as the *Sinkhorn normalization* and it is applied either until convergence of the normalized matrix or a predefined number of times.

Note that, the correspondence indicators $s_{b\beta}$ will tend to discrete values ($s_{b\beta} = \{0, 1\}$) as the control parameter μ of equation (8.23) approaches to ∞.

We also apply the Sinkhorn normalization to the posterior probabilities of the missing data so that they are more correlated with the correspondence indicators.

Since the matrices may not be square (i.e., different number of nodes in the observed and reference graphs), in order to fulfill the law of total probability, we complete the Sinkhorn normalization process with a normalization by rows.

Figure 8.9 shows the pseudo-code implementation of our method.

8.5 Outlier Detection

A node in one graph is considered to be an outlier if it has no correspondent node in the other graph.

Consider, for example, the case of figure 8.3. The rightmost nodes in the right image are outliers originated from the detection of features in the non-overlapping parts of the images. On the other hand, the unmatched nodes in the overlapping parts are outliers originated by differences in the feature detection patterns.

Outliers can dramatically affect the performance of a matching and therefore, it is important to develop techniques aimed at minimizing their influence (Black & Rangarajan, 1996).

According to our purposes, a node $u_b \in \mathcal{U}$ (or $v_\beta \in \mathcal{V}$) will be considered an outlier to the extent that there is no node v_β, $\forall \beta \in \mathcal{J}$ (or u_b, $\forall b \in \mathcal{I}$) which presents a matching benefit $B_{b\beta}^{(n)}$ above a given threshold.

From equations (8.21) and (8.22), the benefit values have the following expression

$$B_{b\beta}^{(n)} = \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} \omega_{a\alpha}^{(n)} \left[D_{ab} M_{\alpha\beta} \ln\left(\frac{1-P_e}{P_e}\right) + \ln\left(\frac{P_{a\alpha b\beta}^{(\Phi_{a\alpha})}}{\rho}\right) \right] \tag{8.24}$$

Note that, the value of ρ controls whether the geometrical compatibility term contributes either positively (i.e., $\rho < P_{a\alpha b\beta}^{(\Phi_{a\alpha})}$) or negatively (i.e., $\rho > P_{a\alpha b\beta}^{(\Phi_{a\alpha})}$) to the benefit measure.

```
while $\mu \leq \mu_f$ do
    while ($S^{(n)}$ does not converge) $\wedge$ (iterations $\leq$ max) do
        # ML Affine Parameters
        $\Phi^{(n)}_{a\alpha} \leftarrow \arg\max_{\Phi_{a\alpha}} \{\ln P\left(v_a, w_\alpha | S^{(n)}, \Phi_{a\alpha}\right)\}, \forall\, a\in\mathcal{I}\ \alpha\in\mathcal{J}$

        # Expectation
        $\omega^{(n)}_{a\alpha} \leftarrow \frac{P\left(v_a, w_\alpha | S^{(n)}, \Phi^{(n)}_{a\alpha}\right)}{\sum_{\alpha'} P\left(v_a, w_{\alpha'} | S^{(n)}, \Phi^{(n)}_{a\alpha'}\right)}, \forall\, a\in\mathcal{I}\ \alpha\in\mathcal{J}$
        $\omega^{(n)} \leftarrow \mathrm{Sinkhorn}\left(\omega^{(n)}\right)$

        # Maximization
        $B^{(n)}_{b\beta} \leftarrow \sum_{a,\alpha} \omega^{(n)}_{a\alpha} \left[ D_{ab} E_{\alpha\beta} \ln\left(\frac{1-P_e}{P_e}\right) + \ln\left(\frac{P^{(\Phi_{a\alpha})}_{a\alpha b\beta}}{\rho}\right)\right], \forall\, b\in\mathcal{I}\ \beta\in\mathcal{J}$
        $s^{(n+1)}_{b\beta} \leftarrow \exp\left(\mu\, B^{(n)}_{b\beta}\right), \forall\, b\in\mathcal{I}\ \beta\in\mathcal{J}$
        $S^{(n+1)} \leftarrow \mathrm{Sinkhorn}\left(S^{(n+1)}\right)$

        $S^{(n)} \leftarrow S^{(n+1)}$
    end
    $\mu \leftarrow \mu \times (1 + \epsilon)$
end
```

Figure 8.9: The outer loop gradually increase the Softassign parameter μ, thereby pushing from continuous to discrete solutions. This reduces the chances of getting trapped in local minima (Gold & Rangarajan, 1996). The body contains the pseudo-code of the E and M steps. Each iteration of the inner loop performs one step of the EM algorithm.

We model the outlier detection process as an assignment to (or from) the *null* node. We consider that the null node has no edges at all and, all the geometrical terms $P^{(\Phi_{a\alpha})}_{a\alpha b\beta}$ involving it are equal to ρ. Under these considerations, the benefit values of equation (8.24) corresponding to the null assignments are equal to zero. We therefore create an augmented benefit matrix $\tilde{B}^{(n)}$ by adding to $B^{(n)}$ an extra row and column of zeros. This extra row and column represent the benefits of the *null* assignments (i.e., $B_{b\emptyset}$, $\forall b\in\mathcal{I}$ and $B_{\emptyset\beta}$, $\forall \beta\in\mathcal{J}$).

We apply the Softassign (exponentiation and Sinkhorn normalization) to the augmented benefit matrix $\tilde{B}^{(n)}$. When performing Sinkhorn normalitzation we keep in mind that the *null* assignments are special cases that only satisfy one-way constraints. This is, there may be multiple assignments to *null* in both graphs. Finally, the extra row and column are removed leading to the resulting matrix of correspondence parameters $S^{(n+1)}$. This process is illustrated in figure 8.10.

As the control parameter μ of the Softassing increases, the rows and columns of $S^{(n+1)}$ associated to the outlier nodes, tend to zero. This fact reduces the influence of these nodes in the maximization phases of the next iteration that, in turn, lead to even lower benefits, and so on.

It is now turn to define the value of the constant ρ. Since ρ is to be compared with $P^{(\Phi_{a\alpha})}_{a\alpha b\beta}$, it is convenient to define it in terms of a multivariate Gaussian

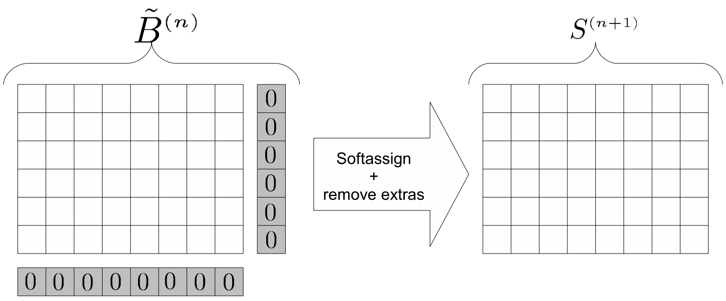


Figure 8.10: The Softassign and outlier detection process.

measurement of a distance threshold. This is,

$$\rho = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\mathbf{d}^{\mathrm{T}}\Sigma^{-1}\mathbf{d}\right] \tag{8.25}$$

where $\Sigma = \operatorname{diag}\left(\sigma_V^2, \sigma_H^2\right)$ is the same diagonal variance matrix as we use in equation (8.8) and $\mathbf{d} = (d^V, d^H)$ is a column vector with the horizontal and vertical thresholding distances.

Note that the quotient within the geometrical compatibility term can be equivalently expressed as

$$\frac{P_{a\alpha b\beta}^{(\Phi_{a\alpha})}}{\rho} = \frac{\tilde{P}_{a\alpha b\beta}^{(\Phi_{a\alpha})}}{\tilde{\rho}} \tag{8.26}$$

where we have removed the constant multiplicative factor $1/2\pi|\Sigma|^{1/2}$ from ρ and $P_{a\alpha b\beta}^{(\Phi_{a\alpha})}$ in order to set $\tilde{\rho}$ and $\tilde{P}_{a\alpha b\beta}^{(\Phi_{a\alpha})}$.

Expressing the thresholding distances as a quantity proportional to the standard deviations of the data, (i.e., $\mathbf{d} = (N\sigma_V, N\sigma_H)$), we get

$$\tilde{\rho} = \exp\left\{-\frac{1}{2}\left[\left(\frac{N\sigma_V}{\sigma_V}\right)^2 + \left(\frac{N\sigma_H}{\sigma_H}\right)^2\right]\right\} = \exp\left(-N^2\right) \tag{8.27}$$

So, we define ρ as a function of the number N of standard deviations permitted in the registration errors, in order to consider a plausible correspondence.

8.6 Experiments and Results

We assess the performance of our method in terms of registration accuracy and recognition ability.

We have not found the parameter ρ to be specially application dependant since the same value for this parameter has offered a fair performance in all the variety of experiments presented. In this sense, the parameter P_e is more application dependant since it establishes the scale of the structural contribution of our model which is to be added to the geometric contribution in order to set up the consistency measure of equation (8.24). Specifically, the value of the structural contribution depends on this scale parameter as well as the mean node degree (i.e., the mean number of incident edges upon each node).

We have tried to be as efficient as possible in the implementations of all the methods. Unless otherwise noted, all the computational time results refer to Matlab® run-times. All the experiments have been conducted on an Intel® Xeon® CPU E5310 at 1.60GHz.

8.6.1 Registration Experiments

These experiments are aimed at testing the ability of our method to locate correct matches.

Performance is assessed by either the correct correspondence rate or the *mean projection error* depending on whether the graphs are synthetically generated (with known ground truth correspondences) or extracted from real images (with known ground truth homography). We compare to other graph matching methods as well as to known point-set registration methods and outlier rejectors.

All the graphs used in this section have been generated by means of Delaunay triangulations over point-sets, where each point has been assigned to a node. We have conducted matching experiments on randomly generated graphs and have experimentally found that the values of $\tilde{\rho} = \exp\left(-1.6^2\right)$ and $P_e = 0.03$ perform well for this type of graphs. Therefore, we have used these values for our method in all the experiments in this section.

The parameters for the rest of the methods have been set using the same procedure.

This section is divided as follows. In sections 8.6.1 and 8.6.1 we use synthetic graphs to evaluate specific aspects of our model. In section 8.6.1 we use real images.

Synthetic Non-Rigid Deformations

In the first set of experiments we evaluate the matching ability in the presence of non-rigid deformations. We have matched randomly generated patterns of 15 points with deformed versions of themselves. Deformations have been introduced by applying random Gaussian noise to the position coordinates of the points.

In the synthetic experiments we assess the performance of each method through the correct correspondence rate. To see how this performance measure is related to our model we measure the ratio $\mathcal{L}_{EM-Soft}/\mathcal{L}_{gtr}$, between the value of the log-likelihood function at the solution found by the EM algorithm and that at the ground truth matching.

From equation (8.4) the expression of log-likelihood function according to our model is the following.

$$\mathcal{L} = \sum_{a \in \mathcal{I}} \ln \left[\sum_{\alpha \in \mathcal{J}} P\left(u_a, v_\alpha | S, \Phi_{a\alpha}\right) \right] \qquad (8.28)$$

Figure 8.11 shows that even though there is an increasing trend in the disagreement between the model hypothesis and the established ground truth as the deformation increases, such a disagreement remains close to the optimum value of 1 for deformations up to 20%.

We have compared the correct correspondence rates of our method (*EM-Soft*) to that of the graph matching + point-set alignment methods *Dual-Step*

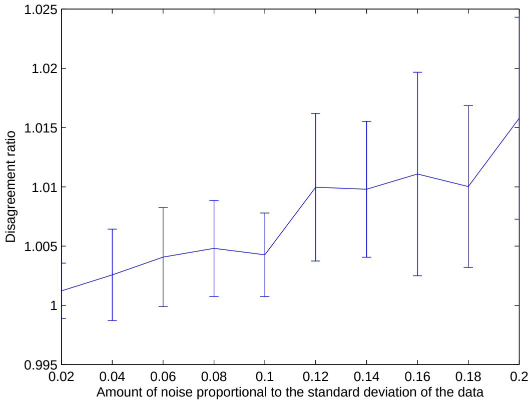


Figure 8.11: Ratio of the log-likelihood of the suboptimal solution found by our method to that of the ground truth solution, according to our model. As the deformation increases the likelihood of the ground truth solution falls below other (partially incorrect) solutions. Each location on the plots is the mean of 25 experiments (5 random patterns of points by 5 random deformations).

(section 4.5.1) and *Unified* (section 4.5.2). The *Dual-Step* has been implemented with an affine geometrical model as well as the capability of detecting outliers. Such an outliers-detection capability increases considerably its required computational time but, evaluating the performance of this feature is an important aspect in our experiments. All the approaches have been initialized with the resulting correspondences of a simple nearest neighbour association. Figure 8.12 shows the correct correspondence rates with respect to the amount of noise. The

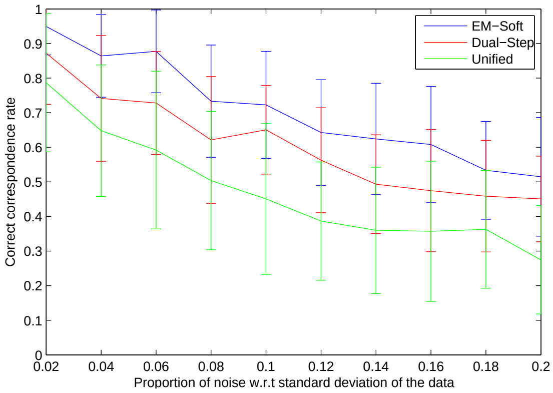


Figure 8.12: Correct correspondence rate with respect to the amount of noise in the point positions (expressed proportionally to the variance of the data). Each location on the plots is the mean of 25 experiments (5 random patterns of points by 5 random deformations).

mean computational times are: 14.6 sec. (*EM-Soft*), 124.5 sec. (*Dual-Step*) and 0.91 sec. (*Unified*).

The computational time obtained with a C implementation of our method is 0.24 sec.

Synthetic Addition of Random Points

The next set of experiments evaluates the matching ability in the presence of outliers. We have randomly added outlying points (with no correspondence in the other side) to *both* synthetic patterns of 15 points. We have preserved a proportion of ground-level non-rigid noise of 0.02 between the inliers of both patterns. In order to contribute positively to the correct correspondence rate, outliers must not be matched to any point while, inliers must be assigned to its corresponding counterpart. The approaches compared in this experiment are those with explicit outlier detection mechanisms. These are *RANSAC* (affine) (Fischler & Bolles, 1981), Graph Transformation Matching (*GTM*) (Aguilar *et al.*, 2009) and *Dual-Step* (Cross & Hancock, 1998) which are explained in sections 2.4, 4.3.4 and 4.5.1, respectively.

GTM is a powerful outlier rejector based on a graph transformation that holds a very intuitive idea. We use the same strategy as Aguilar *et al.* (2009) consisting in using k-NN graphs with $k = 5$ instead of Delaunay triangulations in order to present the results for the *GTM* method. However, similar results are obtained using Delaunay triangulations.

All the methods have been initialized with the resulting correspondences of a simple nearest neighbour association. Figure 8.13 shows the correct correspondence rate with respect to the number of outliers and figure 8.14 shows the computational times.

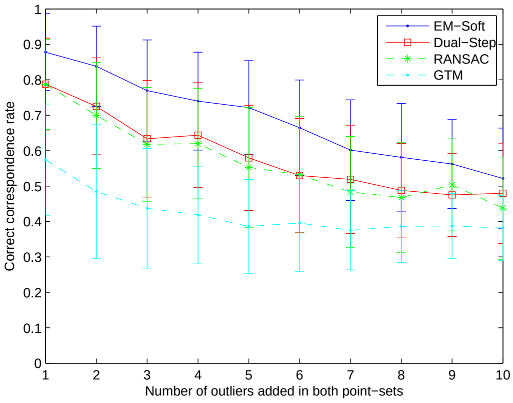


Figure 8.13: Correct correspondence rate with respect to the number of outliers in each side. Each location is the mean of 125 experiments (5 random inlier patterns by 5 random outlier patterns by 5 random ground-level non-rigid noise).

Real Images

We have performed registration experiments on real images from the database in `http://www.featurespace.org/`. Point-sets have been extracted with the Harris operator (Harris & Stephens, 1988) (section 1.2.1). Each pair of images shows two scenes related by either a zoom or a zoom + rotation. They belong

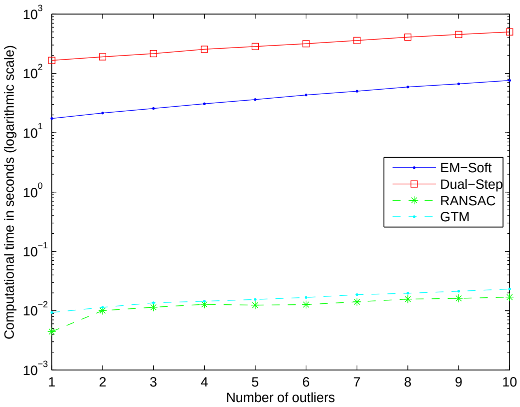
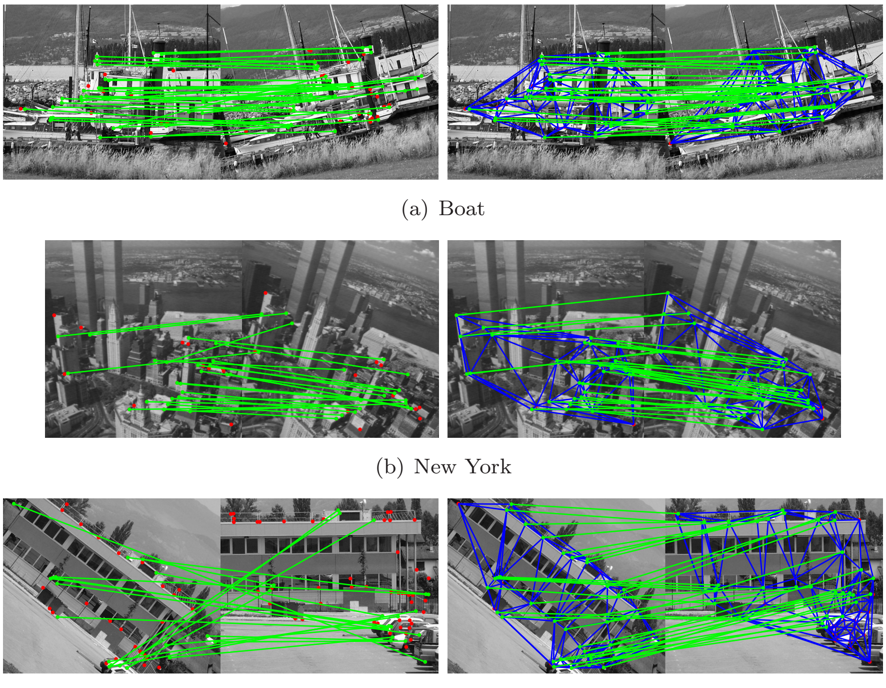


Figure 8.14: Plots of the computational times with respect to the number of outliers. The time (vertical) axis is in logarithmic scale.

to the classes *Resid*, *Boat*, *New York* and *East Park*. All the approaches use the same parameters as in the previous section. Figures 8.3 and 8.15 show the resulting correspondences found by our method as well as the tentative correspondences used as starting point.



(a) Boat

(b) New York

(c) East Park

Figure 8.15: Right column shows the results of our method using the matching by correlation results (left column) as starting point.

We have compared all the methods with outlier detection capabilities of the previous section. The graph-based approaches have been initialized with the matching by correlation results (*Corr*) which is explained in section 1.4.1. In order to avoid the sparsity problem mentioned in figure 8.2, we have applied *ICP* (Besl & McKay, 1992) (section 3.1.1) to the correlation results, as a previous step to the outlier rejectors *RANSAC* and *GTM*.

From the resulting correspondences, we have estimated the corresponding homographies with the DLT algorithm (Hartley & Zisserman, 2000). Since the ground truth homography between each pair of images is available, we have measured the mean projection error (MPE) of the feature-points in the origin images. Table 8.1 shows the mean projection errors as well as the proportion of matched points in the origin images. Table 8.2 shows the computational times in seconds. Table 8.4 shows the computational times of the Matlab and C implementations of our method.

In order to show how methods benefit from outlier rejection in a real world application, we have repeated the above experiments using the *Unified* (section 4.5.2)method and the pure structural method Graduated Assignment (*GradAssig*) (Gold & Rangarajan, 1996) (section 4.3.2), both without explicit outlier rejection capabilities. We have also added modified versions of *EM-Soft* and *Dual-Step* so that outlier rejection is disabled (marked with an asterisk). Table 8.3 shows the results.

8.6.2 Recognition Experiments

In this section we assess the recognition ability of the underlying model in our graph matching method in a series of shape retrieval experiments on the GREC database (Riesen & Bunke, 2008) and a 25-shapes database. In these experiments, the structure of the graphs has been given rise by the morphology of the objects.

Due to numerical reasons, lower values of P_e are needed in the case of this morphologically-induced graphs than in the case of Delaunay triangulations. This is because in this case, resulting graphs are sparser and therefore structural contributions under equation (8.24) need to be amplified so as to play a role comparable to the geometric contributions. We have used the first 5 graphs from each class of the GREC database in order to tune the parameters of all the methods. Due to the similar nature of the graphs in both databases and to the lack of examples in the 25-shapes database in order to perform training, we have used the same parameters in both databases. We have used the values of $P_e = 3 \cdot 10^{-4}$ and $\tilde{\rho} = \exp\left(-1.6^2\right)$ for our method. The parameters for the rest of the methods have been set using the same procedure.

Given a query graph $\mathbf{G}$, we compute its similarity to a database graph $\mathbf{H}$ using the following measure

$$\mathcal{F}_{\mathbf{GH}} = \frac{\max_S \mathcal{F}\left(\mathbf{G},\mathbf{H};S\right)}{\max\left(\mathcal{F}_{\mathbf{GG}},\mathcal{F}_{\mathbf{HH}}\right)} \tag{8.29}$$

where, $\mathcal{F}\left(\mathbf{G},\mathbf{H};S\right) = \ln P\left(\mathbf{G}|S,\Phi\right)$ is the incomplete log-likelihood of the observed graph $\mathbf{G}$, assuming $\mathbf{H}$ as the missing data graph; $\mathcal{F}_{\mathbf{GG}} = \mathcal{F}\left(\mathbf{G},\mathbf{G};I_{\mathbf{G}}\right)$ and $\mathcal{F}_{\mathbf{HH}} = \mathcal{F}\left(\mathbf{H},\mathbf{H};I_{\mathbf{H}}\right)$, being $I_{\mathbf{G}}$ and $I_{\mathbf{H}}$ the identity correspondence indicators defining self-matchings. This results in a normalized measure $\mathcal{F}_{\mathbf{GH}} \in [0,1]$ that equals to one in the case of a self-matching between two identical graphs and, moves towards zero as they become different.

	Resid		Boat		New York		East Park	
Method	MPE	%	MPE	%	MPE	%	MPE	%
Corr	835	27	24.5	76	31	67	463	43
ICP	40.3	87	21	100	19.4	100	88	100
EM-Soft	1.5	69	**0.68**	72	**0.69**	91	**1.13**	75
Dual-Step	**1.3**	72	1.7	62	0.7	91	153	25
ICP+RANSAC	12.3	54	10.7	64	10.9	76	98	41
ICP+GTM	32.5	61	10.5	70	2.9	70	327	45

Table 8.1: Mean Projection Error (MPE) and percentage of matched points in the origin images (%).

Method	Resid	Boat	New York	East Park
Corr	0.54	0.55	0.26	0.55
ICP	0.22	0.18	0.13	0.2
EM-Soft	378	449	73	438
Dual-Step	3616	3794	1429	3027
ICP+RANSAC	0.29	0.24	0.18	0.3
ICP+GTM	0.23	0.25	0.17	0.27

Table 8.2: Computational times (in seconds). The number of points of the origin ($N1$) and destination ($N2$) images in each case are: *Resid* $N1 = 55, N2 = 48$; *Boat* $N1 = 50, N2 = 61$; *New York* $N1 = 34, N2 = 34$; *East Park* $N1 = 44, N2 = 67$.

	Resid		Boat		New York		East Park	
Method	MPE	%	MPE	%	MPE	%	MPE	%
*EM-Soft**	2619	87	25.3	100	23.9	100	56.8	95
*Dual-Step**	26.2	83	19.5	100	1.15	91	332	100
Unified	39.8	69	12.3	86	3.04	88	1104	75
GradAssig	174	85	60.8	100	14.8	94	1716	100

Table 8.3: Mean Projection Error (MPE) and percentage of matched points (%) obtained without outlier rejection mechanisms.

Method	Resid	Boat	New York	East Park
EM-Soft (Matlab)	378	449	73	438
EM-Soft (C)	15.5	19.5	2.1	19.1

Table 8.4: Computational times of the Matlab and C implementations of our method.

Note that the maximization in the numerator of equation (8.29) has the same form as the log-likelihood maximization of equation (8.13) performed by our EM algorithm (section 8.4).

Performance is assessed through precision-recall plots. We compute the pairwise similarities between all the graphs in the database thus obtaining, for each query graph, a list of retrievals ordered by similarity. Suppose that our database contains C classes with N graphs each. We can define a retrieval of depth r as the first r graphs from each ordered list. Note that the number of elements retrieved by such an operation is rCN.

Precision is then defined as the fraction of retrieved graphs that are relevant in a retrieval of depth r. This is,

$$\text{precision} = \frac{\#\text{relevant}\,(r)}{rCN} \tag{8.30}$$

where $\#\text{relevant}\,(r)$ is the number of retrieved graphs that agree with the class of their respective queries, in a retrieval of depth r.

Recall is defined as the fraction of the relevant graphs that are successfully retrieved by retrieval of depth r. This is,

$$\text{recall} = \frac{\#\text{relevant}\,(r)}{CN^2} \tag{8.31}$$

where CN^2 is the maximum number of relevant graphs possible.

Precision-recall plots are generated by varying r in the range $[1 \ldots CN]$.

GREC Graphs

We have performed retrieval experiments on the GREC subset of the IAM Graph Database Repository (Riesen & Bunke, 2008). This subset is composed by 22 classes of 25 graphs each. Figure 8.16 shows an example graph of each class. Some classes show considerable inter-class similarities as well as significant intra-class variations such as missing or extra nodes, non-rigid deformations, scale differences and structural disruptions. See for example, the graphs in figure 8.17.

We have compared our method (*EM-Soft*) to the purely structural method *GradAssig* (section 4.3.2) and the geometric + structural methods *Dual-Step* (Cross & Hancock, 1998) (section 4.5.1) and *Unified* (Luo & Hancock, 2003) (section 4.5.2). We have included two additional pure geometric methods in order to provide evidence of the benefits of the combined methods. On one hand, we have used our method with an ambiguous structural model (i.e., $P_e = 0.5$). On the other hand, we have implemented a point-set registration algorithm (*EM-reg*) using the following EM update rule.

$$\Phi^{(n+1)} = \arg\max_{\Phi} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} P\left(v_\alpha | u_a, \Phi^{(n)}\right) \ln P\left(u_a, v_\alpha | \Phi\right) \tag{8.32}$$

where $P\left(u_a, v_\alpha | \Phi\right)$ is a multivariate Gaussian function of the point-position errors given the alignment parameters.

For each method, we have used the equivalent analog of the normalized similarity measure of equation (8.29) according to their models. All the approaches have been initialized by the tentative correspondences found as explained in

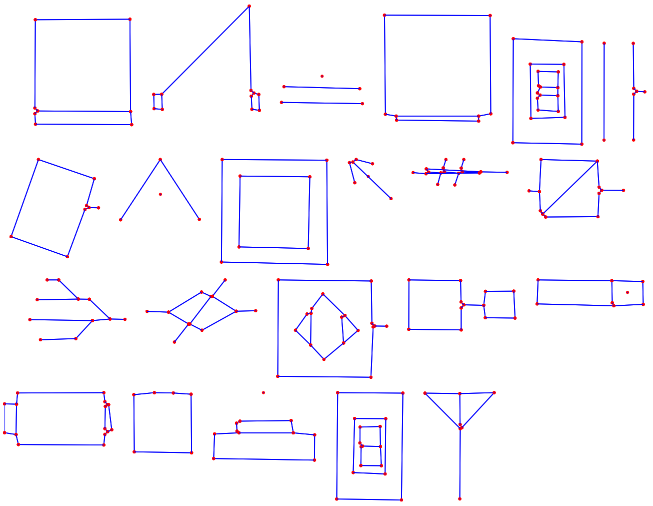
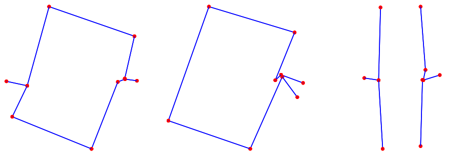


Figure 8.16: An example graph of each class of the GREC database (Riesen & Bunke, 2008). Nodes are represented as red dots, while edges as blue lines.



(a) Class 7 (b) Class 7 (c) Class 6

Figure 8.17: Compared in an affine-invariant way, graphs 8.17(a) and 8.17(c) show a similar node-set arrangement, although they are from different classes. They present however, slight differences between their structure. On the other hand, although graphs 8.17(a) and 8.17(b) are from the same class, we can see missing and extra nodes with respect to each other, while still having some differences between their structure. With these considerations, classification is not straightforward.

appendix C. The methods that do not use correspondence parameters have been initialized by the alignment parameters that minimize the pairwise point-position errors according to the aforementioned correspondences.

Figure 8.18 shows the precision-recall plots obtained by varying the depth of the retrieval from 1 to 550 (the total number of graphs in the database).

25 Shapes Database

We have performed retrieval experiments on the database of 25 binary shape images of figure 8.19. Our aim here is to evaluate the recognition abilities of several general-purpose graph matching approaches. Therefore, we have not used databases containing more specific types of deformations such as articulations because of the limitations imposed by the affine model assumptions.

We have used the skeleton pruning approach by Bai *et al.* (2007) in order to obtain the skeletal representations. Graphs have been generated by placing the

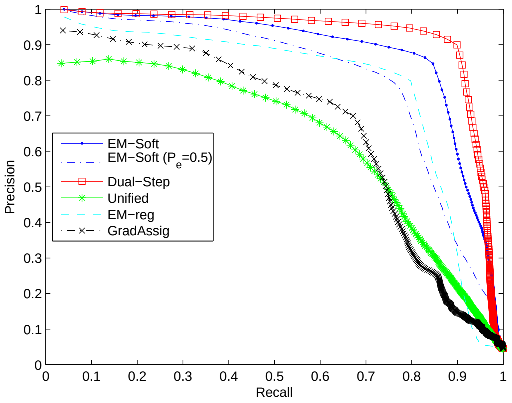


Figure 8.18: Precision-recall plots in the GREC database.

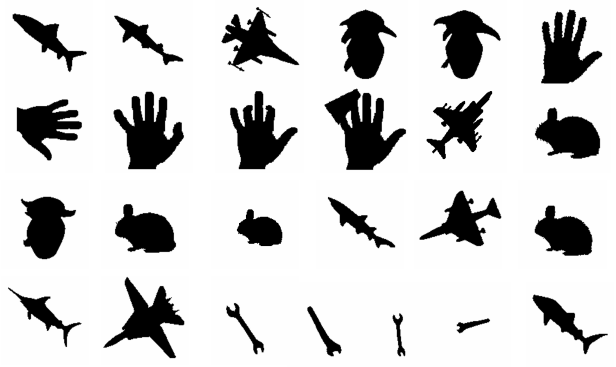


Figure 8.19: This database is divided into 6 classes: *Shark* (5 instances), *Plane* (4 instances), *Gen* (3 instances), *Hand* (5 instances), *Rabbit* (4 instances) and *Tool* (4 instances).

nodes at the end and intersection skeletal points and, the edges so as to fit to the rest of body of the skeleton. Figure 8.20 shows the graphs extracted from the above database.

All the approaches have been initialized with the tentative correspondences found as explained in appendix C.

We have implemented an affine-invariant template matching method in order to evaluate the benefits of using the structural abstractions instead of using directly the binary images. We evaluate the similarity between two registered binary images on the basis of their overlapping shape areas. Affine registration of the binary images is performed according to the tentative correspondences found by the method in appendix C.

Figure 8.21 shows the precision-recall plots of the graph matching approaches *EM-Soft*, *Dual-Step* and *Unified* and, the affine-invariant template matching (*TM*).

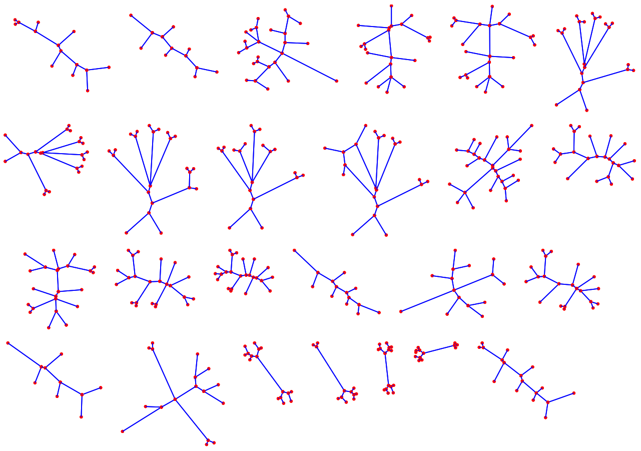


Figure 8.20: Graphs generated from the skeletons of the 25 shapes of figure 8.19.

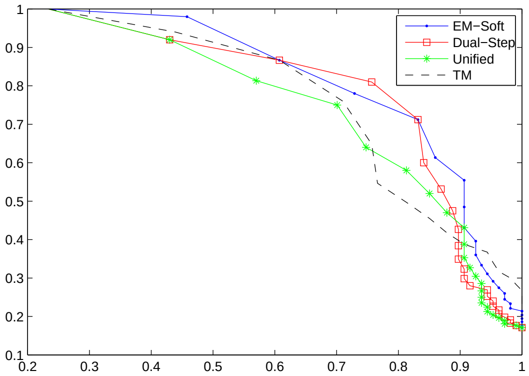


Figure 8.21: Precision-recall plots in the 25-shapes database.

8.7 Discussion

In the matching experiments under non-rigid deformations our method has shown to be the most effective among the compared graph-matching methods. Moreover, it shows a computational time in the typical range of the graph-matching methods. *Dual-Step* obtains a higher correct correspondence rate than *Unified*. However, its computational time is higher as well.

The matching experiments in the presence of outliers show that our method outperforms the compared ones. *Dual-Step* performs as effectively as *RANSAC*. Moreover, while outlier rejectors are specifically designed for these type of experiments, *Dual-Step* has a wider applicability.

The matching experiments on real images show that our method performs generally better than the others. *Dual-Step* performs, in most cases, similarly as ours but with higher computational times. It is worth mentioning that the considerable computational times required by *Dual-Step* are mainly due to the bottleneck that represents its outlier detection scheme. The graph-matching methods generally find more dense correspondence-sets. The ensembles *ICP+RANSAC* and *ICP+GTM* do not perform as effectively as the graph-matching methods but they do it faster.

Furthermore, we show how different methods benefit from outlier rejection in a real world application.

The efficiency shown by the C implementation of our method suggests that, while the outlier rejectors are appropriate for a regular real-time operation, it is feasible to use our method in specific moments when more effectiveness is required.

The dictionary-based structural model of the *Dual-Step* (Cross & Hancock, 1998) has demonstrated to be the most effective in the retrieval experiments on the GREC database. Our method shows a performance decrease with respect to *Dual-Step*. *Unified* is unable to deal with the type of graphs used in this experiment.

Neither the pure geometric methods nor the pure structural (*GradAssig*) are as accurate as *Dual-Step* and *EM-Soft* in the precision-recall scores. This demonstrates the benefits of combining both sources of information as opposed to using them separately. Particularly revealing of this fact is the comparison between the two versions of our method.

In the 25-shapes database the proposed method and the *Dual-Step* obtain similar scores. The affine-invariant template matching method only retrieves correctly the most similar instances of each class. As we increment the depth of the retrieval and more significant intra-class variations appear, the direct comparison of templates experiments a decrease in performance with respect to our structural approach. This shows the benefits of using structural representations as opposed to template-based strategies in the present application. The limitations of the affine model assumptions prevents us from using shape databases presenting further deformations such as larger articulations.

8.8 Conclusions

We have presented a graph matching method aimed at solving the point-set correspondence problem. Our model accounts for relative structural and geometrical measurements which keep parallelism with the compatibility coefficients of the Probabilistic Relaxation approaches. This contrasts with other approaches that use absolute position coordinates (Cross & Hancock, 1998; Luo & Hancock, 2003). Unlike other approaches (Cross & Hancock, 1998; Luo & Hancock, 2003, 2001; Wilson & Hancock, 1997), our underlying correspondence variable is continuous. To that end, we use Softassign to solve the individual assignment problems thus, enforcing two-way constraints as well as being able to control the level of discretization of the solution. Moreover, gradually pushing from continuous to discrete states reduces the chances of getting trapped in local minima (Gold & Rangarajan, 1996). We develop mechanisms to smoothly detect and remove outliers.

In contrast to other approaches such as *Unified* (Luo & Hancock, 2003) (section 4.5.2) and *Dual-Step* (Cross & Hancock, 1998) (section 4.5.1), the proposed approach has the distinguished properties that it uses Softassign to estimate the continuous correspondence indicators and it is based on a model of relational geometrical measurements. Such properties demonstrate to confer the proposed approach a better performance in many of the experiments presented.

Our method is controlled by two parameters, namely, an outlying threshold probability ρ and a probability of structural error P_e. We have not found any particular dependence of the parameter ρ to a specific application and hence, we have used the same value in all the experiments. On the contrary, the parameter P_e scales the contribution of the structural component of our model which is to be compared to the geometric part, and so, there is a dependence of this parameter to the type of graphs (Delaunay triangulations, morphologically-induced graphs,...). For this reason we have needed two different values of P_e in the case of Delaunay triangulations and morphologically-induced graphs in the experimental evaluation.

Chapter 9

Smooth Point-set Registration using Neighboring Constraints

9.1 Introduction

Alignment of point-sets is frequently used in *pattern recognition* when objects are represented by sets of coordinate points. The idea behind is to be able to compare two objects regardless the effects of a given transformation model on their coordinate data. This is at the core of many object recognition applications where the objects are defined by coordinate data (e.g., medical image analysis, shape retrieval, ...), learning shape models (Cootes *et al.*, 1995; Dryden & Mardia, 1998) or reconstructing a scene from various views (Hartley & Zisserman, 2000).

Given that the correspondences are known, there is an extensive work done towards the goal of finding the alignment parameters that minimize some error measure. To cite a few, Dryden & Mardia (1998); Kendall (1984) deal with isometries and similarity transformations; Berge (2006); Umeyama (1991) deal with Euclidean transformations (i.e. excluding reflections from isometries); Haralick *et al.* (1989) deals with similarity and projective transformations; and Hartley & Zisserman (2000) deals exclusively with projective transformations.

However, the point-set alignment problem is often found in the more realistic setting of unknown point-to-point correspondences. This problem becomes then a *registration problem*, this is, one of jointly estimating the alignment and correspondence parameters. Although non-iterative algorithms exist for specific types of transformation models (Ho & Yang, 2011), this problem is usually solved by means of non-linear iterative methods that, at each iteration, estimate correspondence and alignment parameters. Despite being more computationally demanding, iterative methods are more appealing to us than the direct ones due to its superior tolerance to noise and outliers.

We distinguish between two families of approaches at solving this registration problem. In the first family, each point in one point-set is influenced only by its nearest point in the other point-set. This is the case of the popular *Itera-*

tive Closest Point (ICP) algorithm introduced by Besl & McKay (1992) (section 3.1.1). Although ICP is attractive for its efficiency, it can be easily trapped in local minima due to the strict selection of the best point-to-point assignments. This makes ICP to be particularly sensitive to initialization. In the second family of approaches, each point is influenced by all the other points by means of a *multiply-linked* utility measure. This is the case of the approaches based on the *Expectation-Maximization* (EM) algorithm (Dempster *et al.*, 1977), and also of those based on *Softassign* (sections 3.3 and 3.4, respectively). The former ones have the advantage of offering statistical insights of such decoupled estimation processes while the latter ones benefit from the well-known robustness and convergence properties of the Softassign embedded within deterministic annealing procedures.

Within this family of approaches, we can distinguish a branch of methods that generalizes from point-sets to graph-based representations thus allowing to take into account the neighboring relations between points. Such *graph matching* approaches benefit from the extended representation skills of graphs with respect to point-sets.

To cite some examples of graph matching methods using statistical estimation, Cross & Hancock (1998) presented an approach for graph matching and point-set alignment within the EM framework (section 4.5.1). They included two types of geometric transformations, namely, affinities and projectivities. They used a kind of structural entities called *cliques* in order to enforce structural consistency constraints. An important limitation of this approach is the high computational demand of the clique-based structural model. From our personal experience, this approach renders impractical for graphs with more than 50 or 60 nodes. Luo & Hancock (2003) proposed an EM-like approach for graph matching and point-set alignment based on a cross-entropy measure (section 4.5.2). They used Procrustes analysis in order to estimate the similarity transformation parameters. They proposed a model of structural errors based on a Bernoulli distribution.

A remarkable technique aimed at the continuous optimization of a correspondence variable is *Softassign*. This technique combines the relaxation of the discrete constraints on the assignment variables together with a method of two-way normalization (Sinkhorn, 1964). Softassign is run within an annealing procedure that gradually pushes from continuous to discrete solutions, a technique which is known to avoid poor local minima. Two worth mentioning approaches that use this technique are *Graduated Assignment* by Gold & Rangarajan (1996) and *Softassign Procrustes* by Rangarajan *et al.* (1997) (sections 4.3.2 and 3.4, respectively). The former is aimed at structural graph matching by maximizing the number of matched edges between two graphs and the latter is aimed at point-set registration by minimizing the *Procrustes distance* (Dryden & Mardia, 1998) (section 2.3.1) between two point-sets over correspondences and similarity alignment parameters.

We try to bridge the gap between the EM-based and the Softassign-based approaches by formulating the graph matching problem within a principled statistical framework, while benefiting from the desirable properties of the Softassign and deterministic annealing ensemble. To that end, we estimate Maximum Likelihood (ML) alignment and correspondence parameters of a mixture model in dual-steps of an EM algorithm. Our mixture model assumes that geometric and structural errors follow Gaussian and Bernoulli distributions, respectively.

Correspondence problem is approximated as a succession of linear assignment problems which are solved using Softassign. This way, we are able to use continuous correspondence variables as opposed to other approaches that use discrete ones (Cross & Hancock, 1998; Luo & Hancock, 2003). Outlier rejection is modeled as a smooth assignment to the null node within the whole annealing procedure.

The outline of this chapter is the following. In section 9.2 we formulate the matching problem as one of mixture modelling with missing data and propose our mixture model. In section 9.3 we derive the EM algorithm for our model. Section 9.4 presents the methodology used to reject outliers. In section 9.5 we present some experiments and results, and finally in section 9.6 we provide some concluding remarks.

9.2 A Mixture Model

Consider two graph representations $\mathbf{G} = (\mathcal{U}, D, \mathcal{X})$ and $\mathbf{H} = (\mathcal{V}, M, \mathcal{Y})$ extracted from two images.

The node-sets $\mathcal{U} = \{u_a,\ a \in \mathcal{I}\}$ and $\mathcal{V} = \{v_\alpha,\ \alpha \in \mathcal{J}\}$ contain the symbolic representations of the nodes, where $\mathcal{I} = 1, \ldots, |\mathcal{U}|$ and $\mathcal{J} = 1, \ldots, |\mathcal{V}|$ are their index-sets.

The vector-sets $\mathcal{X} = \{\mathbf{x}_a,\ a \in \mathcal{I}\}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha,\ \alpha \in \mathcal{J}\}$, contain the column vectors $\mathbf{x}_a = (x_a^V, x_a^H)$ and $\mathbf{y}_\alpha = (y_\alpha^V, y_\alpha^H)$ of the two-dimensional coordinates (vertical and horizontal) of each node.

The adjacency matrices D and M contain the edge-sets, encoding some kind of relation between pairs of nodes (e.g., connectivity or spatial proximity). Hence, $D_{ab} = \begin{cases} 1 & \text{if } u_a \text{ and } u_b \text{ are linked by an edge} \\ 0 & \text{otherwise} \end{cases}$ (the same applies for $M_{\alpha\beta}$).

We use continuous correspondence indicators S so, we denote as $s_{a\alpha} \in S$, the probability of node $u_a \in \mathcal{U}$ being in correspondence with node $v_\alpha \in \mathcal{V}$.

It is satisfied that

$$\sum_{\alpha \in \mathcal{J}} s_{a\alpha} \leq 1,\ a \in \mathcal{I} \tag{9.1}$$

where, $1 - \sum_\alpha s_{a\alpha}$ is the probability of node u_a being an outlier.

Our aim is to recover the correspondence indicators S and the alignment parameters Φ that maximize the *observed-data* likelihood of the data-graph $P(\mathbf{G}|S, \Phi)$. Within this setting, constraints on the data-graph $\mathbf{G}$ are evaluated on the model-graph $\mathbf{H}$. To make this problem tractable, we introduce the *hidden variables*, namely, the corresponding model graph nodes $v_\alpha \in \mathcal{V}$.

By assuming that the observations are independent and identically distributed, the observed-data likelihood writes

$$P(\mathbf{G}|S, \Phi) = \prod_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} P(u_a, v_\alpha | S, \Phi) \tag{9.2}$$

Following a similar development than Luo & Hancock (2001) we factorize, using the Bayes rules, the *complete-data* likelihood in the right hand side of equation (9.2) into terms depending on individual correspondence indicators, in the following way.

$$P\left(u_a, v_\alpha | S, \Phi\right) = K_{a\alpha} \prod_{b \in \mathcal{I}} \prod_{\beta \in \mathcal{J}} P\left(u_a, v_\alpha | s_{b\beta}, \Phi\right) \tag{9.3}$$

where $K_{a\alpha} = \left[1/P(u_a|v_\alpha,\Phi)\right]^{|\mathcal{I}|\times|\mathcal{J}|-1}$. If we assume that conditional dependence of data-graph node u_a can only be taken into account in the presence of the correspondence matches S, then $P\left(u_a|v_\alpha, \Phi\right) = P\left(u_a\right)$. Further assuming equiprobable priors $P\left(u_a\right)$, we can safely discard these quantities in the maximization of equation (9.2), since they do not depend either on S nor Φ.

We propose a measure for the complete-data likelihood of equation (9.3) that combines a model of structural errors based on a Bernoulli distribution augmented with a model of geometric errors based on a Gaussian distribution.

With regards to the structural relations, Luo & Hancock (2001) proposed to model the likelihood of an observed relation given the hypothesis on the correspondences using a Bernoulli distribution with parameters S. This is, given two corresponding pairs of nodes $u_a, u_b \in \mathcal{U}$ and $v_\alpha, v_\beta \in \mathcal{V}$, they assumed that there will be edge-discordance (i.e., $D_{ab} = 0 \vee M_{\alpha\beta} = 0$) with a fixed (low) probability of error P_e. Otherwise, there will be edge-concordance with probability $1 - P_e$. This is,

$$P\left(u_a, v_\alpha | s_{b\beta}\right) = \begin{cases} (1-P_e) & \text{if } D_{ab}=1 \wedge M_{\alpha\beta}=1 \wedge s_{b\beta}=1 \\ P_e & \text{otherwise} \end{cases} \tag{9.4}$$

With regards to the geometrical measurements, it is reasonable to consider that point-position errors between corresponding points follow a Gaussian density. In the case of no correspondence, we use a fixed probability ρ that will model the outlier process. This is,

$$P\left(u_b | s_{b\beta}, \Phi\right) = \begin{cases} P_{b\beta}^{(\Phi)} & \text{if } s_{b\beta}=1 \\ \rho & \text{otherwise} \end{cases} \tag{9.5}$$

where $P_{b\beta}^{(\Phi)}$ is a Gaussian measurement on the point-position errors with parameters Φ. This is,

$$P_{b\beta}^{(\Phi)} = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\left\|\mathbf{x}_b - \mathcal{T}\left(\mathbf{y}_\beta; \Phi\right)\right\|_\Sigma^2\right] \tag{9.6}$$

where $\mathcal{T}\left(\mathbf{y}_\beta; \Phi\right)$ represents the geometric transformation of model point $\mathbf{y}_\beta$ according to alignment parameters Φ, and $\|\mathbf{d}\|_\Sigma^2 = \mathbf{d}^\top \Sigma^{-1} \mathbf{d}$ is the squared Mahalanobis distance with covariance matrix Σ, with $\mathbf{d}$ a column vector. As opposed to the standard Gaussian modeling approach, here the means are parameterized by the alignment parameters which enforce prior knowledge about the transformation that exists between the two sets of points.

We propose a more fine-grained likelihood measure than that of equation (9.4) by considering that it is appropriate to weight the likelihood of an observed relation with the geometric likelihood term defined in equation (9.5).

In the case of discrete correspondence indicators (i.e., $s_{b\beta} = \{0, 1\}$), the proposed density writes

$$P\left(u_a, v_\alpha | s_{b\beta}, \Phi\right) = \begin{cases} (1-P_e)P_{b\beta}^{(\Phi)} & \text{if } D_{ab}{=}1 \wedge M_{\alpha\beta}{=}1 \wedge s_{b\beta}{=}1 \\ P_e P_{b\beta}^{(\Phi)} & \text{if } (D_{ab}{=}0 \vee M_{\alpha\beta}{=}0) \wedge s_{b\beta}{=}1 \\ P_e \rho & \text{if } s_{b\beta}{=}0 \end{cases} \tag{9.7}$$

We extrapolate to continuous correspondence indicators by exploiting each case of equation (9.7) as exponential indicators. This is,

$$\begin{aligned} &P\left(u_a, v_\alpha | s_{b\beta}, \Phi\right) = \\ &\qquad \left[(1-P_e)\,P_{b\beta}^{(\Phi)}\right]^{D_{ab}M_{\alpha\beta}s_{b\beta}} \left[P_e P_{b\beta}^{(\Phi)}\right]^{(1-D_{ab}M_{\alpha\beta})s_{b\beta}} \left[P_e\rho\right]^{(1-s_{b\beta})} \end{aligned} \tag{9.8}$$

By using the density measurement of equation (9.8), the final expression for the complete-data likelihood of equation (9.3), expressed in the exponential form, is

$$\begin{aligned} &P\left(u_a, v_\alpha | S, \Phi\right) = \\ &\qquad \exp\left[\sum_{b\in\mathcal{I}}\sum_{\beta\in\mathcal{J}} s_{b\beta} D_{ab} M_{\alpha\beta} \ln\left(\tfrac{1-P_e}{P_e}\right) + s_{b\beta}\ln\left(\frac{P_{b\beta}^{(\Phi)}}{\rho}\right) + \ln\left(P_e\rho\right)\right] \end{aligned} \tag{9.9}$$

9.3 Expectation Maximization

The EM algorithm has been previously used by other authors to solve the graph matching problem (Cross & Hancock, 1998; Luo & Hancock, 2001). We seek the optimal alignment parameters $\Phi^\star$ and the correspondence indicators $S^\star$ that maximize our observed-data log-likelihood, i.e., $\ln P\left(\mathbf{G}|S, \Phi\right)$. This is,

$$\{\Phi^\star, S^\star\} = \arg\max_{\Phi, S} \sum_{a\in\mathcal{I}} \ln\left(\sum_{\alpha\in\mathcal{J}} P\left(u_a, v_\alpha | S, \Phi\right)\right) \tag{9.10}$$

Dempster *et al.* (1977) proposed to replace equation (9.10) by its *conditional expectation conditioned by the observed data* (section 3.2). It has been proven that maximizing the conditional expectation is equivalent at maximizing the observed-data log-likelihood. Accordingly, we seek the parameters $S^{(n+1)}, \Phi^{(n+1)}$ that maximize the following objective function

$$\begin{aligned} \{\Phi^{(n+1)}, S^{(n+1)}\} &= \arg\max_{\Phi, S} E_{\mathcal{V}}\left[\ln P\left(\mathbf{G}|S, \Phi\right)|\mathbf{G}\right] \\ &= \arg\max_{\Phi, S} \sum_{a\in\mathcal{I}}\sum_{\alpha\in\mathcal{J}} P\left(v_\alpha | u_a, S^{(n)}, \Phi^{(n)}\right) \ln P\left(u_a, v_\alpha | S, \Phi\right) \end{aligned} \tag{9.11}$$

where $P\left(v_\alpha | u_a, S^{(n)}, \Phi^{(n)}\right)$ are the posterior probabilities of the missing data given the most recent available parameters $S^{(n)}, \Phi^{(n)}$.

The basic idea is to alternate between Expectation and Maximization steps until convergence is reached. The expectation step involves computing the posterior probabilities of the missing data using the most recent available parameters. In the maximization phase, the parameters are updated in order to maximize the expected log-likelihood of the observed data.

9.3.1 Expectation

In the expectation step, the posterior probabilities of the missing data (i.e., the corresponding model-graph v_α estimates) are computed using the current parameter estimates $S^{(n)}, \Phi^{(n)}$.

The posterior probabilities are computed, according to the Bayes rule, using the following expression

$$\begin{aligned} P\left(v_\alpha | u_a, S^{(n)}, \Phi^{(n)}\right) &= \frac{P\left(u_a, v_\alpha | S^{(n)}, \Phi^{(n)}\right)}{\sum_{\alpha'} P\left(u_a, v_{\alpha'} | S^{(n)}, \Phi^{(n)}\right)} \\ &= \frac{\exp\left[\sum_{b,\beta} s_{b\beta} D_{ab} M_{\alpha\beta} \ln\left(\frac{1-P_e}{P_e}\right)\right]}{\sum_{\alpha'} \exp\left[\sum_{b,\beta} s_{b\beta} D_{ab} M_{\alpha'\beta} \ln\left(\frac{1-P_e}{P_e}\right)\right]} \overset{def}{=} \omega_{a\alpha}^{(n)} \end{aligned} \tag{9.12}$$

Note that when substituting the complete-data likelihoods by their expressions of equation (9.9), the last two terms in the summations are canceled out by the quotient since they do not depend either on nodes u_a or v_α. As we have stated in equation (9.5), the hidden corresponding nodes v_α do not affect the point-position errors. As consequence, the missing data posteriors are revealed as strictly structural measurements. Point-position errors, which are conditionally dependant on the correspondence indicators $s_{b\beta}$, will affect to the ML estimate of the correspondence parameters as we will see later.

9.3.2 Maximization

It is a well-established strategy to implement the maximization step into a series of *conditional maximization* steps (Horaud *et al.*, 2011). Then, it turns into an instance of the *expectation conditional maximization* (ECM) (Meng & Rubin, 1993) algorithm which still shares the desirable convergence properties of EM. According to ECM, maximization of equation (9.11) can be decomposed into three steps. First, maximize over the alignment parameters, next compute empirical covariances using the newly estimated alignment parameters $\Phi^{(n+1)}$, and finally maximize over the correspondence indicators while using the newly estimated empirical covariances $\Sigma^{(n+1)}$ and alignment parameters $\Phi^{(n+1)}$.

Maximum Likelihood Alignment Parameters

We seek the alignment parameters $\Phi^{(n+1)}$ that maximize equation (9.11). We use the expressions in equations (9.12) and (9.9) for the posterior probability and conditional likelihood terms, respectively. Discarding the terms constant with respect to the alignment parameters we obtain the following expression

$$\Phi^{(n+1)} = \arg\max_{\Phi} \sum_{a \in \mathcal{I}} \sum_{\alpha \in \mathcal{J}} \omega_{a\alpha}^{(n)} \sum_{b \in \mathcal{I}} \sum_{\beta \in \mathcal{J}} s_{b\beta}^{(n)} \ln\left(\frac{P_{b\beta}^{(\Phi)}}{\rho}\right) \tag{9.13}$$

Rearranging and further removing other terms constant with respect to the alignment parameters, we get

$$
\begin{aligned}
\Phi^{(n+1)} &= \arg\max_{\Phi} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta}^{(n)} \ln\left(\frac{P_{b\beta}^{(\Phi)}}{\rho}\right) \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \\
&= \arg\max_{\Phi} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta}^{(n)} \ln P_{b\beta}^{(\Phi)} \qquad (9.14) \\
&= \arg\min_{\Phi} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta}^{(n)} \left\| \mathbf{x}_b - \mathcal{T}\left(\mathbf{y}_\beta; \Phi\right) \right\|_{\Sigma^{(n)}}^2 \qquad (9.15)
\end{aligned}
$$

In going from equation (9.14) to (9.15) we turned the maximization into a minimization by substituting the geometrical probability term $P_{b\beta}^{(\Phi)}$ by its expression in (9.6) while discarding the constant terms.

Note that the alignment parameters do not depend on the posterior probability terms $\omega_{a\alpha}^{(n)}$ but on the correspondence variables $s_{b\beta}^{(n)}$. This is because, as stated in equation (9.5), point position errors are evaluated on the basis of the correspondence variables instead of the missing-data posteriors.

Optimal transformation parameters are computed from equation (9.15) as explained in section 3.5. In our experiments we will use either a similarity model (section 3.5.1) or a projective model (section 3.5.3).

Empirical Covariances

We compute the variances using the newly estimated registration parameters $\Phi^{(n+1)}$ according to the following expression.

$$
\sigma^2 = \frac{\sum_{b,\beta} s_{b\beta}^{(n)} \left(\mathbf{x}_b - \mathcal{T}\left(\mathbf{y}_\beta; \Phi^{(n+1)}\right)\right)^\top \left(\mathbf{x}_b - \mathcal{T}\left(\mathbf{y}_\beta; \Phi^{(n+1)}\right)\right)}{\sum_{b,\beta} s_{b\beta}^{(n)}} \qquad (9.16)
$$

and set isotropic covariance matrix as $\Sigma^{(n+1)} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$.

Maximum Likelihood Correspondence Indicators

One of the key points in our work is to approximate the solution of the graph matching problem by a succession of easier assignment problems. As it is done in Graduated Assignment (Gold & Rangarajan, 1996), we use *Softassign* to solve the assignment problems in a continuous way.

According to the EM development, we compute the correspondence indicators $S^{(n+1)}$ that maximize equation (9.11). Substituting equations (9.12) and (9.9) into (9.11) and discarding the constant term $(\ln(P_e\rho))$, we obtain

$$
\begin{aligned}
&S^{(n+1)} = \\
&\arg\max_{S} \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta} \left[D_{ab} M_{\alpha\beta} \ln\left(\frac{1-P_e}{P_e}\right) + \ln\left(\frac{P_{b\beta}^{(n+1)}}{\rho}\right) \right] \qquad (9.17)
\end{aligned}
$$

where $P_{b\beta}^{(n+1)}$ is the Gaussian of the point errors of equation (9.6) using the recently estimated alignment parameters $\Phi^{(n+1)}$ and covariance matrix $\Sigma^{(n+1)}$.

Rearranging terms we obtain the following assignment problem (Gold & Rangarajan, 1996)

$$
S^{(n+1)} = \arg\max_{S} \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta} B_{b\beta} \qquad (9.18)
$$

where

$$\begin{align}
B_{b\beta} &= \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} \left[D_{ab} M_{\alpha\beta} \ln\left(\tfrac{1-P_e}{P_e}\right) + \ln\left(\frac{P_{b\beta}^{(n+1)}}{\rho}\right) \right] \nonumber \\
&= \ln\left(\frac{P_{b\beta}^{(n+1)}}{\rho}\right) \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} + \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} D_{ab} M_{\alpha\beta} \ln\left(\tfrac{1-P_e}{P_e}\right) \tag{9.19} \\
&\simeq \ln\left(\frac{P_{b\beta}^{(n+1)}}{\rho}\right) + \sum_{a\in\mathcal{I}} \sum_{\alpha\in\mathcal{J}} \omega_{a\alpha}^{(n)} D_{ab} M_{\alpha\beta} \ln\left(\tfrac{1-P_e}{P_e}\right) \tag{9.20}
\end{align}$$

is the benefit value for the assignment $u_b \to v_\beta$.

We have observed a better stability of the algorithm when removing the summation $\sum_{a,\alpha} \omega_{a\alpha}^{(n)}$ in going from equation (9.19) to (9.20) which acts as a constant amplification term, specially when dealing with large graphs.

Notice how the two motivations underpinning our work, namely, the pure geometric and the pure structural, are clearly identified in the benefit measure of equation (9.20).

Computation of the correspondence indicators with Softassign consists in two steps (section 3.4):

1. Correspondence indicators are updated with the exponential of the benefit coefficients. This is,
$$s_{b\beta} = \exp\left[\mu B_{b\beta}\right] \tag{9.21}$$
where μ is a control parameter.

2. Two-way constraints are imposed by alternatively normalizing across rows and columns the matrix of exponentiated benefits. This is known as the *Sinkhorn normalization* (Sinkhorn, 1964) and, it is applied either until convergence or a predefined number of times. We have observed an improvement in the performance of the algorithm when applying Sinkhorn normalization to the missing data posteriors of equation (9.12) as well.

Softassign is run within an annealing procedure that increases the value of μ at each maximization step. Starting from low values of μ, the correspondence indicators $s_{b\beta}$ are gradually pushed from continuous to discrete values as μ increases.

9.4 Outlier Rejection

It is important to develop techniques aimed at detecting and rejecting outliers in order to minimize their influence.

We consider that a node $u_b \in \mathcal{U}$ (or $v_\beta \in \mathcal{V}$) is an outlier if there is not any node v_β, $\forall_{\beta\in\mathcal{J}}$ (or u_b, $\forall_{b\in\mathcal{I}}$) with a matching benefit $B_{b\beta}$ higher than a predefined threshold.

Outlier detection is handled as an assignment to (or from) the *null* node. Considering that the null node has no edges, and that all the probabilities $P_{b\beta}^{(\Phi)}$ involving the null node are equal to ρ, then the benefit values of equation (9.20) corresponding to the null assignments are equal to zero. We create an augmented benefit matrix $\tilde{B}$ by adding to B an extra row and column of zeros representing the benefits of the null assignments (i.e., $B_{b\emptyset}$, $\forall b \in \mathcal{I}$ and $B_{\emptyset\beta}$, $\forall \beta \in \mathcal{J}$).

Note that ρ establishes the threshold at which the terms $\ln\left(P_{b\beta}^{(\Phi)}/\rho\right)$ contribute positively (i.e., $\rho < P_{b\beta}^{(\Phi)}$) or negatively (i.e., $\rho > P_{b\beta}^{(\Phi)}$) to the benefit measure.

We apply Softassign (i.e., exponentiation and Sinkhorn normalization) to the augmented benefit matrix $\tilde{B}$. When performing Sinkhorn normalization we keep in mind that the null assignments are special cases that only satisfy one-way constraints and thus, there may be multiple nodes assigned to null in both graphs. Finally, the extra row and column are removed leading to the resulting matrix of correspondence parameters $S^{(n+1)}$.

As the control parameter μ of the Softassing increases, the rows and columns of $S^{(n+1)}$ associated to the outlier nodes, tend to zero. This fact reduces the influence of these nodes in the maximization phases of the next iteration that, in turn, lead to even lower benefits, and so on.

We still have to define the value of the outlying threshold ρ. From the first term of equation (9.20), we see that ρ is to be compared with $P_{b\beta}^{(n+1)}$. We consider therefore convenient to define it in terms of a multivariate Gaussian of a distance threshold. This is,

$$\rho = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\,\|\mathbf{d}\|_{\Sigma}^{2}\right] \tag{9.22}$$

where, $\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$, is an isotropic covariance variance matrix and, $\mathbf{d} = \left(d^V, d^H\right)^\top$ is a column vector with the vertical and horizontal thresholding distances.

Note that the quotient $P_{b\beta}^{(\Phi)}/\rho$ can be equivalently expressed as

$$\frac{P_{b\beta}^{(\Phi)}}{\rho} = \frac{\tilde{P}_{b\beta}^{(\Phi)}}{\tilde{\rho}} \tag{9.23}$$

where we have removed the constant multiplicative factor $1/2\pi|\Sigma|^{1/2}$ from ρ and $P_{b\beta}^{(\Phi)}$ in order to set $\tilde{\rho}$ and $\tilde{P}_{b\beta}^{(\Phi)}$.

If we express the thresholding distance proportionally to the standard deviations of the data (i.e., $\mathbf{d} = (N\sigma, N\sigma)$), the expression of ρ to be compared with $P_{b\beta}^{(n+1)}$ becomes

$$\tilde{\rho} = \exp\left\{-\frac{1}{2}\left[\left(\frac{N\sigma}{\sigma}\right)^2 + \left(\frac{N\sigma}{\sigma}\right)^2\right]\right\} = \exp\left(-N^2\right) \tag{9.24}$$

So, ρ is defined as a function of the number N of standard deviations permitted in the alignment errors, in order to consider a plausible correspondence.

9.5 Experiments and Results

We have performed matching experiments with synthetic and real data. Experiments with synthetic data consists on matching point-sets extracted from a fish and a Chinese character templates under nonrigid deformations, noise and outliers. Experiments with real data consists on matching point-sets extracted from images from various scenes across different zooms and rotations. In the following we introduce the graph matching methods used in the comparison.

Cross & Hancock (1998); Luo & Hancock (2003) presented graph-matching approaches to recover the correspondence and alignment parameters with decoupled statistical estimation processes. Cross & Hancock (1998) used the EM algorithm and incorporated a dictionary-based model of structurally consistent mappings and an outlier rejection mechanism grounded on the net effects of a node deletion in the re-configured graph (section 4.5.1). Luo & Hancock (2003) used the cross-entropy between the structural and geometrical error models as utility measure (section 4.5.2). Both approaches enforce the discrete update of the correspondence variable.

Gold & Rangarajan (1996) presented Graduated Assignment (section 4.3.2), a structural graph matching method that updates the correspondence variables $s_{a\alpha} \in S$ following an annealing scheme in the following way

$$s_{a\alpha}^{(n+1)} = \exp\left[\mu \sum_{b\in\mathcal{I}} \sum_{\beta\in\mathcal{J}} s_{b\beta}^{(n)} Q_{a\alpha b\beta}\right] \tag{9.25}$$

where μ is an annealing parameter that is gradually increased and $Q_{a\alpha b\beta}$ is the edge-compatibility coefficient for the assignment $(a, b) \to (\alpha, \beta)$. We have used the commonly adopted value $Q_{a\alpha b\beta} = c\, D_{ab} M_{\alpha\beta}$ that assigns a positive scalar c in the case of edge-concordance, and 0 otherwise. The updating equation (9.25) is followed by a Sinkhorn normalization on the matrix of correspondence variables S.

There is a noticeable parallelism between equation (9.25) from Graduated Assignment and equations (9.20) and (9.21) from our method. If we disregard any geometric measurement in our method by setting $P_{b\beta}^{(n+1)} = \rho$ for all b, β we obtain a pure structural version of our method which is equivalent to the aforementioned implementation of Graduated Assignment given the identifications $c = \ln\left[(1 - P_e)/P_e\right]$ and $s_{b\beta}^{(n)} = \omega_{b\beta}^{(n)}$.

A particular case of our method with a pure geometric motivation consists on using an ambiguous structural model, this is, set the value $P_e = 0.5$. This particular case reduces to iteratively computing the correspondence and alignment parameters according to the following steps: (1) from equations (9.20), (9.21) and (9.24), update S with the following expression

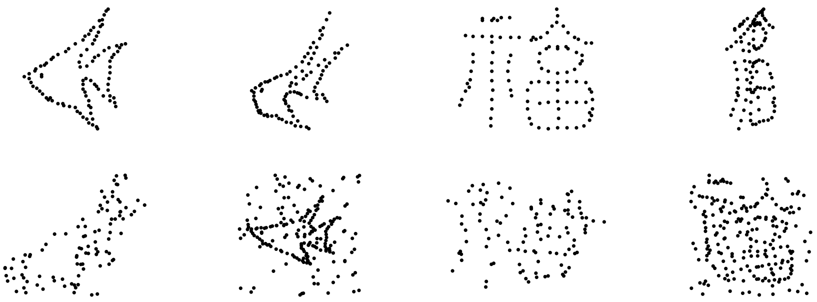
$$\begin{aligned} s_{b\beta} &= \exp\left[\mu\left(-\left\|x_b - \mathcal{T}(y_\beta;\Phi)\right\|_\Sigma^2 - \ln\tilde{\rho}\right)\right] \\ &= \exp\left[\mu\left(-\left\|x_b - \mathcal{T}(y_\beta;\Phi)\right\|_\Sigma^2 + N^2\right)\right] \end{aligned} \tag{9.26}$$

(2) normalize S across rows and columns having into account the extra row and column of the null assignments, (3) compute alignment parameters according to equation (9.15), and (4) increase μ and repeat steps (1-3) until μ reaches a predefined threshold.

It is worth pausing at this point to consider the analogies of this particular case of our method to a well-known method by Rangarajan *et al.* (1997), namely, the *Softassign Procrustes* (section 3.4). The essential difference with the Softassign Procrustes algorithm is that they use the squared Euclidean distance instead of the squared Mahalanobis distance. This way, their "robustness parameter", which is the analog of our N^2 term, is to be compared to the Euclidean distance. Unfortunately, they do not address the estimation of this parameter in their paper. On the contrary, we pose it in terms of the standard deviation which is a well-defined parameter in our method.

9.5.1 Synthetic Data

We have performed matching experiments on the dataset by Chui & Rangarajan (2000). This dataset contains perturbed instances of a fish and a Chinese character templates, consisting of 98 and 105 points, respectively. Perturbation levels range from mild to severe, with 100 different instances for each level. The types of perturbations are, (1) non-rigid deformations based on Gaussian radial basis functions (RBF) (Yuille & Grzywacz, 1989), (2) independent random noise and, (3) and a certain percentage of outliers ranging from 0% to 300%. A certain amount of ground-level non-rigid deformation is maintained in the random noise and outliers. See figure 9.1 for an example.



(a) Fish

(b) Chinese character

Figure 9.1: From top to down and left to right, the model templates and moderately perturbed instances due to non-rigid deformations, noise and outliers for (a) the fish template and, (b) the Chinese character template.

We have generated the graphs for the deformation and noise experiments following a mutual k-nearest-neighbour approach, with $k = 5$. This is, two points are joined by an edge if both points belong to the 5 nearest neighbors of the other point.

In the outlier experiments we follow the strategy by Zheng & Doermann (2006). This is, we place edges between the pairs of points presenting the M lower pair-wise distances. In average we want to place 5 edges for each point. Therefore, we have set $M = (5 \cdot n)/2$, where n is the number of points.

With such a strategy we aim to concentrate most of the edges among the inliers since, in the present data-set, outliers are comparably more scattered. Since graph matching methods try to maximize the number of matched edges, such a strategy allows to better discriminate between inliers and outliers.

In order to see the advantages of graphs with respect to point-sets we present the results of three graph matching methods and two point-set registration methods. All the methods using geometric measurements implement a similarity transformation model which provide a rough approximation to the true transformations undergone by the pattern templates.

The graph matching methods are: the one presented in this chapter, the Unified approach by Luo & Hancock (2003) and the Graduated Assignment by Gold & Rangarajan (1996), which is the equivalent of a pure structural version of our method (sections 4.5.2 and 4.3.2, respectively).

The two point-set registration methods are: a plain ICP (Besl & McKay, 1992) (section 3.1.1) and the approach presented in this chapter using an ambiguous structural model (i.e., $P_e = 0.5$) which is equivalent to the Softassign Procrustes by Rangarajan *et al.* (1997) (section 3.4).

ICP is a popular point-set registration method where each point is only influenced by its nearest neighbor in the other point-set. Points are iteratively associated with the nearest neighbor criterion and transformed using a mean square cost function.

The approach presented in this chapter falls within the category of multiply-linked approaches since each point is influenced by all the other points.

Since the ground truth matchings are available between each perturbed instance and the model templates, we assess the performance of each method by the *mean correspondence error*. For a given correspondence from a perturbed template point, the correspondence error is computed as the Euclidean distance between the ground truth point in the model template and the point which it is actually assigned to. As opposed to the correct correspondence rate, the mean correspondence error provides a qualitative measure for the incorrect matches.

Initial alignment parameters have been defined by the original spatial arrangement of the point-sets except for the Graduated Assignment that does not use them. Regarding the correspondence indicators, each point in one point-set has been initially assigned to its closest point in the other point-set for all the methods except ICP that does not use explicit correspondence variables.

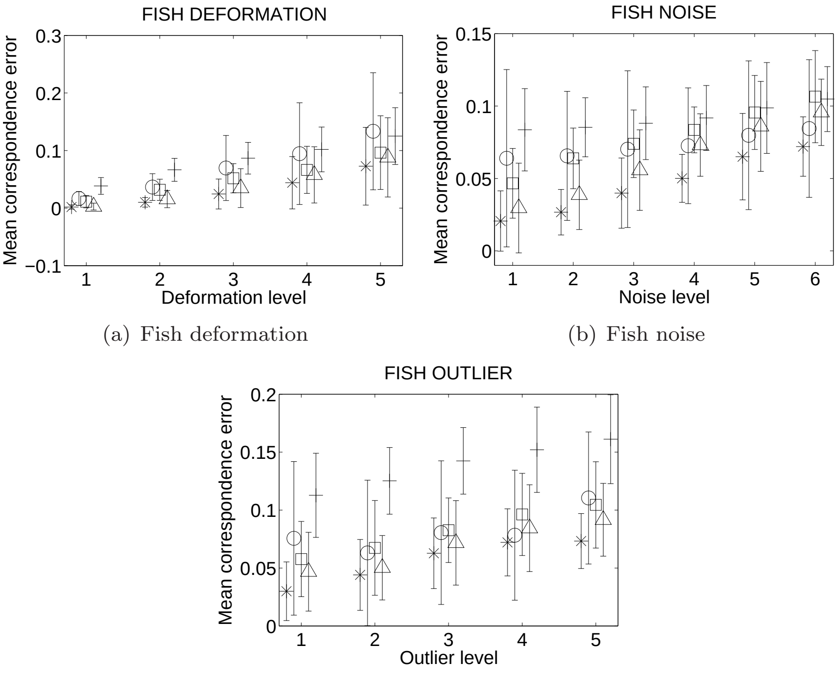
We have experimentally set the parameters $P_e = 0.03$ for the presented method and the Unified approach by Luo & Hancock (2003). With regards to the outlying threshold ρ, we have used the value $N = 1$ from equation (9.24) in the full and the pure geometric versions of our method. In the case of Graduated Assignment, we have experimentally set the compatibility coefficient in case of edge-concordance to $c = 3.47$ (i.e., $P_e = 0.03$).

Figures 9.2 and 9.3 show the results of each method for the fish and Chinese character.

The multiply-linked approaches have performed better than ICP in the presented experiments. Even though the ICP implementation used does not perform outlier rejection, its performance should not be considerably affected since the deformation and noise data-sets contain no outliers at all. Moreover, other methods with superior performance neither implement outlier rejection such as the Unified approach by Luo & Hancock (2003) and Graduated Assignment.

Comparison between the pure geometric version of our method (i.e., Softassign Procrustes) and Graduated Assignment (i.e., the pure structural) reveals that neighboring relations between points have resulted to be more robust than similarity-invariant point errors for matching purposes. Actually the comparison is quite fair since both approaches implement an annealing procedure, one over geometric measurements and the other over structural measurements.

The combined geometric and structural approach by Luo & Hancock (2003) has shown an intermediate performance between the pure geometric and the pure structural method.



(a) Fish deformation

(b) Fish noise

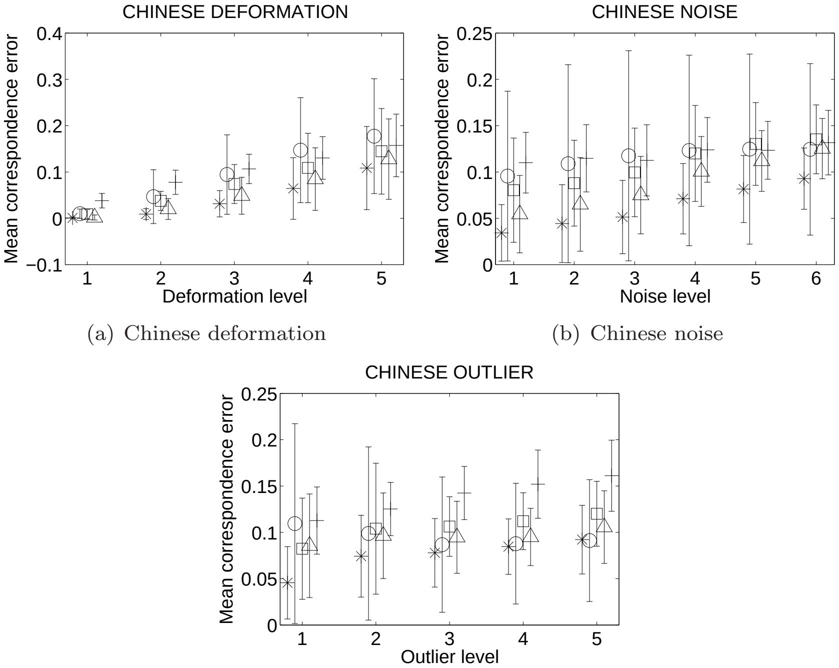
(c) Fish outliers

Figure 9.2: Results for the fish (a) deformation, (b) noise, and (c) outlier experiments of ($*$) our method, ($\bigcirc$) our method with $P_e = 0.5$ (i.e., Softassign Procrustes), ($\square$) Unified approach by Luo & Hancock (2003), ($\triangle$) Graduated Assignment and ($-$) ICP.

In the noise experiments the proposed method has demonstrated a clear superiority with respect to the others. Not surprisingly, the pure geometric methods have experimented a decrease in performance in the deformation experiments. It was expected, since similarity transformations only provide a rough approximation to the non-rigid deformations undergone by the templates. Since non-rigid deformations preserve the local neighborhood structures, methods embodying a model of structural relations have shown the best performance. Specifically, the best methods have been Graduated assignment (i.e., the structural version of our method) and, even slightly better, the full version of our method. The proposed method has performed the best in the outlier experiments followed by the pure structural Graduated Assignment. This confirms the discriminating ability of the used strategy for generating the graphs in the outliers data-set. The pure geometric version of our method (i.e., Softassign Procrustes) achieves fair mean results for some outlier levels, nevertheless the high standard deviations in the distribution of the mean correspondence errors reveals an unstable behaviour.

9.5.2 Real Data

We have performed image matching experiments with some databases from `http://www.featurespace.org` that hosts image databases which are com-



(a) Chinese deformation

(b) Chinese noise

(c) Chinese outliers

Figure 9.3: Results for the fish (a) deformation, (b) noise, and (c) outlier experiments of ($*$) our method, (◯) our method with $P_e = 0.5$ (i.e., Softassign Procrustes), (□) Unified approach by Luo & Hancock (2003), (△) Graduated Assignment and ($-$) ICP.

monly used for performance evaluation of local image detectors and descriptors. We have used the datasets *BOAT*, *EASTPARK*, *EASTSOUTH* and *RESID* from INRIA (France), each one containing a sequence of images showing a scene across different zooms and rotations.

Each sequence, containing between 10 and 11 images, can be ordered according to the variation in zoom. We perform a sort of narrow-baseline matching by using only adjacent image pairs from the ordered sequence. So, results for each dataset are the mean of 18 or 20 experiments. It is very difficult to handle the high amounts of clutter usually present in wide-baseline matching using non-discriminant features such as the arrangement of two sparse sets of points. Such a problem is more accurately driven with the use of discriminant features such as local image feature vectors.

Points are extracted with the *scale-invariant feature detector* by Lowe (2004) that locates points at the scale-space extrema of a *Difference-of-Gaussians* function (section 1.2.2). For each image, we keep the 50 points with the highest scales. We have chosen this point-set size since it represents the limit for our implementation of the Dual-Step method to execute in reasonable time. We have placed the edges between points by using a *Delaunay triangulation*.

We have compared the proposed method, the pure geometric version of our method, the Unified approach by Luo & Hancock (2003), the Dual-Step method by Cross & Hancock (1998) and the Graduated Assignment by Gold & Rangarajan (1996) (sections 4.5.2, 4.5.1 and 4.3.2, respectively).

As in the synthetic experiments, we evaluate the results through the mean correspondence error. Since it is available the ground truth homography between each pair of images we can compute the ground truth projection onto the second image from a point in the first image. The mean correspondence error is then computed as the Euclidean distance between the ground truth projection of a point and the point where it has been actually assigned to.

All the methods have been initialized with a matching by correlation with a fixed window size. We have used the orientation of each point provided by the detector in order to achieve a certain invariance to rotations in the initialization. We have included the results of the matching by correlation in the comparisons.

As said, all the images are related by a similarity transformation. It is expected a moderate amount of structural corruption due to clutter. These two facts lead us to reduce the specific weight of the structural measurements by increasing its uncertainty. Therefore, we have experimentally set $P_e = 0.3$ for the proposed method and the Unified method. We have experimentally set the outlying parameter $N = 0.5$ for the full version of our method and $N = 1.25$ for the pure geometric one. The value of $N = 1.25$ has been set so that it returns a similar number of correspondences than the full version. We have experimentally set the values $P_e = 0.1$ and $\rho = 0.0001$ for the Dual-Step method. We have experimentally set the compatibility coefficient of Graduated Assignment to $c = 3.47$.

Figure 9.4 shows the results obtained by each method in each database.

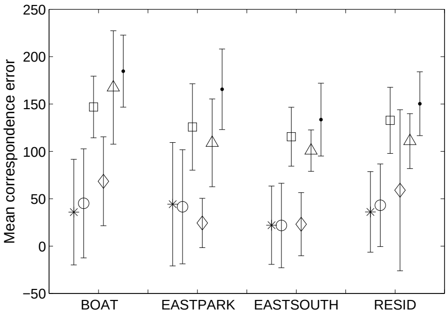
Notice the superiority in terms of mean correspondence errors of the methods incorporating outlier rejection mechanisms. As you can see in figure 9.4(b), all the methods that do not reject outliers tend to match all available points. This penalizes in terms of mean correspondence errors due to the usual presence of clutter in the image matching experiments.

With regards to the methods incorporating outlier rejection, all of them present comparable accuracies with a similar number of matched points. Since all the images used in the experiments are related by a similarity transform, the transformation-invariant geometric model used by our method has succeed in accommodating the underlying transformations. Therefore, we have found no significant advantages from incorporating structural constraints. However, the full version of our method presents slightly better accuracies in the *BOAT* and *RESID* databases.

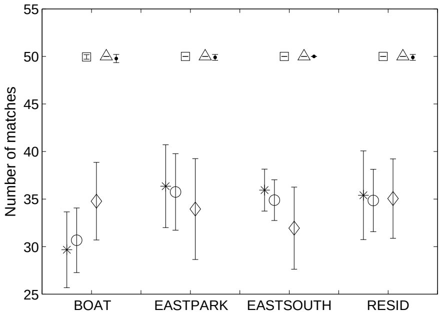
The Dual-Step method presents a similar accuracy than ours but requires considerably higher time to execute. The mean computational times and standard deviations obtained in all the databases by the Dual-Step method are 799.31 ± 78.99 sec.; and those obtained by our method are 1.68 ± 0.2 sec. All the methods have been implemented in Matlab and executed on an Intel Xeon CPU E5310 at 1.60GHz.

9.5.3 Qualitative Experiment with Projective Transformations

We present a registration experiment of two images from the *graf* dataset used in Mikolajczyk & Schmid (2005). The two images show a planar surface from different viewpoints. Therefore, corresponding points in the two images lying on the surface are related by an homography.



(a) Mean correspondence error



(b) Number of correspondences

Figure 9.4: (a) Mean correspondence errors (in pixels) and (b) number of matches returned by (∗) our method, (◯) our method with $P_e = 0.5$ (i.e., Softassign Procrustes), (□) Unified approach by Luo & Hancock (2003), (⋄) Dual-Step by Cross & Hancock (1998), (△) Graduated Assignment and (·) matching by correlation (initialization).

We aim to test the registration accuracy of our method when using a projective transformation model instead of the rigid (i.e., similarity) one used in the previous experiments. Therefore, we use the projective transformation model depicted in section 3.5.3 in the maximization of the alignment parameters of equation (9.15).

Point-sets on the two images are extracted and tentative correspondences are computed using the *Scale-Invariant Feature Transform* (SIFT) by Lowe (2004) (sections 1.2.2, 1.3.1 and 1.4.2). In order to build the graphs, we select a set of inliers found by RANSAC (Fischler & Bolles, 1981) (section 2.4). Edges are extracted by means of Delaunay triangulations over the point-sets. Figure 9.5 show the two graphs superimposed on the images.

Correspondences parameters are initialized to equiprobable values and initial alignment parameters are defined by the original arrangement of the point-sets. Figure 9.6 shows the behaviour of the model point-set in the scale-space under the action of the projective alignment parameters along the different iterations of the presented algorithm.

While the proposed method succeeds in finding the underlying projective

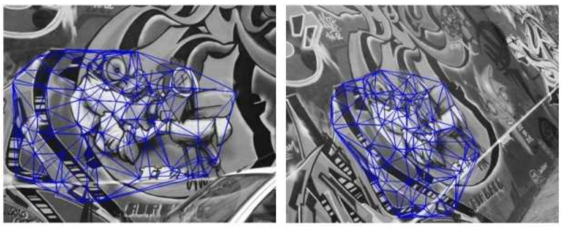


Figure 9.5: Original images with graphs superimposed.

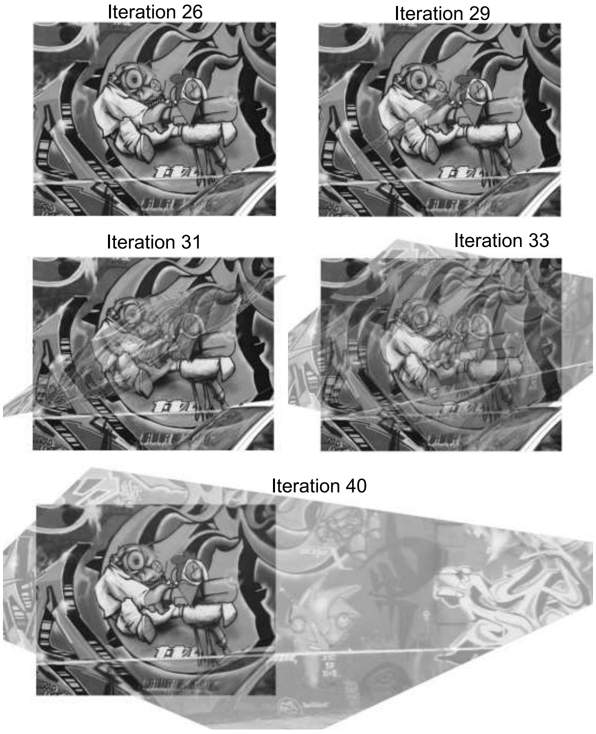


Figure 9.6: Data image and warped model image (shaded) according to the projective transformation parameters estimated at different iterations of our method.

transformation with very little a priori knowledge, it does not show the same robustness against clutter as in the case of similarity transformations. Although it is necessary a specific work for the case of projective transformations aimed at providing robustness against clutter, the proposed attempt shows an interesting behaviour under limited conditions.

9.6 Conclusions

We have presented an ensemble approach for structural graph matching and point-set alignment that benefit from the additional representational facilities of graphs with respect to point-sets. We pose the problem as one of maximum likelihood estimation of correspondence and alignment parameters from a mixture distribution. Our mixture model assumes that point position errors and structural errors follow Gaussian and Bernoulli distributions, respectively. We derive the EM algorithm according to the proposed mixture model where alignment and correspondence parameters are estimated in conditional maximization steps. As opposed to other methods, our method uses a continuous correspondence variable. We use Softassign in order to compute the correspondence indicators at each iteration. An annealing procedure is implemented by updating a control parameter within Softassign at each maximization step. Outliers are gradually rejected on the basis of the number of standard deviations allowed in the alignment errors. We have performed matching experiments on synthetic and real data.

With regards to the synthetic addition of noise and outliers, the combination of both geometric and structural constraints proposed by our method has resulted in a superior performance than any of its parts separately as well as than the rest of the methods. In the presence of nonrigid deformations, both the full version of our method and the purely structural one share the best performance.

In the image matching experiments the methods with outlier rejection capabilities have performed the best, due to the usual presence of clutter in these types of experiments. There are no significant differences between the performance of the full and the pure geometric versions of our method in these experiments. This is because the similarity transformation model adjusts fairly well to the underlying geometry of the problem. The Dual-Step method by Cross & Hancock (1998) present a roughly similar performance than ours but takes a considerably higher time.

With regards to the projective registration, further work is needed aimed at providing some robustness against clutter to our method.

Chapter 10

Conclusions and Future Work

Deciding the correspondences between two images is of pivotal importance in many tasks in the computer vision & pattern recognition field. In this thesis we have focussed on methods that use either local image features or position coordinates to that end.

First of all, we have investigated the benefits of using cross-bin distances such as the Earth Movers' Distance for comparing local features. Afterwards, we have developed a series of methods aimed at finding the correspondences between two images from a graph-theoretic point of view. This is, we have included pairwise relational constraints into the problem.

In order to make explicit the benefits of the proposed methods we have compared to other methods in the literature.

In chapter 5 we have explored the possibility of using a cross-bin measure such as the Earth-Mover's Distance (EMD) in order to compute the distance between local descriptors. We have proposed an efficient algorithm to approximate the EMD based on an heuristic that favours movements involving locations in the boundaries of the histograms. The proposed algorithm presents a theoretical cost lower than similar approaches. Moreover, an empirical study of the time complexity reveals that the actual cost in a practical situation is considerably lower than the theoretical one. In a retrieval rate study the proposed algorithm has shown a similar performance with a lower computational cost than the original EMD algorithm for multiple choices of colour spaces and dimensions of the histograms.

With regards to the matching of image features, registration experiments with Shape Contexts do not show an improvement of using the proposed cross-bin distance with respect to the χ^2 bin-to-bin measure. This is because, in the case of Shape Contexts, the correspondence assumption made by the bin-to-bin measures is fairly accurate.

Concerning other descriptors such as SIFT, the invariance to rotation or affine shape introduced during the description stage suggests us that this correspondence assumption will also hold in these cases.

In chapter 6 we have proposed two graph matching methods aimed at incorporating pairwise constraints to the correspondence problem between SIFT-features. Since the graph matching problem is known to be NP-hard, sub-optimal solutions are sought by Discrete Relaxation and Graduated Assignment, respectively. We have evaluated the robustness of the proposed methods against various types of noise in a series of matching experiments with synthetic images. We have compared to a number of methods falling within various categories, namely, outlier rejection, point-set registration and graph matching. While all the competing methods perform weakly for some specific type of noise, the proposed methods have presented an acceptable performance in all the types of noise, thus behaving the most stably. Specifically, the Graduated-Assignment-based approach performs better than the Discrete-Relaxation-based one.

In chapter 7 we have presented a model aimed at overcoming the limitations of the general models used for graph matching when dealing with sparse graphs such as the ones representing handwritten characters. In order to attain this objective we have introduced position coordinates as nodes' attributes into an existing model for graph matching (Wilson & Hancock, 1997). The mentioned model gauges the structural consistency of a match by means of the Hamming distances between the cliques.

In order to be invariant to some extent to the specific pose of the point-sets, we have included the estimation of the similarity alignment parameters into the problem. As consequence, the proposed model gauges the consistency of the assignments by means of both the Hamming and Procrustes distances at a clique level. We propose two optimization strategies, namely, discrete relaxation and hybrid genetic search.

The proposed method outperforms the original cliques model in terms of correct matching rates in the aforementioned types of graphs. Results suggest that our method would be more appropriate than the original clique method for the task of handwritten character recognition.

In chapters 8 and 9 we have presented approaches that make use of position coordinates as nodes' attributes, as well. These methods formulate the graph matching problem within a principled statistical framework, while benefiting from the desirable properties of the Softassign and deterministic annealing ensemble.

In chapter 8 we have presented a graph matching method that uses both geometric and structural information in a relational way. We have tested the accuracy of the method in a series of image registration and shape retrieval experiments. We have compared to outlier rejectors, point-set registration methods and joint structural graph matching and point-set registration methods.

Matching experiments with synthetic graphs and real images show that our method outperforms most of the compared ones in terms of locating the correct matches. The Dual-Step method (Cross & Hancock, 1998) shows similar matching accuracies than ours but uses higher computational times.

In the shape retrieval experiments with the GREC database, the Dual-Step method shows the best performance followed by our method which presents slightly lower performances. In the shape retrieval experiments with the 25-shapes database our method shares the best performance with the Dual-Step method.

In chapter 9 we have presented a method that addresses the problem of simultaneous structural graph matching and point-set registration. The proposed method has the distinguished property that it encompasses two well-known methods as specific cases, namely, the graph matching method Graduated Assignment (Gold & Rangarajan, 1996) and the point-set registration method Robust Point Matching (Gold *et al.*, 1998).

We have tested the registration accuracy of our method in quantitative experiments using both synthetic data-sets and real images. Our method has usually got the best performance which in some cases has been shared with either its pure structural version (i.e., Graduated Assignment) or pure geometric version (i.e., Robust Point Matching).

The Dual-Step graph matching method obtains similar results than ours in the matching experiments with real images but it requires considerably higher computational times.

We have also presented a qualitative image registration experiment in order to test how our method behaves when dealing with projective transformations. Despite our method has shown to converge to the correct solution in a few iterations it does not show the same stability when outliers are present. Further research is needed in that direction.

Although we have not directly compared the methods proposed in chapters 8 and 9 we deduce that they will achieve similar registration accuracies due to their similar relative performances with respect to the Dual-Step method.

While in this thesis we have addressed separately the combination of structural + geometric information and structural + local image information, we have not addressed the combination of the three features simultaneously. One obvious extension of this thesis would be to provide a unified framework that combines evidence coming from structural relations, geometric positions and local image descriptors.

Recently, Silletti *et al.* (2011) have proposed to solve the correspondence problem by using a combination of Gaussian functions between measurements of these three types. They propose a non-iterative algorithm in which correspondences are decided by the spectral association method by Scott & Longuet-Higgins (1991). The benefit coefficients input to the spectral decision are computed as the product among all the Gaussian functions for each candidate association.

Along the same lines, Dungan & Potter (2010) combine measurements from multiple sources in a multidimensional feature vector which they call *attributed point*. Some attributes in this feature vector may be transformation-dependant such as position coordinates. Correspondence problem between attributed point patterns is faced as a classification problem. The Mahalanobis distance is used to compute the similarity between two of these multidimensional attributed points.

Combining multiple types of measurements can effectively improve the convergence properties. In fact, this is the starting point of this thesis. However, as we increase the number of features it increases the difficulty of *learning* the parameters of the model, as well (e.g., the variance of each feature and the covariances between them). This approach risks of producing too specific methods which, additionally, are cumbersome to train.

One alternative to this view is a multistage approach that uses individual features at separate stages. This is the case of the well-known ensemble SIFT + RANSAC that computes tentative correspondences in one stage using local image descriptors and discards spurious correspondences in another stage using geometric information. Another example is the non-rigid shape registration approach by Belongie *et al.* (2002) that computes tentative correspondences in one stage using Shape Contexts and nonrigidly aligns the two point-sets in another stage using geometric information (these two steps are iteratively repeated until convergence).

Certainly, there are many other successful approaches to solve the correspondence problem that use this concept.

Motivated by this view, we are now working on a hierarchical approach for matching local image features that follows a multilayer idea. Although the proposed method benefits from the accuracy of the SIFT + RANSAC ensemble in the lower layer, it is easily generalizable to other matching methods.

Usually, geometric constraints imposed by RANSAC do not fit to the whole problem. For example, in the case of homography estimation all corresponding points must be coplanar, or in the more general case of fundamental matrix estimation they must have experienced the same 3D rigid displacement. In order to overcome these limitations, we adopt a divide-and-conquer strategy in which the “big” problem is divided into smaller subproblems in a hierarchical structure. This approach is motivated by the following two observations. Strong geometric constraints imposed by RANSAC may hold in the smaller subproblems, and looser structural constraints derived from the hierarchical representations are capable of adapting to a wide range of distortions (e.g., 3D non-rigid deformations, repeated patterns).

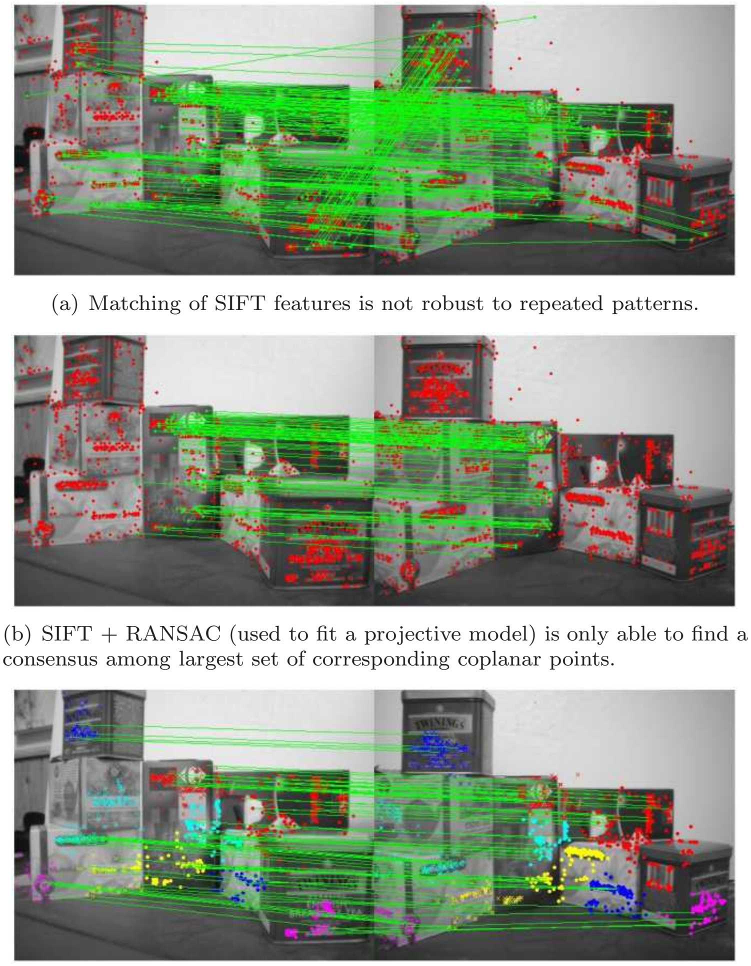
We obtain a hierarchical representation of the scene by applying an agglomerative clustering algorithm over the SIFT keypoint-locations extracted from each image. Trees are obtained by “cutting” the hierarchical representations at a certain level thus obtaining clusters of points as leaf nodes. Trees are matched by using an adaptation of the tree-matching algorithm by Torsello & Hancock (2003) in which the number of SIFT + RANSAC inliers between keypoint-clusters is used as matching coefficients between the leaf nodes. The proposed method is tolerant to oversegmentation (i.e., cutting the tree too low) since it automatically merges sibling nodes when appropriate.

In figure 10.1 we illustrate the case of homography estimation in which our method automatically subdivides the whole set of points into clusters of coplanar corresponding points.

One perennial issue in all graph-based applications for computer vision is how to extract graph representations from images that lead to efficient models for diverse tasks such as matching, characterization, feature extraction and others.

A successful approach applied to non-rigid 3D shapes consists in endowing the shapes with metric spaces and to exploit the diffusion geometry for these tasks (see for example Bronstein *et al.* (2010) for matching, Bronstein *et al.* (2011); Sun *et al.* (2009) for feature extraction, and Bronstein & Bronstein (2011) for characterization, and references therein).

Diffusion geometry presents interesting invariant properties which makes it suitable for 3D shape analysis. Moreover the efficiency of the spectral techniques



(a) Matching of SIFT features is not robust to repeated patterns.

(b) SIFT + RANSAC (used to fit a projective model) is only able to find a consensus among largest set of corresponding coplanar points.

(c) The proposed method divides the whole initial point-set into clusters of coplanar points in which SIFT + RANSAC succeeds (points in different cluster are shown in different colors and markers). The tree-matching algorithm computes the correspondences and performs the appropriate node deletions in both trees that induce a tree isomorphism. Correspondences between individual points are deduced from the correspondences between the leaf nodes of the trees.

Figure 10.1: Two images showing the same scene from different viewpoints. The scene contains several boxes arranged so that no two surfaces are coplanar and containing some repetitive patterns, as well.

used to compute the diffusion metric allow to deal with meshes with a high number of points.

These approaches keep parallelism with graph theory because instead of absolute measurements they use the intrinsic pair-wise distances between elements on a given space (i.e., graph).

We think that facing these problems in the image domain in a similar way than in the 3D shape domain would open the possibility to a lot of interesting applications. Therefore, there is a promising line of research in extracting similar graph representations based on the images' contents.

The use of diffusion processes would allow us to benefit from the powerful results from the spectral graph theory. Moreover, their efficiency would allow for dense representations as opposed to the usual ones based on sparse features.

Matching between two images could be performed by embedding the graph representations into Euclidean spaces through Multidimensional Scaling and then registering the resulting point-sets with existing registration methods.

It would be also interesting to investigate the effects of the Fiedler vector for feature extraction. This way, features would correspond to clusters of highly interconnected regions in the new image representation. Features at different scales would be detected from image representations at different scales. As opposed to other feature detectors that assume circular or elliptic regions, this approach would detect regions of any shape. The study of the graph characteristics from the extracted regions through information-theoretic measurements would provide effective means of describing these regions.

List of Publications Derived from this Thesis

Journal Papers

- G. Sanromà, R. Alquézar & F. Serratosa, "A New Graph Matching Method for Point-Set Correspondence using the EM Algorithm and Softassign", *Computer Vision and Image Understanding* (2011),

  `http://dx.doi.org/10.1016/j.cviu.2011.10.009`

- F. Serratosa & G. Sanromà, "A Fast Approximation of the Earth-Movers Distance between Multi-Dimensional Histograms", *International Journal of Pattern Recognition and Artificial Intelligence*, IJPRAI **22** (8), pp: 1-20, 2008

At this time, one further paper entitled "Smooth Point-set Registration using Neighboring Constraints" is under revision in the journal *Pattern Recognition Letters.*

Conference Papers

- G. Sanromà, R. Alquézar, & F. Serratosa, "Smooth Simultaneous Structural Graph Matching and Point-Set Registration", *Graph based Representations*, GbR2011, Mnster, Germany, LNCS 6658 pp: 142,151 2011.

- G. Sanromà, R. Alquézar & F. Serratosa, "Attributed Graph Matching for Image-Features Association using SIFT Descriptors", *Syntactic and Structural Pattern Recognition*, SSPR2010, Izmir, Turkey, LNCS 6218, pp: 254-263, 2010.

- G. Sanromà, R. Alquézar & F. Serratosa, "A Discrete Labelling Approach to Attributed Graph Matching using SIFT Features", *International Conference on Pattern Recognition*, ICPR2010, Istanbul, Turkey, pp: 954-957, 2010.

- G. Sanromà, R. Alquézar & F. Serratosa, "Graph Matching using SIFT Descriptors, an application to pose recovery of a mobile robot", *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, VISAPP2010, Angers, France, pp: 249-254, 2010.

- G. Sanromà, F. Serratosa & R. Alquézar, "Shape Learning with Function-Described Graphs", *International Congress on Image Analysis and Recognition*, ICIAR2008, LNCS 5112, Povoa de Varzim, Portugal, pp: 475-484, 2008.

- G. Sanromà, F. Serratosa & R. Alquézar, "Hybrid Genetic Algorithm and Procrustes Analysis for Enhancing the Matching of Graphs Generated from Shapes", *Proc. Syntactic and Structural Pattern Recognition*, SSPR2008, LNCS 5342, Orlando, Florida, USA, pp: 298-307, 2008.

- G. Sanromà, F. Serratosa & R. Alquézar, "Improving the Matching of Graphs Generated from Shapes by the Use of Procrustes Distances into a Clique-based MAP Formulation", *19th International Conference on Pattern Recognition*, ICPR2008, Tampa, Florida, USA, vol. 2, pp: 1-4, 2008.

- F. Serratosa, G. Sanromà, & A. Sanfeliu, "A New Algorithm to Compute the Distance between Multi-dimensional Histograms", *12th Iberoamerican Congress on Pattern Recognition*, CIARP2007, LNCS, 4756, Via del Mar-Valparaiso, Chile, pp. 115 - 123, 2007.

- F. Serratosa & G. Sanromà, "Modelling Intermittently Present Features using non-Linear Point Distribution Models", *IEEE, Pacific-Rim on Image and Video Technology*, PSIVT2007, LNCS 4872, Santigao, Chile, pp. 260-273 , 2007.

- F. Serratosa & G. Sanromà, "An Efficient Distance between Multi-dimensional Histograms for Comparing images", *Proc. Syntactic and Structural Pattern Recognition*, SSPR2006, LNCS 4109 , Hong Kong China, pp: 412-421, 2006.

Appendices

Appendix A

$D_{set}\left(\mathcal{A},\mathcal{B}\right) = D_{\textbf{EMD-}g_f}\left(\mathbf{h},\mathbf{k}\right)$

Assume that there are two sets $\mathcal{A}$ and $\mathcal{B}$ that have n elements contained in the domain $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$. Then, the distance between two elements of the sets $\mathcal{A}$ and $\mathcal{B}$ given an assignment f, can be obtained as the distance between bins as follows,

$$d\left(\mathbf{a}_i, \mathbf{b}_{f(i)}\right) = \sum_{j,j'=1}^{p} O^{\mathcal{A}}_{ji} O^{\mathcal{B}}_{j'f(i)} d\left(\mathbf{z}_j, \mathbf{z}_{j'}\right), \; 1 \leq i \leq n, \; f \text{ bijective} \tag{A.1}$$

The demonstration that $D_{\text{EMD-}g_f}\left(\mathbf{h},\mathbf{k}\right)$ is a distance is depicted by Kamarainen *et al.* (2003). Although this paper deal with 1D-histograms, the proof is based on the distance between elements $d\left(\mathbf{a},\mathbf{b}\right)$ independently on the dimension of the elements $\mathbf{a}$ and $\mathbf{b}$.

If we apply equation (A.1) to substitute the distance between elements $d\left(\mathbf{a}_i, \mathbf{b}_{f(i)}\right)$ in the definition of the distance between sets of equation (5.6), we obtain the formula

$$D_{set}\left(\mathcal{A},\mathcal{B}\right) = \min_{f:\mathcal{A}\rightarrow\mathcal{B}} \sum_{i=1}^{n} \sum_{j,j'=1}^{p} O^{\mathcal{A}}_{ji} O^{\mathcal{B}}_{j'f(i)} d\left(\mathbf{z}_j, \mathbf{z}_{j'}\right) \tag{A.2}$$

Then, rearranging the elements, we get

$$D_{set}\left(\mathcal{A},\mathcal{B}\right) = \min_{f:\mathcal{A}\rightarrow\mathcal{B}} \sum_{j,j'=1}^{p} d\left(\mathbf{z}_j, \mathbf{z}_{j'}\right) \sum_{i=1}^{n} O^{\mathcal{A}}_{ji} O^{\mathcal{B}}_{j'f(i)} \tag{A.3}$$

Finally, if we substitute the equation (5.12) of the flow we obtain the final expression

$$D_{set}\left(\mathcal{A},\mathcal{B}\right) = \min_{f:\mathcal{A}\rightarrow\mathcal{B}} \sum_{j,j'=1}^{p} d\left(\mathbf{z}_j, \mathbf{z}_{j'}\right) g_f\left(j, j'\right) = D_{\text{EMD-}g_f}\left(\mathbf{h},\mathbf{k}\right) \tag{A.4}$$

Appendix B

Properties of the Flow g_f

Demonstration of equation (5.8): It is a straightforward property due to equations (5.2) and (5.12).

Demonstration of equation (5.9): Using equation (5.12), we obtain that

$$\sum_{j'=1}^{p} g_f\left(j, j'\right) = \sum_{j'=1}^{p} \sum_{i=1}^{n} O_{ji}^{\mathcal{A}} O_{j' f(i)}^{\mathcal{B}} \tag{B.1}$$

and exchanging the addition, we obtain that

$$\sum_{j'=1}^{p} g_f\left(j, j'\right) = \sum_{i=1}^{n} \sum_{j'=1}^{p} O_{ji}^{\mathcal{A}} O_{j' f(i)}^{\mathcal{B}} \tag{B.2}$$

Then, if we spawn the external addition, we have the following formula,

$$O_{j1}^{\mathcal{A}} \sum_{j'=1}^{p} O_{j' f(1)}^{\mathcal{B}} + O_{j2}^{\mathcal{A}} \sum_{j'=1}^{p} O_{j' f(2)}^{\mathcal{B}} + \ldots + O_{in}^{\mathcal{A}} \sum_{j'=1}^{p} O_{j' f(n)}^{\mathcal{B}} \tag{B.3}$$

that can be reduced to

$$O_{j1}^{\mathcal{A}} + O_{j2}^{\mathcal{A}} + \ldots + O_{in}^{\mathcal{A}} \tag{B.4}$$

due to equation (5.3) and considering that f is bijective. So, we arrive to the expression

$$\sum_{j'=1}^{p} g_f\left(j, j'\right) = \sum_{i=1}^{n} O_{ji}^{\mathcal{A}} = \mathbf{h}\left(j\right) \tag{B.5}$$

Demonstration of equation (5.10) is similar to demonstration of equation (5.9).

Appendix C

Rotation-Invariant Shape Contexts

The following is an adaptation of the Shape Contexts (Belongie *et al.*, 2002) (section 1.3.2) in order to match point-sets regardless of their orientations. Given two point-sets (in our case, the positions of the nodes), $\mathcal{X} = \{\mathbf{x}_a\}, \forall_{a \in \mathcal{I}}$ and $\mathcal{Y} = \{\mathbf{y}_\alpha\}, \forall_{\alpha \in \mathcal{J}}$, we arbitrarily choose one of them (e.g., $\mathcal{Y}$) and create M subsets at different orientations by applying M rigid-body rotations uniformly distributed along the range $[0 \ldots 360]$ degrees[1]. We therefore obtain the subsets $\mathcal{Y}^m, {}_{m \in [1 \ldots M]}$, corresponding to M different orientations of $\mathcal{Y}$. Next, we compute the Shape Contexts, a kind of descriptor (log-polar histograms) that, for each point, encodes how the rest of the points are distributed around it. We do it for the $M+1$ point-sets $\mathcal{X}$ and $\mathcal{Y}^m, {}_{m \in [1 \ldots M]}$. Using the χ^2 test statistic between the Shape Contexts (Belongie *et al.*, 2002), we compute the M matrices of matching costs C^m. Thus, $C^m_{a\alpha}$ indicates the cost of matching the point $\mathbf{x}_a \in \mathcal{X}$ to the point $\mathbf{y}^m_\alpha \in \mathcal{Y}^m$. By applying the Hungarian algorithm (Munkres, 1957) to each C^m, we compute the optimal assignments $f^m : \mathcal{I} \rightarrow \mathcal{J}$ from the points in $\mathcal{X}$ to those in each one of the $\mathcal{Y}^m$'s. We choose as the prevailing orientation $m^\star$, the one with the minimum matching cost, i.e., $m^\star = \arg\min_m \left\{ \sum C^m_{a,f^m(a)} \right\}$ and, the resulting correspondences are those defined by $f^{m^\star}$.

[1] We use $M = 12$

Bibliography

Aguilar, W., Frauel, Y., Escolano, F. & Martinez-Perez, M. E. (2009). A robust graph transformation matching for non-rigid registration. *Image and Vision Computing* **27**, 897–910.

Ahuja, R. K., Magnanti, T. L. & Orlin, J. B. (1993). *Network flows: theory, algorithms, and applications*. Prentice-Hall, Inc.

Bai, X. & Latecki, L. J. (2008). Path similarity skeleton graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(7).

Bai, X., Latecki, L. J. & Liu, W.-Y. (2007). Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(3), 449–462.

Barrow, H. G. & Burstall, R. M. (1976). Subgraph isomorphism, matching relational structures and maximal cliques. *Information Processing Letters* **4**, 83–84.

Belongie, S., Malik, J. & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 509–522.

Berge, J. (2006). The rigid orthogonal procrustes rotation problem. *Psychometrika* **71**(1), 201–205.

Besl, P. J. & McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256.

Bin, L. & R., H. E. (1999). Feature matching with procrustes alignment and graph editing. In: *Proceedings of IEEE Seventh International Conference on Image Processing and its Applications*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Black, M. & Rangarajan, A. (1996). On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision* **19**(1), 57–91.

Blum, H. (1967). A transformation for extracting new descriptors of shape. In: *Models for the Perception of Speech and Visual Form*. MIT Press.

Bookstein, F. L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **11**(6), 567–585.

Borman, S. (2004). The expectation maximization algorithm – a short tutorial. Unpublished.

Bronstein, A. M., Bronstein, M. M., Guibas, L. J. & Ovsjanikov, M. (2011). Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph.* **30**, 1:1–1:20.

Bronstein, A. M., Bronstein, M. M., Kimmel, R., Mahmoudi, M. & Sapiro, G. (2010). A gromov-hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *Int. J. Comput. Vision* **89**, 266–286.

Bronstein, M. M. & Bronstein, A. M. (2011). Shape recognition with spectral distances. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**, 1065–1071.

Brown, M. & Lowe, D. G. (2003). Recognising panoramas. In: *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03. Washington, DC, USA: IEEE Computer Society.

Bunke, H. (1999). Error correcting graph matching: On the influence of the underlying cost function. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 917–922.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698.

Cha, S. (2002). On measuring the distance between histograms. *Pattern Recognition* **35**(6), 1355–1370.

Chan, K. & Cheung, Y. (1992). Fuzzy-attribute graph with application to chinese character-recognition. *IEEE Transactions on Systems Man and Cybernetics* **22**(2), 402–410.

Chapelle, O., Haffner, P. & Vapnik, V. N. (1999). Support vector machines for histogram-based image classification. *Neural Networks, IEEE Transactions on* **10**(5), 1055–1064.

Chetverikov, D., Stepanov, D. & Krsek, P. (2005). Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing* **23**, 299–309.

Christmas, W., Kittler, J. & Petrou, M. (1995). Structural matching in computer vision using probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**(8), 749–764.

Chui, H. & Rangarajan, A. (2000). A new algorithm for non-rigid point matching. In: *in CVPR*.

Chui, H. & Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Underst.* **89**, 114–141.

Chung, F. R. K. (1997). *Spectral Graph Theory.* American Mathematical Society.

Conte, D., Foggia, P., Sansone, C. & Vento, M. (2004). Thirty Years Of Graph Matching In Pattern Recognition. *International Journal of Pattern Recognition and Artificial Intelligence* .

Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. (1995). Active shape models: their training and application. *Comput. Vis. Image Underst.* **61**, 38–59.

Cross, A. (1997). Inexact graph matching using genetic search. *Pattern Recognition* **30**(6), 953–970.

Cross, A. & Hancock, E. (1998). Graph matching with a dual-step em algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(11), 1236–1253.

Cross, A. D. J., Myers, R. & Hancock, E. R. (2000). Convergence of a hill-climbing genetic algorithm for graph matching. *Pattern Recognition* **33**, 1863–1880.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* **39**(1), 1–38.

Di Ruberto, C. (2004). Recognition of shapes by attributed skeletal graphs. *Pattern Recognition* **37**(1), 21–31.

Dryden, I. L. & Mardia, K. (1998). *Statistical shape analysis.* Wiley series in probability and statistics: Probability and statistics. J. Wiley.

Dungan, K. E. & Potter, L. C. (2010). Classifying transformation-variant attributed point patterns. *Pattern Recogn.* **43**, 3805–3816.

Edelman, S., Intrator, N. & Poggio, T. (1997). Complex cells and object recognition. URL `http://cogprints.org/561/`.

Emms, D., Wilson, R. C. & Hancock, E. R. (2009). Graph matching using the interference of continuous-time quantum walks. *Pattern Recogn.* **42**, 985–1002.

Ennesser, F. & Medioni, G. (1995). Finding waldo, or focus of attention using local color information. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 805–809.

Escolano, F., Hancock, E. R. & Lozano, M. A. (2011). Graph matching through entropic manifold alignment. In: *CVPR*.

Eshera, M. A. & Fu, K. S. (1984). A graph distance measure for image analysis. *IEEE Transactions on Systems, Man, and Cybernetics* **14**(3), 398–408.

Fischler, M. A. & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comunications of the ACM* **24**(6), 381–395.

Fogel, D. (1994). An introduction to simulated evolutionary optimization. *IEEE Transactions on Neural Networks* **5**(1), 3–14.

Forssén, P.-E. & Lowe, D. (2007). Shape descriptors for maximally stable extremal regions. In: *IEEE International Conference on Computer Vision*, vol. CFP07198-CDR. Rio de Janeiro, Brazil: IEEE Computer Society.

Frank-Bolton, P., Alvarado-Gonzalez, A. M., Aguilar, W. & Frauel, Y. (2008). Vision based localization for mobile robots using a set of known views. In: *Proceedings of Advances in Visual Computing (LNCS)*, vol. 5358 of *4th International Symposium on Visual Computing*.

Ghahraman, D. E., Wong, A. K. C. & Au, T. (1980). Graph monomorphism algorithms. *Trans. Syst. Man Cybern.* **10**, 189–197.

Godin, G., Rioux, M. & Baribeau, R. (1994). Three-dimensional registration using range and intensity information. In: *Proceedings of the SPIE: Videometrics III* (El-Hakim, S. F., ed.), vol. 2350. SPIE.

Goh, W.-B. (2008). Strategies for shape matching using skeletons. *Computer Vision and Image Understanding* **110**(3), 326–345.

Gold, S. & Rangarajan, A. (1996). A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(4).

Gold, S., Rangarajan, A., ping Lu, C., Pappu, S. & Mjolsness, E. (1998). New algorithms for 2d and 3d point matching: pose estimation and correspondence. *Pattern Recognition* **31**, 1019–1031.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1st ed.

Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M. & Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance functions. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 729–736.

Han, W. & Brady, M. (1995). Real-time corner detection algorithm for motion estimation. *Image and Vision Computing* , 695–703.

Haralick, R. M., Joo, H., Lee, C., Zhuang, X., Vaidya, V. G. & Kim, M. B. (1989). Pose estimation from corresponding point data. *Systems, Man and Cybernetics, IEEE Transactions on* **19**(6), 1426–1446.

Harris, C. & Stephens, M. (1988). A combined corner and edge detection. In: *Proceedings of The Fourth Alvey Vision Conference*.

Hartley, R. I. & Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Ho, J. & Yang, M.-H. (2011). On affine registration of planar point sets using complex numbers. *Comput. Vis. Image Underst.* **115**, 50–58.

Horaud, R., Forbes, F., Yguel, M., Dewaele, G. & Zhang, J. (2011). Rigid and articulated point registration with expectation conditional maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 587–602.

Hummel, R. A. & Zucker, S. W. (1983). On the foundations of relaxation labling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**(3), 267–287.

Ila, V., Porta, J. M. & Andrade-Cetto, J. (2010). Information-based compact pose slam. *IEEE Transactions on Robotics* **26**(1). In press.

Ingrassia, S. & Rocci, R. (2007). Constrained monotone em algorithms for finite mixture of multivariate gaussians. *Comput. Stat. Data Anal.* **51**, 5339–5351.

Jian, B. & Vemuri, B. C. (2005). A robust algorithm for point set registration using mixture of gaussians. In: *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*, ICCV '05. IEEE Computer Society.

Jian, B. & Vemuri, B. C. (2011). Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**, 1633–1645.

Johnson, A. & Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(1), 433 – 449.

Jou, F.-D., Fan, K.-C. & Chang, Y.-L. (2004). Efficient matching of large-size histograms. *Pattern Recogn. Lett.* **25**, 277–286.

Kamarainen, J.-K., Kyrki, V., Ilonen, J. & Kälviäinen, H. (2003). Improving similarity measures of histograms using smoothing projections. *Pattern Recogn. Lett.* **24**, 2009–2019.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* **16**(MAR), 81–121. PT: J.

Kitchen, L. & Rosenfeld, A. (1982). Gray-level corner detection. *Pattern Recognition Letters* **1**(2), 95–102.

Koenderink, L. & Richards, W. (1988). Two-dimensional curvature operators. *Journal of the Optical Society of America: Series A* **5**(7), 1136–1141.

Kolesnik, M. & Fexa, A. (2005). Multi-dimensional color histograms for segmentation of wounds in images. In: *ICIAR*, vol. 3656 of *Lecture Notes in Computer Science*. Springer.

Lazebnik, S., Schmid, C. & Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **27**(8), 1265–1278.

Lee, J.-H. & Won, C.-H. (2011). Topology preserving relaxation labeling for nonrigid point matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 427–432.

Ling, H. & Okada, K. (2007). An efficient earth mover's distance algorithm for robust histogram comparison. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 840–853.

Lowe, D. G. (2001). Local feature view clustering for 3d object recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* **1**, 682.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2).

Lozano, M. A. & Escolano, F. (2004). A significant improvement of softassign with diffusion kernels. In: *SSPR/SPR*.

Lozano, M. A. & Escolano, F. (2005). Local entropic graphs for globally-consistent graph matching. In: *GbRPR* (Brun, L. & Vento, M., eds.), vol. 3434 of *Lecture Notes in Computer Science*. Springer.

Lozano, M. A. & Escolano, F. (2009). Kernelization of softassign and motzkin-strauss algorithms. In: *Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence*, IWANN '09.

Luo, B. & Hancock, E. (2003). A unified framework for alignment and correspondence. *Computer Vision and Image Understanding* **92**(1), 26–55.

Luo, B. & Hancock, E. R. (2001). Structural graph matching using the em algorithm and singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(10).

Luo, B. & Hancock, E. R. (2002). Iterative procrustes alignment with the em algorithm. *Image and Vision Computing* **20**(5), 377–396.

Matas, J., Chum, O., Urban, M. & Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In: *Proceedings of the British Machine Vision Conference*.

McLachlan, G. & Krishnan, T. (1997). *The EM algorithm and extensions*. Wiley series in probability and statistics: Applied probability and statistics. John Wiley.

Meng, X.-L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika* **80**, 267–278.

Mikolajczyk, K. & Schmid, C. (2002). An affine invariant interest point detector. In: *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*. Springer. Copenhagen.

Mikolajczyk, K. & Schmid, C. (2004). Scale & affine invariant interest point detectors. *Int. J. Comput. Vision* **60**, 63–86.

Mikolajczyk, K. & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **27**(10), 1615–1630.

Moravec, H. (1981). Rover visual obstacle avoidance. In: *proceedings of the seventh International Joint Conference on Artificial Intelligence*.

Mori, G., Belongie, S. & Malik, J. (2005). Efficient shape matching using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(11), 1832–1837.

Morovic, J., Shaw, J. & Sun, P.-L. (2002). A fast, non-iterative and exact histogram matching algorithm. *Pattern Recogn. Lett.* **23**, 127–135.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics* **5**(1), 32–38.

Myers, R. & Hancock, E. R. (2001). Least-commitment graph matching with genetic algorithms. *Pattern Recognition* **34**(2), 375–394.

Myronenko, A. & Song, X. (2010). Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(12), 2262–2275.

Nash, J. C. (2000). The (dantzig) simplex method for linear programming. In: *Computing in Science and Engg.*, vol. 2. IEEE Educational Activities Department, pp. 29–31.

Pass, G., Zabih, R. & Miller, J. (1996). Comparing images using color coherence vectors. In: *Proceedings of the fourth ACM international conference on Multimedia*, MULTIMEDIA '96.

Peleg, S., Werman, M. & Rom, H. (1989). A unified approach to the change of resolution: Space and gray-level. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 739–742.

Pelillo, M. (1996). Relaxation labeling networks for the maximum clique problem. *J. Artif. Neural Netw.* **2**, 313–328.

Pelillo, M. (1997). The dynamics of nonlinear relaxation labeling processes. *J. Math. Imaging Vis.* **7**, 309–323.

Rangarajan, A., Chui, H. & Bookstein, F. L. (1997). The softassign procrustes matching algorithm. In: *Proceedings of the 15th International Conference on Information Processing in Medical Imaging*. Springer-Verlag.

Rangarajan, A., Gold, S. & Mjolsness, E. (1996). A novel optimizing network architecture with applications. *Neural Computation* **8**(5), 1041–1060.

Riesen, K. & Bunke, H. (2008). Iam graph database repository for graph based pattern recognition and machine learning. In: *Structural, Syntactic, and Statistical Pattern Recognition*, vol. 5342 of *Lecture Notes in Computer Science*. pp. 287–297.

Robles-Kelly, A. & Hancock, E. R. (2002). String edit distance, random walks and graph matching. In: *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition.*

Robles-Kelly, A. & Hancock, E. R. (2003). Graph matching using spectral seriation and string edit distance. In: *Proceedings of the 4th IAPR international conference on Graph based representations in pattern recognition*, GbRPR'03. Berlin, Heidelberg: Springer-Verlag.

Robles-Kelly, A. & Hancock, E. R. (2005). Graph edit distance from spectral seriation. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 365–378.

Rosenfeld, A., Hummel, R. A. & Zucker, S. W. (1976). Scene labelling by relaxation operations. *IEEE Transactions on Systems, Man and Cybernetics* (6), 420–433.

Rosten, E. & Drummond, T. (2006). Machine learning for high-speed corner detection. In: *In European Conference on Computer Vision*, vol. 1.

Rubner, Y., Puzicha, J., Tomasi, C. & Buhmann, J. M. (2001). Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding* **84**(1), 25–43.

Rubner, Y., Tomasi, C. & Guibas, L. J. (2000a). The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision* **40**, 99–121.

Rubner, Y., Tomasi, C. & Guibas, L. J. (2000b). The earth movers distance as a metric for image retrieval. *International Journal of Computer Vision* **40**.

Russell, E. J. (1969). Extension of dantzig's algorithm to finding an initial near-optimal basis for the transportation problem. *Operations Research* , 187–191.

Sanfeliu, A. & Fu, K. (1983). A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics* **13**, 353–362.

Sanfeliu, A., Serratosa, F. & Alquézar, R. (2004). Second-order random graphs for modeling sets of attributed graphs and their application to object learning and recognition. *IJPRAI* **18**(3), 375–396.

Sanromà, G., Alquézar, R. & Serratosa, F. (2010a). Attributed graph matching for image-features association using sift descriptors. In: *Proceedings of the 2010 joint IAPR international conference on Structural, syntactic, and statistical pattern recognition*, SSPR&SPR'10.

Sanromà, G., Alquézar, R. & Serratosa, F. (2010b). A discrete labelling approach to attributed graph matching using sift features. In: *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR '10.

Scott, G. L. & Longuet-Higgins, H. C. (1991). An Algorithm for Associatin the Features of Two Images. *Proc. Royal Soc. London* (244), 21–26.

Sebastian, T., Klein, P. & Kimia, B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(5), 550–571.

Serratosa, F., Alquézar, R. & Sanfeliu, A. (2002). Synthesis of function-described graphs and clustering of attributed graphs. *IJPRAI* **16**(6), 621–656.

Serratosa, F., Alquzar, R. & Sanfeliu, A. (2000). Efficient algorithms for matching attributed graphs and function-described graphs. In: *In Proceedings of the ICPR2000*.

Serratosa, F. & Sanfeliu, A. (2006). Signatures versus histograms: Definitions, distances and algorithms. *Pattern Recogn.* **39**, 921–934.

Serratosa, F. & Sanromà, G. (2006). An efficient distance between multi-dimensional histograms for comparing images. In: *SSPR/SPR*.

Serratosa, F., Sanromà, G. & Sanfeliu, A. (2007). A new algorithm to compute the distance between multi-dimensional histograms. In: *CIARP*.

Shapiro, L. & Haralick, R. (1981). Structural descriptions and inexact matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **3**, 504–519.

Shapiro, L. S. & Brady, J. M. (1992). Feature-based correspondence: an eigenvector approach. *Image Vision Comput.* **10**(5), 283–288.

Shen, H. C. & Wong, A. K. C. (1983). Generalized texture representation and metric. *Computer Vision, Graphics, and Image Processing* **23**(2), 187–206.

Shi, J. & Tomasi, C. (1994). Good features to track. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*. Seattle.

Shin, D. & Tjahjadi, T. (2010). Clique descriptor of affine invariant regions for robust wide baseline image matching. *Pattern Recogn.* **43**, 3261–3272.

Shokoufandeh, A., Marsic, I. & Dickinson, S. J. (1998). View-based object recognition using saliency maps. *Image and Vision Computing* **17**, 445–460.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J. & Zucker, S. W. (1998). Shock graphs and shape matching. *International Journal of Computer Vision* .

Silletti, A., Abate, A., Axelrod, J. D. & Tomlin, C. J. (2011). Versatile spectral methods for point set matching. *Pattern Recogn. Lett.* **32**, 731–739.

Sinkhorn, R. (1964). Relationship between arbitrary positive matrices + doubly stochastic matrices. *Annals of Mathematical Statistics* **35**(2), 876–&.

Suganthan, P. & Yan, H. (1998). Recognition of handprinted chinese characters by constrained graph matching. *Image and Vision Computing* **16**(3), 191–201.

Sun, J., Ovsjanikov, M. & Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. In: *Proceedings of the Symposium on Geometry Processing*, SGP '09.

Swain, M. J. & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision* **7**, 11–32.

Todorovic, S. & Ahuja, N. (2008). Region-based hierarchical image matching. *Int. J. Comput. Vision* **78**, 47–66.

Tomasi, C. & Kanade, T. (1991). Detection and tracking of point features. Tech. Rep. CMU-CS-91-132, Carnegie Mellon University.

Torsello, A. & Hancock, E. (2004). A skeletal measure of 2d shape similarity. *Computer Vision and Image Understanding* **95**(1), 1–29.

Torsello, A. & Hancock, E. R. (2003). Computing approximate tree edit distance using relaxation labeling. *Pattern Recogn. Lett.* **24**, 1089–1097.

Tsin, Y. & Kanade, T. (2004). A correlation-based approach to robust point set registration. In: *In ECCV*.

Ullmann, J. R. (1976). An algorithm for subgraph isomorphism. *Journal of the ACM* **23**(1), 31–42.

Umeyama, S. (1988). An eigendecomposition approach to weighted graph matching problems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **10**(5), 695–703.

Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 376–380.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In: *The Psychology of Computer Vision*. McGraw-Hill.

Wang, H. & Hancock, E. R. (2006). Graph spectral approach to consistent labelling. In: *Image Analysis and Recognition, PT 2* (Campilho, A and Kamel, M, ed.), vol. 4142 of *Lecture Notes in Computer Science*.

Wang, H. & Hancock, E. R. (2008). Probabilistic relaxation labelling using the Fokker-Planck equation. *Pattern Recognition* **41**(11), 3393–3411.

Wang, Y.-K., chin Fan, K. & tzong Horng, J. (1997). Genetic-based search for error-correcting graph isomorphism. *IEEE Transactions on Systems, Man, and Cybernetics: Part B - Cybernetics* **27**, 588–597.

Werman, M., Peleg, S. & Rosenfeld, A. (1985). A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing* **32**(3), 328–336.

Wilson, R. & Hancock, E. (1997). Structural matching by discrete relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(6), 634–648.

Wilson, R. C., Cross, A. D. J. & Hancock, E. R. (1998). Structural matching with active triangulations. *Comput. Vis. Image Underst.* **72**, 21–38.

Xia, S. & Hancock, E. R. (2009). Graph-based object class discovery. In: *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, CAIP '09. Berlin, Heidelberg: Springer-Verlag.

Yang, J., Williams, J. P., Sun, Y., Blum, R. S. & Xu, C. (2011). A robust hybrid method for nonrigid image registration. *Pattern Recogn.* **44**, 764–776.

Yu, H. & Hancock, E. R. (2005). Graph seriation using semi-definite programming. In: *GbRPR*.

Yuille, A. L. & Grzywacz, N. M. (1989). A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision* **3**(2), 155–175.

Zheng, Y. & Doermann, D. (2006). Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 643–.