IAV2004 - PREPRINTS
5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles
Instituto Superior Técnico, Lisboa, Portugal
July 5-7, 2004

VISUAL TRACKING SYSTEM FOR A MOBILE ROBOT USING COLOUR HISTOGRAMS [1]

Jaume Vergés-Llahí * , **Joan Aranda** ** **and Alberto Sanfeliu** *

* *Inst. de Robòtica i Informàtica Industrial (UPC-CSIC)*
{jverges, asanfeliu}@iri.upc.es
** *Dept. d'Eng. de Sist., Autom. i Inform. Indust. (UPC)*
joan.aranda@.upc.es

Abstract: This work describes the visual system of a mobile robot based on a pan-tilt structure which has been endowed with the ability of tracking moving object using merely colour information. Moving objects in the field of view of the camera are detected and the colour feature of the most relevant regions is selected as the pattern to follow. Colour histograms are used as reliable descriptors to model the appearance of objects. In order to handle with illumination changes a simple adaptation scheme is used. Results show that this system is reliable and fast enough to perform real time tracking of a moving object.

Keywords: mobile robots, tracking systems, real-time image processing, image segmentation, visual pattern recognition, colour histogram, colour adaptation.

1. INTRODUCTION

In this work, we consider the task of tracking moving objects with an active camera. Active vision implies computer vision techniques implemented by means of a movable camera (Aloimonos and Tsakiris, 1991). One of the most important and essential behaviours of an active camera is to keep track of a particular object (the target) in the image by means of computing changes in the position of the target feeding a robotic system which is able to move the camera accordingly.

Two different approaches have been traditionally followed to perform tracking with an active camera: motion-based and feature-based. Classical motion-based techniques only can be applied with a static camera, since background remains unchanged. When a mobile camera is employed, the problem of a moving background can be overcome by stopping the tracker to get measures (Chen and Chang, 1992) or by means of background compensation (Lee *et al.*, 2001).

Both the reliability of the recognition process and the execution time determine the efficiency of a tracking system. Some solutions can reduce the processing time by using expensive computational resources, which reversely limits their generalised applicability to industry. When standard image acquisition hardware is used, the most important limit is the computation time (Arsenio and Santos-Victor, 1997). Colour is an interesting cue which allows a great reduction in the amount of data needed to be processed with an excellent balance between the process time and the lose of robustness in matching the object at each new frame.

In our work, colour histograms have been widely used instead of any other sort of descriptor of

[1] Part. funded by the Gov. of Catalonia and the CICyT DPI2001-2223.

the colour clusters such as Gaussian distributions or mixtures of them, as can be found in the other works (Jang *et al.*, 1998; Heisele *et al.*, 1997; McKenna *et al.*, 1999; Nakamura and Ogasawara, 1999; Schuster, 1994), due to its simplicity, versatility and computational speed, all of them features appreciated in tracking applications.

Moreover, the use of histograms has been vastly proven to be helpful in colour object recognition and object indexing (Swain and Ballard, 1991; Schiele and Crowley, 1996). Colour histograms have also been used in tracking tasks, as in (Birchfield, 1998; Chen *et al.*, 1999). The difference between these works and ours is that we describe a complete robotic system which includes not only the tracking routines but also the head and mobile robot control using just the information brought by colour histograms.

The most important drawback when using colour information is that of its sensitivity to changes that illumination and relative motion between objects and the visual system cause on the appraised colour of tracked objects. In general, colour constancy techniques attempt to avoid these misleading effects. For example, in (Funt and Finlayson, 1995) a correction for the technique of object indexing by colour described in (Swain and Ballard, 1991) is developed.

Other colour adaptation scheme can be found in (Jang *et al.*, 1998; McKenna *et al.*, 1999; Schuster, 1994; Chen *et al.*, 1999), but in this work we only update our histogram description by means of a weight rule filter. In (Vergés-Llahí *et al.*, 2002) an extension of this technique used in human face tracking tasks can be found where other more involved adaptation schemes were applied.

2. THE VISUAL SYSTEM

The goal of this system is to extract the required information from the image sequence grabbed by the head camera to track the course of an object. We assume that initially the robot is stationary with respect to the scene background and we use a model-based approach based on colour histograms to describe the appearance of the tracked object. Next, we present a brief description of the main modules of our visual system.

The system has two different parts: the *model learning* module and the *object tracking* module. As can be seen in Fig. 1, both modules are supplied with images grabbed by the acquisition hardware. The first module aims to segment the moving object out of the background and learning its colour histogram. The second one relocates a previously identified object at each new frame furnishing with updated information about the movement of the object to the control unit of the head and the mobile base. There is a referee unit which switches from one of these modules to the other when the object is lost or when the model has been learnt.

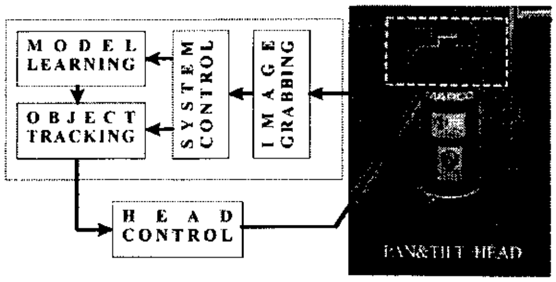


Fig. 1. Visual System Description.

2.1 Object Features

Our object model includes two kinds of features describing both geometry and appearance. First, the geometric features are related to object size and position within the image: area, measured in number of pixels, size of a window delimiting the object, its position and that of the centre of gravity of the object with regard to the image. Secondly, the appearance information describing how the object looks like, i.e., the colour histogram of the object.

Be $\mathcal{I}$ a colour image and $\mathbf{p} = (x, y)$ a pixel of colour $\mathcal{I}(\mathbf{p})$, then a colour histogram is:

$$\mathcal{H} = \{H_{ijk}\}_{1 \le i,j,k \le N} \quad (1)$$

where $H_{ijk} = \#\{\mathcal{I}(\mathbf{p}) \in Bin_{ijk} \mid \forall \mathbf{p} \in \mathcal{I}\}$ is the frequency corresponding to the colour bin Bin_{ijk} and N is the number of bins on each dimension. Then, a set of operations among histograms, such as distance, intersection, union, subtraction, can be defined as follows:

$$dist(\mathcal{H}, \mathcal{T}) = 1 - \sum_{1 \le i,j,k \le N} \min\{\mathcal{H}_{ijk}, \mathcal{T}_{ijk}\} \quad (2)$$

$$\mathcal{H} \cap \mathcal{T} = \{H_{ijk} \cap T_{ijk}\}_{1 \le i,j,k \le N} \quad (3)$$

$$\mathcal{H} \cup \mathcal{T} = \{H_{ijk} \cup T_{ijk}\}_{1 \le i,j,k \le N} \quad (4)$$

$$\mathcal{H} \setminus \mathcal{T} = \{H_{ijk} \text{ if } T_{ijk} = 0\}_{1 \le i,j,k \le N} \quad (5)$$

Those operations are helpful in a histograms-based model to segment images into object components and background components. For example, if $\mathcal{B}$ represents the background and $\mathcal{I}$ the histogram of the image, the histogram of the object can be obtained as $\mathcal{O} = \mathcal{I} \setminus \mathcal{B}$, which is used afterward to segment the object as can be appreciated in Fig. 2.

Moreover, the information obtained from segmenting out the object is used to update the model. Afterward, these results are also employed to compute the corresponding commands to keep the object under track of the pan and tilt head and the robot.

Instead of applying a colour indexing scheme by intersection of histograms to find out where the object is placed, as done in (Swain and Ballard, 1991), the solution we have adopted is very close to the colour histogram backprojection technique in (Chen *et al.*, 1999): we binarize an image by deciding for every pixel whether it belongs or not to the model histogram.

In our case, we compute the intersection in Eq. (3) between a definite neighbourhood $B_\epsilon(\mathcal{I}(\mathbf{p}))$ of the colour of image pixel $\mathbf{p}$ and the model histogram $\mathcal{H}$. If $B_\epsilon(\mathcal{I}(\mathbf{p})) \cap \mathcal{H} \neq \emptyset$, then this colour belongs to the histogram of the model and, therefore, the pixel $\mathbf{p}$ belongs to the object. Fig. 2 b) illustrate how the histogram of the framed area in Fig. 2 a) with a colourful moving object looks like. The main colour of the object (reddish) represents a cluster of the region histogram.

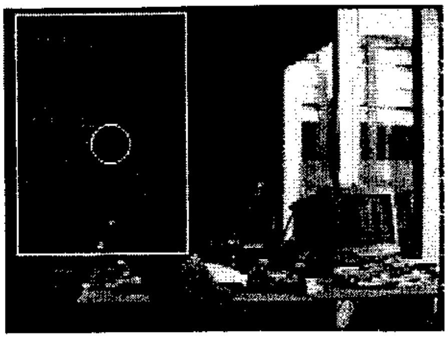
2.2 Model learning

Any ***model based*** tracking algorithm needs to learn a model of the target that it intends to follow. The learning process can be ***supervised*** if the object to track is known before, or ***unsupervised*** otherwise, being tracked the first single moving object in that case. It is also an important issue to realize when a model is outdated and must be re-learnt. In a supervised learning, since we know ***a priori*** what kind of object to track, a ground truth consisting in a fixed model can be established. Then, if the adaptive model goes too far from the ground truth, it is resumed to the fixed one.

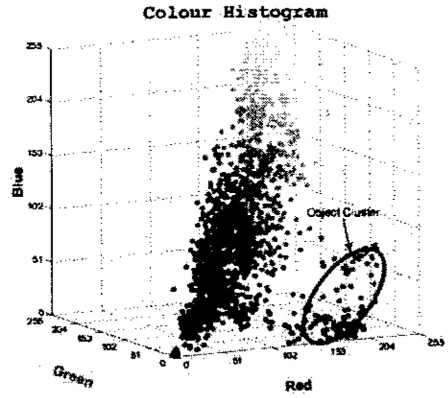
In the unsupervised case, our system tracks the biggest mobile object within its visual field. Unlike the previous case, the system must detect the target before the tracking starts. When the system loses its target, it is resumed again to the previous target-detection step.

In the supervised learning it is necessary to split the histogram of the whole scene from an estimation of the background histogram using Eq. (5) in order to extract the colour histogram object. Area and position can be extracted by running a blob analysis over the binarized image obtained by histogram backprojection of the current image.

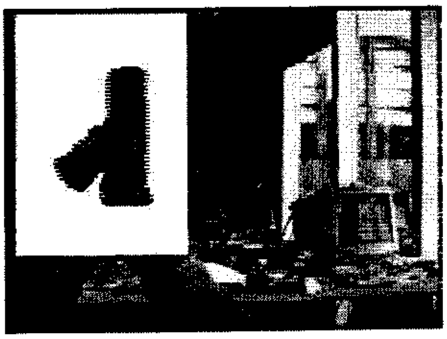
As said, the unsupervised learning approach needs an *a priori* step of detection of independent motion to focus the mobile object search and to learn the new model. The most difficult case is that of a mobile observer because of the self-induced background image motion (Nordlund and Uhlin, 1996).



a) Example of a moving object.



b) Colour histogram of the image.



c) Segmented object.

Fig. 2. Colour histogram of a moving object.

But, for an initially motionless observer, a simple and widespread way to detect the independent motion is frame difference, as done in our work.

The N images of a historical sequence are combined using a set of weights depending on its time position within the sequence: the older an image is, the smaller its weight is. The weights filter out small fluctuations and changes in images through time, and eases to focus on fast changes due to mobile objects (Aranda *et al.*, 1994). The following expression encompasses the process of weighting the set of past images:

$$\mathcal{I}' = \sum_{i=1}^{N} \alpha_i \cdot \mathcal{I}_i \, , \text{ where } \begin{cases} \alpha_i = \dfrac{i}{\sum_{i=1}^{N} i} \\ \displaystyle\sum_{i=1}^{N} \alpha_i = 1 \end{cases} \tag{6}$$

The resulting image $\mathcal{I}'$ is then compared to the actual one at this step to obtain the current difference image $\Delta\mathcal{I}$. If there were some relative motion, differentiated regions would appear and considered to belong to the same object.

2.3 Histogram adaptation

The process of tracking an object has three main steps: object segmentation (colour histogram backprojection), feature extraction (blob analysis) and, finally, histogram adaptation. In this work, we have use a simple adaptation scheme to cope with colour variations based on a weight rule filter as follows:

$$\tilde{\mathcal{H}}^{t+1}_{ijk} = \beta \cdot \tilde{\mathcal{H}}^{t}_{ijk} + (1-\beta) \cdot \mathcal{H}^{t}_{ijk}\,,\ \beta \in [0,1] \quad (7)$$

where $\tilde{\mathcal{H}}^{t+1}$ is the updated histogram value for next time step $(t+1)$, $\tilde{\mathcal{H}}^{t}$ is the updated histogram value at step t and $\mathcal{H}^{t}$ is the computed histogram value at step t. Histogram adaptation should cope with small and slow colour changes due to relative motion between camera, object and illuminant.

3. CONTROL MODULE

The tracking module supplies the head control at 10Hz with the current $[x, y]$ position of the target in the image. The goal consists in keeping the target in the very centre of the image. The error signal $\mathbf{E} = [\varepsilon_x, \varepsilon_y]$ is defined as the difference between the target image position and the image centre pixel coordinates. Both head and camera geometries have been considered to compute the pan and tilt motion (Aranda *et al.*, 1998).

We take advantage of the fact that the camera support keeps the camera rotating around its optical centre, which simplifies to a great extent the head-camera kinematics, although this favourable attribute is not totally fulfilled in practice due to mechanical constraints and calibration errors. Therefore, we define a proportional controller as:

$$\mathbf{\Omega} = \begin{bmatrix} \omega_{pan} \\ \omega_{tilt} \end{bmatrix} = \begin{bmatrix} \frac{k_{pan}}{f} & 0 \\ 0 & \frac{k_{tilt}}{f} \end{bmatrix} \cdot \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \quad (8)$$

where $\mathbf{K}_p = [k_{pan}, k_{tilt}]$ is the vector of proportional gains and f is the camera focal length. Neither the integrative nor the derivative gains have been required in our implementation. Besides the pan and tilt movements of the robotic head, this visual system has been placed on a mobile robot in order to follow about the tracked object.

This platform has only two degrees of freedom, a rotation round the vertical axis and a straight-line movement backward and forward. The rate of this last movement is proportional to the area of the object since – for this experiment – we lack of any other depth estimation. Tracking of the moving object from the two head cameras is in progress, which will allow us to locate the object more accurately.

The rotary motion rate is proportionally adjusted to the difference between the current pan position of the head and a reference angle associated with the direction of the visual system heading forth. This is helpful to compel the robot to being always posed facing the object being tracked. In order to avoid unnecessary robot shaking, a dead zone is defined around the reference angle.

4. EXPERIMENTAL RESULTS

The system overall consists in the pan and tilt unit – a PTU-46-17.5 by Direct Perception Co. – with two orthogonal degrees of freedom moved by stepped motors, a colour camera and a PC Pentium 400MHz. All this equipment is placed on a Pioneer robot by Active Media Robotics. The PTU control unit as well as the robot can be driven by both position and velocity commands.

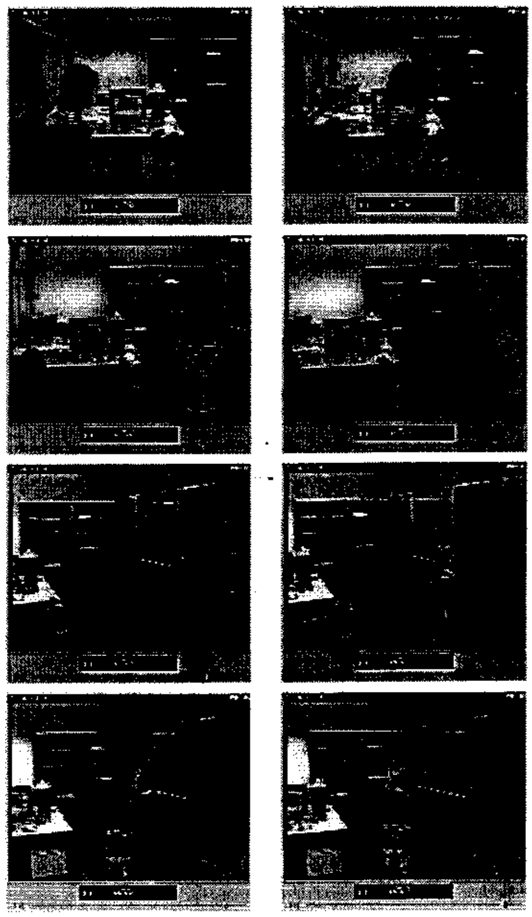


Fig. 3. Mobile robot following a person.

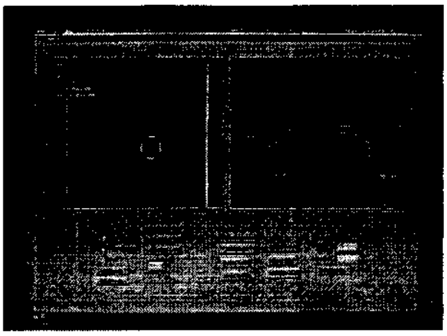


Fig. 4. Object-background segmentation.

The experimental results presented here, including the image processing and visual tracking, are computed in real time (10 Hz). Fig. 3 consists in an excerpt of a sequence where the mobile robot of Fig. 1 follows a person who is walking in a workplace-like environment where no exceptional requirements have been taken.

At a first stage, the person is only followed by the head because she is close to the robot, but as soon as the person progressively goes away and separates from the robot position, it starts the movement towards the person, by rotating and going forth. At the final position, the robot faces the person keeping its distance from her constant.

Fig. 4 shows the obtained results segmenting the person from the background by backprojecting a previously acquired histogram of the person. As can be seen, it is possible to easily extract a complex shape from an unknown background using a histogram where the portion corresponding to the background has been removed.

5. CONCLUSIONS

This work shows the effectiveness of using colour histograms in the complex task of tracking objects, even that of tracking people, using a low cost real-time pan and tilt system based just on the colour information. Such a system has many practical applications, such as robot control feedback, teleoperation, surveillance and security, video-conference systems and home video cameraman. Moreover, it has been shown that colour histograms are also suited for fast object segmentation using histogram backprojection and a simple way to adapt colour information to changes.

REFERENCES

Aloimonos, J.Y. and D.P. Tsakiris (1991). On the visual mathematics of tracking. *Image and Vision Computing* **9**(4), 235–251.

Aranda, J., C. Leon and M. Frigola (1994). *A multi-tracking system for trajectory analysis of people in a restricted area.* Elsevier Science.

Aranda, J., J. Climent and A. Grau (1998). Real-time pan and tilt tracking of moving targets. In: *Controlo98*. pp. 291–295.

Arsenio, A. and J. Santos-Victor (1997). Robust visual tracking by an active observer. In: *IROS97*. Vol. 3. pp. 1342–1347.

Birchfield, S. (1998). Elliptical head tracking using intensity and color histogram. In: *CVPR'98*. pp. 232–237.

Chen, L. and S. Chang (1992). A video tracking system with adaptive predictors. *Pattern Recognition* **25**(10), 1171–1180.

Chen, Tse Min, R.C. Luo and Tsu Hung Hsiao (1999). Visual tracking using adaptive color histogram model. In: *IECON '99*. Vol. 3. pp. 1336–1341.

Funt, B.V. and G.D. Finlayson (1995). Color constant color indexing. *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**(5), 522–529.

Heisele, B., U. Kressel and W. Ritter (1997). Tracking non-rigid, moving objects based on color cluster flow. In: *CVPR'97*. pp. 253–257.

Jang, Daesik and Hyung-Il Choi (1998). Moving object tracking using active models. In: *ICIP98*. Vol. 3. pp. 648–652.

Lee, J.S., K.Y. Rhee and S.D. Kim (2001). Moving target tracking algorithm based on the confidence measure of motion vectors. In: *Proc. of the Int'l Conf. on Image Processing*.

McKenna, S. J., Y. Raja and Shaogang Gong (1999). Tracking colour objects using adaptive mixture models. *Image and Vision Computing* **17**, 225–231.

Nakamura, T. and T. Ogasawara (1999). Online visual learning method for color image segmentation and object tracking. In: *IROS '99*. Vol. 1. pp. 222–228.

Nordlund, P. and T. Uhlin (1996). Closing the loop: Detection and pursuit of a moving object by a moving observer. *Image and Vision Computing* **14**, 265–275.

Schiele, B. and J. L. Crowley (1996). Object recognition using multidimensional receptive field histograms. In: *ECCV'96*. Vol. 1. pp. 610–619.

Schuster, R. (1994). Color object tracking with adaptive modeling. In: *Workshop on Visual Behaviors, International Association for Pattern Recognition*. pp. 91–96.

Swain, M.J. and D.H. Ballard (1991). Color indexing. *Int. Journal of Computer Vision* **7**(1), 11–32.

Vergés-Llahí, J., A. Tarrida and A. Sanfeliu (2002). New approaches for colour histogram adaptation in face tracking tasks. In: *ICPR02*. Vol. 1. pp. 381–384.