

Anàlisi estadístic de la propagació de l'error en estimar el moviment 3D a partir d'imatges

Guillem ALENYÀ¹ Carme TORRAS¹

¹ Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Llorens i Artigas 4-6, 08028-Barcelona, e-mail: {galenya,torras}@iri.upc.edu

Resum—S'analitza la precisió d'un mètode per estimar com es belluga una càmera a partir de les imatges que va captant. D'una vista a una altra es produeix una transformació afí del pla de la imatge caracteritzada per 6 paràmetres, a partir dels quals es calcula la translació i la rotació de la càmera. Mitjançant simulacions de Monte Carlo es valida l'aplicabilitat d'un procediment estadístic menys costós, la *Unscented Transformation*, que permet obtenir les mitjanes i covariances del moviment 3D aplicant el mètode només a 13 mostres per vista. Els resultats indiquen que les translacions i la rotació en el pla de la imatge es recuperen millor que la translació segons l'eix òptic, que a la seva vegada supera les dues rotacions fora del pla. Quant a les covariances, només apareix un cert lligam entre els tres graus de llibertat menys precisos.

Index Terms—propagació de l'error, *Unscented Transformation*, contorns actius

I. INTRODUCCIÓ

Els mètodes per al tractament del soroll en visió per computador han sofert una gran evolució en els darrers anys. Els primers algorismes de geometria de la visió s'ocupaven només de la manipulació de les dades, evitant especificar com s'obtenien. D'aquesta manera, no calia fer consideracions sobre l'error i s'obtenia un estudi teòric del problema en el cas minimal. Aquest és el cas en què s'utilitza el nombre mínim de característiques, generalment punts, que calen per obtenir determinada informació. Obviament l'aplicabilitat d'aquests algorismes en situacions reals és força baixa.

Per tractar amb dades sorolloses l'estrategia tradicional és utilitzar un nombre més elevat de mesures que no pas el mínim necessari. Per exemple, amb 7 correspondències de punts entre dues imatges és possible trobar la matriu fonamental [15], tot i que pot donar lloc a una solució no única. És per això que s'accepta que el nombre mínim de punts per trobar la matriu fonamental és 8 [23], [16]. Així queda definit el mètode algebraic per arribar a la solució. Però aquest algorisme en una situació real amb soroll a les dades d'entrada no retorna un resultat fiable. Per això s'utilitzen més punts, i el problema es converteix llavors en un problema d'estimació on el sistema està sobredeterminat. Per exemple, Liu, Huang i Faugeras [22] van propo-

sar un algorisme d'aquest tipus per localitzar una càmera a partir de punts i línies, primer en 2D i després extenent a 3D. El problema és que normalment en aquests algorismes no es fa cap esforç per tal de modelar l'error, i dades amb molt de soroll introdueixen molt d'error al resultat.

Un dels refinaments possibles és considerar que el conjunt de dades d'entrada conté dades errònies (*outliers*). S'han desenvolupat alguns algorismes per trobar aquests *outliers* i així excloure'ls del conjunt d'entrada. L'estratègia és utilitzar repetides combinacions amb el mínim subconjunt de dades amb què es pot solventar el problema per trobar una solució, i avaluar els resultats amb una funció de distància que cal definir. Les dades errònies produiran les distàncies més grans. Els dos algorismes més coneguts són: Least Median of Squares (LMedS), del qual Zhang [34] en fa una descripció detallada, i Random Sample Consensus (RANSAC), proposat per Fischler i Bolles [12]. Torr i Murray [31] van fer un estudi comparatiu d'ambdues propostes.

Darrerament, diversos autors han mostrat la importància d'assumir que les dades són sorolloses. En la majoria de casos, els algorismes que tracten amb dades sorolloses assumeixen que l'error en les dades es pot modelar amb una distribució gaussiana [15]. Habitualment es representa aquest soroll com una matriu de covariança associada a les dades [25], [6]. Aquesta covariança es pot representar gràficament com un hiperelipsoide d'incertesa. Però la representació gràfica de més de tres dimensions no és fàcil, i per això és habitual fer projeccions 2D de l'hiperelipsoide d'incertesa per tal d'obtenir vistes parcials. En situacions reals s'obtenen millors algorismes quan aquests modelen la incertesa i les solucions s'acompanyen d'una estimació de la covariança. Una bona introducció a l'anàlisi de l'error amb exemples aplicats a casos reals es pot trobar a [7] i a [8]. Brooks i altres [5] utilitzen l'error en forma de covariança en algorismes de correspondències de punts, demostrant la millora que es pot obtenir en relació als algorismes que no tenen en compte la incertesa. També estudien com afecta la incertesa en l'estimació de la pròpia matriu de covariança a la qualitat dels valors estimats. Kanazawa i Kanatani [21]

mostren la importància de poder estimar la incertesa de l'extracció de les característiques en imatges, i presenten un exemple del còmput de la covariança associada a punts en imatges de nivells de gris. Aplicacions de càlcul d'incertesa per localització de robots es poden trobar a [29], [14], [9]. Finalment, Kanatani [20] i Torr [32] estudien la importància de tenir en compte la incertesa del model en si mateix a més de la incertesa en les dades.

Un cop mostrat que és necessari i útil tenir en compte la incertesa en el disseny dels algorismes, cal veure com es pot modelar aquesta incertesa. Una de les maneres més usuals és utilitzar el mètode analític de la propagació de primer ordre. Aquest se serveix de l'expansió en sèrie de Taylor truncada a primer ordre per aproximar linealment els problemes no lineals. Cal un esforç inicial, que sovint no és menyspreable, però s'obté una solució tancada, amb tots els avantatges que això comporta.

En els algorismes d'extracció de moviment a partir d'imatges sovint hi intervé la descomposició en valors singulars (SVD). Per analitzar com es propaga la incertesa es pot estimar el jacobià de la descomposició [26], o analitzar com es propaga la incertesa cap als vectors i valors propis [33].

Un altre camí possible és el d'estimar el jacobià per a tota la transformació no lineal. Drummond i Cipolla [11], utilitzant àlgebres de Lie [27], estimen aquest jacobià per passar de l'espai de deformacions afins en la imatge a l'espai de moviments 3D, i ho apliquen al servocontrol visual. Desgraciadament, aquesta aproximació és vàlida només en determinades condicions i per a moviments infinitesimals. El jacobià obtingut en les condicions més generals del nostre algorisme resulta singular i per tant no és invertible.

També es pot estudiar la propagació de l'error des del punt de vista estadístic. Les simulacions de Monte Carlo són una eina fàcil i potent [10]. És un mètode aplicable quan calen pocs resultats i quan el temps de còmput no representa una restricció important, ja que cal mostrejar densament l'espai d'entrada i executar l'algorisme d'estimació de moviment per a cada mostra. La *Unscented Transformation* [19], proposada per Julier i Uhlmann, es presenta com una millora considerable perquè intenta trobar el mínim nombre de mostres que representen la distribució estadística de les dades d'entrada. Així es disminueix considerablement l'esforç de còmput necessari gràcies a què només s'ha d'estimar el moviment per a un nombre generalment petit de mostres. El mètode es pot validar amb una simulació de Monte Carlo.

Hi ha diverses maneres de definir la incertesa δ_x . Les més comunes són la desviació típica σ_x , i la incertesa relativa (o precisió, mesurada en %). Si no s'indica el contrari, en aquest treball farem servir la desviació típica, que indica que s'estima que el valor té aproximadament el 68%

de possibilitats d'estar en el rang definit per σ_x [30], i el 95% en el rang $2\sigma_x$.

En el cas multivariable, per exemple una deformació afí o bé una posa, la incertesa es pot representar amb una matriu de covariança. Aquesta matriu té a la diagonal la variança de cada variable, i a la resta les covariances de parells de variables.

L'objectiu d'aquest article és fer l'anàlisi de la precisió dels paràmetres de la posa 3D obtinguts, i de les seves covariances. A més es vol obtenir un algorisme que permeti el càlcul en temps real de la covariança associada a cada posa, per tal de fer-lo servir en aplicacions de robòtica mòbil. La Secció II presenta la codificació que es fa de les deformacions afins, i com se'n poden obtenir poses 3D. La següent Secció (III) presenta la representació de les rotacions més convenient. A la Secció IV es mostra un mètode estadístic per a la propagació de la incertesa en funcions no lineals, i a la Secció V es presenten experiments. Finalment les conclusions s'exposen a la Secció VI.

II. DE L'ESPAI DE DEFORMACIONS AFINS DE LA IMATGE A L'ESPAI DE POSES 3D

Sota condicions de perspectiva dèbil (els punts de l'objecte en un rang curt de profunditats, i un objecte centrat a la imatge i petit respecte a la distància de la càmera), el moviment rígid d'un objecte pla a l'espai es projecta al pla de la imatge com una deformació afí de la projecció inicial de l'objecte, que anomenarem *template*.

L'afinitat que relaciona dues vistes normalment es calcula a partir de correspondències de punts. Malauradament, la correspondència de punts és computacionalment molt costosa. En el seu lloc, en aquest treball es fa ús d'un contorn actiu [4]. El contorn actiu es fixa al voltant de la projecció en la imatge de l'objecte que es vol seguir, i els canvis que pateix entre diferents vistes es descriu de la manera que es detalla a continuació. El contorn es representa com una B-Spline, com és usual en Gràfics per Ordinador [13].

S'ha demostrat anteriorment [4], [24] que la diferència en termes dels punts de control $\mathbf{Q}' - \mathbf{Q}$ que quantifica la deformació del contorn es pot escriure com la combinació lineal de sis vectors. Fent servir notació matricial,

$$\mathbf{Q}' - \mathbf{Q} = \mathbf{W}\mathbf{S}, \quad (1)$$

on

$$\mathbf{W} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \mathbf{Q}^y \end{bmatrix}, \begin{bmatrix} 0 \\ \mathbf{Q}^x \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^y \\ 0 \end{bmatrix} \right), \quad (2)$$

i \mathbf{S} és un vector amb els sis paràmetres de la combinació lineal. El vector de forma

$$\mathbf{S} = [t_x, t_y, M_{1,1} - 1, M_{2,2} - 1, M_{2,1}, M_{1,2}], \quad (3)$$

codifica l'afinitat entre dues vistes $\mathbf{d}'(s)$ i $\mathbf{d}(s)$ del contorn pla:

$$\mathbf{d}'(s) = \mathbf{M}\mathbf{d}(s) + \mathbf{t} \quad (4)$$

Es pot veure que si el moviment del robot està restringit, per exemple al pla, la dimensió del vector de forma es pot reduir [3], [2].

Una vegada es té la representació compacta del contorn en termes de punts de control, un filtre de Kalman s'encarrega de seguir el contorn en les diferents imatges de la seqüència [4], actualitzant el nou valor del vector de forma i oferint una estimació de la covariança.

Els paràmetres de l'afinitat es relacionen amb el moviment 3D de la càmera [4], [24] de la següent manera:

$$\mathbf{M} = \frac{Z_0}{Z_0 + T_z} \begin{bmatrix} R_{11} & R_{21} \\ R_{21} & R_{22} \end{bmatrix}, \quad (5)$$

$$\mathbf{t} = \frac{1}{Z_0 + T_z} \begin{bmatrix} T_x \\ T_y \end{bmatrix}, \quad (6)$$

on R_{ij} són els elements de la matriu de rotació 3D \mathbf{R} , i T_i són els elements del vector de translació \mathbf{t} , i Z_0 és la distància des del *template* fins la càmera en la posició inicial.

Veurem tot seguit com s'obtenen cadascuna de les 6 components del moviment 3D. Representant la matriu de rotació en forma d'angles d'Euler,

$$\mathbf{R} = \mathbf{R}_z(\phi)\mathbf{R}_x(\theta)\mathbf{R}_z(\psi), \quad (7)$$

l'equació (5) es pot reescriure com

$$\begin{aligned} \mathbf{M} &= \frac{Z_0}{T_z + Z_0} \mathbf{R}_z|_2(\phi)\mathbf{R}_x|_2(\theta)\mathbf{R}_z|_2(\psi) = \\ &= \frac{Z_0}{T_z + Z_0} \mathbf{R}_z|_2(\phi) \begin{bmatrix} 1 & 0 \\ 0 & \cos\theta \end{bmatrix} \mathbf{R}_z|_2(\psi) \end{aligned} \quad (8)$$

i

$$\mathbf{M}\mathbf{M}^T = \mathbf{R}_z|_2(\phi) \begin{bmatrix} L & 0 \\ 0 & L\cos^2\theta \end{bmatrix} \mathbf{R}_z|_2^{-1}(\phi) \quad (9)$$

on

$$L = \left(\frac{Z_0}{T_z + Z_0} \right)^2.$$

Aquesta darrera equació mostra que θ es pot calcular a partir dels valors propis de la matriu $\mathbf{M}\mathbf{M}^T$, que anomenarem (λ_1, λ_2) :

$$\cos\theta = \sqrt{\frac{\lambda_2}{\lambda_1}}, \quad (10)$$

on λ_1 és el valor propi més gran, i λ_2 , el més petit. L'angle ϕ es pot extreure dels vectors propis de $\mathbf{M}\mathbf{M}^T$; el vector propi \mathbf{v}_1 amb valor propi més gran correspon a la primera columna de $\mathbf{R}_z|_2(\phi)$:

$$\mathbf{v}_1 = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}. \quad (11)$$

Aïllant $\mathbf{R}_z|_2(\psi)$ de l'equació (8),

$$\mathbf{R}_z|_2(\psi) = \left(1 + \frac{T_z}{Z_0}\right) \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\cos\theta} \end{bmatrix} \mathbf{R}_z|_2(-\phi)\mathbf{M}, \quad (12)$$

i observant, a l'equació (9), que

$$1 + \frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}},$$

es pot trobar $\sin\psi$ i llavors ψ .

Un cop coneguts els angles ψ, θ, ϕ , la matriu de rotació \mathbf{R} es pot construir a partir de l'equació (7).

La translació escalada en direcció Z es calcula amb

$$\frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}} - 1. \quad (13)$$

La resta de components de la translació 3D es poden trobar a partir de \mathbf{t} i \mathbf{R} utilitzant l'equació (6):

$$\frac{T_x}{Z_0} = \frac{t_x}{f\sqrt{\lambda_1}}, \quad (14)$$

$$\frac{T_y}{Z_0} = \frac{t_y}{f\sqrt{\lambda_1}}. \quad (15)$$

Així, utilitzant aquestes equacions, es pot transformar la deformació del contorn en el moviment 3D, amb el sistema de referència centrat en l'objecte.

III. REPRESENTACIÓ DE LES ROTACIONS

Fins al moment, les rotacions s'han codificat mitjançant angles d'Euler $\mathbf{R} = \mathbf{R}_z(\phi)\mathbf{R}_x(\theta)\mathbf{R}_z(\psi)$. És conegut que aquesta representació té el problema de la *gimbal lock*: quan dos eixos estan alineats apareix un problema d'indeterminació. En aquesta representació això passa quan la segona rotació $\mathbf{R}_x(\theta)$ és a prop del zero. El resultat és que petites variacions de la posa de la càmera al voltant de la posició inicial no corresponen a valors continus de les components de la rotació ($\mathbf{R}_z(\phi)$ i $\mathbf{R}_z(\psi)$ a la Fig. 1(a)). Aquest problema apareix freqüentment, a l'inici del seguiment i també quan el robot es mou cap a l'objectiu fixat.

Es proposa canviar la representació a la codificació coneguda com *Roll-Pitch-Yaw*. Aquesta representació es fa servir normalment en el camp de la navegació, i també s'anomena *heading, attitude, bank* [28]. Fem servir la forma

$$\begin{aligned} \mathbf{R} &= \mathbf{R}_z(\psi)\mathbf{R}_y(\theta)\mathbf{R}_x(\phi) \\ &= \begin{bmatrix} c\psi c\theta & c\psi s\theta s\phi - s\psi c\phi & c\psi s\theta c\phi + s\psi s\phi \\ s\psi c\theta & s\psi s\theta s\phi - c\psi c\phi & s\psi s\theta c\phi - c\psi s\phi \\ s\theta & c\theta s\phi & c\theta c\phi \end{bmatrix}. \end{aligned} \quad (16)$$

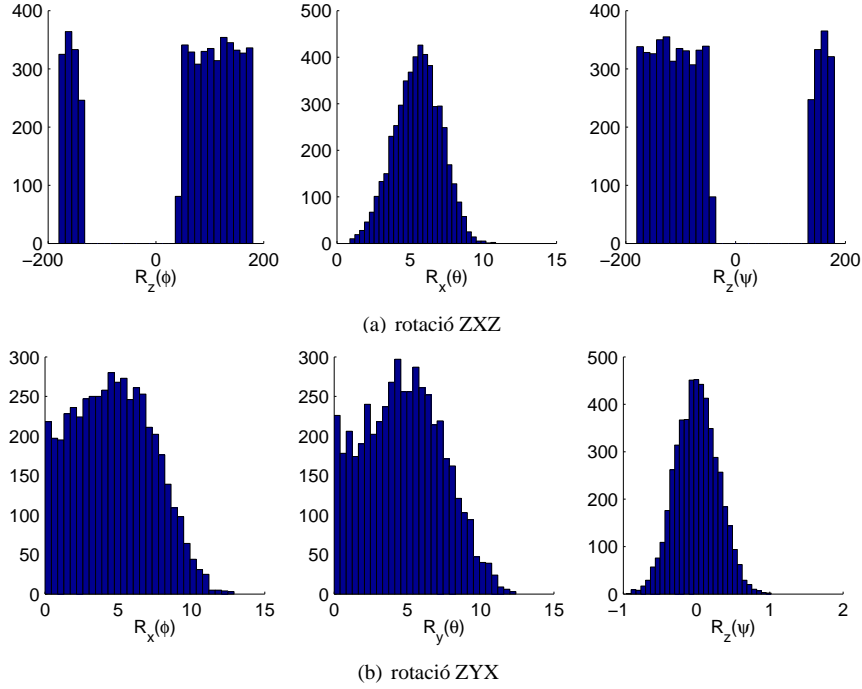


Fig. 1. Histogrames de les components de 5000 rotacions calculades afegint soroll $\sigma = 0.5\text{pixels}$ als punts de control de la projecció d'un contorn. A les rotacions ZXZ , petites variacions de la posa corresponen a valors discontinus de les components $R_z(\phi)$ i $R_z(\psi)$. En canvi, les mateixes variacions corresponen a valors continus en la representació ZYX .

Generalment, per tal de representar tot l'espai de rotacions possibles cadascun dels elements de la rotació es restringeix per pertànyer al rang $[0..2\pi]\text{rad}$ per ψ i ϕ , i a $[0..\pi]\text{rad}$ per θ .

De fet, el seguiment d'un objecte pla per una càmera que rota segons l'eix X o Y més enllà de $\pi/2\text{rad}$ no té sentit, ja que en aquesta posició tots els punts de control estan en una mateixa línia i la forma de l'objecte es perd. A més, degut a l'ambigüitat *Necker reversal*, no és possible determinar el signe de les rotacions entorn d'aquests eixos. Conseqüentment, i sense pèrdua de generalitat, es pot restringir els valors de les rotacions $R_y(\theta)$ i $R_x(\phi)$ per tal que pertanyin al rang $[0..\frac{\pi}{2}]\text{rad}$ i deixar $R_z(\psi)$ al rang de valors $[0..2\pi]\text{rad}$. Amb aquesta representació s'ha aconseguit desplaçar la *gimbal lock* fins a $\cos(\theta) = 0$ que queda fora del rang de valors possibles.

IV. LA Unscented Transformation

La *Unscented Transformation* (UT) va ser proposada per Julier i Uhlmann [18], [19]. És un mètode per propagar els estadístics a través d'una funció general no lineal. La UT no es restringeix a assumir que la distribució d'incertesa de les dades és gaussiana. Es defineix de la següent manera: primer, es trien determinísticament un conjunt de punts (anomenats *sigma points*) que s'espera que capturi els estadístics de la distribució de les dades. Segon, els punts es propaguen fent servir la funció no lineal. Final-

ment, es poden calcular els estadístics amb els punts propagats i així estimar els estadístics del resultat de la funció no lineal.

S'han proposat diferents algorismes per a la tria dels *sigma points*. Farem servir la que originalment es va desenvolupar: la solució Simètrica [18]. Aquest algorisme requereix de $2N_x + 1$ *sigma points*, on N_x és la dimensió del vector de dades d'entrada. Un *sigma point* es situa a la mitjana, i els altres es situen simètricament al voltant, dos en cada dimensió dels eixos principals de la covariància. Es pot representar com un *sigma point* al centre de cada cara d'un hipercub imaginari. A vegades s'anomena UT de segon ordre, ja que garanteix que la mitjana i la covariància (els primers dos moments) són preservats per la transformació.

La variable aleatòria N_x -dimensional \mathbf{x} amb mitjana $\bar{\mathbf{x}}$ i matriu de covariància $\Sigma_{\mathbf{x}}$ es pot aproximar amb el conjunt de punts:

$$\begin{aligned}
 \mathbf{x}^0 &= \bar{\mathbf{x}} \\
 \mathbf{x}^i &= \bar{\mathbf{x}} + \left(\sqrt{\frac{N_x}{1 - W^0} \Sigma_{\mathbf{x}}} \right)_i \quad \text{per } i = 1, \dots, N_x \\
 \mathbf{x}^{i+N_x} &= \bar{\mathbf{x}} - \left(\sqrt{\frac{N_x}{1 - W^0} \Sigma_{\mathbf{x}}} \right)_i \quad \text{per } i = 1, \dots, N_x
 \end{aligned} \tag{17}$$

amb els pesos

$$\begin{aligned} w^0 &= w^0 \\ w^i &= \frac{1 - w^0}{2N_x} \text{ per } i = 1, \dots, N_x \\ w^{i+N_x} &= \frac{1 - w^0}{2N_x} \text{ per } i = 1, \dots, N_x \end{aligned} \quad (18)$$

on $(\sqrt{N_x \Sigma_x})_i$ és la fila o columna i -èsima¹ de l'arrel de la matriu $N_x \Sigma_x$, i w^i és el pes associat amb el i -èsim *sigma point*. Els pesos han de complir la condició $\sum w^i = 1$.

Amb aquest esquema, la mitjana i la covariança es poden calcular a partir dels *sigma points* transformats $\mathbf{y}^i = f(\mathbf{x}^i)$ segons

$$\begin{aligned} \bar{\mathbf{y}} &= \sum_{i=0}^{2n} w^i \mathbf{y}^i \\ \Sigma_y &= \sum_{i=0}^{2n} \{\mathbf{y}^i - \bar{\mathbf{y}}\} \{\mathbf{y}^i - \bar{\mathbf{y}}\}^T. \end{aligned} \quad (19)$$

Per convenció, el primer *sigma point* \mathbf{x}^0 correspon al punt situat a la mitjana. El pes w^0 assignat a aquest punt controla d'alguna manera la situació dels altres punts. Si el pes és positiu, la resta de punts tendeixen a allunyar-se de l'origen, preservant el valor de la covariança. Al contrari, si el pes és negatiu el punt es mouen apropant-se a l'origen [17]. Aquest és un mecanisme per ajustar el moment d'ordre elevat de la distribució.

El reduït nombre de punts que cal transformar per obtenir una estimació de la incertesa associada a cada posa fa que aquest algorisme sigui susceptible de ser aplicat en temps real.

A. Error sistemàtic

Tal i com s'ha presentat anteriorment els *sigma points* es trien simètricament a cada costat de la mitjana en cadascuna de les dimensions de la covariança. Quan la rotació és gairebé nula els dos *sigma points* transformats en la dimensió de la rotació haurien de tenir signe contrari i estar centrats al voltant del 0. Però degut a l'eliminació del signe presentada a la Secció III, els dos *sigma points* tenen el mateix signe, de manera que en aquest cas s'introdueix un biaix en els càlculs de la mitjana i la incertesa de la posa.

Quan es fa servir la UT, la posa no es calcula amb la transformació directa del vector de forma sinó que és una estimació en els *sigma points* transformats. Els dos resultats no tenen per què ser coincidents [19]. En el nostre cas, comparant els dos resultats es pot comprovar que són

¹Depenent de com està formada la matriu Σ cal fer servir les files o les columnes. Si $\Sigma = AA^T$ llavors els *sigma points* es formen amb les files de A. En canvi, si $\Sigma = A^T A$ els *sigma points* es formen amb les columnes de A.

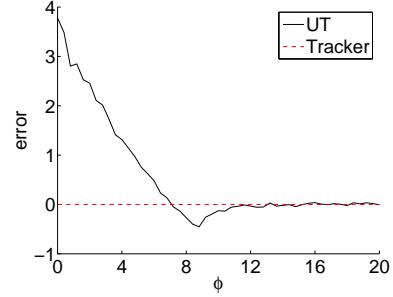


Fig. 2. Error sistemàtic en les rotacions R_x .

sempre coincidents, excepte per a rotacions al voltant de la rotació nul·la. La Fig. 2 mostra l'error del resultat de calcular la posa d'una rotació de 0 fins 20° al voltant de l'eix X. Es pot veure que a mesura que la rotació augmenta l'error disminueix. Al mateix temps, el càlcul de la incertesa també es veu afectada per aquest efecte. Com es veurà més endavant, això no és cap limitació en aplicacions reals.

V. EXPERIMENTS

En aquesta secció es vol avaluar la UT com a mètode per estimar les covariances resultants d'aplicar l'algorisme de càlcul de la posa proposat. Al mateix temps, es vol caracteritzar la incertesa de cada posa per a diferents moviments. S'han dut a terme diversos experiments de simulació en diferents configuracions obtenint una incertesa per simulació de Monte Carlo en cada cas. Aquestes incerteses es consideren certes i es comparen amb les incerteses obtingudes amb la UT. Als punts projectats de l'objectiu se'ls afegeix un soroll $\sigma = 0.5 \text{ pixels}$. Amb una simulació de Monte Carlo es calcula la incertesa de les dades d'entrada (vectors de forma) que es propagarà amb la UT.

A. Recuperació del moviment i incertesa

L'objectiu és determinar experimentalment el grau d'exactitud que s'assoleix en l'estimació de la posa i la seva incertesa per a cada component del moviment de la càmera, és a dir, les translacions T_x, T_y i T_z , i les rotacions R_x, R_y i R_z .

Els primers dos experiments inclouen translacions paral·leles als eixos X i Y. Amb la configuració de càmera triada (5000mm entre càmera i objectiu i una distància focal de 50mm), una translació lateral de la càmera de 250mm fa que la imatge projectada de l'objectiu es desplaça des del centre fins als extrems de la imatge. Els errors en l'estimació de les translacions es presenten a les Figs. 3(a) i 3(b), i tal i com s'esperava són iguals per aquestes dues translacions. Cal observar que a mesura que la càmera es va allunyant de la posició inicial centrada,

l'error en l'estimació de la translació així com la corresponent desviació típica van augmentant. Això es degut a que les condicions de perspectiva dèbil cada cop es compleixen menys (l'objectiu deixa d'estar centrat a la imatge). Amb tot, l'error màxim que s'assoleix és aproximadament del 0.2%, i la pitjor desviació típica és del 0.6%, i per tant, podem dir que les translacions laterals es recuperen amb força precisió. Tal i com s'ha demostrat anteriorment [1], el signe d'aquest error depèn de la forma de l'objectiu i si n'hi ha, de l'orientació de l'eix de rotació.

El tercer experiment involucra translacions sobre l'eix òptic Z : des de la distància inicial $Z_0 = 5000$ la càmera es trasllada fins a $Z = 1500$, és a dir una translació de -3500mm . Els errors i els valors de la incertesa es mostren a la Fig. 3(c). A mesura que la càmera s'acosta a l'objectiu l'error en l'estimació de la translació i la seva incertesa van disminuint. Aquest és un efecte de projecció². Es pot observar que els errors no estan centrats a 0. Més endavant es podrà comprovar que això és degut a la correlació existent entre T_z i R_x , i T_z i R_y quan les rotacions són gairebé nul·les (com és el cas).

Els propers tres experiments inclouen rotacions de la càmera al voltant de l'objectiu. En els dos primers, la càmera es rota al voltant dels eixos X i Y d'un sistema de coordenades centrat a l'objecte. Les Figs. 3(d) i 3(e) mostren els resultats obtinguts. Tal i com s'esperava, els resultats dels dos experiments són molt similars. Cal recordar que es fa servir la representació alternativa presentada a la Secció III, i consegüentment els valors R_x i R_y estan restringits. Tal i com es mostra a la secció IV-A totes les rotacions s'estimen amb el mateix signe i això introdueix un biaix quan les rotacions són petites³. Aquestes rotacions petites produeixen deformacions molt suaus al contorn, i coherentment la incertesa calculada en aquest cas és gran. A mesura que la rotació es va incrementant, la deformació del contorn va creixent i l'error en l'estimació de la rotació així com la incertesa associada van disminuint.

El darrer experiment d'aquesta secció inclou rotacions de la càmera al voltant de l'eix Z . Tal i com s'esperava aquesta component es recupera de manera fiable ja que els errors calculats (Fig. 3(f)) són negligibles i la incertesa és molt petita.

Els experiments presentats són congruents amb la intuïció: les translacions laterals T_x i T_y de la càmera pro-

²La resolució en mil·límetres corresponent a un píxel depèn de la distància de l'objecte a la càmera. Quan l'objectiu és proper a la càmera petites variacions es poden captar fàcilment. En canvi, quan la distància és gran calen moviments més grans per a què la càmera pugui detectar-los.

³A nivell pràctic això no suposa cap limitació ja que, tal i com s'ha vist anteriorment a [1] a la secció d'experiments reals, el soroll present a la fase de captació de la imatge i seguiment emmascara aquestes petites rotacions i l'algorisme no és capaç de recuperar rotacions menors d'aproximadament 10° .

dueixen canvis majors en termes de píxels, i per tant es recuperen millor i amb menor incertesa que les translacions T_z segons l'eix òptic. En canvi, les rotacions segons l'eix òptic R_z provoquen grans deformacions al contorn de manera que es recuperen molt millor i amb menys incertesa que les dues altres rotacions R_x i R_y . Finalment, la incertesa estimada amb la UT és en tots els casos molt semblant a la obtinguda per simulació de Monte Carlo.

B. Covariança

S'han dut a terme diversos experiments que inclouen configuracions de la càmera al voltant de la posició inicial i translacions i rotacions en tots els eixos d'un sistema de coordenades fixat a l'objectiu que es vol seguir. Gairebé tots els resultats han sigut coincidents en no mostrar cap correlació entre les components de la posa, i les incerteses calculades amb la UT i les obtingudes per simulació de Monte Carlo són sempre coincidents, excepte per a rotacions R_x i R_y properes a la rotació nul·la, tal i com ja s'ha presentat a la Secció IV-A. Seguidament es presenten els dos casos especials.

La Fig. 4(a) correspon a un experiment sense moviment, on s'afegeix soroll a les projeccions del contorn. Com es pot observar les components de la posa R_x i T_z estan correlacionades. Quan la rotació R_x es sobreestima, la projecció de l'objectiu a la imatge hauria de tenir mida menor a la que realment té. Per compensar-ho l'algorisme subestima la translació T_z apropant la càmera a l'objectiu. El mateix raonament és aplicable a les components R_y i T_z . A mesura que el valor de la rotació va creixent la correlació entre variables continua existint però el seu efecte és menor. Com es pot observar a la Fig. 4(b) (que correspon a un experiment amb una rotació de 10°) quan el valor de la rotació és prou gran la translació es recupera sense biaix al voltant del 0⁴.

A la Fig. 4(c) es pot veure el segon efecte que es vol presentar: quan es produeix una rotació de la càmera segons l'eix X es computa una petita translació en l'eix Y . S'explica observant el procés de projecció de la càmera de perspectiva dèbil. Quan la càmera es rota, alguns punts 3D de l'objectiu queden més propers a la càmera que d'altres. Els punts més llunyans es projecten lleugerament més aprop del centre òptic que els més propers i consegüentment es computa una petita translació en l'eix Y . De la mateixa manera, si la rotació és entorn de l'eix Y , llavors la translació computada és al llarg de l'eix X .

⁴Aquesta correlació explica la subestimació de la component T_z presentada a la Fig. 3(c), provocada per la sobreestimació que es fa de les rotacions R_x i R_y quan aquestes són petites.

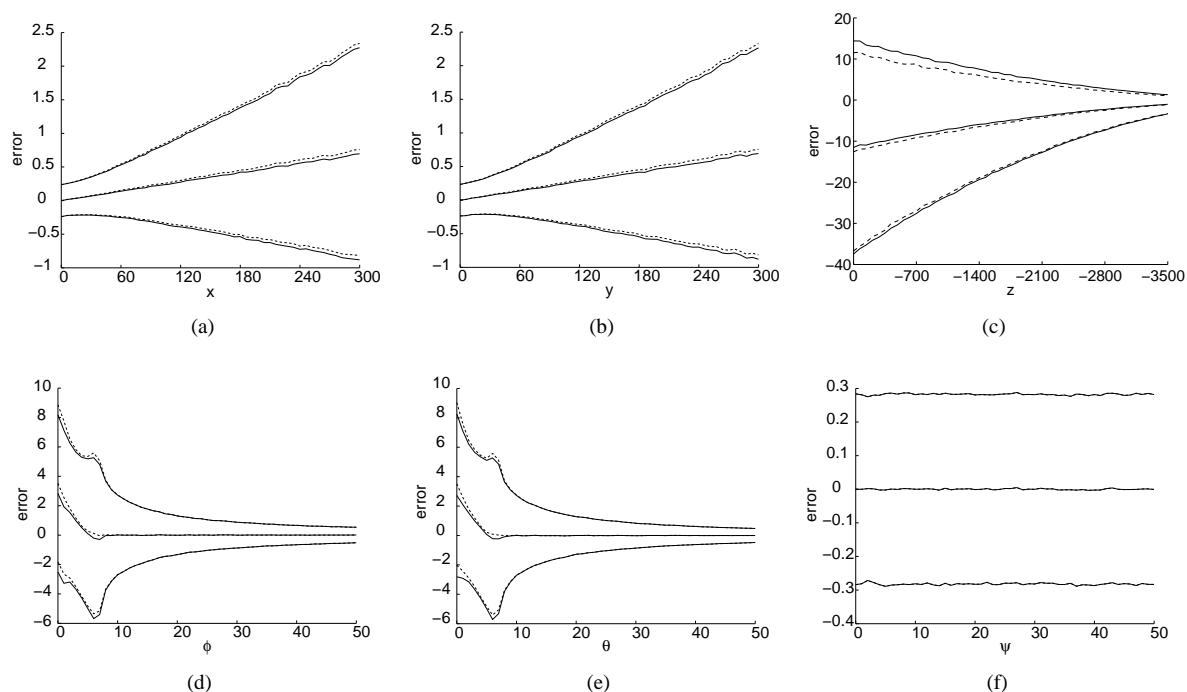


Fig. 3. Mitjana de l'error (línies centrals) i la incertesa 2σ (línies superiors i inferiors) per a moviments purs en cadascun dels eixos coordinats (expressats en mil·límetres les translacions i graus les rotacions). Les línies sòlides corresponen als valors obtinguts amb UT i les línies discontinues a la simulació de Monte Carlo. Els errors per a les translacions T_x i T_y són equivalents, menys quan s'està centrat i creixents a mesura que no s'hi està. Les translacions T_z es recuperen molt pitjor (tot i que milloren a mesura que s'aproximen camera i objectiu). Quan les rotacions R_x i R_y són petites, els errors i les incerteses són grans perquè la deformació del contorn a la imatge és molt petita, però a mesura que van creixent són menys significatives. L'error i la incertesa a les rotacions R_z és negligible.

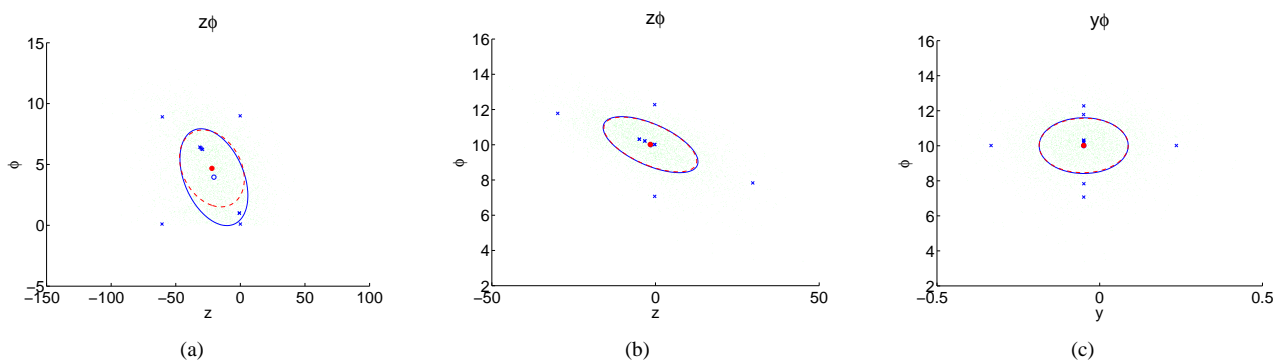


Fig. 4. Representació de les submatrius 2×2 de covariància. Els resultats per a la simulació de Monte Carlo es mostren amb punts verds per a les projeccions dels punts, un punt vermell per a la mitjana i una elipse discontinua per a la incertesa; els resultats de la UT amb creuetes per als *sigma points* i una elipse sòlida per a la incertesa.

VI. CONCLUSIONS

Per a l'estudi estadístic de la propagació de l'error s'han presentat dos mètodes: simulació de Monte Carlo i la *Unscented Transformation*. El primer, que és lent, es fa servir per avaluar el segon, més ràpid. La rapidesa és factor important, ja que l'algorisme ha de respondre en temps real per poder-se aplicar a la robòtica.

S'ha vist que és possible fer servir una representació

ZYX, ja que es pot restringir el rang de valors, i les rotacions en aquest rang són contínues. Això introdueix un biaix, que s'ha vist que es pot detectar comparant el valor de la posa de la mitja dels *sigma points* amb el valor de la posa resultant de transformar el vector de forma. Gràcies a la coincidència entre la posa calculada amb UT, i la posa calculada amb la transformació del vector de forma (que no pateix del biaix esmentat) teòricament és possible estimar correctament rotacions petites sense biaix.

Les translacions en els eixos X i Y es recuperen amb un error molt petit, i tant l'error com la variança augmenten a mesura que la projecció del contorn a la càmera deixa d'estar centrat. Pel que fa a la translació en l'eix Z , es recupera pitjor, i la seva variança també és més gran. S'ha observat que l'error de recuperació i la variança disminueixen a mesura que la càmera s'acosta a l'objectiu.

D'altra banda, les rotacions en els eixos X i Y es recuperen amb un error considerable, sobretot quan les rotacions són petites, però a mesura que es va rotant l'error i la incertesa decreixen. A més, hi ha un esbiaixament del valor de la rotació per valors al voltant de 0 degut a la representació de les rotacions, que va desapareixent a mesura que el valor de la rotació creix. Contràriament, la rotació al voltant de l'eix Z es recupera amb molta precisió i amb una variança molt petita.

S'ha vist que l'única correlació existent es troba entre les rotacions R_x i R_y i la translació T_z . Es pot observar que, quan no hi ha rotació i els valors de les rotacions R_x i R_y estan sobrestimats, el valor de T_z està subestimat (veure la Fig. 4). En canvi, quan R_x i R_y són estimades amb més precisió, T_z també.

S'ha observat que una rotació R_x provoca una lleugera translació en T_y , però com s'ha exposat, això no es produeix d'una correlació sinó de l'assumpció del model de perspectiva dèbil.

Amb el desenvolupament portat a terme i els experiments que s'han realitzat ha quedat provat que és possible calcular la covariança de cada posa amb UT. Els resultats obtinguts s'aproximen molt als esperats per una simulació de Monte Carlo. Finalment, UT compleix amb els requisits de temps de de còmput imposats (ja que només cal transformar 13 *sigma points*), i per tant és possible utilitzar aquest mètode en robots i aplicacions reals.

REFERÈNCIES

- [1] M. Alberich-Carramiñana, G. Alenyà, J. Andrade-Cetto, E. Martínez, and C. Torras. Affine epipolar direction from two views of a planar contour. Technical Report IRI-DT-2005/03, Institute of Robotics (IRI), 2005.
- [2] G. Alenyà, J. Escoda, A.B. Martínez, and C. Torras. Using laser and vision to locate a robot in an industrial environment: A practical experience. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 3539–3544, Barcelona, Apr. 2005.
- [3] G. Alenyà, E. Martínez, and C. Torras. Fusing visual and inertial sensing to recover robot egomotion. *Journal of Robotic Systems*, 21:23–32, 2004.
- [4] A. Blake and M. Isard. *Active contours*. Springer, 1998.
- [5] M. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel. What value covariance information in estimating vision parameters? In *International Conference on Computer Vision*, pages 302–308, 2001.
- [6] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. On the fitting of surfaces to data with covariances. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(11):1294–1303, 2000.
- [7] J.C. Clarke. Modelling uncertainty: A primer. Technical Report 2161/98, University of Oxford. Dept. Engineering science, 1998.
- [8] A. Criminisi. *Accurate visual metrology from single and multiple uncalibrated images*. Springer, 2001.
- [9] A. Davidson. *Mobile robot navigation using active vision*. PhD thesis, University of Oxford, 1999.
- [10] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer, 2001.
- [11] T. Drummond and R. Cipolla. Application of lie algebras to visual servoing. *Int. J. Comput. Vision*, 37(1):21–41, 2000.
- [12] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24:381–385, 1981.
- [13] J. Foley, A. van Dam, S. Feiner, and F. Hughes. *Computer Graphics. Principles and Practice*. Addison-Wesley Publishing Company, 1996.
- [14] N. Gonçalves and H. Araújo. Analysis and comparison of two methods for the estimation of 3d motion parameters. *Robotics and Autonomous Systems*, 45(1):23–50, 2003.
- [15] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2004.
- [16] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(6):580–593, June 1997.
- [17] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Automat. Contr.*, 45(3):477–482, Mar. 2000.
- [18] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In I. Kadar, editor, *Proc. 11th SPIE Int. Sym. Aerospace/Defense Sensing, Simulation, Controls*, pages 182–193, Orlando, Apr. 1997. SPIE.
- [19] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proc. IEEE*, 92(3):401–422, Mar. 2004.
- [20] K. Kanatani. Uncertainty modeling and model selection for geometric inference. *IEEE Trans. Pattern Anal. Machine Intell.*, 26(10):1307–1319, 2004.
- [21] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *Proc. IEEE Int. Conf. Comput. Vision*, pages 301–306, Vancouver, BC, Canada, 2001.
- [22] Y. Liu, T.S. Huang, and O.D. Faugeras. Determination of camera location from 2d to 3d lines and point correspondences. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(1):28–37, 1990.
- [23] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293(11):133–135, 1981.
- [24] E. Martínez. *Recovery of 3D structure and motion from the deformation of an active contour in a sequence of monocular images*. PhD thesis, Universitat Politècnica de Catalunya, 2000.
- [25] D. D. Morris, K. Kanatani, and T. Kanade. Uncertainty modeling for optimal structure from motion. In *Vision Algorithms Theory and Practice*. Springer LNCS, September 1999.
- [26] T. Papadopoulos and I. A. Loukakis. Estimating the jacobian of the svd: theory and applications. Technical Report RR-3961, INRIA, 2000.
- [27] W. Rossmann. *Lie groups: an introduction through linear groups*. Oxford University Press, 2002.
- [28] L. Sciavicco and B. Siciliano. *Modeling and Control of Robot Manipulators*. Springer-Verlag, London, 2000.
- [29] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Int. J. Robot. Res.*, 21(8):735–758, Aug. 2002.
- [30] J. R. Taylor. *An Introduction to Error Analysis The Study of Uncertainties in Physical Measurements*. University Science Books, second edition edition, 1997.
- [31] P. Torr and D. Murray. Outlier detection and motion segmentation. *Sensor Fusion VI, SPIE*, 2059:432–443, 1993.
- [32] P.H.S. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *Int. J. Comput. Vision*, 50(1):35–61, 2002.
- [33] J. Weng, T. S. Huang, and N. Ahuja. *Motion and structure from image sequences*. Springer-Verlag, 1993.
- [34] Z. Zhang. Determining the epipolar geometry and its uncertainty: a review. *Int. J. Comput. Vision*, 27(2):161–195, 1998.