

# Estimació del moviment d'un robot observant contorns actius

Guillem ALENYÀ<sup>1</sup>, Elisa MARTÍNEZ<sup>2</sup>, Carme TORRAS<sup>3</sup>

<sup>1</sup>Robosoft, Teknogunea Izarbel, 64210-Bidart, França, e-mail: galenya@iri.upc.es

<sup>2</sup>CiTS, Universitat Ramón Llull, Pge. Bonanova 8, 08022-Barcelona, e-mail: elisa@salleurl.edu

<sup>3</sup>Institut de Robòtica (CSIC-UPC), Llorens i Artigas 4-6, 08028-Barcelona, e-mail: torras@iri.upc.es

**Abstract**—Es presenta un mètode per estimar el moviment d'un robot a partir del seguiment de contorns actius en una successió d'imatges adquirides amb un sistema monocular i calibrat, muntat sobre el propi robot. Un cop el contorn actiu ha estat ajustat al voltant de la projecció d'un objecte al pla de la imatge, les deformacions afins que pateix el contorn en la seqüència es parametrizen utilitzant un vector de forma de dimensió constant. S'analitzen les implicacions d'utilitzar vectors de forma de dimensió 3, 4 i 6. A partir d'aquests vectors de forma es pot trobar la rotació i translació 3D que ha provocat les deformacions, és a dir, el moviment propi de la càmera. Els resultats preliminars obtinguts amb un robot rodant sobre un terra pla i vectors de forma de 4 components són força encoratjadors.

## I. INTRODUCCIÓ

L'objectiu d'aquest treball és obtenir estimacions del moviment d'un robot. Tradicionalment, les tècniques basades en característiques han fet servir punts, línies, cantonades i combinacions d'elles per analitzar les imatges. Aquest treball proposa els contorns com a característica a utilitzar. El seguiment dels contorns proporciona una estimació del moviment referenciat al contorn inicial.

L'algorisme que permet trobar els paràmetres del moviment consta dels següents quatre passos. Primer, cal trobar un objecte candidat i cal inicialitzar un contorn actiu al voltant de la seva silueta. Aquest contorn es representa en forma de b-spline i es defineix amb una sèrie de punts de control. Actualment el contorn l'inicialitza manualment un operador. També és possible utilitzar algorismes d'inicialització automàtica, com el proposat per Cham i Cipolla [3].

En el segon pas, un algorisme de seguiment reconeix el canvi en la posició de la projecció del contorn a la imatge. El model de càmera amb què es treballa és el de perspectiva dèbil, de manera que només es modelen deformacions afins del contorn inicial [2]. Qualsevol transformació es parametritza com una deformació afí del contorn, sense tenir en compte per exemple efectes perspectius, de manera que una matriu de forma de 6 paràmetres és suficient per definir la família de deformacions possibles. L'algorisme de seguiment [4] fa ús de la tècnica del

filtre de Kalman i utilitza com vector d'estat el vector de forma de 6 posicions que representa la deformació afí obtinguda de la família de deformacions possibles. Això permet, d'una banda, obtenir un seguiment més estable, i, de l'altra, acompanyar cada estimació  $S$  amb una covariança associada  $\Gamma_S$ .

En tercer lloc, a partir dels vectors de forma es dedueix l'estimació de la transformació 3D [1]. Aquesta transformació es refereix a l'instant inicial i es computa de manera no acumulativa. És per això que no pateix del típic biaix que es dona en els algorismes incrementals.

Finalment, si es tenen estimacions de més d'un contorn a la vegada, cal fusionar la informació per tal de donar una resposta global del sistema. També es poden obtenir altres estimacions per mitjà d'altres sensors o tècniques [6] que caldrà incorporar de la mateixa manera a l'algorisme d'estimació del moviment.

La resta de l'article s'estructura de la següent manera: la propera secció introdueix els vectors de forma i les diferents dimensions que poden prendre. La Secció III explica com deduir la informació 3D a partir de vectors de forma, tenint en compte els diferents sistemes de referència que es poden emprar. Els resultats de diversos experiments es presenten a la següent secció, i finalment, algunes conclusions i treballs futurs s'exposen a la Secció V.

## II. DEFORMACIONS D'UN CONTORN

Tal i com s'ha esmentat a la introducció, el model de càmera emprat és el de perspectiva dèbil. Aquest és un model reduït que contempla només a transformacions afins. Això implica fer una sèrie d'assumpcions [5] que caldrà tenir en compte. En primer lloc, l'objecte ha de ser planar. És a dir, la diferència en la profunditat  $\Delta z$  dels punts del seu contorn no pot ser gran en comparació amb la distància  $Z_0$  entre la càmera i l'objecte. En segon lloc, el model de perspectiva dèbil és més inexacte a mesura que els punts del contorn s'allunyen del centre de la imatge. Això implica tenir contorns aproximadament centrats en la imatge. I en tercer lloc, el contorn ha d'envoltar una regió relativament petita dins de la imatge. De fet,

aquesta última és una condició implícita per poder plantejar el seguiment del contorn al llarg d'una seqüència sense perdre'l de vista prematurament.

També cal esmentar que la inicialització del contorn en forma de b-spline determina la família de deformacions afins que es podran donar, de manera que una bona inicialització és clau per poder realitzar un bon seguiment, mentre que una inicialització deficient d'una banda dificultarà el seguiment, i, de l'altra, disminuirà la qualitat dels vectors de forma obtinguts.

Un contorn parametrizat en forma de b-spline es pot representar mitjançant el vector dels seus punts de control,  $Q$ . Considerant  $Q_0$  els del contorn en l'instant inicial, la deformació del contorn es pot definir com

$$Q = WS + Q_0, \quad (1)$$

on  $W$  és l'anomenada matriu de forma, que determina la família de deformacions afins possibles, i  $S$  és el vector de forma, que determina la deformació concreta.

En el cas general, el cas de les deformacions afins, la matriu és de la forma

$$W = \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} Q_0^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^y \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^x \end{bmatrix}, \begin{bmatrix} Q_0^y \\ 0 \end{bmatrix} \right), \quad (2)$$

i el vector de forma s'expressa [1]:

$$S = (t_x, t_y, M_{11} - 1, M_{22} - 1, M_{21}, M_{12})^T. \quad (3)$$

Si restringim el tipus de moviment del robot, és possible definir matrius de forma reduïdes. En el cas d'un robot rodant sobre un terra pla, el nombre de graus de llibertat és 3: dues translacions,  $T_x$  i  $T_z$ , que formen el pla  $XZ$ , i una rotació en l'eix vertical del robot,  $\theta_y$ . Es pot definir la matriu de forma com:

$$W = \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} Q_0^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^y \end{bmatrix} \right), \quad (4)$$

i el vector de forma

$$S = (t_x, M_{11} - 1, M_{22} - 1)^T. \quad (5)$$

Per tal d'assegurar que el moviment planar del robot es tradueix també en un moviment planar de la càmera, cal conèixer els sistemes de referència de cadascun, i assegurar que ambdós estan alineats. Qualsevol error en la calibració robot-càmera pot provocar que un moviment en el pla  $XZ$  del robot es tradueixi en un moviment a l'espai 3D dels objectes vistos per la càmera.

Per tal de fer un seguiment més robust d'aquests tipus de moviments es pot ampliar la matriu de forma anterior

amb un grau de llibertat addicional, les translacions  $T_y$ . D'aquesta manera, la matriu de forma ampliada es pot escriure de la forma:

$$W = \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} Q_0^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^y \end{bmatrix} \right), \quad (6)$$

i el vector de forma com

$$S = (t_x, t_y, M_{11} - 1, M_{22} - 1)^T. \quad (7)$$

### III. RECUPERACIÓ DEL MOVIMENT

#### A. Origen del sistema de referència a l'objecte

Sota condicions de perspectiva dèbil, el moviment rígid d'un contorn a l'espai 3D,  $\mathbf{D}(s)$ , relatiu a un contorn de referència,  $\mathbf{D}_0(s)$ ,

$$\mathbf{D}(s) = \mathbf{R}\mathbf{D}_0(s) + \mathbf{T},$$

es projecta al pla de la imatge com una deformació afí del contorn de referència, que anomenarem *template*. Aquesta deformació respecte el *template* es pot expressar com

$$\mathbf{d}(s) - \mathbf{d}_0(s) = (\mathbf{M} - \mathbf{I})\mathbf{d}_0(s)$$

on  $\mathbf{I}$  és la matriu identitat,

$$\mathbf{M} = \frac{Z_0}{Z_0 + T_z} \begin{bmatrix} R_{11} & R_{21} \\ R_{21} & R_{22} \end{bmatrix}, \quad (8)$$

$$\mathbf{t} = \frac{1}{Z_0 + T_z} \begin{bmatrix} T_x \\ T_y \end{bmatrix}, \quad (9)$$

on  $\mathbf{R}_{ij}$  són els elements de la matriu de rotació 3D,  $\mathbf{R}$ ;  $\mathbf{T}_i$  són els elements del vector de translació,  $\mathbf{T}$ ; i  $Z_0$  és la distància des del *template*  $\mathbf{D}_0(s)$  fins la càmera en la posició inicial.

Representant la matriu de rotació en forma d'angles d'Euler,

$$\mathbf{R} = \mathbf{R}_z(\phi)\mathbf{R}_x(\theta)\mathbf{R}_z(\psi), \quad (10)$$

l'equació (8) es pot reescriure com

$$\begin{aligned} \mathbf{M} &= \frac{Z_0}{T_z + Z_0} \mathbf{R}_z|_2(\phi)\mathbf{R}_x|_2(\theta)\mathbf{R}_z|_2(\psi) = \\ &= \frac{Z_0}{T_z + Z_0} \mathbf{R}_z|_2(\phi) \begin{bmatrix} 1 & 0 \\ 0 & \cos\theta \end{bmatrix} \mathbf{R}_z|_2(\psi) \end{aligned} \quad (11)$$

i

$$\mathbf{M}\mathbf{M}^T = \mathbf{R}_z|_2(\phi) \begin{bmatrix} L & 0 \\ 0 & L\cos^2\theta \end{bmatrix} \mathbf{R}_z|_2^{-1}(\phi) \quad (12)$$

on

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

Aquesta darrera equació mostra que  $\theta$  es pot calcular a partir dels valors propis de la matriu  $\mathbf{M}\mathbf{M}^T$ , que anomenarem  $(\lambda_1, \lambda_2)$ :

$$\cos\theta = \sqrt{\frac{\lambda_2}{\lambda_1}}, \quad (13)$$

on  $\lambda_1$  és el valor propi més gran, i  $\lambda_2$ , el més petit. L'angle  $\phi$  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  $\mathbf{R}_z|_2(\phi)$ :

$$\mathbf{v}_1 = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}. \quad (14)$$

Aïllant  $\mathbf{R}_z|_2(\psi)$  de l'equació (11),

$$\mathbf{R}_z|_2(\psi) = \left(1 + \frac{T_z}{Z_0}\right) \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\cos\theta} \end{bmatrix} \mathbf{R}_z|_2(-\phi)\mathbf{M}, \quad (15)$$

i observant, a l'equació (12), que

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

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

Un cop coneguts els angles  $\psi, \theta, \phi$ , la matriu de rotació  $\mathbf{R}$  es pot construir a partir de l'equació (10).

La translació escalada en direcció  $Z$  es pot calcular com

$$\frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}} - 1. \quad (16)$$

La resta de components de la translació 3D es poden trobar a partir de  $\mathbf{t}$  i  $\mathbf{R}$  utilitzant l'equació (9):

$$\frac{T_x}{Z_0} = \frac{t_x}{f\sqrt{\lambda_1}}, \quad (17)$$

$$\frac{T_y}{Z_0} = \frac{t_y}{f\sqrt{\lambda_1}}. \quad (18)$$

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

#### B. Origen del sistema de referència a la càmera

En aquest cas, les equacions (8) i (9) passen a ser:

$$\mathbf{M} = \frac{Z_0}{R_{33}Z_0 + T_z} \begin{bmatrix} R_{11} & R_{21} \\ R_{21} & R_{22} \end{bmatrix}, \quad (19)$$

$$\mathbf{t} = \frac{1}{R_{33}Z_0 + T_z} \begin{bmatrix} Z_0R_{13} + T_x \\ Z_0R_{23} + T_y \end{bmatrix}. \quad (20)$$

Procedint com abans, podem obtenir  $\psi, \theta$  i  $\phi$ , i construir la matriu de rotació  $\mathbf{R}$  a partir de l'equació (10).

La translació escalada en direcció  $Z$  es pot calcular com

$$\frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}} - R_{33}. \quad (21)$$

La resta de components de la translació 3D es poden trobar a partir de  $\mathbf{t}$  i  $\mathbf{R}$  utilitzant l'equació (20):

$$\frac{T_x}{Z_0} = \frac{t_x}{f\sqrt{\lambda_1}} - R_{13}, \quad (22)$$

$$\frac{T_y}{Z_0} = \frac{t_y}{f\sqrt{\lambda_1}} - R_{23}. \quad (23)$$

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

## IV. EXPERIMENTS

Tots els experiments mostrats en aquesta secció corresponen a la mateixa seqüència de vídeo. Aquesta seqüència ha estat gravada amb una càmera a color muntada sobre un robot planar. Aquest robot disposa també d'un làser de posicionament. Per tal de poder posar en correspondència les estimacions del làser amb les estimacions del sistema de visió s'ha pres la decisió de baixar la freqüència d'adquisició del sistema vídeo de 25 imatges per segon a 8 imatges per segon, que és precisament la freqüència del sistema làser. Així doncs, les seqüències analitzades corresponen a 25 segons de navegació, és a dir, 200 imatges. Cal fer notar però, que la implementació feta del sistema de seguiment i deducció del moviment 3D funciona correctament a freqüències de 20 i 25 imatges per segon fent el seguiment simultani de diversos contorns.

#### A. Sistema de referència centrat a l'objecte

La seqüència de vídeo és analitzada i s'obté un conjunt de vectors de forma de dimensió 6, corresponents a l'equació (3). Utilitzant les equacions (16, 17 i 18) es pot deduir la translació de l'objecte, i fent servir (13, 14 i 15) les rotacions. És clar que aquestes dades no són les buscades, el moviment del robot, però ens permeten observar la naturalesa de les dades obtingudes.

A la Fig. 1 es poden veure les translacions estimades per a la seqüència de vídeo, mentre que a la Fig. 2 es poden observar les rotacions estimades. De les tres rotacions, la rotació  $\theta_z$  es força estable. Contràriament, les rotacions  $\theta_x$  i  $\theta_y$  contenen un nivell molt alt de soroll.

A la Fig. 3 es pot veure la translació  $T_x$  i  $T_z$  relacionades. El moviment recuperat pel que fa a la translació en el pla de moviment del robot sembla plausible, i resta fer la comparació amb el moviment real per veure la seva precisió.

#### B. Sistema de referència a la càmera

En aquest experiment, l'objectiu és obtenir els paràmetres del moviment, però amb el sistema de referència a la càmera. La mida de la matriu de forma em-

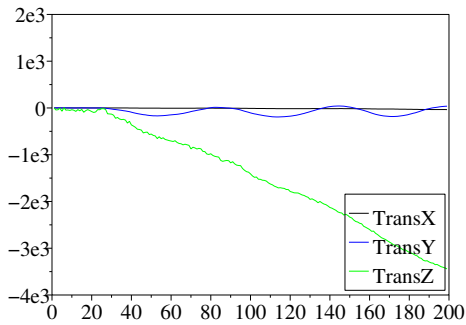


Fig. 1. Translacions a cada imatge SR-Objecte

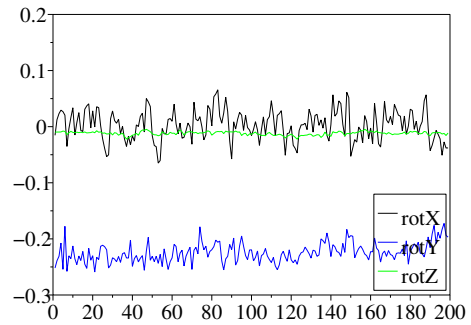


Fig. 4. Rotacions a cada imatge SR-Càmera

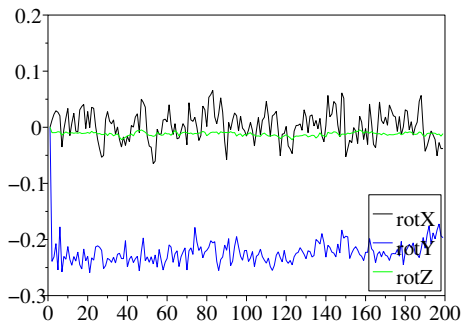


Fig. 2. Rotacions a cada imatge SR-Objecte

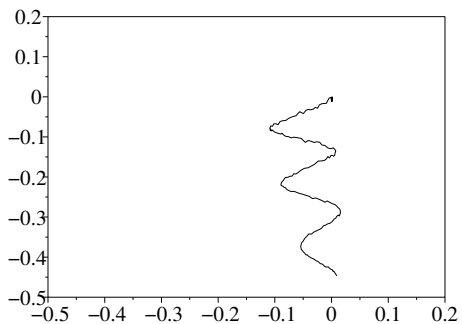


Fig. 3. Moviment estimat al pla SR-Objecte

es transmet a l'estimació de les translacions com es pot observar a la Fig. 5.

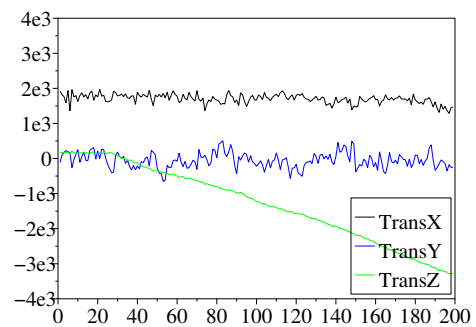


Fig. 5. Translacions a cada imatge SR-Càmera

Posant la translació  $T_x$  en funció de  $T_z$  (Fig.6) es pot observar com aquest soroll distorsiona la sortida, fins al punt de fer-la difícilment vàlida com a estimació del moviment.

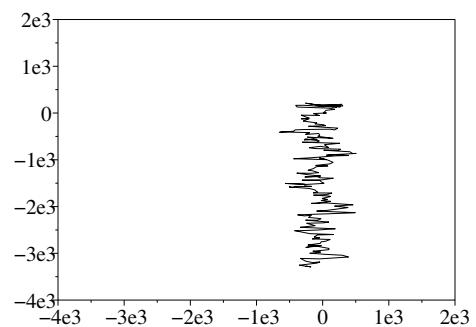


Fig. 6. Moviment estimat al pla ZX SR-Càmera

prada és de 6 components (2), modelant moviments en tot l'espai, incloses les 3 rotacions possibles. A la Fig. 4 es poden veure les tres rotacions estimades per a la seqüència. Així doncs, sabem que les rotacions en  $\theta_x$  i  $\theta_z$  no s'han produït en la realitat, i per tant caldrà considerar els valors obtinguts com soroll del sistema.

Observant les equacions (21, 22 i 23) veiem que per tal d'estimar la translació realitzada és necessari utilitzar els valors de la rotació. Com ja hem explicat anteriorment, aquests valors es recuperen amb un cert soroll, soroll que

### C. Matriu de forma reduïda

Tal i com s'ha vist a la Secció II, quan el moviment 3D és restringit és possible utilitzar una matriu de forma reduïda. Els assajos amb la matriu de dimensió 3 han donat com a resultat la impossibilitat de fer correctament el seguiment del contorn dins de la imatge, ja que en el moment de la presa de dades no es va tenir en compte d'alinear correctament els sistemes de referència de la càmera i el robot. És per això que la matriu de forma utilitzada és la definida a (6), obtenint els vectors de forma de 4 components definits a (7).

A les Figs. 7 i 8 es poden veure respectivament les translacions i la rotació obtingudes pel procés de seguiment.

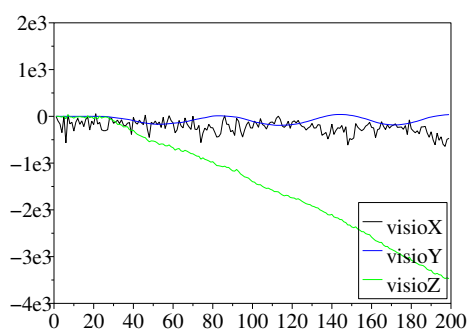


Fig. 7. Translacions amb la matriu reduïda

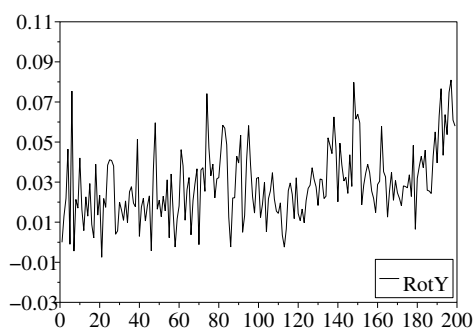


Fig. 8. Rotació amb la matriu reduïda

Quan es posen en correspondència  $T_x$  i  $T_z$  el resultat obtingut (Fig. 9) s'acosta molt més al moviment real, que es podrà observar a la següent secció.

### D. Comparació amb les dades làser

Si es volen implementar aquestes propostes en robots reals caldrà mesurar la precisió que es pot obtenir. El darrer experiment tracta d'analitzar la correctesa del moviment estimat. Per tal d'avaluar la precisió del moviment

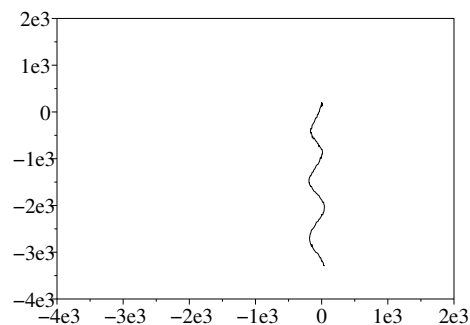


Fig. 9. Moviment en el pla ZX

recuperat, cal disposar d'unes dades de referència. En aquest experiment es fa servir de referència un làser de posicionament per triangulació.

El primer problema és el de calibrar la relació laser-càmera. Cada sensor estima el moviment del robot en el seu propi sistema de referència. Per poder comparar les estimacions, cal obtenir els valors de translació i rotació que posen en correspondència un sistema de referència i l'altre.

A la Fig. 10 es poden veure els errors computats per a les dues translacions. D'una banda, es mostren els errors derivats de fer l'estimació amb la càmera respecte al làser, sense tenir en compte la calibració. De l'altra, les estimacions de la càmera es comparen amb les del sistema làser, un cop aquestes han estat corregides tenint en compte el vector que relaciona un sistema de referència i l'altre.

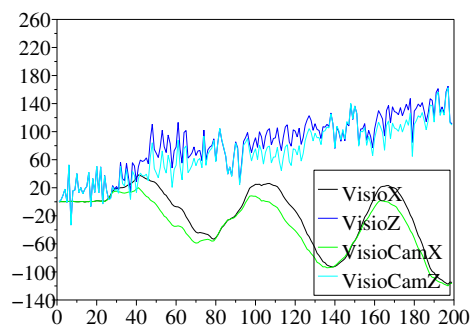


Fig. 10. Errors calculats respecte el làser

Posant una translació respecte l'altra es pot observar l'error que es comet considerant el sistema làser com a referència (Fig. 11).

## V. CONCLUSIONS

S'ha presentat una proposta per tal d'estimar el moviment propi d'un robot a partir de deformacions afins de

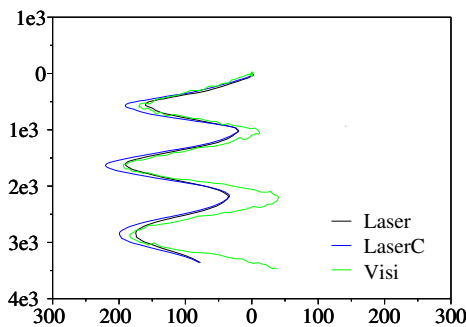


Fig. 11. Posicions en el pla de moviment: làser, làser desplaçat al SR-Càmera i càmera

contorns actius. De l'anàlisi de les imatges s'obté una parametrització d'aquest moviment en forma de vector de forma. S'ha mostrat també com, a partir dels vectors de forma, es pot obtenir l'estimació del moviment 3D.

S'ha mostrat que aquest moviment és possible deduir-lo tant per a un sistema de referència centrat a l'objecte com per a un sistema de referència centrat a la càmera. Aquest darrer cas és el que ens permetrà més endavant utilitzar la informació calculada, per exemple, per realimentar el control del robot, comparar i fusionar estimacions de diversos contorns, o fusionar aquestes dades amb dades d'altres sensors. Els experiments han mostrat la viabilitat d'aquestes propostes en robots reals amb moviment en un pla.

Futurs treballs haurien d'incloure l'aprofitament de la informació de covariància [7] [8] provinent del filtre de Kalman utilitzat. Aquesta informació hauria de permetre una fusió guiada de la informació de diversos contorns, així com la fusió amb informació d'altres tipus de sensors. Cal també fer experiments on les dades de la calibració entre la càmera i la font de les dades de referència, en aquest cas el sistema làser, siguin més precises.

Com s'ha vist, per tal de donar informació mètrica, i no

escalada, cal conèixer la distància inicial entre l'objecte i la càmera. Caldrà trobar una manera d'estimar aquesta distància sense intervenció de l'operador [9]. Finalment, cal provar la validesa de les propostes en els graus de llibertat no explorats fins al moment.

## VI. AGRAÏMENTS

Els experiments d'aquest treball han estat realitzats en el marc del projecte "Sistema de posicionament làser i visual", finançat pel Centre de Referència en Tecnologies Avançades de la Producció (CeRTAP) de la Generalitat de Catalunya. Els autors agraeixen a Antonio B. Martínez i Josep Escoda, del Departament d'Esaii de la UPC, que els hagin proporcionat les dades làser utilitzades com a referència a l'apartat IV.D.

Guillem Alenyà està finançat pel programa Marie Curie FP5 de la Comissió Europea, contracte HPMT-CT-2001-00156. Elisa Martínez i Carme Torras han rebut finançament del Ministeri Espanyol de Ciència i Tecnologia, projecte DPI2003-05193-C02-01.

## REFERENCES

- [1] E. Martínez and C. Torras, "Qualitative vision for the guidance of legged robots in unstructured environments", *Pattern Recognition*, 34 (2001) 1585-1599.
- [2] A. Blake, M. Isard, D. Reynard, "Learning to track the visual motion of contours", *Artificial Intelligence*, 78 (1995) 101-134.
- [3] T. Cham and R. Cipolla, "Automated B-Spline curve representation incorporating MDL and error-minimizing control point insertion strategies", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21 (1999).
- [4] A. Blake and M. Isard, *Active contours*, Springer, 1998.
- [5] E. Martínez, "Recovery of 3D structure and motion from the deformation of an active contour in a sequence of monocular images", PhD Thesis, 2000.
- [6] G. Alenyà, E. Martínez and C. Torras, "Fusing visual and inertial sensing to recover robot egomotion", *Journal of Robotics Systems*, 21 (2004).
- [7] J.C. Clarke, *Modelling uncertainty: A primer*, Tech. Report, University of Oxford. Dept. Engineering Science, 2161/98 (1998).
- [8] J. Clarke and A. Zisserman, "Detection and tracking of independent motion", *Image and Vision Computing*, 14 (1996) 565-572.
- [9] A. Criminisi, *Accurate visual metrology from single and multiple uncalibrated images*, Springer, 2001.