

- Búsqueda por Comunicación y Grupo Temático
- Búsqueda por Autor
- Comités (Créditos Publicación)
- Patrocinadores

5 · 6 · 7 de Septiembre · HUELVA
XXVIII Jornadas de Automática

Salir



CEA
comité
español de
automática



EVALUACION DE DETECTORES DE PUNTOS DE INTERÉS PARA SLAM VISUAL

Mónica Ballesta, Arturo Gil, Óscar Reinoso, Luis Payá
Departamento de Ingeniería de Sistemas Industriales
Universidad Miguel Hernández
03202 Elche (Alicante), SPAIN
Email: {m.ballesta,arturo.gil,o.reinoso,lpaya}@umh.es

Óscar Martínez Mozos
Department of Computer Science
Universidad de Freiburg
79110 Freiburg, Alemania
Email: omartinez@informatik.uni-freiburg.de

Resumen

En este artículo se presentan distintos detectores de puntos de interés y se analiza su aplicación en SLAM basado en visión. Para ello se evalúan los detectores en función de su repetibilidad ante cambios en punto de vista y de escala, puesto que se trata de requerimientos deseables para marcas visuales. Se han llevado a cabo diferentes experimentos con secuencias de imágenes capturadas con elevada precisión. Estas imágenes contienen tanto objetos planos como escenas en tres dimensiones.

Palabras Clave: Detector de puntos de interés, SLAM visual, marcas visuales, robots móviles.

1 INTRODUCCIÓN

La adquisición de mapas es una tarea fundamental para robots móviles autónomos ya que los mapas son necesarios para otras tareas de nivel superior. En consecuencia, el problema de SLAM ha recibido especial atención en los últimos años. Algunos autores usan sensores láser para construir mapas en dos y tres dimensiones (ver [2,5,8][10,14,24]). Recientemente, existe una tendencia creciente en el uso de cámaras y sensores. Este enfoque es conocido como SLAM visual (vSLAM). Las cámaras proporcionan una mayor cantidad de información y son más baratas que los sensores láser. Además, proveen información 3D cuando se utilizan sistemas estéreo.

Normalmente los enfoques que utilizan visión aplican SLAM basado en características, en el que se tratan características visuales como marcas. El punto clave cuando se trabaja con vSLAM es seleccionar características de las imágenes que sean adecuadas para ser integradas como marcas fiables en el mapa.

Además, cuando se trata de mapas en 3 dimensiones, las marcas deben ser robustas ante cambios de escala y de punto de vista. Se han utilizado diferentes características visuales para la construcción de mapas usando visión monocular o estéreo. Por ejemplo, Lemaire y Lacroix [13] usan segmentos como marcas junto con un enfoque de SLAM con EKF. Fintrop et al. [3] extraen regiones de interés usando el sistema VOCUS. Algunos autores utilizan características SIFT como marcas en el espacio 3D [22]. Little et al. [15] y Gil et al.[7] realizaron además el seguimiento de características SIFT para integrar las más robustas al mapa; y Valls Miro et al.[25] han utilizado SIFT para construir mapas en entornos muy extensos. El detector de Harris ha sido también utilizado como marcas en SLAM monocular (Davidson and Murray [3]) o en control reactivo (Hygounenc et al. [11]). Finalmente, Murillo et al. [20] presentan un método de localización usando características SURF. Entre estas características, los puntos de interés han recibido mayor atención en vSLAM. Estos puntos son normalmente invariantes ante rotación, traslación, escala y parcialmente invariantes ante punto de vista, lo cual los hace adecuados para ser usados como marcas visuales. En la práctica, sin embargo, los puntos detectados no son estables y la asociación entre diferentes vistas es difícil. Se han aplicado algunas soluciones para resolver este problema, como mezclar diferentes métodos en un detector [12] o seguir los puntos durante varias imágenes para mantener la estabilidad [7,22]. No obstante, la pregunta de cual es el detector más adecuado para vSLAM está todavía abierta.

En el contexto de asociación y reconocimiento, algunos autores han evaluado varios detectores de puntos de interés. El trabajo presentado por Mikolajczyk y Schmid [18], utiliza diferentes detectores para extraer regiones afines invariantes, pero sólo se centra en la comparación de diferentes métodos de descripción. En [21], se evalúa un

conjunto de detectores. El criterio utilizado mide la calidad de estas características para tareas como reconocimiento de objetos, asociación de imágenes y reconstrucción 3D. Sin embargo, no tienen en cuenta la repetibilidad en varias imágenes sucesivas en una secuencia.

En este artículo se presenta una evaluación de diferentes detectores de puntos que son típicamente utilizados en vSLAM. Los requerimientos exigidos es que sean robustos ante cambios de escala y punto de vista, ya que el robot debe ser capaz de detectar y asociar nuevas marcas con marcas previamente visualizadas. Bajo estas condiciones, se ha evaluado la repetibilidad de los puntos en imágenes consecutivas capturadas para este propósito, y también la probabilidad de que un punto sea detectado en imágenes posteriores.

2 DETECTORES DE PUNTOS DE INTERÉS

El propósito de este trabajo es comparar diferentes métodos de detección con el fin de elegir el más adecuado para su aplicación a tareas de SLAM. Por este motivo, es deseable detectar puntos estables del entorno que sean persistentes a lo largo de un extenso número de imágenes consecutivas. De este modo, se mejoraría la localización del robot. Por ejemplo, cuando dos robots reconocen la misma escena de un entorno se debería realizar una asociación correcta entre las características correspondientes de esta escena. Además, esto debería ocurrir a pesar de ligeras diferencias en cambios de punto de vista u orientación.

Otro requerimiento también importante es que los puntos de interés deben ser suficientemente distintivos para que puntos capturados por dos robots desde localizaciones ligeramente diferentes sean asociados correctamente.

Según este criterio, se han evaluado algunos métodos de detección existentes y que han sido utilizados previamente por otros investigadores en tareas de SLAM visual. Esta evaluación se ha realizado, concretamente, con cinco detectores diferentes que se resumen a continuación.

2.1 DETECTOR DE ESQUINAS DE HARRIS

El detector de esquinas de Harris [9] es, probablemente, el más comúnmente usado, debido a su elevada invarianza ante escala, rotación, cambios de iluminación y ruido en la imagen. Este detector está basado en la siguiente matriz $C(x, y)$ que se calcula sobre una subventana $p \times p$ para cada punto de interés en la posición (x, y) .

$$C(x, y) = \begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \quad (1)$$

donde I_x, I_y son los gradientes de la imagen en dirección horizontal y vertical respectivamente. Sean λ_1 y λ_2 los valores propios de la matriz $C(x, y)$, la función de autocorrelación R tendrá un pico si ambos valores propios son altos. Esto significa que desplazamientos en cualquier dirección producirán un incremento significativo, indicando que se trata de una esquina.

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2) \quad (2)$$

Algunos autores han obtenido buenos resultados experimentales con $k = 0.04$. Este es el caso de Davidson *et al.* [3], que utiliza un detector de esquinas Harris con una subventana de 15×15 .

2.2 HARRIS LAPLACE

Los puntos de interés detectados por Harris-Laplace son invariantes a rotación y escala [19]. Estos puntos son detectados por una función de Harris seleccionada en el espacio de escalas por el operador Laplaciana. La escala seleccionada determina el tamaño de la región de interés. En [12], Jensfelt *et al.* propone utilizar el detector Harris-Laplace en lugar del detector de Diferencia de Gaussianas usado originalmente por SIFT. La razón por la que hace esta elección es porque Harris-Laplace detecta características localizadas de forma más precisa espacialmente, lo cual es mejor cuando estas características son usadas para reconstrucción y localización y no sólo para reconocimiento.

3.3 SIFT

El detector SIFT (Scale Invariant Feature Transform) es un algoritmo que detecta puntos distintivos de imágenes y calcula el descriptor de los mismos. Este algoritmo fue inicialmente presentado por Lowe y usado en tareas de reconocimiento [16]. Los puntos de interés extraídos son invariantes a escala, rotación y parcialmente invariantes a cambios de punto de vista e iluminación. Las características SIFT son localizadas en los máximos y mínimos de una diferencia de Gaussianas (DoG) aplicada en el espacio de escalas. Las características pueden ser calculadas a partir de una pirámide de imágenes con muestreo entre cada nivel [17]. En este trabajo, nos hemos centrado solamente en el detector.

2.3 SURF

SURF (Speeded Up Robust Features(SURF)) es un detector invariante a rotación y a escala y un

descriptor presentado por Bay *et al.* [1]. Este detector se basa en la matriz Hessiana por su precisión y el bajo coste computacional. Según sus autores, este método supera a otros métodos existentes respecto a la repetibilidad, robustez y distinción de los descriptores. Al igual que con las características SIFT, se utiliza sólo el detector SURF dejando a un lado su descriptor, ya que se considera parte de un problema diferente.

2.4 SUSAN

SUSAN (Smallest Univalue Segment Assimilating Nucleus) está orientado al procesamiento de imágenes de bajo nivel [23]. El principio de SUSAN está implementado usando una aproximación digital de máscaras circulares. Si la luminosidad de cada píxel en una máscara es comparada con la luminosidad del núcleo de esa máscara, entonces un área de la máscara puede ser definida teniendo la misma iluminación que el núcleo.

3 MÉTODOS DE EVALUACIÓN

Con el fin de evaluar los diferentes métodos de detección se han utilizado secuencias de imágenes representando la misma escena con diferentes cambios de escala y punto de vista. En este apartado se detalla cómo se han evaluado las secuencias de imágenes. Primero se explican los métodos de seguimiento de puntos utilizados y, a continuación, se describen las medidas de evaluación empleadas. En este artículo no se contempla la invarianza ante cambios de iluminación.

3.1 SEGUIMIENTO DE PUNTOS

Para cada secuencia de imágenes se extraen primeramente los puntos de interés con los detectores presentados en el apartado anterior.

El seguimiento de los puntos en las secuencias de imágenes 2D se ha realizado a partir de la matriz de homografía, calculada en cada par de imágenes consecutivas como en [4]. Sea Y un punto en el espacio 3D que se proyecta en la posición $y_1 = P_1 \cdot Y$ en la imagen I_1 y en la posición $y_i = P_i \cdot Y$ en la imagen I_i , con las matrices de proyección P_1 y P_i . Si suponemos que el punto Y se detecta en dos imágenes, entonces

$$y_i = H_{1i} \times y_1, \text{ with } H_{1i} = P_i P_1^{-1} \quad (3)$$

La matriz de homografía H_{1i} puede ser calculada seleccionando manualmente cuatro correspondencias de puntos coplanares entre las imágenes 1 e i . Dado

un punto detectado en una imagen, se predice su posición en la imagen consecutiva. Si la posición predicha se sitúa a una distancia menor de 2 píxeles, entonces se considera que este punto es correspondiente. Este método se ha aplicado sólo en secuencias con imágenes 2D, ya que la homografía sólo establece correspondencias entre puntos coplanares.

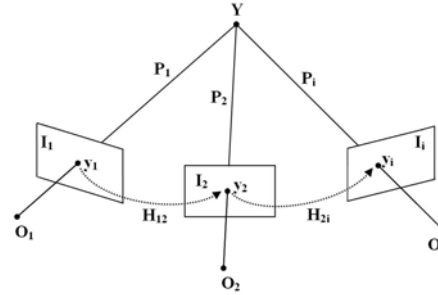


Figura 1: Los puntos y_1 , y_2 e y_i son las proyecciones del punto Y en las imágenes I_1 , I_2 e I_i . Las matrices H_{12} , H_{2i} muestran la relación entre los puntos y_1 e y_2 , e y_2 e y_i respectivamente

En el caso de imágenes con escenas 3D el seguimiento de puntos se ha implementado con la matriz fundamental. Esta matriz 3×3 con rango 2 relaciona puntos correspondientes entre dos imágenes estereó y puede ser calculada con al menos 7 correspondencias iniciales. Dado un punto m_1 de la imagen I_1 , la matriz fundamental F permite calcular la línea epipolar donde debe estar situado el punto correspondiente m_1' en la imagen I_2 . La línea epipolar se calcula como

$$l' = F m_1 \quad (3)$$

En consecuencia, dos puntos correspondientes satisfarán la siguiente ecuación

$$m_i'^T F m_i = 0 \quad (4)$$

En este trabajo, primeramente se ha extraído un conjunto de correspondencias iniciales usando la matriz fundamental con el método de los 7 puntos, seleccionados manualmente. Con ello se obtiene un extenso número de correspondencias que sirven de entrada para calcular otra matriz fundamental en una segunda iteración. En este caso se utiliza el algoritmo de RANSAC como en [26], que calcula la matriz fundamental utilizando como mínimo 8 correspondencias. El hecho de utilizar como entrada un conjunto de numerosas correspondencias permite calcular la matriz de forma más precisa y calcular así las correspondencias definitivas.

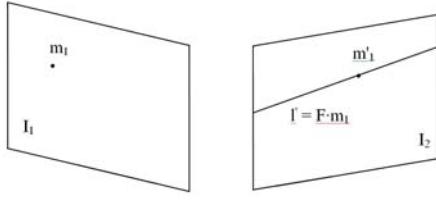


Figura 2: El punto m'_1 es el punto correspondiente de m_1 en la imagen I_2 . Este punto está situado en la línea epipolar l' calculada con la matriz fundamental F

Dado un punto en la imagen I_1 , los posibles candidatos a punto correspondiente se seleccionan estableciendo una ventana de búsqueda de 10×10 píxeles en la imagen consecutiva I_2 , entorno a la posición del punto de la imagen I_1 . Tras haber obtenido la matriz fundamental, el punto correspondiente se ha seleccionado como aquel cuya distancia a la línea epipolar (Ecuación (3)) fuese menor. Esta distancia permite descartar candidatos erróneos y se calcula como sigue

$$d(m'_i, Fm_i) = \frac{|m_i^T Fm_i|}{\sqrt{(Fm_i)_1^2 + (Fm_i)_2^2}} \quad (5)$$

Los métodos utilizados en el seguimiento de los puntos son puramente geométricos, lo cual evita tener que hacer cualquier tipo de descripción de los puntos, y, por tanto, realizar la evaluación únicamente de la etapa de detección.

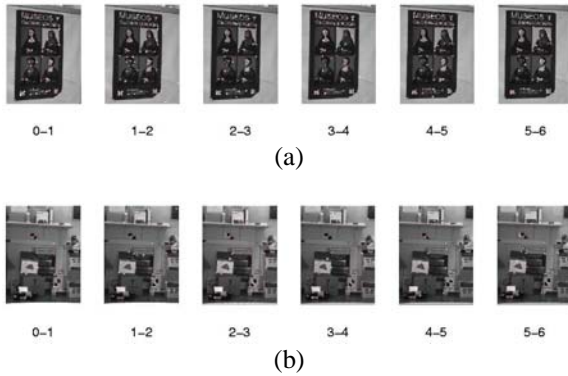


Figura 3: Seguimiento de puntos en una secuencia. (a) Con la matriz de homografía. (b) Con la matriz fundamental.

Una muestra de los métodos de seguimiento utilizados se puede ver en las figuras 3(a) y 3(b) donde aparecen puntos de Harris seguidos en 6 imágenes consecutivas (puntos blancos). Los puntos rojos son aquellos que se siguen en la secuencia. Por tanto, los que aparecen en la última imagen son aquellos que se han seguido a lo largo de la secuencia completa. La figura 3(a) muestra el seguimiento de

imágenes con objetos 2D realizado mediante la matriz de homografía. La figura 3(b) muestra el seguimiento realizado en imágenes con escenas 3D mediante el método de la matriz fundamental explicado anteriormente.

3.2 MEDIDAS DE EVALUACIÓN

El propósito de este trabajo es evaluar los detectores de acuerdo a los requerimientos del SLAM. En este sentido, se ha seguido un criterio de repetibilidad que refleja la estabilidad de los puntos detectados ante cambios en las imágenes, por ejemplo, cambios de escala y de punto de vista.

Se ha definido un ratio de supervivencia S_i en la imagen i de la secuencia como

$$S_i = \frac{np_i}{np_0} \cdot 100 \quad (6)$$

donde np_i y np_0 son el número de puntos detectados en la imagen i y en la primera imagen de la secuencia respectivamente. El detector ideal sería el que detectase los mismos puntos en la primera y última imagen, es decir, $S_i = 100\%$ para cada imagen.

Cuando un robot explora el entorno, se pretende que las marcas visuales extraídas sean estables y puedan ser detectadas en un número p de imágenes consecutivas [7,15]. De este modo, el número de marcas en el mapa se reduce y con ello la complejidad del problema de SLAM. Sin embargo, la elección de este valor p supone un compromiso, puesto que si p es alto, el número de marcas en el mapa será demasiado bajo; y si p es bajo, supondría un gran número de marcas en el mapa con un elevado coste computacional. Por este motivo, se ha analizado durante cuantas imágenes consecutivas se ha de seguir una marca antes de integrarla en el mapa. Se usa la siguiente probabilidad condicionada:

$$P(t_{fa} | t_{fb}) = \frac{t_{fa}}{t_{fb}} \quad (7)$$

donde t_{fi} es el número de puntos seguidos hasta la última imagen f_i . Este valor representa la probabilidad de un punto de interés de que sea seguido hasta la imagen f_a dado que ha sido seguido hasta la imagen f_b , y oscila en un rango entre 0 y 1. Será 0 si todos los puntos seguidos hasta f_b se pierden en f_a , y 1 si ambas imágenes f_a y f_b contienen los mismos puntos seguidos. La expresión [7] proporciona una predicción de la supervivencia

de un punto de interés en las sucesivas instantáneas si el movimiento del robot se mantiene similar.

4 EXPERIMENTOS

La evaluación de los distintos detectores se ha realizado a partir de secuencias de imágenes capturadas para este fin. Concretamente, se han capturado 12 secuencias de 21 imágenes cada una, con cambios de punto de vista, teniendo un incremento de 2.5 grados entre cada imagen consecutiva. Por otro lado, se han capturado 14 secuencias de 12 imágenes cada una, con cambios de escala. En este caso la variación entre imágenes consecutivas es de 0.1 metros. En la figura 4 se muestra ejemplos de estas imágenes.



Figura 4: Secuencia de imágenes tomadas desde diferentes puntos de vista (fila superior). Secuencia con cambios en escala (fila inferior).

Las imágenes han sido capturadas usando un par estéreo STH-MDCS2 de Videre Design. Se ha utilizado una de las imágenes estéreo simultáneamente para cada secuencia. El par estéreo se montó sobre el extremo de un brazo robot para conseguir mayor precisión en la captura de imágenes. Estas imágenes fueron tomadas a varias resoluciones (320x240, 640x480 y 1280x960), para que fuesen lo más representativas posible.

Las figuras 5 y 6 corresponden a secuencias de imágenes de objetos en dos dimensiones y las figuras 7 y 8 imágenes con escenarios 3D. Como se comentó en un apartado anterior, se evalúan los detectores en función de dos parámetros: el ratio de supervivencia (figuras 5(a), 6(a), 7(a) y 8(a)) y la probabilidad condicionada (figuras 5(b), 6(b), 7(b) y 8(b)). En el primero de los casos se analiza la repetibilidad de los detectores calculando la Ecuación (6). Para el cálculo de esta ecuación, se obtiene el total de puntos seguidos hasta la instantánea i considerando todas las secuencias (con cambios de punto de vista o escala) y dividimos este valor entre el número total de números detectados en la primera instantánea. En el caso de la Ecuación (7) el procedimiento es similar, sólo que se divide entre el número total de puntos seguidos hasta la última instantánea.

En cuanto al ratio de supervivencia, se observa que con el detector Harris se obtiene un valor de 30 % aproximadamente, tanto con cambios en el punto de vista (figura 5(a)) como en escala (figura 6(a)). Este resultado es todavía mejor en el caso de las imágenes con escenarios 3D, donde alrededor del 50 % de puntos se siguen hasta el final de la secuencia (figuras 7(a) y 8(a)).

En el caso de los resultados obtenidos con la probabilidad condicionada, se observa que de nuevo Harris es el detector con mejores resultados. La figura 7(b) indica que si un punto es seguido hasta el la imagen 6 de la secuencia, la probabilidad de que se detecte en toda la secuencia será de 0.7.

Harris es el detector que obtiene mejores resultados en todos los casos. Un hecho a destacar es que este detector tiene un comportamiento todavía mejor cuando se trata de imágenes con escenarios en tres dimensiones (figuras 7 y 8) aumentando la diferencia respecto al resto de detectores. En la figura 7(a) se observa que un 55% de los puntos de Harris se siguen durante toda la secuencia, mientras con el resto de detectores se alcanza un 20%, a excepción del detector SUSAN con el que tan solo alrededor de un 10 % de puntos se siguen hasta la última imagen.

Con respecto al resto de detectores, las figuras muestran que el detector de Harris es seguido de SIFT, Harris Laplace y SURF, los cuales tienen un comportamiento similar. En concreto, SURF parece comportarse mejor en el caso de imágenes con escenarios 3D (figuras 7 y 8), así como en imágenes con cambios en escala en general. Por ejemplo, en la figura 6(b) un punto seguido durante 6 imágenes tiene una probabilidad de 0.5 de ser seguido hasta el final de la secuencia. SIFT es el detector que se comporta mejor, después de Harris, en el caso de imágenes con objetos 2D con variaciones en el punto de vista (figuras 5(a) y 5(b)), pero obtiene peores resultados en el resto de casos. Finalmente, SUSAN es el detector con peores resultados en todos los casos.

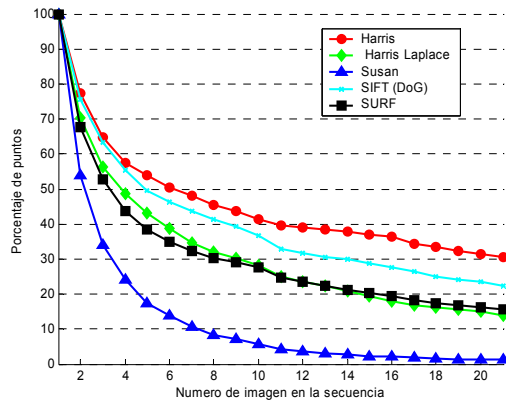
Las tablas 1 y 2 presentan, para cada método de detección, el número total de puntos de interés detectados en la primera imagen y los que se han seguido hasta la última imagen de la secuencia.

5 CONCLUSION

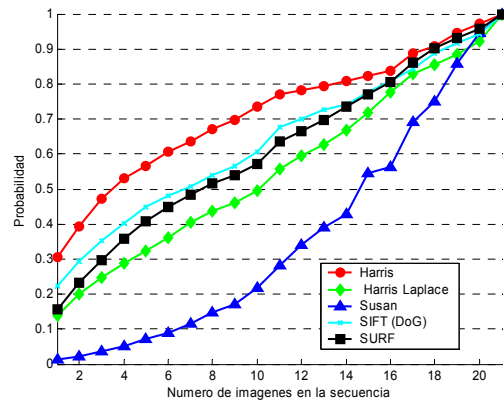
La motivación de este trabajo ha sido encontrar el detector más adecuado para tareas de SLAM. Para ello se ha hecho una evaluación de algunos métodos de detección existentes en base a los requerimientos deseables de una marca visual: repetibilidad y

precisión. Los resultados obtenidos muestran que Harris es el detector que mejor se comporta según los requerimientos evaluados. Por otro lado, cuando se detecta una marca visual, suele ser necesaria una descripción de la misma que permita asociar las observaciones realizadas por el robot con las marcas

del mapa. Sin embargo, este trabajo se ha centrado en encontrar un buen detector para SLAM visual. Como trabajo futuro, se pretende evaluar diferentes métodos de descripción en combinación con el detector Harris.

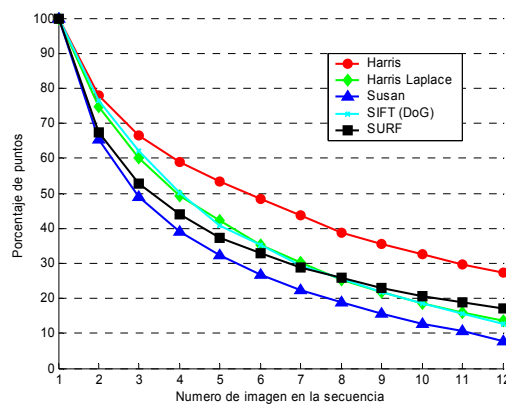


(a)

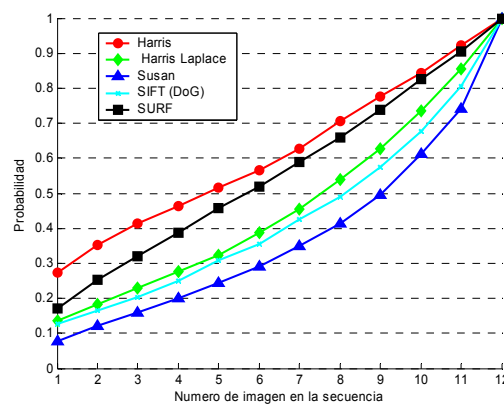


(b)

Figura 5: Secuencias 2D con cambios en orientación. (a) Ratio de supervivencia. (b) Probabilidad condicionada

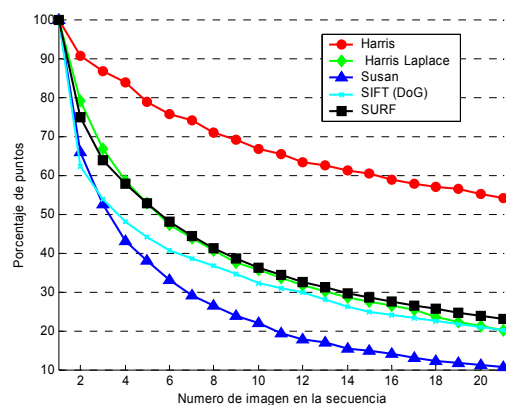


(a)

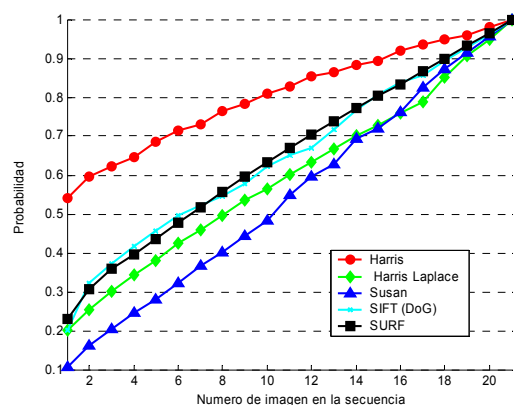


(b)

Figura 6: Secuencias 2D con cambios de escala. (a) Ratio de supervivencia. (b) Probabilidad condicionada.



(a)



(b)

Figura 7: Secuencias 3D con cambios en orientación. (a) Ratio de supervivencia. (b) Probabilidad condicionada.

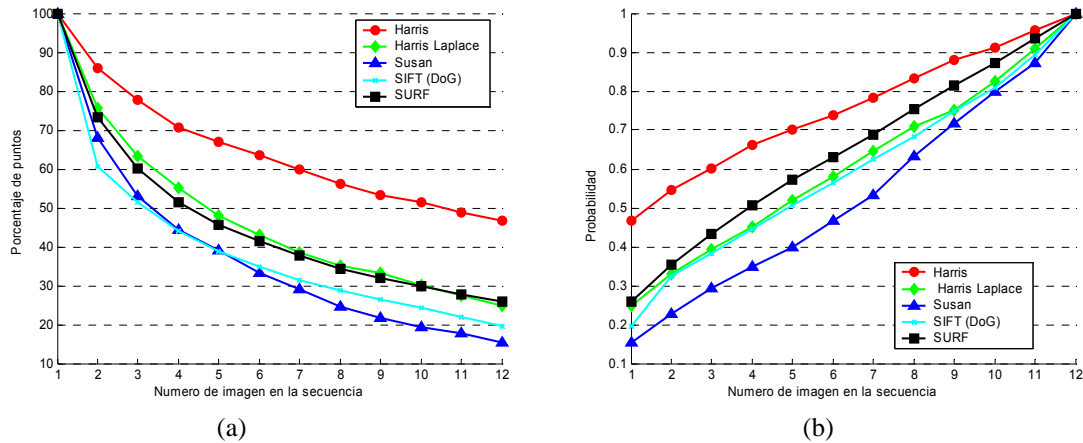


Figura 8: Secuencias 3D con cambios de escala. (a) Ratio de supervivencia. (b) Probabilidad condicionada.

Tabla 1: Número de puntos detectados en la primera y última imagen de cada secuencia de imágenes 2D.

Cambios en el punto de vista	Harris	Harris Laplace	SUSAN	SIFT	SURF
Nº puntos detectados en la primera imagen	1051	1216	1472	1830	5438
Nº puntos seguidos hasta la última imagen	321	170	18	407	853
Cambios en escala	Harris	Harris Laplace	SUSAN	SIFT	SURF
Nº puntos detectados en la primera imagen	4501	4431	5004	6330	20335
Nº puntos seguidos hasta la última imagen	1237	606	392	796	3458

Tabla 2: Número de puntos detectados en la primera y última imagen de cada secuencia de imágenes 3D.

Cambios en el punto de vista	Harris	Harris Laplace	SUSAN	SIFT	SURF
Nº puntos detectados en la primera imagen	1013	1372	1495	1774	4934
Nº puntos seguidos hasta la última imagen	548	276	159	356	1132
Cambios en escala	Harris	Harris Laplace	SUSAN	SIFT	SURF
Nº puntos detectados en la primera imagen	963	1254	1417	1668	4661
Nº puntos seguidos hasta la última imagen	451	313	220	329	1212

Agradecimientos

Esta investigación está patrocinada por el Ministerio de Educación y Ciencia (referencia del proyecto DPI2004-07433-C02-01. Título: Herramientas de Teleoperación colaborativa. Aplicación al control cooperativo de Robots), y proyecto PCT-G54016977-2005 (Fundación quórum: Parque Científico y Empresarial de la Universidad Miguel Hernández. Título: Robots cooperativos para la vigilancia e inspección de edificios e instalaciones industriales).

Referencias

[1] Bay, H., Tuytelaars, T., Van Gool, L., (2006) Object recognition from local scale-invariant features. *European Conference on Computer Vision*.

[2] Biber, P., Andreasson, H., Duckett, T., Schilling, A., (2004). 3D modelling of indoor

environments by a mobile robot with laser scanner and panoramic camera. *IEEE/RSJ Int. Conf. On Intelligent Robots & Systems*.

[3] Davidson, A.J., Murray, D.W., (2002). Simultaneous localisation and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[4] Dork, Gy., Schmid, C., (2003). Selection of scale invariant neighbourhoods for object class recognition. *Int. Conf. On Computer Vision*.

[5] Eustice, R., Singh, H., Leonard, J.J., (2005). Exactly sparse delayed-state filters. *IEEE Int. Conf. On Robotics & Automation*.

[6] Frintrop, S., Jensfelt, P., Christensen, H.I., (2006). Attentional landmark selection for visual slam. *IEEE/RSJ Int. Conf. On Intelligent Robots & Systems*.

- [7] Gil, A., Reinoso, O., Burgard, W., Stachniss, C., Martínez Mozos, O., (2006). Improving data association in rao-blackwellized visual SLAM. *IEEE/RSJ Int. Conf. On Intelligent Robots & Systems*.
- [8] Grisetti, G., Stachniss, C., Burgard, W., (2007). Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Transactions on Robotics*, 23(1).
- [9] Harris, C.G., Mohr, R., Bauckhage, C., (1998). A combined corner and edge detector. *Alvey Vision Conference*.
- [10] Hähnel, D., Burgard, W., Fox, D., Thrun, S.,(2003). An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements.
- [11] Hygounenc, E., Jung, I., Souères, P., Lacroix, S., (2004). The autonomous blimp project of laas-cnrs: Achievements in flight control and terrain mapping. *International Journal of Robotics Research*, 23(4-5).
- [12] Jensfelt, P., Kragic, D., Folkesson, J., Björkman, M., (2006). A framework for vision based bearing only 3D SLAM. *IEEE Int. Conf. On Robotics & Automation*.
- [13] Lemaire, T., Lacroix, S., (2007). Monocular vision based SLAM using line segments. *IEEE Int. Conf. On Robotics & Automation*.
- [14] Leonard, J.J., Durrant-White, H.F., (1991). Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(4).
- [15] Little, J., Se, S., Lowe, D.G., (2002). Global localization using distinctive visual features. *IEEE/RSJ Int. Conf. On Intelligent Robots & Systems*.
- [16] Lowe, D.G., (1999). Object recognition from local scale-invariant features. *Int. Conf. on Computer Vision*.
- [17] Lowe, D.G., (2004). Distinctive image features from scale invariant keypoints. *Int. Journal of computer Vision*, 2(60).
- [18] Mikolajczyk, K., Schmid, C., (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10).
- [19] Mikolajczyk, K., Schmid, C., (2001). Indexing based on scale invariant interest points. *Int. Conf. on Computer Vision*.
- [20] Murillo, A.C., Guerrero, J.J., Sagués, C., (2007). Surf features for efficient robot localization with omnidirectional images. *IEEE Int. Conf. On Robotics & Automation*.
- [21] Schmid, C., Mohr, R., Bauckhage, C., (2000). Evaluation of interest point detectors. *Int. Journal of computer Vision*, 37(2).
- [22] Se, S., Lowe, D.G., Little, J., Se, S., (2001). Vision-based mobile robot localization and mapping using scale-invariant features. *IEEE Int. Conf. On Robotics & Automation*.
- [23] Smith, S.M., (1992). A new class of corner finder. *British Machine Vision Conference*.
- [24] Triebel, R., Burgard, W., (2005). Improving simultaneous mapping and localization in 3d using global constraints. *National Conference on Artificial Intelligence*.
- [25] Valls Miró, J., Zhou, W., Dissanayake, G., (2006). Towards vision based navigation in large indoor environments.
- [26] Zhang, Z., Deriche R., Faugeras, O., Luong, Q., (1995). A robust technique for matching two uncalibrated images through the recovery of the un known epipolar geometry. *Artificial Intelligence*, vol. 78, pp. 87-119.