



Computación y Sistemas

Computación y Sistemas
Instituto Politécnico Nacional
revista@cic.ipn.mx
ISSN (Versión impresa): 1405-5546
MÉXICO

1998

Rafael Murrieta Cid / Maurice Briot / B. Marcel / Héctor González Baños
ASPECTOS DINÁMICOS DE LA VISIÓN: SEGUIMIENTO DE OBJETOS NO
RÍGIDOS Y ESTIMACIÓN DE LA ROTACIÓN DE UNA CÁMARA

Computación y Sistemas, abril-junio, año/vol. 1, número 004

Instituto Politécnico Nacional

Distrito Federal, México

pp. 201-212

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Universidad Autónoma del Estado de México


LA MEMORIA CIENTÍFICA EN LÍNEA
<http://redalyc.uaemex.mx>

Aspectos dinámicos de la visión: Seguimiento de objetos no rígidos y estimación de la rotación de una cámara

R. Murrieta-Cid M. Briot
LAAS-CNRS
7 Av. du Colonel Roche
31077 Toulouse
Cédex 4
{murrieta, briot}@laas.fr

B. Marcel
INP ENSEEIHT GTTSI
2, rue Charles Camichel
31071 Toulouse
Cédex 7
marcel@len7.enseeiht.fr

H. González-Baños
Computer Science Department
Stanford University
Stanford, CA 94305
hhg@robotics.stanford.edu

Artículo recibido en octubre de 1997; aceptado el 20 de marzo de 1998.

Resumen

Dentro del marco de la percepción aplicada a la robótica móvil autónoma, dos tareas son abordadas en este artículo: El seguimiento de un objeto móvil no rígido y la estimación del desplazamiento entre dos imágenes basado en el cambio de la escena percibida. Las principales contribuciones de este trabajo son de dos tipos: En el caso de la estimación del movimiento de la cámara, presentamos un método original, capaz de calcular la translación y rotación de los elementos en la imagen; gracias al cual es posible estabilizar el captor. Por otra parte, para el seguimiento de blanco se desarrollaron estrategias que reducen el tiempo de ejecución del algoritmo, lo que permite efectuar el seguimiento de diversos elementos.

Palabras Clave: *Visión, Robótica, Seguimiento Visual.*

1 Introducción

Nuestro estudio está situado en el contexto de la percepción aplicada a la robótica móvil autónoma.

Particularmente nuestro interés se focaliza en la percepción por cámara video de entornos o elementos no estructurados, tales como ambientes naturales y objetos no rígidos o de formas diversas.

En robótica móvil es particularmente importante considerar que existirá un movimiento relativo entre la escena y los captores embarcados en el robot.

Dentro de dicho contexto, en este artículo presentaremos un método que permitirá realizar el seguimiento visual de objetos no rígidos y de formas diversas, entidades a las que llamaremos blancos. Este método es capaz de seguir a un objeto móvil a pesar de que el **captor** esté también **en movimiento**. La técnica empleada está basada en el cálculo de la **distancia de Hausdorff**, la cual es una medida de similaridad entre dos patrones dados [Serra, 1982].

El objetivo es proporcionar a un robot móvil índices visuales que indiquen la ubicación de un blanco a medida que este se desplaza. Gracias a esta información un robot móvil podrá seguir físicamente un blanco. Es interesante mencionar que utilizando este tipo de métodos será posible que un robot autónomo siga a otro robot gracias a informaciones visuales (para una aplicación específica ver [Becker *et al.*, 1995]).

Por otra parte, presentamos un método que permite **estimar la rotación de una cámara** estudiando las transformaciones que sufre una imagen a través del tiempo. **El cálculo del desplazamiento (rotación y translación) de la imagen**, es obtenido utilizando propiedades de la **transformación de Fourier**.

Uno de los intereses de esta tarea, radica en permitir estabilizar una cámara montada sobre un robot que se

desplaza en terreno irregular. La estabilización consiste en compensar el movimiento errático de la cámara realizando una rotación correctiva en dicha cámara de la misma magnitud y en sentido inverso al desplazamiento calculado.

El método propuesto es global y no necesita identificar un blanco específico (patrón), por lo que está especialmente bien adaptado a la aplicación, dado que a causa del desplazamiento del robot la escena percibida cambiará.

La organización de este artículo será la siguiente: Inicialmente presentaremos el método de seguimiento de blanco, las estrategias de reducción de tiempo de ejecución propuestas así como los resultados experimentales obtenidos. A continuación se describirá el método de estimación de la rotación de una cámara, las adaptaciones efectuadas a fin de utilizar dicha técnica en robótica móvil y los resultados obtenidos. Como conclusión discutiremos los puntos de convergencia de dichos métodos así como sus principales limitaciones en función de una aplicación específica. Finalmente propondremos las futuras perspectivas de nuestro trabajo.

2 Seguimiento de blanco

En los años recientes, el problema del seguimiento de blanco ha recibido mucha atención dentro de la comunidad de la visión artificial. Varias técnicas han sido reportadas en la literatura y una gran cantidad de características han sido propuestas para realizar el seguimiento.

Algunos métodos tratan de seguir objetos 3D, usando múltiples vistas del objeto [Gavrila and Davis, 1996], otros utilizan contornos como características a seguir [Blake, 1993], [Garibotto, 1990]. Entre ellos algunos utilizan la técnica de contornos activos con la intención de manejar las posibles deformaciones del objeto [Delagnes *et al.*, 1994]. Es posible también utilizar el color para facilitar el seguimiento [Du, 1995], o bien usar una combinación de regiones segmentadas por color y contornos como características a seguir [Wren, 1996]. Finalmente algunas otras técnicas detectan el movimiento gracias a variaciones de alguna característica global de la imagen, tal como la textura o invariantes estadísticos [Jiansho and Tomasi, 1994].

En el presente trabajo se presenta un método basado en la comparación entre un modelo y una imagen, el modelo y la imagen son elementos binarios extraídos de una secuencia de imágenes de niveles de gris, usando un extractor de líneas de contraste (bordes) similar al reportado en [Canny, 1986]. Los datos de entrada son manejados como una configuración de elementos binarios donde los pixels¹ característicos (puntos diferentes de ce-

ro pertenecientes a las líneas de contraste en la imagen y el modelo) **no se encuentran concatenados**.

La distancia parcial de Hausdorff es usada para medir el grado de semejanza entre la imagen y el modelo (dicho modelo podrá ser un objeto no rígido).

El método está basado en la suposición de que el movimiento de un objeto no rígido en el espacio tridimensional, puede ser caracterizado usando una representación bidimensional.

El movimiento del blanco en su representación bidimensional (imagen) será descompuesto en dos partes:

- Un movimiento bidimensional correspondiente al cambio de posición del blanco en el espacio imagen.
- Un cambio de forma en dos dimensiones correspondiente al nuevo aspecto del blanco.

Con la intención de reducir el número de operaciones y por lo tanto el tiempo de cálculo, la búsqueda del modelo se efectúa solo en una región contigua a la posición del blanco, en la imagen previa de la secuencia. Esto asume que la velocidad de desplazamiento del blanco estará limitada, de tal forma que la nueva posición del blanco se encuentre dentro de la región de búsqueda.

2.1 Comparación de la imagen con el modelo

Dados dos conjuntos de puntos P y Q , la distancia de Hausdorff está definida como :

$$H(P, Q) = \max(h(P, Q), h(Q, P))$$

donde

$$h(P, Q) = \max_{p \in P} \min_{q \in Q} \|p - q\| \quad (1)$$

y $\| \cdot \|$ es una norma para medir la distancia entre dos puntos p y q . La función $h(P, Q)$ es una medida del grado en el cual cada punto de P está cerca de algún punto en Q (distancia del conjunto P al conjunto Q). Un pequeño valor de $h(P, Q)$ implica que cada punto en P está cerca a un punto en Q . La distancia de Hausdorff está definida como el máximo entre $h(P, Q)$ y $h(Q, P)$. Así, dicha distancia mide el grado en el cual cada punto de P está cerca de uno en Q y vice versa.

Calculando la distancia de Hausdorff de esta manera, se obtiene el punto más "alejado" (que no encuentra correspondiente) entre las dos formas comparadas, por consecuencia esta medida es muy sensible a la presencia de ruido. Dado lo anterior es muy apropiado usar una medida de orden, en remplazo de la operación de maximización. Esta función de medida (distancia parcial) está definida como [Huttenlocher, 1993]:

$$h_k = K_{p \in P}^{th} \min_{q \in Q} \|p - q\| \quad (2)$$

Donde $K_{p \in P}^{th} f(p)$ denota el K -ésimo valor ordenado de $f(p)$ sobre el conjunto P . Así, si nosotros consideramos

¹pixel: contracción proveniente de la de la expresión inglesa picture element el cual designa un punto elemental de una imagen

que los puntos en P se encuentran ordenados secuencialmente por sus valores $f(p_1) \leq \dots \leq f(p_n)$, el K -ésimo elemento en la secuencia $f(p_K)$; será el K -ésimo valor ordenado de dicha secuencia. Por ejemplo el n -ésimo valor ordenado será el más grande de la secuencia.

Una interesante propiedad de la distancia de Hausdorff y de la “distancia parcial” es la **asimetría** inherente a su cálculo. El hecho de que cada punto de P (o subconjunto de P) esté cerca de algún punto en Q no significa que cada punto de Q (o subconjunto de Q) esté cerca de algún punto en P . En otras palabras $h_k(P, Q)$ y $h_k(Q, P)$ pueden tener valores muy diferentes. De hecho cada una da diferente información.

El término $h_k(P, Q)$ es la distancia unidireccional parcial del modelo a la imagen y $h_k(Q, P)$ es la distancia unidireccional parcial de la imagen al modelo. Donde $P = M_t$ es el modelo y $Q = I_t$ es la imagen o región de la imagen en un tiempo t de una secuencia. El máximo de esos dos valores define la distancia parcial de Hausdorff.

2.2 Localización del modelo

La primera tarea a realizar es determinar la posición del modelo M_t en la siguiente imagen de la secuencia. La búsqueda del modelo en la imagen (o región de la imagen) es efectuada en alguna dirección seleccionada (por ejemplo de derecha a izquierda y de arriba hacia abajo).

El mínimo valor de $h_{k1}(M_t, I_{t+1})$ identifica la “mejor” posición de M_t en I_{t+1} bajo la acción de algún grupo de translaciones G . Es también posible identificar el conjunto de translaciones de M_t tales que $h_{k1}(M_t, I_{t+1})$ sea más pequeña que algún valor τ [Huttenlocher et al., 1993]. En este caso existirán múltiples translaciones que tengan esencialmente la misma calidad, lo cual obligará a seleccionar una del conjunto de translaciones.

En lugar de determinar la translación que tenga asociada la mínima distancia h_{k1} o el conjunto de translaciones tales que su h_{k1} sea menor a τ , es posible encontrar la **primera** translación cuya h_{k1} asociada sea menor a τ para una **dirección de búsqueda dada**.

A pesar de que la primera translación cuya h_{k1} asociada es menor que τ , no será necesariamente la mejor, si τ es “pequeño” la translación deberá indicarnos aproximadamente cual fue el desplazamiento del modelo y por consecuencia cual es su posición en la siguiente imagen de la secuencia.

Para calcular el término $\min_{q \in Q} \|p - q\|$ de la ecuación (2), el cual establece la distancia de cualquier punto p al punto más cercano en el conjunto de puntos Q , es necesario calcular la distancia de cualquier posición (i, j) en la imagen Q al punto más cercano diferente de cero en dicha imagen. Como el modelo será “colocado” sobre la imagen será posible calcular la mínima distancia de los puntos del modelo a los puntos en la imagen.

Por analogía con los diagramas de Voronoi esta distancia ha sido llamada distancia de Voronoi.

2.3 La distancia de Voronoi

Existen varios métodos para calcular la distancia de Voronoi o una aproximación de ella [Borgefors, 1986], [Danielsson, 1980].

Nosotros utilizamos un algoritmo similar a [Paglieroni, 1992], el cual produce una precisión igual a la precisión máxima de la computadora utilizada.

Este algoritmo primero procesa independientemente cada renglón de la imagen. Para cada renglón de $Q(i, j)$ se calcula la distancia al más cercano pixel en el renglón. Es decir, para cada posición (i, j) se determina Δi tal que $Q(i + \Delta i, j) \neq 0$ o $Q(i - \Delta i, j) \neq 0$ donde Δi es el mínimo valor no negativo para el cual esto es cierto. Posteriormente se calcula del mismo modo la distancia al más cercano pixel en una columna.

Con la intención de reducir el número de operaciones, la distancia al punto (pixel 1) más cercano para una posición dada (i, j) es calculada comparando las distancias de los posibles puntos candidatos, únicamente en una zona de la imagen determinada por el mínimo valor entre la distancia al pixel más cercano en la columna y el pixel más cercano al renglón. Una tabla de equivalencias que depende de la norma de la distancia es usada para calcular el valor de dichas distancias.

2.4 Actualización del modelo

Habiendo encontrado la primera translación $g \in G$ tal que $h_{k1} < \tau$, tendremos ahora que construir un nuevo modelo M_{t+1} , determinando cuales pixels de la imagen I_{t+1} son parte de dicho nuevo modelo.

El modelo es actualizado usando la distancia $h_{k2}(I_{t+1}, g(M_t))$, como criterio de selección del subconjunto de puntos imagen I_{t+1} que pertenecen a M_{t+1} . El nuevo modelo está definido por:

$$M_{t+1} = \{q \in I_{t+1} \mid h_{k2}(I_{t+1}, g(M_t)) < \delta\} \quad (3)$$

Donde $g(M_t)$ es el modelo en el tiempo t , bajo la acción de la translación g y donde δ controla el grado en el cual el método es capaz de seguir objetos que cambien de forma.

Para permitir que el modelo cambie de tamaño, éste es incrementado si hay un número significativo de puntos diferentes de cero cerca de las fronteras del modelo y reducido en caso contrario. La posición del modelo es corregida de acuerdo con la posición en donde las fronteras de dicho modelo fueron definidas.

Naturalmente el modelo inicial debe ser obtenido de manera diferente. El usuario especifica un rectángulo en la imagen que contiene el blanco a seguir (ver figura 1). Con este modelo se inicia el seguimiento, encontrando progresivamente la nueva posición del blanco y actualizando el modelo.

El seguimiento será exitoso si:

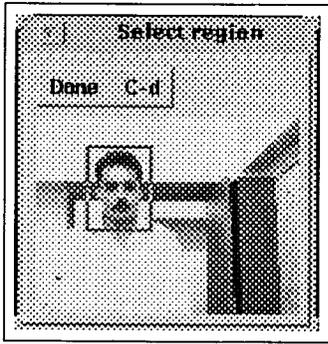


Figura 1: modelo inicial

$$k1 > fM \mid h_{k1}(M_t, I_{t+1}) < \tau \quad (4)$$

y

$$k2 > fI \mid h_{k2}(I_{t+1}, g(M_t)) < \delta, \quad (5)$$

Siendo fM una fracción del número total de puntos del modelo M_t y fI una fracción del número de puntos de la imagen I_{t+1} donde $g(M_t)$ está superpuesto.

Las figuras 2 y 3, muestran dos modelos después de haber actualizado su forma.



Figura 2: modelo a

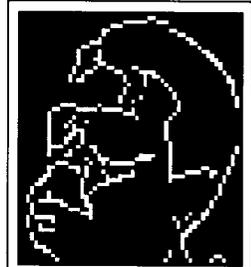


Figura 3: modelo b

2.5 Nuestra contribución al método de seguimiento de blanco

El método de seguimiento de blanco presentado en este artículo está basado principalmente en el reportado en [Huttenlocher, 1993] y [Huttenlocher *et al.*, 1993]. En esta sección mencionamos algunas de las principales extensiones que hemos efectuado sobre el método general [Murrieta-Cid, 1997].

- Para obtener la nueva posición del blanco; solo una región de la imagen es examinada en lugar de analizar la imagen entera.

La idea se basa en que una exploración local de la imagen aportará un buen desempeño del método, si la ejecución del código es suficientemente rápida, ya que el blanco deberá encontrarse en una vecindad próxima a su posición anterior.

Nosotros estamos sacrificando la capacidad de determinar la posición del blanco en todo el espacio imagen a cambio de obtener la nueva posición del blanco con una frecuencia más elevada.

De esta forma estamos incrementando la robustez del método para manejar deformaciones del blanco, dado que es menos probable que el aspecto del blanco cambie significativamente en un menor intervalo de tiempo. Adicionalmente esta estrategia permite reportar la posición del blanco a un sistema externo con mayor frecuencia. En el contexto de la robótica móvil dicho sistema externo será el algoritmo de navegación del robot.

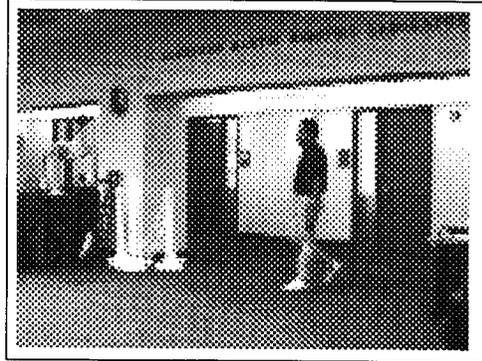


Figura 4.

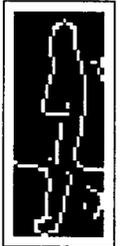


Figura 5

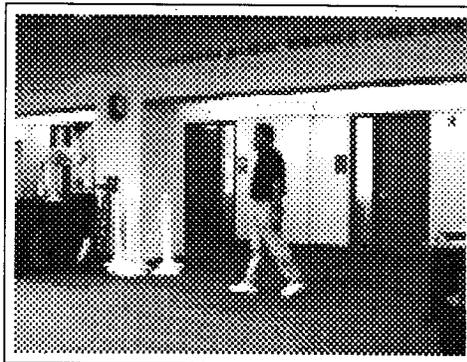


Figura 6



Figura 7

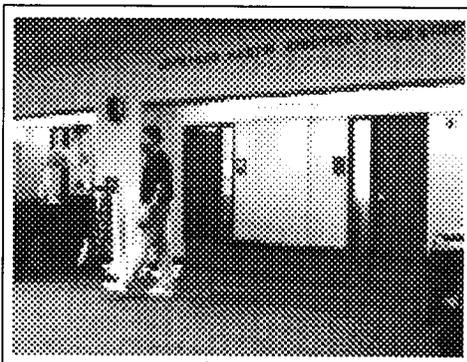


Figura 8



Figura 9

- En lugar de calcular el conjunto de translaciones M_t tales que $h_{k1}(M_t, I_{t+1}) < \tau$, nosotros encontramos la primera $h_{k1}(M_t, I_{t+1}) < \tau$ para una dirección de búsqueda dada. Esta estrategia reduce significativamente el tiempo de cálculo.

Si bien esta técnica puede causar una acumulación de error en la dirección contraria a la dirección de búsqueda, la cual podrá ocasionar una inestabilidad

en el sistema. Es posible evitar dicha inestabilidad redefiniendo las fronteras del modelo al momento de calcular el nuevo modelo. También es posible reducir el error acumulado disminuyendo la tolerancia de la función de comparación entre el modelo y la imagen.

- Para calcular la distancia de Voronoi, usamos un algoritmo, el cual genera una precisión igual a la precisión máxima de la computadora utilizada.

El algoritmo es capaz de calcular la distancia de Voronoi para diferentes métricas, por ejemplo: distancia Euclidiana, "city block" o "chess-board".

2.6 Resultados



Figura 10

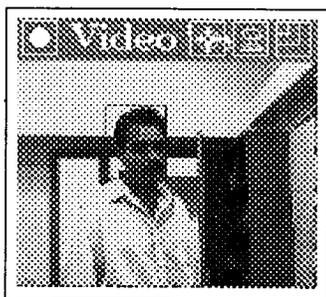


Figura 11

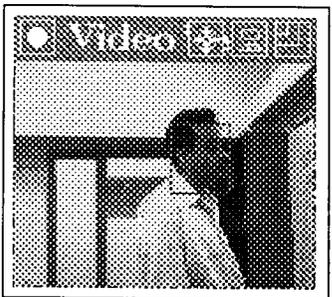


Figura 12

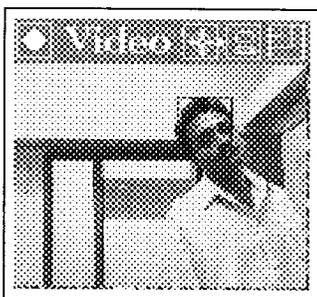


Figura 13

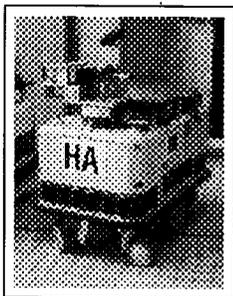


Figura 22: *Hilare*

El método presentado fue implementado en C y C++, y el código es capaz de procesar una imagen en aproximadamente 0.3 segundos. El proceso incluye extracción de líneas de contraste, localización del blanco y actualización del modelo para una imagen de 160×120 pixels, sobre una SPARC 20.

Las figuras 5, 7 y 9 muestran la evolución del modelo para una secuencia de imágenes de niveles de gris. Estas imágenes corresponden al laboratorio de Robótica de la

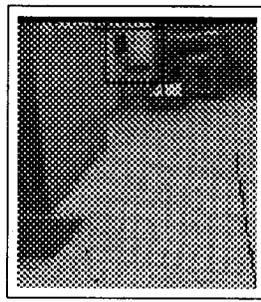


Fig 14: modelo inicial

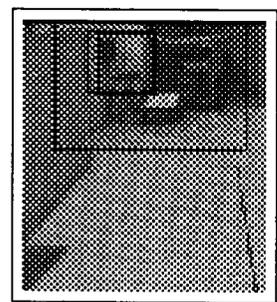


Figura 15

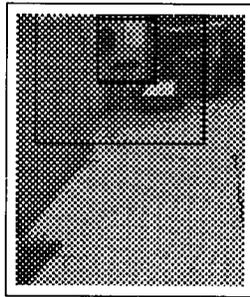


Figura 16

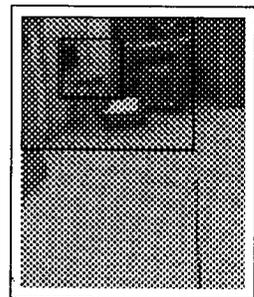


Figura 17

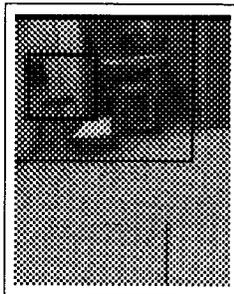


Figura 18

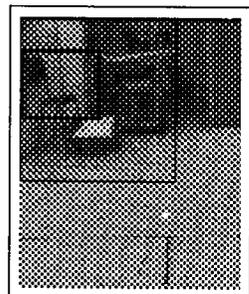


Figura 19

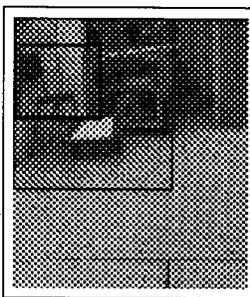


Figura 20

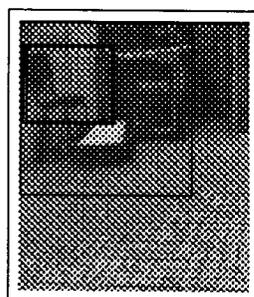


Figura 21

Universidad de Stanford, sitio donde se realizaron estas experimentaciones (figuras 4, 6 y 8).

Las figuras 10, 11, 12 y 13 muestran diferentes escenas de un experimentación ejecutada en tiempo real. Las imágenes corresponden al seguimiento de la cabeza de una persona caminando, la cual está señalada en las figuras con un rectángulo.

A continuación presentaremos otro ejemplo de seguimiento visual, donde el blanco permanece estático y su desplazamiento en la imagen se debe al movimiento del captor.

La figura 14 muestra la región de la imagen tomada como modelo inicial, en este caso una silla. Las figuras 15, 16, 17, 18, 19, 20 y 21 muestran el seguimiento del blanco, las imágenes fueron tomadas por el robot móvil HILARE-2, uno de los robots de interior del grupo de

Robótica e Inteligencia artificial (RIA) del LAAS-CNRS (ver figura 22). El blanco es señalado en las figuras con un rectángulo, un rectángulo de mayor dimensión indica la zona de búsqueda.

En todos los ejemplos mostrados en esta comunicación, es el usuario quien selecciona manualmente el objeto a seguir; con el fin de incrementar la autonomía de los robots, una selección automática del blanco es necesaria. De esta forma será posible que el robot seleccione de manera autónoma un blanco en función de su naturaleza (clase) y de su forma. Una técnica capaz de realizar la selección automática del blanco es presentada en [Murrieta-Cid *et al.*, 1998].

En la segunda parte del artículo presentaremos un método para calcular la translación y rotación de los elementos en una imagen.

3 Cálculo de la rotación de una cámara

Las investigaciones en robots autónomos permiten actualmente la utilización de dichos robots en todo tipo de entornos (ambiente exterior natural, exploración planetaria, exploración submarina, etc).

En el caso de robots de exterior (ver robot ADAM² figura 23) su desplazamiento se llevará a cabo generalmente sobre suelo irregular, por lo que los captosres colocados sobre este tipo de robots, sufrirán movimientos erráticos que perturbarán la escena percibida. Esto podrá ocasionar que la escena no sea útil para las tareas asignadas al robot. Por lo tanto para conservar en el

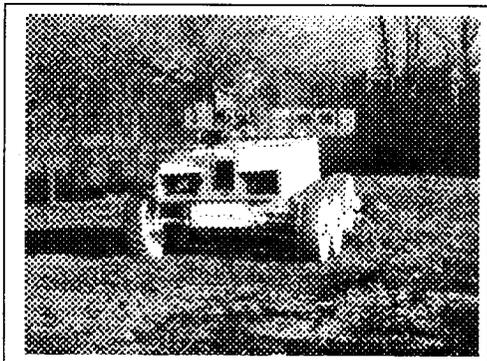


Figura 23: Robot ADAM. MMS, Framatome y LAAS.

campo visual una escena explotable será necesario estabilizar la cámara montada sobre el robot. Estudiando las transformaciones que sufre la imagen en el transcurso del tiempo y utilizando una cámara calibrada, será posible reconstituir los movimientos de rotación de la cámara en el espacio.

El método presentado aquí supone que los movimientos de la cámara se traducen en movimientos planos de la imagen del tipo desplazamientos, es decir, composición de translación y rotación.

Para determinar el desplazamiento que liga dos imágenes sucesivas es posible realizar una operación de alineamiento (“registration” en inglés).

Diversos trabajos han sido presentados para efectuar este tipo de tratamientos [Brown, 1992], [DeCastro and Morandi, 1986].

Nuestro trabajo se basa en la utilización de propiedades de la transformación de Fourier a fin de determinar la translación y rotación de los elementos en la imagen y el método está constituido de dos etapas principales:

- cálculo de la rotación de una imagen abstrayéndose de la translación.
- conocimiento de la rotación para calcular la translación.

La primera etapa del método presentado aquí se inspira en [Messner and Szu, 1985], así mismo nosotros hemos presentado un primer estudio sobre este nuevo método en [Marcel *et al.*, 1997].

3.1 Cálculo de la translación pura: Correlación de fase

Nuestra técnica no trata de medir directamente el desplazamiento de la imagen, si no de encontrar dicho desplazamiento en la transformada de Fourier de la imagen.

La transformación de Fourier conserva las rotaciones y transforma una translación en un desfase frecuencial.

Este desfase deberá ser el de un vector de onda monocromática, donde el vector de onda representa el vector de translación.

La representación matemática de esta propiedad será:

Denotando : (x, y) ó $\begin{pmatrix} x \\ y \end{pmatrix}$ como un punto en el espacio de la señal y (f_x, f_y) ó f_{xy} un punto en el espacio de las frecuencias. Así mismo ($'$) indicará la operación de transposición, por lo que M' será la matriz transpuesta de M .

Sea \mathcal{F} el operador de Fourier y sean $g_0(x, y)$ una imagen y $G_0(f_x, f_y)$ su transformada. Sea t el vector de translación de coordenadas u y v , y t el operador correspondiente. Finalmente siendo g_1 la imagen trasladada. Es posible formular la propiedad anterior como sigue:

$G_0(f_x, f_y)$ transformada de Fourier en dos dimensiones de $g_0(x, y)$. Por definición:

$$\mathcal{F}[g_0(x, y)] = G_0(f_x, f_y) \quad (6)$$

Vector de translación $t = \begin{pmatrix} u \\ v \end{pmatrix}$ y definición de g_1 :

$$g_1(x, y) = tg_0(x, y) = g_0(x - u, y - v) \quad (7)$$

transformada G_1 de la imagen trasladada g_1 :

$$\begin{aligned} G_1(f_x, f_y) &= \mathcal{F}[g_0(x - u, y - v)] \\ &= e^{-2i\pi(uf_x + vf_y)} G_0(f_x, f_y). \end{aligned} \quad (8)$$

²Advanced Demonstrator for Autonomy and Mobility

Sea $\Psi(f_x, f_y)$ la relación de los espectros (llamado *espectro cruzado*, *cross-power spectrum* en inglés) de g_0 et g_1 : $\Psi = \frac{G_1}{G_0}$. El cálculo de Ψ da la ecuación siguiente:

$$\Psi(f_x, f_y) = \frac{G_1(f_x, f_y)}{G_0(f_x, f_y)} = e^{-2i\pi(u f_x + v f_y)}. \quad (9)$$

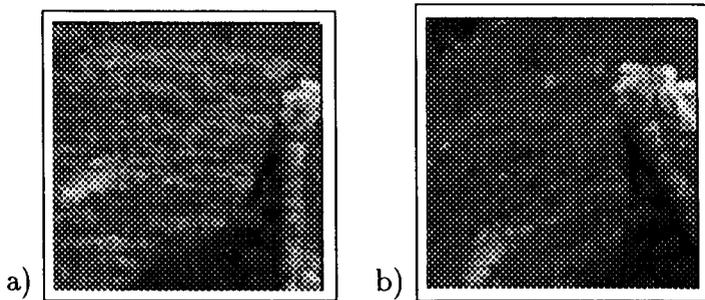
El cálculo del espectro cruzado, obtiene una onda mono-cromática, dicha onda está compuesta de un solo vector que corresponde al vector de translación $-t = (-u, -v)$.

Una vez Ψ obtenido, es posible determinar su vector de onda calculando su transformada de Fourier, que será un pico de Dirac en el punto $-t = (-u, -v)$.

3.2 Extensión a la rotación

Hemos visto que si el desplazamiento entre dos imágenes es una translación pura, los espectros de las imágenes, están ligadas por un desfase puro. Los **módulos** de los espectros son **idénticos** y la transformaciones de Fourier efectuadas sobre dichos módulos serán **invariantes a la translación**. En otras palabras los módulos de los espectros conservan únicamente la información de rotación.

Si se efectúa la transformación de Fourier sobre dos imágenes g_0 y g_1 ligadas por una rotación y translación, (por ejemplo las imágenes de las figuras 24 a y 24 b) los módulos de la transformada de Fourier de dichas imágenes solo conservan la rotación (ver figuras 25 a y 25 b).



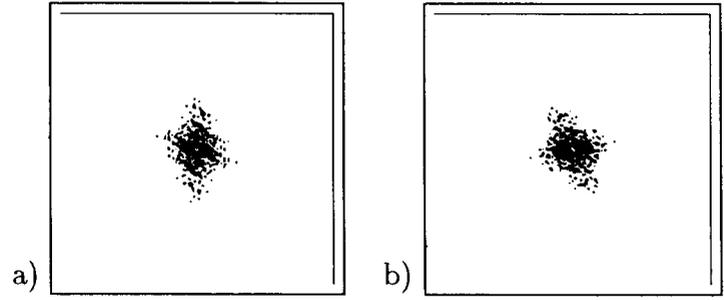
Talla de la imagen 128×128 pixels.

Figura 24: a) Imagen inicial, b) Imagen desplazada. Desplazamientos por transformaciones geométricas sin defectos, obtenidos por un programa. Translación en *pixels* de $(-10, -15)$ y rotación de $0,5$ rad sobre el centro de la imagen.

Expresando estos módulos en **forma polar**, se expresa la rotación bajo la forma de una translación. Por lo tanto será posible calcular la rotación que liga dos imágenes tomando el módulo de la transformada de Fourier de una imagen bajo la forma polar y calculando una translación que corresponderá a una rotación.

Para determinar dicha rotación será necesario calcular el espectro cruzado de la forma polar de los módulos de los espectros de Fourier. El cálculo es como sigue:

Dado un par de imágenes g_0 y g_1 ligadas por una rotación y translación y donde su transformada de Fourier esta dada por:



Talla de los espectros 512×512 pixels.

Figura 25: Espectros de Fourier de las imágenes 24 a y 24 b respectivamente. A pesar que entre las imágenes existe un desplazamiento (translación y rotación), los módulos sólo reflejan la rotación.

$$\mathcal{F}[g_0(x, y)] = G_0(f_{xy}).$$

Si R^{-1} es la matriz de rotación inversa, R el operador de rotación y Pol el operador de conversión de coordenadas rectangulares (*rec*) a polares (*pol*) entonces.

$$\mathcal{F}[g_1(x, y)] = e^{2i\pi(-t' f_{xy})} G_0(R^{-1} f_{xy}).$$

Nosotros calcularemos la transformación de Fourier denotando como $f_{\theta\rho}$ los puntos en el espacio de Fourier expresados en coordenadas polares y $ff_{\theta\rho}$ los puntos en el espacio de la doble transformación de Fourier en coordenadas polares.

Inicialmente se obtendrá el modulo de los espectros de Fourier.

$$\begin{aligned} \mathcal{M}_0(f_{xy}) &= |\mathcal{F}[g_0(x, y)]| \\ &= |G_0(f_{xy})|. \end{aligned} \quad (10)$$

$$\begin{aligned} \mathcal{M}_1(f_{xy}) &= |\mathcal{F}[g_1(x, y)]| \\ &= |G_0(R^{-1} f_{xy})| \\ &= R(\mathcal{M}_0(f_{xy})). \end{aligned} \quad (11)$$

Una vez determinado los módulos de los espectros (considerados ahora como imágenes) se procederá a obtener su transformación en coordenadas polares.

$$\mathcal{M}_{pol_0}(f_{\theta\rho}) = Pol[\mathcal{M}_{rec_0}(f_{xy})]. \quad (12)$$

$$\begin{aligned} \mathcal{M}_{pol_1}(f_{\theta\rho}) &= Pol[\mathcal{M}_{rec_1}(f_{xy})] \\ &= Pol[R(\mathcal{M}_{rec_0}(f_{xy}))] \\ &= t_{pol_\phi} \mathcal{M}_{pol_0}(f_{\theta\rho}). \end{aligned} \quad (13)$$

Para finalmente obtener el espectro cruzado de la forma polar de los módulos (ψ_{pol}).

$$\begin{aligned}
\psi_{pol}(ff_{\theta\rho}) &= \frac{\mathcal{F}[\mathcal{M}_{pol_1}(f_{\theta\rho})]}{\mathcal{F}[\mathcal{M}_{pol_0}(f_{\theta\rho})]} \\
&= \frac{\mathcal{F}[t_{pol_\phi}\mathcal{M}_{pol_0}(f_{\theta\rho})]}{\mathcal{F}[\mathcal{M}_{pol_0}(f_{\theta\rho})]} \\
&= e^{2i\pi(-t_{pol_\phi}ff_{\theta\rho})} \frac{\mathcal{F}[\mathcal{M}_{pol_0}(f_{\theta\rho})]}{\mathcal{F}[\mathcal{M}_{pol_0}(f_{\theta\rho})]} \\
&= e^{2i\pi(-t_{pol_\phi}ff_{\theta\rho})}.
\end{aligned}
\tag{14}$$

El vector de onda del espectro (ψ_{pol}) dará la translación $-t_{pol_\phi}$ y por lo tanto el ángulo ϕ de la rotación entre las dos imágenes.

De esta forma **primero se determinará la rotación**, para **después** (una vez alineadas las imágenes en rotación) **calcular la translación** entre ellas como se explicó en (§ 3.1).

En comparación con otros métodos basados en propiedades de la Transformación de Fourier [DeCastro and Morandi, 1986; Messner and Szu, 1985], este nuevo método tiene la ventaja de utilizar solamente 6 transformaciones de Fourier para estimar rotación y translación; en lugar de realizar un calculo sistemático de transformaciones a fin de maximizar un valor gracias a una serie de pruebas exhaustivas.

3.3 Resultados

Una serie de experimentaciones han permitido validar la robustez del método. Aquí presentaremos solo algunos ejemplos del tipo de pruebas efectuadas.

Las imágenes de la figura 26 presentan un ambiente no estructurado (exterior natural), el desplazamiento en la imagen es ocasionado por una rotación de la cámara sobre su eje óptico y su eje vertical.

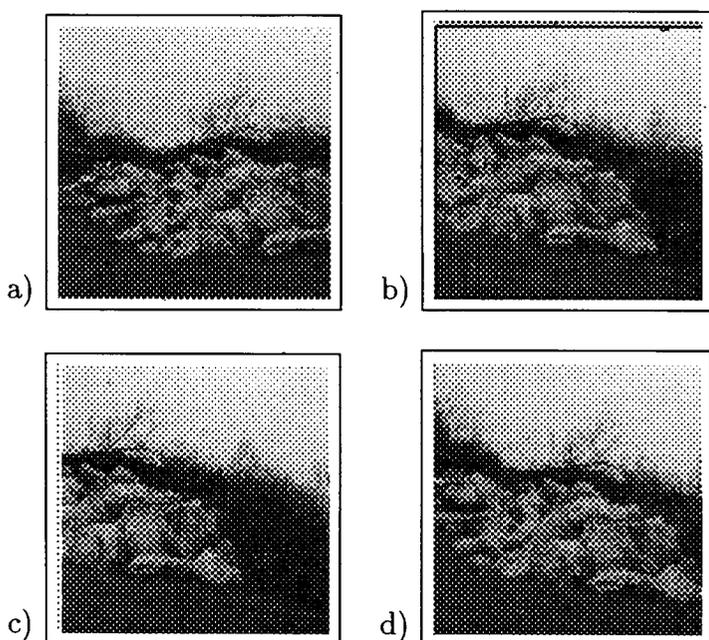


Figura 26: Imágenes de exterior tomadas en el LAAS-CNRS. Distorsión óptica no corregida

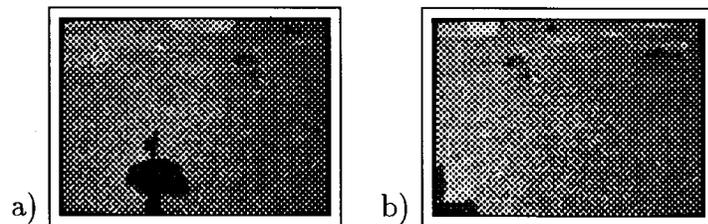


Figura 27: Imágenes tomadas por ADAM sobre un sitio de simulación de ambiente planetario del CNES (Centre National d'Etudes Spatiales). Distorsión óptica corregida

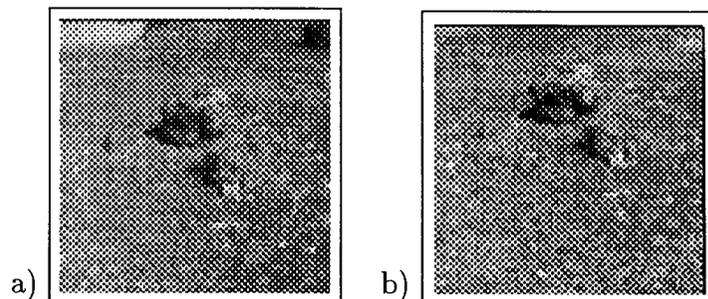


Figura 28: a) y b). Escenas extraídas de la imagen de la figura 27

Las imágenes de figura 27 fueron adquiridas por una cámara montada sobre el robot ADAM. El desplazamiento en las imágenes se debe al movimiento del robot. El análisis se ha realizado sobre sub-regiones de dichas imágenes (ver figuras 28 a y 28 b). El objetivo es tener una zona de imagen común suficientemente importante. Experimentalmente hemos determinado que para garantizar el buen funcionamiento del método será necesario tener al menos 55 % de zona de imagen común.

Una ejemplo de estabilización de una imagen es mostrado en la figura 29. Las figuras 29 (a), 29 (b), 29(c), 29 (d) y 29 (e), presentan las diferentes imágenes de la secuencia, el desplazamiento reflejado en las imágenes es ocasionado por el movimiento de la cámara. La figura 29 (f) muestra la sub-imagen a estabilizar (imagen inicial de referencia). La figura 29 (g) muestra la sub-imagen extraída de la primera imagen desplazada 29 (b), La estimación del desplazamiento es realizada utilizando estas dos sub-imágenes. La sub-imagen 29 (g) a sido extraída en la mismas coordenadas que la imagen 29 (f), mientras que la sub-imagen 29 (k) ha sido extraída tomando en cuenta el valor estimado del desplazamiento y realizando una corrección buscando compensar el movimiento del captor. Es posible observar que la corrección ha sido efectuada. La imagen corregida será utilizada ahora como nueva imagen de referencia a fin de estimar el siguiente desplazamiento. Este proceso es repetido durante toda la secuencia. La columna central de la figura 29 (sub-imágenes g, h, i y j) presenta las imágenes bajo la acción del desplazamiento del captor, la columna de la derecha (sub-imágenes k, l, m y n) presenta la sub-imagen estabilizada.

La tabla 1 indica para cada par de imágenes: los valores nominales de los desplazamientos, los valores calculados por nuestro método y el error resultante.

imágenes fuente desplazada	Angulo nominal (grados)	Δx nominal (pixels)	Δy nominal (pixels)	Angulo obtenido (grados)	Δx obtenido (pixels)	Δy obtenido (pixels)	Error ángulo (grados)	Error $\sqrt{(\Delta x)^2 + (\Delta y)^2}$ (pixels)
26 a 26 b	-11,31	33,30	-3,03	-11,25	34,96	-2,97	0,06	1,66
26 a 26 d	-12,53	9,93	-3,37	-11,95	11,47	-2,81	0,58	1,64
26 b 26 c	0,00	19,00	-3,00	0,00	19,00	-3,00	0,00	0,00
28 a 28 b	8,38	-3,95	-12,53	9,84	-4,34	-13,33	1,46	0,89
29 f 29 g	1,05	38,72	49,88	0,00	39,00	50,00	1,05	0,30
29 k 29 h	-0,50	28,60	-0,71	0,00	28,00	0,00	-0,50	0,92
29 l 29 i	0,00	2,00	7,00	0,00	2,00	7,00	0,00	0,00
29 m 29 j	0,00	22,00	13,00	0,00	22,00	12,00	0,00	1,00

Tabla 1: Resultados obtenidos.

Los errores sobre el cálculo de la translación son todos inferiores a dos pixels y el error para el ángulo es inferior a 1.5° , comportamiento que se observa sobre la totalidad de las pruebas efectuadas.

3.4 Limitaciones del método y sus adaptaciones a la robótica móvil

El método de estimación de la rotación de una cámara está principalmente limitado por dos factores:

La primera limitación se debe a la suposición de que los movimientos de la cámara se traducen por movimientos planos de la imagen (hipótesis de coplanaridad de las escenas). Lo que no necesariamente ocurre en el caso de robótica de exterior, donde la escena podrá contener elementos en diferentes planos.

Gracias a la técnica de focalización, es posible adecuar la escena percibida para que ella contenga un sólo plano. Así será posible seleccionar un plano de la escena que resulte interesante para la aplicación, sea por ejemplo los objetos próximos o un plano lejano al captor. Un estudio al respecto es presentado en [Ayala *et al.*, 1998].

La segunda limitación es la necesidad de tener un área común entre las imágenes analizadas igual o superior a 55 %. Por lo que se deberá alcanzar una frecuencia de tratamiento suficientemente elevada para que el desplazamiento del robot, no ocasione una zona de imagen común menor a la necesaria.

Con imágenes de talla 128×128 pixels y sobre una SPARC-20 la implementación en lenguaje C del método necesita de aproximadamente 15 segundos para realizar el análisis entre un par de imágenes.

Este proceso incluye básicamente el cálculo de 6 transformaciones de Fourier FFT "Fast Fourier Transform" (en imágenes de resolución aumentada en un factor de 2 por la técnica de zero-padding, imágenes resultantes de talla 256×256 pixels), una conversión a polares y una rotación.

El tiempo de ejecución no deberá ser impedimento para su aplicación en robótica móvil, en vista de la existencia de cartas electrónicas que realizan el "cálculo tiempo real" de la transformación de Fourier.

4 Conclusión y perspectivas

Para empezar esta conclusión analizaremos los puntos comunes de ambos métodos presentados:

- Ambas técnicas son capaces de trabajar sobre entidades no estructuras.
- Ambas técnicas están adaptadas para calcular movimientos debidos a cambio de la escena o a movimientos del captor.
- A pesar de que fueron concebidas para aplicaciones diferentes, los dos métodos (teóricamente) pudieran realizar ambas tareas (seguimiento de blanco y estabilización de cámara).

Si bien existen factores comunes entre las técnicas presentadas, hay también diferencias básicas, que son:

- El método de seguimiento de blanco, necesita un patrón a identificar, y a pesar que es capaz de manejar deformaciones del blanco, no será posible su funcionamiento, si no existe una entidad característica (diferente al resto de la escena) que sirva como blanco a seguir o como referencia que permitiera estabilizar la cámara.

Esta técnica puede localizar un objeto aun si éste ha sufrido una rotación (gracias a su capacidad de manejar deformaciones) pero no determina el valor de dicha rotación.

- El método usado para determinar la rotación de la cámara, no necesita un patrón y es capaz de determinar la rotación de los elementos dentro de la imagen, sin embargo necesita una cantidad de zona común entre las imágenes analizadas, y su tiempo de ejecución (en un sistema no dedicado) es significativamente mayor.

En vista de lo anterior consideramos que los métodos son complementarios y que de acuerdo a la tarea genérica asignada al robot deberán de usarse en cooperación.

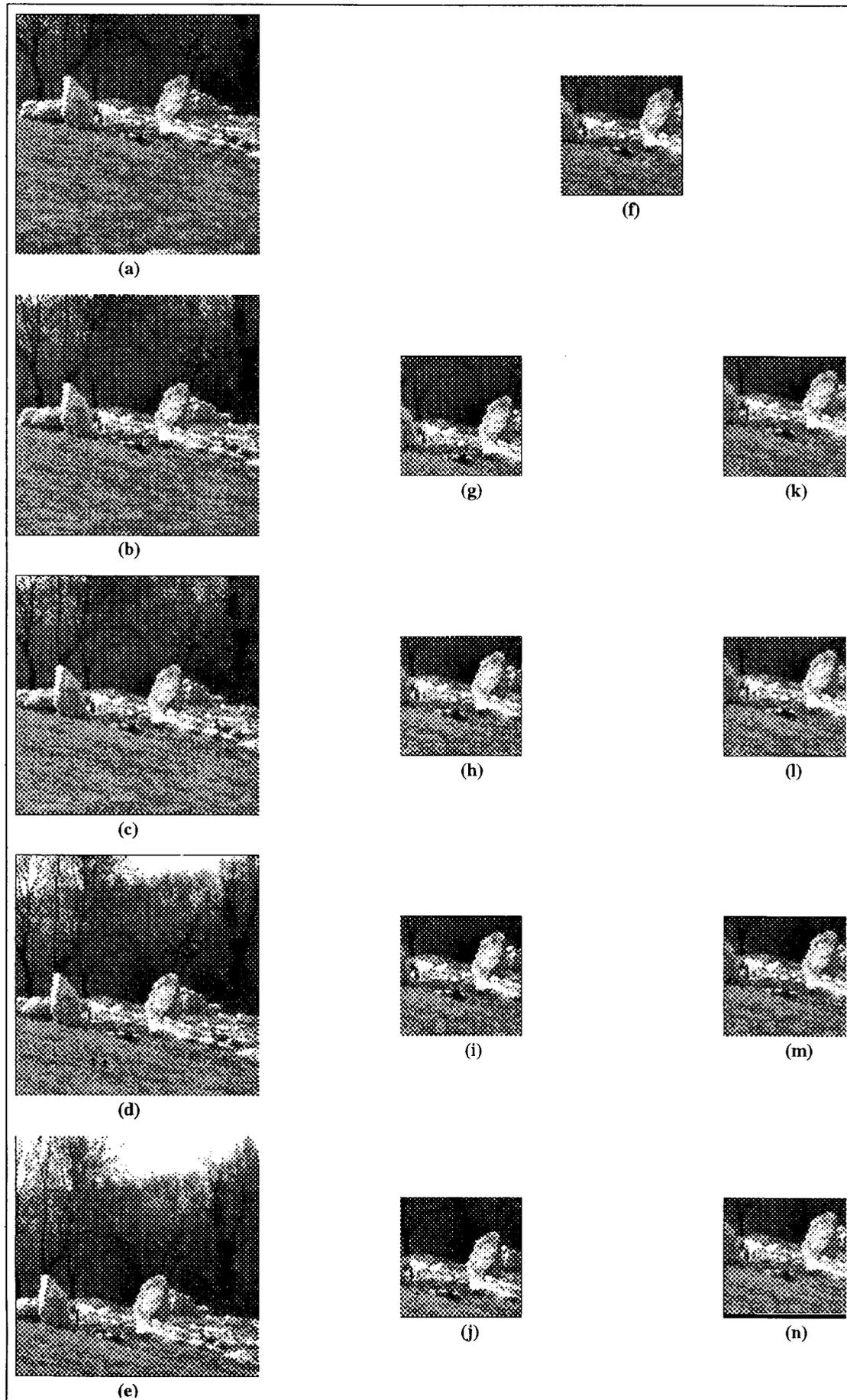


Figura 29 Estabilización de la imagen

Como perspectivas del presente trabajo y en términos de agregar funcionalidad a los sistemas, proponemos varias posibles extensiones.

En el caso del sistema de seguimiento de blanco:

- Extender el algoritmo para manejar diferentes modelos (diferentes vistas) de un mismo blanco. El objetivo es mejorar la robustez del método para manejar cambios de aspecto, así como para recuperar un blanco en caso de pérdida.
- Seguir múltiples blancos en la misma imagen.

Para el caso del sistema de estimación de rotación de la cámara, actualmente estamos efectuando dos extensiones.

- La utilización del método de estimación de desplazamiento entre imágenes en un sistema multi-resolución a fin de obtener el compromiso adecuado de precisión y rapidez de acuerdo a la tarea del robot.
- La implementación del método sobre un sistema Vx-Works, en vista de su aplicación en el robot móvil todo terreno LAMA.

Agradecimientos

El trabajo de Rafael Murrieta está financiado por una beca de CONACyT y la SEP, México. Los autores agradecen las valiosas sugerencias concernientes al método de seguimiento de blanco, del profesor Jean-Claude Latombe, Steven LaValle y David Lin de la Universidad de Stanford.

Nota

Una versión de divulgación de la primera parte de este trabajo, fue publicado en la revista Soluciones Avanzadas.

Referencias

V. Ayala, R. Murrieta-Cid, and M. Briot. A fourier transform based method for estimation of 2-d translation and rotation: an application to outdoor mobile robotics. In *Six International Symposium on Intelligent Robotic Systems (SIRS'98)*, June 1998.

C. Becker, H. González-Banos, J-C Latombe, and C. Tomasi. An intelligent observer. In *Proceedings International Symposium on Experimental Robotics*, 1995.

A. Blake. Affine-invariant contour tracking with automatic control of spatiotemporal scale. In *4th International Conference on Computer Vision*, pages 66–75, USA, April 1993.

G. Borgefors. Distance transforms in digital images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8:344–371, 1986.

L. Gottesfeld Brown. A survey of image registration techniques. *Computing surveys*, Vol. 24(Num. 4):325–347, 1992.

J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8(Num. 6):679–698, 1986.

P. E. Danielsson. Euclidean distance mapping. *Comput. graphics Image processing*, Vol. 14:227–248, 1980.

E. DeCastro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9(Num. 5):700–703, 1986.

P. Delagnes, J. Benois, and D. Barba Adjustable polygons: a novel active contour model for objects tracking on complex background. *Journal on communications*, Vol. 45:85–95, 1994.

Yue Du. A color projection for fast generic target tracking. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pages 73–80, 1995.

G. Garibotto. Motion tracking of connected edge contours. In *Proceedings of 3rd international workshop on Time-varying image processing and moving object recognition*, pages 323–330, 1990.

D. M. Gavrila and L.S Davis. 3-d model-based tracking of humans in action: a multi-view approach. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–80, 1996.

D. P. Huttenlocher, J.J. Noh, and W.J. Rucklidge. Tracking non-rigid objects in complex scenes. In *Proceedings Fourth International Conference on Computer Vision*, pages 93–101, 1993.

D.P. Huttenlocher. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15(Num. 9):850–863, 1993.

S. Jiansho and C. Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

B. Marcel, M. Briot, and R. Murrieta-Cid. Calcul de translation et rotation par la transformation de fourier. *Traitement du Signal*, Vol. 14(Num. 2), 1997.

R. A. Messner and H.H. Szu. An image processing architecture for real time generation of scale and rotation invariant patterns. *Computer Vision, Graphics and Image Processing*, (Num. 31), 1985.

R. Murrieta-Cid, M. Briot, and N. Vandapel. Landmark identification and tracking in natural environment. *Paper accepted to International Conference on Intelligent Robots and Systems IROS'98, to appear*, Victoria, Canada, September 1998.

R. Murrieta-Cid Target tracking method based on a comparison between an image and a model. *Technical Report Num. 97023, LAAS CNRS-Stanford University*, Toulouse, France, 1997.

D.W. Paglieroni. Distance transforms: properties and machine vision applications. *CV-GIP: Graphical models and Image processing*, Vol. 54(Num. 1), 1992.

J. Serra. *Image analysis and mathematical morphology*. Academic Press, London, 1982.

C. Wren. Pfunder: real-time tracking of the human body. In *Proceedings of the SPICE International society for optical engineering*, pages 89-98, 1996.



Maurice Briot

Maurice Briot es doctor de estado de la Universidad Paul Sabatier de Toulouse Francia, lugar donde tiene el grado de profesor y donde dirige el grupo de Sistemas Inteligentes. Asimismo, es investigador dentro del grupo de Robótica e Inteligencia Artificial del LAAS-CNRS. Sus principales áreas de interés son: Visión artificial y micro robótica concerniente a la percepción de robots móviles autónomos.

Baptiste Marcel

Baptiste Marcel es doctor en tratamiento de señales e imágenes del Instituto Politécnico de Toulouse Francia. El realizó su tesis de doctorado en el Grupo de Televisión y tratamiento de Señales e Imágenes del Laboratorio de Electrónica e Informática de Toulouse. El obtuvo el grado de Ingeniero en informática en el Instituto Nacional de Ciencias Aplicadas de Toulouse.

Héctor González

Héctor González es candidato a doctor en Ingeniería Eléctrica en la Universidad de Stanford, donde también se desempeña como asistente de investigación en el departamento de Ciencias Computacionales: Laboratorio de Robótica. Sus áreas de interés incluyen Algoritmos Geométricos, Métodos Numéricos y Sistemas de Control aplicados a Robótica. Su actual tema de investigación es "Construcción y Adquisición Automática de Modelos en 3-D mediante sensores Móviles".

Autores

Rafael Murrieta Cid

Rafael Murrieta es candidato a doctor en Visión artificial aplicada a Robótica móvil en el Instituto Politécnico de Toulouse Francia; realiza su tesis doctoral en el grupo de Robótica e Inteligencia Artificial del LAAS-CNRS. Dentro del marco de su tesis doctoral él ha participado en el proyecto "Observadores Autónomos" en la Universidad de Stanford California. Sus áreas de interés son la interpretación de escenas, la visión activa y la navegación de robots basada en información perceptual.

