

Un Enfoque Neuronal al Seguimiento de un Objeto a través de su Tamaño y su Localización

Antonio Fernández-Caballero ¹, Modesto Montoya ²,
Miguel A. Fernández ¹ y Francisco J. Gómez ¹

Departamento de Informática, Escuela Politécnica Superior
Grupo de Investigación en Computación Neuronal e Inteligencia Artificial (GICoNIA)
Universidad de Castilla-La Mancha, 02071 -- Albacete, España

¹ {caballer, miki, fgomez}@info-ab.uclm.es

² mmontoya@giconia.info-ab.uclm.es

Resumen

El análisis del movimiento en secuencias de imágenes es una disciplina en constante crecimiento debido al gran número de aplicaciones en el que juega un papel primordial. La presentación de un sencillo modo de obtención, en tiempo real y de una forma neuronal, de dos parámetros fundamentales asociados a los objetos en movimiento (su tamaño y su localización), así como su uso para el seguimiento de un objeto es el objetivo de este artículo.

El artículo se completa con un extenso juego de ensayos que permite observar la bondad de los algoritmos expuestos.

El artículo es una versión ampliada del trabajo "Tamaño y localización de un objeto: Una aproximación neuronal", presentado en el IV Simposio Iberoamericano de Reconocimiento de Patrones SIARP'99.

Palabras clave: Detección de movimiento, análisis de movimiento, seguimiento, redes neuronales.

1. Introducción

El análisis del movimiento en imágenes está creciendo en importancia en numerosas aplicaciones. Algunas de estas aplicaciones son: (a) la televisión en cuanto a la codificación de las imágenes de vídeo mediante compensación, (b) la robótica móvil, (c) el tratamiento de las imágenes de satélite, (d) las aplicaciones, tanto civiles como militares, de seguimiento de objetivos y navegación autónoma, (e) el tratamiento de las imágenes biológicas y médicas, (f) la vigilancia de lugares y su supervisión, y, (g) interfaces y realidad virtual (Mítiche y Bouthemy, 1996). El problema de la detección de movimiento es particularmente interesante cuando el objetivo que se persigue es el de la localización espacial de los objetos móviles en la escena. Esta detección de movimiento siempre está fuertemente ligada a la detección de cambios temporales en la imagen. Cuando existen objetos en movimiento en una escena, siempre tendremos cambios en la intensidad de los píxeles de la imagen. Este hecho ha dado lugar a una extensa bibliografía, en la que destacamos algunos trabajos ya clásicos con diferentes enfoques. Para un estudio más extenso del tema, se recomienda la lectura de Fernández (1993) y Fernández-Caballero (1997).

El seguimiento de elementos de una imagen a otra es un procedimiento común, sobre todo en aplicaciones de vigilancia. Algunos procesos de seguimiento han sido definidos por medio de: (a) un modelo representativo de elementos (por ejemplo, las coordenadas de imagen de algunos puntos característicos, la longitud y la orientación de los segmentos del contorno), (b) un modelo cinético de la

evolución de los elementos (por ejemplo, una velocidad constante, una aceleración constante, un movimiento afín), (c) un conjunto de relaciones entre los parámetros del modelo y las medidas de la imagen, y, (d) un filtro temporal para la estimación de los parámetros del modelo a partir de los datos de la imagen.

2. Técnicas de detección de movimiento

Las distintas técnicas de detección de movimiento pueden clasificarse de muchas maneras diferentes. Nuestro enfoque contempla una primera división en métodos 3-D y 2-D. Los métodos 3-D intentan resolver las ecuaciones de las proyecciones utilizando directamente las correspondencias entre las características entre las tramas, mientras que los métodos 2-D estiman el flujo óptico.

2.1. Estimación del movimiento 3-D

El problema de estimar el movimiento 3-D a partir de un conjunto de proyecciones 2-D de la escena es un asunto complejo. La estimación 3-D, en general, se aborda desde dos ópticas distintas: la estimación basada en la correspondencia de características (Huang y Netravali, 1994) y los métodos directos (Horn y Schunck, 1981).

Los métodos directos estiman el flujo óptico 2-D y deducen el movimiento 3-D a partir de un análisis cuidadoso del campo de flujo estimado, tras haber aplicado toda una serie de restricciones adicionales para simplificar el problema. Los métodos directos 3-D pueden considerarse, por tanto, como una extensión de los métodos 2-D de flujo óptico.

Los métodos de correspondencia de características extraen un conjunto altamente discriminante de rasgos en las imágenes y deducen el movimiento resolviendo las ecuaciones de las proyecciones después de haber determinado algunas correspondencias, haciendo un uso explícito de la restricción de rigidez. El problema se descompone en tres pasos: el descubrimiento de las características, la determinación de las correspondencias y la deducción del movimiento y de los parámetros estructurales, dadas las posiciones de los rasgos correspondientes. Algunos ejemplos típicos incluyen puntos, bordes y esquinas (Huang y Netravali, 1994; Liu y Huang, 1988; Liu y Huang, 1989).

El problema de la correspondencia de características puede verse como un problema de búsqueda. Las técnicas de filtrado temporal, como en el caso del filtrado de Kalman (Deriche y Faugeras, 1990; Faugeras *et al.*, 1987), se aplican a menudo en el seguimiento a través del tiempo de características con el fin de reducir el espacio de búsqueda asociado a cada rasgo. Muchos de los algoritmos aplicados hasta la fecha asumen que las correspondencias se conocen de antemano.

La correspondencia de puntos fue usada por Ullman (1979) y Roach y Aggarwal (1980). El segundo de los enfoques, a

modo ilustrativo, contemplaba la resolución de un conjunto de 20 ecuaciones con 27 incógnitas.

La utilización de correspondencias de líneas complica aún más el problema. Puede demostrarse que el movimiento 3-D no puede determinarse de la correspondencia de líneas en dos proyecciones solamente (Weng *et al.*, 1992). El uso de correspondencias de líneas lleva generalmente a un conjunto grande de ecuaciones no lineales que sólo pueden resolverse por métodos iterativos (Holt y Netravali, 1993). Aunque existe una formulación lineal, se produce un considerable número de soluciones espurias (Liu y Huang, 1989).

2.2. Estimación del movimiento 2-D

La mayoría de las aplicaciones reales de estimación del movimiento trabajan, de alguna forma, con un acercamiento 2-D, en el que se obtienen estimaciones del movimiento en el tiempo del movimiento 2-D de patrones de intensidad (Aggarwal y Nandhakumar, 1988; Barron *et al.*, 1994). En estos métodos encontramos la restricción de que el brillo de la imagen a lo largo de la trayectoria del movimiento es constante, es decir, que cualquier cambio en el tiempo en la intensidad de un punto es debido únicamente al movimiento.

Dos visiones diferentes del mismo problema han llevado a dos filosofías diferentes en la estimación del movimiento 2-D: la utilización directa de la ecuación del flujo óptico, y, el cálculo del mínimo de la diferencia del desplazamiento entre tramas en un conjunto de regiones locales, buscando en un conjunto de vectores de movimiento. Basándose en estas dos ideas, se han propuesto múltiples algoritmos de estimación de movimiento 2-D:

- a. Los métodos basados en el gradiente, donde se usa directamente la ecuación del flujo óptico, estimando los gradientes espacial y temporal en cada punto de la imagen, e imponiendo alguna restricción adicional para identificar el movimiento de un modo unívoco.
- b. Los métodos de correspondencia de regiones, donde la diferencia del desplazamiento entre tramas (o cualquier otro criterio de error similar) se minimiza sobre un conjunto de regiones locales empleando algún mecanismo de búsqueda apropiado.
- c. Los métodos en el dominio de frecuencias, en los que se explota la invarianza en el desplazamiento y las propiedades de simetría de la transformada de Fourier para estimar el movimiento.
- d. Los métodos bayesianos, en los que se minimiza globalmente la diferencia del desplazamiento entre tramas por medio de técnicas de optimización estocásticas y el modelado explícito de las propiedades estructurales del campo de movimiento.

2.2.1. Métodos basados en el gradiente

La estimación basada en el gradiente se ha convertido en el principal acercamiento en las aplicaciones de visión por computadora. Ello se debe principalmente a que es computacionalmente eficaz y a que se produce una buena estimación del campo de movimiento. Los métodos basados en el gradiente, para poder usarse en la estimación de movimiento, requieren de algunas restricciones. Se han propuesto muchas de estas restricciones, aunque normalmente ligadas a alguna forma de restricción de suavizado.

El método basado en el gradiente más conocido es el de Horn y Schunck (1981). Su método impuso una restricción de suavizado global en el campo del movimiento. Un ejemplo de método que impone una restricción de suavizado local en el campo del movimiento fue propuesto por Lucas y Kanade (1981).

Se han realizado igualmente algunos esfuerzos para obtener modelos de alto nivel a partir del campo de flujo óptico por medio de técnicas de regresión lineal. Este acercamiento es uno de los más comunes en los modelos de movimiento afín (Black y Anandan, 1996; Meyer y Bouthemy, 1994; Wang y Adelson, 1994).

Las desventajas comunes a todos los métodos basados en el gradiente provienen de los lógicos cambios en la iluminación (Kearney *et al.*, 1987; Verri y Poggio, 1989) y en la dificultad de abordar numéricamente la diferenciación de la función de intensidad de la imagen.

2.2.2. Métodos basados en la correlación de regiones

Obviamente, una forma de evitar las limitaciones de los métodos basados en el gradiente está en considerar regiones de la imagen en lugar de estimar el movimiento pixel a pixel. En general, los métodos basados en la correlación de regiones son menos sensibles al ruido al tomar más datos de la imagen en el proceso de estimación del movimiento.

Las técnicas de correlación se basan en minimizar la diferencia en el desplazamiento entre tramas en un bloque de píxeles. Representan el acercamiento a la estimación de movimiento más ampliamente usado hoy en día (Jain y Jain, 1981). En su forma básica, el algoritmo de correlación de bloques divide una imagen en una serie de regiones de igual tamaño. Para cada una de las regiones se busca, en la siguiente trama, la posible correlación en su vecindad, minimizando un criterio de error como la diferencia en el desplazamiento entre tramas, u otra medida relacionada, sobre un conjunto de vectores de movimiento. Al trabajar con un tamaño fijo de bloque, se está limitando el campo del movimiento estimado, ya que no hay posibilidad de manejar bloques que contengan varios movimientos.

Con el fin de superar esta limitación, algunos investigadores han trabajado con correlación de bloques en entornos de multirresolución. La correlación de bloques en multirresolución puede ser clasificada, a su vez, en dos grandes grupos. El primer grupo utiliza la correlación de bloques en una pirámide gaussiana (Burt y Adelson, 1983). La idea consiste en que los movimientos mayores pueden estimarse utilizando versiones suavizadas y submuestreadas de las imágenes (Anandan, 1989). Este tipo de métodos de multirresolución se denomina también de multiescalado. Las estimaciones más burdas se propagan por la pirámide hacia abajo, donde son refinadas. El segundo grupo, a veces llamado de técnicas multimalla, usa correlación de bloques en una estructura de quadtree, creando un mosaico de bloques espaciales de diferentes tamaños (Dufaux y Moscheni, 1995). Este acercamiento permite que la región analizada se adapte a los datos subyacentes.

También se han presentado generalizaciones a la correlación anteriormente descrita. La idea es crear un tamiz adaptable del plano de la imagen y perturbar este tamiz para minimizar una medida de error sobre las regiones (Bradshaw y Kingsbury, 1997; Nakaya y Harashima, 1994; Seferidis y Ghanbari, 1993).

Otros investigadores han usado la transformada de Hough para la estimación del movimiento (Adiv, 1985; Bober y Kittler, 1994). La transformada de Hough puede concebirse como una técnica de clasificación en el espacio definido por los parámetros que describen la forma. El acercamiento mediante la transformada de Hough puede modificarse para trabajar directamente sobre los niveles de gris de la imagen (Bober y Kittler, 1994).

Una desventaja compartida de los esquemas basados en la correlación de regiones es su dependencia del modelo de movimiento traslacional. Las regiones deben permanecer bastante pequeñas para el que el modelo traslacional permanezca válido.

2.2.3. Métodos en el dominio de frecuencias

Los métodos en el dominio de frecuencias caen en dos grandes clases: los métodos basados en las fases y los métodos basados en la energía. Los métodos basados en las fases se aprovechan del hecho de que un cambio en el dominio espacial produce un cambio de la fase en el dominio de frecuencias (Papoulis, 1984), mientras que los métodos basados en la energía trabajan con la distribución de energía en el espacio de frecuencias al considerar el movimiento en sus orientaciones espacio-temporales.

Se han usado técnicas basadas en las fases para la estimación de la disparidad, sobre todo en el contexto de la estereoscopia (Jepson y Jenkin, 1989; Wilson y Knutsson, 1989). Aunque el problema de la estereoscopia es similar al problema de la estimación del movimiento, se basa en la suposición de que las disparidades son 1-D, o sea, horizontales.

Para Jepson y Jenkin (1989) la disparidad se define en términos de la diferencia de fase en un punto dado entre sus valores filtro paso-banda de dos tramas. Otro acercamiento a la estimación del movimiento basado en fases proviene de Fleet y Jepson (1990), donde se aplica una técnica basada en el gradiente a la componente fase de las salidas de un conjunto de filtros direccionales paso-banda sintonizados a diferentes velocidades.

Calway *et al.* (1992) desarrollaron un enfoque al dominio de frecuencias basado en una técnica de correlación. Su método también se lleva a cabo dentro de un entorno de multiresolución, denominado Multiresolution Fourier Transform (MFT) (Wilson *et al.*, 1992). La correlación en el dominio espacial es muy costosa, pero, llevada al dominio de frecuencias, la carga computacional se torna manejable. Otro método descrito es el de correlación de fases (Kuglin y Hines, 1975), donde el producto de los espectros se normaliza antes de aplicar la transformada de Fourier inversa, obteniéndose en el caso ideal un impulso en la posición correspondiente al vector asociado al movimiento.

Otra técnica en el dominio de frecuencias es la del filtrado espacio-temporal (Adelson y Berger, 1985; Heeger, 1987). La idea básica se basa en la observación de que el movimiento se corresponde con la orientación en el espacio-tiempo.

Los métodos en el dominio de frecuencias presentan algunas ventajas notables respecto de muchos métodos en el dominio espacial. En primer lugar, son menos sensibles a los cambios en la iluminación global, y, en segundo lugar, son bastante robustos con respecto al ruido.

2.2.4. Métodos bayesianos

El problema de la estimación del movimiento puede formularse como un problema de estimación global bayesiano. La idea general consiste en usar un criterio de máximo a posteriori (MAP) para maximizar la probabilidad del campo de movimiento, dada la observación de la intensidad de la imagen en la próxima trama. La estimación del MAP puede lograrse por medio de los métodos estocásticos de relajación, tales como el temple simulado, que garantizan la convergencia en un máximo global. Los métodos bayesianos han evolucionado a partir de los métodos recursivos sobre píxeles (Robbins y Netravali, 1983) y normalmente se formulan usando campos aleatorios de Markov (MRF) (Spitzer, 1971).

Los enfoques bayesianos a la estimación del movimiento son de gran interés al enfrentarse directamente con una de las dificultades fundamentales: cómo segmentar el campo de movimiento en regiones de movimiento coherente trabajando pixel a pixel. Sin embargo, padecen de la lentitud en la convergencia inherente a los métodos estocásticos de relajación.

3. Tamaño y localización de un objeto

Uno de los aspectos primordiales de la estructura de un objeto es su tamaño. Conocer el tamaño de un objeto en términos absolutos es útil para el reconocimiento del objeto (Mutch, 1986; Mutch y Heiny, 1986). A su vez, el aspecto predominante para el cálculo de la velocidad de un objeto sería su localización. Un objeto en traslación o dilatación (no en rotación) puede reducirse a efectos de cálculo del movimiento a la obtención de esos dos parámetros: su localización y su tamaño en todo instante de tiempo t . Véase que cuando hablamos de dilatación queremos englobar igualmente la contracción.

Nuestra propuesta es que ambas variables pueden obtenerse de un modo sencillo, conocida la silueta S del objeto en el instante t (Fernández-Caballero *et al.*, 1999). Para ello vamos a definir, por un lado, el tamaño a partir de la longitud de dos líneas rectas (o cuerdas) determinadas por cuatro puntos conocidos de la superficie del objeto. Los puntos a los que hacemos referencia (x_1, y_1) , (x_2, y_2) , (x_3, y_3) y (x_4, y_4) , tales que:

$$\forall(x,y) \in S(i,j,t), \quad x_1 < x$$

$$\forall(x,y) \in S(i,j,t), \quad x_2 > x$$

$$\forall(x,y) \in S(i,j,t), \quad y_3 < y$$

$$\forall(x,y) \in S(i,j,t), \quad y_4 > y$$

Dicho de otro modo, los cuatro puntos son:

(x_1, y_1) : punto más a la izquierda del objeto en la imagen

(x_2, y_2) : punto más a la derecha del objeto en la imagen

(x_3, y_3) : punto más arriba del objeto en la imagen

(x_4, y_4) : punto más abajo del objeto en la imagen

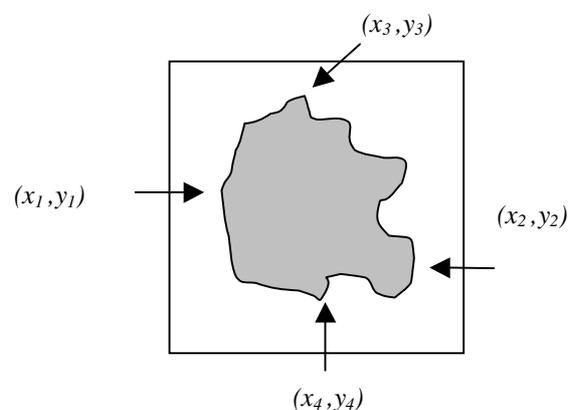


Figura 1. Obtención de los puntos extremos de un objeto.

Las dos cuerdas, que denominamos segmentos de línea máximos del objeto, no van a unir los puntos (x_1, y_1) y (x_2, y_2) , (x_3, y_3) y (x_4, y_4) entre sí, respectivamente, sino sus proyecciones $(X_1, 0)$ y $(X_2, 0)$, $(0, Y_3)$ y $(0, Y_4)$, respectivamente, como puede apreciarse en la figura 2.

En cuanto a la localización del objeto, ésta vendrá determinada por un punto único característico del mismo (X_{obj}, Y_{obj}) , a saber la intersección de los dos segmentos $(X_1, Y_3)(X_2, Y_4)$ y $(X_2, Y_3)(X_1, Y_4)$. Este punto será denominado punto representativo del objeto.

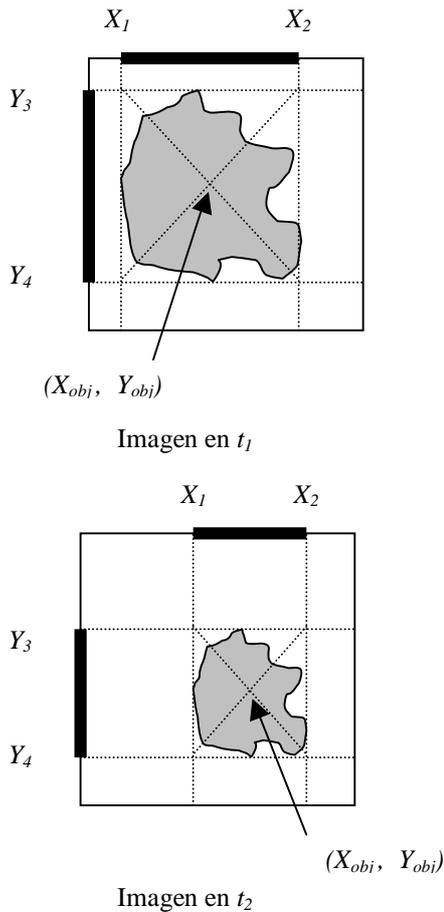


Figura 2. Obtención de los segmentos de línea máximos y del punto representativo del objeto.

4. Implementación neuronal

Se propone una implementación hardware que obtenga los valores X_1, X_2, Y_3 e Y_4 en tiempo real. Para ello se propone implementar una estructura neuronal multifuncional (Alevizos *et al.*, Safonova *et al.*, 1996) de fácil manejo. Partimos de la estructura básica de neurona de la figura 3.

En dicha figura tenemos:

- INH_{in} = señal de inhibición procedente de la neurona anterior
- ACT_{in} = señal de activación de entrada de la neurona
- INH_{out} = señal de inhibición hacia la neurona Posterior
- ACT_{out} = señal de activación de salida de la neurona

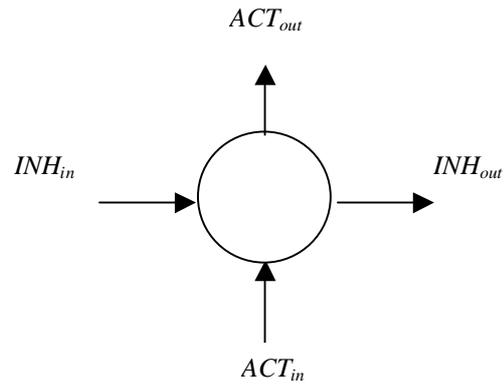


Figura 3. Estructura de la neurona multifuncional.

Esta neurona tiene como característica primaria el poder ser unida en serie con otras neuronas del mismo tipo a través de las señales INH del siguiente modo:

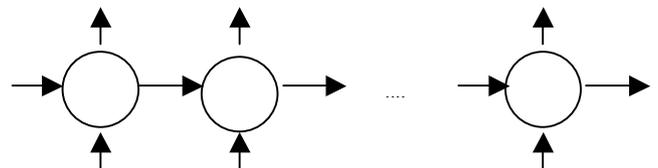


Figura 4. Conexión en serie de las neuronas.

Véase que la señal INH se va propagando con el valor inicial 0 (es decir, no inhibir) hasta una neurona determinada que cumpla la condición adecuada para transmitir a partir de ese momento el valor 1 (inhibir).

En el caso concreto que nos ocupa, se trata de detectar los valores X_1, X_2, Y_3 e Y_4 anteriormente descritos. Para ello se propone utilizar cuatro hileras de neuronas dispuestas según la figura 5, donde las señales ACT (tanto ACT_{in} como ACT_{out}) tienen como finalidad el pasar a través de la neurona la información a ser procesada desde un nivel inferior (obtención de la silueta del objeto) a un nivel superior (cálculo de la trayectoria del objeto).

El algoritmo se presenta aquí para el caso de las neuronas detectoras de las filas Det_{X1} o Det_{X2} . Para los casos Det_{Y3} o Det_{Y4} , cámbiese i por j , fila por columna, y viceversa.

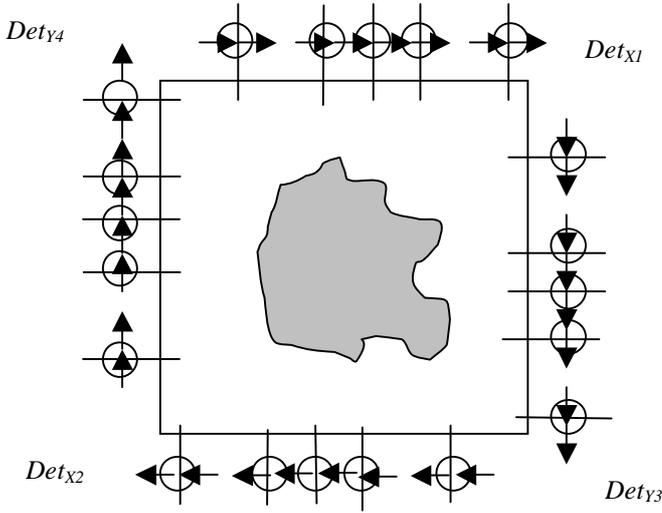


Figura 5. Determinación de los valores X_1 , X_2 , Y_3 , e Y_4 .

$$ACT_{in}(i, t) = \begin{cases} 1, & \text{si } \sum_{j=1}^k S(i, j, t) > 0 \\ 0, & \text{en caso contrario} \end{cases} \quad (1)$$

Esta ecuación (1) viene a decir que la neurona de la columna i tiene una activación en su entrada si cualquiera de los pixeles de la imagen de la columna i y filas j está incluido en la silueta del objeto.

$$INH_{out}(i, t) = \begin{cases} 1, & \text{si } INH_{in}(i, t) = 1 \cup ACT_{in}(i, t) = 1 \\ 0, & \text{en caso contrario} \end{cases} \quad (2)$$

Se va propagando a través de la línea de neuronas la señal INH_{out} con valor 0 hasta que ocurre uno de dos posibles eventos (no tienen por qué ser excluyentes entre sí): (a) a la neurona llega un valor de inhibición de 1, o, (b) la neurona recibe una señal de activación desde el nivel inferior. En ambos casos, comienza a propagarse la señal INH_{out} con un valor de 1.

$$ACT_{out}(i, t) = i * ACT_{in}(i, t) * [1 - INH_{in}(i, t)] \quad (3)$$

Esta última ecuación nos ofrece el comportamiento de la neurona en su salida hacia el nivel superior. En este caso la función elegida permite elevar a instancias superiores el valor de la posición de la neurona dentro de la fila. Vemos como, en cada hilera, a lo sumo una neurona pasará el valor de su posición al nivel superior. Todas las demás valdrán 0. La neurona "ganadora" es, pues, aquella que detecta la primera posición en que se encuentra silueta. Así:

Det_{X1} : consigue obtener la posición de la columna donde aparece silueta más a la izquierda, es decir X_1

Det_{X2} : consigue obtener la posición de la columna donde aparece silueta más a la derecha, es decir X_2

Det_{Y3} : consigue obtener la posición de la columna donde aparece silueta más arriba, es decir Y_3

Det_{Y4} : consigue obtener la posición de la columna donde aparece silueta más a bajo, es decir Y_4

5. Seguimiento del movimiento

Una vez que se han obtenido los segmentos de línea máximos y el punto representativo de un objeto en una secuencia de imágenes, es bastante sencillo detectar un gran número de casos de movimiento. Si consideramos las siguientes posibilidades:

- no hay movimiento (N),
- traslación en el eje X o Y (T),
- dilatación o traslación en el eje Z (D), y,
- rotación (R)

podemos únicamente obtener por combinación los siguientes casos de movimiento:

- | | |
|---------|--|
| (A) N | no se detecta movimiento |
| (B) T | detectada una traslación pura |
| (C) TD | detectada una traslación más una dilatación |
| (D) TR | detectada una traslación más una rotación |
| (E) TDR | detectada una traslación más una dilatación más una rotación |
| (F) D | detectada una dilatación pura |
| (G) DR | detectada una dilatación más una rotación |
| (H) R | detectada una rotación pura |

Consideramos que los estados anteriores aparecen en la mayoría de los casos tal como se muestra en el grafo 1. El grafo 1 muestra en la diferentes posibilidades dependiendo de la evolución entre dos imágenes consecutivas de los dos parámetros fundamentales estudiados hasta el momento, a saber el tamaño y la localización del objeto en movimiento.

(1)	(2)	(3)	(4)	(5)
		Similar	D	TD
	Mayor	Diferente	DR	TDR
Mayor	Igual		R	TR
	Menor		R	TR
	Mayor		R	TR
Igual	Igual		N	T
	Menor		R	TR
	Mayor		R	TR
Menor	Igual		R	TR
	Menor	Similar	D	TD
		Diferente	DR	TDR

Grafo 1. Evaluación del movimiento.

En dicho grafo, debe entenderse:

- (1) comparación entre los segmentos de línea máximos horizontales de la imagen anterior $k-1$ y actual k

$$\begin{cases} \text{Mayor,} & \text{si } (X_2 - X_1)_k - (X_2 - X_1)_{k-1} > l \\ \text{Igual,} & \text{si } (X_2 - X_1)_k - (X_2 - X_1)_{k-1} = l \\ \text{Menor,} & \text{si } (X_2 - X_1)_k - (X_2 - X_1)_{k-1} < l \end{cases}$$

siendo l la diferencia máxima permitida

- (2) comparación entre los segmentos de línea máximos verticales de la imagen anterior $k-1$ y actual k

$$\begin{cases} \text{Mayor,} & \text{si } (Y_4 - Y_3)_k - (Y_4 - Y_3)_{k-1} > l \\ \text{Igual,} & \text{si } (Y_4 - Y_3)_k - (Y_4 - Y_3)_{k-1} = l \\ \text{Menor,} & \text{si } (Y_4 - Y_3)_k - (Y_4 - Y_3)_{k-1} < l \end{cases}$$

siendo l la diferencia máxima permitida

- (3) grado de similitud entre el cambio de escala de los segmentos máximos de las imágenes en $k-1$ y k

$$\begin{cases} \text{Similar,} & \text{si } 1 - \alpha \leq \frac{(X_2 - X_1)_k}{(Y_4 - Y_3)_k} \leq 1 + \alpha \\ \text{Diferente,} & \text{en caso contrario} \end{cases}$$

siendo α la fluctuación permitida en la función de similitud

- (4) estado resultante si el punto representativo del objeto no ha cambiado sustancialmente; un cambio no sustancial se obtiene por medio del algoritmo

$$(X_{objk} - X_{objk-1} \leq d) \cap (Y_{objk} - Y_{objk-1} \leq d)$$

siendo d el desplazamiento máximo permitido

- (5) estado resultante si el punto representativo del objeto sí que ha cambiado suficientemente; se trata de la condición negada del punto anterior

Por supuesto, asumimos la posibilidad de ofrecer algunos resultados erróneos, especialmente ante algunos ejemplos que incluyan rotaciones. No obstante, si el número de imágenes es suficiente en la secuencia de imágenes, el error tenderá a ser pequeño.

6. Pruebas y resultados

Los algoritmos expuestos anteriormente se han aplicado a multitud de secuencias sintéticas como las ofrecidas en las figuras 6 a 9, así como a secuencias reales de las que ofrecemos el ejemplo de la figura 10.

Para la realización de las pruebas que se detallan a continuación, se ha implementado un prototipo software que incorpora cada uno de los pasos descritos con anterioridad en este artículo. Este prototipo, escrito bajo lenguaje MS Visual C++ versión 5.0, ha sido íntegramente desarrollado por Modesto Montoya y se ha denominado *TamYLoca*.

La aplicación *TamYLoca* recibe como entrada la secuencia de imágenes a tratar y ofrece como resultado los datos tal como se ofrecen en las tablas 1 a 5 correspondientes a los ejemplos ofrecidos. Véase que cada una de las líneas de resultado se corresponde con la aplicación de nuestros algoritmos a las imágenes en los instantes de tiempo t y $t-1$.

Se pasa a exponer cada uno de los casos descritos en este artículo, así como a comentar los resultados apreciados para cada uno de ellos.

Ejemplo 1

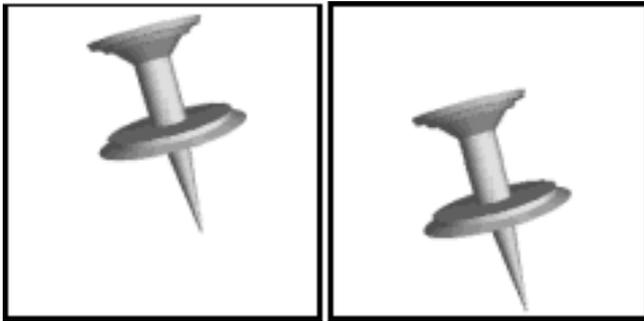


Figura 6. Imágenes 1 y 20 de una chincheta con traslación pura en el eje y .

Ejemplo 2



Figura 7. Imágenes 1 y 20 de un taco con traslación en los tres ejes x , y , z .

Imagen	X_{obj}	Y_{obj}	Seg.H.	Seg.V.	Movimiento
1	128,5	96,0	128,0	181,0	
2	128,5	99,5	128,0	180,0	T
3	128,5	103,0	128,0	181,0	T
4	128,5	106,0	128,0	181,0	T
5	128,5	109,5	128,0	180,0	T
6	128,5	113,0	128,0	181,0	T
7	128,5	116,5	128,0	180,0	T
8	128,5	120,0	128,0	181,0	T
9	128,5	123,0	128,0	181,0	T
10	128,5	126,5	128,0	180,0	T
11	128,5	130,0	128,0	181,0	T
12	128,5	133,5	128,0	180,0	T
13	128,5	136,5	128,0	180,0	T
14	128,5	140,0	128,0	181,0	T
15	128,5	143,5	128,0	180,0	T
16	128,5	147,0	128,0	181,0	T
17	128,5	150,5	128,0	180,0	T
18	128,5	153,5	128,0	180,0	T
19	128,5	157,0	128,0	181,0	T
20	128,5	160,5	128,0	180,0	T

Tabla 1. Resultados para la secuencia de la chincheta.

Como botón de muestra, se ofrece en este primer ejemplo una traslación pura en uno de los ejes, en este caso concreto, en el eje y . Puede apreciarse cómo los algoritmos expuestos funcionan a la perfección en este sencillo caso, aún a pesar de la rebuscada forma del objeto tratado.

Las traslaciones puras en los tres ejes x , y , z con este mismo objeto sintético (y otros) han sido todas probadas y han ofrecido los mismos buenos índices de comportamiento. Evidentemente, este es el caso que con mayor profusión ha sido tratado en el pasado. Era de esperar que funcionase así de bien.

Imagen	X_{obj}	Y_{obj}	Seg.H.	Seg.V.	Movimiento
1	64,0	64,5	61,0	78,0	
2	70,5	69,0	64,0	83,0	TD
3	77,5	74,0	66,0	87,0	TD
4	83,5	78,0	70,0	91,0	TD
5	91,5	82,5	74,0	96,0	TD
6	97,5	87,5	76,0	100,0	TD
7	104,5	91,5	80,0	104,0	TD
8	111,0	96,0	83,0	109,0	TD
9	118,0	101,0	87,0	113,0	TD
10	124,5	105,0	90,0	117,0	TD
11	131,0	110,0	93,0	121,0	TD
12	138,0	114,5	97,0	126,0	TD
13	145,0	119,0	101,0	131,0	TD
14	151,5	123,5	104,0	134,0	TD
15	159,0	128,0	107,0	139,0	TD
16	165,5	133,0	110,0	143,0	TD
17	172,0	137,0	113,0	147,0	TD
18	179,0	141,5	117,0	152,0	TD
19	185,5	146,0	120,0	156,0	TD
20	192,0	150,5	123,0	160,0	TD

Tabla 2. Resultados para la secuencia del taco.

Este segundo ejemplo es el representante de los movimientos traslacionales algo más complejos. Aquí estamos hablando de traslaciones simultáneas en varios ejes. Se han probado todas las combinaciones de traslaciones, obteniéndose para todos los objetos analizados un excelente comportamiento de los algoritmos.

Este ejemplo en concreto ofrece el movimiento traslacional de un taco en los tres ejes de modo simultáneo. De ahí que se arroje el resultado correcto TD en la totalidad de los veinte pasos de que consta la secuencia sintética abordada.

Ejemplo 3

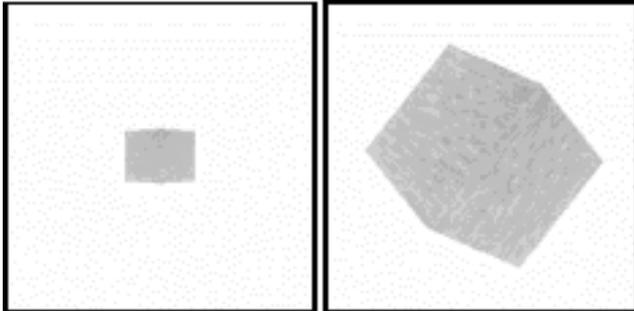


Figura 8. Imágenes 1 y 20 de un cubo con traslación en el eje z y rotación simultánea.

Ejemplo 4

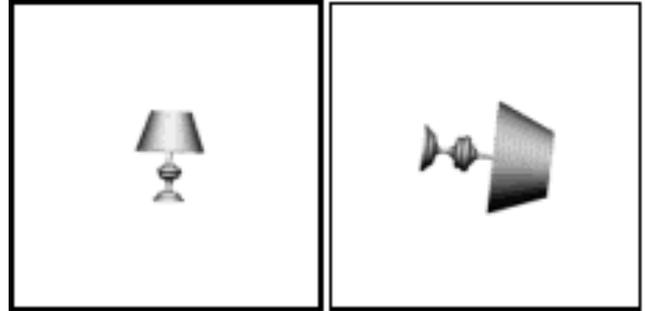


Figura 9. Imágenes 1 y 20 de una lámpara con traslación en el eje z y rotación simultánea.

Imagen	X _{obj}	Y _{obj}	Seg.H.	Seg.V.	Movimiento
1	128,5	128,5	57,0	45,0	
2	128,5	129,0	61,0	46,0	D
3	128,5	128,5	65,0	47,0	D
4	128,5	128,5	67,0	52,0	D
5	128,5	128,5	73,0	55,0	D
6	128,5	128,5	77,0	59,0	D
7	128,5	128,5	81,0	65,0	D
8	128,5	128,5	85,0	69,0	D
9	128,5	128,5	91,0	75,0	D
10	128,5	128,5	95,0	81,0	D
11	128,5	128,5	101,0	87,0	D
12	128,5	128,5	109,0	93,0	D
13	128,5	128,5	115,0	101,0	D
14	128,5	128,5	123,0	109,0	D
15	128,5	128,5	131,0	119,0	D
16	128,5	128,5	141,0	129,0	D
17	128,5	128,5	151,0	141,0	D
18	128,5	128,5	163,0	153,0	D
19	128,5	128,5	177,0	167,0	D
20	128,5	128,5	193,0	185,0	D

Tabla 3. Resultados para la secuencia del cubo.

Como era fácil de prever, los problemas iban a comenzar al incorporar movimientos rotacionales. El ejemplo 3 es una muestra de ello. En efecto, nos encontramos ante el caso de un cubo acercándose por el eje z y rotando simultáneamente. Véase que el algoritmo no arroja el resultado esperado DR, sino un simple D.

La explicación hay que buscarla en la forma del objeto. En efecto, nuestro algoritmo funciona tanto mejor para el caso de rotaciones cuanto más irregular es la forma del objeto cuyo movimiento se está analizando. Desgraciadamente, los segmentos horizontal y vertical siempre tienen el valor para esta figura.

Imagen	X _{obj}	Y _{obj}	Seg.H.	Seg.V.	Movimiento
1	128,0	128,0	55,0	77,0	
2	128,5	128,0	56,0	81,0	D
3	129,0	127,5	57,0	86,0	D
4	129,5	128,0	58,0	89,0	D
5	130,5	127,0	58,0	91,0	D
6	130,5	127,0	64,0	93,0	D
7	130,5	127,0	70,0	95,0	D
8	130,5	126,5	76,0	96,0	D
9	130,5	126,0	82,0	95,0	DR
10	130,0	126,0	87,0	93,0	DR
11	130,0	125,5	93,0	92,0	DR
12	129,5	125,5	96,0	88,0	DR
13	130,0	125,0	99,0	85,0	DR
14	129,5	125,0	102,0	81,0	DR
15	129,5	126,5	104,0	82,0	DR
16	129,0	126,5	105,0	86,0	DR
17	128,5	127,5	106,0	90,0	DR
18	128,5	128,5	106,0	92,0	DR
19	128,5	128,5	106,0	94,0	DR
20	128,5	129,0	112,0	95,0	DR

Tabla 4. Resultados para la secuencia de la lámpara.

El ejemplo 4 analiza un movimiento similar al del ejemplo 3. Aquí, no obstante, estamos ante una forma irregular. De modo que cabe esperar un mejor comportamiento de los algoritmos expuestos en este trabajo. Y, en efecto, obtenemos un buen resultado a partir del análisis de la imagen 9 de la secuencia.

La explicación de por qué las primeras imágenes no arrojan el resultado deseado está en el valor elegido en este caso para la fluctuación permitida $\alpha=0,2$ en la función de similitud del grafo de evaluación de movimiento. Un valor más pequeño permite mejorar los resultados anteriores.

Ejemplo 5

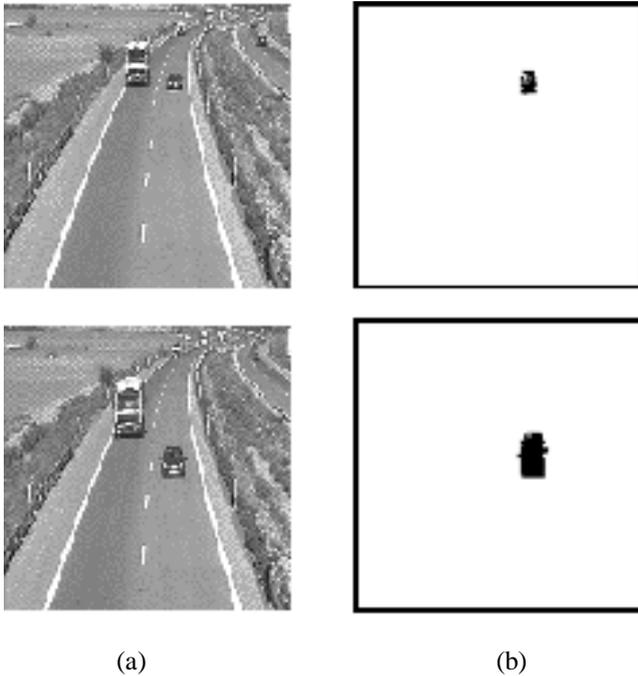


Figura 10. Imágenes 1 y 20 de una secuencia de tráfico.
(a) Las imágenes originales reales. (b) Las imágenes segmentadas para la obtención de uno de los vehículos.

Imagen	X_{obj}	Y_{obj}	Seg.H.	Seg.V.	Movimiento
1	154,5	70,0	14,0	21,0	
2	156,0	71,0	15,0	19,0	R
3	157,0	73,5	17,0	24,0	D
4	154,5	73,5	18,0	24,0	D
5	154,0	75,0	17,0	21,0	TR
6	154,5	78,5	18,0	24,0	TD
7	155,5	78,5	18,0	26,0	TD
8	155,5	80,0	18,0	29,0	TD
9	154,5	84,0	18,0	29,0	TD
10	156,5	86,0	18,0	29,0	TD
11	154,0	91,0	19,0	27,0	TD
12	153,5	92,0	20,0	29,0	TD
13	155,0	92,0	19,0	33,0	TD
14	154,0	94,5	21,0	32,0	TD
15	153,5	100,5	20,0	34,0	TD
16	153,0	101,0	23,0	35,0	TD
17	152,5	105,0	22,0	35,0	TD
18	153,0	110,5	23,0	36,0	TD
19	155,5	113,0	24,0	39,0	TD
20	153,0	118,5	27,0	40,0	TD

Tabla 5. Resultados para la secuencia real de tráfico.

Es sumamente importante poder trabajar con imágenes reales en cualquier algoritmo de reconocimiento de patrones. De ahí que mostremos en este artículo la aplicación de los algoritmos de obtención de tamaño y localización, así como los de seguimiento del movimiento a imágenes directamente obtenidas de un caso real.

La columna (a) de la figura 10 muestra la primera y última imagen de una secuencia de tráfico. La columna (b) de la misma figura ofrece la imagen resultante de aplicar los mecanismos de segmentación desarrollados en el Grupo de Investigación en Computación Neuronal e Inteligencia Artificial a partir de la tesis doctoral de Antonio Fernández-Caballero denominada “*Modelos de interacción lateral en computación acumulativa para la obtención de siluetas*”.

Véase que todos los resultados son coherentes con la situación de tráfico, salvo el primero, que ofrece una inexistente rotación. Ello es debido, seguramente, a la lejanía del vehículo en las dos primeras imágenes, por lo que dicha información es de poca utilidad.

7. Conclusiones

Se ha presentado en este artículo, en primer lugar, un método sencillo, pero eficaz, para la detección de dos parámetros importantes de un objeto en movimiento, como son el tamaño y la localización del mismo. El algoritmo es implementable en hardware, utilizando mecanismos neuronales, persiguiendo el objeto de obtención de los datos buscados en tiempo real.

Dichos parámetros son utilizados a continuación a partir de un amplio conocimiento del dominio del movimiento de objetos en secuencias de imágenes para un seguimiento eficiente de un objeto en movimiento presente en la escena analizada. Los casos de traslación y dilatación funcionan de un modo excelente, mientras que los casos en los que existe rotación pueden presentar unos peores resultados cuanto más regulares sean geoméricamente los objetos tratados.

Agradecimientos

Las imágenes del ejemplo 3 se han obtenido a partir de la secuencia sintética SOFA 3, cortesía de Computer Vision Group, Heriot-Watt University. Agradecemos el permiso de uso de las mismas, disponibles en el siguiente sitio: <http://www.cee.hw.ac.uk/~mtc/sofa>.

Agradecemos a Juan Diego Delicado la toma de las imágenes reales de tráfico del ejemplo 5.

También estamos muy agradecidos por los sustanciosos comentarios vertidos sobre este artículo por parte de los revisores del mismo.

Referencias

- Adelson E.H. & Bergen J.R.**, "Spatiotemporal energy models for the perception of motion", *Journal of the Optical Society of America, A*, Vol. 2, No. 2, 1985, pp. 284-299.
- Adiv G.**, "Determining three-dimensional motion and structure from optical flow generated by several moving objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 7, No. 4, 1985, pp. 384-401.
- Aggarwal J.K. & Nandhakumar N.**, "On the computation of motion from sequences of images -- A review", *Proceedings of the IEEE*, 1988, pp. 917-935.
- Akansu A.N. & Haddad R.A.**, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*, Academic Press Inc., 1992.
- Alevizos et al.**, *Journal of Neurophysiology*, 61, 1989, pp. 1053-1063.
- Anandan P.**, "A computational framework and an algorithm for the measurement of visual motion", *International Journal of Computer Vision*, 2, pp. 283-310.
- Barron J.L., Fleet D.J. & Beauchemin S.S.**, Performance of optical flow techniques, *International Journal of Computer Vision*, Vol. 12, No. 1, 1994, pp. 43-77.
- Black M.J. & Anandan P.**, "Estimating optical flow in segmented images using variable-order parametric models with local deformations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 10, 1996, pp. 973-986.
- Bober M. & Kittler J.**, "Robust motion analysis", en *Proceedings CVPR'94*, Seattle, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994, pp. 947-952.
- Bradshaw D.B. & Kingsbury N.G.**, "Combined affine and translational motion estimation scheme using triangulation tessellations", en *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP'97*, Munich, vol. 4, 1997, 2645-2648.
- Burt P.J. & Adelson E.H.**, "The Laplacian pyramid as a compact image code", *IEEE Transactions on Communications*, Vol. 31, No. 4, 1983.
- Calway A.D., Knutsson H. & Wilson R.**, "Multiresolution estimation of 2-D disparity using a frequency domain approach", en *Proceedings of the British Machine Vision Conference*, 1992.
- Daubechies I.**, "The wavelet transform, time-frequency localization and signal analysis", *IEEE Transactions on Information Theory*, 36, 1990, pp. 961-1005.
- Deriche R. & Faugeras O.**, "Tracking line segments", *Image and Vision Computing*, Vol. 8, No. 4, 1990, pp. 261-270.
- Dufaux F. & Moscheni F.**, "Motion estimation techniques for digital TV: A review and a new contribution", en *Proceedings of the IEEE*, 83 (6), 1995, pp. 858-876.
- Faugeras O., Lustman F. & Toscani G.**, "Motion and structure from motion from point and line matches", en *Proceedings of the 1st International Conference on Computer Vision*, 1987, pp. 25-34.
- Fernández M.A.**, "Análisis de movimiento en secuencias de imagen", en *Notas de Visión y Apuntes sobre la Ingeniería del Software*, III Curso de Verano de Informática, Colección Estudios, 24, Universidad de Castilla-La Mancha, 1993, pp. 99-110.
- Fernández-Caballero A., Lozano M.D. & Pons A.**, "La inspiración biológica de los modelos computacionales de análisis de movimiento de imágenes", en *Ensayos*, 12, Diciembre 1997, Universidad de Castilla-La Mancha, 1998.
- Fernández-Caballero A., Montoya M., Fernández M.A. & Gómez F.J.**, "Tamaño y localización de un objeto: Una aproximación neuronal", en *Memorias del IV Simposio Iberoamericano de Reconocimiento de Patrones SIARP'99*, La Habana (Cuba), pp. 443-451.
- Fernández-Caballero A., Montoya M., Fernández M.A. & Moreno J.**, "Maximum line segments for object's motion estimation", en *Proceedings of the Eighth Turkish Symposium on Artificial Intelligence and Neural Networks TAINN'99*, Estambul (Turquía), pp. 145-153.
- Fleet D.J. & Jepson A.D.**, "Computation of component image velocity from local phase information", *International Journal of Computer Vision*, 5, 1990, pp. 77-104.
- Heeger D.J.**, "Model for the extraction of image flow", *Journal of the Optical Society of America, A*, Vol. 4, No. 8, 1987, pp. 1455-1471.
- Holt R.J. & Netravali A.N.**, "Motion from optic flow: Multiplicity of solutions", *Journal of Visual Communications and Image Representation*, Vol. 4, No. 1, 1993, pp. 14-24.
- Horn B.K.P. & Schunck B.G.**, "Determining optical flow", *Artificial Intelligence*, 17, 1981, pp. 185-203.
- Huang T.S. & Netravali A.N.**, "Motion and structure from feature correspondences: A review", en *Proceedings of the IEEE*, 82, 1994, pp. 252-269.
- Jain J.R. & Jain A.K.**, "Displacement measurement and its application in interframe image coding", *IEEE Transactions on Communications*, Vol. 29, No. 12, 1981, pp. 1799-1808.

- Jepson A.D. & Jenkin M.R.M.**, "The fast computation of disparity from phase differences", en *Proceedings of the IEEE Conference on Computational Vision and Pattern Recognition*, San Diego, 1989, pp. 398-403.
- Kearney J.K., Thompson W.B. & Boley D.L.**, "Optical flow estimation: An error analysis of gradient-based methods with local optimization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 2, 1987, pp. 229-244.
- Kuglin C.D. & Hines D.C.**, "The phase correlation image alignment method", en *Proceedings of the IEEE International Conference on Cybernetics and Society*, New York, 1975, pp. 163-165.
- Liu Y. & Huang T.S.**, "A linear algorithm for determining motion and structure from line correspondences", *Computer Vision, Graphics and Image Processing: Image Understanding*, Vol. 44, No. 1, 1988, pp.35-57.
- Liu Y. & Huang T.S.**, "Motion estimation from corner correspondences", en *Proceedings of the International Conference on Image Processing, ICIP'89*, Singapore, 1989, pp. 785-790.
- Lucas B. & Kanade T.**, "An iterative image registration technique with an application to stereo vision", en *Proceedings DARPA Image Understanding Workshop*, 1981, pp. 121-130.
- Mallat S.G.**, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, 1989, pp. 674-693.
- Meyer F.G. & Bouthemy P.**, "Region-based tracking using affine motion models in long image sequences", *Computer Vision, Graphics and Image Processing: Image Understanding*, Vol. 60, No. 2, 1994, pp. 119-140.
- Mitiche A. & Bouthemy P.**, "Computation and analysis of image motion: A synopsis of current problems and methods", *International Journal of Computer Vision*, Vol. 19, Issue 1, 1996, pp. 29-55.
- Mutch K.**, "Determining object translation information using stereoscopic motion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, 1986, pp. 750-755.
- Mutch K. & Heiny L.C.**, "Calculating object size from stereoscopic motion", en *Proceedings, CVPR86. IEEE Publ. 86CH2290-5*, 1986, pp. 183-187.
- Nakaya Y. & Harashima H.**, "Motion compensation based on spatial transformations", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4, No. 3, 1994, pp. 339-356.
- Papoulis A.**, *Signal Analysis*, McGraw-Hill, 1984.
- Roach J.W. & Aggarwal J.K.**, "Determining the movement of objects from a sequence of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 6, 1980, pp. 554-562.
- Robbins J.D. & Netravali A.N.**, "Recursive motion compensation: A review", en T.S. Huang, ed., *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, 1983, pp. 76-103.
- Safonova, T.A., Zhuravlev, V.L. & Kodirov, S.A.**, "Multimodal and Multifunctional Neurons of Simple Nervous Networks" en *Toward a Science of Consciousness*, Tucson II Conference, 1996.
- Seferidis V. & Ghanbari M.**, "General approach to block-matching motion estimation", *Optical Engineering*, Vol. 32 No. 7, 1993, pp. 1464-1474.
- Spitzer F.**, "Markov random fields and Gibbs ensembles", *American Mathematical Monthly*, 78, 1971, pp. 142-154.
- Ullman S.**, *The Interpretation of Visual Motion*, The MIT Press, 1979.
- Verri A. & Poggio T.**, "Motion field and optical flow: Qualitative properties", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 5, 1989, pp. 490-498.
- Wang J.Y.A. & Adelson E.H.**, "Representing moving images with layers", *IEEE Transactions on Image Processing*, Vol. 3, No 5, 1994, pp. 625-638.
- Weng J., Ahuja N. & Huang T.S.**, "Motion and structure from line correspondences: Closed-form solution, uniqueness and optimization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 3, 1992, pp. 318-336.
- Wilson R. & Knutsson H.**, "A multiresolution stereopsis algorithm based on the Gabor representation", en *Proceedings of the IEE International Conference on Image Processing and its Applications*, Warwick, 1989, pp. 19-22.
- Wilson R., Calway A.D. & Pearson E.R.S.**, "A generalized wavelet transform for Fourier analysis: The multiresolution Fourier transform and its application to image and audio signal analysis", *IEEE Transactions on Information Theory*, 1992, Vol. 38, No. 2.



Antonio Fernández Caballero es Licenciado en Informática por la Universidad Politécnica de Madrid (España) en 1993. Ha ejercido como consultor en empresas públicas y privadas durante 10 años. Es Ayudante de Facultad en el Departamento de Informática de la Universidad de Castilla-La Mancha (España). En la actualidad está realizando su tesis doctoral en el Departamento de Inteligencia Artificial de la Universidad Nacional de Educación a Distancia (España). Pertenece al Grupo de Investigación en Computación Neuronal e Inteligencia Artificial (GICoNIA).



Modesto Montoya Santos es Ingeniero Técnico en Informática de Sistemas por la Universidad de Castilla-La Mancha (España) en 2000. Pertenece al Grupo de Investigación en Computación Neuronal e Inteligencia Artificial (GICoNIA).



Miguel Angel Fernández Graciani es Doctor en Ciencias Físicas por la Universidad Nacional de Educación a Distancia (España) en 1996 y Licenciado en Ciencias Físicas por la Universidad de Granada (España) en 1987. Ingeniero de diseño en la Empresa Nacional de Electrónica y Sistemas INISEL desde 1987 hasta 1991. Profesor Titular de Escuela Universitaria de la Universidad de Castilla-La Mancha (España) desde 1991. Director del Grupo de Investigación en Computación Neuronal e Inteligencia Artificial (GICoNIA) del Departamento de Informática de la UCLM.



Francisco Javier Gómez Quesada es Licenciado en Informática por la Universidad de Granada (España) en 1991. Profesor Titular de Escuela Universitaria de la Universidad de Castilla-La Mancha (España) desde 1991. Pertenece al Grupo de Investigación en Computación Neuronal e Inteligencia Artificial (GICoNIA). Desde Abril de 1998 es el Coordinador Responsable de Prácticas de Alumnos en Empresas de la Escuela Politécnica Superior de Albacete.