

# LA INSPIRACION BIOLOGICA DE LOS MODELOS COMPUTACIONALES DE ANALISIS EN MOVIMIENTO DE IMAGENES

Antonio Fernández Caballero  
María Dolores Lozano Pérez  
Amparo Pons Penedés\*

**Resumen.** Este artículo presenta la inspiración biológica de algunos de los modelos computacionales más conocidos de análisis en movimiento de imágenes. Tras una breve introducción al estado actual del conocimiento fisiológico del análisis de movimiento, se procede a comentar la validez de dichos modelos computacionales.

## 1. UNA DESCRIPCION FISIOLÓGICA DEL ANALISIS DEL MOVIMIENTO.

Los detectores del sistema visual miran solo a una pequeña parte del campo visual. Surge un problema al asignar la velocidad verdadera de un objeto a partir de medidas locales, ya que el movimiento de un único segmento extendido de línea no determina el movimiento de un objeto que contenga ese segmento de línea [Fen79, Hor81, Mar81, Ade82, Hil84, Wal76]. El movimiento paralelo a la línea es invisible. De este modo, una familia de posibles movimientos pueden ser el resultado del movimiento detectado. El problema de la apertura se reduce a la asignación de la velocidad verdadera de todo el modelo a los elementos de una imagen en movimiento en aquellos puntos locales con posibles ambigüedades. Diremos, pues, que el movimiento del componente local o 1-D es el movimiento aparente de líneas extendidas 1-D, mientras

---

\*Antonio Fernández Caballero es profesor del Departamento de Informática, Escuela Politécnica de Albacete, Universidad de Castilla-La Mancha. María Dolores Lozano Pérez y Amparo Pons Penedés son colaboradoras en proyectos de investigación del Departamento.

que el movimiento de la forma o 2-D es el movimiento real de los elementos.

La solución al problema de la apertura se resuelve si existen al menos dos medidas de movimiento de componentes locales que lleven a la estimación de la velocidad de la forma en un punto. En un movimiento simple como la traslación en un plano, el problema está ampliamente resuelto. En efecto, al ser la velocidad 2-D la misma en todo el móvil, hay, en la gran mayoría de los casos, más de dos medidas de componentes locales para estimar la velocidad 2-D. No es éste el caso, sin embargo, para el movimiento 3-D y el movimiento rotacional, en el que la velocidad real 2-D varía de punto a punto. Por ello, la medida del movimiento 3-D es ambigua [Eme81, Tik77] y son necesarias unas restricciones adicionales para encontrar una solución única. Los datos neurofisiológicos y psicofísicos disponibles hasta la fecha sugieren que existen varios niveles en el análisis de movimiento en el sistema visual. La figura 1 presenta una descripción simplificada de estos niveles de proceso.

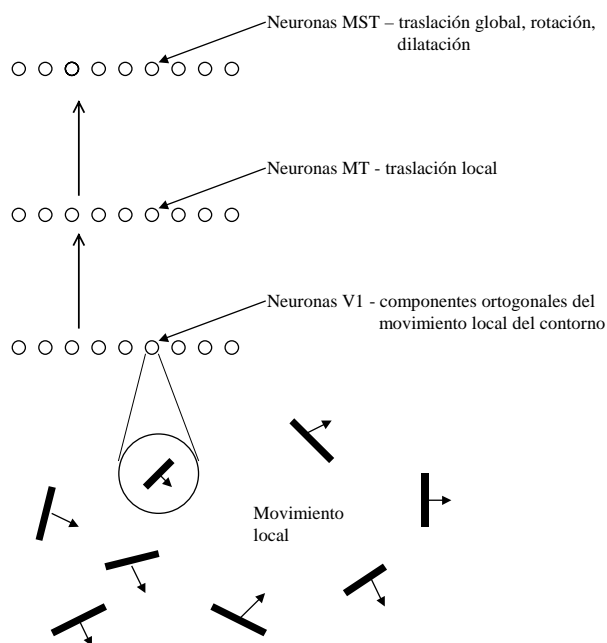


Figura 1. Representación esquemática de los pasos en el análisis del movimiento en la corteza visual primaria.

En primer lugar, se ve que el problema de la apertura para el movimiento traslacional plano queda resuelto en dos niveles. En el primer nivel las medidas de movimiento local extraen los componentes del movimiento que se encuentran en la dirección perpendicular a los elementos en la imagen. El segundo nivel combina las medidas de movimiento local de porciones de la imagen con el fin de calcular un número menor de estimaciones locales de traslación del móvil.

Finalmente, un tercer nivel integra las estimaciones locales de movimiento traslacional para calcular movimientos no locales más complejos (por ejemplo, rotaciones globales). De este modo, en cada nivel, la información de movimiento localizado espacialmente parece estar combinado en un área para calcular menos movimientos locales, pero más movimientos complejos.

Una característica fundamental de la arquitectura cortical de los sistemas sensoriales es la presencia de múltiples mapas topográficos proyectados unos sobre los otros. En el sistema visual, aproximadamente veinticinco de estos mapas están ordenados en una jerarquía de cinco a seis bandas definidas por el modelo de interconexiones hacia adelante y hacia atrás.



para el proceso de importantes características no locales de la visión, como pueden ser la traslación de las formas, la rotación, la dilatación, la obtención de la figura a partir de las sombras, la obtención de la estructura a partir del movimiento. La evidencia neurofisiológica de los estudios sugieren que los tres pasos del análisis de movimiento descritos anteriormente están representadas en al menos tres áreas corticales diferentes que están jerárquicamente unidas unas a otras a través de vías directas e indirectas. Así han sido descubiertas algunas células en las tres grandes áreas (V1, MT y MST) que responden al movimiento del componente (área V1), al movimiento de la forma (área MT) y a los movimientos globales (área MST).

## **2. MODELOS COMPUTACIONALES E INSPIRACION BIOLOGICA.**

Se pasa a comentar, pues, algunos de los más conocidos modelos computacionales de detección inicial de movimiento y de análisis del movimiento en secuencias de imágenes. Básicamente podemos partir de dos grandes enfoques al problema que nos dan las dos orientaciones genéricas más conocidas:

- a) los modelos basados en la detección local de movimiento
- b) los modelos basados en la atención selectiva visual

### **2.1. Modelos de detección local de movimiento.**

Se han propuesto algunos modelos para la detección inicial de movimiento. Estos modelos se dividen, a grandes rasgos, en dos clases:

- (1) los modelos basados en el análisis del gradiente espacio-temporal, que calculan la velocidad a partir de la variación de la intensidad en el espacio y en el tiempo producida por el movimiento de una imagen
- (2) los modelos cuya estructura básica está inspirada en los primeros trabajos acerca del sistema visual del escarabajo y de la mosca, denominado modelo correlacional [Has56], y de la retina del conejo [Bar65]

Los modelos basados en el análisis del gradiente calculan la velocidad tomando la derivada temporal de la intensidad en un punto dividida por la derivada espacial de la intensidad en ese mismo punto

$$V_x = - \frac{dI}{dt} / \frac{dI}{dx}$$

donde  $V_x$  representa la velocidad en la dirección  $x$  [Fen79, Hor81, Mar81]. Estos modelos permiten la determinación punto a punto de la disparidad basándose en criterios puramente locales sin la necesidad de examinar largas secuencias de imágenes.

Sin lugar a dudas, los modelos basados en la obtención del gradiente espacio-temporal, que miden el movimiento a partir de la distribución de la luminiscencia pura [Fen79, Hor81] son modelos biológicamente defendibles. Al menos más que aquellos que procesan en primer lugar la imagen con filtros paso banda de frecuencia espacial orientada o no orientada [Mar81, Law89].

Los modelos que incluyen filtros de frecuencias espaciales miden la velocidad en lugares donde los cambios de intensidad son significativos (por ejemplo, en los bordes). Al contrario, los modelos sin filtros también pueden calcular la velocidad a partir de gradientes lineales de intensidad. La evidencia psicofísica y fisiológica, no obstante,

indica que el sistema visual es bastante insensible a los gradientes lineales de intensidad - los humanos no perciben bien los gradientes de intensidad, ya sean estacionarios [McC74] o móviles [Nak83]. Así pues, las neuronas en el sistema visual (véase LGN y el área V1) no responden bien a gradientes lineales de intensidad o a niveles absolutos de luminiscencia.

Además, un sistema de movimiento que opera en valores de intensidad pura es considerablemente más susceptible de confundir factores tales como los cambios globales en el nivel de iluminación, que son, a su vez, causantes de los cambios locales en la intensidad. De este modo, los cambios en la iluminación ambiental de las superficies pueden ser incorrectamente interpretados como movimientos de las propias superficies. Contrariamente a ésto, los gradientes de intensidad pronunciados en los bordes de los objetos son mucho más estables bajo estas condiciones, así como bajo condiciones de ruido en general.

El segundo tipo de modelos, aquellos basados en la correlación, calcula la dirección a partir de las correlaciones en las actividades de los campos receptivos de entrada vecinos. Los cambios en un campo receptivo de entrada son correlacionados con los cambios en los campos receptivos vecinos después de un intervalo de tiempo, claro está, siempre que el movimiento sea en la dirección de esos vecinos. Estos modelos comparten algunas propiedades organizativas comunes:

- (1) un par de campos receptivos de entrada de selección de la posición espacial, y muchas veces también la frecuencia espacial de un estímulo
- (2) una combinación de señales de los dos campos receptivos, de modo que

se obtiene una respuesta direccional selectiva

(3) una o varias etapas de integración espacial y/o temporal

Todos estos modelos calculan la dirección, pero no calculan explícitamente la velocidad [Has56, Bar65, Ade85, Wat85]. Estos modelos recientes de detección de movimiento son notables en su intento de construir detectores de movimiento a partir de las propiedades básicas de filtrado espacial y temporal de las neuronas. Sin embargo, queda aún un largo camino por recorrer hasta relacionar estos modelos más íntimamente con los componentes de la circuitería neuronal real. Muchos de estos modelos necesitan de cinco a seis pasos jerárquicos, multiplicando, sumando, restando, elevando al cuadrado, y sumando las entradas iniciales.

Una implementación neurofisiológica aceptable de estos pasos no es inmediatamente evidente. En la corteza primaria se encuentra un alto porcentaje de las células de selección direccional en la capa 4B, que es la que provee de la mayor proyección al área MT, un importante área para el análisis del movimiento. Las neuronas en la capa 4B, a su vez, reciben su entrada de las neuronas de la capa 4C $\alpha$ . Así, los cinco o seis pasos propuestos deberían ser implementados entre estas dos capas. Además, estos modelos proponen que un par de campos receptivos de células simples proporcionan la entrada al detector de movimiento. Esta suposición es cuestionable, al existir informes conflictivos de si la entrada de la capa 4C $\alpha$  a las neuronas de selección de movimiento de la capa 4B consiste en neuronas de selección de orientación [Bul80, Haw84] o en neuronas que no seleccionan la orientación [Bla84, Hub77].

Una segunda consecuencia de tener células simples de selección no direccional



alimentando a un detector de movimiento de célula compleja de selección direccional está en la presunción subyacente que las células simples de selección direccional no juegan ningún papel en la percepción y la detección del movimiento. Muchas de las neuronas de la capa 4B podrían ser células simples de selección de la dirección.

En tercer lugar, los modelos postulan que solamente dos campos receptivos (y no un número mayor) sirven como entrada al detector de movimiento, sugiriendo que la activación óptima de un detector puede ser producida por un par de estímulos idénticos, en los cuales la presentación del segundo estímulo se desplaza ligeramente en el espacio y en el tiempo respecto de la presentación del primer estímulo. La evidencia psicofísica contradice esta predicción. Algunos estudios, utilizando una variedad de paradigmas, indican que la presentación de múltiples tramas de un estímulo de movimiento aparente puede:

- (1) incrementar en gran medida la probabilidad de detectar el movimiento en manifestaciones de puntos ocasionales [Lap82]
- (2) incrementar la distancia máxima por desplazamiento para la percepción correcta de la dirección del movimiento en un estímulo de movimiento aparente de rango corto y de puntos ocasionales - una secuencia de dos desplazamientos incrementaron grandemente (50 %) el desplazamiento máximo aceptable por paso comparado con un desplazamiento [Nak84]
- (3) incrementar el rendimiento de la discriminación de la velocidad - una secuencia de cinco a ocho presentaciones de un estímulo de movimiento aparente (un corte de luz) fué necesario para igualar el rendimiento de discriminación de la velocidad obtenido con blancos en movimiento

continuo [McK85]

Otra suposición crucial de los modelos está en que los filtros son lineales. Para activar las unidades de entrada, algunos modelos toman una transformada de Fourier del modelo, que la convierte del dominio espacio-tiempo al dominio de frecuencias espacio-temporales. Las neuronas poseen propiedades no lineales intrínsecas como la saturación y la rectificación. Así, al contrario que los modelos propuestos, las neuronas reales no serían capaces de descomponer un estímulo de movimiento complejo en sus componentes de frecuencias espacio-temporales.

Finalmente, una característica que falta en estos modelos es la inhibición. Esta sección concluye con la discusión de la evidencia de un mecanismo inhibitorio en la detección del movimiento, así como con un modelo que incorpora esta característica. Se ha demostrado por medios fisiológicos [Bar65, Eme81] y farmacológicos [Sil77] que la inhibición desempeña un papel crucial en la obtención de la selección direccional.

## **2.2. Modelos basados en la atención selectiva visual.**

Estos modelos pueden ser agrupados naturalmente entre una de varias clases principales. Antes de discutir estas clases, se da una lista parcial de criterios utilizados para evaluar la bondad de estos modelos. Algunos temas importantes a tener en cuenta son:

- (1) el tipo de movimiento que el modelo soporta (por ejemplo, traslacional plano y/o movimiento general 3-D)

- (2) cómo el modelo se enfrenta a varios objetos en movimiento
- (3) la habilidad del modelo para producir varios fenómenos perceptuales - incluyendo, por ejemplo, el movimiento no rígido percibido de algunas figuras rígidas en movimiento, efectos coherentes (por ejemplo, captura de movimiento y transparencia)
- (4) la validez fisiológica y perceptual de las presunciones tomadas, los algoritmos usados y los resultados obtenidos

Las clases de modelos discutidos incluyen

- (1) modelos que incorporan líneas de restricciones, también denominados de obtención de clases
- (2) modelos de correspondencia o "matching", que seleccionan y correlacionan características resaltadas de las imágenes en secuencias de imágenes estudiadas en intervalos cortos de tiempo
- (3) modelos que utilizan una restricción de uniformidad, que cuando se aplica a las imágenes producidas por el movimiento 3-D, produce un campo velocidad que varía suavemente de punto a punto
- (4) otros modelos que no caen en los anteriores tipos

En una primera clase de modelos, los modelos de obtención de clases ("clustering"), se utilizan restricciones múltiples para el movimiento en el espacio (extraídas a partir de estimaciones locales de la velocidad) para obtener la velocidad de la forma [Fen79].

En la segunda clase de modelos, o sea, los esquemas de correlación y de correspondencia, se mide el movimiento a partir de secuencias discretas de imágenes, identificando e igualando las características en las imágenes correspondientes a lo largo de tramas sucesivas [Tho81, Pra83]. El problema de la apertura no aparece en estos esquemas de imágenes. Las dos clases principales de técnicas de correlación incluyen las técnicas basadas en la correlación cruzada de las intensidades y las técnicas que correlacionan rasgos o características simbólicas.

Las técnicas de correlación simbólica definen estructuras geométricas a ser correlacionadas en la imagen. Al restringir los candidatos a la correlación para las estructuras de imágenes distintas, más que la intensidad de los pixels, se obtiene la ventaja de que estas estructuras muestran una mayor estabilidad a lo largo de las tramas de imágenes y que necesitan menos correlaciones (computacionalmente son menos caros). Los esquemas de correlación de rasgos correlacionan puntos característicos (por ejemplo, terminaciones o puntos de alta curvatura), informaciones de los bordes o combinaciones más complejas de estructuras geométricas de bordes o áreas de la imagen.

En general, es más fácil trabajar con rasgos más simples, pero hay más probabilidad de falsas correlaciones. Las características más complejas son más difíciles de procesar, pero es más fácil encontrar la correspondencia correcta. Sin embargo, ya que la forma de los rasgos complejos cambia en los movimientos complejos, estos esquemas funcionan mejor en movimientos rígidos planos. La mayoría de las evidencias psicofísicas y fisiológicas sugieren que el movimiento no se mide a través de procesos simples de correlación.

El tercer tipo de modelos utiliza una restricción de uniformidad para calcular el campo velocidad producido por un movimiento general 3-D - la proyección del movimiento 2-D de objetos expuestos a rotación o traslación en profundidad [Hor81, Hil84, Koc86, Yui88]. El argumento para la restricción de uniformidad es que, ya que los objetos normalmente tienen superficies uniformes (superficies que cambian gradualmente en su profundidad), el campo velocidad producido por un objeto en movimiento también variará de modo suave.

Puede calcularse el campo velocidad más uniforme obteniendo el campo velocidad de menor variación. La uniformidad del campo velocidad puede medirse tomando las primeras derivadas o derivadas de orden superior de la velocidad a lo largo de las áreas de la imagen o a lo largo de los contornos.

Un problema aparece en este algoritmo de uniformidad cuando una imagen contiene más de un objeto en movimiento, lo que causa discontinuidades abruptas en el campo de flujo. El algoritmo de uniformidad en tipos heterogéneos de movimientos conduce a errores sustanciales en la estimación de la velocidad.

Los algoritmos basados en restricciones de uniformidad son problemáticos debido a su falta de plausibilidad fisiológica. Por ser estos algoritmos iterativos y cooperativos, parece improbable que pudieran procesar lo suficientemente rápido como para ser implementados en un sistema neuronal real. Además, estos algoritmos calculan derivadas espaciales de la velocidad que requieren de la comparación explícita punto a punto de las velocidades adyacentes. Los datos fisiológicos y psicofísicos, sin embargo,

sugieren que la información de la velocidad de los puntos adyacentes en el espacio no pueden ser comparados directamente, por causa de la relativamente mala resolución espacial del sistema de movimiento. Al contrario, parece que las señales de velocidad son integradas en el espacio antes de ser ejecutadas las comparaciones.

Otros modelos son difíciles de catalogar en los vistos hasta el momento.

Los modelos energéticos de movimiento [Hee87] son atractivos en su uso de filtros espacio-temporales que les permiten detectar el movimiento directamente a partir de las secuencias de imágenes. Mientras que está claro que los detectores de movimiento en los diversos niveles del sistema visual son sensibles a los aspectos espaciales y temporales de las imágenes en movimiento, no está tan claro que efectúen una descomposición exacta en frecuencias espaciotemporales, tal como hacen estos modelos. La falta de un conjunto base adecuado de filtros espaciales y temporales, así como la no linealidad inherente en la respuesta neuronal puede estar en el origen de esta falta de descomposición [Val88].

## **Bibliografía.**

- [Ade82] E.H. Adelson & J.A. Movshon. (1982). Phenomenal coherence of moving visual patterns. Nature 300, 523-525.
- [Ade85] E.H. Adelson & J.R. Bergen. (1985). Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America A, 2, 284-299.
- [Bar65] H.B. Barlow & R.W. Levick. (1965). The mechanism of directional selectivity in the rabbit's retina. Journal of Physiology 173, 477-504.
- [Bla84] G.G. Blasdel & D. Fitzpatrick. (1984). Physiological organization of layer 4 in macaque striate cortex. Journal of Neuroscience 46, 880-895.
- [Bul80] J. Bullier & G.H. Henry. (1980). Ordinal position and afferent input of neurons in monkey striate cortex. Journal of Comparative Neurology 193, 913-935.

- [Eme81] R.C. Emerson & L. Coleman. (1981). Does image movement have a special nature for neurons in the cat's striate cortex? Investigative Ophthalmology and Visual Science 20, 766-783.
- [Fen79] C.L. Fennema & W.B. Thompson. (1979). Velocity determination in scenes containing several multiple moving objects. Computer Graphics and Image Processing 9, 301-315.
- [Has56] B. Hassenstein & W.E. Reichardt. (1956). Functional structure of a mechanism of perception of optical movement. Proceedings of the 1st International Congress of Cybernetics in Namar, 797-801.
- [Haw84] M.J. Hawken & A.J. Parker. (1984). Contrast sensitivity and orientation selectivity in lamina IV of the striate cortex of old world monkeys. Experimental Brain Research 54, 367-372
- [Hee87] D.J. Heeger. (1987). Model for the extraction of image flow. Journal of the Optical Society of America A, 4, 1455-1471.
- [Hil84] E.C. Hildreth. (1984). The measurement of visual motion. Cambridge. Ed. MIT Press.
- [Hor81] B.K.P. Horn & B.G. Schunck. (1981). Determining optical flow. Artificial Intelligence 17, 185-203.
- [Hub77] D.H. Hubel & T.N. Wiesel. (1977). Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London, B, 198, 1-59.
- [Koc86] C. Koch, J. Marroquin & A. Yuille. (1986). Analog "neuronal" networks in early vision. Proceedings of the National Academy of Sciences USA 83, 4263-4267.
- [Lap82] J.S. Lappin & M. Fuqua. (1982). Non-linear recruitment in the visual detection of moving patterns. Investigative Ophthalmology and Visual Science Supplement 22, 123.
- [Law89] T.B. Lawton. (1989). Outputs of paired Gabor filters summed across the background frame of reference predict the direction of movement. IEEE Transactions of Biomedical Engineering, 36, 130-139.
- [Liv88] M. Livingstone & D. Hubel. (1988). Segregation of form, color, movement and depth: anatomy, physiology and perception. Science, 240, 740-749.
- [Mar81] D. Marr & S. Ullman. (1981). Directional selectivity and its use in early visual processing. Proceedings of the Royal Society (London) B, 211, 151-180.
- [McC74] J.J. McCann, R.L. Savoy, J.A. Hall & J.J. Scarpetti. (1974). Visibility of continuous luminance gradients. Vision Research 14, 917-927.
- [McK85] S.P. McKee & L. Welch. (1985). Sequential recruitment in the discrimination of velocity. Journal of the Optical Society of America A, 2, 243-251.

- [Nak83] K. Nakayama & G.H. Silverman. (1983). Perception of moving sinusoidal lines. Journal of the Optical Society of America A, 72.
- [Nak84] K. Nakayama & G.H. Silverman. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. Vision Research 24, 293-299.
- [Pra83] J.M. Prager & M.A. Arbib. (1983). Computing the optic flow: The MATCH algorithm and prediction. Computer Vision, Graphics and Image Processing, 24, 271-304.
- [Sil77] A.M. Sillito. (1977). Inhibitory processes underlying the directional specificity of simple, complex and hypercomplex cells in the cat's visual cortex. Journal of Physiology 271, 699-720.
- [Tho81] W.B. Thompson & S.T. Barnard. (1981). Lower-level estimation and interpretation of visual motion. IEEE Computer 14, 20-28.
- [Tik77] A.N. Tikhonov & V.Y. Arsenin. (1977). Solutions of ill-posed problems. Washington D.C. Ed. Winston.
- [Val88] R.L. De Valois & K.K. De Valois. (1988). Spatial Vision. New York. Ed. Oxford University Press.
- [Wal76] H. Wallach. (1976). On perceived identity: 1. The direction of motion of straight lines. In H.Wallach (ed.). On perception. New York. Ed. Quadrangle.
- [Wat85] A.B. Watson & A.J. Ahumada. (1985). Model of visual-motion sensing. Journal of the Optical Society of America A, 2, 322-341.
- [Yui88] A.L. Yuille & N. Grzywacz. (1988). A computational theory for the perception of coherent visual motion. Nature 333, 71-74.