La percepción del movimiento: parte de lo que pasa por tu cabeza en unos milisegundos

Jesús Malo
Profesor Titular de la Universitat de València

1.- Un cortometraje poco original: chico encuentra a chica

Un bar. Dos sujetos. Ella (sujeto 1) está sola en una mesa. Café y tabaco. Una excusa demasiado obvia, piensa él (sujeto 2), pero aun así, saca un cigarrillo... Perdona, ¿tienes fuego? dice él mientras levanta el cigarrillo. Sí, contesta ella dándose la vuelta desde la otra mesa... Aquí está... toma! dice ella cogiendo un mechero y haciendo una oscilación de la mano que es también una pregunta. Él mueve afirmativamente la cabeza indicando que está preparado para el lanzamiento. Ella repite el gesto de nuevo, pero ahora, en el instante preciso, sus dedos dejan escapar el mechero, y éste inicia la clásica parábola. Voilá. Él mira la evolución del mechero durante unos 50 milisegundos, el escote de la chica vuelta hacia él durante otros 100 milisegundos, y la cara de la chica durante los 300 milisegundos siguientes... por último, cuando ya sospecha que no tiene nada que hacer con ella, vuelve a mirar el mechero en la última parte de la trayectoria. Fin.

En vez de abordar lo realmente interesante, este texto sólo va de lo que ocurre en ciertas regiones del córtex visual de los primates 1 y 2 durante los primeros 50 ms del vuelo del mechero: en ese intervalo, ciertas neuronas de V1 y MT (de ambos sujetos) se excitan de tal manera que ellos son capaces de calcular (o dicho de otro modo, interpretar o percibir) la velocidad del mechero. Y ocurre lo mismo en cualquier observador que mire el film.

Decepcionada? Toma y yo. A mí también me interesa más la parte del escote y la interpretación actoral de los otros 400 ms del vuelo del mechero, pero la neurociencia computacional (como el fúrbol) es así: aun no nos dice gran cosa de lo realmente interesante... como dicen los críticos, para el neurocientífico, a este film le sobra metraje.

Si, a pesar de lo restringido de la intención, aun quieres seguir leyendo, que sepas que este texto tiene tres partes: primero veremos el efecto del movimiento de una escena en una secuencia de imágenes (el estímulo), luego veremos cómo las neuronas del córtex visual pueden extraer la información de movimiento a partir de los datos de la secuencia (la percepción de velocidad) y finalmente veremos un ejemplo de lo anterior simulando parte de lo que, lo creas o no, pasa en tu cabeza en 5 ó 10 milisegundos de visionado de un interesantísimo film como éste.

2.- Una superproducción de 3 fotogramas (el estímulo!)

Toda percepción se inicia con un estímulo. Movimiento... cinemática... cinematógrafo... imágenes en movimiento... En el caso de la percepción de movimiento, el estímulo es una secuencia de imágenes.

La óptica nos dice que una escena (como el mechero volador de nuestro cortometraje) genera una distribución de energía en el plano imagen de un sistema de formación de imágenes (como las retinas de los sujetos de nuestro corto, o el plano CCD de la cámara con la que estamos grabando nuestro corto). En definitiva, los ojos que llevamos en la cabeza no son más que dos camaritas que generan sendas imágenes del exterior en nuestras retinas. Una imagen es una distribución de energía en un plano: en un punto cualquiera, $\mathbf{x}_1$, del plano (en un cierto píxel -picture element-) tenemos una cierta cantidad de energía, $E(\mathbf{x}_1)$, y en otro punto (otro pixel), $\mathbf{x}_2$, tendremos otra cantidad de energía diferente, $E(\mathbf{x}_2)$. Si la escena está en movimiento, la distribución de energía, E, no solo depende del espacio, sino que también depende del tiempo: $E(\mathbf{x},t)$, es decir, la energía en un cierto punto del plano imagen cambiará con el tiempo (si la peli no es un muermo).

Ninguno de los hermanos Lumiere (vaya nombre más apropiado!) diría que una película es una función de energía, E, definida en un dominio de tres dimensiones (espacio $\mathbf{x}$, y tiempo t), pero tu profesor de matemáticas sí lo diría. Y como él también te dijo, el movimiento (la velocidad) es una rela-

ción entre el espacio recorrido por un objeto, Δ**x**, y el tiempo que tarda en hacerlo, Δt...

Vamos a verlo: consideremos una interesantísima película (sin duda independiente) que narra el desplazamiento de un mechero (figura 1). A pesar de que esta superproducción sólo tiene 3 fotogramas, algunos críticos han reconocido los elementos básicos de la tragedia griega: planteamiento-nudo-desenlace...

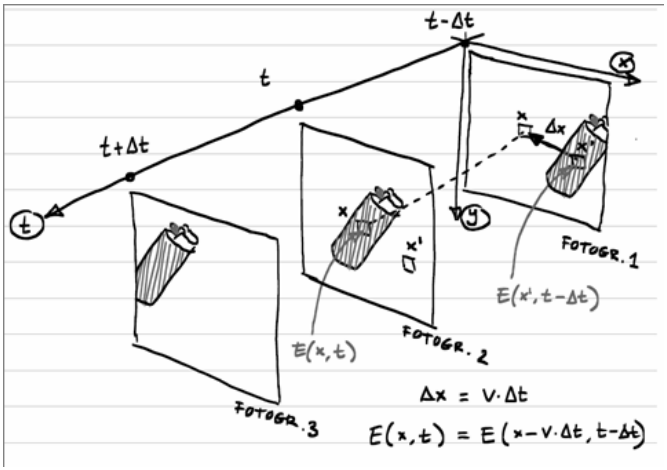


Fig.1: Fotogramas del estímulo (irradiancia en cada posición e instante).

Como el mechero (o más bien su proyección sobre el plano imagen) se mueve con velocidad, **v**, resulta que la energía en un cierto píxel, **x**, en el instante, t, es igual a la energía en un instante anterior, t' = t - Δt, en un píxel diferente, **x**' = **x** - v. t, es decir (véase la figura 1):

$$E(\mathbf{x}, t) = E(\mathbf{x} - \mathbf{v}\,\Delta t,\ t - \Delta t) \qquad (1)$$

Nótese que si el mechero estuviese en reposo (film conceptual donde los haya), **v**=0. En ese caso, tendríamos E(**x**,t+ Δt) = E(**x**,t) para todos los píxels, **x**, y todos los intervalos temporales Δt. Un auténtico tostón con todos los fotogramas iguales, o *movie portrait* que diría Warhol.

El flujo óptico de una secuencia, **v**, es el campo de vectores velocidad que relaciona los valores de las irradiancias en el plano imagen en instantes diferentes de tiempo. La figura 2 muestra los valores de este campo en el cortometraje (extra-corto) de la figura 1.

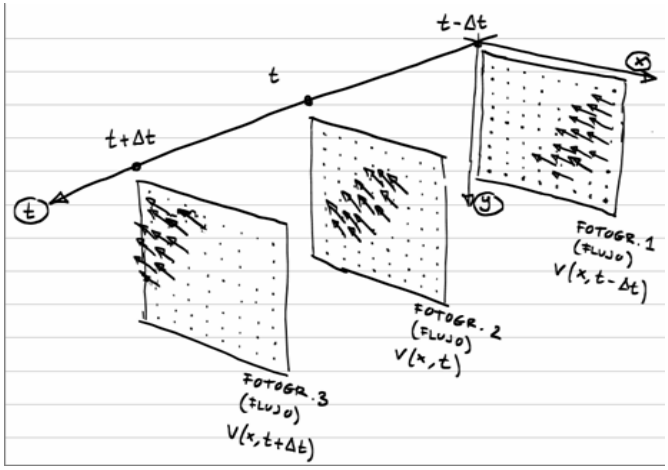


Fig. 2: Fotogramas de flujo óptico (velocidad en cada posición e instante de tiempo)

La visión artificial y la neurociencia dicen que la percepción más elemental del movimiento consiste en el cálculo de estas velocidades, **v**, (el flujo óptico) a partir de los valores del estímulo (la secuencia de imágenes). Conocer estas velocidades nos permite distinguir los objetos en movimiento de aquellos que no lo están, tener sensación de profundidad y hacer inferencias sobre la trayectoria que seguirán los objetos móviles... todo ello muy útil si queremos cruzar la calle o pescar el mechero que nos lanzan desde la mesa de al lado. De puta madre. Sin embargo, ¿cómo se despeja **v** en la ecuación 1?

3.- Sexo, mentiras y cintas de vídeo [1] (excitación neuronal, espectro y movimiento)

Para entender como las regiones V1 y MT de nuestro córtex hacen esa cuenta (perciben esas velocidades), son necesarios dos elementos: (1) hace falta saber que nuestras neuronas (visuales) son sensores sensibles a ciertas frecuencias espacio-

temporales de los estímulos y (2) es necesario saber que las secuencias en movimiento tienen una composición frecuencial muy particular, que está relacionada con la velocidad de los objetos que se mueven. Ambas cosas implican el concepto de representación frecuencial, ya sea para caracterizar la sensibilidad de un sensor (sistema) o para caracterizar el contenido energético de un estímulo (señal).

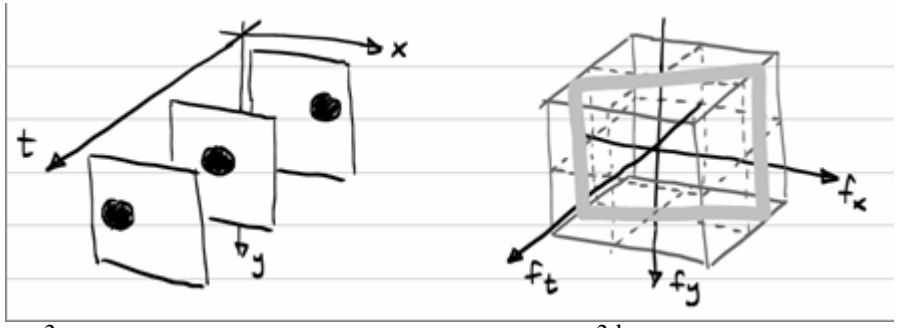
La representación frecuencial de señales y sistemas proviene del matemático francés Joseph Fourier [2] que mal vivió en el siglo XVIII. No obstante, aún interpretamos gran parte de lo que pasa en el mundo gracias a sus contribuciones. Por ejemplo, cuando el Nen de Castefa [3] lleva el loro del buga a toda hostia, le mola mirar las *rayitas* del ecualizador dale que te pego arriba y abajo: tung-quish-tung-quish... Eso es una representación frecuencial (variable con el tiempo) de las ondas de presión que llamamos *música*. En el caso de la música, el estímulo es una función E(t), es decir, una determinada energía (presión acústica) en cada instante de tiempo. No obstante, resulta más intuitiva una representación en frecuencias temporales, e(ft). Cada nota, cada tono, corresponde a la presencia de energía en una determinada frecuencia temporal. La relación matemática entre E(t), la música, y e(ft), su espectro, es una *Transformada de Fourier*. Las *rayitas* que bailan en la pantalla del ecualizador gráfico son el espectro de pequeños fragmentos de la música que estamos escuchando. Como en cada fragmento hay notas diferentes el espectro va cambiando con el tiempo.

Una señal más complicada, por ejemplo el estímulo que nos interesa (la película E(**x**,t)), también puede ser representada frecuencialmente, es decir, también puede ser descompuesta en notas. Como mostraron Watson y Ahumada en un trabajo de 1985 [4], la descomposición frecuencial de una secuencia de imágenes es muy interesante porque, si la secuencia presenta una cierta velocidad **v**, la energía del espectro está concentrada en una cierta región del dominio de frecuencias espacio-temporales, **fx**, ft (o dominio de Fourier 3D), cumpliendo la siguiente ecuación:

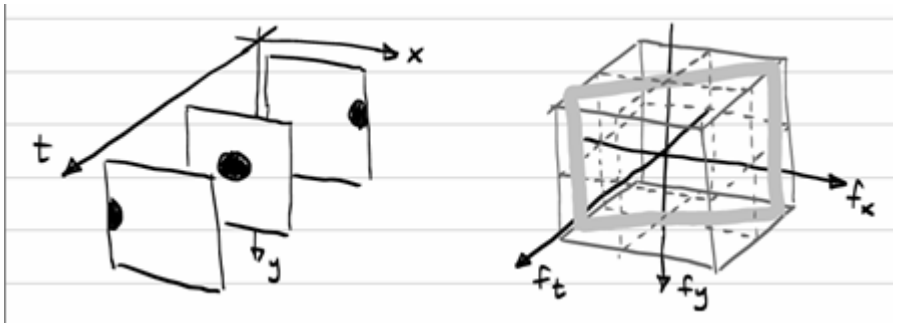
$$\mathbf{fx} \cdot \mathbf{v} + ft = 0 \qquad (2)$$

Esa es la llamada ecuación del flujo óptico en el dominio de Fourier. Es decir, resulta que si en la secuencia un cierto objeto se mueve con velocidad, **v**, su espectro tiene alta energía en una región determinada por la velocidad, **v**, y energía nula en las otras zonas (ver fig. 3).

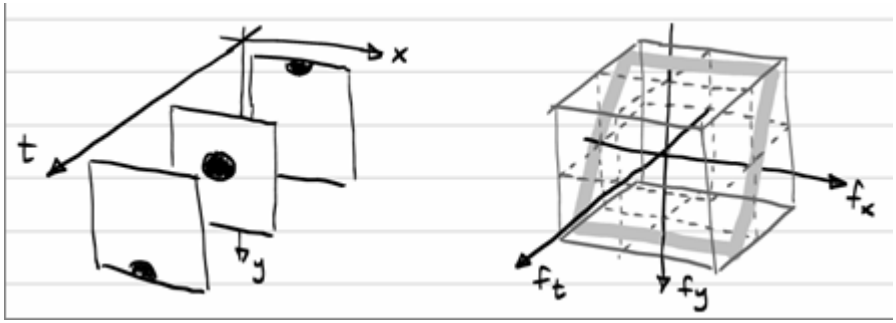
Si en nuestro cerebro dispusiésemos de sensores sintonizados a un conjunto de frecuencias espacio-temporales que recubrieran una amplia región del dominio de Fourier 3D, ante una secuencia con una cierta velocidad **v**, un cierto conjunto consistente de neuronas daría alta respuesta, mientras que el resto de neuronas daría respuesta nula. Eso es justamente lo que ocurre en las regiones V1 y MT de nuestro córtex visual.



3.a 3.b



3.c 3.d



3.e 3.f

Fig. 3:
Fotogramas de unas secuencias en el dominio espacio-temporal (figs. a, c, e), y espectros de dichas secuencias (figs. b, d, f). Las secuencias muestran el movimiento de una bola negra sobre fondo blanco moviéndose con diferentes velocidades y sentidos. Los dos primeros casos (a,c) muestran una bola moviéndose de derecha a izquierda con poca velocidad (3.a), y una velocidad mayor (3.c). El último caso (3.e) muestra una bola cayendo a gran velocidad. En los espectros, el recuadro resaltado en gris corresponde a la región frecuencial donde la secuencia tiene energía. Fuera de ese plano, el estímulo tiene energía nula. La inclinación de dicho plano está dada por la ecuación 2, es decir, para cada velocidad ***v*** *se tiene una inclinación diferente del espectro.*

Hubel y Wiesel [5], recibieron el premio Nobel de medicina en 1985 por el siguiente *error*. Ellos estaban registrando las respuestas de las neuronas de V1 en macacos (que naturalmente morían tras los experimentos) presentándoles un cierto estímulo definido por una pequeña región transparente en una lámina opaca que se iluminaba desde atrás. Lo que veía el mono era la luz que pasaba por esa región transparente. Después de repetidos intentos fallidos (sin obtener respuesta), por error vieron que esas neuronas sólo respondían cuando metían o sacaban la lámina opaca (ver cortometrajes de este tipo en la figura 4). Y no siempre respondían. Resulta que distintos grupos de esas neuronas son exclusivamente sensibles a bordes de una cierta orientación moviéndose con cierta velocidad. ¿Te das cuen?

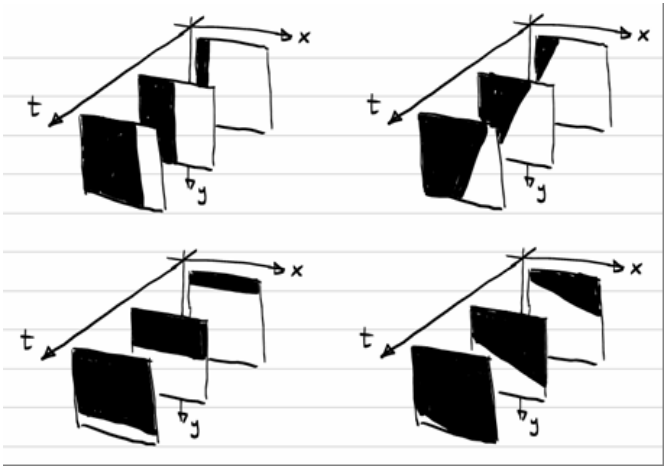


Fig. 4. Ejemplos del tipo de estímulos (secuencias) que Hubel y Wiesel mostraban a los macacos metiendo una placa opaca delante de un fondo iluminado. La neurona que responde al primero de los estímulos (placa con borde vertical avanzando de izquierda a derecha) no responde en los otros casos: las neuronas son selectivas a la orientación y a la velocidad de desplazamiento de los bordes. Este fenómeno está relacionado con la sensibilidad de dicha neurona a frecuencias espacio-temporales.

La investigación neurofisiológica de los años 60 y 70 mató a muchos monos pero también consiguió establecer que las células de V1 son sensibles a pequeñas regiones del dominio de Fourier 3D (responden sólo a ciertas *notas* del estímulo), y que las neuronas de MT recogen la respuesta de las neuronas de V1 con sensibilidades alineadas de forma coherente según la ecuación (2). Véase la figura 5.

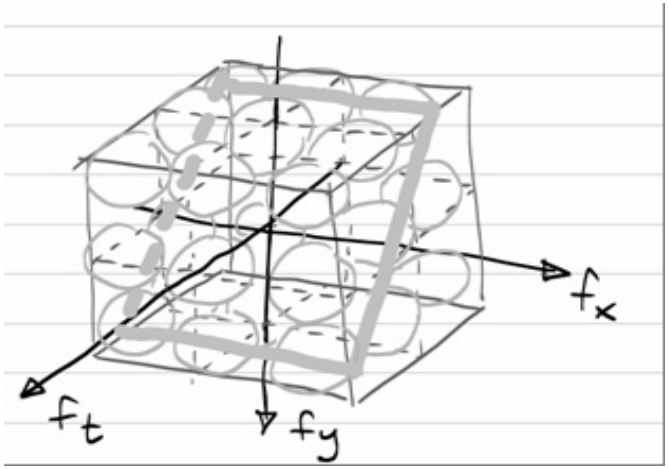


Fig. 5. Las bolas en trazo claro representan las regiones de sensibilidad frecuencial de cada sensor de V1 (la estructura es orientativa, y existe un recubrimiento más denso en la zona de bajas frecuencias espacio-temporales). Lo que implica cada una de estas regiones circulares es que el sensor correspondiente solo responde si el estímulo presenta energía en esa zona. Ya vimos que no todas las secuencias presentaban energía en todas las zonas del dominio (figura 3). Las neuronas de MT recogen las salidas de un conjunto coherente de sensores. Conjunto coherente en el sentido de tener sus sensibilidades alineadas en un cierto plano del dominio de Fourier 3D (como por ejemplo, los cuyas sensibilidades intersectan el plano marcado en gris). Diferentes neuronas de MT recogen respuestas de neuronas de V1 con sensibilidades alineadas según distintos planos, es decir, son sensibles a distintas velocidades.

De esta manera, resulta que ante la presencia del estímulo (secuencia) con velocidad **v**, sólo responde un pequeño conjunto de neuronas de MT, e interpretamos esa excitación como (percibimos) que ahí hay algo moviéndose con velocidad **v**.

4.- Fundido en negro

Con todo lo anterior, se puede construir un modelo sencillo que permite simular la percepción de velocidades tal como ocurre en la región MT de nuestro cerebro [4,6,7]. Los elementos de este modelo incluyen un conjunto de filtros que simulan la sensibilidad de las neuronas de V1 (tales como las que se representan en la figura 5), obteniendo su respuesta representando el estímulo en el dominio de Fourier 3D e integrando la energía del mismo en esas regiones de sensibilidad. La suma ponderada de esas respuestas produce las respuestas de MT que pueden interpretarse como velocidades.
En la figura 6 vemos el resultado de este cálculo para el octavo fotograma de un corto de autor [8] que muestra a un panoli vestido con una camiseta de rayas moviendo las manos arriba y abajo. Algunos críticos han señalado una velada referencia a Chiquito de la Calzada, aunque el autor no se ha manifestado sobre el particular.

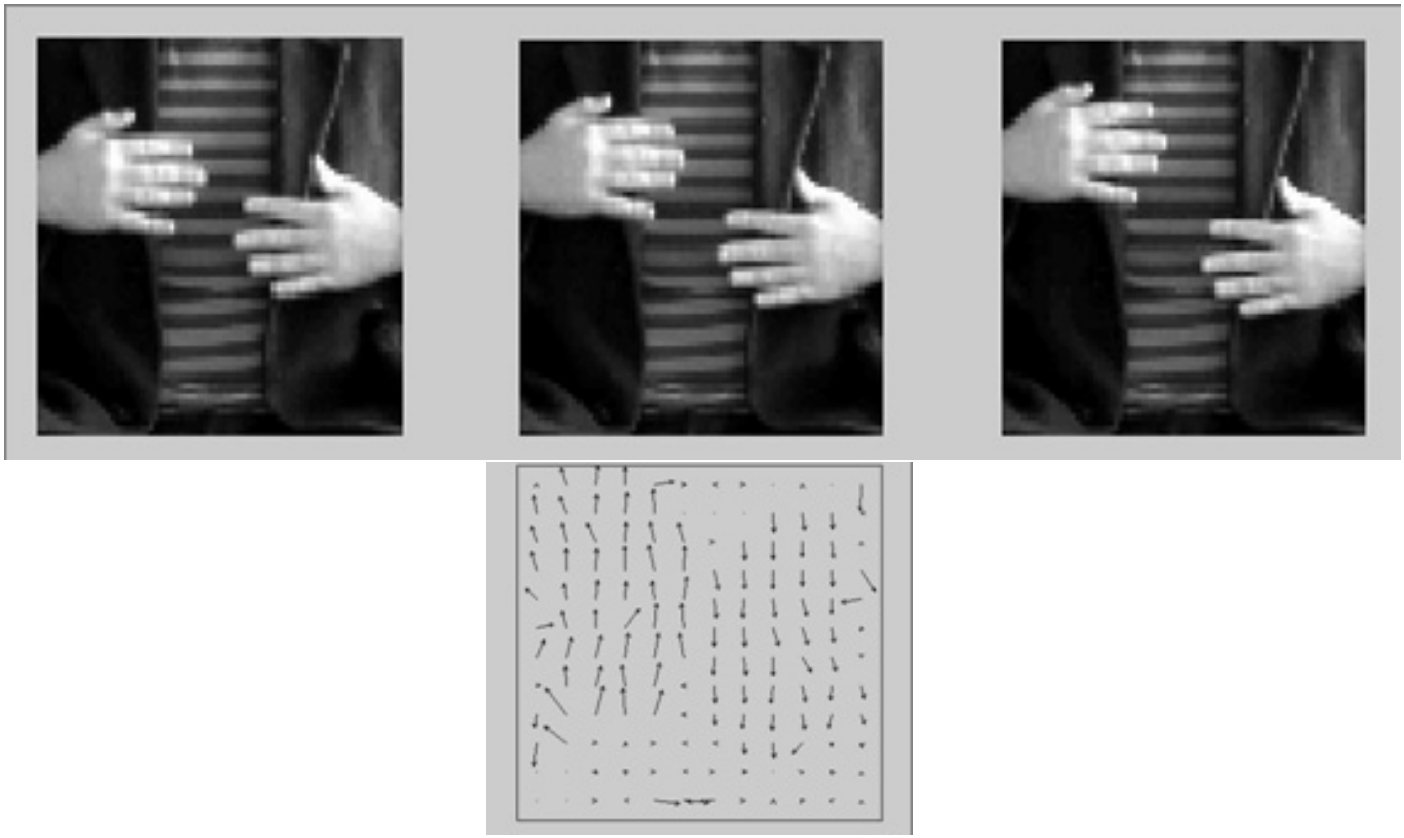
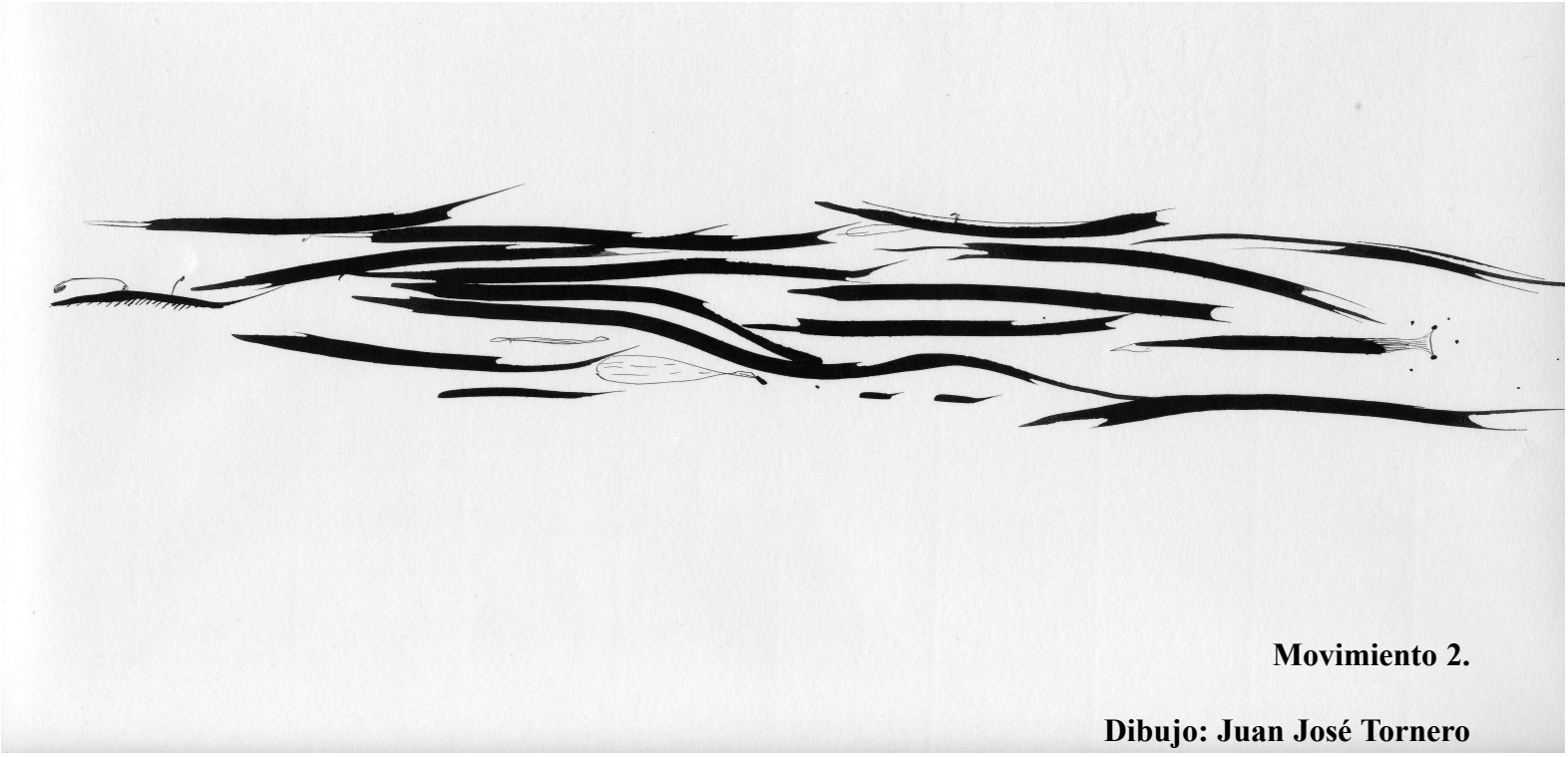


Fig. 6. La parte superior representa los fotogramas 6, 8 y 10 de la secuencia Manos [8]. El recuadro inferior representa el flujo óptico para el fotograma 8 calculado según el procedimiento descrito en el texto (modelo de Heeger del cortex V1 y MT [6]) según la implementación [9].

Nótese cómo el modelo identifica adecuadamente las regiones en movimiento y la magnitud y el sentido de su velocidad: flechas hacia arriba en la zona izquierda y lo opuesto en la otra zona. En algunas zonas se tienen errores de cálculo (percepciones falsas!). Estas cuentas, que a tu córtex le cuestan sólo unos milisegundos, a un ordenador Pentium CoreDuo de 1GB de RAM ejecutando el código de Matlab preparado por el autor [9] le llevan 70 minutos (400000 veces más tiempo!), vaya un programador...! No obstante, independientemente de la calidad de la programación, donde esté tu córtex, que se quite HAL9000 [10] que, además de ordenador, es un cabrón que te lee los labios y luego te asesina. Por cierto, en qué está ocupado el 99% restante de tu cerebro mientras lees ésto? Aún pensando en los otros 400 ms? Eso sí que es fascinante.

5.- Títulos de crédito. Una producción de Jesús Malo.

[1] Steven Soderbergh. Sex, lies and videotape. Virgin and Outlaw Productions 1989
[2] Joseph Fourier. Biografía en http://http://www-history.mcs.st-and.ac.uk/Biographies/Fourier.html
[3] Eduard Soto. El Neng. Buenafuente, Antena 3 TV, 2006.
[4] Andrew Watson & Albert Ahumada. A model of human visual-motion sensing. J.Opt.Soc.Am.A, Vol.2, N.2, pp. 322-342. 1985.
[5] Hubel. Plenary Talk en la 29 European Conference of Visual Perception. A Coruña 2005
[6] David Heeger. Model for the extraction of image flow. J.Opt.Soc.Am.A, Vol.4, pp. 1455-1471. 1987.
[7] Eero Simoncelli & David Heeger. A model of neuronal responses in visual area MT. Vis. Res. Vol.38, N.5, pp. 743-761, 1998.
[8] Jesús Malo. Manos. Cortometraje de 2.5 segundos filmado con JVC Everio GZ-MG27e. Diciembre de 2006
[9] Jesús Malo. Implementación Matlab del modelo de Heeger. Aula virtual de la Universitat de València. Mecanismos y Modelos de Visión de Movimiento. 2007.
[10] Stanley Kubrick. 2001: A Space Odyssey. Metro-Goldwyn-Mayer, 1968.



Movimiento 2.

Dibujo: Juan José Tornero