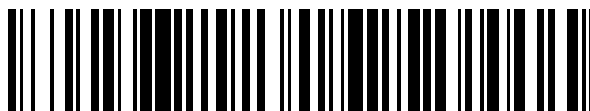


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 478 284**

51 Int. Cl.:

G06K 9/46

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.02.2011** **E 11705177 (1)**

97 Fecha y número de publicación de la concesión europea: **09.04.2014** **EP 2534612**

54 Título: **Extracción y descripción eficaz en el espacio escalado de puntos de interés**

30 Prioridad:

08.02.2010 US 302290 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

21.07.2014

73 Titular/es:

TELEFÓNICA, S.A. (100.0%)

**Gran Vía, 28
28013 Madrid, ES**

72 Inventor/es:

**MARIMON SANJUAN, DAVID;
BONNIN LLOFRIU, ARTURO;
ADAMEK, TOMASZ y
GIMENO HERNÁNDEZ, ROGER**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 478 284 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Extracción y descripción eficaz en el espacio escalado de puntos de interés

Campo técnico

La presente invención se refiere al campo de la visión por ordenador y en particular al subcampo de extracción de "puntos clave".

Descripción de la técnica anterior

Puntos clave o puntos destacados son aquellas muestras en una imagen que son muy repetibles a través de diferentes circunstancias de visión, tales como punto de vista e iluminación. La identificación de estos puntos clave en diferentes instancias de imagen de un objeto o una escena permite efectuar tareas como la reconstrucción en 3D de espacios o el reconocimiento de objetos, entre otras.

En la última década se han desarrollado diversas técnicas de extracción de puntos clave. Entre ellas se encuentran dos procedimientos que han logrado una amplia aceptación en diversos campos de aplicación debido a su robustez, concretamente el procedimiento de transformación de características invariantes a escala, SIFT (usado por ejemplo en la patente estadounidense número US 6,711,293 "*Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image*") y el procedimiento de características robustas acelerado, SURF (usado por ejemplo en la patente europea número 1850270 "*Robust interest point detector and descriptor*").

El procedimiento SIFT busca extremos en la diferencia de Gaussianas a escalas muestreadas de manera consecutiva. En primer lugar se calcula una pirámide de versiones alisadas de la imagen de entrada. Se hallan extremos en cada octavo (duplicación de sigma). En cada uno de estos extremos se describe un punto clave con información extraída de los píxeles del entorno. En el caso del procedimiento SIFT, se construye un descriptor basándose en la orientación del gradiente. En primer lugar, se define una rejilla según la orientación principal del gradiente alrededor del punto clave. Dentro de cada rejilla se calcula un histograma de la orientación del gradiente ponderado por su magnitud. El descriptor se construye, normalmente, con 4x4 de tales regiones e histogramas de 8 intervalos, lo que lleva a un vector de 128 componentes. Este procedimiento ha demostrado tener un buen rendimiento en una amplia variedad de aplicaciones, tales como reconstrucción en 3D, reconocimiento de objetos o localización de robots. Sin embargo, este procedimiento tiene un inconveniente principal, concretamente, el coste computacional de construir la pirámide y determinar el descriptor para cada punto clave.

Por otro lado, el procedimiento SURF está diseñado para una extracción en el espacio escalado mucho más rápida. Los extremos se ubican en el determinante Hessiano aproximado por *wavelets* de Haar. El descriptor se basa en la polaridad de los cambios de intensidad. Se calculan sumas del gradiente (orientadas con la orientación principal del punto clave) y el valor absoluto del gradiente en dirección horizontal y vertical. El descriptor se forma habitualmente por 4 de tales valores calculados en 4x4 regiones alrededor del punto clave, lo que conduce a un descriptor de 64 valores. El beneficio de su procedimiento se encuentra principalmente en la fase de extracción, en la que las *wavelets* de Haar se calculan accediendo a la imagen integral. Esto reduce drásticamente el número de accesos a la memoria y de cálculos, especialmente en un entorno multiescala.

Estudios recientes (como por ejemplo S. Winder y M. Brown. *Learning local image descriptors*. En Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), volumen 0, páginas 1-8, Los Alamitos, CA, EE.UU., 2007. IEEE Computer Society) han evaluado el rendimiento de diferentes esquemas y características de descriptores, entre otros parámetros para describir parches de imagen. En sus estudios, pueden observarse claras mejoras en combinaciones específicas de parámetros (por ejemplo el descriptor DAISY divulgado en E. Tola, V. Lepetit, y P. Fua. *A fast local descriptor for dense matching*. En Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), volumen 0, páginas 1-8, Los Alamitos, CA, EE.UU., 2008. IEEE Computer Society) frente a los descriptores ampliamente aceptados citados anteriormente. Sin embargo, la implementación completa de estas combinaciones lleva a procedimientos con altos requisitos computacionales.

En resumen, nos proponemos un nuevo marco para la rápida extracción y descripción de puntos clave en el espacio escalado con las siguientes aportaciones:

Una aproximación del determinante Hessiano mediante filtros triangulares definidos a trozos que sea más rápida y proporcione un rendimiento similar o mejor que la aproximación, basada en imágenes integrales, de SURF y la diferencia de Gaussianas de SIFT. Una asignación de la orientación de puntos clave más rápida que la realizada en el procedimiento SURF.

Un descriptor de tipo DAISY extraído de manera eficaz reutilizando cálculos realizados para la extracción de puntos clave y optimizando el espacio de muestreo.

Todas estas mejoras conducen a un factor de aceleración de 6 en comparación con el procedimiento SIFT y de 3 en comparación con el procedimiento SURF, con mejores rendimientos de "precisión frente a exhaustividad" que las obtenidas por SIFT y SURF.

Sumario de la invención

5 La presente invención propone un procedimiento, sistema y programa informático para extraer y describir puntos clave en el espacio escalado. En particular, se divulga un procedimiento, sistema y programa informático para extraer puntos clave invariantes en cuanto al punto de vista y la iluminación en el espacio escalado con una configuración particular que muestra mejores resultados de precisión frente a exhaustividad que SIFT y SURF.

10 El algoritmo se diseña con vistas a obtener una complejidad computacional global baja. Por un lado, los datos adquiridos durante la extracción se reutilizan de manera intensa en la fase de descripción. Por otro lado, se propone una optimización algorítmica de la descripción que acelera drásticamente el proceso.

En un primer aspecto, un procedimiento para la extracción y descripción de puntos clave en el espacio escalado en una imagen, comprendiendo el procedimiento las siguientes etapas:

a) filtrar la imagen con filtros de núcleo (filtro kernel) triangular a diferentes escalas,

15 b) calcular una aproximación del determinante Hessiano a cada escala, calculándose esta aproximación a cada escala k como $|\partial_{xx}^k(i, j) \cdot \partial_{yy}^k(i, j) - \partial_{xy}^k(i, j)^2|$, donde

$$\partial_{xx}^k = L(k, i - d_1, j) - 2 \cdot L(k, i, j) + L(k, i + d_1, j)$$

$$\partial_{yy}^k = L(k, i, j - d_1) - 2 \cdot L(k, i, j) + L(k, i, j + d_1)$$

$$\partial_{xy}^k = L(k, i - d_2, j - d_2) - L(k, i + d_2, j - d_2) - L(k, i - d_2, j + d_2) + L(k, i + d_2, j + d_2)$$

20 donde $L(k, i, j)$ es la respuesta de imagen filtrada obtenida en la etapa a) a escala k en el punto (i, j) y en este caso d_1 y d_2 son parámetros de diseño,

c) buscar valores extremos, tanto en una única escala como a través del espacio escalado, de la aproximación del determinante Hessiano obtenido en la etapa b) y calcular los puntos clave a partir de estos valores extremos,

25 d) para cada punto clave, ubicado en un valor extremo, detectar las orientaciones dominantes a partir de información de gradiente calculada usando la respuesta de imagen filtrada obtenida en la etapa a)

e) calcular, para cada orientación dominante, un descriptor de punto clave.

30 La figura 1 muestra un diagrama de bloques del mejor modo del procedimiento. En primer lugar, la imagen se filtra con un núcleo (kernel) triangular a diferentes escalas. Esto va seguido del cálculo del determinante Hessiano a cada escala y, después, de la detección de extremos en este espacio. Los puntos clave se hallan usando dichos extremos (los puntos clave se seleccionan como aquellos píxeles a una escala particular que son los extremos del determinante Hessiano) y, para cada punto clave, las orientaciones dominantes se calculan a partir de información de gradiente extraída usando las imágenes filtradas por el triángulo. Para cada orientación dominante, se calcula su descriptor. Este cálculo utiliza los gradientes orientados también extraídos a partir de las imágenes filtradas por el triángulo.

En otro aspecto, se presenta un sistema que comprende medios adaptados para llevar a cabo el procedimiento anteriormente descrito.

Finalmente, se presenta un programa informático que comprende medios de código de programa informático adaptados para llevar a cabo el procedimiento anteriormente descrito.

40 Para una comprensión más completa de la invención, sus objetos y ventajas, puede hacerse referencia a la siguiente memoria descriptiva y a los dibujos adjuntos,

Breve descripción de los dibujos

45 Para completar la descripción y con el fin de proporcionar una mejor comprensión de la invención, se proporciona un juego de dibujos. Dichos dibujos forman parte integrante de la descripción e ilustran una realización preferida de la invención, que no ha de interpretarse como que restringe el alcance de la invención, sino más bien como un ejemplo de cómo puede implementarse la invención. Los dibujos comprenden las siguientes figuras:

La figura 1 representa un diagrama de bloques del mejor modo del procedimiento de extracción y descripción de puntos clave propuesto. Las imágenes filtradas por el triángulo se reutilizan para la extracción de la orientación dominante y el cálculo del descriptor de tipo DAISY.

La figura 2 muestra un núcleo con forma de triángulo 2D.

- 5 La figura 3 muestra una pila de versiones filtradas $L(k,i,j)$ con $k=1,\dots,K$ creadas mediante el filtrado de la imagen de entrada con un núcleo con forma de triángulo 2D. A cada escala se aumenta el tamaño (ambos ejes horizontal y vertical de igual modo) del núcleo.

La figura 4 muestra la forma de la segunda derivada. Parte superior: Gaussiana. Parte inferior: aproximación con filtros triangulares ponderados.

- 10 La figura 5 muestra la forma del filtro de segunda derivada para los ejes horizontal (o vertical rotados de manera equivalente) usada en SURF.

La figura 6 muestra una representación gráfica de la búsqueda de extremos en las escalas actual, superior e inferior. Se realiza una búsqueda rápida opcional en una ventana de 3x3 en la escala actual.

- 15 La figura 7 muestra la distribución con 2 anillos y 8 segmentos (representados con círculos) por cada anillo orientada a 0 grados. Debido a la distancia de los segmentos respecto al centro pueden ubicarse muchas muestras (representadas con puntos) muy próximas entre sí.

La figura 8 muestra la puntuación de repetibilidad (%) para secuencias de imagen. DART tiene un rendimiento similar o mejor que SIFT y SURF para cambios de punto de vista, borrosidad y compresión jpeg. Los cambios de escala parecen afectar a la fase de extracción de DART más que las otras técnicas.

- 20 La figura 9 muestra una representación gráfica de precisión frente a exhaustividad para secuencias de imagen, de izquierda a derecha y de arriba a abajo, Graffiti, Boat, Bikes y Leuven. DART da mejores resultados en todas las secuencias evaluadas.

Números y símbolos correspondientes en las diferentes figuras se refieren a partes correspondientes a menos que se indique lo contrario.

25 Descripción detallada de la invención

El procedimiento de la invención describe un proceso para detectar en primer lugar puntos clave en el espacio escalado (etapa de extracción de puntos clave) y después describirlos con un descriptor de tipo DAISY (etapa de descripción de puntos clave). La presente patente abarca ambos procesos tanto por separado como conjuntamente.

- 30 El procedimiento de la invención para la detección de extremos en el espacio escalado se compone de tres etapas.

1. Primera, un cálculo del determinante Hessiano de cada escala.
2. Segunda, búsqueda de extremos en el espacio escalado.
3. La última etapa es opcional y consiste en hallar los puntos clave con subpíxel y subescala.

- 35 Se comenzará explicando el algoritmo de extracción de manera genérica y después el proceso exacto que hace que sea eficaz. El determinante Hessiano consiste en calcular $|\partial_{xx}(i, j) \cdot \partial_{yy}(i, j) - \partial_{xy}(i, j)^2|$, donde ∂_{xx} es la segunda derivada horizontal de Gaussian de una imagen, ∂_{yy} es la vertical, y ∂_{xy} son derivadas cruzadas. El presente procedimiento puede usarse con diferentes núcleos de filtro. Al menos la etapa de extracción puede llevarse a cabo con filtros gaussianos bidimensionales, habiéndose aproximado esos filtros con imágenes integrales ponderadas simétricas y también una aproximación con triángulos definidos por trozos. Las imágenes integrales ponderadas
- 40 simétricas son imágenes integrales (suma bidimensional acumulativa) que se ponderan con un factor creciente o decreciente. Calculando varias de tales integrales y accediendo a ellas en diferentes ubicaciones, es posible reproducir un gran número de formas de núcleo. El procedimiento calculado con triángulos definidos por trozos es el modo preferido y explicado a continuación en el presente documento.

- 45 Supóngase que una imagen se filtra con un filtro con forma de triángulo 2D obteniendo $L(i, j)$. La figura 2 traza la forma de este filtro.

La figura 3 muestra la pila de versiones filtradas $L(k,i,j)$ con $k=1,\dots,K$ creadas filtrando la imagen de entrada con núcleo con forma de triángulo 2D. A cada escala aumenta el tamaño (ambos ejes horizontal y vertical de igual forma) del núcleo.

Para cada escala k , todas las derivadas se calculan accediendo a la correspondiente respuesta filtrada $L(k,i,j)$ en

diferentes puntos (i, j) . Teóricamente, la forma de la segunda derivada horizontal gaussiana (véase la figura 4, parte superior) se aproxima mediante respuestas con forma de triángulo trasladadas y ponderadas (véase la figura 4, parte inferior).

Este proceso se lleva a cabo a diferentes escalas $k=1, \dots, K$ lo que lleva a las siguientes aproximaciones:

$$\begin{aligned} \partial_{xx}^k &= L(k, i - d_1, j) - 2 \cdot L(k, i, j) + L(k, i + d_1, j) \\ \partial_{yy}^k &= L(k, i, j - d_1) - 2 \cdot L(k, i, j) + L(k, i, j + d_1) \\ \partial_{xy}^k &= L(k, i - d_2, j - d_2) - L(k, i + d_2, j - d_2) - L(k, i - d_2, j + d_2) + L(k, i + d_2, j + d_2) \end{aligned}$$

donde d_1 y d_2 se eligen experimentalmente y proporcionales a la σ (sigma) de la segunda derivada aproximada de núcleo gaussiano. Obsérvese que esas aproximaciones no son equivalentes a un filtrado con un filtro triangular y después una convolución con un filtro de segunda derivada [1;-2; 1] ya que esto generaría artefactos no deseados. Tal como puede deducirse, calcular las derivadas requiere sólo 9 accesos a L . Esto debe compararse con la aproximación en forma de caja de SURF, en la que el cálculo del determinante hessiano requiere $8 + 8 + 16 = 32$ accesos a la imagen integral. Un ejemplo de esta aproximación para la segunda derivada usando *wavelets* de Haar se muestra en la figura 5.

El espacio escalado está formado con la pila de versiones filtradas $L(k, i, j)$ tal como se muestra en la figura 3. Esto es diferente a lo indicado en SIFT y SURF donde se crea una pirámide de versiones filtradas. En este proceso, en cada octava (duplicado de la sigma σ del filtro gaussiano o el filtro gaussiano aproximado, respectivamente) se realiza un submuestreo. En los experimentos realizados, el submuestreo provoca una pérdida importante de rendimiento y por tanto el procedimiento descrito en la presente invención realiza el filtrado de la imagen de entrada sin submuestreo. A pesar de que esto podría considerarse como muy exigente desde el punto de vista computacional, tal como se describe más adelante, una de las mejoras de este procedimiento es la constante reutilización de las versiones filtradas $L(k, i, j)$. En primer lugar, esto permite grandes ahorros computacionales durante el cálculo del determinante Hessiano (tal como se comentó anteriormente) y en segundo lugar, esto se aprovecha adicionalmente en la descripción de los puntos clave tal como se explicará posteriormente.

La búsqueda de un extremo se realiza en todas las escalas salvo la primera $k = 1$ y la última $k = K$. Los extremos se buscan dentro de una ventana de búsqueda de σ en las escalas actual k , superior $k+1$ e inferior $k-1$. Esta patente también abarca una extensión para acelerar el proceso: se realiza una primera prueba en una ventana de 3×3 en la escala actual k para detectar rápidamente los no máximos y evitar un procesamiento adicional. La figura 6 ilustra una representación gráfica de esta búsqueda.

Ha de observarse que tanto SIFT como SURF buscan extremos sólo dentro de octavos de la pirámide, y generan escalas adicionales para permitir una detección de extremos correcta. La búsqueda de extremos correctos es continua en la pila en el espacio escalado.

La última etapa es opcional y consiste en hallar el punto clave con precisión de subpíxel y subescala. El modo preferido es obtener estas precisiones en *M. Brown y D. Lowe. Invariant features from interest point groups. En Proc. British Machine Vision Conference (BMVC), páginas 656-665, septiembre de 2002*.

Es decir, el procedimiento ajusta una función cuadrática centrada en cada muestra de extremo para determinar la ubicación interpolada, buscando en los píxeles del entorno de la misma escala, las superiores y las inferiores. Los resultados coincidentes y la estabilidad mejoran usando precisión de subpíxel y subescala.

Algunos estudios han evaluado recientemente diferentes características, distribuciones y etapas para describir puntos clave. La que proporciona los mejores resultados es una distribución de DAISY similar a la ilustrada en la figura 7. Motivado por estos resultados, el procedimiento de la invención abarca una variación de un descriptor DAISY que se ajusta a esta finalidad de reutilizar tanta información como sea posible de la etapa de extracción.

Con el fin de generar descriptores invariantes respecto al punto de vista, la primera etapa es detectar la orientación dominante del punto clave. El enfoque seguido en SIFT es calcular el histograma de orientaciones de gradiente ponderadas por su magnitud en una ventana de la correspondiente imagen alisada de manera gaussiana. Los gradientes se calculan mediante diferenciación de píxeles. La orientación dominante se encuentra en el pico del histograma. Si se encuentra más de un pico dominante, se generan varios puntos clave. En el caso de SURF, se calculan derivadas con *wavelets* de Haar (aprovechando la imagen integral) en puntos muestreados en un entorno circular del punto clave. La magnitud de cada muestra ocupa un espacio de derivadas horizontal y vertical. Este espacio se explora y la orientación dominante se halla con la suma mayor de valores en la ventana.

La presente invención aprovecha algunos de los beneficios de ambos enfoques y define un enfoque alternativo.

Para cada punto clave, se accede a un número reducido de muestras en un entorno para calcular el gradiente. Se gana aceleración muestreando un entorno circular. Para cada muestra (i,j) , la presente ventaja es que se calculan gradientes accediendo simplemente a $L(k)$ en dos puntos (en comparación con 6 muestras en SURF):

$$\partial_x^k = L(k, i - d_3, j) - L(k, i + d_3, j)$$

5 para la primera derivada horizontal y de manera equivalente para la vertical. d_3 es proporcional a σ si no se obtiene una precisión de subescala o a σ' en caso contrario. Cada derivada se acumula en un histograma con un peso proporcional a su magnitud y con un núcleo gaussiano centrado en el punto clave. Finalmente, se hallan múltiples orientaciones dominantes buscando picos con valores próximos al máximo.

10 En la presente invención, un descriptor se compone de derivadas de primer orden (gradiente) muestreadas con una distribución específica, seguido de una normalización por todo el vector.

La distribución consiste de varios segmentos y anillos. Cada segmento es una parte del entorno del punto clave. Cada segmento genera varias características. Los anillos son un grupo de segmentos que comparten la siguiente propiedad: el centro del segmento está situado a la misma distancia euclidiana del punto clave. Esta distancia se comenta más adelante en este documento. Además, la distribución tiene un segmento central centrado en el punto clave.

15 Cada segmento contribuye al vector de características que describe el punto clave. Se accede a las muestras en cada segmento para calcular las características de ese segmento correspondiente. El bloque de muestreo de características original consiste en calcular derivadas de primer orden (gradiente) de cada píxel en un entorno del punto clave y obtener un vector de cuatro valores: $\{|\partial_x| - \partial_x; |\partial_x| + \partial_x; |\partial_y| - \partial_y; |\partial_y| + \partial_y\}$, siendo ∂_x la primera derivada horizontal y ∂_y la primera derivada vertical.

20 En la presente invención, se aproximan y se calculan los mismos cuatro valores, pero sólo a partir de muestras seleccionadas en un segmento. En cada una de tales muestras, se evalúan gradientes y las derivadas se orientan según la orientación del punto clave. Obsérvese que SURF calcula también gradientes orientados, pero no puede aprovechar la imagen integral sin introducir artefactos ya que los *wavelets* de Haar están orientados con indexación de píxeles de imagen. Por otro lado, la aproximación de gradientes orientados es sencilla en la presente invención con sólo dos accesos a $L(k)$:

$$\partial_x^\theta = L(k, i - d_3 \cdot \cos\theta, j - d_3 \cdot \sin\theta) - L(k, i - d_3 \cdot \cos\theta, j + d_3 \cdot \sin\theta).$$

donde θ es el ángulo de la correspondiente orientación dominante del punto clave.

30 Tal extracción de característica se realiza sobre muestras con la distribución espacial que sigue a la distribución de segmentos y anillos anteriormente descrita.

Esta patente abarca todas las posibles distribuciones que obtienen las características a partir de las versiones filtradas L independientemente del núcleo de filtro. El modo preferido de esta patente utiliza la distribución que produce los mejores resultados al tiempo que mantiene la longitud del descriptor relativamente corta. Tiene 8 segmentos y 2 anillos. Esto produce un vector de $(1+2 \cdot 8) \cdot 4 = 68$ valores. Además, el número de muestras seleccionadas para cada segmento varía dependiendo del anillo. En particular, los núcleos tienen un tamaño de 3×3 , 5×5 , y 7×7 , para el segmento central, anillos primero y segundo, respectivamente. También se realiza una ponderación gaussiana sobre las muestras de cada segmento. Antes de acceder a las muestras de cada segmento, se rota la distribución completa (centros de segmento y muestras de segmento) con respecto a la orientación dominante del punto clave.

40 Finalmente, se aplica una normalización L2 al vector descriptor. El resultado se cuantifica a sólo 8 bits. La normalización L2 y la cuantificación a 8 bits son ambas opcionales e independientes, pero forman parte del modo preferido. Este vector es lo que se denomina punto clave DART.

45 En el descriptor DAISY original hay dos parámetros que determinan adicionalmente la distribución: la distancia entre muestras del bloque de muestreo de cada segmento y la distancia entre el centro del punto clave y los centros de cada segmento.

Debido a resultados experimentales, las muestras se separan en 2σ , y se toma una distancia respecto a los segmentos de 4σ para el primer anillo y de 8σ para el segundo. Los segmentos en esta distribución se solapan en gran medida tal como puede verse en la figura. *Esta propiedad es realmente deseable en nuestro diseño puesto que el cálculo del descriptor puede optimizarse.*

50 Al mirar las coordenadas de las muestras que contribuyen a cada segmento, puede observarse cómo se accede a

muestras muy próximas (si no exactamente a la misma). Se procede reagrupando muestras cercanas en una única muestra. Las derivadas orientadas X e Y calculadas en esa muestra contribuyen a varios segmentos con el peso correspondiente.

5 El proceso de calcular las muestras a las que va a accederse se realiza sólo una vez independientemente de la escala. La escala (σ) o la subescala (σ'), si se aplica precisión de subescala, es un factor de multiplicación aplicado en el momento de calcular el descriptor de un punto clave dado. El resultado es una rejilla de muestras y los correspondientes enlaces a los segmentos a los que contribuyen.

10 Esta optimización reduce drásticamente el número de accesos. De las $3 \times 3 + 5 \times 5 + 7 \times 7 = 601$ muestras originales (lo que implica 2404 accesos a la memoria para la versión filtrada), el número baja hasta 197 muestras (788 accesos a la memoria), mediante la reagrupación de muestras ubicadas en un radio de $\sigma/2$. Esta reducción no tiene una pérdida significativa de rendimiento tal como se muestra en la siguiente sección. Además, el reagrupamiento con un radio mayor empieza a empeorar los resultados globales. Como puede deducirse, la aproximación de rejilla tiene menos impacto sobre los puntos clave a pequeñas escalas en las que las muestras están incluso más próximas entre sí.

15 La presente invención abarca el muestreo con y sin optimización sobre el número de muestras.

20 El proceso de extracción puede validarse midiendo la repetibilidad de puntos clave a través de diferentes transformaciones de imagen. Para simplificar la comprensión de las figuras, la presente invención se ha identificado usando el modo preferido con el nombre DART. La figura 8 muestra una comparación de la puntuación de repetibilidad para SIFT, SURF y DART en varias secuencias de la base de datos. Para cada secuencia se aplican diferentes distorsiones con una intensidad creciente (por ejemplo compresión JPEG o borrosidad creciente).

Para el punto de vista, la borrosidad y la compresión jpeg, puede observarse que DART tiene un rendimiento similar o mejor que SIFT y SURF. Aunque es comparable con las otras técnicas, los cambios de escala parecen afectar a la fase de extracción de DART. Este comportamiento se atribuye a la aproximación de segundas derivadas gaussianas con triángulos.

25 La variación de la invención del descriptor DAISY en puntos clave detectados con el presente extractor puede validarse midiendo la precisión-1 frente a exhaustividad. La figura 9 ilustra el rendimiento logrado por el procedimiento de la invención con y sin optimización del descriptor, junto con el rendimiento de SIFT y SURF. Como puede verse, DART produce mejores resultados en todas las secuencias evaluadas.

30 El algoritmo propuesto para extraer y describir puntos clave está diseñado con vistas a lograr la eficacia. Con el fin de validar esta hipótesis, se han extraído puntos clave en la primera imagen Graffiti (tamaño de imagen: 800 x 640 píxeles) en una Intel Core 2 Duo CPU a 2,33 GHz con 2GB de RAM. El tiempo que tarda el presente procedimiento se compara con el tiempo que tardan los binarios de SIFT y SURF (a partir de la página web oficial de los autores). La tabla 1 muestra el tiempo que tarda cada procedimiento. Obsérvese que el tiempo que se tarda incluye la carga de imágenes y también la escritura de los puntos clave en un archivo ASCII. Se han usado diferentes umbrales en DART para una comparación más justa en cuanto al número de puntos clave.

Tabla 1. Tiempo que tardan los diferentes procedimientos de extracción de puntos clave en la primera imagen de la secuencia Graffiti (800 x 640 píxeles).

Procedimiento	Longitud del descriptor	Número de puntos clave	Tiempo [s]
SIFT	128	3106	3,356
DART	68	3044	0,536
SURF	64	1557	1,207
DART	68	1540	0,394

40 Los resultados muestran una aceleración por un factor de 6x en comparación con SIFT y 3x en comparación con SURF. Aunque no se implementa en este documento, obsérvese la naturaleza altamente paralela de la etapa de extracción de pila de escalas. Esto indica que es posible una reducción adicional del coste computacional.

Con el fin de validar adicionalmente la aplicabilidad del procedimiento de la invención, se han explorado dos problemas de visión artificial que emplean de manera satisfactoria DART. Obsérvese que DART puede aplicarse también a otros problemas aparte de los validados en el presente documento.

Seguimiento de objetos: El seguimiento de objetos tridimensional consiste en seguir el rastro de la posición 3D de un objeto con respecto a una cámara estática o en movimiento. Esto se utiliza a menudo en aplicaciones tales como la realidad aumentada. En el caso particular de objetos planos, el problema puede resolverse correlacionando puntos clave en una imagen de referencia del objeto frente a puntos clave extraídos en cada fotograma de un flujo de vídeo. Una vez establecidas las correspondencias, puede estimarse la posición del objeto. Se ha implementado una correlación de vecinos más próximos de descriptores DART. Se eliminan aquellas correspondencias en las que la distancia euclidiana sobrepasa un umbral y si la razón entre la distancia respecto a la mejor coincidencia primera y segunda no es mayor que 0,7.

Reconstrucción en 3D: Existen una gran variedad de técnicas de reconstrucción de escenas. En el presente caso, se triangulan puntos clave DART a partir de dos o más vistas coherentes para generar una nube de puntos en 3D. Para realizar esta tarea, se usa geometría epipolar y de estructura a partir de movimiento para construir la representación geométrica de una escena real usando una secuencia de vídeo como entrada.

En resumen, la presente invención ha propuesto un procedimiento novedoso para extraer y describir de manera eficaz puntos clave en el espacio escalado. La invención incluye un procedimiento de extracción que aproxima el determinante Hessiano en el espacio escalado mediante filtros triangulares definidos por trozos calculados de manera eficaz. La invención también incluye una variación del descriptor DAISY con una optimización por el espacio de muestreo.

El procedimiento se ha comparado con técnicas similares en cuanto a la repetibilidad, precisión frente a exhaustividad y también el coste computacional. En cuanto a la de repetibilidad, el presente extractor tiene un rendimiento comparable o mejor que SIFT y SURF. En el caso de precisión-exhaustividad, la distribución del descriptor optimizado de la invención tiene un beneficio claro con respecto a los otros procedimientos. También se muestra una reducción de la complejidad computacional con una aceleración de 6x en comparación con SIFT, y de 3x en comparación con SURF. Junto con esta evaluación, la invención se ha aplicado de manera satisfactoria a aplicaciones de reconocimiento de objetos y de reconstrucción en 3D.

REIVINDICACIONES

1.- Un procedimiento realizado por un sistema para la extracción y descripción en el espacio escalado de puntos clave en una imagen, comprendiendo el procedimiento las siguientes etapas:

- 5 a) filtrar la imagen con filtros de núcleo triangular a diferentes escalas, donde los filtros son filtros de forma triangular 2D
- b) calcular una aproximación de un determinante Hessiano a cada escala, donde esta aproximación a cada escala k se calcula como $|\partial_{xx}^k(i, j) \cdot \partial_{yy}^k(i, j) - \partial_{xy}^k(i, j)^2|$, donde

$$\partial_{xx}^k = L(k, i - d_1, j) - 2 \cdot L(k, i, j) + L(k, i + d_1, j)$$

$$\partial_{yy}^k = L(k, i, j - d_1) - 2 \cdot L(k, i, j) + L(k, i, j + d_1)$$

$$\partial_{xy}^k = L(k, i - d_2, j - d_2) - L(k, i + d_2, j - d_2) - L(k, i - d_2, j + d_2) + L(k, i + d_2, j + d_2)$$

donde L(k, i, j) es la respuesta de imagen filtrada obtenida en la etapa a) a escala k en el punto (i, j) y en este caso d_1 y d_2 son parámetros de diseño, siendo d_1 y d_2 proporcionales a la sigma, σ , de la segunda derivada de núcleo gaussiano aproximada

- 15 c) buscar valores extremos tanto en una única escala como a través del espacio escalado de la aproximación del determinante Hessiano obtenido en la etapa b) y calcular los puntos clave a partir de estos valores extremos,

d) para cada punto clave, ubicado en un valor extremo, detectar las orientaciones dominantes a partir de la información de gradiente calculada usando la respuesta de imagen filtrada obtenida en la etapa a)

- 20 e) calcular, para cada orientación dominante, un descriptor de punto clave.

donde el descriptor está compuesto de gradientes orientados muestreados con una distribución específica, en el que los gradientes orientados se calculan como:

$$\partial_x^\theta = L(k, i - d_3 \cdot \cos \theta, j - d_3 \cdot \sin \theta) - L(k, i + d_3 \cdot \cos \theta, j + d_3 \cdot \sin \theta)$$

25 donde L(k, i, j) es la respuesta de imagen filtrada obtenida en la etapa a) a escala k en el punto (i, j) y d_3 es un parámetro de diseño, y θ es el ángulo de la orientación dominante del punto clave

2.- El procedimiento según la reivindicación 1, en el que la etapa de detectar las orientaciones dominantes comprende las siguientes etapas:

- para cada punto clave, se seleccionan varias muestras en el entorno del punto clave,
- para cada muestra (i,j), el gradiente horizontal se calcula como

$$\partial_x^k = L(k, i - d_3, j) - L(k, i + d_3, j)$$

y el gradiente vertical se calcula como:

$$\partial_y^k = L(k, i, j - d_3) - L(k, i, j + d_3)$$

donde L(k, i, j) es la respuesta de imagen filtrada obtenida en la etapa a) a escala k en el punto (i, j) y d_3 es un parámetro de diseño,

- 35 - se acumula cada gradiente en un histograma con un peso proporcional a su magnitud y con un núcleo gaussiano centrado en el punto clave,
- las orientaciones dominantes se hallan buscando picos con valores próximos al máximo.

3.- El procedimiento según la reivindicación 2, en el que el entorno es un entorno circular.

40 4.- El procedimiento según cualquiera de las reivindicaciones anteriores, en el que la distribución consiste en varios segmentos y anillos, siendo cada segmento una parte del entorno del punto clave y cada anillo un grupo de segmentos que comparten la siguiente propiedad: el centro del segmento está situado a la misma distancia

euclidiana desde el punto clave.

5.- El procedimiento según cualquiera de las reivindicaciones anteriores, en el que el cálculo de los puntos clave se realiza con una precisión de subpíxel y subescala.

5 6.- El procedimiento según la reivindicación 5, en el que el cálculo de los puntos clave se realiza ajustando una función cuadrática centrada en cada valor extremo para determinar la ubicación interpolada y buscando, en los píxeles del entorno de la misma escala, los superiores y los inferiores.

7.- Un sistema que comprende medios adaptados para llevar a cabo el procedimiento según cualquier reivindicación anterior.

10 8.- Un programa informático que comprende medios de código de programa informático adaptados para llevar a cabo el procedimiento según cualquiera de las reivindicaciones 1 a 6, cuando dicho programa se ejecuta en un ordenador, un procesador de señal digital, una disposición de puertas programables en campo, un circuito integrado de aplicación específica, un microprocesador, un microcontrolador, o cualquier otra forma de hardware programable.

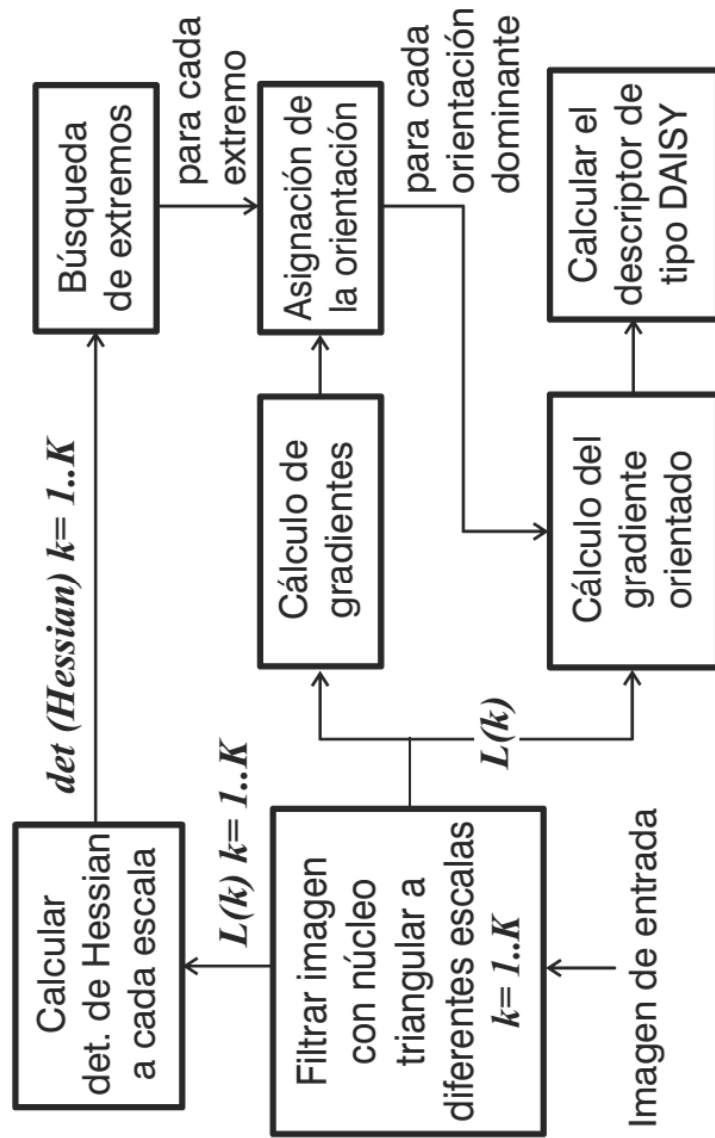


FIG. 1

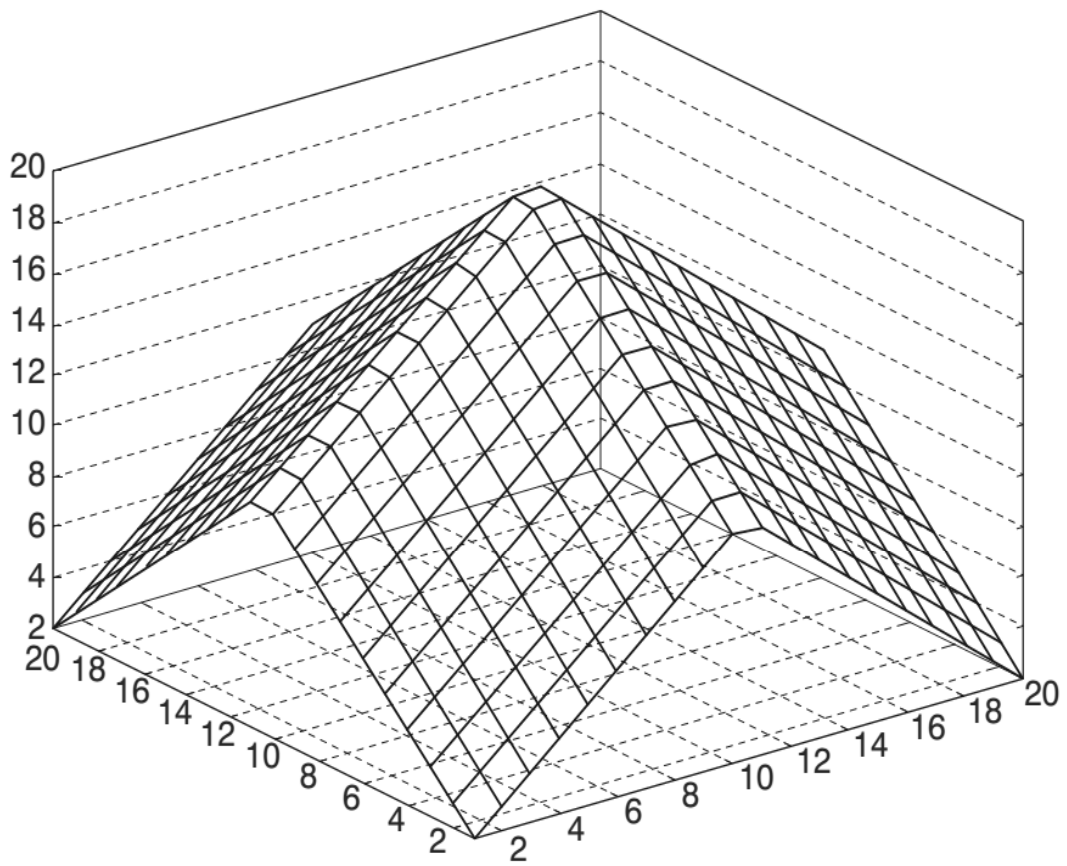


FIG. 2

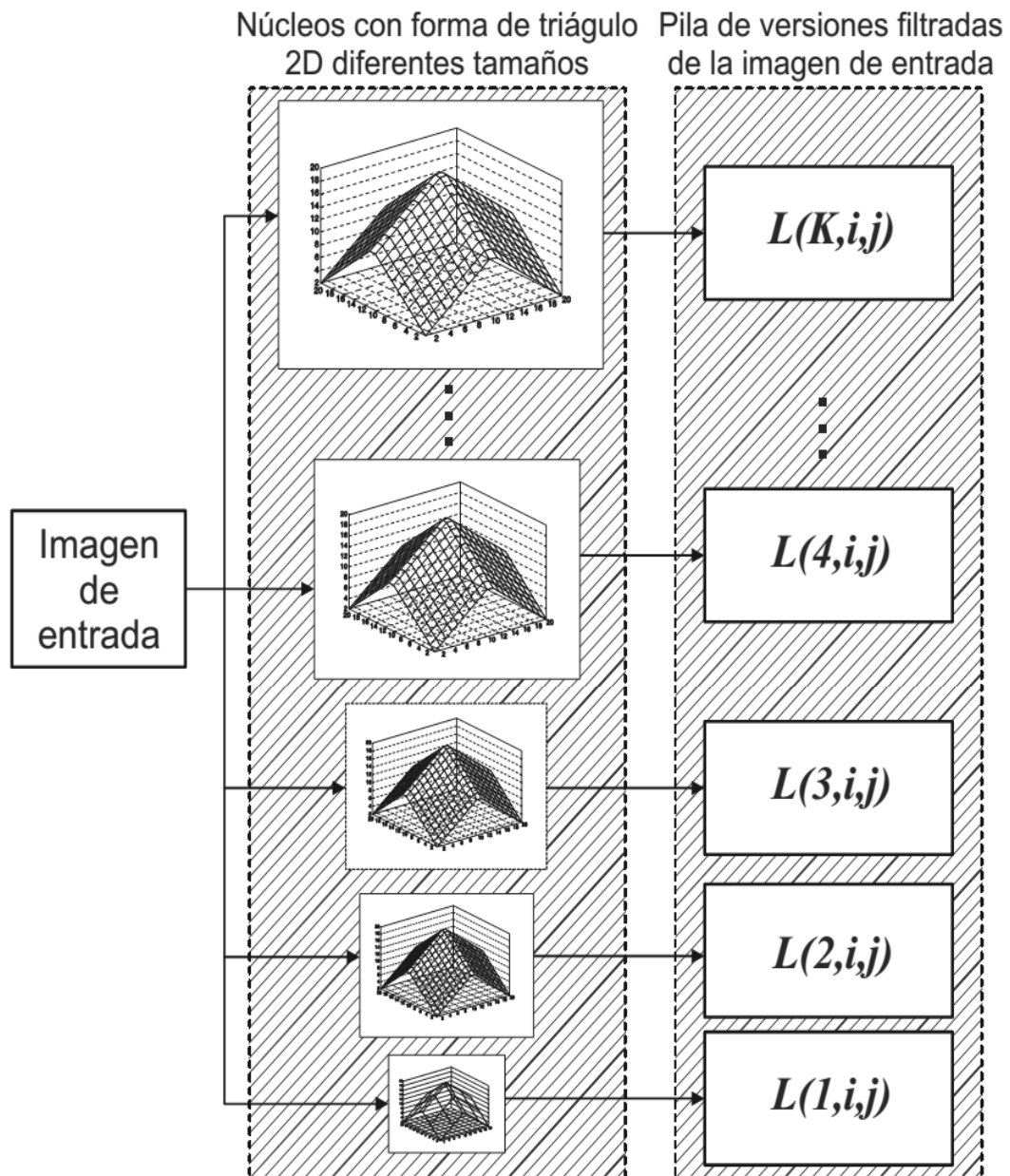


FIG. 3

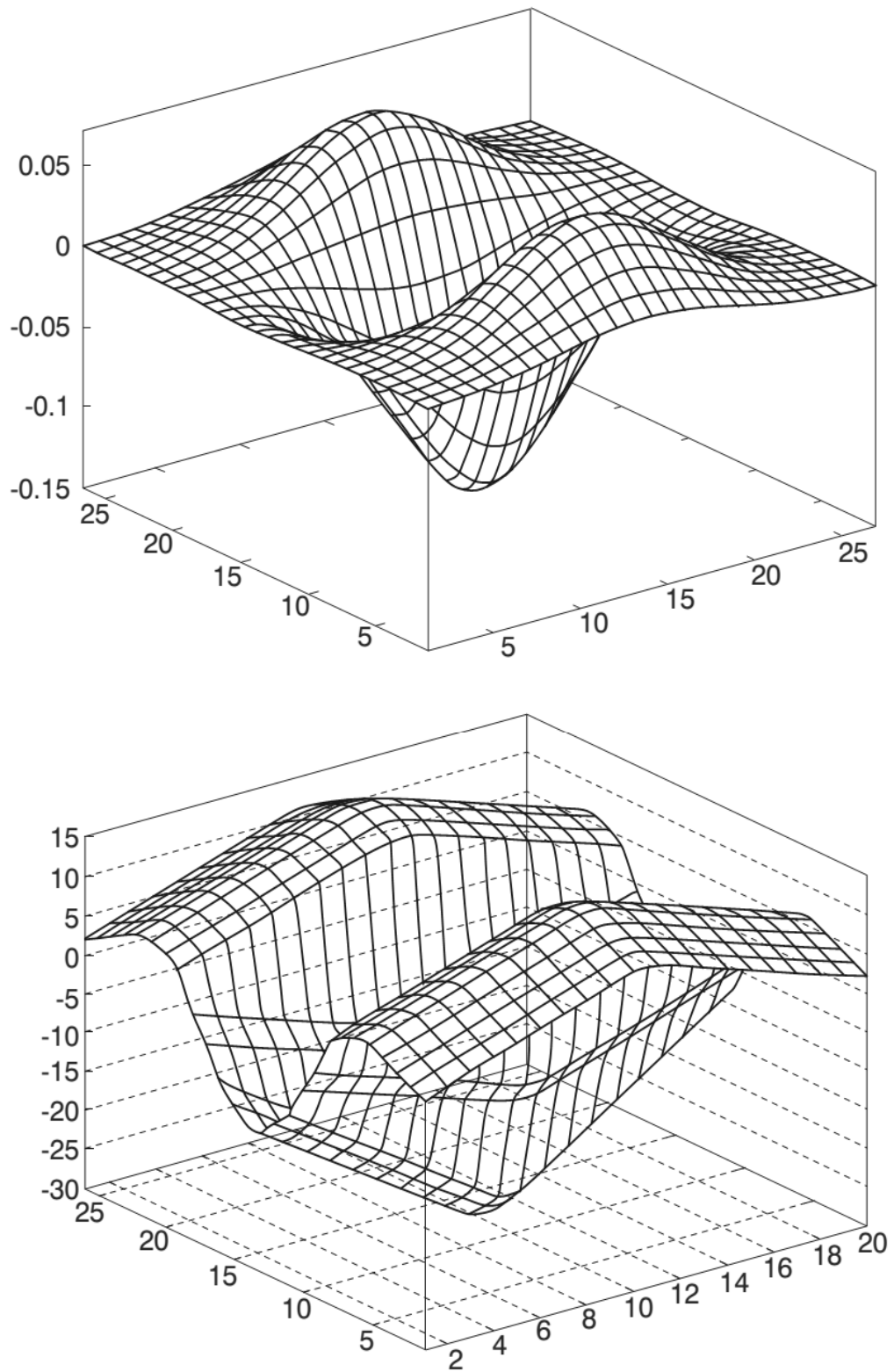


FIG. 4

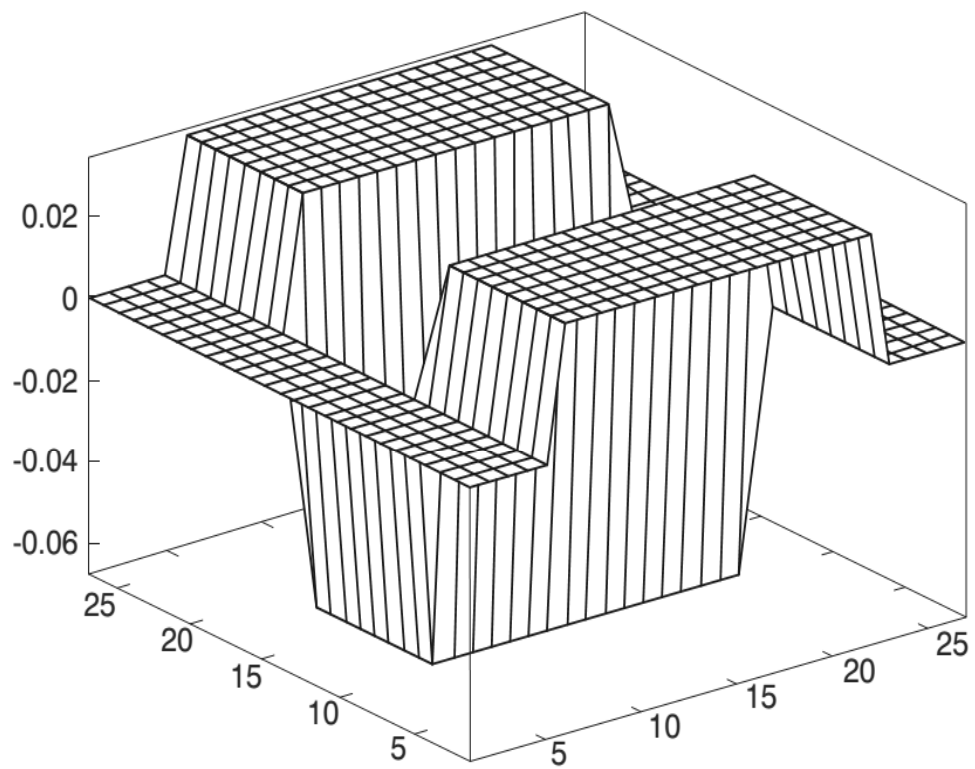


FIG. 5

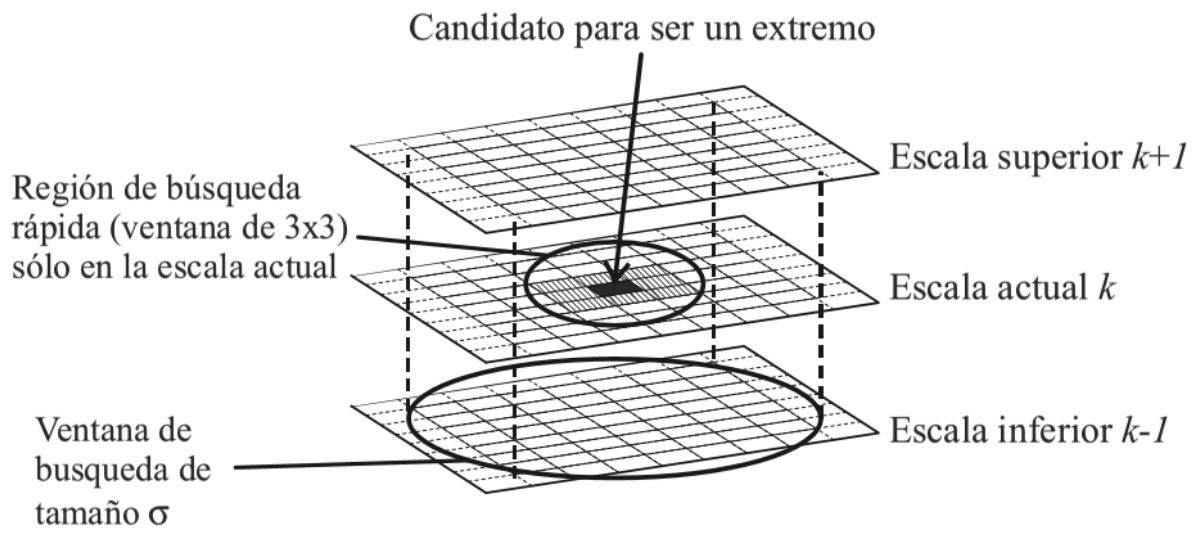


FIG. 6

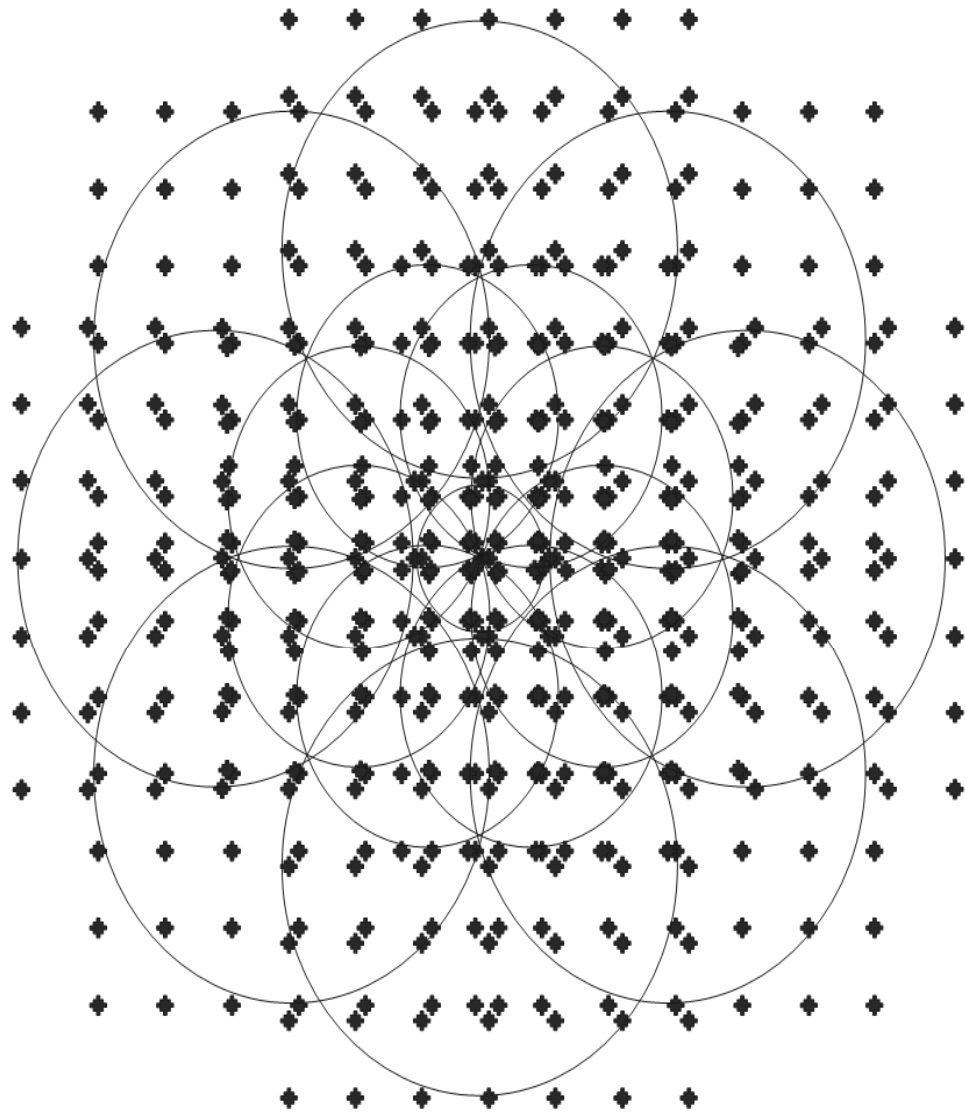


FIG. 7

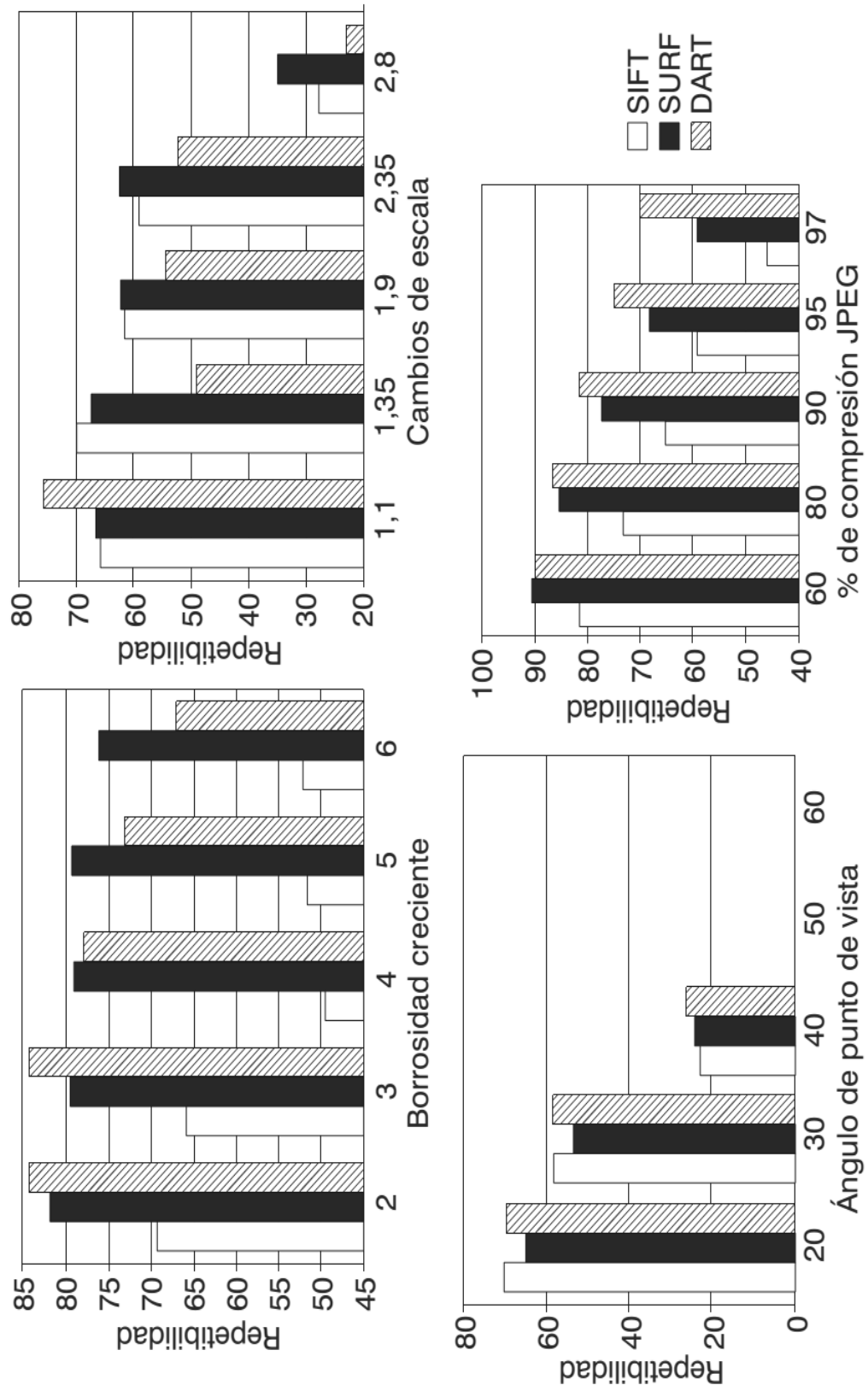


FIG. 8

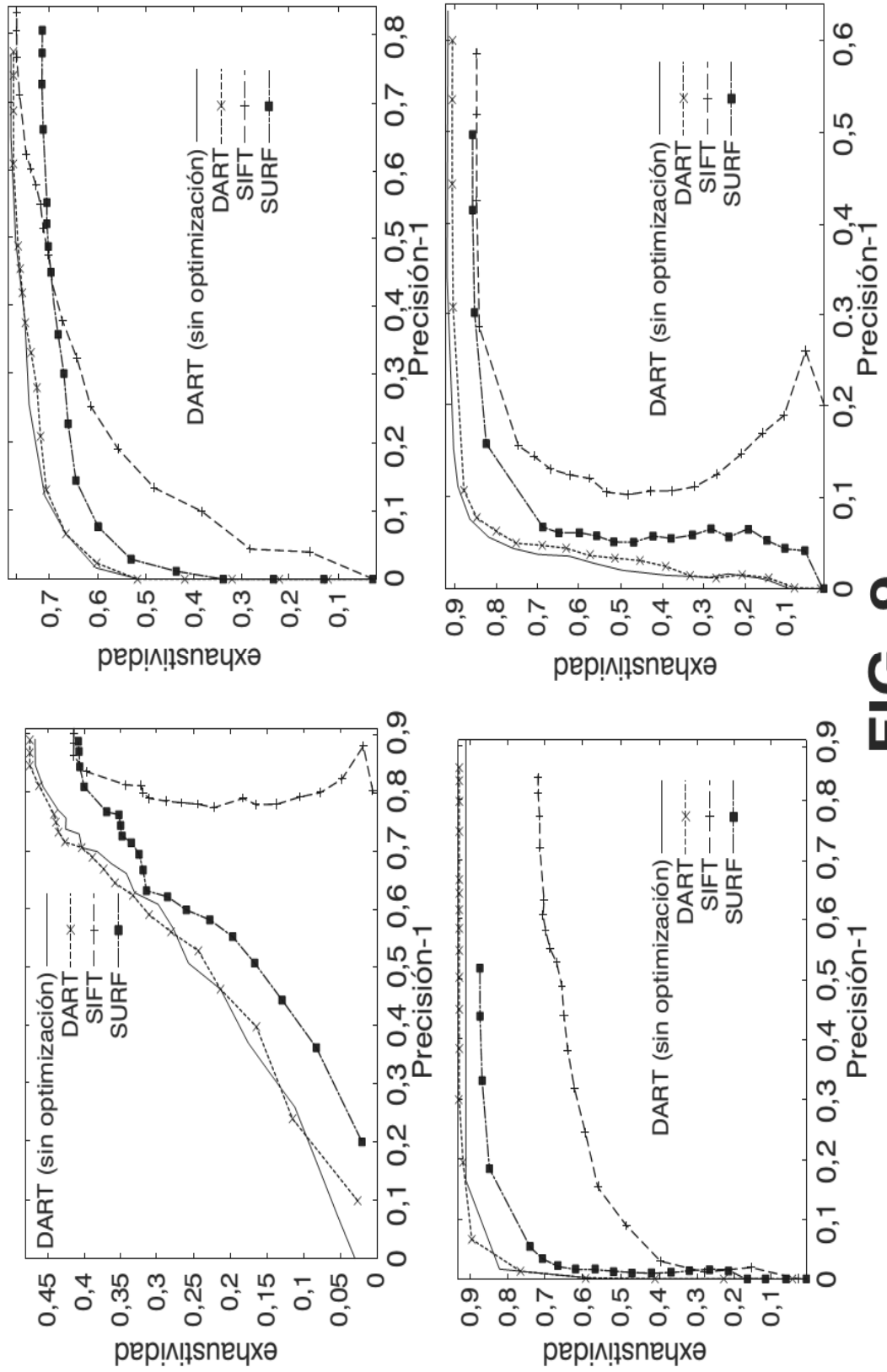


FIG. 9