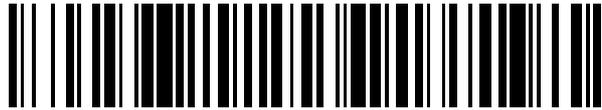


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 421 581**

21 Número de solicitud: 201200090

51 Int. Cl.:

**G06T 1/00** (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

**01.02.2012**

43 Fecha de publicación de la solicitud:

**04.09.2013**

Fecha de la concesión:

**09.12.2014**

45 Fecha de publicación de la concesión:

**16.12.2014**

56 Se remite a la solicitud internacional:

**PCT/ES2013/070037**

73 Titular/es:

**UNIVERSIDAD DE SANTIAGO DE COMPOSTELA (33.3%)**

**Edificio Emprendia- Campus Vida  
15782 Santiago de Compostela (A Coruña) ES;  
UNIVERSIDAD DE SEVILLA (33.3%) y  
CONSEJO SUPERIOR DE INVESTIGACIONES  
CIENTÍFICAS (33.3%)**

72 Inventor/es:

**SUÁREZ CAMBRE, Manuel;  
BREA SÁNCHEZ, Víctor Manuel;  
PARDO SECO, Fernando Rafael;  
RODRÍGUEZ VÁZQUEZ, Ángel y  
CARMONA GALÁN, Ricardo**

74 Agente/Representante:

**PARDO SECO, Fernando Rafael**

54 Título: **Procesador de imágenes para extracción de características**

57 Resumen:

La presente invención se refiere a un procesador de imágenes para la extracción de características que comprende un único chip no planar que contiene una pluralidad de sensores integrados y recursos de procesamiento distribuidos en dos o más capas adaptado a capturar cuadros de imágenes y extraer características de las imágenes. En una realización particular, el chip no planar es un circuito integrado CMOS-3D (CI CMOS-3D) con una distribución vertical de los recursos de sensado y de procesado distribuidos en dos o más capas verticales de circuito integrado. El CI CMOS-3D implementa dos o más detectores de características en un único chip reutilizando una pluralidad de circuitos empleados para la obtención del gradiente y de puntos clave. Los detectores de características incluyen un detector invariante a transformaciones (SIFT), un detector de Harris y un detector basado en la matriz Hessiana.

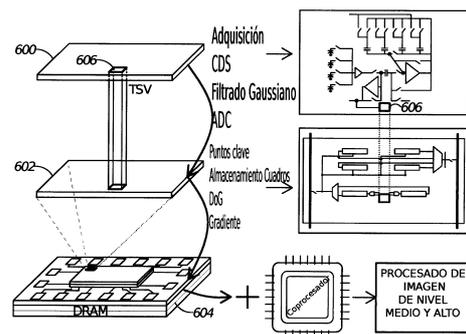


FIG. 6

ES 2 421 581 B1

DESCRIPCIÓN

**Procesador de imágenes para extracción de características**

**Sector de la Técnica**

Las realizaciones presentadas se refieren a procesadores de imágenes. Específi-  
5 camente, se refieren a procesadores para la adquisición de imágenes y detección de características

**Antecedentes de la invención**

Los procesadores de imágenes convencionales siguen una estructura en la que operan cuadro a cuadro, es decir, primero se capturan los cuadros, posteriormente se  
10 codifican al dominio digital y finalmente se procesan. Esta aproximación se beneficia del enorme poder de cálculo de los procesadores digitales en tecnologías de alta integración, pero no es la aproximación más eficiente en términos de velocidad de procesado (tiempo transcurrido entre la entrada y las acciones) ni en términos de consumo de energía.

15 Los detectores de características invariantes a cambios de escala y rotación se emplean en distintas tareas de procesado de imágenes, como son la detección de objetos y su clasificación, recuperación de imágenes, registro de imágenes o tareas de seguimiento (tracking). Su naturaleza invariante les confiere repetibilidad, lo que permite trabajar en ambientes con oclusión, o con escenas adquiridas bajo distintas  
20 condiciones, como iluminación o captura desde distintos ángulos de visión. Los detectores de características modernos invariantes a cambios de escala y rotación, como es el detector de características invariantes a transformaciones de escala, del inglés Scale Invariant Feature Transform(SIFT), son técnicas de procesado de imágenes complejas con un elevado coste computacional, haciendo difícil su uso empleando  
25 procesadores de propósito general y software. Una parte importante de este algoritmo es la extracción de la pirámide Gaussiana, que comprende un conjunto de imágenes de diferentes resoluciones, denominadas octavas. Cada octava es el resultado de un escalado de 1/4 de la octava anterior. A su vez, cada octava está formada

por una serie de imágenes llamadas escalas. Cada escala es el resultado de realizar un filtrado Gaussiano con un ancho dado (nivel- $\sigma$ ) en la escala anterior.

En Harris and Stephens (Proc. Avley Vis. Conf., Manchester, pp. 147-152, 1998) se propuso un algoritmo para detectar características, cuya principal ventaja era  
5 que los cálculos involucrados en la obtención de las características de las imágenes, llamadas puntos de Harris, no requieren un tiempo de cómputo elevado. Sin embargo, ofrece unos resultados pobres cuando trabaja con cambios de escala y rotación en las imágenes.

El SIFT, presentado en la U.S. Pat. No. 6,711,293, es un algoritmo de procesado  
10 de imágenes que obtiene características invariantes a cambios de escala y rotación de imágenes digitales. Su principal limitación viene dada por el elevado coste computacional del método, que dificulta su uso en aplicaciones que requieran operar en tiempo real (por ejemplo operar a una tasa de procesado de cuadros de 24 cuadro por segundo o más).

15 Para evitar las limitaciones el algoritmo SIFT original, en Bonato et al., (IEEE Trans Circuits Syst., 18(12), pp. 1703-1712, 2008) se presenta una implementación sobre una FPGA, del inglés Field Programmable Gate Array. Sin embargo, la adquisición de las imágenes no está integrada con los elementos de procesamiento, lo que ralentiza el procesado. Además, el elevado consumo de potencia de las FPGAs hace  
20 que sea difícil integrarlas en un sistema de bajo consumo de potencia para tareas de visión por computador.

En Yao et al. (International Conference on Field-Programmable Technology, 2009. FPT 2009) se presenta otra implementación FPGA del algoritmo SIFT. Como en el caso de Bonato et al., la adquisición de la imagen no se ha tenido en cuenta a  
25 la hora de desarrollar el sistema.

En Kiyoyama et al., (IEEE International Conference on 3D System Integration, 2009) se lleva a cabo el análisis del rendimiento de un circuito paralelo de procesado de señal, que incluye un circuito a nivel de píxel y un convertidor analógico digital (ADC) con muestreo doble correlacionado, del inglés correlated double sampling  
30 (CDS), jerárquico. Este trabajo se centra en la adquisición de la imagen pero no trata el problema de la creación de un núcleo de procesado.

## Descripción de la invención

Las realizaciones mostradas incluyen un procesador de imágenes para extracción de características que comprende un único chip no planar que contiene una pluralidad de elementos de sensado y procesado distribuidos en dos o más capas adaptados a capturar cuadros de imágenes y extraer características de imágenes. De acuerdo a una realización particular, el chip no planar es un circuito integrado CMOS-3D con una distribución vertical de los elementos de sensado y de procesado a lo largo de dos o más circuitos integrales dispuestos verticalmente. En particular, el circuito integrado CMOS-3D implementa dos o más detectores de características en un único chip. En una realización, y sin limitar su alcance, estas características son invariantes a transformaciones (SIFT), detector de puntos Harris y detector de características basado en la matriz Hessiana, o una combinación de todas ellas. En particular, el circuito integrado CMOS-3D implementa los detectores de características reutilizando una pluralidad de circuitos empleados para la obtención del gradiente y de los puntos clave (keypoints). A modo de ejemplo, en una realización, el circuito integrado implementa una arquitectura procesador-por-píxel para alcanzar un paralelismo masivo a nivel de píxel, e incluye un array de procesadores en píxel y una pluralidad de circuitos que en paralelo realizan el SIFT, detectan puntos de Harris, y puntos basado en la Hessiana. Ciertas realizaciones incluyen un red de capacidades en conmutación (switched capacitor network) multipropósito. Esta red de capacidades en conmutación multipropósito implementa la generación de una pirámide Gaussiana, muestreo doble correlacionado (CDS), y actúa como memoria analógica a nivel de píxel. Una realización particular de la red de capacidades en conmutación emplea una configuración doble *Forward-Euler*. De manera similar, el circuito integrado CMOS-3D incluye circuitos compartidos reutilizados para la conversión analógica-digital en píxel y el almacenamiento de los cuadros. En una realización particular, y sin limitar el alcance, los circuitos del circuito integrado se disponen en 4 capas con (a) una capa para adquisición, muestreo doble correlacionado, filtrado Gaussiano y conversión analógica-digital; (b) un nivel para el almacenamiento de los cuadros, la detección de puntos clave, cálculo de diferencia-de-Gaussianas (DoG), cálculo de gradiente, y conversión analógica-digital compartida; y (c) dos niveles con memoria

dinámica de acceso aleatorio, del inglés Dynamic Random Access Memory (DRAM). De igual manera, se muestra un método para obtener características de las imágenes de forma eficiente usando un circuito integrado. El método comprende: (a) implementar en un único chip no planar una pluralidad de elementos de sensado y  
5 procesado adaptados a la captura de imágenes y extracción de características y distribuidos en dos o más capas; y (b) reutilizar la pluralidad de circuitos en el chip no planar para realizar un cálculo concurrente del gradiente y la detección de los puntos clave; y el muestreo doble correlacionado (CDS) y la conversión analógica-digital. En una realización particular, el chip no planar implementa dos o más detectores de  
10 características en un único chip. Por ejemplo, el detector de características puede ser un detector de características invariante a cambios de escala (SIFT), un detector de puntos Harris, un detector de puntos dados por la matriz Hessiana, o una combinación de todos ellos.

En una realización particular, el procesador de imágenes descrito se puede usar  
15 como un procesador de visión CMOS-3D para ejecutar distintos detectores de características. El procesador de imágenes permite al usuario seleccionar el detector de características más apropiado de acuerdo a las necesidades de la aplicación. En una realización, el procesador de imágenes ejecuta dos modos principales: 1) Modo SIFT, proporcionando una alta precisión pero con el coste de una baja velocidad, y  
20 2) Detectores de puntos Harris y basado en la matriz Hessiana, que son más veloces a costa de ofrecer una menor precisión. Esos dos modos son posibles gracias a: 1) la arquitectura CMOS-3D, y 2) el hecho de que ejecutar el SIFT implica ejecutar algunas de las operaciones requeridas para el detector de puntos Harris o detector basado en la matriz Hessiana. En concreto, la realización particular, y sin limitar  
25 el alcance, incluye una arquitectura de pixel con CDS a nivel de píxel, y conversión analógica-digital a nivel de píxel mediante un convertidor analógico-digital de rampa simple de 8 bits. La reutilización de los diferentes circuitos permite obtener una menor área por píxel. Además, de acuerdo a una realización, la arquitectura incluye la asignación de una estructura de 3 transistores de sensores activos de fotodiodos,  
30 del inglés active photodiode sensor (APS), por procesador, proporcionando una capacidad de procesamiento masivamente paralela, muy adecuada para operaciones a nivel de píxel, operaciones muy frecuentes en cualquier detector de características.

En una realización particular, la arquitectura se implementa con una pila CMOS-3D de dos niveles. El nivel superior contiene los píxeles. Cada píxel se completa con los circuitos necesarios para la red de capacidades en conmutación. Esta red de capacidades en conmutación implementa un filtrado Gaussiano, que se emplea  
5 en varios detectores de características. En un ejemplo de realización más particular, y sin limitación, el circuito integrado se diseña con una arquitectura CMOS-3D de dos niveles para la generación de la pirámide Gaussiana, detección de extremos, y cálculo de las derivadas espaciales de una imagen. Estas tareas están incluidas en muchos detectores de características modernos, que a su vez pueden ser usados para  
10 aplicaciones tales como detección de objetos, registro de imágenes o seguimiento. En una realización del nivel superior, su arquitectura contiene los circuitos de adquisición de imágenes en un array de  $320 \times 240$  de sensores activos de fotodiodos (active photodiode sensors, APS) y un array más reducido de  $160 \times 120$  procesadores analógicos para el procesamiento de la imagen a bajo nivel. El nivel superior comprende  
15 de la realización en píxel CDS, una red de capacidades en conmutación para la generación de la pirámide Gaussiana, memorias analógicas y un comparador para la conversión en píxel analógica digital. La reutilización de circuitos para distintas funciones permite obtener una menor área por cada píxel. De acuerdo a una realización, el nivel inferior de la arquitectura contiene un almacenamiento de cuadros  
20 con una correspondencia uno-a-uno con los procesadores analógicos del nivel superior, la circuitería digital necesaria para la detección de extremos y el cálculo de la primera y segunda derivadas espaciales de la imagen, así como detectores de características Harris y Hessiana. A modo de ejemplo, y sin limitación, la arquitectura de dos niveles se puede implementar en un chip CMOS-3D en tecnología de 130 nm.  
25 Un chip auxiliar puede realizar las tareas de alto nivel así como las comunicaciones. Empleando esta tecnología se puede obtener un área de  $300 \mu m^2$  por píxel o menor ( $1200 \mu m^2$  por procesador analógico) y permite obtener un ratio de cuadros de 180 cuadros/s para un tiempo de conversión analógico digital de  $120 \mu s$ .

## Breve descripción de las Figuras

Las modalidades detalladas se ilustran a modo de ejemplo, y no a modo de limitación, en las figuras de los esquemas adjuntos.

La **FIG. 1** muestra un diagrama de bloques de una realización del procesador de  
5 imágenes.

La **FIG. 2** muestra una realización del bloque de adquisición mostrando la pluralidad de elementos de sensado.

La **FIG. 3** muestra un diagrama de bloques de una realización del bloque de sensado.

10 La **FIG. 4** muestra un diagrama de bloques de una realización del generador del espacio de escalas.

La **FIG. 5** muestra un diagrama de bloques de una realización del extractor de características.

La **FIG. 6** muestra una realización de la arquitecturas CMOS-3D.

15 La **FIG. 7** muestra una realización de un elemento de procesamiento del nivel que realiza la adquisición de un píxel y el filtrado Gaussiano.

La **FIG. 8** muestra un ejemplo del cronograma de las señales de control para el nivel de acuerdo a una realización.

La **FIG. 9** ilustra la conexión entre los niveles de acuerdo a una realización.

20 La **FIG. 10** muestra el esquemático de la red de capacidades en conmutación de acuerdo a una realización.

La **FIG. 11** muestra el esquemático de la red de difusión para una rejilla de 16x16 píxeles de acuerdo a una realización.

La **FIG. 12** muestra una realización del elemento de memoria del almacenamiento  
25 de cuadros.

La **FIG. 13** muestra la arquitectura de un segundo nivel de acuerdo a una realización.

La **FIG. 14** muestra el diagrama de bloques de los elementos que calculan el gradiente, puntos de Harris, puntos Hessianos y DoG.

## Ejemplos de realización de la invención

Las realizaciones mostradas incluyen un procesador de imágenes para la extracción de características que comprende un bloque de adquisición **100**, un bloque de generación del espacio de escalas **102**, y un bloque para detectar características **106**.

5 En esta realización particular, el bloque de adquisición, que obtiene una imagen de una escena para que el procesador la analice, comprende una pluralidad de elementos de sensado **200**. Cada elemento de sensado adquiere una parte de la escena **302** que va a ser analizada **300** que es transformada en un valor analógico **304** y luego se almacena **306**. La pluralidad de valores analógicos adquiridos por la pluralidad  
10 de elementos de sensado forman la imagen de la escena.

El bloque de generación del espacio de escalas genera el espacio de escalas, es decir, un conjunto de imágenes borrosas obtenidas tras la aplicación de filtros Gaussianos con una  $\sigma$  determinada.:

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (1)$$

donde  $G(x, y, k\sigma)$  es un filtro Gaussiano con una desviación estándar  $k\sigma$ ,  $I(x, y)$  es  
15 la imagen adquirida compuesta por la pluralidad de valores analógicos adquiridos por el bloque de adquisición,  $L(x, y, k\sigma)$  es la imagen borrosa obtenida, y  $*$  es la operación de convolución. De acuerdo a una realización, un proceso de difusión **402** consiste en aplicar un filtro Gaussiano de acuerdo a la Eq.(1) al valor analógico adquirido **400** obtenido por el elemento de sensado. La pluralidad de valores obtenidos  
20 de la imagen borrosa **404** es convertida a valores digitales mediante un convertidor analógico-digital **406**. El espacio de escala está compuesto de imágenes borrosas, cuyo nivel de difuminado se incrementa aplicando iterativamente un filtro Gaussiano con una determinada  $\sigma$ .

El valor digital de las imágenes borrosas **500** se transfiere a un bloque de memoria  
25 **502** que comprende una pluralidad de elementos de memoria. Los valores almacenados se envían a los distintos bloques de obtención de características, que implican distintos modos de funcionamiento:

- 5
- **Detector de puntos Harris 504:** En este bloque, a partir de la primera imagen borrosa se lleva a cabo una detección de los puntos de Harris. La detección de esta característica involucra la realización de la siguiente operación sobre una imagen, y un punto es clasificado como punto de Harris si el siguiente valor es mayor que un umbral establecido:

$$Ha = (I_{gx}^2 I_{gy}^2 - I_{gxy}) - K(I_{gx}^2 + I_{gy}^2) \quad (2)$$

donde  $I_{gx}$  y  $I_{gy}$  son imágenes que se obtienen tras aplicar a la imagen adquirida los gradientes en las direcciones x e y respectivamente, seguido por un filtrado Gaussiano,  $I_{gxy}$  es el producto de las imágenes gradiente en las direcciones x e y seguida por un filtrado Gaussiano, y  $K$  es un parámetro predefinido.

- 10
- **Detector de puntos basado en la Hessiana 506:** En este bloque, a partir de las imágenes borrosas, se lleva a cabo una detección de puntos singulares a través de la matriz Hessiana. Un punto es clasificado como punto Hessiano si el siguiente valor es mayor que un umbral establecido:

$$He = I_{xx} I_{yy} - I_{xy}^2 \quad (3)$$

15

donde  $I_{xx}$  y  $I_{yy}$  son, respectivamente, las segundas derivadas espaciales de la imagen en las direcciones x e y, y  $I_{xy}$  es la derivada cruzada en las direcciones x e y.

- **Diferencia de Gaussianas, del inglés Difference of Gaussians, (DoG) 508:** En este caso se lleva a cabo una diferencia entre dos imágenes consecutivas del espacio de escalas, es decir:

$$DoG(x, y, k\sigma) = L(x, y, k\sigma) - L(x, y, (k - 1)\sigma) \quad (4)$$

20 De acuerdo a una realización, el detector de características se implementa como un circuito integrado CMOS-3D. La tecnología CMOS-3D permite integrar más funcionalidad en el mismo chip gracias a los distintos niveles de integración que permite. La arquitectura de tal implementación comprende dos capas también denominadas niveles **600** y **602**, donde la funcionalidad deseada es implementada y dos niveles  
25 adicionales de memoria DRAM **604** para almacenar los datos. La conexión entre

los dos niveles superiores se lleva a cabo mediante conexiones directas denominadas vías a través de silicio, del inglés Through-Silicon-Vias (TSV), **606**. El nivel superior incluye la adquisición de la imagen y la generación del espacio de escalas (generación de la pirámide Gaussiana y su digitalización). El segundo nivel **602** comprende un

5 bloque para detectar características.

El nivel 1 comprende una pluralidad de elementos de procesamiento, en el que cada uno es responsable de la captura de una imagen libre de offset, unión de píxeles y difusión en cooperación con los elementos vecinos, y contribuyen a la realización paralela de la conversión analógica-digital mediante un convertidor de rampa simple. Para reducir el consumo de área, se asignan 4 fotodiodos a cada elemento. La

10 adquisición de la imagen se lleva a cabo usando un Sensor Activo de Píxel de tres transistores, del inglés 3T Active Pixel Sensor (APS), **700**, que junto con los condensadores de estado **702**, y el condensador C **710**, llevan a cabo el muestreo doble correlacionado (CDS) de acuerdo a una realización particular, y sin limitación. El

15 resultado de la adquisición se almacena en la correspondiente memoria analógica  $C_{Si}$  **702** que está dado por:

$$V_{C_{Si}} = V_{ref} + \frac{C}{C_{Si}}(V_S(t_0) - V_S(t_1)) - V_Q \quad (5)$$

donde  $V_{ref}$  es un valor analógico de referencia,  $V_S(t_0)$  y  $V_S(t_1)$  son los valores adquiridos por los fotodiodos  $S_i$ , **700** en el instante de tiempo  $t_0$  y  $t_1$  respectivamente, y  $V_Q$  es el punto de reposo del inversor, **726**. La adquisición se controla mediante

20 las señales  $\phi_{r_{-si}}$ , **701**,  $\phi_{vref_{-si}}$ , **704**,  $\phi_{acq}$ , **712**,  $\phi_{r_{d.m}}$ , **734**,  $\phi_{write_{-si}}$ , **704**, tal y como se muestra en la **FIG. 8**. Tal y como se muestra en el cronograma cuando se lleva a cabo el procesado del píxel 1 se activan las señales  $\phi_{rst\ 1}$ ,  $\phi_{acq}$  y  $\phi_{w\ d\ m}$ , de esta forma se transfiere el valor capturado por el fotodiodo al condensador de estado correspondiente, en este caso  $C_{s1}$ , **702**, donde es almacenado para su posterior digitalización

25 y realización del proceso de difusión. Ajustando el valor de  $\phi_{acq}$  a 0, y de  $\phi_{write_{-si}}$  a 1, los valores almacenados en la salida del inversor se pueden leer como:

$$V_{Si} = V_{ref} + \frac{C}{C_{Si}}(V_S(t_0) - V_S(t_1)) \quad (6)$$

De acuerdo a una realización particular, y sin limitación, se usa un convertidor analógico-digital (ADC) de 8 bits de rampa simple para convertir los valores analógi-

cos almacenados en valores digitales. El comparador **708** que lleva a cabo la conversión analógica-digital (A/D) se lleva a cabo con el inversor **726**, y reutilizando el condensador C **710** cuando la señal  $\phi_{acq}$  **712** está inactiva. El ADC se distribuye en dos niveles: el generador de la rampa analógica **900** y el comparador **902** en el nivel superior y un registro **908** y un contador digital **906** en el nivel inferior. Para llevar a cabo la conversión, el valor dado por la Eq.(6) se escribe en el condensador C, **710**, activando las señales  $\phi_{write_{si}}$  **706**,  $\phi_{write}$  **718** y  $\phi_{comp_{rst}}$  **728**. A continuación, este valor se compara con el valor de la rampa analógica global  $V_{ramp}$  **714** activando  $\phi_{conv}$  **716**. A, la salida del inversor **732** viene dada por:

$$V_{comp.out} = -K(V_{ramp} - V_{Si}) + V_Q \quad (7)$$

10 Cuando se produce un cruce por cero en el primer término de la Eq. (7) el comparador cambia el valor lógico a su salida. La salida del comparador es la señal que habilita o inhabilita la lectura de los registros del nivel inferior. Una TSV envía esta conversión a los registros que almacenan el código digital correspondiente.

La otra funcionalidad de los elementos de procesamiento del nivel superior es la de la realización del filtrado Gaussiano, o generación del espacio de escalas. Los bloques **722** y **724** trabajan conjuntamente con el condensador de estado **702** para llevar a cabo esta tarea. Los bloques periféricos están implementados con una red de capacidades en conmutación. Los conmutadores controlados por la señal  $\phi_{1/4}$  llevan a cabo el submuestreo 1/4, uniendo el valor de los cuatro condensadores de estado para la segunda octava.

El filtrado Gaussiano, necesario para la generación del espacio de escalas, es la solución de la ecuación de calor. De acuerdo a una realización, es implementado usando una red resistiva-capacitiva (RC). Una red capacitiva en configuración doble Forward-Euler tiene el mismo comportamiento que una red RC continua en el tiempo, excepto por el intercambio discreto de carga entre nodos vecinos. Un esquema de la red de capacidades en conmutación se muestra en la **FIG. 10(a)**. El intercambio de carga para llevar a cabo el proceso de difusión se realiza mediante los condensadores de estado **910**, y aquellos vecinos localizados a lo largo de las 4 direcciones cardinales. Este comportamiento discreto permite controlar el proceso de difusión, el valor de  $\sigma$ , de forma simple controlando el número de ciclos de las señales  $\phi_1$  **912**

y  $\phi_2$  **914**.

La anchura de la Gaussiana  $\sigma$  de cada ciclo se fija mediante la relación entre los condensadores de estado e intercambio ( $C/C_E$ ). En particular, el valor de un nodo en un ciclo  $n$  está dado por:

$$V_{ij}(n) = V_{ij}(n-1) + (V_{i-1j}(n-1) + V_{i+1j}(n-1) + V_{ij-1}(n-1) + V_{ij+1}(n-1) - 4V_{ij}(n-1)) \frac{\frac{C}{C_E}}{1 + 4\frac{C_E}{C}} \quad (8)$$

5 Por otro lado, se considera el valor del mismo nodo en una iteración con un núcleo Gaussiano discreto, donde sólo existe interacción con los vecinos cardinales, que es modelado por:

$$V_{ij}(n) = V_{ij}(n-1) + (V_{i-1j}(n-1) + V_{i+1j}(n-1) + V_{ij-1}(n-1) + V_{ij+1}(n-1) - 4V_{ij}(n-1)) \frac{e^{-\frac{1}{2\sigma^2}}}{1 + 4e^{-\frac{1}{2\sigma^2}}} \quad (9)$$

De las Eqs. (8)-(9), el nivel por ciclo se puede escribir como:

$$\sigma_0 = \left( 2ln \frac{C}{C_E} \right)^{-\frac{1}{2}} \quad (10)$$

La aplicación de dos filtros Gaussianos sucesivos o núcleos con  $\sigma_0$  es equivalente a  
 10 un núcleo Gaussiano con una cierta  $\sigma$ . Esta propiedad permite obtener con un nivel de filtrado  $\sigma_0$  fijado por el cociente  $C/C_E$  aproximar cualquier núcleo Gaussiano mediante filtrado recursivo o aplicando núcleos Gaussianos de  $\sigma_0$ . La dependencia de  $\sigma$  con el número de ciclos  $\sigma = \sigma(n)$  viene dada por:

$$\sigma = \sqrt{\frac{2nC_E}{4C_E + C}} \quad (11)$$

Las S escalas de cada octava de la pirámide Gaussiana se generan con los mismos  
 15 S valores de  $\sigma$ . Los bloques periféricos A **724** y B **722** llevan a cabo la interacción con los vecinos a lo largo de las direcciones cardinales. **FIG. 10(b),(c)** muestran su estructura interna. Para cada condensador de estado el elemento de procesamiento tienen cuatro condensadores de intercambio **922**. La terminación o1 **920**, o2, **920** y o3 **918** en los nombres de las señales hacen referencia a las escalas en las

que son usadas. Después de ejecutar las difusiones de la primera octava, los cuatro condensadores de estado  $C_{Si}$  **702** se unen en un único condensador a través de los conmutadores controlados por  $\phi_{1/4}$  **720** llevando a cabo el submuestreo de 1/4 de la imagen. En esta situación, los bloques A **724** están cortocircuitados, y por lo tanto  
 5 el cociente  $C/C_E$  incrementa su valor por un factor de 2, cambiando la relación  $\sigma = \sigma(n)$  del sistema. Para preservar el valor de  $\sigma$ , se inhabilitan los condensadores  $C_{s3}$  y  $C_{s4}$  después de realizar la unión. Esto se lleva a cabo a través del conmutador  $\phi'_{1/4}$ , **730**.

La **FIG. 11** muestra la estructura de una red de  $16 \times 16$  píxeles, de acuerdo a una  
 10 realización. Para la tercera octava esto involucra un nuevo juego de conmutadores. Estos son controlados por la señal  $\phi_{1/16}$  **922**. Cuando  $\phi_{1/16}$  está activada, los valores almacenados en las cuatro celdas son promediados en un único píxel, llevando a cabo el submuestreo de la imagen original a otra con una resolución 1/4 de la original. Después de la unión, sólo una de cada cuatro celdas permanece habilitada **924**. El  
 15 filtrado se lleva a cabo con estas celdas activas en un proceso similar al descrito anteriormente mediante las líneas de conexión punteadas de la **FIG 11 926**.

La conversión A/D se lleva a cabo compartiendo recursos entre los dos niveles superiores. Estando distribuidos el comparador en el nivel superior, y las señales del contador global (generación del código digital) entre los registros del nivel inferior,  
 20 así se necesita una TSV para cada conjunto de 4 píxeles, que forman lo que se denomina un elemento de procesamiento en el nivel superior. La misma estructura en forma de array se replica en el nivel inferior, permitiendo hacer de forma sencilla el emparejamiento de celdas entre el nivel superior e inferior. Así, los píxeles digitalizados se escriben en un conjunto de registros de tamaño  $M/2 \times N/2$ , donde  $M$  y  
 25  $N$  son los tamaños de la imagen adquirida. Cada uno de estos conjuntos de registros comprende 6 registros de 8 bits **928**, **930**, **940**, **942**, **944** y **946**. Dos de ellos realizan la conversión de la escala  $k$ , **928** y **930**, junto con el comparador del nivel superior, donde  $k$  indica la escala dentro de una octava. Se necesitan dos registros para permitir la conversión de un píxel mientras los otros están siendo leídos para  
 30 su posterior procesado. El resto de los registros, **940**, **942**, **944** y **946**, almacenan los valores de la escala previa  $(k-1)$ . De esta forma la imagen completa  $M \times N$  se almacena en el nivel inferior. Estos píxeles se denotan como:  $P1$  ,  $P2$  ,  $P3$  y  $P4$ ,

que corresponden con las localizaciones  $(i,j)$ ,  $(i, j + 1)$ ,  $(i + 1, j)$  y  $(i + 1, j + 1)$ , respectivamente, donde  $i$  indica la fila y  $j$  indica la columna dentro de la imagen. Los cuatro píxeles  $P1 - P4$  se digitalizan en serie, ya que sólo hay un TSV por cada 4 elementos de procesamiento. Esto implica que todos los píxeles  $P1$  se digitalizan

5 en un primer ciclo de conversión,  $P2$  en un segundo ciclo de conversión, y así para los píxeles  $P3$  y  $P4$ . Por lo tanto, se necesitan cuatro ciclos de conversión para la digitalización de toda la imagen en la primera octava. Hay una señal de 1 bit que controla dos puertas AND, **932** y **934**, con  $\phi_{conv13}$  **950**, y  $\phi_{conv24}$  **951** como entradas, controlando la señal de habilitación para los dos registros superiores R13\_K **928** y

10 R24\_K **930** que reciben el valor a través de la TSV, **936**. Los dos registros superiores **928** y **930** almacenan los píxeles de la escala  $k$ . Los 4 registros inferiores, **940**, **942**, **944** y **946**, almacenan los píxeles  $P1 - P4$  de la escala  $k-1$ . Los multiplexores **938** y **948** envían las escalas  $k$  y  $k-1$  de los píxeles correspondientes a los respectivos buses para un posterior procesado, **931** y **933**. La secuencia de operaciones para obtener

15 cada escala en la primera octava se describe a continuación. El píxel  $P1$  se digitaliza en los registros R13\_K, **928**, con  $\phi_{conv13}$ , **960** activada. A continuación, el píxel  $P2$  se digitaliza y almacena en el registro R24\_K **930** siguiendo un proceso similar con la señal  $\phi_{conv24}$  **960** activada. Mientras  $\phi_{conv24}$  está activada, se lleva a cabo el cálculo de la DoG para todos los píxeles  $P1$  y se escriben en la memoria DRAM. Después

20 de la lectura de los píxeles  $P1$  el contenido del registro R13\_K, **928**, se transfiere al registro R1\_K\_1 **940** mediante la activación de la señal  $\phi_{WE1}$ , **952**. Posteriormente, los píxeles  $P3$  se digitalizan en el registro R13\_K **928**, mientras que los píxeles  $P2$  se leen y se transfieren a los registros R2\_K\_1 **942** mediante la señal de activación  $\phi_{WE2}$ , **954**. Las señales  $\phi_{WE3}$ , **956**, y  $\phi_{WE4}$ , **958**, juegan un papel similar en el

25 proceso de transferencia y almacenamiento de los píxeles  $P3$  y  $P4$  respectivamente. El proceso continua hasta llegar a los píxeles  $P4$ , completando la primera octava.

De acuerdo a una realización, el almacenamiento de cuadros **960** es un conjunto de  $M/2 \times N/2$  grupos de registros que almacenan las diferentes escalas del espacio de escalas. Después de cada difusión o filtrado Gaussiano de una imagen y su di-

30 gitalización, se ejecutan varias tareas en paralelo: 1) la digitalización del píxel  $P1$  en la escala  $k$ , 2) escalas  $S(k)$  **982**, 3) Diferencia de Gaussianas entre escalas  $k$  y  $k-1$  DoG( $k$ ) **980**, 4) gradiente horizontal a lo largo de las direcciones  $x$  e  $y$  para la

escala  $k$ ,  $d_x(k)$  **968** y  $d_y(k)$  **970**, así como las derivadas  $d_{xx}(k)$ ,  $d_{xy}(k)$  y  $d_{yy}(k)$  5) detección de puntos de Harris y Hessianos, **966**, en las escalas, **972** y 6) detección de puntos de Harris y Hessianos **976** sobre las DoG, **974**. Después los resultados se ordenan en grupos de 128 bits (16 palabras de 8 bits cada uno) y se transfieren  
 5 en modo ráfaga a la memoria DRAM. En una realización particular, las imágenes del almacenamiento de cuadros se leen en grupos de 20 registros fila a fila para proporcionar las 16 primeras y 16 segundas derivadas al mismo tiempo. Para cada fila  $i$ , las columnas de píxeles  $P_i$  se seleccionan mediante los 4 multiplexores, **962** y **964**. dos de ellos son compartidos por la primera y segunda octavas para las escalas  
 10  $k$  y  $k-1$ , **962**. Se requieren ambas escalas para el cálculo de las DoG. Los otros dos multiplexores **964** se emplean para la tercera octava. Debe tenerse en cuenta que para la primera y segunda octava los multiplexores pueden ser compartidos, ya que el acceso a los registros se hacen a lo largo de las filas. En el caso de la primera octava se necesitan 4 ciclos para realizar la lectura y transferencia de los píxeles  
 15  $P_1 - P_4$ . En la segunda octava y posteriores, la transferencia se lleva a cabo en un único ciclo debido al submuestro de  $1/4$  y posteriores.

El cálculo del gradiente es una operación muy común en el procesado de imágenes. Además, las primeras derivadas se usan en tareas subsiguientes como son la obtención de la orientación y el vector descriptor de cada punto clave en el algoritmo  
 20 SIFT. Además, las primeras derivadas se pueden usar para el detector de puntos Harris.

De acuerdo a una realización particular, debido a la asignación de 4 píxeles a un procesador en el nivel superior los mecanismos de lectura no permiten obtener las derivadas primeras  $d_x$  y  $d_y$  en un único ciclo. Por consiguiente, esto se logra calculando  
 25 el gradiente a lo largo de un conjunto diferente de ejes que han sido rotados  $45^\circ$  con respecto a los ejes convencionales  $x$  e  $y$ . Así, el gradiente se calcula de acuerdo al siguiente conjunto de ecuaciones:

$$d_x(i, j) = I(i + 1, j + 1) - I(i - 1, j - 1) \quad (12)$$

$$d_y(i, j) = I(i + 1, j - 1) - I(i - 1, j + 1) \quad (13)$$

El resultado del detector basado en la Hessiana **982** y el detector de puntos Harris  
 30 **984** son dos imágenes de tamaño  $M \times N$  y de 1 y 2 bits por píxel respectivamente.

"1" implica un extremo y "0" es un punto sin información significativa para la Hessiana. El algoritmo Harris tiene tres estados: "00" es una esquina, "01" es un bordez "1X" un punto sin información relevante. El bloque **986** de la **FIG. 14** lee los valores del los registros del almacenamiento de cuadros, **987**, y realiza un cálculo  
 5 por ciclo (un píxel).

Otros detectores están basados en la localización de puntos característicos a través de la matriz Hessiana **982**. Para obtener esta matriz se necesita realizar las segundas derivadas. El cálculo de estas derivadas requiere una vecindad de orden 4. Como se mencionó previamente, en el proceso de lectura la realización mostrada, un píxel  
 10 no tiene disponibles los vecinos para calcular las primeras derivadas a lo largo de los ejes convencionales x e y (direcciones vertical y horizontal). Sin embargo, es posible realizar esta operación con los píxeles localizados a una distancia de dos píxeles. Una aproximación para la segunda derivada se realiza generando los vecinos localizados a una distancia de un píxel a lo largo de las direcciones horizontal y  
 15 vertical interpolando los píxeles situados a una distancia de dos píxeles con respecto al píxel que está siendo procesado. Así, en este proceso, el vecino en (i+1; j) se genera como  $I(i+1; j) = [I(i+2; j)+I(i; j)]/2$ . Con esta aproximación las segundas derivadas vienen dadas por:

$$d_{xx}(i, j) = I(i, j + u) + I(i, j - u) - vI(i, j) \quad (14)$$

$$d_{yy}(i, j) = I(i + u, j) + I(i - u, j) - vI(i, j) \quad (15)$$

$$d_{xy}(i, j) = I(i+u, j+u)+I(i+u, j-u)+I(i-u, j+u)+I(i-u, j-u)-2vI(i, j) \quad (16)$$

donde  $u = 2$  y  $v = 1$  para la primera octava, y  $u = 1$  y  $v = 2$  para las siguientes octavas, dado que en la segunda y tercera octavas cada píxel tiene los vecinos a lo largo de las direcciones vertical y horizontal para calcular el gradiente a lo largo de los ejes x e y convencionales, y en donde nos es necesario llevar a cabo la interpola-  
 25 ción.

## REIVINDICACIONES

1. Un procesador de imágenes para detección de características que comprende un único chip no planar que contiene una pluralidad de elementos integrados de sensado y procesamiento distribuidos en dos o más capas adaptados para la  
5 captura de cuadros de imágenes y la extracción de características de imágenes.
2. El procesador de imágenes de la reivindicación 1, donde el chip no planar es un circuito integrado CMOS tridimensional con una distribución vertical del sensado y los elementos de procesamiento en dos o más capas de circuitos integrados.
- 10 3. El procesador de imágenes de la reivindicación 2, donde el circuito integrado CMOS tridimensional implementa dos o más detectores de características en un único chip.
4. El procesador de imágenes de la reivindicación 3, donde los detectores de características son características invariantes a cambios de escala (SIFT), detectores  
15 de puntos Harris, detector basado de la matriz Hessiana, o una combinación de todos ellos.
5. El procesador de imágenes de la reivindicación 4, donde el circuito integrado CMOS tridimensional implementa dos o más detectores de características reutilizando una pluralidad de circuitos usados para la extracción del gradiente  
20 y puntos clave.
6. El procesador de imágenes de la reivindicación 5, donde el circuito integrado CMOS tridimensional implementa una arquitectura píxel-por-procesador para obtener un paralelismo masivo a nivel de operación de píxel.
7. El procesador de imágenes de la reivindicación 6, donde el circuito integrado  
25 CMOS tridimensional comprende un array de procesadores en píxel y una pluralidad de circuitos que de forma paralela implementan los detectores de características SIFT, detector de puntos Harris, y detector de puntos basados en la matriz Hessiana.

8. El procesador de imágenes de la reivindicación 7, donde la pluralidad de circuitos incluye una red de capacidades en conmutación multipropósito.
9. El procesador de imágenes de la reivindicación 8, donde la red de capacidades en conmutación multipropósito implementa la generación de la pirámide Gaussianas, muestreo doble correlacionado (correlated double sampling, CDS) en píxel, y una memoria analógica a nivel de píxel.
10. El procesador de imágenes de la reivindicación 9, donde la red de capacidades en conmutación multipropósito presenta una configuración doble Forward-Euler.
- 10 11. El procesador de imágenes de la reivindicación 10, donde el circuito integrado CMOS tridimensional incluye la reutilización de circuitos compartidos para la conversión en píxel analógica-digital y el almacenamiento de cuadros.
12. El procesador de imágenes de la reivindicación 11, donde los circuitos del circuito integrado CMOS tridimensional están dispuestos en capas con (a) un nivel para la adquisición, CDS, filtrado Gaussiano y conversión analógica-digital; (b) un nivel para el almacenamiento de cuadros, extracción de puntos clave (keypoints), cálculo de la diferencia de Gaussianas (Difference-of-Gaussian, DoG), cálculo del gradiente y compartición de la conversión analógica-digital; y (c) dos niveles con memoria dinámica de acceso aleatorio, del inglés Dynamic Random Memory Access (DRAM).
- 15 20 13. Un método para extraer de forma eficiente características de imágenes mediante un circuito integrado que comprende:
  - (a) implementar un único chip no planar que contiene una pluralidad de elementos integrados de sensado y de procesamiento distribuidos en dos o más niveles adaptados a la captura de cuadros de imágenes y extracción de características de imágenes; y
  - (b) reutilizar una pluralidad de circuitos en el chip no planar para obtener una implementación concurrente del cálculo del gradiente y de los puntos clave; CDS y conversión analógica-digital.
- 25

14. El método de la reivindicación 13, donde el chip no planar implementa dos o más detectores de características en un único chip.
15. El método de la reivindicación 14, donde las características son características invariantes a cambios de escala (SIFT), detector de puntos de Harris, detector  
5 basado en la Hessiana, o una combinación de todos ellos.
16. El método de la reivindicación 15, donde el circuito integrado CMOS tridimensional implementa un procesador por píxel para obtener un paralelismo de operación masivo a nivel de píxel; el circuito integrado CMOS tridimensional comprende un array de procesadores en píxel y una pluralidad de circuitos que  
10 realizan en paralelo el SIFT, detección de puntos Harris, detección de puntos basada en la Hessiana; la pluralidad de circuitos incluye una red de capacidades en conmutación multipropósito; la red de capacidades en conmutación multipropósito implementa la generación de la pirámide Gaussiana, el muestreo doble correlacionado (CDS) en píxel, y una memoria analógica a nivel de  
15 píxel; la red de capacidades en conmutación multipropósito tiene una configuración doble Forward-Euler; el circuito integrado CMOS tridimensional incluye recursos compartidos para la conversión analógica-digital en píxel y el almacenamiento de cuadros; los circuitos del circuito integrado CMOS tridimensional están dispuestos en cuatro capas con (a) un nivel para la adquisición, muestreo doble correlacionado, filtrado Gaussiano y conversión analógica-digital;  
20 (b) un nivel para el almacenamiento digital de cuadros, extracción de puntos clave, cálculo de la DoG, cálculo del gradiente, y compartición de la conversión analógica-digital; y (c) dos niveles con memoria DRAM.

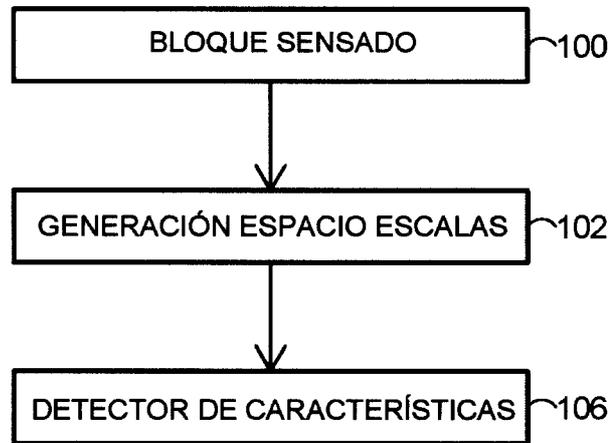


FIG. 1

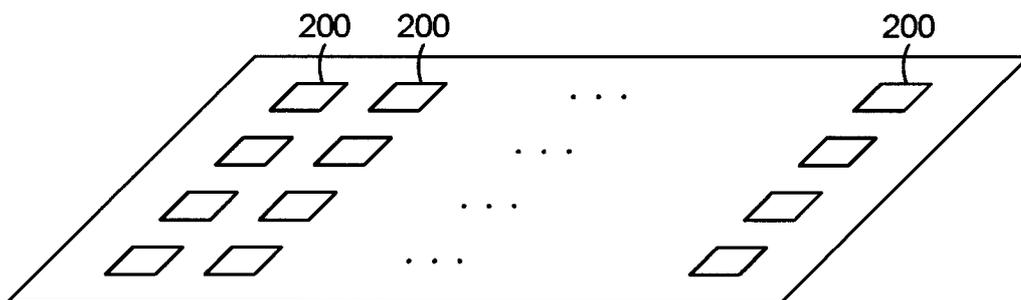


FIG. 2

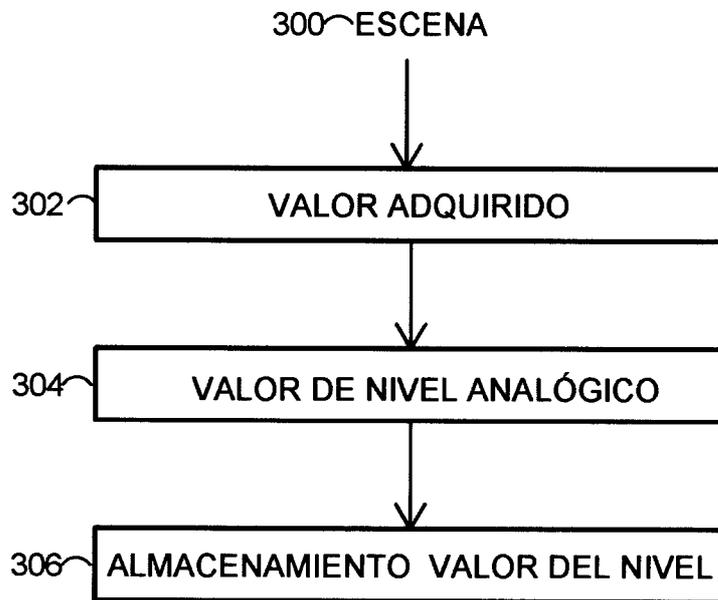


FIG. 3

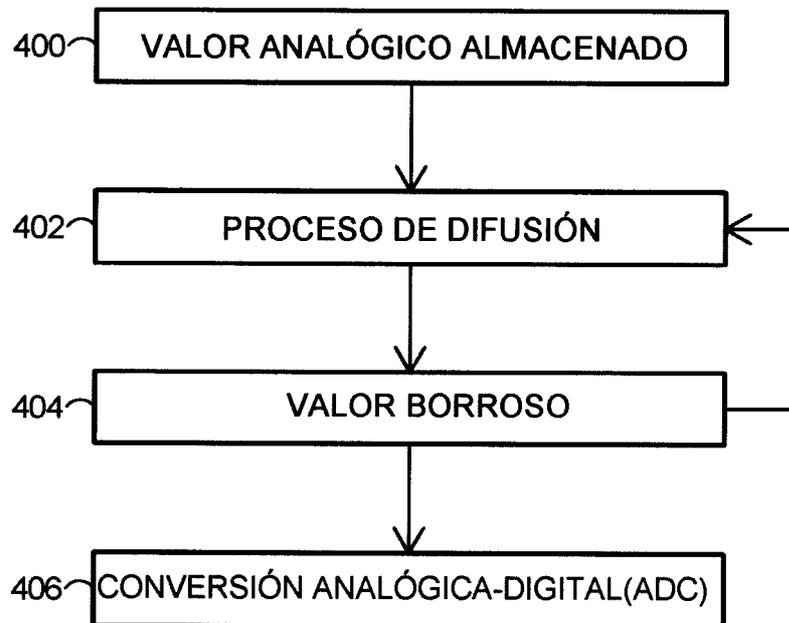


FIG. 4

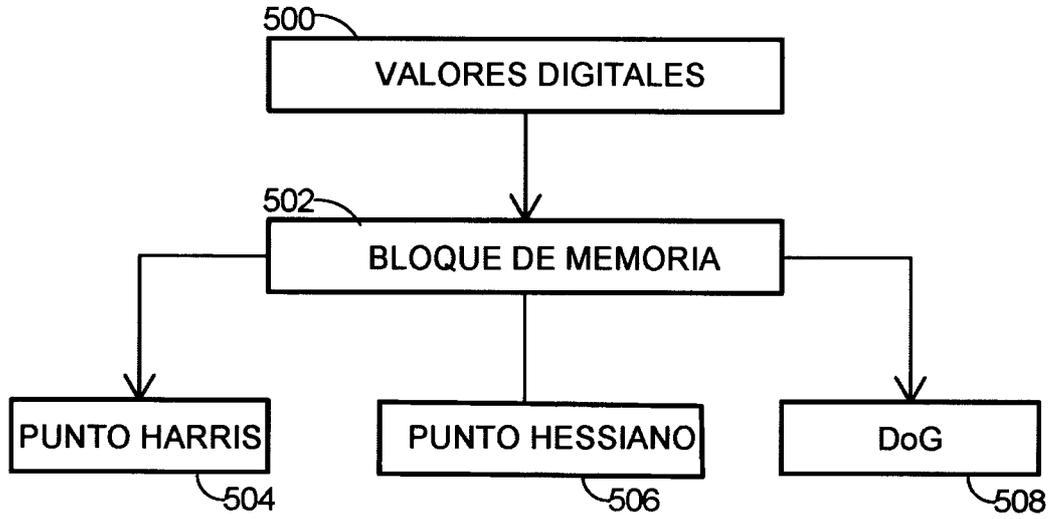


FIG. 5

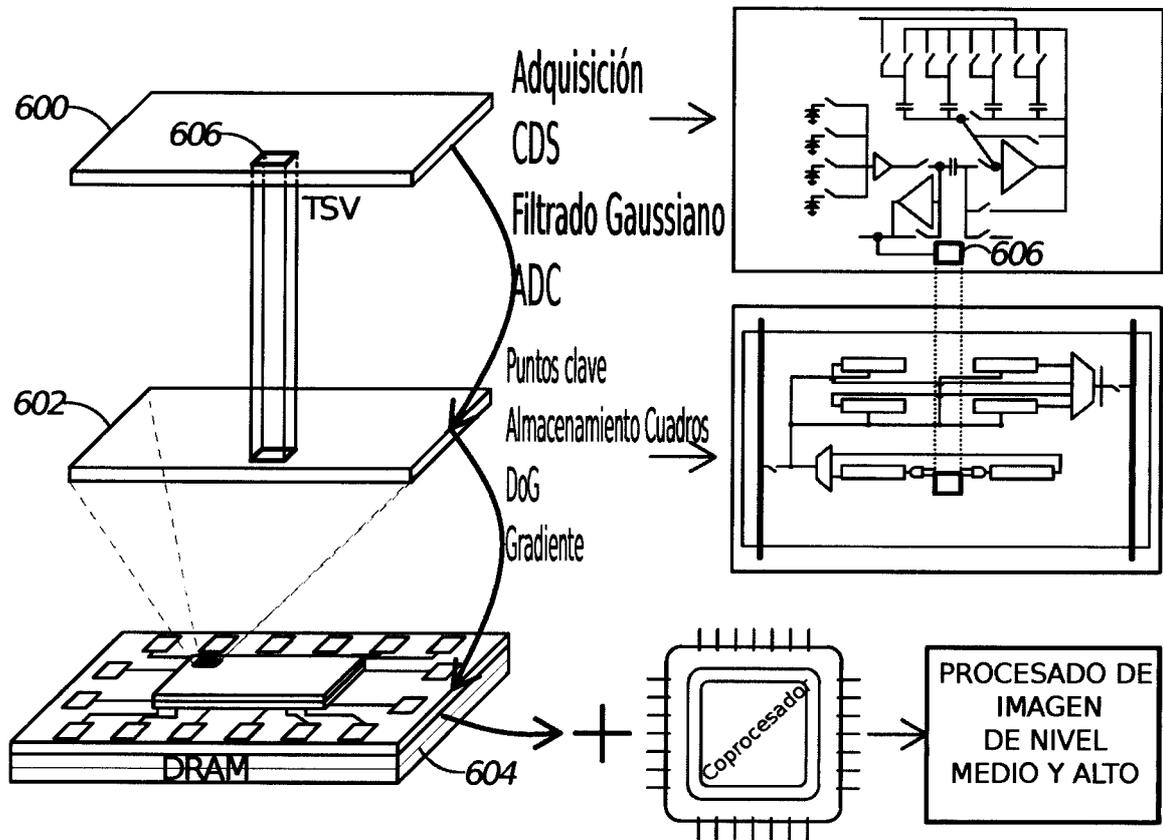


FIG. 6

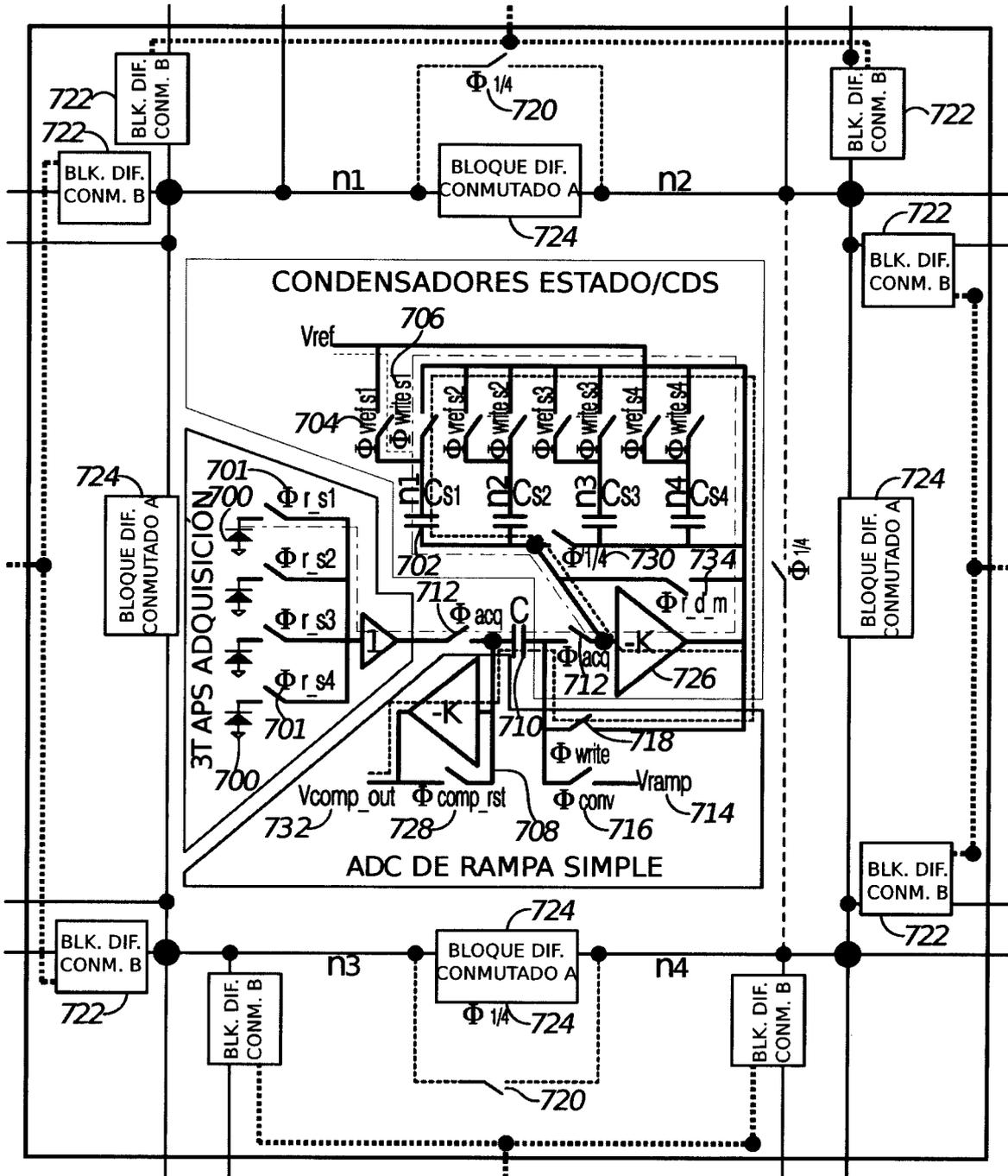


FIG. 7

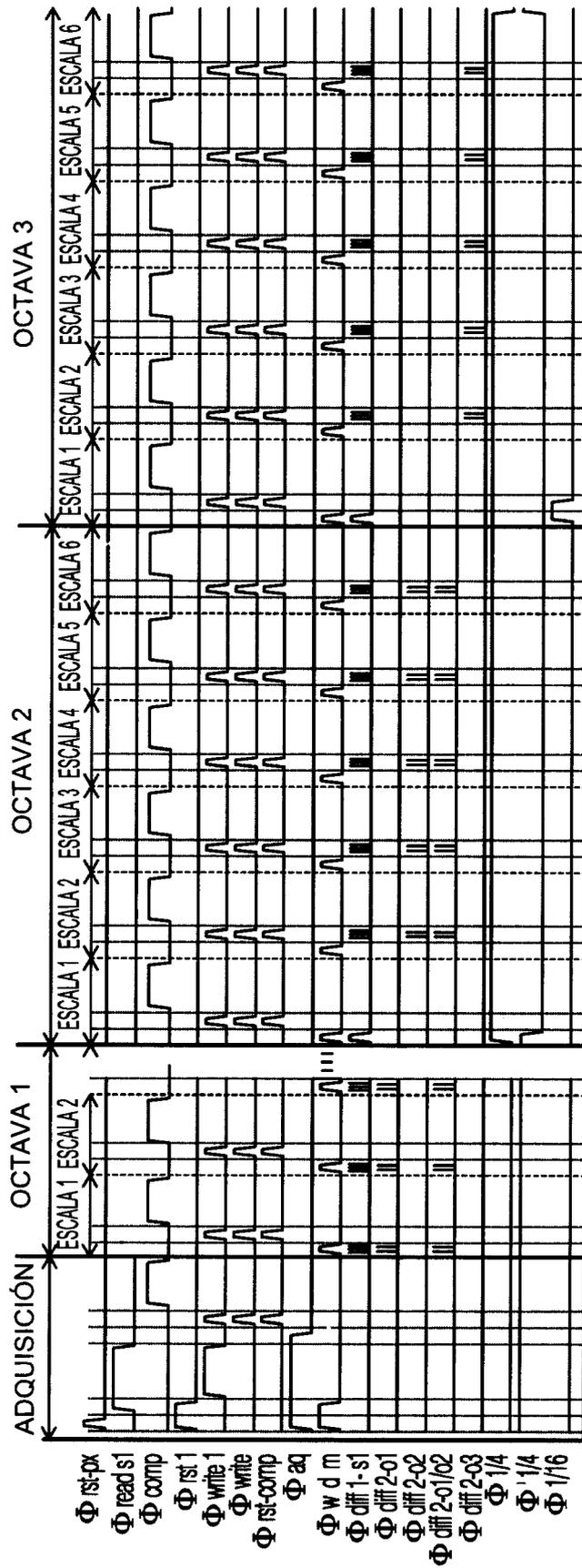


FIG. 8

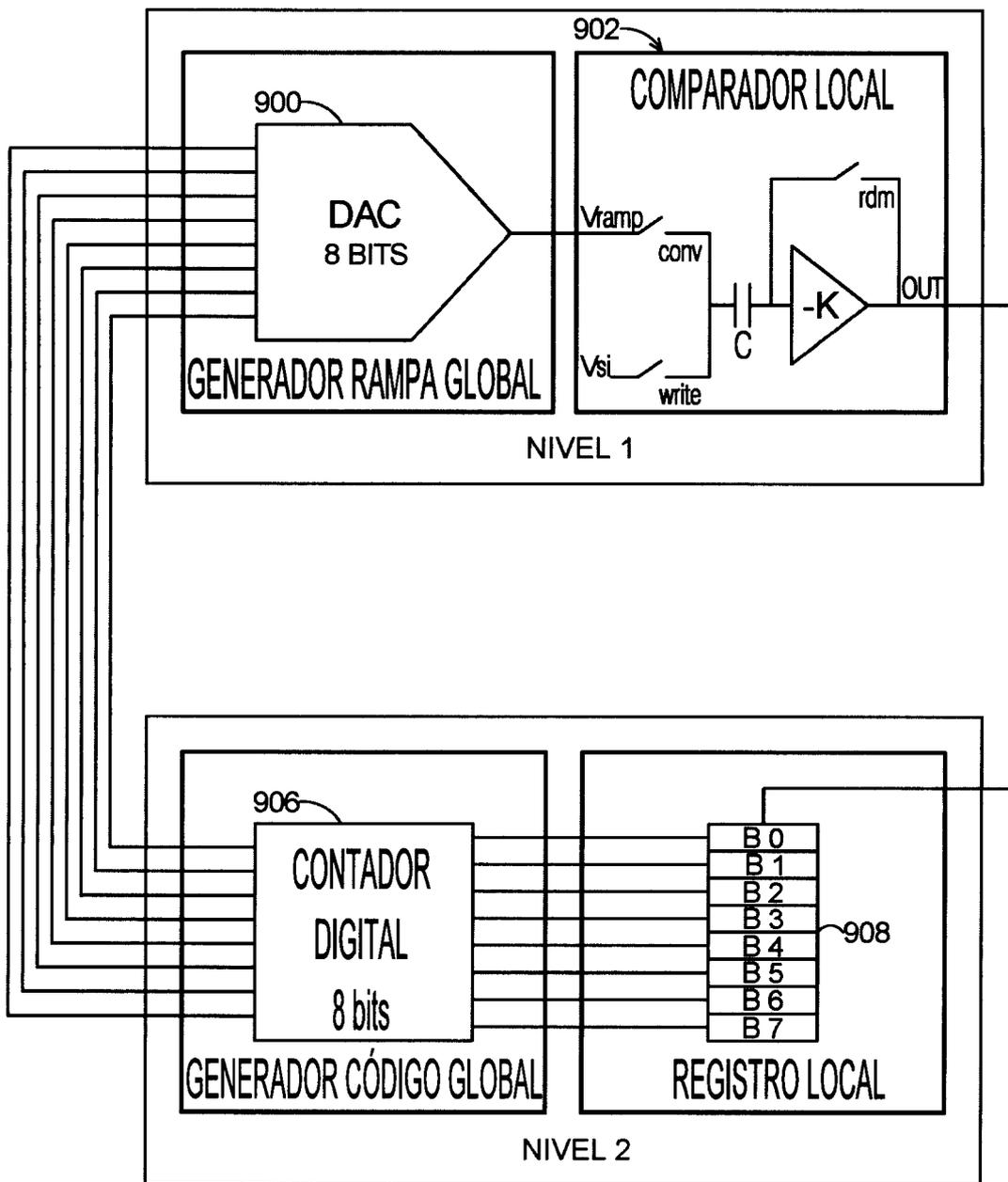


FIG. 9

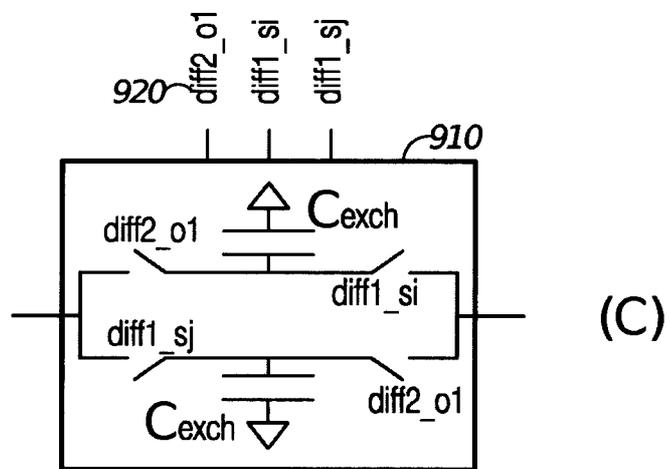
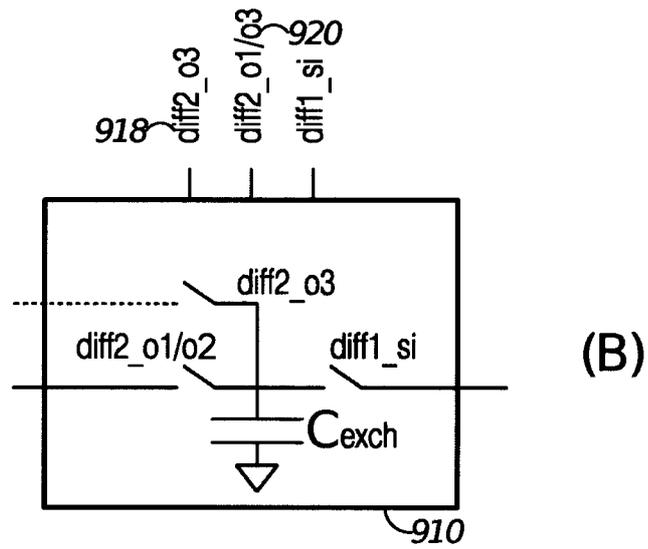
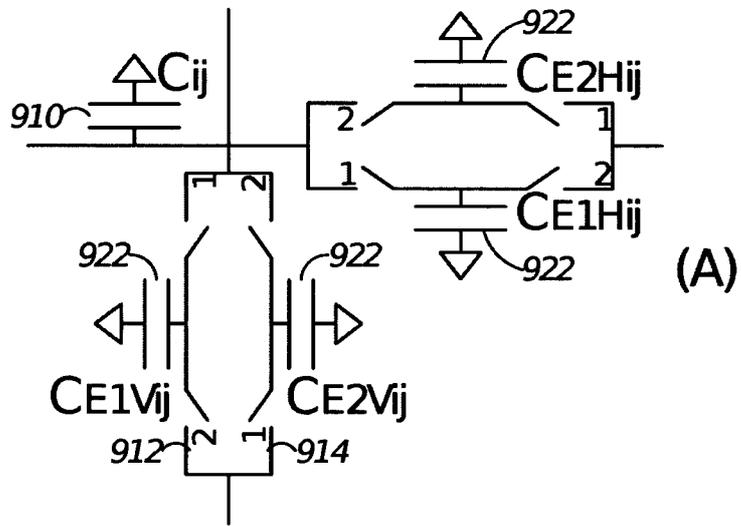


FIG. 10

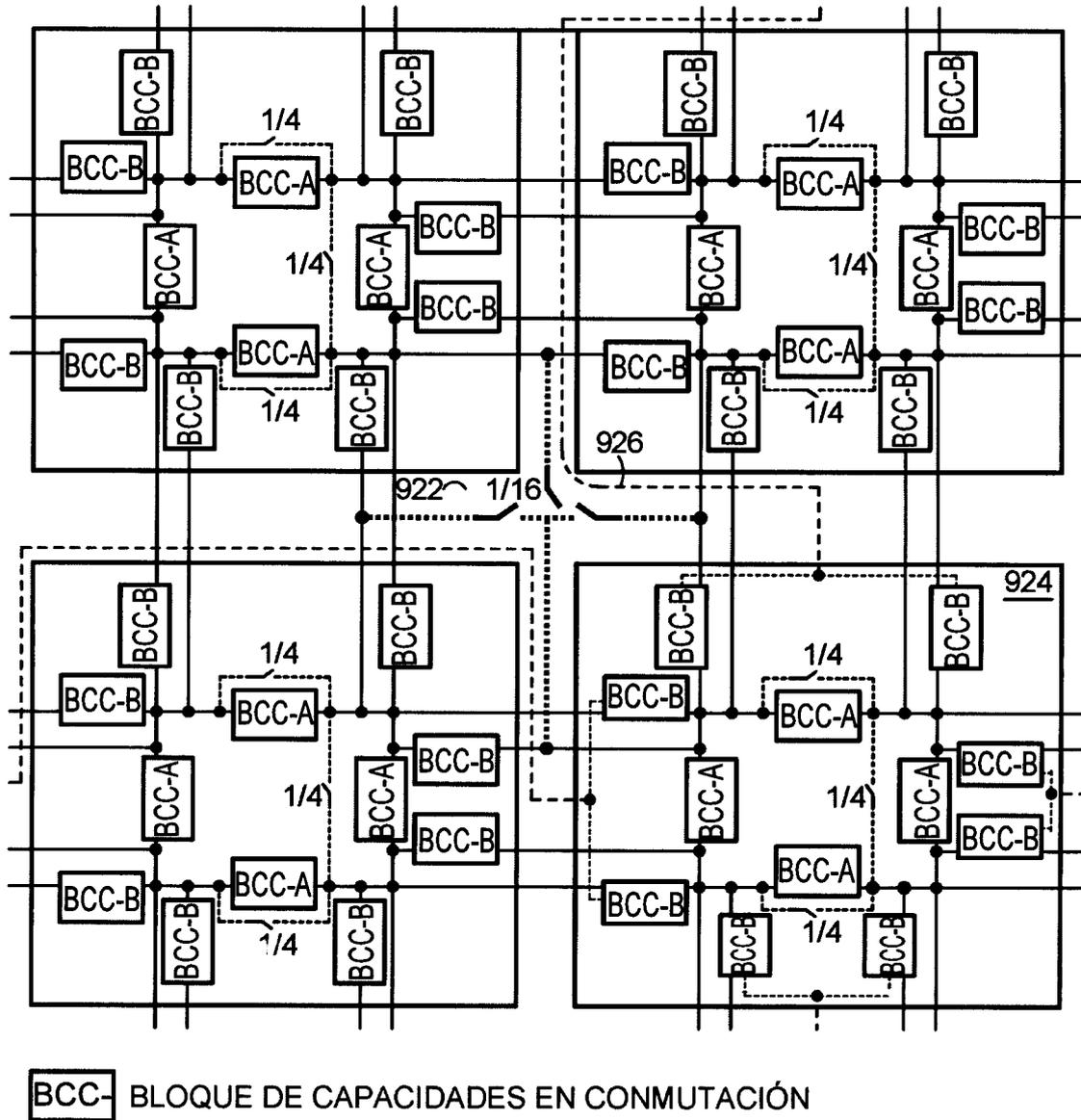


FIG. 11

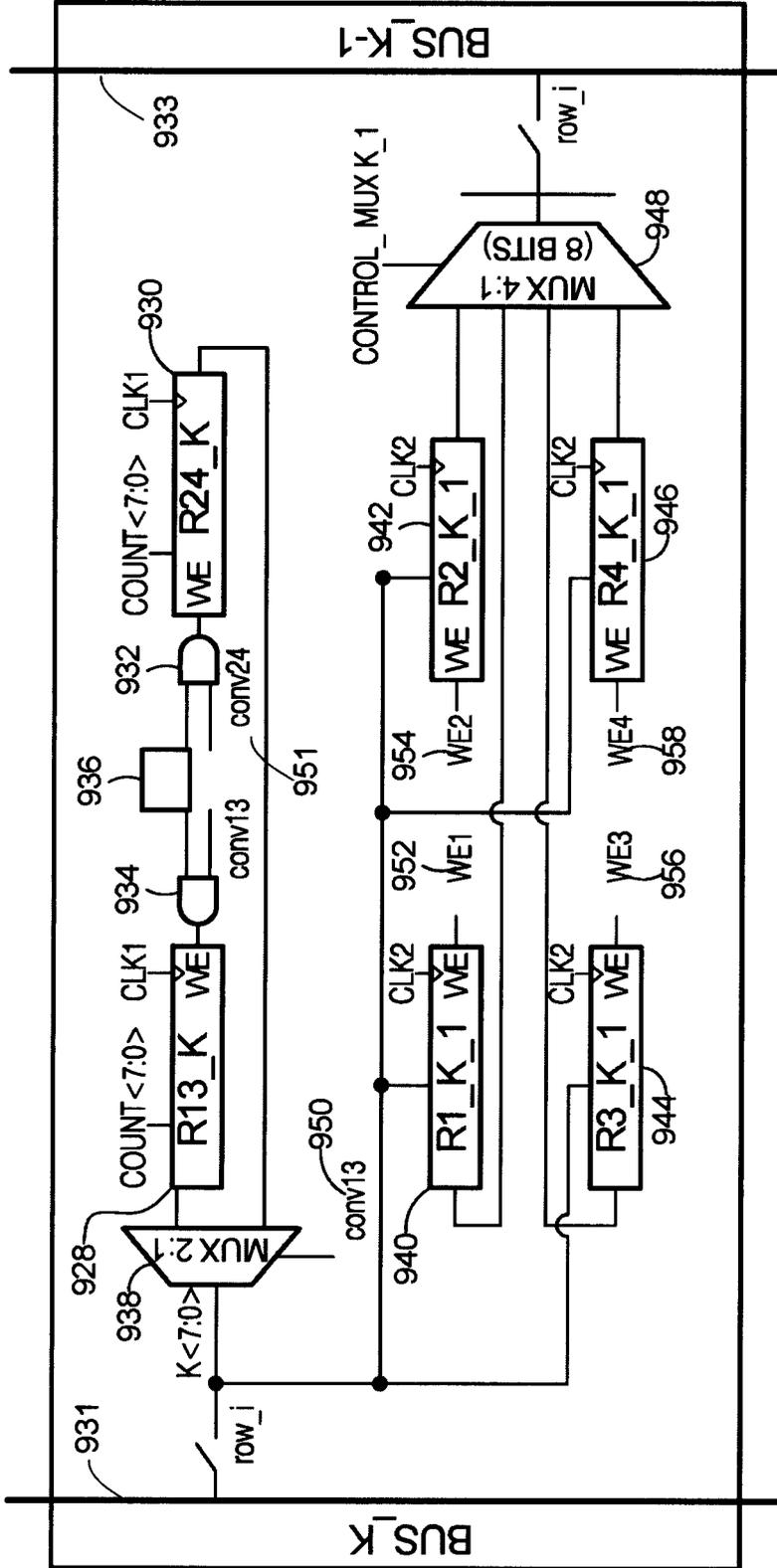


FIG. 12

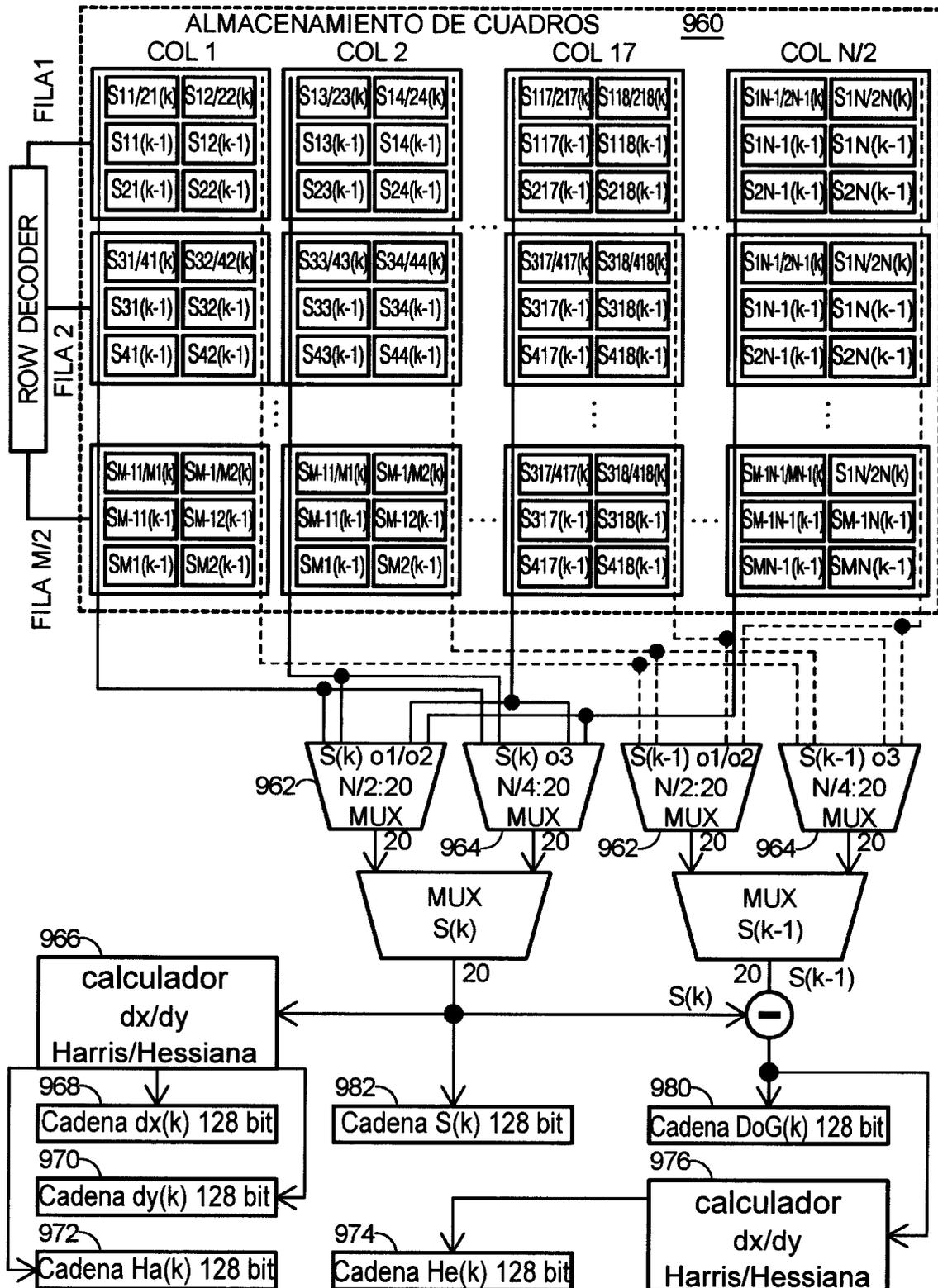


FIG. 13

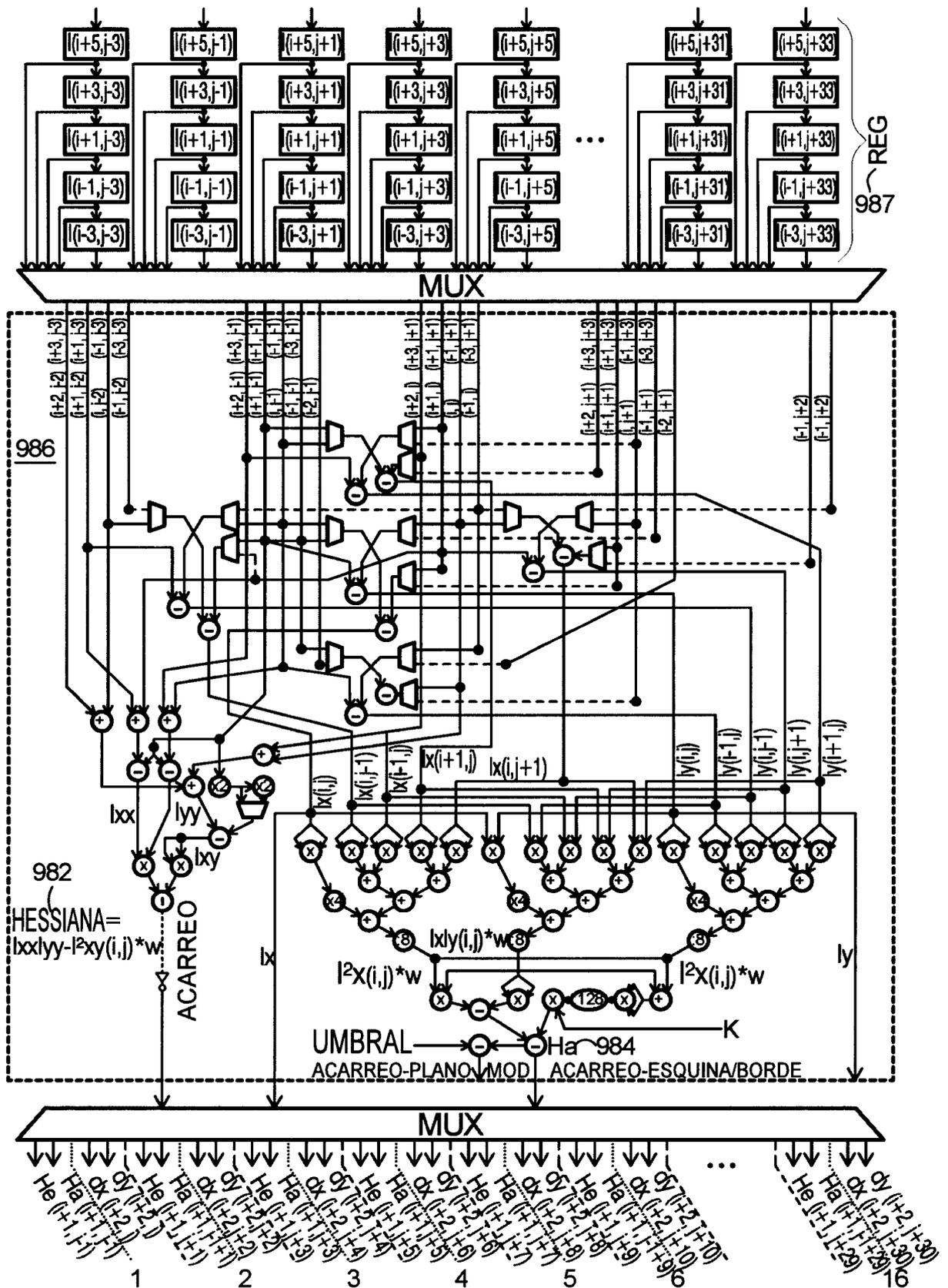


FIG. 14