

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 938 285**

21 Número de solicitud: 202130929

51 Int. Cl.:

G06T 1/00 (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

05.10.2021

43 Fecha de publicación de la solicitud:

05.04.2023

71 Solicitantes:

**CONSEJO SUPERIOR DE INVESTIGACIONES
CIENTÍFICAS (CSIC) (90.0%)**

C/ Serrano, 117

28006 Madrid (Madrid) ES y

UNIVERSIDAD DE SEVILLA (10.0%)

72 Inventor/es:

LINARES BARRANCO, Bernabé y

SERRANO GOTARREDONA, María Teresa

74 Agente/Representante:

PONS ARIÑO, Ángel

54 Título: **SENSOR DE VISIÓN DINÁMICO FOVEADO ELECTRÓNICAMENTE**

57 Resumen:

Sensor de visión dinámico foveado electrónicamente. El sensor de visión comprende una matriz (1) de píxeles (5) sobre los que se aplica un mecanismo de foveación, definiendo una serie de regiones de baja resolución de píxeles agrupados (macro-píxeles), de manera que trabajan como un único pixel (5) aislado, obteniéndose información del grupo de píxeles (5) y no de cada pixel (5) de manera individual. Gracias a las regiones de baja resolución de macro-píxeles se logra un ahorro de energía y ancho de banda de datos en favor de las regiones de alta resolución no agrupadas o foveadas. Las regiones de píxeles agrupados son configurables con señales electrónicas externas. Además, se pueden activar electrónicamente múltiples regiones de alta resolución o foveación y tamaños de región.

ES 2 938 285 A1

DESCRIPCIÓN

SENSOR DE VISIÓN DINÁMICO FOVEADO ELECTRÓNICAMENTE

5 OBJETO DE LA INVENCION

Se trata de un sensor de visión dinámico capaz de agrupar píxeles en macro-píxeles como si funcionaran como uno aislado, reduciendo así la resolución, y por tanto también el consumo energético, la información, y la carga computacional posterior. De esta
10 manera puede operar por defecto a baja resolución y si detecta una zona de interés, activar dicha zona en alta resolución para extraer de ella información más detallada.

ANTECEDENTES DE LA INVENCION

15 Entre la colección de sensores de visión de impulso (*spiking vision sensors*), los DVS (sensores de visión dinámicos, *Dynamic Vision Sensor*) debido a su reducido desapareamiento entre píxeles, alto rango dinámico, gran ancho de banda de píxeles, baja latencia, bajo consumo de energía y propiedades de supresión de redundancia temporal de datos se han hecho muy populares y han suscitado el interés de la industria.

20

Estos sensores difieren sustancialmente de las cámaras de visión más convencionales que se basan en la adquisición de secuencias de imágenes o fotogramas. En un DVS no se adquieren imágenes. En un DVS cada píxel es autónomo y envía un “evento” cuando detecta un cambio relativo en su fotocorriente. El “evento” suele ser la
25 coordenada x,y más el signo del cambio, y se denomina “evento de dirección” (“*Address Event*” en inglés). El “evento” se puede interpretar como uno de los impulsos nerviosos que mandaría una retina biológica a la corteza cerebral por el nervio óptico.

En consecuencia, un píxel en un DVS es mucho más complejo que un píxel en una
30 cámara de adquisición de imágenes/fotogramas. En éstas cada píxel contiene un fotosensor más un número reducido de transistores para escanear la matriz de píxeles desde la circuitería periférica. En un DVS, sin embargo, cada píxel contiene circuitería para calcular en tiempo real la variación relativa de luz, su signo, más circuitería de comunicación con la periferia, comunicación que es iniciada desde el píxel.

35

Los DVS de megapíxeles permiten realizar una detección de alta resolución de objetos pequeños, manteniendo un amplio campo de visión. Sin embargo, la tasa de datos producida por el sensor puede saturar su ancho de banda de salida, especialmente si el campo visual observado tiene un alto contenido dinámico, como ocurre durante la captación visual en conducción de automóviles. Además, al aumentar la resolución del sensor, también aumenta el tamaño y el consumo de energía de los sistemas de procesamiento de visión posteriores.

Se han propuesto algunas alternativas para mejorar el ancho de banda del sensor, como comunicar en paralelo los eventos de un grupo de píxeles. Otras alternativas pretenden reducir la cantidad de datos de salida del sensor incluyendo un bloque lógico de correlación espacio-temporal compartido por un grupo de píxeles para filtrar los ruidosos, así como para limitar la redundancia espacial, o incluyendo un procesador de señales de eventos digitales a la salida del sensor, que puede programarse con algunas capacidades de filtrado espacio-temporal.

Estas últimas alternativas pretenden aumentar la dispersión de los datos de salida, pero la enorme complejidad del sistema de procesamiento posterior sigue siendo la misma.

Los seres humanos están dotados de una visión foveal. Nuestros ojos tienen una región foveal equipada con una alta densidad de fotodetectores sensibles al color, donde se percibe una visión de alta resolución (HR). La densidad de fotorreceptores disminuye logarítmicamente hacia la región periférica, donde se adquiere una parte de la escena de baja resolución (LR) insensible al color.

La localización a alta velocidad de los objetos de interés se realiza mediante cálculo en las regiones periféricas de baja resolución (también conocido como el cálculo del "dónde") para controlar los movimientos sacádicos del ojo que centran el punto de fijación foveal en la región objetivo, que será procesada con alta resolución para su posterior observación y procesamiento por el reconocimiento del subsistema "qué".

De esta manera, no sólo se dedican menos recursos de la retina a adquirir las regiones periféricas, sino que también el área cortical de recursos computacionales dedicada a procesar las regiones periféricas es menor. Mientras que la región foveal corresponde a sólo 5° del campo visual, el 25% de los recursos computacionales corticales se dedican a procesarla.

El uso de la visión foveada para lograr un reconocimiento eficaz de los objetos de alta resolución, combinada con un amplio campo de visión de baja resolución que permite localizar el objetivo en regiones periféricas con una baja carga computacional se ha
5 propuesto anteriormente para cámaras basadas en adquisición de imágenes o fotogramas, con el objetivo de resolver los problemas indicados anteriormente.

En el contexto de cámaras de visión por adquisición de imágenes o fotogramas, se han reportado varias técnicas de foveación. En el área de visión robótica se han propuesto
10 sistemas de hardware que combinan varias cámaras con diferentes campos de visión. Sin embargo, estos sistemas requieren un control mecánico para mover el punto central del sensor de foveación, lo que aumenta el retraso del sistema y el consumo de energía. Otros sistemas se basan en el uso de sensores de alta resolución, y a continuación en etapas posteriores de procesamiento se realiza una transformación foveal para ahorrar
15 potencia de cálculo y energía.

Sin embargo, estos sistemas no alivian la comunicación del ancho de banda de los datos al nivel de los sensores. Ha habido algunas propuestas de sensores con geometría física foveal a nivel de sensor. No obstante, muchas de ellas adolecen de un uso no óptimo
20 del área de silicio. También se ha propuesto un sensor foveado con píxeles uniformes de alta resolución en el centro y píxeles periféricos de detección de movimiento de baja resolución, pero todavía necesita un control mecánico del centro de foveación.

Por otra parte, se ha publicado información relativa a un sensor foveado con control
25 electrónico del punto de foveación en el que la resolución espacial se intercambia con la resolución temporal. Sin embargo, no reduce la tasa de datos de salida del sensor.

Por otra parte, son conocidos del estado de la técnica documentos como el WO2013043259A1, que describe un sistema de imagen con capacidades de foveación
30 de imagen. El documento WO0051345A1 que describe un procedimiento y aparato con un display de doble resolución foveal. El documento US7808528B2 que muestra un procedimiento y aparato la obtención para un conjunto de imágenes de resolución variable en chip que incorpora la medición de los índices de ángulo de balanceo, cabeceo y guiñada. El documento US2020169734A1 que divulga un procedimiento de
35 renderizado variable basado en estimación de movimiento. El documento US2018300951A1 que describe un procedimiento de teselación variable para

renderizado foveado. Por último, el documento US6455831B1 que muestra un sensor chip CMOS de imagen foveada.

5 Todos los documentos anteriores se basan en la utilización de regiones de interés (ROI, del inglés *Regions of Interest*). Algunas cámaras ofrecen la posibilidad de transmitir sólo una región parcial de un sensor, lo que ofrece la opción de concentrarse en las áreas de la imagen que contienen la información relevante para el procesamiento de la misma. Esta llamada región de interés puede ser a veces tan pequeña como un píxel. Al reducir el tamaño de la imagen transmitida por la cámara se reduce el tamaño de la imagen y, por tanto, el ancho de banda necesario para cada imagen. Al configurar un ROI, la ventana de lectura se reduce al área relevante para el análisis, lo que aumenta la frecuencia de la imagen de forma significativa.

DESCRIPCIÓN DE LA INVENCIÓN

15

El sensor de visión dinámico de la presente invención permite resolver los problemas planteados anteriormente para sensores tipo DVS, pues logra implementar un sensor de visión dinámico foveado electrónicamente (EF-DVS, *electronically foveated dynamic vision sensor*).

20

El sensor propuesto comprende una matriz de píxeles complejos DVS sobre los que se aplica un mecanismo de agrupación para formar macro-píxeles. Por defecto, la matriz de sensado se configura en modo de baja resolución agrupando píxeles de manera que trabajan como un único píxel (macro-píxel), obteniéndose información del conjunto de píxeles que constituyen el macro-píxel y no de cada píxel de manera individual.

25

Cuando se define una zona de interés en la matriz, sus píxeles se configuran en alta resolución (es decir, no agrupados), y esta zona se convierte en región foveada. Se pueden definir simultáneamente varias zonas foveadas (o regiones de interés) que estarían configuradas en alta resolución. El resto de las zonas que continúan en baja resolución, serían las regiones no-foveadas de macro-píxeles (o píxeles agrupados).

30

Gracias a las regiones no-foveadas de baja resolución de macro-píxeles se logra un ahorro de energía y ancho de banda de datos en favor de las regiones de alta resolución foveadas. Las regiones foveadas/no-foveadas son configurables con señales

35

electrónicas externas. Además, se pueden activar electrónicamente múltiples zonas de foveación y tamaños de dichas regiones de interés.

5 La foveación múltiple permite capacidades de reconocimiento que pueden superar la visión humana en tareas en las que deben atenderse simultáneamente varias regiones de interés o atención, como los sistemas de vigilancia en tiempo real, la conducción automática o los vuelos autónomos.

10 Particularmente, el sensor comprende una matriz de $N \times M$ píxeles con unos circuitos periféricos correspondientes, para generar los eventos de direcciones de salida. Estos circuitos periféricos pueden ser unos bloques de control para la configuración digital de las regiones no foveadas de macro-píxeles (píxeles agrupados), y unos bloques de lectura de la matriz.

15 Cada uno de los píxeles del sensor que conforman la matriz comprende preferentemente, conectados entre sí en este orden:

- una etapa fotorreceptora que transduce la fotocorriente generada en un fotodiodo a un voltaje (fototensión),
- una etapa de preamplificación que introduce una ganancia de voltaje,
- 20 - un búfer vinculado a una etapa de amplificación y diferenciación de la tensión, desacoplando el búfer la fototensión de una entrada capacitiva de la etapa de amplificación y diferenciación de la tensión,
- unos comparadores de salida, que comparan la tensión amplificada con dos umbrales de tensión para generar unos eventos de salida ON y OFF, y
- 25 - otros circuitos que impulsan la solicitud de fila y columna y reciben señales de acuse de recibo para comunicar los eventos a los circuitos periféricos.

El bloque fotorreceptor comprende preferentemente un fotodiodo conectado a una etapa activa accionada por una fuente que genera una tensión V_{ph} que es logarítmica con la
30 fotocorriente.

Es posible combinar la salida de diferentes píxeles de forma que la señal combinada represente la variación temporal de la media espacial de las fotocorrientes de los píxeles individuales en la vecindad, pero reduciendo la actividad de salida del grupo de píxeles
35 en comparación con la actividad de salida total de los píxeles individuales.

Para ello, cada píxel comprende un nodo de interconexión con los píxeles vecinos, preferentemente con un píxel anterior en su columna y un píxel posterior en su fila. Gracias a que los nodos de interconexión de salida de la fotocorriente preamplificada son compartidos por los píxeles vecinos se logra construir regiones de baja resolución.

5

Existen además tres alternativas de arquitecturas foveadas digitalmente para agrupar los píxeles dentro de la matriz:

- En un primer aspecto de la invención, el sensor comprende una matriz de NxM píxeles con los correspondientes circuitos periféricos para generar el evento de dirección de salida. Bloques adicionales de control foveal X-FOV e Y-FOV se añaden al sensor para la configuración digital de las regiones de macro-píxeles.
- En un segundo aspecto de la invención, los NxM píxeles de la matriz se organizan en grupos de nxm píxeles cada uno. Cada grupo puede configurarse utilizando bloques de control de foveación periféricos X-FOV e Y-FOV en modo de alta resolución (HR) o de baja resolución (LR) utilizando un solo bit almacenado en el grupo.
- En un tercer aspecto de la invención los píxeles vecinos se agrupan de forma predefinida cuando se configuran en baja resolución como se propone en el segundo aspecto de la invención. Sin embargo, en este caso se añade la posibilidad de formar macro-píxeles con forma arbitraria. Para conectar los grupos vecinos, se puede utilizar un esquema similar al considerado en el primer aspecto de la invención. En este caso, tres bits definen la resolución de cada macro-píxel.

25

DESCRIPCIÓN DE LOS DIBUJOS

Para complementar la descripción que se está realizando y con objeto de ayudar a una mejor comprensión de las características de la invención, de acuerdo con un ejemplo preferente de realización práctica de la misma, se acompaña como parte integrante de dicha descripción, un juego de dibujos en donde con carácter ilustrativo y no limitativo, se ha representado lo siguiente:

Figura 1.- Muestra ilustración de un sensor de resolución multifoveal.

10

Figura 2.- Muestra un sistema de visión propuesto con control electrónico de múltiples puntos de foveación y subsistemas paralelos de reconocimiento de objetos.

Figura 3.- Muestra un diagrama de bloques de un píxel DVS del estado de la técnica.

15

Figura 4.- Muestra un diagrama de bloques de un píxel DVS agrupable.

Figura 5.- Detalle del nodo de interconexión de un píxel.

Figura 6.- Muestra una primera arquitectura de la matriz DVS foveada.

20

Figura 7.- Muestra un bloque de control periférico X-FOV.

Figura 8.- Muestra una segunda arquitectura de la matriz DVS foveada.

25

Figura 9.- Muestra una tercera arquitectura de la matriz DVS foveada.

Figura 10.-Esquemático de un macropíxel de la segunda arquitectura propuesta de matriz DVS foveada

30

Figura 11.-Esquemático de un macropíxel de la tercera arquitectura propuesta de matriz DVS foveada

Figura 12.- Esquemático del píxel A correspondiente al macropíxel de la segunda arquitectura propuesta de matriz DVS foveada.

35

Figura 13.- Esquemático del pixel B correspondiente al macropixel de la segunda arquitectura propuesta de matriz DVS foveada.

Figura 14.- Esquemático del pixel G correspondiente al macropixel de la tercera
5 arquitectura propuesta de matriz DVS foveada.

REALIZACIÓN PREFERENTE DE LA INVENCIÓN

Se describe a continuación, con ayuda de las figuras 1 a 14, una realización preferente
10 del sensor de visión dinámico foveado electrónicamente.

El sensor propuesto comprende una matriz (1) de píxeles (5) sobre los que se aplica un mecanismo de agrupación de píxeles, definiendo una serie de regiones de macro-píxeles en las que cada macro-píxel trabaja como un único píxel (5) aislado. El resto de
15 píxeles (5) de la matriz (1) constituye las regiones foveadas de alta resolución.

La figura 1 ilustra la adquisición visual haciendo uso del sensor de la invención. En la parte superior se ilustra una captura de resolución 259,2k píxeles. La captura inferior muestra una representación de menor resolución con 4k píxeles. A partir de esta captura
20 de menor resolución se han obtenido cuatro sub-capturas parciales de objetos de interés (cuatro personas en este caso), a las que como se muestra en la figura 1 se las ha configurado en alta resolución, definiendo así zonas de interés o foveadas.

En la figura 2 se muestra una representación esquemática del sensor vinculado a una
25 óptica de gran angular (2), asociada a la matriz (1) de píxeles (5), equivalente a una retina de alta resolución. Por defecto, toda la matriz se configura en baja resolución, agrupando píxeles en macro-píxeles. Un sistema externo procesaría la información en baja resolución para identificar zonas de interés. Dentro de la matriz (1) de píxeles (5) agrupados (o macro-píxeles) se definen unas regiones de interés del campo visual y se
30 controlan electrónicamente las regiones de foveación (R1, R2, ... R4) que se configuran en alta resolución. Por un lado, se realiza un reconocimiento de objetos simultáneo de alta resolución sobre pequeñas regiones de interés, reduciendo las demandas de cálculo que serían necesarias para procesar el campo visual completo. Por otro lado, se realiza un reconocimiento de baja resolución de las regiones no foveadas.

35

Se describe a continuación los componentes de cada uno de los píxeles (5) de la matriz (1). El diseño cada píxel (5) se basa en un píxel DVS (sensor dinámico de visión) del estado de la técnica, como el que se muestra en la figura 3.

5 Cada píxel (5) comprende una etapa fotorreceptora (6) que transduce la fotocorriente generada en un fotodiodo a un voltaje, una etapa de preamplificación (7) conectada a la etapa fotorreceptora (6), que introduce una ganancia de voltaje, un búfer (8), conectado a la etapa de preamplificación (7), para desacoplar la fototensión de la entrada capacitiva de una etapa de amplificación y diferenciación (9) de la tensión.

10

A continuación, la tensión amplificada se compara con dos umbrales de tensión en unos comparadores de salida (10) para generar unos eventos de salida ON u OFF. Otros circuitos (no mostrados en la figura 3) impulsan la solicitud de fila y columna que señalan las coordenadas del píxel (5) dentro de la matriz (1) y reciben señales de acuse de recibo para comunicar los eventos a unos circuitos periféricos de arbitraje y decodificación.

15

Como se muestra en la Figura 3, la etapa fotorreceptora (6) comprende un fotodiodo conectado a una etapa activa accionada por una fuente que genera una tensión V_{ph} que es logarítmica con la fotocorriente, de manera que:

20

$$V_{ph} = \frac{V_G}{n_p} + U_T \log \frac{I_{ph}}{I_{sp}} \quad (1)$$

Donde n_p e I_{sp} son el factor de pendiente subumbral y el factor de corriente del transistor M_{p1} . Los transistores M_{p1} de la etapa fotorreceptora (6) y M_{p2} de la etapa de preamplificación (7) forman un espejo de corriente que, cuando operan en la región subumbral, puede introducir una ganancia de corriente A_I que puede ajustarse a través de la diferencia de tensión $\Delta V_p = V_G - V_{GA}$, de manera que la corriente que fluye a través del transistor M_{p2} puede expresarse como:

25

$$I_{phA} = I_{ph} e^{\frac{V_G - V_{GA}}{n_p U_T}} = A_I I_{ph} \quad (2)$$

30

Como se muestra en la figura 3, la etapa de preamplificación (7) comprende el transistor de conversión de tensión a corriente M_{p2} conectado a una pila de cuatro transistores NMOS conectados como diodos, que genera una tensión de salida v_o que es aproximadamente igual a:

$$v_o = 4n_n U_T \log \frac{I_{phA}}{I_{sp}} \quad (3)$$

De modo que la tensión v_o también es logarítmica con la fotocorriente, pero se puede obtener un factor de amplificación de la tensión aproximadamente igual al número de transistores apilados conectados como diodos. El número de transistores apilados conectados en modo diodo no está restringido a 4, sino que puede ser cualquier número, según la amplificación que se requiera y lo que permita el rango de alimentación. Esta tensión amplificada se refuerza a v_{ob} en el búfer (8) para evitar la ralentización de la constante de tiempo debido al acoplamiento con el condensador C1, de la etapa de amplificación y diferenciación (9). El búfer opcionalmente podría ser omitido dependiendo de las necesidades de ancho de banda de la aplicación.

La tensión v_{ob} es la entrada a la etapa de amplificación y diferenciación (9), de modo que:

$$\Delta V_{diff} = -\frac{c_1}{c_2} \Delta V_{ob} \cong -\frac{c_1}{c_2} \Delta V_o = -\frac{c_1}{c_2} 4n_n U_T \log \frac{I_{ph2}}{I_{ph1}} = -A_T U_T \log \frac{I_{ph2}}{I_{ph1}} \quad (4)$$

Donde, A_T es la ganancia total de tensión introducida por la etapa de preamplificación (7) y la etapa de amplificación y diferenciación (9).

La tensión V_{diff} se utiliza como entrada a los comparadores de salida (10) ON y OFF. Cuando los comparadores de corriente de salida (10) ON/OFF se activan, el píxel genera un evento de salida positivo/negativo, y la tensión V_{diff} se restablece a su valor de reposo, de manera que:

$$I_{sp} e^{\frac{n_p V_{dd} - V_{rest}}{n_p U_T}} = I_{sn} e^{\frac{V_{REF}}{n_n U_T}} \quad (5)$$

Siempre que haya una excursión de tensión ΔV_{diff} tal que:

$$\begin{aligned} \Delta V_{diff ON} &= -\frac{n_p}{n_n} V_{\theta R}^+ \\ \Delta V_{diff OFF} &= +\frac{n_p}{n_n} V_{\theta R}^- \end{aligned} \quad (6)$$

donde $V_{\theta R} = |V_{REF} - V_{\theta}^{\pm}|$, se genera un evento de salida ON/OFF y la tensión V_{diff} vuelve a su valor de reposo.

Volviendo a la entrada, se denomina "umbral de contraste ON" (o "sensibilidad de contraste ON") al mínimo estímulo de contraste que genera un único evento positivo a través del

canal de salida ON "umbral de contraste ON" (o "sensibilidad de contraste ON"), y el estímulo de contraste mínimo que genera un único evento negativo a través del canal de salida OFF "umbral de contraste OFF umbral de contraste" (o "sensibilidad de contraste OFF"):

5

$$\theta_{ev}^+ = \log \frac{I_{bright}}{I_{dark}} = \frac{\frac{n_p V_{\theta R}^+}{n_n}}{A_T U_T} \quad (7)$$

$$\theta_{ev}^- = \log \frac{I_{dark}}{I_{bright}} = \frac{\frac{n_p V_{\theta R}^-}{n_n}}{A_T U_T}$$

El contraste mínimo detectable del estímulo puede ajustarse mediante las tensiones de control de la sensibilidad al contraste $V_{\theta R}$.

10

Una vez analizada la estructura de cada uno de los píxeles (5) se va a analizar cómo se puede combinar la salida de diferentes píxeles (5) de forma que la combinada represente la variación temporal de la media espacial de las fotocorrientes de los píxeles (5) individuales en la vecindad, pero reduciendo la actividad de salida del grupo de píxeles (5) en comparación con la actividad de salida total de los píxeles (5) individuales.

15

Se han considerado varias opciones para combinar la información de fotocorriente local obtenida por píxeles (5) individuales en una información combinada significativa de baja resolución que promedia la información local en la vecindad mientras que ahorra el ancho de banda de salida, reduciendo el ruido y manteniendo el rango dinámico:

20

- Interconectar los píxeles (5) en el nodo de salida de los búfers (8) de tensión v_{ob}

Al interconectar los píxeles en el nodo v_{ob} se obtiene (donde G es el número de píxeles (5) por grupo):

25

$$\sum_{i=1}^G I_{sn} e^{\frac{v_{oi} - v_{ob}}{nU_T}} = G I_{bb} \quad (8)$$

30

Esta es la ecuación de un par diferencial de entrada múltiple, donde el píxel (5) con la tensión más alta v_{oi} va a tomar la mayor parte de la corriente de polarización $G I_{bb}$ del par diferencial de entrada múltiple y va a determinar el voltaje en el nodo de salida v_{ob} . En consecuencia, el nodo de salida v_{ob} va a seguir sólo los cambios en la máxima fotocorriente del grupo de píxeles (5).

- Añadir tensiones v_{phi} (de forma similar, añadir tensiones v_{oi} o v_{obi} utilizando algún circuito de adición de tensión) para mantener una relación logarítmica con la fotocorriente.

5

Como se conoce que v_{phi} , v_{oi} , v_{obi} son proporcionales a $\log(I_{phi})$, por lo que cualquiera de estas tensiones es proporcional al logaritmo de la fotocorriente. Definiendo una tensión genérica $V_{\log(I_{phi})}$ de forma que $\Delta V_{\log i} \propto \log(I_{phi_{final}} / I_{phi_{initial}})$, y considerando la tensión añadida $\Delta V_{\log} = \sum_{i=1}^K V_{\log i}$, entonces:

10

$$\Delta V_{\log} = V_{\log final} - V_{\log initial} = (\sum_{i=1}^G V_{\log i_{final}} - V_{\log i_{initial}}) \propto$$

$$\left(\sum_{i=1}^G \log(I_{phi_{final}}) - \sum_{i=1}^G \log(I_{phi_{initial}}) \right) = \log \frac{\prod_{i=1}^G I_{phi_{final}}}{\prod_{i=1}^G I_{phi_{initial}}} \quad (9)$$

El efecto de la suma de las tensiones de los píxeles (5) individuales equivale a la multiplicación de sus fotocorrientes. El grupo de píxeles (5) será sensible a la variación relativa del producto de todas las fotocorrientes del grupo de píxeles (5). La sensibilidad del grupo es proporcional al número total de píxeles (5) en el grupo, por lo que la generación de la frecuencia de eventos se multiplicaría por el número de píxeles (5) para una misma variación relativa de la fotocorriente y el ancho de banda de salida consumido por el grupo en modo de baja resolución será el mismo que el de todos los píxeles (5) individuales en modo de alta resolución.

Un circuito sencillo y práctico para sumar las tensiones de los píxeles (5) sería interconectar el nodo flotante de los condensadores C1/C2 de los píxeles (5) del grupo. En este caso:

25

$$\Delta V_{diff} = -\frac{C_1}{C_2} \sum_i \Delta V_{lobi} \cong -\frac{C_1}{C_2} 4n_n U_T \log \frac{\prod_{i=1}^G I_{phi_{final}}}{\prod_{i=1}^G I_{phi_{initial}}} = -A_T U_T \log \frac{\prod_{i=1}^G I_{phi_{final}}}{\prod_{i=1}^G I_{phi_{initial}}} \quad (10)$$

Esta agrupación de píxeles (5) da lugar a una mayor sensibilidad de contraste para el grupo de píxeles (5).

30

- Nodos de interconexión (11) v_{oi} .

Los nodos de interconexión (11) v_{oi} propuestos se muestran en detalle, conectados en el píxel (5) en la figura 4 y en la figura 5. Cada nodo de interconexión (11) v_{oi} se conecta al

búfer (8) y a un píxel inmediatamente anterior por columna e inmediatamente posterior por fila dentro de la matriz (1). Además, ambas conexiones están unidas a una puesta a tierra (V_{reset}) común. En este caso, la tensión v_o resultante del grupo de píxeles (5) interconectados puede expresarse como:

$$V_o = 4n_n U_T \log \frac{\sum_{i=1}^G A_i I_{phi}}{G I_{sp}} \quad (11)$$

En este caso, puede observarse que se obtiene un promedio de la variación total de la corriente sobre los píxeles (5) en el grupo, como se desea. La figura 4 muestra el esquema modificado de un píxel DVS que permite la interconexión de píxeles en los nodos de interconexión (11) v_{oi} .

Las señales digitales m_{up} y m_{right} controlan la reconfiguración del píxel (5). Cuando ambas están desactivadas, el píxel (5) funciona individualmente en modo de alta resolución. Sin embargo, al activar las señales digitales m_{up} y m_{right} , el nodo de interconexión (11) puede ser interconectado con los píxeles (5) vecinos implementando el promedio de las fotocorrientes amplificadas.

La tensión v_o se amortigua y se amplifica y diferencia, de modo que:

$$\Delta V_{diff} = -\frac{c_1}{c_2} \Delta V_o \cong -\frac{c_1}{c_2} 4n_n U_T \log \frac{\sum_{i=1}^G I_{phi_{final}}}{\sum_{i=1}^G I_{phi_{initial}}} = -A_T U_T \log \frac{\sum_{i=1}^G I_{phi_{final}}}{\sum_{i=1}^G I_{phi_{initial}}} \quad (12)$$

El comportamiento es, pues, matemáticamente equivalente a unir los fotodiodos de los píxeles (5) vecinos y sumar sus fotocorrientes.

Basándose en las consideraciones anteriores, se ha seleccionado la tercera opción de agrupación de píxeles (5) para que los nodos de interconexión (11) de la fotocorriente v_o preamplificada sean compartidos por los píxeles (5) vecinos para construir regiones de baja resolución.

Por otra parte, se van a analizar diferentes arquitecturas para agrupar los píxeles (5) dentro de la matriz (1). Se describen a continuación distintas alternativas de arquitecturas foveadas reconfigurables digitalmente:

- Arquitectura foveada digitalmente reconfigurable 1

La figura 4 ilustra una realización particular de un píxel (5) donde los nodos interconexión (11) de la fotocorriente preamplificada son compartidos por los píxeles (5) vecinos. Las señales digitales de control m_{up} y m_{right} se almacenan digitalmente y son reconfigurables individualmente para cada píxel (5) de la matriz (1), como se muestra en la figura 5. En esta realización cada píxel requiere dos memorias de un bit.

Siempre que la señal digital m_{up} esté activa, el píxel (5) se configurará en modo de baja resolución y su señal v_o se conectará al píxel (5) vecino superior. Del mismo modo, cuando la señal m_{right} esté activa, el píxel (5) estará en baja resolución conectado a su píxel (5) vecino derecho. Utilizando este esquema, las regiones no foveadas de baja resolución con formas y tamaños arbitrarios pueden ser programadas externamente en el sensor, como se explicará a continuación.

Como se ilustra en la figura 5, la señal V_{reset} se pone en cortocircuito a tierra para que el píxel (5) se reinicie y no pueda generar eventos de salida, siempre que una de las señales m_{up} o m_{right} esté activa.

Sólo cuando m_{up} y m_{right} son simultáneamente bajos, el píxel (5) genera eventos de salida positivos y negativos. En consecuencia, sólo el píxel (5) superior izquierdo de cada región de macro-píxeles (píxeles agrupados) de baja resolución generará eventos de salida que representen los cambios temporales de las fotocorrientes promediadas en el grupo.

La arquitectura de la matriz (1) para la reconfiguración foveal de control totalmente digital se muestra en Figura 6. El sensor contiene una matriz (1) de NxM píxeles (5) con unos correspondientes circuitos periféricos para generar el evento-dirección de salida. Los circuitos periféricos comprenden unos bloques de control foveal X-FOV (13) e Y-FOV (14) vinculados a la matriz (1), que se añaden al sensor para la configuración digital de los macro-píxeles que definen las regiones no foveadas de baja resolución. La figura 7 muestra el esquema del bloque de agrupación X-FOV (13). Los circuitos periféricos comprenden también unos bloques de lectura (15, 16) conectados asimismo a la matriz (1).

Los bloques de control foveal (13, 14) reciben como entrada la dirección [fax, fay] del píxel (5) que debe ser reconfigurado y una señal digital de modo de configuración de dos bits

$[mdX, mdY]$. Los bloques de control foveal X-FOV (13) e Y-FOV (14) contienen decodificadores que activan las correspondientes señales de columna y fila $[FX_i, FY_j]$.

Como puede observarse en la figura 5, el píxel (5) correspondiente en el que las señales $[FX_i, FY_j]$ se seleccionan simultáneamente, actualiza sus señales de configuración de foveación m_{up} y m_{right} a las entradas digitales $[mdX, mdY]$.

Señales de control adicionales AEX, AEY y AD se añaden a los bloques de control foveal X-FOV (13) e Y-FOV (14) para permitir la configuración simultánea de toda la matriz (1), todos los píxeles (5) de una fila, o de todos los píxeles (5) de una columna.

Durante la configuración de un solo píxel (5), las señales AEX, AEY y AD deben ponerse a 1 lógico. Al poner AEX a 0, todas las señales de selección de columna FX_i están simultáneamente, de modo que todos los píxeles (5) de la fila FY_j seleccionada se configuran paralelamente con la misma entrada $[mdX, mdY]$.

Del mismo modo, al poner AEY a cero, todas las señales de selección de la fila FY_j se activan simultáneamente, de modo que todos los píxeles (5) de la columna FX_i seleccionada se configuran paralelamente con la misma entrada $[mdX, mdY]$. El ajuste de AEX y AEY simultáneamente a 0, permite la configuración paralela de todos los píxeles (5) de la matriz (1) al mismo estado $[mdX, mdY]$. El ajuste de la señal AD a 0 desactiva todas las señales de selección de filas y columnas $[FX_i, FY_j]$ deshabilitando la reconfiguración foveal.

25 - Arquitectura de foveo digitalmente reconfigurable 2

En la Figura 8 se muestra una arquitectura alternativa para definir las regiones no foveadas de baja resolución. Los $N \times M$ píxeles (5) de la matriz (1) se organizan en grupos de $n \times m$ píxeles (5) cada uno. En este caso cada grupo de $n \times m$ píxeles puede operar como un único macropíxel (baja resolución) o píxeles independientes (alta resolución).

La ventaja de esta arquitectura radica en que sólo se requiere una memoria de un bit para el grupo completo. Cada grupo puede configurarse utilizando bloques de control foveal X-FOV (13) e Y-FOV (14) en modo de alta resolución (HR) o de baja resolución (LR) utilizando un solo bit almacenado en el grupo. En cada grupo se usan dos variantes de

pixel: solo uno de los píxeles se realizará con la variante pixel-G (pixel gobernante), mostrado en Figura 14, mientras que el resto se realizará con la variante pixel-A, mostrado en la Figura 12. La posición concreta del pixel-G dentro del grupo no es relevante.

- 5 La Figura 10 muestra cómo se interconectan las distintas señales dentro del macropíxel. Cuando el bit de control almacenado HR es alto, cada píxel (5) funciona individualmente. Por el contrario, cuando el bit de control está desactivado, todos los píxeles (5) del grupo están interconectados y sólo un píxel del grupo genera eventos de salida.
- 10 La determinación de si cada grupo trabaja en alta o baja resolución se realiza gracias a un interruptor controlado por \overline{HR} , y que se activa en cada macro-píxel excepto en el píxel central que genera los eventos de salida del grupo.

Utilizando esta arquitectura, el número de líneas de control para configurar las regiones foveadas (R1, R2, R4) se reduce, así como el número de circuitos adicionales necesarios para almacenar la configuración foveal, que ahora es compartida por todos los píxeles (5) de un grupo. Sin embargo, las regiones no foveadas de baja resolución tienen ahora una resolución fija sin posibilidad de adaptación.

20 - Arquitectura foveada reconfigurable digitalmente 3

Una arquitectura de foveo que combina la mayor flexibilidad para configurar las regiones foveadas (R1, R2, R4) de la primera arquitectura mostrada en la figura 6 con la reducción del número de líneas de control de foveo y la compartición de la memoria de configuración entre los píxeles de un grupo propuesta en la figura 8 se muestra en la figura 9. En esta arquitectura, los píxeles (5) vecinos se agrupan de forma predefinida cuando se configuran en baja resolución como se propone en la segunda arquitectura de foveo.

Sin embargo, en este caso se añade la posibilidad de formar super-macro-píxeles con forma arbitraria. Para conectar los grupos vecinos, se puede utilizar un esquema similar al considerado en la primera arquitectura foveada como se muestra en la figura 9, utilizando nodos de interconexión (11) en cada grupo. En este caso, tres bits definen la resolución de cada grupo de píxeles (5).

En este caso, la topología de los píxeles es siempre la misma, que llamamos pixel-B y se muestra en la Figura 13. La Figura 11 muestra cómo se interconectan los píxeles entre sí.

Todos los píxeles se conectan de la misma manera, excepto uno de ellos (mostrado en la esquina inferior derecha en la Figura 11, aunque podría estar en cualquier posición).

- 5 La señal mdHR configura cada grupo para que funcione de forma individual (modo de alta resolución) o de forma interconectada con grupos de baja resolución interconectada. En caso de que HR esté activa, los píxeles funcionan de manera independiente (alta resolución) y todos los píxeles pueden generar eventos de salida. Cuando la señal HR está desactivada, sólo un píxel (5) del grupo emite eventos de salida.
- 10 Dos bits adicionales mdX y mdY pueden definir la interconectividad con los macropíxeles (5) superior y derecho siempre que la señal HR esté baja. En ese caso, sólo un píxel (5) de la esquina superior derecha del supra-grupo tendría su señal RT desactivada permitiéndole generar eventos de salida.

REIVINDICACIONES

- 1.- Sensor de visión dinámico foveado electrónicamente, que comprende:
- 5 - una matriz (1) NXM píxeles (5), en el que cada píxel (5) de coordenadas (x, y) comprende:
 - una etapa fotorreceptora (6),
 - una etapa de preamplificación (7), conectada a la etapa fotorreceptora (6),
 - un búfer (8), conectado a la etapa de preamplificación (7),
 - 10 - una etapa de amplificación y diferenciación (9) de tensión, conectada al búfer (8),
 - unos comparadores de salida (10) configurados para generar unos eventos de salida ON y OFF en función de unos umbrales de contraste, y
 - 15 - un nodo de interconexión (11) con píxeles (5) vecinos, conectado al búfer (8), comprendiendo cada nodo de interconexión (11):
 - una primera conexión ($v_{o_{up}}$) a un píxel (5) de coordenadas (-x, y-1),
 - una segunda conexión ($v_{o_{right}}$) a un píxel (5) de coordenadas (x+1, y),
 - 20 - una conexión a tierra (V_{reset}) vinculada a la primera conexión ($v_{o_{up}}$) y a la segunda conexión ($v_{o_{right}}$),
 - dos bloques de control foveal X-FOV (13) e Y-FOV (14), que permiten la selección de los píxeles (5) y la configuración de la conexión ($v_{o_{up}}$) y ($v_{o_{right}}$)
 - 25 - una primera señal de configuración (m_{up}),
 - una segunda señal de configuración (m_{right}), y
 - un primer bloque de lectura (15) conectado a los píxeles (5) por columnas, y un segundo bloque de lectura (16) conectado a los píxeles (5) por filas.
- 30 2.- El sensor de la reivindicación 1, en el que la matriz (1) comprende NXM píxeles (5) y en el que los píxeles (5) están organizados en grupos (G_11, G_1L, G_K1, G_KL) de nxm píxeles (5), y el que cada grupo (G_11, G_1L, G_K1, G_KL) comprende un interruptor de selección de estado de configuración del grupo (G_11, G_1L, G_K1, G_KL).

3.- El sensor de la reivindicación 2, en el que los grupos (G₁₁, G_{1L}, G_{K1}, G_{KL}) conforman a su vez supra-grupos unidos a través de un nodo de interconexión adicional, conectado entre cada grupo (G_{kl}) y un grupo superior (G_{kl-1}) y un grupo lateral (G_{k+1l}).

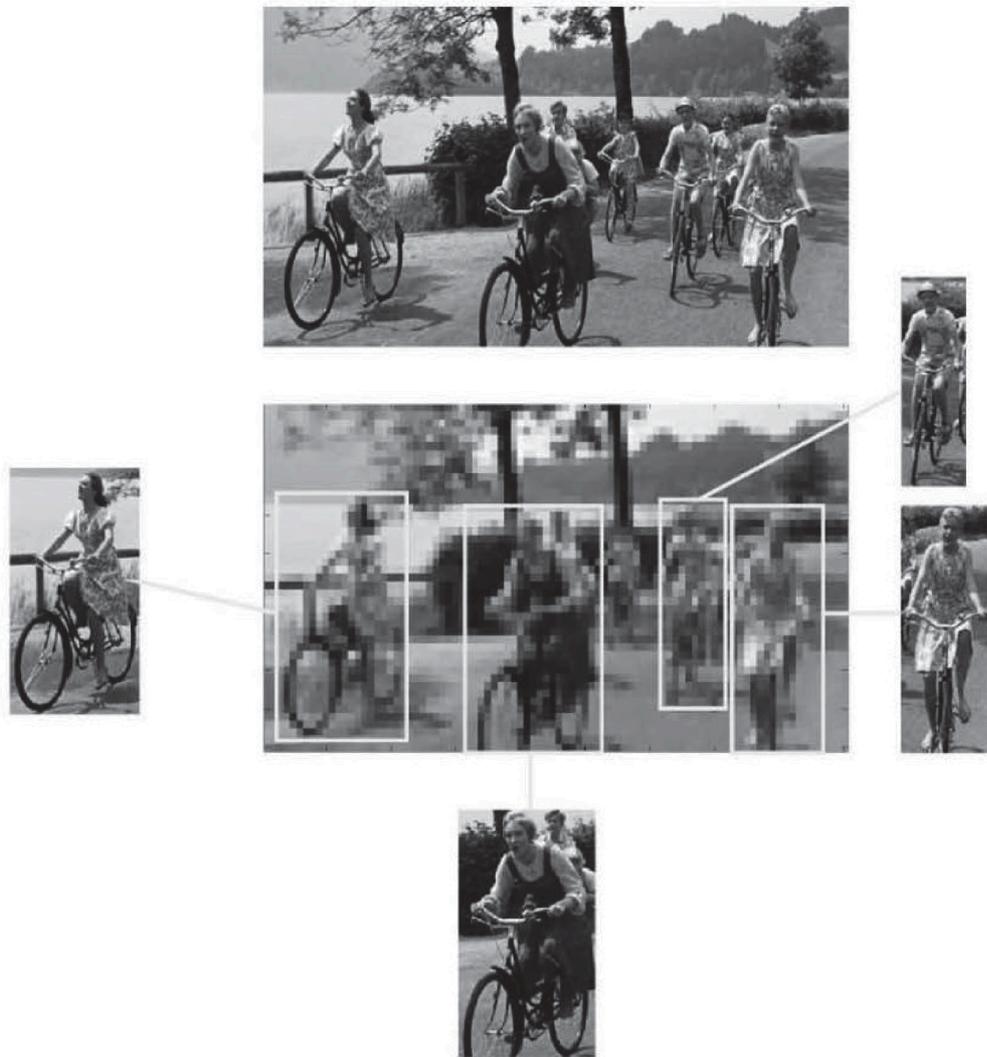


FIG. 1

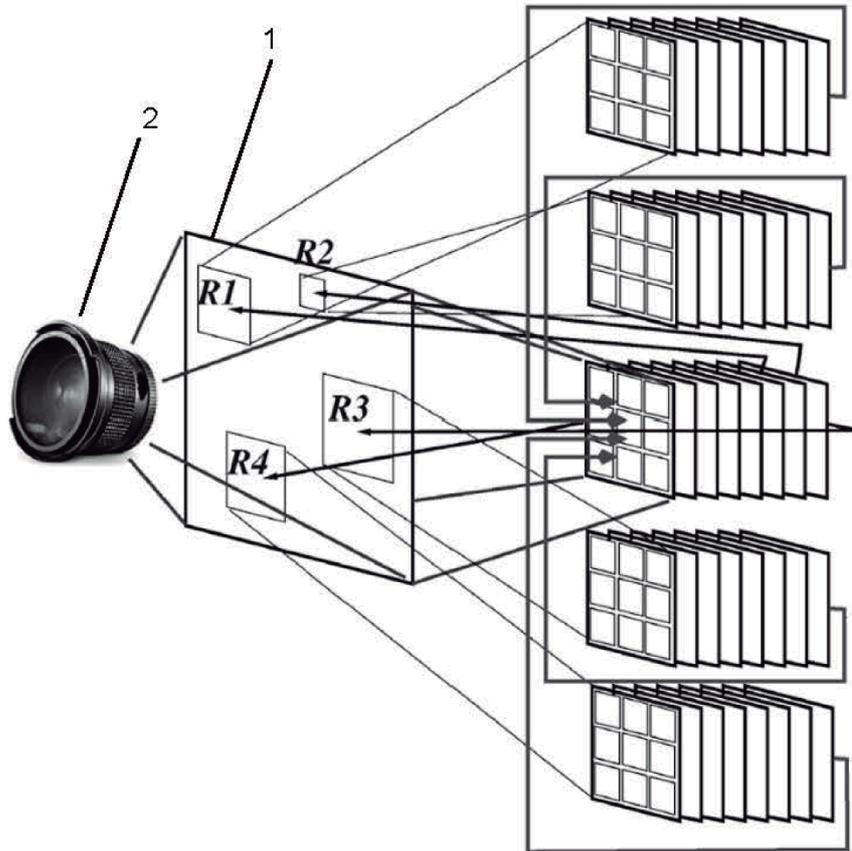
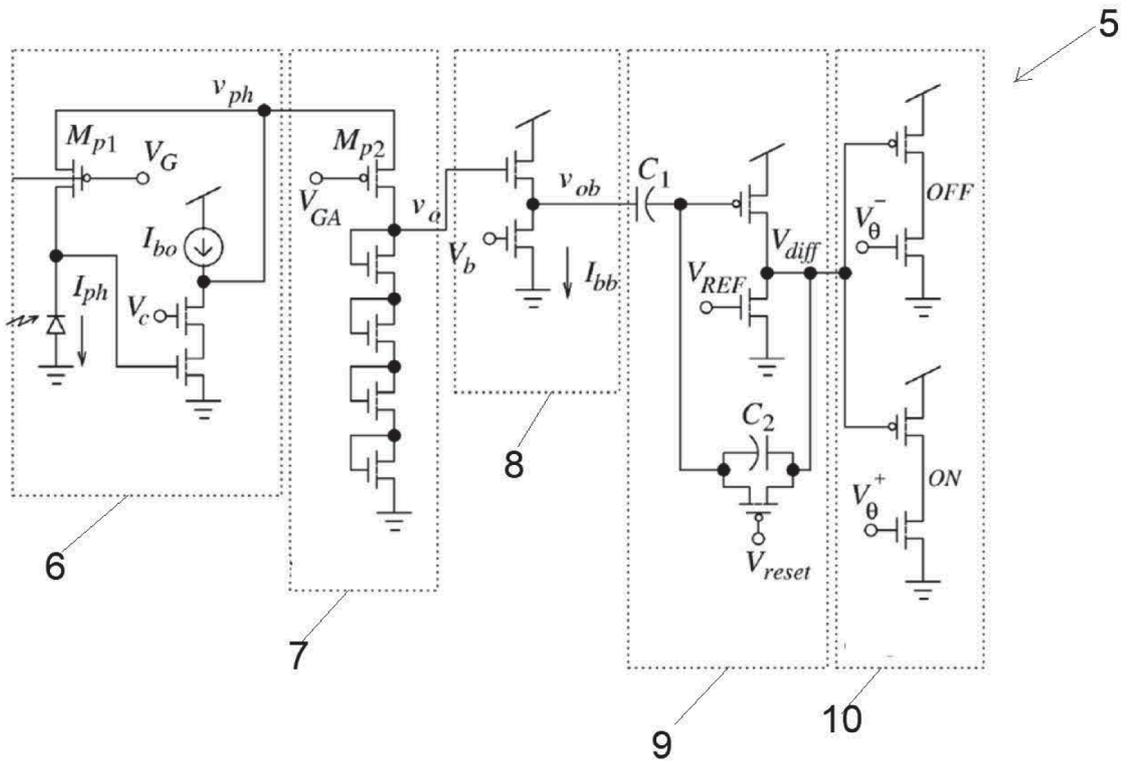


FIG. 2



(ESTADO DE LA TÉCNICA)

FIG. 3

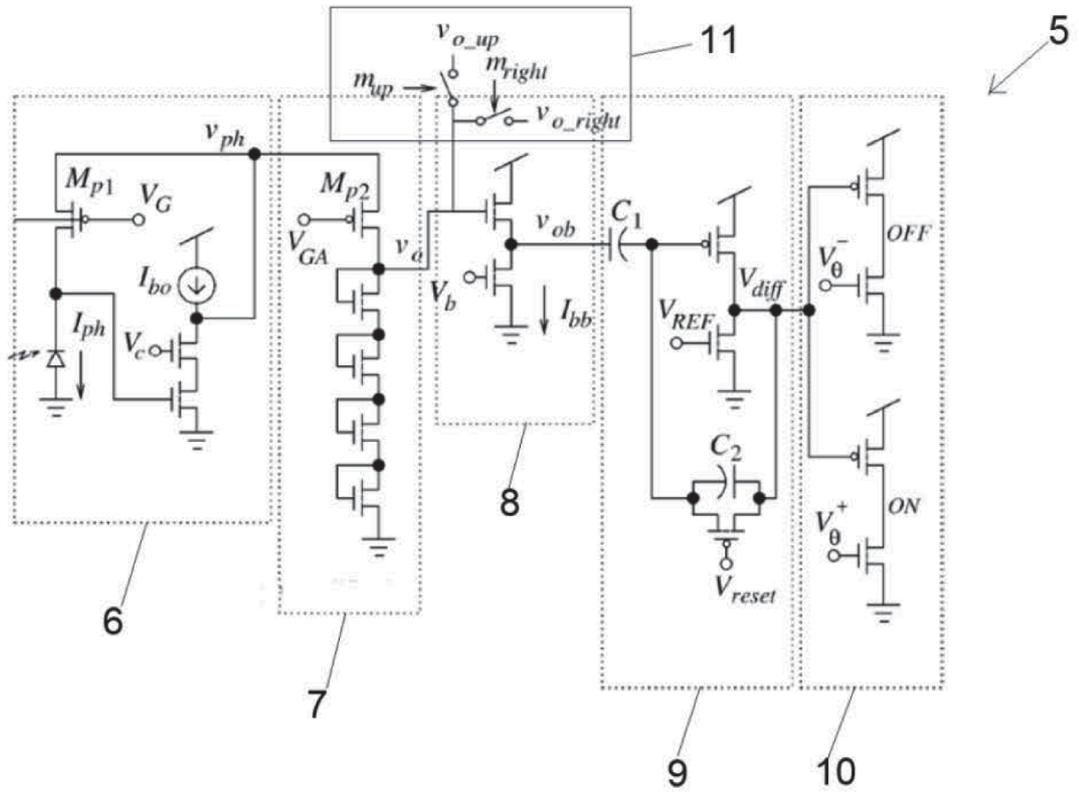


FIG. 4

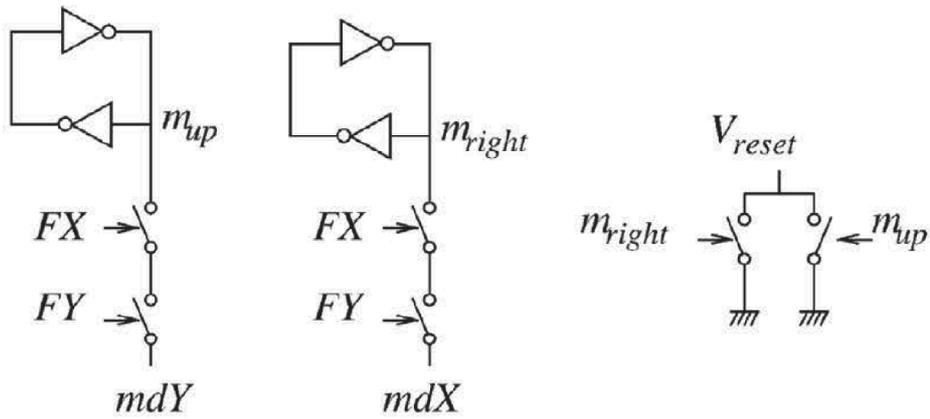


FIG. 5

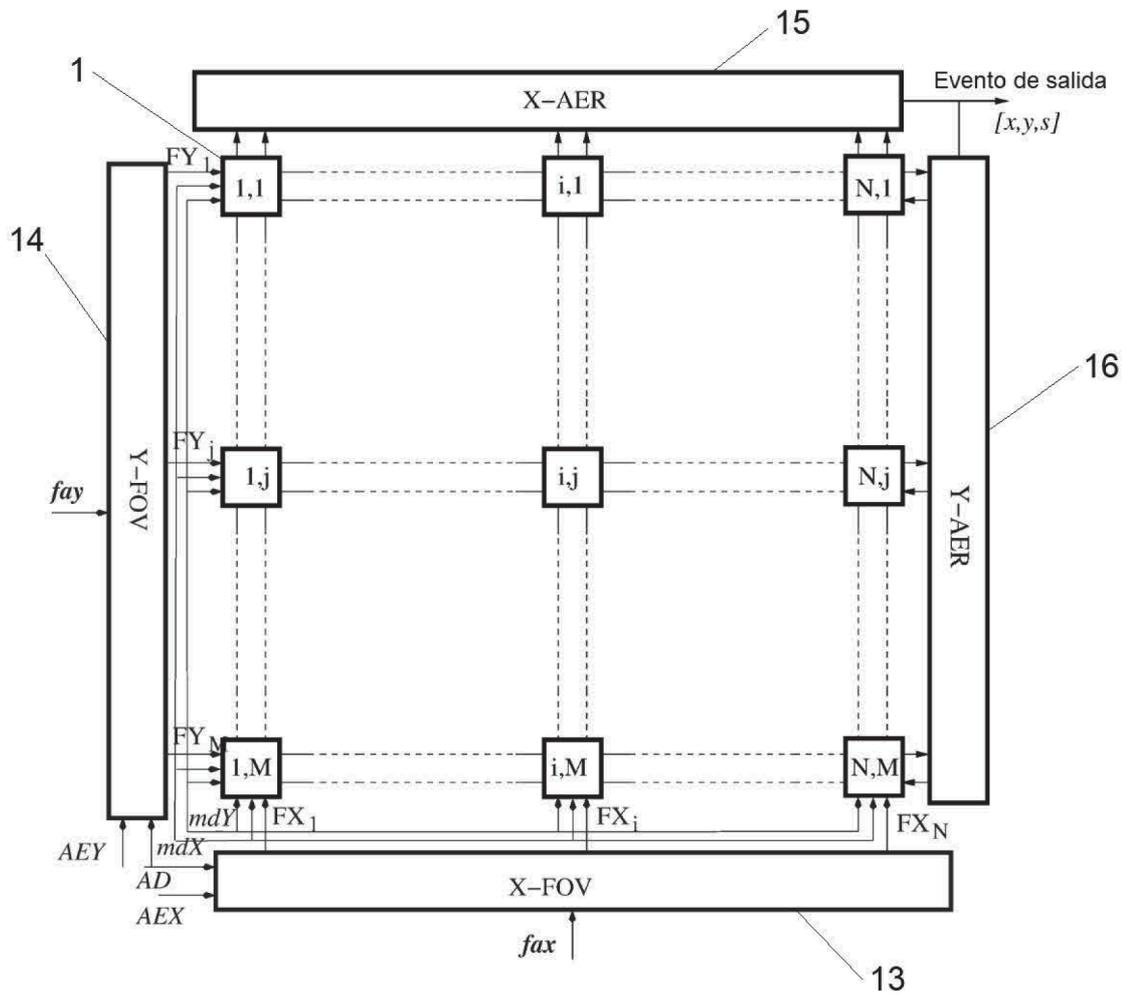


FIG. 6

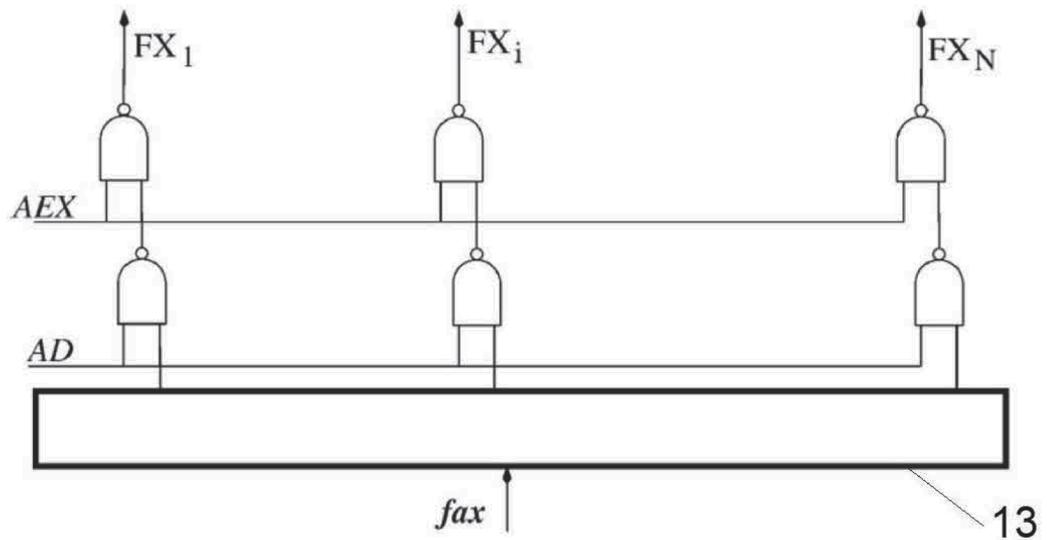


FIG. 7

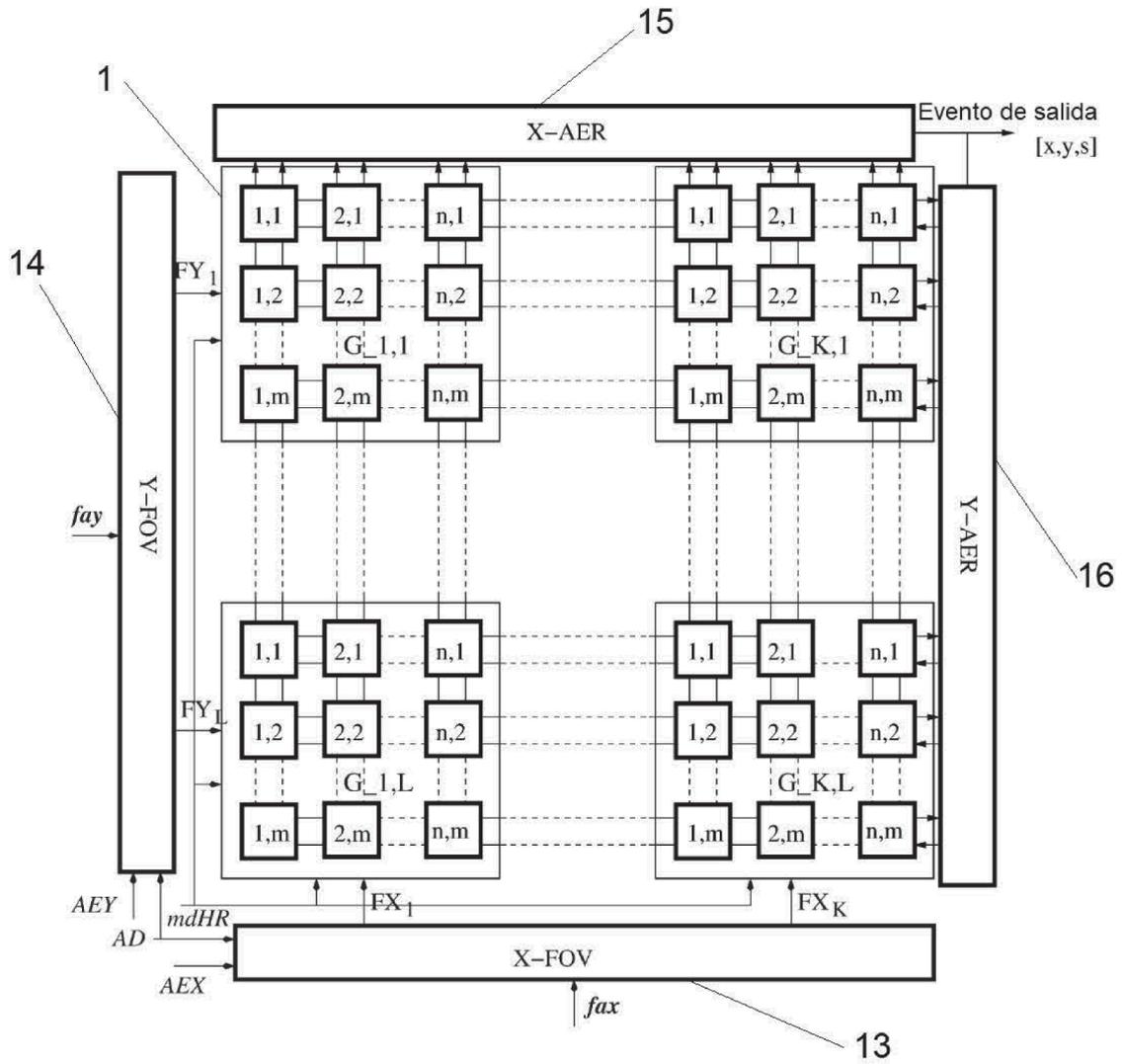


FIG. 8

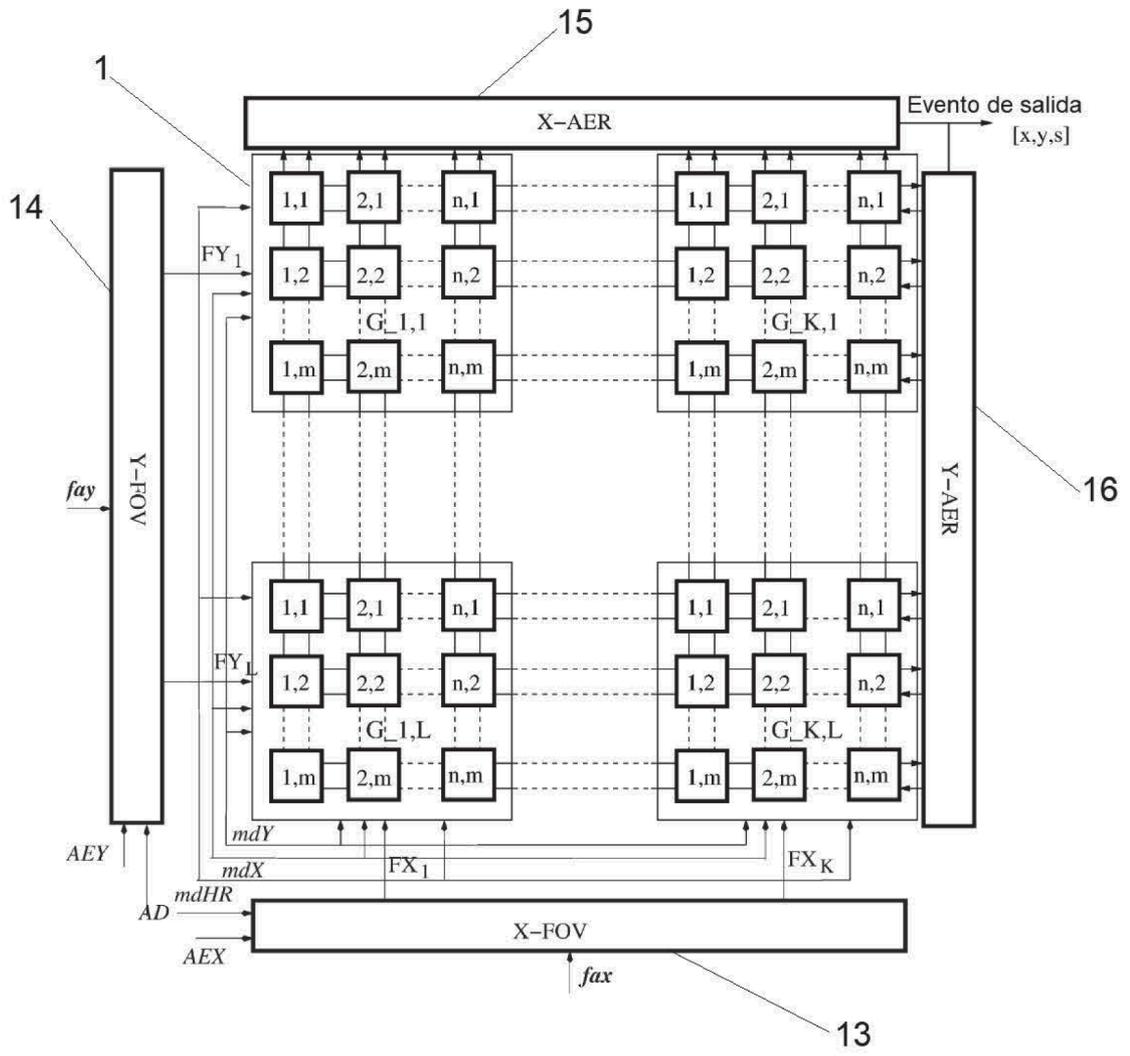


FIG. 9

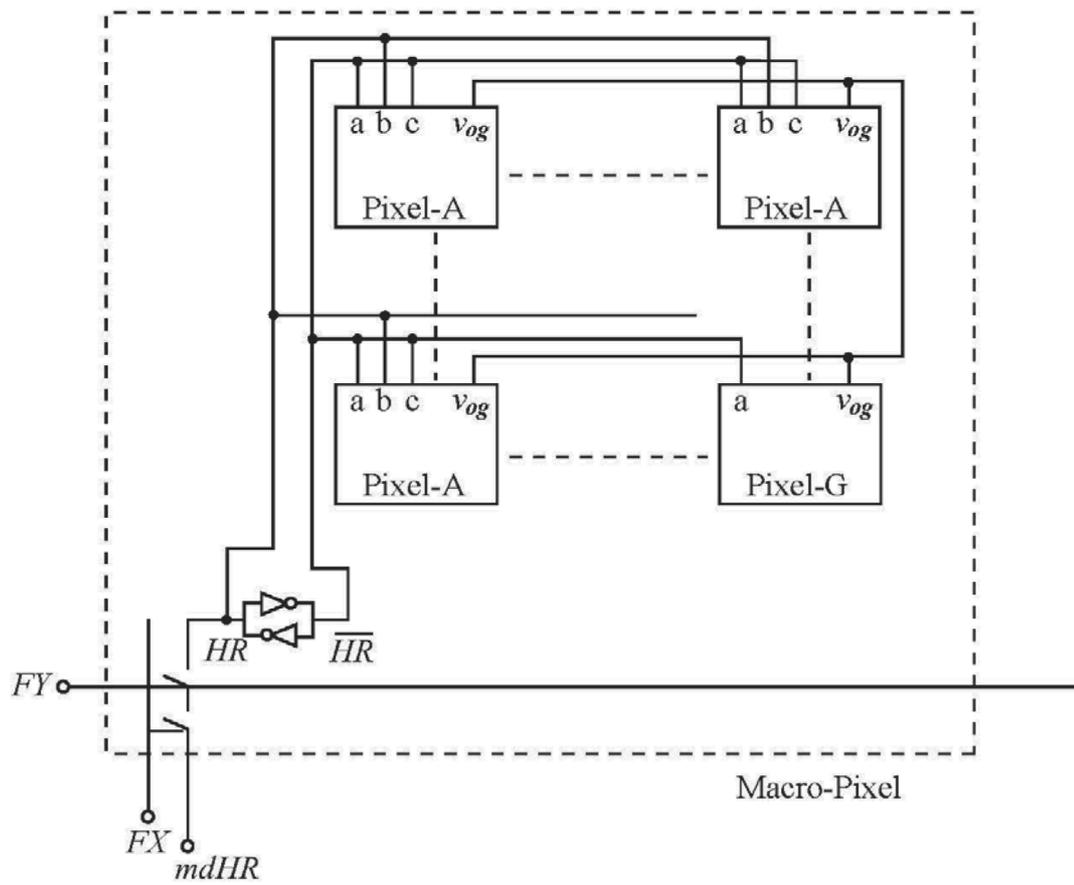


FIG. 10

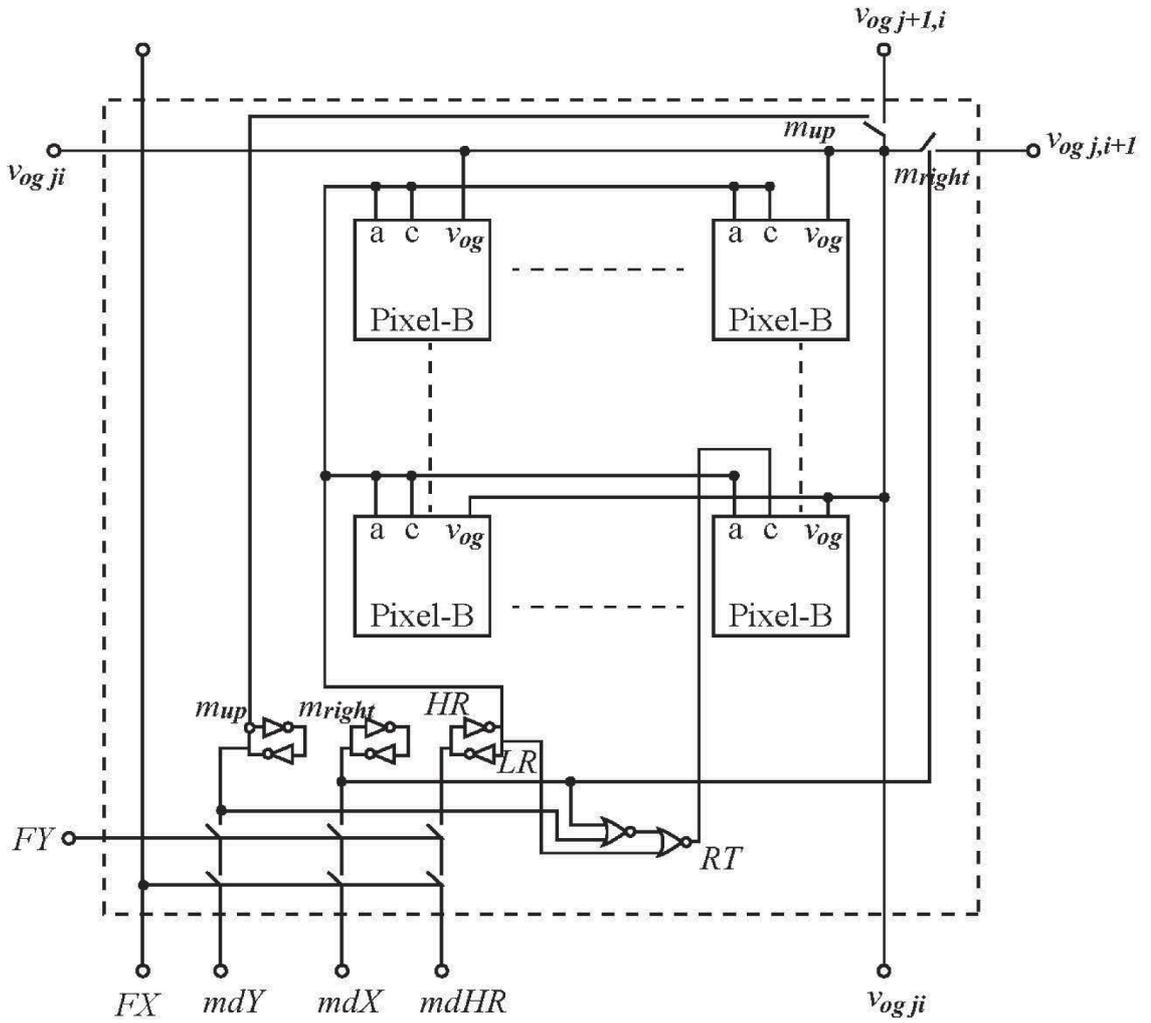


FIG. 11

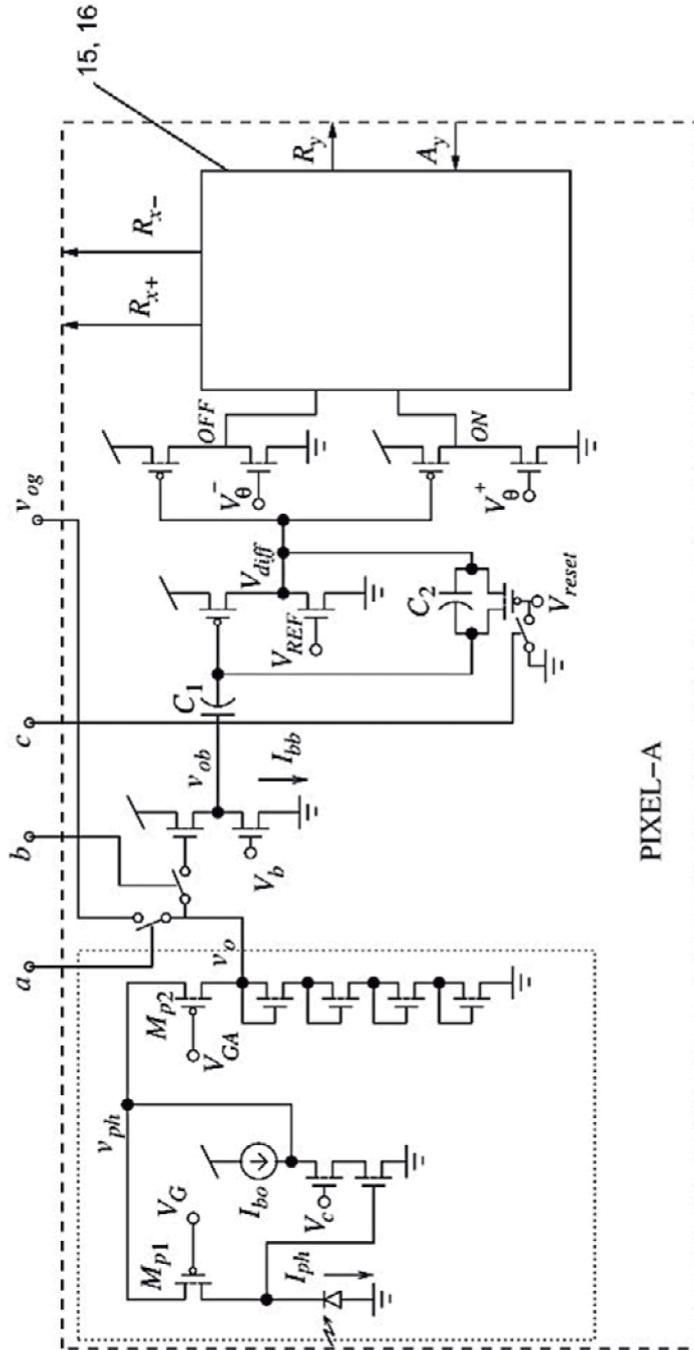


FIG. 12

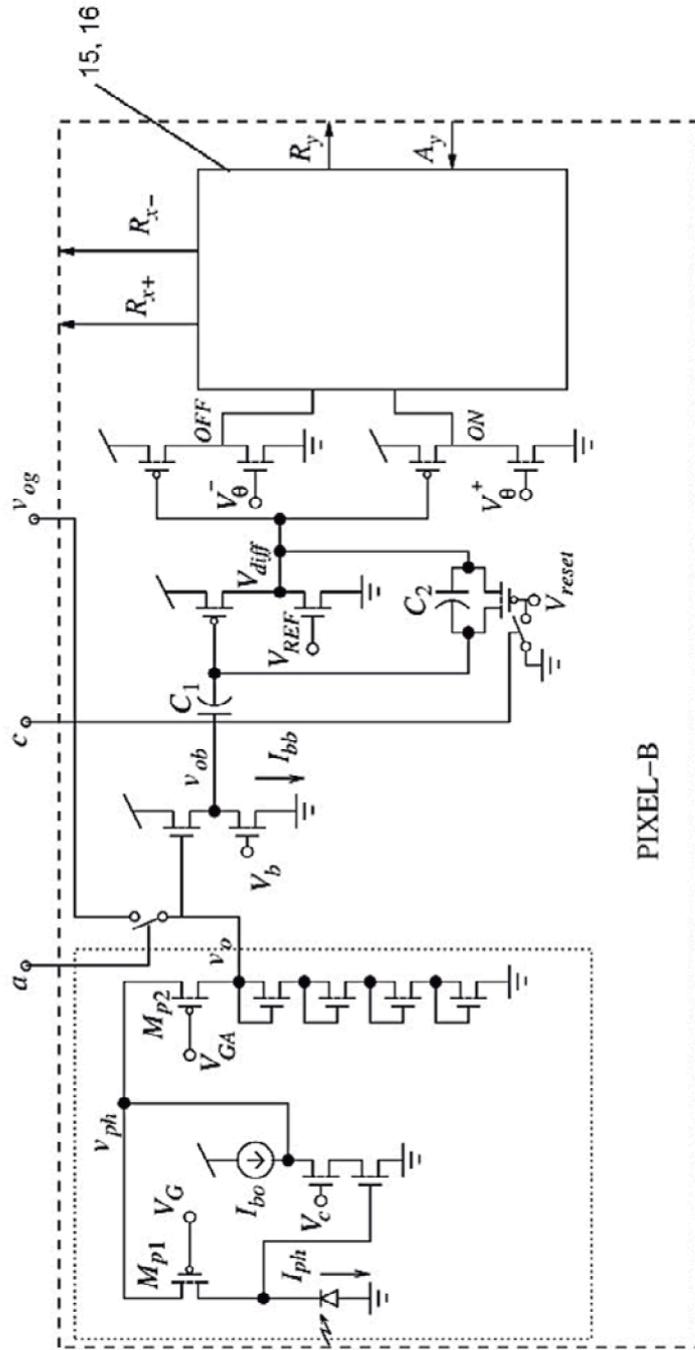


FIG. 13

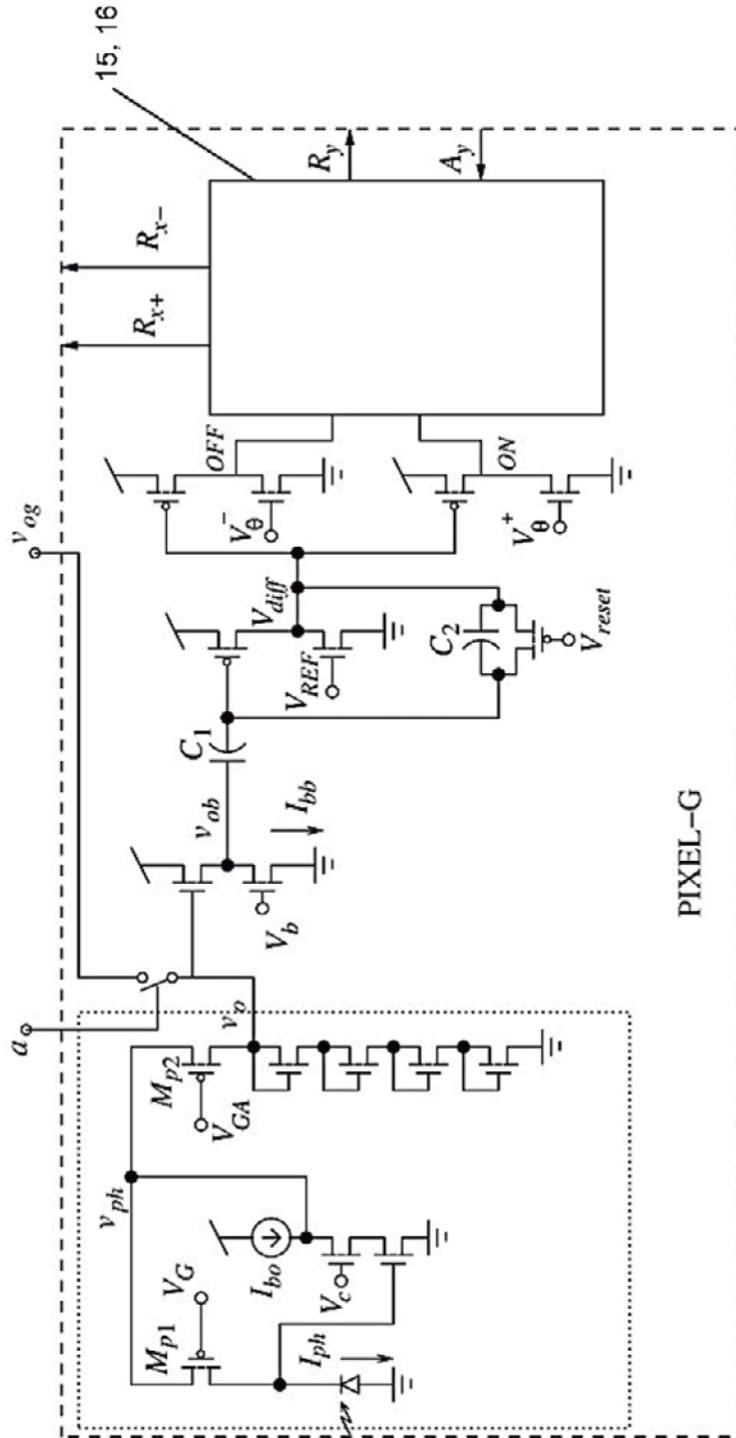


FIG. 14



OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

21 N.º solicitud: 202130929

22 Fecha de presentación de la solicitud: 05.10.2021

32 Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

51 Int. cl.: **G06T1/00** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	56 Documentos citados	Reivindicaciones afectadas
X	US 6839452 B1 (YANG GUANG et al.) 04/01/2005, desc. figs. 1-25	1-3
A	US 5262871 A (WILDER, JOSEPH et al.) 16/11/1993, todo el documento	1-3
A	US 2021089803 A1 (LUBANA, EKDEEP SINGH et al.) 25/03/2021, todo el documento	1-3
A	US 2018270436 A1 (IVARSSON, MAGNUS et al.) 20/09/2018, todo el documento	1-3

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº:

Fecha de realización del informe
23.09.2022

Examinador
G. Madariaga Domínguez

Página
1/2

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06T

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

WPI, EPODOC