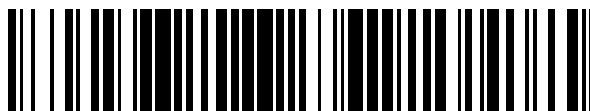


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 391 418**

51 Int. Cl.:  
**G06T 1/20** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09152275 .5**  
96 Fecha de presentación: **06.02.2009**  
97 Número de publicación de la solicitud: **2169614**  
97 Fecha de publicación de la solicitud: **31.03.2010**

54 Título: **Análisis de vídeo**

30 Prioridad:  
**26.09.2008 US 100575 P**  
**21.11.2008 US 275832**

45 Fecha de publicación de la mención BOPI:  
**26.11.2012**

45 Fecha de la publicación del folleto de la patente:  
**26.11.2012**

73 Titular/es:  
**AXIS AB (100.0%)**  
**EMDALAVÄGEN 14**  
**223 69 LUND, SE**

72 Inventor/es:  
**ALMBLADH, JOHAN**

74 Agente/Representante:  
**No consta**

ES 2 391 418 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

Análisis de vídeo.

Campo técnico

5 La presente descripción se refiere a un método para realizar de manera eficaz operaciones de análisis de vídeo, a un programa informático correspondiente y a una cámara.

Antecedentes

10 El análisis de vídeo o análisis de contenido de vídeo varían desde la detección de movimiento en vídeo y detección de audio, hasta sistemas más avanzados que incluyen la detección de manipulaciones de cámara, recuento de personas, detección de objetos que cruzan líneas o áreas de interés, reconocimiento de placas de matrícula de vehículos, segmentación de vídeo en partes en primer plano y segundo plano, seguimiento de objetos, análisis de tráfico, extracción de metadatos, aplicaciones biométricas y reconocimiento facial. El análisis de vídeo también hace que los sistemas de vigilancia sean más inteligentes para reducir las enormes cantidades de datos de imagen a niveles razonables. Los sistemas de vigilancia de vídeo inteligentes pueden por ejemplo analizar y etiquetar automáticamente los vídeos de vigilancia en tiempo real, detectar actividades sospechosas, iniciar la grabación de vídeo, activar alarmas u otras acciones para alertar a los operadores o a otras personas.

15 En aplicaciones de vigilancia, a menudo se usa el análisis de vídeo para detectar movimiento. La detección de movimiento es una manera de definir la actividad en una escena analizando datos de imagen, y puede realizarse en todo el campo visual de la cámara de vigilancia o en un área de interés definida por el usuario. Además, un sistema de vigilancia de vídeo con capacidades de detección de movimiento puede detectar el movimiento de manera más fiable que un operador humano, y por tanto puede liberar a los operadores humanos de tener que mirar fijamente múltiples monitores de vídeo durante muchas horas. En su lugar, el sistema de vigilancia de vídeo con capacidades de detección de movimiento puede alertar al operador usando un indicador visual, un indicador de audio o ambos cuando se detecta el movimiento. Un sistema de vigilancia de este tipo también puede centrarse automáticamente en una cámara de vigilancia en el área en el que se detectó el movimiento para obtener una imagen más detallada.

20 Como reconoce el presente inventor, una cámara de red que pueda realizar un análisis de vídeo reduciría la carga de trabajo de un sistema de procesamiento de imagen centralizado, y conservaría el ancho de banda de red de valor. Una cámara de red de este tipo permitiría sistemas de vigilancia iniciados por eventos verdaderos en los que la detección de movimiento mediante la cámara podría activar procesos automáticos predefinidos, tales como ajustar la temperatura, activar alarmas, bloquear/desbloquear puertas, etc.

25 Sin embargo, como el análisis de vídeo a menudo supone realizar varias operaciones relativamente simples en grandes cantidades de datos de píxel, los métodos actuales no se prestan a aplicaciones móviles o integradas, tales como una cámara de red. Se han desarrollado los métodos convencionales de reducción rápida de la cantidad de datos (número de píxeles) procesados durante el análisis de vídeo para reducir la carga de procesamiento de CPU.

30 Un método convencional de este tipo pasa de manera secuencial por todos los píxeles en una imagen para identificar los píxeles que son de interés. Si el píxel no es de interés, el método pasa al siguiente píxel para su análisis. En este contexto, "de interés" significa que el píxel contiene información relevante para el análisis que está llevándose a cabo, por ejemplo la información de movimiento. Si el píxel es de interés, el método realiza las operaciones relevantes en el píxel antes de avanzar al siguiente píxel. Por tanto, este método anida el procesamiento de vídeo de un píxel dentro de la rutina que identifica píxeles de interés. En el caso de un filtro, por ejemplo, mientras el filtro se aplica de manera selectiva sólo a los píxeles de interés, el filtro se sigue aplicando a sólo un píxel de interés cada vez.

35 Muchos procesadores modernos pueden realizar instrucciones de una instrucción, múltiples datos (SIMD) para procesar múltiples campos de datos en paralelo, y aumentar el rendimiento. En procesadores con capacidad de instrucción SIMD, cada registro se divide en al menos dos campos. Cada campo representa datos que son independientes de los datos en otros campos. Por ejemplo, en un contexto de análisis de vídeo, cada campo puede representar un píxel individual. Como el procesador puede ejecutar una instrucción SIMD en todo un registro, los píxeles contenidos en los campos del registro se procesan simultáneamente. Por tanto, el rendimiento de un procesador que soporta SIMD puede ser significativamente mejor que el rendimiento de un procesador de propósito general.

40 Sin embargo, como reconoce el presente inventor, el método convencional descrito anteriormente para seleccionar píxeles que van a procesarse no es adecuado para el procesamiento de instrucciones SIMD. Debido a que el método anida el procesamiento de análisis de vídeo dentro de la rutina de selección de píxeles, se fuerza a un procesador que soporta SIMD a procesar los píxeles seleccionados de uno en uno más o menos de la misma manera que un procesador de propósito general, invalidando de este modo las ventajas adquiridas empleando instrucciones SIMD.

45 Un resultado de las ineficiencias descritas anteriormente del análisis de vídeo convencional es que los sistemas informáticos potentes que tienen altas capacidades de procesamiento se siguen prefiriendo para realizar funciones de análisis

de vídeo, tales como el filtrado y la detección de movimiento. Por tanto, estos métodos no son adecuados para la implementación local de análisis de vídeo en cámaras de red.

5 Un sistema de vigilancia de vídeo típico incluye múltiples cámaras de vigilancia de vídeo conectadas a una unidad de procesamiento central mediante una red, tal como una red basada en IP. A menudo la red basada en IP no está dedicada exclusivamente al sistema de vigilancia de vídeo, sino que se comparte con otra aplicación basada en la red, tal como correo electrónico, navegación por web, sistemas de bases de datos, y similares. En el caso en el que el sistema de vigilancia de vídeo emplea el análisis de vídeo convencional realizado por la unidad de procesamiento central, cada cámara debe proporcionar un flujo de imagen de vídeo sin procesar a la unidad de procesamiento central. Esto sitúa una enorme cantidad de datos de vídeo en la red con grandes cantidades de tráfico de datos, requiriendo un ancho de banda que por el contrario podría usarse por otra aplicación de red.

10 El documento WO 2005/106786 describe un aparato y método de procesamiento de imagen que identifica una zona de interés en un fotograma de imagen. La zona de interés se define dentro de un cuadro límite en curso, que se usa para definir un cuadro de límite de "mejor ajuste". A continuación vuelve a escanearse el cuadro límite de mejor ajuste en una retícula ortogonal. La retícula ortogonal se procesa entonces mediante un procesador SIMD para el procesamiento paralelo de píxeles.

### Sumario

La presente descripción proporciona un método, un programa informático, y una cámara según las reivindicaciones independientes.

### Breve descripción de los dibujos

20 Una apreciación más completa de la invención y muchas de las ventajas acompañantes de la misma se obtendrán fácilmente a medida que ésta se entienda mejor con referencia a la siguiente descripción detallada considerada en relación con los dibujos adjuntos, en los que:

la figura 1 es un diagrama esquemático de un sistema de análisis de vídeo según una realización a modo de ejemplo de la presente descripción;

25 la figura 2 es un diagrama esquemático de un sistema informático en el que se ejecuta un método de análisis de vídeo según una realización a modo de ejemplo de la presente descripción;

la figura 3 es un diagrama esquemático de un método de análisis de vídeo según una realización a modo de ejemplo de la presente descripción;

30 la figura 4 es un diagrama de flujo de la aplicación de máscara binaria según una realización a modo de ejemplo de la presente descripción;

la figura 5 es un diagrama esquemático de dilatación y aplicación morfológica de la máscara binaria según una realización a modo de ejemplo de la presente descripción;

la figura 6 es un diagrama de flujo de dilatación morfológica según una realización a modo de ejemplo de la presente descripción;

35 la figura 7 es un diagrama de recopilación de píxeles según una realización de la presente descripción;

la figura 8 es un diagrama de dispersión de píxeles según una realización de la presente descripción; y

la figura 9 es un diagrama de flujo del método de análisis de vídeo según una realización a modo de ejemplo de la presente descripción;

la figura 10 es un diagrama de un sistema de vigilancia de vídeo; y

40 la figura 11 es un diagrama de una cámara de red según una realización a modo de ejemplo de la presente descripción.

### Descripción detallada de las realizaciones

La presente descripción describe una realización preferida con referencia a los dibujos adjuntos.

45 La figura 1 es un diagrama de un sistema de análisis de vídeo según una realización a modo de ejemplo de la presente descripción. El sistema incluye una unidad 3 de recopilación para recopilar píxeles de interés, una unidad 4 de dispersión para dispersar los píxeles de interés una vez que se completa el procesamiento, una memoria 2 electrónica para almacenar los píxeles de interés, y un procesador 1 de datos para realizar el procesamiento de análisis de vídeo. Todos los componentes mencionados anteriormente están interconectados mediante un bus 5 de datos.

5 El sistema de procesamiento de análisis de vídeo puede implementarse en un dispositivo de ordenador personal, tal como un PC que emplea un procesador Pentium. El conjunto de instrucciones de la unidad 3 de recopilación y la unidad 4 de dispersión puede proporcionarse como una aplicación de utilidad, un componente o demonio en segundo plano de un sistema operativo, o una combinación de los mismos que se ejecutan junto con un procesador y un sistema operativo tal como Microsoft VISTA<sup>®</sup>, Unix, Solaris, Linux, Apple MAC-OS y otros sistemas conocidos por los expertos en la técnica.

La memoria 2 electrónica puede ser una memoria de acceso aleatorio (RAM) dinámica o estática integrada en el PC, o puede ser otro dispositivo de memoria tal como FLASH, EEPROM, disco óptico, servidores, y similares. La memoria también puede ser extraíble del sistema de procesamiento de análisis de vídeo.

10 La figura 2 es un diagrama esquemático de un ordenador cliente, o un ordenador central de vigilancia según la presente realización. El ordenador incluye un dispositivo 100 informático al que se conectan un teclado 112, un dispositivo 111 indicador y una pantalla 110.

15 El dispositivo 100 informático incluye un procesador 101 que realiza procesos de control y en relación con una memoria 105 principal para el almacenamiento temporal de datos que van a procesarse, y una ROM 106 que almacena las instrucciones ejecutadas por el procesador 101. Un controlador 107 de disco controla al menos un disco 108 y un CD-ROM 109 en los que pueden almacenarse secuencias de vídeo antes y después del procesamiento. Se usa un controlador 104 de pantalla para controlar una pantalla externa, y se usa una interfaz 102 de E/S para controlar otros dispositivos externos, tales como el teclado 112, el dispositivo 111 indicador y la pantalla 110. Un bus 103 interconecta todos estos componentes.

20 El dispositivo 100 informático puede ser, por ejemplo, un PC que emplee un procesador Pentium. El conjunto de instrucciones almacenado en la ROM 106 puede proporcionarse como una aplicación de utilidad, componente o demonio en segundo plano de un sistema operativo, o una combinación de los mismos que se ejecutan junto con un procesador y un sistema operativo tal como VISTA, UNIX, SOLARIS, LINUX, APPLE MAC-OS y otros sistemas conocidos por los expertos en la técnica.

25 La memoria 105 principal del dispositivo 100 informático puede ser una memoria de acceso aleatorio (RAM) dinámica o estática integrada en el PC, o puede ser otro dispositivo de memoria tal como FLASH, EEPROM, disco óptico, servidores, y similares. Adicionalmente, al menos parte de la memoria 105 principal también puede ser extraíble del sistema de detección de movimiento de vídeo.

30 El sistema de procesamiento de análisis de vídeo puede implementarse también en dispositivos de hardware tales como FPGA, ASIC, microcontroladores, PLD u otros dispositivos de este tipo conocidos en la técnica.

35 La figura 3 representa el método de procesamiento de análisis de vídeo según una realización a modo de ejemplo de la descripción. El método de procesamiento de análisis de vídeo tiene tres fases: recopilar píxeles, procesar píxeles y dispersar píxeles. En la fase 22 de recopilación, se usa una máscara 23 binaria para recopilar píxeles seleccionados para el procesamiento de análisis de vídeo de una imagen 21. La máscara 23 binaria es un mapa de bits que define las posiciones de píxel de los píxeles de interés usando unos y ceros. Cada píxel se representa mediante un único bit en la máscara 23 binaria. Un bit que tiene un valor de uno define un píxel seleccionado para el procesamiento (es decir un píxel "de interés") y un bit que tiene un valor de cero define un píxel no seleccionado para el procesamiento. Sin embargo, también puede usarse el caso inverso con la presente descripción, concretamente un píxel de interés puede denotarse por un valor cero y un píxel que puede no procesarse se denota por un valor de uno.

40 La identificación de píxeles de interés usando la máscara 23 binaria es muy eficaz puesto que varias posiciones de píxel de la máscara binaria pueden someterse a prueba simultáneamente. La figura 4 es un diagrama de flujo de la aplicación de la máscara binaria a una imagen. En este ejemplo, se usa un registro de 32 bits, aunque este proceso puede usarse con registros de cualquier longitud. El proceso comienza en la etapa S40. Se carga un registro con una máscara binaria de 32 bits correspondiente a una imagen de 32 píxeles en la etapa S41. En la etapa S42, el registro se compara con cero, ya que en este ejemplo un valor de cero define un píxel no seleccionado para el procesamiento. Si el registro es igual a cero el proceso continúa con la etapa S51 para determinar si se ha alcanzado el final de la máscara. Si se ha alcanzado el final de la máscara, el proceso termina en la etapa S53.

45 Sin embargo, si hay más bits de máscara, y por tanto píxeles, para procesar, se hace avanzar en 32 un puntero que determina el bloque de la máscara de 32 bits que se examina en S52, y el proceso vuelve a la etapa S41 para cargar un nuevo conjunto de máscara de 32 de bits en el registro. Si el registro no es igual a cero en S42, se divide el registro en dos palabras de 16 bits, una palabra más significativa (MSW) y una palabra menos significativa (LSW), en S43. En S44 la MSW se compara con cero. Si la MSW no es igual a cero, se divide en dos bytes de 8 bits, un byte más significativo (MSB) y un byte menos significativo (LSB), en S45. En S46, el MSB se compara con cero, y si el MSB no es igual a cero, se busca linealmente para identificar los bits distintos de cero en S47. A continuación se ejecuta una serie similar de etapas para el LSB en S49 y S47. Los píxeles distintos de cero se almacenan en S48 como "píxeles de interés" antes de volver a S51.

Se ejecuta un proceso similar para la LSW, comenzando con la etapa S50. Una vez que se han recopilado todos los píxeles de interés, y se alcanza el final de la máscara, el proceso termina en S53.

5 Mientras que en el ejemplo anterior, sólo se usan dos niveles de división (es decir desde 32 bits hasta 16 bits y desde 16 bits hasta 8 bits) son posibles niveles de división adicionales, o pueden usarse menos niveles de división. Alternativamente, el registro puede buscarse linealmente de manera directa sin realizar ninguna división. En este contexto, división significa crear dos grupos de píxeles, iguales en el número de píxeles, a partir de un grupo mayor de píxeles. Por ejemplo, un primer nivel de división de un grupo de 8 bits forma dos grupos de 4 bits, etc.

10 En la recopilación de píxeles descrita anteriormente, los píxeles de interés definidos por la máscara 23 binaria se recopilan en un orden predefinido en una matriz para su procesamiento adicional. Por ejemplo, los píxeles pueden recopilarse de izquierda a derecha en una matriz unidimensional. Sin embargo, la información relativa a los píxeles vecinos se pierde en la dirección vertical, y en la dirección horizontal para los píxeles en el límite de un área que está recopilándose. Muchas operaciones de análisis de vídeo requieren tal información de vecinos, y por tanto requieren recopilar píxeles vecinos adicionales adyacentes a los píxeles de interés. Estas operaciones incluyen, por ejemplo, filtros de gradiente Sobel. La recopilación de los píxeles adicionales en la dirección horizontal se consigue de manera eficaz a través de la dilatación morfológica de la máscara 23 binaria en la dirección horizontal para incluir los píxeles vecinos.

15 En el caso de dilatación morfológica en la dirección vertical, se realizan tres pasos de recopilación: un paso para la máscara dilatada horizontalmente trasladada una fila por encima de su posición original; otro paso en su posición original; y un último paso de una fila de píxeles por debajo de su posición original. Esto genera una matriz de 3xN. Los píxeles de interés tienen ahora sus respectivos píxeles vecinos a partir de la imagen original en las mismas posiciones relativas en la matriz. Se añaden píxeles de relleno como píxeles vecinos a los píxeles de interés situados a lo largo de un límite.

20 La figura 5 es un ejemplo simplificado del proceso de recopilación descrito anteriormente. Los píxeles 58 de interés están en gris, los píxeles 52 marcados con una "X" son píxeles recopilados como resultado de la dilatación morfológica, los píxeles marcados con una "X" dentro de un círculo son píxeles 56 de relleno, y los píxeles marcados con un recuadro muestran píxeles 55 vecinos adicionales en una dirección vertical. La figura 5 también incluye la matriz 59 resultante en la que se muestran los píxeles 58 de interés recopilados en el segundo paso en relación con sus vecinos más cercanos. También son posibles otros métodos sin apartarse del alcance de la presente descripción. Por ejemplo, los píxeles pueden recopilarse en un único paso de recopilación.

25 La figura 6 es un diagrama de flujo de la aplicación de la máscara 54 binaria a una imagen según una realización a modo de ejemplo de la descripción. En la etapa S61, la máscara 54 binaria (véase la figura 5) se desplaza hacia arriba una fila de píxeles para recopilar los vecinos verticales por encima de los píxeles 58 de interés. A continuación, la máscara 54 binaria se aplica a la imagen 51 en S62, y los píxeles correspondientes se almacenan en la fila 59a superior de la matriz 59 en la etapa S63. La máscara binaria se desplaza de vuelta a su posición original en la etapa S64 para recopilar los píxeles 58 de interés y sus respectivos vecinos horizontales. En la etapa S65 la máscara 54 binaria se aplica a la imagen 51 tal como se describe anteriormente, y los píxeles correspondientes se almacenan en la segunda fila 59b de la matriz 59 en la etapa S66. La máscara 54 binaria se desplaza entonces hacia abajo una fila de píxeles para recopilar los vecinos verticales por debajo de los píxeles 58 de interés. La máscara 54 binaria se aplica en la etapa S68 y los píxeles se almacenan en la tercera fila 59c de la matriz 59 en la etapa S69.

30 Debe observarse que cualquier descripción de procesos o bloques en los diagramas de flujo deben entenderse como módulos representativos, segmentos, partes de código que incluyen una o más instrucciones ejecutables para implementar funciones lógicas específicas o etapas en el proceso, y se incluyen implementaciones alternativas dentro del alcance de la realización a modo de ejemplo de la presente invención en las que las funciones pueden ejecutarse fuera del orden que se muestra o comenta, incluyendo sustancialmente de manera simultánea o en orden inverso, dependiendo de la funcionalidad implicada, tal como se entenderá por los expertos en la técnica.

35 La recopilación de los píxeles en una matriz de píxeles de 3xN, tal como se describió anteriormente, permite el uso eficiente de las instrucciones SIMD porque la matriz de píxeles contiene sólo algunos de los píxeles de la imagen original. Una instrucción SIMD aplicada a una fila o columna de la matriz de píxeles procesa múltiples píxeles de interés simultáneamente, a diferencia de los métodos convencionales que procesan un píxel de interés cada vez. Ejemplos de instrucciones SIMD realizadas en la matriz de píxeles de 3xN incluyen una instrucción de empalme, una instrucción de doble valor absoluto, una instrucción de promediado que tiene un resultado redondeado, una instrucción de promediado que tiene una instrucción truncada, y una instrucción de interpolación lineal que tiene un resultado redondeado hacia un número predeterminado.

40 En referencia a la figura 7, una imagen 36 se compone de una gran cantidad de píxeles correspondientes a posiciones de píxel (X, Y), donde X define la posición de fila del píxel (A a F) e Y define la posición de columna (1 a 5). Sin embargo, en la imagen sólo se determina que ciertos píxeles son píxeles de interés (31 a 34). En la imagen 36, los píxeles de interés (31 a 34) están esparcidos, no son adyacentes. Durante la recopilación de píxeles, se forma la matriz 35 de píxeles, que sólo contiene píxeles de interés (31 a 34), y cualquiera de los píxeles vecinos relevantes. Debido a que la matriz

35 de píxeles sólo contiene una pequeña fracción de los píxeles contenidos en la imagen 36, el procesamiento basado en SIMD puede conseguirse de manera eficaz.

A continuación, en referencia a la figura 8, se describe la dispersión de píxeles. Durante la dispersión de píxeles, los píxeles procesados (41 a 44) contenidos en la matriz 45 de píxeles se sitúan de nuevo en la imagen 36 en sus respectivas ubicaciones originales (31 a 34). Por tanto, la dispersión es el proceso inverso de recopilación, y usa la misma máscara 23 binaria.

En otra realización a modo de ejemplo de la descripción, se prefiere tener una unidad DMA para su uso en las fases de recopilación y dispersión. La unidad DMA permite el acceso de memoria directo, evitando el procesador 1 de datos y descarga algunas de las demandas de procesamiento del procesador de datos.

La figura 9 es un diagrama de flujo que resume las etapas del método de análisis de vídeo según una realización a modo de ejemplo de la presente descripción. En la etapa S1 la máscara binaria que define las ubicaciones de los píxeles de interés se aplica a la imagen. Los píxeles de interés identificados se recopilan a continuación en la etapa S2, y se almacenan adyacentes entre sí en una matriz de píxeles durante la etapa S3. A continuación se realiza el procesamiento de análisis de vídeo en la matriz de píxeles en la etapa S4. El procesamiento de análisis de vídeo puede incluir procesar la imagen usando instrucciones SIMD. En la etapa S5, los píxeles procesados se dispersan de vuelta en la imagen usando la máscara binaria para definir sus respectivas ubicaciones.

En otra realización a modo de ejemplo se usa el método descrito anteriormente para implementar un filtro de  $N \times N$ . El núcleo de filtro de  $N \times N$  se aplica a la matriz 35 de píxeles después de que los píxeles de interés se hayan recopilado en la fase de recopilación. Esto mejora la eficacia de filtrado porque el filtro funciona sólo en los píxeles de interés, y sus respectivos vecinos, recopilados previamente en la matriz de píxeles, en lugar de en toda la imagen. Una vez que se completa el filtrado los píxeles se dispersan tal como se describió anteriormente.

La eficacia de filtrado puede mejorarse adicionalmente implementando el filtro usando técnicas que evitan la manera convencional de realizar convoluciones entre el núcleo de filtro y los datos de píxel. El resultado intermedio de una convolución convencional contiene el doble del número de bits de los argumentos para la función de convolución. Por ejemplo, una convolución de 8 bits expande los datos a 16 bits en una etapa intermedia. El uso de operaciones SIMD tales como instrucciones de promediado, instrucciones de empalme e instrucciones de interpolación lineal para implementar un filtro de análisis de vídeo elimina la necesidad de realizar convoluciones y mantiene el número de bits por píxel constante durante todo el proceso de filtrado.

A continuación, se describe una cámara de red que emplea el sistema de análisis de vídeo descrito anteriormente y la metodología asociada en referencia a la figura 9. Una cámara 70 de red (o una cámara IP) puede ser cualquier cámara digital que pueda generar secuencias de imagen y comunicar las secuencias de imagen o datos de imagen desde las secuencias de imagen, tales como metadatos de imagen, por una red hasta un cliente usuario.

La figura 10 es una cámara 70 de red con análisis de vídeo integrado. La cámara 70 de red incluye un sistema 71 de lente para recoger la luz incidente, un sensor 72 de imagen (sensor) para registrar la luz incidente, un procesador 73 de imagen, una unidad 76 central de procesamiento (CPU) para gestionar la funcionalidad de cámara así como el procesamiento 76a de análisis de vídeo, un codificador 75 de imagen/vídeo, una memoria 74 y una interfaz 77 de red.

En la cámara 70 de red, el sensor 72 puede ser, por ejemplo un dispositivo acoplado por carga (CCD), un sensor CMOS, o similar. El codificador 75 de imagen/vídeo codifica vídeo en cualquier formato de codificación de vídeo/imagen conocido tal como MPEG 1, MPEG2, MPEG4, H.264, JPEG, M-JPEG, mapa de bits, etc. La memoria 74 almacena información de imagen, instrucciones de programa, etc., y la interfaz 77 de red se usa para transferir información de imagen a través de una red, tal como una red IP, y también para recibir información desde la red.

Además, el procesador 73 de imagen, el análisis 76a de vídeo, y el codificador 75 de imagen/vídeo pueden implementarse como instrucciones de software ejecutadas mediante la CPU 76 o pueden implementarse en hardware tal como microprocesadores individuales, FPGA, ASIC y similares. La CPU 76 puede implementarse también como un microprocesador con o sin capacidad de instrucción SIMD, FPGA, ASIC y similares. La memoria puede ser RAM estática o dinámica, ROM, EPROM, EEPROM, FLASH y similares, y el sistema 71 de lente puede unirse permanentemente a la cámara 70 de red o puede ser desmontable.

Mientras que cámara 70 de red descrita anteriormente se describe como una única unidad, son posibles otras implementaciones. Un sistema de una cámara analógica, que puede generar secuencias de imagen analógicas, y un convertidor, tal como un codificador de vídeo o un servidor de vídeo, puede ser también una cámara de red. El convertidor en un sistema de este tipo transforma las secuencias de imagen analógicas desde la cámara analógica hasta las secuencias de imagen digitales o datos de imagen, y proporciona las secuencias de imagen digitales o datos de imagen a la red. El convertidor realiza también la codificación de vídeo/imagen y el análisis de vídeo local.

Se conocen otros componentes de la cámara 70 de red, tales como motores eléctricos usados para centrar el sistema 71 de lente, medición de luz, suministro de potencia, etc., y por tanto se omiten por motivos de brevedad. Además, son

posibles otras implementaciones sin apartarse del alcance de esta descripción. Por ejemplo, la cámara 70 de red puede ser una cámara infrarroja o la interfaz de red puede ser una interfaz inalámbrica, y similar.

5 Mediante la realización del análisis de vídeo localmente, en la cámara de red, pueden situarse sólo datos de vídeo relevantes en la red, lo que es significativamente menor que la cantidad de datos de vídeo situados en la red por las cámaras de red convencionales.

La figura 11 es un sistema de cámara de red que emplea cámaras de red tal como se describieron anteriormente. Esta red a modo de ejemplo incluye múltiples cámaras (81 - 85) de red conectadas a una red 92 privada o una red 93 pública. La cámara 81 de red incluye una cámara 81a analógica y un convertidor 81b, mientras que las cámaras (82-85) de red integran completamente todas las funcionalidades, incluyendo el análisis de vídeo.

10 La red 92 privada puede ser una red de área local (LAN) o una red de área amplia (WAN) a la que se restringe el acceso. Además, la red 92 privada puede conectarse a una red 93 pública a través de un dispositivo 90 de limitación de acceso que puede proporcionar una función de cortafuegos, traducción de dirección de red, etc. El dispositivo 90 de limitación de acceso puede ser un encaminador, ordenador, servidor dedicado, etc. La red 93 pública, por ejemplo Internet, puede contener segmentos inalámbricos y cableados tales como Wifi, redes celulares, PSTN, ISDN, etc. La red 92 privada puede incluir también segmentos inalámbricos y cableados tales como Wifi, Ethernet, etc.

15 Los ordenadores (87, 88, 91) cliente, un centro 89 de vigilancia, y un grabador 86 de vídeo de red se conectan a las cámaras (81-85) de red por medio de las redes (92 ó 93 respectivamente) privadas o públicas. En el sistema 80 de cámara de red, las interfaces de red de las cámaras (81-85) de red permiten la comunicación entre las cámaras (81-85) de red y los ordenadores (87, 88, 91) cliente. Tal comunicación incluye la transferencia de datos de imagen, datos de audio, información de control (panorámica, panorámica vertical, *zoom*, etc.), configuraciones de vídeo, y similares. Además, las cámaras (81-85) de red pueden dar servicio a más de un cliente, por ejemplo un centro 89 de vigilancia, ordenadores (87, 88, 91) cliente o un servidor para almacenar imágenes y retransmitirlas a otros clientes (no mostrado).

Además, la cámara de red descrita anteriormente puede usarse en un sistema tal como se comenta en la figura 3 del documento US 2010/0080477.

25 Las aplicaciones de análisis de vídeo, tales como la detección de movimiento, están basadas en el análisis de imágenes de las secuencias de imagen capturadas por las cámaras (81-85) de red. Como reconoce el presente inventor, cuando se realiza el análisis de las imágenes en las propias cámaras de red, puede reducirse el tráfico de datos en la red 93 pública o la red 92 privada. Sólo los datos relativos a eventos relevantes pueden transmitirse al centro 89 de vigilancia o los ordenadores (87, 88, 91) cliente, conservando el ancho de banda para otras aplicaciones, reduciendo la cantidad de espacio de almacenamiento requerido para almacenar información de vídeo, y reduciendo la cantidad de datos de vídeo que van a buscarse posteriormente.

Aunque en la descripción anterior se ha comentado la detección de movimiento, el sistema y la metodología asociada descritos en el presente documento también es adecuada para otras formas de análisis de vídeo, tales como reconocimiento facial, reconocimiento de objetos, separación de primer plano/segundo plano, análisis de patrones, etc.

35

**REIVINDICACIONES**

1. Método de procesamiento de análisis de vídeo que comprende:
 

recopilar, en un medio de almacenamiento legible por ordenador, una pluralidad de píxeles de interés de una imagen (S2) según una máscara binaria predeterminada, definiendo la máscara binaria predeterminada ubicaciones de la pluralidad de píxeles de interés en la imagen;

5 disponer la pluralidad de píxeles de interés en una matriz de píxeles que incluye al menos los píxeles de interés, estando almacenado cada píxel de interés en la matriz de píxeles adyacente a otros píxeles de interés (S3);

realizar el procesamiento de análisis de vídeo en la matriz de píxeles, usando instrucciones SIMD en un procesador de datos (S4); y

10 dispersar los píxeles de interés en la imagen después del procesamiento según la máscara binaria predeterminada (S5), en el que

la etapa de recopilación y la etapa de dispersión se realizan independientemente de la etapa de realización del procesamiento de análisis de vídeo.
2. Método de procesamiento de análisis de vídeo según la reivindicación 1, en el que la etapa de recopilación incluye para cada píxel de interés,
 

15 dilatar la máscara binaria predeterminada al menos una posición de píxel en una dirección horizontal,

recopilar al menos un píxel vecino a partir de una fila de píxeles por encima del píxel de interés según la máscara binaria dilatada (S61-S63),

20 recopilar, según la máscara binaria dilatada, al menos un píxel vecino y el píxel de interés a partir de una fila de píxeles del píxel de interés (S64-S66), y

recopilar al menos un píxel vecino a partir de una fila de píxeles por debajo de los píxeles de interés según la máscara binaria dilatada (S67-S69).
3. Método de procesamiento de análisis de vídeo según la reivindicación 1, en el que la etapa de recopilación y la etapa de dispersión incluyen además usar una unidad DMA para recopilar y dispersar respectivamente la pluralidad de píxeles de interés y los píxeles vecinos independientemente del procesador de datos.
 

25
4. Método de procesamiento de análisis de vídeo según cualquiera de las reivindicaciones 1 a 3, en el que la realización del procesamiento de análisis de vídeo incluye filtrar los píxeles de interés con un núcleo de filtro de N x N usando un conjunto de instrucciones SIMD, incluyendo el conjunto de instrucciones SIMD una instrucción de promediado, una instrucción de resta y división por dos con resultado sesgado y una instrucción de empalme.
 

30
5. Método de procesamiento de análisis de vídeo según la reivindicación 4, en el que un número predeterminado de bits que representan cada píxel se mantiene constante durante toda la operación de filtrado.
 

35
6. Programa informático que comprende instrucciones legibles por ordenador, instrucciones que cuando se ejecutan por un procesador hacen que el procesador realice el método de procesamiento de análisis de vídeo que comprende:
 

recopilar, en un medio de almacenamiento legible por ordenador, una pluralidad de píxeles de interés de una imagen (S2) según una máscara binaria predeterminada, definiendo la máscara binaria predeterminada ubicaciones de la pluralidad de píxeles de interés en la imagen;

disponer la pluralidad de píxeles de interés en una matriz de píxeles que incluye al menos los píxeles de interés (S3);

realizar el procesamiento de análisis de vídeo en la matriz de píxeles usando un procesador SIMD (S4); y

40 dispersar los píxeles en la imagen después del procesamiento según la máscara binaria predeterminada (S5), en el que

la etapa de recopilación y la etapa de dispersión se realizan independientemente de la etapa de realización del procesamiento de análisis de vídeo.
7. Programa informático según la reivindicación 6, en el que la etapa de recopilación incluye para cada píxel de interés,
 

45 dilatar la máscara binaria predeterminada al menos una posición de píxel en una dirección horizontal,



recopilar al menos un píxel vecino a partir de una fila de píxeles por encima del píxel de interés según la máscara binaria dilatada (S61-S63),

recopilar, según la máscara binaria dilatada, al menos un píxel vecino y el píxel de interés a partir de una fila de píxeles del píxel de interés (S64-S66), y

5 recopilar al menos un píxel vecino a partir de una fila de píxeles por debajo de los píxeles de interés según la máscara binaria dilatada (S67-S69).

8. Programa informático según la reivindicación 6 ó 7, en el que la etapa de recopilación y la etapa de dispersión incluye además usar una unidad DMA para recopilar y dispersar respectivamente la pluralidad de píxeles de interés y los píxeles vecinos independientemente del procesador de datos.

10 9. Programa informático según cualquiera de las reivindicaciones 6 a 8, en el que la realización del procesamiento de análisis de vídeo incluye filtrar los píxeles de interés con un núcleo de filtro de  $N \times N$  usando un conjunto de instrucciones SIMD, incluyendo el conjunto de instrucciones SIMD una instrucción de promediado, una instrucción de resta y división por dos con resultado sesgado y una instrucción de empalme.

15 10. Programa informático según la reivindicación 9, en el que un número predeterminado de bits que representan cada píxel se mantiene constante durante toda la operación de filtrado.

11. Cámara (70) de red que comprende:

al menos una lente (71) configurada para recoger la luz incidente;

un sensor (72) dispuesto detrás de la lente y configurado para convertir la luz incidente en datos de imagen;

20 una memoria (74) electrónica configurada para almacenar, como datos de imagen, los datos eléctricos correspondientes a la imagen;

un codificador (75) de vídeo configurado para codificar los datos de imagen;

un procesador (76) de datos que incluye

25 una unidad de recopilación configurada para recopilar, en la memoria electrónica, una pluralidad de píxeles de interés de una imagen según una máscara binaria predeterminada, definiendo la máscara binaria predeterminada ubicaciones de la pluralidad de píxeles de interés en la imagen,

una unidad de disposición configurada para disponer la pluralidad de píxeles de interés en una matriz de píxeles que incluye al menos los píxeles de interés,

una unidad de análisis de vídeo configurada para realizar el procesamiento de análisis de vídeo en la matriz de píxeles independientemente de la unidad de recopilación,

30 una unidad de dispersión configurada para dispersar los píxeles de interés en la imagen, independientemente de la unidad de recopilación y la unidad de análisis de vídeo; y

una interfaz de red configurada para su conexión a una red de datos.

12. Cámara de red según la reivindicación 11, en la que la unidad de recopilación está configurada además para, dilatar la máscara binaria predeterminada al menos una posición de píxel en una dirección horizontal,

35 recopilar al menos un píxel vecino a partir de una fila de píxeles por encima del píxel de interés según la máscara binaria dilatada,

recopilar, según la máscara binaria dilatada, al menos un píxel vecino y el píxel de interés a partir de una fila de píxeles del píxel de interés, y

40 recopilar al menos un píxel vecino a partir de una fila de píxeles por debajo de los píxeles de interés según la máscara binaria dilatada.

13. Cámara de red según la reivindicación 11 ó 12, que comprende además:

una unidad DMA configurada para transportar píxeles independientemente del procesador de datos, empleando la unidad de recopilación de píxeles y la unidad de dispersión de píxeles la unidad DMA cuando recopila y dispersa respectivamente píxeles de interés.

45 14. Cámara de red según cualquiera de las reivindicaciones 11 a 13, en la que el procesador de datos está configurado además para realizar el filtrado de los píxeles de interés con un núcleo de filtro de  $N \times N$  usando instrucciones SIMD, incluyendo las instrucciones SIMD una instrucción de promediado, una instrucción de resta y división por dos con resultado sesgado, y una instrucción de empalme.

15. Cámara de red según la reivindicación 14, en la que un número predeterminado de bits que representan cada píxel se mantiene constante durante toda la operación de filtrado.

Figura 1

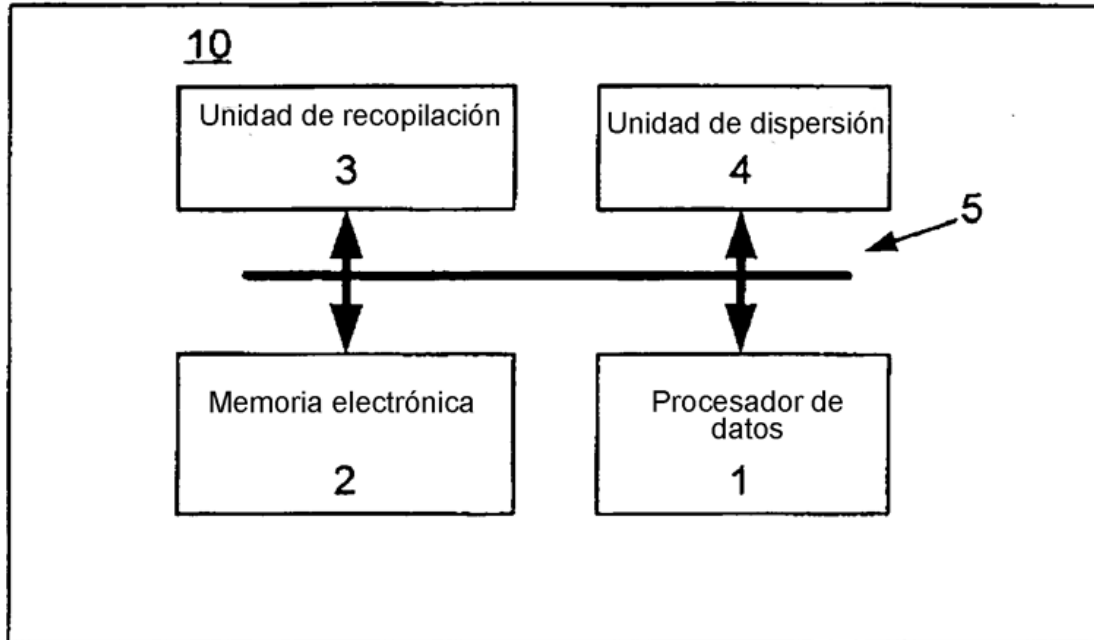


Figura 2

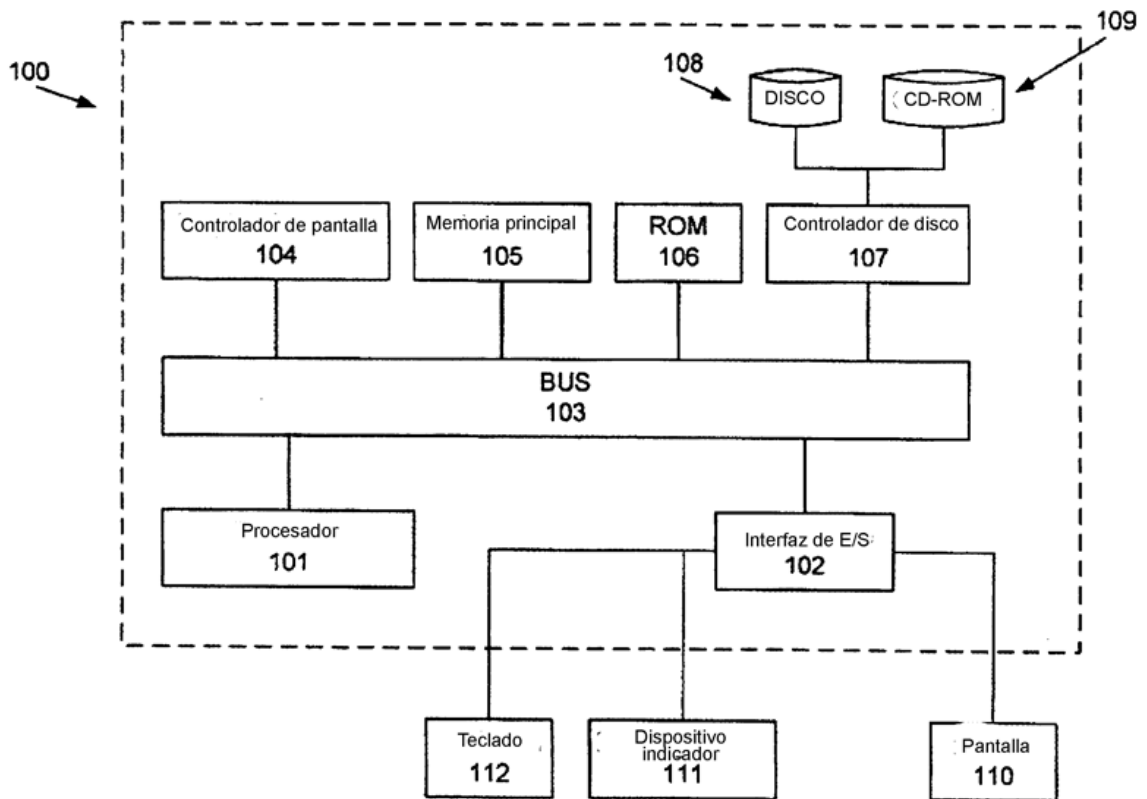


Figura 3

20

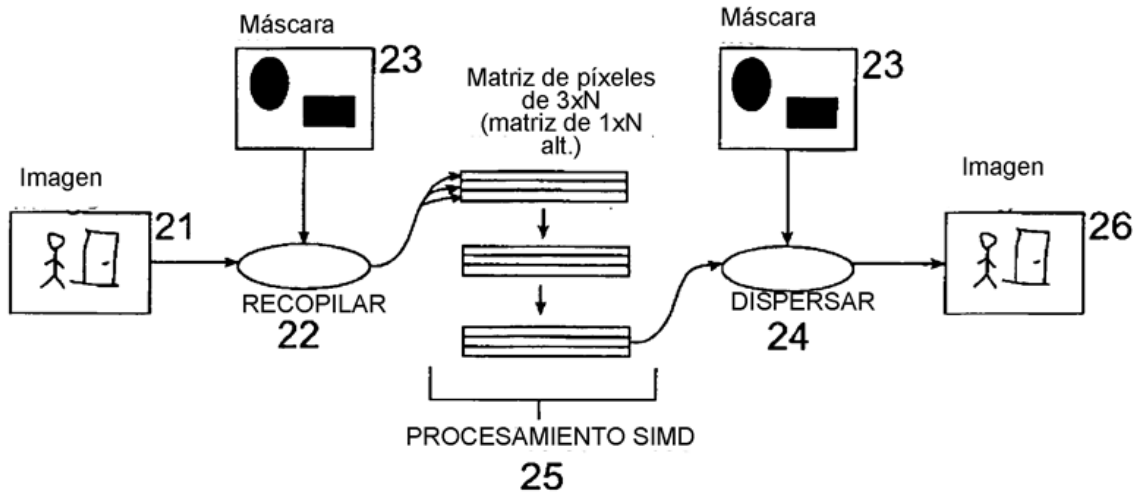
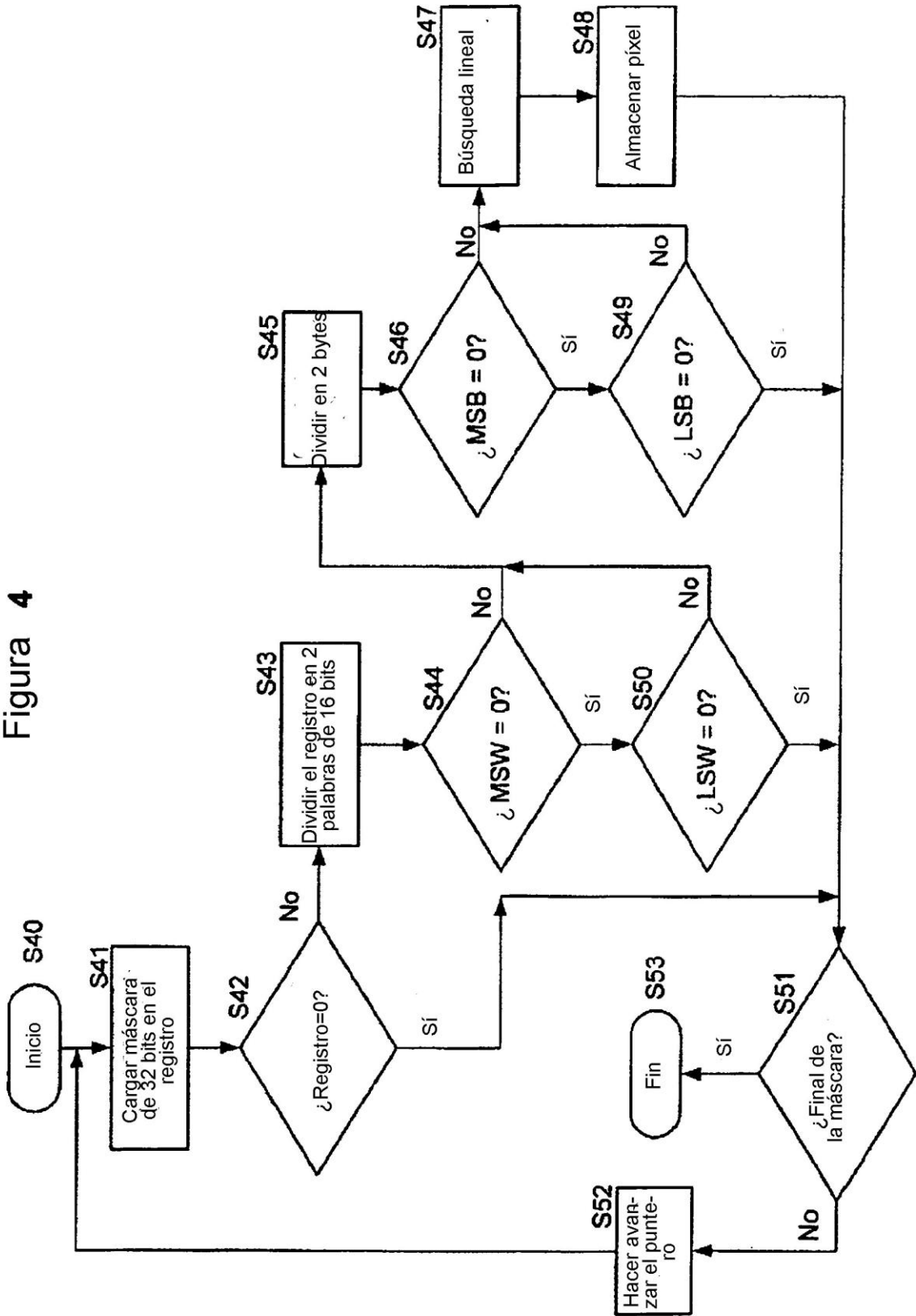
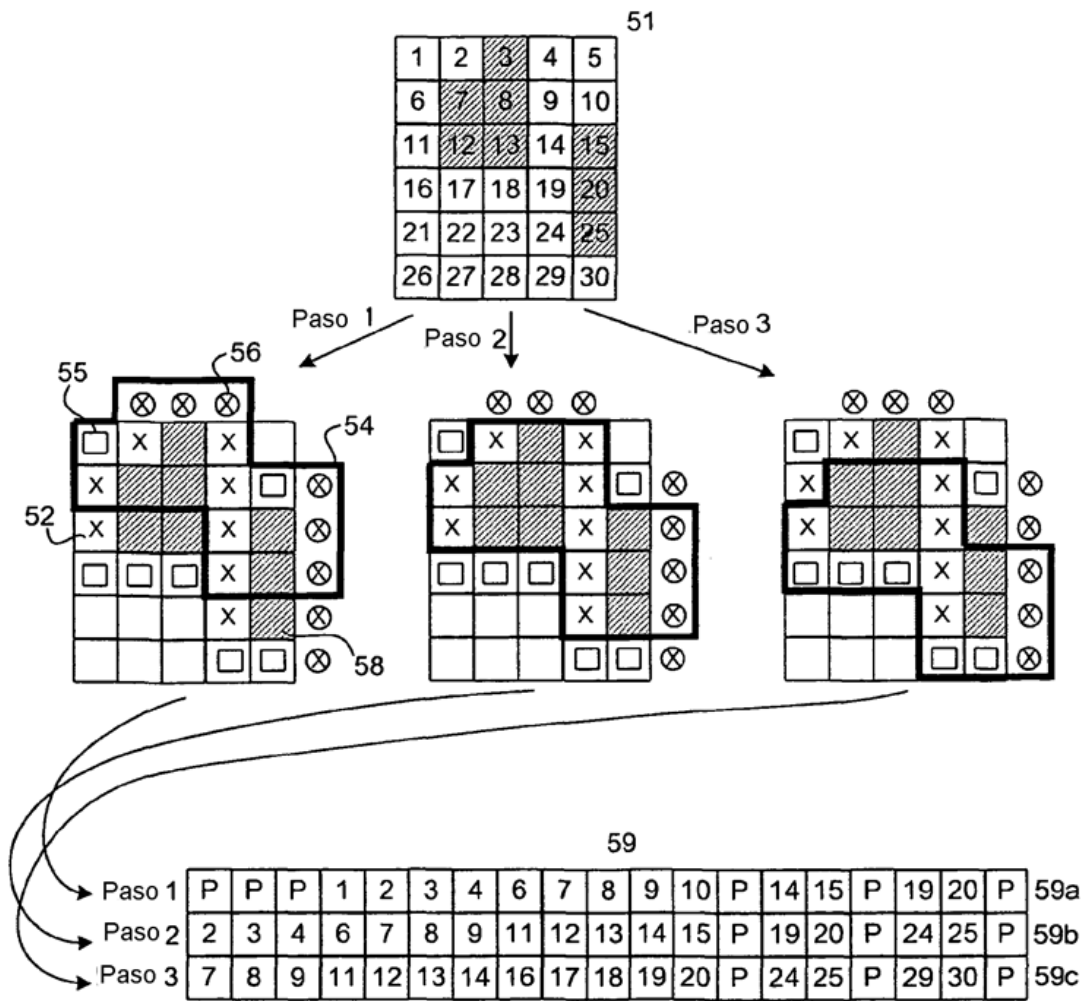


Figura 4





Figuras 5

Figura 6

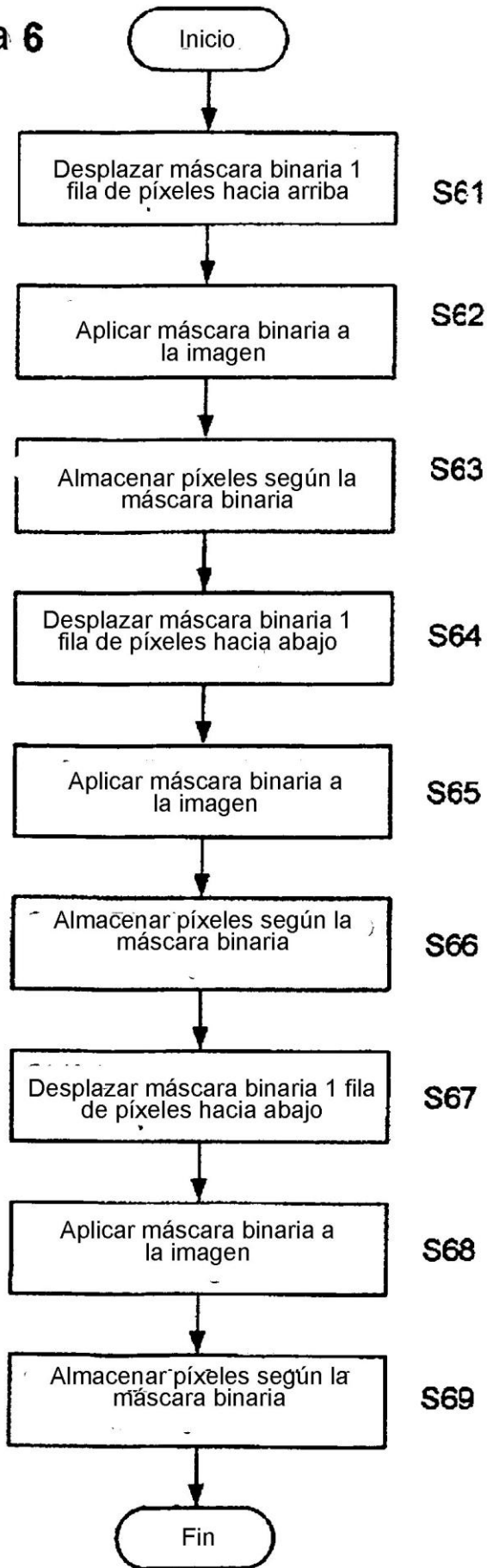




Figura 7

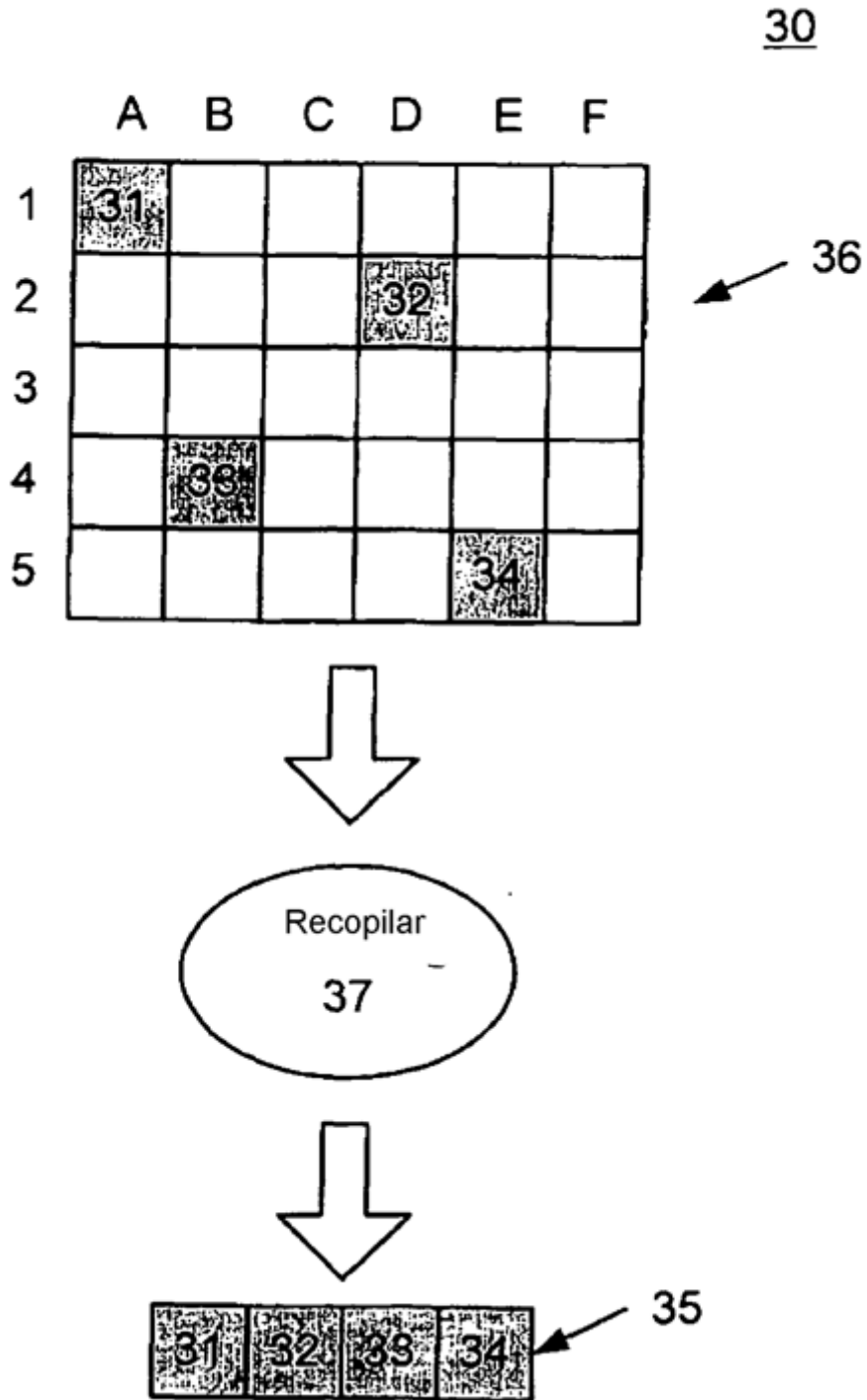
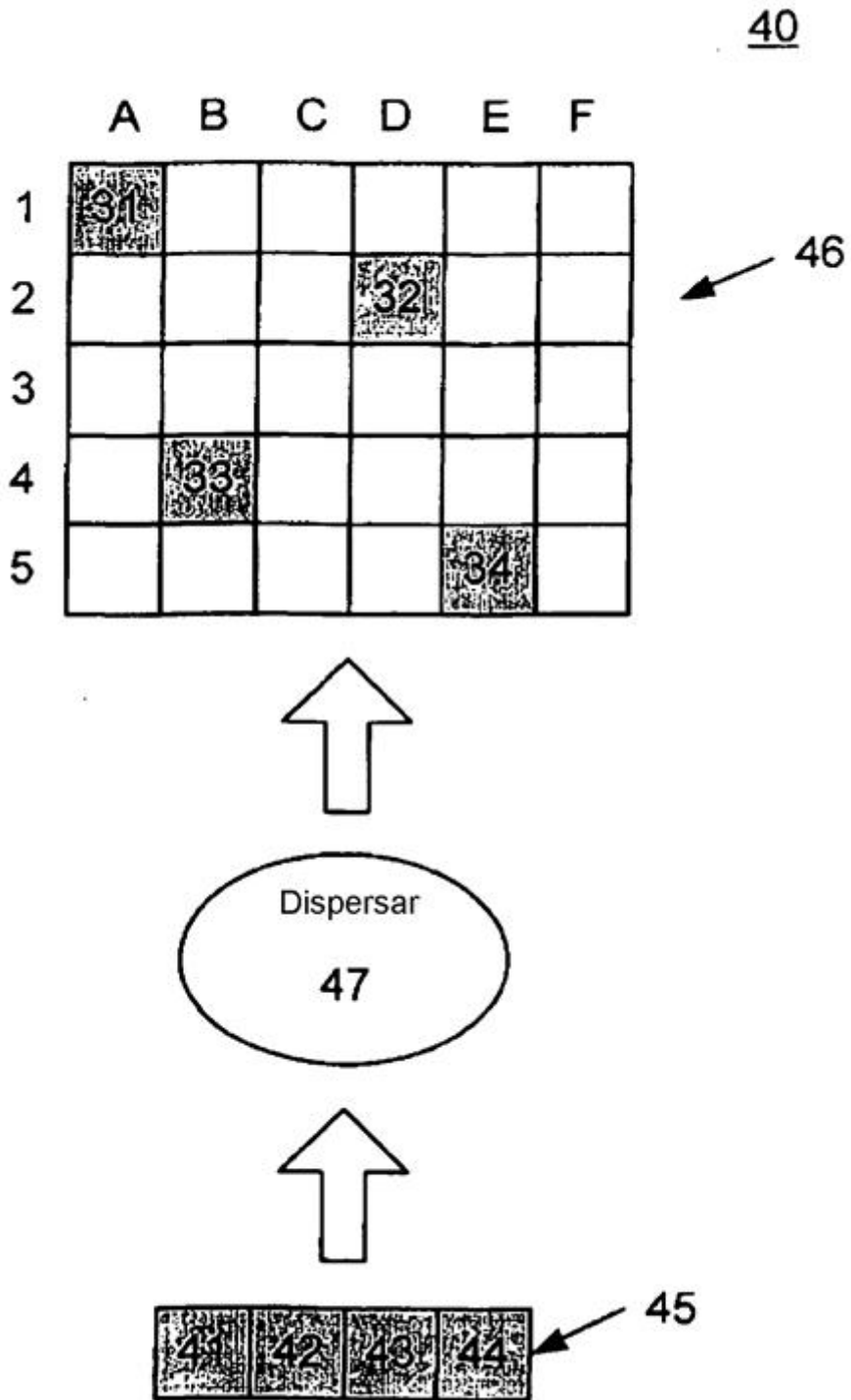


Figura 8



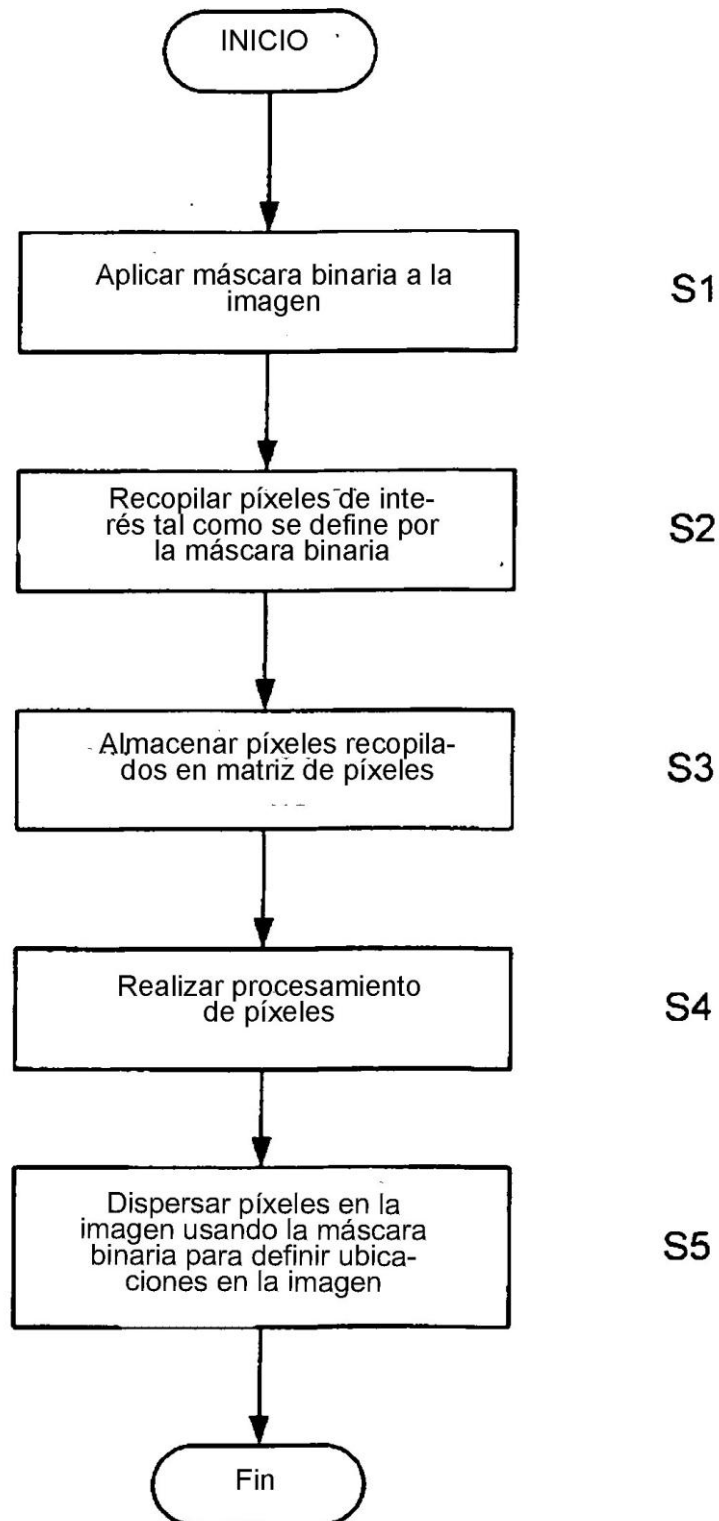
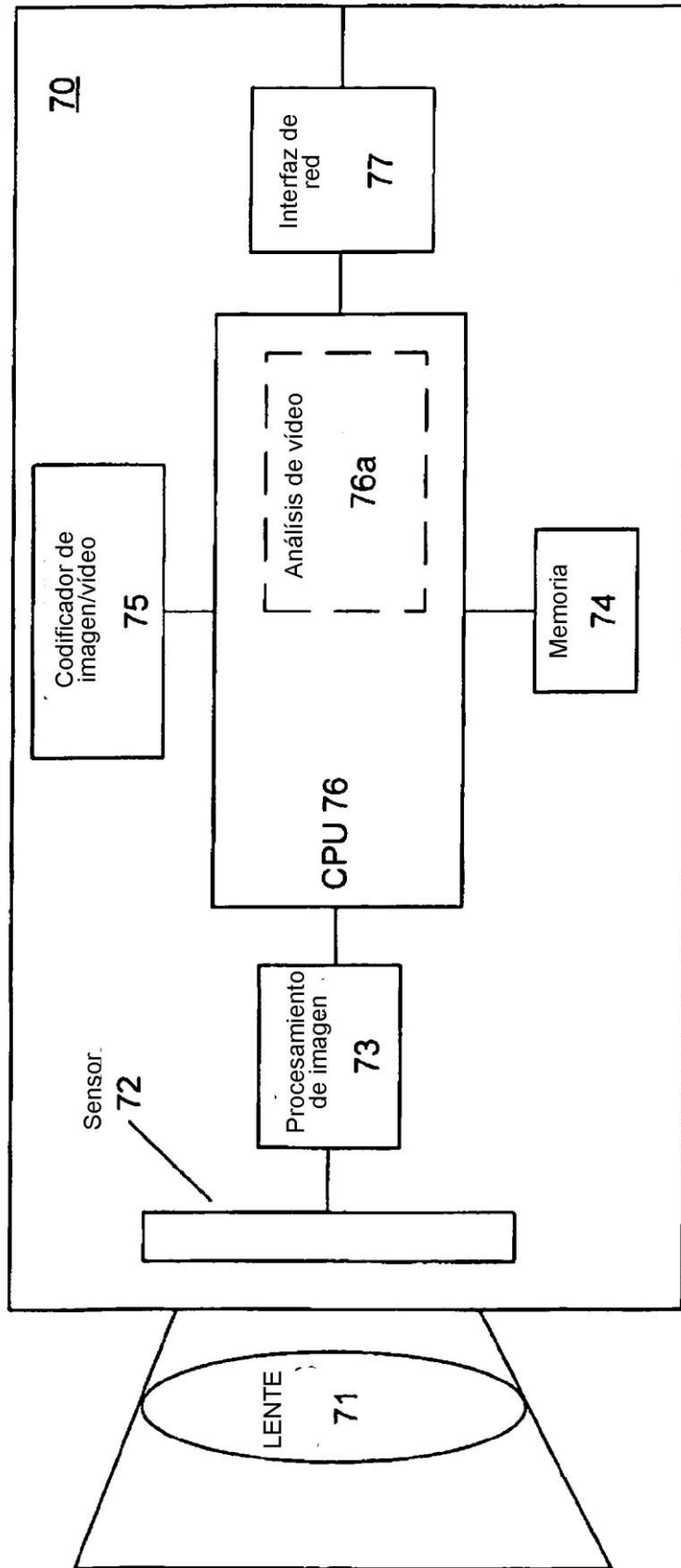


Figura 9

Figura 10



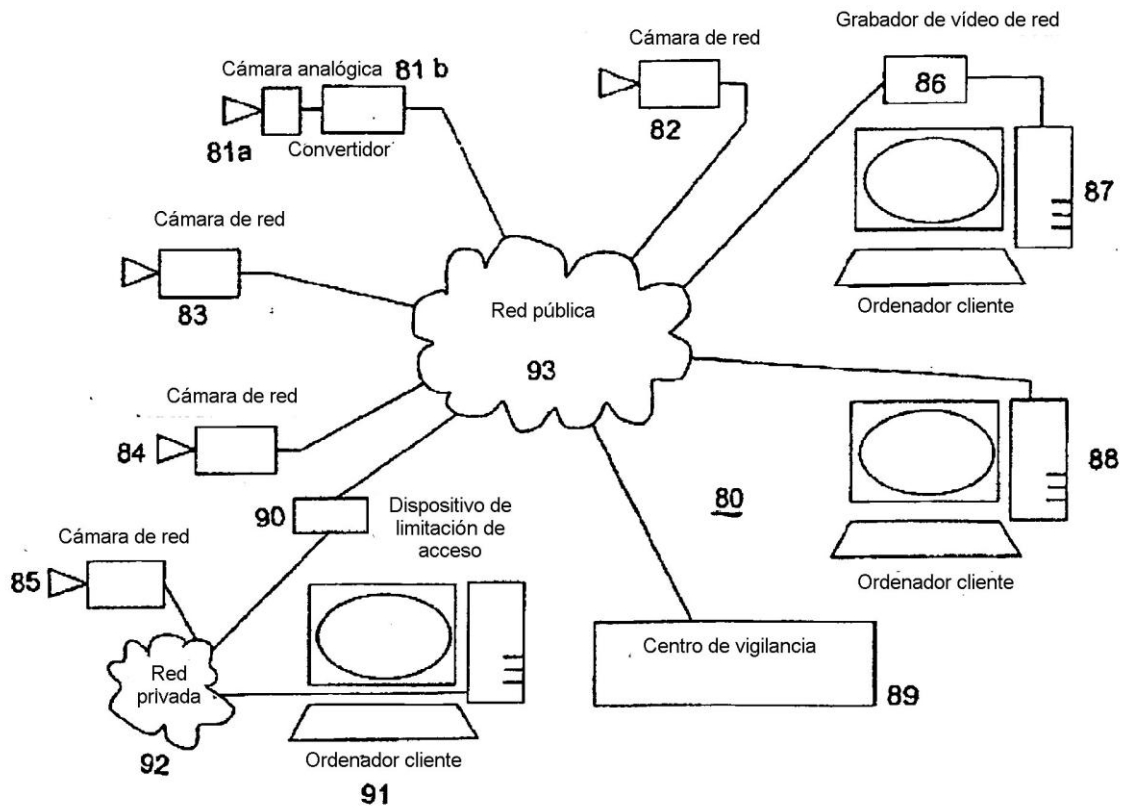


Figura 11