

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 945 657**

51 Int. Cl.:

H04N 7/01 (2006.01)
G06V 20/40 (2012.01)
G06T 7/11 (2007.01)
G06V 20/70 (2012.01)
H04N 19/169 (2014.01)
G06V 10/26 (2012.01)
H04N 19/16 (2014.01)
G09G 5/00 (2006.01)
H04N 1/60 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **20.04.2020** **PCT/US2020/029023**
- 87 Fecha y número de publicación internacional: **29.10.2020** **WO20219401**
- 96 Fecha de presentación y número de la solicitud europea: **20.04.2020** **E 20723760 (3)**
- 97 Fecha y número de publicación de la concesión europea: **19.04.2023** **EP 3959646**

54 Título: **Analizador de rango de PQ con consciencia de contenido, y mapeo de tono en alimentaciones en vivo**

30 Prioridad:

25.04.2019 EP 19171057
25.04.2019 US 201962838518 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.07.2023

73 Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (100.0%)
1275 Market Street
San Francisco, CA 94103, US

72 Inventor/es:

ZANDIFAR, ALI;
CRENSHAW, JAMES E. y
VASCO, CRISTINA MICHEL

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 945 657 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

Analizador de rango de PQ con consciencia de contenido, y mapeo de tono en alimentaciones en vivo

5 Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica la prioridad de la solicitud de patente provisional de EE. UU. con número de serie 62/838,518, presentada el 25 de abril de 2019 y la solicitud de patente EP con número de serie 19171057.3, presentada el 25 de abril de 2019.

10 Campo técnico

Esta solicitud se relaciona generalmente con imágenes. Más específicamente, esta solicitud se relaciona con el conocimiento del contenido en el análisis de rango de cuantificador perceptual (PQ) de video HDR y mapeo de tonos para alimentaciones en vivo.

Antecedentes

Como se usa en el presente documento, el término "rango dinámico" puede relacionarse con la capacidad del sistema visual humano para percibir un rango de intensidad (por ejemplo, luminancia, luma y similares) en una imagen; por ejemplo, el rango de los negros más oscuros ("oscuros") a los blancos más brillantes ("claros"). En este sentido, el rango dinámico se relaciona con una intensidad "referida a la escena". El rango dinámico también puede relacionarse con la capacidad de un dispositivo de visualización para renderizar adecuada o apropiadamente un rango de intensidad de una amplitud particular. En este sentido, el rango dinámico se refiere a una intensidad "referida al visualizador". A menos que se especifique explícitamente que un sentido particular tiene un significado particular en cualquier punto de la descripción del presente documento, se debe inferir que el término puede usarse en cualquier sentido, por ejemplo, indistintamente.

Como se usa en el presente documento, el término "alto rango dinámico" (HDR) se refiere a una amplitud de rango dinámico que abarca los aproximadamente 14-15 órdenes de magnitud del sistema visual humano. En la práctica, el rango dinámico sobre el cual un ser humano puede percibir simultáneamente una gran amplitud en el rango de intensidad puede verse comparativamente truncado, en relación con HDR. Como se usa en el presente documento, los términos "rango dinámico extendido" (EDR) o "rango dinámico visual" (VDR) pueden relacionarse individualmente o de manera intercambiable con el rango dinámico que es perceptible simultáneamente por un sistema visual humano. Como se usa en el presente documento, EDR puede relacionarse con un rango dinámico que abarca de cinco a seis órdenes de magnitud. Por lo tanto, mientras que el EDR puede ser algo más estrecho en relación con el HDR referido a la escena real, el EDR, no obstante, representa una amplia amplitud de rango dinámico y también puede denominarse HDR.

En la práctica, las imágenes comprenden uno o más componentes de color (por ejemplo, luma Y y croma Cb y Cr) en donde cada componente de color está representado por una precisión de n bits por píxel (por ejemplo, n = 8). Usando la codificación de luminancia lineal, las imágenes en las que n < 8 (por ejemplo, imágenes JPEG en color de 24 bits) se consideran imágenes de rango dinámico estándar, mientras que las imágenes en las que n > 8 pueden considerarse imágenes de rango dinámico mejorado. Las imágenes EDR y HDR también se pueden almacenar y distribuir usando formatos de punto flotante de alta precisión (por ejemplo, 16 bits), como el formato de archivo OpenEXR desarrollado por Industrial Light and Magic.

La mayoría de los visualizadores de escritorio para consumidores admiten una luminancia de 200 a 300 cd/m² ("nits"). La mayoría de los televisores de alta definición ("HDTV") de consumo varían de 300 a 1000 nits. Tales visualizadores tipifican así un rango dinámico bajo (LDR), también denominado rango dinámico estándar (SDR), en relación con HDR o EDR. A medida que aumenta la disponibilidad de contenido EDR debido a los avances tanto en los equipos de captura (p. ej., cámaras) como en los visualizadores EDR (p. ej., el monitor de referencia profesional PRM-4200 de Dolby Laboratories), el contenido EDR puede clasificarse por colores y visualizarse en visualizadores EDR que admitan rangos dinámicos mayores (por ejemplo, de 1000 nits a 5000 nits o más).

Como se usa en el presente documento, el término "gestión de visualización" incluye, entre otros, el procesamiento (p. ej., mapeo de tonos y gama) necesario para mapear una señal de vídeo de entrada de un primer rango dinámico (p. ej., 1000 nits) a una visualización de un segundo rango dinámico (por ejemplo, 500 nits).

Los enfoques descritos en esta sección son enfoques que podrían aplicarse, pero no necesariamente enfoques que hayan sido previamente concebidos o perseguidos. Por lo tanto, a menos que se indique lo contrario, no se debe suponer que cualquiera de los enfoques descritos en esta sección se califica como técnica anterior simplemente en virtud de su inclusión en esta sección. Del mismo modo, no se debe suponer que los problemas identificados con respecto a uno o más enfoques han sido reconocidos en ningún técnica anterior sobre la base de esta sección, a menos que se indique lo contrario.

El documento US 2017/161881 A1 se refiere a un equipo de usuario (UE) que incluye un receptor, al menos un sensor y un procesador. El receptor está configurado para recibir un flujo de bits que incluye al menos una imagen codificada y metadatos. El sensor está configurado para determinar la información del punto de vista de un usuario. El procesador está configurado para renderizar al menos dicha imagen codificada basándose en los metadatos y el punto de vista.

El documento US 2014/321746 A9 se relaciona con sistemas y métodos para clasificar videos basándose en el contenido del video. Para un archivo de video dado que incluye una pluralidad de fotogramas, se extrae un subconjunto de fotogramas para su procesamiento. Los fotogramas que son demasiado oscuros, borrosos o candidatos de clasificación deficientes se descartan del subconjunto. Generalmente, los puntuaciones de clasificación de materiales que describen el tipo de contenido de material probablemente incluido en cada fotograma se calculan para los fotogramas restantes del subconjunto. Las puntuaciones de clasificación de materiales se usan para generar vectores de disposición de materiales que representan la disposición espacial del contenido del material en cada fotograma. Los vectores de disposición de materiales se clasifican posteriormente para generar un vector de puntuación de clasificación de escenas para cada fotograma. Los resultados de la clasificación de escenas se promedian (o se procesan de otro modo) en todos los fotogramas del subconjunto para asociar el archivo de video con una o más categorías de escena predefinidas relacionadas con los tipos generales de contenido de escena del archivo de video.

El documento PRATIKAKIS IOANNIS ET AL, "Extracción semántica de imágenes", CONFERENCIA INTERNACIONAL SOBRE CRIPTOGRAFÍA FINANCIERA Y SEGURIDAD DE DATOS; [NOTAS DE CONFERENCIA EN INFORMÁTICA; LECT.NOTES COMPUTER], SPRINGER, BERLÍN, HEIDELBERG, PÁGINA(S) 50 - 88, (20111231), vol. 6050 presenta una descripción general del estado de la técnica en la extracción semántica de imágenes.

El documento Ultra Hd Forum, "Directrices Foro Ultra HD Fase B", <https://ultrahdforum.org/resources/phaseb-guidelines-description/>, (20180406) y documento Etsi, "Señalización HDR y transporte de metadatos dinámicos para transformación de volumen de color; Aplicación n.º 1 para ESPECIFICACIÓN TÉCNICA de sistemas compatibles con DVB ", (20180301), páginas 2018 - 3 analiza temas de la próxima generación de tecnologías UHD.

Breve resumen de la divulgación

Varios aspectos de la presente divulgación se relacionan con circuitos, sistemas y métodos para el procesamiento de imágenes, incluida la conciencia de contenido en análisis de rango de PQ y mapeo de tono para alimentaciones en vivo. Las realizaciones de la presente invención están definidas por las reivindicaciones independientes. Las características adicionales de las realizaciones de la invención se presentan en las reivindicaciones dependientes. A continuación, las partes de la descripción y los dibujos que se refieren a realizaciones anteriores que no necesariamente comprenden todas las características para implementar realizaciones de la invención reivindicada no se representan como realizaciones de la invención sino como ejemplos útiles para comprender las realizaciones de la invención. Las realizaciones de la presente solicitud (o la presente divulgación) proporcionan un sistema de procesamiento de imágenes, un método de procesamiento de imágenes y un medio legible por ordenador no transitorio.

De esta manera, varios aspectos de la presente divulgación proporcionan mejoras en al menos los campos técnicos del procesamiento de imágenes, así como los campos técnicos relacionados de captura, codificación y transmisión de imágenes.

Descripción de los dibujos

Estas y otras características más detalladas y específicas de varios aspectos de la presente divulgación se divulgan más completamente en la siguiente descripción, con referencia a los dibujos adjuntos, en los que:

la figura 1 ilustra una escena de origen y varias escenas renderizadas de acuerdo con varios aspectos de la presente divulgación;

la figura 2 ilustra un diagrama de bloques de un flujo de trabajo de transmisión de ejemplo de acuerdo con varios aspectos de la presente divulgación;

la figura 3 ilustra un diagrama de bloques de una unidad de procesamiento de ejemplo de acuerdo con varios aspectos de la presente divulgación;

la figura 4 ilustra un flujo de proceso para un método de procesamiento de ejemplo de acuerdo con varios aspectos de la presente divulgación;

la figura 5 ilustra un flujo de proceso para un método de clasificación de ejemplo de acuerdo con varios aspectos de la presente divulgación;

la figura 6 ilustra una escena de ejemplo de acuerdo con varios aspectos de la presente divulgación;

la figura 7 ilustra otra escena de ejemplo de acuerdo con varios aspectos de la presente divulgación; y

la figura 8 ilustra otra escena de ejemplo de acuerdo con varios aspectos de la presente divulgación.

Descripción detallada

En la siguiente descripción, se exponen numerosos detalles, como configuraciones de circuitos, temporizaciones de formas de onda, operaciones de circuitos y similares, para proporcionar una comprensión de uno o más aspectos de la presente divulgación. Será evidente para un experto en la técnica que estos detalles específicos son meramente de ejemplo y no pretenden limitar el alcance de esta solicitud.

Esta divulgación puede materializarse en varias formas, incluidos hardware o circuitos controlados por métodos implementados por ordenador, productos de programas de ordenador, sistemas y redes de ordenador, interfaces de usuario e interfaces de programación de aplicaciones; así como métodos implementados en hardware, circuitos de procesamiento de señales, matrices de memoria, circuitos integrados de aplicación específica, matrices de puertas programables en campo y similares. El resumen anterior pretende únicamente dar una idea general de varios aspectos de la presente divulgación, y no limita el alcance de la divulgación de ninguna manera.

En el presente documento se describen la captura, el análisis y la codificación de vídeo. En la siguiente descripción, se exponen numerosos detalles, como configuraciones de circuitos, temporizaciones, operaciones de circuitos y similares, para proporcionar una comprensión de uno o más aspectos de la presente divulgación. Será evidente para un experto en la técnica que estos detalles específicos son meramente de ejemplo y no pretenden limitar el alcance de esta solicitud. Por ejemplo, en algunos casos, varios aspectos de la presente divulgación pueden practicarse sin estos detalles. En otros casos, es posible que las estructuras y los dispositivos bien conocidos no se describan con un detalle exhaustivo, para evitar ocluir, oscurecer u ofuscar innecesariamente la presente invención.

Descripción general

Los ejemplos descritos en el presente documento se relacionan con el procesamiento de imágenes, incluida la generación de metadatos durante una transmisión en vivo de una transmisión de vídeo. Algunos ejemplos descritos en el presente documento se pueden utilizar con la arquitectura "Dolby Vision". Dolby Vision para aplicaciones de consumo es un conjunto de tecnología integral que permite la creación y distribución de contenido masterizado con un alto rango dinámico y una amplia gama de colores. La gestión del visualizador de Dolby Vision coincide con la capacidad de un televisor determinado (que puede que solo sea capaz de visualizar imágenes SDR) mediante el uso de una serie de algoritmos para mapear la señal a cualquier televisor de consumo Dolby Vision. Al visualizar contenido HDR en un visualizador SDR, las imágenes HDR se mapean al rango dinámico comparativamente reducido del visualizador.

La figura 1 ilustra un ejemplo de mapeo de una escena de origen a varias escenas renderizadas. Como se ilustra en la figura 1, una imagen HDR 101 representa la escena de origen con oscuros (por ejemplo, áreas en la parte inferior izquierda y superior izquierda de la imagen HDR 101) y claros (por ejemplo, áreas en la parte superior media y superior derecha de la imagen HDR 101). Cuando se mapea la imagen HDR 101 para visualizar fielmente los claros en un visualizador SDR, se puede crear una imagen subexpuesta 102 como la escena renderizada. En la imagen subexpuesta 102, los claros se reproducen con precisión pero el detalle se reduce o se pierde en las áreas correspondientes a los oscuros. Por el contrario, cuando se mapea la imagen HDR 101 para visualizar fielmente los oscuros en el visualizador SDR, se puede crear una imagen sobreexpuesta 103 como la escena renderizada. En la imagen sobreexpuesta 103, los oscuros ahora se reproducen con precisión, pero las áreas correspondientes a los claros pueden aparecer borradas. Para presentar una imagen convertida que no esté subexpuesta ni sobreexpuesta, se pueden utilizar metadatos (es decir, datos relacionados con los datos de la imagen) para determinar qué características de la imagen HDR 101 deben considerarse áreas de enfoque de la imagen.

La figura 2 ilustra un ejemplo de un sistema 200 de flujo de trabajo de difusión, que incluye captura, producción y posproducción de vídeo, y distribución en directo. La captura de vídeo puede lograrse mediante uno o más bancos 210 de cámaras, cada uno de los cuales incluye una o más cámaras 211. Los bancos 210 de cámaras individuales pueden ubicarse en diferentes ubicaciones físicas para capturar diferentes contenidos de vídeo. Por ejemplo, si el sistema 200 de flujo de trabajo de transmisión se usa para una transmisión deportiva en vivo, un primer banco 210 de cámaras puede colocarse para capturar vídeo del evento deportivo en sí, un segundo 210 banco de cámaras puede colocarse para capturar vídeo de una cabina de transmisión, se puede colocar un tercer banco 210 de cámaras para capturar vídeo de analistas en un estudio y similares. Cada banco 210 de cámaras puede incluir cualquier número de cámaras 211. Las cámaras individuales 211 pueden ser capaces de capturar datos de vídeo HDR o datos de vídeo SDR. Los datos de vídeo capturados por una cámara dada 211 pasan a través de un enlace

212 de contribución correspondiente para su posterior procesamiento.

Como se ilustra en la figura 2, los datos de video que pasan a través de un enlace 212 de contribución se reciben en un convertidor 220 de entrada correspondiente. Cuando los datos de video son datos de video HDR, el convertidor 220 de entrada puede realizar una conversión de HDR a HDR; por ejemplo, la conversión de gamma log híbrida (HLG) o HDR SLog-3 a cuantificador perceptual (PQ) HDR, por ejemplo, como se describe en Rec. UIT-R BT. 2100-1 (06/2017), "Valores de parámetros de imagen para televisión de alto rango dinámico para uso en producción e intercambio internacional de programas".

Cuando los datos de video son datos de video SDR, el convertidor 220 de entrada puede realizar una conversión de SDR a HDR. Mientras que la figura 2 ilustra un convertidor 220 de entrada para cada enlace 212 de contribución, en la práctica pueden estar presentes menos convertidores 220 de entrada. Por ejemplo, cuando los datos de video son datos de video HDR que usan PQ, no puede ocurrir ninguna conversión y, por lo tanto, no puede proporcionarse ningún convertidor 220 de entrada. En cualquier caso, los datos de video se proporcionan a un conmutador 221 de producción.

El conmutador 221 de producción recibe datos de video de cada una de las cámaras 211 y proporciona varias salidas, que incluyen: un flujo 222 de transmisión, que puede corresponder a datos de video recibidos de una de las cámaras 211 seleccionada; una salida a una unidad 223 de control de calidad (QC); una salida a una unidad 224 de mapeo, que a su vez puede proporcionar una salida a una unidad 223 de control de calidad con capacidad SDR; una salida a un servidor 225 de reproducción; y una ingesta 226 de archivos para almacenamiento. Los datos de la ingesta 226 de archivos pueden estar sujetos a procesamiento adicional en una unidad 227 de postproducción y posteriormente proporcionados al servidor 225 de reproducción. Los datos de video almacenados en el servidor 225 de reproducción se pueden utilizar para la reproducción en un momento posterior, como para la reproducción instantánea o el análisis de medio tiempo/intermedio. La salida del servidor 225 de reproducción puede incluir datos de video SDR (en cuyo caso la conversión se puede realizar a través de otro convertidor 220 de entrada), datos de video HDR o ambos.

Para la distribución en vivo, el flujo 222 de transmisión y/o los datos del servidor 225 de reproducción se reciben en un enrutador 230. El enrutador 230 proporciona varias salidas, que incluyen: una o más salidas (HDR y/o SDR) a las unidades 223 de control de calidad; uno o más flujos 231 de distribución HDR, cada uno a un codificador 232 de transmisión respectivo; uno o más flujos 237 de distribución SDR (por ejemplo, una transmisión simultánea SDR); y salida HDR y/o SDR a una unidad 238 de mapeo. El codificador 232 de transmisión respectivo incluye una unidad 233 de procesamiento HDR (HPU) que recibe el flujo 231 de distribución HDR, realiza varios análisis como se describirá con más detalle a continuación y emite una alimentación 234 de video HDR y una alimentación 235 de metadatos. La alimentación 234 de video HDR y la alimentación 235 de metadatos se proporcionan a una unidad 236 de codificación para codificar y transmitir. El flujo 237 de distribución SDR, si está presente, puede enviarse directamente a una unidad 236 de codificación sin la generación de una alimentación 235 de metadatos.

Procesamiento HDR

La figura 3 ilustra un sistema de procesamiento de imágenes de ejemplo de acuerdo con varios aspectos de la presente divulgación. Específicamente, la figura 3 ilustra una HPU 300, que puede ser un ejemplo de la HPU 233 ilustrada en la figura 2. La HPU 300 incluye una unidad 310 de entrada/salida (E/S), una memoria 320, una unidad 330 de comunicación, una interfaz 340 de usuario (IU) y un procesador 350. Varios elementos de la HPU 300 se comunican entre sí a través de un bus 360. La unidad de E/S recibe datos 311 de entrada, que pueden ser un ejemplo del flujo 231 de distribución HDR ilustrado en la figura 2, y emite una alimentación 312 de video y una alimentación 313 de metadatos, que pueden ser ejemplos de la alimentación 234 de video HDR y la alimentación 235 de metadatos ilustrados respectivamente en la figura 2. El procesador 350 incluye una unidad 351 de determinación, una unidad 352 de segmentación y una unidad 353 de extracción, cada una de las cuales se describirá con más detalle a continuación.

Los componentes individuales de la HPU 300 pueden implementarse como hardware, software, firmware o combinaciones de los mismos. Por ejemplo, varias unidades pueden implementarse como circuitos o circuitería, pueden implementarse como módulos de software en la memoria o algoritmos en un procesador, etc., incluidas combinaciones de circuitería y módulos de software.

La unidad 310 de E/S puede incluir uno o más puertos para la entrada o salida de datos a través de un cable, una fibra óptica, un protocolo de comunicación inalámbrica o combinaciones de los mismos. La memoria 320 puede ser una unidad de memoria volátil o una unidad de memoria no volátil, que incluye, entre otros, memoria de solo lectura (ROM) o memoria de acceso aleatorio (RAM), como un disco duro, un almacenamiento flash y similares. La unidad 330 de comunicación puede incluir circuitería para recibir señales de control u otra comunicación desde el exterior a la HPU 300 a través de un cable, una fibra óptica, un protocolo de comunicación inalámbrica o combinaciones de los mismos. La IU 340 puede incluir dispositivos o puertos para recibir instrucciones y/o comunicarse con un usuario local, como un ratón, un teclado, una interfaz de pantalla táctil, un visualizador, una IU gráfica (IUG) y similares.

Varios componentes de la HPU 300, incluidos, entre otros, el procesador 350, pueden implementarse con un sistema de ordenador, sistemas configurados en circuitería y componentes electrónicos, y un dispositivo de circuito integrado (IC) como un microcontrolador, una unidad central de procesamiento (CPU), una unidad de procesamiento de gráficos (GPU), una matriz de puertas programables en campo (FPGA), otro dispositivo lógico programable o configurable (PLD), un procesador de señales digitales o de tiempo discreto (DSP), un IC específico de la aplicación (ASIC) y similares. En un ejemplo, la unidad 351 de determinación, la unidad 352 de segmentación y la unidad 353 de extracción pueden implementarse como circuitería dentro del procesador 350. En otro ejemplo, la unidad 351 de determinación, la unidad 352 de segmentación y la unidad 353 de extracción pueden implementarse como módulos de software dentro del procesador 350. Varios de la unidad 351 de determinación, la unidad 352 de segmentación y la unidad 353 de extracción pueden compartir componentes de circuito, algoritmos y/o subrutinas entre sí.

En las figuras 4-5 se ilustra un ejemplo de un método de procesamiento de imágenes implementado por la HPU 300. En el paso S401, la HPU 300 recibe una señal de imagen; por ejemplo, a través de la unidad 310 de E/S. La señal de imagen incluye una pluralidad de fotogramas de datos de imagen y puede corresponder a una alimentación de vídeo en directo. En el paso S401, la HPU 300 determina automáticamente una clasificación de imagen basada en al menos un fotograma de la pluralidad de fotogramas de datos de imagen incluidos en la señal de imagen. La determinación puede incluir una serie de subprocesos, como se ilustra en la figura 5. Por ejemplo, en el paso S501, la HPU 300 determina un tipo de contenido para la señal de imagen, en el paso S502, la HPU 300 segmenta los datos de la imagen en una pluralidad de regiones de elementos característicos basándose en el tipo de contenido determinado, y en el paso S503, la HPU 300 extrae al menos un valor de aspecto de imagen para las respectivas de la pluralidad de regiones de elementos característicos. La clasificación de imágenes puede ser realizada por el procesador 350, de modo que la determinación del tipo de contenido del paso S501 puede ser realizada por la unidad 351 de determinación, la segmentación de datos de imagen del paso S502 puede ser realizada por la unidad 352 de segmentación, y la extracción del valor del aspecto de la imagen del paso S503 puede ser realizado por la unidad 353 de extracción. Como puede ser claramente entendido y apreciado por el experto en la materia, una clasificación de imágenes generalmente puede implicar, entre otras, la asignación (por ejemplo, mediante el etiquetado o la segmentación) de imágenes en varias categorías (por ejemplo, predefinidas) y/o una única imagen en varias regiones (p. ej., basándose en el contenido de la imagen). En particular, dicha asignación o categorización se puede realizar basándose en cualquier criterio y/o condición adecuada usando cualquier manera adecuada, dependiendo de varias implementaciones y/o requisitos. Por ejemplo, la asignación o categorización puede lograrse basándose en tipos de contenido determinados a partir de imágenes respectivas. Por lo tanto, en la presente divulgación, la serie de subprocesos/subrutinas S501-S503 puede verse colectivamente y denominarse proceso/algoritmo de clasificación de imágenes o, en resumen, una clasificación de imágenes. Basándose en la clasificación de imágenes, en el paso S403 (véase la figura 4), la HPU 300 genera metadatos de mapeo para la salida; por ejemplo, a través de la unidad 310 de E/S.

Generación y uso de metadatos de mapeo

Estos métodos se describirán con más detalle con referencia a las figuras 6 a 8, que ilustran escenas de ejemplo. Específicamente, las figuras 6 a 8 ilustran ejemplos de fotogramas individuales de datos de imagen, que pueden ser fotogramas del flujo 232 de distribución HDR y/o los datos 311 de entrada. La figura 6 ilustra un fotograma 600, en el que el tipo de contenido es vóley playa. La figura 7 ilustra un fotograma 700, en el que el tipo de contenido es cricket. La figura 8 ilustra un fotograma 800, en el que el tipo de contenido es fútbol (fútbol asociación). Mientras que los tipos de contenido de las figuras 6 a 8 están relacionados con deportes en vivo, la presente divulgación no está tan limitada. Por ejemplo, el tipo de contenido puede ser deportes en vivo, películas, programas de noticias, escenas de la naturaleza y similares.

Al recibir un fotograma (o varios fotogramas) de datos de imagen, como los fotogramas 600, 700 u 800, un sistema de procesamiento de imágenes determina una clasificación de imágenes. Este puede ser un ejemplo del paso S402 ilustrado en la figura 4, y puede ser realizado por la HPU 300 ilustrada en la figura 3. Al determinar la clasificación de imágenes, el sistema de procesamiento de imágenes determina el tipo de contenido, que puede ser un ejemplo del paso S501 ilustrado en la figura 5.

El tipo de contenido puede determinarse analizando varias regiones del fotograma de imagen y determinando una o más regiones de confianza. Por ejemplo, el sistema de procesamiento de imágenes puede analizar el fotograma 600 de imagen y determinar que la gran parte que tiene un color comparativamente beige es una región 601 de confianza, y que la región 601 de confianza probablemente corresponde a arena. El sistema de procesamiento de imágenes puede determinar además que la parte superior del fotograma 600 de imagen incluye una región 602 de confianza, y que la región 602 de confianza probablemente corresponde a una cara. De manera similar, el sistema de procesamiento de imágenes puede analizar el fotograma 700 de imagen y determinar que la parte verde grande es una región 701 de confianza, y que la región 701 de confianza probablemente corresponde a césped. El sistema de procesamiento de imágenes también puede diferenciar entre diferentes tonos del mismo color. Por ejemplo, como se ilustra en la figura 8, el sistema de procesamiento de imágenes puede analizar el fotograma 800 de imagen y determinar que la parte izquierda incluye una región 801 de confianza y la parte derecha incluye otra

región 802 de confianza. Si bien el sistema de procesamiento de imágenes puede determinar que las regiones 801 y 802 de confianza probablemente correspondan a césped, el sistema de procesamiento de imágenes puede diferenciar entre el césped sombreado de la región 801 de confianza y el césped iluminado por el sol de la región 802 de confianza. Mientras que las figuras 6 a 8 ilustran las regiones de confianza respectivas como circulares, en la práctica una región de confianza puede ser elíptica, rectangular o de cualquier otra forma.

Basándose en las regiones de confianza, el sistema de procesamiento de imágenes puede generar una lista clasificada o no clasificada de tipos de contenido potencial. Por ejemplo, en la figura 6 el sistema de procesamiento de imágenes puede determinar que existe un 85 % de probabilidad de que el fotograma 600 de imagen muestre vóley playa, un 12 % de probabilidad de que el fotograma 600 de imagen muestre fútbol playa, un 4 % de probabilidad de que el fotograma de imagen muestre tenis de playa y similares. Esta determinación puede basarse en un solo fotograma de datos de imagen, una serie de fotogramas consecutivos de datos de imagen o una serie de fotogramas no consecutivos de datos de imagen (por ejemplo, cada cuarto fotograma). La determinación se puede realizar repetidamente a lo largo de la emisión, tal como cada diez fotogramas, cada treinta segundos y similares.

Una vez que se ha determinado el tipo de contenido, el sistema de procesamiento de imágenes segmenta los datos de la imagen en una o más regiones de elementos característicos. Este puede ser un ejemplo del paso S502 ilustrado en la figura 5. La segmentación puede basarse en el tipo de contenido en sí; por ejemplo, el sistema de procesamiento de imágenes puede determinar un conjunto ordenado de elementos prioritarios en los datos de imagen para buscar y segmentar. En el ejemplo de vóley playa ilustrado en la figura 6, el sistema de procesamiento de imágenes puede buscar y segmentar primero una región de elementos característicos de arena, luego una región de elementos característicos de multitud basándose en la presencia de múltiples caras en estrecha proximidad, y así sucesivamente. En el ejemplo de cricket ilustrado en la figura 7, de manera similar, el sistema de procesamiento de imágenes puede buscar y segmentar primero una región de elementos característicos de césped, luego jugadores de un primer equipo basándose en el color de la camiseta, y así sucesivamente. La segmentación también puede basarse en el color o el tono; por ejemplo, en el ejemplo de fútbol ilustrado en la figura 8, el sistema de procesamiento de imágenes puede buscar y segmentar una región de elementos característicos de hierba sombreada, una región de elementos característicos de hierba iluminada por el sol, etc. La figura 8 ilustra la segmentación explícitamente, donde el fotograma 800 de imagen está segmentado en una primera región 810 de elementos característicos (césped iluminado por el sol) y una segunda región 820 de elementos característicos (césped sombreado). La segmentación puede basarse en un solo fotograma de datos de imagen, una serie de fotogramas consecutivos de datos de imagen o una serie de fotogramas no consecutivos de datos de imagen (por ejemplo, cada cuarto fotograma). La segmentación se puede realizar repetidamente a lo largo de la emisión, como cada diez fotogramas, cada treinta segundos y similares. En algunos aspectos de la presente divulgación, la segmentación ocurre con más frecuencia que la determinación del tipo de contenido. Por ejemplo, el sistema de procesamiento de imágenes puede determinar un tipo de contenido cada cinco segundos, mientras que el sistema de procesamiento de imágenes puede segmentar los datos de la imagen cada medio segundo.

A partir de las regiones de elementos característicos segmentadas, el sistema de procesamiento de imágenes puede extraer al menos un valor de aspecto de imagen para las respectivas regiones de elementos característicos. Este puede ser un ejemplo del paso S503 ilustrado en la figura 5. Los valores de aspecto de la imagen pueden relacionarse con (entre otros) información de luminancia de las respectivas regiones de elementos característicos. Por ejemplo, los valores de aspecto de la imagen pueden incluir, sin limitación, un máximo de luminancia, un mínimo de luminancia, un punto medio de luminancia, una media de luminancia, una varianza de luminancia y similares. Los valores de aspecto de la imagen pueden representarse visualmente o en la memoria como un histograma. La distribución de los valores de aspecto de la imagen puede derivarse basándose en el contenido de la imagen (p. ej., valores de píxeles, valores de luma, valores de croma, valores Y, valores Cb/Cr, valores RGB y similares), la escena, la ganancia/compensación/potencia, y similares. En algunos aspectos de la presente divulgación, la extracción se produce cada vez que se produce la segmentación.

Una o más de las rutinas y subrutinas implementadas por el sistema de procesamiento de imágenes pueden realizarse automáticamente. Por ejemplo, la HPU 300 puede utilizar un algoritmo de aprendizaje automático como el aprendizaje profundo. El aprendizaje profundo, como se usa en el presente documento, se refiere a una clase de algoritmo de aprendizaje automático que usa una cascada de múltiples capas de unidades de procesamiento no lineal para la extracción y/o transformación de características. Cada capa sucesiva puede usar la salida de la capa anterior como entrada. El aprendizaje profundo puede aprender de manera supervisada (p. ej., clasificación) y/o no supervisada (p. ej., análisis de patrones). El aprendizaje profundo se puede usar para aprender múltiples niveles de representaciones que corresponden respectivamente a diferentes niveles de abstracción, de modo que los niveles formen una jerarquía de conceptos. Los ejemplos de tales técnicas incluyen el trabajo de D. Tran et al., "Aprendizaje de características espaciotemporales con redes convolucionales 3d", la Conferencia internacional IEEE sobre visión por ordenador (ICCV), 2015, pp. 4489-4497, y el trabajo de K. Zhang et al., "Detección y alineación de caras conjuntas mediante redes convolucionales en cascada multitarea", Cartas de procesamiento de señal IEEE 23.10, 2016, págs. 1499-1503.

Los resultados de la clasificación de imágenes, incluidos uno o más del tipo de contenido determinado, las regiones de elementos característicos y/o los valores de aspecto de la imagen, pueden usarse para generar dinámicamente metadatos de mapeo, como la alimentación 235 de metadatos ilustrada en la figura 2 y/o la alimentación 313 de metadatos ilustrado en la figura 3. Como apreciará el experto en la materia, los metadatos de mapeo se pueden generar usando cualquier manera adecuada dependiendo de varias implementaciones y/o requisitos. Por ejemplo, la generación de metadatos de mapeo se puede realizar basándose en algunos o todos los tipos de contenido determinados, las regiones de elementos característicos y/o los valores de aspecto de la imagen, como se ilustra arriba. Además, los metadatos de mapeo pueden generarse dinámicamente a lo largo del procesamiento de la señal de entrada. Es decir, al recibir la señal de imagen/video de entrada (por ejemplo, de una alimentación en vivo), los metadatos de mapeo pueden generarse dinámicamente junto con el procedimiento de clasificación de imágenes (o en otras palabras, junto con la determinación del tipo de contenido, las regiones de elementos característicos y/o los valores de aspecto de la imagen), mejorando así la calidad, la precisión y la eficiencia de la imagen/video cuando se reproduce y, al mismo tiempo, reduciendo o incluso evitando retrasos innecesarios o no deseados (por ejemplo, durante una transmisión en vivo). En términos generales, los metadatos de mapeo pueden generarse de tal manera que permitan o faciliten la conversión (por ejemplo, mapeo) de una señal de entrada a una señal de salida. Por ejemplo, la señal de entrada y la señal de salida pueden tener diferentes rangos dinámicos. En este caso, la conversión puede implicar convertir datos (en la señal de entrada) de un primer rango dinámico en datos (en la señal de salida) de un segundo rango dinámico. En otras palabras, los metadatos pueden generarse para (habilitar/facilitar) la conversión de datos de imagen desde un primer rango dinámico a un segundo rango dinámico (que puede ser mayor o menor que el primer rango dinámico). Como apreciará el experto en la materia, la conversión puede incluir, entre otros, mapeo de tonos y/o gama. Los metadatos de mapeo pueden incluir varios componentes o parámetros usados en el procesamiento de imágenes aguas abajo. A modo de ejemplo, pero no de limitación, la presente divulgación (en particular, la clasificación de imágenes de la misma) puede reconocer una transmisión en vivo como un partido de fútbol. Entonces, la presente divulgación puede decidir o seleccionar una lista de prioridad de objetos como regiones de césped y rostros (humanos). Posteriormente, puede calcular, por objeto en la lista, las características del perfil HDR PQ (por ejemplo, promedio, varianza, etc.) dentro de esas regiones. Posteriormente, dichas características pueden usarse para determinar la curva de mapeo de tonos adecuada. Un caso típico para el mapeo sería uno en el que una alimentación HDR varía de 0,001 nits a 1000 nits, y existe el requisito de que podría ser posible mapear a SDR que es de 0,005 nits a 100 nits. Además, el SDR puede codificarse de acuerdo con el estándar BT.1886. Además, los requisitos también pueden establecer que los rostros (humanos) deben estar en aproximadamente el 70 % del valor máximo del código SDR, y que la hierba debe estar en el 18 % del valor máximo del código SDR. A menudo se puede decir que esto es 70% IRE y 18% IRE, donde IRE se refiere al Instituto de Ingenieros de Radio, una antigua organización profesional que estableció prácticas operativas para la televisión. Ahora, en el presente ejemplo (de un partido de fútbol), supóngase que en la señal HDR, la presente divulgación encuentra que las caras pueden estar sombreadas a 200 nits y que el césped puede estar a 40 nits. Luego, se puede impulsar un algoritmo de optimización para seleccionar parámetros para el algoritmo de mapeo de tonos, de modo que los píxeles a 200 nits en HDR se sean mapeados al 70 % de la señal SDR, y los píxeles a 40 nits en HDR se sean mapeados al 18 % de la señal SDR. Debería ser evidente para los expertos en el campo que se pueden agregar restricciones adicionales para el valor de píxel HDR máximo y el valor de píxel HDR mínimo, de modo que estos también se sean mapeados a niveles adecuados en la señal SDR. Como se ilustra en la figura 8, los metadatos de mapeo se pueden usar en el punto de visualización (por ejemplo, una televisión comercial propiedad de un usuario final) para visualizar un fotograma de imagen renderizado 830 que reproduce fielmente la imagen completa, incluidos los oscuros y los claros. En un ejemplo particular, donde la presente divulgación se implementa en la arquitectura Dolby Vision, los metadatos de mapeo pueden incluir parámetros L1, parámetros L2/L8, parámetros L3, parámetros L4, parámetros L11 y similares.

Los metadatos L1 proporcionan o describen información sobre una distribución de valores de luminancia en una imagen de origen, una escena de origen y similares. Como se indicó anteriormente, la distribución de los valores de aspecto de la imagen se puede derivar basándose en el contenido de la imagen (p. ej., valores de píxeles, valores de luma, valores de croma, valores Y, valores Cb/Cr, valores RGB y similares), la escena y similares. Los metadatos L1 pueden incluir cantidades que representan valores de luminancia mínimos ("aplastamiento"), medios tonos ("medio") y máximos ("recorte") representativos de una o más escenas en los datos de la imagen.

Los metadatos L2 proporcionan o describen información sobre los ajustes de las características del video que se originan a partir de los ajustes realizados por un director, un graduador de color, un profesional de video y similares, o se remontan a ellos. Los metadatos L2 pueden basarse, al menos en parte, en el procesamiento realizado en producción y/o posproducción, como por el convertidor 220 de entrada, el conmutador 221 de producción, la unidad 223 de control de calidad, el servidor 225 de reproducción, la ingesta 226 de archivos, y/o la postproducción 227, ilustrados en la figura 2. Los metadatos L8 son similares a los metadatos L2 y, en algunos casos, pueden ser equivalentes a los metadatos L2 (p. ej., dependiendo de las curvas de tono respectivas). Los metadatos L2 y L8 pueden denominarse parámetros de "recorte" y pueden ser indicativos o estar relacionados con la ganancia/compensación/potencia de los datos de imagen. Los metadatos L2 pueden corresponder a un primer visualizador de referencia con un rango dinámico de referencia particular.

Los metadatos L3 proporcionan o describen información sobre los ajustes de las características del video que se originan a partir de los ajustes realizados por un director, un calificador de color, un profesional de video y similares,

o se remontan a ellos. En comparación con los metadatos L2, los metadatos L3 pueden corresponder a un segundo visualizador de referencia con un rango dinámico de referencia diferente del rango dinámico de referencia del primer visualizador de referencia. Los metadatos L3 pueden incluir, por ejemplo, compensaciones o ajustes de los metadatos L1, incluidas compensaciones o ajustes a los valores de luminancia de aplastamiento, medio y/o recorte.

5

Los metadatos L4 proporcionan o describen información sobre las operaciones de atenuación global. Los metadatos L4 pueden ser calculados por un codificador durante el preprocesamiento y pueden calcularse usando colores primarios RGB. En un ejemplo, los metadatos L4 pueden incluir datos que dictan un nivel de brillo de retroiluminación global de un panel de visualización por fotograma. Otros metadatos generados, como los metadatos L11, pueden proporcionar o describir información que se usará para identificar el origen de los datos de video, como contenido cinematográfico, contenido de juegos de ordenador, contenido deportivo y similares. Dichos metadatos pueden proporcionar o describir además la configuración de imagen prevista, como el punto blanco deseado, la nitidez y similares.

10

15

En conjunto, los metadatos de mapeo pueden incluir datos de conversión para convertir de un primer rango dinámico a un segundo rango dinámico que es diferente del primer rango dinámico. En algunos aspectos de la presente divulgación, el primer rango dinámico puede ser mayor que el segundo rango dinámico (por ejemplo, conversión de HDR a SDR). En otros aspectos de la presente divulgación, el segundo rango dinámico puede ser mayor que el primer rango dinámico (p. ej., conversión de SDR a HDR). Con referencia a la figura 1, los metadatos de mapeo se pueden utilizar para evitar la sobreexposición o la subexposición como en las imágenes 102 y 103, respectivamente. Por ejemplo, los metadatos de mapeo pueden codificarse con los propios datos de la imagen para su uso en el mapeo de tonos por parte de una televisión comercial propiedad de un usuario final.

20

Equivalentes, extensiones, alternativas y varios

25

Con respecto a los procesos, sistemas, métodos, heurísticas, etc. descritos en el presente documento, debe entenderse que, aunque los pasos de tales procesos, etc. se han descrito como si ocurrieran de acuerdo con una cierta secuencia ordenada, podrían agregarse otros pasos, o que ciertos pasos descritos en el presente documento podrían omitirse. En otras palabras, las descripciones de los procesos en el presente documento se proporcionan con el propósito de ilustrar ciertas realizaciones, y de ninguna manera deben interpretarse como una limitación de las reivindicaciones.

30

Se pretende que todos los términos usados en las reivindicaciones se les den sus construcciones razonables más amplias y sus significados ordinarios tal como los entienden los expertos en las tecnologías descritas en el presente documento, a menos que se indique explícitamente lo contrario en el presente documento. En particular, el uso de los artículos singulares como "un/una", "el/la/los/las", "dijo", etc. debe leerse para enumerar uno o más de los elementos indicados, a menos que una reivindicación enumere una limitación explícita al contrario.

35

Así se describen aspectos de ejemplo que se relacionan con la captura, el análisis y la transmisión de video. En la memoria descriptiva anterior, se han descrito aspectos de la presente invención con referencia a numerosos detalles específicos que pueden variar de una implementación a otra. Por lo tanto, el único y exclusivo indicador de lo que es la invención, y lo que los solicitantes pretenden que sea la invención, es el conjunto de reivindicaciones.

40

Cualquier definición expresamente establecida en el presente documento para los términos contenidos en dichas reivindicaciones regirá el significado de dichos términos tal como se usan en las reivindicaciones.

45

Por consiguiente, la especificación y los dibujos deben considerarse en un sentido ilustrativo más que restrictivo.

REIVINDICACIONES

1.- Un sistema de procesamiento de imágenes, que comprende:

5 una entrada configurada para recibir una señal de imagen, incluyendo la señal de imagen una pluralidad de fotogramas de datos de imagen; y

un procesador (350) configurado para determinar automáticamente una clasificación de imágenes basada en al menos un fotograma de la pluralidad de fotogramas, y generar dinámicamente metadatos de mapeo basados en la clasificación de imágenes,

en donde el procesador (350) incluye:

15 circuitería (351) de determinación configurada para determinar un tipo de contenido para la señal de imagen, en donde la circuitería (351) de determinación está configurada para determinar el tipo de contenido analizando regiones del fotograma y determinando una o más regiones de confianza; y en donde la determinación del tipo de contenido implica generar una lista clasificada o no clasificada de tipos de contenido potenciales basándose en una o más regiones de confianza;

20 circuitería (352) de segmentación configurada para segmentar los datos de imagen en una pluralidad de regiones de elementos característicos, basándose en el tipo de contenido, en donde la segmentación de los datos de imagen implica, basándose en el tipo de contenido determinado, determinar un conjunto ordenado de elementos prioritarios en los datos de imagen para buscar y segmentar; y

25 circuitería (353) de extracción configurada para extraer al menos un valor de luminancia de imagen para las regiones respectivas de la pluralidad de regiones de elementos característicos,

en donde los metadatos de mapeo se generan dinámicamente basándose en el tipo de contenido, las regiones de elementos característicos y el valor de luminancia de la imagen; y

30 en donde los metadatos de mapeo incluyen datos de conversión para convertir de un primer rango dinámico a un segundo rango dinámico que es diferente al primer rango dinámico.

35 2.- El sistema de procesamiento de imágenes de acuerdo con la reivindicación 1, en donde al menos un valor de luminancia de imagen incluye al menos uno seleccionado de un máximo de luminancia, un mínimo de luminancia, un punto medio de luminancia, una media de luminancia o una varianza de luminancia.

40 3.- El sistema de procesamiento de imágenes de acuerdo con la reivindicación 1 o 2, en donde una región de elementos característicos respectiva indica al menos uno seleccionado de una región de paisaje, una región de sombra, una región de cielo, una región de detección facial o una región de multitud.

4.- El sistema de procesamiento de imágenes de acuerdo con cualquiera de las reivindicaciones 1 a 3, en donde la señal de imagen es una alimentación de video en vivo.

45 5.- El sistema de procesamiento de imágenes de acuerdo con cualquiera de las reivindicaciones 1 a 4, que comprende además un codificador configurado para codificar la señal de imagen y los metadatos de mapeo.

6.- El sistema de procesamiento de imágenes de acuerdo con la reivindicación 1, en donde el primer rango dinámico es mayor que el segundo rango dinámico.

50 7.- Un método de procesamiento de imágenes que comprende:

recibir (S401) una señal de imagen, incluyendo la señal de imagen una pluralidad de fotogramas de datos de imagen;

55 determinar automáticamente (S402) una clasificación de imagen basada en al menos un fotograma de la pluralidad de fotogramas, incluyendo:

60 determinar (S501) un tipo de contenido para la señal de imagen, en donde el tipo de contenido se determina analizando regiones del fotograma y determinando una o más regiones de confianza; y en donde la determinación del tipo de contenido implica generar una lista clasificada o no clasificada de posibles tipos de contenido basándose en una o más regiones de confianza,

65 segmentar (S502) los datos de imagen en una pluralidad de regiones de elementos característicos, basándose en el tipo de contenido, en donde la segmentación de los datos de imagen implica, basándose en el tipo de contenido determinado, determinar un conjunto ordenado de elementos prioritarios en los datos de imagen para buscar y

segmento, y

extraer (S503) al menos un valor de luminancia de imagen para las regiones respectivas de la pluralidad de regiones de elementos característicos; y

5

generar (S403) una pluralidad de fotogramas de metadatos de mapeo basándose en la clasificación de imágenes, en donde los respectivos de la pluralidad de fotogramas de metadatos de mapeo corresponden a los respectivos de la pluralidad de fotogramas de datos de imagen,

10

en donde los metadatos de mapeo se generan dinámicamente basándose en el tipo de contenido, las regiones de elementos característicos y el valor de luminancia de la imagen; y

en donde los metadatos de mapeo incluyen datos de conversión para convertir de un primer rango dinámico a un segundo rango dinámico que es diferente al primer rango dinámico.

15

8.- El método de procesamiento de imágenes de acuerdo con la reivindicación 7, en donde al menos un valor de luminancia de imagen incluye al menos uno seleccionado de un máximo de luminancia, un mínimo de luminancia, un punto medio de luminancia, una media de luminancia o una varianza de luminancia.

20

9.- El método de procesamiento de imágenes de acuerdo con la reivindicación 7 u 8, en donde una región de elementos característicos respectiva indica al menos una seleccionada de una región de paisaje, una región de sombra, una región de cielo, una región de detección facial o una región de multitud.

25

10.- El método de procesamiento de imágenes de acuerdo con cualquiera de las reivindicaciones 7 a 9, en donde la señal de la imagen es una alimentación de video en vivo.

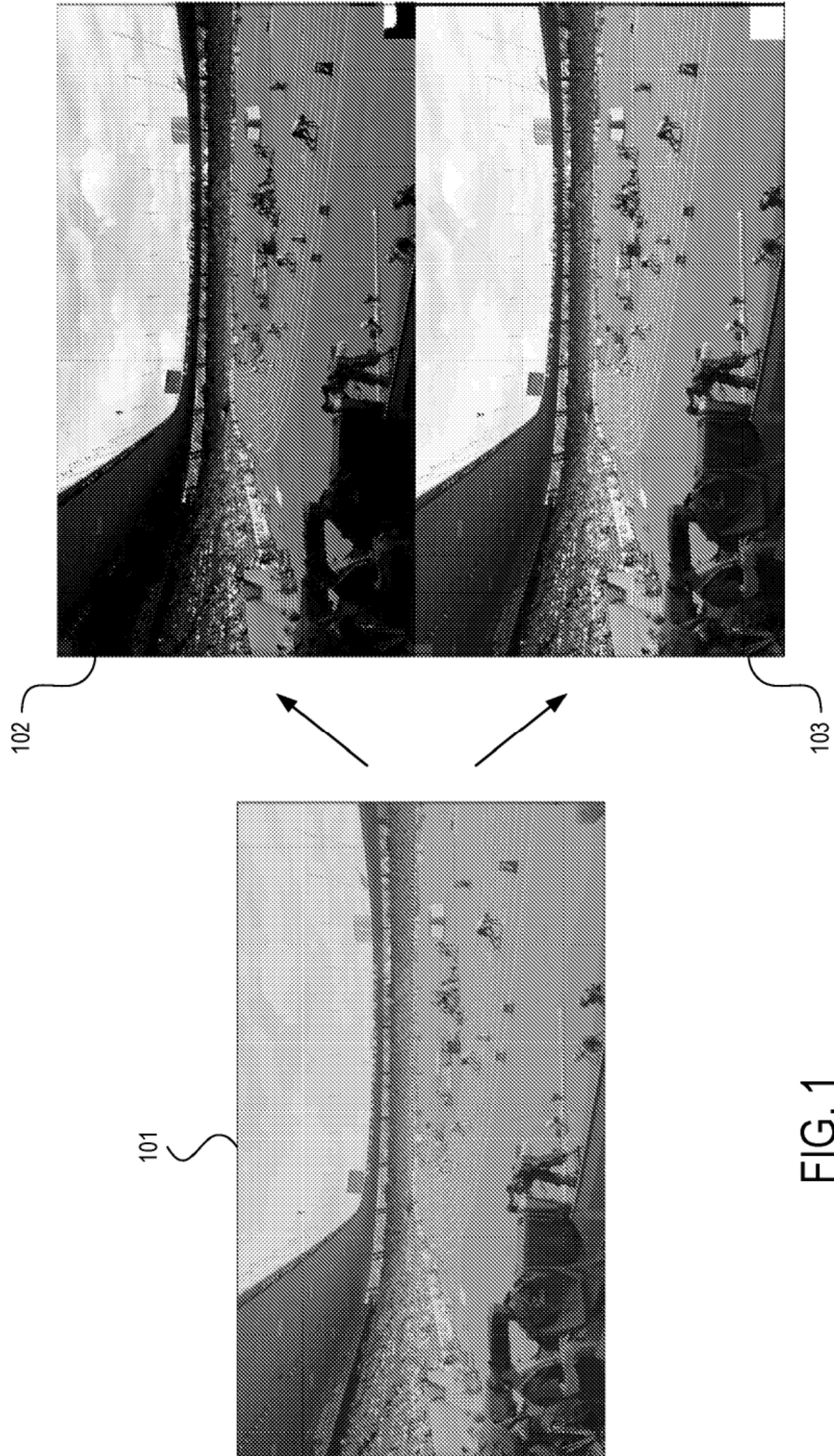
11.- El método de procesamiento de imágenes de acuerdo con cualquiera de las reivindicaciones 7 a 10, que comprende además la codificación de la señal de imagen y los metadatos de mapeo en una señal de salida comprimida.

30

12.- El método de procesamiento de imágenes de acuerdo con la reivindicación 7, en donde el primer rango dinámico es mayor que el segundo rango dinámico.

35

13.- Un medio legible por ordenador no transitorio que almacena instrucciones que, cuando son ejecutadas por un procesador de un sistema de procesamiento de imágenes, hacen que el sistema de procesamiento de imágenes realice el método de acuerdo con cualquiera de las reivindicaciones 7 a 12.



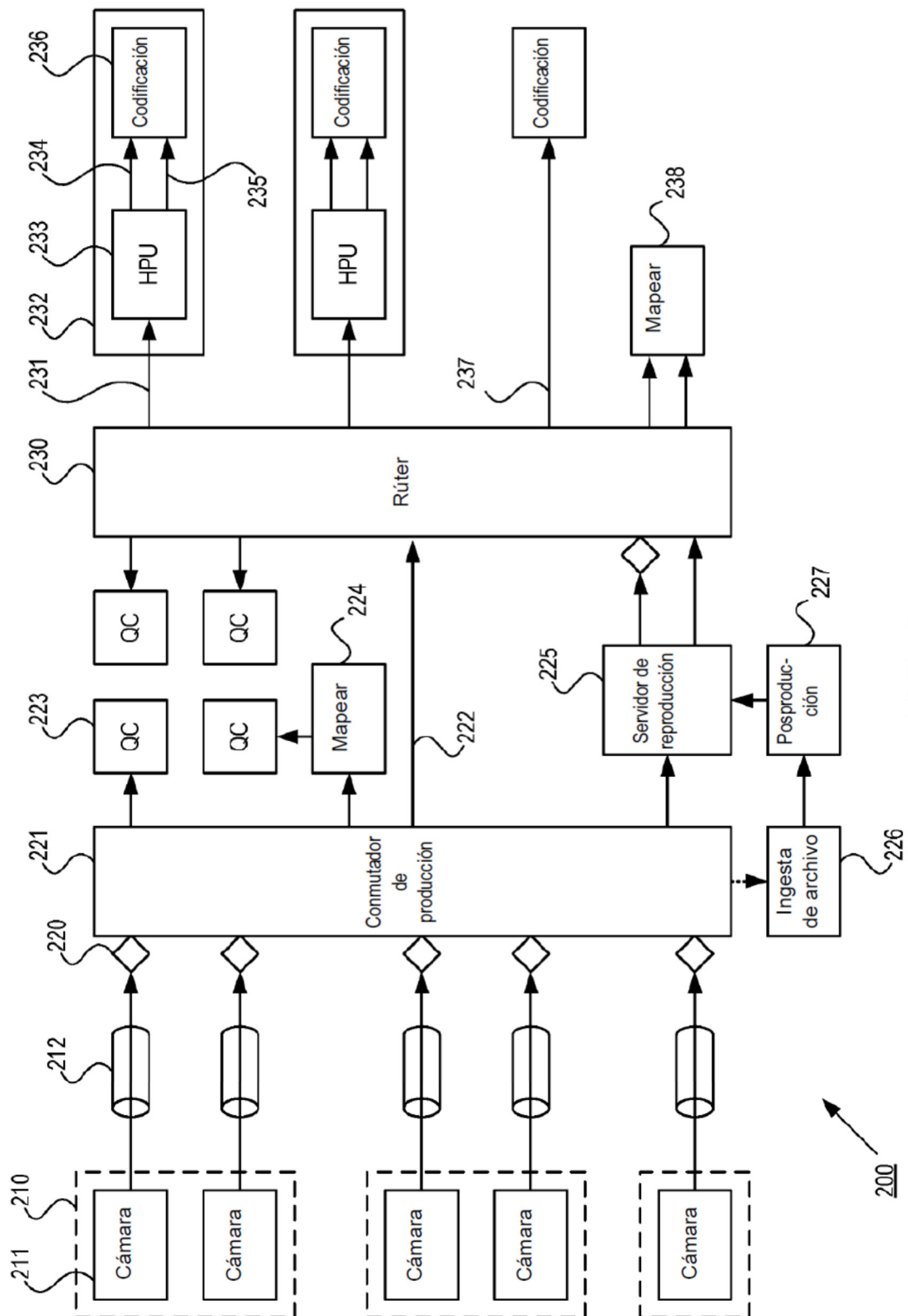


FIG. 2

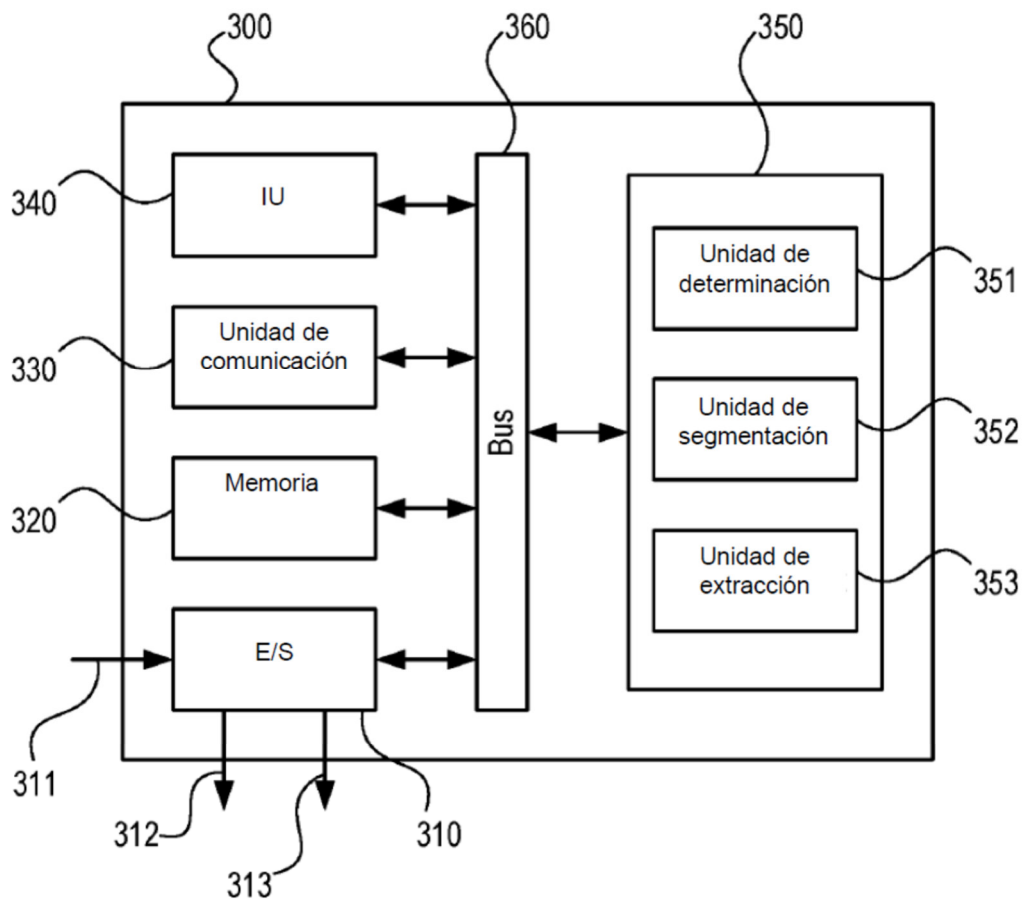


FIG. 3

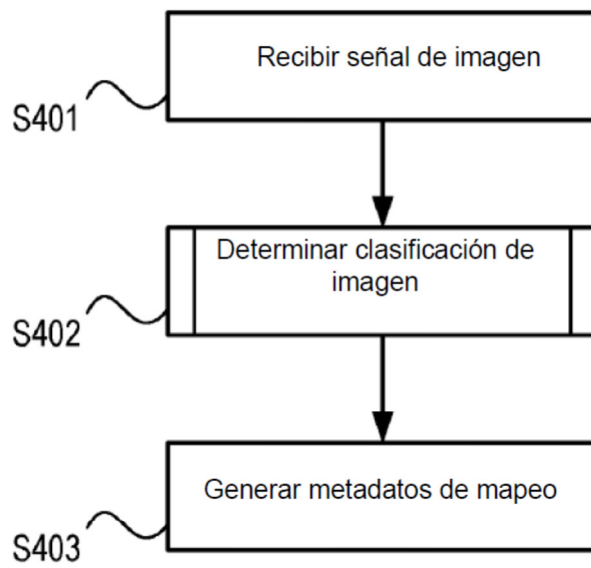


FIG. 4

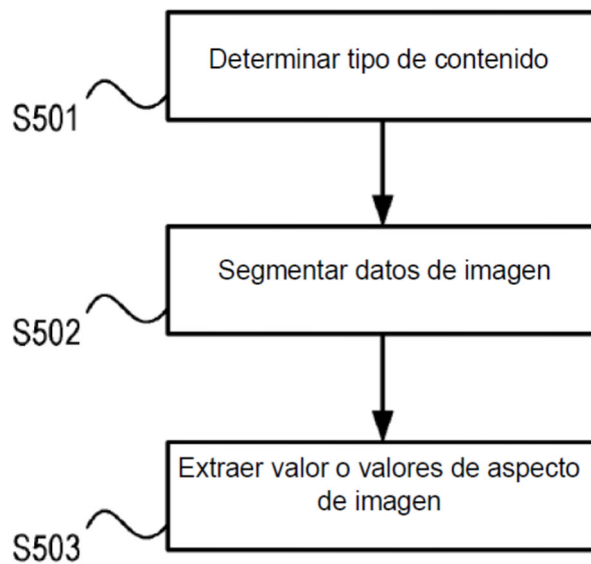


FIG. 5

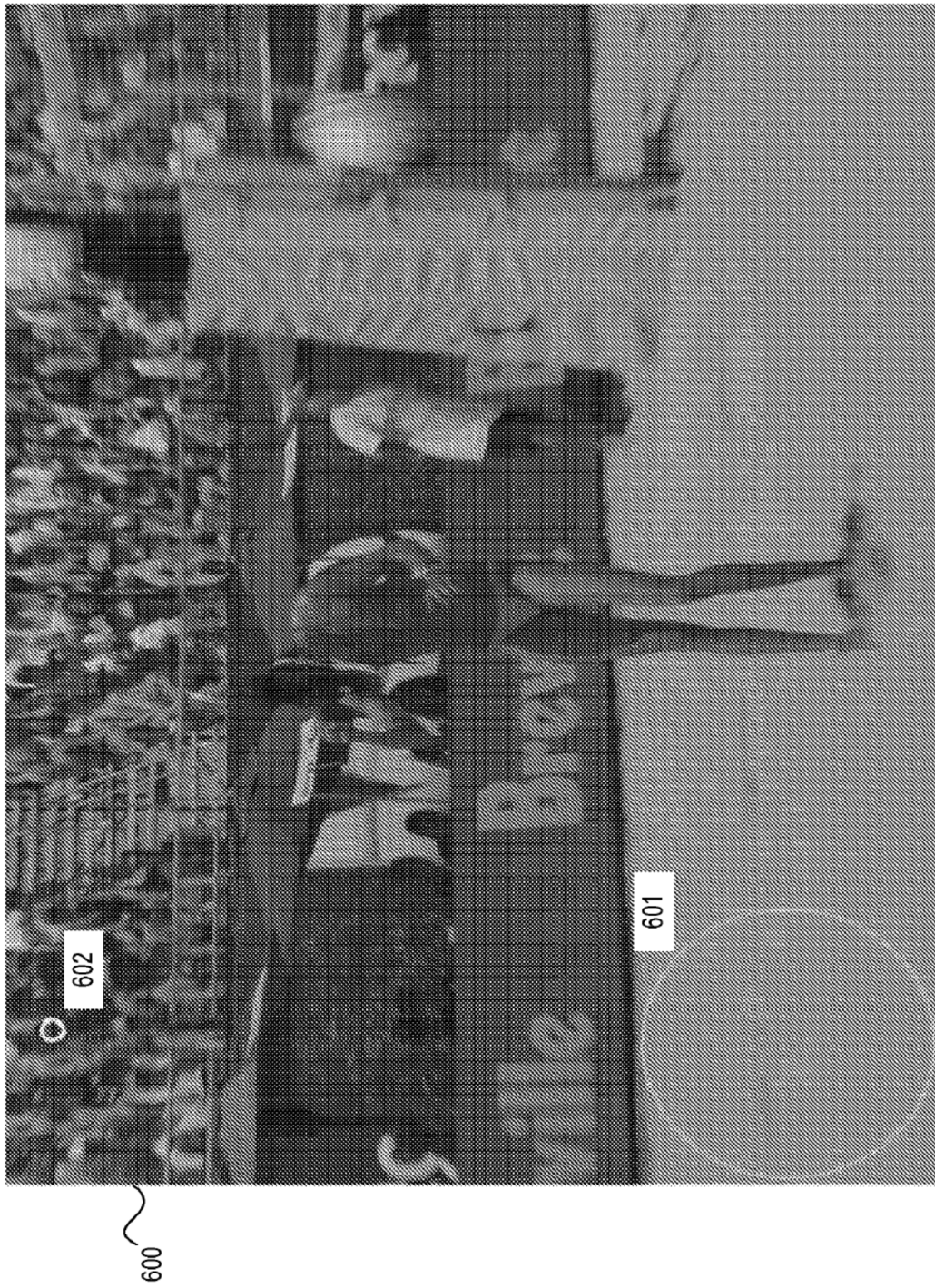


FIG. 6

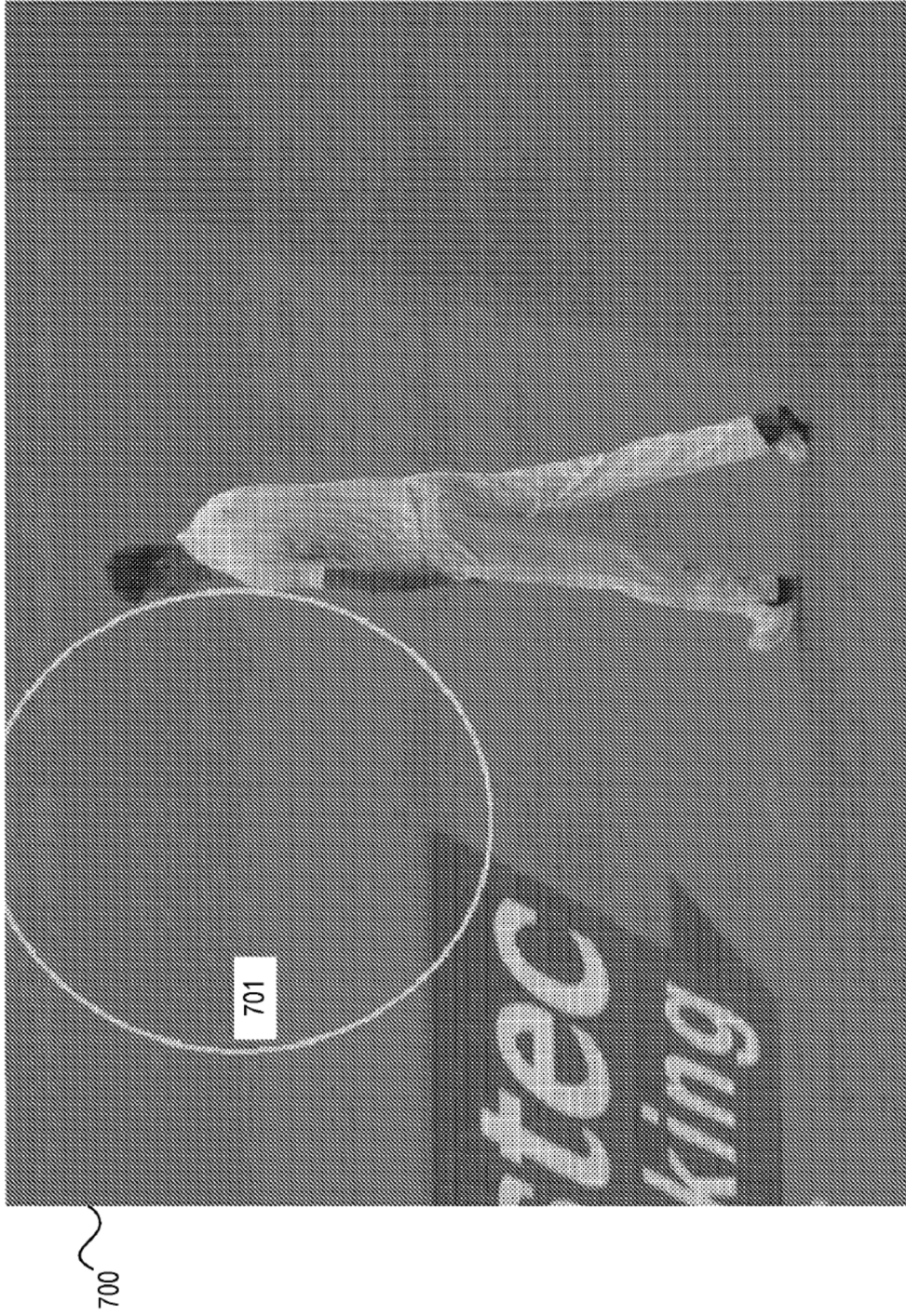


FIG. 7

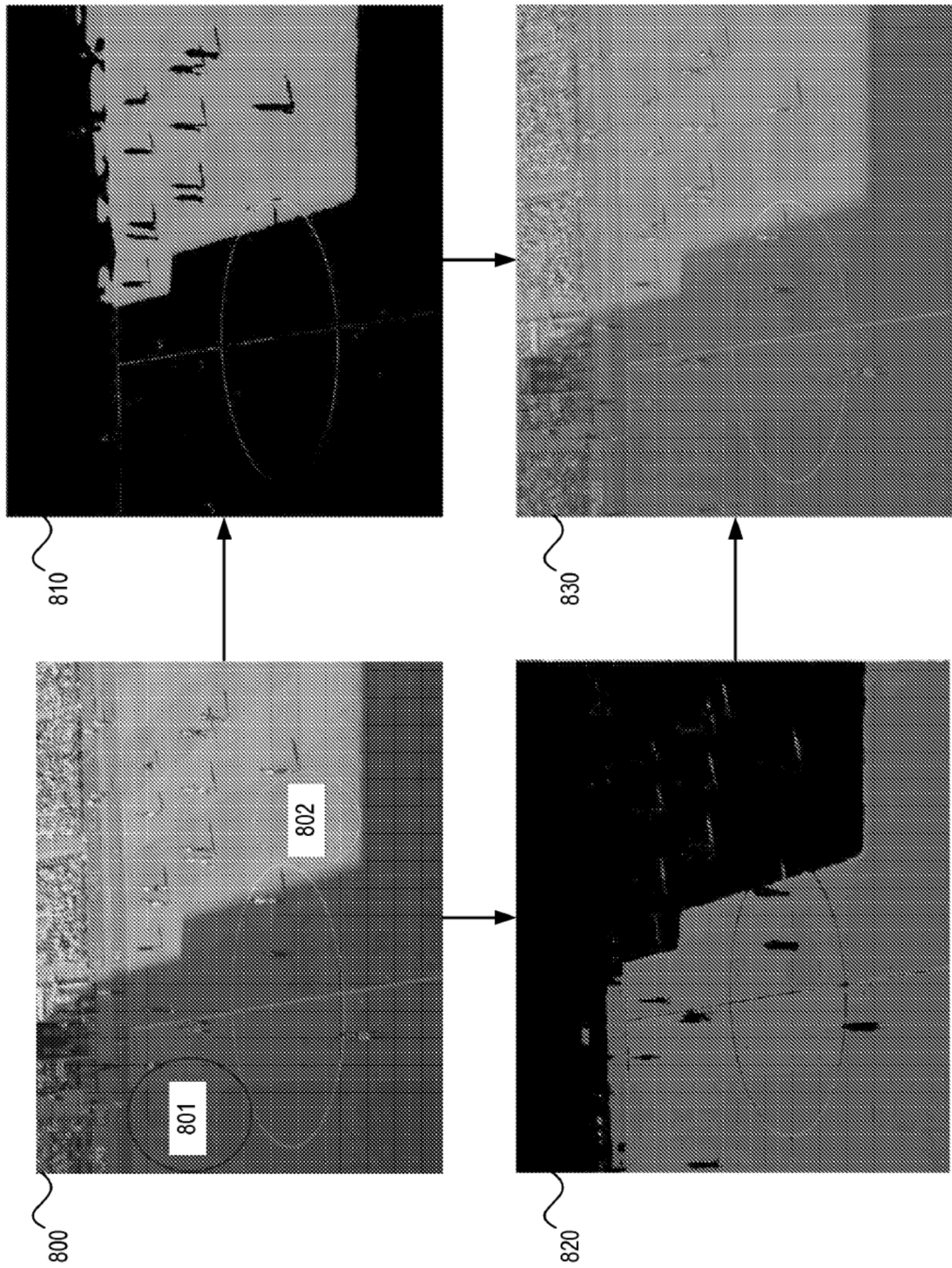


FIG. 8