



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 345 893**

51 Int. Cl.:
H04N 7/15 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **05710946 .4**

96 Fecha de presentación : **11.02.2005**

97 Número de publicación de la solicitud: **1721462**

97 Fecha de publicación de la solicitud: **15.11.2006**

54 Título: **Disposición y procedimiento para generar imágenes de presencia continua.**

30 Prioridad: **13.02.2004 NO 20040661**

45 Fecha de publicación de la mención BOPI:
05.10.2010

45 Fecha de la publicación del folleto de la patente:
05.10.2010

73 Titular/es: **TANDBERG TELECOM AS.**
Philip Pedersens vei, 22
1366 Lysaker, NO

72 Inventor/es: **Lia, Tom, Erik y**
Johansen, Tom-Ivar

74 Agente: **Ponti Sales, Adelaida**

ES 2 345 893 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

Disposición y procedimiento para generar imágenes de presencia continua.

5 **Campo de la invención**

La presente invención se refiere a la videoconferencia y, en particular, a la generación de imágenes de presencia continua (CP, *Continuous Presence*) en una unidad de control multipunto (MCU, *Multipoint Control Unit*).

10 **Antecedentes de la invención**

La transmisión de imágenes en movimiento en tiempo real se utiliza en varias aplicaciones tales como, por ejemplo, videoconferencias, reuniones a través de la red, difusión de TV y videotelefonía.

15 Sin embargo, la representación de imágenes en movimiento requiere una gran cantidad de información ya que el vídeo digital se describe normalmente representando cada píxel de una imagen con 8 bits (1 octeto). Tales datos de vídeo no comprimidos dan como resultado grandes volúmenes de bits y no pueden transferirse a través de redes de comunicación y líneas de transmisión convencionales en tiempo real debido a un ancho de banda limitado.

20 Por lo tanto, permitir la transmisión de vídeo en tiempo real requiere un alto grado de compresión de datos. Sin embargo, la compresión de datos puede comprometer la calidad de las imágenes. Por lo tanto, se han llevado a cabo grandes esfuerzos para desarrollar técnicas de compresión que permitan la transmisión en tiempo real de vídeo de alta calidad a través de conexiones de datos limitadas por el ancho de banda.

25 En los sistemas de compresión de vídeo, el objetivo principal es representar la información de vídeo con la menor capacidad posible. La capacidad se define con bits, ya sea como un valor constante o como una unidad de bits/tiempo. En ambos casos, el objetivo principal es reducir el número de bits.

30 El procedimiento de codificación de vídeo más común se describe en las normas MPEG* y H.26*. Los datos de vídeo pasan por cuatro procesos principales antes de su transmisión, concretamente predicción, transformación, cuantificación y codificación de entropía.

35 El proceso de predicción reduce significativamente la cantidad de bits requerida para cada imagen de una secuencia de vídeo que va a transferirse. Aprovecha la similitud de partes de la secuencia con otras partes de la secuencia. Puesto que la parte de predicción es conocida tanto para el codificador como para el decodificador, solo tiene que transferirse la diferencia. Esta diferencia requiere normalmente mucha menos capacidad para su representación. La predicción se basa principalmente en el contenido de la imagen a partir de imágenes reconstruidas anteriormente en las que la ubicación del contenido se define mediante vectores de movimiento. El proceso de predicción se realiza normalmente en tamaños de bloques cuadrados (por ejemplo, 16x16 píxeles).

40 Los sistemas de videoconferencia permiten el intercambio simultáneo de información de audio, de vídeo y de datos entre múltiples emplazamientos de la conferencia. Los sistemas conocidos como unidades de control multipunto (MCU) llevan a cabo funciones de conmutación para permitir que múltiples emplazamientos se intercomunicen en una conferencia. La MCU conecta los emplazamientos entre sí recibiendo tramas de señales de conferencia desde
45 los emplazamientos, procesando las señales recibidas y retransmitiendo las señales procesadas a los emplazamientos apropiados. Las señales de conferencia incluyen información de audio, de vídeo, de datos y de control. En una conferencia conmutada, la señal de vídeo de uno de los emplazamientos de la conferencia, normalmente el del interlocutor que hable más alto, se difunde a cada uno de los participantes. En una conferencia de presencia continua, las señales de vídeo de dos o más emplazamientos se mezclan espacialmente para formar una señal de vídeo compuesta para su
50 visualización por parte de los participantes de la conferencia. La imagen compuesta o de presencia continua es una imagen combinada que puede incluir flujos de vídeo en directo, imágenes estáticas, menús u otras imágenes visuales de los participantes de la conferencia.

55 En una conferencia de presencia continua típica, la visualización de vídeo se divide en una distribución compuesta que presenta áreas o regiones (por ejemplo, cuadrantes). Los emplazamientos se seleccionan en la configuración de la conferencia a partir de los emplazamientos conectados en la conferencia para su visualización en las regiones. Las distribuciones compuestas comunes incluyen cuatro, nueve o dieciséis regiones. La distribución se selecciona y después se fija durante la duración de la conferencia.

60 Algunas disposiciones de conferencia proporcionan diferentes señales compuestas o una mezcla de vídeo de manera que cada emplazamiento puede ver una mezcla diferente de emplazamientos. Otra disposición utiliza una selección de cuadrantes activados por voz para asociar emplazamientos con cuadrantes particulares. Esa disposición permite a los participantes de la conferencia ver no solamente emplazamientos fijos de mezcla de vídeo, sino también un emplazamiento seleccionado en función de la actividad de voz. Sin embargo, la distribución, en lo que respecta al número
65 de regiones o cuadrantes, es fija para la conferencia.

Haciendo referencia ahora a la Fig. 1, se muestra un diagrama esquemático de una realización de una MCU 10 del tipo desvelado en la patente estadounidense 5.600.646, cuya descripción se incorpora expresamente en este documento

como referencia. La MCU 10 incluye además funcionalidad H.323 tal y como se desvela en la patente estadounidense 6.404.745, cuya descripción también se incorpora expresamente en este documento como referencia. Además, el procesamiento de vídeo en la MCU se ha mejorado, tal y como se describirá en detalle en este documento. Las características descritas en este documento para la MCU 10 pueden representarse en una MCU Tandberg.

La MCU 10 incluye al menos una unidad de interfaz de red (NIU, *Network Interface Unit*) 120, al menos una unidad de procesamiento de puente (BPU, *Bridge Processing Unit*) 122, una unidad de procesamiento de vídeo (VPU, *Video Processing Unit*) 124, una unidad de procesamiento de datos (DPU, *Data Processing Unit*) 126 y una unidad de procesamiento principal (HPU, *Host Processing Unit*) 130. Además de un bus de control principal 132 de la Arquitectura Estándar Industrial (ISA, *Industry Standard Architecture*), la MCU 10 incluye un bus de red 134, un bus BPU 136 y un bus X 138. El bus de red 134 cumple con el protocolo de integración de múltiples fabricantes (MVIP, *Multi-Vendor Integration Protocol*), mientras que el bus BPU 136 y el bus X son derivados de la especificación MVIP. La HPU 130 proporciona una interfaz de gestión para operaciones MCU. Cada uno de los elementos MCU anteriores se describe en detalle en las patentes estadounidenses 5.600.646 y 6.404.745 mencionadas anteriormente.

La funcionalidad H.323 se proporciona mediante la adición de una unidad de procesamiento de pasarela (GPU, *Gateway Processing Unit*) 128 y de una BPU modificada denominada como una BPU-G 122A. La GPU 128 utiliza protocolos H.323 para la señalización de llamadas y la creación y control de flujos de audio, de vídeo y de datos a través de una Ethernet u otra interfaz LAN 140 para terminales finales. La BPU-G 122A es una BPU 122 que está programada para procesar paquetes de audio, de vídeo y de datos recibidos desde la GPU 128.

A continuación se describe el funcionamiento de una MCU a un alto nivel, inicialmente para conferencias por conmutación de circuitos y después para conferencias H.323 por conmutación de paquetes. En conferencias por conmutación de circuitos, las tramas de datos digitales de terminales finales H.320 de conmutación de circuitos se vuelven disponibles en el bus de red 134 a través de una interfaz de red 142 para una NIU 120. Las BPU 122 procesan las tramas de datos a partir del bus de red 134 para generar tramas de datos que se vuelven disponibles para otras BPU 122 en el bus BPU 136. Las BPU 122 también extraen información de audio de las tramas de datos.

Las BPU 122 combinan información de vídeo comprimida e información de audio codificada mezclada en tramas que se colocan en el bus de red 134 para su transmisión a terminales H.320 respectivos.

En casos en los que los terminales audiovisuales funcionan a diferentes velocidades de transmisión o con diferentes algoritmos de compresión o van a mezclarse en una imagen compuesta, múltiples entradas de vídeo se envían a la VPU 124, en donde las entradas de vídeo se descomprimen, se mezclan y vuelven a comprimirse en un único flujo de vídeo. Este único flujo de vídeo se devuelve después a través de la BPU 122, la cual conmuta el flujo de vídeo a los terminales finales apropiados.

Para una conferencia H.323 basada en paquetes, la GPU 128 vuelve disponibles paquetes de audio, de vídeo y de datos en el bus de red 134. Los paquetes de datos se procesan a través de la DPU 126. La BPU-G 122A procesa paquetes de audio y vídeo a partir del bus de red 134 para generar mezclas de difusión de audio y vídeo que se colocan en el bus de red 134 para su transmisión a terminales finales respectivos a través de la GPU 128. Además, la BPU-G 122A procesa paquetes de audio y vídeo para generar tramas de datos que se vuelven disponibles para las BPU 122 en el bus BPU 136. De esta manera, la MCU 10 realiza una función de pasarela mediante la cual las BPU 122 habituales y la BPU-G 122A pueden intercambiar audio y vídeo entre terminales H.320 y H.323 de manera transparente.

Una vez descritos los componentes de la MCU 10 que permiten las funciones puente básicas de una conferencia, a continuación se proporciona una descripción a un alto nivel de la flexibilidad proporcionada por la VPU 124 con referencia al diagrama de bloques funcional de la Fig. 2. En la MCU 10, información de vídeo comprimida de hasta cinco terminales audiovisuales que están en la misma conferencia se encamina hasta una VPU 124 particular a través del bus BPU 136. La VPU 124 comprende cinco procesadores de compresión de vídeo (VCP0 a VCP4), presentando cada uno un par descodificador/codificador de vídeo 102-i, 106-i, y bloques de escalado de píxeles 104-i, 108-i.

Un par descodificador/codificador de vídeo 102-i, 106-i está asignado al flujo de información de vídeo comprimido asociado a cada emplazamiento particular de la conferencia. Cada descodificador de vídeo 102-i descodifica la información de vídeo comprimida utilizando el algoritmo que coincide con el algoritmo de codificación de su emplazamiento asociado. El procesamiento para determinar la estructura de trama, los paquetes y las sumas de control que pueden ser parte del protocolo de transmisión puede estar incluido como parte del descodificador de vídeo 102-i. Debe observarse que un flujo de vídeo codificado por procesador puede asignarse a múltiples emplazamientos (por ejemplo, una aplicación de presencia continua que tenga más de cinco emplazamientos en la conferencia). Además, un par descodificador/codificador 102-i, 106-i puede conmutar entre los emplazamientos de una conferencia.

La información de vídeo descodificada (por ejemplo, píxeles) se escala de manera ascendente o descendente, si fuera necesario, por un bloque de escalado de píxeles 104-i para adaptarse a los requisitos de resolución de píxeles de otros emplazamientos de la conferencia que codificarán los píxeles escalados. Por ejemplo, un sistema de escritorio puede codificar a una resolución de 256x240 píxeles, mientras que un terminal H.320 puede requerir una resolución de píxel de 352x288 píxeles para una imagen de formato intermedio común (CIF, *Common Intermediate Format*). Otros formatos comunes incluyen el formato intermedio común con un cuarto de resolución (QCIF, *Quarter Common*

Intermediate Format) (176x144 píxeles), 4CIF (704x576), SIF (352x240), 4SIF (704x480), VGA (640x480), SVGA (800x600) y XGA (1024x768).

La VPU 124 incluye un bus de píxeles 182 y una memoria 123. El sistema desvelado en la patente estadounidense 5.600.646 utiliza un bus de multiplexión por división de tiempo. En particular, cada decodificador 102-j proporciona píxeles en el bus de píxeles 182 a la memoria 123. Cada codificador 106-j puede recuperar cualquiera de las imágenes de la memoria 123 en el bus de píxeles para una recodificación y/o una mezcla o composición espacial. Otro bloque de escalado de píxeles 108-j está acoplado entre el bus de píxeles 182 y el codificador 106-j para ajustar la resolución de píxeles de la imagen muestreada según sea necesario.

A continuación se describirá una aplicación de presencia continua con referencia a las Figs. 3 y 4. Por motivos de simplicidad, los terminales finales mostrados son terminales H.320. En la Fig. 3, los datos de los emplazamientos 38 llegan a través de una red de comunicaciones a las NIU 120 respectivas. Cinco emplazamientos 38 (A, B, C, D, E) están conectados en la conferencia. Los emplazamientos A y B se muestran conectados a una NIU 120 particular que soporta múltiples conexiones de códec (por ejemplo, una interfaz T1). Los otros emplazamientos C, D y E están conectados a NIU 120 que soportan solamente una única conexión de códec (por ejemplo, una interfaz ISDN). Cada emplazamiento 38 coloca uno o más octetos de datos digitales en el bus de red 134 como datos de trama H.221 no sincronizados. Después, las BPU 122 determinan la alineación de octetos y la estructura de trama H.221. Estos datos alineados se vuelven disponibles para todas las demás unidades en el bus BPU 136. Las BPU 122 extraen además información de audio de las tramas H.221 y descodifican el audio en datos PCM de 16 bits. Los datos de audio descodificados se vuelven disponibles en el bus BPU 136 para mezclarse con datos de audio de otros emplazamientos de la conferencia.

Las tramas H.221 alineadas se reciben por la VPU 124 para procesarse mediante elementos de codificador/descodificador denominados como procesadores de compresión de vídeo (VCP, *video compression processors*). La VPU 124 presenta 5 VCP (Fig. 2) que en este ejemplo están asignados respectivamente a los emplazamientos A, B, C, D, E. Un VCP de la VPU 124 que está asignado al emplazamiento E se ilustra de manera funcional en la Fig. 4. La información de vídeo comprimida (H.261) se extrae de las tramas H.221 y se descodifica por el VCP como una imagen X. La imagen X de vídeo de decodificador se coloca en el bus de píxeles 182 a través de un bloque de escalado. La Fig. 4 muestra el bus de píxeles 182 con tramas de vídeo descodificadas de cada emplazamiento A, B, C, D, E recuperadas sucesivamente de la memoria 123 identificadas mediante sus direcciones RAM respectivas. El VCP asignado al emplazamiento E recibe las tramas de vídeo descodificadas desde los emplazamientos A, B, C y D que después se disponen en mosaico (se mezclan espacialmente) en una única imagen I compuesta. Después, la imagen I en mosaico se codifica como vídeo H.261 en una estructura de trama H.221 y se coloca en el bus BPU 136 (Fig. 3) para su procesamiento BPU descrito anteriormente.

Tal y como puede observarse a partir de la descripción anterior, la transcodificación requiere considerables recursos de procesamiento, ya que los datos de píxel no procesados tienen que mezclarse y después codificarse para formar una vista mezclada o una vista de presencia continua. Para evitar vistas propias, es decir, para evitar que las vistas CP contengan una imagen de los respectivos participantes a los que se transmiten, la MCU debe incluir al menos un codificador para cada imagen de una vista CP. Para permitir 16 CP, la MCU debe incluir entonces al menos 16 codificadores.

Resumen de la invención

La invención se describe en las reivindicaciones independientes 1 y 8.

Objetivos y ventajas adicionales se consiguen mediante las características descritas en las reivindicaciones dependientes.

Breve descripción de los dibujos

Con el fin entender más fácilmente la invención, la siguiente descripción hará referencia a los dibujos adjuntos, en los que:

Figura 1. Diagrama de bloques de una configuración MCU.

Figura 2. Diagrama de bloques esquemático de una realización de una VPU.

Figura 3. Diagrama de bloques de una configuración MCU que ilustra un flujo de datos para una conferencia de presencia continua.

Figura 4. Diagrama de bloques que ilustra la disposición en mosaico de una imagen en una conferencia de presencia continua.

Figura 5. Diagrama de bloques de las disposiciones de un grupo de bloques en una imagen CIF.

Figura 6. Ilustra la capa del grupo de bloques según la norma H.263.

Figura 7. Ilustra la capa de macrobloques según la norma H.263.

Figura 8. Diagramas de bloques que ilustran tres imágenes diferentes de presencia continua utilizadas en una realización de la presente invención.

Figura 9. Diagrama de bloques esquemático de una realización de la presente invención.

Figura 10. Diagrama de flujo esquemático que ilustra una realización del procedimiento según la invención.

Mejor modo de llevar a cabo la invención

La presente invención utiliza la estructura de bits de la norma H.26* de la ITU para reducir el tiempo de procesamiento y los requisitos en una MCU para generar vistas CP sin vistas propias. Para entender las características de las estructuras de bits que se utilizan, a continuación se describirá la estructura de bloques de imágenes según la norma H.263.

Según la norma H.263, cada imagen se divide en bloques que representan 8x8 píxeles. Los bloques están dispuestos en macrobloques, que para la parte de luminancia de los píxeles son 16 (8x8) bloques y para la parte de crominancia de los píxeles son 4 (2x2) bloques. Un grupo de bloques (GOB, *Group Of Blocks*) representa normalmente 22 macrobloques, y el número de GOB por imagen es de 6 para sub-QCIF, 9 para QCIF y 18 para CIF, 4CIF y 16CIF. La numeración de los GOB se realiza utilizando un barrido vertical de los GOB, empezando con el GOB superior (número 0) y terminando con el GOB inferior. Un ejemplo de la disposición de los GOB en una imagen se proporciona para el formato de imagen CIF de la figura 5. Los datos para cada GOB consisten en una cabecera GOB seguida por datos para macrobloques. Los datos para los GOB se transmiten para cada GOB en un número de GOB creciente. El inicio de un GOB se identifica mediante un código de inicio de grupo de bloques (GBSC, *Group of Block Start Code*). La estructura de la capa GOB se muestra en la figura 6.

Los datos para cada macrobloque consisten en una cabecera de macrobloque seguida por datos para bloques. La estructura se muestra en la figura 7. El COD sólo está presente en imágenes que no son del tipo "INTRA" para cada macrobloque de estas imágenes. Un bit que cuando está fijado a "0" indica que el macrobloque está codificado. Si está fijado a "1" no se transmite información adicional para este macrobloque; en ese caso, el descodificador tratará el macrobloque como un macrobloque INTER con un vector de movimiento para todo el bloque igual a cero y sin datos de coeficientes.

Si el COD está fijado a "0", la parte de datos del macrobloque incluye información de los bloques respectivos en el macrobloque, y esta información se representa mediante vectores de movimiento que indican la posición en imágenes anteriores para las que los píxeles incluidos son iguales.

Convencionalmente, la evitación de vistas propias en una imagen CP requiere una codificación especial para cada uno de los participantes que implica un codificador para cada flujo de datos saliente en la MCU, tal y como se indica en la figura 2. La presente invención utiliza la estructura de macrobloques de los datos de vídeo ya codificados para obtener una mezcla personalizada de una imagen CP dependiente del receptor.

En la siguiente realización de ejemplo de la presente invención, considérese una conferencia con cinco emplazamientos finales que capturan imágenes de vídeo del formato CIP y que codifican las imágenes según la norma H.263. En la MCU, el flujo de datos de los respectivos participantes se descodifica mediante descodificadores asociados con entradas MCU respectivas según la norma H.263. Después de la descodificación, los datos de píxel no procesados de los respectivos participantes estarán disponibles en un bus interno de la MCU, listos para mezclarse y transcodificarse.

En caso de cinco participantes, resulta obvio elegir un formato CP 4 de la imagen mezclada que va a devolverse a los respectivos emplazamientos. El formato mezclado puede seleccionarse por la MCU según los principios de "mejor impresión" descritos en la solicitud de patente estadounidense 10/601.095.

Según esta realización de ejemplo de la invención, dos imágenes CP4 diferentes, una imagen 1 CP y una imagen 2 CP, se codifican por cada codificador respectivo tal y como se ilustra en la figura 8. La imagen 1 CP incluye las imágenes recibidas desde los emplazamientos 1, 2, 3 y 4, mientras que la imagen 2 CP incluye las imágenes recibidas desde el emplazamiento 5 en un cuadrante dejando los cuadrantes restantes vacíos. Cuando se codifican las imágenes CP y se disponen los datos codificados en el sistema de bloques descrito anteriormente, los límites de los cuadrantes coinciden con los límites de los macrobloques de los GOB. En lo que respecta a la imagen 1 CP, los primeros 11 macrobloques del primer GOB incluyen datos de la imagen del emplazamiento 1, mientras que los 11 últimos macrobloques del primer GOB incluyen datos de la imagen del emplazamiento 2.

Según la presente invención, la MCU reorganiza los macrobloques de cada imagen CP según el receptor. Como un ejemplo, en la imagen CP transmitida al emplazamiento 4, los 11 últimos macrobloques de cada uno de los 9 últimos GOB de la imagen 1 CP se sustituyen por los 11 primeros macrobloques de cada uno de los 9 primeros GOB de la imagen 2 CP, respectivamente. Esto da como resultado una nueva imagen CP descodificada que incluye la imagen recibida desde el emplazamiento 5 en lugar de la imagen recibida desde el emplazamiento 4. Esta imagen CP se devuelve al emplazamiento 4, evitando de esta manera la vista propia en ese emplazamiento.

Sustituciones o reorganizaciones correspondientes se llevan a cabo para las otras cuatro imágenes CP asociadas respectivamente con los otros emplazamientos.

La figura 9 ilustra un ejemplo de la arquitectura interna de una MCU según la presente invención. Esta arquitectura se utiliza según la invención en lugar de la VPU de la técnica anterior ilustrada en la figura 2. El bus de píxeles, la memoria y las unidades de escalado de píxeles se sustituyen por simplicidad por una unidad de mezcla y escalado 156. Obsérvese que, como alternativa, el primer bus de datos 176 y el segundo bus de datos 177 de la figura 9 pueden fusionarse para formar un bus de datos común. Obsérvese también que la implementación real puede ser diferente y que solo se muestran las unidades relevantes para la presente invención.

Los flujos de datos de entrada 171, 172, 173, 174 y 175 se descodifican con distintos descodificadores 151, 152, 153, 154, 155 respectivamente, con un descodificador para cada emplazamiento. La codificación se lleva a cabo según la norma de codificación que esté utilizándose, en este caso la H.263. Los datos descodificados están en la forma de datos PCM y se vuelven accesibles para una unidad de mezcla y escalado (MSU, *Mixing and Scaling Unit*) 156 en el primer bus de datos 176.

La MSU 156 mezcla espacialmente los datos PCM 171, 172, 173, 174 del primer, segundo, tercer y cuarto emplazamiento, creando una primera imagen CP. Una segunda imagen CP también se crea colocando los datos PCM 175 del quinto emplazamiento en el primer cuadrante, dejando los cuadrantes restantes vacíos o llenos con datos ficticios. Los datos PCM de las dos imágenes mezcladas espacialmente se vuelven después accesibles para el primer codificador 157 y el segundo codificador 158 siguientes en el segundo bus de datos 177.

Los codificadores 157, 158 extraen los datos PCM de las imágenes CP generadas por la MSU 156 a partir del segundo bus de datos 177 y codifican cada imagen respectiva. El resultado del proceso de codificación es una pluralidad de macrobloques ensamblados en GOB como los descritos anteriormente. En el caso de un formato CIF según la norma H.263, un GOB contiene 22 macrobloques y cada imagen consiste en 18 GOB. Después de codificar las imágenes, los GOB se insertan consecutivamente en memorias intermedias 159, 160 asociadas, respectivamente. El tamaño de las memorias intermedias 159, 160 debe ser lo bastante grande como para alojar los GOB de al menos una imagen. Las salidas de las memorias intermedias 159, 160 están conectadas a un tercer bus de datos 178 que también está conectado a las entradas de los dispositivos de reorganización 161, 162, 163, 164, 165.

Sin embargo, el número de bits que representan una imagen codificada no es constante, sino que puede variar sustancialmente según la variación del contenido de imagen y los movimientos de una imagen a otra. El número de bits también depende de si la imagen está INTRA codificada o INTER codificada, es decir, predicción a partir de macrobloques vecinos en la misma imagen o predicción a partir de una imagen/imágenes anterior(es).

Cuando los datos codificados de imágenes síncronas completas se insertan en las memorias intermedias respectivas, los dispositivos de reorganización 161, 162, 163, 164, 165 están listos para reorganizar el orden de los macrobloques para crear las imágenes CP requeridas para las salidas 181, 182, 183, 184 y 185 asociadas, respectivamente. Los dispositivos de reorganización podrán identificar y aislar los macrobloques mediante el GOB y las cabeceras de macrobloque. El inicio de cada GOB se indica mediante un código de inicio único denominado GBSC (código de inicio de grupo de bloques), seguido de un GN (número de grupo, *Group Number*) que indica el número de GOB. En las cabeceras de los macrobloques, COD indica si el macrobloque está codificado o no. Si COD vale "1", no hay información adicional para ese macrobloque, con el vector de movimiento para todo el bloque igual a cero y sin datos de coeficientes. Si COD vale "0", hay datos adicionales del macrobloque. Algunos de los datos adicionales pueden tener una longitud variable, pero los diferentes códigos están definidos de tal manera que la longitud de cada código está bien definida.

Puesto que los macrobloques pueden identificarse y se almacenan temporalmente en una memoria intermedia 159, 160, los dispositivos de reorganización 161, 162, 163, 164, 165 pueden leer los macrobloques en cualquier orden creando cualquier variante de una imagen CP 4 a partir de la imagen de los cinco emplazamientos. Como un ejemplo, considérese el dispositivo de reorganización 164 que crea la imagen CP para el cuarto emplazamiento, es decir, la salida 184 del emplazamiento 4. La primera memoria intermedia 159 contiene los datos codificados de la imagen 1 CP, mientras que la segunda memoria intermedia 160 contiene los datos codificados de la imagen 2 CP. El dispositivo de reorganización 164 extrae los GOB 1 a 9 de la imagen 1 CP en el mismo orden establecido en la primera memoria intermedia 159 para crear la nueva imagen CP. Sin embargo, cuando se crea el GOB 10, el dispositivo de reorganización 164 identifica y extrae los 11 primeros macrobloques del GOB 10 de la primera memoria intermedia 159 seguidos por los 11 primeros macrobloques del GOB 1 de la segunda memoria intermedia 160. Además, el GOB 11 se crea extrayendo los 11 primeros macrobloques del GOB 11 de la primera memoria intermedia 159 seguidos por los 11 primeros macrobloques del GOB 2 de la segunda memoria intermedia 160. Los 7 GOB restantes se crean de manera similar, terminando con el GOB 18 que se crea extrayendo los 11 primeros macrobloques del GOB 18 de la primera memoria intermedia 159 seguidos por los 11 primeros macrobloques del GOB 9 de la segunda memoria intermedia 160.

Los dispositivos de reorganización 181, 182, 182, 184, 185 pueden preprogramarse para extraer los macrobloques de las memorias intermedias 159, 160 en un orden constante, o el orden de los macrobloques puede controlarse por una unidad de control (no mostrada) permitiendo que un dispositivo de reorganización cree varias imágenes CP.

La fig. 10 es un diagrama de bloques esquemático que ilustra una realización del procedimiento según la invención.

Las etapas de procedimiento ilustradas en la fig. 10 están incluidas en un procedimiento para crear una imagen de presencia continua (CP) objetivo codificada según una norma de codificación de vídeo a partir de una pluralidad de señales de vídeo codificadas que incluyen órdenes definidos de macrobloques. Cada macrobloque comprende señales de vídeo codificadas correspondientes a una imagen de vídeo respectiva de terminal final, recibida desde terminales finales que participan en una conferencia de vídeo multipunto.

El procedimiento comienza en la etapa 202.

En primer lugar se lleva a cabo la etapa de descodificación 204, en la que las señales de vídeo se descodifican para las imágenes de vídeo correspondientes de terminal final.

Después, en la etapa de mezcla 206, las imágenes de vídeo de terminal final se mezclan espacialmente para una pluralidad de imágenes CP compuestas por regiones asociadas respectivamente con cada una de las imágenes de vídeo de terminal final.

Después, en la etapa de codificación 208, las imágenes CP se codifican para una pluralidad de imágenes CP codificadas, respectivamente. Esta etapa establece los órdenes definidos de macrobloques correspondientes a la norma de codificación de vídeo y una fusión de los límites de región y los límites de macrobloque.

Después de las tres etapas preparatorias 204, 206 y 208 anteriores, se lleva a cabo la etapa de creación 210. En la etapa de creación 210, la imagen CP objetivo codificada se crea reorganizando dichos órdenes de macrobloques de una manera predefinida o controlada.

De manera ventajosa, la etapa de creación 210 comprende sustituir un primer número m de macrobloques que representan un ancho de región en n números de GOB subsiguientes que representan una altura de región en una primera imagen de la pluralidad de imágenes CP codificadas por m números de macrobloques en n números de GOB subsiguientes de una segunda imagen de la pluralidad de imágenes CP codificadas.

El procedimiento termina en la etapa 212.

Como alternativa, las tres etapas preparatorias del proceso descrito con referencia a la fig. 10, es decir, la etapa de descodificación 204, la etapa de mezcla 206 y la etapa de codificación 210, pueden sustituirse por una única etapa solicitante (no ilustrada). En esta etapa solicitante, se solicita a los terminales finales que codifiquen las imágenes de vídeo respectivas de terminal final según la norma de vídeo y una resolución, velocidad binaria y escalado determinados.

En su forma más básica, el procedimiento según la invención incluye simplemente la etapa de creación 210.

Las realizaciones descritas hasta ahora se han limitado a la creación de imágenes CP 4 de formato CIF según la norma H.263. Sin embargo, los expertos en la materia observarán que los principios básicos de la presente invención también pueden aplicarse a otras imágenes CP de otros formatos. Como un ejemplo, la presente invención puede utilizarse de hecho en la creación de imágenes CP 9. Por tanto, los límites de las imágenes de cada GOB (en el caso de CIF, H.263) se encuentran después del séptimo y del decimocuarto macrobloque (como alternativa, después del octavo y del decimoquinto o del séptimo y del decimoquinto).

En cualquier caso, debe haber un dispositivo de reorganización (o al menos un procedimiento de reorganización) para cada salida MCU. El número de codificadores depende de la relación entre el número de regiones en las imágenes CP y el número de emplazamientos que van a rellenar las regiones de las imágenes CP. Debe haber un número suficiente de codificadores para generar una región para cada emplazamiento. En caso de CP 4 y de ocho emplazamientos, dos codificadores serán suficientes para colocar una región para cada emplazamiento en cada uno de los ocho cuadrantes totales. Sin embargo, aumentando el número de emplazamientos a nueve, se requerirá un codificador adicional para crear una tercera imagen CP 4 en la que pueda residir la novena región.

En una realización más general de la invención, no se lleva a cabo ninguna descodificación y ningún escalado/mezcla en la MCU. En cambio, se solicita a los terminales finales que transmitan una imagen codificada según una norma, resolución, velocidad binaria y escalado determinados. Los macrobloques de los flujos de datos entrantes se cargan después directamente en las memorias intermedias (preferentemente una para cada flujo de datos), y los dispositivos de reorganización reorganizan los macrobloques según procedimientos preprogramados o controlados, creando imágenes CP sin vista propia que se transmiten a los respectivos emplazamientos de conferencia. Como un ejemplo, considérense cinco emplazamientos que participan en una conferencia. Entonces, una vista CP 4 requiere que los terminales finales coloquen sus respectivas imágenes en uno de los cuadrantes de una imagen completa antes de codificarse. Esto se solicita a los terminales finales junto con información de codificación como la norma, la resolución, la velocidad binaria y el escalado. Después, los dispositivos de reorganización pueden reorganizar fácilmente los macrobloques de los flujos de datos entrantes cuando están presentes en las respectivas memorias intermedias tal y como se ha descrito anteriormente.

Referencias citadas en la descripción

Esta lista de referencias citadas por el solicitante está prevista únicamente para ayudar al lector y no forma parte del documento de patente europea. Aunque se ha puesto el máximo cuidado en su realización, no se pueden excluir errores u omisiones y la OEP declina cualquier responsabilidad al respecto.

Documentos de patente citados en la descripción

- US 5600646 A [0011] [0012] [0021]

- US 6404745 B [0011]

- US 6404745 A [0012]

- US 10601095 B [0034]

REIVINDICACIONES

1. Procedimiento para crear una imagen de presencia continua (CP) objetivo codificada según una norma de codificación de vídeo a partir de una pluralidad de señales de vídeo codificadas incluyendo órdenes definidos de macrobloques, comprendiendo cada uno señales de vídeo codificadas correspondientes a una imagen de vídeo respectiva de terminal final recibida desde terminales finales que participan en una conferencia de vídeo multipunto, **caracterizado** por el hecho de que el procedimiento comprende las siguientes etapas:

- decodificar dichas señales de vídeo codificadas, dando como resultado imágenes de vídeo de terminal final,
- mezclar espacialmente dichas imágenes de vídeo de terminal final, dando como resultado una pluralidad de imágenes CP compuestas por regiones asociadas respectivamente con cada una de dichas imágenes de vídeo de terminal final,
- codificar dichas imágenes CP,
- reorganizar los macrobloques de las imágenes CP codificadas, creando de ese modo dicha imagen CP codificada objetivo.

2. Procedimiento según la reivindicación 1, que comprende además la siguiente etapa adicional, antes de la etapa de crear la imagen CP objetivo codificada:

solicitar a los terminales finales que codifiquen las imágenes de vídeo respectivas de terminal final según la norma de vídeo y una resolución, velocidad binaria y escalado determinados.

3. Procedimiento según la reivindicación 1, en el que dicha etapa de codificar dichas imágenes CP comprende establecer órdenes definidos de macrobloques correspondientes a la norma de codificación de vídeo y una fusión de los límites de región y los límites de macrobloque.

4. Procedimiento según la reivindicación 3, en el que las imágenes CP codificadas y la imagen CP objetivo codificada están cada una en un formato CIF con 18 grupos de bloques (GOB), incluyendo cada uno 22 macrobloques, dispuestos en una formación apilada de manera que los 9 primeros GOB representan regiones superiores y los últimos 9 GOB representan regiones inferiores.

5. Procedimiento según la reivindicación 4, en el que la etapa de crear la imagen CP objetivo codificada incluye:

sustituir m números de macrobloques que representan un ancho de región en n números de GOB subsiguientes que representan una altura de región en una primera imagen de la pluralidad de imágenes CP codificadas por m números de macrobloques en n números de GOB subsiguientes de una segunda imagen de la pluralidad de imágenes CP codificadas.

6. Procedimiento según la reivindicación 5, en el que $m=11$, $n=9$ y dichas regiones representan cada una un cuadrante de una imagen CP.

7. Procedimiento según la reivindicación 5, en el que $m=7$ o $m=8$, $n=6$ y dichas regiones representan cada una un octavo de una imagen CP.

8. Disposición en una unidad de control multipunto (MCU) para crear una imagen CP objetivo codificada según una norma de codificación de vídeo a partir de una pluralidad de señales de entrada de vídeo codificadas, correspondiéndose cada una con una imagen de vídeo respectiva de terminal final recibida desde terminales finales que participan en una conferencia de vídeo multipunto,

caracterizada por

- decodificadores para decodificar cada una de dichas señales de vídeo codificadas, dando como resultado imágenes de vídeo de terminal final,
- una unidad de mezcla y escalado, configurada para mezclar espacialmente dichas imágenes de vídeo de terminal final, dando como resultado una pluralidad de imágenes CP compuestas por regiones asociadas respectivamente con cada una de dichas imágenes de vídeo de terminal final,
- una pluralidad de codificadores, configurados para codificar dichas imágenes CP,
- un dispositivo de reorganización, configurado para reorganizar los macrobloques de las imágenes CP codificadas, creando de ese modo dicha imagen CP codificada objetivo.

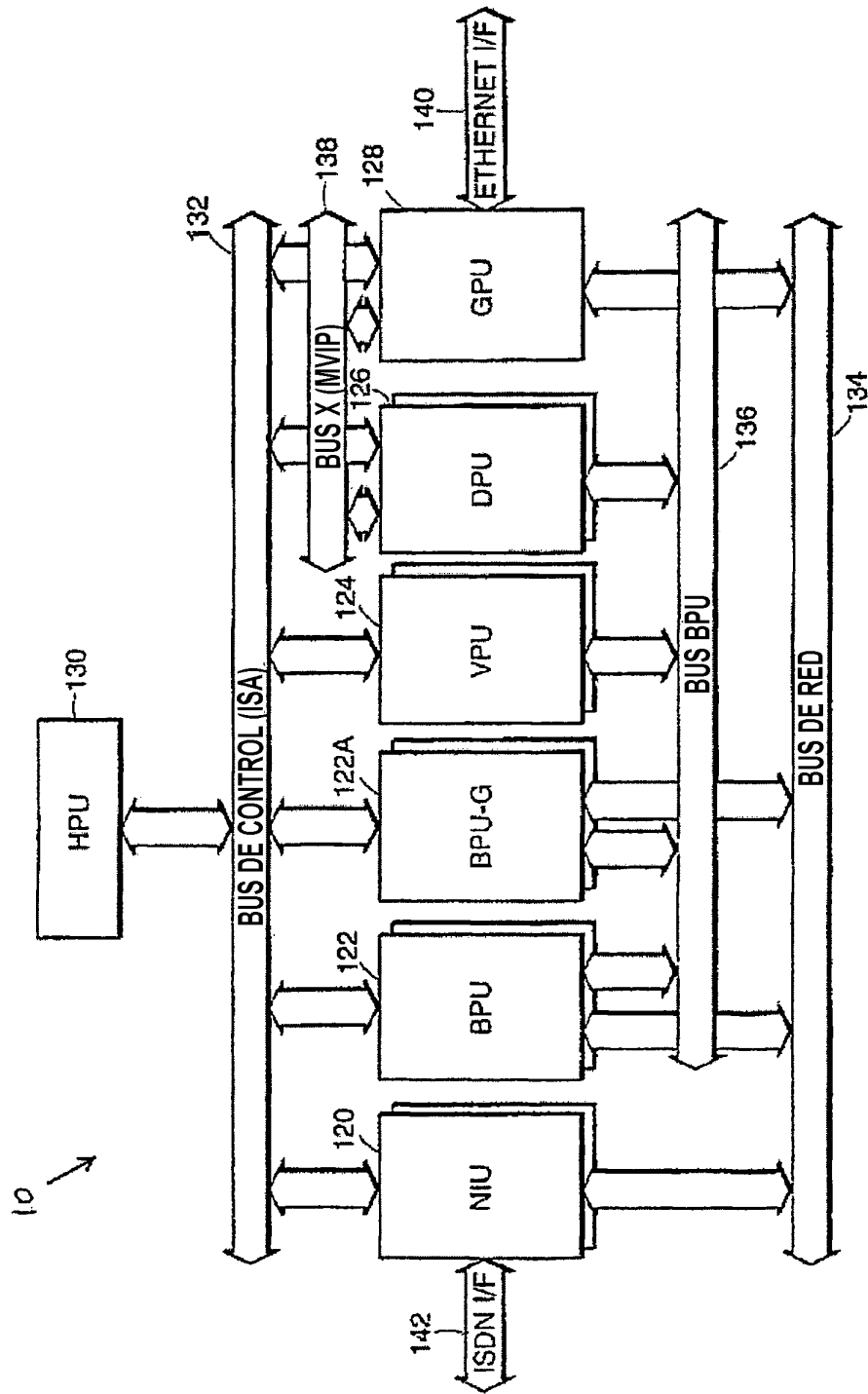
ES 2 345 893 T3

9. Disposición según la reivindicación 8, en la que las imágenes CP codificadas y la imagen CP objetivo codificada están cada una en un formato CIF con 18 grupos de bloques (GOB), incluyendo cada uno 22 macrobloques, dispuestos en una formación apilada de manera que los 9 primeros GOB representan regiones superiores y los últimos 9 GOB representan regiones inferiores.

10. Disposición según la reivindicación 9, en la que dicho dispositivo de reorganización está configurado además para sustituir m números de macrobloques que representan un ancho de región en n números de GOB subsiguientes que representan una altura de región en una primera imagen de la pluralidad de imágenes CP codificadas por m números de macrobloques en n números de GOB subsiguientes de una segunda imagen de la pluralidad de imágenes CP codificadas.

11. Disposición según la reivindicación 10, en la que $m=11$, $n=9$ y dichas regiones representan cada una un cuadrante de una imagen CP.

12. Disposición según la reivindicación 10, en la que $m=7$ o $m=8$, $n=6$ y dichas regiones representan cada una un octavo de una imagen CP.



TÉCNICA ANTERIOR

Figura 1

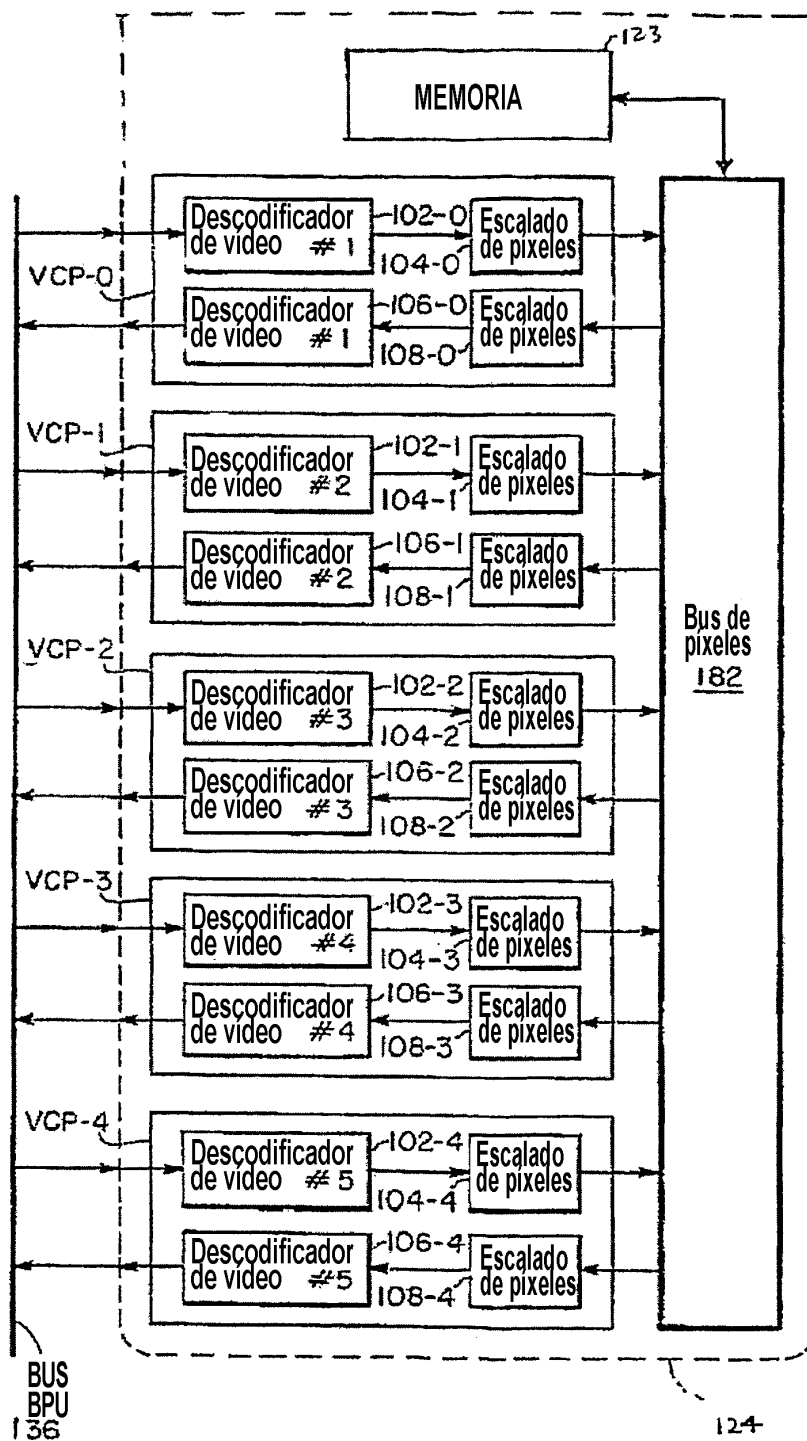
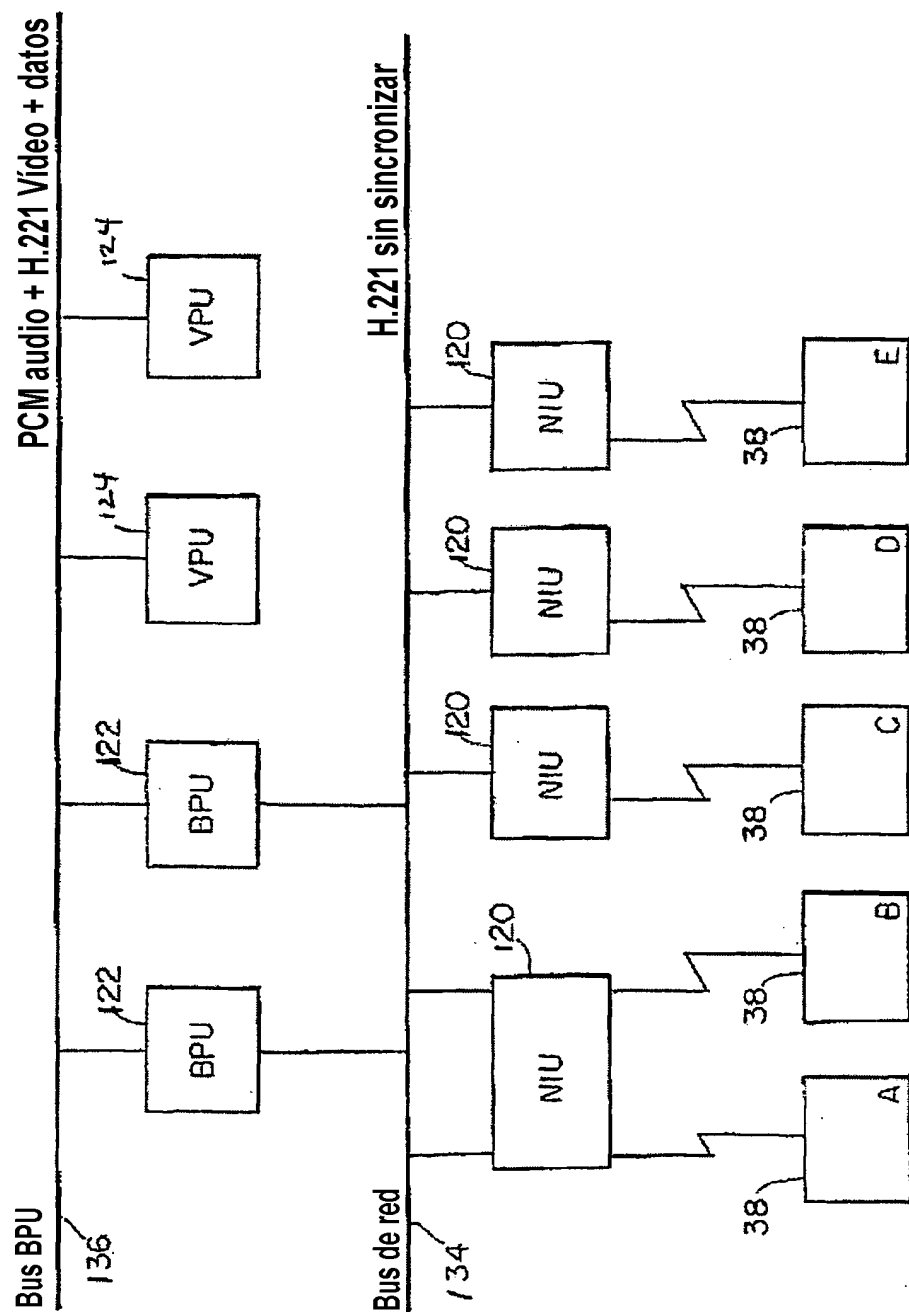


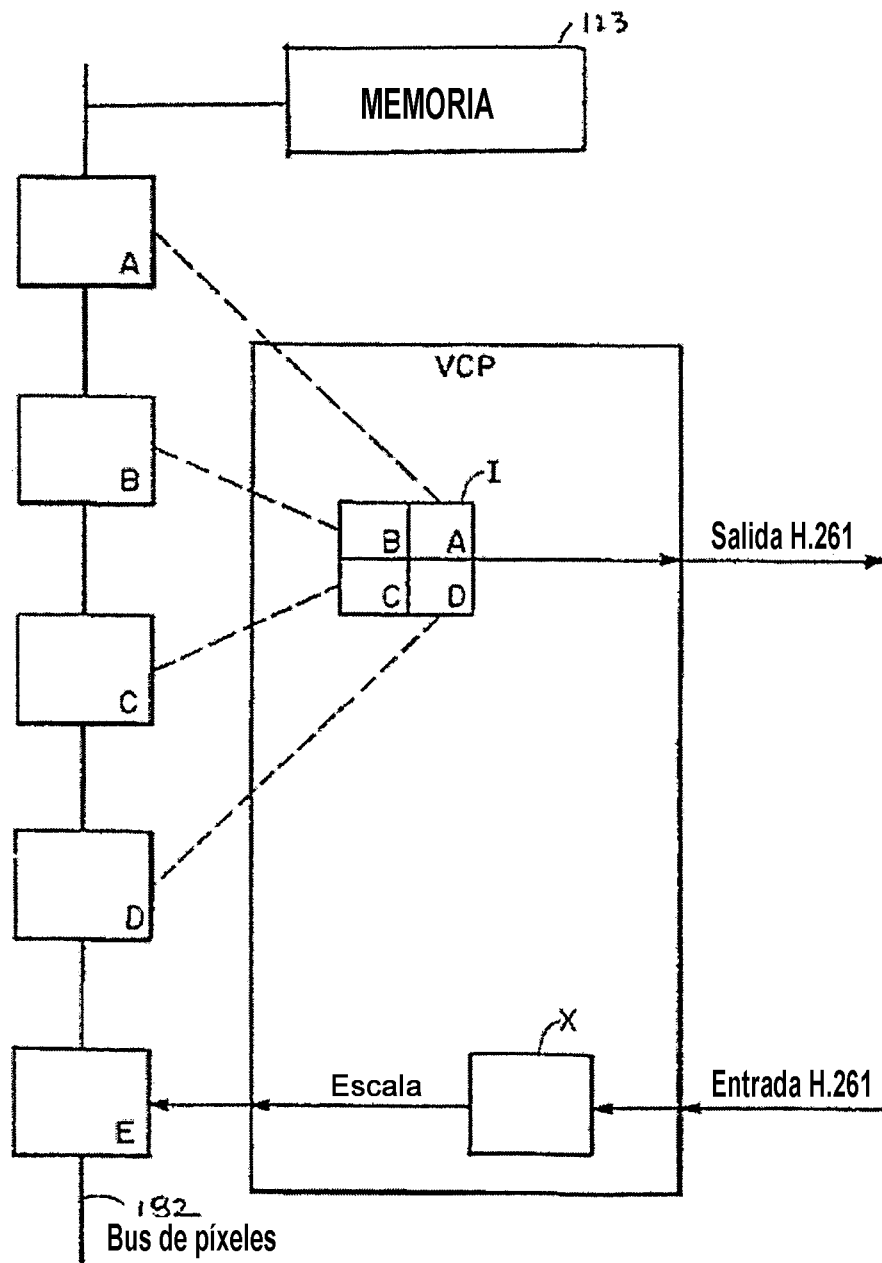
Figura 2

TÉCNICA ANTERIOR



TÉCNICA ANTERIOR

Figura 3



TÉCNICA ANTERIOR

Figura 4

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Figura 5: Disposición de los GOB en una imagen CIF



Figura 6: Estructura de capa GOB



Figura 7: Estructura de capa de macrobloque

Emplazamiento 1	Emplazamiento 2
Emplazamiento 3	Emplazamiento 5

Imagen CP reorganizada transmitida al
emplazamiento 4

Emplazamiento 5	

Imagen 2 CP

Emplazamiento 1	Emplazamiento 2
Emplazamiento 3	Emplazamiento 4

Imagen 1 CP

Figura 8

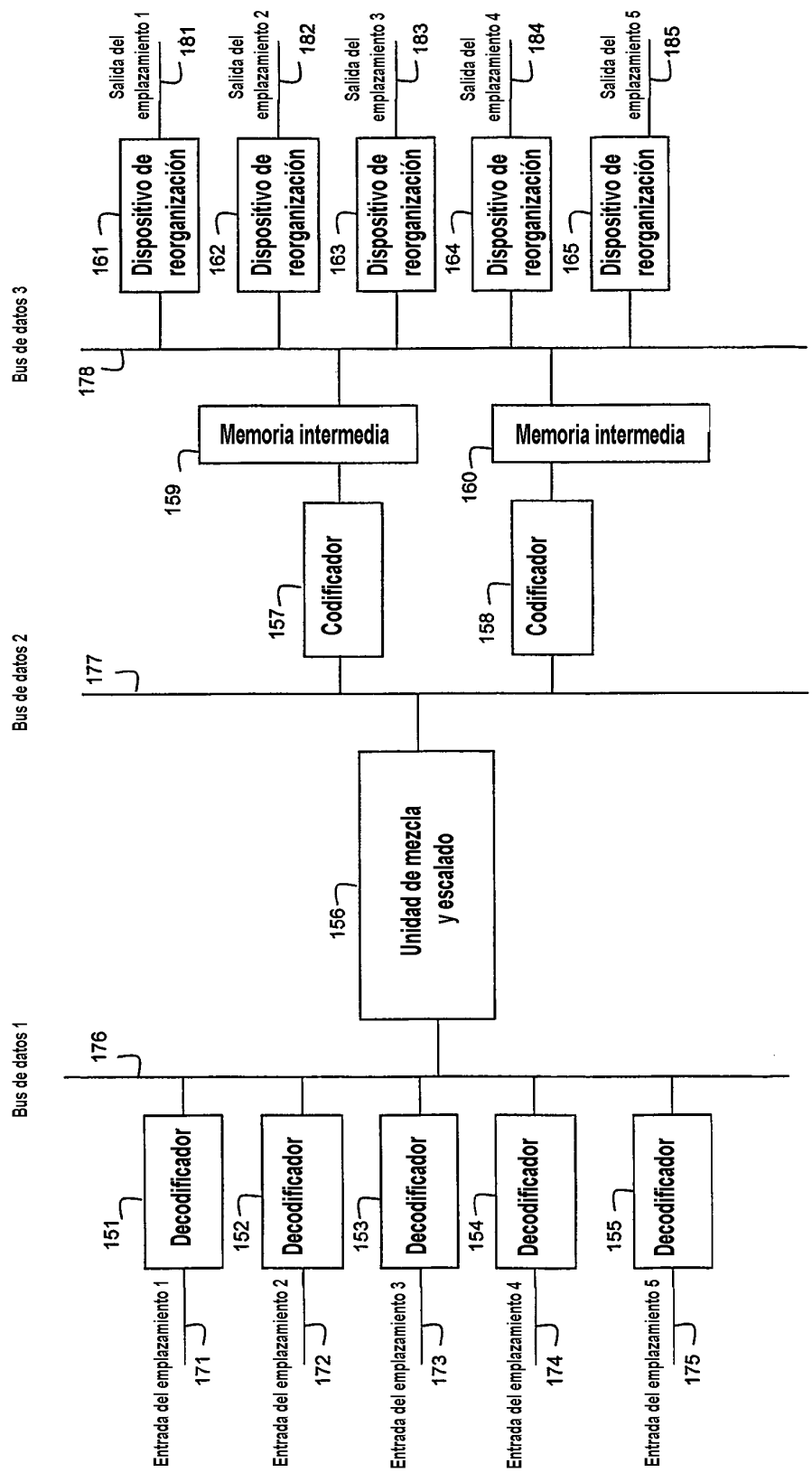


Figura 9

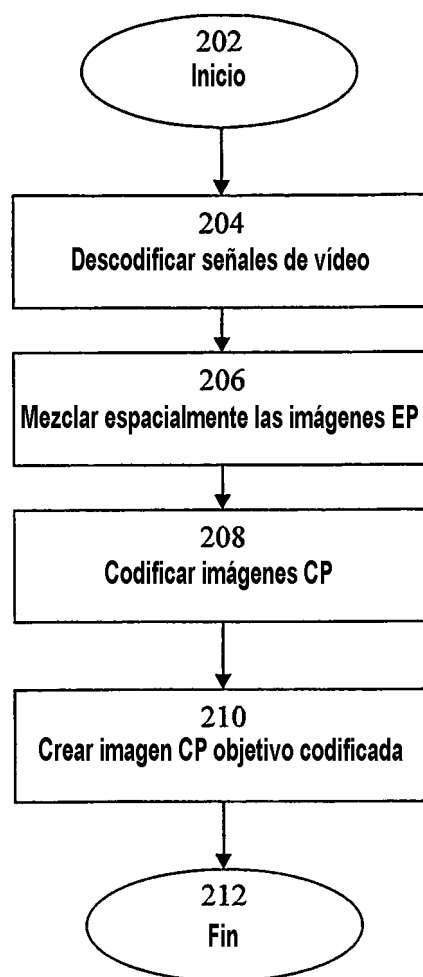


Figura 10