

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 807 259**

51 Int. Cl.:

<b>H04N 19/107</b>	(2014.01)	<b>H04N 19/33</b>	(2014.01)
<b>H04N 19/114</b>	(2014.01)	<b>H04N 19/587</b>	(2014.01)
<b>H04N 19/132</b>	(2014.01)	<b>H04N 19/87</b>	(2014.01)
<b>H04N 19/142</b>	(2014.01)		
<b>H04N 19/14</b>	(2014.01)		
<b>H04N 19/172</b>	(2014.01)		
<b>H04N 19/177</b>	(2014.01)		
<b>H04N 19/179</b>	(2014.01)		
<b>H04N 19/192</b>	(2014.01)		
<b>H04N 19/31</b>	(2014.01)		

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **16.12.2009 PCT/EP2009/067300**
- 87 Fecha y número de publicación internacional: **01.07.2010 WO10072636**
- 96 Fecha de presentación y número de la solicitud europea: **16.12.2009 E 09796697 (2)**
- 97 Fecha y número de publicación de la concesión europea: **22.04.2020 EP 2368367**

54 Título: **Sistema interactivo y procedimiento para transmitir imágenes clave seleccionadas de un flujo de vídeo sobre una red de ancho de banda bajo**

30 Prioridad:

**23.12.2008 FR 0807407**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**22.02.2021**

73 Titular/es:

**THALES (100.0%)  
Tour Carpe Diem - Place des Corolles, Esplanade  
Nord  
92400 Courbevoie, FR**

72 Inventor/es:

**CAPMAN, FRANÇOIS;  
RENAN, ERWANN y  
LE BARZ, CÉDRIC**

74 Agente/Representante:

**GONZÁLEZ PECES, Gustavo Adolfo**

ES 2 807 259 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

Sistema interactivo y procedimiento para transmitir imágenes clave seleccionadas de un flujo de vídeo sobre una red de ancho de banda bajo

5 La invención se refiere a un sistema interactivo y a un procedimiento para transmitir imágenes y vídeos a través de redes de velocidad restringida. Aborda con mayor precisión los problemas de transmisión de imágenes relevantes de alta resolución dentro de una secuencia de vídeo en redes de baja velocidad.

10 La invención puede usarse, entre otras cosas, en aplicaciones que implementan la norma definida conjuntamente por ISO MPEG y el grupo de codificación de vídeo de l'ITU-T conocido como H.264 o MPEG-4 AVC (codificación de vídeo avanzada), que es una norma de vídeo que proporciona una compresión más eficiente que las normas de vídeo anteriores, mientras que tiene una complejidad de implementación razonable y está orientado a aplicaciones de red. También se aplica en el contexto de los sistemas de transmisión de imágenes que utilizan la norma JPEG 2000 y el protocolo de comunicación JPIP (Protocolo Interactivo JPEG2000).

15 Uno de los problemas a resolver es el hecho de poder transmitir secuencias de vídeo de alta resolución a través de redes de velocidad restringida. De hecho, una secuencia de vídeo de alta resolución, incluso después de la compresión a través de un dispositivo de codificación de fuente adecuado, tiene una velocidad de bits útil que a menudo excede la capacidad del canal de transmisión, en particular la de las redes inalámbricas. Una solución consiste en seleccionar y transmitir dentro de la secuencia solo ciertas imágenes a una resolución más baja, pero luego surgen los problemas de seleccionar imágenes relevantes en la secuencia para transmitir casi toda la información útil contenida en el vídeo. Otro problema a resolver se refiere al procedimiento de transmisión que se implementará para transmitir y recuperar, en el lado del receptor, las imágenes de alta resolución. Además, la implementación de una forma de interactividad entre el operador remoto y el transmisor para seleccionar solo una parte del flujo de vídeo para la transmisión tiene ventajas notables para, por ejemplo, adaptar la transmisión a las necesidades del operador, y transmitir solo la información que se considere relevante. Finalmente, la preocupación por la complejidad de la implementación es un punto importante a tener en cuenta en la realización de una solución global y que satisface, en particular, las limitaciones en tiempo real inherentes a las aplicaciones multimedia interactivas.

20 En el resto del texto, la expresión "imágenes relevantes" o "imágenes clave" se referirá a un subconjunto de imágenes seleccionadas dentro de una secuencia de vídeo y que tienen una mayor prioridad desde el punto de vista del usuario final. En el contexto de una transmisión de dicha secuencia de vídeo a través de una red de baja velocidad, las imágenes relevantes son, por ejemplo, aquellas que presentan una diferencia significativa en el contenido entre ellas. En el contexto de una compresión de dicha secuencia de vídeo por un codificador de vídeo adecuado, las imágenes clave son también aquellas que se comprimirán de la manera más eficiente para garantizar su calidad final en términos de representación visual, una vez descomprimidas. Para este fin, un resumen de una secuencia de vídeo corresponde al conjunto de "imágenes relevantes" o "imágenes clave" de dicha secuencia.

#### Selección de imágenes relevantes en una transmisión de vídeo

35 El problema de seleccionar imágenes relevantes dentro de una secuencia de vídeo a menudo se trata en la técnica anterior por medio de soluciones que consisten en crear un resumen de dicha secuencia teniendo en cuenta el contenido general del vídeo. Por ejemplo, la solicitud de patente US2008/0232687 describe un procedimiento para seleccionar imágenes clave dentro de una secuencia de vídeo. Este procedimiento también permite la segmentación temporal de la secuencia para producir varias escenas. Este tipo de procedimiento no es adecuado para la transmisión en tiempo real de una transmisión de vídeo porque requiere el procesamiento de toda la secuencia para producir todas las imágenes clave asociadas. Por otro lado, las restricciones vinculadas a la transmisión de vídeo requieren que las imágenes se procesen sobre la marcha, el procedimiento de selección de imágenes clave solo se beneficia como un punto de entrada de la imagen actual y posiblemente de sus vecinas cercanas. temporalmente, en particular de las imágenes anteriores si se desea minimizar el retraso de transmisión.

#### Técnicas de compresión de transmisión de vídeo

45 Una secuencia de vídeo por su propia naturaleza tiene una redundancia estadística significativa tanto en el tiempo como en el espacio. El deseo de usar siempre de manera más eficiente el ancho de banda de los medios de transmisión por los que pasan estas secuencias y los objetivos de reducir el coste de su almacenamiento plantearon la cuestión de la compresión de vídeo desde el principio. Las técnicas convencionales de compresión de vídeo generalmente se pueden dividir en dos etapas. La primera tiene como objetivo reducir la redundancia espacial y, para ello, comprimir una imagen fija. La imagen se divide en primer lugar en bloques de píxeles (4x4 u 8x8 según los patrones MPEG-1/2/4), un paso en el dominio de la frecuencia y luego una cuantificación hace posible aproximar o eliminar las frecuencias altas a las que el ojo es menos sensible y, finalmente, los datos cuantificados se codifican entrópicamente. La segunda tiene como objetivo reducir la redundancia temporal. Esta técnica permite predecir una imagen a partir de una o más imágenes de referencia previamente procesadas dentro de la misma secuencia (predicción de movimiento). Esta técnica consiste en buscar en estas imágenes de referencia el bloque al que se desea aplicar una predicción, y solo se mantiene un vector de estimación de movimiento correspondiente al desplazamiento del bloque entre las dos imágenes y un error residual que permite refinar la representación visual.

Granularidad temporal en un patrón de vídeo

Un flujo de datos comprimidos de acuerdo con un procedimiento que permite la granularidad temporal, o en inglés "a temporally scalable compressed bit-stream", sigue un esquema de codificación de tipo jerárquico. Esta jerarquía en el esquema de codificación permite la definición de conjuntos de imágenes a las que se puede acceder por paso o resolución temporal. El primer nivel, llamado "resolución básica", es la secuencia mínima que permite la reconstrucción sin degradación de las tramas que la componen. Los otros niveles corresponden a refinamientos de esta secuencia básica. En general, los niveles de refinamiento tienen frecuencias de trama que son múltiples de la base, la relación entre estas frecuencias de trama se denomina factor de escala. En un ejemplo de una secuencia a 30 tramas por segundo siguiendo un esquema de codificación con una granularidad temporal del factor de escala igual a dos y escalonado en tres niveles, obtendríamos un primer nivel de resolución (resolución básica) correspondiente a un contenido de vídeo a 7,5 tramas por segundo. En este ejemplo, si se tiene acceso a los subconjuntos básicos, así como a los del primer nivel de refinamiento, entonces se puede acceder al contenido de vídeo a 15 tramas por segundo. Si se añade el último nivel de refinamiento, el contenido de vídeo en la resolución temporal original (30 tramas por segundo) es accesible. Se supone que cada uno de estos subconjuntos corresponde a una compresión eficiente de la información que contiene. La figura 1 muestra esquemáticamente un ejemplo de granularidad temporal. El nivel base (0) corresponde a la resolución temporal mínima que se transmite. En el contexto de la transmisión de vídeo, el flujo de código correspondiente al nivel base representa la información mínima que el receptor debe recibir y, por lo tanto, debe comprimirse para garantizar la menor pérdida posible durante la transmisión. Típicamente, las imágenes contenidas en esta resolución temporal base se codifican de forma independiente. Las resoluciones temporales de los niveles superiores (nivel de refinamiento 1 y 2) pueden codificarse implementando una predicción con respecto a las imágenes de la resolución base (0). Es posible una predicción entre imágenes que pertenecen a la resolución base, por otro lado, las imágenes contenidas en la resolución temporal base no pueden predecirse a partir de una imagen contenida en otro nivel de refinamiento.

Patrón JPIP

El patrón JPIP (JPEG 2000 Interactive Protocol) define un protocolo dedicado a la transmisión progresiva de imágenes conforme al patrón JPEG 2000. Permite aprovechar los diferentes niveles de granularidad propuestos por la norma JPEG 2000 (granularidad espacial, granularidad en resolución o, en inglés "resolution", y en calidad). De hecho, después de una solicitud formulada por el operador, solo se transmite la información necesaria para satisfacer dicha solicitud y, de forma progresiva, en términos de calidad. El uso del protocolo JPIP combinado con el patrón JPEG 2000 hace posible no retransmitir la información ya transmitida. Por lo tanto, los recursos en términos de rendimiento de la transmisión y la complejidad de procesamiento en ambos lados de la cadena de transmisión se reducen. Además, dado que el envío de información es jerárquico, es posible ver rápidamente una parte de la imagen a baja calidad, calidad que aumenta progresivamente a medida que se recibe nueva información. Este patrón se puede utilizar para realizar una transmisión interactiva con optimización de velocidad de bits en el caso de una transmisión de imágenes JPEG2000, pero no permite implementar el mismo tipo de proceso para una transmisión de vídeo que se basaría en un patrón diferente. En particular, este protocolo no tiene en cuenta la selección de una imagen por parte del operador dentro de un flujo de vídeo transmitido.

La técnica anterior, como se ha descrito anteriormente, no permite resolver un cierto número de problemas, en particular la transmisión de información de alta resolución en una red que tiene una velocidad limitada. El acceso interactivo a una imagen, o al área de una imagen, dentro de un flujo de vídeo transmitido en tiempo real tampoco está cubierto por el estado de la técnica.

La técnica anterior también comprende los documentos.WO2007/078801 et WO 2004/114672.

Para responder a las limitaciones de la técnica anterior, la invención propone un nuevo abordaje que consiste en trabajar solo en imágenes subresueltas y submuestreadas temporalmente de manera inteligente para minimizar la redundancia y adaptarse al ancho de banda requerido. La solución propuesta también permite el análisis de esta secuencia subresuelta interactivamente a través de solicitudes hechas por un operador remoto. La presente invención es compatible con los patrones siguientes. El patrón H.264 está definido por la norma ISO/IEC 14496-10. El patrón JPEG 2000 está definido por la norma ISO/IEC 15444-1. El patrón JPIP está definido por la norma ISO/IEC 15444-9.

Con este fin, el objeto de la invención es un procedimiento de transmisión en línea de una secuencia de vídeo de alta resolución compuesta por una sucesión de imágenes T, caracterizada porque comprende al menos una etapa de selección de imágenes relevantes que comprenden al menos las etapas siguientes:

- Etapa 1: Dividir cada imagen que se transmitirá en el instante t en un número N de zonas,
- Etapa 2: Para cada área n determinada en la etapa 1, calcular un valor  $v_n(t)$  representativo del contenido de dicha área,
- Etapa 3: Para cada imagen I(t) que se va a transmitir, generar un vector  $\mathbf{V}_N(t)$  de tamaño N representativo del contenido de dicha imagen y que contiene los valores obtenidos en la etapa 2 para cada una de las áreas determinadas en la etapa 1  $\mathbf{V}_N(t) = [v_0(t), \dots, v_{N-1}(t)]^T$ ,  $[\cdot]^T$  siendo el operador de transposición para un vector.
- Etapa 4: Calcular un coeficiente de correlación normalizado  $\alpha$  entre el vector de referencia  $\mathbf{R}_N(k) = \mathbf{V}_N(t_1) = [r_0(k), \dots, r_{N-1}(k)]^T$  determinado para una imagen previamente seleccionada en el instante  $t_1$   $r^{sel}(k-1) = I(t_1)$  y dicho

$V_N(t)$  calculado para la imagen actual  $I(t)$  mediante el siguiente cálculo:  $\alpha = \frac{\mathbf{R}_N^T(k-1)\mathbf{V}_N(t)}{\sqrt{\mathbf{R}_N^T(k-1)\mathbf{R}_N(k-1)\sqrt{\mathbf{V}_N^T(t)\mathbf{V}_N(t)}}$ ,

donde  $k$  es el índice temporal de una imagen seleccionada.

➤ Etapa 5: Tomar una decisión sobre la selección (o no) para la transmisión de la imagen actual  $I(t)$  comparando el valor del coeficiente de correlación normalizado  $\alpha$  con uno o más umbrales con valores fijados a priori.

5 En una realización alternativa de la invención, el valor  $v_n(t)$  calculado en la etapa 2 se obtiene calculando uno de los criterios derivados de un histograma que contiene los valores de luminancia o crominancia de los píxeles del área  $n$  de la imagen  $I(t)$ , dichos criterios incluyen: la energía, la varianza, el coeficiente de amortiguamiento (o curtosis), el coeficiente de asimetría (o sesgo, Skewness) o el centro de gravedad de dicho histograma, o realizando una combinación lineal de al menos dos de estos criterios.

10 En una realización alternativa de la invención, la etapa de selección de imágenes relevantes comprende además las etapas siguientes:

➤ Sobre la base de cada par de vectores  $(R_N(k-1), V_N(t))$ , obtenidos para el par de imágenes  $(I^{sel}(k-1), I(t))$ , calcular un segundo vector  $\Delta V_N(t) = (\Delta v_0(t), \dots, \Delta v_N(t))$  tal que  $\Delta v_n(t) = |v_n(t) - r_n(k-1)|^\beta$ , siendo  $\beta$  un número entero positivo dado.

15 ➤ Calcular la norma  $\|\Delta V_N(t)\|$  del vector  $\Delta V_N(t)$  y dividir dicho vector por dicha norma para obtener un tercer vector

$$\Delta'V_N(t) = \frac{\Delta V_N(t)}{\|\Delta V_N(t)\|} = (\Delta v'_0(t), \dots, \Delta v'_N(t))$$

➤ Calcular el siguiente criterio adicional para determinar si existe una distribución uniforme o localizada de las diferencias entre las dos imágenes consideradas:

20 ○

$$DC = (\Delta V_N^{MAX} - \frac{1}{N}) / (1 - \frac{1}{N}),$$

con

$$\Delta V_N^{MAX} = \text{MAX}(\Delta v'_n(t)), 0 \leq n < N,$$

25 ○ aplicar la siguiente prueba, si  $DC < S_{DC}$ , las diferencias entre las dos imágenes consideradas se distribuyen uniformemente, de lo contrario se localizan en un área dada de dichas imágenes.

En una realización alternativa de la invención, la etapa 5 se implementa llevando a cabo una prueba de comparación entre el coeficiente de correlación  $\alpha$  y un umbral predefinido  $S$ , un entero estrictamente positivo, de la siguiente forma:

➤ si  $\alpha < S$ , la imagen actual  $I(t)$  está seleccionada, el vector de referencia se actualiza  $\mathbf{R}_N(k) = \mathbf{V}_N(t)$  y el índice de imagen seleccionado se incrementa  $k = k + 1$

30 ➤ de lo contrario, la imagen no se conserva como una imagen relevante.

En una realización alternativa de la invención, la etapa de selección de imágenes relevantes comprende las etapas complementarias siguientes:

➤ Definir un contador de imágenes  $cpt\_I$ , inicializado a 1 cuando se selecciona una imagen relevante en la etapa 5 y se incrementa en 1 con cada nueva imagen actual  $I(t)$  recibida, así como un valor máximo de dicho contador  $cpt\_I\_max$ ,

35 ➤ Para cada imagen, realizar la siguiente prueba de comparación: si  $\alpha \geq S$  y  $cpt\_I \geq cpt\_I\_max$ , seleccionar dicha imagen como relevante.

En una realización alternativa de la invención, la etapa de selección de imágenes relevantes comprende las etapas complementarias siguientes:

➤ Definir un contador de imágenes  $cpt\_I$ , inicializado a 1 cuando se selecciona una imagen relevante en la etapa 5 y se incrementa en 1 con cada nueva imagen actual recibida  $I(t)$  y un valor mínimo de dicho contador  $cpt\_I\_min$ ,

40 ➤ Para cada imagen, realizar la siguiente prueba de comparación: si  $\alpha < S$  y  $cpt\_I < cpt\_I\_min$ , no conservar dicha imagen como relevante.

En una realización alternativa de la invención, la etapa de selección de imágenes relevantes comprende las etapas complementarias siguientes:

45

- Definir un número  $N_s$  deseado de imágenes seleccionadas por segundo,
- Definir un intervalo de tiempo de duración igual a  $T_h$  para la actualización de dicho umbral  $S$ .
- En cada intervalo de tiempo de duración igual a  $T_h$ , calcular el número de imágenes  $N$  seleccionadas por el procedimiento durante dicho intervalo de tiempo anterior de duración igual a  $T_h$ ,
- En cada intervalo de tiempo de duración igual a  $T_h$ , modificar el valor del umbral  $S$  de la siguiente manera:

Si

$$N < N_s, S = S - \delta S$$

Si  $N > N_s$ ,  $S = S + \delta S$ , donde  $\delta S$  es un valor de número entero positivo incremental inicialmente fijo y/o ajustado dinámicamente.

10 En una realización alternativa de la invención, la etapa 5 consiste en realizar una prueba de comparación entre el coeficiente de correlación  $\alpha$  y un número estrictamente positivo  $J$  de umbrales predefinidos  $S_j$ ,  $1 \leq j < J$ , de la siguiente manera:

- Si  $0 \leq \alpha < S_1$ , la imagen actual se selecciona con un grado máximo de prioridad,
- Si  $S_j \leq \alpha < S_{j+1}$ , para  $1 \leq j < J-1$ , la imagen actual se selecciona con un grado medio de prioridad  $j$ ,
- Si  $\alpha \geq S_J$ , la imagen no se conserva como relevante,  $S_j$  es un número entero estrictamente positivo.

En una realización alternativa de la invención, la etapa 5 se lleva a cabo del siguiente modo:

- Ordenar las imágenes  $T$  de dicha secuencia de vídeo en orden ascendente del valor de su coeficiente de correlación  $\alpha$ ,
- Agrupar las imágenes ordenadas en  $T$  en  $J + 1$  grupos jerarquizados en función del valor de  $\alpha$ , comprendiendo cada uno de dichos grupos de imágenes el mismo número de imágenes y correspondiendo a un grado medio de prioridad  $j'$

En una realización alternativa de la invención, la secuencia de vídeo de alta resolución se comprime antes de la transmisión a través de las siguientes etapas:

- realizar un submuestreo espacial de dicho flujo de vídeo de alta resolución para obtener una secuencia de vídeo de menor resolución,
- aplicar la etapa de seleccionar imágenes relevantes a la secuencia de vídeo de menor resolución obtenida,
- realizar una compresión de dicha secuencia mediante un codificador de vídeo adecuado para generar una estructura particular de grupo de imágenes, en la que las imágenes relevantes seleccionadas definen:
  - una resolución temporal, que contiene el conjunto de imágenes relevantes seleccionadas en el caso en el que la etapa de seleccionar imágenes relevantes libera solo un grupo de imágenes relevantes, todas con el mismo grado de prioridad,
  - varias resoluciones temporales, cada una con las imágenes relevantes seleccionadas con un grado diferente de prioridad, en el caso en que la etapa de seleccionar imágenes relevantes libera varios grupos de imágenes relevantes con diferentes grados de prioridad.

35 En una realización alternativa de la invención, el codificador de vídeo respeta el patrón H.264.

En una realización alternativa de la invención, la secuencia de vídeo de alta resolución que se va a transmitir se guarda en una unidad de almacenamiento y se selecciona una imagen o área de imagen de dicha secuencia de vídeo para transmitir o retransmitir de acuerdo con las siguientes etapas:

- realizar un sobremuestreo espacial de dicha imagen para obtener una imagen de alta resolución,
- calcular la diferencia entre la imagen sobremuestreada y la correspondiente almacenada previamente para obtener un residuo,
- comprimir dicho residuo a través de un codificador de imagen para producir un flujo comprimido que comprenda al menos paquetes que contengan información sobre las frecuencias altas de la imagen y otros paquetes que contengan información sobre las frecuencias bajas de dicha imagen,
- transmitir dicho flujo comprimido a un operador remoto.

En una realización alternativa de la invención, la imagen que se va a retransmitir o la imagen asociada con dicha área de imagen que se va a retransmitir se compara con sus imágenes temporalmente vecinas dentro de dicha secuencia de vídeo de alta resolución para determinar cuál es la menos borrosa de acuerdo con las siguientes etapas:

- calcular en la imagen o el área de imagen seleccionada un criterio representativo de la nitidez de dicha imagen o área de imagen, exhibiendo dicho criterio de nitidez un primer extremo en el que dicha imagen o área de imagen es perfectamente nítida y un segundo extremo opuesto al primero cuando dicha imagen o el área de la imagen está totalmente borrosa,
- realizar el mismo cálculo en las  $M$  imágenes temporalmente vecinas, siendo  $M$  un valor predefinido,

- seleccionar la imagen más nítida según el criterio anterior y retransmitir esta imagen o el área de imagen asociada, en lugar de la seleccionada inicialmente.

5 En una realización alternativa de la invención, el codificador de imagen es adecuado para definir el orden de envío de los paquetes que forman el flujo comprimido en función, al menos, de una información sobre la velocidad de bits disponible para transmisión y fijada a priori.

En una realización alternativa de la invención, el codificador de imágenes respeta el patrón JPEG2000.

En una realización alternativa de la invención, la retransmisión de una imagen o área de imagen se realiza mediante una solicitud del operador remoto.

10 En una realización alternativa de la invención, las solicitudes y las respuestas a dichas solicitudes se implementan a través del patrón JPIP.

La invención también se refiere a un sistema interactivo para transmitir secuencias de vídeo de alta resolución, caracterizado porque comprende un emisor y un receptor que comprende medios para ejecutar las etapas del procedimiento descrito anteriormente.

15 Otras características y ventajas del procedimiento y del dispositivo según la invención aparecerán mejor al leer la siguiente descripción de una realización de ejemplo dada a modo de ilustración y de ningún modo limitante, adjunta a las figuras que representan:

- La figura 1 es una ilustración de la granularidad temporal en un flujo de vídeo,
- la figura 2 es un diagrama de bloques del sistema para la transmisión interactiva de imágenes y vídeos a través de una red de baja velocidad según la invención,
- la figura 3 es una ilustración del procedimiento para seleccionar imágenes relevantes según la invención,
- la figura 4 es una ilustración de una realización alternativa del procedimiento para seleccionar imágenes relevantes según la invención,
- la figura 5 es una ilustración de un ejemplo de implementación en forma de un esquema de granularidad temporal con una restricción en las imágenes clave,
- la figura 6 es una ilustración de la asignación de velocidad y de la distorsión para cada resolución de tiempo.
- la figura 7 es una ilustración de una región de interés en una imagen.

20 La figura 2 representa un esquema del sistema global según la invención. En el lado izquierdo de la figura se ilustra el sistema de adquisición, compresión, almacenamiento y servidor de datos. La parte derecha de la figura corresponde al operador o experto remoto que realiza las solicitudes al transmisor. Las flechas punteadas significan que hay transmisión (ya sea datos de vídeo o imagen, o solicitudes del usuario) a través de una red con una velocidad restringida conocida.

30 Una secuencia de vídeo de alta resolución, 1, de un sensor, 10, se muestrea en primer lugar espacialmente, 11, para reducir la resolución inicial e indirectamente la velocidad de bits útil de la secuencia a transmitir. Al mismo tiempo, la secuencia, 1, también se almacena localmente en una unidad de almacenamiento 12. Un procedimiento, según la invención, para seleccionar imágenes relevantes, 13, recupera el flujo submuestreado, 2, para producir un resumen de la secuencia de vídeo, en tiempo real. Este resumen permite determinar cuáles son las imágenes clave, 3, dentro de la secuencia. Dicha secuencia de imágenes clave, 3, así obtenida es una serie de imágenes cuyo contenido entre cada imagen difiere significativamente. Para cada imagen transmitida, el procedimiento de selección, 13, determina si es una imagen clave y proporciona esta indicación al codificador de vídeo 14. Dicho codificador, por lo tanto, tiene en su entrada un flujo submuestreado, 2, y una indicación de imagen clave, 3 y utiliza estas dos fuentes de información para producir dos flujos de vídeo comprimidos. El primer flujo, 4, corresponde a la resolución temporal base o secuencia de baja resolución, que necesariamente debe contener las imágenes clave determinadas previamente. El segundo, 5, corresponde al flujo de vídeo completo, submuestreado espacialmente y luego codificado. Al menos uno de estos dos flujos se transmite, a través de una red de velocidad restringida, al descodificador de vídeo remoto, 18 que realiza la descodificación y genera una secuencia de vídeo descomprimida que puede visualizarse. La elección de transmitir uno u otro de los flujos disponibles depende del operador y del ancho de banda disponible. El flujo, 4, correspondiente a la resolución temporal base se transmitirá con prioridad. Los dos flujos comprimidos también se almacenan en una unidad de almacenamiento, 15, antes de ser transmitidos. Cuando se utiliza un codificador de vídeo de tipo H.264, solo se almacena el flujo prioritario, 4.

45 En el lado de la recepción, un operador, 19, puede hacer varios tipos de solicitudes al servidor de vídeo. Una primera solicitud, 8, permite al operador solicitar la retransmisión de una imagen específica o de todas las imágenes clave en la resolución original (asociada con el sensor 10). Esta solicitud también se puede llevar a cabo en parte de una imagen, para acercarse a un elemento preciso que el operador desea mostrar en alta resolución. Por lo tanto, la solicitud 8 se transmite a una etapa 17 que realiza la recuperación de la imagen, del grupo de imágenes o de la parte de imagen requerida. Por lo tanto, la imagen recuperada se descodifica desde el flujo almacenado en la unidad de almacenamiento 15 y luego se sobremuestra espacialmente, 16, para volver a la resolución original. La diferencia, 25, entre la imagen original de alta resolución (o parte de una imagen) previamente almacenada en la unidad, 12, y la imagen sobremuestreada (o parte de la imagen), 6, se calcula para obtener un residuo, 9.

Dicho residuo obtenido se comprime luego a través de un codificador de imagen 21, respetando, por ejemplo, el patrón

JPEG2000. Es este residuo el que luego se transmite al operador mediante, por ejemplo, el uso del protocolo JPIP bidireccional, 7. En la recepción, el flujo se descodifica gradualmente a través de un descodificador de imagen, 22, el residuo descodificado así obtenido se comunica a un módulo de restauración de imagen, 23, que a partir de este residuo y de la primera imagen de baja calidad y baja resolución que ha recibido (que le permitió hacer su solicitud de información adicional) proporciona una imagen descomprimida completa al operador. De este modo, se podrá reconstruir gradualmente la parte de la imagen que se desea analizar hasta su resolución completa y máxima calidad.

El operador 19 puede interactuar con el sistema de acuerdo con la invención de acuerdo con varios escenarios. Por ejemplo, se requiere la retransmisión en alta resolución de una imagen o de un área de imagen ya transmitida en baja resolución. El operador 19 también puede requerir la transmisión del flujo de vídeo o de las imágenes de alta resolución entre dos imágenes ya transmitidas en baja resolución.

El procedimiento, de acuerdo con la invención, para seleccionar imágenes relevantes, 13, permite determinar, en tiempo real y durante la transmisión del flujo de vídeo, las imágenes clave de la secuencia de vídeo. El conjunto de estas imágenes representa un resumen de la información de la secuencia inicial de alta resolución. En el resto de la descripción, se utilizarán las siguientes notaciones:

- T es el número total de imágenes de la secuencia de vídeo procesada, t es el índice temporal asociado con una imagen de dicha secuencia,
- N es el número de áreas de la imagen determinadas por el procedimiento, n es el índice de un área,
- K es el número de imágenes seleccionadas por el procedimiento entre las imágenes T disponibles, k es el índice de tiempo asociado con una imagen entre el conjunto de imágenes seleccionadas.
- Se anota la imagen actual de la secuencia de vídeo procesada  $I(t)$

El procedimiento para seleccionar imágenes relevantes implica las etapas siguientes:

Etapla 1: Cada imagen  $I(t)$  recibida en el momento t se divide en un número N de áreas que pueden variar, normalmente N es, por ejemplo, un número entero mayor que 9.

Etapla 2: Cada área n está asociada con un valor  $v_n(t)$  representativo del contenido de esta área de la imagen y obtenido según un criterio preciso.

Etapla 3: Un vector  $V_N(t)$  de tamaño N representativo del contenido de la imagen  $I(t)$  se genera concatenando los valores obtenidos en la etapa 2 para cada área de la imagen:

$$V_N(t) = [v_0(t), \dots, v_{N-1}(t)]^T$$

o  $[\cdot]^T$  es el operador de transposición. Cuando se selecciona una imagen actual del índice t mediante el procedimiento de selección de imágenes relevantes, se actualiza un vector de referencia:

$$R_N(k) = V_N(t) = [r_0(k), \dots, r_{N-1}(k)]^T$$

La imagen seleccionada se anota  $I^{sel}(k) = I(t)$  y para cada nueva imagen seleccionada mediante el procedimiento de acuerdo con la invención, k se incrementa en 1 ( $k = k + 1$ ).

Etapla 4: Se calcula un coeficiente de correlación normalizado  $\alpha$  entre el vector de referencia  $R_N(k-1)$  determinado para la imagen seleccionada previamente  $I^{sel}(k-1)$  y el vector  $V_N(t)$  representativo del contenido de la imagen actual  $I(t)$

$$\alpha = \frac{\mathbf{R}_N^T(k-1)\mathbf{V}_N(t)}{\sqrt{\mathbf{R}_N^T(k-1)\mathbf{R}_N(k-1)} \cdot \sqrt{\mathbf{V}_N^T(t)\mathbf{V}_N(t)}}$$

o mejor

$$\alpha = \frac{\sum_{n=0}^{N-1} r_n(k-1)v_n(t)}{\sqrt{\sum_{n=0}^{N-1} r_n(k-1)r_n(k-1)} \cdot \sqrt{\sum_{n=0}^{N-1} v_n(t)v_n(t)}}$$

Etapla 5: Se realiza una prueba de comparación entre el coeficiente de correlación  $\alpha$  y un umbral predefinido S. Si  $\alpha < S$ , la imagen actual  $I(t)$  está seleccionada, el vector de referencia se actualiza  $R_N(k) = V_N(t)$  y el índice k se incrementa en 1. Por el contrario, la imagen no se retiene como imagen relevante.

El procedimiento implementado mediante las etapas 1 a 5 descritas anteriormente se ilustra en la figura 3. El procedimiento se inicializa seleccionando la primera imagen  $I^{sel}(0) = I(t=0)$  de la secuencia y luego una nueva imagen  $I^{sel}(k) = I(t_1)$  se selecciona cada vez que su criterio asociado  $\alpha$  es menor que el umbral S.

En una realización alternativa, es posible definir J umbrales  $S_j$  ( $1 \leq j \leq J$  y  $S_j < S_{j+1}$ ) para seleccionar J + 1 grupos de

imágenes. Dichos grupos obtenidos se clasificarán según su grado de prioridad. Por ejemplo, para  $J = 3$ , el grupo de imágenes tal que  $0 \leq \alpha < S_1$  es la mayor prioridad, el segundo grupo de imágenes en el orden de prioridad es el obtenido para  $S_1 \leq \alpha < S_2$  y así sucesivamente. Otra posibilidad para obtener grupos de imágenes  $J + 1$  consiste en dividir todas las imágenes disponibles en grupos  $J + 1$ , cada uno con un número igual de imágenes, la distribución de las imágenes en cada grupo se realiza priorizando las imágenes de acuerdo con el valor de su criterio asociado  $\alpha$ . La figura 4 ilustra esta realización alternativa en el caso en el que 12 imágenes están inicialmente disponibles. El procedimiento de selección de acuerdo con la invención define tres grupos de imágenes. En un caso, los umbrales  $S_1$ ,  $S_2$  y  $S_3$  se utilizan para determinar la distribución de imágenes en los tres grupos. El grupo 1 contiene las imágenes (representadas por su índice de tiempo)  $\{1,2\}$ , el grupo 2 contiene las imágenes  $\{3,4,5,6,7,8\}$  y el grupo 3 contiene las imágenes  $\{9,10,11,12\}$ . En el segundo caso, las imágenes se clasifican de acuerdo con el valor de su criterio  $\alpha$  y se agrupan en tres grupos, cada uno con la mayor cantidad de elementos. El grupo 1 contiene imágenes  $\{1,2,3,4\}$ , el grupo 2 contiene imágenes  $\{5,6,7,8\}$  y el grupo 3 contiene imágenes  $\{9,10,11,12\}$ .

El valor  $v_n(t)$  representativo del contenido de un área  $n$  de una imagen recibida en el tiempo  $t$ , así como el vector representativo de dicha imagen  $V_N(t)$  se puede calcular de acuerdo con diferentes criterios como se describe en el resto de este párrafo.

Un criterio posible es un criterio basado en la luminancia promedio de la imagen. En este caso, cada valor  $v_n(t)$  es igual al valor promedio de la luminancia calculada sobre el área  $n$  de la imagen  $I(t)$ . Este criterio se puede utilizar, pero su rendimiento depende en gran medida del paso de cuantificación utilizado para la función de distribución de la luminancia. El procedimiento según la invención permite mejorar el rendimiento de este criterio modelando la distribución de la luminancia más finamente usando, por ejemplo, un modelo basado en varias funciones gaussianas. El modelo utilizado para esto es un modelo de mezcla gaussiana, o "Gaussian Mixture Model" (GMM) en inglés, que conocen los expertos en la materia. La estimación de este modelo se puede hacer con un algoritmo de maximización de la esperanza conocido por el término en inglés "Expectation-Maximisation" (EM). Los coeficientes obtenidos a través de este algoritmo pueden usarse directamente como valores  $v_n(t)$  sin ponderación particular, o ponderados utilizando los pesos asociados con cada gaussiano y determinados por el algoritmo EM. Finalmente, el número de gaussianos utilizados se puede fijar a priori o estimarse dinámicamente mediante un criterio de información conocido. Sin embargo, dado que el procedimiento GMM es complejo de implementar, la invención propone modelar la distribución de la luminancia mediante un histograma. Dicho histograma se calcula de manera convencional, utilizando la resolución inicial de la luminancia, es decir, el número de valores que la luminancia puede tomar en una imagen o una resolución más baja al agrupar varios valores vecinos. Una vez que se ha construido este histograma, se pueden derivar varios criterios, como, por ejemplo, energía, varianza, centro de gravedad, coeficiente de aplanamiento (mejor conocido como kurtosis) o coeficiente de asimetría (conocido con el término Sesgo (en inglés, Skewness)). Estos parámetros permiten caracterizar la distribución de la luminancia más o menos finamente.

El valor  $v_n(t)$  representativo de un área de la imagen puede ser diverso. Depende de la aplicación. Se puede basar en la luminancia, que habitualmente se indica  $Y$ , en las crominancias, que habitualmente se indican como  $Cr$  y  $Cb$ , o en una combinación lineal de la luminancia y las crominancias. Los coeficientes de la combinación lineal se determinan para representar mejor el contenido del área  $n$  de la imagen para una aplicación específica. Del mismo modo, el valor  $v_n(t)$ , en lugar de basarse en los componentes  $YCrCb$  de la imagen, puede basarse en los componentes de los diferentes espacios de color conocidos por el experto (RGB, LUV, Lab, ...). En general, este valor representativo también puede resultar de una transformación aplicada a los componentes de la imagen. En este caso, esta transformación tiene el objetivo de resaltar una característica de la imagen de interés para la aplicación prevista, como, por ejemplo, la textura.

Finalmente, un criterio adicional para el vector representativo  $V_N(t)$  se puede configurar para determinar si la diferencia estimada (a través del vector  $V_N(t)$ ) entre la imagen actual y la imagen de referencia se distribuye uniformemente en toda la imagen o se encuentra en un área de esta imagen solamente. A partir de uno de los criterios propuestos anteriormente, un vector representativo  $V_N(t)$  está disponible para la imagen actual, así como un vector de referencia  $R_N(k - 1)$  correspondiente a la última imagen seleccionada. Cada componente de un vector representa un área de la imagen. Un tercer vector  $\Delta V_N(t) = (\Delta v_0(t), \dots, \Delta v_N(t))$  se puede calcular, con  $\Delta v_n(t) = |v_n(t) - r_n(k-1)|^\beta$ ,  $\beta \geq 0$  el valor absoluto, posiblemente elevado a una potencia  $\beta$ , de las diferencias de los componentes de los vectores. El valor de la potencia  $\beta$  se puede determinar por simulación. El vector  $\Delta V_N(t)$  así obtenido se normaliza. La distribución de las diferencias entre las dos imágenes se puede interpretar observando la homogeneidad de los componentes del vector normalizado: si la distribución de las diferencias es perfectamente uniforme, todos los componentes del vector son iguales a  $1/N$ , donde  $N$  es el tamaño del vector (igual al número de áreas de la imagen considerada), y si las diferencias están perfectamente localizadas en un área única, todos los componentes del vector son cero, excepto uno. A continuación, se define un criterio estándar que permite proporcionar información sobre la homogeneidad o la localización de las diferencias entre dos imágenes consecutivas. Dicho criterio utiliza dos valores:

➤

$$\Delta V_N^{MAX} = MAX(\Delta v_n(t)), 0 \leq n < N$$



$$DC = (\Delta V_N^{MAX} - \frac{1}{N}) / (1 - \frac{1}{N})$$

Si hay una distribución uniforme de las diferencias entre la imagen actual y la imagen de referencia,  $\Delta V_N^{MAX} = \frac{1}{N}$  y  $DC = 0$

Si hay diferencias en la localización de un área determinada,

5 
$$\Delta V_N^{MAX} = 1 \text{ et } DC = 1.$$

El criterio adicional consiste en comparar el valor de la variable DC con un umbral predefinido,  $S_{DC}$  comprendido entre 0 y 1 y cuyo valor se determina de acuerdo con el escenario de uso. Si  $DC < S_{DC}$ , se concluye con la distribución uniforme de las diferencias entre la imagen actual y la imagen de referencia, en el caso contrario, se concluye con la localización de las diferencias en un área determinada.

10 Este criterio adicional se puede implementar en particular en el caso de usar un sensor en una posición fija, para lo cual se intenta seleccionar una imagen en una diferencia localizada, incluso si el criterio inicial no activa la selección.

En una realización alternativa, es posible forzar al sistema a seleccionar una imagen incluso si el coeficiente de correlación  $\alpha$  permanece por encima del umbral S utilizando un contador de imágenes  $cpt\_I$ . Este contador se incrementa para cada imagen recibida y se restablece a uno cuando se selecciona una imagen relevante. El procedimiento consiste en fijar un valor máximo  $cpt\_I\_max$  para dicho contador. Cuando el contador alcanza este valor máximo, la imagen actual  $I(t)$  se selecciona incluso si el criterio definido en la etapa 5 no se verifica. Del mismo modo, es posible establecer un valor mínimo  $cpt\_I\_min$  para el contador de imágenes. Si la imagen actual cumple el criterio de selección (etapa 5) pero el contador de imágenes  $cpt\_I$  no superó el valor mínimo predefinido, la imagen actual  $I(t)$  no está seleccionada. Esta realización alternativa tiene en particular la ventaja de responder a mayores restricciones de velocidad de bits en el enlace de transmisión.

En otra realización alternativa, el umbral S usado para implementar el criterio para seleccionar una imagen relevante puede hacerse adaptativo. En particular, el umbral S puede variar con el tiempo como una función, por ejemplo, de una restricción de velocidad de flujo promedio. La adaptación del umbral a lo largo del tiempo se puede realizar utilizando un algoritmo de tipo gradiente, conocido por los expertos en la materia, que tiene como objetivo minimizar la diferencia entre el número de imágenes N seleccionado, a través de la etapa 5, por segundo y el número de imágenes  $N_s$  deseado seleccionadas por segundo.

Al definir un horizonte temporal  $T_h$  al final del cual se realiza una actualización del umbral S. Por ejemplo, pero no solo, este horizonte temporal  $T_h$  se toma igual a un segundo.

En cada intervalo de tiempo  $T_h$ , se realiza una prueba de comparación y el umbral S se modifica de la siguiente manera:

Si

$$N < N_s, S = S - \delta S$$

Si  $N > N_s, S = S + \delta S$ , donde  $\delta S$  es un incremento inicialmente fijo que corresponde al paso de adaptación.

El paso de adaptación  $\delta S$  también se puede ajustar dinámicamente con el tiempo.

35 Además, un valor mínimo y un valor máximo de dicho umbral S se fijan para evitar los problemas de divergencia, si  $S < S_{min}$  entonces  $S = S_{min}$ , si  $S > S_{max}$  entonces  $S = S_{max}$ .

El valor del incremento  $\delta S$  se puede adaptar en función del valor absoluto del error  $E = |N - N_s|$ . Un ejemplo simple de una función es una función lineal con saturación definida por:

40  $\delta S = \delta S_{max}$  si  $|N - N_s| > E_{max}$ ,  $\delta S = \delta S_{min}$  si  $|N - N_s| < E_{min}$ ,  
 $\delta S$  se interpola linealmente entre  $\delta S_{min}$  y  $\delta S_{max}$ , si  $E_{min} \leq |N - N_s| \leq E_{max}$ , donde  $\delta S_{max}$ ,  $\delta S_{min}$ ,  $E_{max}$  y  $E_{min}$  son datos de entrada.

Las imágenes clave seleccionadas por el procedimiento descrito anteriormente deben transmitirse con la máxima prioridad. El codificador de vídeo, 14, adecuado para implementar el procedimiento de acuerdo con la invención, utiliza dichas imágenes clave definidas en la etapa anterior para generar una estructura particular de un grupo de imágenes, o en inglés, "Group of Pictures" (GOP). El hecho de que las imágenes clave se definan sobre la marcha durante la transmisión del flujo de vídeo requiere que el codificador de vídeo adapte dinámicamente las estructuras de los GOP. El flujo comprimido obtenido a la salida del codificador de vídeo estará compuesto por varias resoluciones temporales, la resolución mínima debe contener todas las imágenes clave y también debe comprimirse para garantizar la mejor calidad con respecto a la velocidad de bits disponible en el canal de transmisión. La resolución temporal que comprende las imágenes clave debe ser recibida como prioridad por el operador. La figura 5 ilustra un ejemplo de implementación en forma de un esquema de granularidad temporal bajo restricción de prioridad en ciertas imágenes previamente identificadas como relevantes. La restricción impuesta al codificador de vídeo reside en el hecho de que cada imagen codificada que pertenece a una resolución temporal dada no puede predecirse a partir de una imagen

codificada que pertenece a resoluciones temporales más altas.

En el modo de realización en el que el módulo de selección de imágenes relevantes proporciona un grado de prioridad que define varias clases de imágenes clave, y no solo una, el codificador de vídeo generará varias resoluciones temporales, jerarquizadas de acuerdo con el grado de prioridad de las imágenes clave. Una vez que se define la estructura de un GOP, el codificador, para cada resolución temporal, define una velocidad y una asignación de bits en distorsión específica que debe permitir cumplir con las restricciones impuestas por la transmisión, en términos de la velocidad de bits disponible en particular. Esta asignación hace posible, para cada tipo de imagen recibida (imagen clave o imagen de menor prioridad) determinar el tipo de codificación que se aplicará, en particular la codificación espacial o la codificación por predicción de otra trama. La figura 6 ilustra esquemáticamente este principio.

Las solicitudes realizadas por el operador para interactuar con el sistema de acuerdo con la invención pueden cumplir con el protocolo de comunicaciones JPIP. Para este propósito, se realiza un recordatorio de las funcionalidades de este protocolo antes de introducir la adaptación que se realiza dentro del alcance de la invención. La arquitectura típica de un sistema JPIP consiste en varios clientes que se comunican con un servidor. La aplicación del cliente tiene una interfaz gráfica donde se muestra la imagen descodificada y la aplicación del servidor devuelve datos cuando el cliente modifica, haciendo zoom o moviéndose, su área de visualización. La comunicación entre el cliente y el servidor consiste en una sucesión de solicitudes y respuestas. La consulta define los parámetros del área de visualización (como la resolución, el tamaño o la posición de la ventana de zoom) de una manera completamente transparente en comparación con la sintaxis del patrón JPEG2000. Una de las propiedades del protocolo JPIP es proporcionar porciones de imágenes JPEG2000 conocidas por los expertos en la materia bajo el término regiones de interés, en respuesta a una solicitud del cliente. Tres pares de parámetros permiten definir una región de interés como se ilustra en la figura 7. El primer par de parámetros ( $F_x$  y  $F_y$ ) permite definir la resolución a la que se desea la imagen, o parte de la imagen. Los otros dos pares de parámetros se aplicarán para esta resolución. ( $X\_offset$  e  $Y\_offset$ ) son los parámetros de posición de la parte de imagen deseada dentro de la resolución definida por la pareja ( $F_x$ ,  $F_y$ ) y ( $R_x$ ,  $R_y$ ) definen su tamaño también dentro de la resolución definida por la pareja ( $F_x$ ,  $F_y$ ).

Convencionalmente, en respuesta a una solicitud de JPIP, el servidor JPEG2000 transmite como prioridad las bajas frecuencias contenidas en la imagen y luego las altas frecuencias de manera incremental. En la implementación del sistema según la invención, el protocolo JPIP se aplica a una imagen residual, 9, obtenida mediante las etapas ilustradas en la figura 2. En este caso, las bajas frecuencias ya se han enviado parcial o completamente en el transmisión de vídeo transmitido, 5. Por lo tanto, es prudente enviar de manera prioritaria las frecuencias altas de la imagen correspondientes a la resolución requerida por el operador, ya que contienen más información. Sin embargo, la restricción de velocidad de bits impuesta en el codificador de vídeo también debe tenerse en cuenta. De hecho, cuanto menor sea la velocidad de bits de la secuencia que contiene las imágenes clave, más se deteriorarán las frecuencias altas y, en el peor de los casos, las frecuencias bajas también pueden dañarse. En este caso, es necesario enviar los residuos de baja frecuencia primero antes de enviar las frecuencias altas correspondientes a la solicitud del usuario. Por lo tanto, el módulo JPIP utilizado en el contexto de la invención es adecuado para determinar el orden de envío de los paquetes JPEG2000 de acuerdo con la velocidad de bits asignada por el codificador de vídeo y los parámetros de resolución, calidad y área espacial solicitados por el operador. El objetivo es reducir el tiempo de transmisión tanto como sea posible, lo que permitirá al operador tomar una decisión rápida.

En una realización alternativa, el operador puede especificar, durante su solicitud, que desea llevar a cabo una búsqueda en las proximidades de la imagen seleccionada para retransmitir. Esta investigación tiene como objetivo encontrar, posiblemente, una imagen cuyo contenido sea similar pero que sea más relevante según un criterio de nitidez. En este caso, las siguientes etapas se implementan cuando el operador lo solicita:

- Se realiza un cálculo en el área seleccionada (por el operador) de la imagen seleccionada de un criterio representativo de la nitidez de dicha área.
- El mismo cálculo se aplica a las mismas áreas de las imágenes ubicadas en una vecindad temporal cercano a la imagen seleccionada.
- La imagen cuya área seleccionada es la más nítida de acuerdo con el criterio elegido se selecciona y recupera mediante la etapa 17 (figura 2). A continuación, se procesa utilizando el procedimiento de acuerdo con la invención descrito anteriormente e ilustrado en la figura 2.

El criterio de nitidez utilizado puede basarse, por ejemplo, en un principio similar al que se puede utilizar para realizar una función de enfoque automático al nivel del zoom óptico de una cámara fotográfica o una cámara. Se puede usar un criterio, conocido por los expertos en la materia, como el descrito en el siguiente artículo: "Autofocusing Algorithm Selection in Computer Microscopy", Yu Sun, S. Duthaler and B. J. Nelson, en Proceedings of IEEE IROS (International Conference on Intelligent Robots and Systems), pp 70-76, 2005. Tal criterio de nitidez debe, para ser efectivo, presentar un valor máximo (respectivamente mínimo) cuando la imagen o el área de la imagen en la que se aplica este criterio es perfectamente nítida y, por el contrario, debe presentar un valor decreciente (respectivamente creciente) a medida que la imagen o el área de la imagen se vuelve menos nítida.

El sistema y procedimiento según la invención tienen en particular las siguientes ventajas:

- La secuencia de alta resolución que se va a transmitir se adapta a las restricciones del medio de transmisión haciendo una selección de la información más relevante para el usuario final.
- El operador puede, gracias a solicitudes interactivas, obtener información temporal o espacial adicional a la

que recibió inicialmente.

- Todas las operaciones de procesamiento realizadas son compatibles en tiempo real y requieren bajos recursos del sistema, además, el tiempo de latencia introducido es mínimo.
- Todo el sistema propuesto permite adaptarse a las restricciones de velocidad, sin sacrificar la resolución y la calidad de las imágenes a las que puede acceder el operador.

5

REIVINDICACIONES

1. Un procedimiento de transmisión en línea de una secuencia de vídeo de alta resolución compuesta por una sucesión de imágenes T, caracterizado porque comprende al menos una etapa para seleccionar imágenes (13) relevantes que comprende al menos las siguientes etapas:

- 5 > Etapa 1: Dividir cada imagen que se transmitirá en el instante t en un número N de áreas,
- > Etapa 2: Para cada área n determinada en la etapa 1, calcular un valor  $v_n(t)$  representativo del contenido de dicha área,
- > Etapa 3: Para cada imagen I(t) que se va a transmitir, generar un vector  $\mathbf{V}_N(t)$  de tamaño N representativo del contenido de dicha imagen y que contiene los valores obtenidos en la etapa 2 para cada una de las áreas
- 10 > Etapa 4: Calcular un coeficiente de correlación normalizado  $\alpha$  entre el vector de referencia  $\mathbf{R}_N(k)=\mathbf{V}_N(t_1)=[\mathbf{r}_{0(k)}, \dots, \mathbf{r}_{N-1(k)}]^T$  determinado para una imagen previamente seleccionada en el instante  $t_1$   $I^{sel}(k-1) = I(t_1)$  y el vector  $\mathbf{V}_N(t)$  calculado para la imagen actual I(t) mediante el siguiente cálculo:

$$\alpha = \frac{\mathbf{R}_N^T(k-1) \cdot \mathbf{V}_N(t)}{\sqrt{\mathbf{R}_N^T(k-1) \cdot \mathbf{R}_N(k-1) \cdot \mathbf{V}_N^T(t) \cdot \mathbf{V}_N(t)}},$$

donde k es el índice temporal de una imagen seleccionada,

- 15 > Etapa 5: Tomar una decisión sobre la selección (o no) para la transmisión de la imagen actual I(t) comparando el valor del coeficiente de correlación normalizado  $\alpha$  con uno o más umbrales con valores fijados a priori,
- > Etapa 6: Transmitir una secuencia submuestreada que comprende solo las imágenes seleccionadas, para al menos una imagen seleccionada,
  - o buscar en una vecindad temporal de la imagen seleccionada una imagen que sea más nítida que la imagen seleccionada,
  - o retransmitir la imagen más nítida.
- 20

2. El procedimiento según la reivindicación 1, **caracterizado porque** el valor  $v_n(t)$  calculado en la etapa 2 se obtiene calculando uno de los criterios derivados de un histograma que contiene los valores de luminancia o crominancia de los píxeles del área n de la imagen I(t), dichos criterios incluyen: la energía, la varianza, el coeficiente de amortiguamiento (o curtosis), el coeficiente de asimetría (o sesgo, Skewness) o el centro de gravedad de dicho histograma, o realizando una combinación lineal de al menos dos de estos criterios.

3. El procedimiento según una de las reivindicaciones 1 a 2, **caracterizado porque** la etapa de seleccionar imágenes relevantes (13) comprende además las etapas siguientes:

- 30 > sobre la base de cada par de vectores  $(\mathbf{R}_N(k-1), \mathbf{V}_N(t))$ , obtenidos para el par de imágenes  $(I^{sel}(k-1), I(t))$ , calcular un segundo vector  $\Delta \mathbf{V}_N(t) = (\Delta v_0(t), \dots, \Delta v_{N-1}(t))$  tal que  $\Delta v_n(t) = |v_n(t) - r_n(k-1)|^\beta$ , siendo  $\beta$  un número entero positivo dado,
- > calcular la norma  $\|\Delta \mathbf{V}_N(t)\|$  del vector  $\Delta \mathbf{V}_N(t)$  y dividir dicho vector por dicha norma para obtener un tercer vector  $\Delta \mathbf{V}'_N(t) = \Delta \mathbf{V}_N(t) / \|\Delta \mathbf{V}_N(t)\| = (\Delta v'_0(t), \dots, \Delta v'_{N-1}(t))$ ,
- 35 > calcular el siguiente criterio adicional para determinar si existe una distribución uniforme o localizada de las diferencias entre las dos imágenes consideradas:

$$DC = (\Delta V_N^{MAX} - \frac{1}{N}) / (1 - \frac{1}{N}),$$

con

$$\Delta V_N^{MAX} = MAX(\Delta v'_n(t)), 0 \leq n < N,$$

- 40 o aplicar la siguiente prueba, si  $DC < S_{DC}$ , las diferencias entre las dos imágenes consideradas se distribuyen uniformemente, de lo contrario se localizan en un área dada de dichas imágenes.

4. El procedimiento según una de las reivindicaciones 1 a 3, **caracterizado porque** la etapa 5 se implementa realizando una prueba de comparación entre el coeficiente de correlación  $\alpha$  y un umbral S predefinido, un número entero estrictamente positivo, de la siguiente manera:

- 45 > si  $\alpha < S$ , la imagen actual I(t) está seleccionada, el vector de referencia se actualiza  $\mathbf{R}_N(k) = \mathbf{V}_N(t)$  y el índice de imagen seleccionado se incrementa  $k = k + 1$
- > en el caso inverso, la imagen no se retiene como imagen relevante.

5. El procedimiento según las reivindicaciones 1 a 4, **caracterizado porque** la etapa de seleccionar imágenes relevantes (13) comprende las siguientes etapas adicionales:

- > definir un contador de imágenes  $cpt\_I$ , inicializado a 1 cuando se selecciona una imagen relevante en la etapa 5 y se incrementa en 1 con cada nueva imagen actual  $I(t)$  recibida, así como un valor máximo de dicho contador  $cpt\_I\_max$ ,  
 > para cada imagen, realizar la siguiente prueba de comparación: si  $\alpha \geq S$  et  $cpt\_I \geq cpt\_I\_max$ , seleccionar dicha imagen como relevante.
- 5
6. El procedimiento según las reivindicaciones 1 a 4, **caracterizado porque** la etapa de seleccionar imágenes relevantes (13) comprende las siguientes etapas adicionales:
- definir un contador de imágenes  $cpt\_I$ , inicializado a 1 cuando se selecciona una imagen relevante en la etapa 5 y se incrementa en 1 con cada nueva imagen actual recibida  $I(t)$  y un valor mínimo de dicho contador  $cpt\_I\_min$ ,
  - para cada imagen, realizar la siguiente prueba de comparación: si  $\alpha < S$  et  $cpt\_I < cpt\_I\_min$ , no conservar dicha imagen como relevante.
- 10
7. El procedimiento según las reivindicaciones 1 a 6, **caracterizado porque** la etapa de seleccionar imágenes relevantes (13) comprende las siguientes etapas adicionales:
- definir un número  $N_s$  deseado de imágenes seleccionadas por segundo,
  - definir un intervalo de tiempo de duración igual a  $T_n$  para la actualización de dicho umbral  $S$ ,
  - en cada intervalo de tiempo de duración igual a  $T_n$ , calcular el número de imágenes  $N$  seleccionadas por el procedimiento durante dicho intervalo de tiempo anterior de duración igual a  $T_n$ ,
  - en cada intervalo de tiempo de duración igual a  $T_n$ , modificar el valor del umbral  $S$  de la siguiente manera:
- 15
- Si
- 20  $N < N_s$ ,  $S = S - \delta S$
- Si  $N > N_s$ ,  $S = S + \delta S$ , donde  $\delta S$  es un valor de número entero positivo incremental inicialmente fijo y/o ajustado dinámicamente.
8. El procedimiento según las reivindicaciones 1 a 3, **caracterizado porque** la etapa 5 consiste en realizar una prueba de comparación entre el coeficiente de correlación  $\alpha$  y un número estrictamente positivo  $J$  de umbrales predefinidos  $S_j$ ,  $1 \leq j < J$ , de la siguiente manera:
- Si  $0 \leq \alpha < S_1$ , la imagen actual se selecciona con un grado máximo de prioridad,
  - Si  $S_j \leq \alpha < S_{j+1}$ , para  $1 \leq j < J-1$ , la imagen actual se selecciona con un grado medio de prioridad  $j$ ,
  - Si  $\alpha \geq S_J$ , la imagen no se conserva como relevante,  $S_j$  es un número entero estrictamente positivo.
- 25
9. El procedimiento según las reivindicaciones 1 a 3, **caracterizado porque** la etapa 5 se realiza de la siguiente manera:
- ordenar las imágenes  $T$  de dicha secuencia de vídeo en orden ascendente del valor de su coeficiente de correlación  $\alpha$ ,
  - agrupar las imágenes ordenadas en  $T$  en  $J + 1$  grupos jerarquizados en función del valor de  $\alpha$ , comprendiendo cada uno de dichos grupos de imágenes el mismo número de imágenes y correspondiendo a un grado medio de prioridad  $j$ .
- 30
- 35
10. El procedimiento según una de las reivindicaciones 8 o 9, **caracterizado porque** la secuencia de vídeo de alta resolución se comprime antes de la transmisión a través de las siguientes etapas:
- realizar un submuestreo espacial (11) de dicho flujo de vídeo de alta resolución para obtener una secuencia de vídeo de menor resolución (2),
  - aplicar la etapa de seleccionar imágenes relevantes (13) a la secuencia de vídeo de menor resolución obtenida,
  - realizar una compresión de dicha secuencia (2) mediante un codificador de vídeo (14) adecuado para generar una estructura particular de grupo de imágenes, en la que las imágenes relevantes seleccionadas definen:
    - una resolución temporal, que contiene el conjunto de imágenes relevantes seleccionadas en el caso en el que la etapa de seleccionar imágenes relevantes (13) libera solo un grupo de imágenes relevantes, todas con el mismo grado de prioridad,
    - varias resoluciones temporales, cada una con las imágenes relevantes seleccionadas con un grado diferente de prioridad, en el caso en que la etapa de seleccionar imágenes relevantes (13) libera varios grupos de imágenes relevantes con diferentes grados de prioridad.
- 40
- 45
11. El procedimiento según la reivindicación 10, **caracterizado porque** el codificador de vídeo (14) se ajusta al patrón H.264.
- 50
12. El procedimiento según una de las reivindicaciones 10 u 11, **caracterizado porque** la secuencia de vídeo de alta resolución que se va a transmitir se guarda en una unidad de almacenamiento (12) y se selecciona una imagen o área de imagen de dicha secuencia de vídeo para transmitir o retransmitir de acuerdo con las siguientes etapas:

- realizar un sobremuestreo espacial (16) de dicha imagen para obtener una imagen de alta resolución (6),
  - calcular la diferencia entre la imagen sobremuestreada (6) y la correspondiente almacenada previamente (12) para obtener un residuo (9),
  - comprimir dicho residuo (9) a través de un codificador de imagen (21) para producir un flujo comprimido que comprenda al menos paquetes que contengan información sobre las frecuencias altas de la imagen (12) y otros paquetes que contengan información sobre las frecuencias bajas de dicha imagen,
  - transmitir dicho flujo comprimido a un operador remoto (19).
- 5
13. El procedimiento según la reivindicación 12, **caracterizado porque** la imagen que se va a retransmitir o la imagen asociada con dicha área de imagen que se va a retransmitir se compara con sus imágenes temporalmente vecinas dentro de dicha secuencia de vídeo de alta resolución para determinar cuál es la menos borrosa de acuerdo con las siguientes etapas:
- 10
- calcular en la imagen o el área de imagen seleccionada un criterio representativo de la nitidez de dicha imagen o área de imagen, exhibiendo dicho criterio de nitidez un primer extremo en el que dicha imagen o área de imagen es perfectamente nítida y un segundo extremo opuesto al primero cuando dicha imagen o el área de la imagen está totalmente borrosa,
  - realizar el mismo cálculo en las M imágenes temporalmente vecinas, siendo M un valor predefinido,
  - seleccionar la imagen más nítida según el criterio anterior y retransmitir esta imagen o el área de imagen asociada, en lugar de la seleccionada inicialmente.
- 15
14. El procedimiento según una de las reivindicaciones 12 o 13, **caracterizado porque** el codificador de imagen (21) es adecuado para definir el orden de envío de los paquetes que forman el flujo comprimido en función, al menos, de una información sobre la velocidad de bits disponible para transmisión y fijada a priori (14).
- 20
15. El procedimiento según una de las reivindicaciones 12 a 14, **caracterizado porque** el codificador de imagen (21) se ajusta al patrón JPEG2000.
- 25
16. El procedimiento según una de las reivindicaciones 12 a 15, **caracterizado porque** la retransmisión de una imagen o área de imagen se realiza mediante una solicitud (8) del operador remoto.
17. El procedimiento según la reivindicación 16, **caracterizado porque** las solicitudes (8) y las respuestas a dichas solicitudes se implementan a través del patrón JPIP.
- 30
18. Un sistema interactivo para transmitir secuencias de vídeo de alta resolución, **caracterizado porque** comprende un emisor y un receptor que comprende medios para ejecutar las etapas del procedimiento según una de las reivindicaciones 1 a 17.

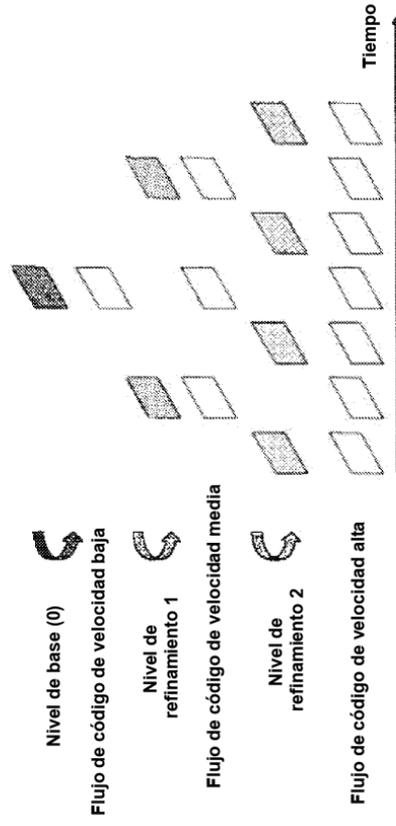


FIG.1

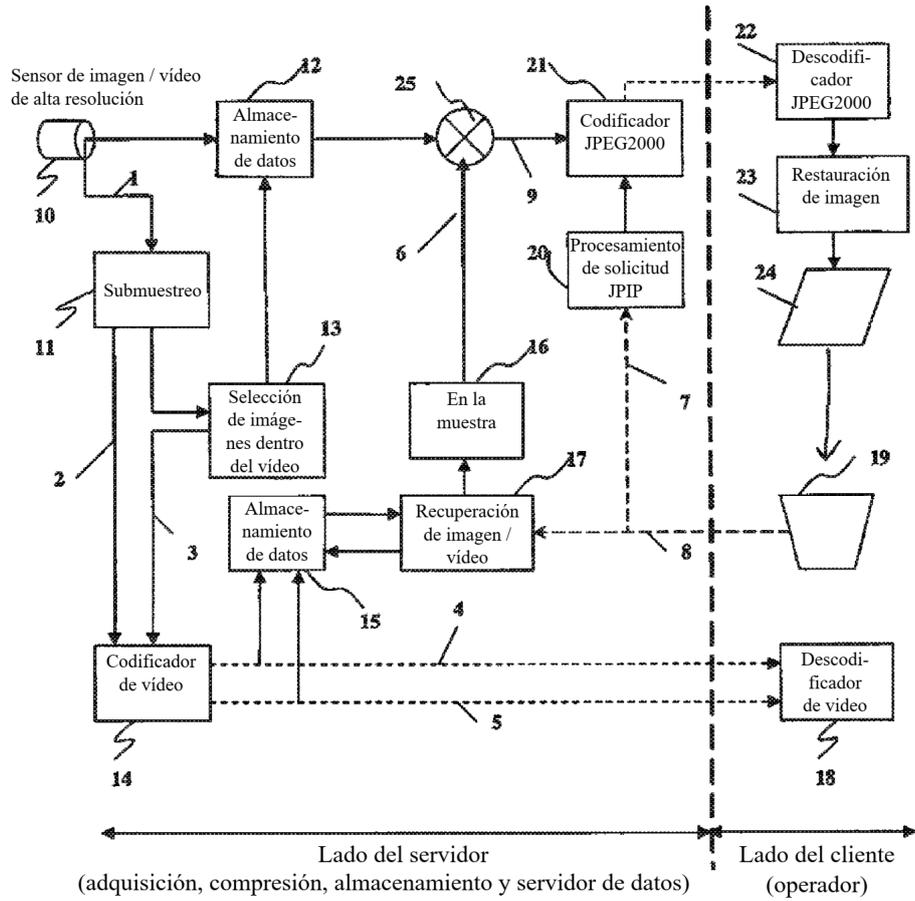
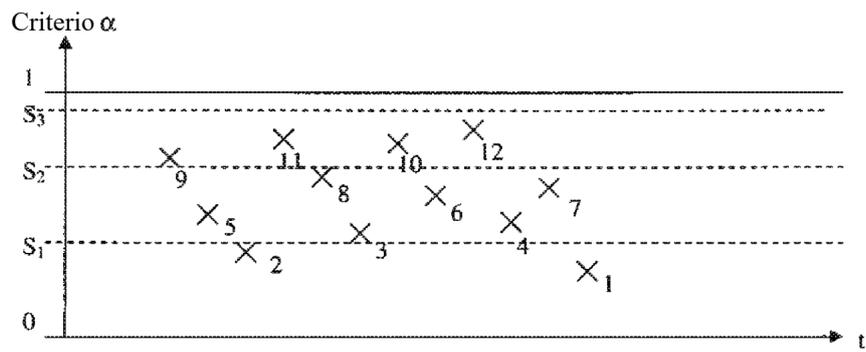
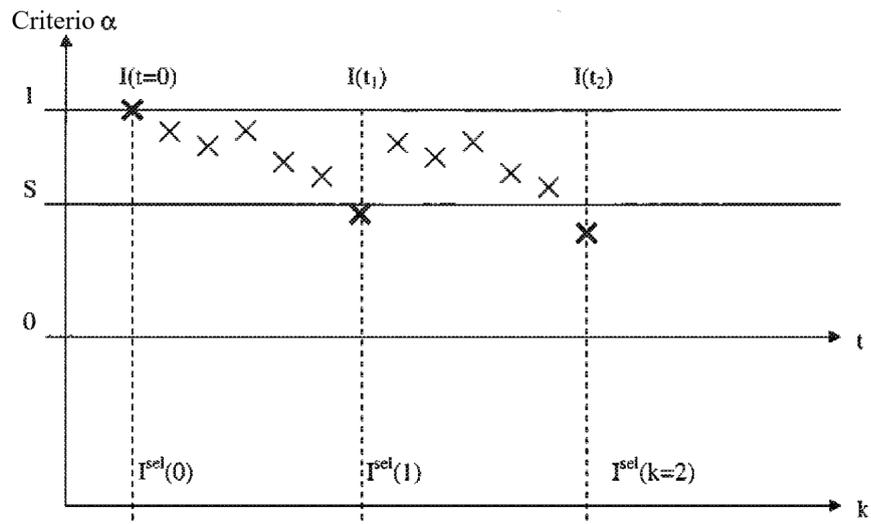
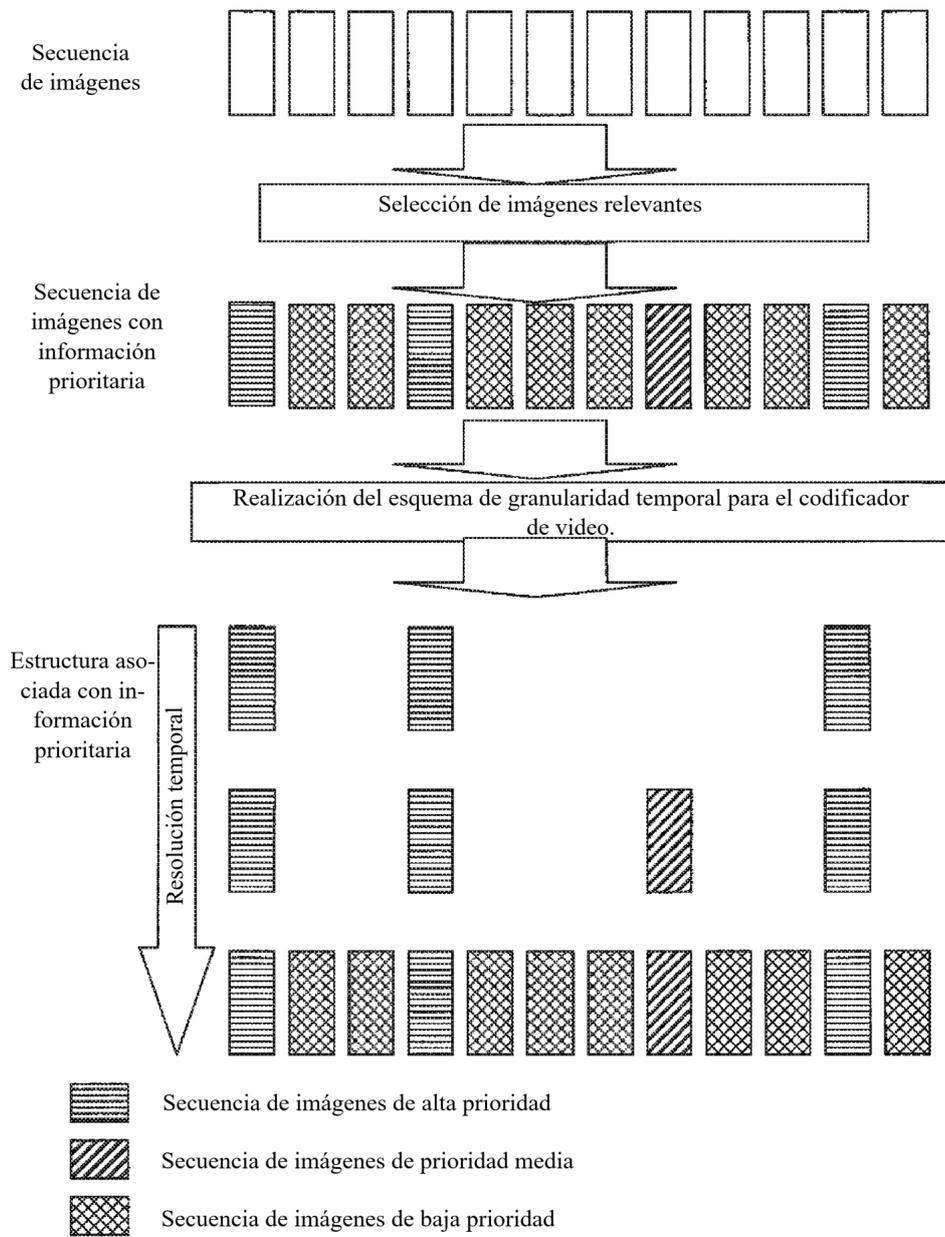
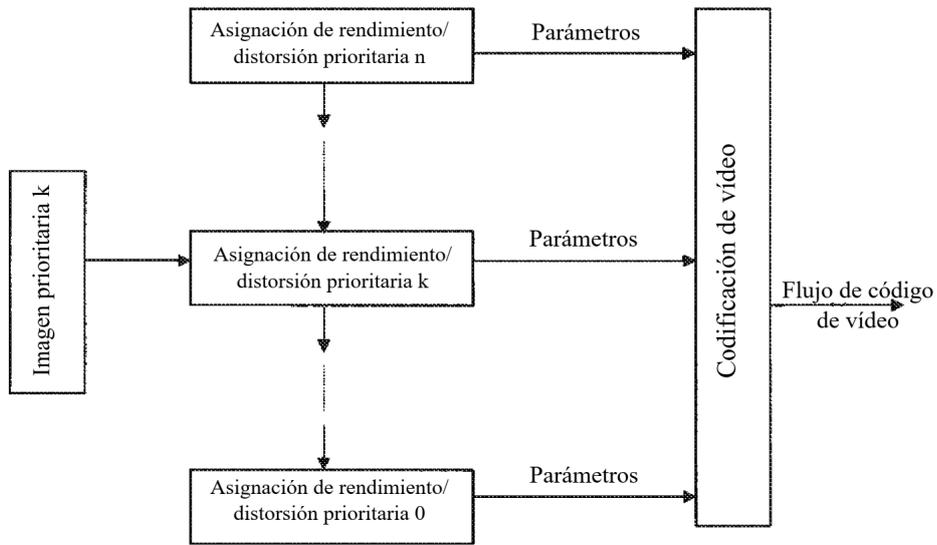


FIG.2

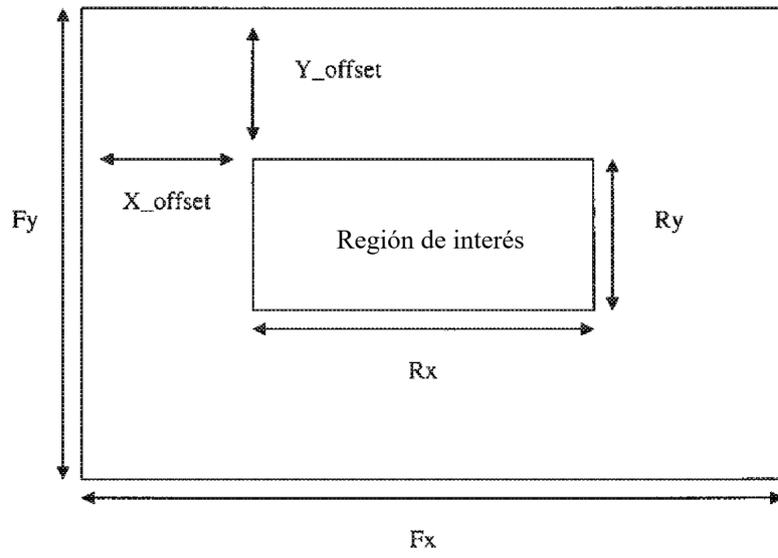




**FIG.5**



**FIG.6**



**FIG.7**