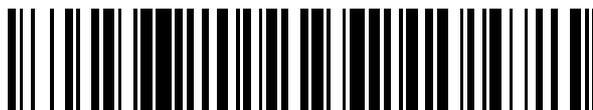


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 545 513**

51 Int. Cl.:

G06T 11/60 (2006.01)

H04N 13/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.04.2012** **E 12165350 (5)**

97 Fecha y número de publicación de la concesión europea: **27.05.2015** **EP 2658266**

54 Título: **Renderización de vista virtual sensible a texto**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
11.09.2015

73 Titular/es:

VESTEL ELEKTRONIK SANAYI VE TICARET A.S.
(100.0%)
Organize Sanayi Bölgesi
45030 Manisa, TR

72 Inventor/es:

CIGLA, CEVAHIR

74 Agente/Representante:

ARPE FERNÁNDEZ, Manuel

ES 2 545 513 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

Renderización de vista virtual sensible a texto

5 Campo técnico

[0001] La presente invención se refiere a un dispositivo y un método para renderización de vista virtual sensible a texto en video estéreo o de vistas múltiples.

10 Técnica anterior

[0002] En las pantallas de presentación estereoscópicas o pantallas de presentación de vistas múltiples, la síntesis de vista virtual es un método utilizado comúnmente para mejorar las capacidades 3D, tales como ajuste de profundidad, cambio del ángulo de visualización para vista libre o extensión de videos mono/estéreo a vistas múltiples, de las pantallas de visualización. En general, con el fin de generar dicha vista virtual, se utiliza la renderización de base de profundidad. En la renderización de base de profundidad, preferentemente se genera por escena un mapa de profundidad, utilizando imágenes reales adquiridas desde diferentes puntos de vista y mediante la utilización del mencionado mapa de profundidad y textura de una escena original (referencia), se genera una vista virtual desde un punto de vista arbitrario (renderizado). Durante este proceso, la textura de dicha escena se ha deformado de acuerdo con la información del mapa de profundidad (por ejemplo los píxeles (o los voxels) más cercanos a la posición del punto de vista de la vista virtual se tienen en cuenta para visualizarse). Cuando toda la textura de la vista de referencia es deformada; se producen algunas posiciones perdidas (huecos) debido a regiones ocluidas. Estas posiciones son rellenadas por diversos métodos, tales como pintura (interpolación de imagen), copiado de fondo o promediado de píxeles vecinos (voxels). Para el caso estéreo, las texturas de las vistas izquierda y derecha también podrían mezclarse para obtener la textura de la vista arbitraria.

[0003] Durante la síntesis de la vista virtual, la información de mapa de profundidad no podrá generarse perfectamente. En otras palabras, dicha información puede incluir errores. Debido a dichos errores la vista virtual generada puede contener objetos con una profundidad inadecuada. Aunque la mayoría de los usuarios no se dan cuenta de este error de profundidad de los objetos de fondo, dichos errores en los textos (tales como subtítulos) a la vista de usuarios degradan su confort ya que los textos son considerados los objetos más sobresalientes en una pantalla. Así, durante la renderización de la vista virtual, los textos deberían ser renderizados con mínimo error.

[0004] El documento de patente WO2008115222A1 describe un sistema y un método para combinar texto con contenido 3D. En este documento, un texto se inserta en el contenido 3D como el valor de profundidad más alto. Aunque en este documento son insertados subtítulos con alta visibilidad en la vista 3D, la imagen de fondo y parte de texto de la imagen se pueden mezclar entre sí porque regiones de texto se detectan poniendo cajas alrededor del texto y procesándose dicha caja como un todo. De la misma manera, la detección de profundidad se proporciona para la imagen original sin subtítulos lo cual es tarea fácil en comparación con las imágenes con texto incorporado.

[0005] Otro documento de patente WO2010151555A1, describe un método para incrustar subtítulos y/o superponer gráficos en un video 3D o de vistas múltiples. Sin embargo, este documento no resuelve el problema de los caracteres erróneamente posicionados en la imagen renderizada, lo cual se produce debido a mapa de profundidad erróneo.

[0006] El documento de patente EP2403254A1 describe un aparato, programa y método para el procesamiento de imagen. De acuerdo con este documento, se reciben dos imágenes por un receptor. Utilizando un módulo de detección de subtítulos se determinan los textos de las imágenes recibidas. En este documento, la misma información de paralaje se utiliza para los subtítulos.

Breve descripción de la invención

[0007] La presente invención describe un método y sistema para la renderización de vista virtual, en el cual un vídeo de entrada con un texto se renderiza de acuerdo con los datos de profundidad. Dicho método incluye las etapas de; determinación del formato de vídeo en términos de número de vistas; detección de dicho texto en el mencionado vídeo de entrada, mediante la ejecución de un algoritmo de detección de texto de acuerdo con un determinado formato del vídeo de entrada; generación de una vista renderizada, llevando a cabo el post-procesamiento de datos de profundidad en el cual el texto detectado es deformado con distribución de profundidad constante de acuerdo con dichos datos de profundidad.

[0008] Además, dicho sistema comprende al menos unos medios para determinar el formato de un vídeo de entrada, en términos de número de vistas; al menos unos medios para la detección de, al menos, un texto del mencionado vídeo de acuerdo con determinado formato; al menos unos medios de post-procesamiento de datos de profundidad, los cuales deforman el texto detectado con un valor de profundidad constante.

[0009] Con el método de renderización de vista virtual y el sistema de la presente invención, a partir del video se extrae un texto de un vídeo. Entonces, dicho texto y la vista de fondo del video son deformados Independientemente de acuerdo con los datos de profundidad del video. Así, de acuerdo a la presente invención, se previenen caracteres desplazados debido a información de profundidad errónea y el fondo del video es deformado sin mezclar el texto.

65 Objeto de la Invención

[0010] El objeto de la invención es proporcionar un método y un sistema para la renderización de vistas virtuales que comprenden textos.

5 **[0011]** Otro objeto de la presente invención es proporcionar un método y un sistema para la renderización de textos en una vista virtual con el mínimo error.

[0012] Otro objeto de la presente invención es proporcionar un método y un sistema para la renderización de vistas virtuales con alta fiabilidad.

Descripción de las figuras

10

[0013]

La figura 1 muestra una renderización de vista virtual errónea de la técnica anterior.

La figura 2 muestra un diagrama de flujo del método de renderización de vista virtual de la presente invención.

La figura 3 muestra un ejemplo de renderización de vista virtual de la presente invención.

15 **[0014]** Los números de referencia que se utilizan en las figuras pueden poseer los siguientes significados.

Vista original (O)

Caracteres de texto originales (O1, O2)

Vista renderizada (R)

20

Caracteres de texto renderizados (R1, R2)

Datos de mapa de profundidad (M)

Información de profundidad de los caracteres (M1, M2)

Profundidad de los caracteres (D1, D2)

Caracteres con información de profundidad (OD1, OD2)

25

Vídeo de entrada (101)

Detector tipo de vídeo (102)

Detección de texto (103)

Análisis de error de renderizado (104)

Post-procesamiento de datos de profundidad (105)

30

Datos de profundidad (106)

Vista renderizada (107)

Descripción detallada de la invención

35 **[0015]** En pantallas de visualización estereoscópica o pantallas de visualización de vistas múltiples, la síntesis de vista virtual es un método comúnmente utilizado para la mejora de las capacidades 3D de dichas pantallas de visualización. En la síntesis de vista virtual, por lo general, se generan unos datos de textura y una información de mapa de profundidad para una pantalla de visualización y dicha información se utiliza para generar (renderizar) una vista virtual desde una ubicación arbitraria. Durante la generación de dicha vista virtual, dicha textura es deformada de acuerdo con la información del mapa de profundidad. Si dicha información de mapa de profundidad comprende valores erróneos, la imagen generada (renderizada) también comprende partes erróneas, que también son llamadas artefactos. Normalmente, estos artefactos no son percibidos por los usuarios. Sin embargo, si un mapa de profundidad de un texto comprende errores, la posición de los caracteres de dicho texto puede desplazarse erróneamente. Un ejemplo de imagen renderizada erróneamente se muestra en la figura 1. En este ejemplo, la vista original (O) implica porciones de texto con caracteres de texto original (O1, O2). Cuando la textura de vista original (O) es deformada con un mapa de profundidad erróneo, la imagen renderizada (R) comprende caracteres de texto renderizados (R1, R2), cuyas posiciones están desplazadas. Por lo tanto, con el fin de evitar este tipo de caracteres posicionados erróneamente en la imagen renderizada (R), la presente invención proporciona un método y un sistema para renderización sin error de textos en una vista virtual.

40 **[0016]** En el método de renderización de vista virtual de la presente invención, cuyo diagrama de flujo se da en la figura 2, en primer lugar, el formato de vídeo de entrada (101) se determina (102) en función de número de vistas (por ejemplo, mono, estéreo o vistas múltiples). Entonces, de acuerdo con el formato determinado, se ejecuta un algoritmo de detección de texto para la detección de, al menos, un texto en dicho vídeo de entrada (101) (103). En la etapa de detección de texto (103), se buscan regiones de texto y se extrae la información de texto (caracteres) de dicha región de texto. Por último, de acuerdo con una información de datos de profundidad (106) del vídeo de entrada (101), se genera una vista renderizada (107) mediante el procesamiento posterior de datos de profundidad (105). Dichos datos de profundidad pueden ser adquiridos con el vídeo de entrada (101) como información separada o pueden ser generados a partir del vídeo de entrada (101). En dicha etapa de procesamiento de datos (105), el texto detectado es deformado con una distribución de profundidad constante e independiente de la imagen de fondo del vídeo. De acuerdo con la presente invención, ya que el texto detectado se procesa independientemente a partir de dicho fondo, a pesar de que la información de mapa de profundidad comprenda errores, los textos del vídeo se representan con mismo valor de profundidad. De este modo, se previene el desplazamiento de los caracteres de la imagen.

50 **[0017]** El sistema de renderización de vista virtual de la presente invención comprende, al menos, unos medios para determinar el formato de un vídeo de entrada (101) en términos de número de vistas; al menos, unos medios para la detección de, al menos, un texto en dicho vídeo de conformidad con el formato determinado; al menos, unos medios

para el procesamiento posterior de datos de profundidad, los cuales deforman el texto detectado con un valor de profundidad constante. En el sistema, el texto de dicho vídeo es detectado por dichos medios de detección y el texto detectado es deformado con datos de profundidad del video de entrada en los medios de procesamiento posterior de datos de profundidad independientemente de la vista de fondo del vídeo.

5 **[0018]** En una realización preferida de la presente invención, si el formato de vídeo detectado de dicho vídeo de entrada (101) es una vista mono (única), se ejecuta un algoritmo de detección para determinar las posiciones de posibles regiones de texto. Entonces, la información de texto de dichas regiones se extrae, utilizando cualquiera de los algoritmos de reconocimiento de texto conocidos de la técnica (es decir, mediante la utilización de un algoritmo OCR (reconocimiento óptico de caracteres).

10 **[0019]** En otra realización preferida de la presente invención, si el formato de vídeo detectado de dicho vídeo de entrada (101) es una vista estéreo o una vista múltiple, se ejecuta cualquier algoritmo de detección para detectar el texto en una vista estéreo o múltiple. En esta realización, antes de la etapa de detección de texto (103), preferiblemente se lleva a cabo una etapa de análisis de error (104). En la etapa de análisis de errores (104), se comparan dos tramas de diferentes vistas de video estéreo o múltiple, para detectar el texto. Más específicamente, una primera trama de video estéreo o múltiple es deformada hasta otra trama pareja (que es una trama del mismo instante con dicha primera vista pero desde otro punto de vista) por medio de los datos de profundidad del vídeo. Durante esta deformación, la limitación de visibilidad de la trama deformada también se utiliza y la textura que pertenece a los píxeles más cercanos son sobre-escritos en la vista renderizada. De esta manera, los objetos de primer plano pueden ocluir el fondo, lo que es un hecho esperado. Entonces, la vista sintetizada (primera vista deformada) se compara con la mencionada trama pareja. Esta comparación se hace tomando la diferencia de trama absoluta de la vista sintetizada y dicha trama pareja. Durante la diferencia de trama, los huecos en la vista sintetizada no se consideran ya que corresponden a las regiones que no pueden ser vistas mediante el punto de vista de dicha trama pareja. Las grandes incongruencias, entre la vista sintetizada y dicha trama pareja, con altas diferencias absolutas RGB corresponden a las regiones erróneas en los datos de profundidad. Estas regiones incongruentes pueden incluir texto, patrones repetidos, objetos delgados donde la mayoría de los métodos de extracción de profundidad pueden fallar. Con el fin de extraer el texto de las incongruencias indicadas, se comparan las diferencias de trama con un umbral y los píxeles que tienen errores de RGB mayores que el umbral se asignan como posibles regiones de texto. Entonces, las posibles regiones de texto son buscadas como texto y se ejecuta un algoritmo de detección de texto. De esta manera, el algoritmo de detección de texto se ejecuta para posibles regiones de texto en lugar de toda una trama. De este modo, se mejora el rendimiento y la fiabilidad de la etapa de detección de texto (103).

20 **[0020]** En otra realización preferida de la presente invención, durante la etapa de post procesamiento de datos de profundidad (105), el texto detectado es deformado con el valor de los datos de profundidad del carácter de texto, que está más cercano al punto de vista. Así, mediante este método todos los caracteres del texto detectado son visionados con la misma profundidad. De este modo, se evita no sólo el desplazamiento de los caracteres, sino también que todos los caracteres son vistos de manera uniforme.

25 **[0021]** En la figura 3 se proporciona una realización de la presente invención. De acuerdo con esta realización, primero se detectan los caracteres del texto original (01, 02) de una vista original (O), mediante la identificación de regiones de texto (dichas regiones de texto se muestran con líneas diagonales). Entonces, de acuerdo a los datos de mapa de profundidad (M) de dicha vista original (O), se determina la profundidad de los caracteres (D1, D2) de dichas regiones de texto. Para cada región de texto, se generan todos los caracteres con la información de profundidad (OD1, OD2), de modo que todos los caracteres (OD1, OD2) en unas regiones de texto tienen la misma profundidad. Por último, la imagen renderizada (R), con los caracteres renderizados correctos (R1, R2) (caracteres sin desplazamiento) son generados mediante la deformación de la imagen original (O) con datos de mapa de profundidad (M).

30 **[0022]** Con el método de renderización de vista virtual y el sistema de la presente invención, un texto (tal como subtítulos, más específicamente un subtítulo incrustado) de un video se extrae a partir del vídeo. Entonces, el mencionado texto y la vista de fondo del video (sin texto) son deformados independientemente de acuerdo con los datos de profundidad del video. Por lo tanto, según la presente invención se previene el desplazamiento de caracteres debido a la información errónea datos de profundidad y el fondo del vídeo es deformado sin mezclar el texto.

REIVINDICACIONES

- 5 **1.** Método de renderización de vista virtual, en el que un vídeo de entrada (101), con al menos un texto, es renderizado de acuerdo a los datos de profundidad (106) para representar una vista virtual, caracterizado porque dicho método comprende las etapas de;
- determinación de un formato del vídeo de entrada (101) en términos de número de vistas (102);
 - detección de, al menos, un texto en dicho vídeo de entrada (101) (103) mediante la ejecución de un algoritmo de detección de texto, de acuerdo con un determinado formato de video de entrada (101); en el que
- 10 i. si el formato de vídeo detectado del mencionado vídeo de entrada (101) es de una vista única, la etapa de detección de texto (103) comprende ejecutar un algoritmo de detección para determinar las posiciones de regiones de texto posibles,
- 15 ii. si el formato de vídeo detectado del mencionado vídeo de entrada (101) es de una vista en estéreo o de vista múltiple, antes de la etapa de detección de texto se lleva a cabo una etapa de análisis de error de renderizado (104) para detectar posibles regiones de texto (103) y la etapa de análisis de error de renderizado comprende comparar dos imágenes procedentes de diferentes vistas de videos estéreo o de vistas múltiples para detectar texto y comprendiendo la etapa de detección de texto (103) ejecutar un algoritmo de detección para detectar el texto de un
- 20 vista estéreo o múltiple,
- generación de una vista renderizada (107) mediante la realización de post-procesamiento de datos de profundidad (105) en el que para cada región de texto, todos los caracteres de texto detectado son deformados con misma distribución de profundidad de acuerdo con dichos datos de profundidad (106).
- 25 **2.** Método de acuerdo con la reivindicación 1 caracterizado porque; la etapa de detección de texto (103) comprende adicionalmente la extracción de información de texto de dichas regiones mediante utilización de un algoritmo de reconocimiento de texto.
- 30 **3.** Método de acuerdo con la reivindicación 1 caracterizado porque la etapa de análisis de error de renderizado (104) comprende las etapas de;
- deformación de una primera trama de dicho vídeo en estéreo o de vistas múltiples en una trama pareja a través de los datos de profundidad (106) de dicho vídeo de entrada (101);
 - comparación de una imagen deformada con dicha trama pareja, tomando la diferencia de trama absoluta de la vista sintetizada y dicha trama pareja;
 - comparación de la diferencia de trama absoluta con un valor umbral;
 - si para un píxel la diferencia de trama absoluta de la trama deformada y dicha trama pareja es mayor que dicho valor umbral, se asigna dicho píxel como una posible región de texto;
 - llevar a cabo la etapa de detección de texto (103) para dicha región de texto posible.
- 35 **4.** Método de acuerdo con la reivindicación 1 caracterizado porque la etapa de post procesamiento de datos de profundidad (105), comprende la deformación del mencionado texto detectado con un valor de datos de profundidad del carácter de texto, que está más próximo a un punto de vista.
- 40 **5.** Sistema de renderización de vista virtual, en el que un vídeo de entrada (101), con al menos un texto, es renderizado de acuerdo con unos datos de profundidad (106) para representar una vista virtual caracterizado porque comprende;
- al menos unos medios para determinar el formato de un vídeo de entrada (101) en términos de número de vistas;
 - al menos unos medios para la detección de, al menos, un texto en dicho vídeo mediante la ejecución de un algoritmo de detección de texto de acuerdo con determinado formato de vídeo de entrada; en el que
- 45 i. si el formato de vídeo detectado de dicho vídeo de entrada es una vista única, la etapa de detección de texto comprende la ejecución de un algoritmo de detección para determinar las posiciones de regiones de texto posibles,
- 50 ii. si el formato de vídeo detectado de dicho vídeo de entrada es una vista estéreo o una vista múltiple antes de la etapa de detección de texto se lleva a cabo una etapa de análisis de error de renderizado para detectar posibles regiones de texto y comprendiendo la etapa de análisis de error de renderizado comparar dos imágenes procedentes de diferentes vistas de video estéreo o de vistas múltiples para la detección de texto y comprendiendo la etapa de detección de texto ejecutar un algoritmo de detección para detectar el texto de un vista estéreo o
- 55 múltiple,
- al menos unos medios para generar una vista renderizada, llevando a cabo un post procesamiento de los datos de profundidad en el que para cada región de texto, todos los caracteres de texto detectado son deformados con misma
- 60 distribución de profundidad de acuerdo con dichos datos de profundidad.

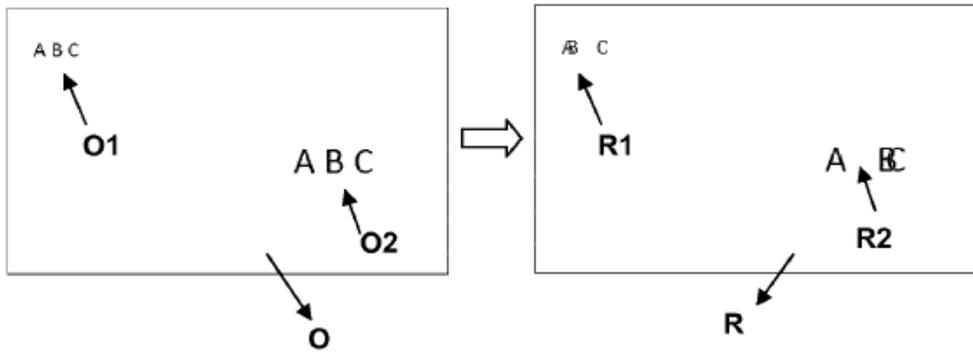


Figura 1

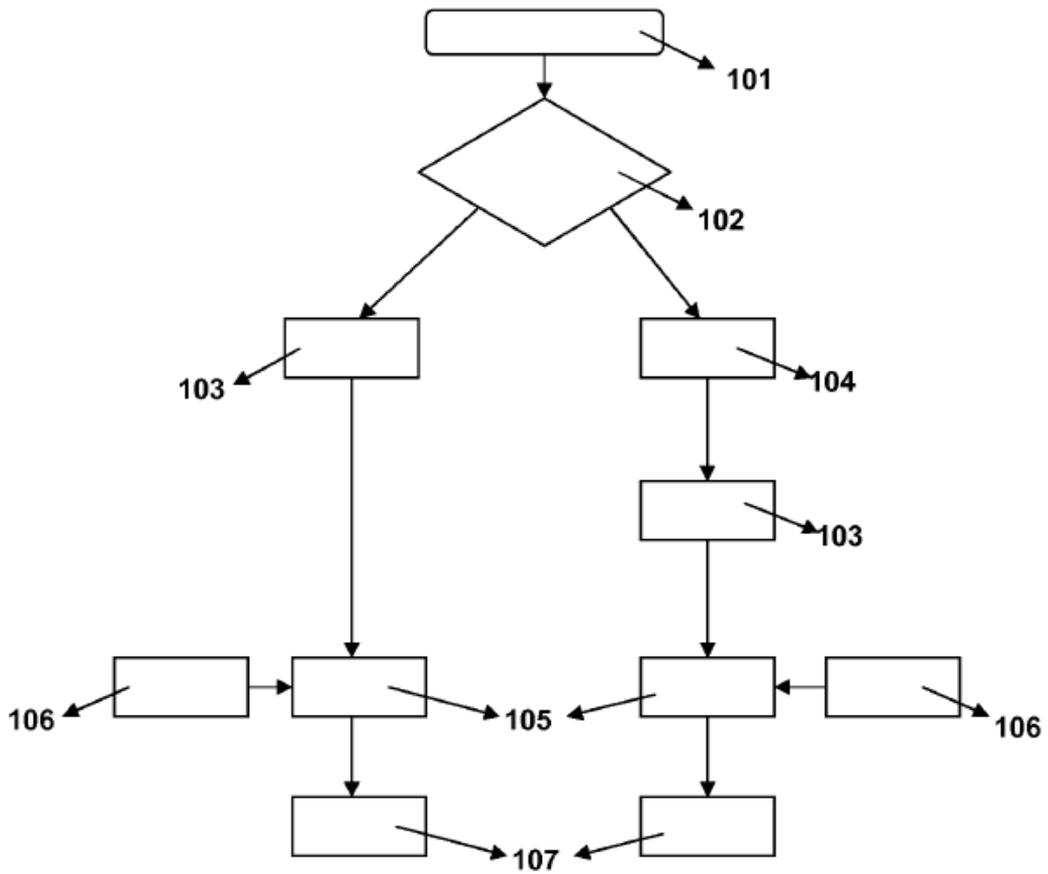


Figura 2

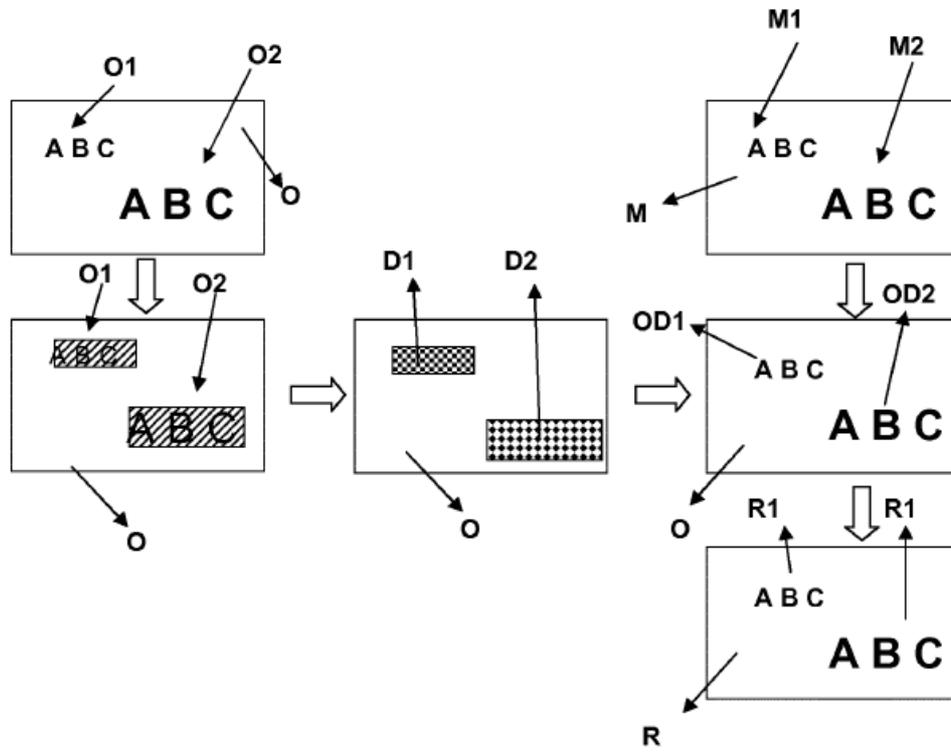


Figura 3

REFERENCIAS CITADAS EN LA DESCRIPCIÓN

5 La lista de referencias citada por el solicitante lo es solamente para utilidad del lector, no formando parte de los documentos de patente europeos. Aún cuando las referencias han sido cuidadosamente recopiladas, no pueden excluirse errores u omisiones y la OEP rechaza toda responsabilidad a este respecto.

Documentos de patente citados en la descripción

- WO 2008115222 A1 [0004]
- WO 2010151555 A1 [0005]
- EP 2403254 A1 [0006]

10