

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 972 232**

51 Int. Cl.:

G06V 30/414 (2012.01)

G06V 10/26 (2012.01)

G06T 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.06.2021** **E 21382500 (3)**

97 Fecha y número de publicación de la concesión europea: **10.01.2024** **EP 4099215**

54 Título: **Método de visión por ordenador para detectar regiones de documentos que serán excluidas de un proceso de incrustación y programas informáticos del mismo**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
11.06.2024

73 Titular/es:

**TELEFONICA CIBERSECURITY & CLOUD TECH
S.L.U. (100.0%)
Edificio Oeste 3, Ronda de la Comunicación s/n
28050 Madrid, ES**

72 Inventor/es:

**BIANZINO, ARUNA PREM;
ELOSUA TOME, JUAN;
PÉREZ VIEITES, DIEGO;
DE LOS SANTOS VILCHEZ, SERGIO y
GARCIA NOGUEIRAS, IVÁN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 972 232 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

Método de visión por ordenador para detectar regiones de documentos que serán excluidas de un proceso de incrustación y programas informáticos del mismo

5

Campo técnico

La presente invención tiene su aplicación en el sector de las telecomunicaciones, dentro del campo de la seguridad de la información digital y el procesamiento de contenidos digitales y, en general, se refiere a la tecnología de marcas de agua digitales.

10

Más concretamente, la presente invención se refiere a un método para detectar, mediante algoritmos de inteligencia artificial (AI), regiones de documentos que serán excluidas de un proceso de incrustación. El método permite detectar contenidos de texto no válidos que se desplazarán mediante un algoritmo de marca de agua de texto digital para mejorar la imperceptibilidad de la información incrustada.

15

Antecedentes de la invención

La inclusión de información oculta en un archivo de texto es una característica deseada en muchos campos y para muchas aplicaciones diferentes. Por un lado, la información oculta incrustada en un documento de texto puede ser decodificada por máquinas para permitir la clasificación automática de los documentos. Al mismo tiempo, dicha información puede permitir la validación automática de la autenticidad, integridad y propiedad de un documento. Además, esa información -incluida en documentos confidenciales- puede permitir su rastreo, disuadir de la fuga de información o ayudar en la identificación del responsable en caso de fuga.

20

25

La información incluida en los documentos con este fin suele estar pensada para que sea imperceptible por los seres humanos, pero apreciable por los sistemas informáticos, que pueden leerlos y decodificarlos. Esta inclusión de información suele denominarse marca de agua. Una marca de agua es una alteración de un documento que puede incluir una imagen o patrón identificativo, como el espaciado o la deformación de caracteres en el caso del texto, o el desplazamiento de píxeles en frecuencia o espacio en el caso de las imágenes. Se pretende que estas alteraciones no sean perceptibles, pero hasta ahora ningún método propuesto resuelve el problema de minimizar los efectos de alteración para una amplia gama de contenidos de documentos diferentes (por ejemplo, documentos con tablas, listas, sangrías diferentes, ecuaciones matemáticas, índice, etc.).

30

35

En el campo de la marca de agua digital para imágenes y vídeo son bien conocidos, por ejemplo, los métodos basados en la frecuencia que lo hacen de forma indirecta, donde la información se incrusta en los componentes de alta frecuencia. En estos componentes de alta frecuencia, el sistema visual humano (SVI) no es capaz de detectar las modificaciones suaves, luego la perceptibilidad de la información incrustada es muy baja.

40

Por otro lado, todos los algoritmos de marca de agua digital para documentos de texto existentes en el estado de la técnica y las patentes, concretamente los basados en el desplazamiento de los elementos como caracteres, palabras o líneas, no muestran ningún mecanismo para hacer frente a ese tipo de problemas, y todos los métodos parten del supuesto de que los documentos se han elaborado únicamente con texto regular, espaciado uniforme entre elementos y otras propiedades que no existen en el mundo real. Cuando estos métodos se utilizan en documentos reales existen altas probabilidades de conseguir efectos visuales no deseados en los documentos utilizados en el proceso de incrustación (principalmente desalineaciones en elementos como índices, tablas y listas, o incluso solapamientos entre elementos como caracteres o palabras). Esta limitación es la principal razón por la que los algoritmos de marca de agua digital de texto no han dado el salto al mercado como lo han hecho con otros tipos de formatos multimedia, tal como imágenes o vídeos.

45

50

Además, la selección errónea de los elementos que se utilizarán en el proceso de incrustación no solo afecta a la parte visual, sino que también puede afectar a la capacidad del sistema para extraer la información. La razón principal es que los algoritmos extraen información de partes en las que el espaciado (información) no es el esperado, de modo que estos espacios inesperados actúan como ruido sobre el resto de la información extraída, reduciendo las posibilidades de extraer información.

55

Un tercer efecto a tener en cuenta es la cantidad de información que el sistema cree que puede incluirse en el documento. La cantidad de información que se puede incrustar está relacionada con la cantidad de texto del documento (páginas para ser más exactos). Si los algoritmos cuentan texto que no es válido para ocultar información en él, la cantidad de información que se puede extraer es menor de la que espera el algoritmo, y es probable que los resultados al extraer la información sean inesperados.

60

Por lo tanto, es muy conveniente desarrollar un método para detectar las regiones que quedarán excluidas del proceso de incrustación de los algoritmos de marca de agua de texto digital por las razones mencionadas: 1) reducir los

artefactos visuales no deseados en los documentos incrustados, 2) aumentar la capacidad del sistema para extraer la información de los documentos con marcas de agua anteriores y 3) conseguir datos más precisos sobre cuánta información se puede insertar y extraer en un contenido específico teniendo en cuenta toda la diversidad que ofrece el contenido de los documentos de texto.

5 Se conocen algunas patentes y solicitudes de patente en este campo.

10 El documento US2020042785A1 divulga métodos, sistemas y productos de programas informáticos para el reconocimiento de tablas en documentos PDF. Un método implementado por ordenador incluye la discretización de una o más áreas contiguas de un documento PDF; identificar una o más líneas separadoras de espacios en blanco dentro de una o más áreas contiguas discretizadas del documento PDF; detectar una o más regiones de tabla candidatas dentro de las una o más áreas contiguas discretizadas del documento PDF agrupando las una o más líneas separadoras de espacios en blanco en una o más cuadrículas; y la salida de al menos una de las regiones candidatas de la tabla como tabla finalizada de acuerdo con las puntuaciones asignadas a cada una de las una o más regiones candidatas de la tabla en función de (i) la información sobre los bordes y (ii) la información sobre la estructura de las celdas.

20 El documento EP2807608A1 proporciona un motor de detección de tablas sin bordes y un método asociado para identificar las tablas sin bordes que aparecen en los datos extraídos de un documento de formato fijo. Debido a la falta de bordes visibles, la detección automatizada fiable de una tabla sin bordes es difícil. El motor de detección de tablas sin bordes utiliza los espacios en blanco, en lugar del contenido, para detectar los candidatos a tabla sin bordes. Aplicando un análisis heurístico, el motor de detección de tablas sin bordes descarta las candidatas a tabla sin bordes con un diseño que carece de las características suficientes de una tabla y es poco probable que sea una tabla sin bordes válida.

25 Estas dos soluciones del estado de la técnica solo identifican tablas en los archivos PDF de entrada, y no otros elementos de texto, como listas con viñetas, texto justificado, texto en imágenes, etc. Además, hacen uso de la heurística para la identificación de tablas, y no de elementos AI, lo que ha demostrado ser mucho más eficaz en la solución de este tipo de problemas.

30 El documento US10817717B2 se refiere a un método y a un dispositivo para analizar una tabla en una imagen de documento. El método comprende las siguientes etapas: introducir en el dispositivo electrónico una imagen de documento que se va a analizar y que incluye una o varias zonas de tablas; detectar, mediante el dispositivo electrónico, un área de mesa en la imagen del documento utilizando un modelo de detección de tablas entrenado previamente; detectar, mediante el dispositivo electrónico, los bloques de texto internos incluidos en el área de la tabla utilizando un modelo de detección de texto preentrenado; determinar, mediante el dispositivo electrónico, una estructura espacial de la tabla; y realizar el reconocimiento de texto en un bloque de texto de cada celda de acuerdo con la estructura espacial de la tabla, a fin de obtener datos estructurados editables mediante el análisis sintáctico. El método y el dispositivo de la presente solicitud pueden aplicarse a diversas tablas, como tablas con líneas incluidas o tablas con líneas excluidas o tablas en blanco y negro. Esta solución solo identifica tablas en los archivos del documento de entrada, y ningún otro elemento de texto, como listas con viñetas, texto justificado, texto en imágenes, etc.

45 El documento US9697423B1 proporciona un método para el tratamiento de imágenes que incluye obtener una imagen que incluye una tabla; identificar una primera pluralidad de líneas geométricas en la imagen; agrupar la primera pluralidad de líneas geométricas en una pluralidad de conglomerados; determinación de una pluralidad de líneas dibujadas a mano en la imagen correspondiente a la tabla a partir de la pluralidad de conglomerados; calcular una pluralidad de puntos para la pluralidad de líneas trazadas a mano; y determinar una geometría de la tabla basada en la pluralidad de puntos. Como las soluciones indicadas anteriormente, este método solo identifica tablas en los archivos PDF de entrada, y no otros elementos de texto, como listas con viñetas, texto justificado, texto en imágenes, etc. Además, hace uso de la heurística para la identificación de tablas, y no de elementos AI.

50 El documento WO2017160654A3 divulga un método para extraer datos de archivos PDF. El método incluye la identificación de al menos un identificador de documento asociado a un primer documento de un archivo PDF. El método incluye además la determinación, mediante el al menos un identificador de documento, de un identificador de punto de referencia para identificar un punto de referencia en el primer documento, un valor de desplazamiento para indicar una ubicación de una primera área de detección en el primer documento, e información de tamaño para indicar un tamaño de la primera área de detección en el primer documento. El método también incluye la identificación, mediante un identificador de punto de referencia, del punto de referencia en el primer documento. El método incluye además la identificación, mediante el valor de desplazamiento y la información sobre el tamaño, de la primera área de detección en el primer documento y la extracción, mediante el procesamiento de datos binarios del archivo PDF, de datos dentro de la primera área de detección del primer documento.

Otras soluciones o estrategias se conocen mediante los documentos US9348848B2, US9268999B2, US8645819B2,

US2008084573A1, US10445615B2 y US10592738B2. Otros ejemplos de divulgaciones del estado de la técnica relacionadas con la materia reivindicada incluyen:

5 HUIJUAN YANG ET AL: "Text document authentication by integrating inter character and word spaces watermarking", 2004 IEEE International Conference on Multimedia and Expo, Junio 27 - 30, 2004, IEEE Operations Center, vol. 2, publicado el 27 de junio de 2004, páginas 955-958, XP010770979, DOI: 10.1109/ICME.2004.1394360, ISBN: 978-0-7803-8603-7 y

10 DUAN XINTAO ET AL: "Reversible Image Steganography Scheme Based on a U-Net Structure", IEEE Access, vol. 7, páginas 9314-9323, XP011707115, DOI: 10.1109/ACCESS.2019.2891247, publicado el 29-01-2019

Por lo tanto, se necesitan métodos más eficaces y rápidos, basados en técnicas AI o de visión por ordenador, capaces de tratar las distorsiones geométricas en una página de un documento, con documentos que incluyan texto o no, y/o con PDF vectoriales.

15 **Descripción de la invención**

Para ello, la presente invención propone, según un aspecto, un método para detectar las regiones del documento que se excluirán de un proceso de incrustación. El método comprende convertir, mediante un módulo adaptador, al menos una página de un documento recibido en una representación visual del mismo, manteniendo la representación visual la posición de los caracteres de la al menos una página; recibir, mediante un detector de texto, la representación visual; procesar, por parte del detector de texto, la representación visual mediante una red neuronal profunda que predice las regiones no válidas del documento correspondientes a un texto no válido que incluye una tabla y/o una lista de viñetas, y, por parte del detector de texto, una lista de regiones no válidas con sus posiciones de página asociadas como resultado del procesamiento; y excluir, mediante un módulo de incrustación de marcas de agua o mediante un módulo de extracción de marcas de agua, la lista de regiones no válidas para proporcionar un documento con marca de agua o un mensaje incrustado en el documento.

30 Según la invención, el documento puede ser un documento digital o digitalizado.

En una realización, cada región inválida del documento de la lista de regiones inválidas tiene asociada una puntuación de confianza.

35 En una realización, la representación visual comprende una imagen.

En una realización, la imagen es una imagen binaria en la que los píxeles asociados a los caracteres se representan utilizando un primer color y en la que los píxeles asociados al fondo se representan utilizando un segundo color. En algunas realizaciones, el primer color puede ser blanco y el segundo color puede ser negro.

40 En una realización, la al menos una página está en formato de imagen, en la que la etapa de conversión comprende además la corrección de una inclinación de la página mediante el uso de un algoritmo de reconocimiento. Por ejemplo, el algoritmo de reconocimiento puede ser la transformada de Hough.

45 En una realización, antes de llevar a cabo la etapa de corrección, el método comprende el preprocesamiento de la al menos una página mediante la aplicación de un algoritmo de detección de bordes Canny y diferentes operaciones morfológicas sobre la al menos una página. Además, puede realizarse un redimensionamiento de la al menos una página a una resolución inferior.

50 En una realización, la posición de los caracteres puede obtenerse mediante el uso de un algoritmo de segmentación basado en la imagen, por ejemplo un algoritmo basado en la proyección.

En una realización, el detector de texto en la etapa de procesamiento utiliza además un algoritmo de aprendizaje supervisado. La red neuronal profunda puede basarse en una arquitectura basada en la detección o en una arquitectura basada en la segmentación.

55 En una realización, la lista de regiones no válidas comprende un conjunto de cajas donde cada caja está definida por: un valor "x" que indica una esquina superior izquierda de la caja en una coordenada horizontal, un valor "y" que indica una esquina superior izquierda de la caja en una coordenada vertical, un valor "w" que indica una anchura de la caja en píxeles, y un valor "h" que indica una altura de la caja en píxeles.

60 Otras realizaciones de la invención que se desvelan en el presente documento también incluyen programas de software para realizar las etapas y operaciones de la realización del método resumidas anteriormente y desveladas en detalle a continuación. Más particularmente, un producto de programa informático es una realización que tiene un medio legible por ordenador que incluye instrucciones de programa informático codificadas en el mismo que, cuando

se ejecutan en al menos un procesador en un sistema informático, provocan que el procesador realice las operaciones indicadas en el presente documento como realizaciones de la invención.

5 Por lo tanto, la invención garantiza un resultado más eficaz gracias al uso de AI en lugar de sólo heurística y un tiempo de procesamiento más rápido gracias al uso de mapas digitales en páginas rasterizadas. Además, la invención es capaz de hacer frente a las distorsiones geométricas de la página; documentos que incluyan o no texto, PDF vectoriales; identificar y etiquetar tanto las estructuras del texto como los elementos del documento, como tablas, listas con viñetas, texto justificado, pies de imagen, imágenes y gráficos, etc.

10 Breve descripción de los dibujos

Las anteriores y otras ventajas y características se entenderán más completamente a partir de la siguiente descripción detallada de las realizaciones, con referencia a las figuras adjuntas, que deben considerarse de una manera ilustrativa y no limitante, en las que:

15 La figura 1 muestra la arquitectura para implementar un método de detección de regiones del documento que se excluirán de un proceso de incrustación, según una realización.

20 La figura 2 muestra otra realización de una arquitectura para aplicar el método propuesto.

La figura 3 muestra un ejemplo de una página y sus respectivos mapas que utilizará el detector de texto. La figura 3A ilustra la página del documento; La figura 3B ilustra el mapa creado con el adaptador de entrada y la imagen utilizada como entrada por el detector de texto; La figura 3C ilustra el solapamiento entre la página y el mapa.

25 La figura 4 muestra una realización de la arquitectura utilizada para crear la representación visual, o imagen cartográfica.

La figura 5 muestra una realización del módulo de corrección de la inclinación.

30 La figura 6 representa gráficamente un ejemplo de detección.

La figura 7 muestra gráficamente una comparación de un algoritmo de marca de agua sin (imagen superior) y con (imagen inferior) el uso de la presente invención.

35 Descripción detallada de realizaciones preferidas

La figura 1 muestra una realización de la arquitectura en la que el método propuesto puede utilizarse como parte de un proceso de marca de agua de texto digital. Un módulo adaptador 18 convierte una o varias páginas de un documento recibido 16 (por ejemplo, un documento digital o un documento digitalizado) en una representación visual 40 10 de las posiciones del contenido de la página, que debe mostrar la representación visual de la página, o cualquier otra representación que mantenga la posición de las palabras de la página. Un detector de texto no válido (o simplemente detector de texto) 12 toma como entrada la representación visual 10 y la procesa utilizando uno o varios algoritmos de inteligencia artificial. El detector de texto 12 devuelve como salida una lista de regiones no válidas 14 que deben eliminarse de un proceso de incrustación. La lista de regiones no válidas 14 lleva asociada su posición en la página y, en particular, una puntuación que indica el nivel de confianza otorgado por el método. La lista de regiones no válidas 14 se utiliza como entrada del módulo de incrustación de marcas de agua 22, que también recibe como entradas el documento 16 y el mensaje a incrustar 20. El módulo de incrustación de marcas de agua 22 puede tener en cuenta la lista de regiones no válidas 14 para crear un documento con marca de agua 24; consiguiendo así mejores resultados relacionados con el aspecto visual y la solidez de la información incrustada.

50 Del mismo modo, la figura 2 muestra el uso de la invención en un proceso de extracción. El documento de entrada 16 puede ser procesado por el módulo adaptador 18 para adaptar el documento 16 a la entrada requerida por el detector de texto 12, que devuelve la lista de regiones no válidas 14 que pueden ser tenidas en cuenta por un módulo de extracción de marcas de agua 26, que también recibe el documento 16, y devuelve como salida un mensaje 28 incrustado en el documento 16.

60 Es decir, el módulo adaptador 18 se utiliza para adaptar el documento 16 a la entrada requerida por el detector de texto 12. En algunas realizaciones, la entrada (es decir, la representación visual 10) puede ser un formato de imagen, y el contenido de esta imagen puede ser cualquier tipo de información que represente la posición de las palabras. Puede ser la página visual tal cual o puede ser otro tipo de representación.

En una realización, se propone el uso de mapas abstractos como representación visual 10. En concreto, estos mapas abstractos son imágenes binarias con fondo a píxeles negros y palabras representadas como rectángulos blancos. La figura 3 muestra un ejemplo de página y su respectivo mapa que utilizará el detector de texto 12. En la figura 3 también

se puede ver el solapamiento entre la página y el mapa utilizado para comprender mejor lo que representa la entrada utilizada. La razón principal para trabajar con este tipo de representaciones es evitar el proceso de rasterización cuando los documentos digitales están en formato vectorial (conversión de formato vectorial a imagen) o evitar el uso de algún método de Reconocimiento Óptico de Caracteres (OCR) en el caso de documentos rasterizados (la invención no requiere conocer el valor de los caracteres o palabras del contenido, lo que, en realidad, no añade valor, ya que un carácter "normal" puede permanecer tanto en un área de texto seleccionable como en un área de texto no válido como una tabla, lista de viñetas, etc.). Ambos procesos mencionados (rasterización o métodos OCR) consumen mucho tiempo en comparación con el tiempo adicional de los sistemas de marca de agua digital, luego, al utilizar esos mapas en lugar de páginas rasterizadas el tiempo total necesario para procesar documentos vectoriales y rasterizados será menor. Por ejemplo, el tiempo necesario en un ordenador Intel(R) Core(TM) i7-6700K CPU @ 4,00 GHz y 16 Gb RAM para rasterizar una página es de aproximadamente ~300 ms para una resolución de 300 dpi, y el tiempo consumido por el conocido software Tesseract OCR para la misma resolución es de unos ~1500 ms. Las técnicas de marca de agua digital de texto conocen la posición de los caracteres para poder cambiar sus posiciones, por lo que la creación de una imagen de mapa utilizando esta información como la que proponemos puede realizarse en unos pocos milisegundos.

La figura 4 muestra una realización de cómo puede crearse la representación visual 10. Según esta realización, el módulo adaptador 18 puede procesar pdf vectoriales de cualquier tipo de imagen (que incluye pdf rasterizados o de cualquier tipo el formato que utiliza imágenes). Dependiendo del tipo de entrada (vectorial o imagen) 30 el método puede variar para lograr el resultado esperado. La salida de este módulo es una imagen $M(r,c)$ 10 donde r es la fila r -ésima, c es la columna c -ésima, y $M(r,c)$ es la imagen binarizada del documento.

Quando la página está en formato de imagen, hay múltiples maneras de obtener la posición de las palabras utilizando un enfoque de segmentación basado en la imagen 34. Los métodos de reconocimiento óptico de caracteres suelen devolver la posición de las palabras, pero en esta realización se propone preferentemente un método basado en proyecciones, ya que ofrece mejores respuestas temporales. Antes de aplicar el método basado en proyecciones se corrige la inclinación 32 de la imagen mediante un algoritmo de reconocimiento, por ejemplo la transformada de Hough.

La figura 5 muestra una realización del algoritmo de corrección de la inclinación 32. En primer lugar se preprocesa el documento 16 recibido en formato de imagen. La página podría digitalizarse mediante un escáner o con un dispositivo móvil, donde el color y las distorsiones geométricas encontradas por el algoritmo pueden ser muy variados. Para abordar este problema se realiza una etapa de preprocesamiento. Consiste en una primera etapa en la que se aplica un algoritmo de detección de bordes Canny 40. La detección de bordes Canny es un algoritmo de varias etapas que puede detectar bordes suprimiendo el ruido al mismo tiempo. En primer lugar, la imagen se suaviza con un filtro gaussiano para reducir el ruido y los detalles y texturas no deseados:

$$g(r,c) = G(r,c) * I(r,c),$$

donde el filtro gaussiano G se define como:

$$G = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{r^2+c^2}{2\sigma^2}\right)$$

El Gradiente de $g(r,c)$ se calcula como:

$$M(r,c) = \sqrt{g_x^2(r,c) + g_y^2(r,c)},$$

y

$$\theta(r,c) = \tan^{-1}\{g_x(r,c)/g_y(r,c)\}.$$

Se aplica un umbral T a $M(r,c)$:

$$M_T = \begin{cases} M(r,c) & \text{if } M(r,c) > T \\ 0 & \text{de lo contrario} \end{cases},$$

donde T se elige de forma que se conserven todos los elementos de borde y se suprima la mayor parte del ruido.

A continuación, el método suprime los píxeles no máximos en los bordes en la $M_T(r,c)$ obtenida anteriormente para adelgazar las crestas de los bordes. Para ello, el método comprueba si cada $M_T(r,c)$ distinto de cero es mayor que

sus dos vecinos a lo largo de la dirección del gradiente $\theta(r,c)$. Si es así, MT (r,c) no se modifica; si no, se pone a 0.

El resultado anterior puede umbralizarse mediante dos umbrales diferentes ζ_1 y ζ_2 (donde $\zeta_1 < \zeta_2$) para obtener dos imágenes binarias T_1 y T_2 . Observe que T_2 con mayor ζ_2 tiene menos ruido y menos bordes falsos pero mayores huecos entre los segmentos de borde, en comparación con T_1 con menor ζ_1 .

El método une los segmentos de borde en T_2 para formar bordes continuos. Para ello, el método rastrea cada segmento en T_2 hasta su final y luego busca sus vecinos en T_1 para encontrar cualquier segmento de borde en T_1 que sirva de puente hasta alcanzar otro segmento de borde en T_2 . La salida del algoritmo de detección de bordes Canny 40 se denota aquí como $I^c(r,c)$.

Las siguientes etapas son las operaciones morfológicas para unir y rellenar los huecos entre caracteres y palabras e intentar convertir las líneas de texto en líneas rectas. Esto se realiza en un primer momento con una operación de dilatación 42, y posteriormente con una operación de erosión 44. Estas operaciones morfológicas proporcionan una imagen $I^m(r,c)$ más limpia con líneas rectas más fáciles de procesar con la transformada de Hough.

Como la transformada de Hough puede ser lenta, es preferible redimensionar la imagen 46 para trabajar con una imagen de menor resolución (75 ppp funciona bien). Esto permite conseguir el ángulo correcto para desestabilizar la imagen al mismo tiempo que reduce el tiempo necesario para obtener el resultado.

La transformada de Hough 50 se aplica sobre la imagen redimensionada para detectar los ángulos de las líneas de texto. La transformada de Hough 50 convierte el espacio euclidiano en el espacio de Hough. El espacio de Hough es un plano 2D que tiene un eje horizontal que representa la pendiente y un eje vertical que representa la intercepción de una línea en la imagen de borde. Una línea en una imagen de borde se representa en la forma $y = ax + b$ (Hough, 1962). Una línea en la imagen de borde produce un punto en el Espacio de Hough, ya que una línea se caracteriza por su pendiente y su intercepto b. Normalmente, el Espacio de Hough utiliza una forma diferente de representar las líneas rectas llamada líneas normales que pasa por el origen y es perpendicular a esa línea recta. La forma de la recta normal es $\rho = x \cos(\theta) + y \sin(\theta)$ donde ρ es la longitud de la recta normal y θ es el ángulo entre la recta normal y el eje x. De este modo, cada punto de borde (x_i, y_i) genera ahora una curva coseno en el espacio de Hough en lugar de una línea recta. Una línea en la imagen de borde sigue produciendo un punto en el espacio de Hough. Las líneas se detectan buscando estos puntos en el espacio de Hough, para esta tarea debe fijarse un umbral Th . Si el umbral Th es bajo, el número de líneas obtenidas será elevado y podría resultar difícil determinar la dirección más probable de las líneas reales. Para evitar este problema, se propone una estrategia recursiva. El umbral se fija en un valor alto, si las líneas encontradas son inferiores a 10, se repite el proceso con un nuevo umbral calculado como $Th_{i+1} = \frac{3}{4} * Th_i$, donde i indica el número de la iteración. Se pueden utilizar otros valores diferentes a 10, pero se ha comprobado empíricamente que este número funciona realmente bien. Cuando el número de líneas es superior a 10, el proceso se detiene. La condición de parada 62 también debe evitar los bucles infinitos (es posible encontrar menos de 10 líneas en una página), por lo que el método sólo permite repetir el proceso un número finito de veces.

En un caso normal, tras aplicar la transformada de Hough 50, se obtienen casi diez líneas. El método obtiene el vector de ángulos θ_s de estas líneas y calcula el histograma de estos ángulos. El máximo del histograma 54 se toma como el ángulo estimado (ángulo de inclinación) θ_f .

El enderezamiento 60 se realiza simplemente tomando $-\theta_f$ y girando la imagen. Tras la rotación, la imagen obtenida $Irot(r',c')$ tiene más anchura y altura que la imagen de entrada. Para corregir esta distorsión se aplica la operación de recorte sobre la imagen. Una forma fácil de hacerlo es calcular el nuevo punto de origen (r_0, c_0) como:

$$r_0 = R' - \left\lfloor \frac{R'}{2} \right\rfloor$$

$$c_0 = C' - \left\lfloor \frac{C'}{2} \right\rfloor$$

donde R' y C' son el número de filas y columnas de la imagen $Irot(r',c')$, y R y C son el número de filas y columnas de la imagen de entrada $I(r,c)$. El recorte se realiza sobre la imagen $Irot(r',c')$ en la región delimitada por las filas $[r_0, R]$ y las columnas $[c_0, C]$.

Tras esta operación de recorte se obtiene la salida del módulo de corrección de la inclinación 32; es decir, una imagen $I'(r,c)$ en la que se ha corregido la rotación de la página y con el mismo tamaño de la imagen original. Esto permite que la segmentación basada en imágenes alcance una gran precisión. El rendimiento de este módulo 32 de corrección de la inclinación se comparó con otros métodos del estado de la técnica, como el método Leptonica (Dan S. Bloomberg *et al.* Measuring document image skew and orientation. In Document Recognition II (páginas 302 - 316). SPIE.), mostrando un mejor rendimiento en la precisión del ángulo obtenido, el tiempo necesario para realizar la estimación y

el rango de ángulos que el algoritmo puede corregir.

Tras corregir la inclinación de la imagen, se puede aplicar la segmentación basada en la imagen 34. Como ya se ha explicado, se puede utilizar un método basado en proyecciones. El método comienza binarizando el documento en formato de mapa de bits $I(r,c)$ y sumando los píxeles en dirección horizontal, la proyección horizontal $Ph(r)$ se obtiene como:

$$Ph(r) = \sum_c I^b(r,c),$$

10 donde la imagen binarizada se denota como $I^b(r,c)$, r es la fila r -ésima, c es la columna c -ésima, e $I(r,c)$ es la imagen binarizada del documento.

Fijando un valor de umbral adecuado, el documento en formato de mapa de bits puede segmentarse en líneas de texto, encontrando umbrales de binarización óptimos a partir del histograma de la imagen. Una vez calculadas las proyecciones e identificadas las líneas de texto, se segmenta cada una de ellas. Para cada línea, su proyección vertical $Pvi(c)$ corresponde a:

$$Pvi(c) = \sum_i I^b(r,c),$$

20 donde, $I^b(r,c)$ es la imagen binarizada del documento recortada a la i -ésima línea.

Aplicando la misma técnica utilizada para aislar la línea de texto mediante la proyección horizontal $Ph(r)$, se pueden identificar los espacios de cada línea y medirlos aplicando un umbral adecuado a la proyección vertical $Pvi(c)$.

25 El resultado de la segmentación 34 es un conjunto de longitudes de palabra y un conjunto de sus ubicaciones correspondientes en el documento, denotados como conjunto W y conjunto L respectivamente 38.

Si el documento está en formato vectorial, el sistema puede leer la información del Lenguaje de Descripción de Páginas (PDL) basado en la norma ISO 32000 conocida como PDF utilizando un intérprete de PDF 36. Un documento PDF consiste en una colección de objetos que, en conjunto, describen el aspecto de una o varias páginas, posiblemente acompañados de elementos interactivos adicionales y datos de aplicación de nivel superior. El sistema lee la sección de tráiler del archivo, obtiene la información de la raíz y, a continuación, lee el objeto de páginas. Dentro de los objetos de página se encuentran los objetos de contenido de la página. El contenido puede tener objetos de texto que consisten en operadores que pueden mostrar cadenas de texto, mover la posición del texto y establecer el estado del texto y algunos otros parámetros. El sistema es capaz de leer e interpretar todos los operadores de estado del texto, los operadores de posicionamiento del texto y los operadores de visualización del texto para obtener el conjunto de longitudes de las palabras y un conjunto de sus ubicaciones correspondientes en el documento, denotados como conjunto W y L , respectivamente. En este caso, la operación resulta más rápida.

40 Los valores W y L 38 se utilizan para crear una imagen cartográfica (binaria) $M(r, c)$ 10 en el módulo de creación de imágenes binarias 39, en la que los píxeles de las regiones asociadas a la ubicación L y las longitudes de palabra W se establecen en blanco, y el resto de los píxeles en negro. L es un vector de posiciones de píxeles $\{(lx1, ly1), (lx2, ly2) \dots (lxn, lyn)\}$ que indica la esquina inferior izquierda de la palabra, y W es un vector de pares de valores que indican la anchura y la altura de una palabra $\{(w1, h1), (w2, h2) \dots (wn, hn)\}$ donde la longitud del vector L y W es la misma ya que cada par está relacionado con una palabra de la página. El mapa $M(r,c)$ se crea con el mismo tamaño que la imagen de entrada $I(r,c)$ y con todos sus píxeles con valor cero (píxeles negros). Entonces, para cada valor $i \in [1, n]$ las regiones delimitadas por $[L_i, L_i + W_i]$ se ponen en blanco (esto es lo mismo que la región $[(lx_i, ly_i), (w_i, h_i)]$ para todo $i \in [1, n]$). Un ejemplo del mapa de entrada puede verse en la figura 3 y en la figura 6.

50 Una vez generada la entrada al detector de texto 12, en una realización se utiliza una red neuronal profunda para predecir en el proceso de inferencia las regiones correspondientes al texto no válido. En particular, se propone el uso de una red de pirámide de características (FPN), aunque pueden utilizarse igualmente otras redes, como un detector de disparo múltiple (SSD), entre otras. La red FPN funciona bien con objetos de diferentes escalas, lo que puede ser útil para los datos procesados aquí, en los que el tamaño del objeto depende del tamaño de la lista, la tabla, etc. Por supuesto, estos modelos tienen que ser entrenados para aprender las ponderaciones de la red antes de ser aplicados en la etapa de inferencia. Para ello se utiliza el aprendizaje supervisado, por lo que los datos utilizados deben estar anotados en etapas anteriores.

60 Una posible aplicación de la invención puede basarse en el uso de la Pérdida Focal y la utilización de tareas de Aprendizaje por Transferencia, tomando como referencia el conjunto de datos COCO. La pérdida focal evita que el gran número de negativos fáciles abrume al detector de texto 12 durante el entrenamiento. La invención también puede utilizar las arquitecturas Efficientnet o Mobilenet como red troncal para reducir los términos de cómputo.

La salida de la red será el conjunto de regiones no válidas 14 (un conjunto de cajas que representan regiones no

válidas) B que representa detecciones múltiples, donde cada caja está definida por valores $\{x,y, w, h\}$ y x indica la esquina superior izquierda en la coordenada horizontal, y indica la esquina superior izquierda de la caja en la coordenada vertical, w indica la anchura en píxeles y h indica la altura de la caja en píxeles. La salida también devuelve una puntuación asociada a cada casilla. Esta puntuación indica el nivel de confianza de la detección (figura 6).

5 Cabe señalar que el problema puede resolverse como un problema de detección de objetos o de segmentación de imágenes. En este caso, en particular, se considera el escenario del objeto de detección. La segmentación de imágenes es una ampliación de la detección de objetos en la que el método marca la presencia de un objeto mediante máscaras de píxeles generadas para cada objeto de la imagen. Esta técnica es más granular que la generación de cuadros delimitadores porque puede ayudar a determinar la forma de cada objeto presente en la imagen. Esta granularidad ayuda en diversos campos como el procesamiento de imágenes médicas, las imágenes por satélite, etc., pero es menos relevante para las tareas relacionadas aquí, en las que las cajas delimitadoras rectangulares son suficientes para cubrir los objetivos del sistema, ya que hemos corregido la distorsión geométrica de las páginas.

15 Una de las aportaciones más importantes de la presente invención es cuando estas predicciones se utilizan como parte de un proceso de marca de agua de texto digital basado en desplazamientos de los elementos del texto. La solución también puede utilizarse para detectar objetos específicos en documentos de texto, como tablas, etc., de forma similar a los algoritmos más avanzados, pero con un tiempo de procesamiento mucho más rápido. La solución puede adaptarse a distintos algoritmos, ya que las zonas no deseadas se etiquetan correctamente en la etapa de entrenamiento de la red neuronal. Se puede realizar una adaptación específica utilizando el método de marca de agua descrito en el documento EP3477578B1, de los mismos autores de la presente invención.

25 En cuanto a la figura 7, muestra la salida del algoritmo de marca de agua con y sin el uso del método propuesto. La línea "91" muestra un espaciado incorrecto entre el "91" y la palabra "Sección". La segunda imagen muestra una alineación perfecta de los contenidos, haciendo imperceptibles los cambios realizados sobre la posición del contenido del documento.

30 La presente invención se ha descrito con especial detalle con respecto a posibles realizaciones específicas. Los expertos en la materia apreciarán que la invención se puede poner en práctica en otras realizaciones. Por ejemplo, la nomenclatura utilizada para los componentes, el uso de mayúsculas en las designaciones y términos de los componentes, los atributos, las estructuras de datos o cualquier otro aspecto de programación o estructural no es significativo, obligatorio ni limitante, y los mecanismos que implementan la invención o sus características pueden tener varios nombres, formatos y/o protocolos diferentes. Además, el sistema y/o la funcionalidad de la invención pueden implementarse mediante diversas combinaciones de software y hardware, como se ha descrito, o enteramente en elementos de software. Asimismo, las divisiones particulares de funcionalidad entre los diversos componentes aquí descritos son meramente de ejemplo, y no obligatorias ni significativas. En consecuencia, las funciones realizadas por un único componente pueden, en otras realizaciones, ser realizadas por múltiples componentes, y las funciones realizadas por múltiples componentes pueden, en otras realizaciones, ser realizadas por un único componente.

40 Ciertos aspectos de la presente invención incluyen etapas de proceso u operaciones e instrucciones descritas en el presente documento de forma algorítmica y/o similar a un algoritmo. Debe tenerse en cuenta que las etapas del proceso y/o las operaciones e instrucciones de la presente invención pueden plasmarse en software, firmware y/o hardware, y cuando se plasman en software, pueden descargarse para residir en y ser operados desde diferentes plataformas utilizadas por sistemas operativos de red en tiempo real.

45 El alcance de la presente invención se define en el siguiente conjunto de las reivindicaciones.

REIVINDICACIONES

1. Un método para detectar regiones de documentos que se excluirán de un proceso de incrustación, que comprende:
 - 5 convertir, mediante un módulo adaptador (18), al menos una página de un documento recibido (16) en una representación visual (10) del mismo, manteniendo la representación visual (10) la posición de los caracteres de la al menos una página;
 - recibir, mediante un detector de texto (12), la representación visual (10);
 - 10 procesar, mediante el detector de texto (12), la representación visual (10) utilizando una red neuronal profunda que prediga las regiones inválidas del documento correspondientes a un texto inválido que incluya una tabla y/o una lista de viñetas;
 - devolver, mediante el detector de texto (12), una lista de regiones no válidas (14) con sus posiciones de página asociadas como resultado del procesamiento; y
 - 15 excluir, mediante un módulo de incrustación de marcas de agua (22) o mediante un módulo de extracción de marcas de agua (26), la lista de regiones no válidas (14) para proporcionar un documento con marca de agua (24) o un mensaje (28) incrustado en el documento (16).

2. El método de la reivindicación 1, en el que el documento (16) comprende un documento digital o digitalizado, y en el que cada región inválida del documento de la lista de regiones inválidas (14) tiene asociada una puntuación de confianza.

- 20 3. El método de la reivindicación 1 o 2, en el que la representación visual (10) comprende una imagen.

4. El método de la reivindicación 3, en el que la imagen es una imagen binaria en la que los píxeles asociados a los caracteres se representan utilizando un primer color y en la que los píxeles asociados al fondo se representan utilizando un segundo color.

- 25 5. El método de la reivindicación 4, en el que el primer color es blanco y el segundo negro.

- 30 6. El método de una cualquiera de las reivindicaciones anteriores, en el que la al menos una página está en formato de imagen, y en el que la etapa de conversión comprende además la corrección de una inclinación de la página mediante el uso de un algoritmo de reconocimiento.

- 35 7. El método de la reivindicación 6, en el que el algoritmo de reconocimiento comprende una transformada de Hough.

8. El método de la reivindicación 6 o 7, en el que antes de realizar la etapa de corrección, el método comprende el preprocesamiento de la al menos una página mediante la aplicación de un algoritmo de detección de bordes Canny y diferentes operaciones morfológicas sobre la al menos una página.

- 40 9. El método de la reivindicación 8, que comprende además redimensionar la al menos una página a una resolución inferior.

10. El método de una cualquiera de las reivindicaciones anteriores 6 a 9, que comprende además la obtención de la posición de los caracteres mediante el uso de un algoritmo de segmentación basado en la imagen, comprendiendo este último un algoritmo basado en la proyección.

- 45 11. El método de una cualquiera de las reivindicaciones anteriores, en el que el detector de texto (12) en la etapa de procesamiento utiliza además un algoritmo de aprendizaje supervisado.

- 50 12. El método de la reivindicación 1, en el que la lista de regiones no válidas (14) comprende un conjunto de cajas en el que cada caja está definida por: un valor "x" que indica una esquina superior izquierda de la caja en una coordenada horizontal, un valor "y" que indica una esquina superior izquierda de la caja en una coordenada vertical, un valor "w" que indica una anchura de la caja en píxeles, y un valor "h" que indica una altura de la caja en píxeles.

- 55 13. El método de la reivindicación 1, en el que la red neuronal profunda se basa en una arquitectura basada en la detección.

14. El método de la reivindicación 1, en el que la red neuronal profunda se basa en una arquitectura basada en la segmentación

- 60 15. Un medio no transitorio legible por ordenador que contiene instrucciones de ordenador almacenadas en el mismo para hacer que un procesador de ordenador realice un método según cualquiera de las reivindicaciones 1 a 14.

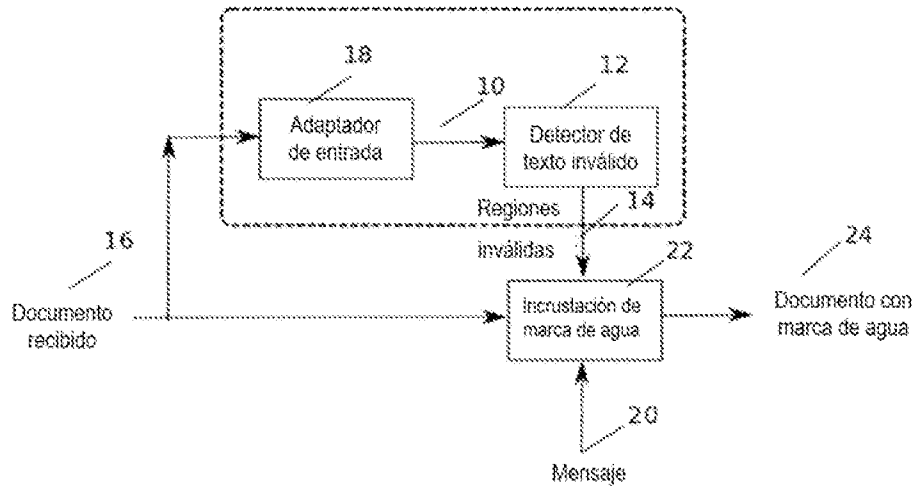


Fig. 1

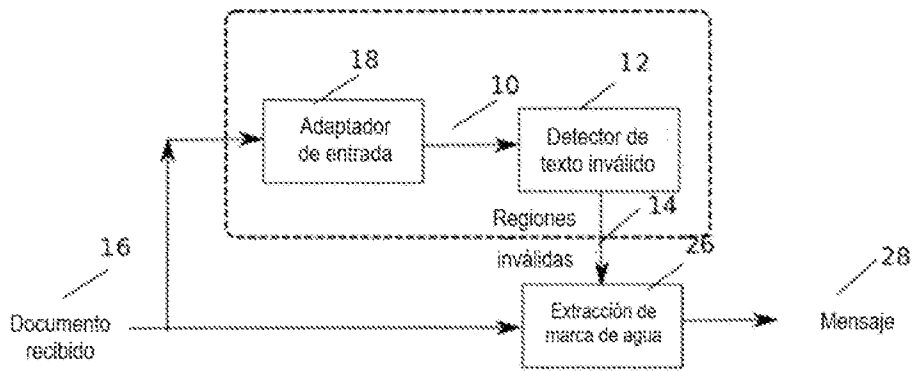
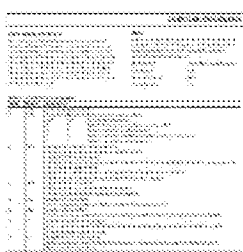


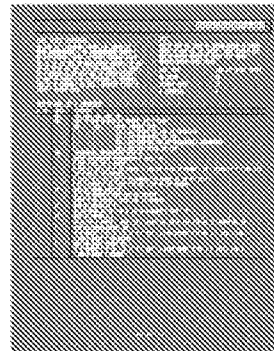
Fig. 2



(A)



(B)



(C)

Fig. 3

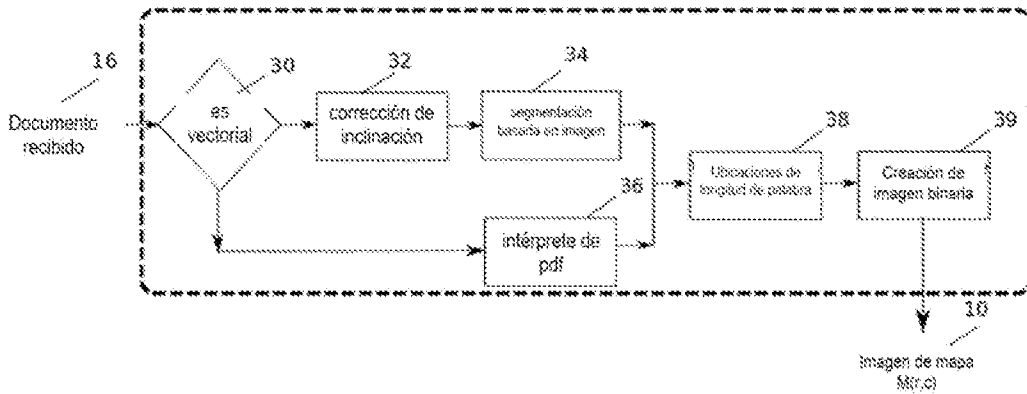


Fig. 4

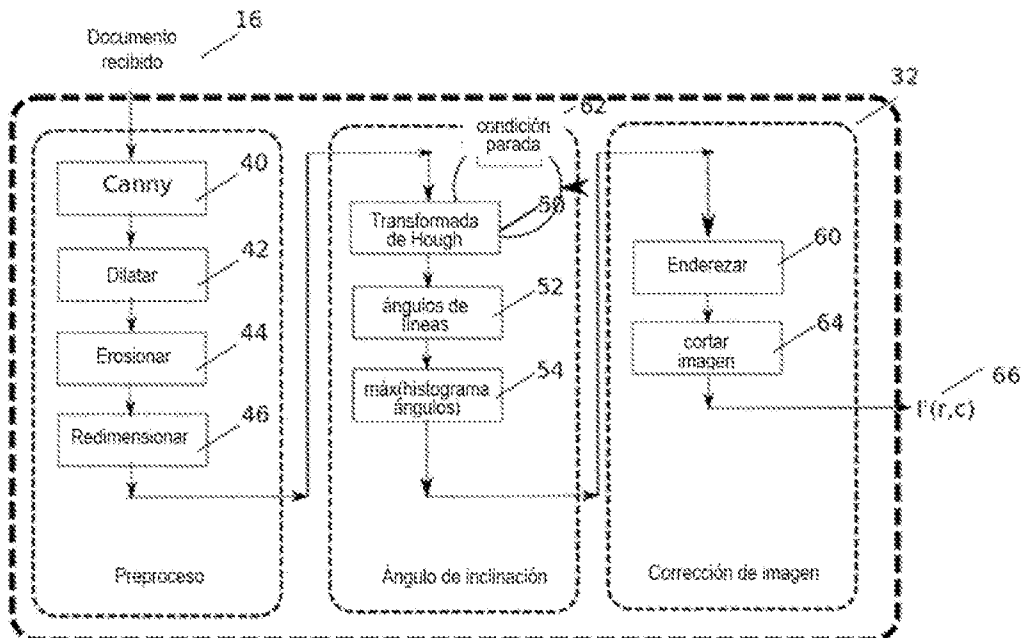


Fig. 5

Sección	A	La relación de consejo
Sección	B	Confidencialidad y privacidad
Sección	C	Responsabilidad profesional
Sección	D	Relaciones con otros profesionales
Sección	E	Evaluación, valoración e interpretación
Sección	F	Supervisión, entrenamiento y enseñanza
Sección	G	Investigación y publicación
Sección	H	Consejo a distancia, tecnología y medios sociales
Sección I:		Resolución de aspectos éticos
<hr/>		
Sección A:		La relación de consejo
Sección B:		Confidencialidad y privacidad
Sección C:		Responsabilidad profesional
Sección D:		Relaciones con otros profesionales
Sección E:		Evaluación, valoración e interpretación
Sección F:		Supervisión, entrenamiento y enseñanza
Sección G:		Investigación y publicación
Sección H:		Consejo a distancia, tecnología y medios sociales
Sección I:		Resolución de aspectos éticos

Fig. 7