

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 610 467**

51 Int. Cl.:

G06K 9/34 (2006.01)

G06K 9/20 (2006.01)

G06K 9/72 (2006.01)

G06K 9/46 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.05.2005 E 05104250 (5)**

97 Fecha y número de publicación de la concesión europea: **12.10.2016 EP 1598770**

54 Título: **Reconocimiento óptico de caracteres de baja resolución para documentos adquiridos con cámara**

30 Prioridad:

20.05.2004 US 850335

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

27.04.2017

73 Titular/es:

**MICROSOFT TECHNOLOGY LICENSING, LLC
(100.0%)
One Microsoft Way
Redmond, WA 98052, US**

72 Inventor/es:

**JACOBS, CHARLES E.;
RINKER, JAMES R.;
SIMARD, PATRICE Y. y
VIOLA, PAUL A.**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 610 467 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

Reconocimiento óptico de caracteres de baja resolución para documentos adquiridos con cámara

Esta invención se refiere al reconocimiento óptico de caracteres (OCR), y más específicamente, al software de OCR que mejora el procesamiento de documentos captados por dispositivos de baja resolución.

5 Los avances tecnológicos en la electrónica están reduciendo rápidamente el tamaño, el peso, el coste y el consumo de energía de las cámaras. Por lo tanto, los dispositivos informáticos móviles tales como los teléfonos móviles, PDA y ordenadores portátiles, pueden capturar información usando unas cámaras digitales pequeñas, económicas y de baja resolución que hoy en día están diseñadas como subcomponentes de los mismos. Tales cámaras baratas y versátiles permiten actualmente fotografiar fácilmente una amplia variedad de documentos sin usar dispositivos periféricos engorrosos como un escáner. Documentos que van desde libros y documentos legales a billetes y tarjetas de visita pueden fotografiarse ahora instantáneamente en teléfonos móviles, PDA y ordenadores portátiles.

10 Sin embargo, el reconocimiento óptico de caracteres (OCR) de tales documentos fotografiados presenta un desafío diferente. Las conversiones de papel a representaciones digitales y viceversa son sencillas cuando se utilizan ordenadores de escritorio y dispositivos de exploración periféricos adecuados. Por el contrario, la rápida evolución de las tecnologías adecuadas para soportar la informática móvil hace ahora más complicado captar las imágenes de tales documentos por tales dispositivos. Por ejemplo, es bastante difícil imprimir y explorar documentos cuando no están conectados a un dispositivo periférico adecuado. Las impresoras móviles son pesadas, gastan mucha energía y son caras. Los escáneres portátiles son igualmente difíciles de manejar. Además, estas cámaras de baja resolución, ya sean independientes o incrustadas en un dispositivo informático móvil, presentan un nuevo reto al OCR de tales documentos fotografiados. Los sistemas de caracteres impresos tradicionales proceden primero binarizando la imagen, segmentando el carácter y a continuación reconociendo los caracteres. Debido a que estos sistemas separan las etapas, se logran tasas de reconocimiento mucho más bajas que las deseadas, mitigando de este modo el incentivo para usar tales capacidades del dispositivo. El procedimiento de OCR es mucho más complicado cuando se usan estas cámaras de baja resolución porque es prácticamente imposible realizar una binarización de calidad o una segmentación de caracteres independiente del procedimiento de reconocimiento. La segmentación es donde el motor OCR organiza los píxeles de una imagen explorada píxelada en caracteres.

15 Cuando una tecnología móvil de este tipo puede proporcionar ahora la capacidad al OCR de capturar y procesar los datos del documento, se espera que las personas usen las cámaras para tomar fotografías de muchos tipos diferentes de documentos, que van desde facturas de restaurantes, artículos interesantes, informes, portadas de libros, tarjetas de visita, capturas de pantalla, diapositivas proyectadas en una pared, mapas, etc. El incentivo para capturar una amplia variedad de documentos es alto, ya que tales documentos digitales pueden almacenarse, indexarse, archivar, e incluso imprimirse de nuevo más tarde, una vez en forma electrónica. Además, para el trabajador de la información, esto presenta una gran cantidad de oportunidades.

20 Las cámaras baratas actuales (por ejemplo, de aproximadamente \$ 50) no proporcionan suficiente resolución para capturar una página completa de documento con una fuente de 11 puntos en una exposición y realizar la segmentación de caracteres independientemente de reconocer los caracteres. Las imágenes de una cámara de baja resolución son tan borrosas y de tan mala calidad que la binarización, la segmentación y el reconocimiento no pueden realizarse de manera independiente y aun así mantener altas tasas de reconocimiento. Por ejemplo, a baja resolución, la columna central de carácter "o" es a menudo un buen candidato para cortarse que otros muchos tales cortes entre letras. No puede esperarse un buen rendimiento sin abordar el problema de segmentación de manera efectiva. Las mejoras en el OCR de baja resolución también benefician a las cámaras de alta resolución, permitiendo de este modo a los usuarios tomar fotografías desde más lejos y fuera de foco.

Por lo tanto, lo que se necesita es una capacidad de OCR que pueda resolver los símbolos de baja resolución.

25 30 35 40 45 SIMARD PY Y COL.: "Best practices for convolutional neural networks applied to visual document analysis" ANÁLISIS Y RECONOCIMIENTO DE DOCUMENTOS, 2003. PROCEDIMIENTOS. SÉPTIMA CONFERENCIA INTERNACIONAL del 3 al 6 de agosto de 2003, PISCATAWAY, NJ, USA, IEEE, 3 de agosto de 2003 (2003-08-03), páginas 958-963, XP010656898 ISBN: 978-0-7695-1960-9 se refiere al uso de redes neuronales para la clasificación de entradas visuales generadas a partir de documentos.

50 BURGESS C J C Y COL.: "OFF LINE RECOGNITION OF HANDWRITTEN POSTAL WORDS USING NEURAL NETWORKS" REVISTA INTERNACIONAL DE RECONOCIMIENTO DE PATRONES E INTELIGENCIA ARTIFICIAL, PUBLICACIÓN CIENTÍFICA MUNDIAL, SINGAPUR, SI, vol. 7, n.º 4, 1 de agosto de 1993 (1993-08-01), páginas 689-704, XP000571180 ISSN: 0218-0014 desvela un procedimiento, la segmentación de trayecto más corto (SPS), que combina programación dinámica y un reconocedor de redes neuronales para segmentar y reconocer cadenas de caracteres.

55 BUNKE H Y COL.: HANDBOOK OF CHARACTER RECOGNITION AND DOCUMENT IMAGE ANALYSIS, 1 de enero de 1997 (1997-01-01), XP002410978 Singapur, PUBLICACIÓN MUNDIAL DE CIENCIAS ISBN: 981-02-2270-X desvela el reconocimiento de la información de elevación en los sistemas OCR. El procedimiento desvelado reconoce el valor de elevación reconociendo los números y los símbolos predefinidos por la correspondencia MAP.

Además, una combinación de la correspondencia MAP con la operación MAP se usa para reconocer los símbolos "l" y "-". Además, los candidatos de punto de referencia se distinguen de los candidatos de punto decimal contando el número de píxeles contenidos en las características puntuales.

5 BISHOP C M: "NEURAL NETWORKS FOR PATTERN RECOGNITION" REDES NEURALES PARA RECONOCER PATRONES, OXFORD: OXFORD UNIVERSITY PRESS, GB, 1 de enero de 1995 (1995-01-01), XP002387293 ISBN: 978-0-19-853864-6 desvela un procedimiento para adiestrar una red neuronal con ruido.

10 Yi Lu y M. Shridhar, Character segmentation in Handwritten Words - An Overview. Reconocimiento de patrones, 29 (1), pág. 77-96, 1996 (XP4000406) se refiere a la segmentación de caracteres en palabras manuscritas. Una etapa de segmentación inicial usa el procedimiento de componente conectado para calcular los segmentos de caracteres iniciales. Posteriormente, se detectan y se eliminan todas las marcas de signo de puntuación. El resultado del procedimiento de segmentación es una secuencia de primitivas segmentadas, cada una de las cuales contiene o un solo carácter o un carácter parcial. Por último, se usa un algoritmo de programación dinámica asistido por un clasificador de redes neuronales para hacer coincidir la secuencia de primitivas del procedimiento de segmentación con las palabras de un léxico.

15 Seni G. y Cohen E., External word segmentation of off-line handwritten text lines, 27 (1), p. 41-52, 1994, XP007947521 desvela técnicas para separar una línea de texto manuscrito sin restricciones en palabras. Cuando el estilo de escritura no está restringido, el reconocimiento de los componentes individuales puede no ser fiable por lo que deben agruparse entre sí en hipótesis de palabras, antes de que puedan usarse los algoritmos de reconocimiento. Este sistema usa algoritmos originales para determinar las distancias entre los componentes en una
20 línea de texto y para detectar el signo de puntuación. Para dividir las líneas de texto detectadas en palabras, se describen tres algoritmos separados, que son: un algoritmo para detectar las distancias entre los componentes adyacentes conectados, un algoritmo para detectar el signo de puntuación que detecta las marcas de signo de puntuación y un algoritmo que combina los dos primeros algoritmos para clasificar los huecos desde el más grande hasta el más pequeño, para determinar cuáles son los huecos entre caracteres y los huecos entre palabras.

25 Es el objeto de la presente invención proporcionar un procedimiento mejorado para el reconocimiento de caracteres ópticos de símbolos de baja resolución, así como un sistema correspondiente y un medio legible por ordenador.

Este objeto se resuelve con el objeto de las reivindicaciones independientes.

Las realizaciones preferidas están definidas por las reivindicaciones dependientes.

30 A continuación, se presenta un resumen simplificado de la invención con el fin de proporcionar una comprensión básica de algunos aspectos de la invención. Este resumen no es una visión general extensa de la invención. No se pretende identificar elementos clave/críticos de la invención o delinear el ámbito de la invención. Su único fin es presentar algunos conceptos de la invención de una forma simplificada como un preámbulo a la descripción más detallada que se presenta más adelante.

35 La presente invención desvelada y reivindicada en el presente documento, en un aspecto de la misma, comprende una arquitectura que resuelve estos problemas simultáneamente, combinando una tecnología de aprendizaje mecánica (por ejemplo, una red neuronal convolucional), una programación dinámica y un modelo de lenguaje que en combinación, proporcionan las etapas necesarias para lograr las capacidades de OCR de baja resolución para los símbolos y documentos adquiridos por una cámara. El uso de un enfoque de aprendizaje de máquina basado en la red neuronal convolucional y adiestrado en una gran cantidad de datos facilita la máxima robustez.

40 La presente invención está compuesta por varias partes que trabajan juntas para reconocer los símbolos en una página fotografiada. El reconocedor de caracteres basado en redes neuronales convolucionales se usa para predecir qué caracteres pueden estar presentes en una localización dada en la imagen de entrada. Un componente de análisis de diseño global realiza la tarea de encontrar las localizaciones de las líneas de texto en la página y devolver una lista de rectángulos delimitadores para estas líneas de texto. El analizador de diseño fragmenta las
45 líneas en palabras individuales, indicando los límites de cada palabra. Al tratar todos estos entre sí es el reconocedor de palabras el que encuentra la palabra más probable dentro de un rectángulo dado en la página. El reconocimiento de palabras es un problema de optimización resuelto usando la programación dinámica.

En otro aspecto de la presente invención, se proporciona un algoritmo que detecta las líneas de texto directamente desde la imagen de nivel de grises.

50 Para la realización de las conclusiones anteriores y relacionadas, se describen en el presente documento algunos aspectos ilustrativos de la invención en relación con la siguiente descripción y los dibujos adjuntos. Estos aspectos son indicativos, sin embargo, de solo algunas de las diversas maneras en que pueden emplearse los principios de la invención y se pretende que la presente invención incluya todos estos aspectos y sus equivalentes. Otras ventajas y características novedosas de la invención se pondrán de manifiesto a partir de la siguiente descripción detallada de
55 la invención cuando se considera junto con los dibujos.

Breve descripción de los dibujos

- 5 la figura 1 ilustra un sistema que es una infraestructura de optimización global para un OCR de baja resolución de documentos adquiridos por cámara de acuerdo con la presente invención.
- la figura 2 ilustra un diagrama de flujo de las operaciones principales del sistema OCR de baja resolución de la presente invención.
- la figura 3 ilustra una muestra de un documento sin procesar y el documento captado después de la captura usando una cámara.
- la figura 4 ilustra un texto de muestra y una palabra de un documento antes y después del preprocesamiento, de acuerdo con la presente invención.
- 10 la figura 5 ilustra una metodología para el adiestramiento de preprocesamiento de acuerdo con la presente invención.
- la figura 6 ilustra un diagrama de flujo de un procedimiento para generar datos etiquetados de acuerdo con la presente invención.
- la figura 7 ilustra un texto de muestra que se ha explorado y alineado con el mapa de bits de realidad sobre el terreno.
- 15 la figura 8 ilustra una metodología para el análisis de diseño de acuerdo con la presente invención.
- la figura 9 ilustra un diagrama de flujo de una metodología para detectar texto de acuerdo con la presente invención.
- la figura 10 ilustra un diagrama que representa un resultado de los filtros de detección de texto, de acuerdo con la presente invención.
- 20 la figura 11 ilustra una imagen de cámara original en comparación con una salida del detector de texto añadida a la imagen de cámara, como en una segunda representación, de acuerdo con la presente invención.
- la figura 12 ilustra un diagrama de flujo de un procedimiento para detectar huecos entre letras, de acuerdo con la presente invención.
- 25 la figura 13 ilustra un texto de muestra procesado para la extracción de líneas usando la detección de huecos, de acuerdo con la presente invención.
- la figura 14 ilustra una metodología para detectar los espacios entre palabras, de acuerdo con la presente invención.
- 30 la figura 15 ilustra un diagrama de flujo de una implementación para adiestrar una red neuronal convolucional de acuerdo con la presente invención.
- la figura 16 ilustra un diagrama de flujo de un procedimiento para reconocer caracteres usando un reconocedor de caracteres basado en una red neuronal convolucional de acuerdo con la presente invención.
- la figura 17 ilustra un diagrama de la red neural convolucional para un OCR de texto impreso de baja resolución de acuerdo con la presente invención.
- 35 la figura 18 ilustra una metodología de adiestramiento para una red neuronal convolucional de acuerdo con la presente invención.
- la figura 19 ilustra una metodología de programación dinámica para reconocer una palabra de acuerdo con la presente invención.
- 40 la figura 20 ilustra una segmentación de una palabra de muestra de acuerdo con la presente invención.
- la figura 21 ilustra una tabla de programación dinámica de acuerdo con la presente invención que se usa para contener soluciones parciales de un problema de programación dinámica cuando no se usa ningún modelo de lenguaje.
- la figura 22 ilustra una tabla de programación dinámica de muestra para una optimización del modelo de diccionario de acuerdo con la presente invención que se usa para determinar una asignación óptima de las observaciones para las letras de una palabra específica.
- 45 la figura 23 ilustra un diagrama de búsqueda de diccionario basándose en un árbol de prefijos de acuerdo con la presente invención.
- la figura 24 ilustra una metodología de detección de signo de puntuación de acuerdo con la presente invención
- 50 la figura 25 ilustra los resultados de una muestra captada de acuerdo con la arquitectura de la presente invención.
- la figura 26 ilustra una comparación de modelos de lenguaje a lo largo de una parte del texto captado de acuerdo con la arquitectura de la presente invención.
- la figura 27 ilustra un diagrama de bloques de un ordenador operativo para ejecutar la arquitectura desvelada.
- 55 la figura 28 ilustra un diagrama de bloques esquemático de un entorno de cálculo a modo de ejemplo de acuerdo con la presente invención.

Descripción detallada de la invención

La presente invención se describe a continuación con referencia a los dibujos, en los que se usan números de referencia similares para referirse a elementos similares en toda su extensión. En la siguiente descripción, con fines de explicación, se exponen numerosos detalles específicos con el fin de proporcionar una comprensión completa de la presente invención. Puede ser evidente, sin embargo, que la presente invención puede practicarse sin estos detalles específicos. En otros casos, se muestran estructuras y dispositivos bien conocidos en forma de diagramas de bloques para facilitar la descripción de la presente invención.

Tal como se usan en esta solicitud, se pretende que los términos “componente” y “sistema” se refieran a una entidad relacionada con la informática, o un hardware, una combinación de hardware y software, software o software en ejecución. Por ejemplo, un componente puede ser, pero no está limitado a ser, un procedimiento que se ejecuta en un procesador, un procesador, un objeto, un ejecutable, un subproceso, un programa y/o un ordenador. A modo de ilustración, tanto una aplicación que se ejecuta en un servidor como el servidor pueden ser un componente. Uno o más componentes pueden residir dentro de un procedimiento y/o de un subproceso, y un componente puede localizarse en un ordenador y/o distribuirse entre dos o más ordenadores.

Tal como se usa en la presente memoria, el término “inferir” o “inferencia” se refiere en general al procedimiento de razonamiento sobre o inferir estados del sistema, el entorno y/o el usuario a partir de un conjunto de observaciones capturadas a través de eventos y/o datos. La inferencia puede emplearse para identificar un contexto o acción específica, o puede generar por ejemplo, una distribución de probabilidad a lo largo de los estados. La inferencia puede ser probabilística, es decir, el cálculo de una distribución de probabilidad a lo largo de los estados de interés basándose en una consideración de datos y eventos. La inferencia también puede referirse a las técnicas empleadas para componer eventos de nivel superior a partir de un conjunto de eventos y/o datos. Esta inferencia da lugar a la construcción de nuevos eventos o acciones a partir de un conjunto de eventos observados y/o datos de eventos almacenados, si los eventos están correlacionados o no en una proximidad temporal cercana y si los eventos y datos provienen de una o varias fuentes de eventos y datos.

OCR DE DOCUMENTOS ADQUIRIDOS CON CAMERA

La presente invención facilita la captura y el reconocimiento óptico preciso de caracteres (OCR) de símbolos y texto cuando se usan símbolos y/o imágenes de documento de baja resolución. Tal capacidad también encuentra aplicación en escenarios que implican símbolos, texto y/o documentos degradados, cuyo medio se ha dañado o contaminado de alguna manera y los símbolos asociados son más difíciles de reconocer.

Haciendo referencia ahora a la figura 1, se ilustra un sistema 100 que es una infraestructura de optimización global para un OCR de baja resolución de los documentos adquiridos con cámara de acuerdo con la presente invención. El sistema 100 resuelve el problema de la técnica anterior al combinar simultáneamente al menos una red neuronal convolucional, un componente de programación dinámica y un modelo de lenguaje. Por lo tanto, para un documento 108 que se capturado y se procesa por el sistema 100, la tasa de reconocimiento es muy alta, dando como resultado una representación 110 digital de salida adecuada para el usuario.

La red neuronal convolucional puede calcular eficientemente una función de clasificación en múltiples posiciones reutilizando los resultados parciales de una posición en el cálculo de la posición adyacente. Esto se facilita además, ya que una red neural puede procesar una entrada de nivel de grises. Un sistema robusto se proporciona usando un enfoque de aprendizaje de máquina que se adiestra con una cantidad grande de datos. La red neuronal convolucional se emplea porque calcula eficientemente una función de clasificación en múltiples posiciones, se escala bien con grandes cantidades de datos de adiestramiento, emite probabilidades que hacen que sea fácil de combinar con un algoritmo de modelo de Markov oculto (HMM), toma la entrada de nivel de grises, que hacen innecesaria la binarización. Recopilando los datos en diferentes condiciones de iluminación, fuentes, tamaños, cámaras, ángulos, enfoques, etc., el algoritmo de aprendizaje de máquina realiza la tarea de construir un clasificador invariante a estos factores.

El componente de programación dinámica se utiliza para resolver la optimización del reconocimiento de palabras. Un reconocedor de palabras usa el componente de programación dinámica para determinar qué palabra se localiza en un rectángulo delimitador de palabra dado devuelta por el análisis de disposición. La programación dinámica encuentra una solución óptima para un problema construyendo una serie de soluciones óptimas para los subproblemas del problema original. Esto facilita la reutilización de gran parte del cálculo para encontrar los subproblemas óptimos cuando se determinar la solución óptima global.

El componente de modelo de lenguaje facilita el reconocimiento de caracteres y palabras.

Haciendo referencia ahora a la figura 2, se ilustra un diagrama de flujo de las operaciones principales del sistema OCR de baja resolución de la presente invención. En primer lugar, se realiza una función de preprocesamiento para capturar una imagen de documento en preparación para su procesamiento adicional. Una función de análisis de diseño global realiza la tarea de encontrar las localizaciones de las líneas de texto en el documento capturado y de devolver una lista de rectángulos delimitadores para estas líneas de texto. A continuación, el analizador de diseño fragmenta las líneas en palabras individuales, indicando los límites de cada palabra. Se usa un reconocedor de caracteres para predecir qué carácter es probable que esté presente en una localización dada en la imagen de entrada. En una implementación, un reconocedor de caracteres basado en redes neuronales convolucionales se usa para facilitar el reconocimiento de caracteres y otras funciones. Tratando todos estos entre sí está un reconocedor de palabras que encuentra la palabra más probable en el interior de un rectángulo dado en el documento. El reconocimiento de palabras es un problema de optimización que se resuelve usando la programación dinámica.

Haciendo referencia ahora a la figura 3, se ilustra una muestra de un documento 300 sin procesar y el documento 302 captado después de la captura usando una cámara. El documento 302 captado exhibe un color de fondo más oscuro y un texto borroso en respuesta a las capacidades de baja resolución de la cámara.

PREPROCESAMIENTO

5 Haciendo referencia ahora a la figura 4, se ilustra un texto de muestra y una palabra de un documento antes y después del preprocesamiento, de acuerdo con la presente invención. Un documento explorado puede presentar una serie de problemas que normalmente se encontrarían cuando un usuario captura el documento usando una cámara (por ejemplo, una cámara web). Por ejemplo, si la cámara estaba demasiado cerca del documento, tenía una alineación errónea del eje focal y rotacional en relación con la superficie del documento y una iluminación incorrecta o irregular. La alineación rotacional hace que las líneas de texto no estén perfectamente horizontales en la imagen. La deformación de la perspectiva introduce una distorsión debido a la perspectiva cuando la cámara no se enfrenta al documento directamente. Además, puede producirse la distorsión de barril debido a la relativamente pobre calidad de la óptica. Algunos de los problemas de iluminación se alivian preprocesando las imágenes con un algoritmo adaptativo de balance de blancos que persigue maximizar el contraste localmente sobre unas regiones o unos bloques de la imagen. Este procedimiento puede ejecutarse cada vez que se presenta una imagen de documento al sistema para su procesamiento de OCR.

ADIESTRAMIENTO

20 Haciendo referencia ahora a la figura 5, se ilustra una metodología para el adiestramiento de preprocesamiento de acuerdo con la presente invención. Aunque, con fines de simplificación de la explicación, la una o más metodologías mostradas en el presente documento, por ejemplo, en forma de un diagrama de flujo, se muestran y describen como una serie de acciones, debería entenderse y apreciarse que la presente invención no está limitada por el orden de las acciones, ya que algunas acciones pueden, de acuerdo con la presente invención, producirse en un orden diferente y/o concurrentemente con otras acciones de las mostradas y descritas en la presente memoria. Por ejemplo, los expertos en la materia entenderán y apreciarán que una metodología podría representarse alternativamente como una serie de estados o eventos interrelacionados, tal como en un diagrama de estado. Además, no todas las acciones ilustradas pueden ser necesarias para implementar una metodología de acuerdo con la presente invención.

30 El siguiente preprocesamiento es diferente al de la figura 4, en que se realiza una vez para inicializar el sistema basado en redes neuronales. En 500, se genera un dato etiquetado. En 502, se adiestra el reconocedor de caracteres de redes neuronales. A continuación, el procedimiento llega a un bloque de parada. Para adiestrar al reconocedor de caracteres basado en redes neuronales, se requiere una gran recopilación de datos de adiestramiento etiquetados. Una recopilación de este tipo podría construirse teniendo un humano que anote las páginas exploradas, e identificando la localización y el código ASCII para cada carácter en la página. Sin embargo, esto puede ser un procedimiento que consuma mucho tiempo y propenso a errores. Por lo tanto, se proporciona un mecanismo semiautomático para generar los datos etiquetados.

40 Haciendo referencia ahora a la figura 6, se ilustra un diagrama de flujo de un preprocedimiento para generar datos etiquetados de acuerdo con la presente invención. El sistema de adquisición de datos de la presente invención consiste en un controlador de impresora especial que registra el cuadro delimitador y el código de carácter para cada uno de los caracteres de una página y una herramienta para ayudar a alinear estos cuadros delimitados con una imagen explorada. En 600, se proporciona el controlador de impresora. En 602, se procesa el documento con el controlador de impresora. En 604, se generan los datos de verdad del terreno (GTD) creando un mapa de bits de alta resolución del documento y registrando la información geométrica de cada carácter en el documento. El documento y el mapa de bits de alta resolución son la "verdad del terreno", donde cada carácter está correctamente etiquetado y su posición conocida. En 606, se imprime y se explora el mapa de bits de alta resolución con el sistema de cámara. En 608, se alinea la imagen explorada con el mapa de bits de verdad de tierra para obtener la posición de cada uno de los caracteres de la página explorada. Con el fin de hacer esta etapa de rectificación, en 610, se identifican las localizaciones correspondientes entre la imagen explorada y el mapa de bits de verdad de tierra para cada uno de los cuatro puntos del documento. En 612, se calcula la transformación de la mejor perspectiva de ajuste para mapear la imagen explorada en la imagen reproducida. En 614, se deforma la imagen explorada, si es necesario, para alinearse con la imagen reproducida. A continuación, el procedimiento llega a un bloque de parada.

55 Desafortunadamente, al usar una transformación de perspectiva para rectificar las imágenes exploradas no siempre es suficiente para alinear las cajas delimitadoras con los caracteres explorados. Para mejorar las coincidencias, cada carácter se traduce por hasta cinco píxeles en cada dirección, y la mejor coincidencia se registra usando una medida de distancia por mínimos cuadrados. A continuación, las correspondientes cajas delimitadoras de desplazamiento se usan como la localización en la imagen deformada y explorada para cada ejemplo de adiestramiento.

Haciendo referencia ahora a la figura 7, se ilustra una muestra de texto que se ha explorado y alineado con el mapa de bits de verdad de tierra. Lo mostrado, es el resultado de usar la transformación de perspectiva (descrita en la figura 6) para alinear las cajas delimitadoras de caracteres con la imagen explorada. El texto 700 explorado está

alineado con los datos 702 de verdad de tierra.

ANÁLISIS DE DISEÑO

Haciendo referencia ahora a la figura 8, se ilustra una metodología para el análisis de diseño de acuerdo con la presente invención. En 800, se recibe un documento para captarse por una cámara. En 802, el análisis de diseño detecta el diseño del texto en la página encontrando las palabras, líneas y columnas. Usando la información de línea, los caracteres a lo largo de la línea pueden segmentarse, escalarse y presentarse a un algoritmo de clasificación de caracteres, tal como se indica en 804. A continuación, el procedimiento llega a un bloque de parada. En una imagen explorada, este tipo de análisis de diseño puede continuar usando unas operaciones de procesamiento de imágenes sencillas pero confiables tal como el análisis de componentes conectados. Estos componentes conectados se usan a continuación como las entradas básicas para el procedimiento de diseño.

Como se ha indicado anteriormente, las imágenes de documento capturadas desde una cámara en un entorno de oficina son muy diferentes de los documentos captados en un dispositivo de cálculo periférico tal como un escáner. La iluminación usada en las fotografías no está controlada, dando lugar a grandes variaciones de contraste y brillo. Estas variaciones presentan un reto para los algoritmos de detección de texto simple. El análisis de componentes conectados requiere una imagen binaria (es decir, blanco y negro) como entrada. Por lo tanto, puede ser difícil construir un algoritmo de umbral adaptable que separe de manera fiable el texto del papel de fondo. Las variaciones tanto en el contraste como en la intensidad pueden hacer que parte del texto desaparezca en el fondo mientras que el otro texto se fusiona en bloques muy grandes de caracteres.

Haciendo referencia ahora a la figura 9, se ilustra un diagrama de flujo de una metodología para detectar texto de acuerdo con la presente invención. En una implementación, la detección de texto procede de la siguiente manera. En 900, se aplican dos filtros lineales a cada localización en la imagen de entrada. En 902, se asigna un umbral a cada filtro. Las localizaciones cuyo valor está por encima de este umbral se denominan "texto", las que están por debajo son "no texto". En 904, las localizaciones que se determinan como texto por ambos filtros se usan y se retienen como posibles regiones de texto. A continuación, el procedimiento llega a un bloque de parada.

En una implementación, los umbrales se determinan experimentalmente (100 y 6000, para una altura de línea de siete píxeles y un hueco de carácter de un píxel). En otra implementación, pueden incorporarse características adicionales y los umbrales pueden calcularse automáticamente usando un "boosting". Un Boosting es un procedimiento general para mejorar el rendimiento de cualquier algoritmo de aprendizaje que genera consistentemente clasificadores que necesitan realizarse solo un poco mejor que la suposición aleatoria.

Haciendo referencia ahora a la figura 10, se ilustra un diagrama 1000 que representa un resultado de los filtros de detección de texto, de acuerdo con la presente invención. El primer filtro opera en el texto "Now is" para encontrar los espacios entre los caracteres. En el presente documento, se detecta un hueco entre las letras "o" y "w". El segundo filtro opera en el mismo texto para encontrar las líneas asociadas. Por lo tanto, el segundo filtro detecta los espacios por encima y por debajo de la palabra "Now".

Haciendo referencia ahora a la figura 11, se ilustra una imagen 1100 de cámara original en comparación con una salida del detector de texto añadida a la imagen de cámara, como en una segunda reproducción 1102, de acuerdo con la presente invención. Las zonas blancas de la segunda representación indican el texto detectado.

Haciendo referencia ahora a la figura 12, se ilustra un diagrama de flujo de un procedimiento para detectar huecos entre las letras, de acuerdo con la presente invención. El resultado es que ambos filtros detectan aproximadamente huecos entre los caracteres, como se indica en 1200. En 1202, se calcula un análisis de componentes conectados (CC) en los "píxeles de hueco". En 1204, el análisis de diseño se realiza en los "CC de hueco" resultantes. Los CC de hueco se agrupan a continuación probando sus relaciones geométricas relativas, como se indica en 1206. Por ejemplo, si los CC de hueco son relativamente del mismo tamaño y están horizontalmente cercanos y alineados, se detecta una línea. Las pruebas usan estadísticas robustas tales como las alturas y anchuras medias de los CC en la línea actual, y el algoritmo realiza múltiples pasadas. En 1208, los umbrales se determinan a lo largo de un conjunto cada vez más grande de CC y sobre muchos documentos. A continuación, el procedimiento llega a un bloque de parada.

A medida que las estadísticas se recopilan en un conjunto cada vez más grande de CC, las pruebas de agrupación se hacen de manera más agresiva. Los umbrales se determinan experimentalmente a lo largo de una recopilación de docenas de documentos. Dados los CC de hueco, la etapa de detección de línea es invariante para la escala y el alfabeto porque solo se comparan las posiciones relativas de las cajas delimitadoras de los CC de hueco.

Haciendo referencia ahora a la figura 13, se ilustra el texto 1300 de muestra procesado para la extracción de líneas usando la detección de huecos, de acuerdo con la presente invención. En esencia, el sistema circunscribe una línea detectada de texto con un rectángulo delimitado. En este documento de dos columnas, la detección de líneas da lugar a que cada línea detectada de cada columna esté circunscrita al rectángulo delimitado. Si se tratara de un documento de columna única, cada línea de texto que se extiende a través de la página estaría circunscrita al rectángulo delimitado, lo que indicaría que se ha detectado la línea.

Haciendo referencia ahora a la figura 14, se ilustra una metodología para detectar los espacios entre palabras, de acuerdo con la presente invención. En 1400, se detectan las líneas de texto. Después de que se han detectado las líneas de texto, estas líneas se fragmentan adicionalmente en las regiones de la página que contienen cada una de las palabras del documento, como se indica en 1402. Se identifican las localizaciones de los espacios entre palabras en cada línea y la línea se fragmenta usando estos espacios, como se indica en 1404. Para encontrar los espacios, se detectan las columnas verticales de espacios en blanco que cortan a través de la caja delimitadora de una línea. En 1406, una columna se clasifica como un espacio en blanco si todos los píxeles de esa columna son más brillantes que algún umbral. El umbral se determina explorando a través de un segmento horizontal a través del centro de la línea actual, encontrando los valores de brillo mínimo y máximo de los píxeles a lo largo de esta línea de exploración, y calculando un umbral de la misma. En una implementación, un píxel cuyo brillo es mayor que $(\text{Brillomin} + 3 * \text{brillomáx})/4$, se clasifica como un píxel de espacio en blanco. En 1408, el buscador de espacios, a continuación explora a través de la línea de texto que detecta columnas de píxeles de espacio en blanco, y usa estas localizaciones espaciales para dividir la línea en palabras. En 1410, las palabras se envían a continuación al reconocedor de caracteres. A continuación, el procedimiento llega a un bloque de parada.

15 RECONOCEDOR DE CARACTERES

Haciendo referencia ahora a la figura 15, se ilustra un diagrama de flujo de un procedimiento para adiestrar a un reconocedor de caracteres basado en una red neural convolucional de acuerdo con la presente invención. En el núcleo del sistema de OCR de baja resolución está el reconocedor de caracteres basado en una red neural convolucional. En 1500, se proporciona una primera capa convolucional para extraer las características simples a una resolución más alta. En 1502, se proporciona una segunda capa convolucional para convertir las características simples en características complejas a una resolución más gruesa. En 1504, se proporcionan dos capas totalmente conectadas adicionales para formar un clasificador de fines múltiples para el aprendizaje. En 1506, las cuatro capas se adiestran simultáneamente. En 1508, la red se adiestra para minimizar la entropía cruzada. A continuación, el procedimiento llega a un bloque de parada.

Haciendo referencia ahora a la figura 16, se ilustra un diagrama de la red 1600 neuronal convolucional para OCR de textos impresos de baja resolución de acuerdo con la presente invención. Una estrategia general de una red convolucional es extraer características simples a una resolución mayor, y a continuación convertir las características simples en características más complejas en una resolución más gruesa. Una forma de generar una resolución más gruesa es sub-muestrear una capa en un factor de dos. La red 1600 incluye una imagen 1602 de entrada que tiene un campo receptivo de píxeles de 29x29 para los datos de recepción. Una primera capa 1604 extrae normalmente unas características muy ordinarias, tales como unas derivadas de X e Y, un paso bajo, y una derivada de X-Y. La primera capa 1604 explora la imagen 1602 de entrada usando cinco neuronas (o mapas) de detector de características de píxel de 13x13. Debido a las características que se aprenden en los datos, es difícil predecir lo que realmente rendirán hasta que se logre el adiestramiento. Una segunda capa 1606 convolucional extrae unas características mucho más complejas a una resolución ordinaria. La segunda capa 1606 explora la salida de la primera capa 1604 usando cincuenta neuronas de detector de características de píxel de 5x5. El número de características no es crítico, siempre que haya suficientes características para llevar la información a las capas (1608 y 1610) de clasificación. En este punto, las características son difíciles de interpretar, debido a que se distribuyen en tantas unidades. Puede hacerse la hipótesis de que las características son bucles, intersecciones, curvaturas, etc. Las dos primeras capas (1604 y 1606) pueden verse como un extractor de características adiestrable. Las dos últimas capas (1608 y 1610) están totalmente conectadas, y pueden verse como formando un clasificador de usos múltiples, ya que una red neuronal totalmente conectada de dos capas puede aprender cualquier función. Las cuatro capas (1604, 1606, 1608, y 1610) se adiestran de manera simultánea. A continuación, la red 1600 se adiestra para minimizar la entropía cruzada.

Las experiencias anteriores con esta arquitectura convolucional indican que en una implementación, una selección de cinco características para la primera capa 1604 convolucional y cincuenta características para la segunda capa 1606 convolucional son adecuadas para un amplio intervalo de reconocimiento de caracteres basado en imágenes, incluyendo el OCR impreso de baja resolución de la presente invención. En esta implementación, se usaron setenta y seis unidades de salida y setecientos sesenta unidades ocultas para las capas (1608 y 1610) totalmente conectadas. El número de unidades ocultas es variable, y es controlando este número por lo que se controla la capacidad y la generalización del clasificador general. Sin embargo, debería apreciarse por un experto en la materia que pueden emplearse otros valores para una optimización adicional.

Haciendo referencia ahora a la figura 17, se ilustra un diagrama de flujo de un procedimiento para reconocer caracteres usando un reconocedor de caracteres basado en una red neural convolucional de acuerdo con la presente invención. En 1700, se proporciona una red neuronal adiestrada. En 1702, se introduce un mapa de bits de la imagen de carácter. En este caso, el mapa de bits es el campo de píxel receptivo de 29x29. En 1704, la red emite una lista de probabilidades que indica qué caracteres son susceptibles de representarse por la imagen. A continuación, el procedimiento llega a un bloque de parada.

Haciendo referencia ahora a la figura 18, se ilustra una metodología de adiestramiento de la red neuronal convolucional de acuerdo con la presente invención. En 1800, el sistema se adiestra tomando caracteres de ejemplo de una base de datos de imágenes de adiestramiento. A continuación, en 1802, las imágenes de adiestramiento se

hacen fluctuar aleatoriamente dentro de la ventana de entrada. En 1804, el brillo y el contraste de las imágenes de adiestramiento se alteran de manera aleatoria. En 1806, para cada tiempo (o exposición para completar el conjunto de patrones) de adiestramiento, se presenta aleatoriamente un ejemplo ajustado de fluctuación y de contraste de cada carácter a la red. A continuación, el procedimiento llega a un bloque de parada. En una implementación, la base de datos de adiestramiento contiene 57555 caracteres tomados de quince páginas del texto del documento.

RECONOCEDOR DE PALABRAS

Haciendo referencia ahora a la figura 19, se ilustra una metodología de programación dinámica para reconocer una palabra, de acuerdo con la presente invención. El reconocedor de caracteres proporciona, en cada localización, la probabilidad de que cada carácter esté presente y centrado en la ventana de entrada de la red neuronal. El analizador de diseño identifica los rectángulos delimitadores de cada palabra en la página. Desgraciadamente, las localizaciones de las letras individuales de la página son desconocidas, por lo tanto, no es posible el centrado del reconocedor de caracteres sobre cada carácter para leer su mejor suposición. De hecho, el reconocedor de caracteres estará feliz de informar de qué 'letra' es más probable para cualquier mapa de bits dado, si es o no es en realidad una letra.

En 1900, se reciben una serie de observaciones del reconocedor de caracteres en la entrada del reconocedor de palabras. Las observaciones se realizan mediante la exploración a través de la palabra e invocando al reconocedor de caracteres en cada posible localización de carácter válido. En 1902, se definen las localizaciones de caracteres válidos. Para definir las localizaciones de caracteres válidos, el sistema fragmenta la palabra en trozos pequeños, o huecos, que se establecen para ser el tamaño mínimo posible permitido para un carácter, como se indica en 1904. En una implementación, los huecos son dos píxeles de ancho. Ya que la palabra se divide en huecos que representan el tamaño mínimo posible para un carácter, puede necesitarse que un número de huecos (por ejemplo, uniendo hasta cuatro huecos) se concatenen entre sí para encontrar una secuencia combinada de los huecos que representen una letra, como se indica en 1906. En 1908, se rellena una tabla con los resultados del reconocedor de caracteres para todas las secuencias válidas de huecos. En 1910, el reconocedor de palabras a continuación, intenta reconciliar esta secuencia de salidas del reconocedor con una palabra específica. A continuación, el procedimiento llega a un bloque de parada.

El reconocedor de palabras usa un algoritmo de programación dinámica para determinar qué palabra está localizada en un rectángulo que delimita una palabra dada devuelta por la etapa de análisis de diseño. La programación dinámica encuentra una solución óptima para un problema construyendo una serie de soluciones óptimas para los subproblemas del problema original. Esto permite la reutilización de la mayor parte del cálculo para encontrar subproblemas óptimos para determinar la solución óptima global.

Haciendo referencia ahora a la figura 20, se ilustra una segmentación 2000 de palabra de muestra de acuerdo con la presente invención. Las pequeñas marcas de control 2002 indican el tamaño del hueco para segmentar una palabra. Las marcas 2004 de control más grandes indican los huecos usados realmente para cada una de las letras de las palabras "private" y "certainly".

MODELOS DE LENGUAJE

En una implementación, el reconocedor de palabras no tiene un modelo de lenguaje específico construido en (lenguaje neutro), sino que simplemente intenta producir la interpretación más probable de una secuencia de observaciones del reconocedor de caracteres. Este reconocedor usa un simple algoritmo de programación dinámica unidimensional donde la función objetivo a maximizarse es simplemente la suma de las puntuaciones de caracteres. La puntuación de caracteres es la probabilidad de que el carácter más probable (como el devuelto por el reconocedor de caracteres), multiplicado por un factor de escalado que depende de cuánto difiere la anchura media del carácter más probable de la anchura de la parte de la imagen que se está considerado para este carácter. Cada celda de la tabla de programación dinámica contiene la letra de ese punto final de la solución, un enlace de vuelta a la letra anterior y la puntuación acumulativa para la solución hasta el momento. La cadena (invertida) más probable se recupera remontando a través de la tabla desde la última entrada.

Haciendo referencia ahora a la figura 21, se ilustra una tabla de programación dinámica de acuerdo con la presente invención que se usa para contener soluciones parciales de un problema de programación dinámica cuando no se usa modelo de lenguaje. Por lo general, la programación dinámica procede rellenando una tabla, donde cada celda de la tabla representa un subproblema del problema original. Por ejemplo, considérese una tabla 2100 (unidimensional) usada para contener soluciones parciales de un problema de programación dinámica. La tabla 2100 contiene una celda 2102 para cada hueco en la secuencia de entrada. Cada una de estas celdas 2102 representa un punto final, que es la solución óptima para la parte de la palabra que termina en el hueco que corresponde a la celda. Cada solución se expresa en términos de la solución óptima para un subproblema más pequeño (como se indica por la flecha asociada). En este ejemplo, la palabra se reconoce como "FOOD". Cada celda 2102 contiene una medida del estado de ese subproblema específico y un puntero de vuelta a una celda anterior de la tabla 2100 que representa la solución óptima para la parte de la palabra anterior a esta letra. Cuando la tabla 2100 se ha llenado completamente, la última celda 2104 representa la palabra óptima para toda la secuencia de observaciones.

Haciendo referencia ahora a la figura 22, se ilustra una tabla 2200 de programación dinámica de muestra para una optimización del modelo de diccionario de acuerdo con la presente invención que se usa para determinar una asignación óptima de observaciones 2202 para las letras de una palabra específica. En una implementación, el reconocedor de palabras intenta encontrar qué palabra de un diccionario de palabras es la coincidencia más probable para una imagen de entrada dada. Si no hay una palabra lo suficientemente plausible en el léxico para describir la imagen, el sistema recurre a un procedimiento anterior.

En una primera implementación, se emplea una versión del reconocedor basado en el diccionario que simplemente explora linealmente a través de todo el léxico, evaluando la probabilidad para cada palabra, y emitiendo la palabra más probable. Una segunda implementación describe una organización alternativa que permite el intercalado de la optimización de programación dinámica con el recorrido del diccionario para calcular más rápidamente la palabra más probable.

En este problema, cada columna de la tabla 2200 de programación dinámica representa los subproblemas que terminan en una posición específica (hueco) en la secuencia de entrada, y cada fila de la tabla representa una letra de la palabra en cuestión. Una entrada en la columna *c* y la fila *r* de la tabla 2200 representa una palabra parcialmente reconocida que termina en el hueco *c*, y que termina con la letra representada por fila *r*. En esta localización de la tabla se almacena un puntero de vuelta a la letra y al hueco anterior donde termina la letra, así como, a una puntuación acumulada.

Un procedimiento de puntuación similar se usa para la segunda aplicación del reconocedor de palabras, la probabilidad de que la observación coincida con la letra implícita de la celda actual, los tiempos de un factor de escalado que dependen de la anchura de hueco y la anchura media del carácter. Una vez más, la puntuación acumulada es la puntuación de la celda actual más la puntuación acumulada para la celda que representa la solución parcial anterior. Una vez que se ha completado el llenado de la tabla, la puntuación óptima para la palabra se almacena en la celda final (superior derecha). Esta puntuación se normaliza a continuación, dividiendo por el número de letras de la palabra. Sin esta normalización, las palabras largas con letras con relativamente mal puntuación pueden acumular puntuaciones altas y vencer a palabras más cortas que tienen muy buenas puntuaciones de letra. Por lo tanto, se desea maximizar la puntuación para cada letra.

Ya que muchas palabras en el diccionario comparten prefijos con otras palabras, hay una duplicación del trabajo calculando esta información compartida para cada palabra. Por ejemplo, considérese una tabla de programación dinámica empleada para encontrar la puntuación de la palabra "FOOL". Esta tiene las mismas primeras tres filas en el ejemplo anterior "FOOD". Por lo tanto, se desea una manera de compartir estas filas idénticas al calcular las puntuaciones de las palabras con prefijos comunes.

Cabe señalar que algunas celdas no necesitan llenarse, ya que es imposible para cualquier carácter que no sea el primero en ocupar el primer hueco, y cada letra solo puede ser responsable de hasta cuatro huecos. La ruta óptima está representada por las flechas que se mueven a lo largo de las celdas con los más altos valores observados.

Haciendo referencia ahora a la figura 23, se ilustra un diagrama 2200 de diccionario de búsqueda basado en un árbol de prefijos de acuerdo con la presente invención. Un árbol de prefijos (derivado de recuperación) es una estructura de datos o estructura de árbol de posición *k*-aria para almacenar cadenas en las que hay un nodo para cada prefijo común. Las cadenas se almacenan en los nodos hoja adicionales. Los árboles de prefijos se usan para implementar un tipo de datos de abstracción (ADT) de diccionario, donde pueden realizarse operaciones básicas como anular, buscar, insertar y suprimir. Pueden usarse para codificar y comprimir y, buscar expresiones regulares y hacer coincidir cadenas aproximadas.

Una solución es crear y evaluar una fila de la tabla 2200 (de la figura 22) a la vez que se recorre el diccionario. Para recorrer el diccionario en un orden que maximiza la cantidad de cálculos reutilizados, el diccionario está dispuesto en la estructura 2300 de árbol de prefijos. Cualquier nodo 2302 en el árbol 2300 de prefijos representa o una palabra parcial o una palabra completa (o, ambos - "fool" es una palabra y también un prefijo de "foolish"). A medida que se recorre el árbol de prefijos de diccionario, se construye la tabla de programación dinámica. Durante la visita a un nodo, se crea una nueva "fila" 2304 para esta tabla 2300 virtual que corresponde a la letra representada por ese nodo en el árbol de prefijos, y la fila se llena con las puntuaciones.

Por lo tanto, el árbol 2300 de prefijos resultante tiene un nodo "F" de nivel superior al que está asociada la fila "F" de los valores de la tabla 2200 de diccionario de la figura 22. Recorriendo hacia abajo, un segundo nodo (un primer nodo "O") tiene una primera fila "O" asociada de la tabla 2200. Continuando, un tercer nodo (un segundo nodo "O") tiene la fila correspondiente asociada con el mismo de la tabla 2200. Este tercer nodo tiene una primera rama que corresponde a un nodo "D" de la palabra "FOOD" y una segunda rama que corresponde a un nodo "L" de la palabra "FOOL". El nodo "D" se asocia con la línea "D" de la tabla 2200. Se realiza una valoración de fila similar, pero más baja y se asocia con el nodo "L".

Cabe señalar que el único contexto necesario para esta operación es la fila anterior, que se pasa como parámetro a la rutina de recorrido de árbol de prefijos recursiva. Si el nodo en cuestión representa una palabra completa, puede observarse la última entrada de la fila para encontrar la suma de las puntuaciones de las letras de esa palabra. Una

vez más, la suma de las puntuaciones de letra se divide por la longitud de la palabra para obtener la puntuación final de la palabra. Cuando termina el recorrido del árbol de prefijos, se devuelve la palabra de mayor puntuación encontrada.

5 Hay algunas optimizaciones heurísticas que aceleran inmensamente el cálculo. En primer lugar, solo visitar las palabras que comienzan con las letras que son propensas a ser la letra inicial de la palabra. Esta optimización proporciona un aumento de varias veces en la velocidad, en especial para las palabras que comienzan con las letras poco comunes. Otro aumento de velocidad proviene del recorte de la búsqueda de manera que no se siguen los enlaces en el árbol de prefijos que son poco probables que resulten en una palabra de alta puntuación. Si la puntuación media de las letras del prefijo de palabra en un nodo dado (el valor en la entrada final de la columna, dividido por el número de letras) es peor que un umbral, se supone que no importa cuán buena sea la puntuación de palabra de las letras restantes, ya que nunca será lo suficientemente buena para hacer que el promedio supere la mejor palabra vista hasta ahora. Esta segunda optimización proporciona un aumento significativo en la velocidad sin comprometer notablemente los resultados.

15 El signo de puntuación plantea un problema al reconocedor de palabras basado en el diccionario, en que las palabras con un signo de puntuación de arrastre pueden no estar presentes en el léxico. Haciendo referencia ahora a la figura 24, se ilustra una metodología de detección de signo de puntuación de acuerdo con la presente invención. Antes de que una secuencia de localizaciones de reconocimiento se pase al reconocedor de palabras, el sistema determina la probabilidad de que el último carácter de la palabra sea un signo de puntuación, como se indica en 2400. Si es así, se usa el reconocedor de caracteres para identificar el signo de puntuación específico. Con el fin de determinar si el carácter final es un signo de puntuación, el sistema busca el carácter más probable para cada posible posición de carácter final, como se indica en 2402. En 2404, se genera una puntuación para cada carácter más probable. En 2406, cada puntuación se compara con un umbral predeterminado. En 2408, el sistema determina si la puntuación ha superado el umbral. Si la puntuación más alta entre estas salidas del reconocedor de caracteres está por encima de un umbral, el flujo es hacia 2410 para considerar también si el carácter es más probable un signo de puntuación. Si se dan ambas situaciones, el flujo es hacia 2412 donde se supone que es una palabra puntuada. El resto de la palabra se envía a continuación al reconocedor basado en el diccionario, como se indica en 2414. En 2416, el signo de puntuación se añade a continuación al resultado del reconocedor de palabras. A continuación, el procedimiento llega a un bloque de parada. Si la puntuación no supera el umbral, o si lo hace, pero no se considera un signo de puntuación, el flujo es de 2408 y 2410 hacia el bloque de parada.

30 Haciendo referencia ahora a la figura 25, se ilustran los resultados 2500 de una muestra 2502 captada de acuerdo con la arquitectura de la presente invención. El procesamiento se ve facilitado por una interfaz de usuario que produce una ventana 2506 para visualizar una comparación de lado a lado del texto 2502 captado y los resultados 2500 procesados. En este ejemplo, el sistema realiza el OCR correctamente de 104 de 118 caracteres y signos de puntuación. Los resultados son mejores si se ignora la separación de sílabas, en donde los fragmentos de palabras asociadas no están en el diccionario. En esta interfaz, el usuario puede seleccionar entre no modo de lenguaje, un modelo de frecuencia, y el modelo de diccionario.

40 Haciendo referencia ahora a la figura 26, se ilustra una comparación de modelos de lenguaje a lo largo de una parte del texto captado de acuerdo con la arquitectura de la presente invención. A la izquierda, se usa el modelo de diccionario, y se muestra una tasa de reconocimiento mucho más alta que la mostrada a la derecha, que no usa un modelo.

CLASIFICADOR

45 Como se ha indicado anteriormente en el presente documento, la presente invención puede emplear diversos esquemas basados en inteligencia artificial para realizar diversos aspectos de la invención objeto. Por ejemplo, puede facilitarse un procedimiento para determinar un umbral para determinar si un carácter es un signo de puntuación a través de un sistema y procedimiento clasificador automático.

Un clasificador es una función que mapea un vector de atributo de entrada, $x = (x_1, x_2, x_3, x_4, x_n)$, a una confianza en que la entrada pertenece a una clase, es decir, $f(x) = \text{confianza}(\text{clase})$. Tal clasificación puede emplear un análisis probabilístico y/o basado en la estadística (por ejemplo, la factorización en las utilidades de análisis y los costes) para pronosticar o deducir una acción que un usuario desea que se realice automáticamente.

50 Una máquina de vectores de soporte (SVM) es un ejemplo de un clasificador que puede emplearse. La SVM opera encontrando una hipersuperficie en el espacio de posibles entradas, hipersuperficie que intenta dividir los criterios de activación a partir de los eventos no activados. Intuitivamente, esto hace que la clasificación correcta de los datos de prueba sea próxima, pero no idéntica a los datos de adiestramiento. Se incluyen otros enfoques dirigidos y no dirigidos del modelo de clasificación, por ejemplo, puede emplearse el bayesiano ingenuo, las redes bayesianas, los árboles de decisión, y los modelos de clasificación probabilísticos que proporcionan diferentes patrones de independencia. La clasificación como se usa en el presente documento, incluye también la regresión estadística que se utiliza para desarrollar modelos de prioridad.

Como se apreciará fácilmente a partir de la especificación objeto, la invención objeto puede emplear clasificadores que están adiestrados explícitamente (por ejemplo, a través de unos datos de adiestramiento genérico), así como adiestrados implícitamente (por ejemplo, a través de la observación del comportamiento del usuario que recibe la información extrínseca). Por ejemplo, las SVM se configuran a través de una fase de aprendizaje o de adiestramiento dentro de un constructor clasificador y un módulo de selección de características. Por lo tanto, el clasificador(s) puede usarse para realizar automáticamente una serie de funciones, que incluyen, pero no se limitan a lo siguiente: determinar el contexto sustantivo del documento con el fin de resolver mejor el símbolo reconocido; y realizar un procesamiento de bigrama, trigramas etc., para aumentar la probabilidad de que la palabra actual sea correcta. Por ejemplo, si la palabra se reconoce inicialmente como "wosd", la palabra anterior o siguiente (bigrama) también podría procesarse. Si la palabra siguiente es procesador, es más probable que "wosd" deba ser "word".

En otra implementación, el contenido de la palabra puede analizarse para aumentar la tasa de reconocimiento. Por ejemplo, si la palabra es money, y una palabra reconocida posterior es securities, y puede hacerse una inferencia de que el contenido puede tratarse de finanzas, valores, o similares. De este modo, el sistema puede activar automáticamente uno o más modelos de lenguaje que incluyen términos y/o capacidades relacionadas con la resolución de estos términos. Por otra parte, tal modularización del modelo de lenguaje facilita los requisitos de datos reducidos en que el dispositivo informático en el que se emplea la novedosa arquitectura se utiliza principalmente en un entorno financiero, solo el módulo(s) de lenguaje de este tipo necesita instalarse o ponerse a disposición en el dispositivo. Esto está en contraste con proporcionar un modelo de lenguaje de cajón de sastre que puede sobrecargar las capacidades de almacenamiento y/o las capacidades de procesamiento del dispositivo en el que está instalado.

En otra aplicación más, el clasificador pueden emplearse para estimar el número de símbolos a resolver y deducir, de manera que si el número de símbolos es dos, por ejemplo, es más probable que la palabra sea un pronombre o artículo. En consecuencia, tal preprocesamiento de la palabra relacionada puede emplearse antes de que se habiliten otros modelos de lenguaje.

El clasificador también puede usarse para ajustar automáticamente los umbrales usados para determinar el signo de puntuación, y otros símbolos a reconocerse. Este ajuste puede basarse en el contenido correctamente reconocido por el que avanza el procedimiento de OCR. Es decir, existe un mecanismo de retroalimentación que automáticamente impacta en el umbral basado en la historia.

El umbral también puede ajustarse automáticamente basándose en la importancia del contenido. Por ejemplo, a medida que avanza el procedimiento de OCR, si se determina que las palabras reconocidas son cada vez más relacionadas con un documento contractual legal, el clasificador puede hacer que el sistema eleve el umbral para mejorar la precisión (también, para registros médicos, por ejemplo), e incluso volver a ejecutar el procedimiento, si se desea. Considerando que, si el contenido está determinado a estar más relacionado con una lista de la compra, el umbral podría relajarse a medida que el procedimiento de OCR continúa.

Muchos otros usos para el clasificador incluyen hacer inferencias basadas en el tamaño del símbolo, el tipo de signo de puntuación, duración de las frases, el tipo de símbolo en relación a su localización en una cadena de símbolos, etc.

Obsérvese que la novedosa arquitectura no se limita al empleo de una red neuronal convolucional para el reconocimiento de caracteres. Puede emplearse un procedimiento de cálculo o de clasificación total de los resultados en cada posición en el documento, y obtener la programación dinámica sin el procesamiento convolucional. Sin embargo, en una implementación menos robusta, esto podría afectar al rendimiento del sistema.

Haciendo referencia ahora a la figura 27, se ilustra un diagrama de bloques de un ordenador que puede operar para ejecutar la arquitectura desvelada. Con el fin de proporcionar un contexto adicional para diversos aspectos de la presente invención, la figura 27 y la siguiente exposición están destinados a proporcionar una breve descripción general de un entorno 2700 informático adecuado en el que pueden implementarse los diversos aspectos de la presente invención. Aunque la invención se ha descrito anteriormente en el contexto general de las instrucciones ejecutables por ordenador que puede ejecutarse en uno o más ordenadores, los expertos en la materia reconocerán que la invención también puede implementarse en combinación con otros módulos de programa y/o como una combinación de hardware y software.

En general, los módulos de programa incluyen rutinas, programas, componentes, estructuras de datos, etc., que realizan tareas específicas o implementan tipos de datos abstractos específicos. Por otra parte, los expertos en la materia apreciarán que los procedimientos de la invención puedan practicarse con otras configuraciones de sistemas informáticos, incluyendo sistemas informáticos multiprocesador o de un solo procesador o, miniordenadores, ordenadores centrales, así como los ordenadores personales, dispositivos informáticos de mano, electrónica de consumo basada en microprocesadores o programable, y similares, cada uno de los cuales puede acoplarse operativamente a uno o más dispositivos asociados.

Los aspectos ilustrados de la invención también pueden ponerse en práctica en entornos informáticos distribuidos donde ciertas tareas se realizan por dispositivos de procesamiento remoto que están enlazados a través de una red

de comunicaciones. En un entorno informático distribuido, los módulos de programa pueden estar localizados tanto en dispositivos de almacenamiento de memoria locales como remotos.

Un equipo incluye normalmente una variedad de medios legibles por ordenador. Los medios legibles por ordenador pueden ser cualquier medio disponible al que puede accederse por el ordenador e incluye medios tanto volátiles como no volátiles, medios extraíbles como no extraíbles. A modo de ejemplo, y no de limitación, los medios legibles por ordenador pueden comprender medios de almacenamiento informáticos y medios de comunicación. Los medios de almacenamiento informáticos incluyen tanto medios volátiles como no volátiles, extraíbles como no extraíbles implementados en cualquier procedimiento o tecnología para el almacenamiento de información tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento informáticos incluyen, pero no se limitan a, RAM, ROM, EEPROM, memoria flash u otra tecnología de memoria, CD-ROM, disco de vídeo digital (DVD) u otro almacenamiento en disco óptico, cassetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnéticos, o cualquier otro medio que pueda usarse para almacenar la información deseada y que pueda accederse por el ordenador.

Los medios de comunicación incorporan normalmente instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos en una señal de datos modulada tal como una onda portadora u otro mecanismo de transporte, e incluye cualquier medio de entrega de información. La expresión "señal de datos modulada" significa que una señal tiene una o más de sus características establecidas o cambiadas de una manera tal como para codificar la información en la señal. A modo de ejemplo, y no de limitación, los medios de comunicación incluyen medios cableados tales como una red cableada o una conexión de cableado directo, y medios inalámbricos tales como acústicos, de RF, infrarrojos y otros medios inalámbricos. Las combinaciones de cualquiera de los anteriores también deberían incluirse dentro del ámbito de los medios legibles por ordenador.

Haciendo referencia de nuevo a la figura 27, se ilustra un entorno 2700 a modo de ejemplo para implementar diversos aspectos de la invención que incluye un ordenador 2702, incluyendo el ordenador 2702 una unidad 2704 de procesamiento, una memoria 2706 de sistema y un bus 2708 de sistema. El bus 2708 de sistema acopla los componentes del sistema, incluyendo, pero no limitado a, la memoria 2706 de sistema a la unidad 2704 de procesamiento. La unidad 2704 de procesamiento puede ser cualquiera de los diversos procesadores disponibles comercialmente. Los microprocesadores duales y otras arquitecturas multi-procesador también pueden emplearse como la unidad 2704 de procesamiento.

El bus 2708 de sistema puede ser cualquiera de diversos tipos de estructuras de bus que pueden interconectarse además a un bus de memoria (con o sin un controlador de memoria), un bus periférico, y un bus local que usa cualquiera de una variedad de arquitecturas de buses disponibles comercialmente. La memoria 2706 de sistema incluye una memoria 2710 de solo lectura (ROM) y una memoria 2712 de acceso aleatorio (RAM). Un sistema de entrada/salida básico (BIOS) se almacena en una memoria 2710 no volátil tal como ROM, EPROM, EEPROM, BIOS que contiene las rutinas básicas que ayudan a transferir información entre los elementos dentro del ordenador 2702, tal como durante el arranque. La RAM 2712 puede incluir también una RAM de alta velocidad tal como una RAM estática para almacenar los datos en caché.

El ordenador 2702 incluye además, una unidad 2714 de disco duro interna (HDD) (por ejemplo, EIDE, SATA), una unidad 2714 de disco duro externa que también puede configurarse para su uso externo en un chasis adecuado (no mostrado), una unidad 2716 de disquete magnético (FDD), (por ejemplo, para leer de o escribir en un disquete 2718 extraíble) y una unidad 2720 de disco óptico, (por ejemplo, para leer un disco 2722 CD-ROM o, para leer de o escribir en otros medios ópticos de gran capacidad tal como el DVD). La unidad 2714 de disco duro, la unidad 2716 de disco magnético y la unidad 2720 de disco óptico puede conectarse al bus 2708 de sistema por una interfaz de unidad 2714 de disco duro, una interfaz de unidad 2726 de disco magnético y una interfaz 2728 de unidad óptica, respectivamente. La interfaz 2724 para implementaciones de unidad externa incluye al menos uno o ambos de entre el bus serie universal (USB) y las tecnologías de interfaz IEEE 1394.

Las unidades y sus medios legibles por ordenador asociados proporcionan un almacenamiento no volátil de datos, estructuras de datos, instrucciones ejecutables por ordenador, y así sucesivamente. Para el ordenador 2702, las unidades y los medios adaptan el almacenamiento de cualquier dato en un formato digital adecuado. Aunque la descripción del medio legible por ordenador anterior se refiere a un HDD, un disquete magnético extraíble y un medio óptico extraíble tal como un CD o un DVD, debería apreciarse por los expertos en la materia que otros tipos de medios que son legibles por un ordenador, tales como las unidades zip, los cassetes magnéticos, las tarjetas de memoria flash, los cartuchos, y similares, también pueden usarse en el entorno operativo a modo de ejemplo, y, además, que cualquiera de tales medios pueden contener instrucciones ejecutables por ordenador para realizar los procedimientos de la presente invención.

Una serie de módulos de programa pueden almacenarse en las unidades y la RAM 2712, incluyendo un sistema 2730 operativo, uno o más programas 2732 de aplicación, otros módulos 2734 de programa y datos 2736 de programa. Todo o partes del sistema operativo, las aplicaciones, módulos, y/o los datos también pueden almacenarse en caché en la RAM 2712.

Se apreciará que la presente invención puede implementarse con diversos sistemas operativos comercialmente disponibles o combinaciones de sistemas operativos.

Un usuario puede introducir órdenes e información en el ordenador 2702 a través de uno o más dispositivos de entrada cableados/inalámbricos, por ejemplo, un teclado 2738 y un dispositivo de puntero, tal como un ratón 2740. Otros dispositivos de entrada (no mostrados) pueden incluir un micrófono, un mando a distancia IR, un palanca de mando, un controlador para juegos, un lápiz óptico, una pantalla táctil, o similares. Estos y otros dispositivos de entrada están a menudo conectados a la unidad 2704 de procesamiento a través de una interfaz 2742 de dispositivos de entrada que está acoplada al bus 2708 de sistema, pero pueden estar conectados por otras interfaces, tales como un puerto paralelo, un puerto serie IEEE 1394, un puerto de juegos, un puerto USB, una interfaz de infrarrojos, etc.

Un monitor 2744 u otro tipo de dispositivo de visualización también están conectados al bus 2708 de sistema a través de una interfaz, tal como un adaptador 2746 de vídeo. Además del monitor 2744, un ordenador incluye normalmente otros dispositivos de salida periféricos (no mostrados), tales como altavoces, impresoras, etc.

El ordenador 2702 puede operar en un entorno de red usando conexiones lógicas a través de comunicaciones cableadas y/o inalámbricas a uno o más ordenadores remotos, tales como un ordenador(s) 2748 remoto. El ordenador(s) 2748 remoto puede ser una estación de trabajo, un equipo servidor, un encaminador, un ordenador personal, un ordenador portátil, un dispositivo de entretenimiento basado en un microprocesador, un dispositivo parecido u otro nodo de red común, y que normalmente incluye muchos o todos los elementos descritos en relación con el ordenador 2702, aunque, con fines de brevedad, solamente se ilustra un dispositivo 2750 de almacenamiento de memoria. Las conexiones lógicas representadas incluyen una conectividad cableada/inalámbrica a una red 2752 de área local (LAN) y/o redes más grandes, por ejemplo, una red 2754 de área amplia (WAN). Tales entornos de red LAN y WAN son comunes en oficinas y empresas, y facilitan las redes amplias de ordenadores de empresa, tales como las intranets, todas las cuales pueden conectarse a una red de comunicaciones global, por ejemplo, Internet.

Cuando se usa en un entorno de red LAN, el ordenador 2702 se conecta a la red 2752 local a través de una interfaz de red de comunicaciones cableada y/o inalámbrica o un adaptador 2756. El adaptador 2756 puede facilitar la comunicación cableada o inalámbrica con la LAN 2752, que también puede incluir un punto de acceso inalámbrico dispuesto en la misma para comunicarse con el adaptador 2756 inalámbrico. Cuando se usa en un entorno de red WAN, el ordenador 2702 puede incluir un módem 2758, o se conecta a un servidor de comunicaciones de la red, o tiene otros medios para establecer las comunicaciones sobre la WAN 2754, tal como a través de la Internet. El módem 2758, que puede ser interno o externo y un dispositivo cableado o inalámbrico, está conectado al bus 2708 de sistema a través de la interfaz 2742 de puerto serie. En un entorno de red, los módulos de programa representados en relación con el ordenador 2702, o partes de los mismos, pueden almacenarse en el dispositivo 2750 de memoria/almacenamiento remoto. Se apreciará que las conexiones de red mostradas son a modo de ejemplo y pueden usarse otros medios de establecer un enlace de comunicaciones entre los ordenadores.

El ordenador 2702 puede operarse para comunicarse con cualquier dispositivo inalámbrico o entidades dispuestas operativamente en comunicación inalámbrica, por ejemplo, una impresora, un escáner, un ordenador de escritorio y/o portátil, un asistente de datos portátil, un satélite de comunicaciones, cualquier pieza de un equipo o localización asociada con una etiqueta detectable de manera inalámbrica (por ejemplo, un quiosco, un puesto de periódicos, unos aseos), y un teléfono. Esto incluye al menos las tecnologías inalámbricas Wi-Fi y Bluetooth. Por lo tanto, la comunicación puede ser una estructura predefinida como con la red convencional o simplemente una comunicación ad hoc al menos entre dos dispositivos.

Wi-Fi o fidelidad inalámbrica, permite la conexión a Internet desde un sofá en casa, una cama en una habitación de hotel o una sala de conferencias en el trabajo, sin necesidad de cables. Wi-Fi es una tecnología inalámbrica como un teléfono móvil, que permite a tales dispositivos, por ejemplo, ordenadores, enviar y recibir datos en interiores y exteriores; cualquier lugar dentro del alcance de una estación base. Las redes Wi-Fi utilizan unas tecnologías de radio llamadas IEEE 802.11 (a, b, g, etc.) para proporcionar una conectividad inalámbrica segura, fiable y rápida. Una red Wi-Fi puede usarse para conectar ordenadores entre sí, a Internet, y a las redes cableadas (que usan IEEE 802.3 o Ethernet). Las redes Wi-Fi operan en las bandas de radio sin licencia de 2,4 y 5 GHz, con una velocidad de datos de 11 Mbps (802.11b) o 54 Mbps (802.11 a) o con productos que contengan las dos bandas (banda dual), por lo que las redes pueden proporcionar un rendimiento del mundo real similar a las redes Ethernet cableadas 10BaseT básicas usadas en muchas oficinas.

Haciendo referencia ahora a la figura 28, se ilustra un diagrama de bloques esquemático de un entorno 2800 informático a modo de ejemplo de acuerdo con la presente invención. El sistema 2800 incluye uno o más clientes 2802. El cliente(s) 2802 puede ser hardware y/o software (por ejemplo, subprocesos, procedimientos, dispositivos informáticos). El cliente(s) 2802 puede alojar una galleta(s) informática y/o información contextual asociada empleando, por ejemplo, la presente invención. El sistema 2800 también incluye uno o más servidores 2804. El servidor(s) 2804 también puede ser hardware y/o software (por ejemplo, subprocesos, procedimientos, dispositivos informáticos). Los servidores 2804 pueden albergar subprocesos para realizar transformaciones empleando, por ejemplo, la presente invención. Una posible comunicación entre un cliente 2802 y un servidor 2804 puede ser en forma de un paquete de datos adaptado para transmitirse entre dos o más procedimientos informáticos. El paquete

de datos puede incluir, por ejemplo, una galleta informática y/o información contextual asociada. El sistema 2800 incluye una infraestructura 2806 de comunicaciones (por ejemplo, una red de comunicación global, tal como Internet) que puede emplearse para facilitar las comunicaciones entre el cliente(s) 2802 y el servidor(s) 2804.

Las comunicaciones pueden facilitarse a través de una tecnología cableada (incluida la fibra óptica) y/o inalámbrica.

- 5 El cliente(s) 2802 se conecta operativamente a uno o más almacenes 2808 de datos de cliente que puede emplearse para almacenar información local para el cliente(s) 2802 (por ejemplo, una galleta(s) informática y/o una información contextual asociada). Del mismo modo, el servidor(s) 2804 se conecta operativamente a uno o más almacenes 2810 de datos de servidor que pueden emplearse para almacenar información local para los servidores 2804.
- 10 Lo que se ha descrito anteriormente incluye unos ejemplos de la presente invención. Por supuesto, no es posible describir cada combinación concebible de componentes o metodologías con fines de describir la presente invención, pero un experto en la materia puede reconocer que son posibles muchas otras combinaciones y permutaciones de la presente invención. Por consiguiente, la presente invención pretende abarcar todas las alteraciones, modificaciones y variaciones que están dentro del ámbito de las reivindicaciones adjuntas. Además, en la medida en que el término "incluye" se usa, o en la descripción detallada o en las reivindicaciones, tal término pretende ser inclusivo de una manera similar a la expresión "que comprende", en la medida "que comprende" se interpreta cuando se emplea como una palabra transitoria en una reivindicación.
- 15

REIVINDICACIONES

1. Un sistema que facilita el reconocimiento óptico de caracteres, OCR, de símbolos de baja resolución, en el que una cadena de símbolos es representativa de una palabra, y en el que los símbolos representan caracteres, que comprende:

- 5 un componente de segmentación para detectar espacios entre los símbolos para determinar las líneas de texto, y para fragmentar las líneas de texto en palabras individuales; y
- un componente de reconocimiento para reconocer caracteres (206) usando un reconocedor de caracteres basado en el aprendizaje de máquina para explorar a través de cada una de las palabras individuales para predecir qué carácter es probable que se presente en una localización determinada, para reconocer el signo de puntuación y para reconocer las palabras;

10 en el que dicho reconociendo de un signo de puntuación se usa para identificar si un carácter final de una palabra es un signo de puntuación, que comprende:

- determinar un carácter más probable para cada posible posición de carácter final de la palabra;
- generar una puntuación para cada carácter más probable;
- 15 determinar si la palabra es una palabra puntuada, en el que la palabra es una palabra puntuada si el carácter más probable con la puntuación más alta es un signo de puntuación y si la puntuación del carácter más probable con la puntuación más alta está por encima de un umbral predeterminado; y

en el que dicho reconocimiento de palabras comprende:

- reconocer la palabra usando el resto de la palabra sin el signo de puntuación, y añadir el signo de puntuación a la palabra reconocida; y
- 20 reconocer las palabras (208) individuales reconciliando una secuencia de salidas del reconocedor de caracteres con una palabra específica usando una programación dinámica y un diccionario.

25 2. El sistema de la reivindicación 1, que comprende además un algoritmo de clasificación de aprendizaje de máquina para procesar una entrada de nivel de grises y para facilitar el reconocimiento del símbolo calculando una función de clasificación en una localización de símbolos.

3. El sistema de la reivindicación 1, que comprende además un algoritmo de clasificación de aprendizaje de máquina que es una red (102) neural convolucional para procesar la entrada de nivel de grises y para calcular la función de clasificación en múltiples localizaciones de símbolos.

30 4. El sistema de la reivindicación 1, que comprende además un componente de detección de línea para facilitar la detección de las líneas de texto de una imagen de nivel de grises de los símbolos.

5. El sistema de la reivindicación 1, que comprende además al menos uno de entre un modelo (106) de lenguaje y un modelo (104) de programación para facilitar la interpretación del símbolo como una palabra o una parte de la misma.

35 6. El sistema de la reivindicación 1, en el que el componente de reconocimiento es para reconocer además tanto un símbolo como una cadena de símbolos, en el que el símbolo o la cadena de símbolos son representativos de una palabra.

7. El sistema de la reivindicación 1, en el que el componente de reconocimiento usa al menos uno de entre una red (102) neural convolucional, un modelo (106) de lenguaje, y un algoritmo (104) de programación dinámica.

40 8. El sistema de la reivindicación 1, en el que el componente de reconocimiento es para construir además un clasificador invariante para al menos una de las diferentes condiciones de iluminación, fuentes, tamaños de símbolo, tipo de cámara, ángulo y enfoque.

9. El sistema de la reivindicación 1, en el que el componente de reconocimiento es para predecir además qué carácter se representa por el símbolo en una localización dada en la imagen.

45 10. El sistema de la reivindicación 1, en el que el reconocedor es para extraer además una característica simple a una resolución más alta, y para convertir la característica simple en una característica más compleja a una resolución más gruesa.

11. El sistema de la reivindicación 1, que comprende además:

- un modelo (106) de lenguaje para facilitar el procesamiento de un carácter en una cadena de caracteres; y
- 50 un componente (104) de programación dinámica para facilitar el reconocimiento de la cadena de caracteres como una palabra.

12. El sistema de la reivindicación 11, en el que el componente de programación dinámica es para determinar además qué palabra se localiza en un rectángulo que delimita la palabra dada.
13. El sistema de la reivindicación 11, en el que el modelo de lenguaje es un modelo de diccionario para explorar a través de todo un léxico, para evaluar una probabilidad para cada palabra del léxico, y para emitir la palabra más probable.
14. El sistema de la reivindicación 11, en el que el modelo de lenguaje es un lenguaje neutro para producir la interpretación más probable de una secuencia de observaciones de reconocedor de caracteres.
15. El sistema de la reivindicación 11, en el que el modelo de lenguaje es para intercalar además la optimización de programación dinámica con el recorrido de un léxico para calcular la palabra más probable.
16. El sistema de la reivindicación 11, que comprende además un clasificador para hacer automáticamente una inferencia basándose en una o más observaciones asociadas con el reconocimiento de al menos uno de los caracteres y la palabra.
17. Un medio legible por ordenador que tiene instrucciones ejecutables por ordenador para un procedimiento de realizar un reconocimiento óptico de caracteres, OCR, de símbolos de baja resolución, en el que una cadena de símbolos es representativa de una palabra, y en el que los símbolos representan caracteres, comprendiendo el procedimiento:
- detectar espacios entre los símbolos para determinar las líneas de texto;
 - fragmentar las líneas de texto en palabras individuales;
 - reconocer los caracteres (206) usando un reconocedor de caracteres basado en el aprendizaje de máquina para explorar a través de cada una de las palabras individuales para predecir qué carácter es probable que esté presente en una localización dada;
 - reconocer el signo de puntuación, en el que reconocer un signo de puntuación se usa para identificar si un carácter final de una palabra es un signo de puntuación, que comprende:
 - determinar un carácter más probable para cada posible posición de carácter final de la palabra;
 - generar una puntuación para cada carácter más probable;
 - determinar si la palabra es una palabra puntuada, en el que una palabra es una palabra puntuada si el carácter más probable con la puntuación más alta es un signo de puntuación y si la puntuación del carácter más probable con la puntuación más alta está por encima de un umbral predeterminado;
 - reconocer la palabra usando el resto de la palabra sin el signo de puntuación y añadir el signo de puntuación a la palabra reconocida; y
 - reconocer las palabras (208) individuales reconciliando una secuencia de salidas del reconocedor de caracteres con una palabra específica usando una programación dinámica y un diccionario.
18. El medio legible por ordenador de la reivindicación 17, en el que el procedimiento comprende además las acciones de:
- extraer las características simples de los símbolos separados a una resolución más alta; y
 - convertir las características simples en características más complejas a una resolución más gruesa;
- en el que al menos una de las acciones de extraer y convertir se realiza por una red neuronal convolucional.
19. El medio legible por ordenador de la reivindicación 17, en el que el procedimiento comprende además localizar el conjunto de símbolos en la disposición asociada, siendo la disposición asociada una línea, que usa la programación dinámica.
20. El medio legible por ordenador de la reivindicación 17, en el que las acciones de segmentar, extraer y convertir se realizan sustancialmente de manera simultánea.
21. El medio legible por ordenador de la reivindicación 17, en el que el procedimiento comprende además adiestrar a un algoritmo de aprendizaje de máquina para generar los símbolos reconocidos a partir de los símbolos captados en las posiciones horizontales dadas.
22. El medio legible por ordenador de la reivindicación 21, en el que el procedimiento comprende además generar un conjunto de adiestramiento para el algoritmo de aprendizaje de máquina, en el que la acción de generar comprende además al menos una de las acciones de:
- imprimir una recopilación de documentos tanto en soporte papel como en soporte electrónico; y
 - hacer coincidir una posición de cada carácter del soporte papel con una posición de un símbolo captado correspondiente para generar una base de datos de imágenes de caracteres etiquetados.

23. Un procedimiento de realizar el reconocimiento óptico de caracteres, OCR, de símbolos de baja resolución, en el que una cadena de símbolos es representativa de una palabra, y en el que los símbolos representan caracteres, comprendiendo el procedimiento:

5 detectar los espacios entre los símbolos para determinar las líneas de texto;
 fragmentar las líneas de texto en palabras individuales;
 reconocer los caracteres (206) usando un reconocedor de caracteres basado en el aprendizaje de máquina para explorar a través de cada una de las palabras individuales para predecir qué carácter es probable que esté presente en una localización dada;
 10 reconocer el signo de puntuación, en el que reconocer un signo de puntuación se usa para identificar si un carácter final de una palabra es un signo de puntuación, que comprende:

 determinar el carácter más probable para cada posible posición de carácter final de la palabra;
 generar una puntuación para cada carácter más probable;
 determinar si la palabra es una palabra puntuada, en el que una palabra es una palabra puntuada si el carácter más probable con la puntuación más alta es un signo de puntuación y si la puntuación del carácter más probable con la puntuación más alta es superior a un umbral predeterminado;
 15 reconocer la palabra usando el resto de la palabra sin el signo de puntuación y añadir el signo de puntuación a la palabra reconocida; y
 reconocer las palabras (208) individuales reconciliando una secuencia de salidas del reconocedor de caracteres con una palabra específica usando una programación dinámica y un diccionario.

20 24. El procedimiento de la reivindicación 23, que comprende además preprocesar un documento fotografiado con un algoritmo de balance de blancos para maximizar el contraste sobre las regiones del documento.

25. El procedimiento de la reivindicación 23, en el que el algoritmo de clasificación de aprendizaje de máquina es uno de una red (102) neural convolucional y una máquina de vectores de soporte.

26. El procedimiento de la reivindicación 23, que comprende además al menos una de las acciones de:

25 disponer el diccionario en una estructura que maximice la cantidad de cálculos reutilizados; y
 generar una tabla de programación dinámica a medida que se recorre la estructura para determinar una asignación óptima de una observación a un carácter de la palabra individual.

27. El procedimiento de la reivindicación 26, en el que la estructura es una estructura de árbol de prefijos.

30 28. El procedimiento de la reivindicación 23, que comprende además analizar una primera palabra y una segunda palabra como un par con el fin de reconocer la primera palabra.

29. El procedimiento de la reivindicación 23, que comprende además emplear los modelos de lenguaje primero y segundo, de tal manera que si el uso del primer modelo de lenguaje falla al generar una palabra de salida, el segundo modelo de lenguaje se emplea automáticamente.

35 30. El procedimiento de la reivindicación 23, que comprende además detectar las líneas del texto probando unas relaciones geométricas relativas de los componentes conectados.

31. El procedimiento de la reivindicación 23, que comprende además adiestrar el algoritmo de clasificación de aprendizaje de máquina con unas imágenes de adiestramiento, incluyendo el adiestramiento al menos uno de:

 hacer fluctuar aleatoriamente las imágenes de adiestramiento en una ventana de entrada; y
 alterar aleatoriamente el brillo y el contraste de las imágenes de adiestramiento en la ventana de entrada.

40

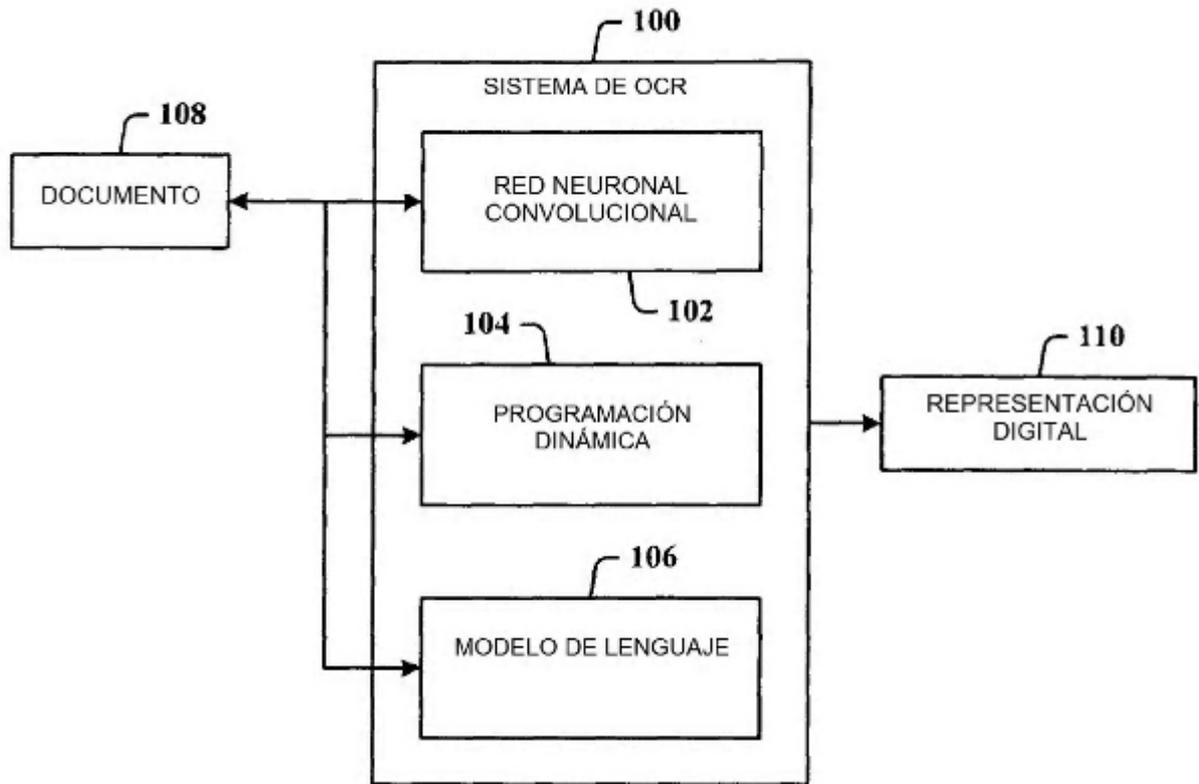


FIG. 1

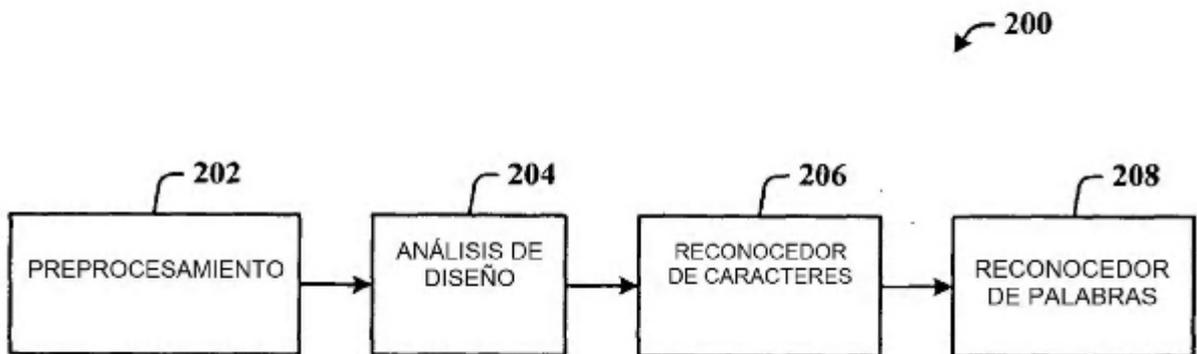


FIG. 2

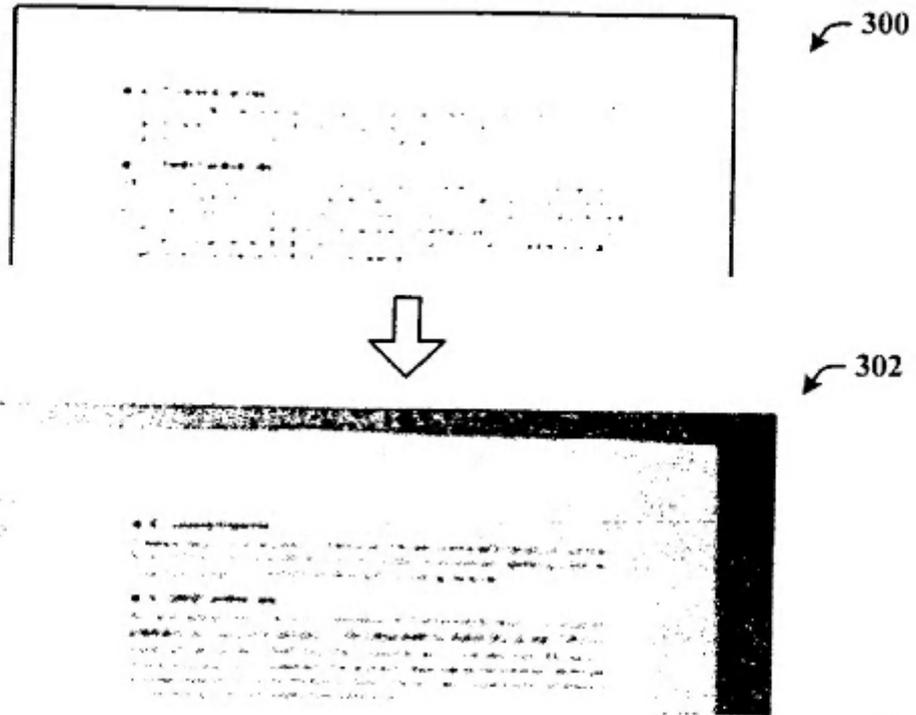
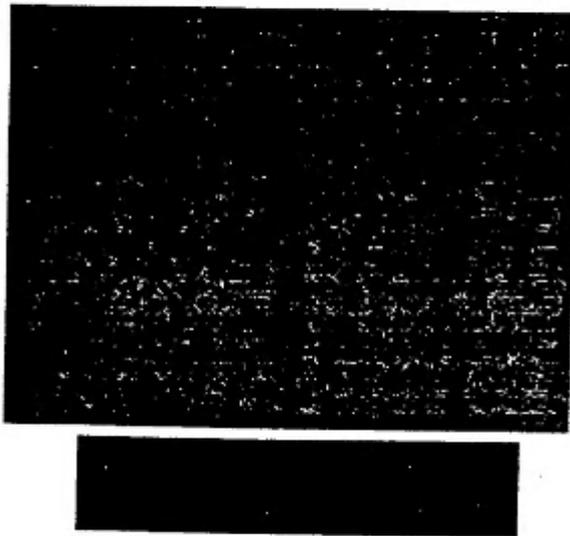


FIG. 3



... for a private gentleman to keep the secret under such circumstances would be very difficult and for a State whose every act is particularly watched by posterity to do this certainly impossible.

Upon my arrival in New York several persons did me the honor of consulting me on the phenomenon in question. I had published in France a work in quarto, in two volumes, entitled *Mysteries of the Great Submarine Grounds*. This book, highly approved of in the learned world, gained for me a special reputation in the rather obscure branch of Natural History. My advice was asked. As long as I could deny the reality of the fact, I continued myself to a decided negative. But soon, being myself driven into a corner, I was obliged to explain myself point by point. I discussed the question in all its forms, politically and scientifically, and I give here an extract from a carefully studied article which I published in the number of the 30th of April. It ran as follows:

... the sea, after attaining a length of sixty feet, increase its size fivefold or tenfold, give it strength proportionate to its size, lengthen its destructive weapons, and you obtain the monster required. It will have the proportions determined by the officers of the Shannon, the instrument required by the penetration of the *Scyra*, and the power necessary to pierce the hull of the steamer.

"Indeed, the monster is armed with a sort of ivory sword, a halberd, according to the expression of certain naturalists. The principal task has the handles of steel. Some of these tasks have been found buried in the bodies of whales, which the unicorn always attacks with success. Others have been drawn out, not without trouble, from the bottoms of ships, which they had pierced through and through, as a gimlet pierces a barrel. The Museum of the Faculty of Medicine of Paris possesses one of these destructive weapons, two yards and a quarter in length, and fifteen inches in diameter in

part of the body

FIG. 4

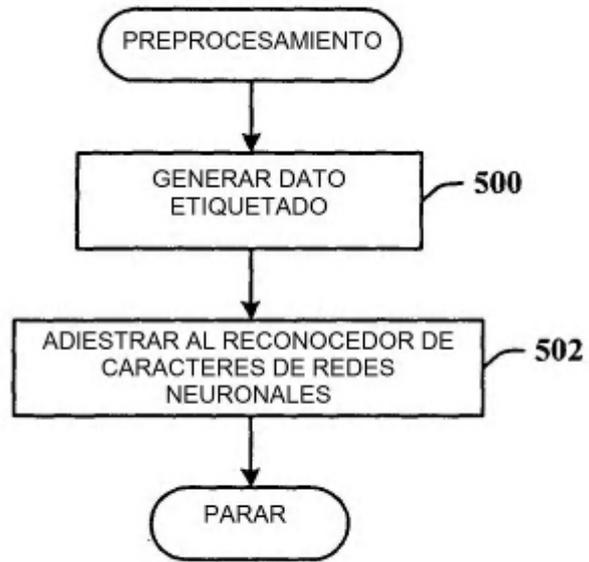


FIG. 5

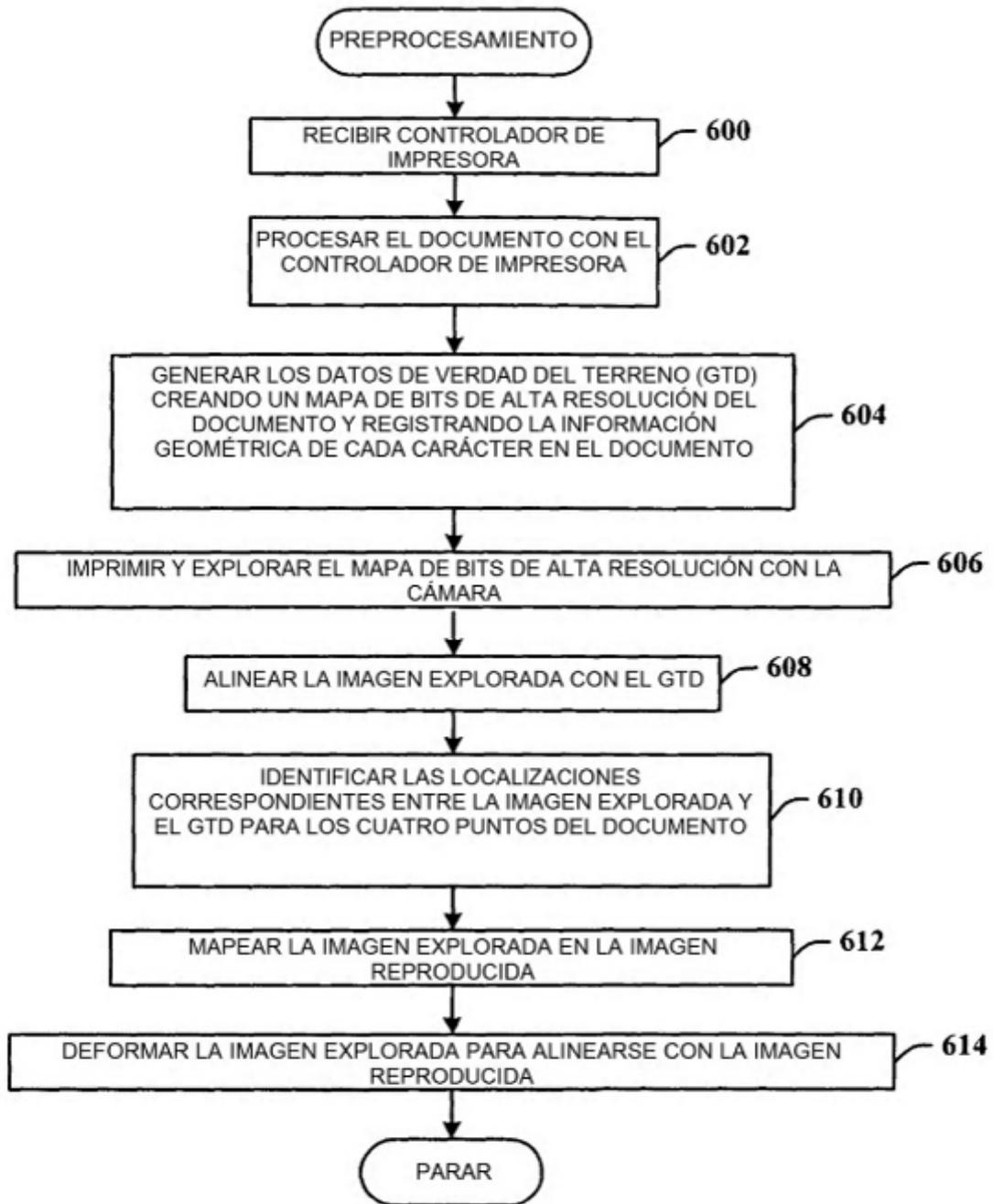


FIG. 6

- 8. **Controlled Employees**
 A person, including an applicant or transferee, who performs a safety-sensitive function for an entity subject to this part however, a criterion is covered only regarding a vehicle designed to transport sixteen or more passengers, including the driver.
- 9. **CRBS Certified Tests**
 All drug testing under the U.S. Department of Transportation regulations must be completed in a laboratory certified by the Department of Health and Human Services. These laboratories have been previously inspected and tested and meet the highest standards for analytical competence. The laboratory must use an immunoassay technique to screen urine specimens for the specific drugs. The laboratory must confirm all positive screens with gas chromatography/mass spectrometry.
- 10. **Disabling Damage**

- 8. **Controlled Employees**
 A person, including an applicant or transferee, who performs a safety-sensitive function for an entity subject to this part however, a criterion is covered only regarding a vehicle designed to transport sixteen or more passengers, including the driver.
- 9. **CRBS Certified Tests**
 All drug testing under the U.S. Department of Transportation regulations must be completed in a laboratory certified by the Department of Health and Human Services. These laboratories have been previously inspected and tested and meet the highest standards for analytical competence. The laboratory must use an immunoassay technique to screen urine specimens for the specific drugs. The laboratory must confirm all positive screens with gas chromatography/mass spectrometry.
- 10. **Disabling Damage**

FIG. 7

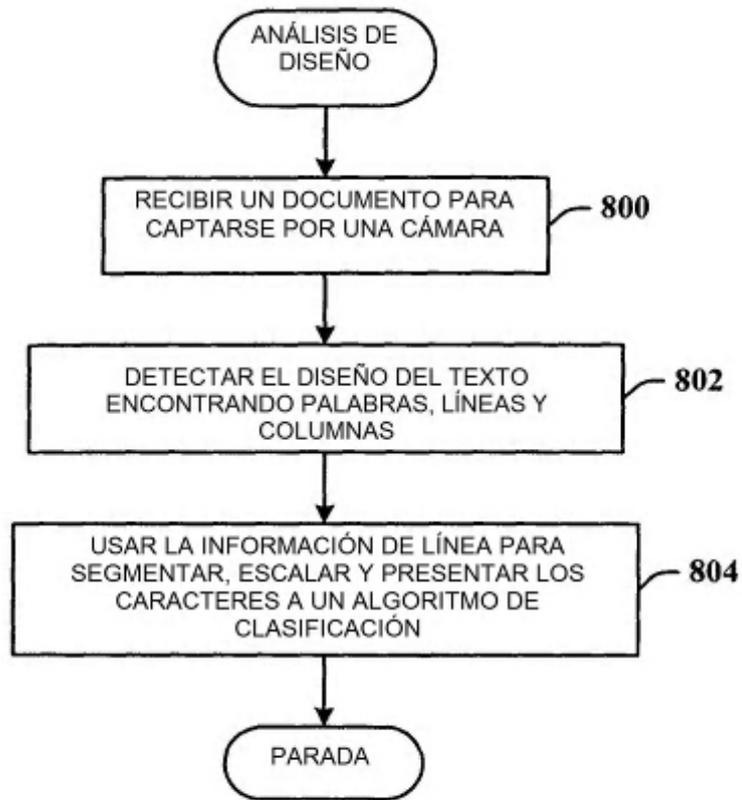


FIG. 8

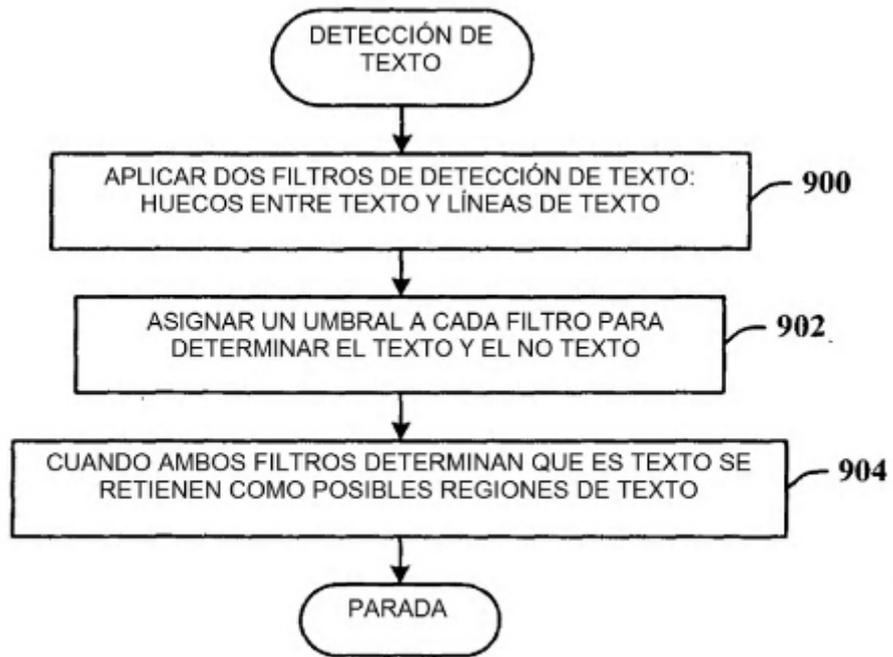


FIG. 9

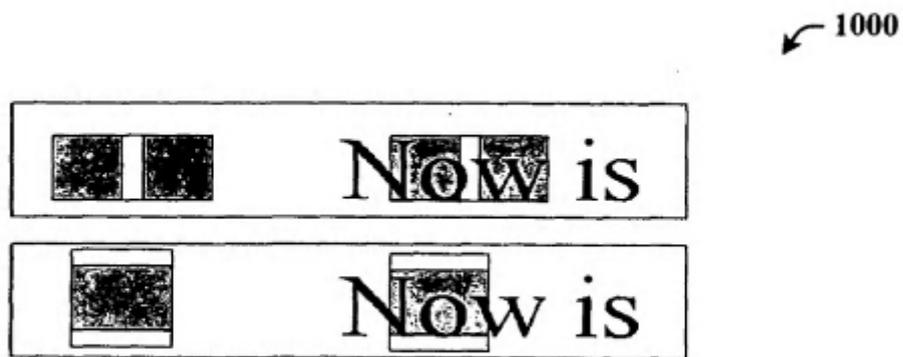


FIG. 10

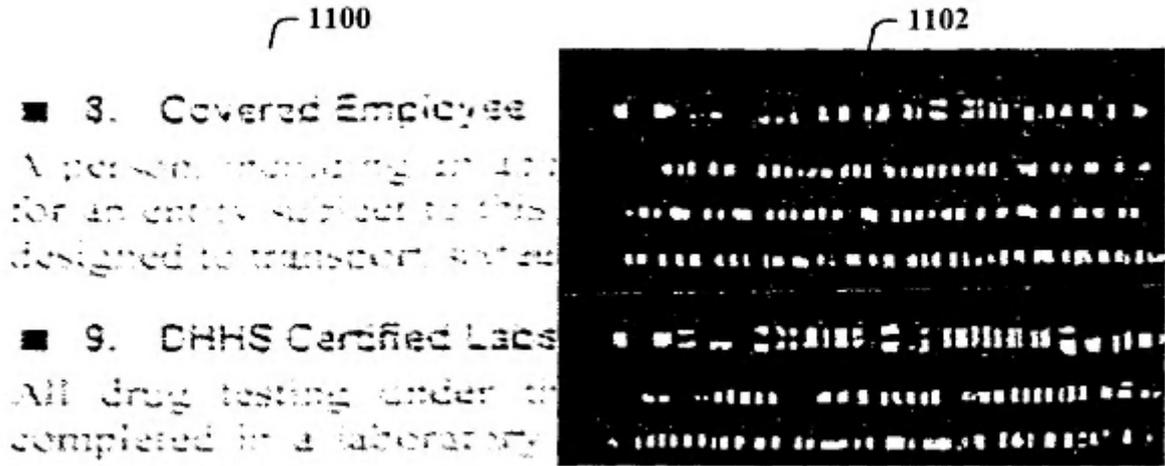


FIG. 11

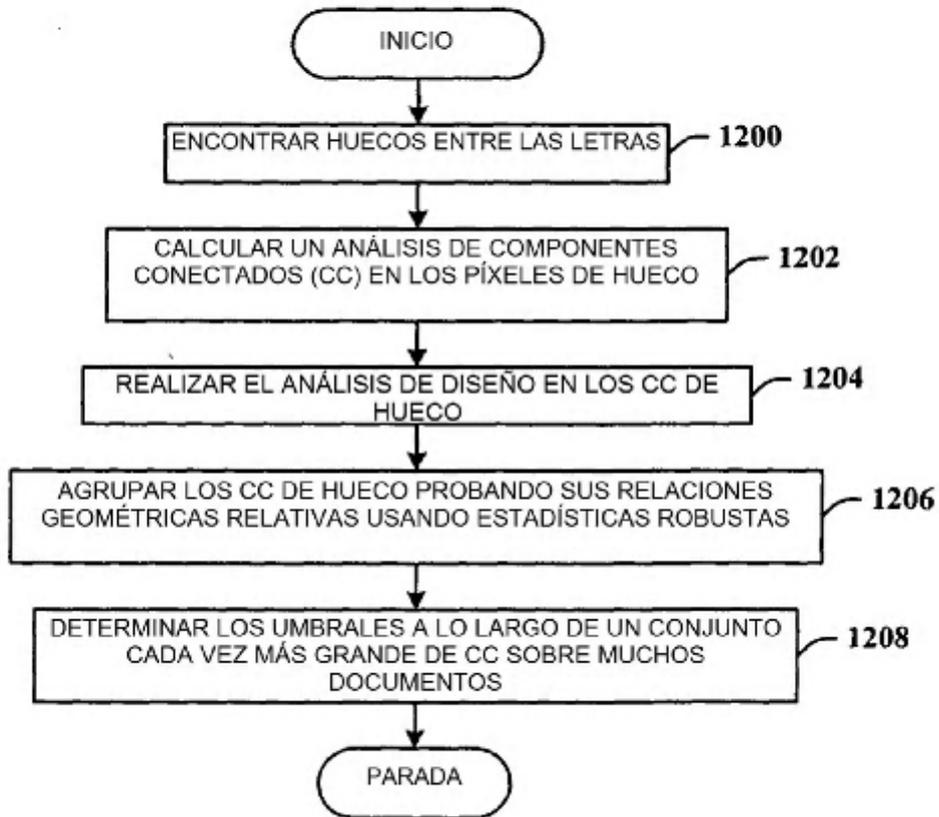


FIG. 12

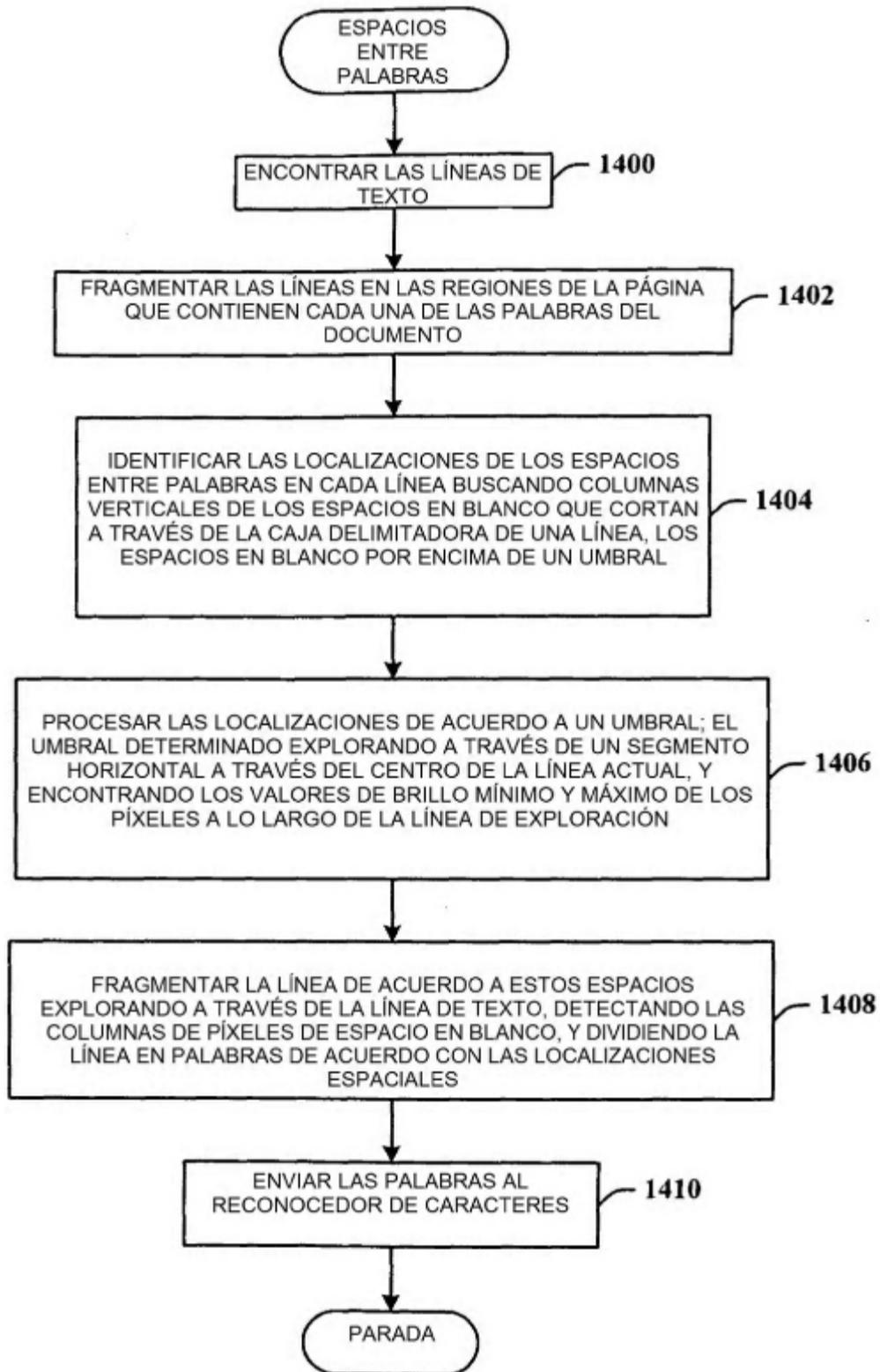


FIG. 14

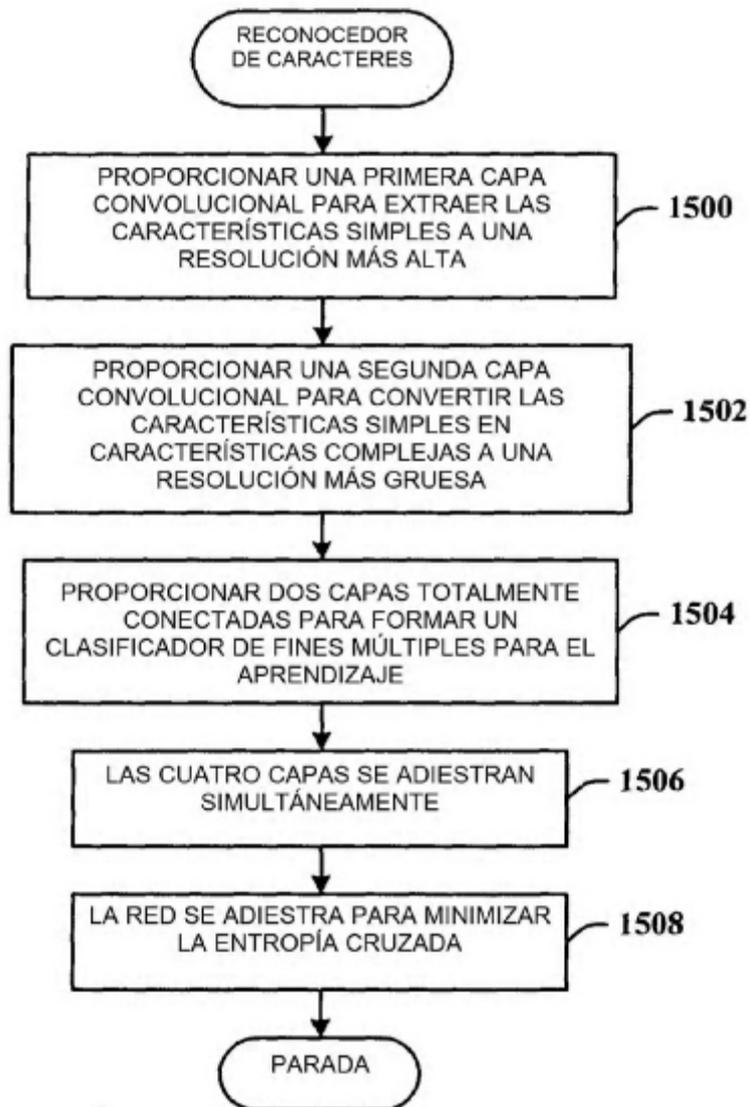


FIG. 15

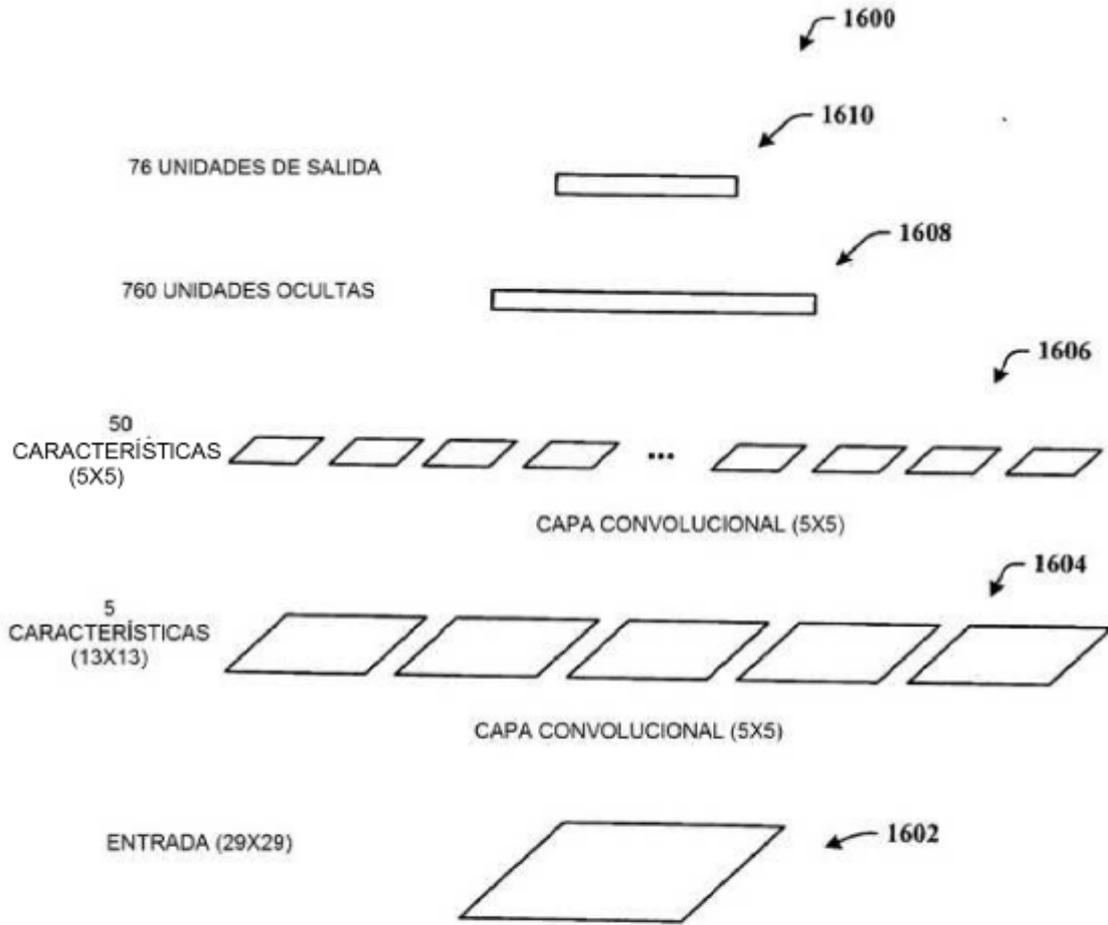


FIG. 16

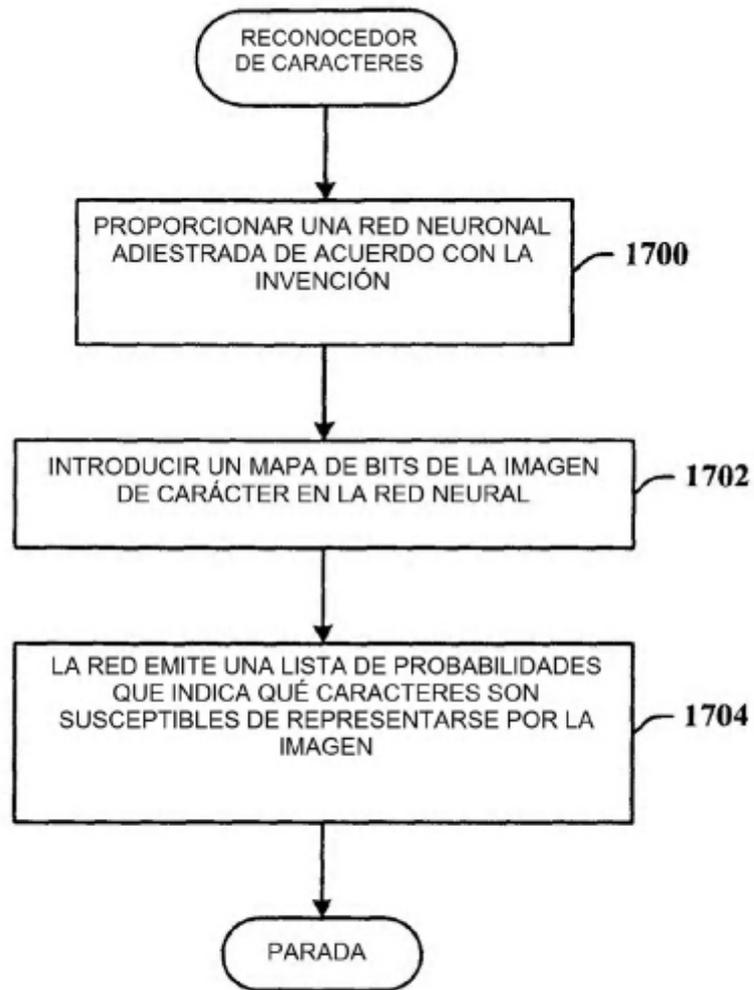


FIG. 17

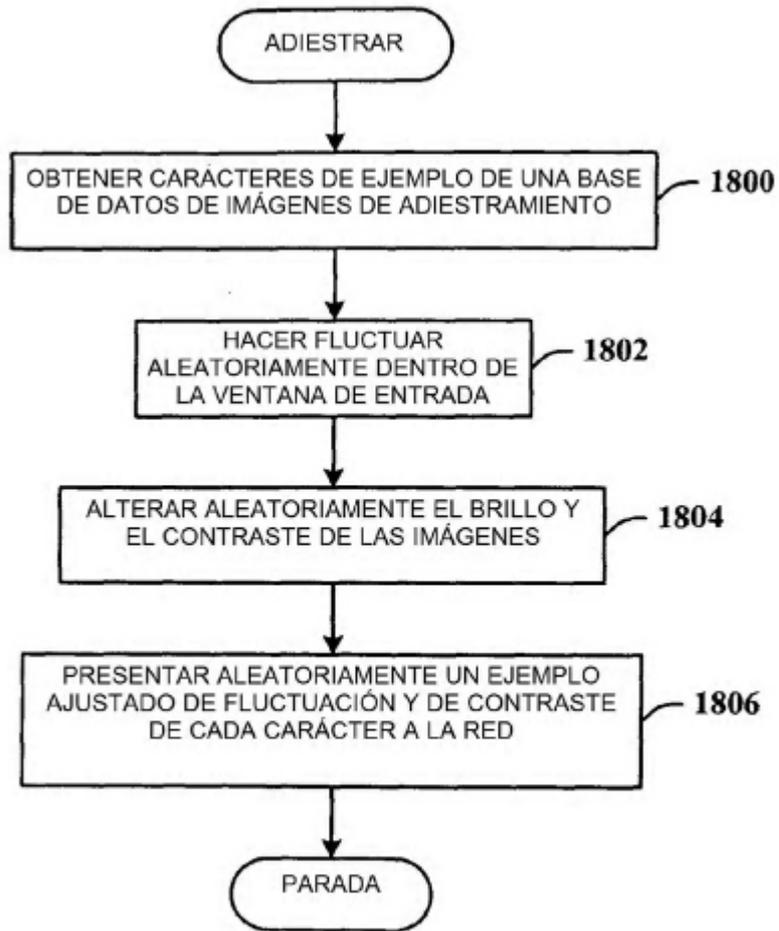


FIG. 18

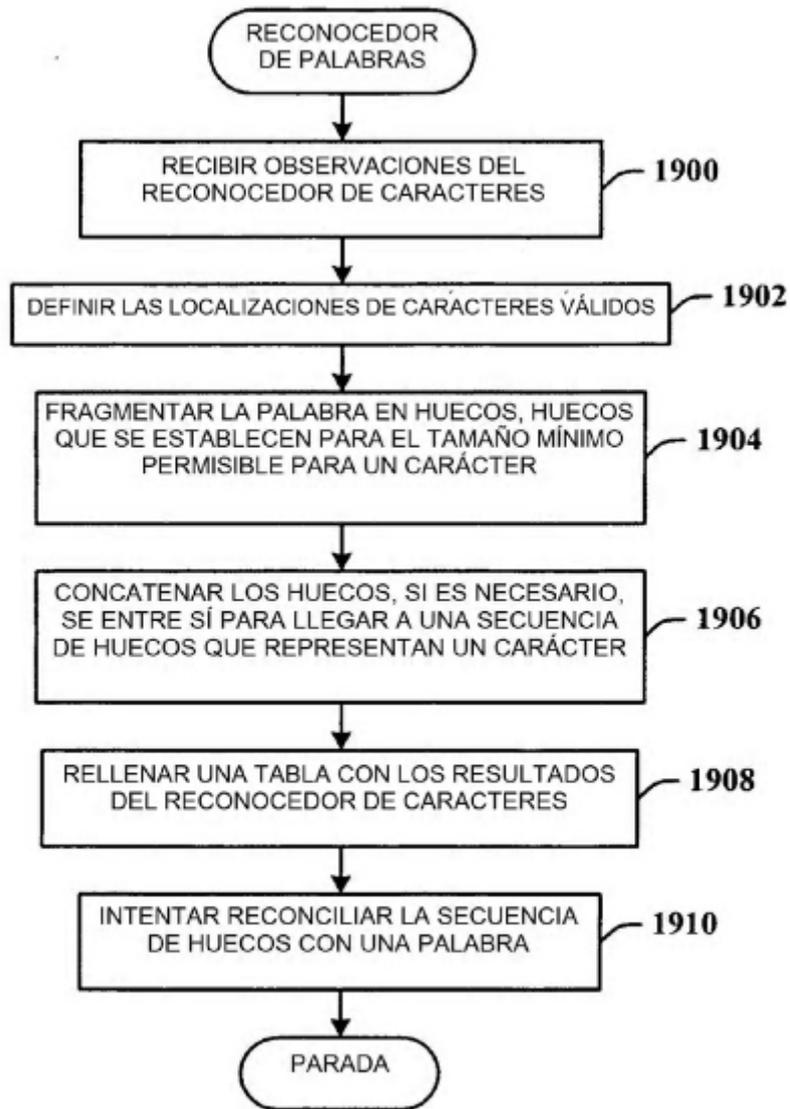


FIG. 19



FIG. 20

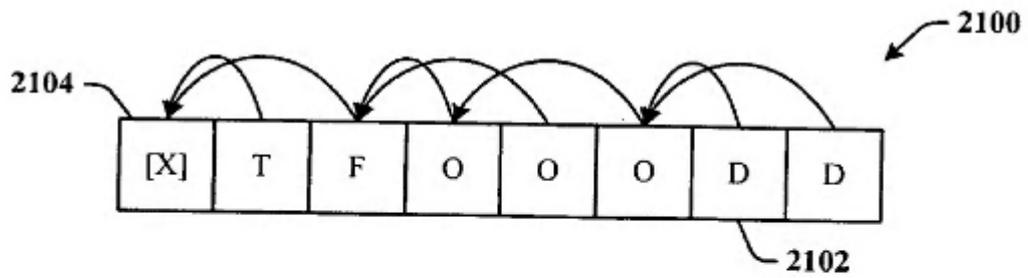


FIG. 21

D	X	X	X	0.1	0.1	0.5	0.7
O	X	X	0.1	0.1	0.1	0.8	0.1
O	X	0.1	0.1	0.9	0.1	0.1	0.1
F	0.1	0.2	0.9	0.5	X	X	X
	1	2	3	4	...	7	

NÚMERO DE HUECOS

FIG. 22

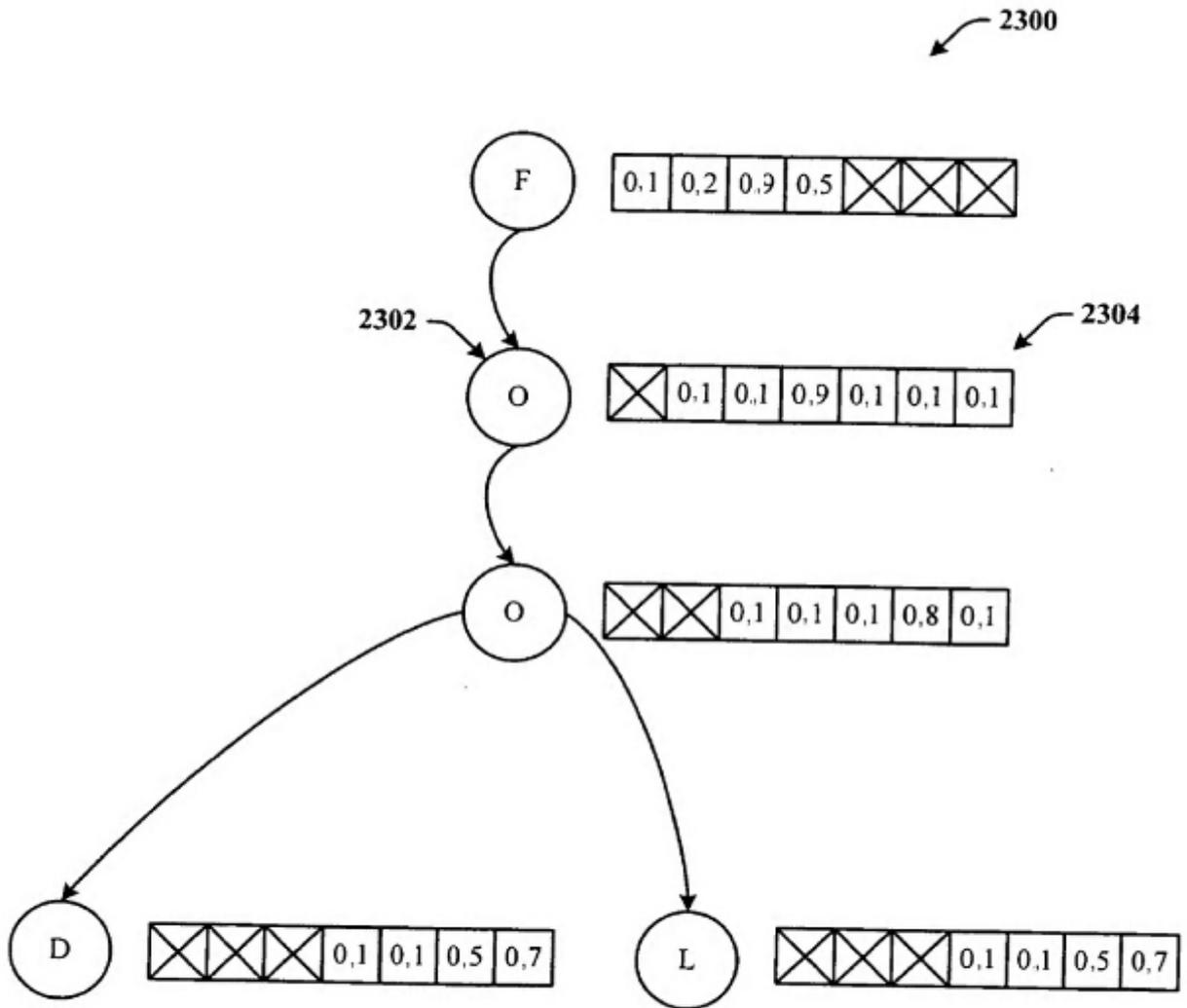


FIG. 23

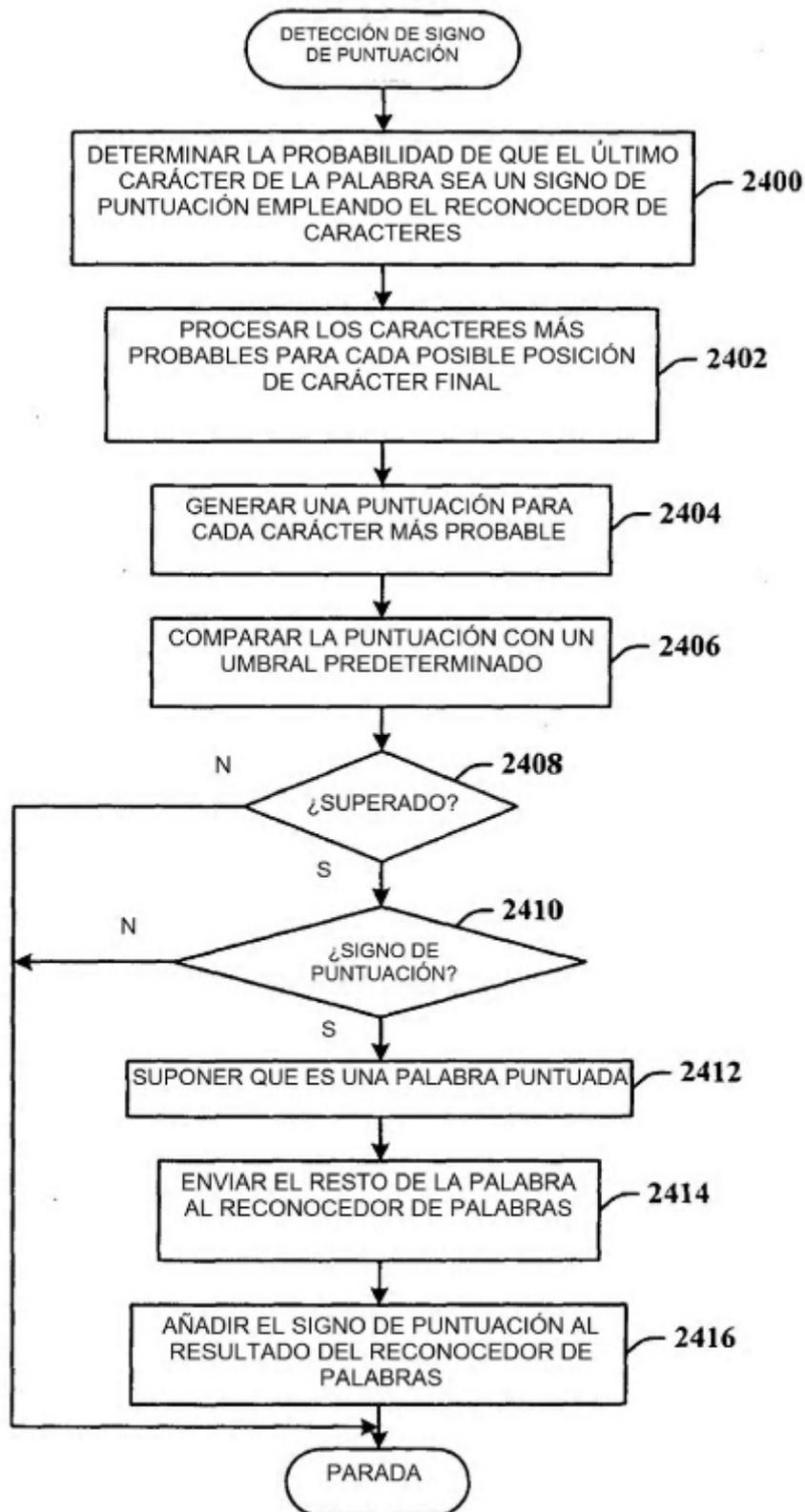


FIG. 24

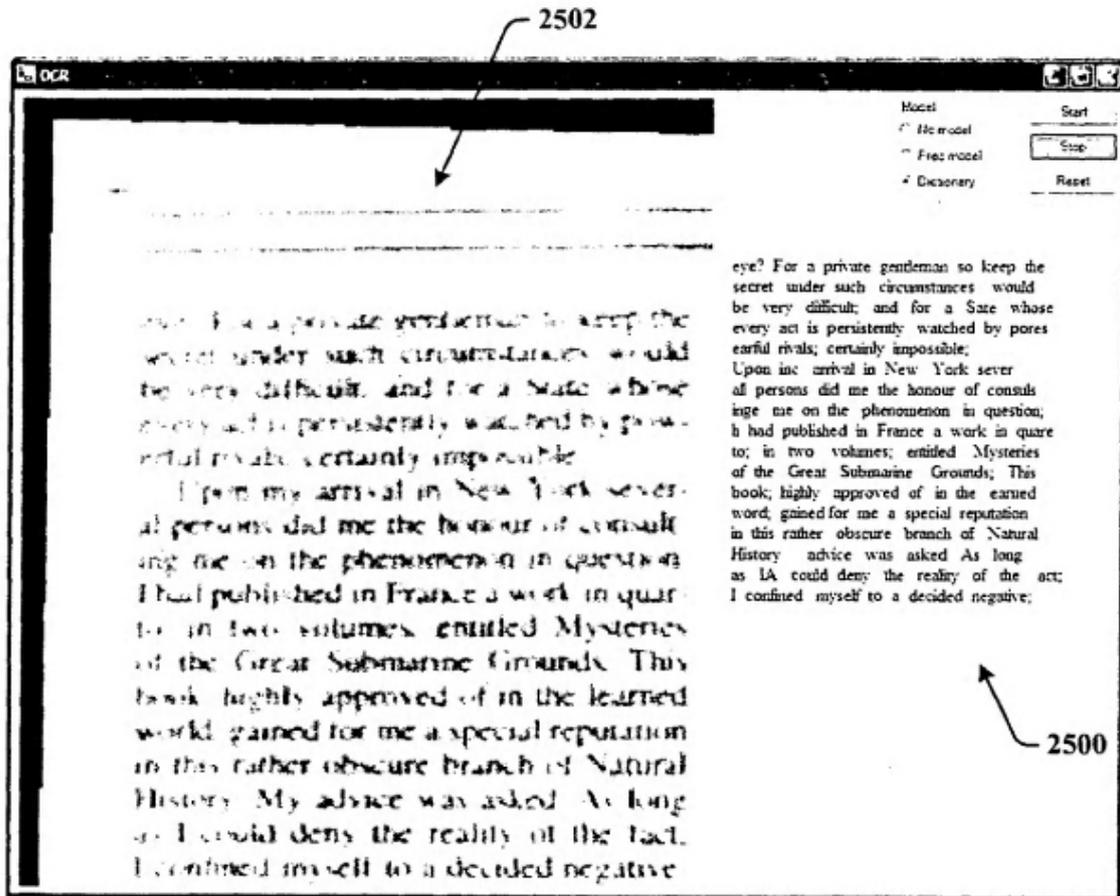


FIG. 25

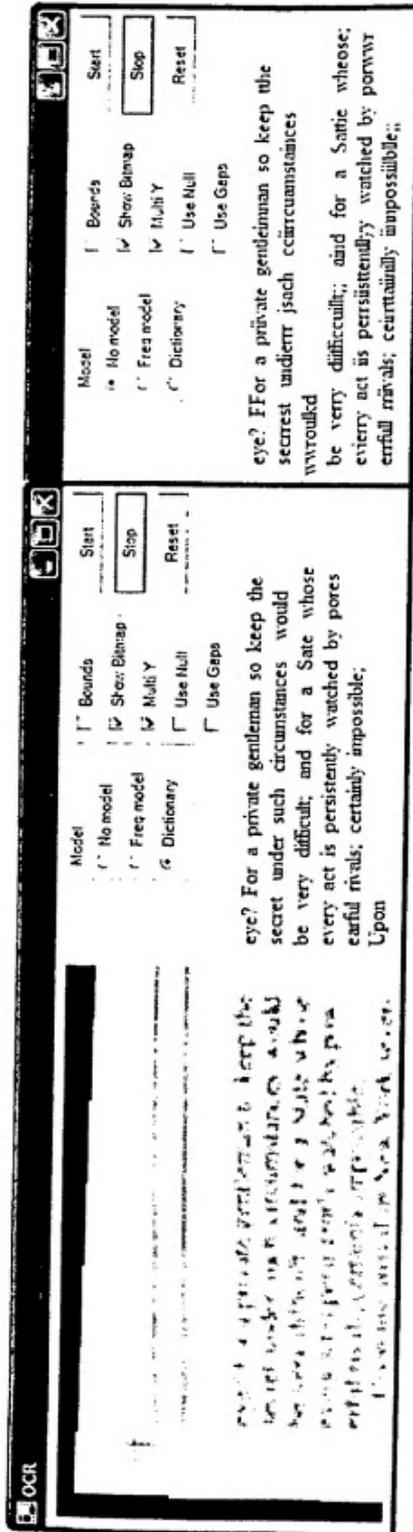


FIG. 26

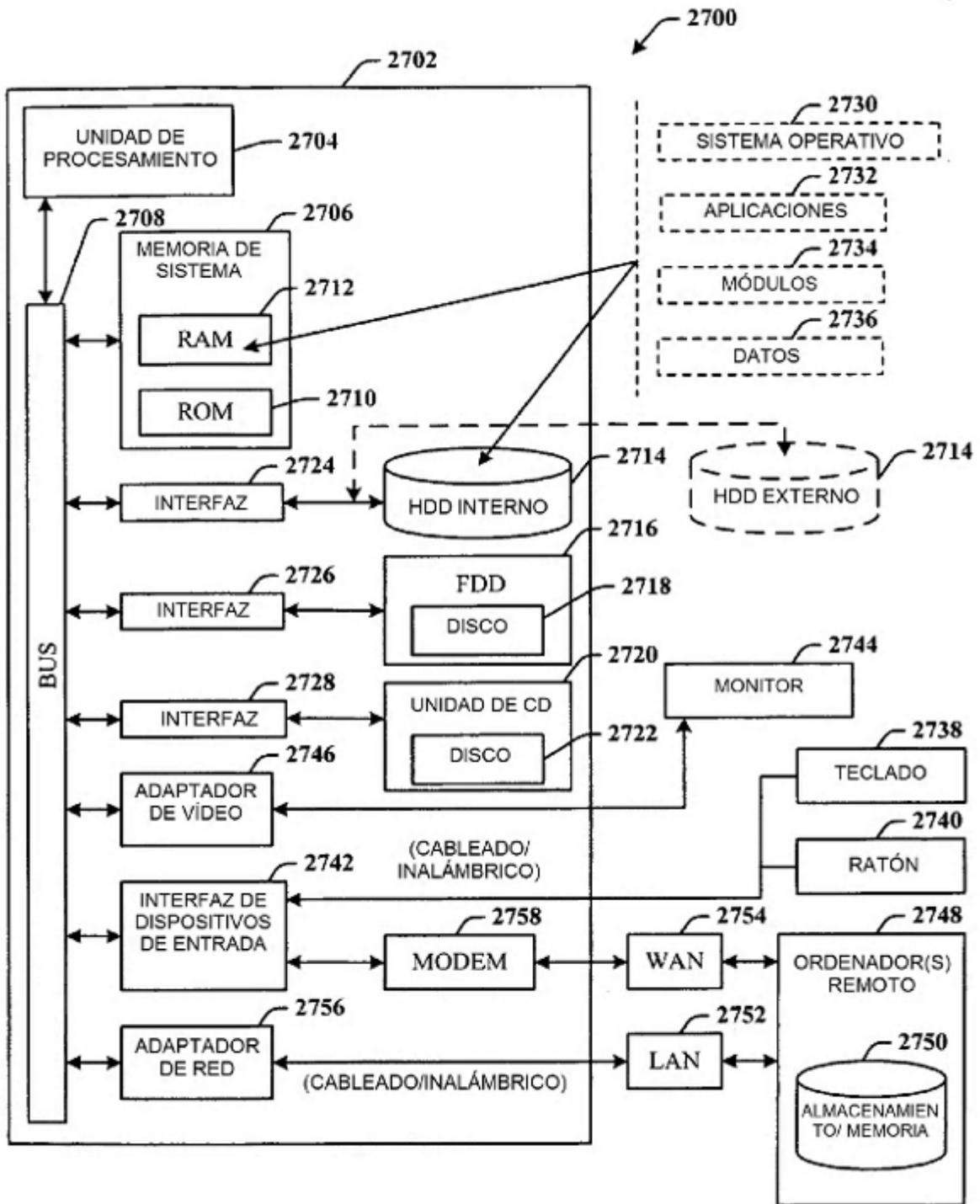


FIG. 27

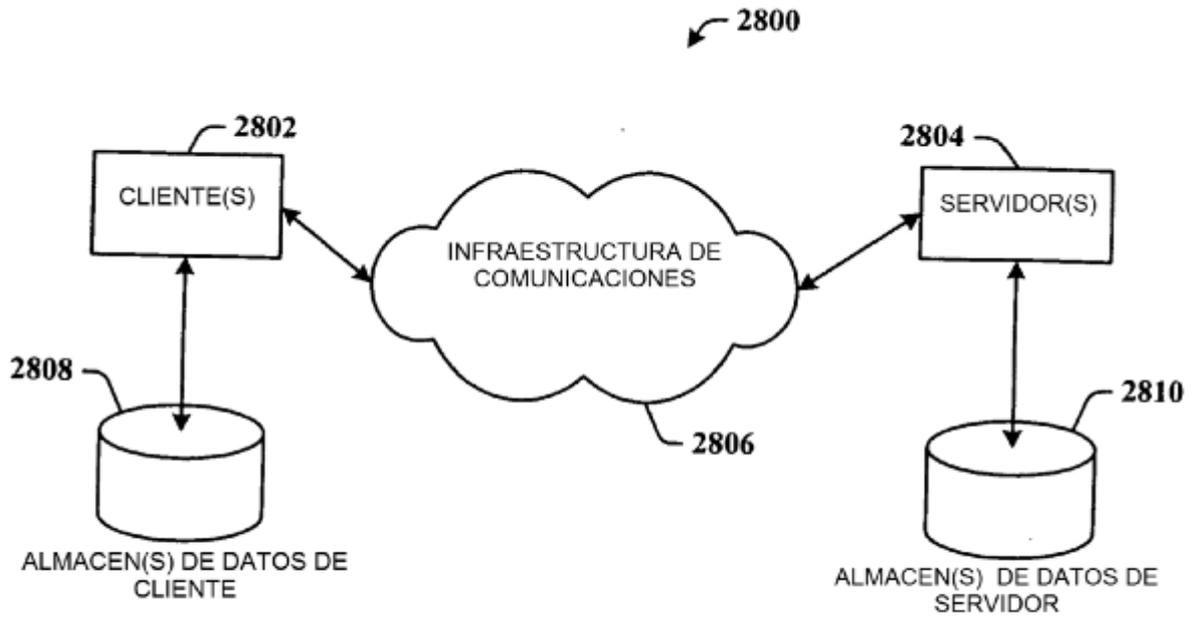


FIG. 28