

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 370 306**

51 Int. Cl.:
H04N 1/32

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04779752 .7**

96 Fecha de presentación: **08.08.2004**

97 Número de publicación de la solicitud: **1654620**

97 Fecha de publicación de la solicitud: **10.05.2006**

54 Título: **FILTRO PARA FAX MASIVO NO SOLICITADO.**

30 Prioridad:
08.08.2003 US 637306

45 Fecha de publicación de la mención BOPI:
14.12.2011

45 Fecha de la publicación del folleto de la patente:
14.12.2011

73 Titular/es:
J2 Global Ireland Limited
Woodford Business Park Unit 3 Santry
Dublin 17, IE

72 Inventor/es:
EL-GAZZAR, Amin;
HARTMUT, Wolf;
PUENTE, Adrian y
NOLASCO, Alexander

74 Agente: **Ungría López, Javier**

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).

ES 2 370 306 T3

DESCRIPCIÓN

Filtro para fax masivo no solicitado

5 Antecedentes de la invención

Estado del campo de la técnica

10 La presente invención se refiere al campo de los dispositivos de telefax y más concretamente al procesamiento de imágenes de telefax masivas no solicitadas en un dispositivo de telefax.

Descripción de la técnica relacionada

15 En segundo lugar sólo tras el teléfono, el dispositivo de telefax sigue siendo un modo principal en las comunicaciones comerciales. En la actualidad, se han desplegado más de ochenta y cinco millones de dispositivos de telefax en todo el mundo y más de mil quinientos millones de números de teléfono de telefax se han difundido de manera pública. Por consiguiente, no ha de ser una sorpresa que la industria de la publicidad comercial directa haya incorporado la imagen de telefax como un medio para la difusión en masa de mensajes de publicidad comercial de la misma forma en la que la industria de la publicidad comercial directa ha adoptado el teléfono y el correo electrónico como un modo de publicidad directa.

20 Históricamente, los medios impresos servían como el modo principal de publicidad en masa no solicitada por parte de la industria de la publicidad comercial directa. Los materiales de publicidad comercial impresa no solicitados, a los que normalmente se hace referencia como "correo basura", podían entregarse de forma masiva a una vasta selección de destinatarios, con independencia de si los destinatarios habían solicitado los materiales de publicidad comercial. Con una tasa de respuesta promedio de un uno a un dos por ciento, el correo basura ha sido una herramienta efectiva en la generación de nuevos ámbitos de venta. Sin embargo, los destinatarios de los correos basura encuentran que en general esta práctica les resulta irritante. Adicionalmente, el franqueo para el envío del correo basura puede ser costoso para "distribuciones de correo" significativas. Por consiguiente, la industria de la publicidad comercial directa busca constantemente unas modalidades igualmente efectivas, pero menos costosas para la entrega de los materiales de publicidad comercial no solicitados.

25 La llegada del correo electrónico ha proporcionado un alivio que necesitaban en gran medida las empresas de publicidad comercial directa, debido a que la entrega de correo electrónico a un vasto número de destinatarios objetivo no precisa de franqueo. Además, la entrega de correo electrónico no solicitado puede hacerse de forma instantánea y el correo electrónico no solicitado puede incluir hiperenlaces integrados a información de servicios o de productos o por lo tanto facilitar una velocidad de respuesta aumentada para la "distribución de correo". Aún así, como es el caso en el ámbito de los medios impresos, el correo electrónico no solicitado, al que se hace referencia habitualmente como "mensaje masivo no solicitado", sigue siendo una molestia para los consumidores en todo el mundo. Como resultado, ha surgido todo un sector industrial de "filtros de mensaje masivo no solicitado" cuya tarea exclusiva es la erradicación de los mensajes masivos no solicitados.

35 Al igual que el correo electrónico, el medio de telefax sigue siendo un medio popular para difundir la publicidad comercial. A diferencia del correo electrónico, no obstante, hay un coste real asociado a la recepción por parte del consumidor del mensaje masivo no solicitado en el medio de telefax. En concreto, una única imagen de telefax masiva no solicitada puede consumir recursos de papel y de tóner y puede consumir ancho de banda de telecomunicaciones que de otro modo puede usarse para el envío mensajes de telefax salientes, o para recibir mensajes de telefax entrantes legítimos. En el Reino Unido, se estima que los mensajes masivos no solicitados en el campo del telefax le cuestan a cada compañía británica más de trescientas de setenta y seis libras al año. En los Estados Unidos, algunas estimaciones sitúan el coste de los mensajes masivos no solicitados en el campo del telefax en unos niveles tan altos como ¡nueve mil millones de dólares al año! Tan grave y objeto de tan intensas disputas se ha convertido el tema de los faxes masivos no solicitados, que el Estado de Missouri demandó con éxito a dos de las más grandes organizaciones de envío fax masivo no solicitadas de acuerdo con la sección 227 de la Ley de protección al consumidor de servicios telefónicos de 1991 (*Telephone Consumer Protection Act*, a continuación en el presente documento "TCPA"), demanda que apoyaron los Estados de Alaska, Arkansas, California, Colorado, Connecticut, Florida, Idaho, Illinois, Iowa, Kentucky, Maryland, Michigan, Minnesota, Nuevo México, Oregón, Dakota del Sur, Texas, Vermont, Virginia del Oeste y el Distrito de Columbia. Merece resaltarse que la sección 227 de la TCPA hace ilegal "el envío de un anuncio no solicitado a una máquina de telefax por teléfono".

45 Sin embargo, el resultado de esta demanda está a la espera de resolución y la parte demandada en la demanda ha puesto en duda la constitucionalidad de la TCPA a la vista de las protecciones a la libertad de expresión comercial que otorga la Constitución de los Estados Unidos. Adicionalmente, sigue además estando pendiente la aplicación de la sección 227 de la TCPA a las comunicaciones de fax por Internet. En concreto, ya no se requiere que dos máquinas de telefax por teléfono coexistan para intercambiar datos de telefax. De hecho, la recomendación T.37 de la ITU-T especifica la transmisión de una imagen de telefax a través de Internet en ausencia de máquinas de telefax por teléfono.

Merece resaltarse que los filtros de mensaje masivo no solicitado, pese a ser efectivos en el campo del correo electrónico, no tienen aplicación en el campo del telefax. En concreto, el correo electrónico se diferencia de las imágenes de telefax de varias maneras distintas e importantes. Merece resaltarse sobre todo que el correo electrónico puede entregarse en un formato de texto legible — normalmente el formato MIME. Como resultado, los filtros de mensaje masivo no solicitado pueden actuar directamente sobre la naturaleza, el contenido y la disposición del texto en el correo electrónico. Más concretamente, el filtro de mensaje masivo no solicitado puede buscar la aparición de palabras concretas que pueden indicar la presencia de un mensaje masivo no solicitado. Las imágenes de telefax, en comparación, se entregan en un formato de imagen procesada por exploración de trama cuyos contenidos no pueden interpretarse más allá de los meros píxeles en blanco y negro o en color. Por lo tanto, las técnicas de filtrado de mensaje masivo no solicitado por correo electrónico siguen siendo poco efectivas en el campo del telefax y se han considerado otras soluciones de forma alternativa.

Varias tecnologías, a pesar de que no tratan el mensaje masivo no solicitado directamente, pueden verse como que proporcionan unas soluciones rudimentarias al problema de fax masivo no solicitado. Por ejemplo, en la Patente de los Estados Unidos con n.º US-A-5 999 600 a nombre de Dong-Myong Shin para Facsimile Mail System And Method For Controlling The Same (Sistema de correo por telefax y procedimiento para controlar el mismo), una clave secreta que se suministra a través de señales de DTMF puede restringir la entrega de un “mensaje literal” cuando la clave secreta que se proporciona a través de las señales de DTMF no se corresponde con una clave secreta almacenada previamente. Como otro ejemplo, en la Publicación de Solicitud de Patente de los Estados Unidos US-A-2002/0191226 para Call-Screening Device For A Facsimile Machine (Dispositivo de detección de llamadas para una máquina de telefax), el número de teléfono de la llamada entrante puede compararse con una base de datos de números de teléfono para determinar si se permite el establecimiento de una conexión telefónica tras la cual puede proporcionarse la transmisión de datos de fax. Curiosamente, en la Patente de los Estados Unidos con n.º US-B-6 553 494 a nombre de Randal Glass para *Method And Apparatus For Applying And Verifying A Biometric-Based Digital Signature To An Electronic Document* (Método y aparato para la aplicación y verificación de una firma digital de base biométrica a un documento electrónico), unos datos biométricos que se acoplan un documento pueden usarse para autenticar el documento por un destinatario.

Haciendo una referencia más concreta al campo del telefax, se han comercializado varias tecnologías anti-mensaje masivo no solicitado específicas para fax. A modo de ejemplo, GFI Software Ltd. ha desarrollado la línea de productos GFI FAXmaker™ en la que pueden enviarse y recibirse faxes a través de clientes de correo electrónico ordinario. Como parte del producto GFI FAXmaker, se proporciona un “filtro de fax basura” a través del cual los faxes masivos no solicitados pueden “borrarse de forma automática” en función del número de teléfono del remitente del fax basura. Específicamente, una base de datos de números de teléfono asociada con el remitente de un fax que se identifica manualmente como un fax basura puede usarse para examinar los faxes recibidos posteriormente.

En comparación, MFM Communication Software, Inc. de Cincinnati, Ohio, los Estados Unidos produce el producto Fax Everywhere™ en el que los faxes basura identificados manualmente pueden etiquetarse como “mensaje masivo no solicitado”. Una vez que se ha etiquetado, los contenidos del fax etiquetado pueden compararse con faxes posteriormente recibidos para reconocer un mensaje masivo no solicitado. Cuando el noventa y cinco por ciento (95 %) del fax recibido se corresponde con una imagen procesada por exploración de trama de un fax etiquetado almacenado como un mensaje masivo no solicitado, el fax recibido puede declararse como un mensaje masivo no solicitado. Por lo tanto, el producto Fax Everywhere puede verse como un procesador de detección de fax masivo no solicitado asistido manualmente que carece de capacidades de examen de mensaje masivo no solicitado de autoaprendizaje o preventivo.

Aún así, la mera comparación de una imagen de telefax con una imagen almacenada en un depósito de software de imágenes de fax masivo no solicitado puede ser un proceso intensivo en cuanto a recursos y que consume mucho tiempo. Lo que es más importante, en un sistema de detección y de bloqueo de mensaje masivo no solicitado en el que un mensaje masivo no solicitado ha de identificarse manualmente, el “daño” se habrá “hecho” una vez que el mensaje masivo no solicitado se ha reconocido manualmente como tal — concretamente en un entorno con un alto volumen de faxes en el que un único caso de difusión de fax masivo no solicitado puede alcanzar a miles de usuarios finales en una única “distribución”. Por consiguiente, sería deseable detectar de forma dinámica los faxes masivos no solicitados de una forma con autoaprendizaje sin requerir una intervención manual.

El documento EP-A-1 001 593 da a conocer procedimientos y aparato para operar sobre un mensaje que no es de texto (voz o fax) que incluye para cada uno de un número de mensajes de voz, buscar en cada mensaje de voz muestras de señal que tengan uno de una pluralidad de conjuntos de las características predefinidas (tal como cualquier número de siete o de diez dígitos que se supone a continuación que es un número de teléfono) tal como se da a conocer. Para cada muestra de señal que se encuentra que tiene uno de los conjuntos de las características predefinidas, la muestra de señal encontrada se asocia con una cadena de texto equivalente y esta cadena de texto se almacena en una envolvente para el mensaje. Basándose en una entrada de usuario, puede buscarse en los mensajes una cadena de texto dada y cuando se encuentra una envolvente de mensaje que contiene la cadena de texto dada, puede operarse sobre tal mensaje.

El documento GB-A-2 373 130 da a conocer un procedimiento para el procesamiento de correo electrónico para detectar correo electrónico en masa y/o comercial no solicitado ("mensaje masivo no solicitado") y un sistema del mismo, comprendiendo el procedimiento las etapas de monitorizar uno o más nodos de una red para patrones de tráfico de correo electrónico indicativos de una difusión masiva de mensajes de correo electrónico no solicitado y tras la detección de tales patrones iniciar una acción correctiva automática y/o alertar a un operador. El procedimiento puede comprender la etapa de descomponer cada correo electrónico recibido en sus partes constituyentes para buscar un contenido indicativo de una difusión masiva de mensajes y registrar datos del correo electrónico descompuesto en una base de datos. El procedimiento puede comprender la etapa de ejecutar de forma continua o sin interrupción un algoritmo frente a entradas en una base de datos para identificar patrones de correo electrónico tráfico indicativos de una difusión masiva de mensajes. Un algoritmo que se da a conocer es: si el correo es más grande que un cierto tamaño, no hacer el registro (un correo de mensaje masivo no solicitado es en general pequeño, para mantener a un nivel bajo las taras del remitente).

El documento US-A-5 911 139 da a conocer un sistema y procedimiento para la búsqueda y recuperación basada en el contenido de objetos visuales. Un motor de recuperación de información de visual (VIR, *visual information retrieval*) de base utiliza un conjunto de primitivas universales para operar sobre los objetos visuales. Un motor de VIR extensible permite que se definan y que se registren unas primitivas personalizadas y modulares. Una primitiva personalizada se dirige a los problemas específicos de dominio y puede utilizar cualquier técnica de comprensión de imagen. Los atributos de objeto pueden extraerse a lo largo de la totalidad de la imagen o a lo largo de sólo una parte del objeto. Un esquema se define como una colección específica de primitivas. Un esquema específico implica un conjunto específico de características visuales que han de procesarse y un vector de características correspondiente que ha de usarse para una puntuación de similitud basada en el contenido. Una interfaz de registro de primitivas registra las primitivas personalizadas y facilita el almacenamiento de una función de análisis y de una función de comparación a una tabla de esquemas. Cualquiera de varias técnicas estadísticas puede usarse en la función de análisis para una primitiva dada. Por ejemplo, puede usarse un análisis por histogramas, tal como un histograma en color. Como otro ejemplo, una primitiva de intensidad media puede definirse como la suma de la intensidad de la totalidad de los píxeles en una imagen dividida por el número de píxeles en la imagen.

El documento EP-A-0 856 980 da a conocer un procedimiento de encaminamiento para un sistema de distribución de información (10), tal como un sistema de distribución de comunicaciones automática, que permite una libertad de contenido sustancial en la formulación de un mensaje al sistema. Los mensajes se convierten a un formato que puede buscarse por medios informáticos y se someten a un análisis de contenido para identificar habilidades ventajosas para responder a los mensajes. En una aplicación, el mensaje entrante es una transmisión de telefax que se convierte a una información de texto generada por ordenador usando un módulo de reconocimiento óptico de caracteres.

El documento WO-A-03/010680 da a conocer un sistema, un procedimiento, y un producto de programa informático que se prevé para detectar un mensaje no deseado. En primer lugar, se recibe un mensaje de correo electrónico. Se descompone el texto en el mensaje de correo electrónico. Se recopilan unas estadísticas asociadas con el texto usando un analizador de estadísticas. Se enseña a un motor de red neural que se acopla al analizador de estadísticas a reconocer los mensajes no deseados basándose en unos indicadores de estadísticas. Los indicadores de estadísticas se analizan utilizando el motor de red neural para determinar si el mensaje de correo electrónico es un mensaje no deseado.

Sumario de la invención

La presente invención se dirige a las deficiencias de la técnica con respecto a la transmisión de imágenes de telefax no solicitadas (a continuación en el presente documento, "fax masivo no solicitado") en el campo del telefax, y proporciona un procedimiento, sistema y aparato novedoso y no obvio para el filtrado de fax masivo no solicitado con el fin de identificar una imagen de fax como un fax masivo no solicitado de forma automática tras la detección de un intento para transmitir fax masivo no solicitado a un dispositivo de salida de fax. Es importante que, a diferencia de las tecnologías de reducción y de eliminación de fax masivo no solicitado convencionales que se conocen en la técnica, en la presente invención, las formas procesadas por exploración de trama de conjunto de imágenes de fax no se procesan directamente para detectar un fax masivo no solicitado. Más bien, la forma procesada por una exploración de trama del conjunto de imágenes de fax puede transformarse en una representación no procesada por exploración de trama a fin de facilitar la detección de fax masivo no solicitado sin consumir unos recursos de sistema no necesarios y sin acometer unas comparaciones de píxeles que consumen mucho tiempo como es el caso en la técnica anterior. Lo más importante es que la presente invención es un sistema de autoaprendizaje dinámico que no requiere el etiquetado manual por adelantado de mensaje masivo no solicitado para evitar el efecto debilitante de una distribución de fax masivo no solicitado.

El alcance de la invención se define en la reivindicación adjunta.

Un filtro para fax masivo no solicitado que se describe en el presente documento puede incluir un procesador de transformación de fax configurado para transformar unas imágenes de fax procesadas por exploración de trama en

unas representaciones no procesadas por exploración de trama de imágenes de fax. Las representaciones no procesadas por exploración de trama pueden incluir estar basadas en números, basadas en texto, basadas en vectores o en otra representación de datos de una imagen de este tipo, lo que incluye un número de páginas que forman una imagen de fax, un número de píxeles blancos y negros que forman una imagen de fax, o la intensidad de una imagen de fax. El filtro para fax masivo no solicitado puede incluir además una base de conocimientos de características de fax que incluye unas características de faxes masivos no solicitados, y unas características de faxes que no son mensajes masivos no solicitados. Por último, el filtro para fax masivo no solicitado puede incluir un comparador configurado para comparar unas representaciones no procesadas por exploración de trama de imágenes de fax que se producen por el procesador de transformación de fax con las características en la base de conocimientos para detectar un fax masivo no solicitado en el filtro para fax masivo no solicitado.

Merece resaltarse que el filtro para fax masivo no solicitado puede incluir opcionalmente un procesador de reconocimiento óptico de caracteres (OCR, *optical character recognition*) que se acopla al procesador de transformación de fax. El procesador de OCR puede configurarse para convertir las imágenes de fax procesadas por exploración de trama en unas representaciones de texto de las imágenes de fax procesadas por exploración de trama. El filtro para fax masivo no solicitado puede incluir opcionalmente además un filtro de perfilado de bordes que se acopla al procesador de transformación de fax. El filtro de perfilado de bordes puede configurarse para producir un conjunto de imágenes intermedio para las imágenes de fax procesadas por exploración de trama que tiene unos bordes resaltados que definen la imagen procesada por exploración de trama de fax. Por consiguiente, puede calcularse un conjunto de vectores para los bordes resaltados en el conjunto de imágenes intermedio.

Como aún una opción adicional, el filtro para fax masivo no solicitado puede incluir un filtro de difuminado que se acopla al procesador de transformación de fax; y, un procesador de hash robusto configurado para producir un valor de hash robusto para una forma difuminada de la imagen procesada por exploración de trama de fax. Por último, el filtro para fax masivo no solicitado puede incluir opcionalmente unos medios para detectar un valor de color para una selección de píxeles en una zona de las imágenes de fax procesadas por exploración de trama. Basándose en los valores de color detectados, el procesador de transformación puede codificar de forma numérica cada una de las imágenes de fax procesadas por exploración de trama con al menos un valor numérico que representa un número de píxeles negros para una única imagen procesada por exploración de trama de fax y un número de píxeles blancos en las únicas imágenes de fax procesadas por exploración de trama.

Un procedimiento que se describe en el presente documento para el filtrado de fax masivo no solicitado puede incluir la transformación de una forma procesada por exploración de trama de una imagen candidata de fax en una o más formas no procesadas por exploración de trama; sometiendo a prueba la forma no procesada por exploración de trama de la imagen candidata de fax frente a unas características de fax tales como características de fax masivo no solicitado, o unas características de imágenes de fax que se conoce que no son de mensaje masivo no solicitado; y, la identificación de la imagen candidata de fax como un fax masivo no solicitado cuando la imagen candidata de fax somete a prueba favorablemente a al menos una de las características del fax masivo no solicitado conocido. A este respecto, cuando una transformación de una imagen de fax muestra una o más de las características de un mensaje masivo no solicitado más allá de un nivel umbral, la imagen de fax puede declararse como un mensaje masivo no solicitado. De forma similar, cuando una transformación de una imagen de fax muestra una o más de las características que no son de mensaje masivo no solicitado más allá de un nivel umbral, puede declararse que la imagen de fax no es un mensaje masivo no solicitado.

La etapa de transformación puede incluir la etapa de procesamiento la imagen candidata de fax en un procesador de reconocimiento óptico de caracteres para convertir la imagen candidata de fax en un texto. De forma alternativa, la etapa de transformación puede incluir las etapas de resalte de bordes en la imagen candidata de fax; y, el cálculo de un conjunto de vectores para los bordes resaltados. Como otra alternativa, la etapa de transformación puede incluir tanto el difuminado de la imagen candidata de fax; y, la producción de un valor de hash robusto para la imagen candidata de fax difuminada. Por último, como otra alternativa más, la etapa de transformación puede incluir la detección de un valor de color para una selección de píxeles en una zona de la imagen candidata de fax. La imagen candidata de fax posteriormente puede codificarse de forma numérica con al menos un valor numérico que representa un número de píxeles negros y un número de píxeles blancos en la imagen candidata de fax.

Es importante que la etapa de realización de prueba pueda incluir la etapa de aplicar las características a la transformación para producir un resultado. Específicamente, la transformación puede compararse con una selección de las características para determinar si la transformación cumple un nivel umbral de los criterios de característica. Cada comparación puede producir una puntuación. Cuando la puntuación supera uno correspondiente de los valores umbral, la puntuación puede ponderarse y las puntuaciones ponderadas para las comparaciones pueden sumarse. La suma puede producir el resultado definitivo para determinar si la imagen candidata de fax es un fax legítimo o un fax masivo no solicitado.

De manera significativa, produciendo una decisión dinámica en línea acerca de si una imagen candidata es un fax masivo no solicitado o legítimo, el impacto costoso de una distribución de fax masivo no solicitado puede evitarse debido a que la presente invención es predictiva por naturaleza. Adicionalmente, debido a que en la presente invención, se usan unas representaciones no procesadas por exploración de trama de la imagen de fax para

procesar la determinación de fax masivo no solicitado en lugar de una forma procesada por exploración de trama de la imagen de fax, pueden evitarse los retardos y el consumo excesivo de los recursos de cálculo. Por último, debido a que la presente invención incorpora una base de conocimientos de características con la que puede identificarse un fax masivo no solicitado, la invención puede ser de tipo autoaprendizaje en la que las características de un fax masivo no solicitado identificado pueden usarse para ajustar y refinar los contenidos de la base de conocimientos.

Se expondrán aspectos de la invención en parte en la descripción que sigue, y serán obvios en parte a partir de la descripción, o pueden aprenderse mediante la práctica de la invención. Los aspectos de la invención se llevarán a cabo y se alcanzarán por medio de los elementos y combinaciones que se señalan concretamente en las reivindicaciones adjuntas. Ha de entenderse que tanto la descripción general anterior como la siguiente descripción detallada son sólo a modo de ejemplo y explicativas y no son restrictivas de la invención, tal como se reivindica.

Breve descripción de los dibujos

Los dibujos adjuntos, que se incorporan en y constituyen una parte de la presente memoria descriptiva, ilustran las realizaciones de la invención y junto con la descripción, sirven para explicar los principios de la invención. Las realizaciones que se ilustran en el presente documento se prefieren en la actualidad, entendiéndose, no obstante, que la invención no se limita a las disposiciones y a los instrumentos concretos que se muestran, en las que:

la figura 1 es una ilustración esquemática de un filtro para fax masivo no solicitado que se ha configurado de acuerdo con las disposiciones de la invención;

la figura 2 es un diagrama de flujo que ilustra un proceso para el filtrado de mensaje masivo no solicitado en el filtro para fax masivo no solicitado de la figura 1;

la figura 3 es un diagrama de bloques que ilustra un proceso para someter a prueba un conjunto de imágenes de fax candidato frente a una base de conocimientos de unas características de fax masivo no solicitado conocidas; y,

la figura 4 es un diagrama de bloques que ilustra múltiples configuraciones preferidas para el procesador de transformación de imagen de fax de la figura 1.

Descripción detallada de las realizaciones preferidas

La presente invención es un sistema, procedimiento y aparato de filtro para fax masivo no solicitado. De acuerdo con las disposiciones de la invención, un filtro para fax masivo no solicitado puede procesar un conjunto de imágenes de fax candidato a través de una comparación de unas representaciones de transformación que no se han procesado por exploración de trama del conjunto de imágenes candidato con una base de conocimientos de unas características de fax conocidas. Cuando la transformación del conjunto de imágenes de fax candidato muestra de forma suficiente las características de fax almacenado en la base de conocimientos para indicar que el fax es un mensaje masivo no solicitado, el fax candidato puede declararse como un mensaje masivo no solicitado y por consiguiente puede manejarse el fax candidato. Además, la base de conocimientos puede ajustarse con los puntos de datos adicionales que se prevén por el conjunto de imágenes candidato. De otro modo, cuando el conjunto de imágenes de fax candidato no muestra las características suficientes de mensaje masivo no solicitado almacenado en la base de conocimientos, el fax candidato puede declararse legítimo y puede procesarse por consiguiente el fax candidato.

La figura 1 es una ilustración esquemática de un filtro para fax masivo no solicitado 120 que se ha configurado de acuerdo con las disposiciones de la invención. El filtro para fax masivo no solicitado 120 puede acoplarse a un dispositivo de salida de fax 170 y puede incluir tanto un procesador de transformación de fax 130 como un comparador 140. El comparador 140 puede acoplarse a una base de conocimientos de mensaje masivo no solicitado 150 que puede almacenar los conjuntos de datos de unas características de fax conocidas 160. Las características de fax conocidas 160 pueden incluir una o ambas de las características de fax masivo no solicitado y características que no son de fax masivo no solicitado. Las características de fax masivo no solicitado pueden incluir unos datos no procesados por exploración de trama con respecto a las características de mensaje masivo no solicitado, tal como las longitudes de página habituales de una transmisión de fax, las intensidades de color, los patrones comunes dentro de la transmisión de fax, y similares. De forma similar, las características que no son de fax masivo no solicitado pueden incluir unos datos no procesados por exploración de trama con respecto a las características de faxes legítimos, tal como las longitudes de página habituales de una transmisión de fax, las intensidades de color, los patrones comunes dentro de la transmisión de fax, y similares.

Volviendo a continuación a la figura 1, una imagen procesada por exploración de trama de fax 110 puede recibirse en el filtro para fax masivo no solicitado 120. Por procesada por exploración de trama, el experto ha de entender que el conjunto de imágenes de fax puede incluir una representación de exploración de trama que se almacena normalmente en uno de varios formatos normalizados, por ejemplo el formato TIFF, el formato JBIG y el formato JPEG, por nombrar unos pocos. La imagen de fax 110 puede convertirse en el procesador de transformación de fax 130 en una o más representaciones no procesadas por exploración de trama de la imagen de fax 110, a las que se hace referencia en el presente documento como la transformación 180, tal como un conjunto de vectores que definen la imagen de fax 110. De forma alternativa, el procesador de transformación de fax 130 puede reducir la forma procesada por exploración de trama de la imagen de fax 110 a una representación numérica codificada o a un

hash robusto numérico. De forma todavía más alternativa, el procesador de transformación de fax 130 puede convertir la forma procesada por exploración de trama de la imagen de fax 110 a una forma de texto a través de un proceso de reconocimiento óptico de caracteres.

En cualquier caso, una vez que el procesador de transformación de fax 130 ha producido una o más de las transformaciones no procesadas por exploración de trama 180 de la imagen de fax, las transformaciones producidas 180 pueden procesarse en el comparador 140. En concreto, el comparador 140 puede comparar una de las transformaciones 180 con los conjuntos de las características conocidas de mensaje masivo no solicitado 160 almacenado en la base de conocimientos 150. Cuando la imagen de fax transformada 180 muestra uno o más de los conjuntos de las características de mensaje masivo no solicitado 160 en la base de conocimientos 150, la imagen de fax 110 puede identificarse como un caso de fax masivo no solicitado y por consiguiente manejarse, por ejemplo descartarse, ponerse en cuarentena, etc. De otro modo, el filtro para fax masivo no solicitado 120 puede permitir el procesamiento de salida de fax adicional de la imagen de fax 110 a través de un dispositivo de salida de fax 170 tal como una impresora de fax, o un encaminador de fax. Es importante que cuando se identifica la imagen de fax 110 como un mensaje masivo no solicitado, las características de la imagen de fax 110 pueden tenerse en cuenta en la base de conocimientos 150 para refinar y ajustar adicionalmente los conjuntos de las características conocidas de mensaje masivo no solicitado 160.

En una ilustración más concreta, la figura 2 es un diagrama de flujo que ilustra un proceso para el filtrado de imágenes de telefax masivas no solicitadas en el filtro para fax masivo no solicitado de la figura 1. Comenzando en el bloque 210, la imagen candidata procesada por exploración de trama de fax puede recibirse para su procesamiento. En el bloque 215, la imagen candidata puede transformarse a una representación no procesada por exploración de trama. El proceso de transformación del bloque 215 no necesita limitarse a ningún tipo de reducción. A este respecto, la figura 4 es un diagrama de bloques que ilustra múltiples configuraciones preferidas para el procesador de transformación de imagen de fax de la figura 1.

Tal como se muestra en la figura 4, el proceso de transformación 410 puede producir una forma no procesada por exploración de trama de la imagen de fax a través del uso de al menos una de cuatro metodologías que se ilustran a pesar de que las metodologías que se ilustran de ninguna forma representan el universo exclusivo de metodologías para producir una forma no procesada por exploración de trama reducida de una imagen de fax. Más bien, el experto en la técnica del procesamiento de imagen reconocerá que existen varias técnicas para convertir una imagen de exploración de trama generalizada a una imagen que no es de exploración de trama. En la figura 4, se muestran cuatro metodologías a modo de ejemplo de este tipo.

Tal como se muestra en el bloque 420, un filtro de perfilado de bordes puede aplicarse a la imagen procesada por exploración de trama para producir una imagen intermedia que tiene unos bordes resaltados que definen la imagen de fax original. Los vectores pueden producirse para los bordes y el conjunto de vectores resultante puede almacenarse como una representación única de la imagen procesada por exploración de trama. De acuerdo con la presente invención en el bloque 430 las intensidades de color relativas de la imagen procesada por exploración de trama pueden calcularse para codificar de forma numérica la imagen procesada por exploración de trama con un único vector de intensidad. Por ejemplo, el número de píxeles blancos y negros puede enumerarse en la imagen de fax, o la proporción de los píxeles blancos respecto de los negros en una zona de la imagen procesada por exploración de trama puede calcularse para producir una única proporción de intensidad para la zona. La imagen procesada por exploración de trama, en sí misma, puede subdividirse en una o más zonas, observando el experto que si tiene más zonas puede producir una más alta probabilidad de generar una única proporción de intensidad, a la vez que se aumenta de forma correspondiente el procesamiento que se requiere para calcular las intensidades relativas para las numerosas zonas.

Como una tercera alternativa adicional, en el bloque 440, la imagen procesada por exploración de trama puede difuminarse pasando la imagen procesada por exploración de trama a través de un filtro de difuminado. Posteriormente, un hash robusto tal como el hash MD-5 que se conoce bien puede aplicarse a la imagen difuminada para producir un único valor de hash para la imagen. Por último, en el bloque 450, la imagen puede procesarse a través de OCR para generar un conjunto de palabras y de símbolos de texto. Los símbolos pueden procesarse previamente para las correlaciones entre símbolos tal como a través de un análisis bayesiano. En un caso de este tipo, cuando se compara la representación de texto de la imagen con otra imagen transformada de forma similar, puede suponerse una correspondencia en la que unas correlaciones idénticas subsisten en ambas imágenes transformadas.

Volviendo a continuación a la figura 2, una vez que la imagen candidata de fax se ha transformado a una forma no procesada por exploración de trama, en el bloque 220 puede recuperarse un primer conjunto de las características conocidas de mensaje masivo no solicitado, tal como unas intensidades o patrones de color habituales, vectores de imagen habituales dentro de una imagen de fax masivo no solicitado, una cierta apariencia de palabras o de combinaciones de palabras, un número habitual de páginas en una transmisión de fax, etc. En el bloque de decisión 225, la representación transformada puede compararse con el primer conjunto de las características de mensaje masivo no solicitado para determinar si la imagen candidata muestra las características definidas por el primer conjunto. Si es así, en el bloque 240 puede ponderarse el resultado para indicar una importancia relativa del

conjunto. En el bloque 245, puede sumarse una puntuación total de suma con la puntuación ponderada para el conjunto y el proceso puede continuar a través de bloque de decisión 230.

En el bloque de decisión 230, si más conjuntos de las características de mensaje masivo no solicitado siguen teniendo que someterse a prueba en la imagen candidata, en el bloque 235 el siguiente conjunto de las características de mensaje masivo no solicitado puede recuperarse y el proceso puede repetirse en los bloques entre el 225 y el 245. Una vez que no siguen teniendo que aplicarse más conjuntos de las características de mensaje masivo no solicitado, en el bloque de decisión 250 puede determinarse si la puntuación total de suma supera un valor umbral. Si no es así, en el bloque 265 la imagen candidata puede procesarse como una transmisión de fax que puede permitirse. De otro modo, si la puntuación total de suma supera el valor umbral, la imagen candidata puede etiquetarse como un fax masivo no solicitado y descartarse en el bloque 255. Adicionalmente, en el bloque 260 el conjunto de las características de mensaje masivo no solicitado en la base de conocimientos puede ajustarse de nuevo con los datos de imagen no procesada por exploración de trama presentes en el fax candidato.

Es importante que la determinación de si una imagen candidata es un mensaje masivo no solicitado puede realizarse a través de una ponderación configurable por el usuario de unas características de mensaje masivo no solicitado diferentes tal como se define dentro de la base de conocimientos. Con este fin, la figura 3 es un diagrama de bloques que ilustra el proceso para identificar un mensaje masivo no solicitado probable. Tal como se muestra en la figura 3, una transformación no procesada por exploración de trama 310 de una imagen candidata de fax puede procesarse de forma concurrente en una o más pruebas de mensaje masivo no solicitado 320. Cada prueba de mensaje masivo no solicitado individual 320 puede aplicar un conjunto de las características de mensaje masivo no solicitado a la transformación no procesada por exploración de trama para determinar si la transformación no procesada por exploración de trama muestra la característica de prueba dentro de un valor umbral configurable por el usuario. A modo de ejemplo, una prueba a modo de ejemplo puede incluir el número de páginas dentro de la transmisión de fax. Cuando la cuenta de las páginas supera las cinco páginas, puede concluirse que el fax candidato cumple los criterios para la prueba a modo de ejemplo.

El experto reconocerá que no todas las pruebas de mensaje masivo no solicitado 320 se considerarán de igual manera en la evaluación de la suma total de si una imagen candidata de fax es un fax masivo no solicitado. Por consiguiente, cada uno de los resultados 330 puede ponderarse 340 de acuerdo con la importancia de la prueba de mensaje masivo no solicitado 320. La suma ponderada 350 puede evaluarse frente a un umbral 350 para determinar si la imagen candidata es un fax masivo no solicitado o un fax legítimo.

La presente invención puede llevarse a cabo en hardware, en software, o en una combinación de hardware y de software. Una realización del procedimiento y sistema de la presente invención puede llevarse a cabo de una forma centralizada en un sistema informático, o de una forma distribuida en la que elementos diferentes se dispersan a lo largo de varios sistemas informáticos conectados entre sí. Cualquier tipo de sistema informático, u otro aparato adaptado para llevar a cabo los procedimientos que se describen en el presente documento, es adecuado para realizar las funciones que se describen en el presente documento.

Una combinación habitual de hardware y de software puede ser un sistema informático de propósito general con un programa informático que, cuando está cargándose y ejecutándose, controla el sistema informático de tal modo que lleva a cabo los procedimientos que se describen en el presente documento. La presente invención puede también integrarse en un producto de programa informático, que comprende la totalidad de las características que habilitan la realización de los procedimientos que se describen en el presente documento, y que, cuando se carga en un sistema informático es capaz de llevar a cabo estos procedimientos.

Programa informático, o su aplicación en el presente contexto, quiere decir cualquier expresión, en cualquier lenguaje, código o notación, de un conjunto de instrucciones que tengan como objetivo dar lugar a un sistema que tiene una capacidad de procesamiento de información para realizar una función concreta o bien directamente o bien después de una o ambas de las siguientes a) conversión a otro lenguaje, código o notación; b) reproducción en una forma material diferente. De manera significativa, la presente invención puede realizarse en otras formas específicas sin alejarse del alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un filtro para fax masivo no solicitado (120) que comprende:

- 5 un procesador de transformación de fax (130) configurado para transformar unas imágenes de fax procesadas por exploración de trama en unas representaciones no procesadas por exploración de trama (180) de imágenes de fax;
- 10 una base de conocimientos (150) de características de fax que comprende al menos una característica de fax que se selecciona del grupo que consiste en características de mensaje masivo no solicitado y características de mensaje no masivo no solicitado; y,
- un comparador (140) configurado para comparar unas representaciones no procesadas por exploración de trama (180) de imágenes de fax que se producen por dicho procesador de transformación de fax (130) con al menos una de dichas características de fax para detectar un fax masivo no solicitado en el filtro para fax masivo no solicitado (120);
- 15 **caracterizado por que** dicha base de conocimientos (150) incluye características de fax que se obtienen de la intensidad de imagen de fax.

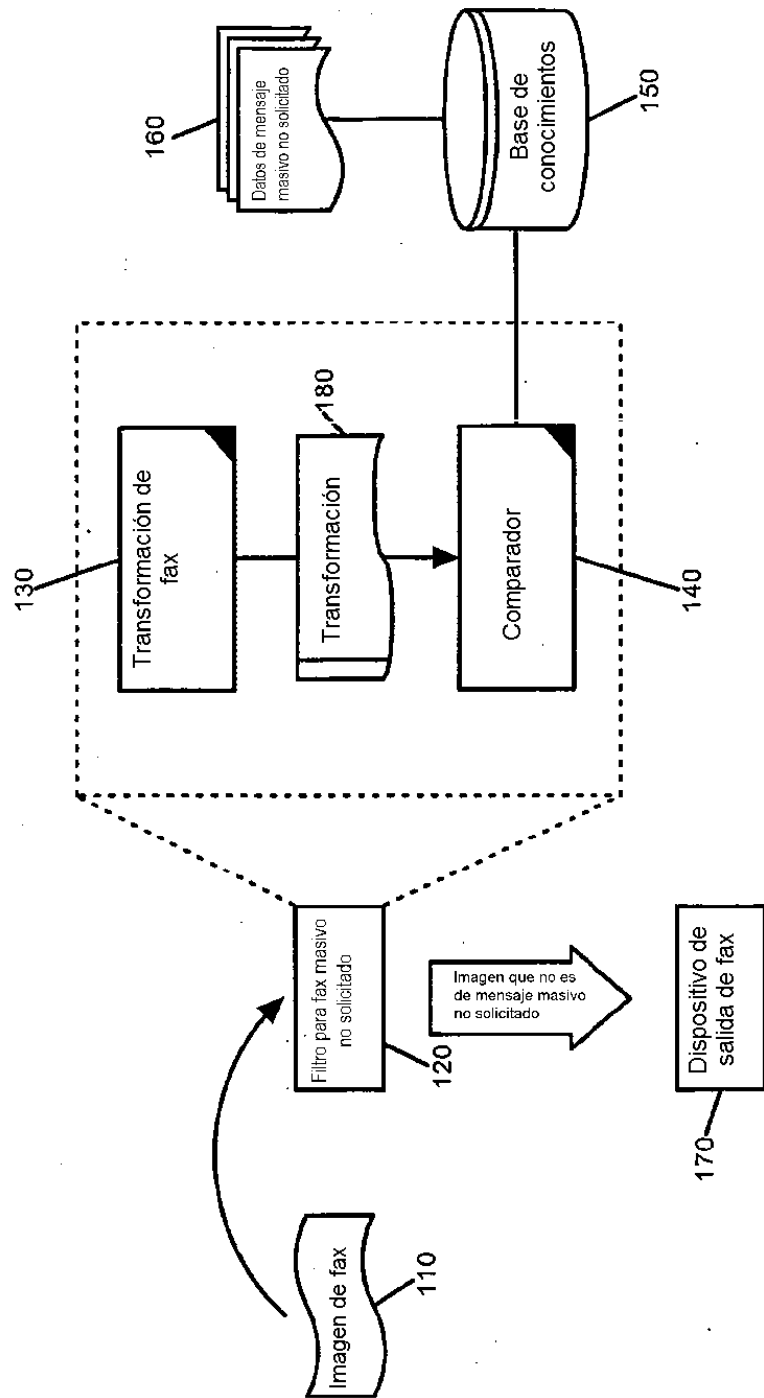


FIG. 1

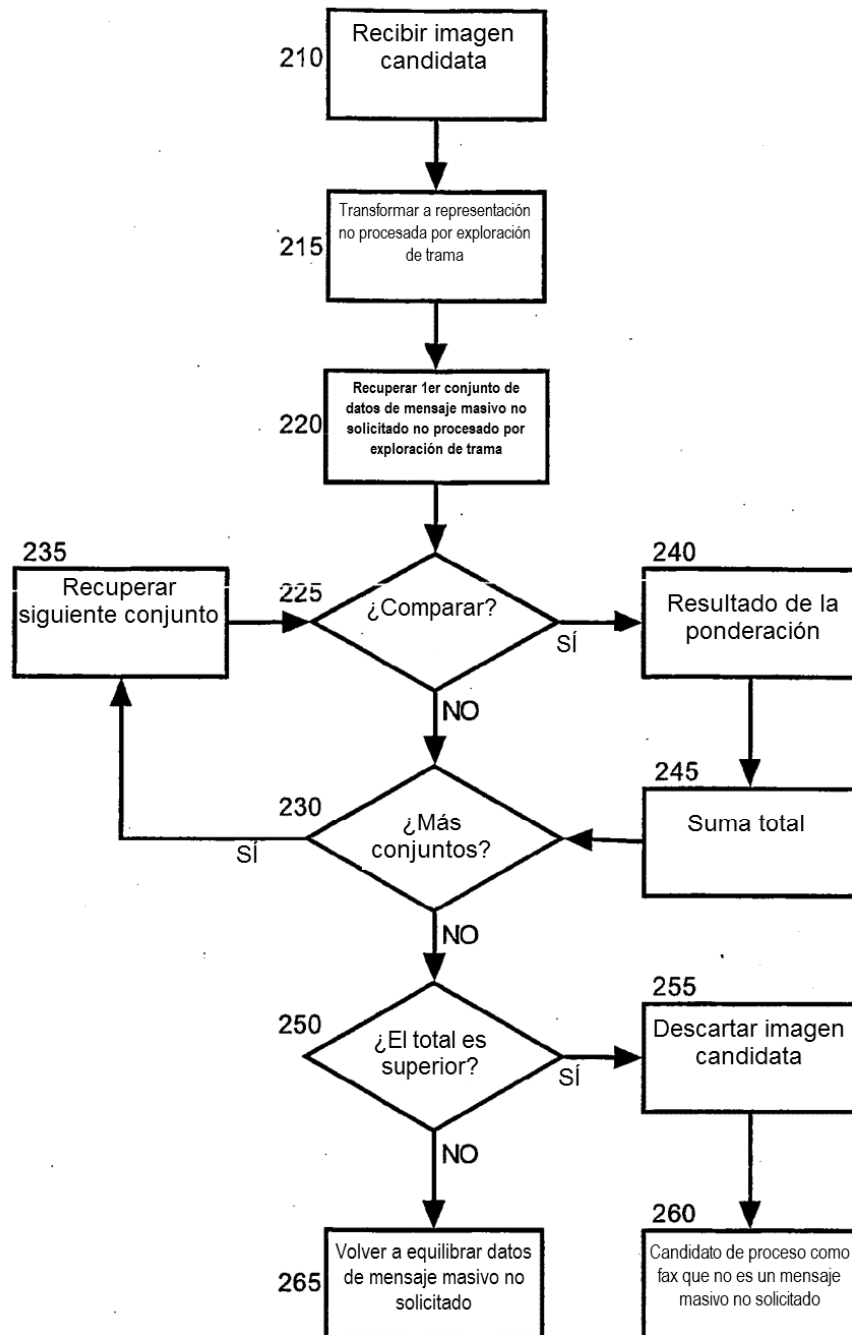


FIG. 2

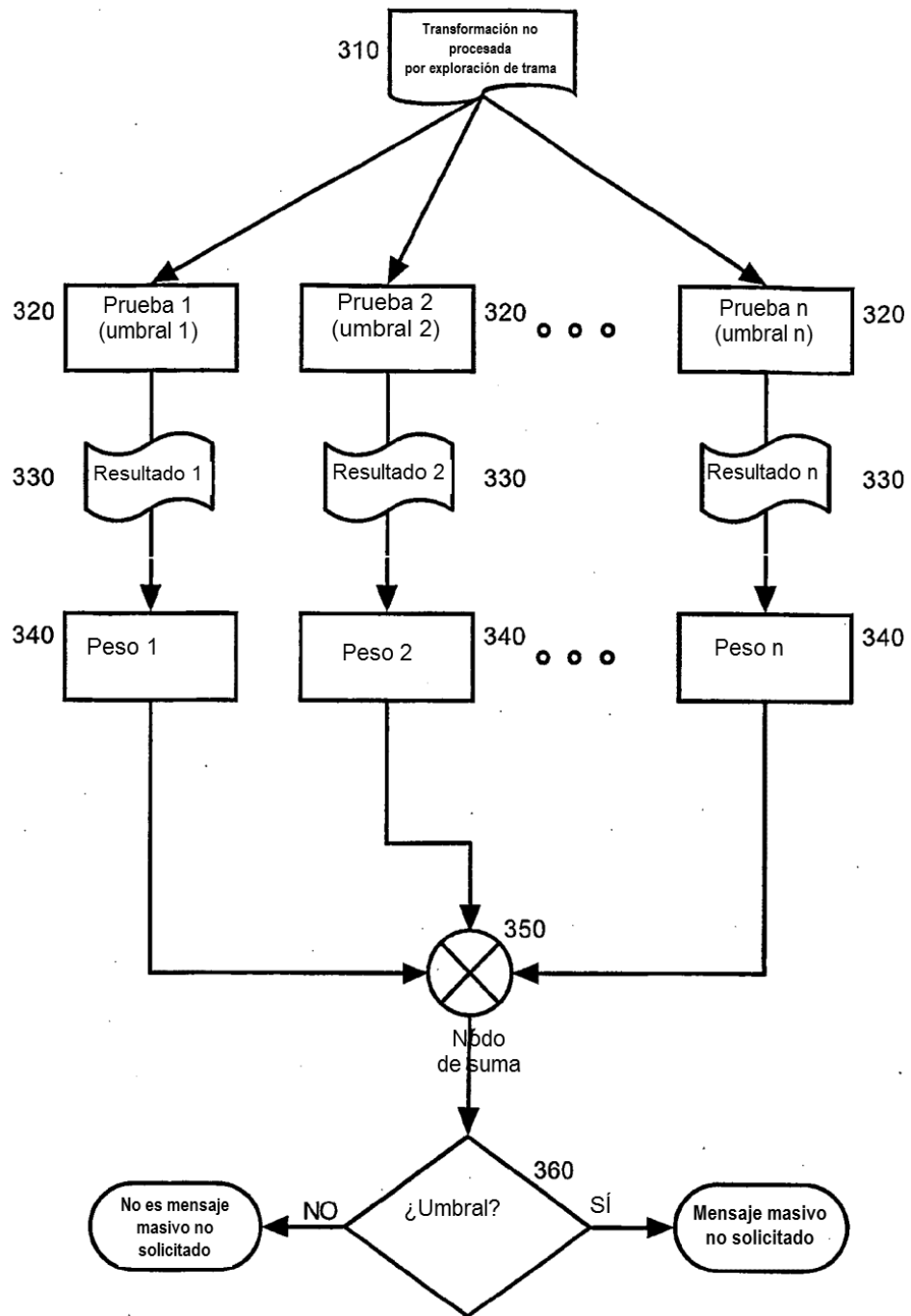


FIG. 3

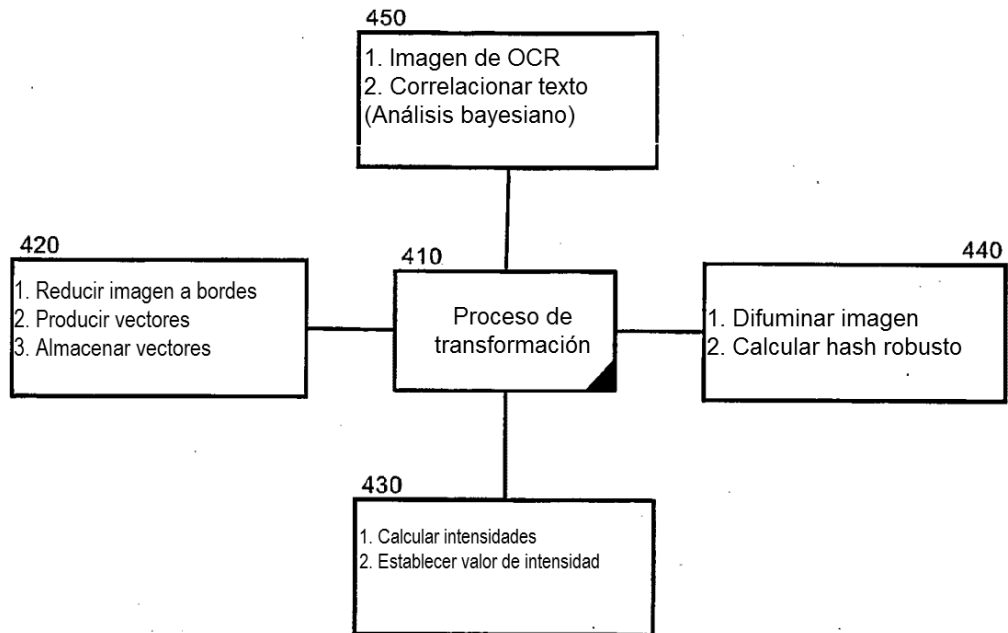


FIG. 4