(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 327 853**

(51) Int. Cl.:
G06K 9/20 (2006.01)
G06K 9/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **05300482 .6**
(96) Fecha de presentación : **16.06.2005**
(97) Número de publicación de la solicitud: **1622065**
(97) Fecha de publicación de la solicitud: **01.02.2006**

(54) Título: **Procedimiento para la detección de series continuas.**

(30) Prioridad: **20.07.2004 FR 04 51591**

(45) Fecha de publicación de la mención BOPI:
04.11.2009

(45) Fecha de la publicación del folleto de la patente:
04.11.2009

(73) Titular/es: **Solystic**
14, avenue Raspail
94257 Gentilly Cédex, FR

(72) Inventor/es: **Letombe, Mathieu y**
Benyoub, Belkacem

(74) Agente: **Elzaburu Márquez, Alberto**

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).

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Procedimiento para la detección de series continuas.

La invención se refiere a un procedimiento de tratamiento de envíos postales por el que se genera una imagen digital de cada envío postal que comprenda una información de dirección, y se realiza, a partir de la imagen del envío postal, una operación de reconocimiento automático de dirección por OCR (reconocimiento de caracteres merced a medios ópticos).

Más concretamente, la invención se refiere a un procedimiento de tratamiento de envíos postales que permite detectar series continuas. Una serie continua es, típicamente, una sucesión o un grupo de envíos postales homogéneos, con frecuencia en número muy grande, que presenta características físicas visibles idénticas. Los envíos postales de una serie continua provienen de un mismo emisor de correo y se dirigen a destinatarios diferentes. Estos envíos postales se suelen denominar "envíos de gran usuario" o también "de envíos de cliente con referencia".

En la figura 1 se ha ilustrado, de manera muy esquemática, los envíos postales E de una serie continua F. En la figura 1 se muestran, de manera más precisa, dos envíos postales E, de los que puede verse la cara delantera, que comprende informaciones de dirección. Los dos envíos postales E comprenden, de la misma manera, arriba a la izquierda, el nombre y la dirección del expedidor, indicados mediante la referencia x, abajo a la izquierda el nombre del grupo (indicado mediante la referencia y) al que pertenece el expedidor, y, en el centro, ligeramente desplazado a la derecha, un bloque de dirección, indicado mediante la referencia z, estando dispuesto este bloque de dirección, en este caso, en un recuadro cubierto de plástico. Todos los envíos postales E de esta serie continua F están previstos de la misma manera y comprenden, cada uno, los elementos x, y, z característicos de la serie continua F, situados en las mismas posiciones, en la cara delantera de los envíos. Únicamente la dirección del destinatario, en el bloque z del envío postal E, cambia de un envío postal a otro de la serie continua F, como puede verse en la figura 1.

Actualmente, las series continuas son identificadas visualmente por los operarios cuando alimentan las máquinas de clasificación con envíos postales.

Más concretamente, cuando un operario haya identificado una serie continua, puede ajustar manualmente los parámetros del proceso automático de OCR de la dirección del destinatario en la máquina de clasificación. Pero, con frecuencia, esta parametrización no es óptima y no se utiliza, pues requiere cierta competencia en la utilización del OCR que, en general, no posee el operario. Cuando el operario tiene la competencia necesaria para esta parametrización manual, ésta se utiliza, generalmente, sólo con series continuas muy grandes, es decir series continuas compuestas por miles de envíos postales, pues su puesta en práctica requiere operaciones de medición, optimización y toma de datos que ralentizan el funcionamiento global de la máquina de clasificación. El ajuste de los parámetros de OCR puede realizarse, por ejemplo, de la manera propuesta por el documento de patente europea EP 1 159 705.

El documento DE-198 36 767 describe un procedimiento de tratamiento de envíos postales por el que se utiliza un proceso de detección de series continuas. Este proceso se utiliza para devolver, automáticamente, correo en gran cantidad a un expedidor.

Generalmente, los envíos postales que constituyen series continuas son enviados a las agencias postales por clientes dotados de referencia que disfrutan de contratos de "bonificación". Los contratos de bonificación prevén una reducción de la tarifa de expedición en función de la tasas de lectura de encaminamiento y distribución de direcciones por OCR. Generalmente, las tasas de lectura se determinan por medio de pruebas de lotes de envíos postales en el momento de la firma del contrato y no se verifican mientras dure el mismo, porque la determinación de las tasas de lectura lleva mucho tiempo y, por tanto, es costosa. Por otro lado, las agencias postales no tienen modo de controlar el respeto a los comportamientos especificados en el contrato por parte de los clientes dotados de referencia.

El objeto de la presente invención consiste en ofrecer un procedimiento para detectar automáticamente series continuas, con el fin de permitir una parametrización automática del lector de dirección por OCR y un control automático de los contratos de bonificación.

Con este fin, la invención tiene por objeto un procedimiento de tratamiento de envíos postales tal como se define mediante la reivindicación 1.

De acuerdo con un modo de puesta en práctica particular del procedimiento de la invención, el tratamiento para la identificación de series continuas comprende una etapa que consiste en obtener componentes conexas a partir de la imagen del envío postal.

Estas componentes conexas ya se extraen para realizar el reconocimiento de dirección por OCR. Mediante el procedimiento de la invención se utilizan estas mismas informaciones para detectar series continuas y añadir así, de manera sencilla, una nueva funcionalidad a un proceso de clasificación automática de envíos postales.

Merced al procedimiento de la invención, el sistema de reconocimiento automático de dirección por OCR puede regularse automáticamente por cada nueva identificación de serie continua. De la misma manera, merced a la identificación automática de serie continua de acuerdo con la invención, resulta posible organizar una determinación siste-

mática y automática de la tasa de lectura de series continuas identificadas para la gestión de contratos de bonificación por parte de las agencias postales.

De acuerdo con otro modo de puesta en práctica del procedimiento de la invención, las series continuas se incluyen en un catálogo en forma de conjunto de componentes conexas predeterminadas para cada serie continua. De ese modo, al tratar un envío postal en curso, se comparan las componentes conexas obtenidas a partir de la imagen digital del envío postal en curso con las componentes conexas predeterminadas de las series continuas del catálogo, con el fin de detectar si el envío postal en curso forma parte de una de las series continuas.

De acuerdo con otro modo de puesta en práctica del procedimiento de la invención, el tratamiento para la identificación de series continuas comprende una etapa que consiste en obtener atributos de imagen multinivel de gris (MNG) a partir de la imagen del envío postal, y las series continuas se incluyen en un catálogo a modo de atributos de imagen multinivel de gris predeterminados por cada serie continua.

De acuerdo con otro modo de puesta en práctica del procedimiento de la invención, el catálogo contiene parámetros de regulación del OCR asociados con las series continuas, mejorándose estos parámetros de regulación de manera automática.

De acuerdo con otro modo de puesta en práctica del procedimiento de la invención, se identifica que un envío postal en curso forma parte de una serie continua y se registra en una base de datos la imagen digital del mismo mediante los resultados de la operación de reconocimiento de dirección por OCR, con el fin de efectuar un control de la tasa de lectura de la serie continua.

La presente invención incluye, también, una máquina de clasificación postal prevista, especialmente, para llevar a la práctica el procedimiento definido en lo que antecede.

La invención se comprenderá mejor a partir de la lectura de la descripción que sigue y del examen de las figuras que la acompañan. Esta descripción se ofrece sólo a título de ejemplo indicativo de la invención, sin carácter limitativo alguno.

La figura 1 consiste en una representación de dos envíos postales que pertenecen a una serie continua.

La figura 2 es un cuadro sinóptico, muy general, que muestra una detección de serie continua durante el proceso de reconocimiento de dirección por OCR.

La figura 3 es un cuadro sinóptico del procedimiento para detectar una serie continua de acuerdo con la invención.

La figura 4 muestra, de manera más detallada, el proceso de comparación de las componentes conexas de la figura 3.

La figura 5 muestra, con más detalle, el catálogo de la figura 3.

La figura 6 es un cuadro sinóptico detallado de las diferentes etapas que permiten la decisión de serie continua y la determinación de acciones a realizar.

La figura 7 es un cuadro sinóptico, muy general, que muestra otro modo de puesta en práctica de la detección de serie continua.

La figura 8 es una representación esquemática de la obtención de atributos a partir de una imagen multinivel de gris.

En la figura 2 se ha representado, de manera muy esquemática, un modo de puesta en práctica del procedimiento de la invención para detectar o identificar, automáticamente, series continuas durante un proceso de reconocimiento automático de dirección por OCR.

El proceso de tratamiento automático de un envío postal en una máquina de clasificación comprende un tratamiento de reconocimiento automático de dirección por OCR, por el que se genera una imagen digital 1 de la superficie del envío postal en curso, que comprende una información de dirección. Se binariza 2 la imagen digital del envío postal y se obtienen 3, a partir de la imagen digital binarizada, componentes conexas (CC) cuya composición se estudiará de manera más detallada en lo que sigue. Las componentes conexas CC del envío postal se utilizan, entonces, para localizar el bloque de dirección, las líneas y los caracteres de la dirección postal y para el reconocimiento automático de dirección 4 por OCR.

El procedimiento de la invención utiliza también las componentes conexas obtenidas 3 a modo de datos destinados a distinguir, durante la etapa 5, un grupo de envíos postales homogéneos o serie continua. Como se ha representado en la figura 2, la detección de serie continua 5 puede realizarse en paralelo al reconocimiento automático de dirección por OCR, de manera que el proceso de tratamiento se repita, L, por cada envío postal subsiguiente, con el resultado de la detección de serie continua. Las informaciones obtenidas a partir de la detección de serie continua y del reconocimiento

de dirección por OCR del envío postal en curso se utilizan durante el proceso de tratamiento por OCR del envío postal subsiguiente y la detección de serie continua asociada.

En la figura 3 se ha representado la detección de serie continua 5 de la figura 2 de manera más detallada, a partir de la etapa 6. En la figura 3, se ha representado, igualmente, la generación 1 de la imagen digital de la superficie del envío postal en curso, la binarización 2 y la obtención de las componentes conexas 3 a partir de esta imagen digital. Durante la etapa 6, se comparan las componentes conexas del envío postal en curso con las componentes conexas de modelos registrados previamente e incluidos en un catálogo dinámico 7 registrado en memoria. Se denomina modelo a un conjunto de componentes conexas predeterminadas que definan (caractericen) una serie continua. El proceso de comparación 6 de las componentes conexas se detallará más en lo que sigue con referencia a la figura 4.

Si durante la etapa 6 se encuentra que las componentes conexas del envío postal en curso concuerdan 8 con las componentes conexas de un modelo del catálogo, se enriquece 9 el modelo correspondiente con las componentes conexas del envío postal en curso, con el fin de consolidar la robustez del modelo. En cambio, si se encuentra que las componentes conexas del envío postal en curso no concuerdan 10 con las componentes conexas de ninguno de los modelos del catálogo 7, se genera un nuevo modelo 11 en el catálogo 7, por ejemplo, que reemplace un modelo antiguo, de manera que las componentes conexas del envío postal en curso constituyan las componentes conexas del nuevo modelo. Se puede prever, fácilmente, asociar con cada modelo del catálogo un número de identificación de modelo que se conserve en una memoria 12 del OCR. Igualmente, las sanciones de OCR determinadas durante la etapa 4 podrán registrarse en la memoria 12 del OCR en relación con el número de identificación de un modelo de serie continua.

De acuerdo con el procedimiento de la invención, un módulo de decisión del OCR analiza, de manera recurrente, la memoria 12 del OCR para detectar una serie continua y decide 13 la ejecución de funciones que tengan su origen en la detección de series continuas.

Más particularmente, una vez que el módulo de decisión haya decidido 13 que el envío postal en curso pertenece a una serie continua en curso, puede aplicar acciones 14 destinadas a optimizar la lectura automática de dirección por OCR. Igualmente, si se detecta que un envío postal en curso forma parte de una serie continua, el módulo 13 puede ordenar 15 el registro de informaciones, en una base de datos 16, relacionadas con el envío postal en curso, y, eventualmente, con los envíos postales precedentes que pertenezcan, también, a esta serie continua en curso, con fines de control de bonificaciones a grandes usuarios.

Un ejemplo de puesta en práctica del procedimiento de la invención toma como base la utilización de las componentes conexas a modo de atributos de imagen destinados a distinguir un grupo de envíos postales homogéneos, con el fin de caracterizar series continuas. De manera en sí conocida, una componente conexa de una imagen binaria está constituida por todos los puntos negros adyacentes que se puedan asociar por cercanía inmediata sin pasar por un punto blanco. En la figura 1, dos componentes conexas 17, “S” y “Raspail” (en letra cursiva), que concuerdan, respectivamente, con cada uno de los envíos postales E de la serie continua F, se muestran rodeadas, cada una, por un cuadro de línea discontinua 18. Por tanto, la obtención 3 de las componentes conexas a partir de una imagen digital binarizada consiste en la identificación de cada una de las componentes conexas en una tabla de componentes conexas, por distintos criterios, por ejemplo, por la posición en la imagen del cuadro rectangular 18 que rodea a la componente conexa 17, por la longitud y la anchura del cuadro 18, por la densidad de negro en el cuadro 18 y por el número de píxeles conexos.

La figura 4 presenta, de manera más detallada, el proceso de comparación 6 de las componentes conexas de la figura 3 con el fin de determinar si un envío postal en curso puede asociarse con un modelo del catálogo 7. Durante la etapa 3 se obtienen todas las componentes conexas a partir de la imagen del envío postal en curso. De todas estas componentes conexas sólo se conservan las más significativas, por eliminación de las demás durante la etapa 20; por ejemplo, pueden conservarse sólo las cien que presenten más píxeles. La eliminación 20 de las componentes conexas de menor tamaño permite acelerar el proceso de comparación 6, y, también, algunas veces, eliminar la dirección del destinatario, que, en general, es diferente en los distintos envíos postales que pertenezcan a una serie continua. A continuación, se compara 21 cada componente conexa discriminante del envío postal en curso con cada componente conexa discriminante de cada uno de los modelos. Para ello, se procede por cascada de filtros de acuerdo con los diferentes criterios de la tabla de componentes conexas, definiéndose un umbral de compatibilidad por cada uno de los criterios. Por ejemplo, en primer lugar, pueden compararse las posiciones de los cuadros que rodeen a las componentes conexas antes de comparar, en caso de superarse el umbral de compatibilidad seleccionado, las longitudes y anchuras de los cuadros, etc. De ese modo pueden conocerse las componentes conexas, y, por tanto, el número de componentes conexas del envío postal en curso que concuerden con las componentes conexas de cada uno de los modelos del catálogo (etapa 22). Se selecciona 23 el modelo con el que concuerden el número mayor de componentes conexas y se compara este número con un umbral 24, por ejemplo, de 30 componentes conexas concordantes. Si esta última etapa de comparación es concluyente, se habrá detectado que el envío postal en curso forma parte de una serie continua identificada mediante un modelo del catálogo 7.

La figura 5 muestra, con más detalle, un ejemplo de estructura del catálogo 7 ilustrado en la figura 2. El catálogo 7 consiste en un fichero estructurado y comprende, principalmente, cuatro elementos de información por cada modelo 30, a saber: un número de identificación 31, un conjunto de componentes conexas 32, parámetros 33 de optimización para la regulación del OCR y un indicador de interés 34 del modelo.

Por otro lado, el catálogo 7 está formado por dos partes. Una primera parte 35 del catálogo es dinámica, y, por tanto, su contenido evoluciona continuamente. En esta primera parte 35 del catálogo 7 se registra (genera) 11 un nuevo modelo 30 asociado con un envío postal en curso, cuando las componentes conexas del envío postal en curso no concuerden 10 con las componentes conexas de los modelos del catálogo 7. Una vez ocupada esta primera parte 35 del catálogo 7, se puede reemplazar un modelo 30 de esta primera parte 35 con el fin de generar 11 un nuevo modelo 30. El modelo reemplazado puede ser, por ejemplo, el que presente el indicador de interés 34 menor. De acuerdo con el procedimiento de la invención, el indicador de interés 34 de un modelo 30 depende del interés del modelo 30, es decir, del tiempo transcurrido desde su último reconocimiento, del número de veces que haya sido reconocido, si ha generado una decisión de detección de serie continua, de la utilidad de los parámetros de OCR del modelo, etc. El registro de un nuevo modelo 30 en la primera parte 35 consiste en asignar un número de identificación 31 al modelo 30. En correspondencia con este número de identificación 31 se registran, por ejemplo, las cien componentes conexas 32 más significativas del envío postal en curso, los parámetros 33 de OCR utilizados para el reconocimiento automático de dirección del envío postal en curso y el indicador de interés 34.

Por otro lado, cuando un envío postal en curso haya sido reconocido como asociado 8 con un modelo 30 de la primera parte 35 del catálogo 7, los elementos de información del modelo se enriquecen con el envío postal en curso. El número de identificación 31 del modelo 30 no cambia, con el fin de permitir al módulo de decisión reconocer la asociación del envío postal en curso con el modelo 30. El conjunto de componentes conexas 32 del módulo se enriquece 9 con las componentes conexas del envío postal en curso. Los valores de cada uno de los criterios de la tabla de componentes conexas para las componentes conexas del conjunto de componentes conexas 32 del modelo 30 concordantes con las componentes conexas del envío postal en curso son promediados. Las componentes conexas del envío postal en curso que no concuerden con las componentes conexas del modelo se registran en el conjunto de componentes conexas 32 del modelo, hasta que el conjunto de componentes conexas contenga un número máximo de componentes conexas para un modelo 30. Este enriquecimiento 9 de las componentes conexas del modelo 30 permite consolidarlo haciéndolo más fiable y, por ejemplo, más tolerante a los ruidos de las componentes conexas durante la binarización de la imagen digital. Además, en el modelo 30, los parámetros 33 de OCR del envío postal en curso se registran sobre los antiguos y el indicador de interés 34 se actualiza.

En una segunda parte 36 del catálogo, no borrable, se registran, inicialmente, modelos 30 de series continuas de clientes dotados de referencia que hayan sido objeto de acuerdos comerciales, o modelos de series continuas que se repitan con mucha frecuencia y que comprendan un número muy grande de envíos postales. Una vez registrados en esta segunda parte 36 del catálogo 7, los elementos de información de los modelos 30 ya no cambian. El conjunto 32 de las componentes conexas de estos modelos está completo y ya no puede enriquecerse con componentes conexas. Los parámetros 33 de OCR de los envíos postales asociados con los modelos 30 se optimizan perfectamente, típicamente, por muestreo, y no se modifican. Los modelos 30 de la segunda parte 36 del catálogo 7 presentan números de identificación 31 consecutivos. El indicador de interés 34 no es necesario. Además de los modelos 30 registrados inicialmente en la segunda parte 36 del catálogo 7 se registran, 37, modelos 30 provenientes de la primera parte 35 del catálogo 7 cuando éstos presenten parámetros de OCR optimizados perfectamente y se repitan frecuentemente, es decir, cuando sus indicadores de interés 34 hayan superado cierto umbral. Estos modelos 30 son registrados en espacios libres de la segunda parte 36 del catálogo 7 y no reemplazan modelos 30 registrados anteriormente.

La figura 6 muestra, de manera detallada, las etapas del análisis de las informaciones contenidas en la memoria 12 del OCR y en el catálogo 7 realizado por un módulo de decisión 40 de la máquina de tratamiento postal con el fin de detectar 13 una serie continua, decidir la ejecución de funciones que sean consecuencia del paso de una serie continua, y, por tanto, de la detección de una serie continua, y determinar 14 las acciones apropiadas a poner en práctica. El módulo de decisión 40 analiza, durante la etapa 41, a partir de la memoria 12 del OCR, los números de identificación 31 de los modelos 30 asociados con los últimos envíos postales. Si el número de identificación 31 del modelo asociado con el envío postal anterior al envío postal en curso es diferente 42 del número del modelo asociado con el envío postal en curso, no se pone en práctica ninguna acción relativa a una serie continua durante la etapa 43. En caso contrario, 44, un examen del catálogo 7 durante la etapa 45 permite saber si el modelo 30 asociado con el envío postal en curso es un modelo 30 de la segunda parte 36 del catálogo 7 o si este modelo ha sido ya objeto de una serie continua. En caso afirmativo 46, el módulo de decisión decide (detecta), durante la etapa 13, que el envío postal en curso pertenece a una serie continua y activa la ejecución de funciones previstas en caso de detección de una serie continua. En caso negativo, 47, el módulo de decisión 40, durante la etapa 48, cuenta el número de envíos postales consecutivos, incluyendo el envío postal en curso, asociados con el mismo modelo 30. Si este número es inferior o igual, por ejemplo, a cuatro, como en el caso de la salida 49, no se pone en práctica ninguna decisión relativa a una serie continua. Si este número es superior a cuatro, como en el caso de la salida 50, el módulo de decisión decide, durante la etapa 13, que el envío postal en curso pertenece a una serie continua y pone en práctica la ejecución de funciones previstas en caso de detección de una serie continua.

Una función que resulta de la detección de una serie continua consiste en el control de las bonificaciones, que pasa por la transmisión 15 de instrucciones a la base de datos 16, de manera que ésta recupere y registre, en ficheros identificados por ejemplo, mediante fecha-hora-minuto-segundo de registro, entre otras cosas, el número de identificación 31 del modelo 30 asociado con los envíos postales de la serie continua, las sanciones de OCR de todos los envíos postales de la serie continua y las imágenes de algunos envíos postales de la serie continua, por ejemplo, de uno por cada cincuenta.

Otra función que resulta de la detección de una serie continua consiste en la optimización automática de los parámetros del OCR para la lectura de la dirección de los envíos postales subsiguientes al envío postal en curso. Los parámetros 33 de OCR registrados en correspondencia con el número de identificación 31 del modelo asociado con los envíos postales de la serie continua se recuperan y su nivel de optimización se verifica 52. Si estos parámetros de OCR están optimizados perfectamente 53 (es el caso, en general, de las series continuas de los clientes dotados de referencia), se transmiten al OCR cuando se repita el bucle, L, para la lectura de la dirección del envío postal subsiguiente. Si estos parámetros 33 de OCR no están perfectamente optimizados 54 (es el caso, en general, de las series continuas cuyo modelo acabe de generarse en el catálogo), el módulo de decisión 40 analiza, además de los parámetros 33 de OCR del catálogo 7, las sanciones de OCR y los parámetros de OCR de los envíos postales asociados con este modelo contenidos en la memoria del OCR, y determina 14 nuevos parámetros de OCR para mejorar automáticamente el reconocimiento automático de dirección por OCR. Entonces, los nuevos parámetros de OCR son transmitidos al OCR durante la repetición de bucle, L, para la lectura del envío postal subsiguiente. Probablemente, estos parámetros de OCR serán mejorados durante numerosos envíos postales antes de permitir un reconocimiento de dirección óptimo.

Más concretamente, un parámetro de OCR que permita acelerar la lectura de OCR es el que re refiere a la distinción entre "escritura manuscrita" y "escritura mecanografiada". Después de haber localizado el bloque de dirección y las líneas, la unidad de reconocimiento de dirección por OCR investiga si la escritura en el bloque de dirección está manuscrita o mecanografiada. Algunas veces la decisión es unívoca a partir de esta etapa y el tratamiento inútil se interrumpe, pero, en ocasiones, la decisión entre escritura manuscrita y mecanografiada no se toma hasta el final del tratamiento de OCR, durante la sanción de OCR. De acuerdo con la invención, un análisis de las sanciones de OCR del envío postal en curso y de los envíos postales anteriores de la serie continua en curso puede dar lugar a detectar que la escritura de la serie continua en curso está mecanografiada y, en este caso, ya no es necesario efectuar el reconocimiento de escritura manuscrita, lo que evita errores de lectura y rechazos.

Por otro lado, si el tratamiento detecta que los destinatarios de todos los envíos postales de la serie continua en curso son los mismos, lo que, típicamente, es un caso de error de localización del bloque de dirección por parte del OCR (probablemente, el OCR lee un bloque que contenga la dirección del expedidor de la serie continua), merced a la invención, resulta posible poner fin a este error al impedir al OCR leer el bloque de dirección erróneo y al localizar otro bloque de dirección en la imagen. Las sanciones de OCR de los envíos postales que pertenezcan a la serie continua en curso tratados anteriormente pueden ser reconsideradas y los envíos postales pueden ser redirigidos, en caso de que la máquina de clasificación esté dotada de una línea con retardo que permita no clasificar instantáneamente un envío postal una vez conocida su sanción de OCR.

La optimización de los parámetros de OCR, también, puede tener relación con:

- la posición del bloque de dirección cuando no haya sido leída dirección alguna;

- los parámetros de binarización;

- la orientación de la dirección en el envío postal;

- la fuente tipográfica utilizada;

- la ordenación de los elementos de la dirección (código postal, ciudad, calle);

- el modo de tratamiento a utilizar, por ejemplo, nacional o internacional.

El procedimiento de la invención descrito en lo que antecede para la optimización de los parámetros de OCR se apoya en la presunción de que el envío postal subsiguiente pertenece a la serie continua en curso. Es posible, también, optimizar los parámetros de OCR para la lectura de la dirección del envío postal en curso realizando un tratamiento de OCR con dos pasos de reconocimiento de dirección por OCR.

Puede conseguirse, también, una optimización de los parámetros de OCR de la manera descrita por el documento de patente europea EP 1159705 mediante codificación de vídeo. En ese caso, los parámetros de OCR optimizados se registran en el modelo correspondiente del catálogo con vistas a una utilización ulterior eventual.

Las informaciones contenidas en la base de datos 12 sirven para determinar las tasas de lectura de envíos postales de clientes dotados de referencia, con el fin de controlar el respeto a las clausulas establecidas en contratos de bonificación. Esta operación se efectúa, normalmente, de manera independiente del tratamiento de los envíos postales, por ejemplo, mediante un ordenador, a partir de las bases de datos 12 mantenidas en memoria en distintas máquinas de clasificación, cuyas bases de datos 12 pueden contener datos que identifiquen la máquina de clasificación y la desapiladora por las que sean hechas pasar las series continuas. La puesta en correspondencia entre clientes dotados de referencia y series continuas detectadas puede ser directa si se ha registrado inicialmente un modelo de estas series continuas en la segunda parte del catálogo. En caso contrario, la determinación del expedidor puede ser realizada por un operario a partir de las imágenes digitales de los envíos postales de las series continuas, almacenadas en la base de datos 12.

De acuerdo con el procedimiento de la invención descrito en lo que antecede, se ha introducido una noción de contexto inter envíos postales y la base de esta detección de series continuas consiste en un módulo que pueda comparar dos imágenes y determinar si concuerdan o no. En lo que antecede, esta comparación se ha realizado sobre la base de componentes conexas a modo de atributos de imagen destinados a distinguir un grupo de envíos postales homogéneos con el fin de detectar series continuas. Pero la invención se extiende a otros tipos de atributos de imagen, como el posicionamiento y el tamaño de los diferentes bloques en la imagen, la densidad de negro en la imagen, la tabla de símbolos resultante de un OCR. Es posible, también, obtener atributos a partir de una imagen multinivel de gris o utilizar las densidades de colores en una imagen digital en color para constituir estos atributos de imagen destinados a distinguir series continuas.

En la figura 7 se ha representado otro modo de puesta en práctica del procedimiento de la invención para detectar automáticamente series continuas, por el que los atributos de imágenes destinados a distinguir series continuas son atributos obtenidos a partir de una imagen multinivel de gris. El proceso de tratamiento automático por el que se genera una imagen digital 1 del envío postal en curso, se binariza 2 esta imagen y se obtienen 3, a partir de la imagen binarizada, componentes conexas utilizadas para el reconocimiento automático de dirección 4 no se modifica. Durante la etapa 60, se trata la imagen digital generada durante la etapa 1 con el fin de obtener atributos de imagen multinivel de gris y se utilizan estos atributos de imagen multinivel de gris para realizar una detección 61 de serie continua. La detección 61 de serie continua a partir de los atributos de imagen multinivel de gris se realiza como la detección de serie continua 5 a partir de las componentes conexas en paralelo con el reconocimiento de dirección por OCR, de manera que el proceso de tratamiento por OCR se repita L por cada envío subsiguiente, con el resultado de la detección de serie continua 60.

La figura 8 detalla la obtención de los atributos de imagen multinivel de gris a partir del envío postal en curso realizada durante la etapa 60. La imagen digital 65 generada durante la etapa 1 es una imagen multinivel de gris del envío postal en curso. En primer lugar, durante la etapa 66, se genera una imagen multinivel de gris de baja resolución 67 del envío postal en curso por submuestreo de la imagen 65 del envío postal en curso. A partir de la imagen de baja resolución 67 se obtienen atributos globales multinivel de gris 68. Más concretamente, los atributos globales multinivel de gris 68 consisten en la altura y la anchura del envío postal, el mínimo, el máximo y la media de la escala de gris de los píxeles, la varianza y la entropía de la imagen. A continuación se aplican distintas mallas a la imagen de baja resolución 67, durante la etapa 69, para obtener imágenes reticuladas, por ejemplo, una primera imagen 70 con estructura de red 4*4 y una segunda imagen 71 con estructura de red 5*4, en la figura 8. A partir de las diferentes imágenes reticuladas 70, 71 se obtienen atributos locales 72 multinivel de gris que, en particular, consisten en el mínimo, el máximo y la media de la escala de gris de los píxeles, así como la varianza y la entropía calculadas por cada bloque 73 de la malla de cada imagen reticulada 70, 71. Los atributos globales 68 con los atributos locales 72 multinivel de gris del envío postal en curso constituyen los atributos de imagen multinivel de gris 74 destinados a la detección 61 de serie continua.

La utilización de los atributos multinivel de gris 74 es sensiblemente idéntica a la de las componentes conexas. En este caso, el conjunto de componentes conexas 32 de un modelo 30 del catálogo 7 se reemplaza por los atributos multinivel de gris 74. La comparación entre los atributos multinivel de gris de un modelo y los atributos multinivel de gris del envío postal en curso se efectúa de la manera siguiente. Se comparan los atributos globales del envío postal en curso con los del modelo y se determina, mediante umbrales, si el envío postal concuerda con el modelo. En caso de concordancia, se comparan los atributos locales, se calculan los coeficientes de correlación entre los atributos locales de todos los bloques, por pares, que se promedian para obtener los coeficientes de correlación medio y mediano de cada malla, y al promediar los coeficientes de correlación de las distintas mallas, se obtiene un único coeficiente de correlación entre envío postal en curso y modelo, que, merced a la definición de umbrales, indique si el modelo y el envío postal en curso están asociados. El enriquecimiento de un modelo para hacerlo más robusto se consigue al promediar los atributos multinivel de gris del modelo con los del envío postal en curso.

REIVINDICACIONES

1. Procedimiento de tratamiento de envíos postales por el que se genera (1) una imagen digital de un envío postal en curso que comprenda una información de dirección, y se realiza, a partir de la imagen del envío postal, una operación de reconocimiento automático de dicha información de dirección por OCR (2, 34), comprendiendo este procedimiento, además, un tratamiento (5) para la identificación automática de grupos de envíos postales homogéneos o series continuas que comprende una etapa consistente en obtener (3), a partir de la imagen del envío postal en curso, atributos de serie continua destinados a constituir una característica distintiva de una serie continua, y en comparar (6) estos atributos de serie continua asociados con el envío en curso con atributos de modelo de serie continua prerregistrados en memoria, en un catálogo, con el fin de detectar una concordancia entre dichos atributos de serie continua asociados con el envío en curso y los atributos de modelo de serie continua asociados con un modelo particular de serie continua, **caracterizado** porque

- en respuesta a una detección de concordancia, se continúa el tratamiento mediante las etapas siguientes:

  - se enriquecen (9) dichos atributos de modelo de serie continua asociados con dicho modelo de serie continua particular mediante dichos atributos de serie continua asociados con el envío en curso;

  - se registran resultados del OCR del envío en curso en correspondencia con dicho modelo de serie continua particular con el fin de generar datos de regulación optimizados del OCR;

  - y se controla (28), en este modelo de serie continua particular, si un valor de contabilización, indicativo del número de veces que se haya detectado una concordancia de envíos consecutivos anteriores al envío en curso con los atributos de modelo de serie continua de este modelo de serie continua particular, es superior a cierto valor de umbral, con el fin de determinar los parámetros de funcionamiento del OCR mediante dichos datos de regulación optimizados asociados con este modelo de serie continua particular;

- y en el caso de que no se detecte concordancia alguna, se registran en memoria, en el catálogo, dichos atributos de serie continua asociados con el envío en curso, a modo de atributos de modelo de serie continua asociados con un nuevo modelo (11) de serie continua.

2. Procedimiento de tratamiento de envíos postales de acuerdo con la reivindicación 1, por el que dichos atributos de serie continua generados a partir de la imagen del envío en curso están constituidos por componentes conexas obtenidas a partir de esta imagen, y por el que se clasifican las componentes conexas asociadas con el envío en curso por número de píxeles, con el fin de seleccionar sólo un número limitado predefinido de componentes conexas discriminantes (20) que presenten el mayor número de píxeles antes de comparar estas componentes conexas discriminantes con componentes conexas asociadas con modelos particulares de serie continua.

3. Procedimiento de acuerdo con la reivindicación 1, por el que para generar dichos atributos de serie continua a partir de la imagen del envío en curso se transforma (66) dicha imagen en una imagen multinivel de gris de baja resolución con el fin de obtener atributos globales, como, por ejemplo, mínimo-máximo-media de los niveles de gris del conjunto de la imagen, y atributos locales, como, por ejemplo, mínimo-máximo-media de los niveles de gris de diversas particiones de la imagen.

4. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 3, por el que los datos de regulación comprenden un dato indicativo del funcionamiento del OCR para el reconocimiento de escritura manuscrita o escritura mecanografiada.

5. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 3, por el que los datos de regulación comprenden datos indicativos de la posición y/o la orientación y/o la fuente tipográfica de la información de dirección en el envío.

6. Procedimiento de acuerdo con cualquiera de las reivindicaciones precedentes, por el que si se identifica que un envío postal en curso forma parte de una serie continua se registra en una base de datos (16) la imagen digital del envío postal en curso mediante los resultados del OCR, con vistas a efectuar un control de la tasa de lectura de la serie continua.

FIG_1

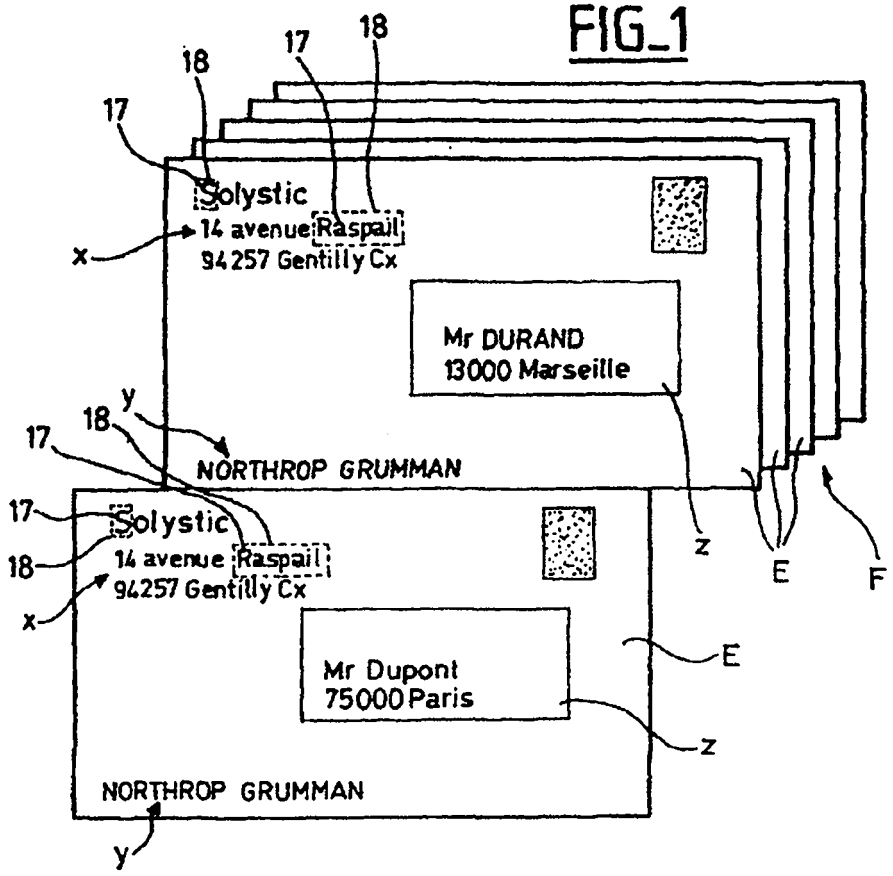
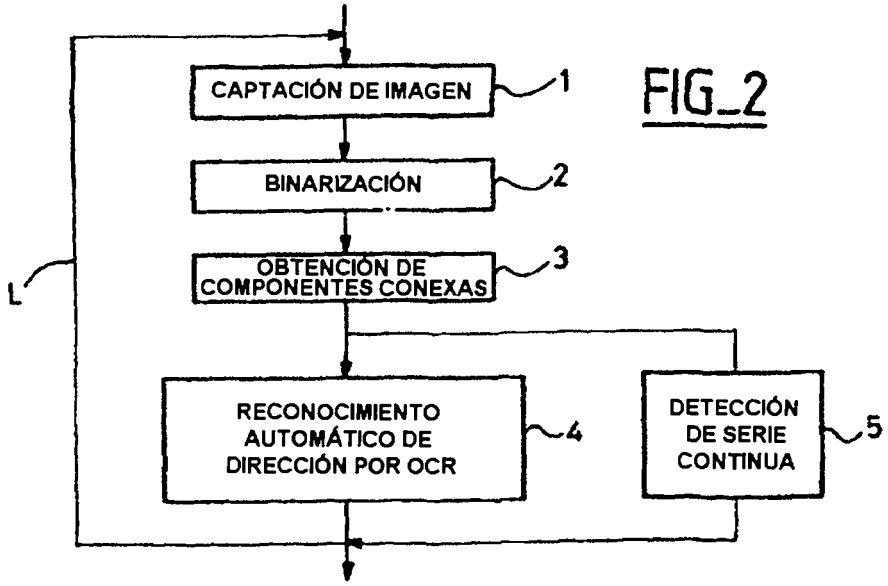

18
17
18
17
Solystic
14 avenue Raspail
94257 Gentilly Cx
x
Mr DURAND
13000 Marseille
y
18
17
NORTHROP GRUMMAN
17
Solystic
18
14 avenue Raspail
94257 Gentilly Cx
x
z
E
F
Mr Dupont
75000 Paris
E
z
NORTHROP GRUMMAN
y

FIG_2


CAPTACIÓN DE IMAGEN
1
BINARIZACIÓN
2
OBTENCIÓN DE COMPONENTES CONEXAS
3
L
RECONOCIMIENTO AUTOMÁTICO DE DIRECCIÓN POR OCR
4
DETECCIÓN DE SERIE CONTINUA
5

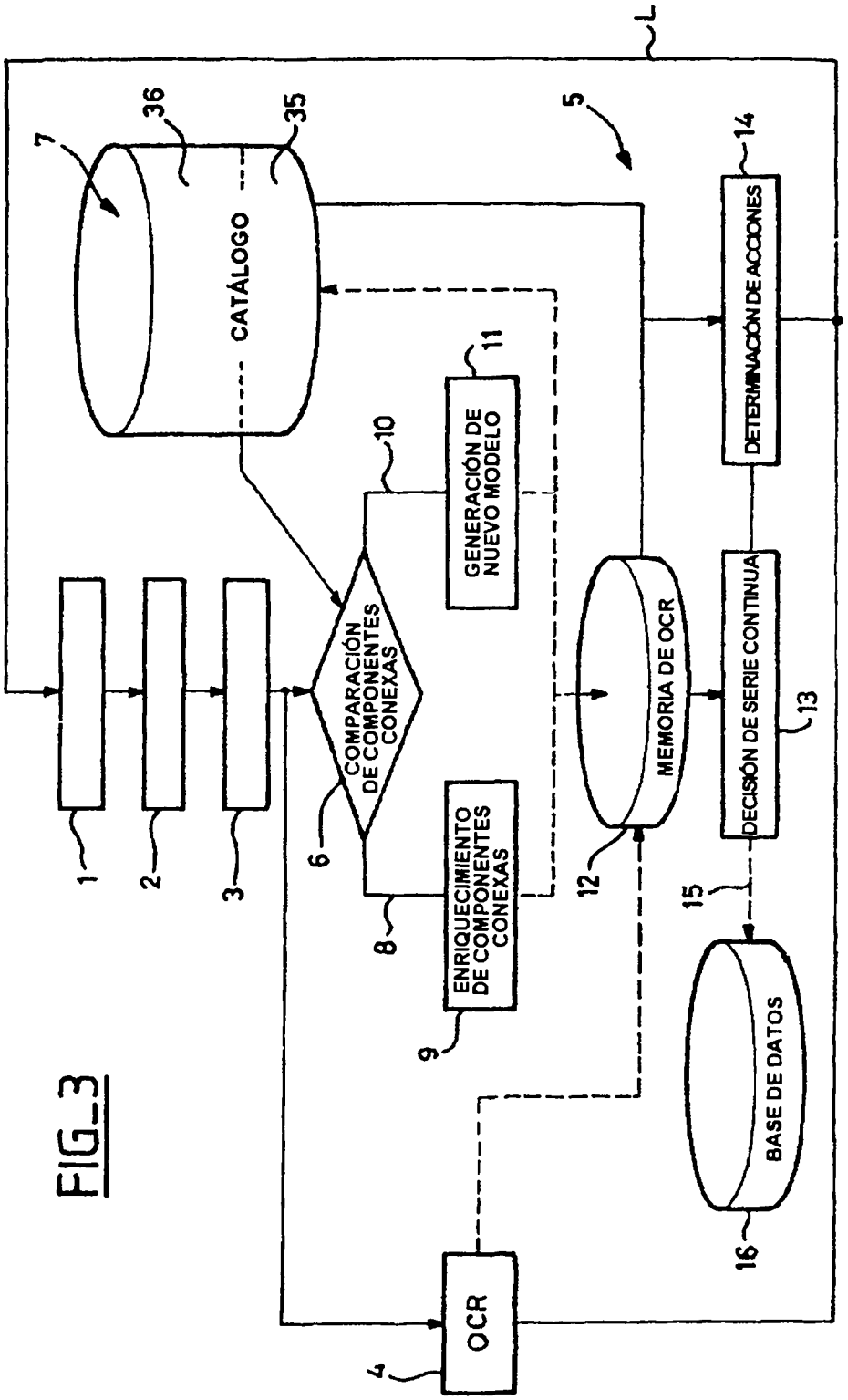

L
5
7
36
35
CATÁLOGO
14
DETERMINACIÓN DE ACCIONES
11
GENERACIÓN DE NUEVO MODELO
10
COMPARACIÓN DE COMPONENTES CONEXAS
MEMORIA DE OCR
DECISIÓN DE SERIE CONTINUA
13
1
2
3
6
12
15
8
ENRIQUECIMIENTO DE COMPONENTES CONEXAS
9
BASE DE DATOS
16
OCR
4

FIG_3

FIG_4

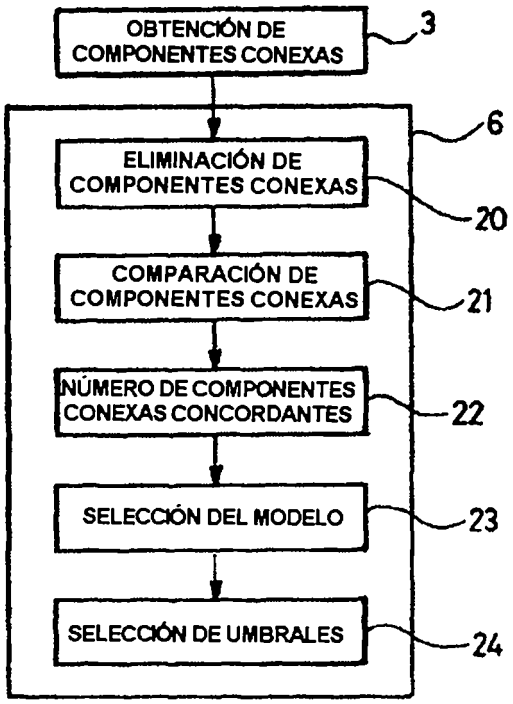
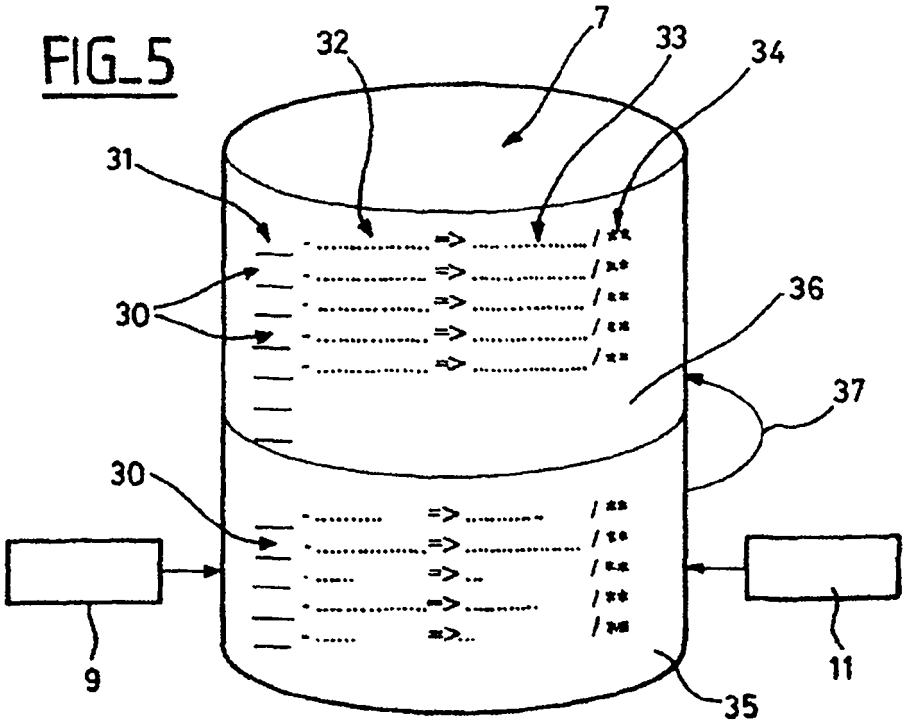

OBTENCIÓN DE COMPONENTES CONEXAS
3
6
ELIMINACIÓN DE COMPONENTES CONEXAS
20
COMPARACIÓN DE COMPONENTES CONEXAS
21
NÚMERO DE COMPONENTES CONEXAS CONCORDANTES
22
SELECCIÓN DEL MODELO
23
SELECCIÓN DE UMBRALES
24

FIG_5


7
32
33
34
31
30
36
37
30
9
11
35

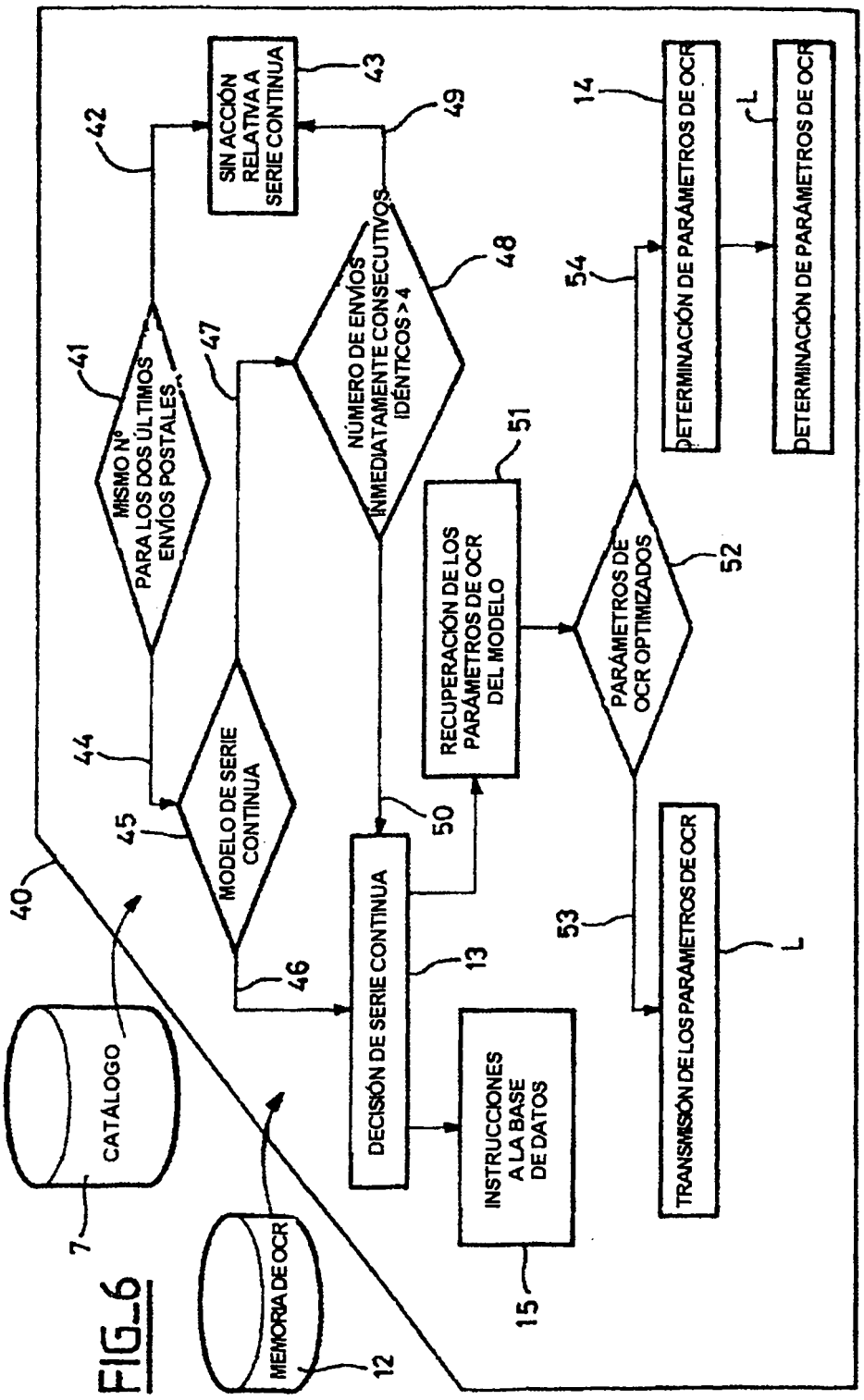

SIN ACCIÓN RELATIVA A SERIE CONTINUA
43
42
49
14
L
DETERMINACIÓN DE PARÁMETROS DE OCR
DETERMINACIÓN DE PARÁMETROS DE OCR
NÚMERO DE ENVÍOS INMEDIATAMENTE CONSECUTIVOS IDÉNTICOS > 4
48
54
MISMO Nº PARA LOS DOS ÚLTIMOS ENVÍOS POSTALES
41
47
RECUPERACIÓN DE LOS PARÁMETROS DE OCR DEL MODELO
51
PARÁMETROS DE OCR OPTIMIZADOS
52
MODELO DE SERIE CONTINUA
44
45
50
40
DECISIÓN DE SERIE CONTINUA
13
46
53
L
CATÁLOGO
7
INSTRUCCIONES A LA BASE DE DATOS
15
TRANSMISIÓN DE LOS PARÁMETROS DE OCR
MEMORIA DE OCR
12

FIG_6

FIG_7

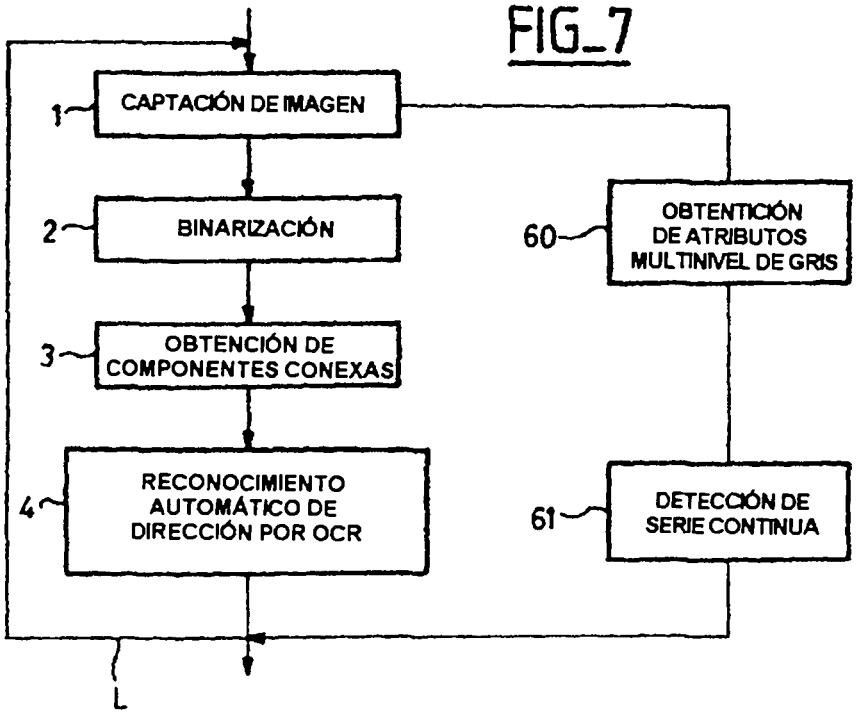
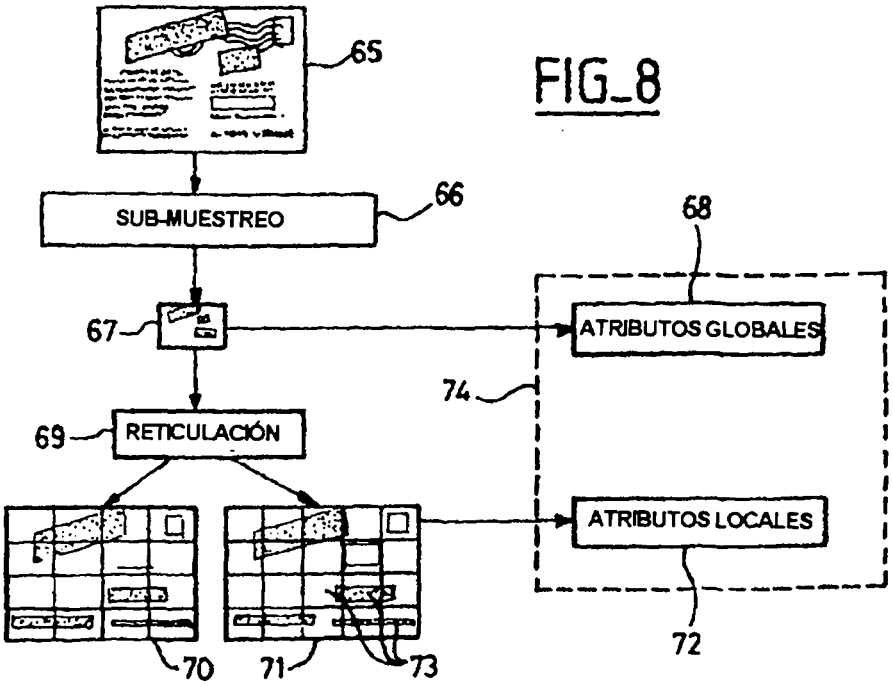

1
CAPTACIÓN DE IMAGEN
2
BINARIZACIÓN
3
OBTENCIÓN DE COMPONENTES CONEXAS
4
RECONOCIMIENTO AUTOMÁTICO DE DIRECCIÓN POR OCR
60
OBTENTICIÓN DE ATRIBUTOS MULTINIVEL DE GRIS
61
DETECCIÓN DE SERIE CONTINUA
L

FIG_8


65
SUB-MUESTREO
66
67
68
ATRIBUTOS GLOBALES
74
69
RETICULACIÓN
ATRIBUTOS LOCALES
70
71
73
72