

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 773 719**

51 Int. Cl.:

**G06K 9/20** (2006.01)

**G06K 7/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.03.2011 PCT/US2011/027439**

87 Fecha y número de publicación internacional: **15.09.2011 WO11112522**

96 Fecha de presentación y número de la solicitud europea: **07.03.2011 E 11753880 (1)**

97 Fecha y número de publicación de la concesión europea: **08.01.2020 EP 2545499**

54 Título: **Mejora de texto de una imagen textual sometida a un reconocimiento óptico de caracteres**

30 Prioridad:

**10.03.2010 US 720732**

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

**14.07.2020**

73 Titular/es:

**ZHIGU HOLDINGS LIMITED (100.0%)  
Harneys Services (Cayman) Limited, 4th Floor,  
Harbour Place, 103 South Church Street, George  
Town, P.O. Box 10240  
Grand Cayman KY1-1002, KY**

72 Inventor/es:

**GALIC, SASA;  
NIJEMCEVIC, DJORDJE y  
DRESEVIC, BODIN**

74 Agente/Representante:

**LÓPEZ CAMBA, María Emilia**

ES 2 773 719 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**

Mejora de texto de una imagen textual sometida a un reconocimiento óptico de caracteres

**5 Antecedentes**

El reconocimiento óptico de caracteres (ROC) es una traducción basada en ordenador de una imagen de texto en forma digital como texto editable por máquina, generalmente en un esquema de codificación estándar. Este procedimiento elimina la necesidad de escribir manualmente el documento en el sistema informático. Pueden surgir varios problemas diferentes debido a la mala calidad de imagen, las imperfecciones causadas por el procedimiento de escaneo y similares. Por ejemplo, un motor de ROC convencional puede estar acoplado a un escáner de superficie plana que escanea una página de texto. Debido a que la página se coloca al ras contra una cara de escaneo del escáner, una imagen generada por el escáner generalmente muestra incluso contraste e iluminación, asimetría y distorsión reducidas, y alta resolución. Por tanto, el motor de ROC puede traducir fácilmente el texto de la imagen al texto editable por máquina. Sin embargo, cuando la imagen es de menor calidad con respecto al contraste, la iluminación, la asimetría, etc., el rendimiento del motor de ROC puede verse degradado y el tiempo de procesamiento puede aumentar debido al procesamiento de todos los píxeles de la imagen. Este puede ser el caso, por ejemplo, cuando la imagen es generada por un escáner basado en cámara, porque el texto/imagen se escanea desde una distancia, desde diferentes orientaciones y con una iluminación variable. Incluso si el rendimiento del procedimiento de escaneo es bueno, el rendimiento del motor de ROC puede degradarse cuando se escanea una página de texto de calidad relativamente baja.

Los procedimientos de mejora de la calidad del texto para ROC se describen en KASAR T Y COL: "Font and background color independent text binarization", 2º TALLER INTERNACIONAL SOBRE ANÁLISIS Y RECONOCIMIENTO DE DOCUMENTOS BASADO EN CÁMARA (CBDAR2007), CURITIBA, BRASIL, 22 de septiembre de 2007, páginas 3-9 y en QIANG CHEN Y COL: "A double-threshold image binarization method based on edge detector", RECONOCIMIENTO DE PATRONES, vol. 41, no. 4 de abril de 2008, páginas 1254-1267.

**RESUMEN**

La precisión de un procedimiento de ROC puede mejorarse significativamente si se detecta y filtra el fondo de la imagen original mientras se consolidan los píxeles de texto restantes. En una implementación, el fondo se detecta y se filtra calculando los perfiles de color de fondo y texto. Por lo tanto, la mejora de texto se realiza utilizando una técnica de recuperación de trazo que identifica los píxeles de trazo de texto con alta precisión.

En una implementación particular, un procedimiento para mejorar una imagen textual para el reconocimiento óptico de caracteres comienza por recibir una imagen que incluye líneas de texto nativas. Se determina un perfil de línea de fondo que representa una intensidad de fondo promedio a lo largo de las líneas nativas en la imagen. Del mismo modo, se determina un perfil de línea de primer plano que representa una intensidad de fondo de primer plano promedio a lo largo de las líneas nativas en la imagen. Los píxeles de la imagen se asignan a una parte de fondo o primer plano de la imagen, basada al menos en parte en el perfil de línea de fondo y el perfil de línea de primer plano. La intensidad de los píxeles designados para la parte de fondo de la imagen se ajusta a un brillo máximo para representar una parte de la imagen que no incluye texto.

A continuación, se determina un ancho de trazo promedio o mediano de los componentes conectados que representan texto en la imagen después de que se han determinado los perfiles de línea de fondo y primer plano. Si el ancho promedio o mediano del trazo es menor que un valor umbral, la calidad del texto se mejora.

Este resumen se proporciona para introducir una selección de conceptos en una forma simplificada que se describen adicionalmente más adelante en la descripción detallada. Este resumen no está destinado a identificar características clave o características esenciales de la materia objeto reivindicada, ni pretende ser utilizado como una ayuda para determinar el alcance de la materia objeto reivindicada.

**BREVE DESCRIPCIÓN DE LOS DIBUJOS**

La figura 1 muestra un ejemplo ilustrativo de un sistema para el reconocimiento óptico de caracteres.  
 La figura 2 muestra un ejemplo de una línea textual que no es adecuada para el procesamiento.  
 La figura 3 muestra una transformación de la línea textual en la figura 2 que es más adecuada para el procesamiento de ROC.  
 La figura 4 muestra una transformación de imagen para garantizar que el texto sea más oscuro que el fondo.  
 La figura 5 muestra un ejemplo de una representación de línea de texto en el borde del espacio.  
 La figura 6 es una versión en escala de grises de la imagen en la figura 2.  
 La figura 7 muestra la parte más a la izquierda de la línea nativa en la figura 6 en la que el procesamiento local se realiza sobre un rectángulo que tiene dimensiones  $2 * h \times h$ .  
 La figura 8 muestra la línea nativa en la figura 7 representada en el borde del espacio.  
 La figura 9 muestra recuadros que rodean los componentes conectados que se muestran en la figura 8 que definen

un área (oscura) que contiene los bordes y otra (clara) que no contiene bordes.

La figura 10a es el histograma de la intensidad para la región rectangular localizada en la figura 7; la figura 10b es un histograma que ilustra la distribución de píxeles con respecto a su intensidad de color para las áreas gris oscuro en la figura 9; la figura 10c es un histograma obtenido restando los valores del segundo histograma en la figura 10b de los valores correspondientes del histograma en la figura 10a.

La figura 11 muestra los perfiles de línea calculados de fondo (línea superior) y primer plano (línea inferior) para la línea de texto en la figura 6.

La figura 12 muestra el resultado de aplicar el operador de inversión a texto claro sobre un fondo oscuro.

La figura 13 muestra un ejemplo de una imagen de texto de baja calidad.

La figura 14 muestra el resultado de aplicar un procedimiento de umbral a la imagen que se muestra en la figura 13.

La figura 15 muestra la manera en que se determina un ancho de trazo local para el único píxel de primer plano en el que se originan las flechas.

La figura 16 muestra un píxel que se asigna a un conjunto de cuatro píxeles en una imagen que se va a interpolar.

La figura 17 ilustra las etapas en el procedimiento de interpolación.

La figura 18 ilustra los resultados de aplicar el procedimiento de interpolación a la imagen de texto representada en la parte superior de la figura.

La figura 19 ilustra una palabra a la que se aplicará el procedimiento de mejora de texto.

La figura 20 representa el procedimiento de mejora de texto agregando píxeles a la primera letra de la palabra que se muestra en la figura 20.

La figura 21a muestra un ejemplo de un componente conectado; la figura 21b muestra los píxeles de borde para el componente conectado en la figura 21a; la figura 21c muestra el esqueleto del componente conectado en la figura 21a.

La figura 22 muestra los resultados de aplicar el procedimiento de adición de píxeles de la figura 20 a los esqueletos de un componente conectado. La figura 23 muestra los resultados de aplicar el procedimiento de mejora de texto a la imagen representada en la figura 13.

La figura 24 muestra un ejemplo de un aparato de procesamiento de imágenes que puede realizar una mejora de texto en una imagen textual sometida a reconocimiento óptico de caracteres.

La figura 25 es un diagrama de flujo que muestra un ejemplo de un procedimiento general para mejorar una imagen textual que se someterá a reconocimiento óptico de caracteres.

## **DESCRIPCIÓN DETALLADA**

La figura 1 muestra un ejemplo ilustrativo de un sistema 5 para el reconocimiento óptico de caracteres (ROC) en una imagen que incluye una disposición de captura de datos (por ejemplo, un escáner 10) que genera una imagen de un documento 15. El escáner 10 puede ser un escáner basado en cámara que utiliza un dispositivo de carga acoplada como un sensor de imagen para generar la imagen. El escáner 10 procesa la imagen para generar datos de entrada y transmite los datos de entrada a una disposición de procesamiento (por ejemplo, un motor de ROC 20) para el reconocimiento de caracteres dentro de la imagen. En este ejemplo particular, el motor de ROC 20 se incorpora al escáner 10. Sin embargo, en otros ejemplos, el motor de ROC 20 puede ser una unidad separada tal como una unidad independiente o una unidad que se incorpora a otro dispositivo tal como un PC, servidor o similar.

La precisión de un procedimiento de ROC puede mejorarse significativamente si se detecta y filtra el fondo de la imagen original mientras se consolidan los píxeles de texto restantes. Como se detalla a continuación, el fondo se detecta y se filtra calculando los perfiles de color de fondo y texto. La mejora de texto se realiza utilizando una técnica de recuperación de trazo que identifica los píxeles de trazo de texto con alta precisión.

La entrada a este procedimiento de mejora de texto es una imagen de las líneas de texto nativas (véase más abajo para una definición de "línea nativa"). Sin embargo, antes de proceder a describir el procedimiento real, se introducirán algunos conceptos matemáticos. Estos conceptos se utilizarán generalmente a lo largo de este análisis para facilitar la comprensión de la invención y, en algunos casos, para formalizar determinadas explicaciones.

**Definición 1.** La imagen digital en color de ancho  $w$  y altura  $h$  es la función vectorial de dos argumentos  $1: W \times H \rightarrow GS^3$  donde  $GS = [0, 1, \dots, 255]$ ,  $W = [0, 1, \dots, w-1]$ ,  $H = [0, 1, \dots, h-1]$  y  $\times$  indica producto cartesiano.

Esta definición se deriva del sistema de color RGB y los componentes  $r, g, b$  en  $1 (r, g, b)$  corresponden a los componentes rojo, verde y azul, respectivamente.

**Definición 2.** La imagen digital en escala de grises de ancho  $W$  y altura  $H$  es la función escalar de dos argumentos  $1: W \times H \rightarrow GS$  donde  $GS$  puede ser:

- $GS = [g1, g2]$ , donde la imagen en escala de grises se conoce como imagen binaria, binivel o bi-tonal
- $GS = [g1, g2, g3, \dots, g16]$  donde la imagen en escala de grises se conoce como imagen en escala de grises de 16 niveles
- $GS = [g1, g2, g3, \dots, g256]$  donde la imagen en escala de grises se conoce como imagen en escala de grises de 256 niveles.

En este punto, debe introducirse una convención utilizada en todo el documento. Como la imagen se considera como una función, se debe definir el sistema de coordenadas de su presentación gráfica. Por lo general, la esquina superior izquierda de la imagen se toma como punto de referencia. Esta convención se seguirá en esta invención y, por lo tanto, la esquina superior izquierda de la imagen se tratará como si estuviera colocada en el origen.

**Definición 3.** El triplete  $(I(x, y), x, y)$  se llama píxel. El par  $(x, y)$  se denomina coordenadas de píxel, mientras que  $I(x, y)$  se denomina valor de píxel.

Por lo general, el término "píxel" se usa indistintamente para las coordenadas, el valor, así como las coordenadas y el valor. El término "píxel" también se usará en esta invención de esta manera si no es probable que surja confusión, de lo contrario se usará el término exacto. Además, la notación  $I(x, y)$  se usará indistintamente cuando no haya confusión.

Las tres definiciones anteriores proporcionaron una comprensión de lo que es la imagen digital. La tarea habitual del procesamiento de imágenes es la serie de transformaciones que conducen a una presentación de la imagen original que es más conveniente para un análisis posterior para que se puedan extraer diversas conclusiones. Las siguientes definiciones proporcionan los medios matemáticos para la formalización de estas transformaciones.

**Definición 4.** Sea  $\Omega$  un conjunto de todas las imágenes con dimensiones  $w$  y  $h$ . La función  $T\Omega^n \rightarrow \Omega$  se llama operador de imagen  $n$ -aria. Si  $n = 1$ , el operador es unario, mientras que para  $n = 2$ , el operador es binario.

La definición anterior implica que el operador es una función que transforma una imagen (o varias imágenes) en otra imagen usando algún conjunto de reglas de transformación. Los operadores de imagen más importantes son los operadores basados en filtros. El filtro (a veces llamado núcleo o máscara) es la matriz  $A_{nm}$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

de tamaño  $n \times m$ . Por lo general,  $n$  y  $m$  son iguales e impares, produciendo, por ejemplo,  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$  filtros. El operador basado en filtro transforma una imagen de entrada usando la regla de que el píxel  $I_0(x, y)$  en la imagen de salida se calcula usando la fórmula:

$$I_0(x, y) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} I\left(x - \frac{n}{2} + i - 1, y - \frac{m}{2} + j - 1\right)$$

donde todas las divisiones son divisiones enteras.

En otras palabras, el píxel en la imagen de salida se construye convolucionando el vecindario del píxel correspondiente en la imagen de entrada con el filtro.

**Definición 5.** Sea  $I$  una imagen de ancho  $w$  y alto  $h$ , y sea  $I(x, y)$  un píxel arbitrario. El conjunto de píxeles  $\{I(x+1, y), I(x-1, y), I(x, y+1), I(x, y-1)\}$  se llama los 4 vecinos de  $I(x, y)$ . Del mismo modo, el conjunto de píxeles  $\{I(x+1, y), I(x-1, y), I(x, y+1), I(x, y-1), I(x-1, y-1), I(x-1, y+1), I(x+1, y-1), I(x+1, y+1)\}$  se llama 8 vecinos de  $I(x, y)$ .

Existen diferentes definiciones de adyacencia en la bibliografía, pero se utilizará la siguiente definición, ya que es la más conveniente para el análisis posterior.

**Definición 6.** Dos píxeles  $I(x_1, y_1)$  e  $I(x_2, y_2)$  son adyacentes si  $I(x_1, y_1)$  e  $I(x_2, y_2)$  es el elemento del conjunto de 8 vecinos de  $I(x_1, y_1)$  y sus valores de píxel son "similares".

La palabra similar se coloca entre comillas porque no existe una definición estricta de similitud. Más bien, esta definición se adopta según las demandas de una aplicación particular. Por ejemplo, se puede decir que dos píxeles son similares si sus valores de píxeles son iguales. En todo el documento se asumirá esta definición, si no se indica lo contrario.

**Definición 7.** Dos píxeles  $I(x_1, y_1)$  e  $I(x_n, y_n)$  están conectados si el conjunto  $\{I(x_2, y_2), I(x_3, y_3), \dots, I(x_{n-1}, y_{n-1})\}$  existe, de

modo que  $I(x_i, y_i)$  e  $I(x_{i+1}, y_{i+1})$  son adyacentes para  $i = 1, 2, \dots, n-1$ .

**Definición 8.** Un componente conectado es el subconjunto de píxeles de la imagen donde cada píxel del conjunto está conectado con todos los píxeles restantes del conjunto.

Para filtrar el fondo en una imagen de texto, es necesario entender la relación entre el texto y el fondo para proporcionar texto en escala de grises sobre un fondo limpio (por ejemplo, blanco). Este requisito puede aclararse con referencia a la figura 1.

La imagen de la línea textual representada en la figura 2 no es adecuada para el procesamiento sucesivo. Si pudiera transformarse en la imagen que se muestra en la figura 3 es probable que las etapas posteriores en el procedimiento de ROC sean más simples y precisas. Además, es conveniente suponer que el texto es más oscuro que el fondo. Si este es el caso, entonces todas las combinaciones texto-fondo se pueden manejar de manera uniforme. En consecuencia, cuando sea necesario, la transformación de la imagen representada en la figura 4 debería de ser realizada.

El análisis anterior debe dejar claro que, para filtrar el fondo de una imagen textual, se debe determinar la relación texto-fondo (en términos de color) y que el procedimiento de mejora del texto produce una imagen en la que el texto oscuro se presenta en un fondo blanco. En la siguiente sección se describe la detección de la relación texto-fondo.

### Perfil de línea

Lo primero a tener en cuenta es que una imagen de línea textual tiene dos tipos de píxeles, uno que se origina del texto y otro que se origina del fondo. Además, hay un cambio repentino de color en el borde entre el texto y el fondo. Esto lleva a la conclusión de que se puede observar una línea de texto en el espacio del borde. Se puede obtener una representación de línea en el espacio del borde utilizando cualquiera de una diversidad de técnicas de detección de bordes bien conocidas. Un ejemplo de una representación de línea en el espacio del borde se muestra en la figura 5.

Está claro que todas las características de texto importantes se conservan en la figura 5. Las letras se representan principalmente como componentes conectados en el espacio del borde. Además, la presentación de características textuales en el espacio de borde es independiente de la relación texto-fondo (por ejemplo, es irrelevante cuál es más oscuro). Por lo tanto, se puede formular la siguiente definición.

**Definición 9.** La línea nativa es un conjunto de todos los componentes conectados en el espacio del borde de la imagen de la línea textual.

Si  $CC_i$  es el componente conectado  $i$ -enésimo y  $nl$  es una línea nativa, entonces contiene:

$$nl = \{CC_1, CC_2, \dots, CC_n\}$$

El procedimiento de perfilado de línea se realiza en la línea nativa que produce los perfiles de línea definidos en la siguiente definición.

**Definición 10.** Sea  $nl$  una línea nativa y sea  $I(x, y)$ ;  $x \in \{0, 1, \dots, w-1\} \wedge y \in \{0, 1, \dots, h-1\}$  la imagen de línea nativa correspondiente con ancho  $w$  y altura  $h$ . Entonces las funciones  $FLF_{nl}: W \rightarrow GS$  y  $BLF_{nl}: W \rightarrow GS$  donde  $W = \{0, 1, \dots, w-1\}$  y  $GS = \{0, 1, \dots, 255\}$  se llaman perfil de línea de primer plano y perfil de línea de fondo si se cumplen las siguientes condiciones:

$$FLP_m(x_0) = \frac{\sum_{j=0}^{h-1} I(x_0, j) |_{I(x_0, j) \in FP(x_0)}}{\text{card}(FP(x_0))}$$

$$BLP_m(x_0) = \frac{\sum_{j=0}^{h-1} I(x_0, j) |_{I(x_0, j) \in BP(x_0)}}{\text{card}(BP(x_0))}$$

donde  $FP(x_0)$  es el conjunto de todos los píxeles de primer plano de la imagen de línea nativa con coordenada  $x = x_0$  y  $BP(x_0)$  es el conjunto de todos los píxeles de fondo de la imagen de línea nativa con coordenada  $x = x_0$ .

En otras palabras, las funciones de perfil reflejan la intensidad promedio de fondo o primer plano para una columna de imagen determinada.

La importancia de introducir perfiles de línea debe quedar clara después de consultar la figura 6, que es una versión en escala de grises de la imagen en la figura 2. Aunque a menudo es una suposición razonable que la intensidad del color del texto y la intensidad del color de fondo son bastante uniformes a lo largo del eje horizontal, esta suposición falla en el caso de la figura 6. La intensidad de fondo promedio en el lado izquierdo de la imagen es de alrededor de 190, mientras que en el lado derecho de la imagen es de alrededor de 100 (la diferencia es más de un tercio de todo el intervalo de escala de grises 0-255).

Así, en el caso de la figura 6, aproximar la intensidad de fondo con un solo valor en toda la imagen produciría una pérdida significativa de precisión, evitando que las etapas de procesamiento posteriores obtengan la información más confiable. Esta es la razón para introducir los perfiles de línea. El procedimiento de perfilado de línea es un procedimiento para estimar los perfiles de línea y se describirá a continuación.

Los perfiles de línea introducidos en la definición 10 son difíciles de calcular utilizando las fórmulas dadas debido a la falta de conocimiento sobre los píxeles de primer plano y fondo en este punto del procedimiento. Por lo tanto, se debe realizar algún tipo de estimación.

Lo primero que debe observarse es que, aunque puede haber una gran variación de color en toda la línea nativa, en una región localizada relativamente pequeña, el color debe ser bastante constante. Si esa región localizada se selecciona para ser, por ejemplo, un rectángulo con un ancho igual a varios múltiplos de  $h$  y una altura igual a  $h$  (donde  $h$  es la altura de la línea), entonces el procesamiento local puede realizarse en esa pequeña porción.

Las dimensiones del rectángulo se eligen para cumplir con el supuesto de que la variación de color no puede ser relativamente grande en una distancia proporcional al ancho de algunos caracteres. El procesamiento local implica que el procesamiento realizado en el rectángulo produce dos valores; uno para el color de fondo dominante y otro para el color de primer plano dominante. Estos dos resultados pueden usarse para aproximar los valores del perfil de línea sobre el rectángulo correspondiente. Un ejemplo de uno de esos rectángulos se representa en la figura 7.

La figura 7 muestra la parte más a la izquierda de la línea nativa en la figura 6 en la que el rectángulo seleccionado tiene dimensiones  $2^i h \times h$ . Si el procesamiento local produce  $df$  y  $db$  como los colores de primer plano y de fondo dominantes, respectivamente, entonces se puede usar la siguiente aproximación:

$$FLP_{ni}(x) \cong df; x = \{0, 1, \dots, 2^i h\}$$

$$BLP_{ni}(x) \cong bf; x = \{0, 1, \dots, 2^i h\}$$

Esta aproximación da lugar a una aproximación escalonada para el perfil de línea. Para calcular los valores de  $df$  y  $db$ , el rectángulo en la figura 7 debe representarse en el espacio de borde, como se muestra en la figura 8. También se pueden usar otras aproximaciones. Por ejemplo, la aproximación lineal por partes puede dar mejores resultados, aunque puede requerir un mayor tiempo de procesamiento.

Los tres rectángulos internos en la figura 8 corresponden a los recuadros delimitadores de los componentes conectados. Estos recuadros definen dos áreas diferentes, una que contiene los bordes y otra que no contiene los bordes. Estas áreas se ilustran en la figura 9. Las áreas de color gris oscuro en la figura 9 corresponden a áreas de borde, mientras que las áreas de color gris claro corresponden a áreas sin borde.

A continuación, se calcula un histograma que ilustra la distribución de píxeles con respecto a su intensidad de color para la región rectangular localizada en la figura 7. Se calcula otro histograma que ilustra la distribución de píxeles con respecto a su intensidad de color solo para las áreas gris oscuro en la figura 9. Sin embargo, se obtiene otro histograma restando los valores del segundo histograma de los valores correspondientes del primer histograma. Los histogramas calculados se muestran en las figuras 10a, 10b y 10c.

El histograma en la figura 10c es el histograma del área de color gris claro en la figura 9 (esto se debe a que fue creado por el procedimiento de sustracción mencionado anteriormente). Esta es el área que no incluye bordes, lo que implica el área sin letras ya que los bordes se encuentran solo en un límite de letra-fondo. En otras palabras, es el área correspondiente al fondo. Esto se demostrará analizando los histogramas en las figuras 10a y 10b con mayor detalle.

El histograma en la figura 10a representa la distribución de la intensidad de píxeles para todo el rectángulo en la figura 7. Como se muestra, hay un pico dominante causado por los píxeles de fondo. Estrictamente hablando, la presencia de este pico dominante no es una regla (considere un texto muy en negrita), por lo que aún no se supondrá que esto realmente representa el fondo. También hay una parte relativamente plana del histograma a la izquierda del pico que es causada por el texto y el fondo, así como por su combinación.

El histograma 10b representa la distribución de la intensidad de píxeles para píxeles dentro de los rectángulos internos que se muestran en la figura 9, que corresponde al área de borde. Este histograma parece bastante similar al de 10a, excepto que el pico de fondo no es tan dominante. El tercer histograma (figura 10c) es lo más importante. Como se señaló anteriormente, corresponde a los píxeles de fondo. Debido al procedimiento de sustracción del histograma, los

componentes que se originan del texto y la combinación de texto-fondo se cancelan por completo (tenga en cuenta la ausencia del área plana de la figura 10a). El bin particular en el histograma correspondiente a los máximos del pico da la intensidad del color de fondo dominante. En este ejemplo es 190.

5 El intervalo dinámico de fondo también se puede obtener del histograma en la figura 10c. Se puede encontrar examinando la intensidad comenzando en 0 y encontrando el valor de intensidad para el primer bin que tiene un número de píxeles distinto de cero. En la figura 10c este valor es 150. A continuación, el valor de intensidad se encuentra para el bin posterior en el que el número de píxeles vuelve a cero. En la figura 10c este valor es 230. Por tanto, el intervalo dinámico es [150, 230]. Los valores de intensidad en el histograma 10b que están fuera de este intervalo son causados por píxeles de texto en primer plano. El bin en el histograma de la figura 10b correspondiente al valor máximo del pico secundario que surge de los píxeles de primer plano corresponde a la intensidad del color de primer plano dominante, que en este ejemplo es 47.

15 Las estimaciones para los perfiles de línea de primer plano y de fondo se pueden obtener aplicando este procedimiento a cada región localizada (por ejemplo, el rectángulo en la figura 7) de la línea nativa. Cabe señalar que todo el procesamiento necesario para obtener estas estimaciones se basa en la simple manipulación del histograma; a saber, cálculo de histograma, sustracción de histograma y extracción de valor máximo. No se realizan operaciones complejas (por ejemplo, detección de pico, estimación de ancho de pico, detección de áreas planas), lo que deja pocas oportunidades para que surjan errores potenciales de un análisis complejo. Los perfiles de línea de fondo y primer plano calculados para la línea de texto en la figura 6 están representados en la figura 11, donde la línea superior representa el perfil de la línea de fondo y la línea inferior representa el perfil de la línea de primer plano.

25 Se debe realizar una etapa adicional en este punto. Es decir, al observar la relación entre los dos perfiles de línea, la relación texto-fondo se puede extraer fácilmente para determinar cuál es más oscura. Si se encuentra texto claro sobre un fondo oscuro, se puede aplicar el operador de inversión de imagen  $Inv(I(x,y))$ . Este operador se define como

$$Inv(I(x,y)) = 255 - I(x,y)$$

30 El resultado de aplicar este operador se representa en la figura 12. Después de realizar esta inversión, cuando sea necesario, se puede suponer con seguridad que aparece texto oscuro sobre un fondo claro. Una vez que se han calculado los perfiles de las líneas de fondo y primer plano (y se realiza la inversión de la imagen cuando es necesario), el fondo se puede filtrar mediante el umbral, que se describirá a continuación.

### 35 Umbral

Una vez que se han calculado los perfiles de línea, el procedimiento de filtrado del fondo es bastante sencillo. La función  $Thresh(I(x,y)) : GS \rightarrow GS$  se aplica a cada píxel y se define como:

$$Thresh(I(x,y)) = \begin{cases} 255, & I(x,y) > (FLP(x) + BLP(x)) / 2 \\ I(x,y), & otherwise \end{cases}$$

40 En otras palabras, cada píxel con un valor de intensidad de píxel más pequeño que el promedio de los perfiles de primer plano y fondo se deja como está, mientras que otros píxeles se cambian para que parezcan tener una intensidad máxima. Es decir, se cambian a blanco. Este tipo de umbral puede verse como un procedimiento que clasifica los píxeles como primer plano o fondo en función de sus valores de píxeles. Si un píxel se clasifica como un píxel de primer plano, su valor no se modifica, y si el píxel se clasifica como un píxel de fondo, su valor cambia a blanco.

Más generalmente, en algunas implementaciones, en lugar de establecer el umbral como el promedio de los perfiles de primer plano y de fondo, se puede seleccionar algún otro parámetro que se determina a partir de los perfiles de primer plano y de fondo.

50 La imagen representada en la figura 3 se obtuvo utilizando la técnica de umbral mencionada anteriormente. Esta técnica produce muy buenos resultados. Sin embargo, su aplicación en imágenes de baja calidad no siempre es satisfactoria y se puede aplicar alguna técnica de mejora de texto a dichas imágenes, como se analiza a continuación.

### 55 Mejora de texto de baja calidad

La figura 13 muestra un ejemplo de una imagen de texto de baja calidad. El resultado de aplicar el procedimiento de umbral mencionado anteriormente en esta imagen se muestra en la figura 14. Evidentemente, la imagen resultante es de baja calidad. Todas las etapas de procesamiento posteriores están destinadas a tomar decisiones incorrectas, ya que la imagen es casi ilegible incluso para un ser humano. Las razones de los malos resultados son claras. Los valores de píxel que componen los trazos de texto en la figura 13 tienen grandes variaciones. Algunos de ellos son aún más

similares a los píxeles de fondo que los píxeles de primer plano, y algunos píxeles ruidosos tienen valores similares a los valores de píxeles de primer plano. Además, el ancho del trazo es extremadamente pequeño, lo que resulta en poca integridad del texto.

- 5 El procedimiento de umbral descrito anteriormente usa un valor fijo para la función de umbral que está a medio camino entre los perfiles locales de fondo y primer plano. Una estrategia para superar el problema relacionado con las imágenes textuales de baja calidad es hacer que el procedimiento de umbral sea más adaptable para que el umbral sea ajustable para diferentes imágenes. Desafortunadamente, generalmente se necesita más información de contexto para lograr esto de manera efectiva. Una alternativa al procedimiento de umbral implica detectar una imagen de baja calidad utilizando un procesamiento más inteligente (y más intensivo) para proporcionar un mejor filtrado de fondo. Dicho procedimiento alternativo se describirá después de analizar cómo se pueden detectar imágenes textuales de baja calidad.

### DetECCIÓN DE TEXTO DE BAJA CALIDAD

- 15 La primera etapa para detectar imágenes textuales de baja calidad implica identificar las características que hacen que tal imagen sea de baja calidad. Como se mencionó anteriormente, dicha imagen textual se caracteriza por un ancho de trazo pequeño y una gran variación de color textual. Si el ancho del trazo fuera mayor, el procedimiento de umbral descrito anteriormente produciría una interpretación "más delgada" de los caracteres, pero aún así serían legibles. Esto también es cierto con respecto a pequeñas variaciones de color. Se puede concluir que las características clave de una imagen textual de baja calidad son un ancho de trazo pequeño en combinación con grandes variaciones de intensidad de color.

- 25 Otra cosa a tener en cuenta es el resultado del procedimiento de umbral representado en la figura 14. Es evidente que este procedimiento produce trazos en partes de los caracteres que tienen un ancho pequeño (típicamente menos de un par de píxeles). Al calcular la media de los anchos de "trazo" en esta imagen, se puede evaluar la calidad del texto.

La siguiente definición es necesaria para describir un procedimiento para estimar el ancho del trazo.

- 30 **Definición 11.** Sea  $cc = \{(x_1, y_1), \dots, (x_n, y_n)\}$  un componente conectado. Cada  $(x_i, y_i) \in cc$  de modo que existe el  $(x_j, y_j) \notin cc$  que pertenece al conjunto de 8 vecinos de  $(x_i, y_i)$  se llama el *píxel del borde de cc*.

- 35 El procedimiento de estimación del ancho del trazo comienza calculando el ancho del trazo local para cada píxel de primer plano de borde en la figura 14. Esto se puede lograr calculando los anchos de trazo en todas las direcciones posibles, como se ilustra en la figura 15 para el único píxel de primer plano en el que se originan las flechas.

- 40 Comenzando en el primer plano seleccionado, se examinan los píxeles vecinos en la dirección de 90 grados. En la dirección vertical hay una serie de 3 píxeles continuos en primer plano antes de llegar a un píxel de fondo. Del mismo modo, hay 2 píxeles de primer plano en la dirección de 0 grados, 1 píxel de primer plano en la dirección de 45 grados y 3 píxeles de primer plano que se extienden en la dirección de 135 grados. De entre estos cuatro valores, el más pequeño (en este caso 1) se selecciona como el ancho del trazo local para el píxel seleccionado. Este mismo procedimiento se aplica a todos los píxeles de borde, produciendo así un conjunto de anchos de trazo locales. Una buena estimación del ancho total del trazo del texto es la mediana de este conjunto de valores, que se indica  $SW_{or}$ .

- 45 Una vez que se calcula  $SW_{or}$  todos los datos necesarios están disponibles para examinar la calidad de la imagen de texto. Si  $SW_{or}$  es más pequeño que algún umbral predefinido  $SW_{fr}$ , la imagen puede considerarse una imagen de mala calidad. La elección de  $SW_{fr}$  influye en el resto del procedimiento. A medida que se eligen valores más pequeños de  $SW_{fr}$ , se considerará que la menor cantidad de imágenes de líneas nativas es de baja calidad. Si bien esto implica una cantidad menor de procesamiento posterior, también deja algunas imágenes de líneas nativas sin mejorar. Por otro lado, si se elige un valor mayor para  $SW_{fr}$ , se mejorarán más líneas nativas al precio de una mayor potencia de procesamiento. En algunas implementaciones, un valor de compromiso adecuado está en el intervalo  $SW_{fr} \in [2, 4]$ .

- 55 Una vez que se detecta una imagen de baja calidad, se puede aplicar un procedimiento de mejora de texto. A continuación, se describirán dos procedimientos ilustrativos para mejorar el texto. El primero es muy preciso pero costoso en términos de la potencia de procesamiento requerida, mientras que el segundo es más eficiente pero algo menos preciso.

### MEJORA PRECISA DE TEXTO

- 60 Las imágenes que se pasan a un componente de mejora de texto tienen un ancho de trazo  $SW_{or}$ , que es más pequeño que un umbral predefinido. Esto implica que solo hay un pequeño grado de libertad en el procesamiento. Una forma de superar este problema es realizar una interpolación de imágenes. La interpolación de imágenes es una técnica para aumentar la resolución de una imagen. Aunque hay un gran número de técnicas de interpolación disponibles, un ejemplo adecuado es la interpolación bilineal, que es muy simple pero lo suficientemente potente como para satisfacer las necesidades del procedimiento de mejora de texto.

Supóngase, a modo de ejemplo, que una imagen debe ser interpolada para que su resolución se incremente en un factor de dos en ambas direcciones (x e y). Esto significa que cada píxel en la imagen original se asigna a un conjunto de cuatro píxeles en la imagen interpolada. Este procedimiento se ilustra en la figura 16.

5 La primera etapa en el procedimiento de interpolación bilineal es asignar los valores de píxel en las coordenadas (x, y) en la imagen original a las coordenadas (2x,2y) en la imagen de salida. Este procedimiento se representa en la figura 17a. Los píxeles que incluyen un pequeño rectángulo sombreado son píxeles a los que no se les asigna un valor. De esta manera, a un cuarto de los píxeles de salida se les asignan los valores de píxel adecuados. Tenga en cuenta que la asignación o copia produce algunos píxeles en la imagen de salida que se encuentran entre dos píxeles asignados (estos son los píxeles que tienen al menos una coordenada par). El valor de estos píxeles puede calcularse como el valor promedio asignado a sus píxeles vecinos (véase la figura 17b). En este punto, solo los píxeles con coordenadas impares en la imagen de salida permanecen sin asignar. Los valores de estos píxeles pueden calcularse como el promedio de todos los valores de píxeles de los píxeles en su conjunto de 4 vecinos. Esta última etapa en el procedimiento de interpolación bilineal produce la imagen interpolada (véase la figura 17c). La imagen final en la figura 17c tiene algunos píxeles sin valores asignados, ya que sus valores dependen de los valores de píxeles que no se muestran en la figura.

10 La parte inferior de la figura 18 muestra el resultado de aplicar el procedimiento de interpolación a la imagen de texto representada en la parte superior de la figura.

20 Un efecto secundario ventajoso que surge de la aplicación de la interpolación de imagen a una imagen de texto de baja calidad es el efecto de suavizado inherente del procedimiento de interpolación. Esto produce una imagen con una variación menor en la intensidad del color del texto, que fue uno de los problemas con el texto de baja calidad.

25 En este punto del procedimiento de mejora de texto, se conoce la siguiente información sobre la imagen interpolada:

- Perfil de línea de fondo  $BLP(x)$
- Perfil de línea de primer plano  $FLP(x)$
- Ancho de trazo  $SW_{hr} = 2 * SW_{or} + 1$

30 Dado que el algoritmo de umbral produce una versión erosionada del texto, todos los valores de píxeles más pequeños que  $(BLP(x)+FLP(x))/2$  se pueden considerar de forma segura como parte del texto y asignarse un valor de 0. Del mismo modo, los valores mayores que  $BLP(x)$  pueden clasificarse como píxeles de fondo y asignarse un valor de 255. Sin embargo, los píxeles con valores dentro del intervalo de  $((BLP(x)+ FLP(x))/2$  to  $BLP(x)) = (L_1(x), L_2(x))$  son cuestionables en términos de clasificación, y se necesita algún procesamiento adicional para determinar si pertenecen al texto o al fondo.

35 Para clasificar estos píxeles de una manera manejable y conveniente, la resolución del color del píxel se puede dividir en un número finito de intervalos de intensidad n sobre el intervalo indeterminado mencionado anteriormente usando la ecuación

$$I(x, y) = \begin{cases} I_1 = \Delta I, & I(x, y) \in (L_1(x), L_1(x) + \Delta L) \\ I_2 = 2\Delta I, & I(x, y) \in (L_1(x) + \Delta L, L_1(x) + 2\Delta L) \\ \dots \\ I_{n-1} = (n-1)\Delta I, & I(x, y) \in (L_1(x) + (n-1)\Delta L, L_1(x) + n\Delta L) \end{cases}$$

donde  $\Delta I = 255/n$  y  $\Delta L = (L_2(x) - L_1(x))/n$ .

45 El parámetro que determina el tamaño de los intervalos de intensidad es n, que debe elegirse para producir un pequeño número de intervalos de intensidad. En algunas implementaciones, una buena opción puede ser  $n \in [3,5]$ , por ejemplo. Disminuir la resolución del color dividiendo la intensidad en n intervalos puede usarse para mejorar la calidad del texto de una manera que se describirá en relación con las figuras 19 y 20.

50 El procedimiento de mejora de texto ahora se aplicará a la letra F en la figura 19. La figura 20 representa el procedimiento agregando píxeles a esta letra. Los píxeles que tienen una intensidad dentro del primer intervalo,  $\Delta I$ , se agregan a la letra más a la izquierda en la figura 20. Los píxeles que tienen una intensidad dentro del segundo intervalo,  $\Delta I$ , se agregan a la segunda letra desde la izquierda en la figura 20. Este procedimiento se repite agregando píxeles dentro de cada intervalo de mayor intensidad posterior a cada letra subsiguiente en la figura.

55 La figura 20 ilustra claramente los resultados de agregar píxeles de esta manera. En particular, la letra F se vuelve más y más pronunciada a medida que se agregan píxeles. Si bien está claro que este procedimiento puede usarse para mejorar o reconstruir letras individuales, surgen problemas cuando las letras comienzan a "filtrarse" en el fondo

o comienzan a tocarse entre sí. Para evitar esto, se debe emplear algún tipo de criterio de detención para determinar cuándo no se deben agregar píxeles adicionales. La siguiente definición es útil para obtener un criterio de detención apropiado.

5 **Definición 12.** Sea  $ccbe$  un componente conectado. Un conjunto de píxeles que pertenecen a un componente conectado que está igualmente distante de dos píxeles de borde diferentes de  $ccis$  llamados esqueleto de  $cc$ .

La figura 21a muestra un ejemplo de un componente conectado. Los píxeles más oscuros en la figura 21b son los píxeles de borde para este componente conectado y los píxeles sombreados en la figura 21c representan el esqueleto del componente conectado.

Una cosa a tener en cuenta es que el procedimiento de adición de píxeles casi no tiene impacto en los esqueletos. Esto se puede ver en la figura 22, que representa los esqueletos de los componentes conectados correspondientes al procedimiento de adición de píxeles de la figura 20. En cada etapa de adición de píxeles, el esqueleto del componente conectado permanece casi igual. Lo que sí cambia es la distancia entre los píxeles del esqueleto y los píxeles del borde. Cuantos más píxeles se agreguen, mayor será la distancia entre los píxeles del esqueleto y los píxeles del borde.

En este punto es importante recordar que el ancho de trazo  $SW_{hr}$  de este componente conectado ya se ha calculado. Es razonable exigir que el procedimiento de adición de píxeles finalice cuando una pequeña cantidad de píxeles de esqueleto excede una distancia de  $SW_{hr}/2$  desde los píxeles de borde. Si se excediera este valor, el componente conectado comenzaría a "filtrarse" en el fondo. En consecuencia, no se deben agregar más píxeles a la letra en este momento.

25 Teniendo en cuenta el análisis anterior, un algoritmo de mejora de texto puede formularse de la siguiente manera:

- 1) Interpolación de una imagen de texto de baja calidad en una resolución dos veces mayor.
- 2) Disminuir la resolución de color de la imagen de texto definiendo un número finito de intervalos de intensidad.
- 3) Identificar todos los componentes conectados en la imagen que se forman solo a partir de los píxeles más oscuros (por ejemplo, aquellos píxeles con una intensidad de cero). Estos componentes conectados se denominarán componentes conectados a la base.
- 4) Para cada componente conectado a la base, realizar las siguientes etapas:
  - a. Calcular el esqueleto del componente conectado.
  - b. Si los píxeles en el esqueleto están a menos de una distancia  $SW_{hr}/2$  de los píxeles de borde, crear un nuevo componente conectado agregando al componente conectado actual los píxeles vecinos que pertenecen al siguiente intervalo de intensidad (*corriente*  $I + \Delta I$ ).
  - c. Repetir las etapas a y b hasta que se cumpla el criterio de detención. Es decir, el procedimiento se detiene cuando hay píxeles de esqueleto que están a más de una distancia de  $SW_{hr}/2$  de los píxeles de borde.

Una vez que todos los componentes conectados a la base se procesan de esta manera, el procedimiento de mejora de texto finaliza. El resultado de aplicar este procedimiento a la imagen representada en la figura 13 se ilustra en la figura 23.

#### 45 Mejora de texto eficiente

El algoritmo descrito anteriormente para mejorar la imagen textual utilizando la recuperación de trazo a menudo puede proporcionar muy buenos resultados. Sin embargo, un inconveniente es su complejidad computacional. En esta sección se describirá una variante del algoritmo que es menos costosa (en términos de potencia de procesamiento), pero con una calidad comparable. Esta variante no está incluida por el alcance de las reivindicaciones adjuntas.

Como en la estrategia anterior, esta estrategia realiza la interpolación de imágenes y divide la resolución del color en un número finito de intervalos de la manera descrita anteriormente. Sin embargo, el algoritmo descrito en la sección anterior lleva mucho tiempo debido al procedimiento recursivo aplicado a los componentes conectados a la base, que implica el cálculo del esqueleto. Dado que los esqueletos son una parte integral del criterio de detención, se puede emplear alguna otra variante del criterio de detención para reducir la complejidad computacional del algoritmo. Una estrategia se describe a continuación.

Después de disminuir la resolución del color, queda un pequeño conjunto de colores. Todos los píxeles que tienen una intensidad (píxeles blancos) se consideran parte del fondo. Todos los píxeles con intensidad de 0 (píxeles negros) se consideran píxeles de texto. Son los píxeles grises los que son cuestionables. Con referencia nuevamente a la figura 19, una forma de evitar la necesidad de realizar un cálculo exacto del esqueleto mientras se obtiene una estimación del mismo es encontrar (para cada píxel gris) el rectángulo máximo, centrado en ese píxel, que está formado completamente por píxeles que no son blancos. Sea  $a(x, y)$  la dimensión del rectángulo para el píxel  $I(x, y)$ . Todos los píxeles que tienen un rectángulo que satisface  $a(x, y) \geq SW_{hr}$  son casi seguramente parte del texto o un área plana y ruidosa en el fondo. Este conjunto se indica como  $A = \{(x_1, y_1), \dots, (x_j, y_j)\}$ . Ahora, todos los píxeles  $(x_k, y_k) \notin A$  de modo

que  $\forall(x, y), d((x_k, y_k), (x, y)) \leq SW_{hr}$  donde  $d((x_k, y_k), (x, y)) = \min(|x_k - x|, |y_k - y|)$  se agregan al conjunto. En otras palabras, se agregan todos los píxeles de la vecindad de los píxeles dentro del conjunto A, que es una alternativa al procedimiento de adición de píxeles en el algoritmo descrito anteriormente. Para garantizar que no se produzca una "fuga" en el fondo, el conjunto A puede sufrir un filtrado. De esta forma, los píxeles del conjunto se pueden eliminar si su distancia desde el borde más cercano es mayor que  $SW_{hr}$ . De esta forma las áreas planas y ruidosas se eliminan del texto.

Este algoritmo es similar al algoritmo más preciso descrito anteriormente. Sin embargo, en lugar de realizar un cálculo recursivo antes de agregar cada conjunto de píxeles dentro de un intervalo, se aplica una aproximación de una etapa, lo que resulta en una menor cantidad de tiempo de procesamiento. Se ha descubierto que los resultados logrados al aplicar este algoritmo en una imagen de texto de baja calidad son similares a los resultados logrados al aplicar el algoritmo preciso, que se representan en la figura 23.

La figura 24 muestra un ejemplo de un aparato de procesamiento de imágenes 300 que puede realizar el procedimiento de mejora de texto descrito anteriormente en una imagen textual sometida a reconocimiento óptico de caracteres. El aparato, que puede incorporarse en un motor de ROC, incluye un componente de entrada 202 para recibir una imagen de entrada que tiene líneas de texto nativas. El aparato también incluye un componente de perfilado de línea de fondo y primer plano 204, un componente de detección de calidad de imagen 206, un componente de umbral 208 y un componente de mejora de trazo 210. Un componente de salida 212 genera la imagen mejorada como salida, que puede ser empleada por componentes posteriores del motor de ROC.

La figura 25 es un diagrama de flujo que muestra un ejemplo de un procedimiento general para mejorar una imagen textual que se someterá a reconocimiento óptico de caracteres. El procedimiento puede implementarse mediante el aparato de procesamiento de imágenes que se muestra en la figura 24. El procedimiento comienza en la etapa 305 cuando el componente de entrada 202 recibe una imagen de líneas nativas de texto. El componente de perfilado de línea 204 realiza la detección de borde en las líneas nativas en la etapa 310 para producir una representación de las líneas nativas en el espacio de borde. En la etapa 315, el componente de perfilado de línea 204 identifica a continuación cada componente conectado que surge en la representación de las líneas nativas en el espacio de borde. Usando la información obtenida en las etapas 310 y 315, el componente de perfilado de línea 204 determina en la etapa 320 los perfiles de línea de fondo y primer plano de la manera descrita anteriormente. Una vez que se han determinado los perfiles de línea, el componente de detección de calidad de imagen 206 determina en la etapa 325 si la calidad de la imagen es suficiente para asignar píxeles a la parte de fondo o primer plano de la imagen en función de los perfiles de línea. Como se describió anteriormente, un criterio que puede usarse para determinar la calidad de la imagen es el ancho del trazo de los componentes conectados. Si se encuentra que la calidad de imagen en la etapa de decisión 330 es suficientemente alta, los píxeles se asignan a la parte de primer plano o de fondo de la imagen por el componente de umbral 208 en la etapa 335. Si, por otro lado, la calidad de imagen no es suficientemente alta, la mejora de imagen se realiza mediante el componente de mejora de trazo 210 en la etapa 340 para mejorar los trazos de los componentes conectados. Finalmente, el componente de salida 212 genera en la etapa 345 la imagen mejorada, que puede ser empleada por componentes posteriores del motor de ROC.

**REIVINDICACIONES**

1. Un procedimiento de mejora de una imagen textual para el reconocimiento óptico de caracteres, que comprende:

5 recibir una imagen que incluye líneas de texto nativas;  
realizar la detección de bordes en las líneas nativas para producir una representación de las líneas nativas en el espacio de borde; detectar componentes conectados en las líneas nativas representadas en el espacio de borde; determinar un perfil de línea de fondo que representa una intensidad de fondo promedio a lo largo de las líneas nativas en la imagen;  
10 determinar un perfil de línea de primer plano que representa una intensidad de primer plano promedio a lo largo de las líneas nativas en la imagen, donde determinar los perfiles de línea de fondo y primer plano incluye determinar una pluralidad de valores de intensidad de fondo y primer plano para una pluralidad de regiones localizadas a lo largo de cada una de las líneas nativas representadas en el espacio de borde;  
15 estimar un ancho de trazo promedio o mediano de componentes conectados que representan texto en la imagen; mejorar la calidad del texto si el ancho de trazo medio o mediano es menor que un valor umbral, comprendiendo la mejora:

20 asignar píxeles en la imagen que tiene una intensidad menor que un parámetro seleccionado a una parte de primer plano de la imagen, donde el parámetro seleccionado se determina al menos en parte a partir del perfil de línea de fondo y el perfil de línea de primer plano;  
asignar píxeles en la imagen que tiene una intensidad mayor que el perfil de línea de fondo a la parte de fondo de la imagen; y asignar píxeles seleccionados que tienen una intensidad entre el parámetro seleccionado y el perfil de la línea de fondo para basar los componentes conectados de las líneas nativas representadas en el espacio de borde agregando incrementalmente conjuntos de píxeles con intensidades cada vez más altas, donde cada conjunto define píxeles que tienen una intensidad dentro de uno de un conjunto finito de intervalos de intensidad; y  
25 terminar la asignación de píxeles seleccionados a al menos uno de los componentes conectados a la base cuando una distancia entre un píxel de borde y un píxel del esqueleto del componente conectado a la base es mayor que la mitad del ancho de trazo promedio o mediano.  
30

2. El procedimiento de la reivindicación 1, que comprende además estimar un ancho de trazo promedio o mediano de los componentes conectados que representan texto en la imagen determinando un ancho de trazo local para cada píxel asignado a los valores de intensidad de primer plano y mejorando la calidad del texto si el ancho de trazo promedio o mediano es menor de un valor umbral.  
35

3. El procedimiento de la reivindicación 1, donde mejorar la calidad del texto incluye realizar una interpolación de imagen para aumentar la resolución de la imagen.

40 4. Un sistema que mejora una imagen textual sometida a un reconocimiento óptico de caracteres, adaptado para realizar el procedimiento de cualquiera de las reivindicaciones 1 a 3.

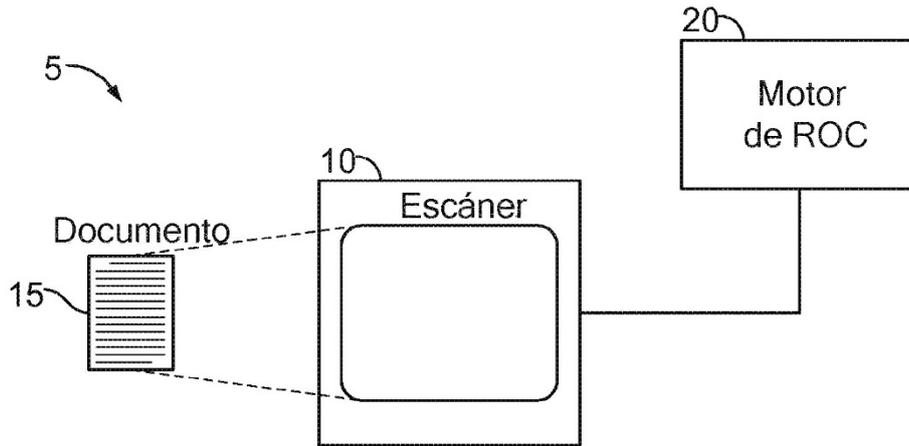


FIG. 1



FIG. 2  
(Texto de muestra)



FIG. 3  
(Texto en escala de grises sobre fondo blanco)

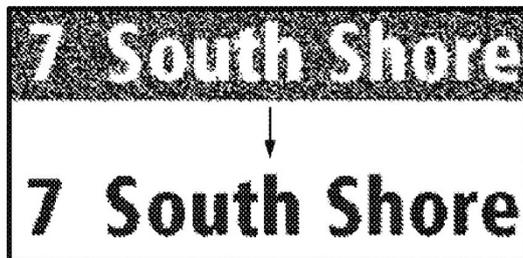


FIG. 4  
(Transformación de fondo oscuro)



FIG. 5  
(Representación de línea textual en el espacio de borde)



FIG. 6  
(Versión en escala de grises de la imagen de texto)



FIG. 7  
(Parte de línea nativa)

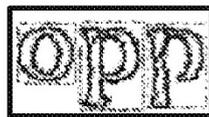


FIG. 8  
(Parte de línea nativa en el espacio de borde)

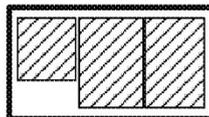


FIG. 9  
(Parte de línea nativa en el espacio de borde)

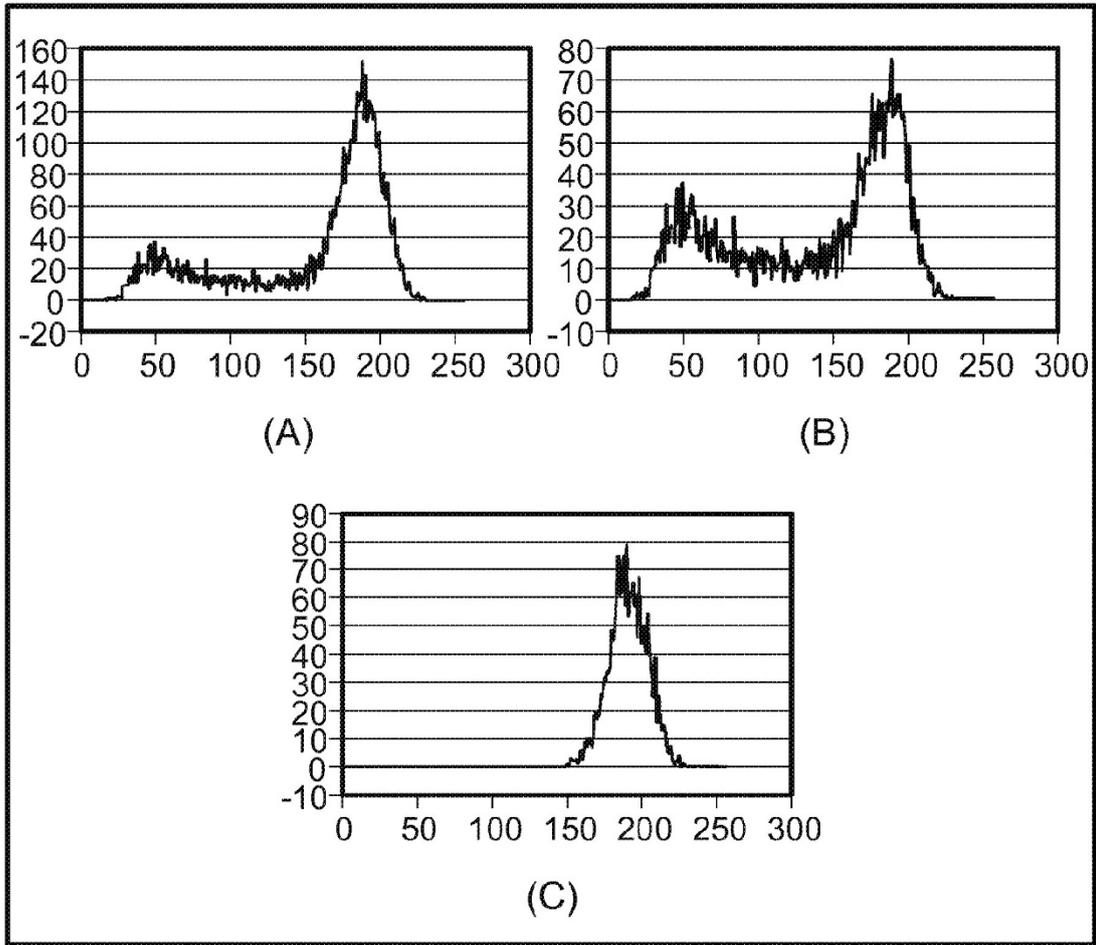


FIG. 10  
(Histogramas)

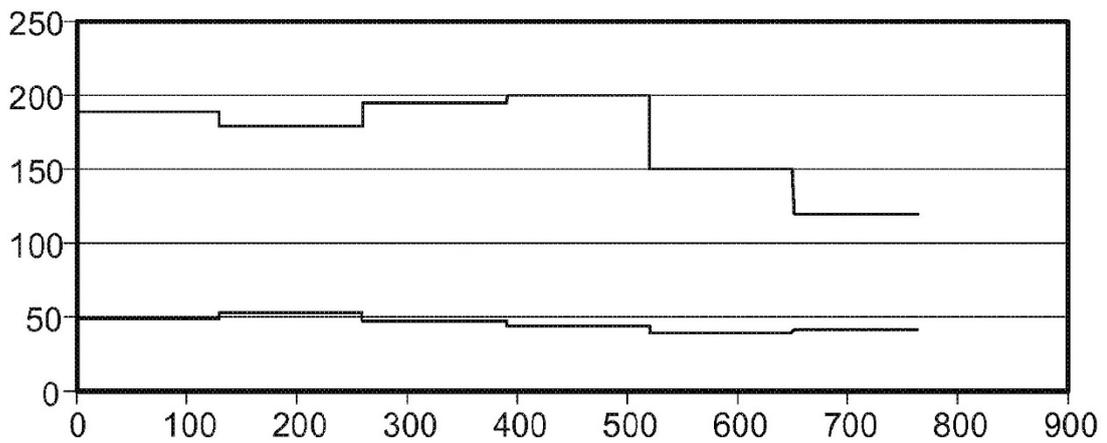


FIG. 11  
(Perfiles de línea)

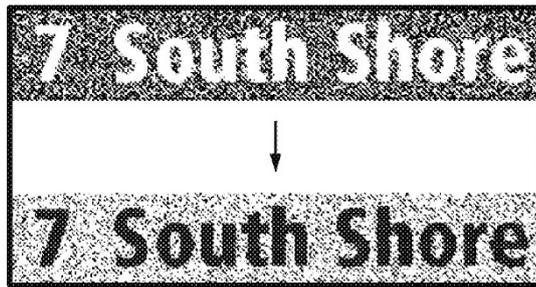


FIG. 12  
(Inversiones de imagen)



FIG. 13  
(Imagen de texto de baja calidad)

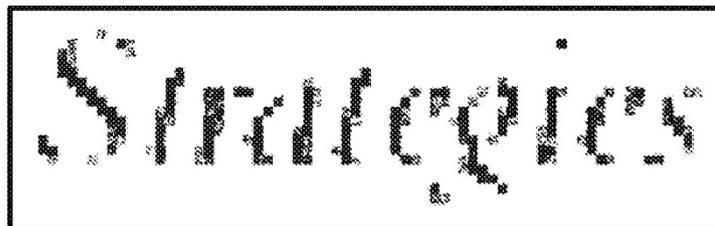


FIG. 14  
(Resultado de umbral en la imagen de texto de baja calidad)

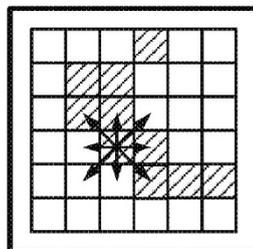


FIG. 15  
(Cálculo del ancho de trazo para el único píxel de primer plano)

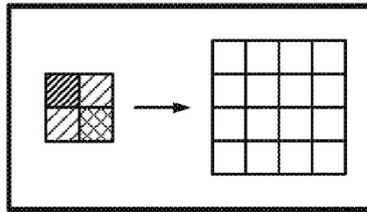


FIG. 16

(Interpolación de imagen)

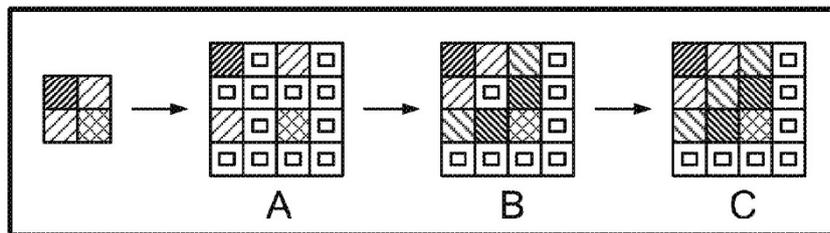


FIG. 17

(Etapas de interpolación)

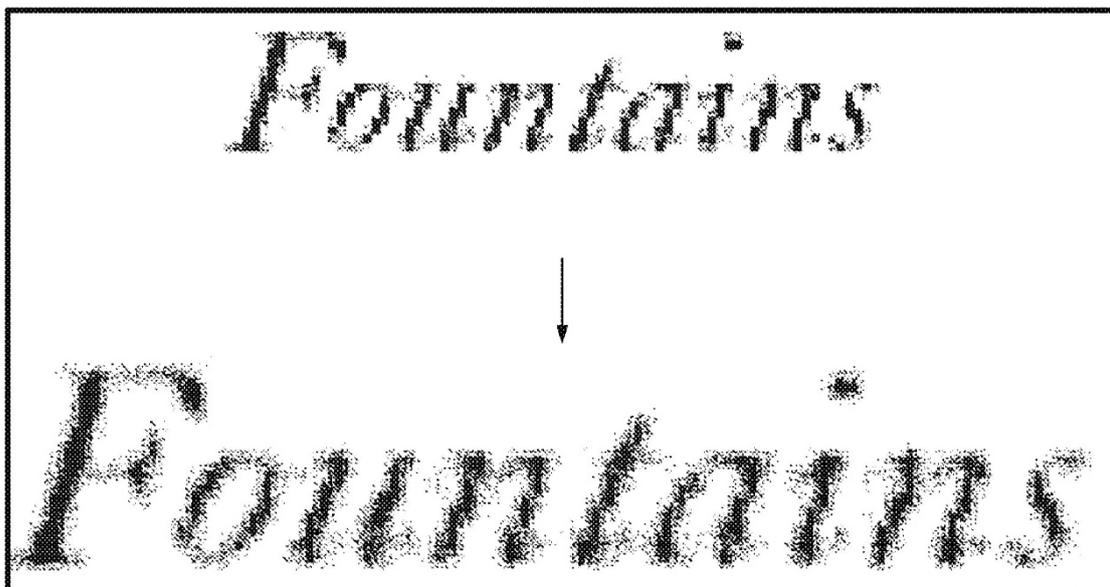


FIG. 18

(Interpolación de la imagen de texto)

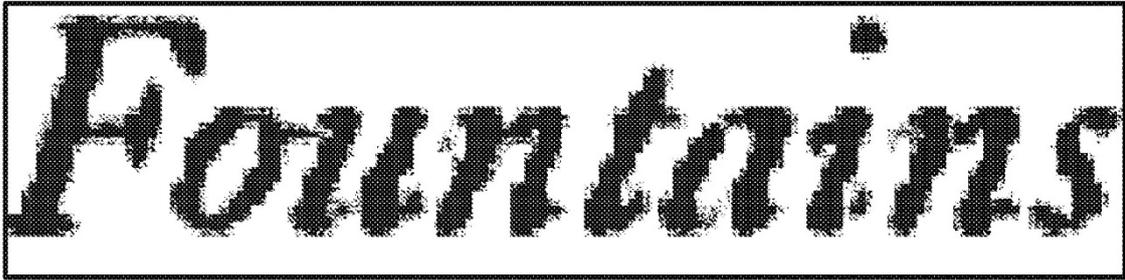


FIG. 19  
(Imagen de texto con resolución de color disminuida)

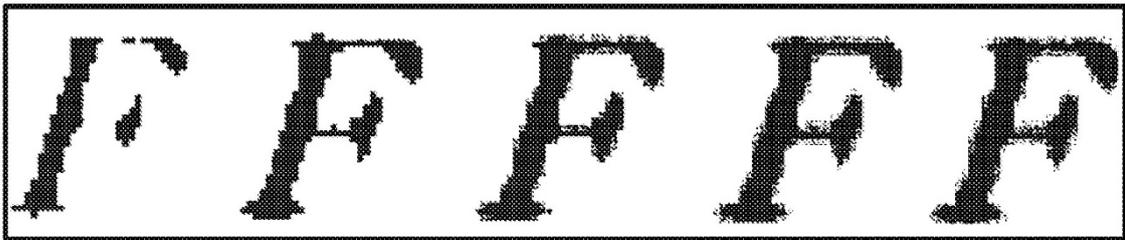


FIG. 20  
(Procedimiento de adición de píxeles)

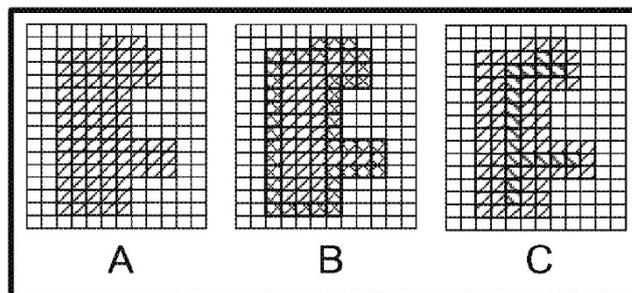


FIG. 21  
(Ilustración del esqueleto)

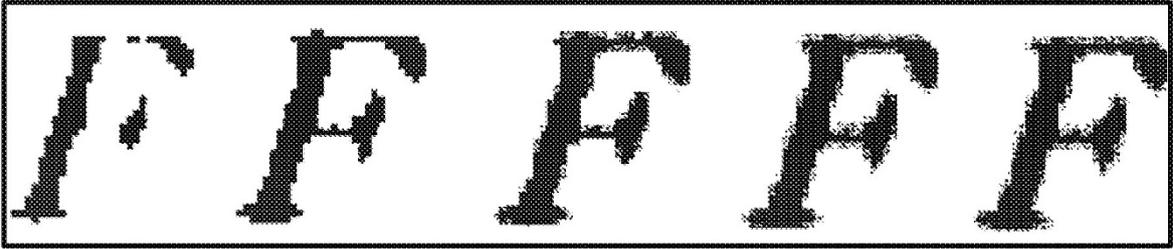


FIG. 22  
(Esqueletos de letra)



FIG. 23  
(Texto mejorado)

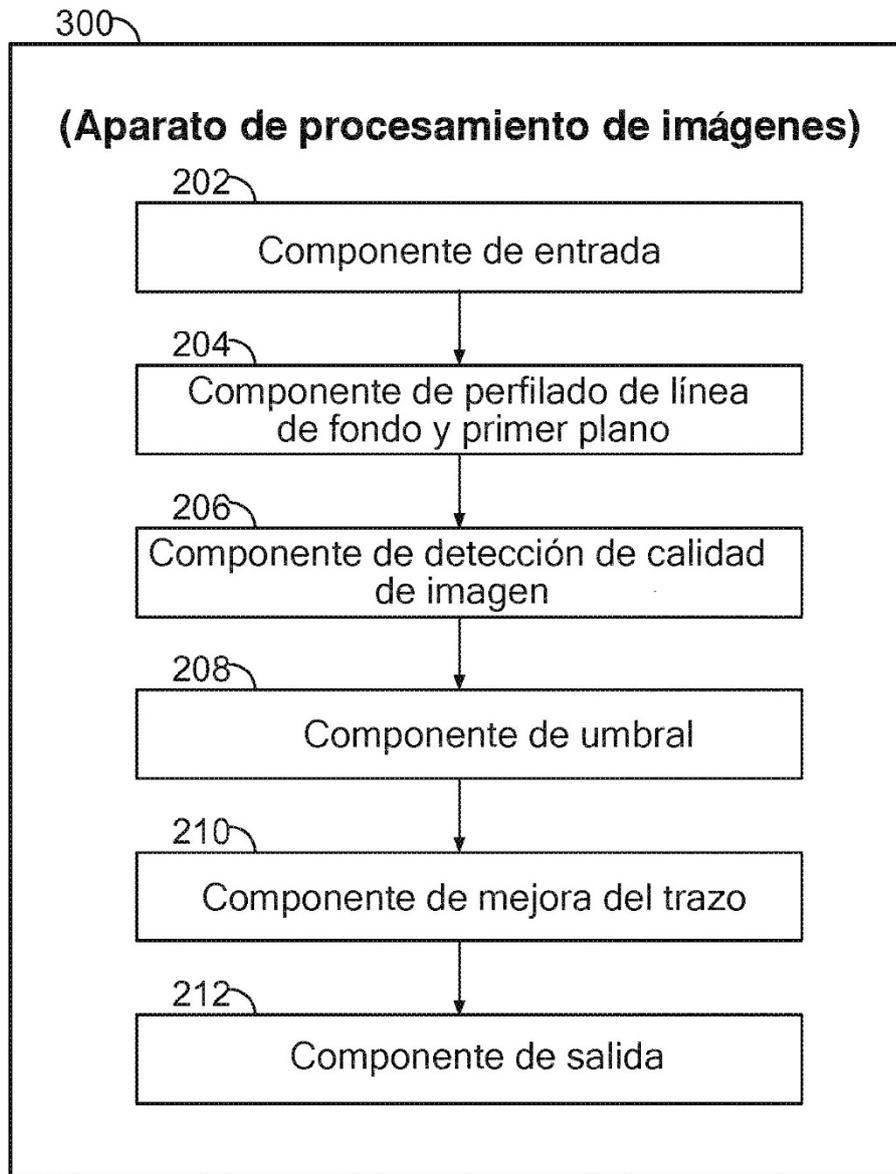


FIG. 24

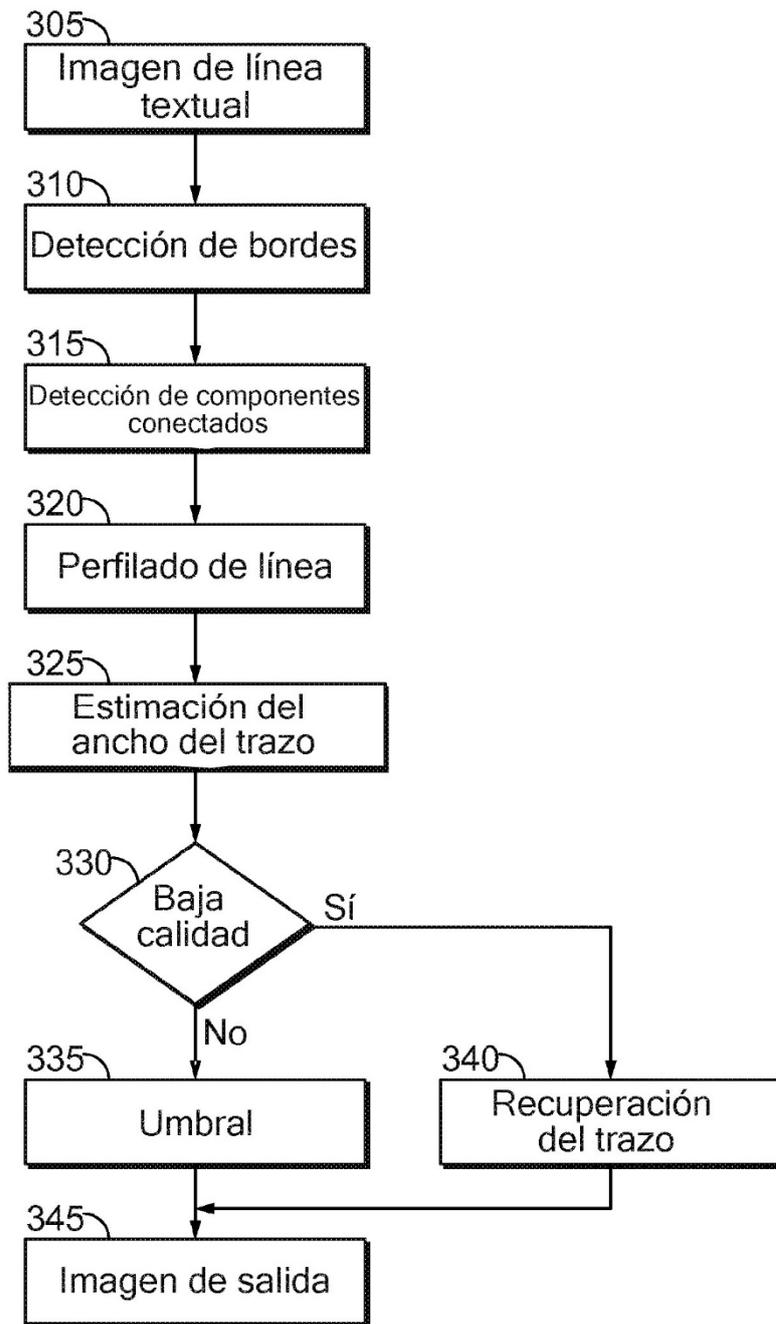


FIG. 25