

Algoritmos de Binarización Robusta de Imágenes con Iluminación No Uniforme

Edgar Molina, Julia Diaz, Hugo Hidalgo-Silva*, Edgar Chávez

CICESE-Ciencias de la Computación, Carretera Ensenada-Tijuana No. 3918, Zona Playitas, C.P. 22860, Ensenada, B.C. México

Resumen

La binarización es una técnica básica de procesamiento digital de imágenes. Los métodos de binarización han sido progresivamente mejorados para tratar con imágenes en condiciones más realistas. De particular importancia es considerar la condición de iluminación no uniforme de las imágenes a binarizar, presente cuando se realiza la adquisición de imágenes con equipo no especializado (e.g. un teléfono celular). En este trabajo se propone un nuevo enfoque para la binarización de imágenes de texto que fueron adquiridas bajo iluminación no uniforme. La idea central es obtener una iluminación más uniforme (apropiada para binarización) a partir de la fusión de dos imágenes iluminadas complementariamente. Para probar el desempeño de nuestra propuesta capturamos un conjunto grande de imágenes de texto reales, obtenidas con un teléfono celular, con diversas condiciones de iluminación no uniforme. Comparamos nuestra propuesta con diversas técnicas del estado del arte y mostramos que tenemos mejores resultados usando distintas métricas de binarización.

Palabras Clave: Binarización, Iluminación, Fusión de Imágenes.

Robust Binarization of Non-Uniform Illuminated Images

Abstract

Image binarization is a basic image processing technique with many applications. Binarization methods have been incrementally improved to deal with real world conditions. A particularly important consideration is non-uniform illumination, present when using general purpose devices for image acquisition (e. g. mobile phones). In this paper we present a new binarization method for non uniformly illuminated text images. The main idea is to improve illumination using two complementary illuminated images. To achieve this we estimate the luminance of the original image and we create a new complementarily illuminated image. We then fuse the two images and for each pixel we estimate in which of the two images the pixel is correctly illuminated. The proposed method is evaluated with synthetic and real images, using several metrics, showing better performance than state-of-the-art techniques.

Keywords: Image binarization, non-uniform illumination, image processing.

1. Introducción

Una de las aplicaciones del procesamiento de imágenes que ha cobrado mayor importancia en las últimas décadas, es la técnica de binarización de imágenes a color o en escala de grises. Esta técnica es comúnmente utilizada cuando se requiere agrupar los píxeles de una imagen en dos tipos de clases: píxeles del fondo y píxeles del primer plano. Visualmente se representan como imágenes digitales en blanco y negro. A pesar de parecer una tarea trivial, en algunas ocasiones puede ser que no lo sea. La presencia de ruido, daños físicos, cambios en la

iluminación, baja resolución, entre otros, son factores que dificultan el proceso.

El método mas simple de binarización es la determinación de umbrales. Este proceso, llamado comúnmente umbralización en castellano, es una transformación puntual que produce imágenes en blanco y negro $B(x, y)$ asignando valores de 1 y 0 a los píxeles de una imagen $I(x, y)$, dependiendo si están por debajo o sobre un valor de umbral T . La obtención del valor de umbral puede hacerse de manera global o de manera local o adaptativa. Discutimos a continuación el trabajo relacionado.

*Autor para correspondencia: hugo@cicese.mx

To cite this article: E. Molina, J. Diaz, Hugo Hidalgo-Silva, Edgar Chavez. 2018. Robust Binarization of Non-Uniform Illuminated Images.

Revista Iberoamericana de Automática e Informática Industrial 15, 252-261. <https://doi.org/10.4995/riai.2017.8847>

Attribution-NonCommercial-NoDerivatives 4,0 International (CC BY-NC-ND 4,0)

La referencia mas antigua para los métodos de binarización es de Otsu (1979), quien propone un método en el que se supone que en la imagen a binarizar se pueden diferenciar claramente el fondo y el primer plano; es decir, que las imágenes tienen histogramas bimodales. El procedimiento consiste en la búsqueda de un umbral global óptimo por medio de la minimización de la varianza intra-clase y la maximización de la varianza inter-clase del histograma de la imagen:

$$\sigma_W^2 = \omega_0(k)\sigma_0^2(k) + \omega_1(k)\sigma_1^2(k), \quad (1)$$

$$\sigma_B^2 = \omega_0(k)(\mu_0 - \mu_T)^2 + \omega_1(k)(\mu_1 - \mu_T)^2, \quad (2)$$

donde σ_W^2 y σ_B^2 son las varianzas intra-clase e inter-clase respectivamente, $\sigma_{0,1}^2(k)$ son las varianzas de las clases 0 y 1, $\omega_{0,1}(k)$ son las probabilidades de ocurrencia de cada clase, μ_0, μ_1 son las medias de cada clase y μ_T es la media total de la imagen.

En su trabajo, Otsu demostró que la minimización de la varianza intra-clase es equivalente a maximizar la varianza inter-clase por medio de un umbral óptimo (k^*). A partir de aquí que la varianza total inter-clase esté dada por

$$\sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{w(k)[1 - w(k)]} \quad (3)$$

donde $\omega(k)$ y $\mu(k)$ son los momentos acumulativos de cero y primer orden del histograma en un nivel (k).

El umbral óptimo (k^*) está dado por el maximizador de

$$\sigma_B^2(k^*) = \max_{1 \leq k \leq L} \sigma_B^2(k), \quad (4)$$

donde L es el valor máximo de niveles de gris. Este método de umbralización es muy popular debido a su fácil implementación y alta eficacia en imágenes en general. Sin embargo, una de sus desventajas es que está pensado para imágenes con histogramas bimodales, con poca degradación, además de ser susceptible a variaciones drásticas en iluminación. A pesar de ello, es uno de los algoritmos más rápidos para binarización de imágenes que existen hoy en día.

Bernsen (1986) propone un algoritmo de umbralización local que trabaja obteniendo el valor de umbral $T(x, y)$ por medio de la información de contraste en vecindarios de píxeles $w(x, y)$, donde (x, y) son las coordenadas de los píxeles en el vecindario. Se obtienen los valores máximo y mínimo de intensidad en cada vecindario de píxeles y el valor de umbral corresponde al promedio de estos dos valores. La imagen es binarizada comparando si la posición de los valores de los píxeles se encuentran por debajo o sobre el valor de umbral en cada vecindario $w(x, y)$. Por lo tanto, el umbral en cada vecindario de píxeles está dado por

$$T(x, y) = \frac{\max(w(x, y)) + \min(w(x, y))}{2}. \quad (5)$$

Niblack (1986) propone un algoritmo que obtiene umbrales locales para cada vecindario de píxeles $w(x, y)$, empleando un vecindario deslizante $w(x, y)$ para obtener la media $\mu(x, y)$ y la varianza $\sigma(x, y)$ en el mismo. Este método además emplea un valor constante k para decidir la cantidad de texto que se encuentra alrededor de cada vecindario deslizante. El valor de umbral $T(x, y)$ en los vecindarios $w(x, y)$ se obtiene por

$$T(x, y) = \mu(x, y) + k \cdot \sigma(x, y) \quad (6)$$

donde $k = -0,2$. Este método falla o decrece su rendimiento en documentos con fondos texturizados o con ligera textura e iluminación no uniforme. Khurshid et al. (2009) presentan otro algoritmo basado en ventanas locales deslizantes, el cual es una versión mejorada del método de Niblack.

Por otro lado, Sauvola and Pietikainen (2000) proponen una mejora al algoritmo de W. Niblack añadiendo una constante R como rango dinámico de la desviación estándar en cada vecindario de píxeles $w(x, y)$. Para esto utilizan la hipótesis de que la distribución de los niveles de gris esta asociada con las clases pertenecientes a fondo y primer plano en las imágenes.

Este método resulta ser más efectivo en la binarización de imágenes severamente degradadas con cambios en la componente de iluminación. El valor del umbral $T(x, y)$ es calculado con

$$T(x, y) = \mu(x, y) \cdot \left[1 + k \left(\frac{\sigma(x, y)}{R} - 1 \right) \right] \quad (7)$$

ellos proponen usar $k = 0,5$ y $R = 128$.

Bradley and Roth (2007) proponen un algoritmo adaptativo de búsqueda de umbral mediante la obtención de la media $\mu(x, y)$ en cada vecindario local $w(x, y)$ de 15×15 con imágenes integrales $I_\Sigma(x, y)$ (Viola et al., 2005). Esto permite la obtención de imágenes binarias en tiempo real. Los umbrales adaptativos permiten hacer uso de las variaciones espaciales de la iluminación presente en las imágenes. El valor de umbral es calculado por medio de la media $\mu(x, y)$ de cada vecindario deslizante $w(x, y)$ y un valor constante $\tau = 15$ que permite calcular el porcentaje de separación del valor de un pixel (x, y) con respecto al promedio de cada vecindario $w(x, y)$. Así el umbral adaptativo está dado por

$$T(x, y) = \mu(x, y) \cdot \left[I_\Sigma(x, y) \left(\frac{100 - \tau}{100} \right) \right]. \quad (8)$$

Singh et al. (2011) proponen una variante del trabajo de Sauvola and Pietikainen (2000), en la cual la desviación estándar $\sigma(x, y)$ es substituida por la desviación media $\delta(x, y)$ y la constante R por el complemento de la desviación media $1 - \delta(x, y)$. Además, emplean una constante $k = 0,06$ para controlar el sesgo y manejar el valor de adaptación del umbral. Este método obtiene una mayor velocidad en la binarización que Sauvola and Pietikainen (2000), el umbral está dado por

$$T(x, y) = \mu(x, y) \cdot \left[1 + k \left(\frac{\delta(x, y)}{1 - \delta(x, y)} - 1 \right) \right]. \quad (9)$$

Es importante notar que este método produce ruido con características similares al ruido *sal y pimienta*.

Los trabajos anteriores no consideran afectaciones en los documentos a binarizar. Con el fin de hacer una comparación más justa con métodos que consideran degradaciones de algún tipo a la imagen del documento, a continuación se describen brevemente los siguientes algoritmos.

En el trabajo de Gatos et al. (2006) se propone una metodología adaptable para la binarización de documentos degradados. Las degradaciones que se consideran son sombras, iluminación no uniforme, bajo contraste, relación señal/ruido grande y difuminado. En su propuesta, aplican un preprocesado mediante un filtro Weiner para estimar el fondo, umbralización mediante la

técnica de Sauvola and Pietikainen (2000) para extraer el fondo, seguido de interpolación para implementar super-muestreo, luego examinan el contraste entre píxeles y finalmente realizan post-procesado. Su aplicación esta dirigida al tratamiento de documentos manuscritos históricos, periódicos antiguos y documentos modernos de baja calidad.

Howe (2013) propone tres estrategias para la binarización de documentos: primero definen la binarización como un problema de etiquetado de píxeles que se resuelve minimizando una función de energía global basada en un modelo de campo aleatorio Markoviano. En seguida, al formular el término de fidelidad para la energía se aplica el Laplaciano de la intensidad de la imagen para distinguirla del fondo. Finalmente se incorporan discontinuidades por bordes en el término de suavidad de la función energía. La funcional de energía es definida como

$$E_I(B) = \sum_{i=0}^M \sum_{j=0}^N [L_{ij}^0(1 - B_{ij}) + L_{ij}^1 B_{ij}] + \sum_{i=0}^{M-1} \sum_{j=0}^N C_{ij}^h(B_{ij} \neq B_{i+1,j}) + \sum_{i=0}^M \sum_{j=0}^{N-1} C_{ij}^v(B_{ij} \neq B_{i,j+1}) \quad (10)$$

donde B etiqueta cada pixel indexado por (i, j) como tinta o fondo, $B_{ij} \in \{0, 1\}$. Los componentes de la energía corresponden a la fidelidad del etiquetado, y otros términos que representan la suavidad o regularidad de la solución. Los costos L_{ij}^0, L_{ij}^1 capturan la cercanía de una etiqueta B_{ij} con su valor en la imagen, y los costos de regularidad C_{ij}^h, C_{ij}^v penalizan a los píxeles que difieren de sus vecinos vertical o horizontal.

Wolf et al. (2002) proponen un sistema para la localización, mejora y binarización de texto en documentos multimedia. La detección se realiza aplicando una medida de gradientes acumulados, los cuales se binarizan mediante el método de Otsu, para después aplicar morfología matemática y con ello reducir el ruido. Una vez detectado el texto, se aplica interpolación bilineal entre marcos de video.

Por otro lado, Feng and Tan (2004) proponen un método adaptable, en el cual, primero se obtiene la media local (m), el mínimo (M), y la desviación estándar (s) de una ventana primaria de tamaño suficiente para cubrir 1-2 caracteres. Para compensar el efecto de la iluminación, la desviación estándar dinámica (R_S) se obtiene sobre una ventana mayor secundaria. Finalmente el umbral se calcula por

$$T = (1 - \alpha_1)m + \alpha_2(s/R_S)(m - M) + \alpha_3M \quad (11)$$

donde $\alpha_2 = k_1(s/R_S)^\gamma$ y $\alpha_3 = k_2(s/R_S)^\gamma$, con α_1, k_1 , y k_2 constantes positivas.

Valizadeh and Kabir (2012) proponen un algoritmo para la binarización de imágenes degradadas. Primero se mapea la imagen en un espacio de características 2D para separar el texto del fondo, para luego particionar este espacio en regiones pequeñas. Esas regiones se etiquetan como texto o fondo usando un algoritmo básico de binarización (Niblack), y luego cada pixel se clasifica como texto o fondo en base a la etiqueta de la región correspondiente en el espacio de características.

2. Metodología

2.1. Formación de Imágenes

Desde un punto de vista físico, una imagen $I(x, y)$ puede ser caracterizada por la multiplicación de dos componentes básicos: la cantidad de iluminación incidente en una escena y la cantidad de iluminación reflejada por los objetos en dicha escena. Estos dos elementos pueden ser llamados *Iluminación* $i(x, y)$ y *Reflectancia* $r(x, y)$ (Gonzalez and Woods, 2002).

$$I(x, y) = i(x, y) \cdot r(x, y) \quad (12)$$

donde la iluminación esta dentro del intervalo $0 < i(x, y) < \infty$ y la reflectancia $0 < r(x, y) < 1$ donde 0 indica una absorción total de luz y 1 una reflexión total de una escena. La naturaleza de la iluminación $i(x, y)$ es determinada por la fuente de luz que incide sobre una escena, mientras que la reflectancia $r(x, y)$ depende de las características de los objetos en las imágenes.

Para poder hacer una correcta binarización de una imagen mal iluminada, es necesario poder obtener una estimación de los componentes de iluminación y reflectancia de una imagen. Esto es indispensable, ya que al trabajar por separado cada componente, es posible realizar operaciones más complejas para normalizar y adecuar la iluminación de una imagen.

2.2. Procesamiento de luminancia en imágenes

Se puede considerar a la luminancia desde el punto de vista óptico, como la luz procedente de una superficie. Sin embargo, cuando trabajamos con imágenes digitales, la luminancia $l(x, y)$ en una imagen en escala de grises nos proporciona información de brillo.

En el método propuesto es vital el cálculo de la luminancia $l(x, y)$ de la imagen para realizar operaciones más complejas. Por medio de la luminancia de la imagen es posible obtener un estimado de la distribución de la no-uniformidad de la iluminación presente en la imagen a binarizar.

De acuerdo con Soille (2004), la aplicación del operador morfológico apertura (\odot) a una imagen no uniformemente iluminada con alto contraste entre los objetos de la misma, permite remover todos los objetos relevantes de su estructura utilizando un elemento estructural grande, preservando la función de iluminación. Por otra parte, si el contraste entre los objetos de interés y el fondo de una imagen decrece, la aplicación de cerradura (\bullet) puede remover dichos objetos y preservar la función de iluminación. La operación cerradura es empleada para obtener una función de iluminación aproximada de la imagen en escala de grises, tal como se muestra en la Figura 1.

$$l(x, y) = I(x, y) \bullet H(x, y) \quad (13)$$

2.3. Estimación de reflectancia

La reflectancia $r(x, y)$ es una de las componentes básicas de una imagen desde el punto de vista óptico. La reflectancia también puede ser vista como una medida de reflexión, siendo una relación entre el flujo de la iluminación reflejada y el flujo de iluminación incidente en una imagen.

Podemos obtener la reflectancia aproximada $r(x, y)$ mediante el cociente de la imagen original $I(x, y)$ y su luminancia $l(x, y)$ en imágenes en escala de grises

$$r(x, y) = \frac{I(x, y)}{l(x, y)} \quad (14)$$

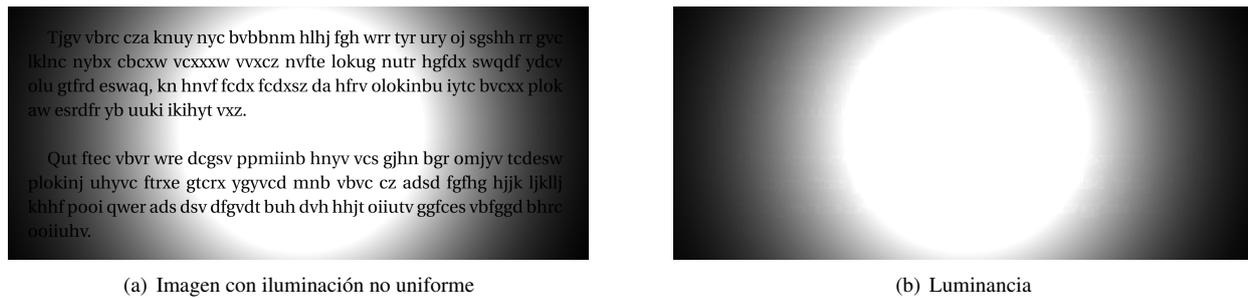


Figura 1: Imagen de texto y su luminancia estimada $l(x, y)$

El cálculo de la reflectancia nos permite obtener las propiedades fotométricas y radiométricas de los objetos que conforman una imagen digital. El obtener esta componente es de suma importancia para el método propuesto, ya que es la encargada de contener la información de reflexión con los que están formados los caracteres de texto de las imágenes de texto.

2.4. Estimación de luminancia complementaria

La hipótesis principal de este trabajo es que las imágenes que no están uniformemente iluminadas pueden ser corregidas si se amplía su rango dinámico. El ojo humano percibe un rango de luminancias mayor que el de los dispositivos de captura en cámaras digitales, puede percibir simultáneamente zonas de penumbra y zonas muy iluminadas sin saturar la imagen. Para obtener resultados parecidos, la técnica de fotografía de alto rango dinámico utiliza dos fotografías; una sobre-iluminada (para distinguir zonas oscuras) y una sub-iluminada (para distinguir zonas claras). Estas dos imágenes se fusionan dejando los píxeles correctamente iluminados en el resultado final. Hay una técnica de menor calidad, llamada de pseudo alto rango dinámico, en donde con una sola fotografía se obtienen dos complementarias manipulando la iluminación relativa de las imágenes. Nuestra técnica sería más parecida a la segunda, como quedará claro más adelante, dado que contamos con una sola imagen; por lo que utilizaremos la iluminación complementaria para obtener una segunda imagen. La idea posterior es luego decidir por cada píxel, cual se encuentra en la imagen más *nítida* utilizando un método de fusión discutido más adelante.

La luminancia inversa $l^{-1}(x, y)$ de la imagen $I(x, y)$ se obtiene invirtiendo el histograma de la función de iluminación $l(x, y)$. Cuando se trabaja con imágenes en escala de grises de 8 bits se tienen imágenes con 256 niveles de cuantización. Por lo tanto, es posible obtener la luminancia complementaria por medio de

$$l^{-1}(x, y) = |255 - l(x, y)|. \quad (15)$$

En la Figura 2(a) puede observarse la iluminación complementaria obtenida de la imagen 1(a).

2.5. Estiramiento y suavizado de la luminancia Inversa

Es importante observar que la luminancia inversa obtenida hasta aquí es una estimación de la función de brillo inversa o luminancia de la imagen original. Es por esta razón que agregamos una etapa de suavizado y estiramiento de contraste para tener una transición suave al iluminar inversamente la imagen

original con esa información de luminancia. Esta etapa será importante a la hora de fusionar las dos imágenes complementarias.

2.6. Imagen inversamente iluminada

Para obtener la imagen inversamente iluminada debemos hacer uso de la reflectancia $r(x, y)$ y la luminancia inversa suavizada $l^{-1}(x, y)$ para formar una imagen con los mismos objetos en la escena, pero con un patrón de iluminación complementario. Este patrón puede ser visto como inverso, ya que lo que se desea es obtener una imagen final con regiones correctamente iluminadas después del proceso de fusión. Siguiendo este enfoque podemos obtener una imagen $I_2(x, y)$ utilizando

$$I_2(x, y) = r(x, y) \cdot l^{-1}(x, y). \quad (16)$$

Una vez realizado este paso obtenemos una imagen en la cual las secciones que en la imagen $I(x, y)$ estaban oscuras, en la imagen resultante $I_2(x, y)$ serán más claras y viceversa.

En la Figura 2(b) se observa la imagen inversa suavizada obtenida para la figura 1(a).

2.7. Fusión de imágenes

En el método propuesto, con los pasos descritos anteriormente, tendremos dos imágenes iluminadas complementariamente. El objetivo es obtener una sola imagen mejor iluminada, sin perder detalles. Para cada píxel, el problema final consiste en decidir si este pertenece al fondo, asignándole el color blanco; o si este pertenece a un objeto y se le asignará el color negro. Una suposición adicional es que un método de binarización de umbral global funciona adecuadamente si la imagen está correctamente iluminada. Lo que haremos a continuación es fusionar las dos imágenes obtenidas, cada píxel de la imagen resultante será escogido de la imagen que tenga mejor respuesta a las altas frecuencias; es decir, lo escogeremos de la imagen que mejor preserve los detalles. Este problema se ha tratado anteriormente en la literatura, en Calderon et al. (2016) se propone un método de fusión de imágenes basado en la respuesta local en frecuencia. El objetivo del trabajo citado de Calderon et al. (2016) es obtener una imagen correctamente enfocada, a partir de dos imágenes enfocadas complementariamente. Para hacer auto-contenido este documento, describimos a continuación el algoritmo desarrollado por Calderon et al. (2016).

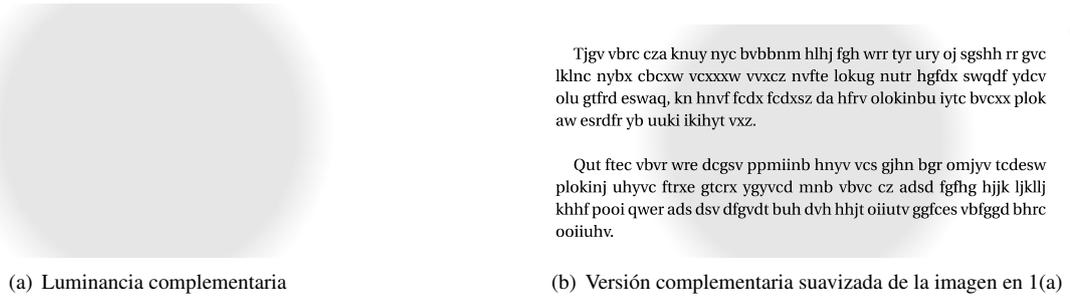


Figura 2: Luminancia complementaria $I^{-1}(x, y)$ de la presentada en la Figura 1 y su versión complementaria suavizada

2.8. Detección de regiones nítidas

Las secciones nítidas de las imágenes pueden ser encontradas localizando las secciones con mayor respuesta a la alta frecuencia. Para esto se aplica el filtro laplaciano. La respuesta a la alta frecuencia F_k de cada imagen fuente es calculada por el valor absoluto de la convolución de cada imagen $I_k(x, y)$ con el kernel Laplaciano h (17) mediante (18).

$$h = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (17)$$

$$\begin{aligned} F_1(x, y) &= |I_1(x, y) \star h(x, y)| \\ F_2(x, y) &= |I_2(x, y) \star h(x, y)| \end{aligned} \quad (18)$$

Una vez obtenidas las respuestas a la alta frecuencia F_k de cada imagen, es necesario obtener la respuesta a la alta frecuencia de la imagen fusionada final. Esto permite encontrar aquellas secciones que conformarán la imagen final. Un pixel nítido solo corresponderá a una sección nítida. Por esto, la respuesta a la alta frecuencia es calculada como la diferencia de las dos respuestas de cada imagen.

$$\Delta F(x, y) = F_1(x, y) - F_2(x, y) . \quad (19)$$

2.9. Función peso binaria

La función peso $p(x, y)$ selecciona los componentes de las imágenes a fusionar, si el valor de $p(x, y) = 1$ en las coordenadas (x, y) de la imagen $I_1(x, y)$, este pixel tiene mayor respuesta a la alta frecuencia que el pixel en las mismas coordenadas para la imagen $I_2(x, y)$. Es decir, es más nítido. El caso contrario ocurre cuando $p(x, y) = 0$, lo cual quiere decir que ese pixel es más nítido en la imagen $I_2(x, y)$. Se propone el uso de ventanas deslizantes de tamaño $w \times w$ para encontrar la función de peso para cada vecindario $w(x, y)$. Al tomar ventanas deslizantes en lugar de tomar directamente los pixeles, se logra que la función de peso satisfaga restricciones de coherencia espacial; es decir, que los los pixeles vecinos comparten niveles de enfoque similar. Es otra manera de establecer que el nivel de nitidez o claridad de los pixeles vecinos está correlacionado.

Para acelerar el proceso anterior se propone el uso de imágenes integrales (Viola et al., 2005). Este método se describe a continuación. El objetivo es calcular el signo de la suma de los valores de respuesta a la alta frecuencia sobre vecindarios de tamaño $w \times w$ utilizando

$$S(x, y) = \sum_{k=-w/2}^{w/2} \sum_{l=-w/2}^{w/2} \Delta F(x+k, y+l) \quad (20)$$

Tjgv vbrc cza knuy nyc bvbnnm hlhj fgh wrr tyr ury oj sgshh rr gvc
klnc nybx cbexw vcxxxw vxzc nvfte lokug nutr hgfdx swqdf ydvc
olu gtrfd eswaq, kn hmvf fedx fedxsz da hfrv olokinbu iytc bvcxx plok
aw esrdfr yb uuki ikihyt vxz.

Qut ftec vbvr wre dcgsv ppmiinb hnyv vcs gjhn bgr omjyv tcdesw
plokiny uhyvc frxe gtrcx ygyved mnb vbvc cz adsd fgfhg hjjk lkllj
khfh pooi qwer ads dsv dfgvdt buh dvh hhjt oiutv ggfcsv vbfegd bhrc
oiuihv.

(b) Versión complementaria suavizada de la imagen en 1(a)

El uso de imágenes integrales puede aumentar considerablemente la velocidad del cálculo de la suma $S(x, y)$ para cada vecindario de pixeles. La complejidad del cálculo de imágenes integrales es de $O(N)$, con N el número de pixeles. La imagen integral para la respuesta de alta frecuencia de la imagen fusionada está dada por

$$\Delta F_{\Sigma} = \sum_{k=0}^x \sum_{l=0}^y \Delta F(k, l) \quad (21)$$

Es posible calcular $S(x, y)$ para cada pixel de la imagen integral ΔF_{Σ} con el uso de ventanas deslizantes utilizando

$$\begin{aligned} S(x, y) &= \Delta F_{\Sigma}(x+w/2, y+w/2) \\ &\quad - \Delta F_{\Sigma}(x+w/2, y-w/2-1) \\ &\quad - \Delta F_{\Sigma}(x-w/2-1, y+w/2) \\ &\quad + \Delta F_{\Sigma}(x-w/2-1, y-w/2-1) \end{aligned} \quad (22)$$

Una vez que se tiene la matriz de pesos binaria, la imagen fusionada $I_F(x, y)$ es construida por medio de la siguiente expresión:

$$I_F(x, y) = I_1(x, y) \cdot p(x, y) + I_2(x, y) \cdot (1 - p(x, y)) \quad (23)$$

En la Figura 3 se presenta el resultado de la fusión de 1(a) con 2(b), y el resultado de aplicar el método de Otsu a esta imagen fusionada.

El pseudocódigo del esquema propuesto se muestra en la Tabla 1.

Tjgv vbrc cza knuy nyc bvbbnm hlhj fgh wrt tyr ury oj sgshh rr gvc
lklnc nybx cbcxw vxxxxw vvxzc nvfte lokug nutr hgfdx swqdf ydcv
olu gtfrd eswaq, kn hnvf fcdx fcdxsx da hfrv olokinbu iytc bvxxx plok
aw esrdfr yb uuki ikihyt vxz.

Qut ftec vbvr wre dcgsv ppmiimb hnyv vcs gjhn bgr omjyv tcdesw
plokiny uhycv ftrxe gtrcx ygyvcd mnb vbvc cz adsd fgfhg hjjk lkklj
khhf pooi qwer ads dsv dfgvdt buh dvh hhjt oiiutv ggfcsv vbfggd bhrc
oiiuhv.

(a) Imagen resultado de la fusión

Tjgv vbrc cza knuy nyc bvbbnm hlhj fgh wrt tyr ury oj sgshh rr gvc
lklnc nybx cbcxw vxxxxw vvxzc nvfte lokug nutr hgfdx swqdf ydcv
olu gtfrd eswaq, kn hnvf fcdx fcdxsx da hfrv olokinbu iytc bvxxx plok
aw esrdfr yb uuki ikihyt vxz.

Qut ftec vbvr wre dcgsv ppmiimb hnyv vcs gjhn bgr omjyv tcdesw
plokiny uhycv ftrxe gtrcx ygyvcd mnb vbvc cz adsd fgfhg hjjk lkklj
khhf pooi qwer ads dsv dfgvdt buh dvh hhjt oiiutv ggfcsv vbfggd bhrc
oiiuhv.

(b) Resultado de aplicar el método de Otsu

Figura 3: (a) Resultado después de obtener la fusión de 1(a) con 2(b), y (b) aplicando el umbral obtenido por Otsu al resultado de la fusión

Tabla 1: Algoritmo Propuesto

Algoritmo Propuesto
Entrada: $I_1(x, y), w$
Calcular $l(x, y)$ con (13)
Calcular $r(x, y)$ con (14)
Calcular $l^{-1}(x, y)$ con (15)
Construir $I_2(x, y)$ con (16)
Calcular $F_1(x, y)$ y $F_2(x, y)$ con (18)
Calcular $\Delta F(x, y) = F_1(x, y) - F_2(x, y)$
Calcular $\Delta F_{\Sigma}(x, y)$ con (21)
para $x = 0$ hasta $n_c - 1$ hacer
para $y = 0$ hasta $n_c - 1$ hacer
Calcular $S(x, y)$ con (22)
si $S(x, y) \geq 0$ entonces
$p(x, y) = 1$
si no
$p(x, y) = 0$
fin si
fin para
fin para
Construir $I_F(x, y)$ con (23)
Binarizar $I_F(x, y)$ utilizando un método tradicional
Salida: $I_F(x, y)$

3. Métricas de Evaluación

La evaluación cuantitativa de la calidad de binarización del algoritmo propuesto comparada con los algoritmos mencionados en este trabajo (Pratikakis and Gatos, 2010, 2013) se presenta usando diferentes métricas. La primera métrica de evaluación fue la medida F , la cual es definida en función de los píxeles clasificados como falsos positivos (FP), falsos negativos (FN) y verdaderos positivos (VP) en una imagen binaria.

Esta métrica es obtenida por medio de la relación de la *exhaustividad* (E) y la *precisión* (P) con el uso de los valores correctos de cada píxel, representados en $GT(x, y)$

$$E = \frac{VP}{VP + FN} \quad (24)$$

$$P = \frac{VP}{VP + FP} \quad (25)$$

Por lo cual la medida F se define como

$$F = \frac{2 * E * P}{E + P} \quad (26)$$

La medida F mide el porcentaje de acierto en la clasificación de los píxeles de acuerdo a la clase a la que pertenecen. Un número o valor cercano a uno en esta métrica indica que el algoritmo de binarización obtiene una mayor aproximación al $GT(x, y)$ de cada imagen y por lo tanto, obtiene una mejor calidad de binarización.

La segunda métrica que se tomó en cuenta es el pico de la relación señal-a-ruido ($PSNR$) (Lu et al., 2004), la cual indica la relación de similitud entre dos imágenes de tamaño $M \times N$, definida como

$$PSNR = 10 \log_{10} \frac{C^2 MN}{\sum_{x=1}^M \sum_{y=1}^N (B(x, y) - GT(x, y))^2} \quad (27)$$

con C la máxima variación de la señal, (255 para imágenes de 8 bits).

Para binarización de imágenes, un valor elevado de $PSNR$ indica una mayor similitud entre la imagen binaria $B(x, y)$ obtenida por un algoritmo de binarización y el $GT(x, y)$.

La tercera medida de evaluación es la métrica de tasa negativa (NRM), la cual evalúa los errores de empatamiento entre la imagen binaria $B(x, y)$ y su $GT(x, y)$. Esta métrica combina el rango de falsos positivos NR_{FP} y el rango de falsos negativos NR_{FN} , expresados como

$$NR_{FP} = \frac{FP}{FP + VN} \quad (28)$$

$$NR_{FN} = \frac{FN}{FN + VP} \quad (29)$$

La NRM está dada por

$$NRM = \frac{NR_{FN} + NR_{FP}}{2} \quad (30)$$

En esta métrica, se observa que entre más cercanos se encuentren los resultados a cero, la calidad de la binarización es mayor. La cuarta métrica que se empleó fue la métrica de distorsión recíproca de distancia (DRD) (Lu et al., 2004). Esta métrica esta asociada con la medición de la distorsión visual que el humano puede apreciar en documentos de texto binarios. En este sentido, esta métrica asocia y evalúa la percepción del ser humano en un conjunto S de píxeles invertidos en un documento binario

$$DRD = \frac{\sum_{k=1}^S DRD_k}{NUBN} \quad (31)$$

donde $NUBN$ estima el área válida (no vacía) en la imagen, definida como el número de bloques 8×8 no uniformes en $GT(x, y)$.

La distorsión DRD_k del k -ésimo pixel distorsionado representa la suma de los pesos de los pixeles presentes en un vecindario $w(x, y)$ de tamaño 5×5 dentro del $GT(x, y)$ de la imagen, la cual matemáticamente se representa por

$$DRD_k = \sum_{i=-2}^2 \sum_{j=-2}^2 |GT_k - B_k(x, y)| \cdot w(x, y) \quad (32)$$

donde este conjunto de pixeles difieren del k -ésimo pixel distorsionado dentro de la imagen binaria $I(x, y)$. El rendimiento de un algoritmo de binarización es mejor cuanto más baja o cercana a cero se encuentre el resultado de la DRD .

4. Experimentos y resultados

Para poder evaluar la calidad de la binarización utilizamos un arreglo experimental en el que pudimos establecer con exactitud una línea base (ground truth) muy precisa. La primera parte de los experimentos se realizaron con imágenes sintéticas con diferentes patrones de iluminación, también sintéticos. Esto permite saber con precisión como debe ser clasificado un pixel, independientemente de la aplicación en la que estará inmersa la técnica.

Jsdjfsdjffgig fvnfkddjdj jhdfguydsfbsj dgudfgcejg ufdki
Dksdhd fjdgf ugyfusdgy fusbgsbes dgfavhgt yftywq gulli
Jdj hsdfguysyd fgsjfhgs dkfhsd jfhgsdth dkfhs dkfhiefi
jkjhh fhdshtfr fqtfbwbw jheudhdu hyefg hdjdfhj dffji

Sdhg sdjhhgq uwdm enivkhydygd bdsfjdj fufjkwgl
Yuejed boiyuywgdw wdgveydgude uydgduydhddhdfi
Eifghgfugdftu dfuihdfidufowu wuwymkzkb wetskij

Figura 4: Imagen usada para generar las imágenes sintéticas de prueba.

Las imágenes empleadas fueron creadas a partir de la imagen mostrada en la Figura 4. Se aplicó degradación con iluminación no uniforme sintéticamente con la herramienta comercial *Adobe Photoshop CS6* usando diferentes niveles de intensidad y exposición para obtener las imágenes de prueba mostradas en la Figura 5.

Se aplicaron los algoritmos de los métodos de binarización tradicionales, incluyendo el método de Otsu, Bernsen, Bradley, Singh, Niblack, Sauvola y Kurshid. Además se implementaron los métodos que consideran degradaciones en la imagen como Howe, Gatos, Wolf, Feng y Valizadeh. Todos los algoritmos fueron programados en C++ usando Visual studio y la biblioteca OpenCV2.4.12, exceptuando el método de Howe, para el cual se utilizó el programa en *Matlab* proporcionado por el autor. Las Figuras 6 y 7 muestran los resultados de la evaluación en términos de las métricas medida F, PSNR, NRM y DRD, en imágenes sintéticas utilizando sólo los métodos tradicionales y los obtenidos aplicando el método propuesto seguido de un binarizador.

De las Figuras 6 y 7 se puede observar que los mejores valores en las métricas de evaluación se obtienen utilizando el método propuesto seguido de una técnica de binarización tradicional. En la tabla 2 se muestran los tiempos promedio de aplicar los métodos a las 15 imágenes sintéticas, usando una computadora Dell Precision T3610 con procesador Intel Xeon E5-1603 a

2.8GHz y 16 GB de memoria RAM. De la tabla podemos concluir que es claro que la propuesta seguida de una técnica de binarización toma más tiempo que el binarizador por sí solo, sin embargo su tiempo de ejecución es mucho menor que los métodos que consideran degradaciones.

Tabla 2: Tiempos de ejecución promedio de los algoritmos aplicados a las imágenes en la Figura 5.

Método	tiempo (s)
Otsu	0.0004
Bradley	0.004667
Niblack	0.005733
Sauvola	0.00733
Singh	0.0091
Bernsen	0.4872
Feng	0.092
Kurshid	0.0062
Gatos	0.035
Valizadeh	5.28
Wolf	0.0059
Howe	0.2159
Prop+Otsu	0.0172
Prop+Bradley	0.022
Prop+Gatos	0.0527

Adicionalmente a las imágenes sintéticas, se utilizaron imágenes reales impresas en una impresora laser y capturadas con un teléfono móvil, con una resolución de 2.4Mpixeles e iluminación no uniforme de ambiente natural.

La suavización de la luminancia complementaria fue realizada con la función *imadjust* de Matlab. Para todas las figuras sintéticas se analizaron los valores cuantitativos de cada una de las métricas empleadas.

En la Figura 8 se observan varias imágenes tomadas bajo diferentes condiciones de iluminación no uniforme con la cámara digital de un teléfono móvil. En las Figuras 9 y 10 se muestran los resultados de los algoritmos para las diferentes imágenes reales. En este caso la iluminación que afecta al texto presente en la imagen es ocasionada directamente por la luz ambiental y algunas obstrucciones y degradaciones leves en la textura de la hoja que sirve de imagen. Esto hace que se produzcan ligeras variaciones en la iluminación, lo cual es hasta cierto punto habitual encontrar en la práctica al capturar imágenes y documentos directamente.

En este caso, los resultados de la inspección visual muestran que los esquemas propuestos tienen buen tratamiento de la iluminación y textura de la imagen. Los resultados de las Figuras 9 y 10 muestran que los mejores resultados se obtienen usando la propuesta seguida del binarizador de Gatos et. al..

En la Figura 11 se presentan los resultados para una de las imágenes reales mostrada en la Figura 8. Los resultados corresponden a los binarizadores de Otsu, Bradley, Sauvola, y los métodos que consideran degradación de algún tipo como Gatos y Howe, además del método propuesto seguida de binarizadores Otsu, Bradley y Gatos. De la Figura 11 se puede observar que el mejor resultado es obtenido por el método propuesto en combinación con el método de Gatos. Sin embargo, al aplicar el binarizador de Gatos por sí solo, no se obtienen resultados



Figura 5: Imágenes sintéticas generadas a partir de la imagen de texto mostrada en la Figura 4.

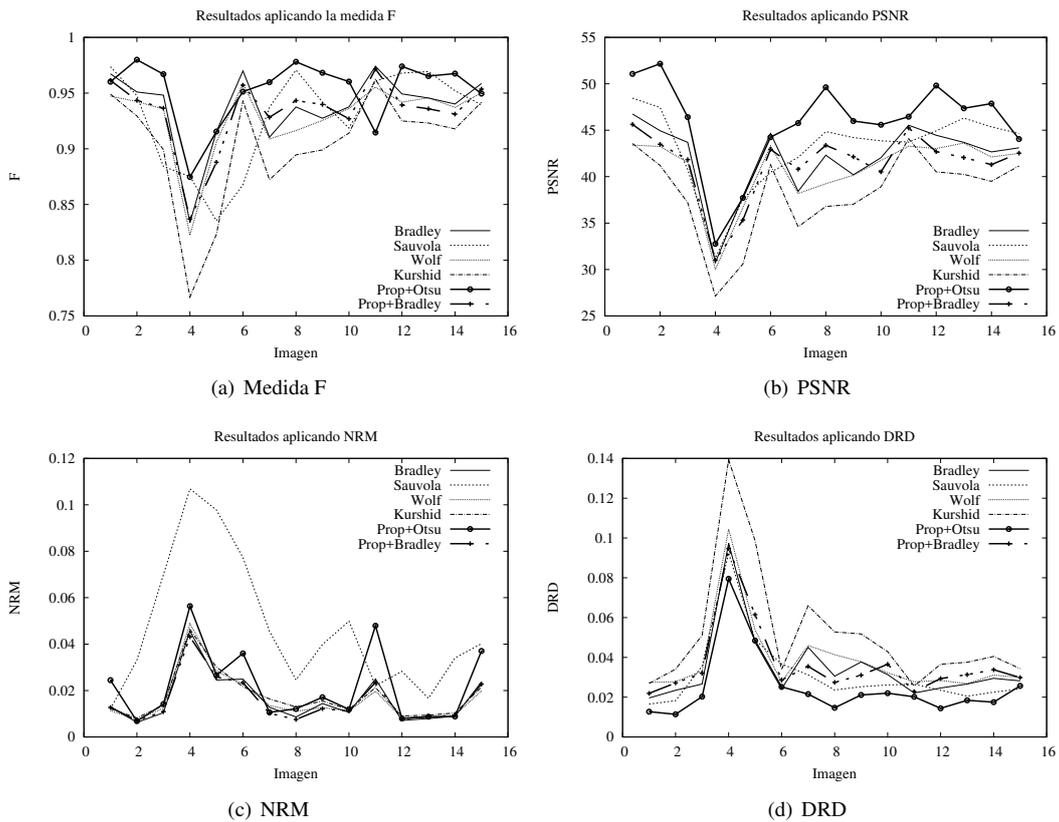


Figura 6: Mejores resultados de binarización de imágenes sintéticas

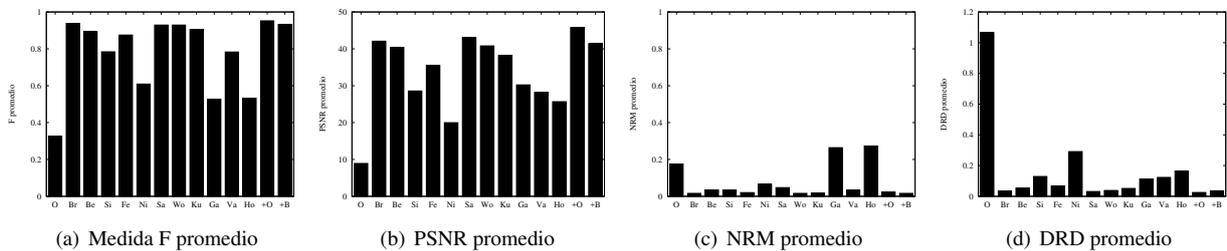


Figura 7: Resultados promedio en las métricas de evaluación al binarizar imágenes sintéticas, O=Otsu, Br=Bradley, Be=Bernsen, Si=Singh, Fe=Feng, Ni=Niblack, Sa=Sauvola, Wo=Wolf, Ku=Kurshid, Ga=Gatos, Va=Valizadeh, Ho=Howe, +O=propuesta + Otsu, +B=propuesta+Bradley.

acceptables como se puede observar en la Figura 11e. El método de Howe tampoco es capaz de resolver el problema de degradación por iluminación no uniforme.

5. Conclusiones

Hemos mostrado de manera experimental, dando además una justificación teórica, que es posible binarizar adecuadamente imágenes de texto con iluminación no uniforme y con degra-

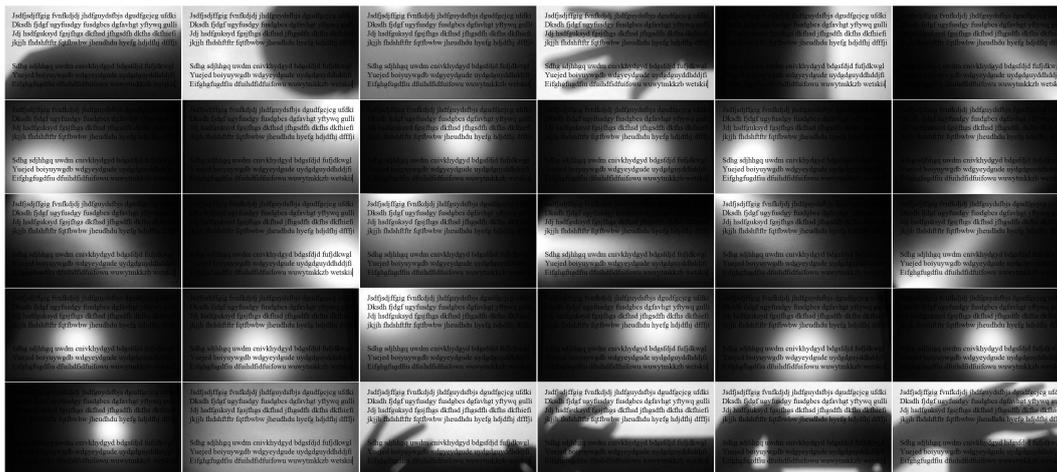


Figura 8: Imágenes reales tomadas con una cámara digital.

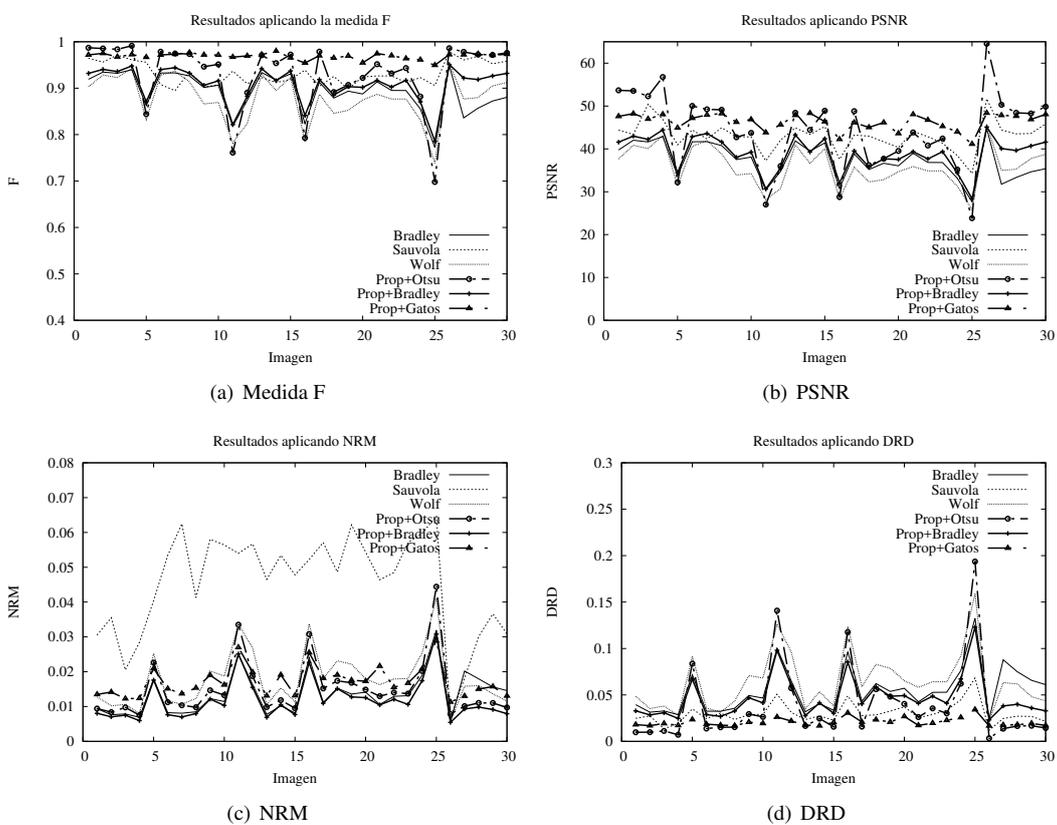


Figura 9: Mejores resultados de binarización de imágenes reales

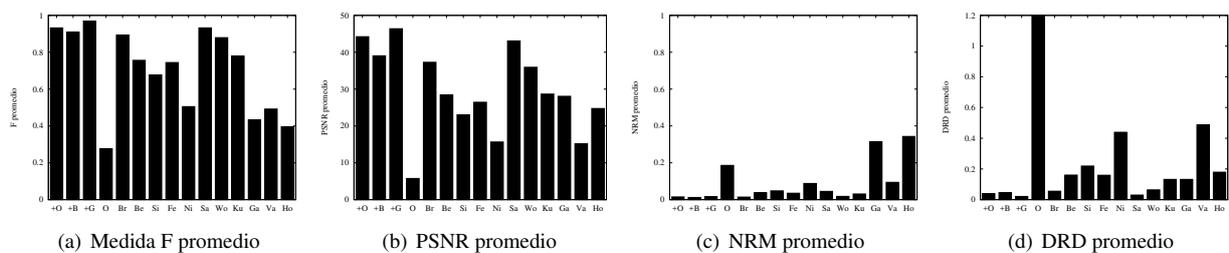


Figura 10: Resultados promedio en las métricas de evaluación al binarizar imágenes reales, O=Otsu, Br=Bradley, Be=Bernsen, Si=Singh, Fe=Feng, Ni=Niblack, Sa=Sauvola, Wo=Wolf, Ku=Kurshid, Ga=Gatos, Va=Valizadeh, Ho=Howe, +O=propuesta + Otsu, +B=propuesta+Bradley.

daciones ligeras en textura. Mostramos que estimando la iluminación y la reflectancia mediante el filtro de cerradura, es posi-

nación y la reflectancia mediante el filtro de cerradura, es posi-



Figura 11: Resultados de binarización de imagen real

ble iluminar de manera complementaria una imagen, de tal modo que cada pixel está correctamente iluminado en la imagen original o en la imagen complementaria. Para decidir en cual de las dos imágenes se encuentra el pixel correcto, adaptamos un método de fusión de imágenes recientemente publicado. La comparación de los esquemas propuestos con los del estado del arte, muestra la eficiencia del esquema propuesto en imágenes con condiciones extremas de iluminación.

Agradecimientos

Este trabajo ha sido realizado parcialmente gracias al apoyo del Consejo Nacional de Ciencia y Tecnología (CONACyT), con el proyecto 179795-Y. E. Molina y J. Diaz fueron financiados por CONACyT.

Referencias

Bernsen, J., 1986. Dynamic thresholding of grey-level images. Proceedings of International Conference on Pattern Recognition, 1251–1255.
 Bradley, D., Roth, G., 2007. Adaptive thresholding using the integral image. Journal of Graphics, GPU and Game Tools 12 (2), 13–21.
 Calderon, F., Garnica-Carrillo, A., Flores, J. J., 2016. Fusión de imágenes multi foco basado en la combinación lineal de imágenes utilizando imágenes incrementales. Revista Iberoamericana de Automática e Informática Industrial RIAI 13 (4), 450–461.
 Feng, M.-L., Tan, Y.-P., 2004. Adaptive binarization method for document image analysis. In: Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on. Vol. 1. IEEE, pp. 339–342.

Gatos, B., Pratikakis, I., Perantonis, S. J., 2006. Adaptive degraded document image binarization. Pattern recognition 39 (3), 317–327.
 Gonzalez, R., Woods, R., 2002. Digital image processing. Pearson Education.
 Howe, N. R., 2013. Document binarization with automatic parameter tuning. International Journal on Document Analysis and Recognition (IJ DAR) 16 (3), 247–258.
 Khurshid, K., Siddiqi, I., Faure, C., Vincent, N., 2009. Comparison of niblack inspired binarization methods for ancient documents. DRR 7247, 1–10.
 Lu, H., Kot, A. C., Shi, Y. Q., 2004. Distance-reciprocal distortion measure for binary document images. IEEE Signal Processing Letters 11 (2), 228–231.
 Niblack, W., 1986. An Introduction to Digital Image Processing. Prentice-Hall.
 Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on System, Man and Cybernetics 6 (3), 62–66.
 Pratikakis, I., Gatos, Basilis. Konstantinos, N., 2010. H-dibco 2010 - handwritten document image binarization competition. International Conference on Frontiers in Handwriting Recognition, 727–732.
 Pratikakis, I., Gatos, Basilis. Konstantinos, N., 2013. Icdar 2013 document image binarization contest (dibco 2013). International Conference on Document Analysis and Recognition, 1471–1476.
 Sauvola, J., Pietikainen, M., 2000. Adaptive document image binarization. Pattern Recognition 33, 225–236.
 Singh, T., Roy, Sudipta. Singh, O., Sinam, T., Singh, M., 2011. A new local adaptive thresholding technique in binarization. IJCSI International Journal of Computer Science Issues 8 (6), 271–277.
 Soille, P., 2004. Morphological Image Analysis: Principles and Applications. Springer.
 Valizadeh, M., Kabir, E., 2012. Binarization of degraded document image based on feature space partitioning and classification. International Journal on Document Analysis and Recognition 15 (1), 57–69.
 Viola, P., Jones, M. J., Snow, D., 2005. Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision 63 (2), 153–161.
 Wolf, C., Jolion, J.-M., Chassaing, F., 2002. Text localization, enhancement and binarization in multimedia documents. In: Pattern Recognition, 2002. Proceedings. 16th International Conference on. Vol. 2. IEEE, pp. 1037–1040.