

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE CIENCIAS



**DESCRIPTOR DE IMÁGENES INVARIANTE A POSICIÓN,
ROTACIÓN Y ESCALA, BASADO
EN FIRMAS 1D: CIRCULAR, RADIAL Y MARGINAL POR COLUMNA**

TESIS
QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA:
ALDO ANTONIO MICHEL VILLALOBOS

Ensenada, Baja California, México. Junio 2016

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE CIENCIAS

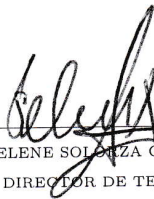
DESCRIPTOR DE IMÁGENES INVARIANTE A POSICIÓN,
ROTACIÓN Y ESCALA, BASADO EN FIRMAS 1D: CIRCULAR,
RADIAL Y MARGINAL POR COLUMNA

TESIS PROFESIONAL

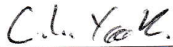
QUE PRESENTA

ALDO ANTONIO MICHEL VILLALOBOS

Y APROBADA POR EL SIGUIENTE COMITÉ:



DRA. SELENE SOLÓRZANO CALDERÓN
DIRECTOR DE TESIS



DR. CARLOS YEE ROMERO
SINODAL



M.C. ADINA JORDANA ARÁMBURO
SINODAL

*Para todos aquellos que no se dieron por
vencidos conmigo...*

Agradecimientos

No hay mayor agradecimiento que el que le debo a la Dra. Selene Solorza Calderón, si no fuera por ella no habría logrado desarrollar, ni terminar este trabajo de tesis. Me brindo más paciencia de la que merezco, por lo cual siempre le agradeceré. Su dedicación, determinación y esfuerzo a su trabajo destacan, es una persona admirable y respetable.

A mis padres, José Antonio y Rosa Adriana que me apoyaron en todo momento, y por el amor y paciencia constante que me brindan.

A mi hermana Claudia Adriana, por ayudarme en los peores de los momentos.

A mis profesores y compañeros que me ayudaron a aprender y crecer, tanto como persona como en el sentido académico.

A mis amigos y amigas por compartir su tiempo, conocimiento y amor.

Al CONACyT por la beca otorgada para la realización de esta tesis bajo el proyecto de Ciencia Básica SEP-CONACyT 2011, con número 169174 y titulado reconocimiento de patrones en imágenes digitales mediante máscaras binarias de anillos concéntricos y firmas unidimensionales invariantes a posición, rotación y escalamiento.

RESUMEN de la Tesis de **Aldo Antonio Michel Villalobos** presentada como requisito parcial para la obtención del grado de Licenciado en Matemáticas Aplicadas. Ensenada, Baja California, México, Junio 2016.

**DESCRIPTOR DE IMÁGENES INVARIANTE A POSICIÓN,
ROTACIÓN Y ESCALA, BASADO EN FIRMAS 1D: CIRCULAR,
RADIAL Y MARGINAL POR COLUMNA**

En este trabajo se presenta un descriptor de imágenes invariante a rotación, escala y traslación (RST) basado en firmas 1D. El descriptor es invariante a traslación al utilizar el espectro de amplitud de la transformada de Fourier de la imagen. Ese espectro de amplitud de Fourier es introducido en la transformada analítica de Fourier-Mellin (TAFM) para obtener la invariancia a escala. Del espectro de amplitud normalizado de la TAFM se construyen tres firmas 1D: una firma circular obtenida a partir de máscaras binarias de trayectorias circulares con las que se filtra el espectro; una firma radial generada filtrando el espectro con filtros binarios de trayectorias radiales; y para la tercer firma se usan los marginales de las columnas del espectro de la TAFM normalizada, por lo que se le ha denominado firma columna. El sistema trabaja solamente con imágenes en blanco y negro y en escala de grises. Para determinar el nivel de confianza del sistema se usa la metodología de espacios de clasificación de ortoedros de confianza del 95.4%.

Resumen aprobado por:


DRA. SELENE SOLÓRZUA CALDERÓN

Contenido

	Página
Agradecimientos	I
Resumen	II
Contenido	III
I. Introducción	1
II. Objetivo	5
III. Antecedentes	6
IV. Sistema de reconocimiento de patrones invariante a traslación, escala y rotación	8
IV.1 La firma 1D circular	14
IV.2 La firma 1D radial	16
IV.3 La firma 1D marginal por columna	19
V. Ortoedro de confianza para el descriptor invariante a traslación, escala y rotación	21
V.1 Imágenes en blanco y negro	22
V.2 Imágenes en escala de grises	26
VI. Conclusiones	29

Capítulo I

Introducción

En estos días, con el gran avance en la tecnología, el reconocimiento de patrones en imágenes digitales es un área muy productiva. Sin embargo, el proceso de extracción de características para generar un descriptor invariante a transformaciones geométricas del objeto (traslación, rotación, escala, entre otros) no es un problema trivial, por lo que se han desarrollado diversas metodologías de extracción de características en imágenes digitales, una de las más utilizadas son las basadas en funciones de correlación. Un descriptor de una imagen es una formulación matemática que representa a una imagen por medio de sus características elementales, como son la forma, el color, textura, entre otros. Debe contener suficiente información como para representar a esa imagen, independientemente de las transformaciones o distorsiones geométricas que presenten los objetos. En 2012, T.V. Hoang y colaboradores[1] utilizaron la transformada de Radon y la transformada 1D de Fourier-Mellin para construir un descriptor 2D invariante a traslación, escala y rotación, en el cual la comparación entre imágenes la realizaron mediante la correlación cruzada de la imagen de referencia y la imagen problema, es decir una correlación 2D. Como la transformada de Radon genera un desplazamiento circular en la variable angular, tuvieron que realizar 180 correlaciones cruzadas 2D de la imagen problema con la imagen de referencia (una correlación 2D para cada $\theta = 0, \dots, 179$), por lo que se emplea mucho tiempo computacional en el proceso de comparación. Por ejemplo, si la base de datos de imágenes de referencia tiene 100 muestras, entonces se realizarán $(180)(100) = 18,000$ correlaciones 2D solamente en la etapa de comparación de la imagen problema con las imágenes de referencia. En 2001, S. Derrode

y colaboradores[2] propusieron otro tipo de descriptores usando la transformada 2D de Fourier-Mellin (TFM), pero debido al factor $\frac{1}{r}$ en esta transformada, la invariancia en traslación generalmente la obtienen en el dominio espacial usando el centroide o el centro de masa de los objetos[2]. Una práctica que se llevó a cabo para tratar de reducir el efecto del factor $\frac{1}{r}$, el cual le da mayor peso a los datos alrededor del centroide (o centro de masa), es eliminar un disco de datos de radio constante y con origen en el centroide (o centro de masa). El problema de dicha acción radica, en que al no tener control de la escala de la imagen se podrían estar eliminando datos relevantes. Para resolver dicho problema, F. Ghorbel[3] propone la transformada analítica de Fourier-Mellin (TAFM) donde las imágenes son ponderadas por el factor r^σ , con $\sigma > 0$. Sin embargo, esta transformada no mantiene la invariancia a escala, pero normalizando apropiadamente la TAFM, se recupera dicha invariancia.

Recientemente se han estado desarrollando sistemas de reconocimiento de patrones basados en máscaras binarias de anillos concéntricos[4-8]. Estas metodologías son robustas y eficientes en el reconocimiento de patrones para imágenes en escala de grises, independientemente de la posición, rotación y escala que el objeto presente. Además, la respuesta de estos sistemas es muy buena bajo iluminación no homogénea y ruido. En 2015, A. Solís-Ventura y colaboradores[7] obtienen la invariancia a escala por medio de la transformada 2D de escala no separable. Esta transformada 2D no es invariante bajo traslación, por lo que el centro de masa del objeto es utilizado como el centro de la imagen. En 2016, C. Barajas-García y colaboradores[8] utilizan la transformada analítica de Fourier Mellin normalizada para la invariancia a escala, lo que les permite utilizar el espectro de amplitud de Fourier para la invariancia a traslación. Ambas metodologías funcionan con firmas 1D construidas por medio de máscaras binarias de anillos concéntricos. El uso de firmas 1D reduce considerablemente el tiempo computa-

cional invertido en el momento de la comparación de las firmas mediante la función de correlación 1D.

En este trabajo de tesis se propone un sistema de reconocimiento de patrones que utiliza tres firmas 1D invariantes a traslación, escala y rotación. Además del uso de la metodología de espacios de clasificación por medio de ortoedros, que permite construir un sólo espacio de salida en vez de los múltiples planos generados con los descriptores de correlación. El sistema digital de reconocimiento de patrones utiliza el espectro de amplitud de la transformada de Fourier de la imagen para obtener la invariancia a traslación. Luego, dicho espectro de amplitud es introducido en la transformada analítica de Fourier-Mellin normalizada (TAFMN), esto con el fin de lograr que el espectro de amplitud de la TAFMN sea invariante a posición y escala. Después, éste se filtra con máscaras binarias circulares para construir la firma 1D circular invariante a traslación, escala y rotación. La segunda firma 1D invariante a traslación, escala y rotación, denominada firma radial, se obtiene al filtrar el espectro de la TAFMN con máscaras binarias radiales. Una tercera firma 1D, invariante a traslación, escala y rotación, se genera mediante los marginales por columna de la imagen del espectro de la TAFMN, por lo que se le ha llamado firma marginal. El espacio de clasificación de ortoedros se construye usando el intervalo de confianza del 95.4% de los valores de las potencias de las firmas: circular, radial y marginal, de cada imagen de entrenamiento. Al tener un sólo espacio de clasificación en el que los valores de las potencias de las firmas de las imágenes están localizadas en una región específica, se reduce considerablemente el costo computacional en el paso de la clasificación.

Este trabajo de tesis está organizado como sigue: en el Capítulo II se presenta el objetivo de esta tesis. En el Capítulo III se definen los conceptos básicos de imagen

digital. En el Capítulo IV se explican los fundamentos matemáticos del sistema de reconocimiento de patrones invariante a traslación, escala y rotación. En el Capítulo V se expone el procedimiento para construir el espacio de clasificación de ortoedros con un nivel de confianza del 95.4%. Finalmente, se presentan las conclusiones en el Capítulo VI.

Capítulo II

Objetivo

Desarrollar e implementar en lenguaje de programación MATLAB un sistema de reconocimiento de patrones en imágenes digitales invariante a traslación, escala y rotación basado en las firmas circular, radial y marginal.

Capítulo III

Antecedentes

La información contenida en imágenes puede ser representada en formas muy distintas, las más importantes son: la espacial y la de número de onda. Dichas representaciones son completas y equivalentes, por lo que es posible pasar de una representación a otra mediante la transformada de Fourier[11].

La representación espacial de una imagen se define como una función $I(x, y)$, en donde x e y son coordenadas espaciales, y al valor de esta función I en cualquier par de coordenadas (x, y) se le llama la intensidad en ese punto[9]. En forma de ecuación, una imagen digital es representada por una matriz $M \times N$ como,

$$I(x, y) = \begin{bmatrix} I(0, 0) & I(0, 1) & \cdots & I(0, N - 1) \\ I(1, 0) & I(1, 1) & \cdots & I(1, N - 1) \\ \vdots & \vdots & & \vdots \\ I(M - 1, 0) & I(M - 1, 1) & \cdots & I(M - 1, N - 1) \end{bmatrix}, \quad (1)$$

en la que a la coordenada (x, y) se le denomina, pels o píxeles[12]. Para una imagen de 8 bits, los valores de intensidad son enteros que están en un rango de $[0, \dots, 255]$ [13].

En el procesamiento de imágenes se trabaja con imágenes a color e imágenes monocromáticas. Una imagen mono-cromática es una imagen que está formada por sólo un color. Este tipo de imágenes utilizan una escala acromática, que es una escala de grises (una modulación continua del blanco al negro), y se utiliza para establecer com-

parativamente tanto el valor de la luminosidad de los colores puros como el grado de claridad de las correspondientes gradaciones de este color puro. Imágenes a color son formadas por una combinación de imágenes individuales. Por ejemplo, en el sistema de colores RGB, una imagen consiste de tres imágenes mono-cromáticas, el canal rojo (R), el canal verde (G) y el canal azul (B), de esta forma se puede trabajar con imágenes a color procesando cada canal individualmente utilizando las técnicas para imágenes mono-cromáticas[9].

Capítulo IV

Sistema de reconocimiento de patrones invariante a traslación, escala y rotación

Si consideramos una imagen $I(x, y)$ de tamaño $M \times N$, cuyo objeto está localizada en el centro de la imagen, entonces el espectro de amplitud de la transformada de Fourier de dicha imagen está dado por

$$A(u, v) = |F(u, v)| = \sqrt{R_F^2(u, v) + I_F^2(u, v)}, \quad (2)$$

donde

$$F(u, v) = R_F(u, v) + iI_F(u, v) = \mathcal{F}\{I(x, y)\} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) e^{-i2\pi(\frac{ux}{M} + \frac{vy}{N})}. \quad (3)$$

Por otra parte, supongamos que se tiene una imagen de tamaño $M \times N$ que tiene el mismo objeto que la imagen $I(x, y)$ pero que no está centrado, matemáticamente se representa por $I(x + x_0, y + y_0)$, tal que $0 \leq x + x_0 \leq M - 1$, $0 \leq y + y_0 \leq N - 1$, pero con la restricción de que se tiene que seleccionar a x_0 y y_0 de manera que el objeto siempre esté contenido en su totalidad dentro de la región fundamental, dada por $\{(x, y) | 0 \leq x \leq M - 1, 0 \leq y \leq N - 1\}$; además dicha región solamente contendrá un objeto. Para este sistema de reconocimiento de patrones no se incluyen los casos de fragmentos de objetos.

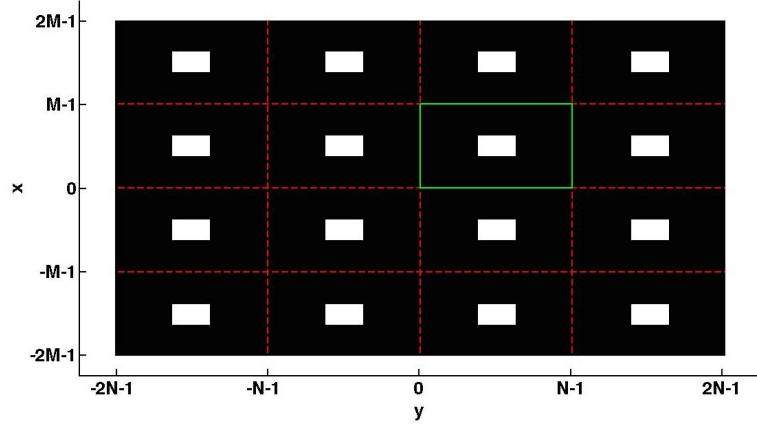


Figura 1. Función periódica $I(x, y)$ de tamaño $M \times N$ con el objeto centrado. El recuadro en verde indica la región fundametal.

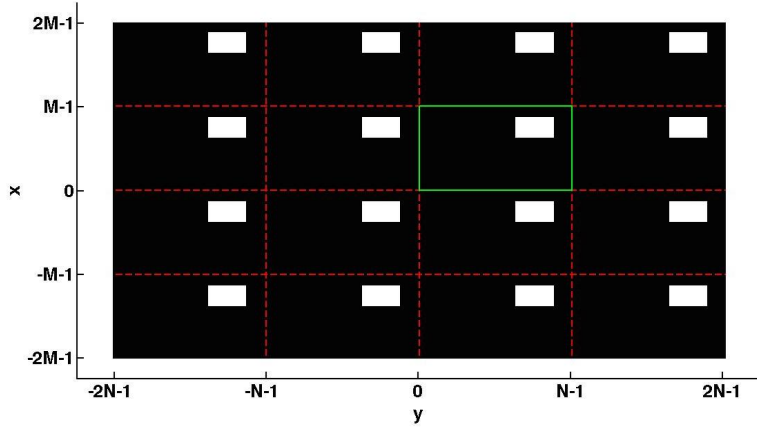


Figura 2. Función periódica $I(x, y)$ de tamaño $M \times N$ con el objeto sin centrar. El recuadro en verde indica la región fundametal.

De la ec.(3) se tiene que la transformada de Fourier de $I(x + x_0, y + y_0)$ es,

$$\mathcal{F}\{I(x + x_0, y + y_0)\} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x + x_0, y + y_0) e^{-i2\pi(\frac{ux}{M} + \frac{vy}{N})}, \quad (4)$$

al hacer el cambio de variables $\alpha = x + x_0$ y $\beta = y + y_0$ en la ec.(4), se tiene que

$$\mathcal{F}\{I(x + x_0, y + y_0)\} = \sum_{\alpha=x_0}^{M-1+x_0} \sum_{\beta=y_0}^{N-1+y_0} I(\alpha, \beta) e^{-i2\pi[\frac{u(\alpha-x_0)}{M} + \frac{v(\beta-y_0)}{N}]}. \quad (5)$$

La ec.(5) toma valores fuera de la región fundamental, pero al extender la propiedad

de la integral de una función periódica en el caso continuo[14],

$$\int_{a-T/2}^{a+T/2} f(t)dt = \int_{-T/2}^{T/2} f(t)dt = \int_0^T f(t)dt, \quad (6)$$

al caso discreto bidimensional, y suponiendo que $I(x, y)$ tiene periodo bidimensional $M \times N$ (Fig. 1 y Fig. 2), por lo que la ec.(5) queda como,

$$\begin{aligned} \mathcal{F}\{I(x + x_0, y + y_0)\} &= \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} I(\alpha, \beta) e^{-i2\pi \left[\frac{u(\alpha-x_0)}{M} + \frac{v(\beta-y_0)}{N} \right]}, \\ &= \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} I(\alpha, \beta) e^{-i2\pi \left(\frac{u\alpha}{M} + \frac{v\beta}{N} \right)} e^{-i2\pi \left(\frac{-ux_0}{M} + \frac{-vy_0}{N} \right)}, \\ &= e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} I(\alpha, \beta) e^{-i2\pi \left(\frac{u\alpha}{M} + \frac{v\beta}{N} \right)}, \\ &= e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} F(u, v), \\ &= e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \mathcal{F}\{I(x, y)\}. \end{aligned} \quad (7)$$

Entonces, el espectro de amplitud de la imagen con el objeto trasladado (x_0, y_0) píxeles con respecto al centro de la imagen está dado por

$$\left| \mathcal{F}\{I(x + x_0, y + y_0)\} \right| = \left| e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \mathcal{F}\{I(x, y)\} \right| = \left| e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \right| \left| \mathcal{F}\{I(x, y)\} \right|, \quad (8)$$

y como

$$\left| e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \right|^2 = \left[e^{i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \right] \left[e^{-i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right)} \right] = e^0 = 1, \quad (9)$$

así que,

$$\left| \mathcal{F}\{I(x + x_0, y + y_0)\} \right| = \left| \mathcal{F}\{I(x, y)\} \right|, \quad (10)$$

por lo tanto, el espectro de Fourier de una imagen cuyo objeto no está centrado es igual al espectro de Fourier de una imagen del mismo objeto pero ubicado en el centro, por lo que en la ec.(10) se ha demostrado que el espectro de amplitud es invariante a traslación.

Una demostración visual de la propiedad dada en la ec.(10) se presenta en la Fig. 3. La Fig. 3(a) muestra una imagen en blanco y negro de 1001×1001 píxeles con

un cuadrado centrado. La Fig. 3(b) muestra el mismo cuadrado pero desplazado del centro. En la Fig. 3(c) y Fig. 3(d) se despliegan los respectivos espectros de amplitud de la Fig. 3(a) y Fig. 3(b), a simple vista se observa que ambas imágenes de los espectros son iguales, las pequeñas diferencias que se podrían presentar se deben a los errores de redondeo generados en los cálculos de la transformada de Fourier de las imágenes. Sólo para su visualización, los espectros se muestran en escala logarítmica.

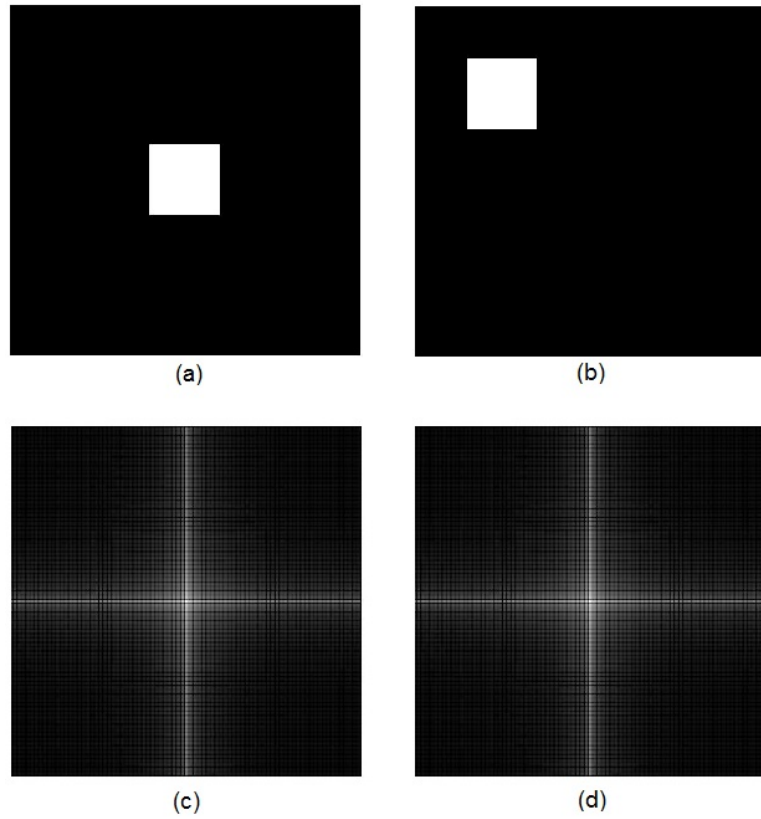


Figura 3. (a) Imagen de un cuadro blanco en el centro. (b) Imagen de un cuadro blanco trasladado a la esquina superior izquierda. (c) Espectro de Fourier de la Fig. 3(a). (d) Espectro de Fourier de la Fig. 3(b).

Después, el espectro de amplitud de la imagen $A(u, v)$ es introducido en la transformada analítica de Fourier-Mellin (TAFM),

$$M(k, \omega) = \mathcal{M}\{A(e^\rho, \theta)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_0^{2\pi} A(e^\rho, \theta) e^{\rho\sigma} e^{-i(k\theta + \rho\omega)} d\theta d\rho, \quad (11)$$

donde $\rho = \ln(r)$ y $\sigma > 0$, en este trabajo se utilizará el σ obtenido experimentalmente por S. Derrode y colaboradores en 2001[2], donde reportan que el valor más apropiado para las aplicaciones al área de reconocimiento de patrones en imágenes digitales es $\sigma = 0.5$. El espectro de amplitud de Fourier en la ec.(2) se reescribe en coordenadas logarítmico-polar, esto se debe a que de esa forma la TAFM, que se exhibe en la ec.(11), se puede calcular vía la transformada rápida de Fourier. La ec.(11) no es invariante a escala ni rotación, pero normalizando $M(k, \omega)$ por su valor en el píxel central (c_k, c_ω) , el espectro de amplitud es invariante a escala [2], esto es,

$$S(k, \omega) = \left| \frac{M(k, \omega)}{M(c_k, c_\omega)} \right|. \quad (12)$$

Las Fig. 4(a) muestra una imagen en blanco y negro con la letra A en tipografía arial, la letra no tiene ninguna transformación geométrica y está centrada en la imagen. La Fig. 4(b) exhibe la misma letra A pero ahora rotada 315° , escalada -25% y ligeramente trasladada del centro de la imagen. En la Fig. 4(c) se presenta la letra B sin transformaciones geométricas, la cual morfológicamente es muy distinta de la A. La Fig. 4(d) es el espectro de Fourier de la Fig. 4(a). El espectro de Fourier de la Fig. 4(b) está dado en la Fig. 4(e), donde se observa que el espectro de esta imagen es muy semejante al espectro de la Fig. 4(d) pero rotado y dilatado. El efecto de la escala es inversamente proporcional en el dominio de Fourier, esto es, si se disminuye el tamaño del objeto, el espectro presentará una dilatación, pero si se incrementa se observará una contracción. La Fig. 4(f) es el espectro de Fourier de la Fig. 4(c), el cual es muy distinto a los espectros de Fourier anteriores. En la Fig. 4(g) se presenta el espectro de amplitud en coordenadas log-polares de la Fig. 4(d). En la Fig. 4(h) se muestra el espectro de amplitud en coordenadas log-polares de la Fig. 4(e), donde se observa que es muy semejante a la Fig. 4(g) pero de picos de menor longitud y

partes convexas a menor altura, este efecto se produce ya que la Fig. 4(b) está escalada -25% . Al definir al eje horizontal como la coordenada angular, la Fig. 4(e) también muestra un desplazamiento angular lateral de 315° debido al efecto de rotación. En la Fig. 4(i) se presenta el espectro de amplitud de la Fig. 4(f) en coordenadas log-polares, que también es muy distinto a los espectros de las Figs. 4(g) y 4(h). En la Fig. 4(j) se muestra el módulo de la transformada analítica de Fourier-Mellin normalizado de la Fig. 4(g). La Fig. 4(k) muestra el módulo de la transformada analítica de Fourier-Mellin normalizado de la Fig. 4(h). Por último, en la Fig. 4(l) se presenta el módulo de la transformada analítica de Fourier-Mellin normalizado de la Fig. 4(i), que aunque muy semejante a los anteriores, a simple vista se puede notar que no son iguales. Hasta este momento el sistema de reconocimiento de patrones solamente es invariante a traslación y escala.

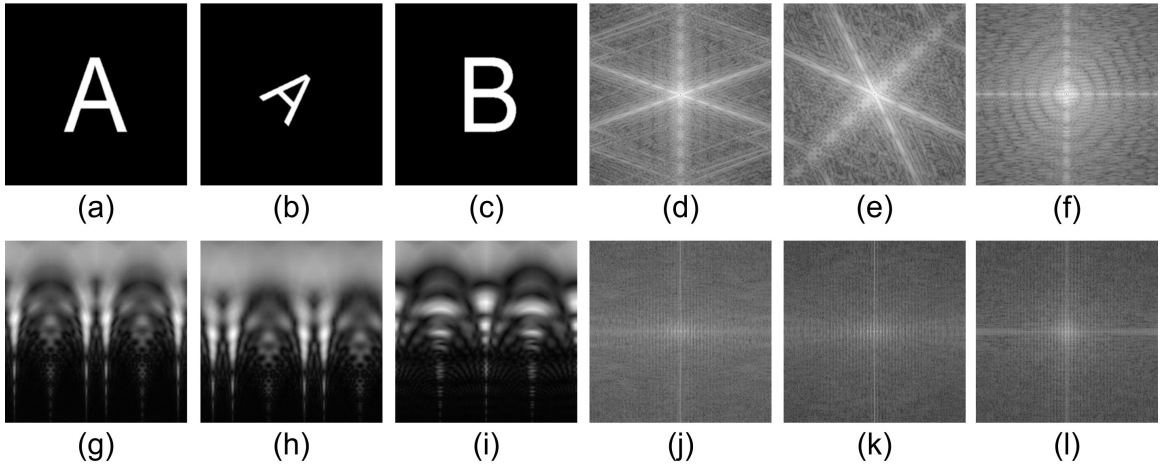


Figura 4. Ejemplos de espectros de amplitud de TAFMN. (a) Imagen I_1 : la letra A con fuente Arial sin transformaciones geométricas. (b) Imagen I_2 : la letra A con fuente Arial con un ángulo de rotación 315° y escalada -25% . (c) Imagen I_3 : la letra B con fuente Arial sin transformaciones geométricas. (d) $A_1(u, v) = |\mathcal{F}\{I_1(x, y)\}|$. (e) $A_2(u, v) = |\mathcal{F}\{I_2(x, y)\}|$. (f) $A_3(u, v) = |\mathcal{F}\{I_3(x, y)\}|$. (g) $A_1(e^\rho, \theta)e^{\rho\sigma}$. (h) $A_2(e^\rho, \theta)e^{\rho\sigma}$. (i) $A_3(e^\rho, \theta)e^{\rho\sigma}$. (j) La $S_1(k, \omega)$ de la Fig. 4(g). (k) La $S_2(k, \omega)$ de la Fig. 4(h). (l) La $S_3(k, \omega)$ de la Fig. 4(i).

IV.1 La firma 1D circular

Para construir la firma 1D circular invariante a traslación, escala y rotación, la imagen digital S , que es el espectro de la TAFMN, dada por la ec.(12) con $\sigma = \frac{1}{2}$, se filtra con máscaras binarias que contienen un círculo de radio r . El círculo está centrado en el píxel (c_k, c_ω) , el cual es el centro de la imagen S . Entonces, la máscara binaria circular se define como,

$$C_r(x, y) = \begin{cases} 1, & \text{si } r = \lceil \sqrt{(x - c_k)^2 + (y - c_\omega)^2} \rceil, \\ 0, & \text{caso contrario,} \end{cases} \quad (13)$$

donde $x = 1, \dots, M, y = 1, \dots, N, r = 1, \dots, n$ y $\lceil \cdot \rceil$ representa la operación de redondeo de un número real. La Fig. 5 muestra algunos ejemplos de máscaras circulares. Matemáticamente, el procedimiento del filtrado se representa por

$$H_r = S \otimes C_r, \quad (14)$$

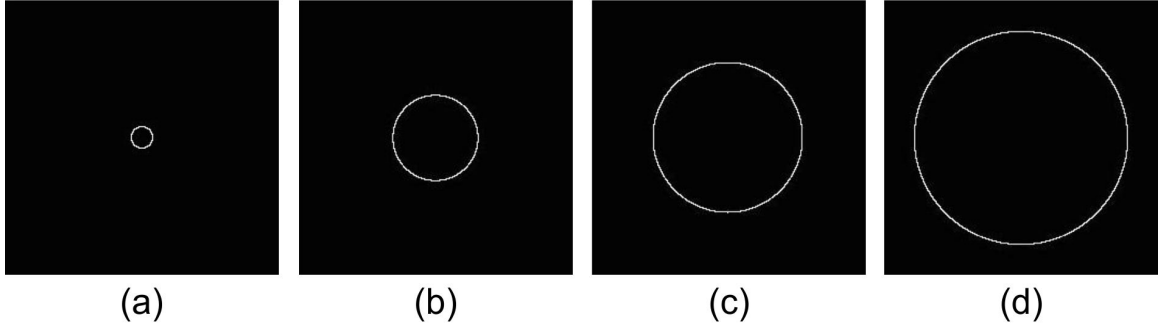


Figura 5. Ejemplos de máscaras circulares binarias. (a) C_{10} . (b) C_{40} . (c) C_{70} . (d) C_{100} .

donde C_r es la máscara binaria circular, que tiene el mismo tamaño de S , $r = 1, 2, \dots, n$, $n = \min \{c_k, c_\omega\}$ (radio de la circunferencia contenida en la máscara C_r), y $\otimes$ representa el producto punto a punto (multiplicación Hadamard). Posteriormente, se suman los valores de intensidad en H_r , es decir

$$k_r = \sum_{x=1}^M \sum_{y=1}^N H_r(x, y). \quad (15)$$

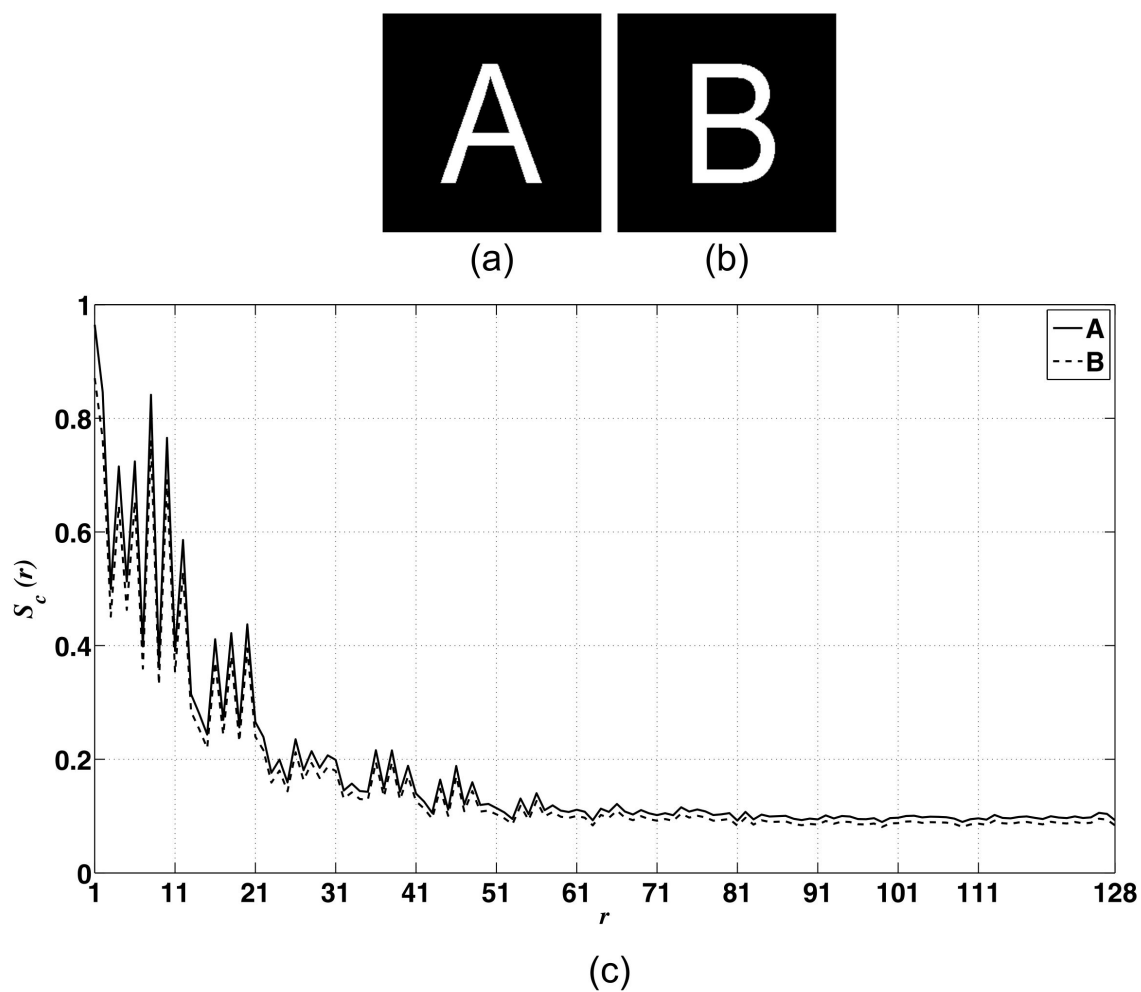


Figura 6. Ejemplos de firmas circulares. (a) Imagen en blanco y negro de la letra A con fuente Arial. (b) Imagen en blanco y negro de la letra B con fuente Arial. (c) Firmas circulares de las Figs. 6(a) y 6(b).

Finalmente, la firma 1D circular a la que denotaremos por S_c , se construye mediante

$$\begin{aligned} S_c : \{1, \dots, n\} &\rightarrow \mathbb{R}, \\ S_c(r) &= k_r. \end{aligned} \tag{16}$$

La Fig. 6(c) muestra la firma circular de las imágenes en las Figs. 6(a) y 6(b). El tamaño de la firma es el número de círculos con los que se muestrea el disco de radio $r = n$, de centro en el píxel (c_k, c_ω) y $\Delta r = 1$.

IV.2 La firma 1D radial

Por otra parte, para obtener la firma 1D radial invariante a traslación, escala y rotación, la imagen digital S del espectro de la TAFMN, dada por la ec.(12) con $\sigma = \frac{1}{2}$, se filtra con máscaras binarias radiales R_θ , las cuales tienen el mismo tamaño que la imagen S . Entonces, la máscara binaria radial se define como

$$R_\theta(x, y) = \begin{cases} 1, & \text{si } y = \lceil m(x - x_1) + y_1 \rceil, \\ 0, & \text{caso contrario,} \end{cases} \tag{17}$$

donde $x = 1, \dots, n$, $m = \frac{y_2 - y_1}{x_2 - x_1}$, $(x_1, y_1) = (c_k + r \cos \theta, c_\omega + r \sin \theta)$, $(x_2, y_2) = (c_k, c_\omega)$, $r = n$, y $\lceil \cdot \rceil$ representa la operación de redondeo de un número real. La Fig. 7 muestra algunos filtros binarios radiales. Matemáticamente, dicho procedimiento se representa por

$$G_\theta = S \otimes R_\theta, \tag{18}$$

donde $\theta = 0, \dots, 359$. El rayo es de tamaño $n = \min \{c_k, c_\omega\}$ y con origen en el píxel central (c_k, c_ω) de la imagen S .

Luego, se suman los valores de intensidad en G_θ , es decir,

$$\bar{k}_\theta = \sum_{x=1}^M \sum_{y=1}^N G_\theta(x, y), \quad (19)$$

después, se construye la firma 1D radial mediante

$$\begin{aligned} \bar{S}_r : \{0, \dots, 359\} &\rightarrow \mathbb{R}, \\ \bar{S}_r(\theta) &= k_\theta. \end{aligned} \quad (20)$$

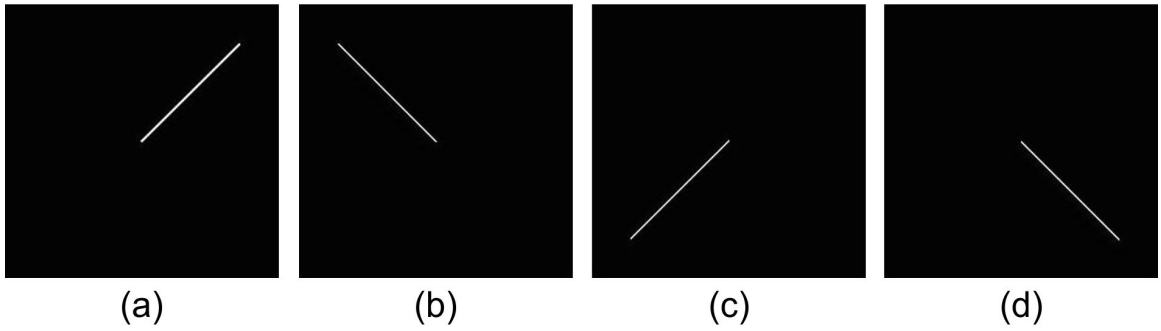


Figura 7. Ejemplos de máscaras binarias radiales. (a) R_{45} . (b) R_{135} . (c) R_{225} . (a) R_{315} .

Debido al desplazamiento circular en el espectro de la TAFM por la transformación log-polar (por ejemplo las Figs. 4(g) y 4(h)), los valores de $\bar{S}_r(\theta)$ se tienen que ordenar ya sea de forma ascendente o descendente. En este caso se optó por ordenarlos de manera descendente para obtener la siguiente firma 1D radial S_r invariante a traslación, escala y rotación

$$S_r(\alpha) = \bar{k}_\alpha, \quad (21)$$

donde $\bar{k}_0 \geq \bar{k}_2 \geq \dots \geq \bar{k}_{359}$. En la Fig. 8(c) se muestran las firmas radiales para las Figs. 8(a) y 8(b). El tamaño de la firma es igual al número de rayos que se utilizaron para muestrear todo el círculo con $\Delta\theta = 1^\circ$.

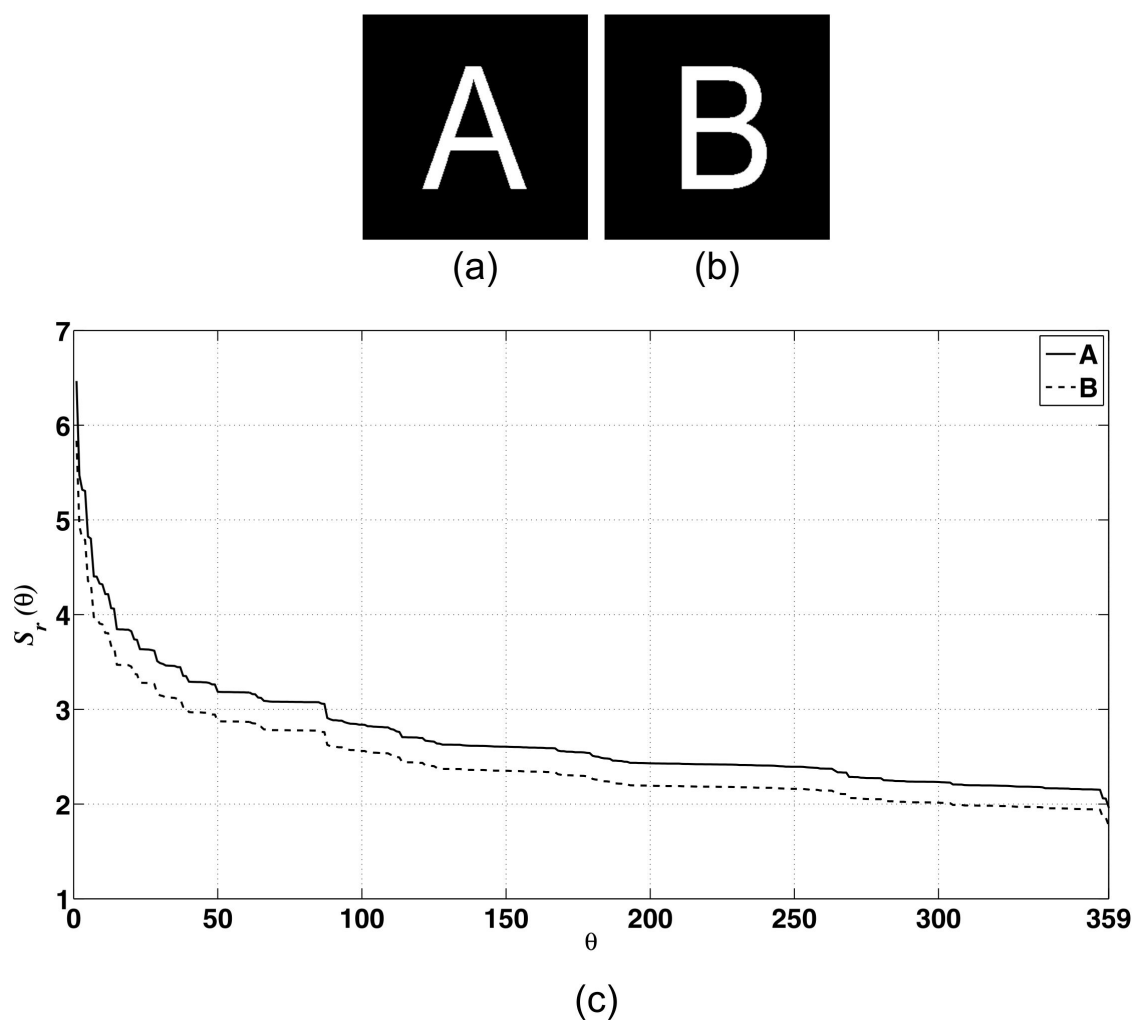


Figura 8. Ejemplos de firmas radiales. (a) Imagen en blanco y negro de la letra A con fuente Arial. (b) Imagen en blanco y negro de la letra B con fuente Arial. (c) Firmas radiales de las Figs. 8(a) y 8(b).

IV.3 La firma 1D marginal por columna

Para obtener la firma 1D marginal por columna invariante a traslación, escala y rotación, a diferencia de las otras firmas, donde se utilizan máscaras binarias, aquí simplemente se suman por columna los valores de intensidad de la imagen S del espectro de la TAFMN, dada por la ec.(12) con $\sigma = \frac{1}{2}$. Matemáticamente dicho procedimiento se representa por

$$\widehat{k}(x) = \sum_{y=1}^N S(x, y), \quad (22)$$

donde, $x = 1, \dots, M$. Después, la firma 1D marginal se construye como,

$$\begin{aligned} \widehat{S}_{col} : \{1, \dots, M\} &\rightarrow \mathbb{R}, \\ \widehat{S}_{col}(x) &= \widehat{k}(x). \end{aligned} \quad (23)$$

En la Fig. 9(c) se muestran las firmas marginales por columna para las Figs.9(a) y 9(b). El tamaño de la firma es igual al número de columnas que tiene la imagen S .

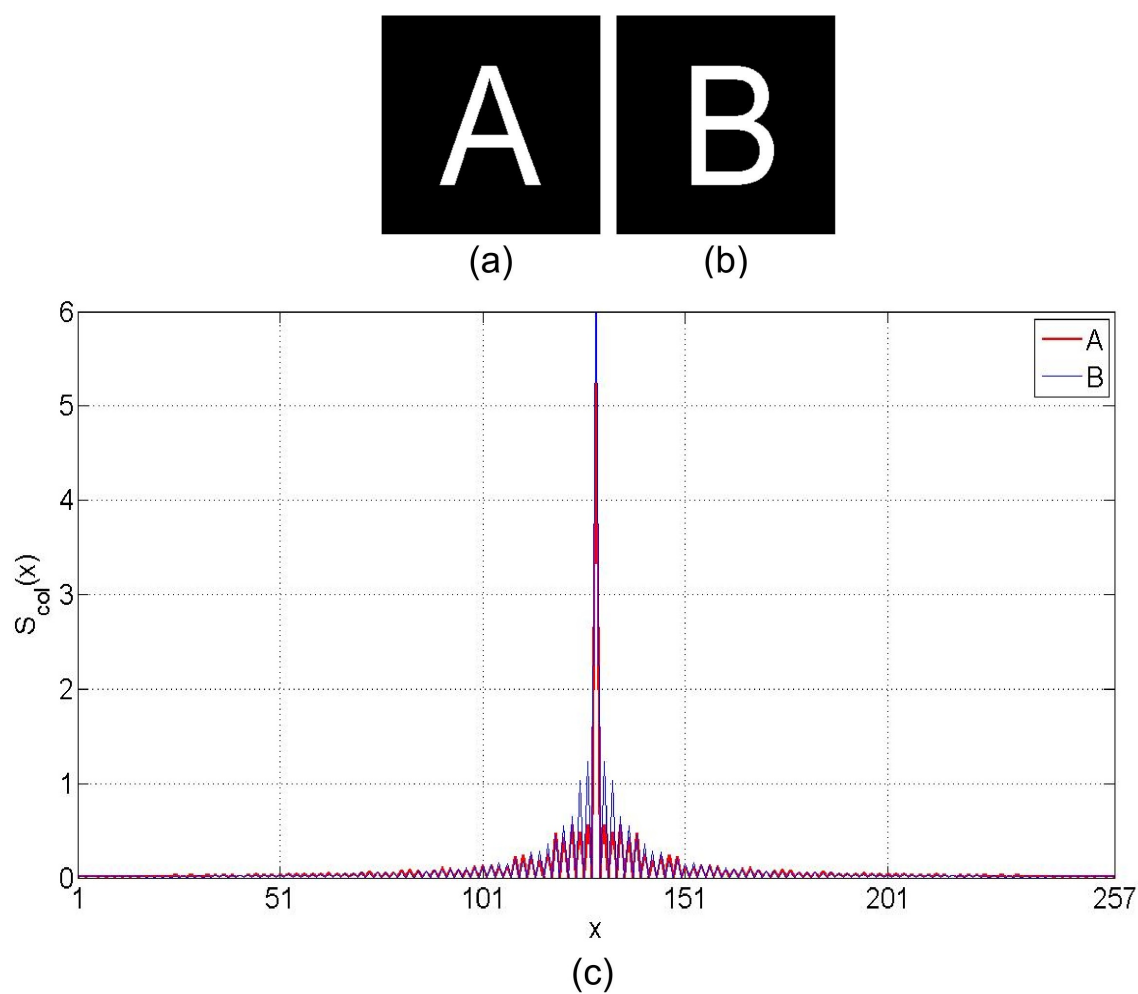


Figura 9. Ejemplos de firmas marginales. (a) Imagen en blanco y negro de la letra A con fuente Arial. (b) Imagen en blanco y negro de la letra B con fuente Arial. (c) Firmas marginales de las Figs. 9(a) y 9(b).

Capítulo V

Ortoedro de confianza para el descriptor invariante a traslación, escala y rotación

En 2015, J. Verdugo-Olachea[15] introdujo una metodología para la creación de un espacio de clasificación por medio de ortoedros de confianza, esto con la finalidad de reducir el tiempo de cómputo en la etapa de clasificación. Tradicionalmente la clasificación se lleva a cabo mediante funciones de correlación o métricas, lo que implica realizar la comparación del vector de características de la imagen problema con el vector de características de una imagen de referencia, entonces dicho procedimiento conlleva a que si, por ejemplo, se trabaja con una base de datos de imágenes de referencia de 1000 elementos, se obtendrán 1000 planos de salida de correlación o de la métrica utilizada. En 2016, R. Vélez-Rábago[16] utilizó de forma exitosa dicha metodología para crear un espacio de clasificación invariante a traslación, escala y rotación para imágenes en blanco y negro, y en escala de grises. En 2016, C. Barajas-García y colaboradores[17] también emplearon de forma exitosa la metodología de espacios de clasificación de ortoedros de confianza, y extendieron su uso a imágenes a color.

Para determinar la eficacia del descriptor de imágenes invariante a traslación, escala y rotación presentado en este trabajo de tesis, se realizaron pruebas con conjuntos de imágenes de referencia en: blanco y negro y en escala de grises.

V.1 Imágenes en blanco y negro

El conjunto de imágenes en blanco y negro (BW) consta de 26 elementos de 257×257 píxeles, conteniendo cada imagen una letra del alfabeto Latino con tipografía Arial Fig. 10. Para generar la base de datos de imágenes problema, cada una de las 26 imágenes de referencia fueron rotadas 360° con tamaño de paso $\Delta\theta = 1^\circ$, luego, cada una de esas imágenes fueron escaladas $\pm 20\%$, con un tamaño de paso $\Delta h = 1\%$, obteniendo de este modo un total de $(360)(41) = 14,760$ imágenes diferentes de cada una de las letras en la base de datos de imágenes de referencia de la Fig. 10, en la Fig. 11 se muestran algunos ejemplos.

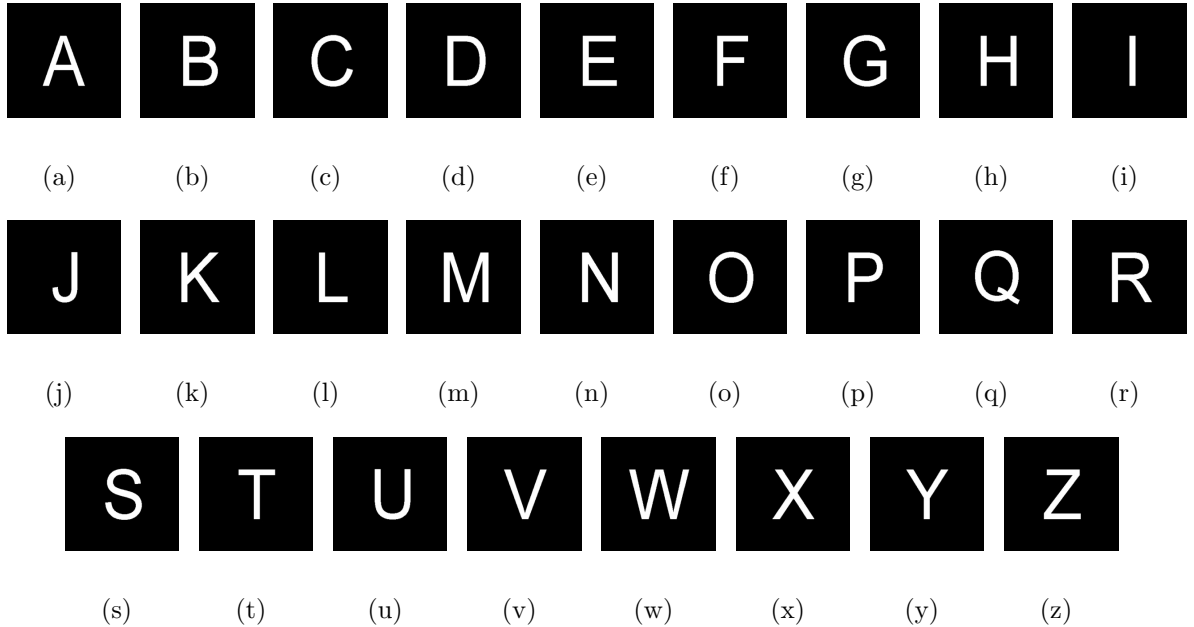


Figura 10. Base de datos de imágenes de referencia en blanco y negro que contienen las letras del alfabeto Latino.

Definiendo al conjunto de imágenes problema generadas a partir de las imágenes de referencia I^j (para el ejemplo de la Fig. 10, $j = a, b, \dots, z$) como

$$\Psi = \{I_{\theta,h}^j | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}, \quad (24)$$

se puede definir a los conjunto de las firmas 1D invariantes a traslación, escala y rotación

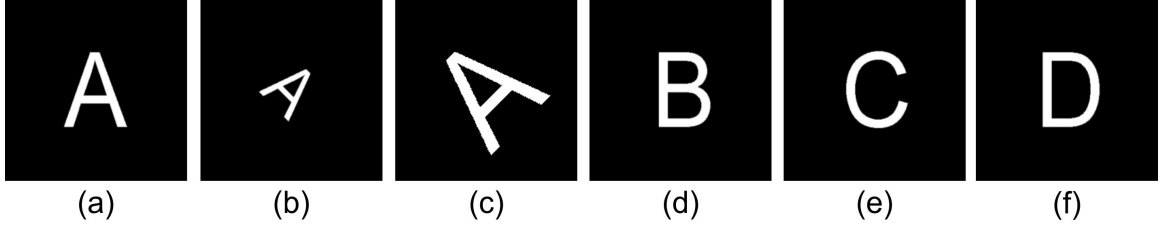


Figura 11. Ejemplos de elementos de la base de datos de imágenes problema. (a) La letra A sin transformaciones geométricas. (b) La letra A con un ángulo de rotación de 315° y escalada -25% . (c) La letra A con ángulo de rotación de 45° y escalada 25% . (d) La letra B sin transformaciones geométricas. (e) La letra C sin transformaciones geométricas. (f) La letra D sin transformaciones geométricas.

S_c , S_r y S_{col} como

$$\Phi^{j,c} = \{S_{\theta,h}^{j,c} | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}, \quad (25)$$

$$\Phi^{j,r} = \{S_{\theta,h}^{j,r} | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}, \quad (26)$$

$$\Phi^{j,col} = \{S_{\theta,h}^{j,col} | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}. \quad (27)$$

En los cuales las firmas se obtienen mediante el procedimiento descrito en el capítulo 3.

Para construir el espacio de clasificación se hace uso del valor de las potencias de las firmas S_c , S_r y S_{col} de cada imagen, que se calcula como sigue,

$$P_{\theta,h}^{j,c} = \frac{1}{N_c} \sum_{r=1}^{N_c} (S_{\theta,h}^{j,c}(r))^2, \quad (28)$$

$$P_{\theta,h}^{j,r} = \frac{1}{N_r} \sum_{\phi=0}^{N_r} (S_{\theta,h}^{j,r}(\phi))^2, \quad (29)$$

$$P_{\theta,h}^{j,col} = \frac{1}{N_{col}} \sum_{x=1}^{N_{col}} (S_{\theta,h}^{j,col}(x))^2, \quad (30)$$

donde N_c , N_r y N_{col} son las longitudes de las firmas S_c , S_r y S_{col} , respectivamente.

Como el conjunto Ψ contiene 14,760 imágenes, que se generaron a partir de la imagen de referencia I^j , los conjuntos

$$P_c^j = \{P_{\theta,h}^{j,c} | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}, \quad (31)$$

$$P_r^j = \{P_{\theta,h}^{j,r} | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}, \quad (32)$$

$$P_{col}^j = \{P_{\theta,h}^{j,col} | \theta = 0, 1, \dots, 359; h = -20, -19, \dots, 20\}, \quad (33)$$

contienen 14,760 elementos escalares cada uno. Utilizando el teorema del límite central, podemos decir que la distribución muestral de medias, provenientes del conjunto P_k donde $k = c, r, col$, tiene un comportamiento aproximadamente normal, con medias muestrales μ_k y errores estándar $(EE_k) \frac{\sigma_k}{\sqrt{n}}$, con desviación estándar σ_k y n el tamaño de la muestra, en este caso $n = 14,760$. Utilizando el método estadístico de diagrama de cajas, se construyeron los intervalos de confianza del 95.4% con $\mu_k \pm 2EE_k$. Después, se crearon ortoedros cuyos lados están dados por los intervalos de confianza $\mu_c \pm 2EE_c$, $\mu_r \pm 2EE_r$, $\mu_{col} \pm 2EE_{col}$ y con vértices en las coordenadas: $(\mu_c - 2EE_c, \mu_r - 2EE_r, \mu_{col} - 2EE_{col})$, $(\mu_c + 2EE_c, \mu_r - 2EE_r, \mu_{col} - 2EE_{col})$, $(\mu_c - 2EE_c, \mu_r + 2EE_r, \mu_{col} - 2EE_{col})$, $(\mu_c + 2EE_c, \mu_r + 2EE_r, \mu_{col} - 2EE_{col})$, $(\mu_c - 2EE_c, \mu_r - 2EE_r, \mu_{col} + 2EE_{col})$, $(\mu_c + 2EE_c, \mu_r - 2EE_r, \mu_{col} + 2EE_{col})$, $(\mu_c - 2EE_c, \mu_r + 2EE_r, \mu_{col} + 2EE_{col})$, $(\mu_c + 2EE_c, \mu_r + 2EE_r, \mu_{col} + 2EE_{col})$. En la Fig. 12 se muestra el ortoedro de confianza del 95.4%, construido a partir de las 14,760 imágenes problema de la letra A, el vector normal al plano que contiene el punto $(\mu_c, \mu_r, \mu_{col} + 2EE_{col})$ se indica en azul, el vector normal al plano que contiene la coordenada $(\mu_c - 2EE_c, \mu_r, \mu_{col})$ se presenta en verde y el vector normal al plano que contiene el punto $(\mu_c, \mu_r - 2EE_r, \mu_{col})$ está dado en amarillo.

En la Fig. 13 se muestra el espacio de clasificación de ortoedros de confianza para las imágenes de la Fig. 10 y en la Fig. 14 se presenta una ampliación de este espacio para poder observar que los ortoedros no se traslapan. Dado que se asignaron espacios tridimensionales a cada imagen vía los ortoedros, y ninguno de estos se traslapa, se concluye que el sistema de reconocimiento de patrones invariante a traslación, escala y rotación presenta un nivel de confianza de al menos 95.4%.

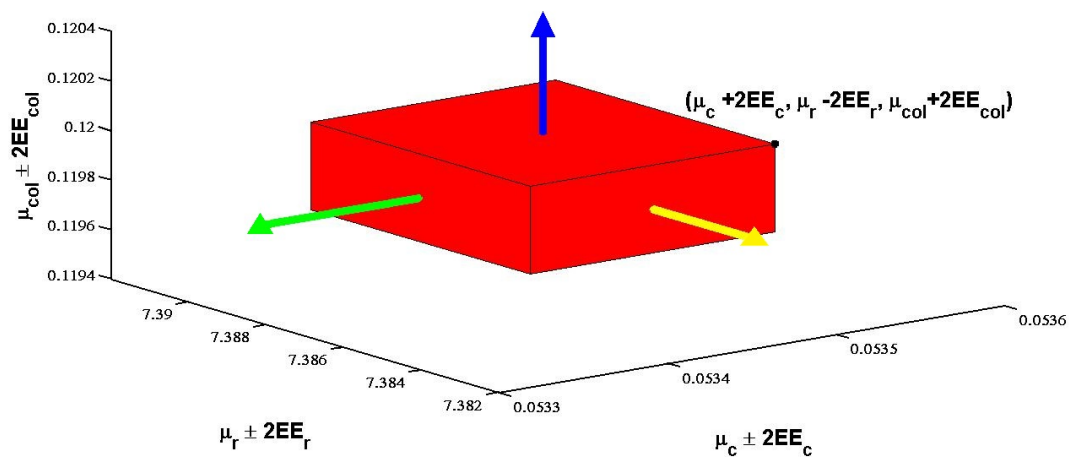


Figura 12. Ortoedro de confianza del 95.4% de la letra A con fuente Arial.

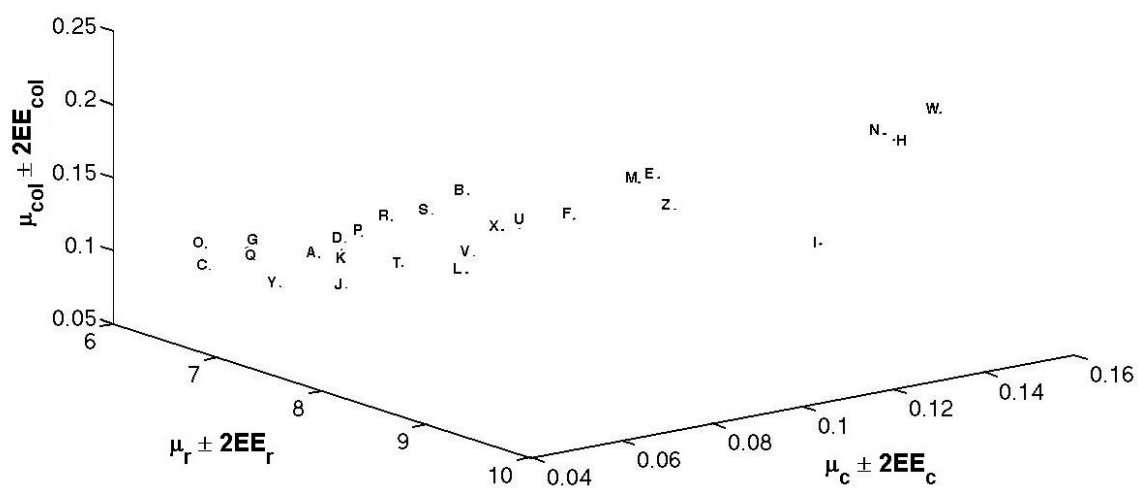


Figura 13. Espacio de clasificación de ortoedros de confianza para la base de datos de imágenes de referencia de la Fig. 10.

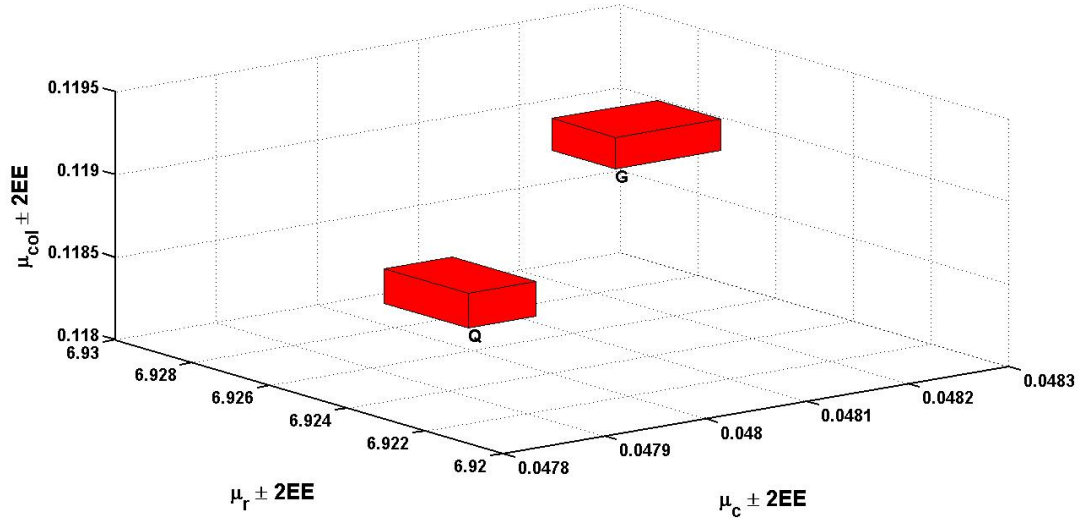


Figura 14. Amplificación de una zona del espacio de clasificación de ortoedros de confianza de la Fig. 13

Los descriptores basados en funciones de correlación [1, 5, 6] o funciones de distancia [2, 10] generan un plano de salida para cada imagen objetivo, por lo tanto, con esos sistemas se generarían 26 planos de salida (uno por imagen de referencia) para clasificar las imágenes digitales. Este trabajo utiliza una nueva metodología para reducir considerablemente el costo computacional en el paso de la clasificación utilizando solamente un espacio de salida.

V.2 Imágenes en escala de grises

Una vez que se cuenta con un descriptor que clasifica imágenes en blanco y negro con un nivel de confianza del 95.4%, el siguiente paso es determinar si el sistema también es eficaz cuando se trabaja con imágenes en escala de grises. El conjunto de imágenes utilizado para dicha prueba consta de 17 elementos de 257×257 píxeles, conteniendo cada imagen una especie de diatomea distinta, pero de morfología similares, Fig. 15. Como

en el Capítulo IV.1, se generó la base de datos de imágenes problema a partir de las 17 imágenes de referencia, las cuales fueron rotadas 360° con tamaño de paso $\Delta\theta = 1^\circ$, y luego, cada una de esas imágenes fueron escaladas $\pm 20\%$, con un tamaño de paso $\Delta h = 1\%$, obteniendo de este modo un total de $(360)(41) = 14,760$ imágenes diferentes de cada una de las diatomeas en la base de datos de imágenes de referencia de la Fig. 15.

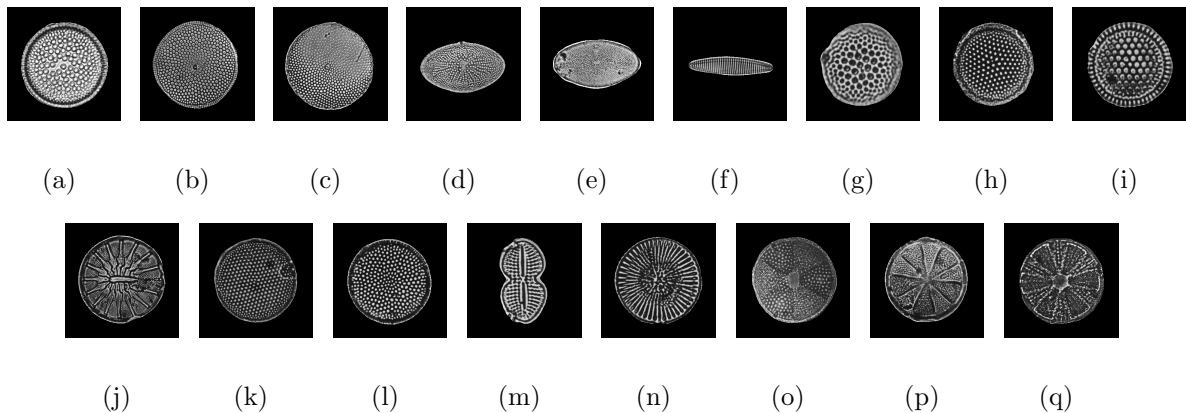


Figura 15. Base de datos de imágenes de referencia de fósiles de diatomeas. (a) A: *Actinocyclus ingens* - Rattray. (b) B: *Azpeitia* sp. (c) C: *Azpeitia nodulifera* - (Schmidth) Fryxell et Sims. (d) D: *Actinocyclus ellipticus* - Grunow in van Heurck. (e) E: *Actinocyclus ellipticus* var *moronensis* - (Deby ex Rattray) Kolbe. (f) F: *Nitzschia praereinholdii* - Schrader. (g) G: *Thalassiosira oestruppii* var 1. (h) H: *Thalassiosira oestruppii* var 2. (i) I: *Thalassiosira domifacta* - (Hendey) Jouse. (j) J: *Asteromphalus imbricatus* - Wallich. (k) K: *Thalassiosira kozlovii* - Makarova. (l) L: *Coscinodiscus radiatus* - Ehrenberg. (m) M: *Diploneis bombus* - Cleve-Euler in Backman et Cleve-Euler. (n) N: *Stephanodiscus* sp. (o) O: *Actinopterychus undulatus* - (Bailey) Ralf. (p) P: *Actinopterychus bipunctatus* - Lohman. (q) Q: *Actinopterychus splendens* - (Shadbolt) Ralf ex Pritchard.

Posteriormente se calcularon las firmas de cada una de las diatomeas, y luego se obtuvieron las potencias de las firmas, para finalmente construir los ortoedros de confianza con el procedimiento que fue descrito en la Capítulo IV, sección 1, para las imágenes en blanco y negro. En la Fig. 16 se muestra el espacio de clasificación de ortoedros de confianza para las imágenes de la Fig. 15. Dado que se asignaron espacios tridimensionales a cada imagen vía ortoedros, y ninguno de estos se traslapa, se concluye que el sistema de reconocimiento de patrones invariante a traslación, escala y rotación

presenta un nivel de confianza de al menos 95.4%.

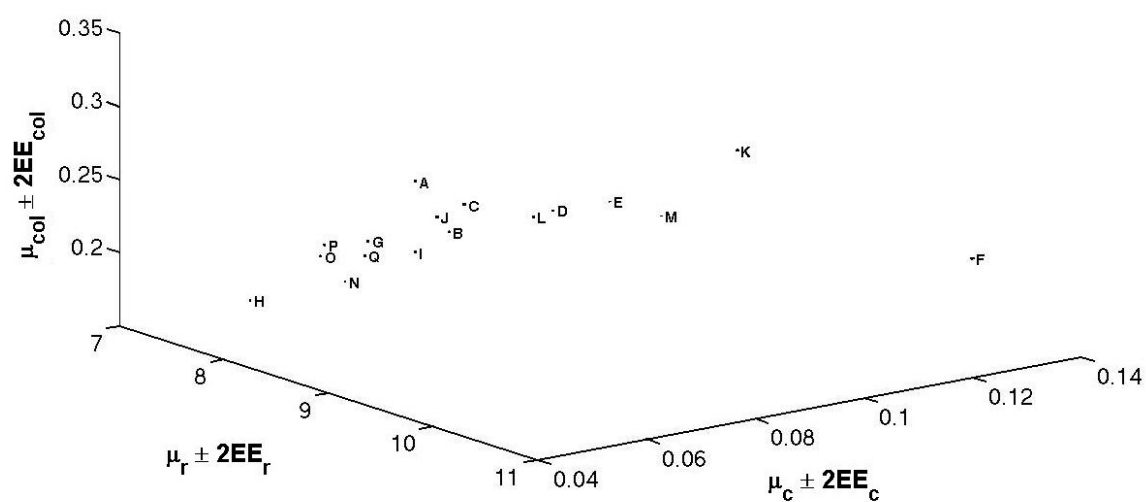


Figura 16. Espacio de clasificación de ortoedros de confianza para la base de datos de imágenes de referencia de la Fig. 15.

Capítulo VI

Conclusiones

Basado en la teoría de la transformada de Fourier y la transformada de Fourier-Mellin, se desarrolló un sistema de reconocimiento de patrones que es invariante a traslación, escala y rotación. Dicho sistema fue diseñado para trabajar con imágenes en blanco y negro, y en escala de grises.

El sistema de reconocimiento de patrones hace uso de las firmas 1D invariantes a traslación, escala y rotación, para reducir el tiempo de cómputo de forma considerable en la etapa de clasificación. Además, en lugar de trabajar con múltiples planos de salida, como se hace en los sistemas de reconocimiento de patrones que utilizan funciones de correlación, aquí se genera solamente un espacio de clasificación de ortoedros, lo que permite reducir la cantidad de operaciones, y por ello, el tiempo de cómputo.

De los espacios de clasificación de ortoedros obtenidos en las pruebas hechas al sistema de reconocimiento de patrones usando imágenes digitales en blanco y negro, y en escala de grises (con objetos de morfología muy similares), se llega a la conclusión de que el sistema tiene un nivel de confianza de al menos el 95.4%.

Referencias

- [1] T.V. HOANG and S. TABBONE, “Invariant pattern recognition using the RFM descriptor”, *Pattern Recognition* **45**(1), págs. 271-284, 2012.
- [2] S. DERRODE and F. GHORBEL, “Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description”, *Computer Vision and Image Understanding* **83**(1), págs. 57-78, 2001.
- [3] F. GHORBEL, “Towards unitary formulation for invariant image description; application to towards unitary formulation for invariant image description; application to image coding”, *Annals of Telecommunication* **53**, págs. 242-260, 1998.
- [4] S. SOLORZA, and J. ÁLVAREZ-BORREGO, “Digital system of invariant correlation to position and rotation”, *Optics Communications* **283**, págs. 3613-3630, 2010.
- [5] S. SOLORZA, and J. ÁLVAREZ-BORREGO, “Translation and rotation invariant pattern recognition by binary rings masks”, *Journal of Modern Optics* **62**(10), págs. 851-864, 2015.
- [6] J. ÁLVAREZ-BORREGO, S. SOLORZA, and M. BUENO-IBARRA, “Invariant correlation to position and rotation using a binary mask applied to binary and gray images”, *Optics Communications* **294**, págs. 105-117, 2013.
- [7] A. SOLÍS-VENTURA, J. ÁLVAREZ-BORREGO, and S. SOLORZA, “An adaptive nonlinear correlation with binary mask invariant to rotation and scale applied to identify phytoplankton”, *Optics Communications* **339**, págs. 185-793, 2015.
- [8] C. BARAJAS-GARCÍA, S. SOLORZA, and J. ÁLVAREZ-BORREGO, “Classification

of fragments of objects by the Fourier masks pattern recognition system”, *Optics Communications* **367**, págs. 335-345, 2016.

- [9] R. GONZALEZ, R. WOODS and S. EDDINS, *Digital Image Processing using Matlab*, Tata McGraw Hill Education Private Limited, 2010.
- [10] J. LERMA-ARAGÓN and J. ÁLVAREZ-BORREGO, “Vectorial signatures for invariant recognition of position, rotation and scale pattern recognition”, *Journal of Modern Optics* **56**(14), págs. 1598-1606, 2009.
- [11] B. JAHNE, *Digital Image Processing*, Springer, 2002.
- [12] R.C. GONZALEZ, and R.E. WOODS, *Digital Image Processing*, Pearson Prentice Hall, 2008.
- [13] M.R. RODRÍGUEZ, A.J.H. SOSSA, *Procesamiento y Análisis Digital de imágenes*, Alfaomega, 2012.
- [14] P. H. HSU, *Análisis de Fourier*, Prentice Hall, New York, 1970.
- [15] J. VERDUGO-OLACHEA, “Reconocimiento de patrones en imágenes digitales usando la transformada de Radon”, Tesis de Licenciatura, Universidad Autónoma de Baja California, Ensenada, México, 2015.
- [16] R. VÉLEZ-RÁBAGO, “Reconocimiento de patrones en imágenes digitales usando la transformada de Fourier normalizada y la transformada analítica de Fourier-Mellin normalizada”, Tesis de Licenciatura, Universidad Autónoma de Baja California, Ensenada, México, 2016.
- [17] C. BARAJAS-GARCÍA, S. SOLORZA, and J. ÁLVAREZ-BORREGO, “Rotation, Scale and Translation Invariant Pattern Recognition System for Color Images”,

12th International Conference of Computational Methods in Science and Engineering: Symposium 4 Pattern Recognition in Digital Images. En proceso de publicación en el AIP Proceeding, 2016.