

A Numerical Technique for Breast Medical Research Based on the FSS Transform

F. A. Uribe, *Senior Member, IEEE*, E. Lozada, *Student*, J. R. Morales, *member, IEEE*,
A. Alanís, *Senior Member, IEEE* and N. G. Arana, *member, IEEE*

Abstract— Breast cancer is one of the more aggressive diseases with a high worldwide death rate in women. The early detection and timely treatment are the best methods to save life patients. Thus, the development of a new technique for medical research can represent an invaluable tool. The aim of this paper is to present an accurate and efficient numerical technique based in *The Fast Slant Stack Transform* as an aid to detect spiculated masses and architectural tissue distortion hidden in digital mammograms. This technique is efficient because performs this task in a very short computer processing time. The technique is validated testing 50 study cases taken from a scientific medical mammogram database. Finally, a sensitivity analysis confirms the accuracy of the here proposed technique as a promising one option particularly in developing countries where digital mammograms are highly preferred for the reduced diagnostic costs.

Index Terms— The Radon Transform, Fourier Transform, image analysis, breast cancer detection.

I. INTRODUCCIÓN

El cáncer de mama es una de las mayores causas de muerte de mujeres tanto en países desarrollados como en vías de desarrollo [1]. Estadísticas del 2019 presenta cifras que 2.08 millones de nuevos casos son diagnosticados anualmente en mujeres de todo el mundo, relacionando un total de 626,679 muertes debidas a esta patología [2, 3]. En México, la tasa actual de cáncer de mama en mujeres es de 27,283 con un número de muertes de 6,884 aumentando a diario [1-4].

Un problema al que se enfrenta un médico radiólogo en el estudio del cáncer de mama es la identificación de masas espiculadas (SM, por sus siglas en inglés) y distorsión de la arquitectura del tejido mamario (AD, también por sus siglas en inglés), ambos escondidos en mamografías digitales con un alto contenido de ruido por la medición del aparato. Cuando esto ocurre, usualmente el paciente puede terminar en una biopsia dolorosa y costosa que puede producirle un daño permanente fisiológico y psicológico.

F. A. Uribe and J. R. Morales are with the Electrical Engineering Graduate Program, The University of Guadalajara, Guadalajara, Jalisco, Mexico, PC. 44430. (e-mail: Felipe.uribe@cucei.udg.mx).

E. Lozada obtained the M. Sc. E. E. from The Electrical Engineering Graduate Program, The University of Guadalajara, Guadalajara, Jalisco, México, PC. 44430.

A. Alanís and N. G. Arana are from the Electronics and Computing Engineering Graduate Program, The University of Guadalajara, Guadalajara, Jalisco, México, PC. 44430.

La probabilidad de aciertos de un radiólogo en la detección de estos problemas es aproximadamente del 75% [5- 8]. El problema aquí es que cierto tipo de masas espiculadas a menudo pueden tener un grado avanzado de evolución interna, pero muestran de manera simultánea una estructura externa intacta y saludable [5-9]. En algunos casos el cáncer de mama no siempre exhibe un patrón de trauma. Incluso, durante las tomas de las mamografías, cuando el equipo médico requiere que la paciente comprima con mucha presión su pecho contra las placas del aparato, lo cual provoca que se oculten los síntomas iniciales. En esta situación, la interpretación de un radiólogo suele ser una tarea muy complicada [8, 10]. Por lo tanto, el diagnóstico asistido por computadora es una herramienta muy útil, sin embargo, con múltiples retos de investigación aún poco explorados. A pesar de que muchos investigadores se han dedicado a reducir el índice de error en la probabilidad de aciertos para la detección del cáncer utilizando técnicas computacionales complejas aún no se han podido reducir los índices de falsos positivos en dichos estudios [9-10].

En este artículo se presenta una nueva técnica basada en la combinación de la Fast Slant Stack Transform (FSST) y filtros de espiculación radial (SRF) para detectar masas espiculadas (SM) y distorsión arquitectónica del tejido mamario (AD) de forma clara como una herramienta de apoyo para la investigación en el diagnóstico asistido por computadora.

En las 50 pruebas realizadas en este trabajo la técnica propuesta obtuvo una precisión mejorada en la probabilidad de aciertos del 92.1% para SM y 85% para AD con respecto a otras técnicas previamente propuestas basadas en la Transformada Discreta de Radon (DRT) [11]. La técnica propuesta consiste en los siguientes pasos:

- Pre-procesamiento de datos de las mamografías para suavizar, homogeneizar y reducir el ruido.
- Mejoramiento en identificación de tejidos espiculados adicionalmente al uso de la DRT con la FSST [11].
- Detección de patrones espiculados en el tejido mamario con distorsión arquitectónica a través de filtros de espiculación radial [12, 13].

Finalmente, la técnica propuesta se valida cuantitativamente con respecto a *Sampat, et al.*, [6-8] y *Karssemeijer et al.* [14] considerando 50 mamografías digitales de pacientes con distintos trastornos mamarios obtenidas de la base de datos (*The mini-MIAS database of mammograms*) [15].

II. TEORÍA DE LA TÉCNICA MEJORADA

Los datos obtenidos de la imagen mamográfica en estudio se deben pre-procesar para suavizar y homogeneizar las estructuras formadas por los píxeles de la imagen. Primero se utiliza un *filtro Gaussiano* con una ventana de conectividad de 3×3 y posteriormente se reduce el ruido con un filtro de mediana con la misma conectividad.

A. Umbral Optimo

El principal problema en el análisis de imágenes médicas consiste en extraer las regiones de píxeles del fondo de la mamografía original. Por lo tanto, se aplica un proceso de segmentación para separar la región de la mama del fondo de la imagen eliminando el área del músculo pectoral de la con un umbral numérico [12, 13]. La condición matemática del umbral T para la imagen binaria en escala de grises $f(x,y)$ se logra al etiquetar los píxeles correspondientes al objeto principal de interés con '1' y con '0' para los píxeles que están en el fondo:

$$B(x, y) = \begin{cases} 1, & \text{si: } f(x, y) > T \\ 0, & \text{si: } f(x, y) \leq T \end{cases} \quad (1)$$

Este problema se resuelve aplicando el criterio del umbral óptimo de *Otsu* [12]. Este es un método en el que se elige un valor de umbral óptimo minimizando la varianza de los niveles de gris para cada clase, que se definen como la suma ponderada de las varianzas:

$$\sigma_B^2 = \omega_0(t)\sigma_0^2(t) + \omega_1\sigma_1^2(t) \quad (2)$$

donde la probabilidad de la ocurrencia de los niveles de gris $\omega_{0,l}$ se define como:

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i) \quad \omega_1(t) = \sum_{i=t}^{L-1} p(i). \quad (3)$$

En términos de la distribución de probabilidades $p(i)$ como también de ω_k el momento de orden cero de la clase k -ésima y del momento de primer orden μ_k se asocian a los valores de los niveles de gris, hasta el k -ésimo nivel como sigue:

$$\sigma_b^2 = \sigma^2 - \sigma_w^2(t) = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \quad (4)$$

$$= \omega_0(t)\omega_1(t)[\mu_0(t) - \mu_1(t)]^2 \quad (5)$$

donde la intensidad promedio de cada clase y la intensidad promedio de toda la imagen se definen por:

$$\mu_0(t) = \sum_{i=0}^{t-1} i \cdot \frac{p(i)}{\omega_0} \quad (6)$$

$$\mu_1(t) = \sum_{i=t}^{L-1} i \cdot \frac{p(i)}{\omega_1} \quad (7)$$

$$\mu_T = \sum_{i=0}^{L-1} i \cdot p(i) \quad (8)$$

B. La región Sembrada de Crecimiento (SRG)

Para obtener un diagnóstico preciso de una mamografía digital se requieren diferentes tipos de enfoques del tejido mamario de la paciente tomados desde distintos ángulos [1-4].

La primera vista es la proyección Media Lateral Oblicua (MLO), que representa el área del músculo pectoral con diferentes niveles de gris. Por lo tanto, es importante eliminar las etiquetas innecesarias de la imagen original para evitar falsas detecciones. Sin embargo, un inconveniente de las técnicas de segmentación es la homogeneidad de las regiones de la imagen [12, 13]. En este trabajo se implementa el método de Región Sembrada de Crecimiento (SRG, del inglés, *seeded region growing*) [13]. El crecimiento de las regiones consiste en formar grupos de píxeles de la imagen para construir una imagen mejorada. Este procedimiento comienza con la elección de un conjunto de píxeles llamados semillas. Aquí se considera que la dirección de la posición del seno (izquierda o derecha) permite el recorte de los bordes en la imagen. Posteriormente, se evalúan las intensidades de los píxeles de las primeras 10 columnas del lado izquierdo y lado derecho, respectivamente mediante $\lambda_L = \sum_{y=1}^m \sum_{x=1}^n f(x,y)$ y $\lambda_R = \sum_{y=1}^m \sum_{x=N-n}^N f(x,y)$. Si la suma del lado izquierdo es menor que la del lado derecho $\lambda_L < \lambda_R = D_L$, esto indica que el pecho tiene una dirección hacia la izquierda. Por lo contrario, la dirección sería hacia la derecha. Con este paso se determina el lugar de siembra para el píxel semilla con mayor intensidad de nivel de gris. Este proceso comienza con la selección de los píxeles P , que aún no han sido asignados y que son adyacentes en al menos en una región:

$$P = \left\{ (x, y) \notin \bigcup_{i=1}^k R_i \mid N(x, y) \cap \bigcup_{i=1}^k R_i \neq \emptyset \right\} \quad (9)$$

Entonces, la región de prueba seguiría creciendo si no se detectan cambios repentinos en la intensidad de los píxeles. Posteriormente, los píxeles que no están dentro de la proximidad de la región de segundo orden $N(x, y)$ se comparan y se re-asignan a una nueva región. Es importante enfatizar que el éxito de esta técnica depende de la posición inicial donde se cultiven los píxeles semilla por primera vez.

C. Fast Slant Stack Transform

La contribución más importante en este trabajo, es la búsqueda de espículas mediante un proceso de imágenes que se interpreta como un problema de detección de estructuras lineales [6-8]. La Transformada Discreta de Radón (DRT) es muy utilizada en el análisis de imágenes de radar, sísmicas, geofísicas y médicas [11]. La característica principal de esta técnica es que, un evento lineal en el espacio coordenado se puede representar mediante un simple punto en el dominio de Radón [11]. Normalmente, la DRT es una función de $f(x,y) \in \mathbb{R}^2$, para $R_f(\rho, \theta)$ con $\rho \in \mathbb{R}$ y $\theta \in [0, \pi)$ de la siguiente forma:

$$R_f(\rho, \theta) = \iint_{-\infty}^{\infty} f(x, y) \delta(\rho - x \cdot \cos(\theta) + \dots + y \cdot \sin(\theta)) dx dy \quad (10)$$

En [11] se menciona la importancia de contar con una versión discreta de R_f a partir de $I=(I(u,v):-n/2 \leq u, v < n/2)$. La versión discreta de la función $f(x,y)$, para la $DR_f(x,y)$ de la imagen de $N \times N$ sería:

$$DR_f(\rho, \theta) = \sum_{y=-N/2}^{N/2} \sum_{x=-N/2}^{N/2} f(x,y) \delta(\rho - x \cos(\theta) - y \sin(\theta)) \quad (11)$$

A. *Averbuch et al.* presentó una versión de la Transformada de Radon que puede procesar datos digitales. Esta herramienta es mejor conocida como la *Fast Slant Stack Transform* (FSST, por sus siglas en inglés) [11]. Este enfoque se basa en agregar píxeles a lo largo de líneas de proyección con pendiente absoluta menor a "1" en función de los ejes x o y . La transformada (11) se aplica para mover puntos discretos desde el dominio espacial hacia una nueva representación de líneas discontinuas en el dominio de Radon, denominadas como linogramas. Esta variante de la DRT, se aplicó por primera vez en la ciencia sísmológica. La DRT se aplica a varias sinusoides intersectadas en un punto las cuales se transforman a un espacio $R_f(\rho, \theta)$, denominado como *sinograma*. En esta variante de la DRT, $\theta \in [-\pi/4, \pi/4]$:

$$S^1 f(\rho, \theta) = \int f(x,y) \delta(\rho - x - y \tan(\theta)) dx dy \quad (12)$$

mientras, que para $\theta \in [\pi/4, 3\pi/4]$, se tiene que:

$$S^2 f(\rho, \theta) = \int f(x,y) \delta(\rho - \cotan(\theta)x - y) dx dy \quad (13)$$

Ajustando a un solo objeto Sf se define como:

$$Sf(\rho, \theta) = \begin{cases} S^1 f(\rho, \theta) & \theta \in [-\pi/4, \pi/4] \\ S^2 f(\rho, \theta) & \theta \in [\pi/4, 3\pi/4] \end{cases} \quad (14)$$

D. Filtros de Espiculación Radial

Combinando la FFST con un conjunto de filtros de espiculación radial (SRF) de coincidencia directa es posible detectar distintos patrones de espiculación y distorsión en la arquitectura del tejido mamario. Estos filtros se propusieron por primera vez en *M. Sampat et al.*, [6-8]. Sin embargo, en este trabajo se descomponen en dos funciones ortogonales, denominados como filtro de espiculación radial coseno (*CSRf*) y filtro radial sinusoidal (*SSRF*) como:

$$CSRf = G(r_0, \sigma) * \cos(\theta * f) \quad (15)$$

$$SSRF = G(r_0, \sigma) * \sin(\theta * f) \quad (16)$$

la envolvente Gaussiana es:

$$G(r_0, \sigma) = \exp(-(r - r_0)^2 / (2 * \sigma^2)) \quad (17)$$

con

$$r = \sqrt{x^2 + y^2}$$

donde en las expresiones anteriores r_0 es el tamaño del parámetro en píxeles, $\theta = \arctan(y/x)$, σ es la desviación *standard* en píxeles y f es la frecuencia en ciclos por segundo. Posteriormente, se aplican iterativamente nuevos bancos de filtros *CSRf* y *SSRF* para obtener una detección precisa con las mismas propiedades de σ , r_0 y f .

III. PROCESAMIENTO NUMÉRICO DE MAMOGRAFÍAS

Las mamografías digitales procesadas en este trabajo se obtuvieron de la base de datos *The mini-MIAS database of mammograms* [15]. El tipo de tejidos en estudio son el graso, graso-glandular y el denso-glandular. La serie de pruebas se realizan en mamografías con masas normales, espiculadas y con presencia de distorsión de arquitectura del tejido.

La Tabla I muestra en parte ciertos casos de estudio aquí tratados según el tipo de tejido y el grado de severidad o trastorno. Las imágenes se digitalizaron en un borde de píxeles de 200 micrones y se recortaron a un tamaño de 1024×1024 píxeles en un formato de Mapa Gris Portátil (PGM) [15].

A. Filtrado de Imágenes

El filtrado se aplica para suavizar los datos en las imágenes, reducir el ruido y mejorar el contraste usando el filtro de la mediana con un núcleo de conectividad de 3×3. Con esto se minimiza el problema de los bordes con la convolución utilizando una máscara de filtro y relleno de ceros.

La Fig. 1 (a) muestra la mamografía original en la que se puede observar el ruido presente en forma de líneas verticales cerca del área del pezón de la paciente y de las líneas horizontales en la parte inferior derecha. La Fig. 1 (b) muestra la secuencia de la misma imagen después de aplicar el filtro de la mediana.

B. Eliminación de Objetos

En un paso inicial se deben eliminar los objetos innecesarios en las mamografías como datos del paciente o la prueba acerca del tipo de mamografía: área, perímetro y el centroide de todos los objetos presentes. Estos podrían interferir en las siguientes etapas de procesamiento.

La Fig. 2 (a) muestra una mamografía original con objetos (etiquetas y números) [15]. Entonces, se aplica un umbral de dos niveles para obtener una imagen binaria mejorada. Por lo tanto, un valor de umbral T se calcula a través del método de *Otsu* obteniendo la nueva imagen binaria en la Fig. 2 (b) [12].

TABLA I
CASOS DE ESTUDIOS TOMADOS DE MINI-MIAS [15]

Tipo de Estudio	Anormalidad		Tipo de tejido			Total de Mamografías
	Malig no	Benig no	Gras o	Graso-Glandular	Denso-Glandular	
Masas espiculadas	8	11	5	7	7	19
Distorsión de la arquitectura	10	9	6	6	7	19
Casos normales	0	31	10	11	10	31

Para identificar los elementos conectados en la imagen binaria (según su proximidad espacial), se utiliza el algoritmo de *Run length Encoder* (RLE, por sus siglas en inglés) de *Matlab R2016a (9.0.0.34)*. En la Fig. 3 (a) se muestran los elementos conectados a un mismo objeto digital dentro de un rectángulo en color verde. La dimensión del área de estos elementos conectados entre sí, se estima para poder definir los objetos que no pertenecen al pecho de la paciente mediante una comparación entre áreas. El área más grande es la región de la mama la cual en promedio tiene 78,600 píxeles.

La Fig. 3 (b) muestra los objetos detectados en rectángulos en color rojo con un área menor a la del pecho. Después, estos objetos se eliminan de la imagen binaria. La Fig. 3 (c) muestra la imagen binaria obtenida del proceso de extracción de objetos, después de multiplicar la imagen binaria con la imagen original filtrada. Posteriormente, se segmenta con respecto al fondo, lo que da como resultado un área del pecho bien delimitada. Esto se observa en la Fig. 3 (d), la cual se usa posteriormente para segmentar el músculo pectoral del seno.

C. Segmentación del Músculo Pectoral

Para segmentar el músculo pectoral se implementa un algoritmo basado en la región sembrada de crecimiento (SRG, tratada en la sección 2.2 de este trabajo) [13]. El SRG tiene un alto rendimiento numérico cuando las regiones de la imagen a eliminar se pueden delimitar previamente.

La Fig. 4 (a) muestra una imagen binaria del seno con parte del músculo pectoral segmentado. Posteriormente, esta imagen se multiplica con la imagen original filtrada en la Fig. 3 (d) para obtener la Fig. 4 (b). Esta última se utiliza posteriormente en el proceso de mejora en la identificación de espículas.

D. Mejoramiento en Identificación de Espículas

Este proceso se realiza aplicando la FSST en la Fig. 4 (b) en la cual cada línea dentro de la imagen se representa apilando otras líneas que se intersectan en un punto común. Se observa que dichas líneas tienen un comportamiento "*Bajo-Alto-Bajo*" por la presencia de espículas [6-8]. El término "*Bajo*" corresponde a un valor menor en magnitud y "*Alto*" corresponde a valores mayores en intensidad dentro del área de estudio. Posteriormente a la medición de 19 mamografías con distintas masas espiculadas se determinó que en promedio el grosor (en valores de píxeles) para una espícula es de 12 píxeles.

La Fig. 5 (a) muestra la imagen de aplicar la DRT con un valor de umbral global según el método de *Otsu* [12]. En esta

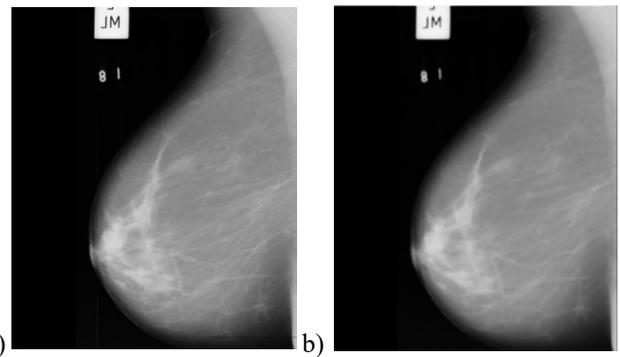


Fig. 1. Imagen filtrada. (a) Imagen original (mdb195 tomada de [15]). (b) Imagen con el filtro de la mediana aplicado.

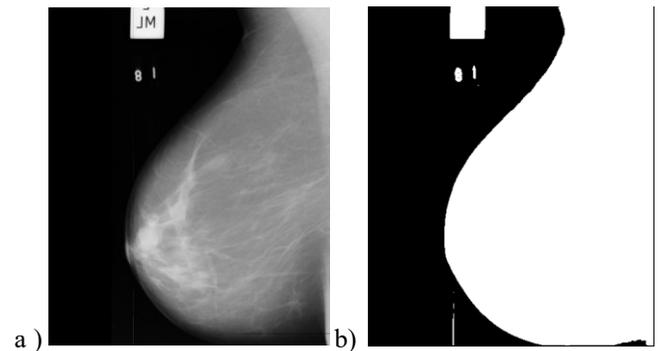


Fig. 2. Imagen conteniendo objetos. (a) Imagen original. (b) Imagen binaria.

imagen se observa que todavía existen regiones de tejido que no cumplen con la restricción de grosor y comportamiento lineal. Entonces, se debe aplicar por segunda vez un umbral de discriminación. Aunque se continúan observando cierto tipo de discontinuidades, pero ahora tienen un grosor específico correspondiente al valor inicial del umbral óptimo aplicado

En la Fig. 5 (b) se muestran los resultados de aplicar la DRT inversa con un valor de umbral óptimo T definido en [11, 12]. En esta imagen los tejidos tienen ahora una cierta forma y grosor correspondientes al tipo de espículas detectadas en esta región de estudio de la mama.

Con el objetivo de obtener una imagen mejorada con ruido reducido y espículas bien definidas, se aplica entonces la transformada FSST inversa en conjunto con filtros con umbral de selección óptima para la reconstrucción de la imagen en la Fig. 5 (c) [11].

Se observa que el método de la FSST es selectivo y eficiente para reducir el ruido con filtros lineales despejando las espículas en el área, lo cual ayuda a reducir el número de falsos positivos en la detección.

E. Detección de Masas Espiculadas

Considere la imagen la Fig. 5 (c) como prueba inicial para el algoritmo de detección de masas espiculadas mediante la técnica aquí combinada de filtros de espiculación radial y la FSST.

Este tipo de filtros se encuentran configurados de forma adaptativa lo que aumenta la precisión en la identificación de

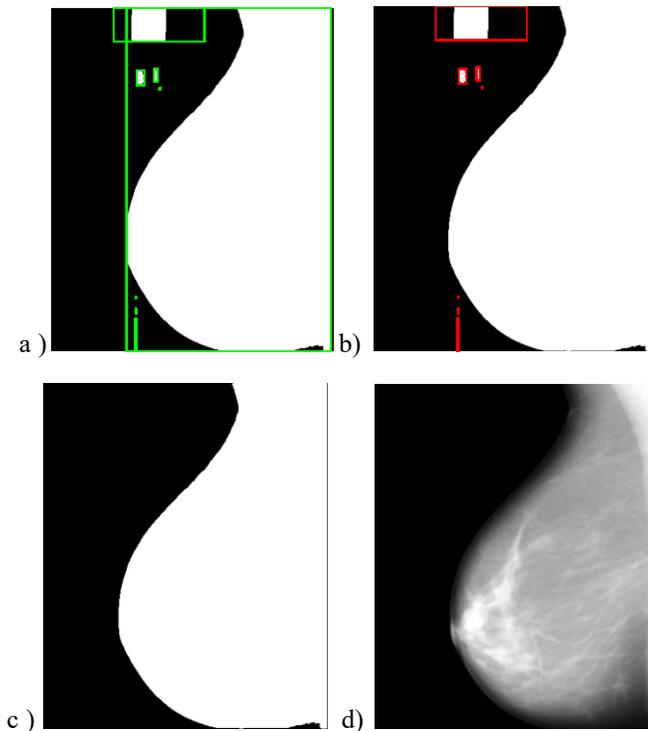


Fig. 3. Resultados con el RLE en la imagen binaria. (a) Objetos detectados en la imagen. (b) Objetos con área menor que el seno. (c) Imagen binaria con objetos borrados. (d) Región del seno segmentada.

cualquier tipo de masa estelar o espiculada en la región de estudio. Una relación promedio del tamaño de las masas diagnosticadas en este trabajo a partir de la base de datos *mini-MIAS* se determina como de 30 píxeles [15]. Por lo que los bancos de filtros están diseñados para este tipo de datos, clasificando el tipo de masa a través de variar sus parámetros a diferentes escalas, tamaños y resoluciones.

La Fig. 5 (c) se filtra nuevamente para detectar las formas lineales del tejido. Cada región en la imagen está pre-definida por una escala y magnitud para los bancos de filtros.

Una vez identificadas las N posibles masas existentes, se procede a obtener las coordenadas de sus puntos máximos, los cuales se marcan en la imagen mamográfica que se muestra en la Fig. 6 (a) para su análisis.

La Fig. 6 (b) muestra la mamografía resultante con 4

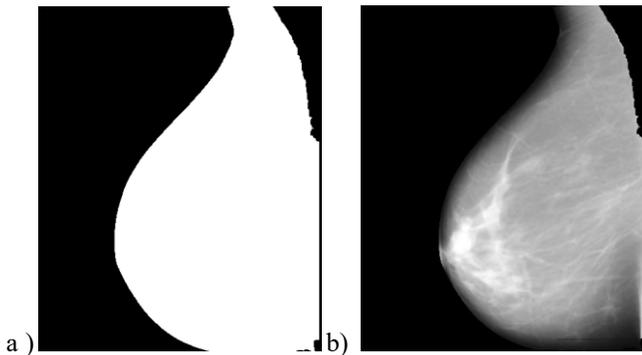


Fig. 4. Imagen del músculo pectoral segmentado. (a) Imagen segmentada binaria. (b) Imagen mejorada con músculo pectoral segmentado.

posibles masas espiculadas detectadas. Estas aún no se han confirmado como casos patológicos de estudio. Sin embargo, los cuatro puntos en color azul en la imagen son una evidencia de la convergencia del método propuesto para detectar las espículas dentro de la mama. Se marca entonces, un círculo de color amarillo indicando la relación aproximada del tamaño de la masa, mientras que los círculos en color azul determinan la posición de las masas detectadas en la región de estudio.

A partir de las regiones detectadas, se procede a estimar la región de interés (ROI, por sus siglas en inglés) de las masas.

Con el fin de extraer las principales características para cada tipo de lesión, se agrega una etapa de clasificación adicional en la cual se debe determinar si la región marcada corresponde a un tumor maligno, o a una lesión benigna del tejido.

La Fig. 6 (c) muestra la ROI de la tercer masa detectada, la imagen corresponde a la segmentación de la lesión obtenida con el método de *Otsu* [12].

La dimensión de la región de interés es de 80×80 píxeles. Este tipo de lesión coincide con el diagnóstico obtenido en la base de datos *mini-MIAS* [15]. La región de la lesión se muestra identificada en la Fig. 6 (d), en la cual se observa que el método enmarca de forma adecuada las estructuras que se irradian hacia el exterior desde las "espículas" en la masa correspondiente [15].

IV. RESULTADOS

Los resultados en este trabajo se obtienen variando el nivel del umbral de detección en las regiones de las mamografías analizadas. Esto se determina mediante las curvas FROC de rendimiento (por sus siglas en inglés, *free response operating characteristic*) [5-12, 14, 15].

Para obtener dichas curvas se traza primero una curva de referencia de la probabilidad de aciertos en función del número promedio de falsos positivos obtenidos en el estudio, los cuales se encuentran definidos en cada imagen en términos de sensibilidad [5-10, 14, 15].

La precisión en la validación de la técnica propuesta se obtiene primero con respecto a los resultados obtenidos por *Sampat et al.* y *Karssemeijer et al.* [6-8, 14]. Estos trabajos son una referencia en la aplicación de las transformadas y filtros aplicados con métricas bien reconocidas de rendimiento numérico para la DRT y FSST, de acuerdo con la validación

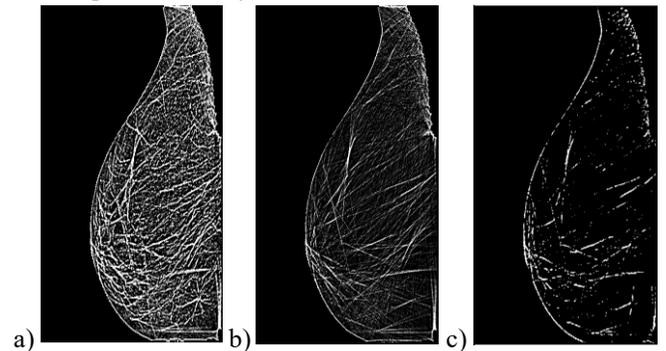


Fig. 5. Ajuste de transformadas inversas. (a) DRT inversa con umbral global. (b) DRT inversa con umbral óptimo. (c) FSST inversa con umbral óptimo.

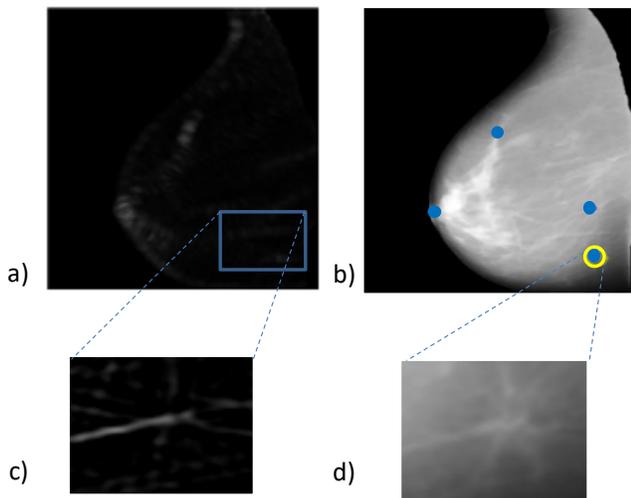


Fig. 6. Detección de masas espiculadas (mdb195). (a) Imagen obtenida, (b) Imagen con posibles masas detectadas. (c) ROI con masa espiculada detectada. (d) Masa diagnosticada como positivo a espiculas.

crucada propuesta previamente por Rangayyan *et al.* [9].

A. Validación de las Curvas FROC

La técnica aquí propuesta se evalúa con respecto a 50 casos tomados de la base de datos *mini-MIAS* que se analizan en este trabajo [15]. De los casos analizados, 10 tienen problemas de distorsión arquitectónica del tejido mamario (AD), 9 casos evidencian la presencia de masas espiculadas (SM) y 31 casos corresponden a tejidos sanos.

La Fig. 7 (a) muestra las curvas FROC resultado del análisis comparando el desempeño numérico de la DRT y FSST [6-15]. En estas curvas se puede observar una clara reducción en el número de falsos positivos mediante la técnica FSST, con una sensibilidad del 80% presentando 8.5 falsos positivos por imagen (FPi). Por otro lado, cuando se aplica la DRT se tiene un índice de 11 FPi con una sensibilidad similar.

Sin embargo, si se combinan el máximo nivel de etapas de agrupación con la técnica FSST, se observa una sensibilidad del 92.1% con 10.5 FPi, mientras que con la DRT se obtiene el 87.4% con 12 FPi, como lo muestra la Fig. 7 (b).

En la Fig. 7 (c) se muestran las curvas FROC para los casos positivos diagnosticados con AD. En esta figura la técnica FSST produce resultados más precisos que la DRT al reducir el número de falsos positivos. En este caso, la sensibilidad de la FSST es del 80% y con un índice de 10.5 FPi mientras que la DRT requiere de un 13.5 FPi con una sensibilidad mayor.

La Fig. 7 (d) combina la etapa de agrupación máxima con la FSST reduciendo el número de FPi con una sensibilidad del 85% para un índice del 8.5 FPi. Esta reducción es más alta en comparación con la obtenida con la DRT del 81% pero con aproximadamente un 3.5 de FPi menor.

B. Desempeño Numérico

La Fig. 8 muestra el desempeño numérico de los métodos implementados con la DRT y FSST con respecto al tiempo de

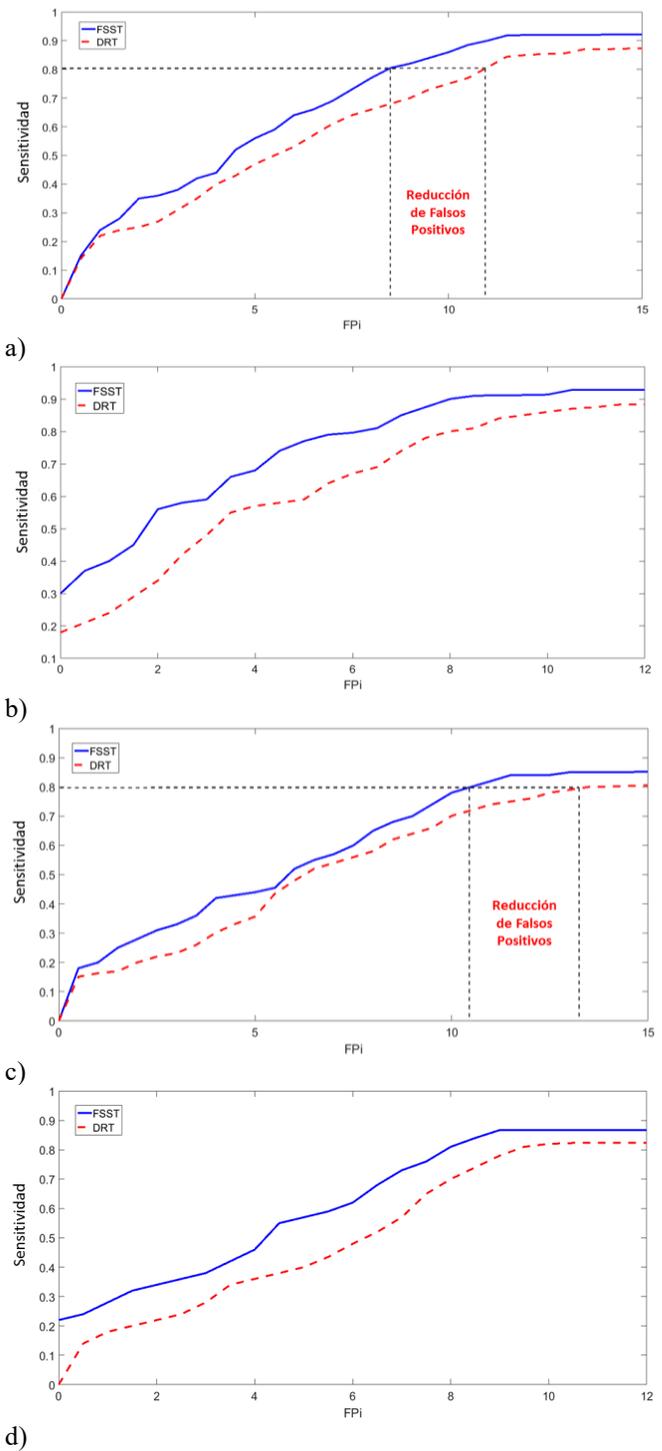


Fig. 7. Curvas FROC. a) Masa Espiculada (SM). b) SM+agrupamiento de regiones máximas c) Distorsión arquitectónica del tejido mamario (AD). d) AD+agrupamiento de regiones máximas.

procesamiento $t(s)$ en el eje vertical de la imagen. En el eje radial externo se correlacionan la complejidad de cada técnica (la cantidad de operaciones realizadas) por cada algoritmo en la estación de trabajo utilizada. [6-8, 11].

El rendimiento numérico se obtuvo de una estación de trabajo (WS) corriendo *Matlab R2016a* (9.0.0.34) Procesador

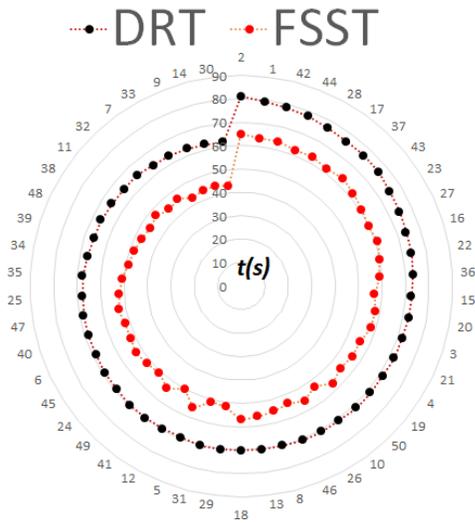


Fig. 8. Desempeño numérico del DRT y FSST. Tiempo de procesamiento (eje vertical) y complejidad (eje radial externo).

Intel (R) Core (TM) i7-4770 CPU de 3.40GHz Memoria RAM 16 GB, Sistema Operativo Windows 7 de 64 bits (win64).

De estas curvas es evidente que el tiempo requerido con la FSST es notablemente menor que el utilizado por la DRT en este tipo de aplicación.

La Tabla II muestra el desempeño computacional de dichas técnicas. Estas pruebas numéricas se obtuvieron para el caso de la detección de anomalías en los tejidos adiposo, grasoglandular y denso-glandular obtenidos de la Tabla I.

C. Eficacia del método

El análisis de datos permite obtener una comprensión del comportamiento y tendencias de las curvas FROC en la Fig.7 aplicado a los 50 casos bajo estudio de *mini-MIAS* [15].

Por el tipo de distribución cuasi-exponencial de los datos en

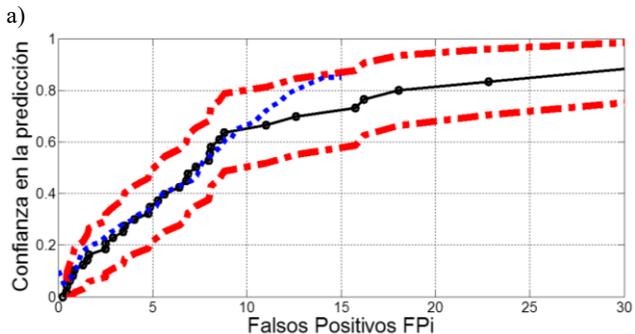
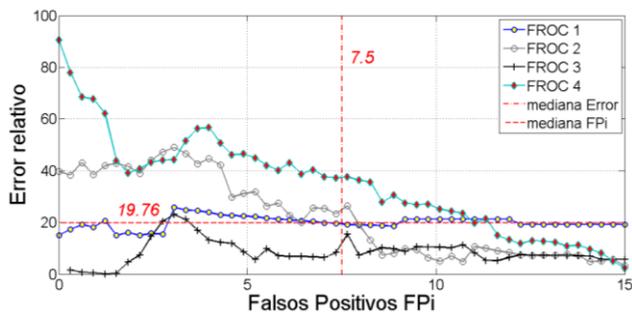


Fig. 9. Eficacia del método. a) Error relativo. b) Modelo pronosticado

la Fig. 9a se evalúa la eficacia del método FSST con respecto a la DRT mediante el criterio del error relativo porcentual ($\%Error_{rel} = [1 - (f_{DRT}/f_{FSST})] \times 100$) donde f_{DRT} es cada una de las curvas FROC calculadas con la DRT, mientras que f_{FSST} son las curvas FROC calculadas con la FSST.

De la curva FROC 3 con el menor error relativo porcentual en la Fig. 9a se obtiene un modelo de pronóstico preciso que permite extender el análisis de 50 a 100 mamografías duplicando el rango de 15 FPi a 30 FPi como se muestra en la Fig. 9b.

En la Fig. 9b se compara el comportamiento de los modelos inicial para 15 FPi (en la línea azul punteada) y el modelo extendido (línea negra continua) para 30 FPi según la generalización de la distribución exponencial de Weibull [16]. Ambos modelos se encuentran dentro del intervalo de confianza (líneas rojas de frontera) las cuales enmarcan la probabilidad de falla en la estimación. Esto implica que es posible extender el análisis a otros conjuntos de datos mayores de mamografías digitales sin aumentar la tasa de riesgo por error en el modelo pronosticado [16].

D. DISCUSIÓN Y CONCLUSIONES

En este artículo se presenta una técnica de apoyo para métodos basados en clasificación de multi-dominio y multi-objetivo en la detección oportuna del cáncer de mama en mamografías digitales.

La técnica aquí propuesta basada en la FSST es eficiente y numéricamente precisa, logrando un índice de sensibilidad para la detección de masas espiculadas del 92.1% en comparación con la DRT del 87.4%.

En el problema de detección de distorsión de la arquitectura del tejido, la sensibilidad alcanzada con la FSST es del 85% mientras que con la DRT es del 81%. Por lo tanto, en este sentido, el uso de la FSST es una opción prometedora con la ventaja de no necesitar una etapa de preclasificación.

Por otro lado, un inconveniente de aplicar la técnica aquí propuesta ocurre cuando se procesa la información que rodea el área del pezón en la mamografía digital. Cuando esta área no se encuentra segmentada previamente, puede dar lugar a un falso positivo, debido a la similitud que se tiene con un tejido sano. Sin embargo, esto se puede evitar aplicando la segmentación de imágenes sucesivas.

Los casos de estudio aquí tratados arrojaron resultados más precisos que con respecto a otros algoritmos previamente propuestos para el caso del tejido con un alto contenido grasoglandular.

TABLA II
DESEMPEÑO COMPUTACIONAL

Método utilizado	Sensibilidad Máxima	Sensibilidad Promedio	FPi	Tiempo-CPU
Sampat et. al.	87%	83.6	8.2	70.4s
Karssemeijer et. al.	96%	82.6	2.86	93.6s
Técnica Propuesta	92%	86	5	54.2s

En cambio, el peor desempeño numérico se obtuvo en el caso del tejido denso-glandular. Lo cual se debe a los niveles

de color gris combinados en las espículas y presentes también en las glándulas mamarias. Tomando en cuenta estos aspectos, los resultados en estudios con procesamiento digital de imágenes pueden confundirse con ciertos tejidos sanos. Sin embargo, este problema se puede resolver agregando una etapa de pre-clasificación a la mamografía original.

REFERENCIAS

- [1] World Health Organization Website (2019). *Breast cancer: prevention and control report*. [Online]. Available: <http://www.who.int/cancer/detection/breastcancer/en/>
- [2] World Health Organization Website (2019). *International Agency for Research on Cancer. Cancer Today: Data visualization tools for exploring the global cancer burden in 2018*. [Online]. Available: <http://gco.iarc.fr/today/home>.
- [3] INEGI-México (2018). "Estadísticas a propósito del día mundial contra el cáncer (4 de febrero)" [Online]. Available: https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2018/cancer2018_Nal.pdf
- [4] NOM-041-SSA2-2011. *Norma Oficial Mexicana para la prevención, diagnóstico, tratamiento, control y vigilancia epidemiológica del cáncer de mama (2011)*. [Online]. Available: http://dof.gob.mx/nota_detalle.php?codigo=5194157&fecha=09/06/2011&print=true. Accessed on: February 12, 2019.
- [5] Minavathi, S. Murali, M. Dinesh, "Model Based Approach for Detection of Architectural Distortions and Spiculated Masses in Mammograms", *Int. Jour. of Comp. Sc. and Eng.*, vol. 3, no. 11, pp: 3534-3546, Nov. 2011.
- [6] M. P. Sampat, G. Whitman, M. Markey, A. Bovik, "Evidence Based Detection of Spiculated Masses and Architectural Distortions", *Proc. of SPIE-Medical Imag.*, vol. 5747, no. 1, pp: 26-37, 2005. DOI: 10.1117/12.595331
- [7] M. P. Sampat, G. J. Whitman, A. C. Bovik and M. K. Markey, "Comparison of algorithms to enhance spicules of spiculated masses on mammography", *Journal of Digital Imag.*, vol. 21, no. 1, pp. 9-17, March 2008. DOI:10.1007/s10278-007-9015-x
- [8] M. P. Sampat, A. C. Bovik, G. J. Whitman, and M. K. Markey, "A model-based framework for the detection of spiculated masses on mammography", *Med. Phys.*, vol. 35, no. 5, pp.2110-2123, May 2008. DOI: 10.1118/1.2890080
- [9] R. Rangayyan, S. Banik, J. Desautels, "Detection of Architectural Distortion in Prior Mammograms via Analysis of Oriented Patterns", *Journal of Visualized Exp.*, vol. 78, no. e50341, pp:1-17, Aug. 2013. DOI:10.3791/50341
- [10] *Computerized Analysis of Mammographic Images for Detection and Characterization of Breast Cancer*, Paola Casti, Arianna Mencattini, Marcello Salmeri and Rangaraj Rangayyan, eds., Morgan & Claypool publishers, 2017. [Online]. Available: DOI:10.2200/S00776ED1V01Y201704BME056
- [11] A. Averbuch et al., "Fast Slant Stack: A notion of Radon Transform for Data in a Cartesian Grid which is Rapidly Computable, Algebraically Exact, Geometrically Faithful and Invertible". Department of Statistics, Stanford University, California, USA. Report Number 2001-11, May 2001. [Online]. Available: <https://statistics.stanford.edu/research/fast-slant-stack-notion-radon-transform-data-cartesian-grid-which-rapidly-computable>
- [12] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", *IEEE Trans. on Syst. Man and Cyb.*, vol. smc-9, no. 1, pp: 62-66, Jan. 1979. Accessed on: November 27, 2019, DOI: 10.1109/TSMC.1979.4310076
- [13] J. Fan *et al.*, "Seeded region growing: an extensive and comparative study", *Pattern Recognition Letters*, vol. 26, no. 8, pp: 1139-1156, Jan. 2005. Accessed on: November 27, 2019, DOI:10.1016/j.patrec.2004.10.010
- [14] N. Karssemeijer and G. M. te Brake, "Detection of stellate distortions in mammograms", *IEEE Trans. on Med Imag.*, vol. 15, no. 5, pp: 611-619, Oct. 1996. Accessed on: November 27, 2019, DOI: 10.1109/42.538938
- [15] J. Suckling *et al.*, (1994). *The Mammographic Image Analysis Society Digital Mammogram Database*. Presented at Excerpta Medica. Int. Congress Series 1069. [Online]. Available: <http://peipa.essex.ac.uk/info/mias.html>

[16] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, 2nd edition, Hoboken, NJ, John Wiley & Sons, Inc. 2003.



F. A. Uribe (M'98–SM'15). Received the B. Sc. and the M. Sc. Electrical Engineering degrees, both from The University of Guadalajara, in 1994 and 1998, respectively. During 2001 he was a visiting researcher at The University of British Columbia, B. C. Canada. In 2002 received the Dr. Sc. Electrical Engineering degree from The Center for

Research and Advanced Studies of México (CINVESTAV-Campus Guadalajara). The dissertation was awarded with The Arturo Rosenblueth prize as the author of the best doctoral dissertation of Mexico in 2002. From 2003 to 2006 Dr. Uribe, was a full professor with The Electrical Graduate Program at The Autonomous University of Nuevo Leon, México. From May 2006, Dr. Uribe joined The Electrical Engineering Graduate Program at The University of Guadalajara, México, where he is currently a full-time researcher. Since 2004, Dr. Uribe is a member of the National System of Researchers of México. His primary interests are the signal and image processing techniques for harmonic and transient analysis of smart grids.



Edwin Lozada (Student M'2016) was born in Oaxaca de Juarez, México in April 2, 1988. He received the B. Sc. Electrical Engineering degree from National Technologic of México campus Oaxaca. In 2017 received the M. Sc. degree in Electrical Engineering from the University of Guadalajara.



J. R. Morales He received the Mechanical and Electrical Eng. M. Sc. degrees of Elect. Eng. and PhD in the Electronic and Computing program all of them from the University of Guadalajara, in 2010, 2013 and 2017, respectively. He was a visiting researcher at the Western Michigan University, USA. He is an

assistant professor with the Department of Mechanical and Electrical Engineering, University of Guadalajara, México. His main research interests are the Electromagnetic Transients in Underground Cables.



Alma Y. Alanis was born in Durango, Durango, Mexico, in 1980. She received the B. Sc. degree from Instituto Tecnológico de Durango (ITD), Durango Campus, Durango, Durango, in 2002, the M.Sc. and the Ph.D. degrees in electrical engineering from the Advanced Studies and Research Center of the National Polytechnic Institute (CINVESTAV-IPN),

Guadalajara Campus, Mexico, in 2004 and 2007, respectively. Since 2008 she has been with University of Guadalajara, where she is currently a Chair Professor in the Department of Computer Science and since 2016 is dean of PhD Program of Electronic and Computer Sciences. She is also member of the Mexican National Research System (SNI-2) and member of the Mexican Academy of Sciences. She has published papers in recognized International Journals and Conferences, besides five International Books. She is a Senior Member of the IEEE and Subject and Associated Editor of the Journal of Franklin Institute (Elsevier) and Intelligent Automation & Soft Computing (Taylor & Francis), moreover she is currently serving on a number of IEEE and IFAC Conference Organizing Committees. In 2013 she receives the grant for women in science by L'Oreal-UNESCO-AMC-CONACYT-CONALMEX. In 2015, she receives the Research Award Marcos Moshinsky. Since 2008 she is member for the Accredited Assessors record RCEA-CONACYT, evaluating a wide range of national research projects, besides she has belonged to important project evaluation committees of national and international research projects. Her research interest centers on neural control, backstepping control, block control, and their applications to electrical machines, power systems and robotics.



Nancy Arana-Daniel received the B. Sc. Degree from the University of Guadalajara in 2000, and her M. Sc. and Ph.D. degrees in electric engineering with the special field in computer sciences from Research Center of the National Polytechnic Institute and Advanced Studies, CINVESTAV, in 2003 and 2007 respectively. She is currently a research

fellow at the University of Guadalajara, in the Department of Computer Science México, where she is working at the Laboratory of Intelligent Systems and the Research Center for Control Systems and Artificial Intelligence. She is IEEE Senior member and a member of National System of Researchers (SNI-1). She has published several papers in International Journals and Conferences and she has been technical manager of several projects that have been granted by the Nacional Council of Science and Technology (CONACYT). Also, she has collaborated in an international project granted by OPTREAT”, She is Associated Editor of the Journal of Franklin Institute (Elsevier). Her research interests focus on applications of Geometric Algebra to Geometric computing, machine learning, bio-inspired optimization, pattern recognition and robot navigation.