

# Procesamiento de imágenes de electroforesis bidimensional: una revisión

## Two-dimensional electrophoresis image processing: a review

G. A. Villegas-Rivera , M. C. Torres-Madronero , S. Röthlisberger-Booth,  E. Delgado-Trejos 

**Abstract**— This paper presents a review of the techniques and algorithms applied to the processing and analysis of two-dimensional gel electrophoresis (2DGE) images. With two-dimensional gel electrophoresis, it is possible to separate hundreds of proteins in one gel, revealing a distinctive pattern. An adequate analysis of these images is dependent on the protein spots in the image being correctly detected, as mistakes in this phase of the analysis could lead to the detection of false proteins, or to important low abundance proteins being overlooked, which can skew the results of the analysis. Segmentation techniques are used to separate proteins from the background and to find anomalies. The techniques used for 2DGE image segmentation are classified as: edge detection based, morphological, thresholding, and region-based. Many proteomic studies require the fusion or registration of several images to identify and compare patterns from different samples. For the fusion process, it is possible to use the original image or the segmented image. Despite significant advances in 2DGE image processing, there are no fully automated techniques found in the literature. The commercially available tools for processing and analyzing 2DGE images require the user to set parameters based on expertise, therefore the results of the analysis depend on the correct selection of parameters. In this paper, we review several techniques for processing 2DGE images, with special attention given to segmentation and fusion techniques.

**Index Terms**— Image processing, proteomic analysis, two-dimensional gel electrophoresis.

**Resumen**— Este artículo presenta una revisión de las técnicas y algoritmos aplicados al procesamiento y análisis de imágenes de electroforesis bidimensional (2DGE). La electroforesis bidimensional permite separar cientos de proteínas en un único gel, mostrando un patrón característico. En el análisis de imágenes es importante una correcta detección de las proteínas presentes, ya que cualquier error en esta etapa puede llevar a la detección de falsas proteínas, o a obviar proteínas importantes, pero de baja abundancia, lo cual afectaría los resultados del análisis. Técnicas de segmentación son empleadas para separar las proteínas del

fondo y encontrar anomalías. Los métodos empleados para la segmentación de imágenes 2DGE se pueden clasificar como: basados en detección de bordes, morfológicos, umbralización y basados en regiones. En muchos estudios proteómicos se hace necesario la fusión o registro de imágenes para la identificación y comparación de patrones de varias muestras diferentes. Para este proceso de fusión se pueden usar las imágenes originales o los resultados de la segmentación. A pesar de los avances significativos en el campo de procesamiento de imágenes 2DGE, no se encuentran en la literatura métodos completamente automatizados. Las herramientas comerciales disponibles para el análisis y procesamiento de imágenes 2DGE requieren que el usuario ajuste por su experiencia los parámetros de ajuste, de los cuales dependen los resultados arrojados por el software. Este artículo revisa diferentes técnicas para el procesamiento de las imágenes 2DGE. Especial atención es dada a las técnicas de segmentación y registro de imágenes.

**Palabras claves**— Procesamiento de imágenes, análisis proteómico, electroforesis bidimensional.

### I. INTRODUCTION

La proteómica caracteriza y compara el perfil de proteínas expresadas por una célula, tejido u organismo en un momento o condición particular (proteoma). El análisis sistemático del proteoma permite identificar proteínas que están correlacionadas con determinados estados fisiológicos [1]. El análisis proteómico se puede dividir en tres etapas: la separación de complejas mezclas de proteínas, la identificación de estas proteínas y el análisis de los datos [2]. Para la separación de las proteínas, la electroforesis bidimensional en gel (2DGE) es una de las técnicas más comúnmente utilizadas a nivel mundial, debido al gran número de proteínas que se puede analizar en un único gel, a la capacidad de detectar proteínas de alto peso molecular y de diferenciar isoformas de una misma proteína que ha sufrido modificaciones [1]. En esta

This manuscript was sent on November 30, 2018 and accepted on March 25, 2019. This work was supported by the Instituto Tecnológico Metropolitano (ITM) of Medellín, Colombia, under grant P14227.

Gerardo Antonio Villegas-Rivera is with the AEyCC Research Group, MIRP Lab (Parque i), Instituto Tecnológico Metropolitano (ITM), Cl 54A No. 30-01, Medellín, Colombia (e-mail: gerardo.villegas@epm.com.co).

María Constanza Torres-Madronero is with the AEyCC Research Group, MIRP Lab (Parque i), Instituto Tecnológico Metropolitano (ITM), Cl 54A No. 30-01, Medellín, Colombia (e-mail: mariatorres@itm.edu.co).

Sarah Röthlisberger-Booth is with the GI2B Research Group, Biomedical Sciences Lab (Parque i), Instituto Tecnológico Metropolitano (ITM), Cl 73 No. 76 A 354, Medellín, Colombia (e-mail: sarahrothlisberger@itm.edu.co).

Edilson Delgado-Trejos is with the CM&P Research Group, AMYSOD Lab (Parque i), Instituto Tecnológico Metropolitano (ITM), Cl 73 No. 76 A 354, Medellín, Colombia (e-mail: edilsondelgado@itm.edu.co).

Corresponding author: M. C. Torres-Madronero (Phone: +57 4 460 07 27 Ext. 5513, e-mail: mariatorres@itm.edu.co).

técnica la separación de proteínas se logra en dos etapas. En primera instancia las proteínas son separadas en función de su punto isoeléctrico; en la segunda dimensión las proteínas son separadas de acuerdo a su peso molecular [3]. A partir de esta separación se genera un arreglo de proteínas que es digitalizado mediante el escaneo del gel. Estas imágenes permiten el análisis e identificación de las proteínas presentes en la muestra analizada [4], [5], a partir de lo cual se pueden determinar cambios en la migración de las proteínas y abundancias de las mismas [6], así como comparar patrones proteicos en diferentes muestras o bajo condiciones distintas [1].

Para el análisis de este tipo de imágenes existen varios softwares comerciales, tales como: PDQuest (BioRad) [7], ImageMaster (GE Healthcare) [8], Melanie II [9], ProteomeWeaver [10], GELLAB [29] y Delta2D [11]. Estos programas usan técnicas básicas de pre-procesamiento para mejorar la apariencia de las imágenes, así como técnicas de segmentación y registro que asisten en la detección de las proteínas [12]. Sin embargo, estas herramientas son en su mayoría costosas y requieren la intervención manual de un experto, por lo cual el procesamiento de imágenes 2DGE está lejos de ser un proceso automático [13]. El avance significativo en las capacidades computacionales para el procesamiento de imágenes y el creciente interés por el análisis proteómico basado en imágenes 2DGE, crean la necesidad de estudiar y desarrollar una metodología que permita procesar la información contenida en múltiples imágenes 2DGE de forma automática y confiable.

En el análisis de imágenes 2DGE se presentan muchos problemas para obtener una buena caracterización de sus componentes, razón por la cual se deben utilizar técnicas de procesamiento automático para reducir el ruido, corregir el fondo, disminuir el efecto de “streaking” (rayas verticales y horizontales), y mejorar la detección de proteínas de baja abundancia (puntos difusos o con baja intensidad). Este artículo presenta una revisión sobre las corrientes, técnicas y algoritmos empleados para el procesamiento y análisis de imágenes 2DGE. Inicialmente, se describe el análisis proteómico y el proceso para la adquisición de estas imágenes. El procesamiento de imágenes 2DGE incluye: pre-procesamiento, detección de proteínas y fusión de imágenes. Esta revisión presenta las principales técnicas y algoritmos empleados en cada una de estas etapas.

## II. CONTENIDO

### A. *Análisis Proteómico Mediante Electroforesis Bidimensional*

La proteómica es el estudio de las proteínas expresadas por una célula, tejido u organismo en un momento dado [14], [15]. Hoy en día es claro que el estudio de los genes es muy valioso, pero que aislado de un contexto en el que miles de proteínas juegan un papel importante, esta información no puede ser interpretada adecuadamente y sin sesgo. Las proteínas ejecutan y controlan la gran mayoría de los procesos celulares, por lo tanto, la expresión anormal de proteínas es un indicador de un

estado patológico, y su estudio proporciona un análisis más inmediato de la fisiología celular [1].

La electroforesis bidimensional es una técnica de alta resolución que permite la separación de complejas mezclas de proteínas, para su posterior análisis [1]. Como lo dice su nombre, la separación ocurre en dos dimensiones o etapas consecutivas [34]. En primera instancia las proteínas son separadas en función de su carga a lo largo de un gel con gradiente de pH [1]. Cada proteína avanza en el campo eléctrico hasta alcanzar un valor de pH donde su carga neta es igual a cero (punto isoeléctrico). Luego, en la segunda dimensión las proteínas son separadas de acuerdo a su peso molecular. A partir de esta separación se genera un patrón de proteínas que es digitalizado como una imagen mediante el escaneo del gel.

Cada proteína se caracteriza por su posición vertical y horizontal en la imagen 2DGE, así como por su tamaño e intensidad. Sin embargo, la posición de una proteína puede verse afectada por diversas causas. La variación en la posición de una misma proteína en un conjunto de geles se debe a una combinación de efectos globales (cambio y rotación del gel en el proceso de escaneado, expansión y contracción durante la rehidratación y secado) y los efectos locales (falta de homogeneidad en el gel o variaciones en las condiciones debido a cambios de temperatura o corriente de fuga) [16].

En general, en el proceso de electroforesis, la preparación y manipulación del gel, así como en la digitalización del mismo, se pueden generar algunas anomalías en las imágenes 2DGE. Entre las anomalías más comunes se encuentran: ruido generado por polvo o residuos que quedan depositados en el gel y que puede llegar a confundirse con pequeñas proteínas, diferentes niveles de saturación debido a la iluminación del sistema de adquisición que puede resaltar algunas proteínas y atenuar otras, proteínas de baja abundancia (puntos difusos o con baja intensidad), saturaciones de puntos, rayas horizontales causadas por el isoelectroenfoco incompleto de algunas proteínas (en la primera dimensión), y rayas verticales causadas por una mala separación en la segunda dimensión, además el fondo de las imágenes puede variar [17]. En la Fig. 1, se muestran algunas de las anomalías más comunes en las imágenes 2DGE. Estas anomalías hacen necesario el uso de técnicas de procesamiento de imágenes digitales para mejorar la calidad de las imágenes y reducir el efecto de las anomalías en la detección e identificación de las proteínas. El procesamiento de imágenes se considera una etapa fundamental para el análisis proteómico basado en imágenes 2DGE. La siguiente sección presenta las principales corrientes de procesamiento empleadas en imágenes 2DGE.

### B. *Procesamiento de Imágenes 2DGE*

El procesamiento de imágenes 2DGE es muy importante para el análisis proteómico, sin embargo, las anomalías comunes a este tipo de imágenes dificultan el análisis. En la literatura, se encuentran dos corrientes de procesamiento de imágenes 2DGE [5], [18]. La primera consiste en la detección e identificación de las proteínas en cada gel de forma independiente [5], [19]. Una vez esta detección es realizada, se identifican proteínas

comunes en el conjunto de geles analizados (Fig. 2(a)). La segunda realiza un registro de las imágenes, que mediante transformaciones geométricas alinean las imágenes de tal manera que se obtiene una correspondencia punto a punto [1], [5]. Una vez se tienen las imágenes registradas, se obtiene una

única imagen que contiene toda la información del conjunto de geles analizados (Fig. 2(b)). Sobre esta única imagen se hace la detección e identificación de proteínas, y sirve de plantilla para identificar las proteínas en cada gel analizado.

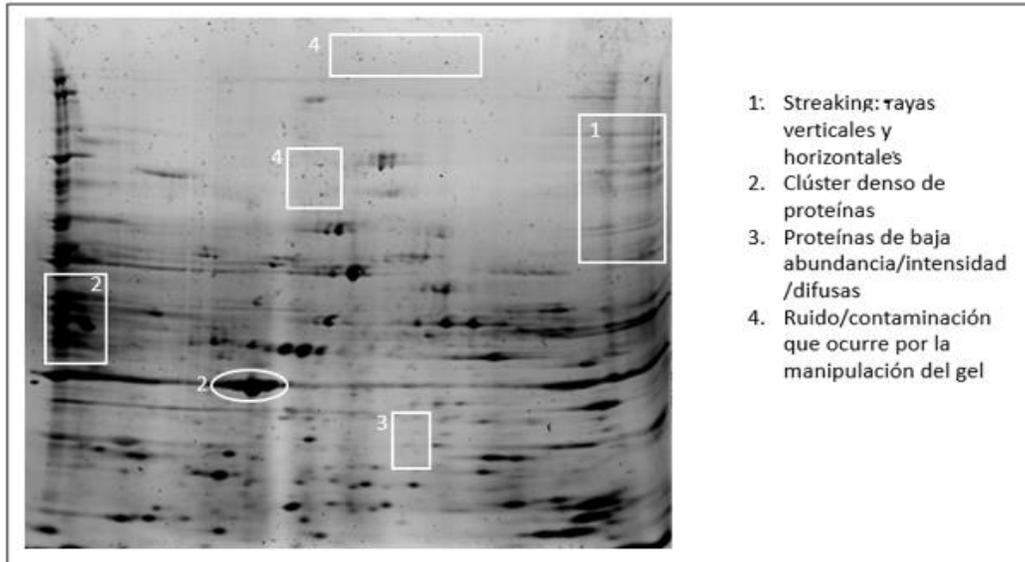


Fig. 1. Proteoma obtenido de células mononucleares de sangre periférica (Fuente: Autores).

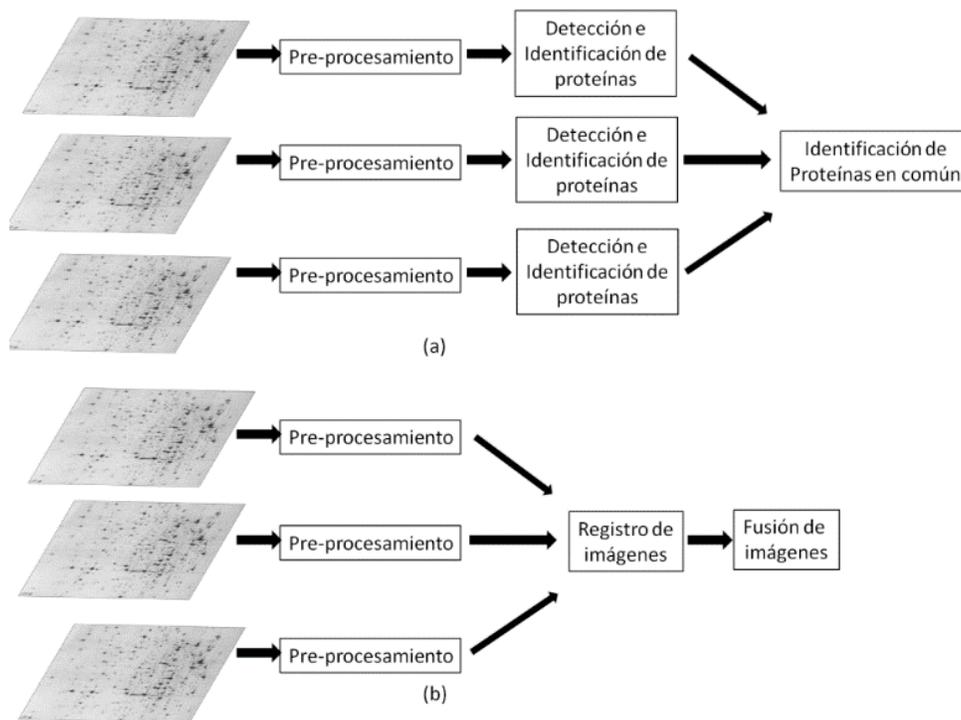


Fig. 2. Diagrama de bloques del procesamiento de imágenes 2DGE estándar usando software especializado (Fuente: Autores).

A partir de las dos corrientes presentadas en la Fig. 2, se nota que en el procesamiento de imágenes 2DGE se realizan tres etapas: pre-procesamiento, identificación de proteínas y fusión de imágenes. Estas dos últimas se alternan de acuerdo a la corriente de procesamiento que se esté empleando.

C. Pre-procesamiento de Imágenes 2DGE

El pre-procesamiento de imágenes busca mejorar la calidad de la imagen de forma que aumente la posibilidad de éxito en las etapas posteriores [20]. Wu y colaboradores publicaron un estudio sobre las técnicas de transformación de ajuste de brillo, contraste y rango de escala de grises para identificar su impacto en la información contenida en las imágenes 2DGE empleando

*TotalLab 100*, una herramienta computacional que soporta el análisis de geles [21]. Este tipo de transformaciones modifican la información contenida de las imágenes 2DGE, principalmente la relacionada con la abundancia de las proteínas.

En general, los métodos de pre-procesamiento para imágenes 2DGE se pueden clasificar en: métodos de reducción de ruido, normalización de intensidades, y corrección de fondo [18], los cuales buscan mejorar los detalles de los puntos, reducir el ruido y homogeneizar el fondo para facilitar la detección de las proteínas. En la reducción de ruido, se pueden emplear filtros lineales o no lineales. Sin embargo, los filtros lineales reducen la intensidad de las proteínas y hacen que los bordes difuminados de las proteínas tengan un efecto secundario no deseado [20]. Por su parte, los filtros no lineales permiten preservar los bordes, pero al ser técnicas más robustas requieren de más parámetros. Los filtros no lineales más empleados para la reducción de ruido en imágenes 2DGE son basados en la transformada Wavelet [22], la transformada Contourlet [23] y Variación Total [24].

Por otra parte, la normalización de imágenes 2DGE busca mejorar el contraste de las proteínas. Las técnicas más empleadas de normalización de imágenes 2DGE se pueden dividir en dos categorías. La primera usa la información de los píxeles en un único gel [4], y la segunda emplea la información desde múltiples geles para normalizar la intensidad [18]. Ambas categorías se basan en el histograma para realizar la normalización de la imagen.

Las técnicas de corrección de fondo de imágenes 2DGE permiten obtener un fondo más uniforme que favorece la detección de las proteínas. Idealmente, se espera que los píxeles en el fondo de una imagen 2DGE sean cercanos al valor máximo de intensidad, correspondiente al color blanco. Las técnicas más comunes para la corrección del fondo se basan en ajuste por polinomios [25], ajuste por histogramas [26], y mínimos locales y globales [18].

#### D. Detección y Segmentación de Proteínas

Los resultados del análisis proteómico dependen en gran medida de la etapa de detección de proteínas [17]. Detectar falsas proteínas o no reconocer otras, conduce a conclusiones erróneas sobre las características proteómicas de una célula o tejido. Adicionalmente, dado que las imágenes 2DGE tienen cientos de proteínas, la detección permite reducir la cantidad de datos desde una gran cantidad de píxeles a algunos puntos característicos [27]. Sin embargo, la detección de proteínas debe enfrentar varios desafíos. En muchos casos, los puntos tienen un bajo contraste que dificulta la visualización de las proteínas. Además, los puntos adyacentes en algunos casos se superponen y por lo tanto son difíciles de separar. Por último, la calidad de estas imágenes se degrada debido a la existencia de ruido, manchas y fondo no homogéneo. Para la detección de estos puntos característicos existen dos metodologías en la literatura: la detección de proteínas independientemente en una imagen 2DGE o identificación de características comunes desde una imagen fusionada de varias imágenes de geles (ver

Fig. 2). En ambos casos se emplean técnicas de segmentación para separar las proteínas del fondo y las anomalías en la imagen 2DGE.

La segmentación divide una imagen en sus partes constituyentes u objetos. Típicamente estos algoritmos se basan en determinar discontinuidades o en algún criterio de similaridad [20]. En la primera categoría, la imagen se divide teniendo en cuenta los cambios bruscos de intensidades. Ejemplos de estas técnicas son la detección de puntos aislados, líneas y bordes de una imagen. Por otra parte, los métodos basados en similaridad dividen la imagen de acuerdo a la homogeneidad de la intensidad. Los métodos de similaridad más comunes son la umbralización, crecimiento de regiones, unión y división de regiones y métodos morfológicos [20].

Particularmente para imágenes 2DGE, los métodos de segmentación que se encuentran con más frecuencia en la literatura son: la detección de bordes [9], [28]–[30], métodos morfológicos [31]–[37], umbralización [31], [38]–[40], y métodos basados en regiones [41]–[43]. A continuación, se presenta una breve revisión de cada una de estas corrientes en la segmentación de imágenes 2DGE.

##### 1) Detección de bordes

La detección de bordes es el método más común para detectar discontinuidades de intensidad [20]. Un borde se define como la frontera entre dos regiones con intensidades de gris relativamente distintas. Para la detección de discontinuidades se emplea una máscara que se desplaza a través de una imagen calculando un operador local de derivación, tales como: el gradiente, Laplaciano, Roberts, Pwett, Sobel, Kirsch, Robinson, entre otros [20].

Varios programas especializados en el procesamiento de imágenes 2DGE emplean detectores de borde para la segmentación de las proteínas [9], [28]–[30], [42]. Sin embargo, la segmentación de proteínas usando detectores de borde presenta el problema de no diferenciar entre proteínas y anomalías, tales como manchas verticales y horizontales, ruido o roturas del gel [27], [42]. Adicionalmente, tampoco son capaces de diferenciar proteínas superpuestas.

##### 2) Métodos morfológicos

En 1977, la transformada Watershed fue introducida en morfología matemática, y por su efectividad en problemas de segmentación de imágenes en escala de grises, se ha empleado ampliamente para mejorar los resultados de segmentación de imágenes 2DGE. Una Watershed puede definirse como una línea divisoria que separa diferentes cuencas (*cathment basin*) en una imagen en escala de grises. Una cuenca es un conjunto de píxeles asociado a un punto mínimo tal que desde cada píxel se puede alcanzar el mínimo siguiendo un camino descendente [44].

La principal desventaja de la transformada Watershed para la segmentación de imágenes 2DGE es su alta sensibilidad al ruido que resulta en una sobre-segmentación. Establecer los mínimos locales se dificulta por la presencia del ruido en las imágenes 2DGE [35]. Para solucionar este problema existen diferentes alternativas que combinan la transformada

Watershed con otras técnicas como la umbralización [33], [37], la unión y separación de regiones [36], y el uso de marcadores [35].

En [37], se usa la transformada Watershed en combinación con umbralización para la segmentación de imágenes 2DGE, con el inconveniente de establecer los puntos mínimos en la imagen, dado que las regiones usualmente se forman por nivel de gris continuos, llevando a una sobre-segmentación de la imagen. Para solucionar este problema, se realiza la umbralización de la imagen con un valor alto de nivel de gris produciendo una región muy grande, a la cual se aplica la transformada Watershed. Si la transformada obtiene dos regiones, el proceso de umbralización-Watershed se vuelve aplicar en la región más grande. Con este proceso, se busca eliminar la necesidad de establecer mínimos y realizar un post-procesamiento para unir o separar regiones.

En [36], se parte de la transformada Watershed para dividir la imagen 2DGE en múltiples regiones. Cada región es segmentada usando *K*-mean identificando el fondo y píxeles que pertenecen a proteínas. Posteriormente, se identifica el centroide de cada proteína, a partir de los cuales se realiza un proceso de división o unión de regiones para evitar problemas de sobre-segmentación.

El uso de marcadores que restringen la transformada Watershed ayuda a mejorar los problemas de sobre-segmentación que se obtienen al aplicar esta técnica en imágenes 2DGE. En [35] usan una técnica de filtrado basada en la distribución de intensidad de los píxeles para definir los marcadores, los cuales se emplean como los mínimos locales para calcular las Watersheds. Adicional a la transformada Watershed, se han introducido otros métodos morfológicos para la segmentación automática de proteínas en imágenes 2DGE. Por ejemplo, en [32], se introduce el uso de contornos activos para la segmentación de proteínas y demuestra que esta técnica ayuda a segmentar proteínas superpuestas.

### 3) Umbralización

La umbralización es uno de los métodos más importantes y sencillos para la segmentación de imágenes. Esta técnica permite separar objetos del fondo a partir de un umbral sobre la intensidad de los píxeles [20]. En procesamiento de imágenes 2DGE es común encontrar enfoques multi-umbral para la detección de las proteínas [38], [45]. Por ejemplo, Kostopoulou y colaboradores dividen la imagen en ventanas de tamaño fijo, en cada una de las cuales se aplica la umbralización [39], [45].

El principal problema que presentan los algoritmos de umbralización es su alta sensibilidad al ruido, anomalías y fondo no homogéneo de las imágenes 2DGE, lo cual resulta en la detección de falsas proteínas [45]. Por esta razón, se encuentra en la literatura la combinación de las técnicas de umbralización con otros métodos de procesamiento de imágenes. En [31], se combina un filtro basado en la información de textura multidireccional con un detector de la máxima intensidad de regiones (umbralización local). Por su parte, en una publicación por Sengar y colaboradores [38] se emplea un método de umbralización que no usa el valor de intensidades para separar las proteínas del fondo, sino que emplea los coeficientes wavelets para separar los puntos, su borde y el fondo de las imágenes 2DGE.

### 4) Métodos basados en regiones

El crecimiento de regiones busca agrupar píxeles o subregiones dentro de regiones más grandes de acuerdo algún criterio de similaridad. El método se basa en la agregación de píxeles similares y espacialmente próximos a partir de un conjunto de semillas aleatorias. La similaridad entre píxeles se determina a partir del nivel de intensidad, color o textura [20]. En segmentación de proteínas se usa el crecimiento de regiones combinado con otras técnicas. Kostopoulou [45] utiliza un algoritmo de corte-crecimiento combinado con crecimiento de regiones para segmentar proteínas en imágenes 2DGE.

Para la segmentación de proteínas en imágenes 2DGE también encontramos métodos paramétricos, los cuales modelan las proteínas a través de una función, usualmente una distribución Gaussiana bidimensional [42]. Sin embargo, este tipo de modelos paramétricos tiene problemas para detectar proteínas saturadas, así como proteínas superpuestas. Se propone el uso de múltiples funciones para modelar las proteínas en las imágenes 2DGE, empleando un algoritmo genético para la estimación de los parámetros que mejor se adecuen a la imagen. A través del multi-modelos logran detectar proteínas superpuestas, sin embargo, el costo computacional del proceso de optimización es bastante alto.

### 5) Registro y fusión de imágenes

La fusión de imágenes consiste en la combinación de datos de varias imágenes. Los objetivos de la fusión de imágenes son: realzar elementos que no son visibles, complementar datos para mejorar una clasificación, detectar cambios en una zona usando datos multitemporales, reemplazar datos anómalos, y adicionar datos faltantes. El proceso de fusión debe satisfacer tres condiciones: preservación de toda la información pertinente, eliminación de información irrelevante y ruido, y minimización de los artefactos y las inconsistencias en la imagen fusionada.

En el análisis proteómico de imágenes 2DGE, es altamente deseable conocer un conjunto significativo de posiciones estándar que permitan identificar proteínas en una colección de geles [46]. Una posición estándar debe cumplir los siguientes requisitos: (1) debe estar razonablemente cerca de la posición de la especie de proteína sobre cualquier imagen de gel; (2) las relaciones espaciales entre los puntos se deben preservar lo mejor posible; (3) debería ser posible encontrar puntos correspondientes en un conjunto de geles basados en sus posiciones normales; y (4) debe permitir la predicción de la especie de proteína en función de su posición. Sin embargo, la posición de una proteína en la imagen 2DGE, se puede ver afectada por anomalías presentes en la imagen debido al proceso de separación de las proteínas y adquisición de la imagen. Por esta razón, el uso de técnicas de procesamiento digital de imágenes debe permitir la detección de proteínas de una manera precisa y confiable dentro de un gel [46].

Existe una amplia variedad de investigaciones realizadas sobre el registro de imágenes 2DGE, principalmente enfocadas al problema de la selección automática de puntos de referencia para la posterior transformación de la imagen (algunos ejemplos son [47], [48]). Srinark y colaboradores proponen un algoritmo de identificación de puntos que es una integración de

un método jerárquico que encuentra correspondencia entre pares de puntos [36]. Otro grupo propuso *Pinnacle*, un método de detección de puntos y cuantificación [11]. Se elabora una imagen promedio sin ruido de un conjunto de imágenes 2DGE correctamente alineadas y esto se logra mediante la detección de los mínimos locales en la imagen y la combinación de la media sin ruido dentro de una proximidad definida. La principal ventaja de *Pinnacle* se encuentra sobre la detección de focos que se superponen. Sin embargo, en la detección resultan puntos no esenciales, es decir, puntos de falsos positivos [21].

A pesar de los avances en el campo de registro y fusión de imágenes, a la fecha todavía es necesario para los investigadores seleccionar manualmente puntos de referencia a fin de realizar el registro y la fusión de las imágenes [12]. Por lo tanto, surge la necesidad, desde hace mucho tiempo de desarrollar un procedimiento, junto con el software, que sea simple y robusto que facilite el proceso de fusión de imágenes 2DGE.

### III. CONCLUSIONES

El procesamiento de imágenes de electroforesis bidimensional 2DGE, permite realizar un análisis de las proteínas contenidas en las células, que son las que ejecutan y controlan la mayoría de procesos en los organismos y una expresión anormal de proteínas indica un estado patológico, cuyo estudio y comprensión proporciona grandes contribuciones en el campo de la medicina, en la exploración de eventos biológicos y de nuevos biomarcadores.

El procesamiento de imágenes 2DGE es necesario para discriminar manchas o ruido de las proteínas reales, cuantificar la abundancia de las proteínas, analizar formas y tamaños, además de estimar su punto isoeléctrico y peso molecular de acuerdo a su ubicación en el gel. Existen dos corrientes en el procesamiento de imágenes 2DGE. El primer enfoque detecta las proteínas de forma independiente en cada muestra y posteriormente registra y fusiona esta información. Por otra parte, el segundo enfoque realiza una alineación y registro de las imágenes y detecta las proteínas sobre la imagen fusionada. Ambos enfoques emplean tres etapas para el procesamiento de imágenes 2DGE: pre-procesamiento, detección de proteínas y fusión de imágenes. El primero busca mejorar la calidad de la imagen, aplicando técnicas de transformación y ajuste, de disminución de ruido, normalización de intensidades y corrección de fondo. Por su parte, los métodos de detección se basan principalmente en técnicas de segmentación de imágenes que buscan identificar falsas proteínas y reduce los puntos de interés. La segmentación divide una imagen en sus objetos y partes constituyentes. Los métodos de segmentación más usados en el procesamiento de imágenes 2DGE se dividen en métodos basados en umbralización, detección de bordes, crecimiento de regiones y métodos morfológicos. A pesar de los avances que se encuentran en la literatura para la detección de proteínas en imágenes 2DGE, aún persiste el problema de detectar proteínas que se traslapan y diferenciar proteínas de anomalías introducidas en el proceso de adquisición. Por último, el registro y fusión de imágenes permite realzar los elementos que no son visibles, complementar datos para una mejor clasificación, reemplazar datos anómalos y adicionar

datos faltantes. La finalidad del proceso de fusión es preservar toda la información pertinente, eliminación de información irrelevante y ruido y minimización de inconsistencias en la imagen fusionada. El principal problema para la aplicación de esta técnica es la selección de puntos de control que permitan una óptima alineación de los datos.

A pesar de existir dos enfoques para el procesamiento de imágenes 2DGE, como los presentados en la Fig. 2, no es frecuente en la literatura una comparación cualitativa que determine cuál de los enfoques es el más adecuado para el análisis proteómico.

De acuerdo con lo evidenciado en este artículo, se hace necesario trabajar en la búsqueda de nuevos procesos con filtros lineales que no reduzcan las intensidades originales de las proteínas y que no difuminen los bordes de éstas. En los filtros no lineales, generar procesos que no exijan mucha complejidad y parametrización. Es efectivo el empleo de la combinación de técnicas como wavelet, watershed, detección de bordes, umbralización y reconocimiento de regiones, generando patrones comunes de normalización para las diferentes combinaciones, logrando mejorar la identificación de las proteínas y el fondo. También es de importancia encontrar y seleccionar técnicas que mejoren el contraste de las proteínas basados en los histogramas, la detección de los bordes que permitan identificar las fronteras entre regiones con intensidades diferentes y que se identifiquen proteínas superpuestas mediante la utilización de contornos activos en la segmentación, técnicas con transformadas watershed que sean menos sensibles al ruido, evitando sobre-segmentación. Otro de los objetivos de estudio debe ser el tratar de resolver el problema de la umbralización, que se presenta en la detección de falsas proteínas, debido a la alta sensibilidad al ruido y fondo no homogéneo y que se puede lograr en combinación con la técnica de coeficientes wavelet. *Pinnacle* es una técnica cuya ventaja se encuentra sobre la detección de focos que se superponen, pero tiene la desventaja de poder detectar puntos falsos positivos (no esenciales) [49], [50]. En la mayoría de los casos, para las tareas de registro y fusión, hay que seleccionar manualmente los puntos de referencia, por lo que se hace necesario desarrollar técnicas automáticas que ejecuten dichas tareas.

### IV. AGRADECIMIENTOS

Este trabajo, se enmarca en el proyecto P14227, financiado por el Instituto Tecnológico Metropolitano (ITM) de Medellín.

### REFERENCIAS

- [1] A. Görg, W. Weiss, and M. J. Dunn, "Current two-dimensional electrophoresis technology for proteomics," *Proteomics*, vol. 4, no. 12, pp. 3665–3685, 2004. DOI:10.1002/pmic.200590007
- [2] W. P. Blackstock and M. P. Weir, "Proteomics: quantitative and physical mapping of cellular proteins," *Trends Biotechnol.*, vol. 17, no. 3, pp. 121–7, Mar. 1999. DOI:10.1016/S0167-7799(98)01245-1
- [3] A. Cieslak and I. Ribera, "Aplicaciones de proteómica en ecología y evolución," *Rev. Ecosistemas*, vol. 18, no. 1, 2009.

- [4] M. Berth, F. M. Moser, M. Kolbe, and J. Bernhardt, "The state of the art in the analysis of two-dimensional gel electrophoresis images," *Appl. Microbiol. Biotechnol.*, vol. 76, no. 6, pp. 1223–1243, Sep. 2007. DOI:10.1007/s00253-007-1128-0
- [5] J. E. Bandow, J. D. Baker, M. Berth, C. Painter, O. J. Sepulveda, K. A. Clark, I. Kilty, and R. A. VanBogelen, "Improved image analysis workflow for 2-D gels enables large-scale 2-D gel-based proteomics studies - COPD biomarker discovery study," *Proteomics*, vol. 8, no. 15, pp.3030–3041, Aug.2008. DOI:10.1002/pmic.200701184
- [6] D. Calligaris, C. Villard, and D. Lafitte, "Advances in top-down proteomics for disease biomarker discovery," *J. Proteomics*, vol. 74, no. 7, pp. 920–934, 2011. DOI:10.1016/j.jprot.2011.03.030
- [7] E. Marengo, E. Robotti, F. Antonucci, D. Cecconi, N. Campostrini, and P. G. Righetti, "Numerical approaches for quantitative analysis of two-dimensional maps: A review of commercial software and home-made systems," *Proteomics*, vol. 5, no. 3, pp. 654–666, Feb. 2005. DOI:10.1002/pmic.200401015
- [8] G. Caruso, C. Cavaliere, C. Guarino, R. Gubbiotti, P. Foglia, and A. Laganà, "Identification of changes in *Triticum durum* L. leaf proteome in response to salt stress by two-dimensional electrophoresis and MALDI-TOF mass spectrometry," *Anal. Bioanal. Chem.*, vol. 391, no. 1, pp. 381–390, May 2008. DOI:10.1007/s00216-008-2008-x
- [9] R. D. Appel, P. M. Palagi, D. Walther, J. R. Vargas, J. C. Sanchez, F. Ravier, C. Pasquali, and D. F. Hochstrasser, "Melanie II—a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface," *Electrophoresis*, vol. 18, no. 15, pp. 2724–34, Dec. 1997. DOI:10.1002/elps.1150181506
- [10] P. Shen, X. Fan, Z. Zeng, and Y. Cheng, "Two-dimensional electrophoresis analysis of proteomics based on image feature and mathematical morphology," *Sci. China Ser. B*, vol. 48, no. 4, pp. 361–367, 2005. DOI:10.1360/04yb0128
- [11] B. N. Clark and H. B. Gutstein, "The myth of automated, high-throughput two-dimensional gel analysis," *Proteomics*, vol. 8, no. 6, pp. 1197–1203, Mar. 2008. DOI:10.1002/pmic.200700709
- [12] F. Li, F. Seillier-Moiseiwitsch, and V. R. Korostyshevskiy, "Region-based statistical analysis of 2D PAGE images," *Comput. Stat. Data Anal.*, vol. 55, no. 11, pp. 3059–3072, 2011. DOI: 10.1016/j.csda.2011.05.013
- [13] A. R. Brandão, H. S. Barbosa, and M. A. Z. Arruda, "Image analysis of two-dimensional gel electrophoresis for comparative proteomics of transgenic and non-transgenic soybean seeds," *J. Proteomics*, vol. 73, no. 8, pp. 1433–40, Jun. 2010. DOI: 10.1016/j.jprot.2010.01.009
- [14] R. V. Pando-Robles and H. Lanz-Mendoza, "The significance of proteomics in public health," *Salud Publica Mex.*, vol. 51 Suppl 3, pp. S386-94, 2009. DOI:10.1590/S0036-36342009000900004
- [15] G. Roti and K. Stegmaier, "Genetic and proteomic approaches to identify cancer drug targets," *Br. J. Cancer*, vol. 106, no. 2, pp. 254–261, Jan. 2012. DOI:10.1038/bjc.2011.543
- [16] P. G. Righetti, "Real and imaginary artefacts in proteome analysis via two-dimensional maps," *J. Chromatogr. B*, vol. 841, no. 1–2, pp. 14–22, Sep. 2006. DOI: 10.1016/j.jchromb.2006.02.022
- [17] A. W. Dowsey, M. J. Dunn, and G.-Z. Yang, "The role of bioinformatics in two-dimensional gel electrophoresis," *Proteomics*, vol. 3, no. 8, pp. 1567–1596, Aug. 2003. DOI:10.1002/pmic.200300459
- [18] M. Rye and E. M. Fargestad, "Preprocessing of electrophoretic images in 2-DE analysis," *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 70–79, 2012. DOI: 10.1016/j.chemolab.2011.09.012
- [19] A. Efrat, F. Hoffmann, K. Kriegel, C. Schultz, and C. Wenk, "Geometric algorithms for the analysis of 2D-electrophoresis gels," *J. Comput. Biol.*, vol. 9, no. 2, pp. 299–315, Apr. 2002. DOI:10.1089/10665270252935476
- [20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Second ed., vol. 14, no. 3. 2002. DOI:10.2307/1574313
- [21] Y. Wu and L. Zhang, "Comparison of two academic software packages for analyzing two-dimensional gel images," *J. Bioinform. Comput. Biol.*, vol. 9, no. 6, pp. 775–94, Dec. 2011. DOI:10.1142/S0219720011005665
- [22] M. T. Faheem, S. Rashwan, A. Sarhan, and B. A. Youssef, "Dennoising 2D gel images using Wavelet transform," *Proc. 10th WSEAS Int. Conf. Appl. Comput. Appl. Comput. Sci.*, no. 1, pp. 187–194, 2011.
- [23] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, 2005. DOI:10.1109/TIP.2005.859376
- [24] G. Oliveri, N. Anselmi, and A. Massa, "Compressive Sensing Imaging of Non-Sparse 2D Scatterers by a Total-Variation Approach Within the Born Approximation," *IEEE Trans. Antennas Propag.*, vol. 62, no. 10, pp. 5157–5170, Oct. 2014. DOI:10.1109/TAP.2014.2344673
- [25] K. Kaczmarek, B. Walczak, S. De Jong, and B. G. M. Vandeginste, "Preprocessing of two-dimensional gel electrophoresis images," *Proteomics*, vol. 4, no. 8, pp. 2377–2389, 2004. DOI:10.1021/ci0256337
- [26] S. Sarkar and S. Das, "Multilevel Image Thresholding Based on 2D Histogram and Maximum Tsallis Entropy— A Differential Evolution Approach," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4788–4797, Dec. 2013. DOI:10.1109/TIP.2013.2277832
- [27] T. Aittokallio, J. Salmi, T. A. Nyman, and O. S. Nevalainen, "Geometrical distortions in two-dimensional gels: applicable correction methods," *J. Chromatogr. B*, vol. 815, no. 1, pp. 25–37, 2005. DOI: 10.1016/j.jchromb.2004.07.037
- [28] B. Raman, A. Cheung, and M. R. Marten, "Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie," *Electrophoresis*, vol. 23, no. 14, pp. 2194–202, Jul. 2002. DOI:10.1002/1522-2683(200207)23:14<2194::AID-ELPS2194>3.0.CO;2-#
- [29] P. F. Lemkin, Y. Wu, and K. Upton, "An efficient disk based data structure for rapid searching of quantitative two-dimensional gel databases," *Electrophoresis*, vol. 14, no. 12, pp. 1341–50, Dec. 1993. DOI:10.1002/elps.11501401207
- [30] J. E. Solomon and M. G. Harrington, "A robust, high-sensitivity algorithm for automated detection of proteins in two-dimensional electrophoresis gels," *Comput. Appl. Biosci.*, vol. 9, no. 2, pp. 133–9, Apr. 1993. DOI:10.1093/bioinformatics/9.2.133
- [31] E. Zacharia, E. Kostopoulou, D. Maroulis, N. P. Anagnou, and K. I. Pappa, "2D-GE spot detection combining multidirectional texture and spatial intensity cues," in 13th IEEE International Conference on Bioinformatics and BioEngineering, 2013, pp. 1–4. DOI:10.1109/BIBE.2013.6701555
- [32] M. A. Savelonas, E. A. Mylona, and D. Maroulis, "Unsupervised 2D gel electrophoresis image segmentation based on active contours," *Pattern Recognit.*, vol. 45, no. 2, pp. 720–731, 2012. DOI: 10.1016/j.patcog.2011.08.003
- [33] A. Anjos, A. L. B. Møller, B. K. Ersbøll, C. Finnie, and H. R. Shahbazkia, "New approach for segmentation and quantification of two-dimensional gel electrophoresis images," *Bioinformatics*, vol. 27, no. 3, pp. 368–75, Feb. 2011. DOI: 10.1093/bioinformatics/btq666
- [34] P. Peer and L. G. Corzo, "Local pixel value collection algorithm for spot segmentation in two-dimensional gel electrophoresis research," *Comp. Funct. Genomics*, p. 89596, Jan. 2007. DOI: 10.1155/2007/89596
- [35] C. Sun and X. Wang, "Spot segmentation and verification based on improve marker controlled watershed transform," in 2010 3rd International Conference on Computer Science and Information Technology, 2010, pp. 63–66. DOI:10.1109/ICCSIT.2010.5563982
- [36] T. Srinark and C. Kambhamettu, "An image analysis suite for spot detection and spot matching in two-dimensional electrophoresis gels," *Electrophoresis*, vol. 29, no. 3, pp. 706–715, Feb. 2008. DOI:10.1002/elps.200700244
- [37] Y. Kim, J. Kim, Y. Won, and Y. In, "Segmentation of Protein Spots in 2D Gel Electrophoresis Images with Watersheds Using Hierarchical Threshold," *Springer, Berlin, Heidelberg*, 2003, pp. 389–396. DOI:10.1007/978-3-540-39737-3\_49
- [38] R. S. Sengar, A. K. Upadhyay, M. Singh, and V. M. Gadre, "Analysis of 2D-gel images for detection of protein spots using a novel non-separable wavelet based method," *Biomed. Signal Process. Control*, vol. 25, pp. 62–75, 2016. DOI: 10.1016/j.bspc.2015.10.013
- [39] E. Kostopoulou, S. Katsigiannis, & D. Maroulis, "A custom grow-

- cut based scheme for 2D-gel image segmentation,” In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2407–2410, 2015. DOI:10.1109/EMBC.2015.7318879
- [40] P. Cutler, G. Heald, I. R. White, and J. Ruan, “A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection,” *Proteomics*, vol. 3, no. 4, pp. 392–401, Apr. 2003. DOI:10.1002/pmic.200390054
- [41] E. Kostopoulou, E. Zacharia, and D. Maroulis, “Accurate segmentation of 2D-PAGE images,” in *Signal Processing Conference (EUSIPCO), 2012 roceedings of the 20th European, 2012*, pp. 2258–2262.
- [42] D. K. Iakovidis, D. Maroulis, E. Zacharia, and S. Kossida, “A genetic approach to spot detection in two-dimensional gel electrophoresis images,” *Proc. 5th IEEE EBMS Int. Spec. Top. Conf. Inf. Technol. Biomed. (ITAB)*, Ioannina, Greece, 2006.
- [43] R. S. Sengar, A. K. Upadhyay, M. Singh, and V. M. Gadre, “Segmentation of two dimensional electrophoresis gel image using the wavelet transform and the watershed transform,” in *2012 National Conference on Communications (NCC), 2012*, pp. 1–5. DOI:10.1109/NCC.2012.6176861
- [44] L. Vincent and P. Soille, “Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, 1991. DOI: 10.1109/34.87344
- [45] E. Kostopoulou, S. Katsigiannis, and D. Maroulis, “2D-gel spot detection and segmentation based on modified image-aware grow-cut and regional intensity information,” *Comput. Methods Programs Biomed.*, vol. 122, no. 1, pp. 26–39, 2015. DOI:10.1016/j.cmpb.2015.06.007
- [46] S. Luhn, M. Berth, M. Hecker, and J. Bernhardt, “Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images,” *Proteomics*, vol. 3, no. 7, pp. 1117–1127, Jul. 2003. DOI:10.1002/pmic.200300433
- [47] M. Daszykowski, M. S. Wróbel, A. Bierczynska-Krzysik, J. Silberring, G. Lubec, and B. Walczak, “Automatic preprocessing of electrophoretic images,” *Chemom. Intell. Lab. Syst.*, vol. 97, no. 2, pp. 132–140, 2009. DOI:10.1016/j.chemolab.2009.03.002
- [48] K. Kaczmarek, B. Walczak, S. de Jong, and B. G. M. Vandeginste, “Matching 2D Gel Electrophoresis Images,” *J. Chem. Inf. Model.*, vol. 43 (3), pp. 978–986, 2003. DOI:10.1021/ci0256337
- [49] J. S. Morris, B. N. Clark, & H. B. Gutstein, “Pinnacle: a fast, automatic and accurate method for detecting and quantifying protein spots in 2-dimensional gel electrophoresis data. *Bioinformatics*,” *Bioinformatics.*, vol. 24(4), pp. 529–536, 2008. DOI:10.1093/bioinformatics/btm590
- [50] C. W. Marean, “Pinnacle Point Cave 13B (Western Cape Province, South Africa) in context: the Cape floral kingdom, shellfish, and modern human origins”. *Journal of Human Evolution*, vol.59(3-4), pp. 425-443, 2010. DOI:10.1016/j.jhevol.2010.07.011



**Gerardo A. Villegas-Rivera** received a B.S. degree in Electronic Engineering from the Universidad San Buenaventura, in 2001, and a M.S. degree in Automation and Industrial Control from the Instituto Tecnológico Metropolitano (ITM) of Medellín, Colombia, in 2018.

Since 2009, he has been a Professional Technician with the Electrical Engineering Department of the Empresas Públicas de Medellín (EPM), in Medellín. His research interests include pattern recognition, industrial automation and biomedicine.



**Maria C. Torres-Madronero** received a B.S. degree in Electronic Engineering from the Universidad Nacional de Colombia, Manizales, in 2006 and a M.S. degree in Electrical Engineering and a Ph.D. degree in Computing and Information Science

and Engineering from the University of Puerto Rico Mayaguez (UPRM), in 2008 and 2013, respectively.

Since 2014, she has been a Full-time Lecturer with the Informatic Systems Department, Instituto Tecnológico Metropolitano (ITM) of Medellín Colombia. She is the author of more than 20 papers in scientific peer reviewed journals. Her research interests include hyperspectral data, image processing, unmixing analysis, pattern recognition, and machine learning. Currently, She is an Associate Professor and the leader of the Automation Electronics and Computational Sciences Research Group.



**Sarah Röthlisberger Booth** received a B.S. degree in Genetics and Biochemistry, and a M.S. degree in Biochemistry, from Massey University, Palmerston North, New Zealand, in 2006 and 2008, respectively.

From 2010 to 2011, she was a researcher with the Molecular Biology Lab, Instituto Tecnológico Metropolitano (ITM) of Medellín, Colombia. From 2012 to 2013, she was the Ethics Committee Administrator, Ministry of Health, Wellington, New Zealand. Since 2013, she has been a Full-time Lecturer with the Biomedical Engineering Program, Instituto Tecnológico Metropolitano (ITM) of Medellín, Colombia. She is the author of more than 10 papers in scientific peer reviewed journals. Her research interests include proteomics, bioinformatics and molecular biology.



**Edilson Delgado-Trejos** received a B.S. degree in Electronic Engineering, a M.S. degree in Industrial Automation, and a Ph.D. degree in Engineering Sciences from the Universidad Nacional de Colombia, in 2000, 2003, and 2008, respectively.

Since August 2008, he has been a Full-time Lecturer and Senior Researcher at the Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia. Currently, he is an Associate Professor and Vice-Chancellor of Research and Academic Extension at the same institute. He has published more than 50 papers, 10 book chapters and 3 books in indexed scientific journals and editorials. His current research interests include pattern recognition, machine learning, multivariate data analysis, nonlinear analysis, signal processing and soft metrology.