



Institución Universitaria

Pre-Procesamiento de Imágenes de Electroforesis Bidimensional en Gel Aplicado al Análisis Proteómico

Tesis de maestría realizada por:

Manuel Mauricio Goez Mora

Instituto Tecnológico Metropolitano

Facultad de Ingenierías

Medellín, Colombia

2017

Pre-Procesamiento de Imágenes de Electroforesis Bidimensional en Gel Aplicado al Análisis Proteómico

Manuel Mauricio Goez Mora

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magíster en Automatización y Control Industrial

Director:

Prof. Maria Constanza Torres Madroño, MSc, PhD ¹

Co-directores:

Prof. Edilson Delgado Trejos, MSc, PhD ²

Prof. Sarah Röthlisberger Booth, MSc ³

Línea de Investigación:

Máquinas inteligentes y reconocimiento de patrones

Grupos de Investigación:

¹ Grupo de investigación en Automática, Electrónica y Ciencias Computacionales (AEyCC).

² Grupo de Investigación en Calidad, Metrología y Producción (CMyP).

³ Grupo de Investigación e Innovación Biomédica (GI2B).

Instituto Tecnológico Metropolitano

Facultad de Ingenierías

Medellín, Colombia

2017

Resumen

El uso de la proteómica es uno de los tópicos de investigación más activos en la biotecnología, la cual se refiere a toda aplicación tecnológica que utilice sistemas biológicos y organismos vivos o sus derivados para la creación o modificación de productos o procesos para usos específicos. Dentro de la proteómica, la electroforesis bidimensional en gel (2DGE) es la herramienta más utilizada para la separación e identificación de proteínas. Esta técnica se ha mantenido como la forma más viable de obtener resultados confiables y reproducibles en el análisis proteómico. En la actualidad, existen numerosos métodos y software para el procesamiento de imágenes de geles obtenidos a partir de la electroforesis bidimensional. Sin embargo, no existe aún un sistema fiable, automatizado y de alta reproducibilidad, debido a las anomalías presentes en este tipo de imágenes. Entre las anomalías más comunes se encuentran líneas verticales, puntos difusos y el ruido, las cuales dificultan el procesamiento de las imágenes. Este proyecto de investigación desarrolló una metodología de pre-procesamiento de imágenes de geles bidimensionales aplicada al análisis proteómico, utilizando técnicas de procesamiento digital de imágenes como normalización de imagen, corrección de fondo y filtrado no lineal. Para el desarrollo de este trabajo se realizó una comparación cuantitativa y cualitativa de diferentes técnicas de pre-procesamiento utilizando imágenes 2DGE tanto sintéticas como reales. A partir de la comparación de las técnicas de pre-procesamiento encontradas en la literatura, se propuso una metodología multinivel, que permite mitigar diferentes tipos de anomalías de las imágenes 2DGE. Con la metodología multinivel propuesta se logró una mejora del 34 % respecto a la aplicación de técnicas de manera individual sobre imágenes 2DGE sintéticas.

Palabras clave: Pre-procesamiento, imágenes 2DGE, reducción de ruido, corrección de fondo, normalización de imagen.

Abstract

Proteomics is one of the most active research areas in biotechnology due to its wide, diverse and important applications. High resolution two-dimensional gel electrophoresis (2DGE) is a key technique for protein separation and analysis. With this technique it is possible to obtain reliable and reproducible results for proteomic analysis. Currently, there are several methods and software for processing two-dimensional gel electrophoresis images. However, there is still not a reliable, automated and highly reproducible system for this, due to the anomalies present in this type of image. The most common anomalies present in 2DGE images are vertical streaking, noise, and low abundance spots, which complicate the image processing. In this research project a methodology for pre-processing two-dimensional gel images applied to proteomic analysis has been applied, using digital image processing techniques such as normalization, background correction and non-linear filtering. A quantitative and qualitative comparison was carried out of several pre-processing techniques using real and synthetic 2DGE images. Based on the results of this comparison of pre-processing techniques for background correction, normalization and noise reduction, a new multilevel methodology is proposed, that reduces the effects of several types of anomalies on 2DGE images. The proposed combined methodology displays a 34 % improvement compared to individual pre-processing techniques on 2DGE synthetic images.

Keywords: Pre-processing, 2DGE images, noise reduction, background correction, image normalization.

Contenido

Resumen	v
1. Introducción	1
1.1. Justificación	1
1.2. Descripción del problema	2
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos específicos	3
1.4. Organización del documento	4
2. Estado del arte	5
2.1. Imágenes 2DGE	6
2.1.1. Adquisición	7
2.1.2. Pre-procesamiento	7
2.1.3. Procesamiento	7
2.2. Pre-procesamiento de imágenes	8
2.2.1. Reducción de ruido	9
2.2.2. Corrección de fondo	12
2.2.3. Normalización de imágenes	13
2.3. Recomendaciones	14
3. Marco experimental	15
3.1. Bases de datos	15
3.1.1. Imágenes sintéticas	15
3.1.2. Imágenes reales	18
3.2. Métodos	20
3.2.1. Indicadores cuantitativos de desempeño	20
3.2.2. Técnicas de reducción de ruido	22
3.2.3. Técnicas de corrección de fondo	24
3.2.4. Técnicas de regularización de imagen	25
3.3. Metodología propuesta	25
4. Resultados y discusión	28
4.1. Reducción de ruido	28
4.1.1. Comparación de técnicas usando imágenes sintéticas	28
4.1.2. Filtrado de imágenes sintéticas usando GNDF	38
4.1.3. Comparación de técnicas usando imágenes reales	44

4.2. Corrección de fondo	51
4.2.1. Comparación de técnicas usando imágenes sintéticas	51
4.2.2. Comparación de técnicas usando imágenes reales	54
4.3. Normalización de imagen	57
4.3.1. Comparación de técnicas usando imágenes sintéticas	57
4.3.2. Comparación de técnicas usando imágenes reales	58
4.4. Metodología para integrar técnicas de pre-procesamiento	61
4.5. Validación de resultados por expertos en imágenes 2DGE	66
5. Conclusiones y recomendaciones	70
5.1. Conclusiones	70
5.2. Recomendaciones	71
5.3. Discusión académica de los resultados	72
Bibliografía	74

1 Introducción

El uso de la proteómica es uno de los tópicos de investigación más activos en la biotecnología, debido a sus crecientes campos de aplicación y la importancia de los mismos, tales como: el desarrollo y evaluación de los efectos de los fármacos, identificación de biomarcadores, entre otros [1]. Dentro de la proteómica, la electroforesis bidimensional en gel (2DGE) es la herramienta más utilizada para la separación e identificación de proteínas [2]. En la actualidad, existen numerosos métodos y software para el procesamiento de imágenes 2DGE. Sin embargo, no existe aún un sistema fiable, automatizado y de alta reproductibilidad dada las anomalías presentes en este tipo de imágenes [3]. Entre las anomalías más comunes se encuentran líneas verticales, puntos difusos y ruido, las cuales dificultan el procesamiento de las imágenes [1]. En la literatura se encuentran técnicas de pre-procesamiento como normalización de imágenes, corrección de fondo y filtrado, empleadas para mejorar la apariencia de las imágenes 2DGE. No obstante, no se encuentra una evaluación cuantitativa que permita evaluar el desempeño de estas técnicas y el efecto de cada una de ellas sobre la detección de las proteínas.

Este proyecto presenta, en primer lugar, un estudio comparativo de las técnicas de pre-procesamiento que se encuentran en la literatura. Para esta comparación fue necesario definir una metodología de evaluación cuantitativa que emplea imágenes sintéticas y reales. A partir de los resultados obtenidos se determina la necesidad de desarrollar una metodología de pre-procesamiento que combina la normalización de imágenes, corrección de fondo y filtrado. Con la cual se busca mejorar la detección de proteínas en imágenes 2DGE.

Este trabajo de tesis maestría está enmarcado en el proyecto de investigación *Caracterización del veneno de la abeja africanizada mediante análisis proteómico y procesamiento digital de imágenes orientado a la exploración de sus propiedades (P14227)*. Este proyecto es desarrollado en conjunto por el Grupo de Investigación e Innovación Biomédica (GI2B) y el grupo de investigación en Automática, Electrónica y Ciencias Computacionales (AEyCC) del Instituto Tecnológico Metropolitano. para el cual es necesario el pre-procesamiento de las imágenes con el objetivo de obtener resultados representativos.

1.1. Justificación

Las enfermedades cardiovasculares, cáncer, diabetes y enfermedades pulmonares, son identificadas como enfermedades crónicas, las cuales son la principal causa de muerte en el mundo [4]. Para 2008, causaron el 63 % de las muertes a nivel global [5]. Además se estiman un incremento del 15 % entre 2010 y 2020, alcanzando el 73 % de la mortalidad [5]. Una herramienta fundamental para enfrentar

esta problemática es la proteómica, la cual permite vislumbrar nuevas aplicaciones en la biomédica como la identificación de proteínas que intervienen en diferentes etapas de una enfermedad por medio de la caracterización proteómica. Este tipo de aplicaciones ayudan a comprender la naturaleza de estas enfermedades, a desarrollar nuevas drogas y a seleccionar medicamentos mediante la verificación de su eficacia y efectos secundarios [4].

Un ejemplo representativo del uso de la proteómica en bio-indicadores se da en la detección de cáncer. A nivel mundial, más de 11 millones de personas se diagnostican con cáncer cada año y se calcula que representan 13% del total de muertes por año (7.5 millones). Con el incremento de la expectativa de vida, la prevalencia de muchos tipos de cáncer se incrementará, a tal punto que para el año 2030, 11.5 millones de personas morirán por esta enfermedad [6].

Las aplicaciones en el sector salud posicionan el análisis proteómico como un campo de interés para la comunidad científica nacional. La salud aparece como un sector estratégico a nivel nacional y regional para el avance de la ciencia, la tecnología y la innovación. Colciencias en su Programa Nacional de CTI en Salud contempla las líneas de acción: Investigación básica biomédica e Investigación en Salud Pública [7]. De la misma manera, este campo de investigación es relevante dentro del Plan Estratégico en Electrónica, Telecomunicaciones e Informática, dentro del área de Control y Automatización. A nivel regional, el plan CT+i de la ciudad de Medellín establece que la salud es una de las cadenas productivas prioritarias y la bioinformática es un camino para disminuir la brecha tecnológica de este sector [8].

La finalidad de este proyecto es mejorar la reproducibilidad del análisis proteómico a partir de imágenes de electroforesis bidimensional en gel, permitiendo obtener resultados más precisos, incentivando los desarrollos obtenidos a partir de las técnicas de proteoma, como lo son la detección de bio-indicadores, desarrollo de fármacos, caracterización proteómica, entre otros. Principalmente, este proyecto pretende desarrollar una metodología de pre-procesamiento que permita una detección y análisis más preciso de las proteínas desde las imágenes de electroforesis bidimensional en gel (2DGE).

1.2. Descripción del problema

Las imágenes 2DGE presentan varias anomalías que dificultan su análisis [3, 9], tales como: altos niveles de ruido, rayas verticales y horizontales, saturación de las muestras, entre otras. Ejemplos de estas anomalías se presentan en la Figura (1-1). Este tipo de defectos alteran la veracidad y reproducibilidad de la caracterización proteómica de las muestras, generando de esta forma un retraso significativo en el avance del análisis proteómico [10]. Por ejemplo, el ruido superficial, debido al proceso de adquisición, puede ser detectado como alguna proteína cuando realmente no existe, es decir un falso positivo. También se presentan de manera consecutiva líneas horizontales y verticales que se generan por el desplazamiento de las proteínas dentro de la muestra. Las anomalías pueden disminuir la probabilidad de detección de proteínas tenues, ya que pueden afectar sus características estructurales degradando el contrastes de sus bordes.

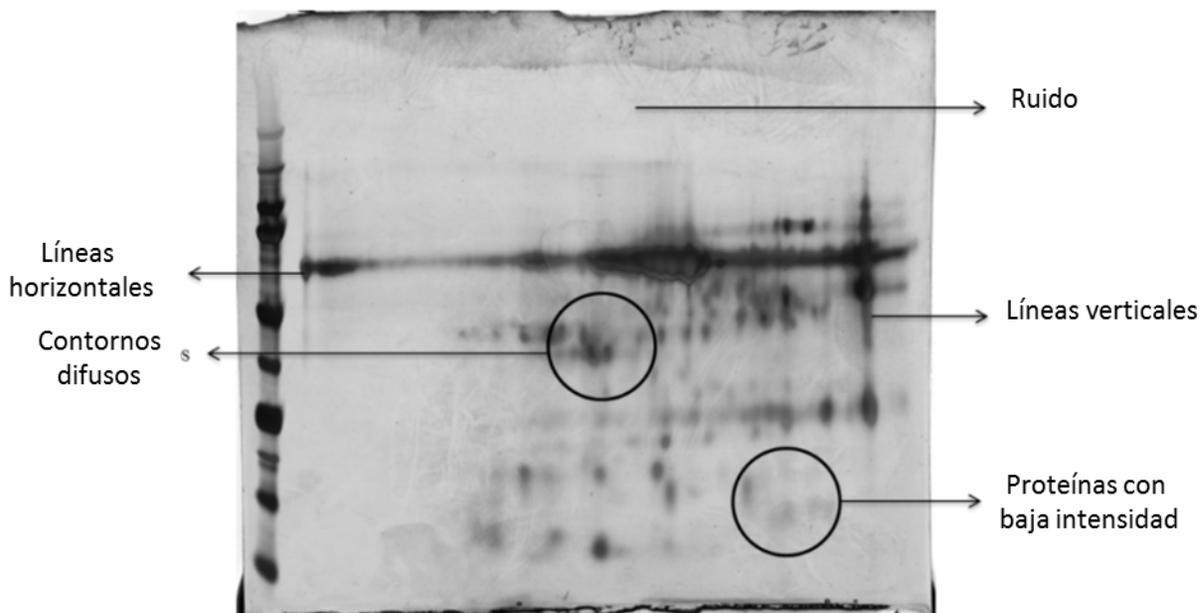


Figura 1-1: Anomalías presentes en una imagen de electroforesis bidimensional en gel de una muestra de próstata humana [10].

En la literatura se establece la necesidad de un procesamiento automático de imágenes 2DGE, que requiera la mínima intervención humana y permita obtener resultados confiables y reproducibles. En la literatura (ver Capítulo 2) se encuentran tres corrientes para el pre-procesamiento de imágenes 2DGE: reducción de ruido, normalización de imagen y corrección de fondo. Para cada una de estas corrientes se encuentran diversas técnicas. Sin embargo, no se encuentra en la literatura una evaluación cuantitativa que permita medir su impacto en la detección de proteínas en imágenes 2DGE. De igual forma, no se encuentra en la literatura una metodología de pre-procesamiento que combine las tres corrientes para mitigar al mismo tiempo diferentes tipos de anomalías.

1.3. Objetivos

1.3.1. Objetivo general

Proponer una metodología para el pre-procesamiento de imágenes en geles bidimensionales, que permita reducir los efectos de las anomalías en la extracción de información en estudios proteómicos, utilizando técnicas de filtrado no lineal, normalización de imagen y corrección de fondo.

1.3.2. Objetivos específicos

- Comparar técnicas de pre-procesamiento reportadas en la literatura para reducir los efectos de las anomalías en las imágenes en geles bidimensionales, proponiendo indicadores cuantitativos

que permitan una adecuada evaluación de los métodos.

- Proponer una metodología de análisis para evaluar el desempeño de las técnicas de pre-procesamiento de imágenes 2DGE de manera cuantitativa.
- Analizar la metodología para el pre-procesamiento de imágenes 2DGE apropiada, a partir de la combinación de técnicas basadas en filtrado no lineal, regularización de imagen y corrección de fondo, orientadas a reducir de forma eficiente los efectos de las anomalías en las imágenes de geles bidimensionales.

1.4. Organización del documento

El Capítulo 1 presentó una introducción de los aspectos generales de este trabajo, con el fin de contextualizar al lector dentro del marco del proyecto. El Capítulo 2 presenta una revisión detallada del estado del arte sobre el tratamiento de las imágenes 2DGE, en la cual se profundiza en la revisión de las técnicas de pre-procesamiento más empleadas en el área. En el Capítulo 3 se introduce el marco experimental, el cual describe los métodos y bases de datos utilizadas en el desarrollo del trabajo, adicionalmente se presenta la metodología propuesta para el pre-procesamiento de imágenes 2DGE. En el Capítulo 4 se exponen los resultados a partir de la comparación de las técnicas de pre-procesamiento empleando datos sintéticos y reales. Finalizando el cuerpo de este trabajo se encuentra la sección de conclusiones y recomendaciones.

2 Estado del arte

El término proteómica se reportó ante la comunidad científica por primera vez en 1995 [11]. La proteómica es una rama de la genómica que estudia los proteomas, es decir, la identificación de las proteínas. Las técnicas de proteómica tienen como objetivo determinar el conjunto de proteínas presentes en una muestra de manera cuantitativa, permitiendo caracterizar la muestra de interés en función de las proteínas presentes [12].

El método más usado en la proteómica es la electroforesis bidimensional 2DGE [13]. Esta técnica se emplea para describir el contenido de proteínas presentes en un tipo particular de muestra, en un momento determinado, bajo un conjunto particular de condiciones [14]. Su característica más deseada es permitir la identificación de proteínas manteniendo la muestra intacta. El principio básico de esta técnica consta de separar las proteínas inmersas en un gel en dos dimensiones, a partir del punto isoelectrónico y el peso molecular. Para resaltar las proteínas se realiza la tinción de la muestra con una pigmentación negra. Las muestras son digitalizadas a través de escáneres especializados, permitiendo el uso de herramientas computacionales para su procesamiento y análisis. La identificación final de las proteínas se realiza por medio de espectrometría de masas.

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. El análisis de geles 2-DGE no solo permite la caracterización proteómica de una muestra en un determinado momento, sino que también permite hacer un análisis comparativo de varias condiciones, por ejemplo, enfermo vs. sano, con el fin de definir proteínas con expresión diferencial (biomarcador) que se puedan medir de manera objetiva y evaluar como un indicador de procesos biológicos, procesos patogénicos o respuestas farmacológicas a un agente terapéutico [15].

Las proteínas de mayor interés, o potenciales biomarcadores, a menudo se encuentran en baja abundancia. Por lo tanto, en el pre-procesamiento de la imagen es de vital importancia eliminar las anomalías presentes en la imagen, como ruido y líneas verticales y horizontales, para facilitar su análisis, pero sin eliminar los spots de baja abundancia que puedan tener gran importancia clínica. La electroforesis bidimensional en gel se ha mantenido como la forma más viable de obtener resultados confiables y reproducibles en el análisis proteómico [16]. Existen numerosos software que permiten el procesamiento de las imágenes 2DGE [17]. Sin embargo, debido a las anomalías que se pueden presentar en este tipo de imágenes bidimensionales, como líneas verticales, puntos difusos,

ruido en la captura de imagen, entre otros; no existe aún un sistema fiable y automatizado para la caracterización proteómica de las imágenes de geles 2-D [17]. Adicionalmente, las anomalías presentes en las imágenes 2DGE afectan la reproducibilidad de los resultados [10]. Estas dificultades dan cabida a incluir en el análisis de imágenes 2DGE un procesamiento previo que permita eliminar estas condiciones no deseadas en las muestras, buscando obtener resultados precisos y con la menor variación posible [18]. A continuación, se presenta los fundamentos teóricos del pre-procesamiento de imágenes, así como la adquisición y procesamiento de imágenes de geles 2DGE.

2.1. Imágenes 2DGE

La electroforesis bidimensional en gel, abreviado como 2DGE, es una técnica ampliamente utilizada para el análisis de proteínas, permitiendo la separación de los componentes del proteoma en un único gel. Esta técnica incluye los procesos de: preparación de la muestra, separación de proteínas de acuerdo al peso molecular y punto isoeléctrico sobre el gel, hasta la adquisición de la imagen y su procesamiento [13]. Su alto nivel de acogida se debe principalmente a que es una técnica de alta resolución, que permite un acoplamiento transparente con la técnica de espectrometría de masas. Sin embargo, cabe notar las limitaciones principales de esta técnica que son la reproducibilidad en los geles y la presencia de proteínas de baja densidad en la muestra, las cuales se pueden mal interpretar como ruidos [9]. La Figura (2-1) muestra tres imágenes 2DGE obtenidas desde muestras de próstata humana y otras sustancias por el Grupo de Investigación e Innovación Biomedica del ITM, Medellín.

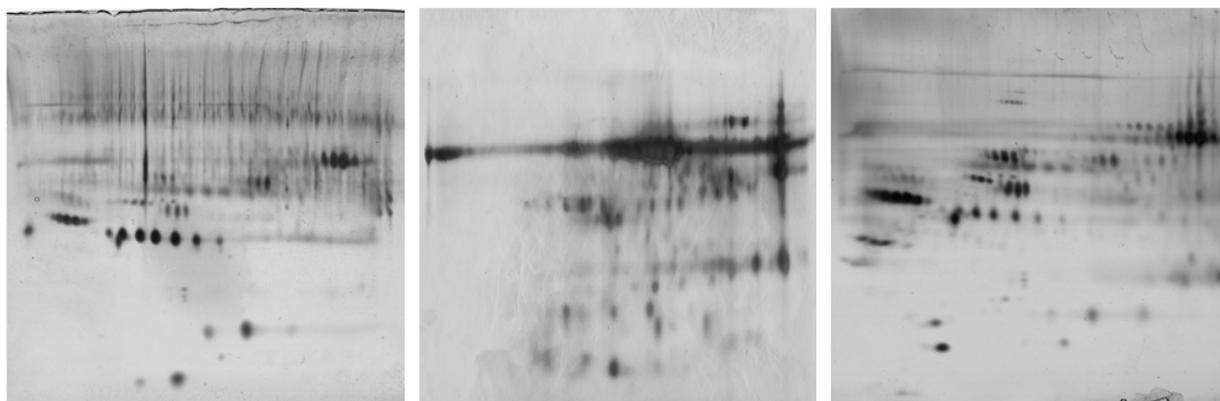


Figura 2-1: Imágenes 2DGE obtenidas por el Grupo de Investigación e Innovación Biomédica.

El diagrama de flujo del procesamiento de imágenes 2-DGE es descrito en la Figura (2-2), el cual contempla desde la recolección de muestras y su acondicionamiento hasta el análisis de ofrecido por el procesamiento de la imágenes.

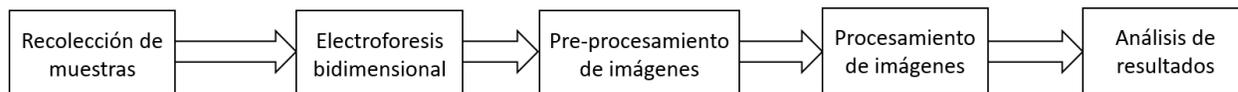


Figura 2-2: Diagrama de flujo procesamiento imágenes 2-DGE.

2.1.1. Adquisición

La adquisición de imágenes 2DGE inicia con el proceso de preparación de la muestra, del cual dependerán la calidad y resolución de las proteínas separadas. En esta preparación se deben considerar las condiciones particulares de la muestra, en donde se debe obtener una completa solubilización y desnaturalización por medio de agentes como urea o tiourea [19]. Una vez preparada la muestra, se realiza la separación de las proteínas, primero por el punto isoelectrónico, etapa conocida como el isoelectro enfoque, y luego de acuerdo a su peso molecular. La separación se realiza sobre un gel, su ubicación depende del peso molecular y punto isoelectrónico, facilitando la identificación de las mismas. Ya obtenida la muestra, es necesario someterlas a un proceso que permita su visualización. Para esto, las proteínas en el gel deben ser teñidas, donde el método de tinción debe ser determinado por factores específicos como: el rango, la sensibilidad esperada, el precio y las características del equipo de adquisición de imagen [19]. Algunos de los métodos más implementados para la identificación de proteínas son la tinción fluorescente y la tinción en plata, con los cuales se modifica selectivamente el color de las proteínas, permitiendo revelar su ubicación espacial dentro del gel. Como último paso se realiza la digitalización del gel a través de dispositivos como densitómetros, escaners e incluso cámaras de alta resolución. La imagen digital del gel es procesada con software especializado que permite la detección e identificación de proteínas.

2.1.2. Pre-procesamiento

Debido al proceso de adquisición de las imágenes 2DGE, se encuentran varias anomalías en los datos digitales que requieren ser reducidos o corregidos antes de iniciar el proceso de detección e identificación de las proteínas. Estas anomalías se deben tanto a la manipulación manual que se realiza sobre el gel (por ejemplo: para transportarlo hasta el dispositivo de digitalización), así como al proceso de separación de las proteínas. El pre-procesamiento de las imágenes 2DGE consta de una serie de técnicas que permiten mejorar la calidad de la imagen digital permitiendo aumentar el desempeño en las técnicas de identificación de las proteínas, tanto en eficiencia como en costo computacional [20]. Una revisión detallada de las técnicas empleadas para el pre-procesamiento de imágenes 2-DGE se presenta en la sección 2.2.

2.1.3. Procesamiento

La etapa de procesamiento de imágenes 2DGE está enfocada a la detección de las proteínas y caracterización de las mismas a través de su tamaño, densidad, peso molecular y punto isoelectrónico. Estas características ayudan a la identificación y análisis de las proteínas. Por esta razón se han

creado una variedad de software científicos y comerciales, enfocados a servir como una herramienta que facilite el análisis proteómico de las muestras. Ejemplos de estos software son: MELANIE, PDQuest, Z2, Proteome Weaver, y Delta 2D [20]. Estas herramientas permiten la cuantificación de proteínas individuales; cuentan con interfaces gráficas que ayudan a solucionar problemas como la superposición de proteínas, puntos de baja densidad, errores de cuantificación y corrección de movimiento. Sin embargo, estas herramientas son muy costosas y usualmente requieren la intervención manual de un experto. Por esta razón, el procesamiento de las imágenes 2DGE esta lejos de ser un proceso automatizado [17], haciendo necesario el desarrollo de herramientas y metodologías automáticas que soporten esta tarea.

2.2. Pre-procesamiento de imágenes

El procesamiento de imágenes se ha convertido en uno de los principales focos de atención para la comunidad científica en ciencias computacionales e ingeniería, debido al creciente número de aplicaciones en diversos campos de estudio. Ésto se da gracias al crecimiento de la tecnología digital, capacidad de computo, resolución de los dispositivos para la adquisición e interpretación de imágenes [21]. El concepto de procesamiento de imágenes es muy amplio, ya que no solo se considera dentro de éste el análisis de las imágenes, si no también el gran número de transformaciones espaciales y espectrales que se pueden aplicar con diversas finalidades, como son: la segmentación de la imagen, identificación de características, reconstrucción tridimensional, interpretación de la información, etc. El pre-procesamiento de imágenes se entiende como una etapa previa al procesamiento de imágenes que busca facilitar la etapa de procesamiento. De manera general, la etapa de pre-procesamiento consta de una serie de técnicas que buscan realizar una adecuación de la imagen, resaltando las características estructurales de interés para su procesamiento [22].

La electroforesis bidimensional en gel (2DGE) requiere de una herramienta eficaz en el procesamiento de la información para obtener los resultados esperados, teniendo en cuenta que se estima que una sola célula contiene alrededor de 10.000 proteínas [23]. Sin embargo, no basta con realizar un procesamiento de caracterización proteómica de manera adecuada, ya que la calidad de los resultados están directamente relacionada con la calidad de la imagen ingresada al proceso [24].

Debido a algunos problemas técnicos inherentes de la electroforesis bidimensional (por ejemplo: el sistema no lineal obtenido en la formación de gel, la calidad de la adquisición de imágenes), se presentan de forma inevitable anomalías como: ruido en el contorno de los puntos debido a la susceptibilidad del gel al polvo [20], saturaciones de puntos, puntos tenues, intensidad no lineal, y líneas estrechas verticales y horizontales en los geles (Figura 1-1). Otras anomalías comunes no deseadas en las imágenes 2DGE son los diferentes tipos y niveles de ruidos; por ejemplo, el ruido impulsivo puede diferir significativamente en el barrido realizado al digitalizar la muestra [25].

Otros efectos no deseados se presentan en el fondo de las imágenes, ya que puede variar entre las muestras, dependiendo en gran medida de la calidad de los equipos para obtener y digitalizar la muestra [13]. Estas anomalías dificultan el procedimiento y análisis de las imágenes 2DGE [9, 3]. La

omisión de un procesamiento previo, que elimine una parte significativa del ruido, puede concebir una representación errada en la detección de proteínas, alterando el resultado al agregar proteínas falsas [26, 27]. La Figura 1-1 presenta un ejemplo de una imagen digital obtenida, a partir de imágenes 2DGE desde una muestra de próstata humana, en ésta se presentan las anomalías típicas en este tipo de imágenes.

Por lo anterior, es evidente la necesidad de un pre-procesamiento para imágenes 2DGE. El concepto de pre-procesamiento de imágenes aplicado en el área de geles bidimensionales para análisis proteómico se encuentra en una etapa temprana, es por ello que existen una gran cantidad de vacíos de alto interés. Sin embargo, se han establecido técnicas que satisfacen las necesidades básicas en el pre-procesamiento de este tipo de imágenes. Se pueden identificar tres objetivos comunes en el pre-procesamiento de imágenes 2DGE: filtrado, regulación de intensidad de píxeles y corrección de fondo [2]. Para alcanzar cada uno de estos objetivos se han establecido diferentes corrientes de pre-procesamiento de imágenes denominadas: reducción de ruido, normalización de imagen y corrección de fondo (ver Figura 2-3) [9].

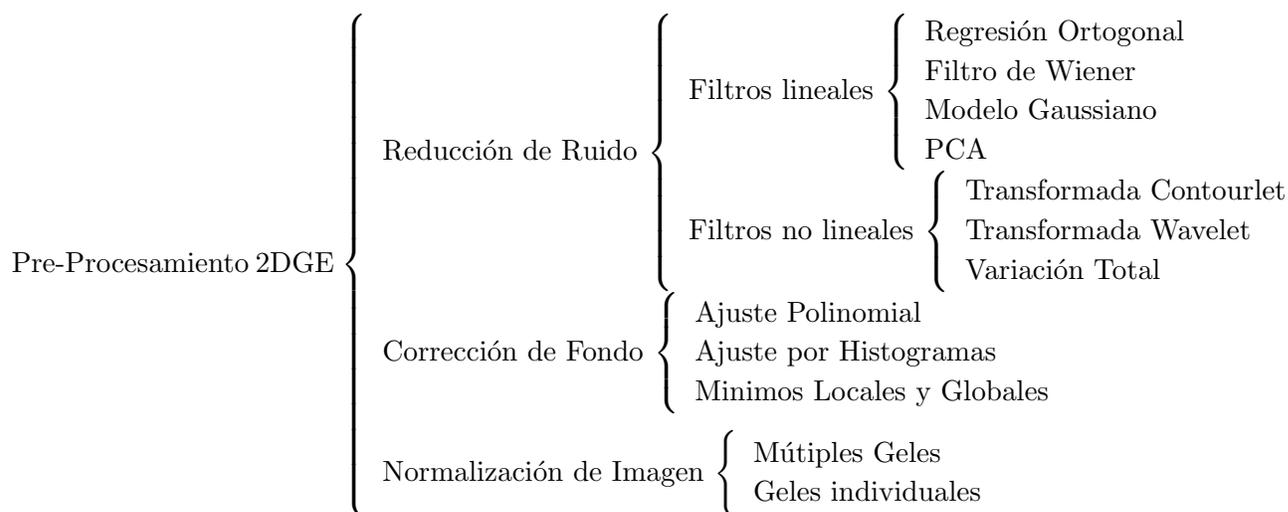


Figura 2-3: Estado del arte en el pre-procesamiento de imágenes 2DGE.

2.2.1. Reducción de ruido

El proceso de reducción de ruido tiene como objetivo mejorar el porcentaje de información de interés de una imagen respecto al ruido presente. La reducción de ruido está compuesto de un conjunto de técnicas que, de manera general, buscan eliminar anomalías. Estas técnicas tienen un mayor desempeño cuando se enmarcan en una aplicación específica, ya que hay ciertos algoritmos que permiten mejorar características de alta relevancia de acuerdo al campo de aplicación [28]. Los objetivos específicos del proceso de reducción de ruido se pueden categorizar de la siguiente manera: suavizar contrastes, mitigación de ruido puntual, mejorar procesos de detección de bordes, entre

otros [22]. Estas técnicas logran su objetivo por medio de una serie de transformaciones espaciales y/o espectrales sobre las imágenes de orígenes.

El filtrado de imágenes de geles bidimensionales es la corriente con mayor uso dentro de la comunidad científica para el pre-procesamiento de imágenes 2DGE, y también la corriente de pre-procesamiento con mayor diversificación en técnicas [13]. Estas técnicas de filtrado pueden ser divididas en dos grandes categorías: filtros lineales y no lineales (ver Figura 2-3).

Filtros lineales

Los filtros lineales fueron una de las primeras técnicas reportadas para el pre-procesamiento de imágenes de geles bidimensionales y la de mayor uso en los software de procesamiento comerciales. Estos filtros determinan el valor del pixel de salida de acuerdo a una relación lineal con los pixeles en una ventana espacial [29]. En general, los filtros lineales tienden a difuminar las imágenes, afectando los bordes, líneas y detalles finos; el nivel de difusión depende en gran medida del tamaño del vecindario o ventana empleado. Si bien las técnicas de filtrado lineales, como modelos de Gauss, han presentado resultados aceptables en condiciones generales, también son susceptibles a disminuir la densidad de las proteínas al deformar su contorno [20, 30].

Las técnicas de regresiones ortogonales permiten obtener resultados rápidos y la aplicación de sus métodos se puede realizar fácilmente debido a su simplicidad. El método de regresión ortogonal [2] es una opción sencilla y eficaz para adaptarse a un modelo lineal cuando no hay distinción entre las variables independientes y dependientes. Sin embargo, en el filtrado de imágenes de geles 2DGE, se reportan dificultades al distorsionar el contorno de los bordes de los puntos de interés [31, 32]. La técnica de análisis de componentes principales -PCA es una versión más robusta de la técnica de regresión ortogonal, reporta una gran ventaja en la conservación de contornos adecuadamente, sin embargo, se pueden eliminar detalles como proteínas de baja densidad en el proceso [33].

Por medio de los filtros de Wiener se obtienen resultados satisfactorios para las imágenes 2DGE, con un bajo nivel de distorsión de la imagen, pero compromete la reducción de ruido [27, 29]. En este tipo de filtros se presenta una eliminación de ruido aislada y la imagen resultante se torna más borrosa, con baja definición de bordes. Sin embargo, cierto grado de degradación en la imagen es por lo general aceptable, en comparación con el beneficio de la eliminación de ruido [13]. En estos métodos se puede graduar el nivel de suavizado y eliminación del ruido, por medio del ajuste de la dimensión de la ventana de convolución y el número de veces sucesivas que se aplique [34].

Filtros no lineales

Los filtros no lineales tienen menores limitaciones, pero en general aumenta el grado de complejidad y la necesidad de definir los parámetros de una manera más extenuante. Algunas técnicas de filtrado no lineal hacen sus primeras incursiones en este campo de aplicación, tal como la transfor-

mada Watershed [2] y la transformada Contourlet [19]. La transformada Watershed es una técnica que permite obtener contornos cerrados, adyacentes y precisos, pero es necesario parametrizar sus variables de manera meticulosa y sus parámetros son susceptibles a cambios entre muestras [35]. Por otra parte, la transformada Contourlet realiza una descomposición de los datos en frecuencia y espacio por medio de una nueva transformada multi-resolución con menos limitaciones, por medio de dos métodos de descomposición [19]. Estos métodos le permiten una escala múltiple de espacio y frecuencia, además de un alto grado de direccionalidad que ha presentado resultados satisfactorios, tanto en reducción de la relación señal-ruido, como en la pérdida de información relevante, superando de acuerdo a los indicadores propuestos a la transformada Wavelet [36]. La técnica Variación Total [37] presenta una gran ventaja en la extracción de las características geométricas de la imagen y conservación de bordes, debido a su nivel de suavizado variante, que permite optimizar la extracción de los bordes y su información direccional circundante. Sin embargo, presenta problemas considerables en la eliminación del ruido [27].

La transformada Wavelet se ha vuelto cada vez más popular en este campo de aplicación, desde su introducción en el pre-procesamiento de imágenes de geles bidimensionales por Kaczmarek [3]. Su popularidad se debe a que los métodos de supresión de ruido basados en filtrado espacial, a pesar de su simplicidad, tienden a distorsionar gravemente la imagen, ya que las imágenes presentan características locales que varían rápidamente, haciendo imposible distinguir la información de interés del ruido si solo se considera en el dominio espacial o espectral [2]. Por esta razón, es apropiado utilizar un filtro con dominio conjunto que supera el filtrado espacial, tanto en la relación señal-ruido como en distorsiones generadas [38]. La transformada Wavelet tiene la capacidad de descomponer la imagen, tanto en componentes espaciales, como componentes de frecuencia y realizar una evaluación de la imagen punto a punto. Además, permite una combinación de varias imágenes para obtener un valor promedio y eliminar eficazmente el ruido. Para un correcto funcionamiento es necesario determinar la escala de la transformada Wavelet. Sin embargo, se presenta una dificultad de definir los contornos adecuadamente, generando una distorsión significativa de la información de interés [39].

El desempeño de las técnicas Variación Total y transformada Watershed es menor, comparados con la transformada Wavelet [2, 39]. De acuerdo con la literatura consultada, la técnica más ampliamente utilizada para reducción de ruido es la transformada Wavelet, esto debido a su alto porcentaje de filtrado y aceptables niveles de pérdida de información relevante [40, 41, 42]. En la literatura se encuentran diferentes derivaciones de la transformada Wavelet [43, 44]. Por ejemplo, Barlaud et al. [45] presentaron una técnica denominada Wavelet Transform Quincunx, la cual permite obtener una escala multiresolución mejorando su desempeño alrededor de los bordes y presentando un mejor filtrado [2].

También se ha reportado una técnica que surge al combinar las técnicas transformada de Wavelet y Variación Total denominada WTTV [39], con la cual se logra disminuir la pérdida de información relevante, por medio de una eliminación de ruido que contempla un esquema híbrido entre ambas técnicas. Esta técnica consiste en aplicar primero la transformada Wavelet y posteriormente Varia-

ción Total, con el objetivo de minimizar las distorsiones causadas al aplicar la transformada. Para este proceso es necesario definir los parámetros en ambas técnicas de manera meticulosa, ya que se puede confundir el ruido con información relevante. Las limitaciones de esta técnica se deben principalmente a la falta de información direccional, que afecta la captura geométrica de bordes, generando una distorsión en las fronteras de los puntos de interés [19].

2.2.2. Corrección de fondo

El proceso de corrección de fondo tiene como objetivo la homogenización del fondo de la imagen, favoreciendo el contraste entre éste y las características u objetos dentro de la imagen. Por lo cual, es común que muchas de las técnicas de corrección de fondo busquen la sustracción del fondo, parcial o completamente, permitiendo disminuir el costo computacional de los procesos de detección o identificación de características. Muchas de estas técnicas contemplan el uso de base de datos y estimaciones matemáticas que representen el fondo de la imagen [22].

La corrección de fondo surge bajo la premisa de que la señal de fondo en las imágenes 2DGE no es uniforme y varía localmente por la intensidad de proteínas; a mayor intensidad de proteínas, mayor es la señal de ruido alrededor de ella en el fondo, permitiendo un error en la detección de proteínas. Además, se encuentran anomalías como la presencia de rayas verticales en las imágenes que deben ser removidas, ya que las proteínas en vetas no son identificables ni cuantificables [46].

Considerando que el objetivo de procesamiento de las imágenes 2DGE es la de detección de puntos, se debe identificar la ubicación, la frontera y la intensidad de cada proteína. Sin embargo, existen condiciones donde la definición de los límites de frontera es un reto, debido al fondo variable, manchas y saturación [47, 48]. La corrección de fondo busca reducir el efecto de este tipo de anomalías en la detección de proteínas. Dentro de esta corriente de pre-procesamiento se establecen tres tipos de técnicas para la corrección de fondo: ajuste por polinomios, mínimos locales y globales y por ajuste por histogramas (ver Figura 2-3).

La corrección de fondo con ajuste por polinomios cuenta con una gran aceptación en la comunidad científica [49, 50]. Esta técnica consiste en realizar un ajuste polinomial del fondo que permite homogeneizarlo, eliminando ruidos sobrepuestos y efectos de degradación de intensidad. La mayor limitación es la selección de la muestra del fondo para realizar el ajuste; una mala selección de esta muestra puede generar una distorsión significativa de la imagen o bien no suavizar de manera representativa la superficie del fondo. La técnica de mínimos locales y globales es la manera más simple para realizar la corrección de fondo, permite mitigar de manera eficaz el ruido en las zonas más planas (sin cambios significativos entre píxeles) y suavizar discontinuidades [13]. Sin embargo, la corrección de fondo por mínimos locales presenta un suavizado muy tenue, siendo en algunos casos insuficiente. Por su parte, el ajuste por histogramas utiliza la distribución de intensidades de una imagen para identificar el fondo, el cual corresponde al máximo pico en la distribución de intensidades de una imagen. El gran reto de esta técnica radica en establecer el umbral de intensidades que corresponden al fondo [51].

Estos métodos son muy eficaces, pero requieren la optimización de los parámetros de entrada. Aunque para los científicos familiarizados con los principios del método, esta optimización es relativamente fácil de realizar, la selección de parámetros resulta ser un serio inconveniente para los usuarios inexpertos [47].

2.2.3. Normalización de imágenes

El proceso de regularización de imagen busca modificar los valores de intensidad de los píxeles, con el objetivo de obtener una imagen más homogénea entre los contrastes, favoreciendo los procesos de detección de características y objetos. También permite modificar los píxeles que presenten anomalías notorias respecto a sus vecinos. Estas técnicas mayormente basadas en procesos estocásticos presentan de manera general una fácil implementación [22]. Este proceso se aplica de manera general, con el fin de disminuir la complejidad y el costo computacional en el procesamiento de las imágenes, por esta razón es común que se incluya este proceso de pre-procesamiento de manera directa en los algoritmos de procesamiento de imágenes [21].

La normalización de imágenes de geles bidimensionales busca regularizar la intensidad de los puntos de concentración de interés para facilitar el procesamiento. Estas intensidades puede variar, entre algunas razones, por la variabilidad en la exposición del gel en el proceso de tinción y la saturación de datos [31, 52, 53]. La normalización de intensidades suele estar presente en el procesamiento de la imagen [44], sin embargo, sigue siendo una técnica de pre-procesamiento. De hecho, varios autores recomiendan eliminar los puntos saturados antes del procesamiento para identificar las proteínas presentes en los geles [54].

Para evitar la saturación causada por una alta densidad de proteínas, los operadores deben garantizar que la mayoría de los datos están representados por niveles ligeramente debajo de la intensidad máxima posible (definida por la resolución a implementar). De igual forma, se debe considerar que ciertas concentraciones de proteínas que pueden comprometer la presencia de manchas débiles, que representan proteínas en baja densidad [55], las cuales a menudo son biológicamente relevante, pero en general demasiado débiles para ser cuantificados con precisión [55].

La normalización de intensidades en la que se emplean múltiples geles, es la más usada y con mejor desempeño [13]. La técnicas que emplean múltiples geles, usualmente, obtienen la información estadística del nivel de intensidad en cada pixel de las imágenes alineadas, con estos promedios estadísticos se lleva a cabo una normalización de intensidad en toda la imagen, pixel por pixel. Con estas técnicas se han obtenido resultados satisfactorios, ya que los cambios en los patrones de expresión de proteínas entre muestras son pequeños, en comparación con el patrón general. La mayor limitación de este enfoque es obtener y alinear un número representativo de muestras. También se puede realizar el proceso de normalización con un solo gel por ecualización de histogramas, resaltando el contorno de los puntos de interés. Sin embargo, en algunas circunstancias la relación de intensidad de píxeles no presenta una tendencia clara [56], dando cabida a la eliminación no

deseada de proteínas tenues [13].

2.3. Recomendaciones

A partir de la revisión del estado del arte se establece algunas necesidades en el campo de procesamiento de imágenes 2DGE. La primera de ellas es la necesidad de una metodología de pre-procesamiento robusta y eficiente, que permita mitigar los efectos de las anomalías presentes en las imágenes. En la literatura se encuentran tres corrientes para el pre-procesamiento: reducción de ruido, regularización de imagen y corrección de fondo. A pesar de existir varias técnicas para cada una de estas corrientes, no se encuentra en la literatura una evaluación cuantitativa que permita medir su desempeño en la detección de proteínas a partir de imágenes 2DGE. De igual forma, no se encuentra en la literatura una metodología de pre-procesamiento que combine las tres corrientes para mitigar al mismo tiempo diferentes tipos de anomalías. A partir de estas necesidades, este proyecto propone una metodología para evaluar cuantitativamente las técnicas de pre-procesamiento, así como una metodología para combinar las diferentes corrientes y disminuir el efecto de las anomalías inherentes en las imágenes 2DGE.

3 Marco experimental

El desarrollo de este trabajo se realizó en el Software MATLAB donde se implementaron las técnicas de pre-procesamiento, debido a que este Software es ampliamente usado y aceptado por la comunidad científica, permitiendo una alta reproducibilidad de resultados [57]. Este Software expresa los resultados obtenidos de manera matricial lo cual permite la generación de imágenes en los formatos de mayor uso en los programas de procesamiento de imágenes 2-DGE, brindando un alto nivel de acoplamiento de la metodología desarrollada en este trabajo con los diversos Software comerciales y académicos en este campo de aplicación.

3.1. Bases de datos

Para el desarrollo del proyecto se contó con dos bases de datos: una sintética, con características controlables, y una base de datos de imágenes reales. La base de datos sintética permite evaluar los algoritmos de pre-procesamiento bajo un ambiente controlado. A continuación se describe la generación de los datos sintéticos y se presenta las imágenes reales empleadas en el proyecto.

3.1.1. Imágenes sintéticas

La base de datos sintética permite realizar una comparación de los métodos de pre-procesamiento bajo parámetros conocidos y controlables, ya que para cada imagen sintética se conoce el número y ubicación de proteínas, así como el nivel y tipo de ruido. Las proteínas se modelaron como una distribución Gaussiana en dos dimensiones [58]. La media y la desviación estándar se asumen iguales para ambas dimensiones. La desviación estándar se relaciona con el tamaño y la medida de dispersión. Un ejemplo de proteína sintética se presenta en la Figura **3-1-a**, generada con una desviación estándar de 0.4 y una media de 0.6. La distribución espacial de las proteínas sobre la imagen sintética se realizó de forma aleatoria a partir de una distribución uniforme. Debido a la aleatoriedad de la distribución espacial se pueden presentar fenómenos de superposición. Un ejemplo de este fenómeno se presenta en la Figura **3-1-c**, en la cual se generaron aleatoriamente dos proteínas, una de ellas con desviación estándar 0.3, y la otra con desviación estándar 0.4. En la Figura **3-1-d** se resaltan los contornos de las proteínas.

Adicionalmente, se empleó un rango de variabilidad aleatorio de los parámetros de la distribución Gaussiana. Para el valor máximo de intensidad se definió el intervalo $[0.6, 1]$ y para la desviación estándar se uso el intervalo $[0.3, 0.8]$. Los límites de estos rangos se seleccionaron de forma experimental. La Figura **3-1-b** presenta 5 proteínas aleatoriamente generadas, en la cual se observa la

variabilidad en el tamaño de las proteínas.

Para las imágenes sintéticas utilizadas en la comparación de las técnicas de filtrado se seleccionó un tamaño de 512x512 píxeles con 250 proteínas. Con el fin de evaluar a fondo las técnicas de pre-procesamiento, se adicionaron diferentes tipos de ruidos. Para este proyecto se seleccionaron los ruidos: blanco Gaussiano, tipo Rayleigh y ruido Exponencial [22], distribuciones que son descritas en la Figura 3-2.

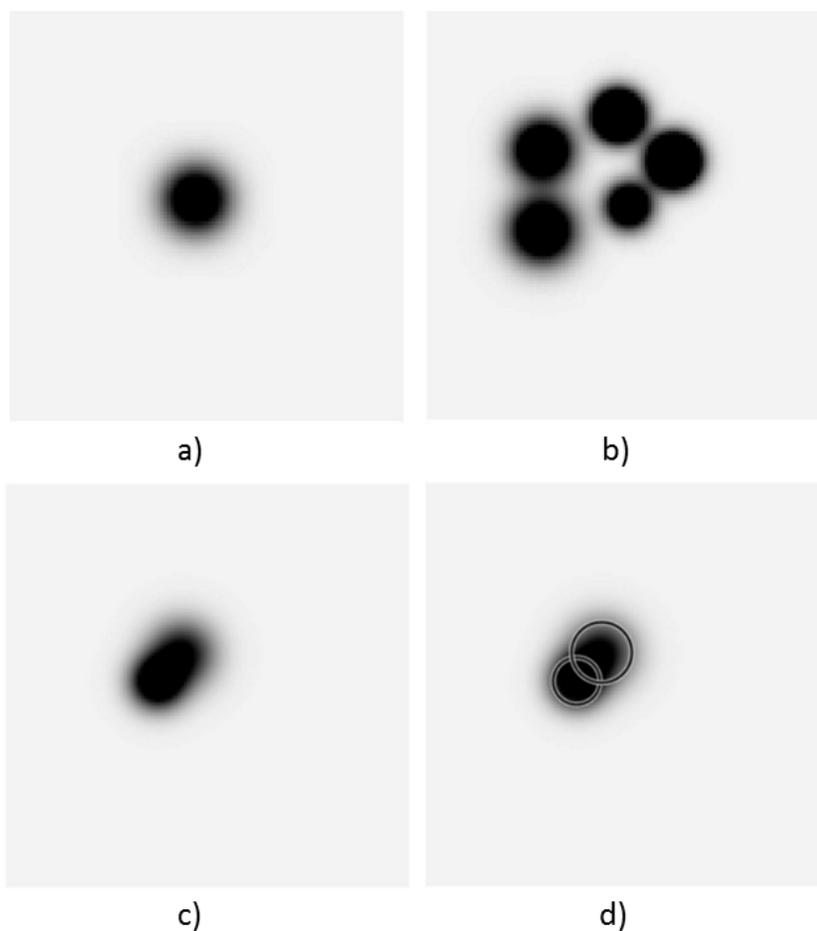


Figura 3-1: Imágenes sintéticas de 512x512 píxeles. a) Proteína sintética con distribución Gaussiana con media 0.6 y desviación estándar 0.4 b) Proteínas sintéticas con rango de [0.3 , 08] para la desviación estándar c) Proteínas sintéticas con superposición. d) Proteínas sintéticas con superposición detectadas.

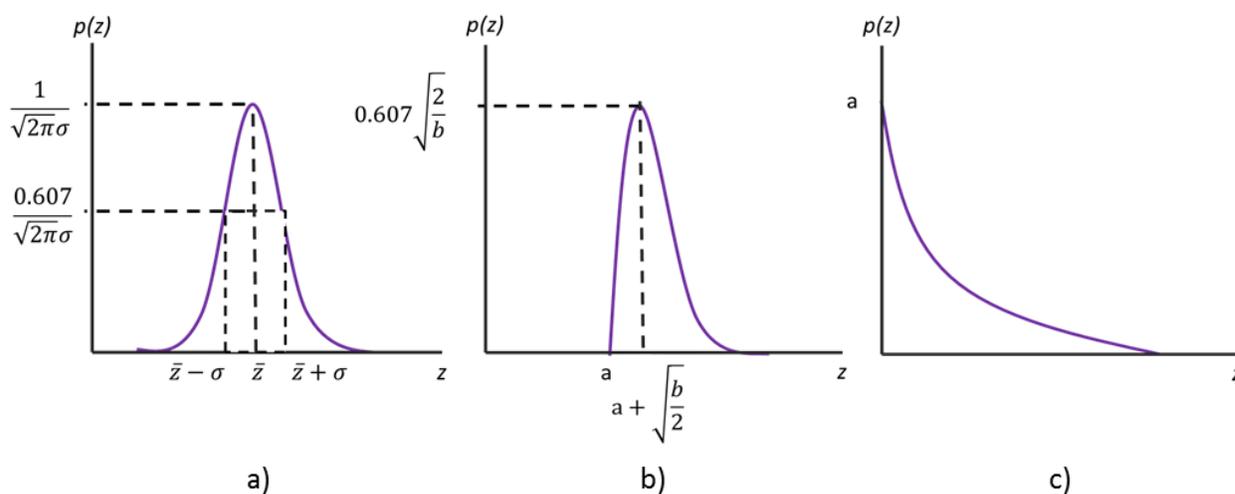


Figura 3-2: Función de distribución para los ruidos: a) blanco Gaussiano b) Rayleigh c) Exponencial.

Los parámetros utilizados para la generación de ruido se presentan en la Tabla **3-1**. Para el ruido Gaussiano se tomó una media cero y una desviación estándar, σ , que varía entre 0.0766 y 0.3091. El ruido Rayleigh se generó con $a = 0$ y variando b entre 0.02771 y 0.1858. Por último, el ruido exponencial se generó con el parámetro a entre 3.8549 a 14.0361. Estos parámetros se ajustaron de tal manera que las imágenes sintéticas tengan la misma relación señal a ruido (SNR) con los diferentes tipos de ruido. Ejemplos la imagen sintética con diferentes tipos de ruido es presentado en la Figura **3-3**.

Ruido Gaussiano con media 0							
SNR	20	18	16	14	12	10	8
σ	0.0766	0.0966	0.1214	0.1530	0.1929	0.2430	0.3091
Ruido Rayleigh con parámetro $a=0$							
SNR	20	18	16	14	12	10	8
b	0.02771	0.0398	0.0539	0.0725	0.098	0.1335	0.1858
Ruido Exponencial							
SNR	20	18	16	14	12	10	8
a	14.0361	10.0361	8.3389	6.8749	5.6862	4.6894	3.8549

Tabla 3-1: Parámetros empleados para generar ruido Gaussiano, Rayleigh y Exponencial.

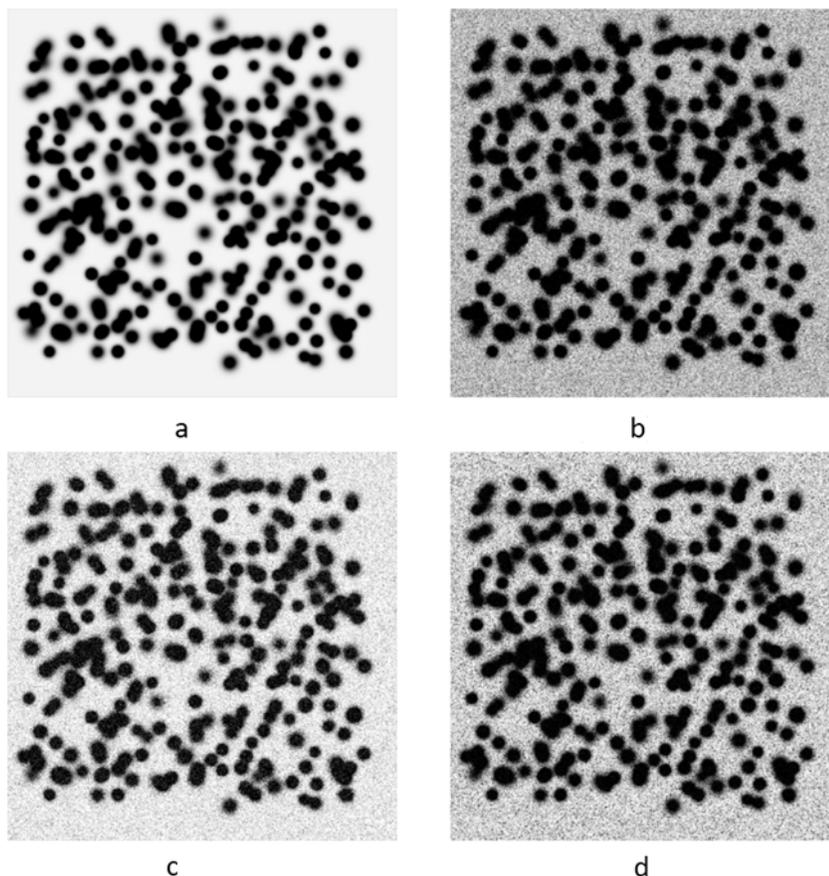


Figura 3-3: Adición de ruido. a) imagen sin ruido b) imagen con ruido Gaussiano $\sigma = 0.3091$ c) imagen con ruido Rayleigh $b = 0.1858$ d) imagen con ruido Exponencial $a = 3.8549$.

3.1.2. Imágenes reales

Para el desarrollo de este proyecto se emplearon imágenes 2DGE reales obtenidas desde diferentes bases de datos:

- Tejidos de la próstata: se emplean imágenes 2DGE obtenidas de tejidos de la próstata. Estas imágenes fueron facilitadas por el Grupo de Investigación e Innovación Biomédica (GI2B). En la Figura 2-1 se muestran una serie de imágenes 2DGE utilizadas como base de datos en este trabajo.
- Human Myocardial Two-Dimensional Electrophoresis Protein Database [7]: Esta base de datos incluye imágenes de electroforesis bidimensional en gel de proteínas presentes en muestras del corazón humano, las características de este tipo de imágenes se pueden apreciar en la Figura 3-4, es de notar que presentan las anomalías a mitigar con la metodología propuesta. Adicionalmente, la base de datos cuenta con la identificación de proteínas sobre las imágenes.

- Apitoxina: base de datos obtenida a partir del proyecto *Caracterización del veneno de la abeja africanizada mediante análisis proteómico y procesamiento digital de imágenes orientado a la exploración de sus propiedades* del ITM, desarrollado por el Grupo de Investigación e Innovación Biomédica (GI2B) y el grupo de investigación en Automática, Electrónica y Ciencias Computacionales (AEyCC). Las muestras corresponden a imágenes 2DGE obtenidas desde el veneno de abejas, conocido como apitoxina (ver Figura 3-5). Para la adquisición de las imágenes se empleó el sistema de electroforesis de Invitrogen, las cuales se encuentran en el laboratorio de Biología Celular y Molecular del ITM.

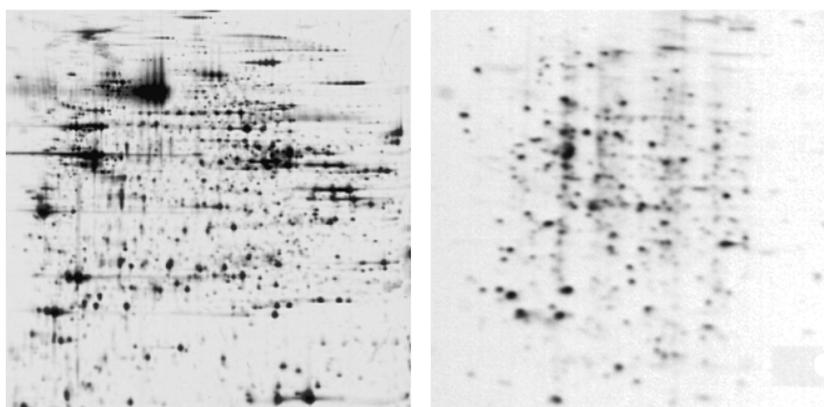


Figura 3-4: Imágenes 2DGE con proteínas identificadas. .

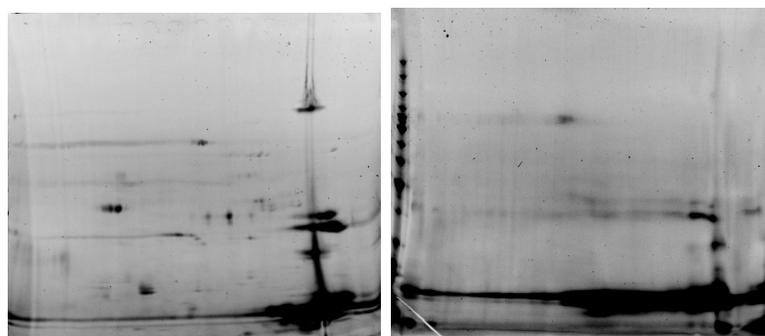


Figura 3-5: Imágenes 2DGE obtenidas desde una muestra de apitoxina.

Se cargaron 50 μg de proteína en cada gel, usando tirillas IPG con rango de pH 3-10 NL de 7 cm (Thermo Fisher Scientific). El isoelectroenfoque se hizo mediante un protocolo de aumento progresivo de voltaje: 200 - 450 - 600 - 750 - 950 v durante 25 minutos, luego 1200 - 1400 - 1600 v durante 30 minutos y finalmente, 2000 v durante 45 minutos [59]. Para la segunda dimensión se utilizaron geles preformados (NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels, 1,5 mm de Thermo Fisher Scientific) corriendo a 200 v durante 40 minutos. Se hizo la tinción de proteínas en los geles usando SYPRO® Ruby Protein Gel Stain, y se visualizaron con el ChemiDoc MP System

(BioRad). Las imágenes obtenidas se sometieron a procesamiento digital de imágenes, usando el software comercial PdQuest Advanced (BioRad) para identificar las proteínas que se encuentran diferencialmente expresadas en los diferentes grupos de estudio.

3.2. Métodos

3.2.1. Indicadores cuantitativos de desempeño

El indicador más utilizado para comparar filtros, de acuerdo a la literatura, es la relación señal-ruido (signal-to-noise ratio) - SNR , obtenido a partir del error cuadrático medio (mean square error) - MSE [50]. El SNR da una apreciación de la calidad de la imagen respecto al ruido presente en la misma con un rango de valores de $(0, \infty)$ y esta dado por la ecuación:

$$SNR = 10 * \log_{10}\left(\frac{1}{MSE_n}\right) \quad (3-1)$$

donde el MSE_n se define como:

$$MSE_n = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{\sum_{i=1}^n (x_i)^2} \quad (3-2)$$

donde que x_i y \bar{x}_i son el valor en intensidad de cada pixel de la imagen y el valor estimado respectivamente.

En la literatura se encuentra un indicador basado en teoría de detección desarrollado para evaluar el procesamiento de imágenes 2DGE. El *Spot efficiency* (SE) [2] establece una relación entre los puntos falsos que se generaron o prevalecieron y puntos que se perdieron al aplicar una técnica de detección. El SE se define como:

$$SE = \frac{P_d - P_f}{P_d + P_p} \quad (3-3)$$

donde P_d , P_f y P_p son el número de puntos detectados, puntos falsos y puntos no detectados respectivamente. SE es utilizado principalmente para evaluar el procesamiento de las imágenes 2DGE con un rango de valores de $(-\infty, 1]$, pero no se ha encontrado que se utilice como un indicador para las técnicas de pre-procesamiento. En este proyecto se introdujo como un Índice cuantitativo para evaluar los efectos de las técnicas de filtrado en la detección de proteínas con imágenes 2DGE.

Para poder usar SE es necesario un algoritmo de detección. En este proyecto se usa un algoritmo de detección basado en la transformada Hough modificada. Esta transformada es ampliamente utilizada en visión por computador y minería de datos en imágenes bidimensionales [60]. La transformada Hough modificada busca un conjunto de datos que se ajusten a la ecuación objetivo, para este caso, se uso la ecuación de una elipse. Antes de aplicar la transformada Hough modificada, se aplica la transformada Watershed, la cual favorece el contorno de los objetos, especialmente de los objetos hiperbólicos. Para este proyecto, se define un valor de radio entre los intervalos de mínimos y máximos, definidos como parámetros en el algoritmo de generación de imágenes sintéticas. Se

tomó como radio inferior el 75 % del radio mínimo y como radio superior el 125 % del radio máximo ingresado al algoritmo de generación de imágenes sintéticas.

Para evaluar las técnicas de corrección de fondo se emplea el Índice de Sustracción $-IS$. Éste proporciona una medida cuantitativa del porcentaje del fondo identificado, por lo cual, es útil para evaluar los métodos de corrección de fondo con un rango de valores de $[0,1]$, definiéndose como:

$$IS = \frac{FI}{FR} \quad (3-4)$$

donde FI es el número de pixeles identificados como fondo y FR es el número de pixeles correspondientes al fondo de la imagen. Para la implementación de este indicador es de suma importancia tener un fondo plenamente caracterizado, condición que se cumple con las imágenes sintéticas.

Finalmente, para evaluar las técnicas de regularización de imagen se propone un indicador relacionado con las proteínas tenues detectadas $-DPT$ con un rango de valores de $[0,1]$. Las proteínas se define como tenue cuando no son detectadas por el método de detección de proteínas implementado (Figura 3-6). El DPT se define como la razón entre el número de proteínas tenues detectadas $-PTD$ y el número total de proteínas tenues $-PT$:

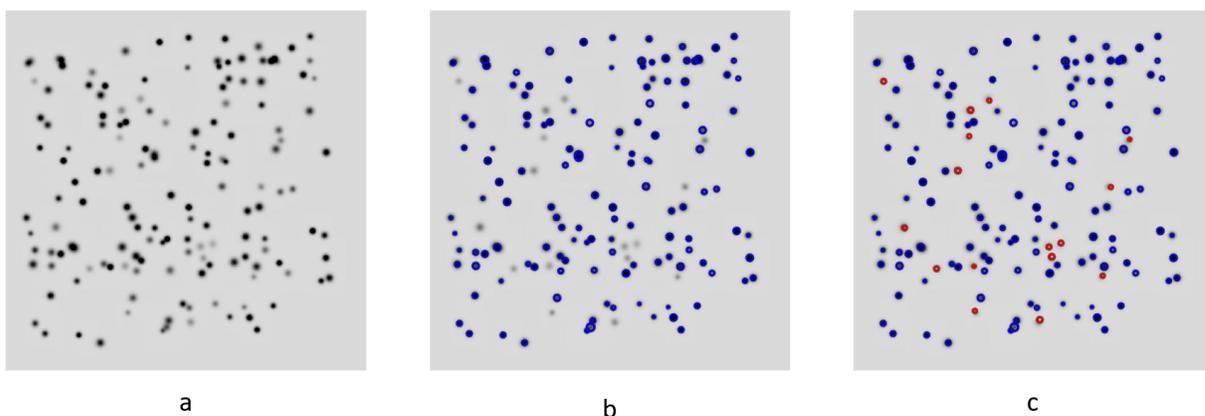


Figura 3-6: a) Imagen sintética b) proteínas identificadas por el algoritmo resaltadas en color azul c) proteínas no identificadas resaltadas en color rojo.

$$DPT = \frac{PTD}{PT} \quad (3-5)$$

Los índices cuantitativos anteriormente nombrados, fueron implementados para evaluar el desempeño en las imágenes sintéticas, porque para poder aplicarlos correctamente se necesita de una imagen completamente caracterizada.

3.2.2. Técnicas de reducción de ruido

Existen numerosas y diversas técnicas de filtrado no lineal. Para el desarrollo de este proyecto se consideraron las técnicas de filtrado no lineal con mayor uso de acuerdo a la literatura: la transformada Contourlet [19], Variación Total [61] y la transformada Wavelet [3].

La técnica más utilizada para la reducción de ruido es la transformada Wavelet por su alto porcentaje de filtrado y aceptables niveles de pérdida de información relevante [40]. De acuerdo a la literatura, la familia Wavelet más apropiada para el procesamiento de imágenes 2DGE es la Wavelet discreta tipo Daubechies, debido a su fácil implementación y conservación de la información espectral [2]. La ecuación de la transformada Wavelet Daubechies esta dada por:

$$\Psi(x) = \sum_{\kappa=-\infty}^{\infty} (-1)^{\kappa} a_{N-1-\kappa} \Psi(2x - \kappa) \quad (3-6)$$

donde N es un numero par, equivalente al conjunto de transformaciones ortogonales utilizados para descomponer la señal y x representa la información de la imagen, para la cual es necesario definir el umbral y el nivel de descomposición.

La transformada Contourlet tiene como objetivo conservar el contorno de los bordes, al realizar una descomposición de los datos en frecuencia y espacio por medio de dos métodos de descomposición [19]. La Figura 3-7 presenta una descripción general de la transformada Contourlet. Estos métodos le permiten una escala múltiple de espacio y frecuencia, además de un alto grado de direccionalidad que ha presentado resultados satisfactorios tanto en reducción del ruido [26], como en los bajos niveles de perdida de información relevante [36]. Para esta transformada los parámetros a definir son el número de pasos en la descomposición piramidal, número de pasos de la descomposición direccional y número de niveles del banco de filtros.

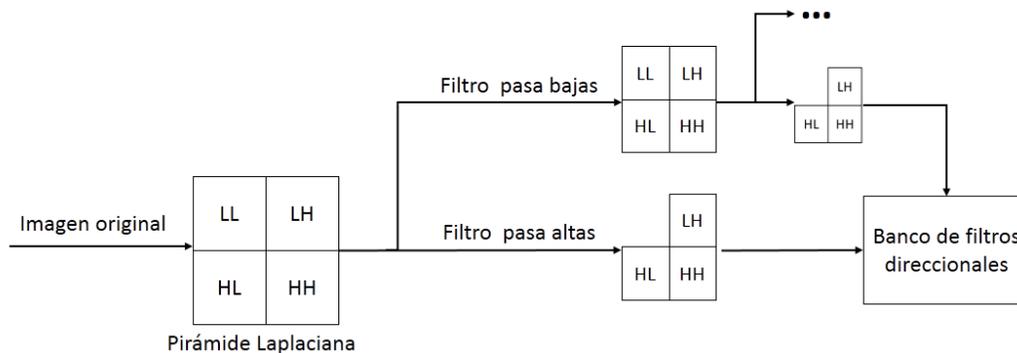


Figura 3-7: Ilustración transformada Contourlet.

La técnica Variación Total [62], consiste en la conservación de bordes por medio de la extracción de las características geométricas de la imagen, con las cuales se logra un suavizado variante que permite optimizar la extracción de los bordes y su información direccional circundante. Sin embargo,

presenta problemas considerables en la eliminación del ruido puntual [63]. El objetivo base de esta transformada es minimizar 3-7, donde E es la norma Euclidiana bidimensional, λ es el parámetro de regularización y $V(y)$ esta dado por 3-8, donde $x(i, j)$ es el pixel con coordenadas (i, j) en la imagen. El desempeño de Variación Total es menor, comparados con la transformada Wavelet de acuerdo al índice SNR [2, 39]. Este filtro es aplicado un número determinado de veces y con un cambio de magnitud por iteración definido por el usuario.

$$\min_y E(x, y) + \lambda * V(x) \quad (3-7)$$

$$V(x) = \sum_{i,j} \sqrt{|y(i+1, j) - x(i, j)|^2 + |x(i, j+1) - x(i, j)|^2} \quad (3-8)$$

En la literatura se encuentran diferentes derivaciones de la transformada Wavelet. Una técnica que se destaca por su desempeño es la que se obtiene al combinar las técnicas transformada Wavelet y Variación Total, denominada WTTV [39]. La técnica logra disminuir la perdida de información relevante y aumentar el nivel de filtrado. La técnica consiste en aplicar primero la transformada Wavelet y posteriormente Variación Total, con el objetivo de minimizar las distorsiones causadas al aplicar la transformada Wavelet. Para esta técnica es necesario definir los parámetros de ambas técnicas de manera meticulosa, ya que se puede agregar información falsa al confundir el ruido con información relevante, debido principalmente a sus limitaciones de información direccional, que afecta la captura geométrica de bordes, generando una distorsión en las fronteras de los puntos de interés [19].

Filtro de difusión no líneal geométrico

El método de filtrado de difusión no lineal geométrico (GNDF, por sus siglas en ingles *Geometric Nonlinear Difusion Filtering*) se basa en la solución de la ecuación diferencial parcial no lineal expresada por 3-9, donde la condición inicial se asume como la imagen original (3-10) [64]. Este método se usa en este proyecto debido al amplio uso que tiene esta técnica en el procesamiento de imágenes médicas. GNDF se caracteriza por disminuir significativamente el ruido, al mismo tiempo que preserva los bordes, lo cual facilita el proceso de detección en imágenes con distribuciones de ruido blanco gaussiano y exponencial [64].

$$\frac{\partial I}{\partial t} = d/dt [C(|\nabla I|) * \nabla I] \quad (3-9)$$

$$I(t=0) = I_o \quad (3-10)$$

En 3-9, C corresponde a los coeficientes de difusión y ∇I es el gradiente de la imagen. Los coeficientes de difusión, C , esta dado por:

$$C(x) = \frac{1}{1 + (x/k)^2} \quad (3-11)$$

donde k es el umbral de ruido que determina la intensidad del filtrado, por tanto este debe ser seleccionado cuidadosamente, ya que de este dependerá la eliminación de manera significativa del ruido y la conservación de los bordes. Existen un gran número de estimadores para el valor de k , en este caso es estimado a partir del valor SNR presente en la imagen [65]. En la aplicación del algoritmo se debe ingresar el número de iteraciones, la constante de iteración y el tamaño de la ventana del vecindario.

3.2.3. Técnicas de corrección de fondo

Las técnicas de corrección de fondo basadas en histogramas son ampliamente utilizadas en el procesamiento de imágenes biomédica, debido a que estas imágenes, así como las imágenes 2DGE, presentan un fondo único. Una de las técnicas empleadas en este proyecto se basa en el histograma de la imagen en escala de grises [51]. Con el histograma de la imagen y un estimado porcentual de la relación de los objetos de interés y fondo, se estima el fondo de la señal para poder ser sustraído, resaltando los objetos presentes en la misma. Para la identificación del fondo se emplea la relación dada por 3-12, donde G_f agrupa los pixeles con niveles de grises por debajo del percentil x . Este procedimiento permite adaptarse a las características de la imagen. La ecuación 3-12 define como intensidades de fondo todos los valores que se encuentren debajo de la media del grupo contenido en el percentil \bar{P}_x . Donde I , son las intensidades de cada pixel.

$$G_f = I \epsilon (I_i < \bar{P}_x) \quad (3-12)$$

A partir de esta técnica de corrección de fondo, Sarkar [66] realiza una modificación que permite obtener 2 niveles de fondo. Esta modificación se describe por 3-13 y 3-14, donde G_{f1} es el primer nivel que contiene los pixeles que se encuentren entre el valor mínimo y la media del percentil objetivo (\bar{P}_x) y G_{f2} es el segundo nivel compuesto de los pixeles con valores entre la media del percentil y el máximo valor agrupado en el percentil objetivo ($max P_x$). Esta técnica da la posibilidad de separar los datos en diferentes grupos, después de tener estos grupos se realiza una comparación entre ellos para separar el fondo de los objetos de interés. Este método realiza una mejora identificación del fondo cuando este presenta irregularidades espaciales, tales como cambios en intensidad lumínica en el fondo. Donde n_1 y n_2 son el número de pixeles contenidos en G_1 y G_2 respectivamente

$$G_{f1} = I \epsilon (I_i < G_1/n_1); G_1 = \sum P_x \epsilon (0, \bar{P}_x) \quad (3-13)$$

$$G_{f2} = I \epsilon (G_1/n_1 < I_i < G_2/n_2); G_2 = \sum P_x \epsilon (\bar{P}_x, max P_x) \quad (3-14)$$

Otra técnica empleada para la corrección de fondo es la aproximación del fondo como una superficie. Esta técnica se basa en la premisa de que el fondo puede asumirse como una sola superficie, en la cual se encuentran los objetos de interés [67]. El primer paso de esta técnica es encontrar una función que describa el fondo como una única superficie, esto se logra por medio de una rutina iterativa, que ajusta los parámetros de una función polinomial o un plano tipo B-Spline [67].

3.2.4. Técnicas de regularización de imagen

Una técnica recurrente en el pre-procesamiento de imágenes para análisis biológico es la ecualización del histograma [68]. La ecualización del histograma busca una distribución de intensidades uniforme, permitiendo resaltar los contrastes, sin perder la información estructural de la imagen [69]. Este procedimiento inicia con el calculo del histograma dado por 3-15, donde L es el número de niveles de intensidad, X_k es un pixel de la imagen con intensidad k , N_k es el número de pixeles con intensidad k y n es el número total de pixeles.

$$P[X_k] = \frac{N_k}{n}; k = 0, \dots, L - 1 \quad (3-15)$$

Una vez calculado el histograma normalizado, se calcula S_k dado por 3-16. La ecualización del histograma se obtiene al mapear cada pixel con nivel de gris k al correspondiente valor de gris dado por S_k .

$$S_k = \sum_{j=0}^k P[X_j] \quad (3-16)$$

Ya que las imágenes 2DGE presentan un comportamiento aleatorio dependiendo de las técnicas de adquisición y tipo de muestra, en [70] se plantea el uso de una técnica adaptativa de ecualización del histograma, que por medio de una estimación recurrente estima los niveles de intensidad adecuados para realizar la distribución uniforme. Esta técnica adaptativa permite tener en cuenta los niveles máximos y mínimos presentes en la imagen. Este método adaptativo se basa en los indicadores dados por 3-17 y 3-18, donde $\Phi(x)$ es una medida de intensidad de los pixeles y $\Upsilon(n_x)$ es el nivel de contraste promedio entre objetos.

$$PM_X = \max \left(\frac{1}{\sum_{x=1}^L n_x} n_x \Phi(x) \right) \quad (3-17)$$

$$M_X = \min \left(\frac{1}{\sum_{x=1}^L n_x} \sum_{x=1}^L n_x \Upsilon(n_x) \right) \quad (3-18)$$

La técnica de histograma descrita en la sección 3.2.3 también se aplica para la regularización de imágenes, con la diferencia que la técnica aplicada en regularización de imagen diferencia el fondo de los objetos de interés para realzar el contraste, disminuyendo la intensidad del fondo e incrementando la intensidad en los bordes de los objetos.

3.3. Metodología propuesta

Las técnicas de pre-procesamiento de reducción de ruido, normalización de imagen y corrección de fondo mitigan anomalías de manera particular, mas no de manera general. Para entender las limitaciones y ventajas de estas técnicas, el Capítulo 4 presenta una comparación de los métodos más empleados en cada corriente de acuerdo a la literatura. En el Capítulo 4 se demuestra que de forma individual estas técnicas no permiten disminuir de manera satisfactoria el efecto de todas las anomalías en las imágenes 2DGE. Por esta razón, se hace evidente la necesidad de una integración

de las técnicas de pre-procesamiento, que permitan reducir las anomalías de manera complementaria.

En este proyecto se propone una metodología de pre-procesamiento de imágenes 2DGE multinivel, que integra técnicas de reducción de ruido, normalización de imagen y corrección de fondo. La metodología propuesta se presenta en la Figura 3-8.

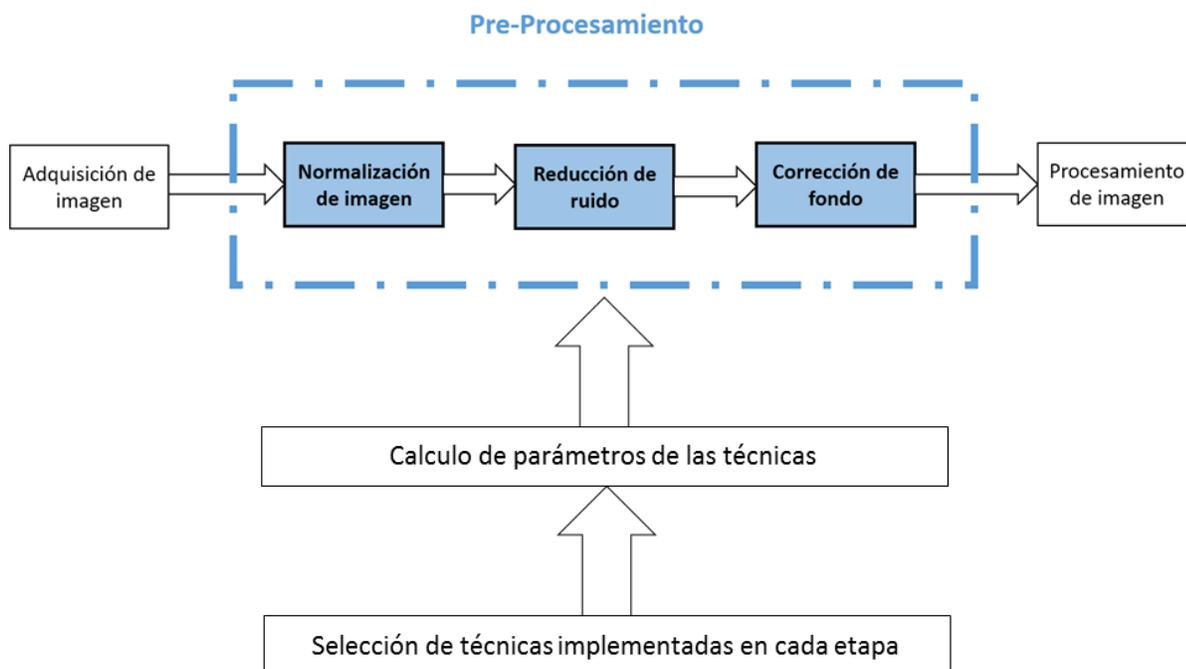


Figura 3-8: Metodología de pre-procesamiento propuesta para la integración de técnicas de reducción de anomalías en imágenes 2DGE.

De acuerdo a lo reportado en la literatura, las técnicas de corrección de fondo mejoran su desempeño cuando las imágenes ingresadas al proceso poseen un fondo con poco ruido, facilitando su identificación y posterior sustracción. Por esta razón se plantea el uso de técnicas de corrección de fondo en el último paso del pre-procesamiento. Debido a que una de las mayores dificultades en el proceso de detección de proteínas es la identificación de proteínas tenues, y considerando que el proceso de filtrado o corrección de fondo puede perjudicar las débiles características morfológicas y espectrales de estas proteínas, se propone realizar como primer paso el proceso de normalización de imagen. Las técnicas de normalización mejora el contraste, lo cual resalta proteínas de baja densidad, disminuyendo su susceptibilidad de atenuación de bordes en una etapa de filtrado. Sin embargo, la normalización al mismo tiempo que mejora el contraste, también intensifica los ruidos puntuales como puntos aislados. De esta manera, como segundo paso se propone usar el proceso de filtrado. Al igual que en otro tipo de imágenes reportadas en la literatura, se sugiere el uso de técnicas de suavizado que no distorsionen el contorno de los objetos de interés [20]. Las técnicas

de filtrado evaluadas en este proyecto ayuda a disminuir de forma efectiva los ruidos puntuales causados por la normalización, así como permiten preservar los bordes de las proteínas.

La metodología propuesta en la Figura 3-8 permite el uso de diferentes técnicas de pre-procesamiento usando el mismo esquema. Se recomienda elegir técnicas a partir de un estudio previo comparativo. Así mismo, se deben definir apropiadamente los parámetros de las técnicas seleccionadas que permitan optimizar la eficacia de la metodología propuesta.

Para realizar la validación de la metodología propuesta se empleó tanto imágenes sintéticas como imágenes reales. La Figura 3-9 presenta el proceso de validación empleado. Las imágenes sintéticas permiten tener control sobre las propiedades morfológicas de las proteínas sintéticas, el fondo de la imagen y las características de las anomalías. Con estas imágenes se realiza el pre-procesamiento y posterior comparación cuantitativa. Posteriormente, se realiza el pre-procesamiento de imágenes reales seleccionadas acorde a la presencia de las anomalías de interés y la calidad de las mismas, finalmente, los resultados de pre-procesamiento de imágenes reales son sometidos a la evaluación de un experto en el tema de imágenes 2DGE

Proceso de Validación

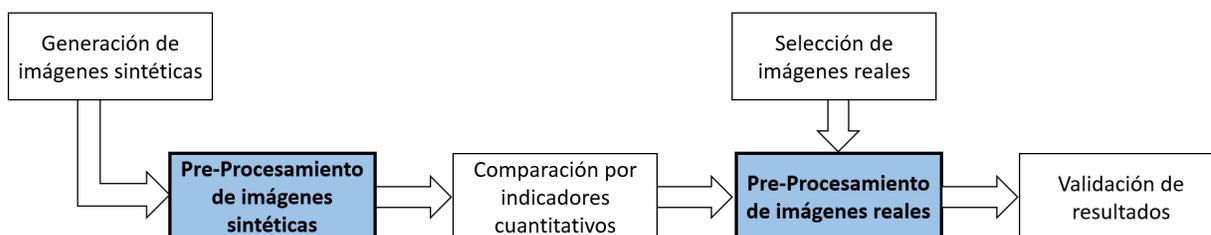


Figura 3-9: Metodología de validación.

4 Resultados y discusión

En este capítulo se presenta la comparación de técnicas de pre-procesamiento reportadas en la literatura para reducir los efectos de anomalías en imágenes 2DGE. Para evaluar estas técnicas se emplea la metodología de validación propuesta en la Figura 3-9. Adicionalmente, en la sección 4.4 se presenta el análisis de la metodología propuesta que integra las tres corrientes de pre-procesamiento de imágenes 2DGE: reducción de ruido, corrección de fondo y normalización de imagen. El desarrollo de este trabajo se realizó en la plataforma Matlab, debido a sus prestaciones computacionales.

4.1. Reducción de ruido

4.1.1. Comparación de técnicas usando imágenes sintéticas

Las técnicas más empleadas para la reducción de ruido en imágenes 2DGE son Wavelet, Contourlet, Variación Total y WVTV (ver Sección 2.2.1). Por esta razón son seleccionadas para este estudio comparativo. Sin bien para la transformada Wavelet, existe una gran variedad de familias que pueden ser utilizadas en este proceso, se exploraron una serie de alternativas para la función base de la transformada, sin embargo, la familia que presentó mejor desempeño con el modelo de optimización de parámetros implementado en este estudio fue la función Daubechies. Los parámetros de las técnicas a evaluar se definen a partir de un proceso de optimización. Se usó el método simplex, con un límite de 50 iteraciones, y con el SNR como función objetivo. Los parámetros iniciales se asumieron iguales al valor medio del rango. Se usó la imagen en la Figura 1-1 para establecer los parámetros para todos los experimentos. Se seleccionó esta imagen debido a que presenta una gran cantidad de anomalías características de imágenes 2DGE. A partir de este proceso se obtienen los siguientes valores: para la transformada Wavelet se consideró un umbral de 0.4532 y un nivel de descomposición de 5, es importante aclarar que la selección de funciones Kernel para el desarrollo de la transformada varían los resultados, por lo cual se conservaron los valores por defecto del paquete de software; para la transformada Contourlet se implementó el código desarrollado por Minh N. Do [18] para el cual se tomó 9 pasos para la descomposición piramidal, 16 pasos para la descomposición direccional tipo Vanishing y 5 niveles para el banco de filtros; para la técnica de Variación total se consideraron 15 iteraciones y un máximo cambio de magnitud por iteración de 0.06 por pixel, finalmente para WVTV, se emplearon los parámetros definidos para las técnicas Variación total y Wavelet. Los estudios se realizaron a partir del conjunto de imágenes sintéticas presentadas en la Figura 3-3.

Para evaluar el desempeño de las técnicas de filtrado no lineal se emplearon los indicadores: SNR (3-1) y SE (3-3). Los resultados de las técnicas Wavelet, Contourlet, Variación Total (VT) y WVTV

respecto a los indicadores SE y SNR se presentan en las Figuras 4-2 a 4-7.

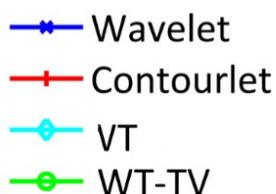


Figura 4-1: Leyendas para la sección 4.1.1 Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WWTV (verde).

Para identificar los resultados de cada técnicas se usa el color azul para Wavelet, rojo para Contourlet, azul claro para Variación Total y verde para WWTV Figura 4-1.

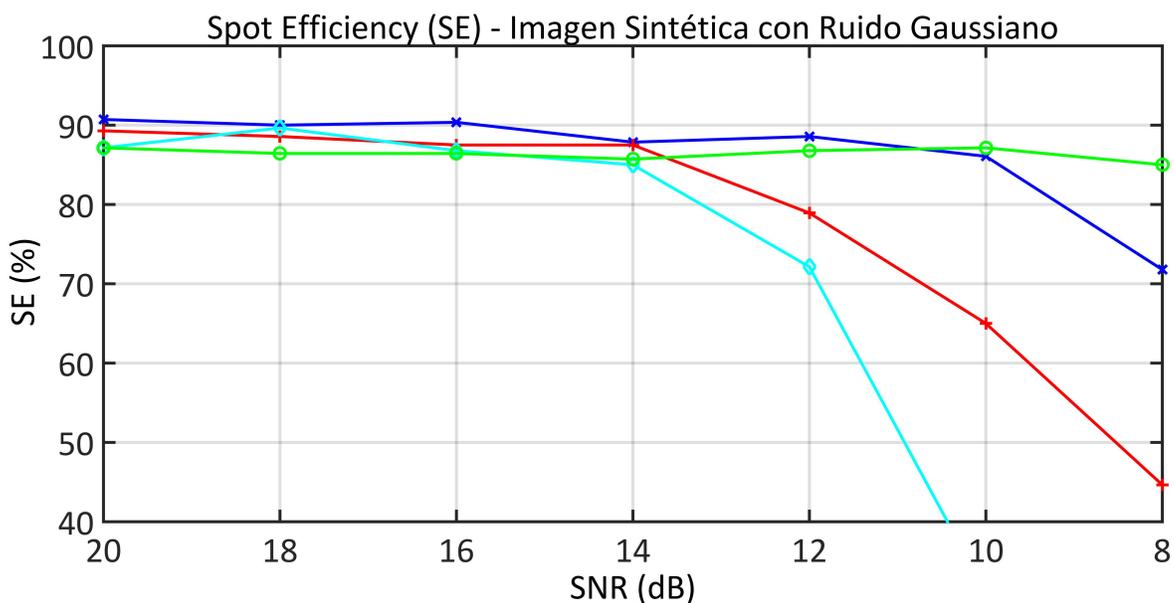


Figura 4-2: SE para imágenes sintéticas con ruido Gaussiano filtradas usando Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WWTV (verde).

La Figura 4-2 presenta el indicador *Spot Efficiency* - SE calculado para las imágenes sintéticas con ruido Gaussiano. Para este tipo de ruido se obtiene resultados similares para las cuatro técnicas evaluadas, con valores de SE entre 85% y 90%, hasta los 14 decibelios (db) de SNR . A partir de 14 db , decrece el desempeño de Contourlet y Variación Total alcanzando valores de SE por debajo de 50%. Por su parte Wavelet mantiene una eficiencia con un valor de SE alrededor del 90% hasta los 10 db y el indicador decae cerca de 70% en 8 db . La técnica WWTV presentan una menor variabilidad, preservando un valor de SE alrededor del 85% para los diferentes niveles de ruido. Por

lo tanto, WVTV es la técnica con menor sensibilidad ante las variaciones de nivel de ruido gaussiano.

En la Figura 4-3 se muestra el comportamiento del indicador SE de las técnicas de filtrado con respecto al ruido Rayleigh. Wavelet obtiene los mejores resultados con un promedio del 90,15 % de SE . Por su parte, Contourlet y Variación Total obtienen valores de SE superiores al 85 % hasta los 12 db . Variación total obtiene valores de SE entre 85 % y 87 % para imágenes con SNR entre 10 y 20 db . De esta manera, la única técnica que no presentó una reducción significativa en su desempeño respecto al nivel de ruido Rayleigh fue Wavelet lo cual es una característica muy deseable para los filtros de pre-procesamiento.

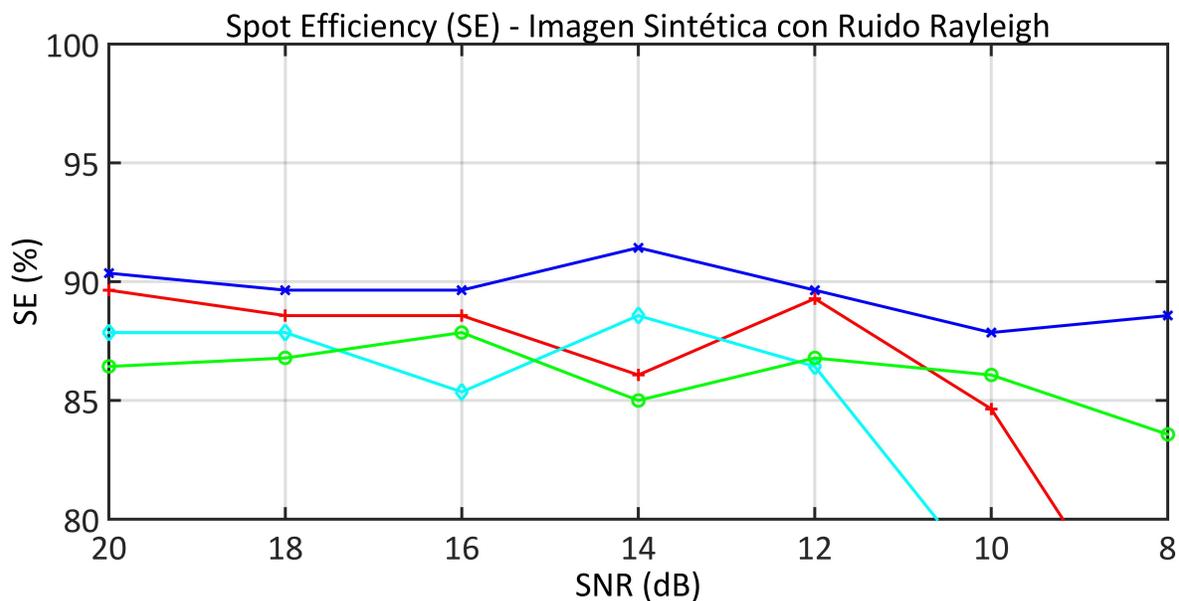


Figura 4-3: SE para imágenes sintéticas con ruido Rayleigh filtradas usando Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WVTV (verde).

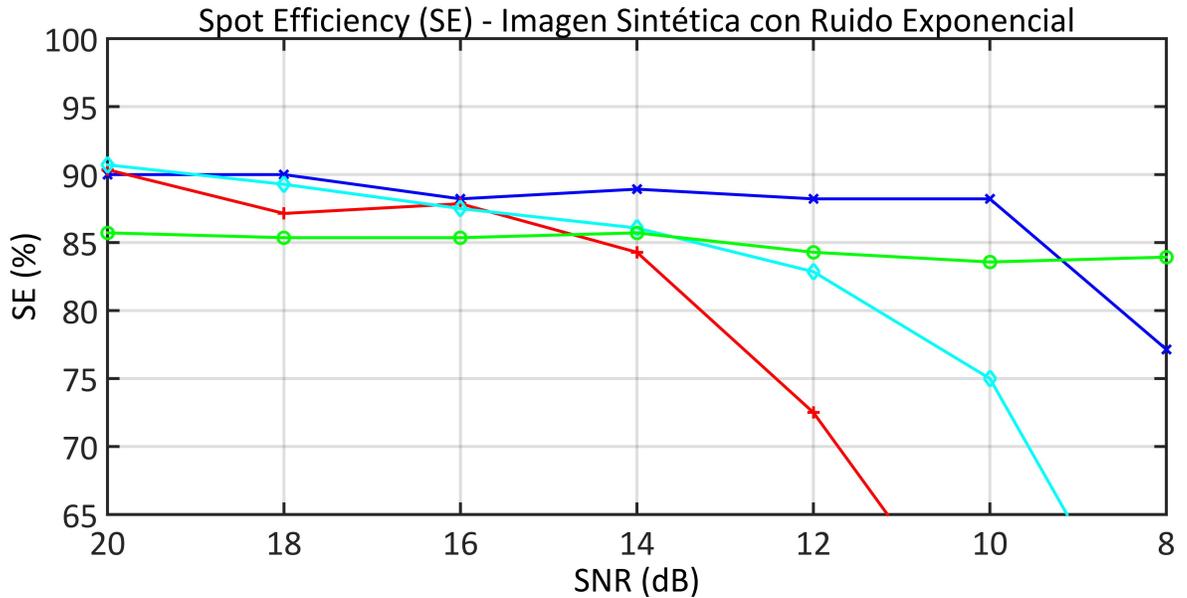


Figura 4-4: SE para imágenes sintéticas con ruido exponencial filtradas usando Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WVTV (verde).

Wavelet presentó el mejor desempeño de acuerdo SE en las imágenes con ruido Exponencial con SNR entre los 10 y 20 db acorde a los resultados obtenidos para las imágenes con ruido Gaussiano, siendo este un filtro de un alto potencial ante niveles de ruidos bajos, por su parte el filtro WVTV presenta una baja sensibilidad al nivel del ruido; mientras que Contourlet y Variación Total, presentan una reducción significativa del desempeño después de los 14 db . WVTV obtiene valores de SE cercanos al 85% para los diferentes niveles de ruido.

La Figura 4-5 presenta la evaluación del desempeño de las técnicas de filtrado en las imágenes con ruido gaussiano usando SNR . El filtro Wavelet permite recuperar en gran medida las propiedades respecto a relación señal a ruido de manera eficaz en imágenes con valores de SNR entre los 19 y 27 db para los diferentes niveles de ruido, presentando el mejor desempeño respecto a las otras técnicas evaluadas. Contourlet y Variación Total obtiene imágenes con SNR superiores a 15 db para todos los niveles de ruido. Por su parte, los resultados de WVTV presenta una baja variabilidad respecto al nivel de ruido, obteniendo imágenes con SNR entre 18 y 20 db .

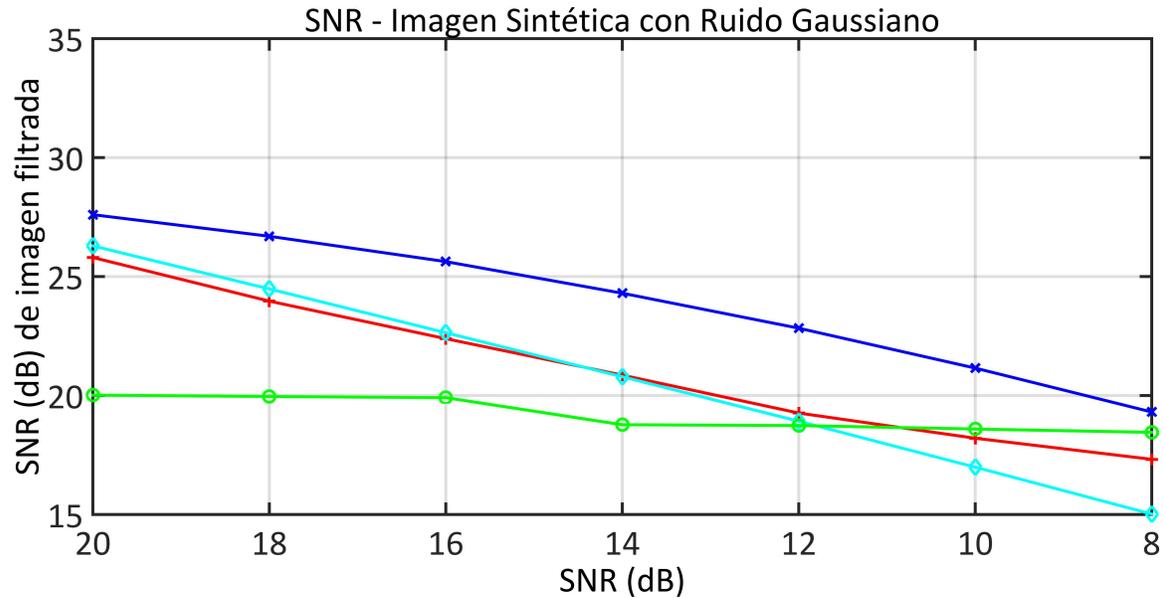


Figura 4-5: *SNR* para imágenes sintéticas con ruido Gaussiano filtradas usando Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WVTV (verde).

Con respecto al ruido Rayleigh, la Figura 4-6 muestra que todas las técnicas evaluadas (Wavelet, Contourlet, Variación Total y WVTV) obtienen imágenes filtradas con valores similares de *SNR* para imágenes con *SNR* entre 14 y 8 *db*. Sin embargo, para imágenes con ruido entre los 20 y 16 *db* de *SNR*, Wavelet obtiene los mejores resultados con valores de *SNR* entre 20 y 27 *db*. Por otra parte, WTTV obtiene los valores más bajos de *SNR*. A diferencia de resultados anteriores, las imágenes filtradas con WVTV presentan una mayor variabilidad respecto al indicador *SNR*, resaltando que ambos indicadores no poseen una relación directa.

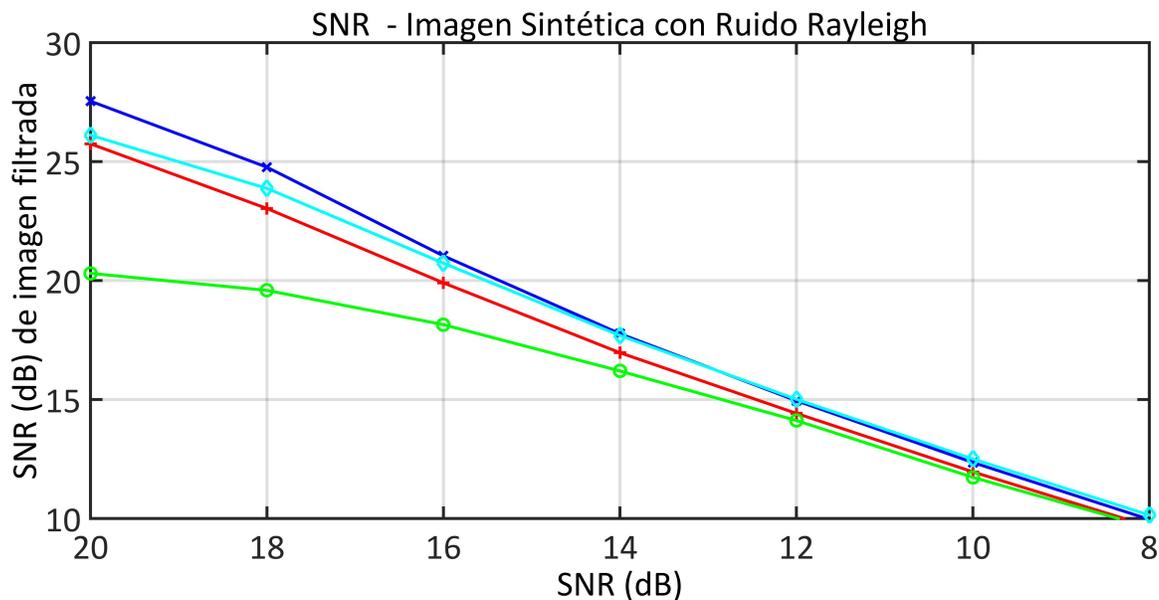


Figura 4-6: *SNR* para imágenes sintéticas con ruido Rayleigh filtradas usando Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WVTV (verde).

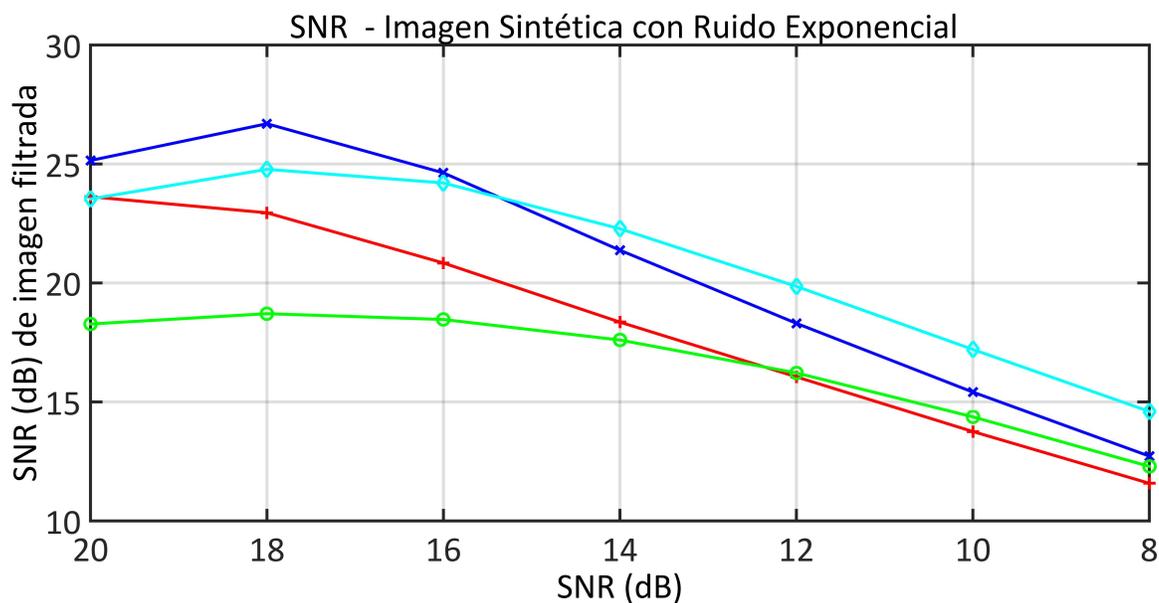


Figura 4-7: *SNR* para imágenes sintéticas con ruido Exponencial filtradas usando Wavelet (azul), Contourlet (rojo), Variación Total (azul claro) y WVTV (verde).

Para las imágenes con ruido Exponencial, el mejor desempeño respecto al *SNR* se obtiene con Wavelet para imágenes con *SNR* entre los 20 y 16 *db*, Variación Total para imágenes con *SNR* entre los 14 y 8 *db* (ver Figura 4-7). Por otra parte, la Figura 4-7 muestra que el peor desempeño se

obtiene con WVTV para imágenes con SNR entre los 20 y 14 db y Contourlet para imágenes con SNR entre 12 y 8 db .

Las tablas 4-1 y 4-2 resumen los indicadores SE y SNR calculados a partir de las imágenes filtradas con Wavelet, Contourlet, Variación Total y WVTV para los tres tipos de ruido. En general, el mejor desempeño respecto al indicador SE para los diferentes ruidos se obtiene con la técnica de filtrado Wavelet (Tabla 4-1). Resultados similares se obtienen para el indicador SNR en imágenes con ruido gaussiano (Tabla 4-2). Sin embargo, para imágenes con ruido Rayleigh y Exponencial, Variación Total obtiene mejores resultados para altos niveles de ruido, es decir, para imágenes con SNR entre los 12 y 8 db .

Nivel de Ruido	20 db	18 db	16 db	14 db	12 db	10 db	8 db
SE : Imágenes con Ruido Gaussiano [%]							
Wavelet	90.71	90	90.36	87.86	88.57	86.07	71.79
Contourlet	89.29	88.57	87.5	87.51	78.93	65	44.64
Variación total	87.14	89.64	86.79	85	72.21	37.89	25.12
WVTV	87.14	86.43	86.43	85.71	86.79	87.14	85.16
SE : Imágenes con Ruido Rayleigh [%]							
Wavelet	90.36	89.64	89.64	91.43	89.64	87.86	88.57
Contourlet	89.64	88.57	88.57	86.07	89.29	84.64	72.36
Variación total	87.86	87.86	85.36	88.57	86.43	87.59	51.25
WVTV	86.43	86.79	87.86	85	86.79	86.07	83.57
SE : Imágenes con Ruido Exponencial [%]							
Wavelet	90	90	88.93	88.21	88.21	88.21	77.14
Contourlet	90.36	87.14	87.86	84.29	72.5	61.84	48.25
Variación total	90.71	89.29	87.5	86.07	82.86	75	62.37
WVTV	85.71	85.36	85.36	85.71	84.29	83.57	83.93

Tabla 4-1: SE para imágenes filtradas usando Wavelet, Contourlet, Variación Total y WVTV.

Nivel de Ruido	20 <i>db</i>	18 <i>db</i>	16 <i>db</i>	14 <i>db</i>	12 <i>db</i>	10 <i>db</i>	8 <i>db</i>
<i>SNR</i> : Imágenes con Ruido Gaussiano [<i>dB</i>]							
Wavelet	27.61	26.7	25.55	24.24	22.82	21.15	19.29
Contourlet	25.8	24.08	22.36	20.79	19.41	18.26	17.33
Variación total	26.29	24.48	22.64	20.79	18.92	16.99	15.02
WVTV	20.01	19.97	19.9	18.76	18.75	18.63	18.46
<i>SNR</i> : Imágenes con Ruido Rayleigh [<i>dB</i>]							
Wavelet	27.54	24.76	21.04	17.77	14.95	12.35	9.86
Contourlet	25.75	20.03	19.9	16.97	14.42	11.96	9.25
Variación total	26.11	23.87	20.73	17.69	15.01	12.5	10.5
WVTV	20.3	19.59	18.15	16.2	14.12	11.73	8.78
<i>SNR</i> : Imágenes con Ruido Exponencial [<i>dB</i>]							
Wavelet	25.15	6.69	24.63	21.38	18.3	15.41	12.72
Contourlet	23.63	22.95	20.84	18.35	16.05	13.76	11.58
Variación total	23.53	24.78	24.2	22.28	19.86	17.2	14.6
WVTV	18.28	18.71	18.47	17.6	16.22	14.37	12.29

Tabla 4-2: *SNR* para imágenes filtradas usando Wavelet, Contourlet, Variación Total y WVTV.

En la Figura 4-8 se puede observar la imagen con ruido mientras que la Figura 4-9 presenta ejemplos de las imágenes filtradas usando Wavelet, Contourlet, Variación Total y WVTV. A partir de la imagen dispuesta en Figura 3-3 b), la cual cuenta con un nivel de ruido de 10 *db* (ver Tabla 3-1). En la Figura 4-9 se puede observar que tanto el filtro Variación Total y como el filtro WVTV generan una distorsión significativa en la imagen, reduciendo el ruido puntal pero disminuyendo la nitidez de la imagen. Las transformadas Wavelet y Countourlet preservan mejor los bordes, favoreciendo la detección de proteínas. Sin embargo, la transformada Countourlet presenta dificultades ante niveles elevados de ruido, generando unas vetas en las imágenes que puede llegar a disminuir el desempeño de las técnicas de detección.

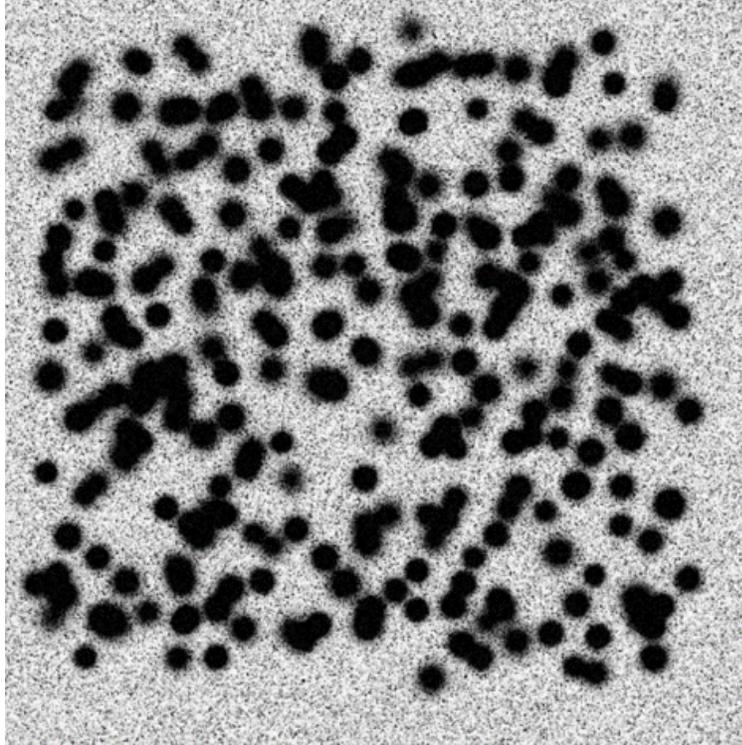


Figura 4-8: Imágenes sintéticas con ruido Gaussiano de media 0 y desviación estándar 0.2430 (10 *db*)

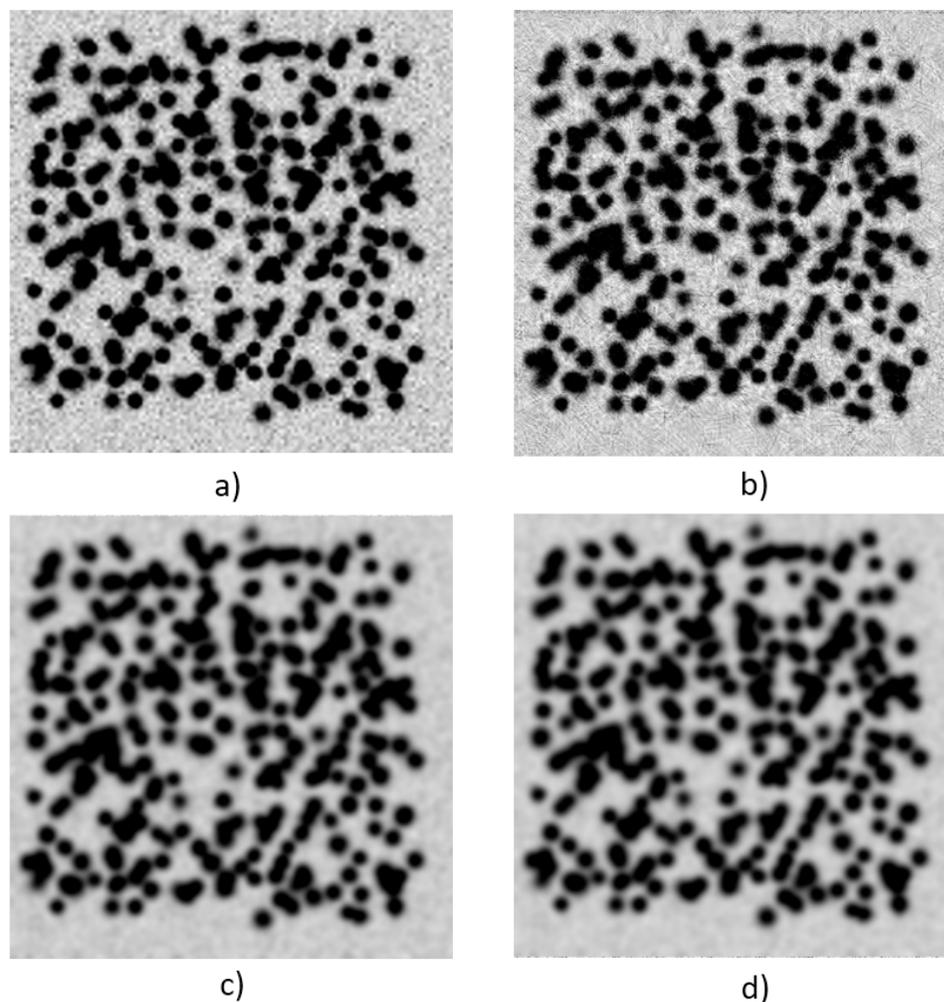


Figura 4-9: Imágenes sintéticas con ruido Gaussiano de media 0 y desviación estándar 0.2430 (10 *db*) filtradas usando: a) Wavelet b) Contourlet c) Variación total y d) WVTV.

Discusión

A partir de la comparación cuantitativa de acuerdo a los indicadores *SNR* y *SE* entre las técnicas de reducción de ruido ya utilizadas de acuerdo a la revisión bibliográfica: Wavelet, Contourlet, Variación Total y WVTV empleando las imágenes sintéticas y los indicadores *SE* y *SNR*, se determina que Wavelet presentó el mejor desempeño de manera general para los diferentes tipos y niveles de ruido.

La técnica de pre-procesamiento WTTV se presenta como opción tentativa para la reducción de ruido en imágenes 2DGE, debido a su baja sensibilidad respecto al ruido de acuerdo al indicador *SE* permitiendo obtener resultados mas homogéneos. Sin embargo, respecto al indicador *SNR* no obtuvo resultados significativos. Por su parte, Variación Total presentó los mejores resultados

respecto al indicador SNR para las imágenes con ruido Rayleigh y Exponencial para niveles altos (entre 14 y 8 db de SNR). Sin embargo, usando el indicador SE , Variación Total obtuvo un bajo desempeño.

Wavelet mostró un desempeño consistente tanto para el indicador SE como para SNR . Como se puede notar en la **4-9**, Wavelet preserva los bordes en las proteínas, pero también se alcanza a apreciar un nivel de ruido en el fondo de la imagen después de su procesamiento. Por esta razón, en el siguiente Capítulo se explora el uso del filtro de difusión no lineal geométrico - GNDF. El cual acuerdo a la literatura, GNDF permite reducir el ruido de forma significativa sin perder la información de los bordes y demás características espaciales de los objetos de interés.

4.1.2. Filtrado de imágenes sintéticas usando GNDF

En la comparación entre los filtros GNDF y Wavelet, los parámetros de las técnica Wavelet se mantuvieron iguales. Respecto a GNDF, se usaron 35 iteraciones con una constante de integración de 0.2, que garantiza estabilidad y convergencia dentro del algoritmo, además el tamaño de ventana fue de 5×5 . Se emplearon las mismas imágenes sintéticas que en el experimento anterior.

Las Figuras **4-10** a **4-12** presentan los resultados de desempeño de acuerdo a los indicadores SE y SNR . Las etiquetas empleadas son: Wavelet (línea azul) y GNDF (línea roja). De acuerdo a los resultados obtenidos para el ruido Gaussiano (Figura **4-10**), las técnicas Wavelet y GNDF presentaron un comportamiento comparable, mostrando esta última una menor sensibilidad al nivel de ruido. Para la imagen con SNR igual a 8 db , es decir con el mayor nivel ruido, GNDF obtuvo un SE superior al 75 %, mientras para Wavelet el SE fue menor al 70 %.

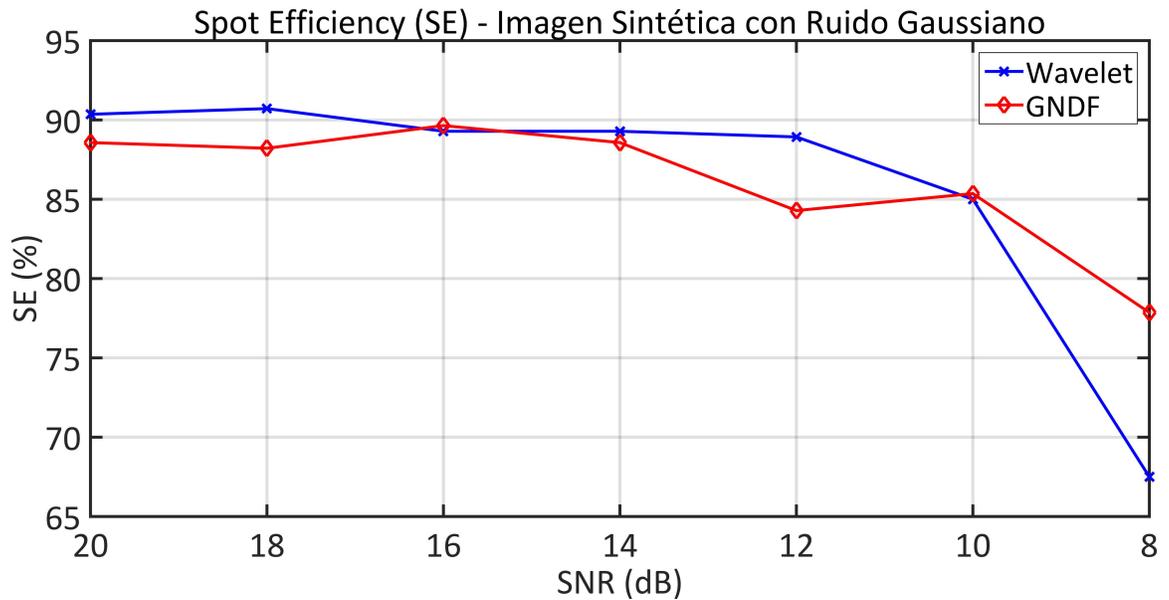


Figura 4-10: SE para imágenes sintéticas con ruido gaussiano usando Wavelet (azul) y GNDF (rojo).

Los resultados para las imágenes sintéticas con ruido Rayleigh muestran un mejor desempeño con el filtro Wavelet (Figura 4-11). Sin embargo, la diferencia entre los valores de SE obtenidos a partir de la aplicación de los filtros Wavelet y GNDF son menores al 5% con un comportamiento consistente. Por su parte, Wavelet y GNDF muestran un desempeño similar respecto al SE con las imágenes sintéticas con ruido Exponencial ambas con un desempeño superior al 85% (Figura 4-12).

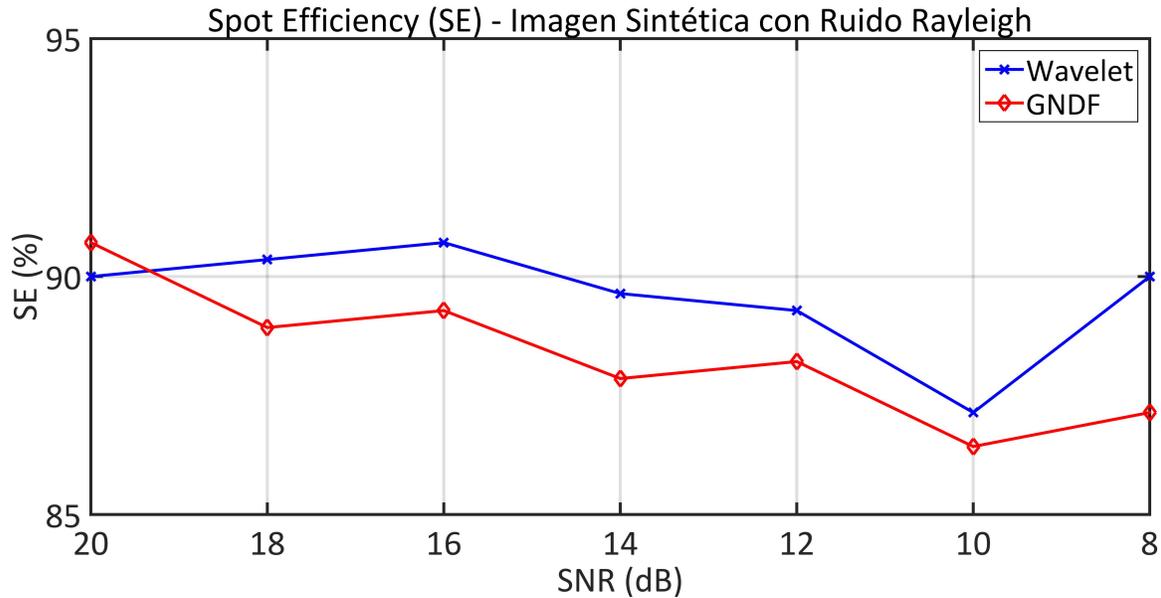


Figura 4-11: *SE* para imágenes sintéticas con ruido Rayleigh usando Wavelet (azul) y GNDF (rojo).

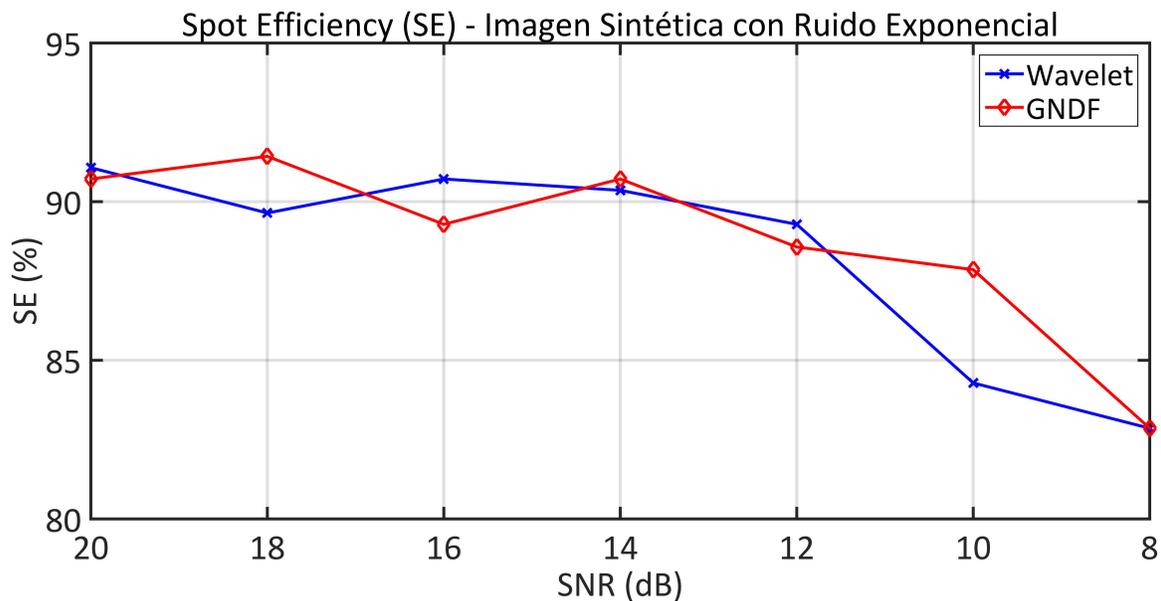


Figura 4-12: *SE* para imágenes sintéticas con ruido Exponencial usando Wavelet (azul) y GNDF (rojo).

La diferencia mas contundente a favor de la técnica de pre-procesamiento GNDF se presenta respecto al indicador *SNR*, GNDF obtiene un mejor filtrado de las imágenes sintéticas con los diferentes tipos de ruido. Para la imagen sinética con ruido Gaussiano, se logra una mejora de cerca 2 db en

la imagen filtrada con GNDF respecto al resultado con Wavelet (ver Figura 4-13). Para el caso del ruido Rayleigh, la diferencia de los valores de SNR obtenidos es menor a 1 db para los diferentes niveles. Sin embargo, GNDF presenta los mejores resultados (ver Figura 4-14). Por su parte, los resultados con respecto al ruido Exponencial, los valores de SNR obtenidos para GNDF y Wavelet alcanzaron 3 db de diferencia (ver Figura 4-15), indicando que esta técnica permite obtener una alta conservación de la información disminuyendo ampliamente el nivel de ruido presente en la imagen y conservando la información morfológica alrededor de los puntos de interés.

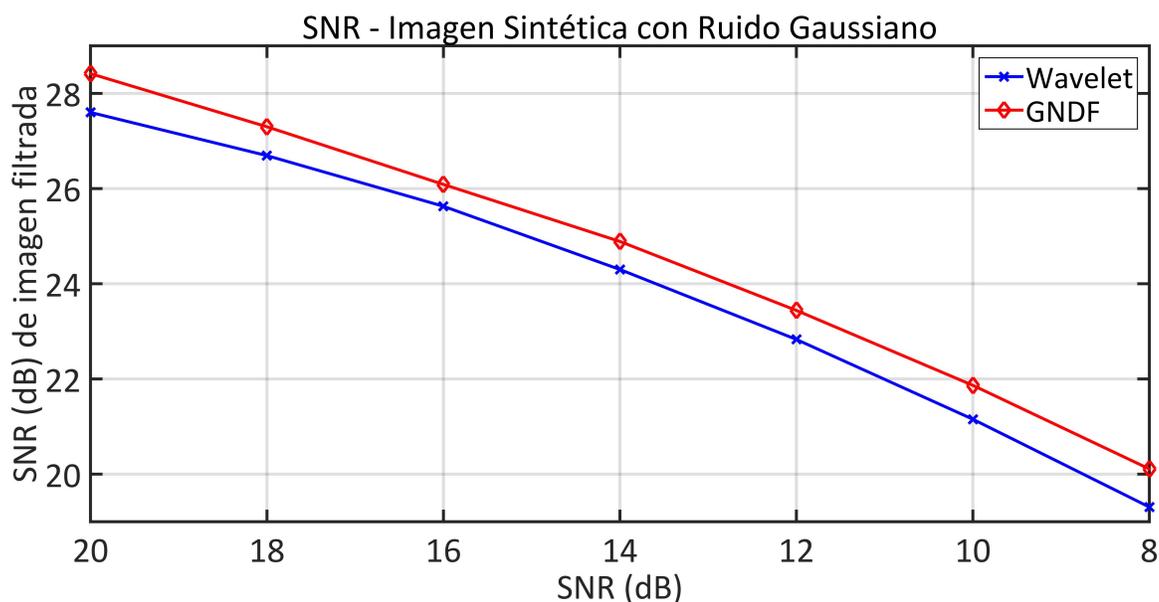


Figura 4-13: SNR para imágenes sintéticas con ruido Gaussiano usando Wavelet (azul) y GNDF (rojo).

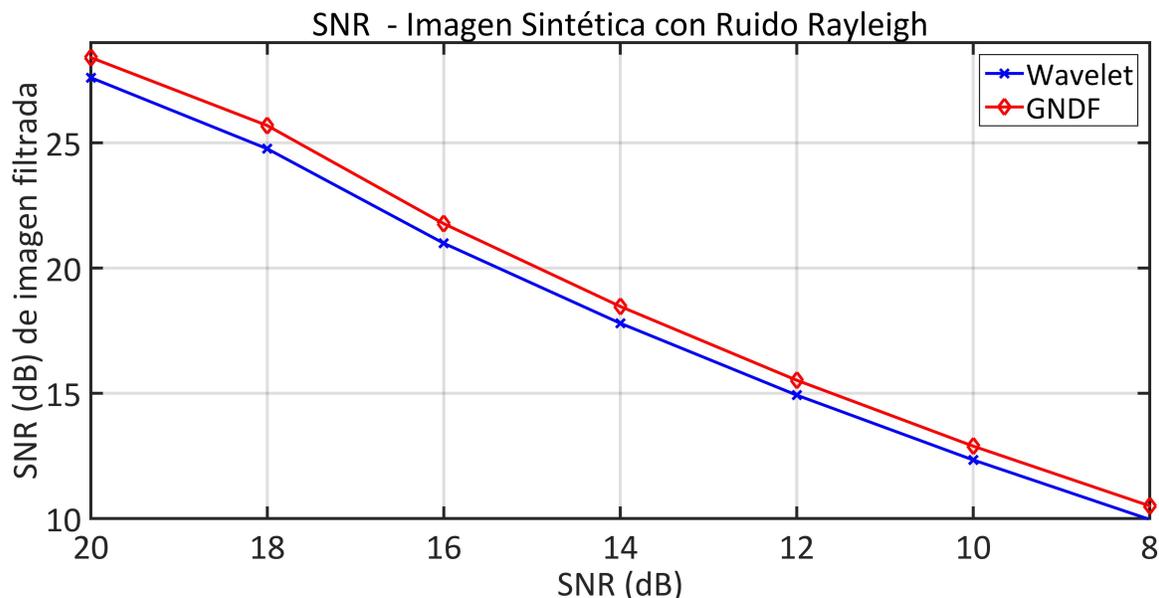


Figura 4-14: *SNR* para imágenes sintéticas con ruido Rayleigh usando Wavelet (azul) y GNDF (rojo).

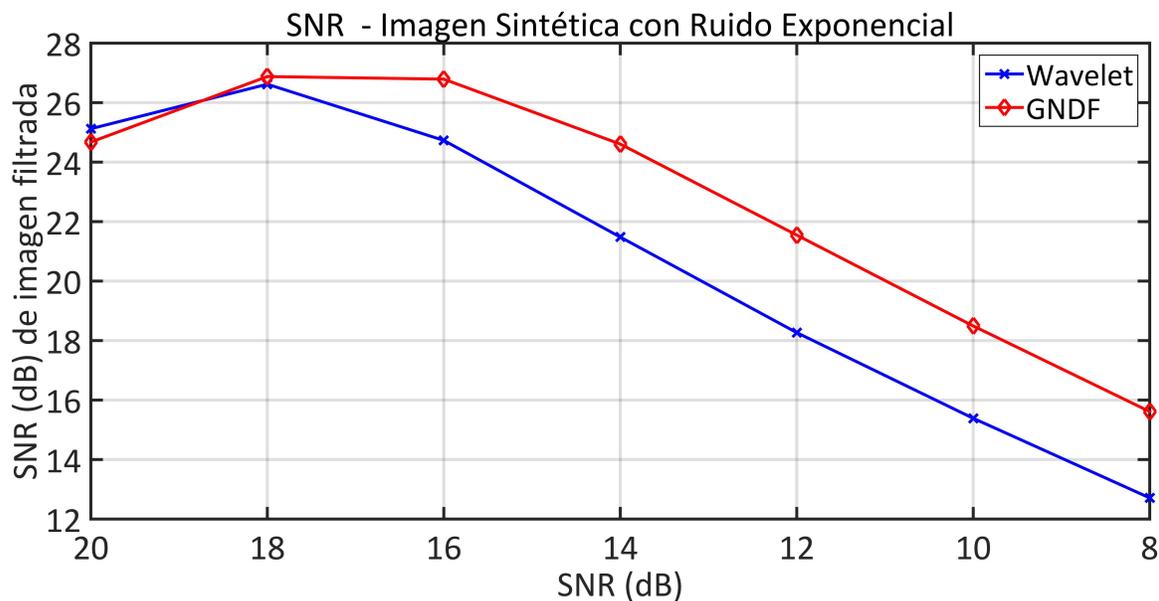


Figura 4-15: *SNR* para imágenes sintéticas con ruido Exponencial usando Wavelet (azul) y GNDF (rojo).

En la Figura 4-16 se muestra el resultado obtenido al aplicar las técnicas de reducción de ruido Wavelet y GNDF a una imagen con adición de ruido Gaussiano con media de cero y $\sigma = 0.3091$ ($SNR = 10db$). Se puede apreciar la reducción de ruido que se obtiene con GNDF, conservando las propiedades morfológicas de los objetos.

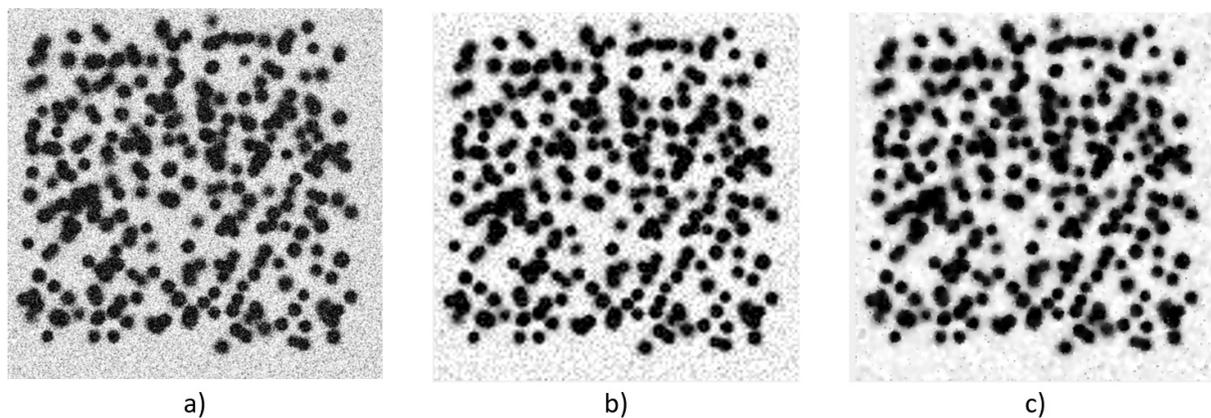


Figura 4-16: Imágenes sintéticas con ruido Gaussiano de media 0 y desviación estándar 0.3091 (10 *db*) filtradas usando b) Wavelet y c) GNDF

La tabla 4-3, presenta los valores de *SE* y *SNR* graficados en las Figuras 4-10 a 4-12. Se resaltan los mejores valores de *SE* y *SNR* para cada tipo y nivel de ruido. Se puede observar que GNDF es una técnica competitiva respecto a ambos indicadores, principalmente para mejorar el *SNR*. Indicado una alta eficiencia de esta técnica, de alta competitividad respecto a la técnica Wavelet la cual se presenta en la bibliografía consultada como una técnica con amplias ventajas en este campo de aplicación.

<i>SE</i> : Imágenes con Ruido Gaussiano [%]							
Wavelet	90.36	90.71	89.29	89.29	88.93	85	67.5
GNDF	88.57	88.21	89.64	88.57	84.29	85.36	77.86
<i>SE</i> : Imágenes con Ruido Rayleigh [%]							
Wavelet	90	90.36	90.71	89.64	89.29	87.14	90
GNDF	90.71	88.93	89.29	87.86	88.21	86.43	87.14
<i>SE</i> : Imágenes con Ruido Exponencial [%]							
Wavelet	91.07	91.43	90.71	90.36	89.29	84.29	82.86
GNDF	90.71	89.64	89.29	90.71	88.57	87.86	82.87
<i>SNR</i> : Imágenes con Ruido Gaussiano [dB]							
Wavelet	27.6	26.69	25.63	24.3	22.83	21.15	19.31
GNDF	28.42	27.3	26.09	24.89	22.98	21.87	20.11
<i>SNR</i> : Imágenes con Ruido Rayleigh [dB]							
Wavelet	27.6	24.76	20.99	17.8	14.93	12.34	9.78
GNDF	28.4	25.69	21.77	18.47	15.52	12.89	10.5
<i>SNR</i> : Imágenes con Ruido Exponencial [dB]							
Wavelet	25.08	26.62	24.73	21.48	18.27	15.39	12.71
GNDF	24.67	26.88	26.79	24.61	21.55	18.49	15.61

Tabla 4-3: *SE* y *SNR* para imágenes filtradas usando Wavelet y GNDF

Discusión

La comparación entre las técnicas de filtrado Wavelet y GNDF mostraron resultados comparables respecto al indicador *SE*, con diferencias inferiores al 2% (Tabla 4-3). Por otra parte, se mejoró el *SNR* de las imágenes filtradas usando GNDF entre 1 y 2 db (Tabla 4-3). A pesar de las pequeñas diferencias obtenidas respecto a los indicadores *SE* y *SNR*, en las imágenes filtradas se observa que GNDF reduce significativamente el ruido de la imagen y no causa distorsiones en las formas de las proteínas sintéticas. Por lo cual, GNDF es una técnica competitiva que puede ayudar a mejorar la detección de proteínas en imágenes 2DGE, donde si bien, en este estudio se demuestra su utilidad aun hay un gran campo de investigación para la optimización de esta técnica en este campo de aplicación ya que es una técnica no antes aplicada para este tipo de imágenes, presentando una desventaja respecto a su proceso de optimización de parámetros comparada con las técnicas tradicionales en este campo como lo es la transformada Wavelet y la transformada Contourlet.

4.1.3. Comparación de técnicas usando imágenes reales

Con el objetivo de validar los resultados obtenidos con las imágenes sintéticas, se emplea las imágenes 2DGE presentadas en las Figura 2-1 b y c. Estas imágenes presentan anomalías clásicas de ruido, causadas por el sistema de adquisición, además de presentar ruidos puntuales y proteínas de baja densidad. Para los experimentos con las imágenes 2DGE reales se conservaron los parámetros

de las técnicas de filtrado usados en los experimentos con datos sintéticos.

Los resultados ante las diferentes técnicas de filtrado para la imagen 2DGE de la Figura **2-1-b** se muestran en la Figura **4-17**. Adicionalmente, se presenta el gradiente morfológico en la Figura **4-18**. Las Figuras incluyen la imagen original y las imágenes filtradas con Wavelet, Contourlet, Variación Total, WVTV y GNDF. En la Figura **4-17** se puede apreciar el efecto de las técnicas de filtrado principalmente en el fondo de la imagen. Para las técnicas WVTV y GNDF la reducción de ruido en el fondo es más evidente. En la Figura **4-18** se aprecia un cambio en el nivel de contraste en el gradiente morfológicos de las imágenes filtradas. A partir de la imagen original, se observa que la técnica GNDF presenta un alto nivel de gradiente morfológico, el cual se asocia con el nivel de contraste y una mayor conservación de los bordes. Por lo cual se espera mejorar el proceso de detección de las proteínas en las imágenes 2DGE al filtrarlas usando GNDF.

El procedimiento se replicó con la imagen 2DGE de la Figura **2-1-c**. Esta imagen tiene un alto nivel de distorsión, causado por un cambio decreciente en la homogeneidad del fondo. Este efecto resalta el ruido en la zona inferior y atenúa el contraste de las proteínas ubicadas en la parte superior de la imagen. Los resultados ante las diferentes técnicas de filtrado para la imagen 2DGE de la Figura **2-1-c** se muestran en la Figura **4-19**. Nuevamente, se presenta el gradiente morfológico en la Figura **4-20**. Las Figuras incluyen la imagen original y las imágenes filtradas con Wavelet, Contourlet, Variación Total, WVTV y GNDF. En la Figura **4-19** se pueden notar como de manera significativa las técnicas WVTV y GNDF reducen el ruido. De igual forma, en el gradiente morfológico presentado en la Figura **4-20**, GNDF y WVTV presentan un alto contraste en las zonas media e inferior de la imagen.

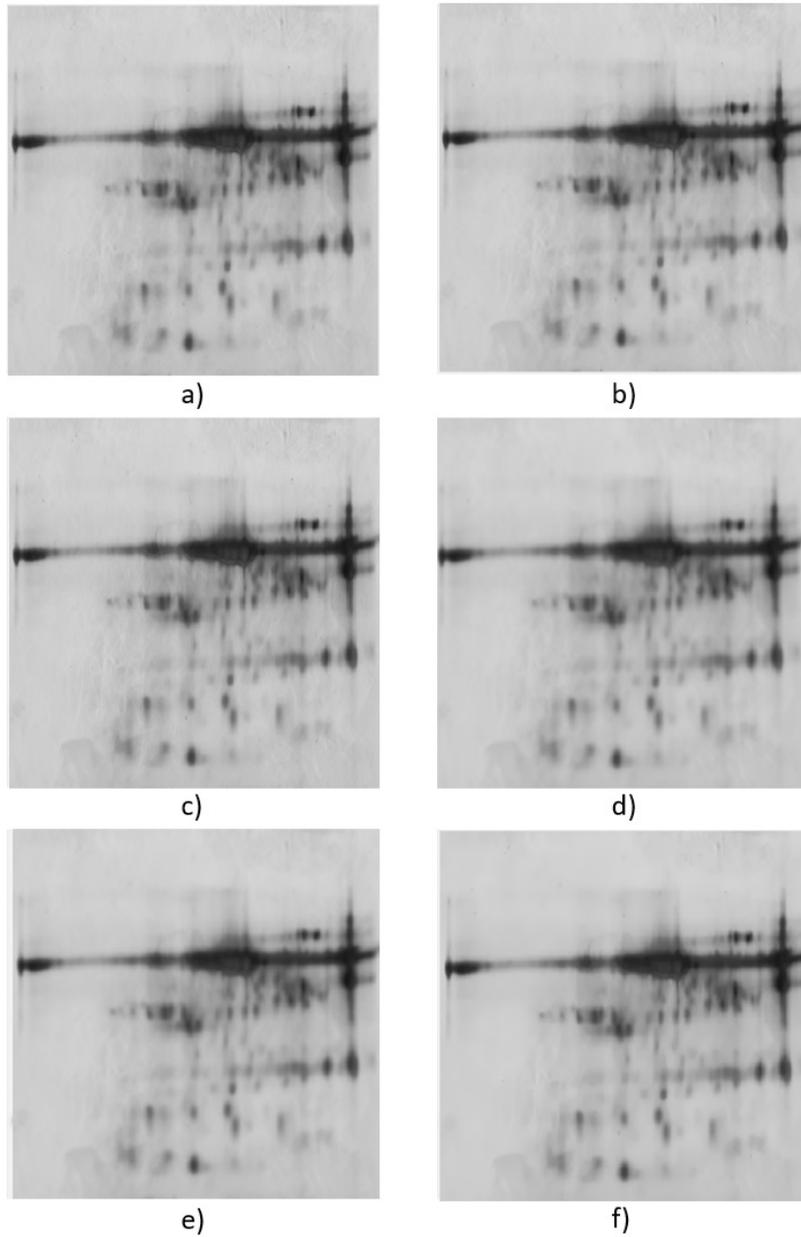


Figura 4-17: Comparación técnicas de filtrado para imágenes 2DGE reales. a) Imagen Original b) Wavelet c) Contourlet d) Variación Total e) WTV f) GNDF

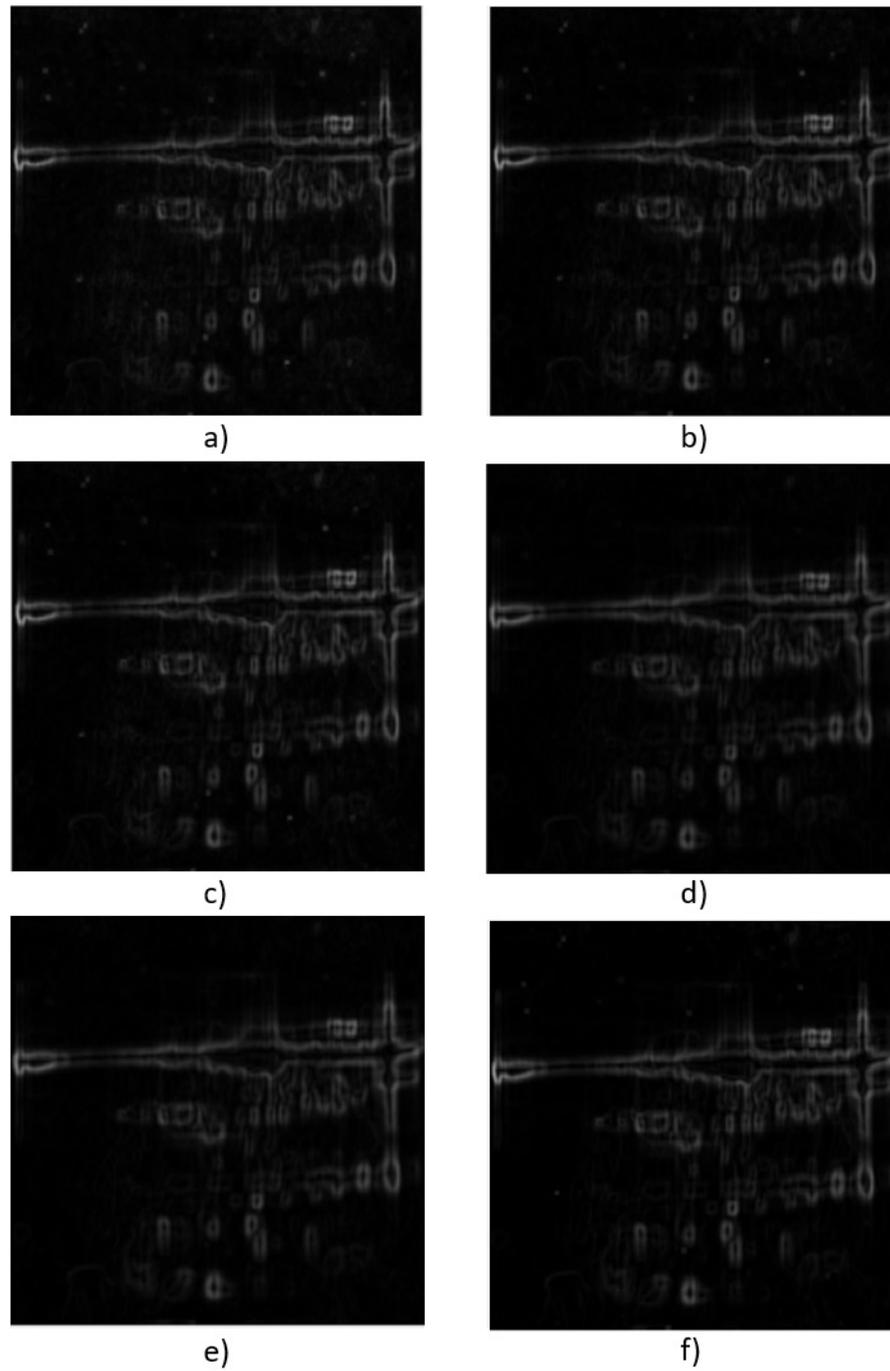


Figura 4-18: Gradiente Morfológico de las imágenes filtradas. a) Imagen Original b) Wavelet c) Contourlet d) Variación Total e) WVTV f) GNDF

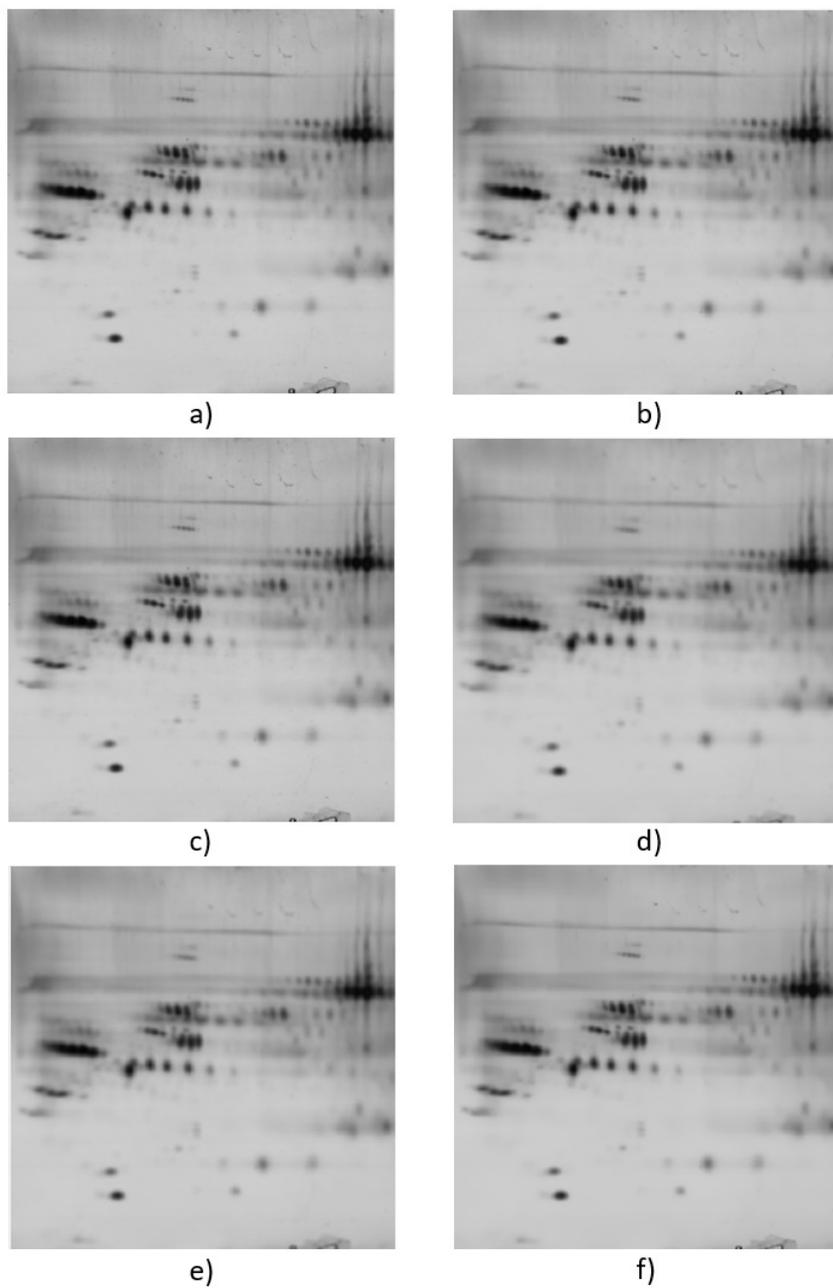


Figura 4-19: Comparación técnicas de filtrado para imágenes 2DGE reales. a) Imagen Original b) Wavelet c) Contourlet d) Variación Total e) WVTV f) GNDF

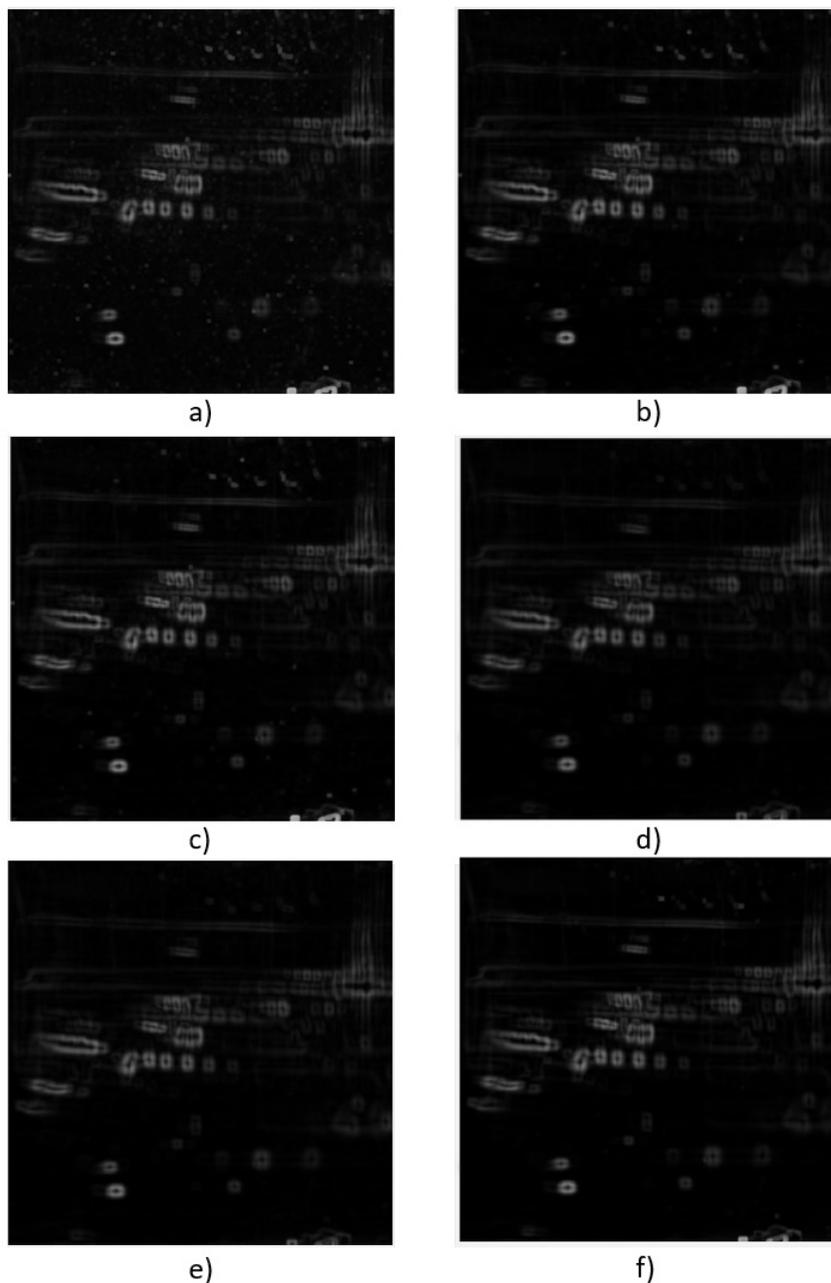


Figura 4-20: Gradiente Morfológico de las imágenes filtradas. a) Imagen Original b) Wavelet c) Contourlet d) Variación Total e) WVTV f) GNDF

Para realizar una comparación cuantitativa de las técnicas de filtrado y su efecto sobre la detección de proteínas en imágenes 2DGE se usó la imagen de la Figura 3-4. Esta imagen cuenta con identificación de proteínas tal como se muestra en la Figura 4-21, lo cual permite calcular el porcentaje de proteínas detectadas desde las imágenes filtradas por las cinco técnicas. La tabla 4-4 resume el desempeño de las técnicas acorde al número de proteínas detectas respecto al número de proteínas conocidas. A partir de la imagen sin ningún tipo de filtrado se detectaron el 75.8%

de las proteínas. Para las imágenes filtradas se obtienen mejores resultados de detección, excepto para Variación Total con la cual se detecto solo el 71.2% de las proteínas. Filtrando con WVTV se obtuvo el 78.8% de las proteínas. Desde las imágenes filtradas con Wavelet y Contourlet se obtuvo el 87.9% y 81.8% de las proteínas respectivamente. Por su parte, el mejor resultado se obtiene a partir de la imagen filtrada con GNDF, con un 90.9% de proteínas detectadas.

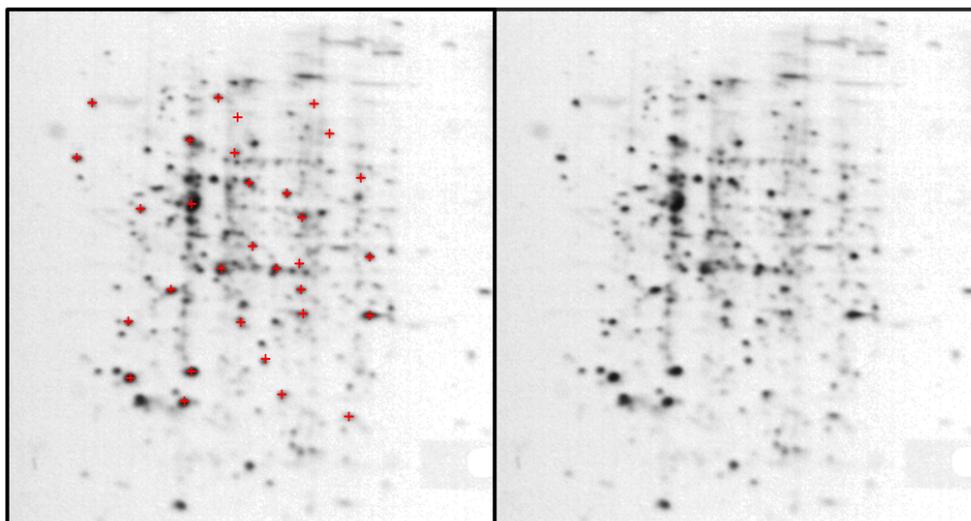


Figura 4-21: Imagen 2DGE con proteínas indentificadas.

Procentaje de Detección [%]	
Control	75.8
Wavelet	87.9
Contourlet	81.8
Variación Total	71.2
WVTV	78.8
GNDF	90.9

Tabla 4-4: Comparación técnicas de filtrado usando imagen con proteínas identificadas.

Discusión

La comparación analítica de las imágenes filtradas con las técnicas Wavelet, Contourlet, Variación Total, WVTV y GNDF, muestran que WVTV y GNDF presentan una reducción significativa en el ruido presente en el fondo de las imágenes 2DGE. Esta comparación se soporta con el cálculo del gradiente morfológico, en el cual se puede notar que WVTV y GNDF conservan un mayor nivel de contraste, lo cual se espera facilite la detección, especialmente en el caso de proteínas tenues. Al realizar una comparación cuantativa de las técnicas de filtrado usando una imagen 2DGE real con sus proteínas identificadas, se encontró que el mejor nivel de detección se logra filtrando la

imagen con GNDF. La técnica de filtrado GNDF, además de lograr una adecuada reducción del ruido, presenta una conservación de los contornos, lo que facilita la detección de las proteínas.

4.2. Corrección de fondo

4.2.1. Comparación de técnicas usando imágenes sintéticas

Se realizó la comparación de las tres técnicas de corrección de fondo descritas en la sección 3.2.3: estimación del fondo a partir del histograma, histograma modificado y aproximación por superficie empleando las imágenes sintéticas presentadas en las Figuras 4-22-a y 4-23-a. Para generar las imágenes sintéticas se adicionó una anomalía que simula el cambio de intensidad de iluminación. Por medio de un cambio creciente en el eje x de hasta un 155 % del valor inicial del fondo y decreciente de hasta un 25 % se generó el fondo no homogéneo de las Figuras 4-22-a y 4-23-a respectivamente. Esta anomalía se presenta debido al instrumento de adquisición de imagen, especialmente en escáneres. Para las técnicas por histogramas se utilizó un percentil 60, ya que de acuerdo a las pruebas realizadas, en las imágenes 2-DGE el fondo equivale máximo a un 70 % de la imagen. Para el caso de la aproximación de fondo como superficie se realizó 150 pasos para ajustar los parámetros de una curva B- Spline. La comparación cuantitativa se realiza calculando el índice IS dado por la ecuación 3-4.

La Figura 4-22 presenta los resultados de corrección de fondo para la primera imagen sintética. Usando el histogramas se logra identificar el fondo de manera parcial. Se puede notar Figura 4-22-b que este método divide el fondo en dos regiones. Por otro lado, la técnica de histogramas modificados logra un fondo homogéneo con una sola intensidad de gris (Figura 4-22-c). En cambio, la técnica de aproximación por superficie realiza una sustracción del fondo, resaltando los contornos de la imagen (4-22-c)). La Tabla 4-5 presenta el indicador IS calculado para cada una de las técnicas de corrección de fondo. Usando únicamente el histograma se logra identificar apenas 71.8 % del fondo. Para la técnica de histogramas modificados se se obtiene una identificación del fondo del 98.5 % y para la aproximación por superficie del 97.9 %.

Indicador IS [%]	
Histogramas	71.8
Histogramas Modificados	98.5
Superficie	97.9

Tabla 4-5: IS para la corrección de fondo de una imagen sintética

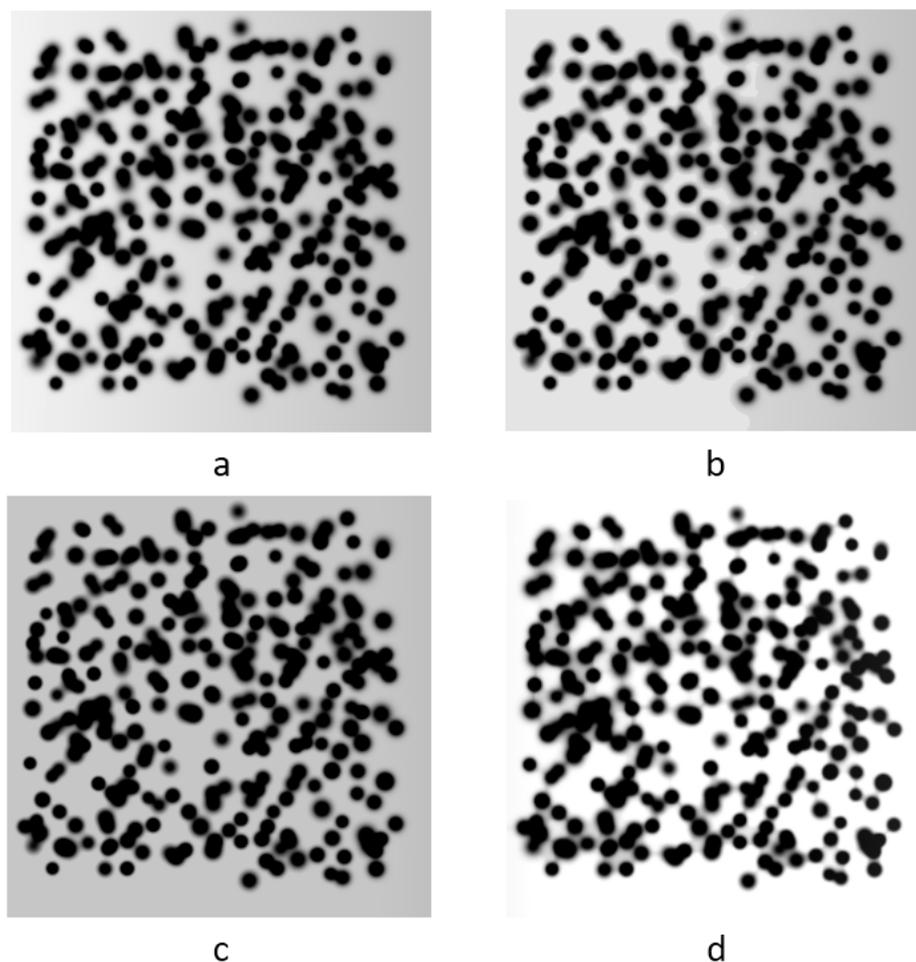


Figura 4-22: Corrección de fondo imagen sintética. a) imagen de control, b) Histogramas c) Histogramas Modificados d) Superficie

La Figura 4-23 presenta los resultados de corrección de fondo para la segunda imagen sintética. Para este caso, nuevamente se puede apreciar que la técnica del histograma no caracteriza adecuadamente el fondo dividiéndolo en dos regiones (Figura 4-23-b). Por su parte, las técnicas de histogramas modificados y aproximación por superficie presentan una mejor identificación del fondo. Sin embargo, la aproximación por superficie mantiene el fondo en algunas partes cercanas a proteínas, lo cual puede causar un agrupamiento de las mismas que puede dificultar la detección. La Tabla 4-6 presenta el índice IS para la segunda imagen sintética. La técnica del histograma solo logra identificar el 68.8% del fondo. Por su parte, las técnicas de histogramas modificados y aproximación por superficie logran una identificación del 97.8% y del 96.4% del fondo.

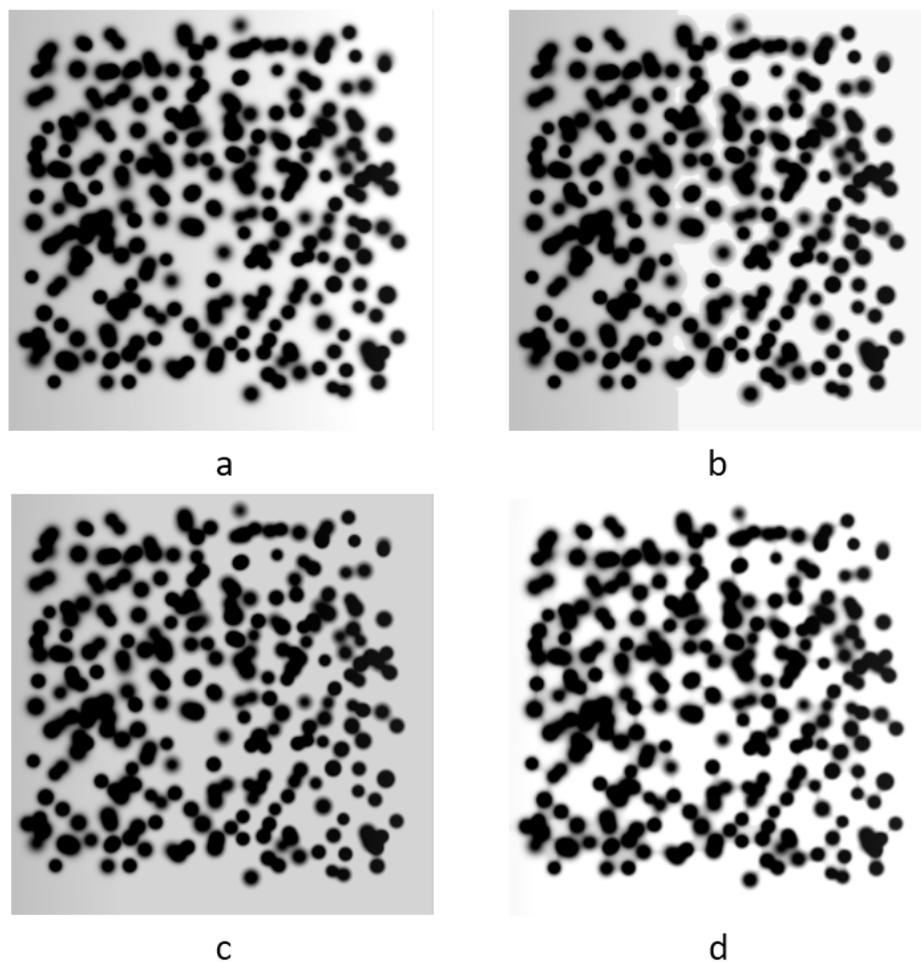


Figura 4-23: Resultados corrección de fondo imagen sintética 2. a) imagen de control, b) Histogramas c) Histogramas Modificados d) Superficie.

Indicador IS [%]	
Histogramas	68.8
Histogramas Modificados	97.8
Superficie	96.4

Tabla 4-6: *IS* para la corrección de fondo de una imagen sintética

Discusión

Debido a la naturaleza de las imágenes 2DGE, particularmente en las anomalías relacionadas con el fondo, como lo es el cambio de intensidad y el cambio de contraste alrededor de las proteínas, la técnicas de histogramas no debe ser usada. Esta técnica no tiene la capacidad para identificar fondos altamente variables como los presentes en imágenes 2DGE. De acuerdo a los resultados

obtenidos con la corrección de fondo en las imágenes sintéticas, las técnicas de histogramas modificados y aproximación por superficie presenta los mejores resultados de acuerdo al porcentaje de identificación del fondo. La técnica de histogramas modificado es la más idónea, no solamente por su ligera superioridad, sino también, por la conservación de los contrastes adecuadamente como se observa en la Figura 4-23-c. La técnica de aproximación por superficie mantiene el fondo alrededor de algunas proteínas, lo cual puede causar que se realice una identificación errada de las proteínas.

4.2.2. Comparación de técnicas usando imágenes reales

Con el objetivo de validar los resultados obtenidos con las imágenes sintéticas, se emplea nuevamente las imágenes 2DGE de las Figura 2-1 b y c. Para los experimentos con las imágenes 2DGE reales se conservaron los parámetros de las técnicas de filtrado usados en los experimentos con datos sintéticos.

La Figura 4-24 presenta los resultados de corrección de fondo para la imagen 2-1-b. Empleando la técnica de histogramas se observa que no logra reducir el fondo alrededor de las proteínas, lo cual sería perjudicial para el proceso de detección automático de proteínas (Figura 4-24-b). Por su parte, las técnicas de histogramas modificados y aproximación por superficie logran homogeneizar en gran parte el fondo. Sin embargo, la aproximación por superficie resalta más el contraste de las proteínas al disminuir significativamente la intensidad del fondo.

La Figura 4-25 presenta los resultados de corrección de fondo para la imagen 2-1-c. Esta imagen presenta un gran nivel de distorsión en la parte superior, asociado a diferentes anomalías como líneas verticales y horizontales, ruido general, degradación lumínica, entre otros. Por esta distorsión, los resultados obtenidos con las técnicas de corrección de fondo no son óptimos. En la Figura 4-25-b se observa que la técnica de de corrección de fondo basada en el histograma divide el fondo en varias secciones. La técnica de histogramas modificados presenta una corrección parcial que solo identifica como fondo la parte media y baja de la imagen (4-25-c). Por su parte, la técnica de aproximación por superficie realizó una ligera corrección de fondo en la parte superior de la imagen y completa en la parte media e inferior de la imagen.

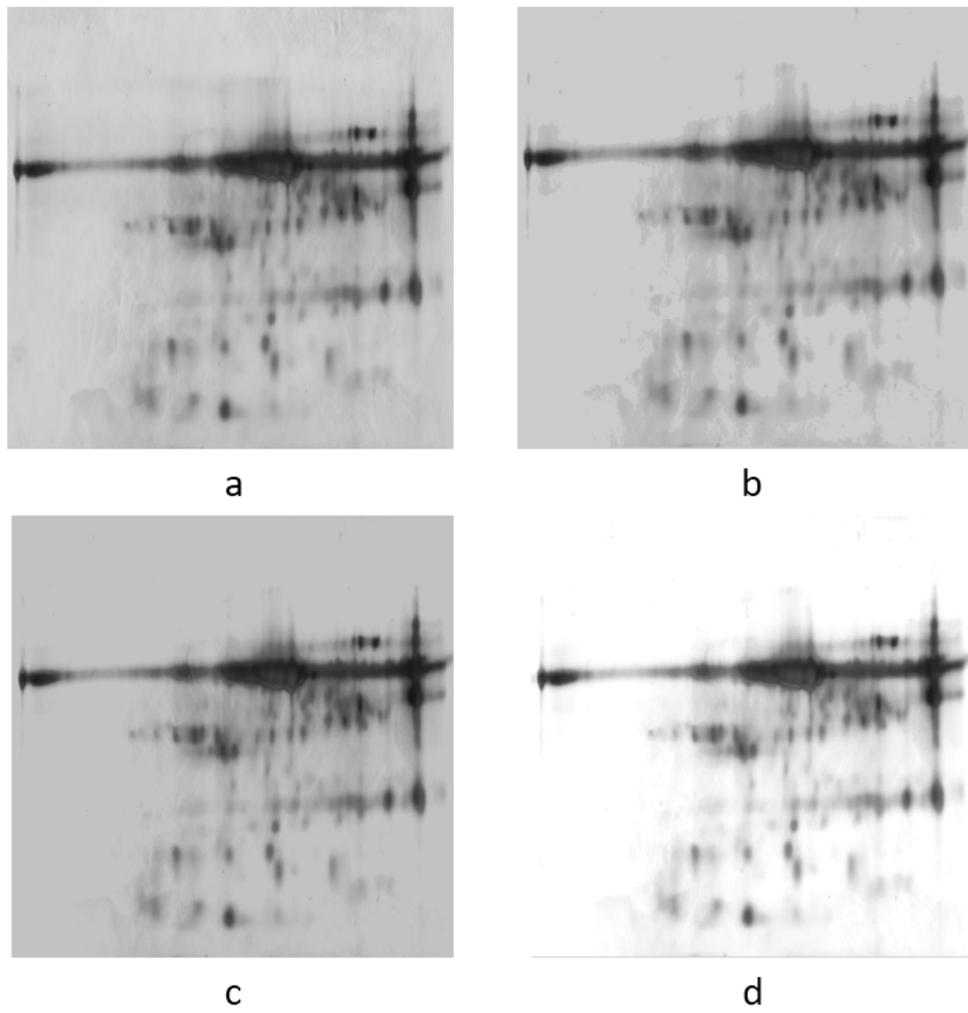


Figura 4-24: Corrección de fondo imagen real. a) imagen de control, b) Histogramas c) Histogramas Modificados d) Superficie.

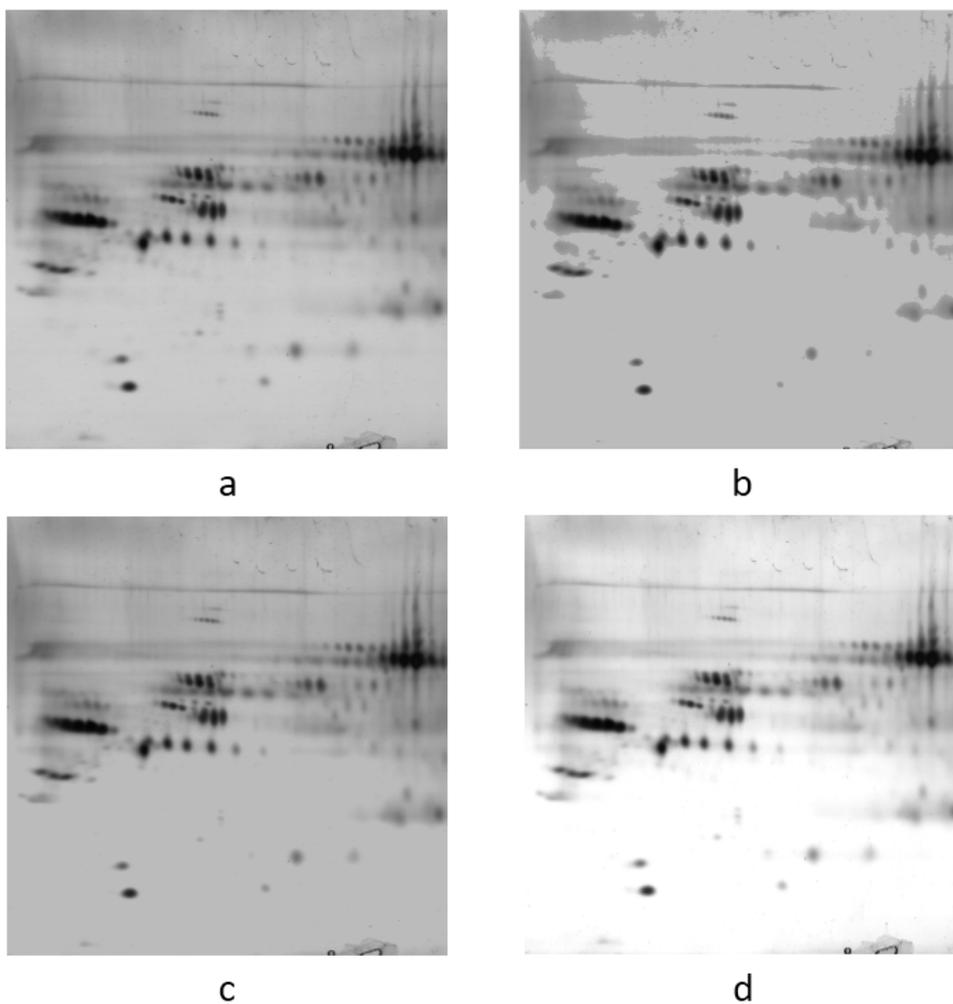


Figura 4-25: Corrección de fondo imagen real. a) imagen de control, b) Histogramas c) Histogramas Modificados d) Superficie.

Discusión

De acuerdo a los resultados obtenidos en el estudio, la técnica de histogramas para corrección de fondo no demuestra un potencial para este campo de aplicación, debido a la irregularidad del fondo presente en este tipo de imágenes. La técnicas de corrección de fondo de histogramas modificados y de aproximación por superficie, de manera cualitativa, se pueden recomendar para su aplicación en imágenes 2DGE. Sin embargo, debido a la baja calidad de las imágenes en relación señal a ruido, se recomienda realizar un filtrado o suavizado en las imágenes 2DGE antes de aplicar alguna técnica de corrección de fondo.

4.3. Normalización de imagen

4.3.1. Comparación de técnicas usando imágenes sintéticas

Dado que la regularización de imágenes busca mejorar la apariencia de puntos tenues y saturados, se generó una imagen sintética con este tipo de puntos para la comparación de las técnicas encontradas en la literatura. La nueva imagen de control tiene 1024x1024 píxeles, con un fondo opaco y 150 proteínas de baja densidad. La desviación estándar para la distribución Gaussiana se encuentra en el intervalo de [0.3 , 0.8]. Para el valor máximo de intensidad por proteína se usó valores en el intervalo [0.1 , 0.8]. La Figura 4-26-a muestra la imagen sintética generada. Se realizó la comparación entre las técnicas de equalización del histograma, ecualización adaptativa y por histogramas.

Los resultados presentados en la Figura 4-26-c muestran que la técnica de regularización basada en el histograma no logra mejorar la apariencia de los puntos tenues en comparación a las demás proteínas, aunque elimina el fondo opaco. Por su parte, las técnicas basadas en la ecualización logran resaltar las proteínas tenues.

Para evaluar de forma cuantitativa estos resultados se emplea el indicador *DPT* dado por la ecuación 3-5. Los resultados obtenidos con *DPT* se presentan en la Tabla 4-7). La técnica basada en el histograma solo permite identificar 48.7% de las proteínas tenues. Por su parte, la ecualización del histograma y la ecualización adaptativa logran identificar 82.1% y 88.9% de las proteínas tenues.

Indicador <i>DPT</i> [%]	
Ecualización	82.1
Histogramas	48.7
Ecualización adaptativa	88.9

Tabla 4-7: *DPT* para la normalización de imágenes sintéticas.

Discusión

Bajo las condiciones evaluadas, la técnica de histogramas presentó un desempeño inferior al 50% para la normalización de la imagen. En cambio, las técnicas basadas en la ecualización del histograma permite identificar más del 80% de las proteínas. Sin embargo, de acuerdo al indicador *DPT* basado en el número de proteínas tenues detectadas, la ecualización adaptativa resulta más efectiva para la regularización de imágenes 2DGE sintéticas.

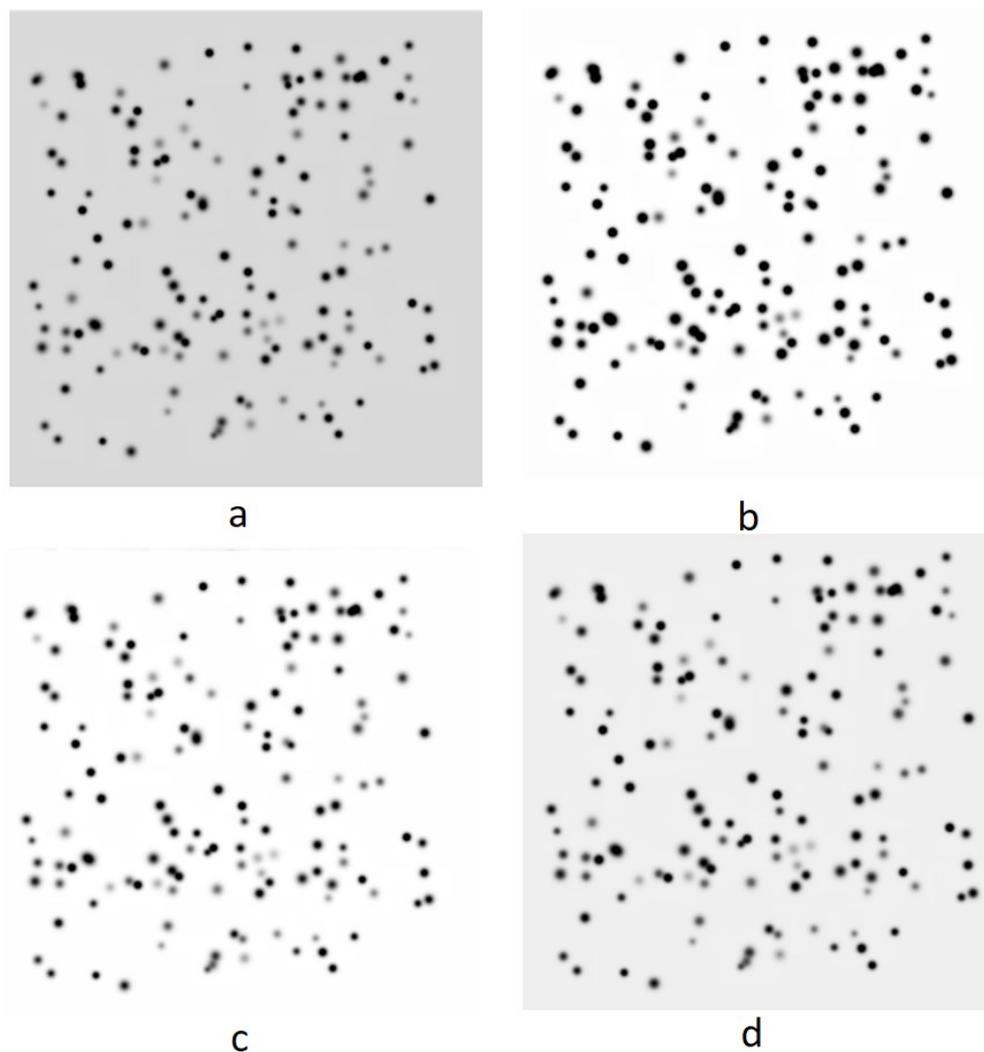


Figura 4-26: Normalización de imágenes sintéticas. a) Imagen de control, b) Ecuación c) Histogramas d) Ecuación adaptativa.

4.3.2. Comparación de técnicas usando imágenes reales

Para las imágenes 2DGE de las Figura 2-1 b y c se compararon también las tres técnicas de normalización del experimento anterior. La Figura 4-27 presenta los resultados de regularización para la imagen 2-1-b. A partir de estos resultados se debe notar que las tres técnicas mejoran el contraste de la imagen. La ecualización y la ecualización adaptativa incrementan el nivel de intensidad de las regiones donde se encuentran las proteínas. Sin embargo, estas técnicas también obtienen un contraste muy marcado del ruido, lo cual podría favorecer la detección de proteínas falsas.

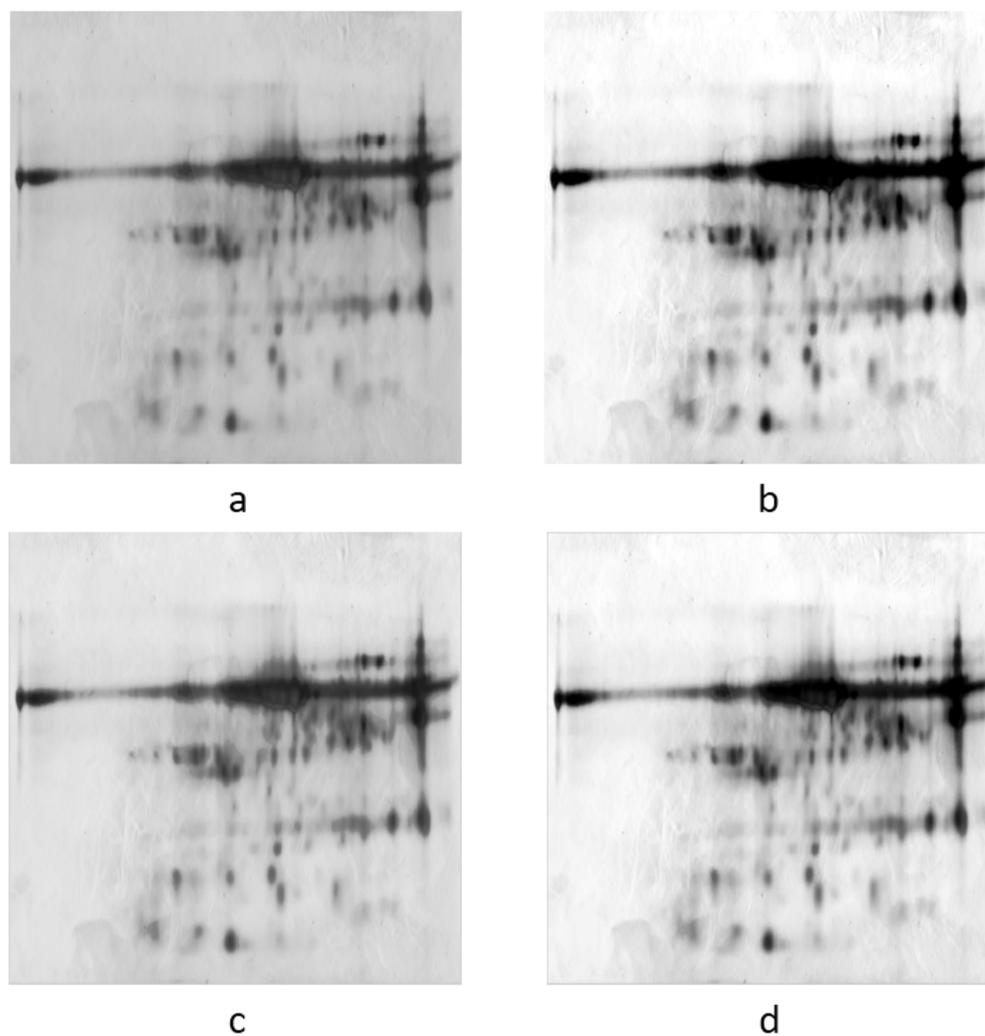


Figura 4-27: Regularización de Imágenes Reales. a) Imagen de control, b) Ecuación c) Histogramas d) Ecuación adaptativa

La Figura 4-28 presenta los resultados de regularización para la imagen 2-1-c. En este caso es evidente como los métodos basados en la ecuación, a la vez que incrementan la intensidad de proteínas tenues, también resaltan el ruido del fondo no uniforme. Por su parte, la técnica del histograma no logra mejorar el contraste de las proteínas tenues.

La técnica de ecuación adaptativa presentó un desempeño notorio para ambas imágenes reales, demostrando una menor sensibilidad en los ruidos ocasionados por el sistema de adquisición y favoreciendo el nivel de contraste de objetos de interés. Como se puede ver en la Figura 4-28-d, la ecuación adaptativa mitiga los efectos no deseados causados por la saturación de las proteínas, sin generar distorsión significativa en el fondo de la imagen.

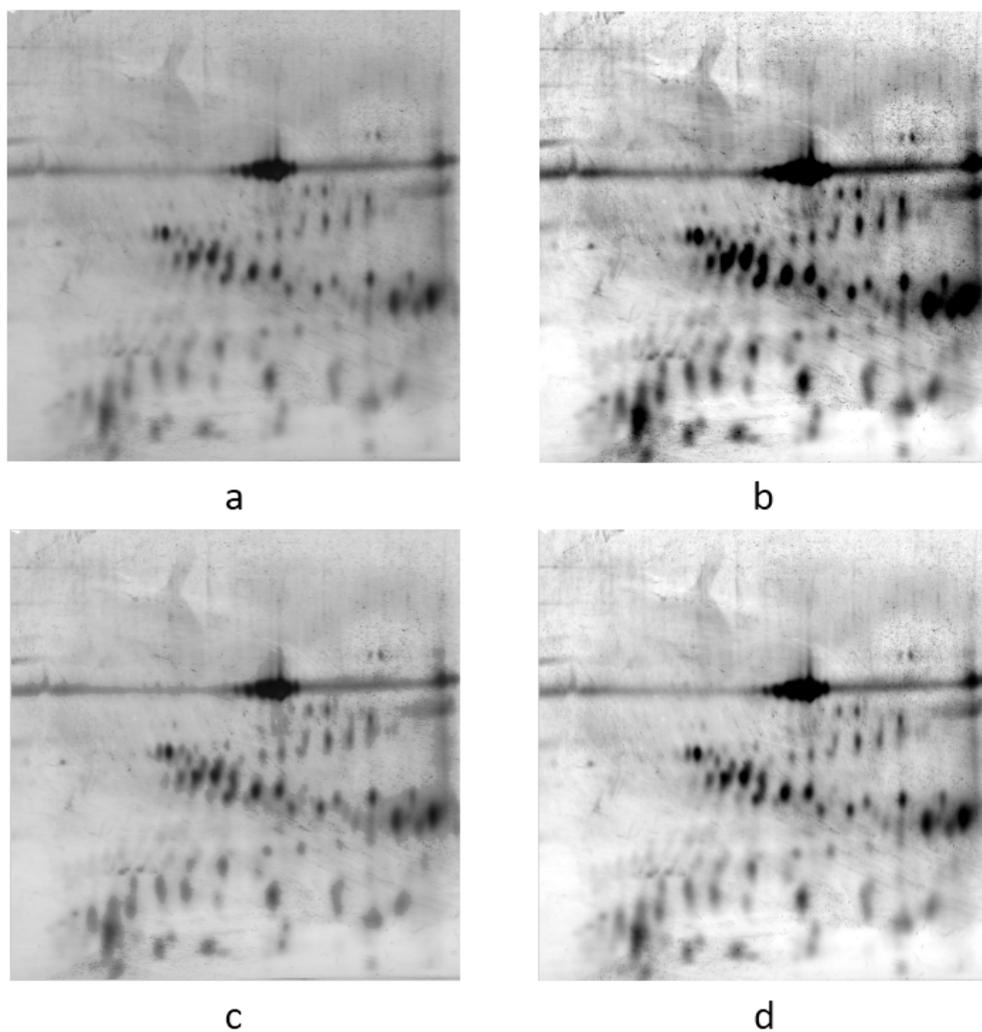


Figura 4-28: Resultados Regularización de Imágenes. a) Imagen de control, b) Ecuación c) Histogramas d) Ecuación adaptativa

Discusión

En general, las técnicas de pre-procesamiento de normalización de imágenes resaltan el contraste de las proteínas, lo cual favorece la detección de los objetos presentes en la imagen. Sin embargo, intensifican igualmente anomalías como los ruidos puntuales y del fondo. Debido a la naturaleza de las imágenes 2DGE, la técnica más apropiada es la ecualización adaptativa, al presentar la menor sensibilidad a los ruidos, mejores resultados cuantitativos y menor distorsión la imagen.

4.4. Metodología para integrar técnicas de pre-procesamiento

De acuerdo a los resultados obtenidos en la comparación de las técnicas de pre-procesamiento para imágenes 2DGE presentado en la secciones 4.1 a 4.3, se evidenció la necesidad de una metodología multinivel, que permita la reducción de diferentes tipos de anomalías. Por esta razón se propuso la metodología descrita en la Figura 3-8. El primer paso consistió en la normalización de la imagen; se busca en esta etapa mejorar el contraste de las proteínas tenues, para que no se vean afectadas por los pasos siguientes del procesamiento. El segundo paso se realizó la reducción del ruido a través de una técnica no lineal que permitió la preservación de los bordes. Por último, se realizó la corrección del fondo.

Para realizar una combinación efectiva, se debe seleccionar las técnicas a implementar en cada etapa. De acuerdo a los resultados previos, se seleccionó la técnica de ecualización por histogramas adaptativa para la primera etapa de normalización de imágenes. La ecualización del histograma adaptativa, resaltó el contraste entre las proteínas y presentó los mejores resultados cuantitativos con base a las imágenes sintéticas. En la segunda etapa, se seleccionó la técnica de filtrado no lineal GNDF. Esta técnica presentó un alto desempeño en la comparación cuantitativa, además de una clara tendencia a conservar bordes y mitigar ruidos puntuales, ruidos que se esperan sean resaltados por la etapa de normalización. Finalmente, se seleccionó para la etapa de corrección de fondo la técnica de histogramas modificados. Esta técnica obtuvo fondos homogéneos que se esperan ayuden a la detección de proteínas. Los parámetros empleados para cada una de las etapas fueron los mismo que se usaron para la comparación de las técnicas con imágenes sintéticas y reales.

La evaluación de la metodología integrada de pre-procesamiento de imágenes 2DGE se realizó usando imágenes sintéticas y reales. Para lo cual, se generó una imagen sintética que incluyera ruido, fondo no homogéneo y proteínas tenues. La imagen generada se presenta en la Figura 4-29. La imagen tiene 1024x1024 píxeles, 150 proteínas tenues y un fondo que varía de intensidad a lo largo del eje x . La desviación estándar para la distribución Gaussiana que modela las proteínas se seleccionó en el intervalo $[0.1, 0.8]$. El valor máximo de intensidad de cada proteína se encuentra en el intervalo de $[0.3, 0.8]$. Posteriormente, se adicionó un ruido Gaussiano con media cero y desviación estándar de 1.535 y un ruido tipo Rayleigh con parámetros $a = 0$ y $b = 0.0539$.

Para validar el comportamiento de la metodología integrada de pre-procesamiento, se comparó el resultado obtenido al aplicar las técnicas de manera individual respecto aplicarlas de manera conjunta. En la Figura 4-30 se muestran los resultados al aplicar las tres técnicas de pre-procesamiento de forma individual y la metodología propuesta. En la Figura 4-30-d se evidencia como la metodología integrada reduce el efecto de las diferentes anomalías, obteniendo una imagen con fondo uniforme, sin ruido y con mejor contraste entre las proteínas.

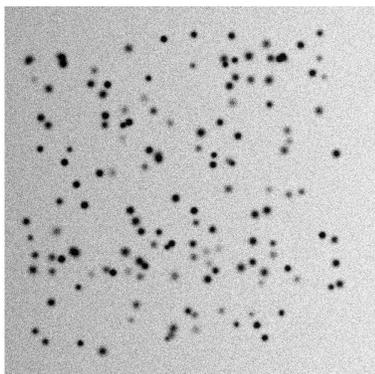
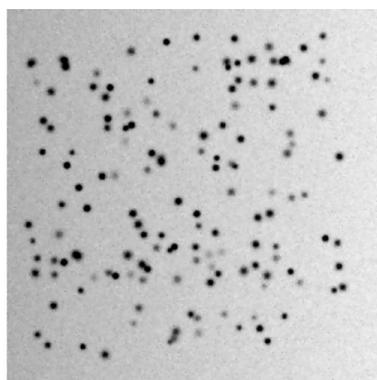
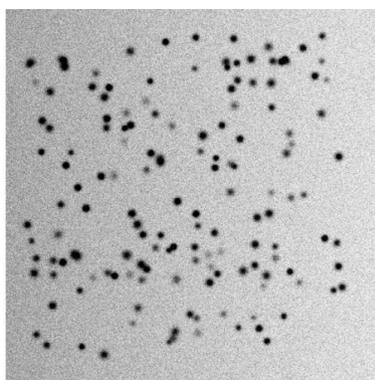


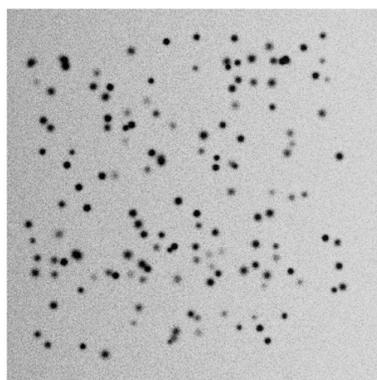
Figura 4-29: Imagen sintética con anomalías de ruido, fondo irregular y presencia de proteínas tenues.



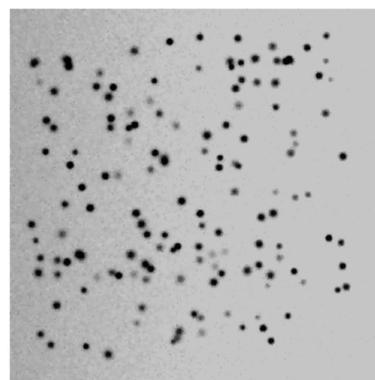
a



b



c



d

Figura 4-30: Comparación de técnicas de pre-procesamiento. Resultado obtenido para la imagen sintética usando a) reducción de ruido, b) regularización de imagen, c) corrección de fondo y d) metodología integrada propuesta.

Para los resultados obtenidos desde cada una de las técnicas de pre-procesamiento, así como la

metodología propuesta, se calcularon los indicadores de desempeño. La Tabla 4-8 presenta el indicador SE obtenido desde la detección de proteínas sobre la imagen original, la imagen filtrada con GNDF, la imagen normalizada con ecualización adaptativa, la imagen corregida por histogramas modificados, y la imagen obtenida por la combinación de las tres etapas. Como se puede observar en la Tabla 4-8, la metodología propuesta es la única que obtiene un porcentaje de SE mayor al 60%. Adicionalmente, la Tabla 4-8 presenta el indicador de identificación del fondo, IS , para la imagen corregida por histogramas modificados y la metodología propuesta. Para este indicador, la metodología propuesta también obtiene un desempeño superior, identificando hasta el 78% del fondo en la imagen. Por último, la Tabla 4-8 presenta el indicador relacionado con las proteínas tenues, DPT . Se presenta este indicador para cada una de las técnicas de pre-procesamiento aplicadas y la metodología propuesta. Con la técnica de regularización basada en la ecualización adaptativa solo se identifica el 40% de las proteínas tenues. Por su parte, la metodología propuesta permite identificar el 60% del fondo.

	SE [%]	IS [%]	DPT [%]
Imagen Original	5.8742	N.A	0
GNDF	17.6923	N.A	3.3333
Ecualización adaptativa	3.5749	N.A	40
Histogramas Modificados	6.6987	11.3658	0
Metodología propuesta	63.8462	78.6232	60

Tabla 4-8: Indicadores de desempeño

Adicionalmente, se realizó la validación de la metodología integrada de pre-procesamiento con diferentes imágenes reales. En la Figura 4-31 se presenta el comportamiento de la metodología propuesta para una imagen 2DGE de apitoxina. La Figura 4-31-a presenta la imagen original. La Figura 4-31-b presenta el resultado al realizar la regularización de la imagen usando la ecualización adaptativa. En esta imagen nuevamente se ve el efecto de la regularización, mejorando el contraste de las proteínas, al mismo tiempo que el ruido del fondo se resalta. La Figura 4-31-c presenta el resultado de la etapa de filtrado usando GNDF. En esta imagen se nota como el filtro no lineal reduce el ruido del fondo, así como ruidos puntuales que se encontraban en la parte superior de la imagen. Por último, Figura 4-31-d presenta el resultado de metodología completa, al aplicar la corrección de fondo por histograma modificado a la imagen filtrada previamente. Comparando analíticamente las imágenes en las Figuras 4-31-a y -d es claro el efecto de la metodología propuesta, mejorando el fondo de la imagen, así como el contraste de las proteínas débiles.

La Figura 4-32 presenta el resultado obtenido con otra imagen 2DGE. Esta imagen se caracteriza por un fondo poco homogéneo, con grandes regiones de fondo con nivel de intensidad similares. En la imagen 4-32-a se resaltan los artefactos en el fondo. La 4-32-b presenta el resultado de la metodología integrada propuesta. En esta imagen se nota como los artefactos del fondo son eliminados o mitigados de manera significativa. También es evidente la mejora en el contraste, con un mínimo

de pérdidas de proteínas tenues. Los niveles de contraste obtenidos resaltan las características estructurales de la imagen lo cual permitiría una mejor detección de proteínas.

La Figura 4-33 presenta un experimento adicional con una imagen real 2DGE. En esta figura se presenta cada una de las etapas de pre-procesamiento. El resultado final de la metodología se muestra en la Figura 4-33-d. En este caso, también se evidencia un mayor nivel de contraste entre las proteínas y el fondo, indicando un beneficio inmediato para la detección de proteínas. También se disminuye de manera significativa los artefactos presentes en la imagen como ruidos causados por las suciedades de las lentes y de las anomalías identificadas como líneas verticales y horizontales, que disminuirían la eficiencia en el procesamiento de la imagen.

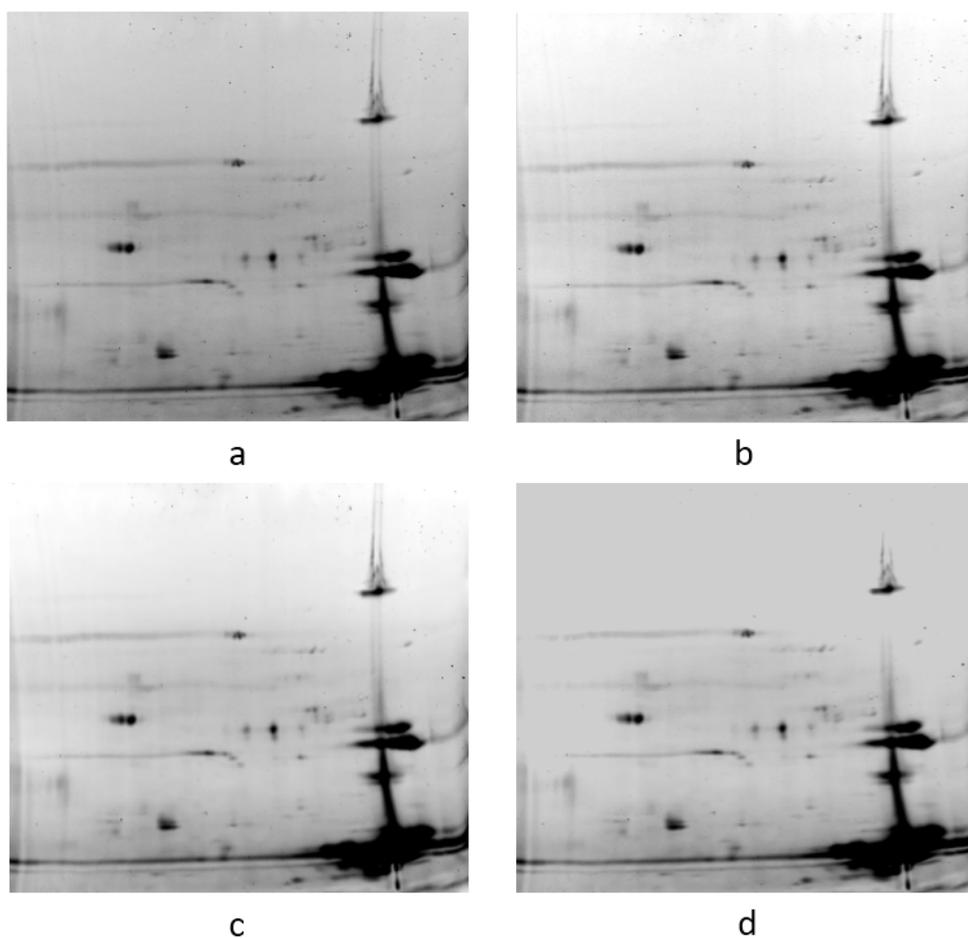


Figura 4-31: Resultados pre-procesamiento apitoxina. a) Imagen de control y resultados de cada etapa de la metodología propuesta: b) regularización c) reducción de ruido y d) corrección de fondo.

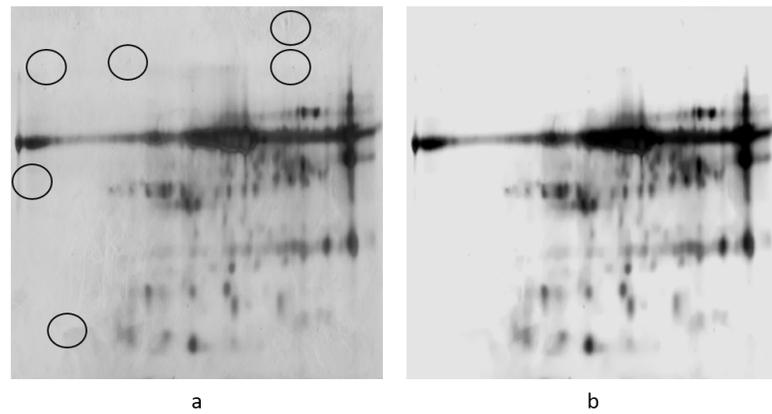


Figura 4-32: Resultados pre-procesamiento a) Imagen de control, b) Metodología propuesta.

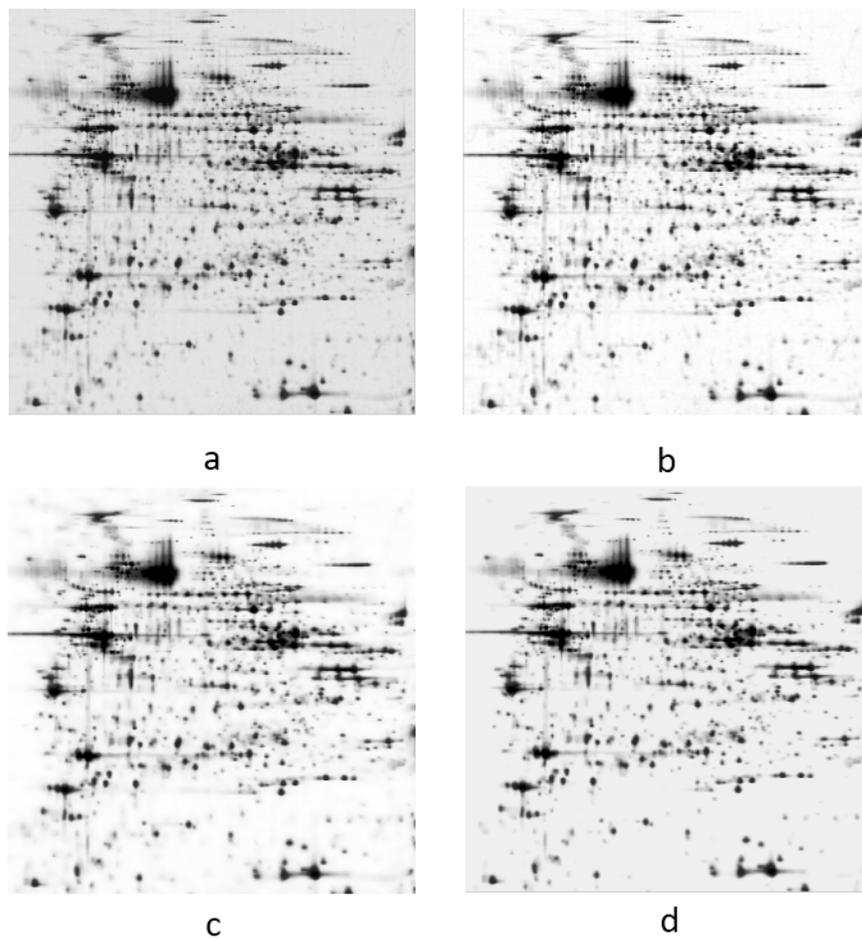


Figura 4-33: Resultados pre-procesamiento. a) Imagen de control, b) regularización c) reducción de ruido y d) corrección de fondo.

4.5. Validación de resultados por expertos en imágenes 2DGE

Para el proceso de validación de resultados se contó con el apoyo de la profesora Sarah Röthlisberger Booth y el Grupo de Investigación e Innovación Biomédica (GI2B). Para esta validación se analizaron los resultados de la metodología de pre-procesamiento en muestras de apitoxinas obtenidas en el proyecto *Caracterización del veneno de la abeja africanizada mediante análisis proteómico y procesamiento digital de imágenes orientado a la exploración de sus propiedades (P14227)*. Ejemplos de este análisis se presenta en las Figuras 4-34, 4-35, 4-36 donde se resaltan las proteínas en color rojo. En cada figura se presenta la muestra original y el resultado del pre-procesamiento. Se puede observar una disminución significativa del ruido generado por la tinción. También se hace evidente la dificultad para discriminar las proteínas que se encuentran en el clúster (recuadro amarillo). Para este ejemplo no se presentó diferencia en el número de proteínas.

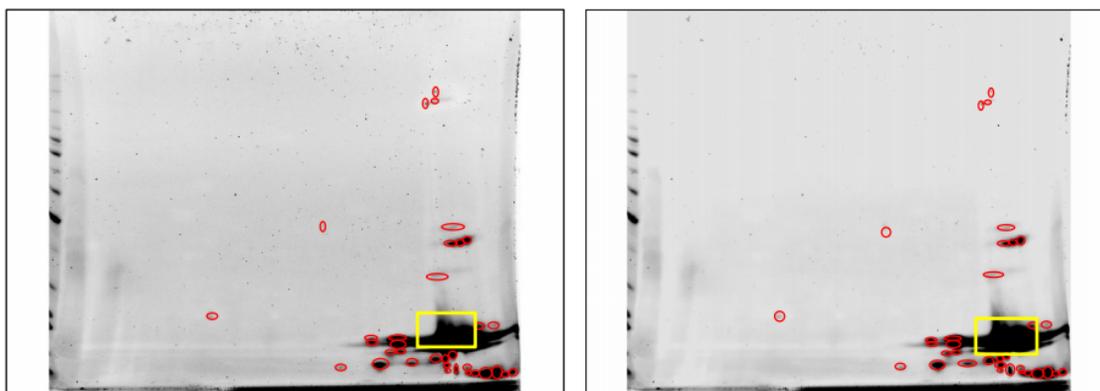


Figura 4-34: Muestra 1: original (lado izquierdo) y resultado del pre-procesamiento (lado derecho).

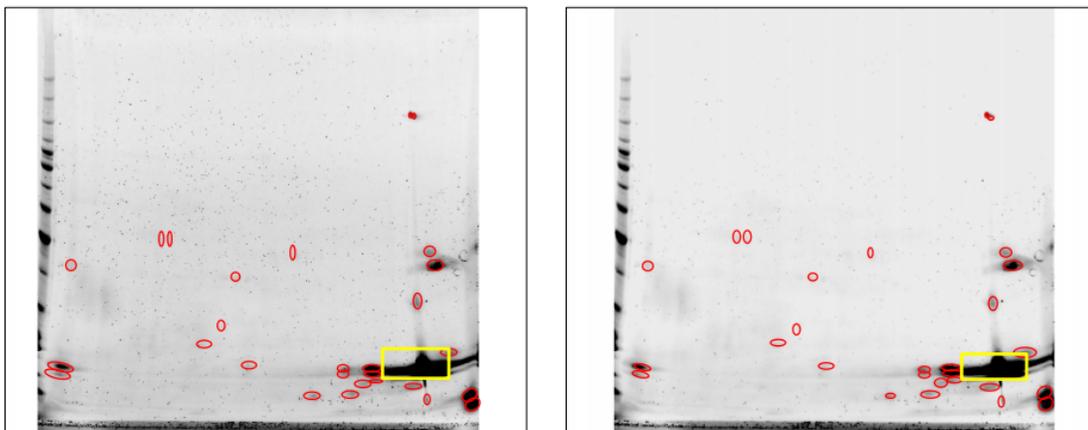


Figura 4-35: Muestra 2: original (lado izquierdo) y resultado del pre-procesamiento (lado derecho).

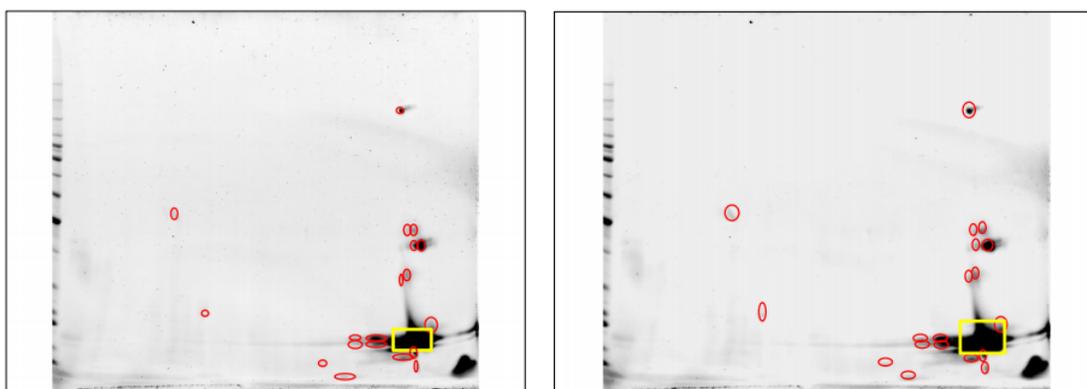


Figura 4-36: Muestra 5: original (lado izquierdo) y resultado del pre-procesamiento (lado derecho).

En la Figura 4-37 se muestra en mismo análisis realizado con la muestra 4. En este caso, se logra una mejor detección a partir de la imagen obtenida después del pre-procesamiento. Por ejemplo, se logra disminuir las falsas proteína como la resaltada en el recuadro morado. En la Figura 4-38 se presenta los resultados para la muestra 6. Se observa un comportamiento similar, logrando identificar 5 proteínas adicionales respecto a la imagen original, enmarcadas en el recuadro morado, las cuales fueron catalogadas como proteínas por poseer las propiedades espaciales presentes en las proteónas caracterizadas.

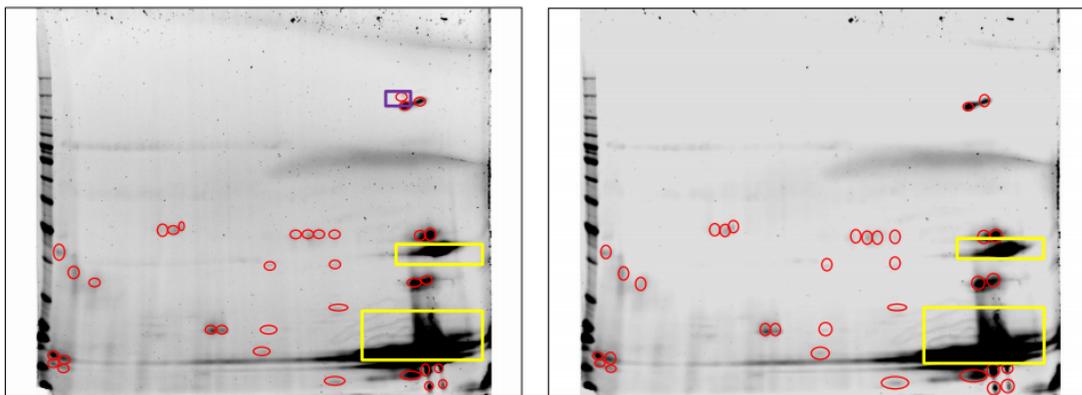


Figura 4-37: Muestra 4: original (lado izquierdo) y resultado del pre-procesamiento (lado derecho).

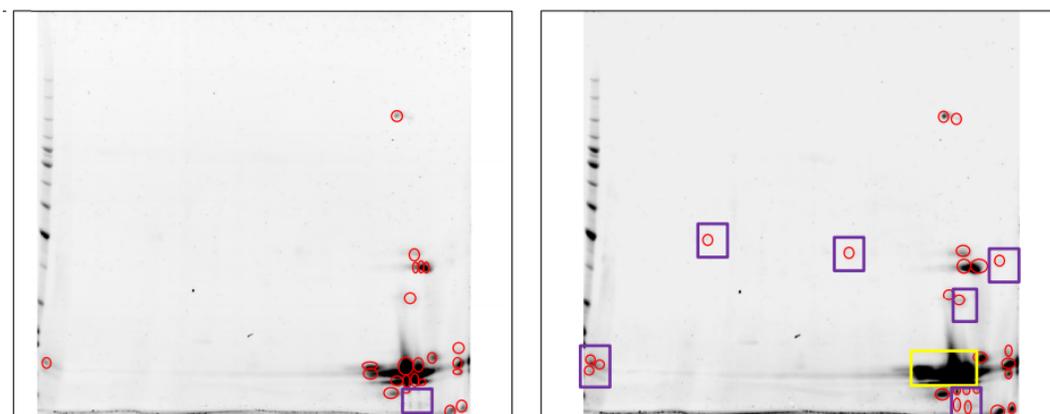


Figura 4-38: Muestra 6: original (lado izquierdo) y resultado del pre-procesamiento (lado derecho).

En el caso de la muestra 3, presentada en la Figura 4-39, la metodología planteada presenta una atenuación de una proteína de baja densidad, resaltada en el cuadro naranja. En la Tabla 4-9 se presentan de manera cuantitativa los resultados obtenidos, donde podemos observar que el número de proteínas reales se preserva o mejora con la aplicación de la metodología propuesta de pre-procesamiento y existe una reducción significativa de las proteínas falsas detectadas.

Muestra	# Spot detectados	# Reales detectados	# Spot Falsos	Reducción de ruido
Muestra 1 Original	1894	34	1860	72 %
Muestra 1 Pre-procesada	549	34	515	
Muestra 2 Original	1822	28	1794	43 %
Muestra 2 Pre-procesada	1043	28	1015	
Muestra 3 Original	1239	27	1212	49 %
Muestra 3 Pre-procesada	646	28	618	
Muestra 4 Original	1698	33	1664	65 %
Muestra 4 Pre-procesada	611	34	578	
Muestra 5 Original	797	20	777	40 %
Muestra 5 Pre-procesada	487	20	467	
Muestra 6 Original	541	21	520	12 %
Muestra 6 Pre-procesada	486	26	460	

Tabla 4-9: Resultados cuantitativos

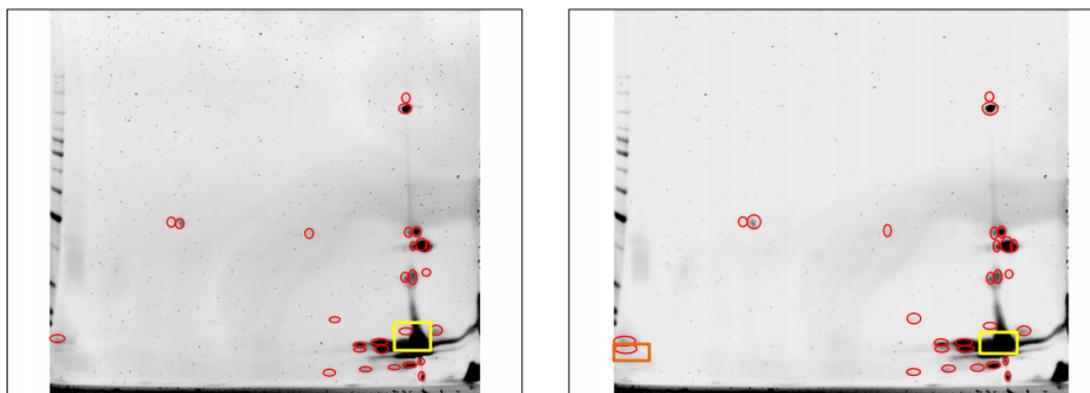


Figura 4-39: Muestra 3: original (lado izquierdo) y resultado del pre-procesamiento (lado derecho).

5 Conclusiones y recomendaciones

5.1. Conclusiones

Las imágenes 2DGE presentan varias anomalías que dificultan la detección y análisis de proteínas. Ejemplos de estas anomalías son el ruido, fondo no homogéneo, proteínas tenues o saturadas, rayas verticales u horizontales. En la literatura se establece la necesidad de un procesamiento automático de imágenes 2DGE, que requiera la mínima intervención humana y permita obtener resultados confiables y reproducibles. Se encuentran tres corrientes para el pre-procesamiento de imágenes 2DGE: reducción de ruido, normalización de imagen y corrección de fondo. A pesar de la gran diversidad de técnicas para cada una de estas corrientes, no se encuentra una evaluación cuantitativa que permita medir su impacto en la detección de proteínas, como tampoco, se encuentra una metodología que combine las tres corrientes para reducir el efecto de diferentes tipos de anomalías al mismo tiempo. Por lo cual, este trabajo se centró en realizar una comparación cuantitativa de las técnicas de pre-procesamiento más comunes definiendo una metodología de análisis de desempeño, y posteriormente, proponiendo una metodología que combine las diferentes corrientes de pre-procesamiento para lograr una reducción significativa del efecto de las anomalías en la detección de proteínas en imágenes 2DGE.

A partir de la revisión de literatura se encontró que: las técnicas más empleadas en la reducción de ruido en imágenes 2DGE son las técnicas no lineales basadas en Wavelet, Contourlet, Variación Total y una técnica que combina Variación Total con Wavelet. Adicionalmente, se estudió el potencial de usar un filtro de difusión no lineal geométrico (GNDF), dado su desempeño para reducir ruido y preservar bordes en otro tipo de imágenes biomédicas. Para la corrección de fondo se encontró que entre las técnicas más empleadas esta la estimación de fondo usando el histograma, el histograma modificado y aproximación del fondo por superficie. Adicionalmente, como técnicas de normalización se encontró que la ecualización del histograma, la ecualización adaptativa y el método basado en el histograma son las más comunes para imágenes 2DGE.

Para realizar la comparación de las técnicas de pre-procesamiento encontradas en la literatura, fue necesario definir una metodología de evaluación basada en una etapa cuantitativa y una etapa cualitativa. Para la evaluación cuantitativa se generó una base de datos de imágenes 2DGE sintéticas, con diferentes tipos de anomalías como ruido, fondo no homogéneo, y proteínas tenues. Para medir el desempeño de las técnicas se emplearon cuatro indicadores. La relación señal-ruido (SNR) y el Spot Efficiency (SE) permitieron medir el desempeño de las técnicas de reducción de ruido. Por su parte, para evaluar las técnicas de corrección de fondo se usó un índice de sustracción (IS) para medir el porcentaje de fondo identificado. Por último, se usó el porcentaje de proteínas tenues

detectadas (*DPT*) para evaluar las técnicas de normalización de imagen. La evaluación cualitativa se basó en el uso de las diferentes técnicas en imágenes 2DGE reales, evaluando analíticamente los resultados.

Desde la comparación de técnicas de reducción de ruido se demostró que Wavelet y GNDF presentaron el mejor desempeño para los diferentes tipos y niveles de ruido con las imágenes sintéticas. Sin embargo, GNDF presentó los mejores resultados con las imágenes reales, logrando una adecuada reducción de ruido, al mismo tiempo que presenta una conservación de los bordes. Para la corrección de fondo, la comparación tanto con imágenes sintéticas como reales, mostró que la técnica basada en histogramas modificados es la más idónea, principalmente por la conservación de los contrastes. Por último, la ecualización del histograma presentó el mejor resultado para la normalización de imágenes 2DGE, resaltando las proteínas tenues, al mismo tiempo que intensifica anomalías como ruidos puntuales y del fondo.

A partir de los resultados compartivos de técnicas de pre-procesamiento se propuso la metodología para combinar las tres corrientes. Como primera etapa se propone la normalización de la imagen usando la ecualización del histograma. La ecualización mejora el contraste de las proteínas, pero también resalta ruidos puntuales. Como segunda etapa se selecciono la técnica de filtrado no lineal GNDF, permitiendo la reducción de ruido al mismo tiempo que se preserva la información de los bordes. Como última etapa, se selecciono la corrección de fondo por histogramas modificados, lo cual permite obtener fondos homogéneos que ayudan a la detección de proteínas. Con la metodología propuesta se logró una mitigación eficiente de las anomalías consideradas, de acuerdo a los indicadores propuestos y la perspectiva del experto en el tratamiento de imágenes 2DGE.

5.2. Recomendaciones

Este trabajo sirve como punto de partida para mejorar la calidad de imágenes 2DGE de entrada en procesos de detección e identificación de proteínas. Esta metodología permite mitigar las anomalías inherentes de este tipo de imágenes. Está claro que la metodología propuesta no excluye a otras técnicas de pre-procesamiento. Como trabajo futuro se propone integrar la metodología de pre-procesamiento desarrollada en este trabajo como un paso inicial en el procesamiento y análisis de imágenes 2DGE.

Una vez obtenida una metodología eficaz de pre-procesamiento de imágenes 2DGE, el siguiente paso a seguir es empaquetar el algoritmo obtenido en un aplicativo de software, que permitirá el fácil acceso a esta metodología de pre-procesamiento de imágenes a los especialistas de dicho tema, permitiéndoles minimizar los tiempos de procesamientos de imágenes, a través de una interfaz sencilla y fácil de manejar, no solo facilitando el trabajo de caracterización, sino también teniendo la posibilidad de obtener resultados acertados significativos.

De acuerdo al panorama descrito en este trabajo y la bibliografía actualizada utilizada como respal-

do del mismo, queda abierto un amplio trabajo futuro no solo en la etapa de pre-procesamiento, si no también, en las etapas de procesamiento y análisis de imágenes 2-DGE, que permitan optimizar los tiempos de los especialistas y aumentar la eficacia, más aún teniendo en cuenta la justificación de la importancia de este trabajo y sus implicaciones en el campo de la salud.

En el desarrollo de este trabajo no se abordó de manera satisfactoria la mitigación de la anomalía descrita como líneas verticales, ya que ninguna de las técnicas contempladas dentro de las ramas de pre-procesamiento para imágenes 2-DGE permite realizar dicha mitigación, no obstante actualmente existen algoritmos de procesamiento que permiten mitigar anomalías de características similares en otros campos de acción, por lo tanto existe una clara oportunidad para incluir este tipo de etapas en el esquema integral de pre-procesamiento.

Es necesario realizar una caracterización extensiva de las características de las anomalías presentes, con el fin de determinar que posibles técnicas de pre-procesamiento pueden tener un rol importante en el pre-procesamiento de imágenes 2DGE, como por ejemplo los tipos de distribución e intensidades del ruido esperados en este tipo de imágenes.

Aún queda abierta la discusión sobre la metodología más apropiada para el pre-procesamiento de imágenes 2DGE, ya que si bien queda establecida la necesidad sobre un pre-procesamiento multinivel, se pueden explorar diferentes estrategias para resolver las anomalías presentes, como el uso de redes neuronales y algoritmos evolutivos, que permitan determinar pesos para cada una de las técnicas de acuerdo con las características particulares de estas imágenes.

5.3. Discusión académica de los resultados

A partir de esta tesis de maestría se han realizado las siguientes publicaciones:

- Manuel M. Goez, Maria C. Torres-Madronero, Sarah Rothlisberger-Booth, Edilson Delgado-Trejos. A comparative study of noise reduction techniques in preprocessing of two-dimensional gel electrophoresis image. *VIII Seminario Internacional de Ingeniería Biomédica*. Bogotá, Colombia, 2016.
- Maria C. Torres-Madronero, Manuel M. Goez, Gerardo Villegas, Sarah Rothlisberger-Booth, Edilson Delgado-Trejos. Procesamiento de imágenes de electroforesis bidimensional para análisis proteómico. *Coloquios de Investigación en Ingeniería Biomédica: Primera edición*. Medellín, Colombia, 2016.

A la fecha de la presentación de este documento, se encuentran en elaboración las siguientes publicaciones:

- Manuel M. Goez, Maria C. Torres-Madronero, Sarah Rothlisberger-Booth, Edilson Delgado-Trejos. Pre-processing of 2-dimensional gel electrophoresis image applied to proteomic analysis: A review. Sometido a *Computer methods and programs in biomedicine*. Noviembre (2016).

- Manuel M. Goetz, Maria C. Torres-Madronero, Sarah Rothlisberger-Booth, Edilson Delgado-Trejos. Two-dimensional gel electrophoresis image pre-processing using geometric nonlinear diffusion.

Bibliografía

- [1] E. M. Faergestad, M. Rye, B. Walczak, L. Gidskehaug, J. P. Wold, H. Grove, X. Jia, K. Hollung, U. G. Indahl, F. Westad, F. Van Den Berg, and H. Martens, "Pixel-based analysis of multiple images for the identification of changes: A novel approach applied to unravel proteome patterns of 2-D electrophoresis gel images," *Proteomics*, vol. 7, no. 19, pp. 3450–3461, 2007.
- [2] R. S. Sengar and A. Kumar, "Segmentation of Two Dimensional Electrophoresis Gel Image Using the Wavelet Transform and the Watershed Transform," *IEEE Communications (NCC), 2012 National Conference*, pp. 0–4, 2012.
- [3] 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.
- [4] World Health Organization, "Global status report on non communicable diseases 2010," tech. rep., 2011.
- [5] World Health Organization, "Global atlas on cardiovascular disease prevention and control," tech. rep., 2011.
- [6] V. Kulasingam and E. P. Diamandis, "Strategies for discovering novel cancer biomarkers through utilization of emerging technologies," *Nature clinical practice. Oncology*, vol. 5, no. 10, pp. 588–599, 2008.
- [7] J. Yuan, L. Zhu, X. Liu, T. Li, Y. Zhang, T. Ying, B. Wang, J. Wang, H. Dong, E. Feng, *et al.*, "A proteome reference map and proteomic analysis of *Bifidobacterium longum* ncc2705," *Molecular & Cellular Proteomics*, vol. 5, no. 6, pp. 1105–1118, 2006.
- [8] ruta n medellín, "Plan de ciencia tecnología e innovación de medellín 2010. [Online]. Available: http://rutanmedellin.org/images/programas/plan_cti/Documentos/Plan_de_CTi_de_Medellin.pdf,"
- [9] P. M. Nugues, "Two-dimensional electrophoresis image interpretation," *IEEE Transactions on biomedical engineering*, vol. 40, no. 8, pp. 760–770, 1993.
- [10] R. Westermeier, *Electrophoresis in Practice: A Guide to Theory and Practice*. VCH, Weinheim, 1993.
- [11] V. C. Wasinger, S. J. Cordwell, a. Cerpa-Poljak, J. X. Yan, a. a. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith, "Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*," *Electrophoresis*, vol. 16, no. 7, pp. 1090–1094, 1995.

- [12] T. Aittokallio, J. Salmi, T. Nyman, and O. Nevalainen, "Geometrical distortions in two-dimensional gels: applicable correction methods," *Journal of Chromatography B*, vol. 815, no. 1, pp. 25–37, 2005.
- [13] M. Rye and E. M. Fargestad, "Preprocessing of electrophoretic images in 2-DE analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 117, pp. 70–79, 2012.
- [14] S. Beranova-Giorgianni, "Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: Strengths and limitations," *TrAC - Trends in Analytical Chemistry*, vol. 22, no. 5, pp. 273–281, 2003.
- [15] R. Frank, R. Hargreaves, "Clinical biomarkers in drug discovery and development," *Nature Reviews*, no. 2, pp. 566–580, 2003.
- [16] S. Mazzara, A. Conti, S. Olivieri, S. Iannaccone, M. Alessio, and L. Pattini, "Evaluation of Two-Dimensional Gel Electrophoresis Maps by Local Tangent Space Alignment : an Application to Neuroproteomics," *IEEE International Conference BioInformatics and BioEngineering*, vol. 1, no. 5, pp. 8–10, 2008.
- [17] 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, 2008.
- [18] M. Savelonas, D. Maroulis, and E. Mylona, "Segmentation of two-dimensional gel electrophoresis images containing overlapping spots," *Final Program and Abstract Book - 9th International Conference on Information Technology and Applications in Biomedicine, ITAB 2009*, no. November, pp. 5–7, 2009.
- [19] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.
- [20] A. W. Dowsey, M. J. Dunn, and G.-Z. Yang, "The role of bioinformatics in two-dimensional gel electrophoresis," *Proteomics*, vol. 3, pp. 1567–96, Aug. 2003.
- [21] R. L. Easton, "Fundamentals of Digital Image Processing," no. November, 2010.
- [22] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, vol. 14. second edi ed., 2002.
- [23] H. J. Issaq *et al.*, "The role of separation science in proteomics research," *Electrophoresis*, vol. 22, no. 17, p. 3629, 2001.
- [24] A. Görg, W. Weiss, and M. J. Dunn, "Current two-dimensional electrophoresis technology for proteomics," *Proteomics*, vol. 4, no. 12, pp. 3665–3685, 2004.
- [25] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, no. 6928, pp. 198–207, 2003.

- [26] P. Tsakanikas and E. S. Manolakos, "Improving 2-DE gel image denoising using contourlets," *Proteomics*, vol. 9, no. 15, pp. 3877–3888, 2009.
- [27] C. V. Cannistraci, F. M. Montevecchi, and M. Alessio, "Median-modified Wiener filter provides efficient denoising, preserving spot edge and morphology in 2-DE image processing," *Proteomics*, vol. 9, no. 21, pp. 4908–4919, 2009.
- [28] T. Acharya and A. K. Ray, *Image Processing: Principles and Applications*. John Wiley & Sons, 2005.
- [29] 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, 2003.
- [30] B. Munsell and R. Barsanti, "Electrophoresis 2D gel residual encoding using adaptive wavelets for image segmentation," *Proceedings. IEEE SoutheastCon, 2005.*, no. 1, 2005.
- [31] M. Daszykowski, M. S. Wróbel, a. Bierzynska-Krzysik, J. Silberring, G. Lubec, and B. Walczak, "Automatic preprocessing of electrophoretic images," *Chemometrics and Intelligent Laboratory Systems*, vol. 97, no. 2, pp. 132–140, 2009.
- [32] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics, 1991.
- [33] M. Daszykowski, I. Stanimirova, a. Bodzon-Kulakowska, J. Silberring, G. Lubec, and B. Walczak, "Start-to-end processing of two-dimensional gel electrophoretic images," *Journal of Chromatography A*, vol. 1158, no. 1-2, pp. 306–317, 2007.
- [34] R. D. Appel, J. R. Vargas, P. M. Palagi, D. Walther, and D. F. Hochstrasser, "Melanie II—a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms," *Electrophoresis*, vol. 18, no. 15, pp. 2735–2748, 1997.
- [35] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communication and Image Representation*, vol. 1, no. 1, pp. 21–46, 1990.
- [36] P. Tsakanikas and I. Manolakos, "Effective Denoising of 2D Gel Proteomics Images using Contourlets," in *2007 IEEE International Conference on Image Processing*, vol. 6, pp. VI – 269–VI – 272, IEEE, 2007.
- [37] T. F. Chan, S. Osher, and J. Shen, "The digital tv filter and nonlinear denoising," *IEEE Transactions on Image processing*, vol. 10, no. 2, pp. 231–241, 2001.
- [38] Mallat Stephane G., "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.
- [39] H. Xin and F. Zhao, "Effective denoising methods for two-dimensional gel electrophoresis images," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, vol. 3, pp. 1571–1574, IEEE, Oct. 2011.

- [40] M. T. Faheem, S. Rashwan, A. Sarhan, and B. A. Youssef, "De-noising 2D gel images using Wavelet transform," *Proceedings of the 10th WSEAS international conference on Applied computer and applied computational science*, no. 1, pp. 187–194, 2011.
- [41] K. Bartušek, J. Přinosil, and Z. Smékal, "Wavelet-based de-noising techniques in mri," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. 480–488, 2011.
- [42] S. Kockanat and N. Karaboga, "The design approaches of two-dimensional digital filters based on metaheuristic optimization algorithms: a review of the literature," *Artificial Intelligence Review*, vol. 44, no. 2, pp. 265–287, 2015.
- [43] M. Rogers and J. Graham, "Robust and accurate registration of 2-D electrophoresis gels using point-matching," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 624–635, 2007.
- [44] J. Chang, H. Van Remmen, W. F. Ward, F. E. Regnier, A. Richardson, and J. Cornell, "Processing of data generated by 2-dimensional gel electrophoresis for statistical analysis: Missing data, normalization, and statistics," *Journal of Proteome Research*, vol. 3, no. 6, pp. 1210–1218, 2004.
- [45] M. Barlaud, P. Sole, T. Gaidon, M. Antonini, and P. Mathieu, "Pyramidal lattice vector quantization for multiscale image coding," *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 367–381, 1994.
- [46] J. Bossinger, M. Miller, K. Vo, E. Geiduschek, and N.-H. Xuong, *Quantitative analysis of two-dimensional electrophoretograms.*, vol. 254. ASBMB, 1979.
- [47] 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, 2008.
- [48] S. Rashwan, T. Faheem, A. Sarhan, and B. A. B. Youssef, "A Fuzzy-Watershed Based Algorithm for Protein Spot Detection in 2DGE images," vol. 10, no. 5, pp. 254–260, 2010.
- [49] P. H. Eilers, "Parametric time warping," *Analytical chemistry*, vol. 76, no. 2, pp. 404–411, 2004.
- [50] K. Kaczmarek, B. Walczak, S. De Jong, and G. Vandeginste, "Baseline reduction in two dimensional gel electrophoresis images," *Acta Chromatographica*, pp. 82–96, 2005.
- [51] N. OTSU, "Threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 1, no. 9, pp. 62–66, 1979.
- [52] R. Millionsi, L. Puricelli, S. Sbrignadello, E. Iori, E. Murphy, and P. Tessari, "Operator- and software-related post-experimental variability and source of error in 2-DE analysis," *Amino Acids*, vol. 42, no. 5, pp. 1583–1590, 2012.

- [53] M. Unlü, M. E. Morgan, and J. S. Minden, "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts.," *Electrophoresis*, vol. 18, no. 11, pp. 2071–2077, 1997.
- [54] M. Berth, F. M. Moser, M. Kolbe, and J. Bernhardt, "The state of the art in the analysis of two-dimensional gel electrophoresis images," *Applied Microbiology and Biotechnology*, vol. 76, no. 6, pp. 1223–1243, 2007.
- [55] G. L. Hortin and D. Sviridov, "The dynamic range problem in the analysis of the plasma proteome," *Journal of Proteomics*, vol. 73, no. 3, pp. 629–636, 2010.
- [56] C. M. Smales, J. R. Birch, A. J. Racher, C. T. Marshall, and D. C. James, "Evaluation of individual protein errors in silver-stained two-dimensional gels," *Biochemical and Biophysical Research Communications*, vol. 306, no. 4, pp. 1050–1055, 2003.
- [57] J. Tranquillo, *MATLAB for Engineering and the Life Sciences*. Morgan Claypool, 2011.
- [58] M. Natale, A. Caiazzo, E. M. Bucci, and E. Ficarra, "A Novel Gaussian Extrapolation Approach for 2D Gel Electrophoresis Saturated Protein Spots," *Genomics, Proteomics and Bioinformatics*, vol. 10, no. 6, pp. 336–344, 2012.
- [59] Z. L. F. Y. H. B. L. X. Z. T. Li, R., "Functional analysis of the honeybee (*apis mellifera* l.) salivary system using proteomics," *Biochemical and Biophysical Research Communications*, no. 397, pp. 740–744., 2010.
- [60] E. J. Stollnitz, A. D. DeRose, and D. H. Salesin, "Wavelets for computer graphics: a primer. 1," *IEEE Computer Graphics and Applications*, vol. 15, no. 3, pp. 76–84, 1995.
- [61] 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 Transactions on Antennas and Propagation*, vol. 62, pp. 5157–5170, Oct. 2014.
- [62] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 9, pp. 1532–46, Jan. 2000.
- [63] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, "Image Selective Smoothing and Edge Detection by Nonlinear Diffusion," *SIAM Journal on Numerical Analysis*, vol. 29, pp. 182–193, Feb. 1992.
- [64] E. Michel-González, M. H. Cho, and S. Y. Lee, "Geometric nonlinear diffusion filter and its application to X-ray imaging.," *Biomedical engineering online*, vol. 10, p. 47, Jan. 2011.
- [65] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data.," *IEEE transactions on medical imaging*, vol. 11, pp. 221–32, Jan. 1992.

-
- [66] S. Sarkar and S. Das, "Multilevel image thresholding based on 2D histogram and maximum Tsallis entropy—a differential evolution approach.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, pp. 4788–97, Dec. 2013.
- [67] J. Huang, B. Wang, W. Wang, and P. Sen, "A Surface Approximation Method for Image and Video Correspondences.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 24, pp. 5100–13, Dec. 2015.
- [68] B. Gupta, M. Tiwari, and M. Shrivastava, "High-speed quantile-based histogram equalisation for brightness preservation and contrast enhancement," *IET Image Processing*, vol. 9, pp. 80–89, Jan. 2015.
- [69] C. Zuo, Q. Chen, and X. Sui, "Range Limited Bi-Histogram Equalization for image contrast enhancement," *Optik - International Journal for Light and Electron Optics*, vol. 124, pp. 425–431, Mar. 2013.
- [70] Y. Wang, H. Shen, Y. Liang, X. Lu, and Z. Ling, "Adaptive extended piecewise histogram equalisation for dark image enhancement," *IET Image Processing*, vol. 9, pp. 1012–1019, Nov. 2015.