

Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica



Estimación de posición, cantidad y ancho de bandas en carriles de imágenes de electroforesis en gel utilizando espacios de escala

Informe de Proyecto de Graduación para optar por el título de
Ingeniero en Electrónica con el grado académico de Licenciatura

Edison Fernández Alvarado

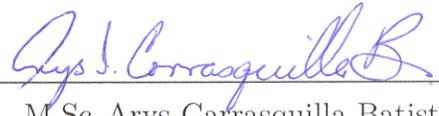
Versión de 30 de noviembre de 2010

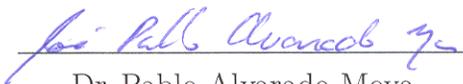
Instituto Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Proyecto de Graduación
Tribunal Evaluador

Proyecto de Graduación defendido ante el presente Tribunal Evaluador como requisito para optar por el título de Ingeniero en Electrónica con el grado académico de Licenciatura, del Instituto Tecnológico de Costa Rica.

Miembros del Tribunal


M.Sc. Eduardo Interiano Salguero
Profesor Lector


M.Sc. Arys Carrasquilla Batista
Profesora Lectora


Dr. Pablo Alvarado Moya
Profesor Asesor

Los miembros de este Tribunal dan fe de que el presente trabajo de graduación ha sido aprobado y cumple con las normas establecidas por la Escuela de Ingeniería Electrónica.

Declaro que el presente Proyecto de Graduación ha sido realizado enteramente por mi persona, utilizando y aplicando literatura referente al tema e introduciendo conocimientos propios.

En los casos en que he utilizado bibliografía he procedido a indicar las fuentes mediante las respectivas citas bibliográficas. En consecuencia, asumo la responsabilidad total por el trabajo de graduación realizado y por el contenido del correspondiente informe final.



Edison Fernández Alvarado

Cartago, 30 de noviembre de 2010

Céd: 3-0422-0672

Resumen

El presente proyecto está ubicado en el campo del análisis de imágenes digitales y tiene como meta el contribuir a la tarea de análisis de las imágenes obtenidas mediante la técnica de geles de electroforesis utilizada por biólogos moleculares.

Generalmente estas imágenes, debido a múltiples factores externos, presentan una calidad que dificulta su análisis por parte del experto. Es por ello que surge la necesidad de implementación de una herramienta de software que automatice esta labor para apoyar al científico a la hora de interpretar sus resultados. Actualmente se tienen avances en la implementación de ésta y el presente proyecto contribuye en su desarrollo con la implementación de un algoritmo para la estimación del ancho de las bandas y su cantidad presentes en un carril de una imagen de geles de electroforesis haciendo uso de la información obtenida mediante la aplicación de la técnica de espacios de múltiples escalas.

Palabras clave: electroforesis en geles, espacio de escalas, detección de bandas.

Abstract

This project is located in the field of digital image analysis and aims to contribute to the task of analyzing the resulting images on the technique of electrophoresis gel used by molecular biologists.

Usually the images resulting from this technique, due to many external factors, have a poor quality which makes difficult its analysis by the expert. That is why there is a need to implement a software tool that automates this task support the scientist in the interpretation of his results.

Currently there are advances in the development of this tool and this project aims to contribute to its development with the implementation of an algorithm that automates the estimation of the width, position and amount of bands present in a lane of an electrophoresis gel image using the information obtained by applying the scale-space technique.

Keywords: electrophoresis gel, scale-space, band detection.

a mi familia

Agradecimientos

Primero que nada a Dios por haberme dado las fuerzas y haberme permitido llegar hasta este momento.

A mis padres por su incondicional apoyo aún en los momentos más difíciles sin el cual nada de esto hubiera sido posible. A mis hermanos por estar siempre conmigo.

A mis compañeros por todo el apoyo brindado y al Dr. Pablo Alvarado Moya por su invaluable ayuda.

Edison Fernández Alvarado

Cartago, 30 de noviembre de 2010

Índice general

Índice de figuras	iii
Índice de tablas	v
1 Introducción	1
1.1 Electroforesis en gel	1
1.2 Imágenes de Geles de Electroforesis	2
1.3 Estrategia propuesta de detección de bandas	4
1.4 Objetivos y estructura del documento	5
2 Marco Teórico	7
2.1 Imágenes en escala de grises y perfil de intensidades	7
2.2 Filtrado de imágenes	8
2.3 Generación del espacio de escalas	11
3 Sistema de detección de bandas	15
3.1 Perfil de intensidades	15
3.2 Generación del espacio de escalas	16
3.3 Detección de blobs	19
3.3.1 Aplicación de la segunda derivada	19
3.3.2 Generación del nuevo espacio de escalas	20
3.4 Caso de una función gaussiana	24
4 Resultados y Análisis	29
4.1 Determinación de la desviación estándar de las bandas	29
4.2 Determinación de la cantidad de bandas presentes en el carril	32
4.3 Determinación de la posición de las bandas en el carril	35
4.4 Comparación con el algoritmo previo	37
5 Conclusiones y Recomendaciones	41
5.1 Conclusiones	41
5.2 Recomendaciones	41
Bibliografía	43

Índice de figuras

1.1	Proceso de electroforesis (Tomado de [5])	2
1.2	Imagen de un gel de electroforesis	2
1.3	Preparación de un gel de electroforesis	3
1.4	Captura de la imagen de un gel de electroforesis	3
1.5	Ejemplo de poca separación entre bandas	4
1.6	Bloques principales del módulo de procesamiento de imágenes (Tomado de [1])	4
1.7	Diagrama de bloques de la solución	5
2.1	Representación de píxeles en una imagen	7
2.2	Ejemplo de imagen en escala de grises	8
2.3	Perfil de intensidades de la imagen de la figura 2.2	8
2.4	Distribución normal o gaussiana	9
2.5	Distribución normal para diferentes varianzas	10
2.6	Representación de las funciones de la figura 2.5 en el dominio de la frecuencia	10
2.7	Representación de una función $f(x)$ en un espacio de escalas	11
2.8	Ejemplo de una imagen en varias escalas. Imagen original a , $t = 1$ b , $t = 8$ c , $t = 64$ d (Tomado de [8])	12
2.9	Imagen de ejemplo	13
2.10	Espacio de escalas de la imagen en la figura 2.9	13
3.1	Ejemplo de un carril de una imagen de un gel de electroforesis	15
3.2	Perfil de intensidades del carril mostrado en la figura 3.1	16
3.3	Ejemplo del proceso de generación de el espacio de escalas	17
3.4	Espacio de escalas para el perfil de intensidades de la figura 3.1	18
3.5	Segunda derivada de una función gaussiana	19
3.6	Segunda derivada del Espacio de Escalas	20
3.7	Localización del punto más alto de la banda	22
3.8	Ubicación de los máximos en el espacio de escalas	23
3.9	Ubicación de las bandas en el carril	23
3.10	Segunda derivada de una función gaussiana	26
3.11	Normalización de la ecuación 3.18 para valores de $n = 0$ (línea roja), $n = 1$ (línea verde), $n = 1, 5$ (línea morada), $n = 2$ (línea café) y $n = 3$ (línea azul) para un $\sigma_i = 1$	27
3.12	Ubicación del valor máximo de (3.17) normalizado por $\sigma_f^{1.5}$ para un $\sigma_i = 1$.	27

4.1	Ejemplo de dos bandas con poca separación	31
4.2	Comparación del error en la determinación de la desviación estándar para el caso de 1 y 20 bandas	32
4.3	Ejemplo de poca separación entre bandas	33
4.4	Ejemplo de no detección de bandas por poca separación	34
4.5	Espacio de escalas para la figura 4.4	34
4.6	Carril generado para la determinación la evaluación de l a determinación de la posición	35
4.7	Error en la determinación de la posición	36
4.8	Carril utilizado para el análisis de un caso real	38
4.9	Resultado de analizar un carril real sin filtrar	39
4.10	Resultado de analizar un carril real filtrado	39

Índice de tablas

4.1	Desviación estándar obtenida para carriles con 20 bandas	30
4.2	Medidas estadística para la determinación de la desviación estándar obtenida para carriles con 20 bandas	30
4.3	Desviación estándar obtenida para carriles con 1 banda	31
4.4	Medidas estadística para la determinación de la desviación estándar obtenida para carriles con 1 banda	31
4.5	Valores experimentales de la determinación de la cantidad de bandas en el carril	33
4.6	Valores teóricos de la determinación de la cantidad de bandas en el carril . .	33
4.7	Datos teóricos de la determinación de la posición (parte 1)	35
4.8	Datos teóricos de la determinación de la posición (parte 2)	35
4.9	Datos experimentales de la determinación de la posición (parte 1)	36
4.10	Datos experimentales de la determinación de la posición (parte 2)	36
4.11	Valores determinados por un observador humano de la determinación de la cantidad de bandas en el carril para la determinación del error en la detección	37
4.12	Valores experimentales de la determinación de la cantidad de bandas en el carril para la determinación del error en la detección	37
4.13	Error en la detección de bandas	37
4.14	Comparación de los resultados en la determinación de σ	38
4.15	Comparación de los resultados en la determinación de σ para el caso de 1 banda	38
4.16	Resultados del análisis de un carril real	39

Capítulo 1

Introducción

Si bien el proceso de análisis de las imágenes resultantes en la técnica de electroforesis ha tenido considerables avances hasta la actualidad la mayor parte de la tarea sigue quedando en manos del experto; es por eso que se hace necesaria la implementación de una herramienta de software que permita automatizar el análisis de dichas imágenes para que de esa forma el experto pueda concentrarse en las tareas concernientes a su especialidad sin tener que lidiar con imágenes de mala calidad o actividades rutinarias automatizables.

1.1 Electroforesis en gel

La electroforesis es el proceso mediante el cual se da el movimiento de partículas eléctricamente cargadas inmersas en un medio, gracias a la acción de un campo eléctrico. Este proceso es particularmente útil en la caracterización molecular (proceso mediante el cual se determinan atributos conspicuos de una molécula en particular [1]).

Con esta técnica se consigue la separación de moléculas (proteínas o ácidos nucleicos) según su movilidad en un campo eléctrico a través de una matriz porosa, la cual finalmente las separa por tamaños moleculares y carga eléctrica, dependiendo de la técnica utilizada. La velocidad con que las moléculas se desplazan está determinada por su relación carga-masa; así por ejemplo, si dos moléculas poseen la misma masa, aquella que tiene la mayor carga se desplazará más rápidamente hacia un electrodo[3]. En la figura 1.1 se observa un ejemplo de este proceso: las moléculas con carga positiva son atraídas hacia el terminal negativo y las negativas hacia el terminal positivo siendo las moléculas más pequeñas (las de menor masa) las que alcanzan un mayor desplazamiento; esto permite la caracterización de moléculas, ya que de los resultados obtenidos se puede extraer, por ejemplo, tamaños moleculares.

Si bien existen sistemas comerciales que automatizan algunas partes del análisis de las imágenes no se cuenta con un sistema que realice la totalidad de la tarea; en la figura 1.2 se presenta una imagen de un gel de electroforesis.

La identificación de las bandas se dificulta debido al ruido presente y al característico bajo contraste, ocasionando que se tenga que realizar un esfuerzo visual para su adecuada iden-



Figura 1.1: Proceso de electroforesis (Tomado de [5])

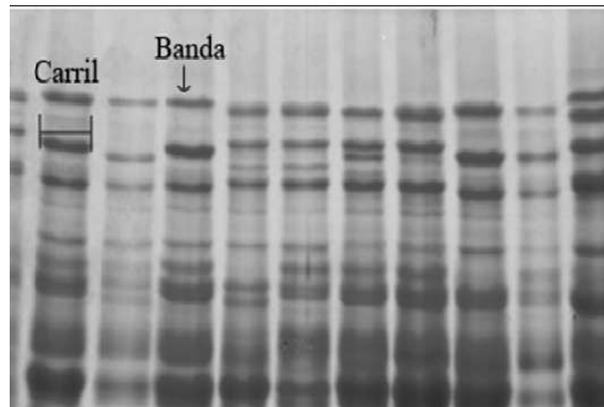


Figura 1.2: Imagen de un gel de electroforesis

tificación (la cual no siempre se puede lograr). Con el mencionado proyecto se persigue automatizar dicha tarea con el fin de reducir el error que podría darse por diversos factores como por ejemplo la fatiga del investigador. Lo que se pretende es brindar una herramienta computacional que permita el procesamiento de las imágenes de geles de electroforesis para así hacer más sencillo su análisis.

La complejidad de dicho procesamiento es alta ya que se tiene que lidiar con obstáculos como la eliminación del fondo (“background”) y el hecho de que las bandas no se encuentran completamente alineadas o demasiado juntas (aparentando una única banda) lo que dificulta su identificación. En proyectos anteriores [1, 9] se han logrado avances en dicha tarea y actualmente se está trabajando en la optimización de la detección de las bandas; sin embargo, para este caso se está dando por conocido el ancho de las mismas.

Como el fin del mencionado proyecto es la detección automática de las bandas se hace necesaria la investigación en métodos para determinar automáticamente el ancho y número de éstas en un carril.

1.2 Imágenes de Geles de Electroforesis

La Figura 1.3 ilustra el proceso de preparación de un gel de electroforesis. Este se realiza en una membrana porosa colocada en medio de dos cristales separados por espaciadores a los cuales se les coloca un peine con el fin de delimitar los distintos carriles.

En la Figura 1.4 se ilustra el método de captura de la imagen del gel.

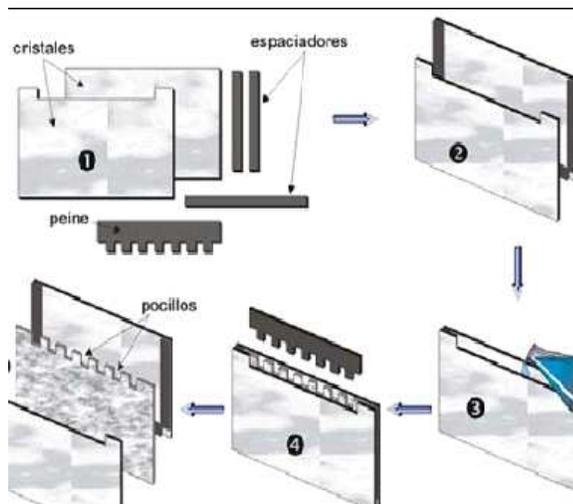


Figura 1.3: Preparación de un gel de electroforesis

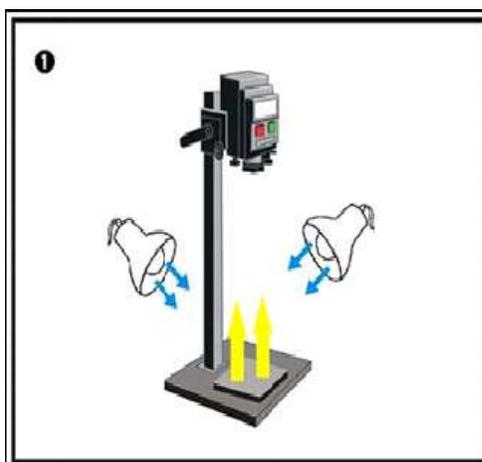


Figura 1.4: Captura de la imagen de un gel de electroforesis

Idealmente el proceso de electroforesis produce una muestra con carriles perfectamente separados y bandas fácilmente identificables. Sin embargo en circunstancias reales se obtienen imágenes como la mostrada en la figura 1.2, donde aunque cada carril es identificable (no siempre es así) se observa cómo la separación entre ellos no es constante y en ciertas regiones se dificulta la identificación de las bandas, elevando la complejidad de automatización de su detección. A esto hay que agregar factores externos a la hora de la captura de la imagen como lo es el nivel de iluminación o posibles reflejos en la superficie del cristal que introducen “ruido” en la imagen el cual debe ser correctamente eliminado. En un proyecto anterior [9] se realizaron algoritmos para la detección de bandas basados en optimización para poder separar bandas que se encuentran juntas, como en la imagen mostrada en la figura 1.5, sin embargo en estos casos se da por conocido el ancho de las bandas.

Es por ello que su detección no es completamente automática ya que se está partiendo de un ancho dado. Para que esta detección sea completamente automática y como paso previo

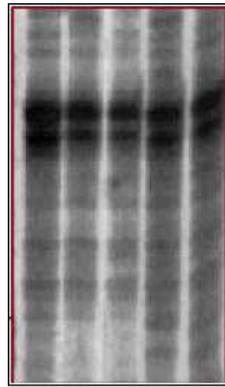


Figura 1.5: Ejemplo de poca separación entre bandas

a los mencionados algoritmos se hace necesaria la implementación de la estimación tanto del ancho como de la cantidad de bandas presentes en un carril lo cual daría completa autonomía a la determinación de la posición de las mismas.

1.3 Estrategia propuesta de detección de bandas

En la figura 1.6 se ilustra el diagrama de bloques general del proyecto de *Análisis por computador de imágenes de geles de electroforesis para la caracterización molecular de organismos* [1]. El actual proyecto se enfoca en el bloque de detección de bandas, específicamente en el proceso de estimación del ancho y la cantidad de estas.

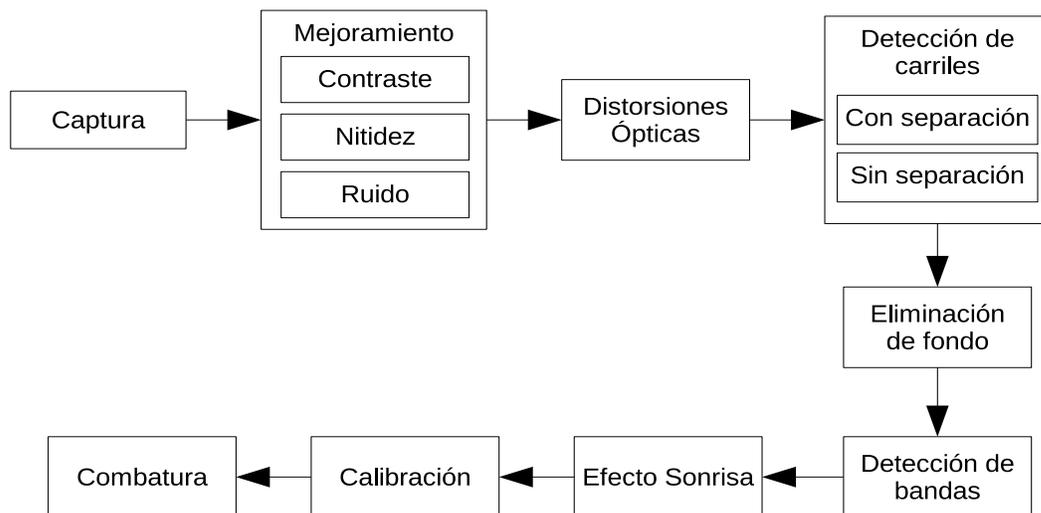


Figura 1.6: Bloques principales del módulo de procesamiento de imágenes (Tomado de [1])

Para la estimación del ancho de las bandas se parte del supuesto de que el perfil de intensidades del carril presenta una distribución normal o gaussiana en los sectores de la imagen donde se localiza una banda presentando una desviación estándar común (la misma para todas las bandas). Para la estimación de la desviación estándar se va a hacer uso de la técnica de espacio de escalas (scale-space) y filtrado de imágenes (ver capítulo 2) mediante

el uso de filtros gaussianos.

La figura 1.7 presenta el diagrama de bloques del procedimiento a seguir. Partiendo de una función en una dimensión (el perfil de intensidades), la cual se analiza utilizando técnicas de espacio de escalas, se genera una nueva sobre la cual se debe trabajar para la determinación del ancho (desviación estándar) promedio de las bandas en el carril. Y sabiendo esto es posible además estimar el número de bandas presentes.

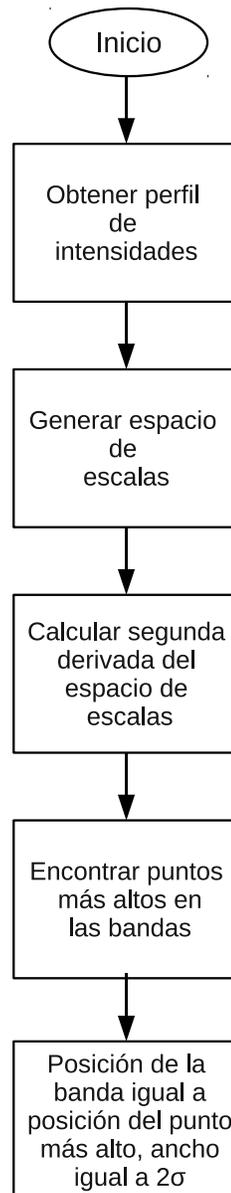


Figura 1.7: Diagrama de bloques de la solución

1.4 Objetivos y estructura del documento

El objetivo principal de este proyecto es implementar un algoritmo que permita determinar correctamente y de manera automática la desviación estándar y cantidad de las bandas

presentes en los carriles de imágenes de geles de electroforesis para lo cual se procede a generar una representación del carril en un espacio de escalas el cual permite la estimación de la desviación estándar y con ello la cantidad de bandas presentes. Todo esto se evaluará aplicándolo a carriles generados sintéticamente lo cual permite conocer los valores teóricos con los cuales se comparan los resultados obtenidos mediante la aplicación del algoritmo.

En el capítulo 2 se cubren los aspectos teóricos necesarios para una completa comprensión de los temas aquí tratados, en el capítulo 3 se da una explicación detallada de los pasos seguidos para la implementación del algoritmo propuesto y por lo tanto cumplir con los objetivos planteados; en el capítulo 4 se analizan los resultados obtenidos tras la realización del proyecto y finalmente en el capítulo 5 se dan las conclusiones obtenidas tras el análisis de los resultados así como algunas recomendaciones para futuras implementaciones.

Capítulo 2

Marco Teórico

En esta sección se presentan los conceptos teóricos que soportan la solución propuesta.

Como punto de partida se da una definición de imagen, para posteriormente poder definir el proceso de filtrado y el proceso de generación del espacio de escalas.

2.1 Imágenes en escala de grises y perfil de intensidades

Una imagen puede definirse como una función de la forma $I = f(x, y)$ donde el par ordenado (x, y) representan la posición de un píxel de la imagen e I representa el valor asociado a esa posición.

Una imagen en escala de grises es aquella que está dada por un conjunto de píxeles definidos por su posición dentro de la imagen y su intensidad que va del negro al blanco, donde el negro está representado por el número menor y el blanco por el número mayor en la escala utilizada. En la figura 2.1 se muestra un ejemplo de esto, si se utiliza una escala de 0 a 255 (8 bits) un píxel ubicado en la posición (10, 10) con un valor de 0 (*b*) se verá como un punto de color negro, mientras que uno ubicado en la misma posición con un valor de 255 (*a*) se verá como un punto blanco.

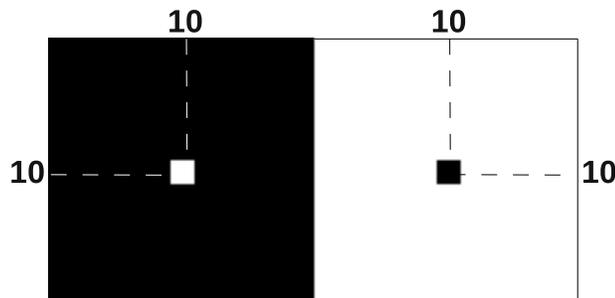


Figura 2.1: Representación de píxeles en una imagen

La figura 2.2 muestra un ejemplo de una escala de grises utilizando 8 bits (escala de 0 a 255). Como se puede ver la intensidad del color va del negro al blanco pasando por todos los colores intermedios cuando la escala va de 0 a 255. En la figura 2.3 se muestra el perfil de intensidades de una fila de la imagen en la figura 2.2; como se puede ver esta es una gráfica del valor del píxel (intensidad) en función de su posición en una fila dada de la imagen la cual, en este caso, presenta un crecimiento constante.

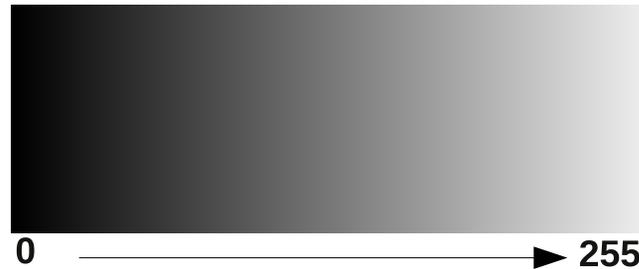


Figura 2.2: Ejemplo de imagen en escala de grises

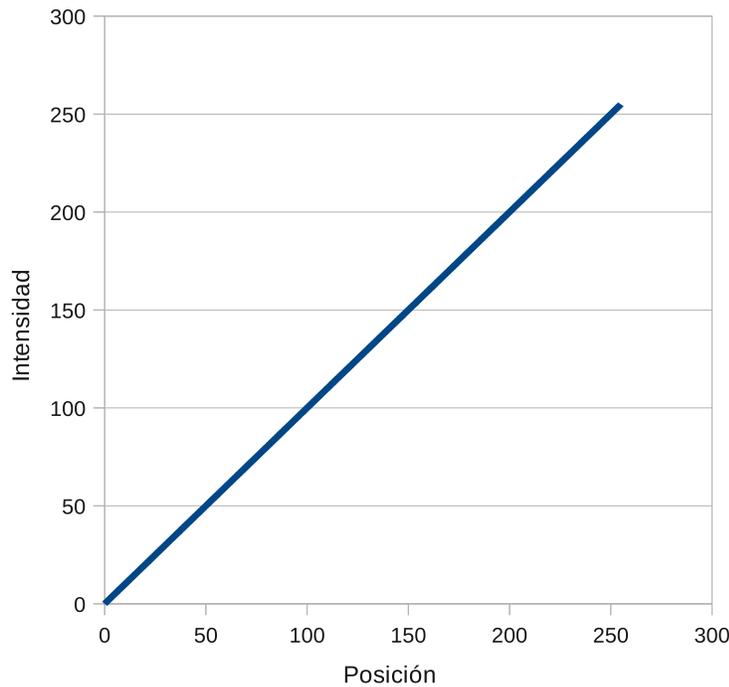


Figura 2.3: Perfil de intensidades de la imagen de la figura 2.2

La figura 2.4 muestra un ejemplo del perfil de intensidades de una distribución gaussiana. En esta distribución el 99.74% del área bajo la curva se encuentra contenida en el intervalo de -3σ a $+3\sigma$ por lo que fuera de ahí se puede considerar como cero.

2.2 Filtrado de imágenes

Por filtrado de una imagen entiéndase la técnica por medio de la cual se elimina determinada información según las especificaciones del filtro a utilizar. Se pueden definir tres tipos de

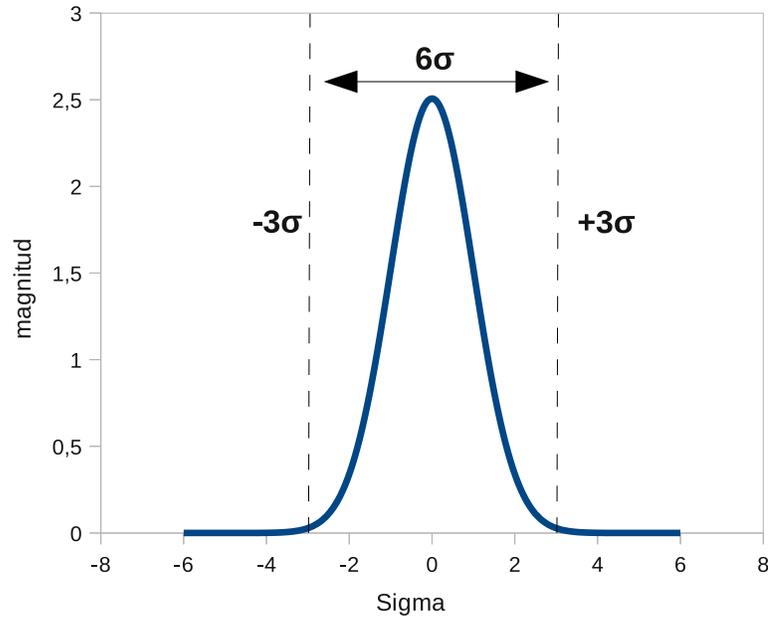


Figura 2.4: Distribución normal o gaussiana

filtro: pasa-bajas, paso-alto y pasa-bandas según sea la banda de frecuencia que se desea “dejar pasar”. En la sección 2.1 se definió una imagen como una función. Además, mediante la transformada discreta de Fourier toda función finita discreta puede representarse como una sumatoria discreta de funciones periódicas, dando como resultado un espectro (función en el dominio de la frecuencia) discreto periódico[4]. Teniendo esto en cuenta se sabe que es posible expresar una imagen en el dominio de la frecuencia como una sumatoria de funciones armónicamente relacionadas.

Un caso particular de filtros paso-bajo utilizado en este trabajo es el filtro gaussiano, el cual debe su nombre a que su respuesta en frecuencia presenta una distribución normal o gaussiana definida como

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \quad (2.1)$$

En la figura 2.5 se muestra la representación de esta función en el dominio del espacio.

Como se puede ver en dicha representación un aumento en la desviación estándar σ ocasiona una expansión de la función en el espacio. En la figura 2.6 se puede ver la representación de esta función en el dominio de la frecuencia y al igual que su representación en el espacio presenta una distribución normal descrita por

$$g(\omega) = e^{-\frac{1}{2}(\omega\sigma)^2} \quad (2.2)$$

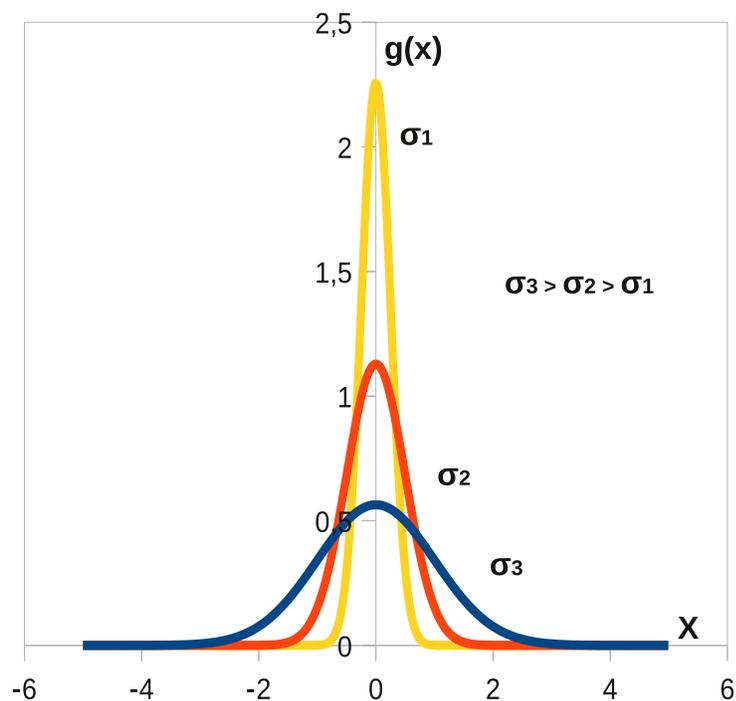


Figura 2.5: Distribución normal para diferentes varianzas

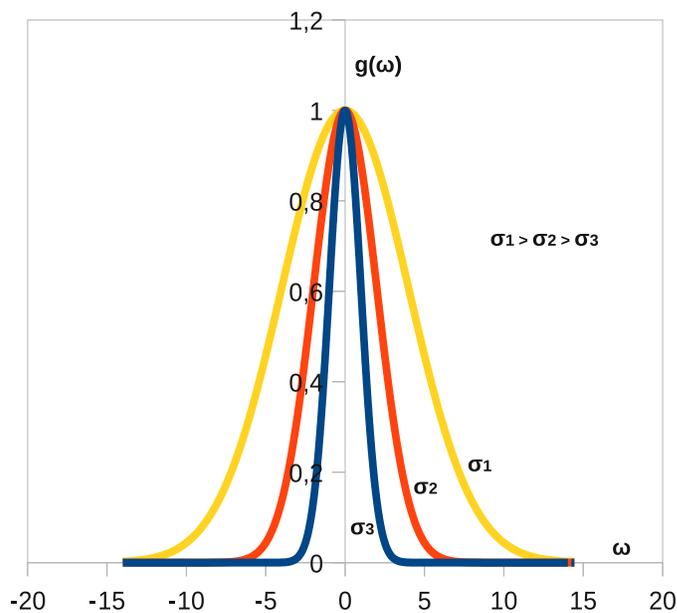


Figura 2.6: Representación de las funciones de la figura 2.5 en el dominio de la frecuencia

Si se comparan (2.1) y (2.2) se puede notar como la función en el dominio de la frecuencia presenta una desviación estándar $\frac{1}{\sigma}$, para una señal en el espacio con desviación estándar σ , lo cual indica que un aumento de σ ocasiona una expansión de la función en el espacio pero una contracción en la frecuencia[2].

2.3 Generación del espacio de escalas

Con la técnica de espacio de escalas se incorpora la señal original en una familia con un parámetro de señales derivadas, donde el parámetro indicado, t , describe el nivel actual de escala o el nivel de detalle[7].

En la figura 3.4 se ejemplifica el proceso de generación del espacio de escalas. En este caso, por ser el perfil de intensidades una función unidimensional se utiliza un espacio de escalas bidimensional.

El proceso de generación del espacio de escalas consiste en aplicar sucesivas transformaciones (filtrar) a la imagen original mientras se varía el parámetro de escala, con lo que se genera una nueva imagen con un nivel de detalle menor a la original.

En este caso la transformación realizada es la convolución de la imagen con una función gaussiana a la cual se le aumenta su varianza (escala):

$$L(x; t) = f(x) * g(x; t) \quad (2.3)$$

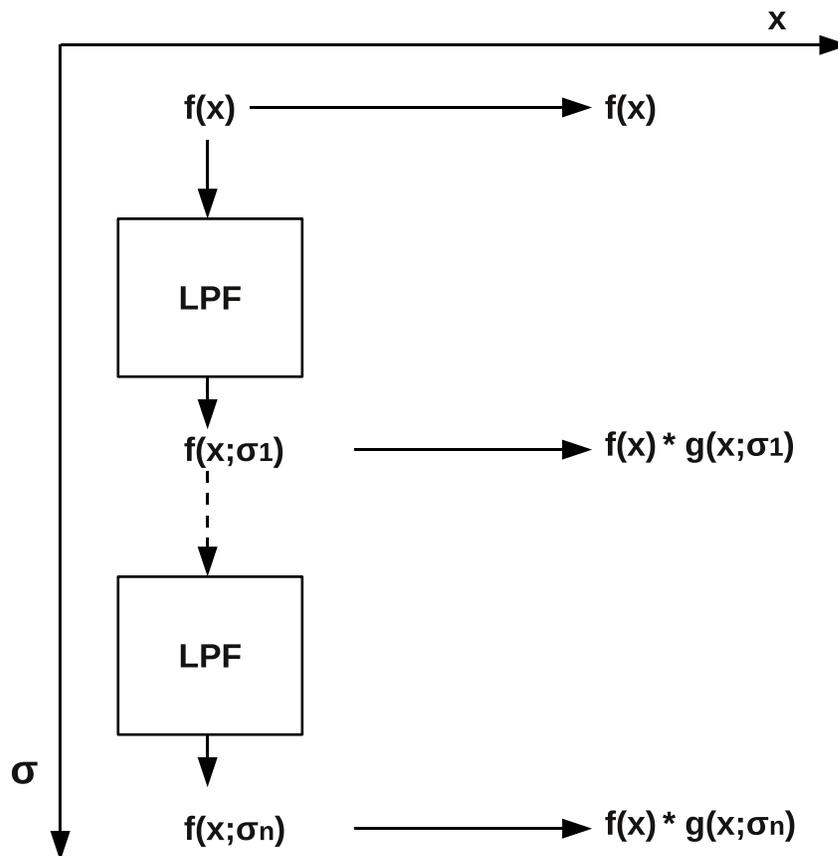


Figura 2.7: Representación de una función $f(x)$ en un espacio de escalas

Como punto de partida se tiene el perfil de intensidades de la imagen original $f(x)$, el cual corresponde a la escala $t = 0$, o lo que es igual, a filtrar el perfil original con una función

gaussiana de varianza cero ($g(x; 0)$). Posteriormente se incrementa la escala (σ) y nuevamente se filtra el perfil original con la función gaussiana ($g(x; \sigma_1)$), cuyo resultado corresponde a la siguiente fila de la imagen del espacio de escalas. Una vez hecho esto se incrementa nuevamente la escala y el resultado corresponde a la siguiente fila y este proceso se repite hasta tener una nueva imagen bidimensional formada a partir del perfil de intensidades original.

La figura 2.8 muestra un ejemplo de cómo al variar la escala se varía el nivel de detalle contenido en la imagen.



Figura 2.8: Ejemplo de una imagen en varias escalas. Imagen original a , $t = 1$ b , $t = 8$ c , $t = 64$ d (Tomado de [8])

En (2.8a) se puede ver la imagen original, en (2.8b) se observa la misma imagen filtrada utilizando una escala de 1, en (2.8c) se utiliza una escala de 8 y finalmente en (d) una escala de 64. Como se observa en (a) se pueden apreciar claramente los detalles en la imagen, mientras en (b) el nivel de detalle es levemente menor, en (c) se puede distinguir aún menos de la imagen mientras en (d) apenas se logran distinguir siluetas. Esto ejemplifica el nivel de detalle contenido en la imagen depende de la escala utilizada.

El ejemplo mostrado en la figura 2.8 corresponde a una imagen en dos dimensiones, ahora bien, para el caso de una imagen en una dimensión (el perfil de intensidades) se tiene el mismo principio solo que aplicado al perfil y el resultado final es una nueva imagen donde conforme se avanza en la fila se tiene una versión de la original pero con un nivel de detalle menor. En las figuras 2.9 y 2.10 se muestra un ejemplo de la generación del espacio de escalas a partir de una imagen unidimensional.



Figura 2.9: Imagen de ejemplo

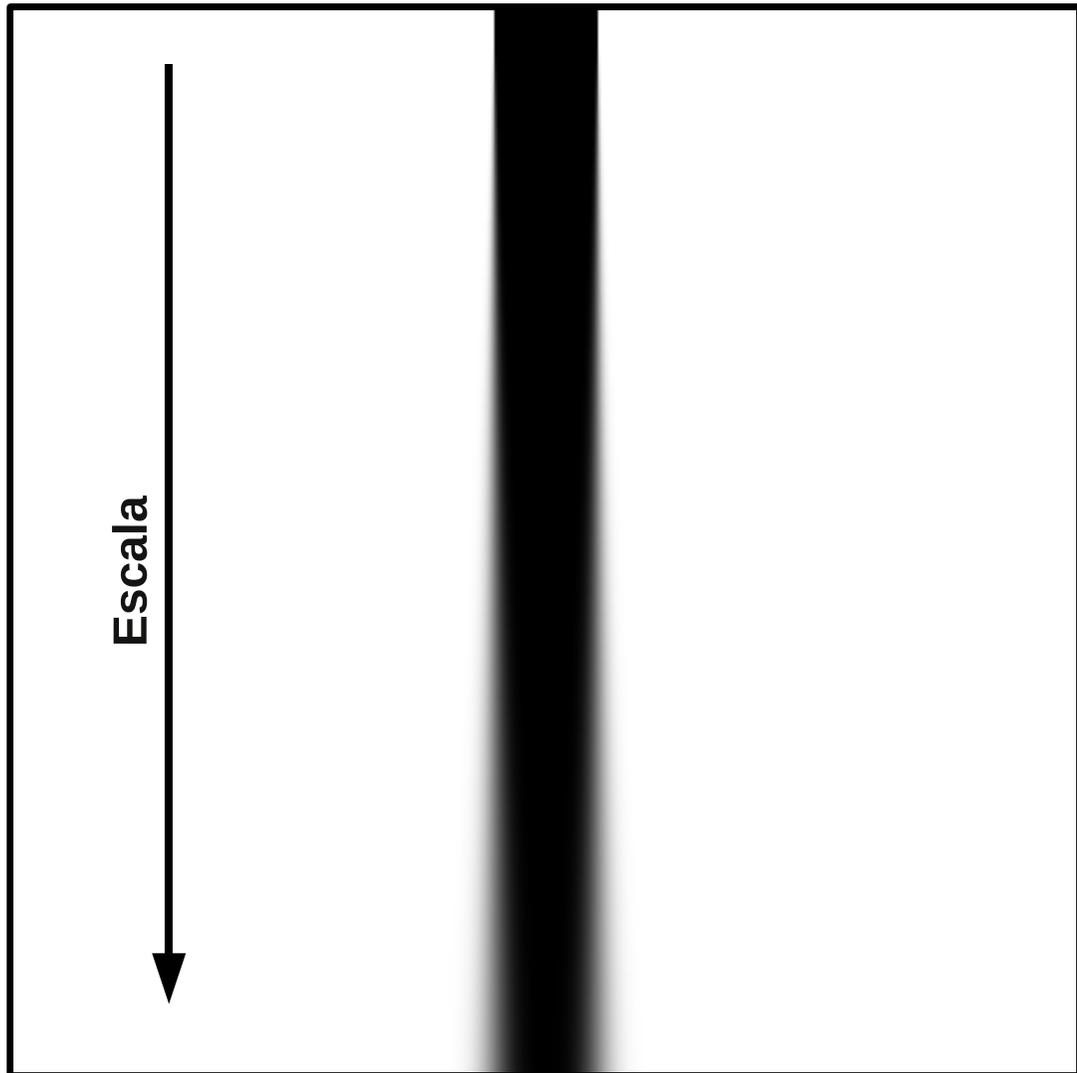


Figura 2.10: Espacio de escalas de la imagen en la figura 2.9

En la figura 2.9 se observa la imagen original de la cual se toma el perfil de intensidades para generar el espacio de escalas mostrado en la figura 2.10. Como se aprecia, conforme se aumenta en la escala se tiene una nueva versión de la imagen original con un menor nivel de detalle, en este caso, los bordes de la franja se difuminan.

Capítulo 3

Sistema de detección de bandas

En este capítulo se presentan los algoritmos implementados para la detección de la posición, cantidad y ancho de las bandas. Para la realización de dicha tarea se hace uso de las técnicas de espacio de escalas y “detección de bordes” utilizando un filtro pasa-bajo de tipo gaussiano.

Con el uso de estas dos técnicas se logra la generación de una nueva imagen del perfil de intensidades del carril bajo análisis, con la cual se determina la presencia o ausencia de bandas y con ello la cantidad y ancho (desviación estándar) de estas.

A continuación se presenta una explicación detallada sobre la implementación de dichos algoritmos.

3.1 Perfil de intensidades

En la figura 3.1 se presenta un ejemplo de carril en una imagen de electroforesis en gel. En ella se puede observar cómo la separación entre bandas no es constante además de que las bandas no presentan la misma intensidad pudiéndose observar bandas perfectamente delimitadas (con una alta intensidad) y bandas que tienden a confundirse con el fondo (background) del carril.

En la figura 3.2 se puede ver la gráfica del perfil de intensidades del carril mostrado en la figura 3.1. En ella se puede observar con mayor claridad la relación de intensidades que existe entre las distintas bandas presentes en el carril. Como se observa la gráfica presenta un valor constante en su mayoría el cual corresponde al fondo y los puntos más bajos representan las bandas con mayor intensidad mientras los puntos altos las de menor intensidad, siendo la segunda banda la de menor intensidad.



Figura 3.1: Ejemplo de un carril de una imagen de un gel de electroforesis

De aquí en adelante todo el análisis se hará sobre el perfil de intensidades del carril bajo estudio. Como primer paso se genera una nueva imagen utilizando la técnica de espacio de

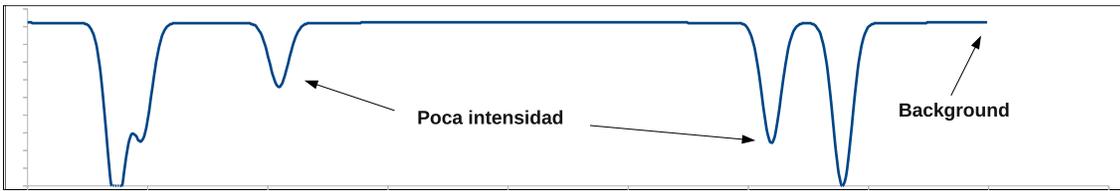


Figura 3.2: Perfil de intensidades del carril mostrado en la figura 3.1

escalas a la cual se le aplica la técnica de detección de blobs con lo cual se trabaja para ubicar la posición de cada una de las bandas así como su ancho y cantidad.

3.2 Generación del espacio de escalas

La figura 3.4 muestra el proceso de generación del espacio de escalas de un carril. Este consiste en tomar el perfil de intensidades del carril bajo estudio y aplicar un filtro gaussiano. Con esto se tiene un nuevo perfil. Posteriormente se aplica nuevamente al carril original un filtro gaussiano pero esta vez aumentando su desviación estándar y su tamaño. Cada resultado del filtrado del perfil pasa a conformar una nueva fila de la imagen genera con lo que al final del proceso se tiene una imagen en dos dimensiones cuyas filas son siempre el resultado de filtrar el carril original con una función gaussiana cuya desviación estándar aumenta según se avanza en las filas y cuyo tamaño aumenta en función de su desviación estándar siendo éste 6 veces el valor de la desviación estándar σ ya que es en esta sección del filtro donde se encuentra contenido el 99.74% del área.

El proceso de generación del espacio de escalas se puede comprender mejor matemáticamente estando este descrito por

$$f(n) = f(0) * g(x; t) \quad (3.1)$$

donde $f(0)$ representa el perfil de intensidades del carril original, el cual es filtrado con una función gaussiana $g(x; t)$ cuyo parámetro t aumenta al descender en las filas.

La figura 3.3 ilustra gráficamente este proceso. En ella se pueden distinguir tres ejes: el eje izquierdo representa la escala (desviación estándar) utilizada en el filtro, el eje superior representa el espacio, o lo que es igual, la posición de cada uno de los píxeles del perfil de intensidades y por último el eje derecho representa la fila correspondiente a la nueva imagen (espacio de escalas) en la cual se “coloca” el nuevo perfil de intensidades.

Como se puede observar, en la escala $t = 0$ el perfil de intensidades correspondiente es el del carril original, mientras que el perfil correspondiente a la fila número uno de la imagen de espacio de escalas corresponde a la convolución del perfil original con una función gaussiana con desviación estándar σ_1 , el perfil correspondiente a la segunda fila corresponde al perfil original convolucionado con una gaussiana con desviación estándar σ_2 y así sucesivamente hasta formar una nueva representación bidimensional.

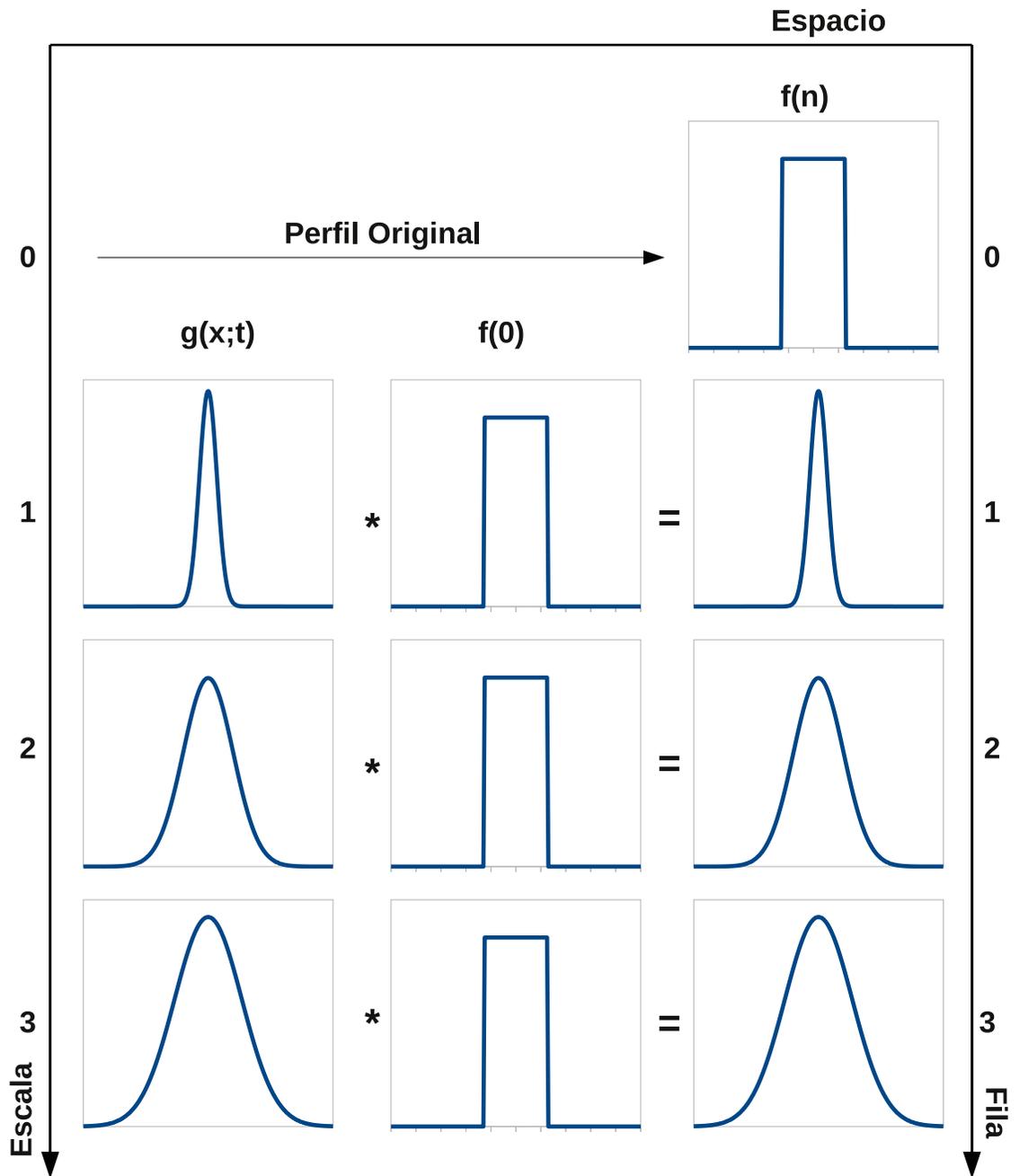


Figura 3.3: Ejemplo del proceso de generación de el espacio de escalas

Otro punto a rescatar de esta imagen es que si bien, como se mencionó antes el tamaño del filtro utilizado es siempre igual a 6σ el tamaño de éste va en aumento, ya que σ aumenta en cada fila. Además en (3.3) se está haciendo uso de una representación de la función en el dominio del espacio lo cual hace que al aumentar la desviación estándar del filtro éste se expanda; sin embargo como se mencionó en el capítulo 2 en el dominio de la frecuencia al aumentar el valor de σ el filtro se contrae lo cual reduce la banda pasante conforme se aumenta la desviación estándar con lo que se explica la reducción de información presente con forme se aumenta en la escala.

En la figura 3.4 se puede ver la imagen de espacio de escalas generada para el perfil de intensidades mostrado en la figura 3.2. Para esta imagen se utilizó un incremento en la desviación estándar del filtro de 0.015625 partiendo de una desviación estándar de 0 para la fila 0 (perfil original) e incrementando su valor conforme se avanza en cada fila.



Figura 3.4: Espacio de escalas para el perfil de intensidades de la figura 3.1

Como se puede ver claramente conforme se desciende en la generación del espacio de escalas el perfil de intensidades se va “difuminando” o lo que es igual, se va disminuyendo la cantidad de información presente en la imagen dejando únicamente información con contenido de bajas frecuencias.

En la siguiente sección se define el porqué de la selección del incremento antes mencionado así como la forma en que esta nueva imagen (espacio de escalas) generada a partir del perfil de intensidades del carril original se utiliza para determinar la ubicación de cada una de las bandas así como su ancho y por lo tanto la cantidad de bandas presente en el carril.

3.3 Detección de blobs

Para la determinación de la desviación estándar de las bandas presentes en el carril se aplicó la técnica de detección de blobs a la imagen de espacio de escalas generada a partir del perfil de intensidades del carril analizado.

Para la realización de dicha tarea se procede a la generación de un nuevo espacio de escalas calculando la segunda derivada normalizada del espacio original.

3.3.1 Aplicación de la segunda derivada

En este caso se aplica una derivación unidimensional, es decir, se deriva cada una de las filas de la imagen de espacio de escalas por separado con lo cual se genera la nueva imagen.

En la figura 3.5 *a* se puede ver un ejemplo de una representación del perfil de intensidades de un carril, en *b* se observa la primera derivada del perfil y en *c* la segunda derivada.

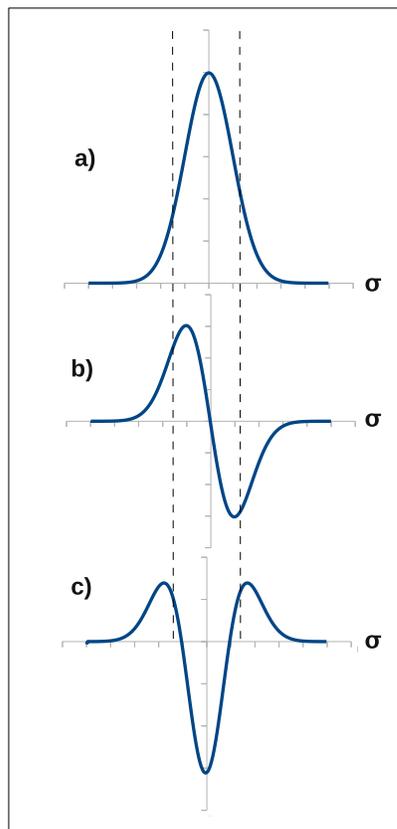


Figura 3.5: Segunda derivada de una función gaussiana

Como se puede observar, al aplicar la segunda derivada al perfil de intensidades de la banda, se generan dos máximos en sus extremos y un mínimo en el centro; característica aprovechada para determinar su ubicación dentro del carril.

La figura 3.5 equivale a una banda de color blanco en un fondo oscuro, sin embargo las imágenes de geles de electroforesis utilizadas presentan bandas de color negro sobre un fondo

claro. No obstante el único cambio es que se presentan mínimos en los extremos de la banda y un máximo en la parte central (la gráfica se invierte).

3.3.2 Generación del nuevo espacio de escalas

El nuevo espacio de escalas generado y sobre el cual se trabaja para la determinación de la desviación estándar asociada a las bandas consiste en aplicar la segunda derivada al generado previamente en el cual se va a cumplir lo mencionado en la sección anterior, es decir, en los extremos de las bandas se presentan mínimos mientras que en el centro se presenta un máximo.

En la figura 3.6 se puede ver el resultado de aplicar la segunda derivada a la imagen presentada en 3.4.

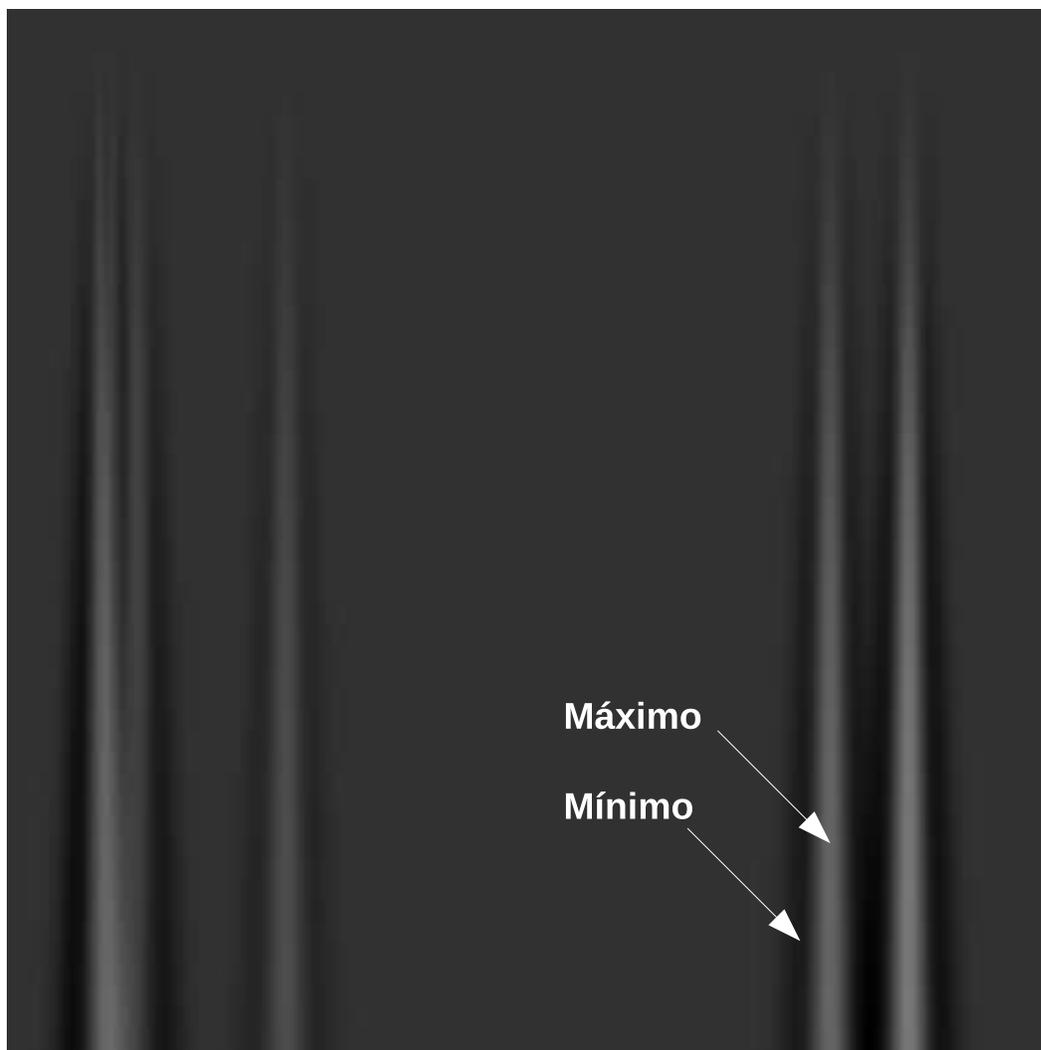


Figura 3.6: Segunda derivada del Espacio de Escalas

Cómo se observa, claramente en la parte central de cada una de las bandas se da un máximo (parte clara de la imagen), mientras en los extremos se distinguen áreas oscuras que corresponden a los mínimos tal y como se explicó en la sección anterior.

Todo esto se puede explicar matemáticamente de la siguiente forma: si se expresa un borde cualquiera con la función escalón unitario:

$$f(x) = u(x - x_0) \quad (3.2)$$

donde x_0 representa la posición del borde y se genera un espacio de escalas como el mencionado anteriormente (aplicando la segunda derivada):

$$f(x) * \frac{d^2G(x)}{dx} = \frac{d}{dx} \left(u(x - x_0) * \frac{dG(x)}{dx} \right) \quad (3.3)$$

$$= \frac{dG(x - x_0)}{dx} \quad (3.4)$$

$$= -\frac{(x - x_0)}{\sigma^3 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - x_0}{\sigma} \right)^2} \quad (3.5)$$

donde $G(x)$ es el filtro gaussiano utilizado. Como se puede observar, en $x = x_0$ la función se hace cero lo cual explica la “sombra” que se mencionó anteriormente en los extremos de las bandas.

Ahora bien, para encontrar la posición donde se da el máximo (anteriormente se dedujo analíticamente que es en el centro de la banda) se procede a calcular e igualar a cero la derivada de (3.3):

$$0 = \frac{d}{dx} \left(u(x - x_0) \frac{d^2G(x)}{dx^2} \right) \quad (3.6)$$

$$0 = - \left(1 - \frac{(x - x_0)^2}{\sigma^2} \right) \frac{1}{\sigma^3 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - x_0}{\sigma} \right)^2} \quad (3.7)$$

$$0 = 1 - \frac{(x - x_0)^2}{\sigma^2} \quad (3.8)$$

$$x = x_0 \pm \sigma \quad (3.9)$$

Como se puede ver en (3.9) a una distancia de $\pm\sigma$ de la posición del borde se da un máximo.

Si se evalúa (3.5) en $x_0 \pm \sigma$ se puede ver como el valor del máximo es de $\pm \frac{1}{\sigma^2 \sqrt{2\pi}} e^{-\frac{1}{2}}$ lo cual es claro que depende del valor de σ ; es por ello que se normaliza multiplicando por σ^2 lo cual elimina la dependencia de σ en el valor de el máximo.

Como se concluyó en [9] es posible modelar la distribución de intensidades que caracteriza el perfil de una banda mediante una función gaussiana para la cual su valor medio determina la ubicación de la misma en el carril, sin embargo esta puede ser modelada mediante un impulso rectangular.

Sabiendo esto, una banda puede ser representada como la sumatoria de dos escalones unitarios de la forma:

$$I(x) = u(x - x_0) - u(x - x_1) \quad (3.10)$$

lo cual define una banda que va de x_0 a x_1 .

De la misma manera que se procedió para un solo borde se puede proceder para el caso de la banda (dos bordes) lo cual resulta en máximos en $x_0 \pm \sigma$ y $x_1 \pm \sigma$. De aquí se puede extraer que cuando la banda presenta un ancho igual a 2σ ambos máximos se darán en el mismo punto.

Esto implica que cuando $\sigma = \frac{x_1 - x_0}{2}$ (la mitad del ancho de la banda) se da el punto más alto dentro de la banda por lo que basta con ubicarlo para poder estimar tanto su ancho como su posición[6]. En la figura 3.7 se ilustra cómo se da el punto más alto dentro de la banda.

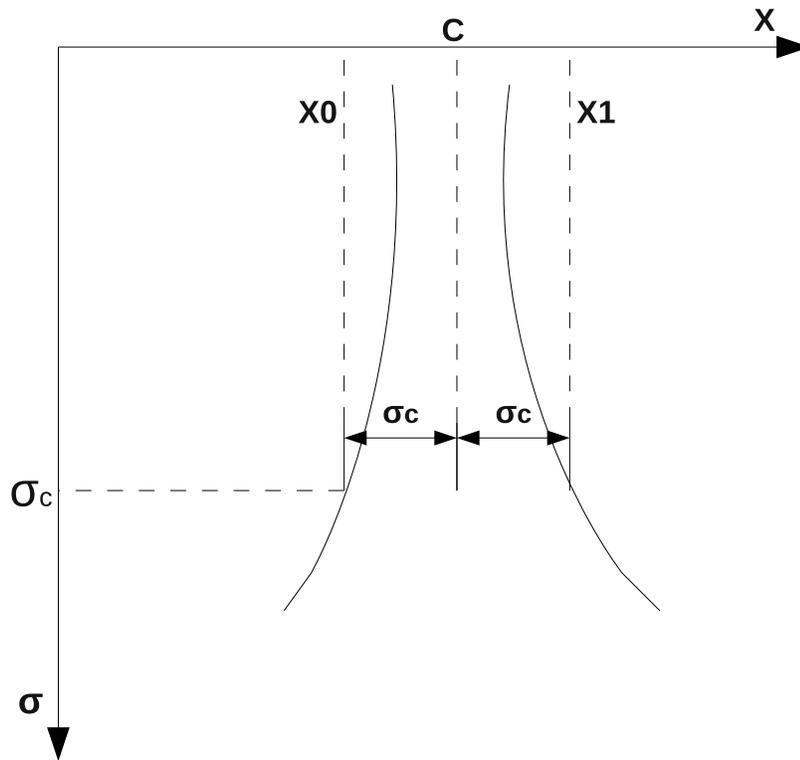


Figura 3.7: Localización del punto más alto de la banda

La figura 3.7 representa el espacio de escalas correspondiente a una de las bandas. Como se puede observar, la fila que fue generada con un σ del filtro exactamente igual a la mitad del ancho de la banda presenta un máximo en la parte central de esta debido a que en ese punto se suman los máximos correspondientes a $x_0 + \sigma$ y $x_1 - \sigma$, lo cual genera el punto más alto dentro de la banda. Esta característica se repite en todas las bandas por lo que solo basta con ubicar los valores máximos en el espacio de escalas, los cuales se dan en la parte central de las bandas (posición de estas) con un sigma que es igual a la mitad de su ancho.

La figura 3.8 muestra la imagen de la figura 3.6 con los puntos donde se dan los máximos mencionados anteriormente resaltados, y en la figura 3.9 se puede ver la posición estimada por el algoritmo.

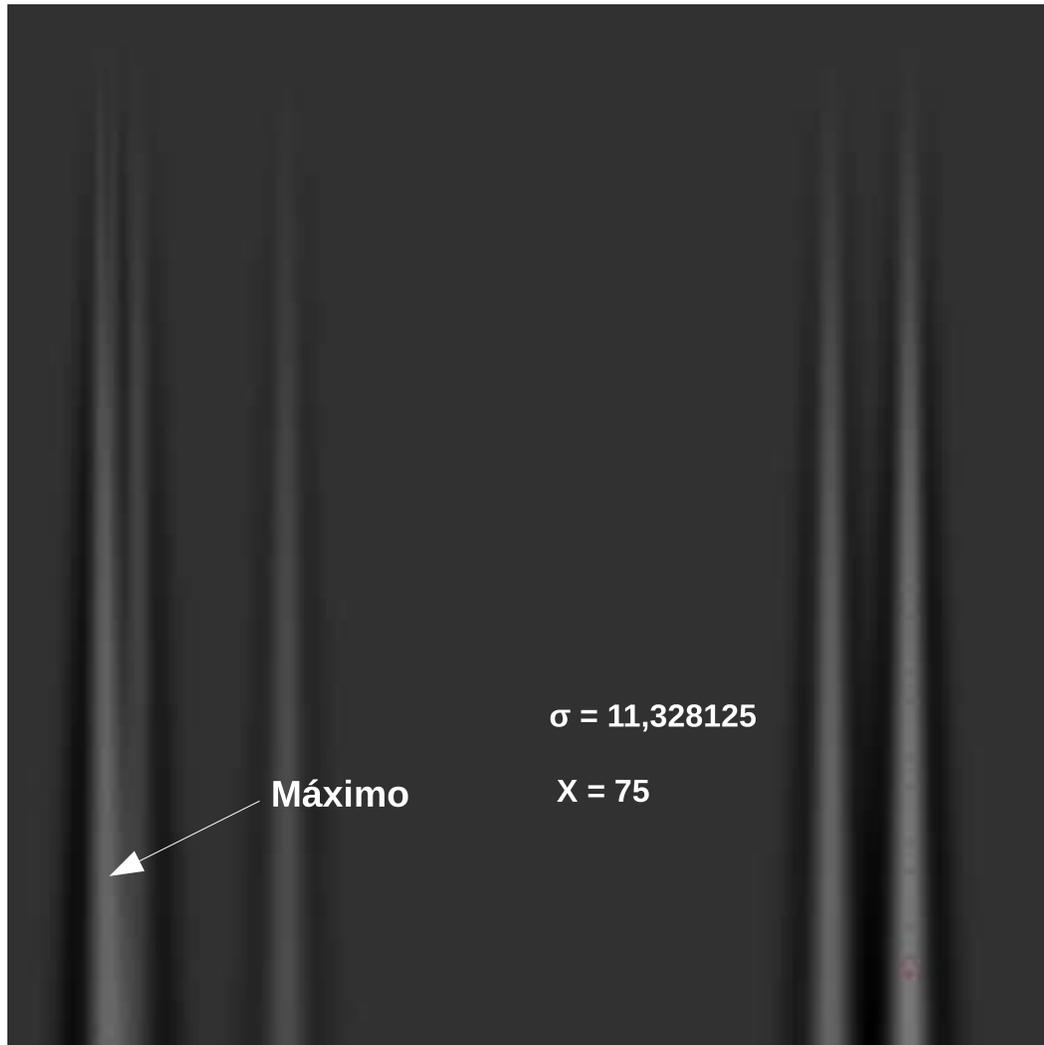


Figura 3.8: Ubicación de los máximos en el espacio de escalas



Figura 3.9: Ubicación de las bandas en el carril

Como se puede ver, basta con ubicar los puntos más altos de la gráfica para determinar tanto la posición como el ancho de la banda y por lo tanto la cantidad de estas dentro del carril.

En la figura 3.8 se puede ver como en una de las bandas el punto más alto se da con un valor de σ del filtro de 11,328125 en la posición 75 lo cual indica que dicha banda tiene un ancho de $2\sigma = 22,62655$ píxeles y se ubica centrada en $x = 75$. En la figura 3.9 se pueden ver las posiciones de las bandas determinadas por el algoritmo; como se observa claramente se detectó correctamente la presencia de 5 bandas.

En resumen, el procedimiento seguido para la caracterización de las bandas es el siguiente: se obtiene el perfil de intensidades del carril a analizar, con ese perfil de intensidades se genera un espacio de escalas aplicando un filtro gaussiano cuya desviación estándar aumenta conforme se avanza en la generación del espacio de escalas. Una vez obtenido este, se calcula la segunda derivada normalizada y es en este nuevo espacio donde se ubican los puntos más altos (máximos) que se dan en el centro de las bandas con un valor del σ del filtro que es igual a la mitad del ancho de la banda; por lo que el ancho estimado es de 2σ . Teniendo esto se conoce la posición y ancho de cada una de las bandas del carril y por lo tanto se puede calcular su cantidad.

3.4 Caso de una función gaussiana

En la sección anterior se demostró cómo definiendo el perfil de intensidades de las bandas como un impulso rectangular se da un máximo ubicado en el centro de la banda cuando se utiliza un filtro con σ igual a la mitad del ancho de esta, lo cual se aprovecha para detectar su posición, ancho y por lo tanto la cantidad en el carril. En esta sección se demuestra cómo el mismo concepto puede ser aplicado a un perfil de intensidad descrito por una función gaussiana, ya que según se concluye en [9] las bandas en las imágenes de geles de electroforesis presentan esta distribución.

Como se definió en el capítulo 2 el filtrado de una imagen está descrito por la convolución de la imagen con el filtro. Ahora bien, si tanto la imagen como el filtro son una función gaussiana el resultado de esa convolución es una nueva función gaussiana como se demuestra a continuación.

Si se tiene el perfil de intensidades de una banda descrito por

$$f(x) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x}{\sigma_i} \right)^2} \quad (3.11)$$

donde σ_i es la desviación estándar de la banda, y un filtro de la forma

$$g(x) = \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x}{\sigma_f} \right)^2} \quad (3.12)$$

donde σ_f es el parámetro de escala (σ) usado en el filtro, y se calcula la transformada de Fourier de ambos, se puede realizar la convolución de las dos funciones como

$$H(\omega) = e^{-\frac{1}{2}(\omega\sigma_i)^2} e^{-\frac{1}{2}(\omega\sigma_f)^2} \quad (3.13)$$

$$= e^{-\frac{1}{2}\omega^2(\sigma_i^2 + \sigma_f^2)} \quad (3.14)$$

$$= e^{-\frac{1}{2}(\omega\sigma_r)^2} \quad (3.15)$$

donde $\sigma_r = \sqrt{\sigma_i^2 + \sigma_f^2}$, y volviendo al dominio espacial se tiene

$$h(x) = \frac{1}{\sigma_r\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma_r}\right)^2} \quad (3.16)$$

que como se ve, es una función gaussiana ubicada en la misma posición de la banda pero con un nuevo σ .

Si se calcula la segunda derivada de (3.16) se obtiene

$$\frac{d^2h(x)}{dx^2} = -\frac{(\sigma_r^2 - x^2)}{\sigma_r^5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma_r}\right)^2} \quad (3.17)$$

la cual, como se puede ver en la figura 3.10 presenta su valor máximo (en magnitud) en $x = 0$ el cual tiene un valor de

$$\max\{h(x)\} = h(0) = \frac{1}{\sigma_r^3\sqrt{2\pi}} \quad (3.18)$$

$$= \frac{1}{\sqrt{\sigma_i^2 + \sigma_f^2}^3 \sqrt{2\pi}} \quad (3.19)$$

Como se puede ver en (3.18) el valor del máximo presenta una dependencia de la desviación estándar tanto del filtro como de la banda, por lo que si se le normaliza multiplicando por un σ_f^n se tiene la familia de curvas mostrado en la figura 3.11 para valores de $n = 0; 1; 1,5; 2$ y 3.

La línea roja representa un valor de $n = 0$, la verde $n = 1$, la morada $n = 1,5$, la café $n = 2$ y la azul $n = 3$, todo esto para un valor de $\sigma_i = 1$.

La figura 3.12 muestra cómo se comporta el valor de este máximo cuando se normaliza con $n = 1.5$. En ella se aprecia claramente cómo al centro de la distribución, con un valor de $\sigma_f = \sigma_i$ se presenta el punto más alto de la gráfica, es decir, el máximo. Sin embargo, en el presente proyecto se utiliza la normalización por σ^2 como se mencionó en la sección anterior, por lo que el máximo aparece en $\sigma_i\sqrt{2}$ (línea café de la figura 3.11).

En resumen, si el perfil de intensidades de una banda está descrito por una función gaussiana, al calcular la segunda derivada del espacio de escalas de dicho perfil normalizado por $\sigma_f^{1.5}$,

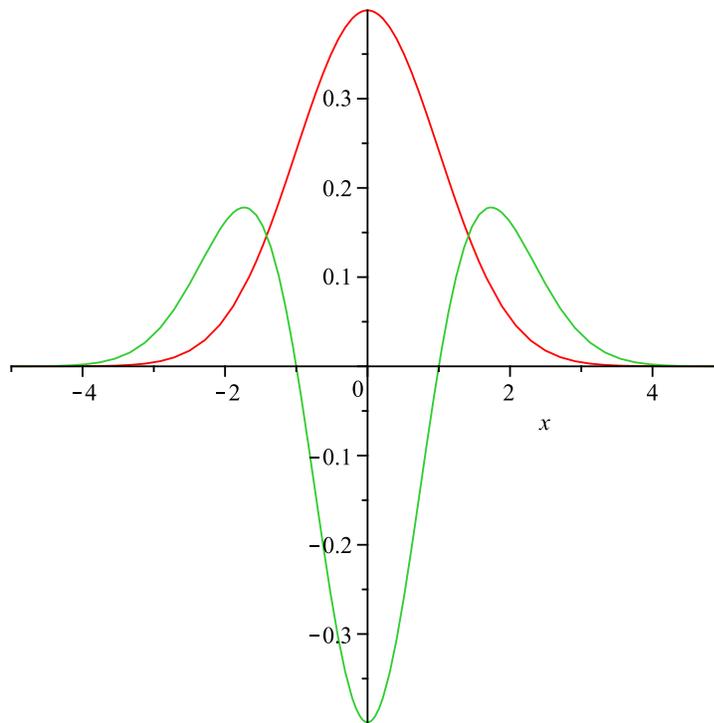


Figura 3.10: Segunda derivada de una función gaussiana

basta con ubicar el punto más alto dentro del espacio correspondiente a la banda para determinar la posición de esta, el cual se da con un $\sigma_f = \sigma_i$, por lo que se logra obtener el ancho, posición y con esto la cantidad de bandas en el carril.

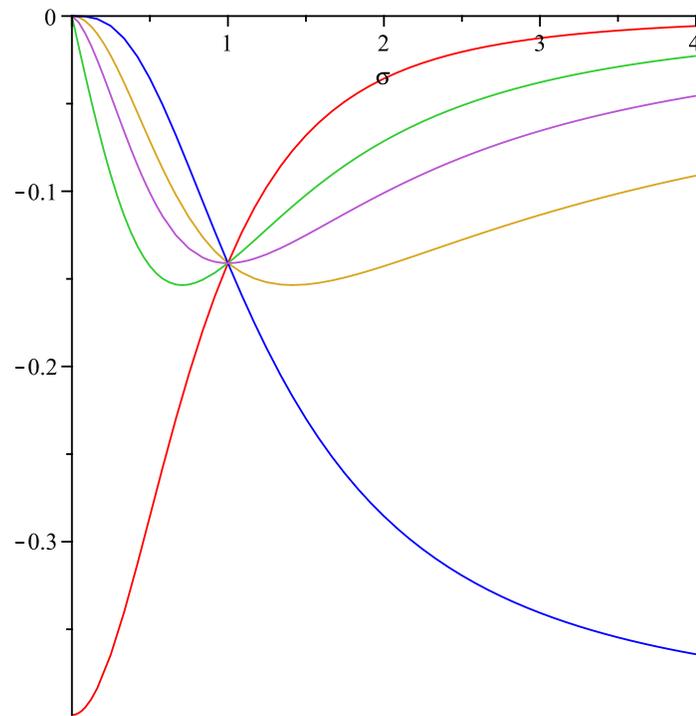


Figura 3.11: Normalización de la ecuación 3.18 para valores de $n = 0$ (línea roja), $n = 1$ (línea verde), $n = 1,5$ (línea morada), $n = 2$ (línea café) y $n = 3$ (línea azul) para un $\sigma_i = 1$

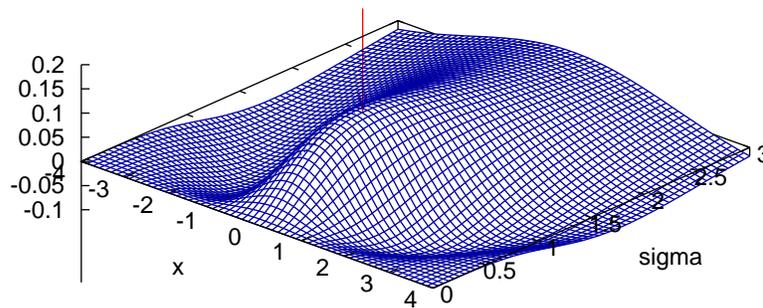


Figura 3.12: Ubicación del valor máximo de (3.17) normalizado por $\sigma_f^{1.5}$ para un $\sigma_i = 1$

Capítulo 4

Resultados y Análisis

En este capítulo se hace un análisis de los resultados obtenidos tras la implementación de los algoritmos propuestos. Los puntos a analizar son: la efectividad en la determinación del ancho de las bandas, determinación de la cantidad de bandas presentes, determinación de la posición de las bandas para finalmente realizar una comparación con el algoritmo ya implementado en [9].

Para evaluar los resultados se utilizan carriles sintéticos generados con un algoritmo desarrollado por David SotoVásquez en [9]. La utilización de dichos carriles permite conocer con exactitud la posición, desviación estándar y cantidad de bandas del carril a analizar contra los cuales se comparan los valores obtenidos mediante el algoritmo implementado.

Para realizar el análisis se consideró un conjunto de datos que consiste en 5 mediciones de carriles generados aleatoriamente con un total de 20 bandas para 10 valores diferentes de σ que van de 1 a 5,5 con un incremento de 0,5 píxeles.

4.1 Determinación de la desviación estándar de las bandas

En esta sección se analizan los resultados obtenidos para la determinación de la desviación estándar de las bandas presentes en el carril. Para la interpretación de los datos se hace uso de la estadística mediante el cálculo de valores medios ($\bar{\sigma}$), desviación estándar (SD) y coeficiente de variación (CV) para cada uno de los valores de σ analizados.

Para realizar el cálculo de CV se utiliza la fórmula:

$$CV = \frac{SD * 100}{\bar{\sigma}} \quad (4.1)$$

En la tabla 4.1 se resumen los resultados obtenidos tras la determinación de la desviación estándar (σ) para el conjunto de datos bajo estudio, mientras que en la tabla 4.2 se presentan las medidas estadísticas de los datos obtenidos.

Tabla 4.1: Desviación estándar obtenida para carriles con 20 bandas

σ teórico \ Medición	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
1	1,37	1,63	2,43	2,60	3,61	4,32	4,35	5,21	4,38	5,50
2	1,34	1,74	2,26	2,97	3,14	3,39	3,53	3,96	5,24	5,04
3	1,37	1,73	2,22	2,52	3,02	3,84	4,51	4,05	5,25	5,87
4	1,33	1,81	2,34	2,47	3,03	3,23	3,94	4,27	4,60	4,87
5	1,72	1,69	2,31	2,499	2,97	3,98	4,87	3,73	5,10	5,09

Tabla 4.2: Medidas estadística para la determinación de la desviación estándar obtenida para carriles con 20 bandas

σ teórico (píxeles)	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
$\bar{\sigma}$	1,43	1,72	2,31	2,61	3,15	3,75	4,24	4,24	4,91	5,28
SD (píxeles)	0,15	0,06	0,07	0,18	0,23	0,39	0,46	0,51	0,36	0,36
CV (%)	10,37	3,45	3,11	7,05	7,44	10,56	10,94	12,09	7,27	6,88
Error (píxeles)	0,43	0,22	0,31	0,11	0,15	0,25	0,24	0,26	0,09	0,22

Como se puede apreciar en la tabla 4.1 los resultados, aunque si bien están dentro del margen de error planteado difieren ligeramente de los valores teóricos, notándose las mayores diferencias para los valores de σ más bajos como por ejemplo para el caso de $\sigma = 1$ donde la diferencia va desde 0,37 píxeles hasta 0,72. Esta diferencia se debe principalmente a 2 factores: el primero es el incremento utilizado para la generación del espacio de escalas y el otro es la forma en que se genera el carril sintético.

En el caso de la generación del espacio de escalas se está utilizando un incremento de 0,015625 lo cual significa que se puede distinguir un cambio en la desviación estándar no menor a 0,1 píxeles. Además este incremento representa el 10 % del valor de σ utilizado lo cual eleva el porcentaje de error en la determinación del σ para valores menores a 2 píxeles.

En el caso de la forma en que se genera el carril sintético en algunos ocasiones se da el caso de dos bandas que se pueden considerar como juntas (con una separación entre ellas de por ejemplo 2 píxeles) las cuales a simple vista lucen como una única banda. Esto se refleja en los valores presentes en la tabla 4.1 ya que los valores allí presentados corresponden al promedio de los σ encontrados para la totalidad del carril por lo que si se da el caso de 2 bandas juntas el σ determinado para ellas será mayor al de las demás bandas lo cual eleva el promedio y por lo tanto el error. La figura 4.1 muestra un ejemplo de dos bandas que fueron generadas con muy poca separación.

Esta imagen corresponde a la primera muestra para un σ de 2 presentada en la tabla 4.1. Como se puede ver, la banda marcada con **a** es más gruesa que la marcada con **b**, esto



Figura 4.1: Ejemplo de dos bandas con poca separación

se debe a que en **a** fueron generadas 2 bandas en las posiciones 55 y 61 las cuales, por su cercanía, aparentan ser una sola. En este caso el σ determinado por el algoritmo fue de 3,552, mientras para el caso en **b** se determinó un σ de 1,9886.

En la tabla 4.2 se observan los resultados estadísticos para los datos en la tabla 4.1. Como se puede ver en la mayoría de los casos el coeficiente de dispersión se encuentra cercano al 10 % lo cual se debe, como se mencionó anteriormente, a la forma en que se genera el carril sintético, ya que al tratarse de valores de σ promedio el hecho de que en algunos casos se generen algunas bandas más gruesas que las demás dentro del mismo carril ocasiona una dispersión de los datos alrededor de su valor medio. Sin embargo, como se puede ver, en ninguno de los casos el error supera 1 píxel.

Es por lo mencionado anteriormente que se recurre a un segundo conjunto de datos, pero en este caso se generaron carriles con una sola banda, lo cual elimina la posibilidad de bandas más gruesas dentro del mismo carril. Los resultados obtenidos se resumen en la tabla 4.3, mientras en la tabla 4.4 se presenta el análisis estadístico.

Tabla 4.3: Desviación estándar obtenida para carriles con 1 banda

Medición \ σ	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
1	1,323	1,771	2,218	2,656	3,104	3,552	4,218	4,437	5,323	5,552
2	1,323	1,771	2,218	2,604	3,104	3,552	4,218	4,565	4,885	5,104
3	1,323	1,750	1,989	2,625	3,104	3,771	4,218	4,437	5,323	5,771
4	1,323	1,717	2,167	2,656	3,104	3,771	3,989	4,885	4,885	5,323
5	1,323	1,771	2,218	2,656	2,104	3,552	3,989	4,656	4,885	5,771

Tabla 4.4: Medidas estadística para la determinación de la desviación estándar obtenida para carriles con 1 banda

σ (píxeles)	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
$\bar{\sigma}$	1,323	1,752	2,162	2,639	3,104	3,639	4,126	4,596	5,060	5,504
SD (píxeles)	0	0,21	0,089	0,021	0	0,107	0,112	0,166	0,214	0,259
CV (%)	0	1,203	4,104	0,810	0	2,948	2,718	3,623	4,240	4,721
Error (píxeles)	0,323	0,252	0,162	0,139	0,104	0,139	0,126	0,096	0,060	0,004

Como se puede ver en la tabla 4.4 para este caso la dispersión de los datos es muy baja o nula del todo, siendo el mayor de ellos de 4.721. Además el error en la determinación del σ

es significativamente menor que para el caso determinado con 20 bandas lo cual comprueba que la mayor parte del error se debe al caso mencionado de bandas más gruesas en el carril.

En la figura 4.2 se presenta una comparación del error para el caso de 1 y 20 bandas.

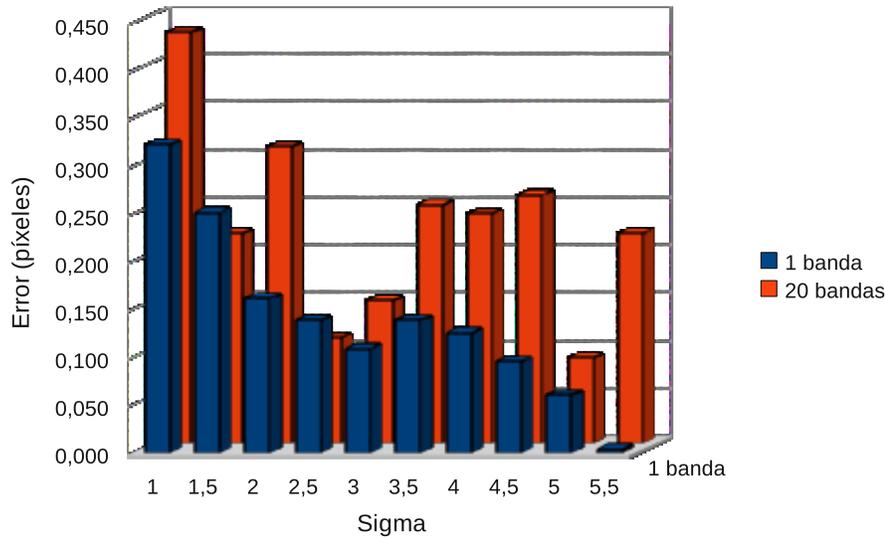


Figura 4.2: Comparación del error en la determinación de la desviación estándar para el caso de 1 y 20 bandas

Como se aprecia claramente para el caso del carril con 20 bandas (barras rojas) el error en la estimación de la desviación estándar (σ) es mucho mayor que para el caso de una sola banda (barras azules). Además se nota cómo el error obtenido es menor para valores mayores de σ ; sin embargo, se puede ver cómo el error máximo obtenido para el caso del carril con una sola banda fue de 0.323 para el caso de un $\sigma = 1$ lo cual se encuentra muy por debajo de la meta planteada de 1 píxel como máximo.

4.2 Determinación de la cantidad de bandas presentes en el carril

En esta sección se analizan los resultados obtenidos en cuanto a la determinación de la cantidad de bandas presentes en el carril. Para ello se utiliza el mismo conjunto de datos analizados en la sección 4.1.

Como se mencionó anteriormente, aunque se están generando carriles con un total de 20 bandas, en los casos de muy poca separación entre ellas estas se presentan como un única banda, por lo que en estos casos el número de bandas identificado es menor al esperado. Es por ello que se realizan dos tablas: una para los datos experimentales (4.5) y otra para los datos teóricos (4.6). Esta última contiene el número de bandas identificados en el carril mediante una inspección visual.

Como se puede ver en la tabla 4.5 solo en dos de los casos se identificó un total de 20 bandas.

Tabla 4.5: Valores experimentales de la determinación de la cantidad de bandas en el carril

Medición \ σ teórico	σ teórico									
	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
1	18	18	15	17	14	12	15	13	15	15
2	20	18	16	15	16	17	20	16	15	17
3	18	17	18	19	16	15	13	16	17	13
4	18	17	16	17	16	17	16	16	14	17
5	16	18	16	17	15	14	14	18	15	14

Tabla 4.6: Valores teóricos de la determinación de la cantidad de bandas en el carril

Medición \ σ teórico	σ teórico									
	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
1	18	18	15	17	14	12	15	13	14	15
2	20	18	16	15	16	17	20	16	15	17
3	18	17	18	19	16	15	13	16	17	13
4	18	17	16	17	16	17	16	16	14	17
5	18	18	16	17	16	14	14	18	15	14

Sin embargo, al comparar estos resultados con los valores teóricos en la tabla 4.6 se puede ver cómo solo en uno de los casos el número de bandas determinado por el algoritmo fue diferente al valor teórico. Esto se debe a que dos bandas en el carril se encuentran demasiado juntas.

En la figura 4.3 se presenta la imagen del carril para el cual se determinó un número menor de bandas al que realmente existe.

**Figura 4.3:** Ejemplo de poca separación entre bandas

Como se puede ver, los dos casos en que no se detectó la presencia de una banda se dio para bandas con poca separación entre ellas. Si bien a simple vista es fácil identificar la existencia de las dos bandas, en el caso del algoritmo, cuando dos bandas se encuentran demasiado juntas da como resultado que al momento de generar el espacio de escalas estas dos se fusionan en una sola, dando como resultado que el algoritmo solo identifique a una de las dos. En la figura 4.4 se presenta un ejemplo de no detección de bandas por poca separación y en la figura 4.5 se presenta el espacio de escalas correspondiente.

Como se puede ver en la zona señalada en el caso en que dos bandas se encuentran muy



Figura 4.4: Ejemplo de no detección de bandas por poca separación

cercanas una de la otra el algoritmo detectó únicamente una de las dos. En la figura 4.5 se indica la zona correspondiente a las bandas señaladas en la figura 4.4. En ella se señala el punto donde se detecta únicamente una de las dos bandas, el cual se da después de la zona marcada como unión de bandas, que es la región donde se unen los espacios de escalas correspondientes a ambas bandas. Es por ello que solo una de ellas fue detectada, ya que el algoritmo localiza el máximo dentro de la zona correspondiente a la banda dentro del espacio de escalas, y en este caso este punto se encuentra dentro de la zona de unión por lo que el algoritmo ve una única banda.

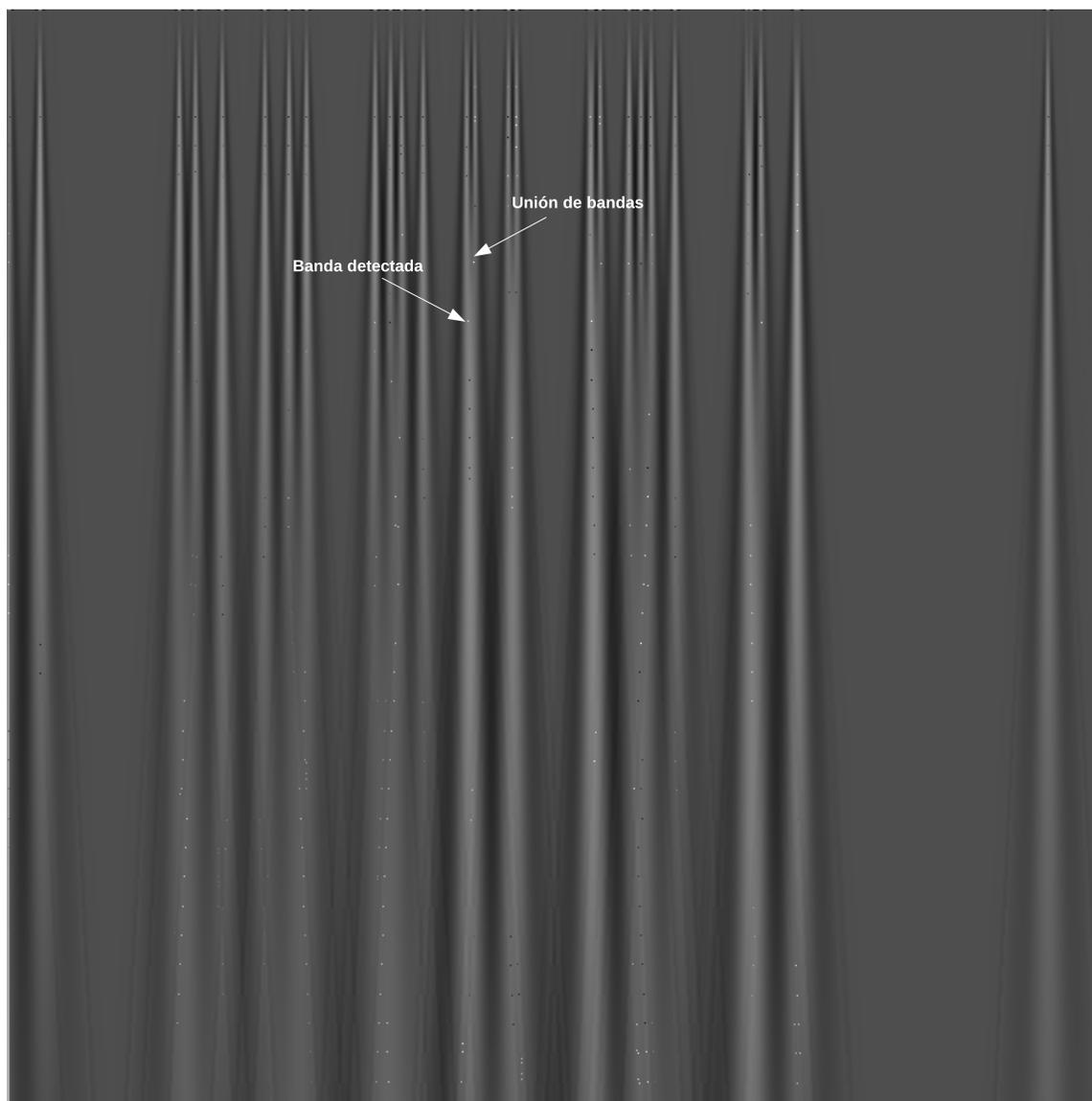


Figura 4.5: Espacio de escalas para la figura 4.4

4.3 Determinación de la posición de las bandas en el carril

En esta sección se analizan los resultados obtenidos en la detección de la posición de las bandas dentro del carril. Primero se analiza con un único carril donde se presentan los resultados en la determinación de la posición de cada una de las bandas dentro del carril para posteriormente analizar un conjunto de datos para determinar el porcentaje de error en la detección de la presencia y ausencia de bandas. En este caso el error máximo esperado es del 10 % de la cantidad de bandas en el carril.

En las tablas 4.7 y 4.8 se presentan los datos teóricos de la posición para un carril sintético generado con 20 bandas y una desviación estándar de 2 y en la figura 4.6 se presenta el carril generado.

Tabla 4.7: Datos teóricos de la determinación de la posición (parte 1)

Banda	1	2	3	4	5	6	7	8	9	10
Posición	9,6	32,3	45,5	56,2	79,2	85,6	123,9	215,3	219,1	375,3

Tabla 4.8: Datos teóricos de la determinación de la posición (parte 2)

Banda	11	12	13	14	15	16	17	18	19	20
Posición	391,1	400,2	467,3	505,0	512,8	540,9	587,8	720,8	762,4	787,3



Figura 4.6: Carril generado para la determinación la evaluación de l a determinación de la posición

Como se puede ver, aunque se generaron 20 carriles en la imagen solo se logran identificar 18, esto debido a la cercanía que existe entre algunas de las bandas, como por ejemplo la número 5 y 6, que estén separadas por 6,4 píxeles y en la imagen aparecen como una sola.

En las tablas 4.9 y 4.10 se resumen los resultados obtenidos para las 18 bandas identificables y en la figura 4.7 se presenta una gráfica del error en la determinación de la posición.

Como se puede ver el máximo error en la determinación de la posición se dio para la banda número 7 la cual fue identificada en la posición 217. Este error se debe a que como se ve en 4.7 las bandas 8 y 9 se encuentran en las posiciones 215,3 y 219,1 respectivamente, con una separación entre ellas de apenas 3,8 píxeles apareciendo en la imagen como una única banda la cual fue detectada por el algoritmo en la posición 217, o sea, en medio de las dos bandas generadas.

En la tabla 4.11 se presentan los resultados teóricos de la cantidad de bandas detectadas en el carril, mientras en la figura 4.12 se presentan los resultados obtenidos en la determinación

Tabla 4.9: Datos experimentales de la determinación de la posición (parte 1)

Banda	1	2	3	4	5	6	7	8	9
Posición	10	32	45	56	80	124	217	375	391

Tabla 4.10: Datos experimentales de la determinación de la posición (parte 2)

Banda	10	11	12	13	14	15	16	17	18
Posición	401	467	505	513	541	588	721	762	787

de la posición de las bandas en el carril para un nuevo conjunto de muestras generadas con una desviación estándar que va de 1 a 5,5 con un incremento de 0,5 para un total de 20 bandas.

Como se puede ver del conjunto de datos analizado solamente en dos ocasiones se detectó un número de bandas menor al real, esto por razones ya analizadas anteriormente.

En los dos casos en que se falló lo hizo por un máximo de 2 bandas lo cual representa un error máximo del 11,11 %, sin embargo si se calcula un promedio del error en la totalidad de las mediciones realizadas el valor obtenido es de 0,36 % lo cual es significativamente menor al máximo deseado de 10 %. En la tabla 4.13 se resumen estos resultados.

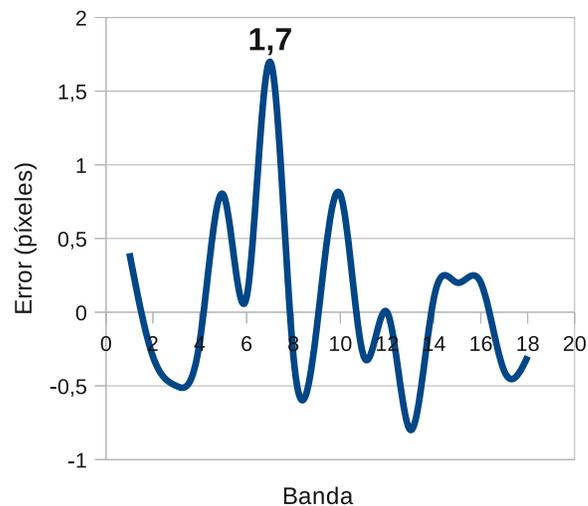
**Figura 4.7:** Error en la determinación de la posición

Tabla 4.11: Valores determinados por un observador humano de la determinación de la cantidad de bandas en el carril para la determinación del error en la detección

Medición \ σ	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
1	18	18	17	17	16	18	16	17	13	16
2	18	16	18	17	17	15	17	15	15	19
3	20	18	16	16	14	15	15	17	13	18
4	20	18	18	18	17	13	15	15	17	11
5	19	20	17	17	12	16	18	16	15	14

Tabla 4.12: Valores experimentales de la determinación de la cantidad de bandas en el carril para la determinación del error en la detección

Medición \ σ	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5
1	18	18	17	17	16	18	16	17	13	16
2	18	16	18	17	17	15	17	15	15	19
3	20	18	16	16	14	14	15	17	13	18
4	20	16	18	18	17	13	15	15	17	11
5	19	20	17	17	12	16	18	16	15	14

4.4 Comparación con el algoritmo previo

En esta sección se presenta una comparación de los resultados obtenidos en la determinación de σ con los presentados en [9] y por último se presenta el rendimiento del algoritmo en un caso con un carril real.

En las tablas 4.14 y 4.15 se resume la comparación de los datos obtenidos. El conjunto de datos denominado como (a) corresponde a los resultados en [9], mientras el conjunto denominado (b) corresponde a los datos aquí obtenidos.

Tabla 4.13: Error en la detección de bandas

Total de mediciones	50
Error máximo	11,11 %
Error mínimo	0 %
Error promedio	0,36 %

En el caso de la tabla 4.14 se comparan los resultados para el caso de carriles con 20 bandas, mientras en 4.15 se presenta el caso de una única banda.

Tabla 4.14: Comparación de los resultados en la determinación de σ

σ \ Proyecto	a				b			
	$\bar{\sigma}$	SD	CV (%)	E (píxeles)	$\bar{\sigma}$	SD	CV (%)	E (píxeles)
1	0,88	0,19	22,11	0,12	1,43	0,15	10,37	0,43
1,5	1,2	0,45	37,45	0,3	1,72	0,06	3,45	0,22
2	0,74	0,74	61,16	1,23	2,31	0,07	3,11	0,31

Como se puede ver en 4.14 para el conjunto de datos (a) se obtiene un mejor resultado para valores de σ menores a 2 píxeles, mientras en (b) se obtiene un mejor resultado para el caso de 2 píxeles, además en este último el error nunca sobrepasa 1 píxel, mientras en (a) se tiene un error de 1,23 píxeles para el caso de $\sigma = 2$. Otro punto a destacar es que el coeficiente de dispersión obtenido en (b) es significativamente menor al obtenido en (a), siendo el mayor de ellos (10,37) menos de la mitad del menor en (a) (22,11) lo cual indica que los resultados obtenidos son mucho más consistentes.

Por último en la tabla 4.15 se presenta una comparación de los datos con el caso analizado de 1 banda. En esta se ratifica una mayor exactitud en la determinación de σ .

En la figura 4.8 se presenta un ejemplo de un carril real y en la tabla 4.16 se resumen los resultados obtenidos al analizar dicho carril.



Figura 4.8: Carril utilizado para el análisis de un caso real

Como se puede ver la imagen esta dividida en dos secciones. Una con la imagen sin filtrar y la otra después de pasar por el filtro desarrollado en [9].

En las figuras 4.9 y 4.10 se pueden ver los resultados tras analizar la imagen sin y con el filtro respectivamente.

Tabla 4.15: Comparación de los resultados en la determinación de σ para el caso de 1 banda

σ \ Proyecto	a				b			
	$\bar{\sigma}$	SD	CV (%)	E (píxeles)	$\bar{\sigma}$	SD	CV (%)	E (píxeles)
1	0,88	0,19	22,11	0,12	1,323	0	0	0,323
1,5	1,2	0,45	37,45	0,3	1,752	0,21	1,203	0,252
2	0,74	0,74	61,16	1,23	2,162	0,089	4,104	0,162



Figura 4.9: Resultado de analizar un carril real sin filtrar

Tabla 4.16: Resultados del análisis de un carril real

Banda	Posición	σ
1	13	5,10
2	407	3,10
3	501	4,88
4	623	1,55
5	653	2,43
6	835	2,22
7	893	2,02
8	931	2,44

Como se puede ver, para el caso sin filtrar se detectaron un total de 8 bandas con un σ que va desde 1,55 hasta 5,10 píxeles con un promedio de 2,97 píxeles lo cual demuestra como el algoritmo determina la desviación estándar para cada uno de las bandas de manera independiente.

Por último la figura 4.10 muestra los resultados obtenidos tras analizar la imagen filtrada. Para este caso se detectaron un total de 37 bandas con un σ que va desde 0,85 hasta 1,55, con un promedio de 1,22 píxeles lo cual indica que el comportamiento del algoritmo es mejor cuando el ruido es menor.

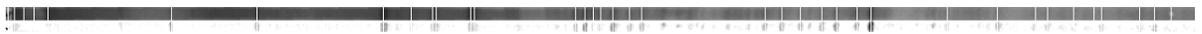


Figura 4.10: Resultado de analizar un carril real filtrado

Este comportamiento se debe a que, como se puede apreciar, en el caso de la imagen sin filtrar el tono de las bandas tiende a confundirse con el del fondo del carril, lo cual ocasiona que al momento de generar el espacio de escalas no se detecte la diferencia entre fondo y banda, por lo que no se generan los máximos buscados para detectar las bandas, mientras en el caso de la imagen filtrada, al eliminar el fondo, se nota una diferencia más marcada entre banda y fondo lo que ocasiona que en este caso si se generen los máximos mencionados.

Capítulo 5

Conclusiones y Recomendaciones

5.1 Conclusiones

Con la realización del presente proyecto se comprobó que mediante la utilización de la técnica de espacio de escalas y detección de bordes es posible estimar de forma independiente el ancho de cada una de las bandas que conforman el carril, ubicando los valores máximos dentro de la imagen generada y mediante la utilización de la segunda derivada es posible la localización de esos máximos. Además ubicando estos mismos puntos es posible conocer la ubicación exacta de cada una de las bandas que conforman el carril y por lo tanto, sabiendo esto, es posible determinar la cantidad de bandas que lo conforman. Sin embargo, factores como la cercanía entre bandas influyen en la correcta determinación de su posición, dando como posición, el punto central o la ubicación de una de las dos bandas cercanas.

Como se comprobó, el aproximar el perfil de intensidades de la banda mediante un modelo rectangular resulta suficiente para su identificación y la ubicación del máximo dentro de ésta depende de su ancho. En el caso en que dos bandas se unen aparentando una solo el resultado obtenido es la detección de una banda cuyo ancho es la unión de los ancho de las dos bandas.

5.2 Recomendaciones

- Es necesaria la optimización el algoritmo para que logre discernir cuándo un ancho detectado corresponde a la unión de dos bandas, para que de esta manera se logre obtener un mejor resultado en la estimación de la desviación estándar.
- Disminuyendo el incremento utilizado para la generación del espacio de escalas se lograría obtener una mayor exactitud a la hora de estimar el ancho de las bandas; sin embargo, se debe considerar una forma de optimizar el rendimiento del algoritmo ya que al disminuir el incremento se aumenta el tiempo requerido para realizar el análisis.

Bibliografía

- [1] P. Alvarado M. Análisis por computadora de imágenes de geles de electroforesis para caracterización molecular de organismos. Reporte técnico, Instituto Tecnológico de Costa Rica, 2010.
- [2] P. Alvarado M. *Señales y sistemas. Fundamentos Matemáticos*. Centro de Desarrollo de Material Bibliográfico, ITCR 2008.
- [3] A. Berk and H. Lodish. *Biología Celular Y Molecular/ Molecular and Cellular Biology*. Editorial Médica Panamericana, 2005.
- [4] Jähne. Bernd. *Digital Image Processing*. Springer-Verlang, Germany, 6th edition, 2005.
- [5] Univeridad Javeriana. Facultad de Neurobioquímica. Electroforesis. URL <http://javeriana.edu.co>.
- [6] Langer. Michael. Scale space 1 (normalized derivatives, blobs). In *Fundamentals of Computer Vision*, page 6. oct 2009. URL <http://www.cim.mcgill.ca/~langer/558/2009/558.html>.
- [7] Lindeberg. Tony. *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*. PhD thesis, Computational Vision and Active Perception Laboratory (CVAP), Royal Institute of Technology, S-100 44 Stockholm, Sweden, May 1991. URL <http://ftp.nada.kth.se/pub/CVAP/reports/>.
- [8] Lindeberg. Tony. Scale-space. In *Encyclopedia of Computer Science and Engineering*, pages 2495–2504. SE-100 44 Stockholm, Sweden, sep 2008. URL <http://www.csc.kth.se/~tony/abstracts/Lin08-EncCompSci.html>.
- [9] David Soto Vásquez. Detección automática de bandas en imágenes de geles de electroforesis por medio de la optimización de una función objetivo. Tesis de licenciatura, Escuela de Ingeniería en Electrónica, ITCR, Cartago, Junio 2010.

