

GELES DE ADN: ADQUISICIÓN Y ANÁLISIS MEDIANTE TÉCNICAS DE
PROCESAMIENTO DIGITAL DE IMÁGENES

MIGUEL ANGEL SOTAQUIRÁ GUTIÉRREZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
MAESTRÍA EN INGENIERÍA ÁREA INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2005

GELES DE ADN: ADQUISICIÓN Y ANÁLISIS MEDIANTE TÉCNICAS DE
PROCESAMIENTO DIGITAL DE IMÁGENES

MIGUEL ANGEL SOTAQUIRÁ GUTIÉRREZ

Trabajo de Investigación para optar al título
de Magister en Ingeniería área Ingeniería Electrónica

Director: Arturo Plata Gómez, Doctor en Ciencias para Ingeniería
Codirectores: Dra. Adriana Castillo Pico; MPE Jaime Guillermo Barrero

UNIVERSIDAD INDUSTRIAL DE SANTANDER
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
MAESTRÍA EN INGENIERÍA ÁREA INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2005

AGRADECIMIENTOS

El autor expresa sus agradecimientos a:

- Los profesores Arturo Plata Gómez (Escuela de Física), Adriana Castillo Pico (Laboratorio de Genética, Facultad de Salud) y Jaime Guillermo Barrero Pérez (Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones), por su orientación durante el desarrollo del presente trabajo.
- Los integrantes de los grupos de Óptica y Tratamiento de Señales, de la Escuela de Física en la Universidad Industrial de Santander, por sus contribuciones durante esta etapa de mi formación académica.
- Al personal del Laboratorio de Genética, de la Facultad de Salud – UIS, por el apoyo oportuno y adecuado, necesario en la elaboración de este trabajo de investigación.

CONTENIDO

	pág.
INTRODUCCION	12
1. EL PROCESO DE ELECTROFORESIS Y LOS GELES DE ADN	15
1.1 LA ELECTROFORESIS DE MOLÉCULAS DE ADN	15
1.2 GELES DE ADN	17
1.2.1 Utilidad de los geles de ADN	19
1.2.2 Estimación del tamaño de los fragmentos presentes en un gel de ADN	20
1.3 PROCEDIMIENTO LLEVADO A CABO EN EL LABORATORIO DE GENÉTICA	23
2. MÉTODOS EXISTENTES PARA EL PROCESAMIENTO DE GELES DE ADN	24
2.1 TRABAJOS EN EL CAMPO ACADÉMICO	24
2.1.1 Segmentación de carriles	24
2.1.1.1 Análisis bidimensional	25
2.1.1.2 Análisis unidimensional	26
2.1.2 Detección de bandas	27
2.2 SISTEMAS COMERCIALES	28
2.2.1 Principales características	28
2.2.2 Costo	30
3. ALGORITMO PARA EL ANÁLISIS DE GELES DE ADN, BASADO EN TÉCNICAS DE PROCESAMIENTO DIGITAL DE SEÑALES E IMÁGENES	31
3.1 SEGMENTACIÓN DE CARRILES	31
3.1.1 Cálculo de la orientación de la imagen	32
3.1.2 Los mapas de distancia	36
3.1.3 Cálculo de la proyección de la imagen sobre el eje x'	39
3.1.4 Suavizado del patrón de proyección	41

3.1.5 Corrección de la tendencia de la iluminación en la proyección	46
3.1.6 Estimación de los mínimos locales	47
3.2 DETERMINACIÓN DE LA UBICACIÓN DE LAS BANDAS	48
3.2.1 Obtención de la coordenada y' de los máximos locales en la proyección del carril	49
4. IMPLEMENTACIÓN DEL ALGORITMO E INTERFAZ GRÁFICA	50
4.1 MÓDULO "LECTURA/ADQUISICIÓN"	51
4.2 MÓDULO "EDITAR IMAGEN"	51
4.3 MÓDULO "SEGMENTAR CARRILES"	52
4.4 MÓDULO "DETECTAR BANDAS"	52
4.5 MÓDULO "CÁLCULO DE TAMAÑOS"	52
5. RESULTADOS EXPERIMENTALES	53
5.1 SEGMENTACIÓN DE CARRILES	53
5.2 DETECCIÓN DE BANDAS	55
5.2.1 Resultados obtenidos sobre un grupo de nueve imágenes	56
5.2.2 Resultados obtenidos en una imagen específica	57
5.3 FUENTES DE ERROR	59
5.3.1 Exceso de bandas	61
5.3.2 Eliminación de algunas bandas existentes	61
5.3.3 Errores en la estimación del tamaño de las bandas	62
6. CONCLUSIONES Y OBSERVACIONES	65
BIBLIOGRAFÍA	67

LISTA DE TABLAS

	pág.
Tabla 1. Tamaño de algunos fragmentos obtenidos tras la digestión de ADN con algunas enzimas de restricción	20
Tabla 2. Resultados de la segmentación de carriles tras la aplicación del algoritmo planteado	54
Tabla 3. Resultados de la detección de bandas tras la aplicación del algoritmo planteado (método 1) y su comparación con los resultados obtenidos mediante el uso del algoritmo convencional (método 2)	57
Tabla 4. Parámetros de desempeño evaluados para una imagen con diferentes orientaciones, usando dos métodos diferentes	58

LISTA DE FIGURAS

	Pág.
Figura 1. Estructura de la molécula de ADN	15
Figura 2. Molécula de ADN tras ser digerida por la bacteria	16
Figura 3. Ilustración del proceso de electroforesis (considerando la interacción entre el campo eléctrico y la molécula de ADN únicamente)	17
Figura 4. Imagen obtenida tras aplicar el proceso de electroforesis a un gel con una muestra de ADN	18
Figura 5. Un gel de ADN típico	20
Figura 6. El gel de ADN y sus correspondientes ejes coordenados	22
Figura 7. Ilustración del proceso involucrado en la obtención de la curva Tamaño vs. Posición	22
Figura 8. Ilustración del concepto de "frontera" entre uno y otro carril	24
Figura 9. Obtención de las zonas límite intercarril tras el procesado de la imagen del gel	25
Figura 10. Proyección de la imagen original sobre el eje x	26
Figura 11. Un carril y su proyección	27
Figura 12. Ubicación de los límites intercarril a) Asumiendo carriles verticales; b) Teniendo en cuenta la correcta orientación de la imagen	32
Figura 13. Un gel de ADN cuyos carriles no están alineados verticalmente (a) y su correspondiente imagen de autocorrelación (b)	33
Figura 14. Los nuevos ejes x', y' y su relación con los ejes x, y tradicionales	34
Figura 15. Algoritmo usado para determinar la orientación de la imagen	35
Figura 16. Los dos marcadores utilizados en la obtención de los mapas de distancia: a) Marcador paralelo al eje x' ; b) Marcador paralelo al eje y'	36
Figura 17. Ejemplo para una imagen binaria con un marcador (tamaño 5x3)	37
Figura 18. La imagen binaria y su matriz <i>Grilla</i> correspondiente	37
Figura 19. Matriz A cuyas filas corresponden a las coordenadas de los pixeles pertenecientes al marcador	37
Figura 20. Matriz B , resultado de evaluar la diferencia entre la matriz <i>Grilla</i> y la matriz A	38
Figura 21. Mapa de distancias obtenido para el ejemplo de una matriz de 5x3	38
Figura 22. Algoritmo utilizado en el cálculo de la proyección de la imagen sobre el eje x'	39
Figura 23. Ilustración del algoritmo usado para determinar la proyección de la imagen sobre el eje x'	40
Figura 24. Proyección sobre el eje x' de un gel de ADN típico	41
Figura 25. Resultado de aplicar el filtrado por la media a la señal original	42
Figura 26. Resultado de aplicar el filtrado morfológico a la señal original	43
Figura 27. Ilustración del proceso de difusión no lineal	44
Figura 28. Proceso de difusión lineal (modificado) aplicado a la proyección sobre el eje x'	45
Figura 29. Ilustración del procedimiento que permite corregir la tendencia en la iluminación de la proyección	46
Figura 30. Los diferentes mínimos locales presentes en la proyección están determinados por el cambio en el signo de la pendiente (negativo/positivo)	47

Figura 31. El algoritmo utilizado en la segmentación de carriles	48
Figura 32. Un carril y su proyección sobre el eje y'	48
Figura 33. Diagrama de bloques del sistema implementado en Matlab	51
Figura 34. Una banda detectada correctamente (izquierda) y otras detectadas de forma incorrecta (derecha)	56
Figura 35. Desempeño de los métodos 1 y 2 para una imagen específica, en términos del porcentaje de aciertos	58
Figura 36. Una imagen cuyos carriles no son completamente verticales, y su correspondiente proyección sobre el eje x (izquierda) y sobre el eje x' (derecha)	59
Figura 37. Nueva segmentación de carriles, completamente verticales, para corregir la ausencia de algunos	60
Figura 38. Ilustración del exceso de bandas debido a la incorrecta segmentación de los carriles	61
Figura 39. Los carriles 2 y 3 (izquierda) y el carril 2 con su respectiva proyección sobre el eje "y" (derecha)	62
Figura 40. Estimación del tamaño de dos fragmentos de ADN asumiendo carriles y bandas completamente verticales y horizontales (a), y teniendo en cuenta la orientación de la imagen (b)	64

RESUMEN

TÍTULO: GELES DE ADN: ADQUISICIÓN Y ANÁLISIS MEDIANTE TÉCNICAS DE PROCESAMIENTO DIGITAL DE IMÁGENES*.

AUTOR: MIGUEL ÁNGEL SOTAQUIRÁ GUTIÉRREZ**

PALABRAS CLAVES: Geles de ADN, electroforesis, mapas de distancia, autocorrelación, ejes principales, difusión no lineal, procesamiento digital de imágenes.

DESCRIPCIÓN:

El análisis de geles de ADN es de gran importancia en estudios de genética poblacional y en el diagnóstico de enfermedades de carácter genético. El objetivo principal de dicho análisis consiste en la estimación del tamaño (medido en pares de bases) de todos o algunos fragmentos presentes en el gel, procedimiento que hasta el momento es llevado a cabo de forma manual en el Laboratorio de Genética de la Universidad Industrial de Santander.

El presente trabajo describe un método, basado en técnicas de procesamiento digital de imágenes, que permite segmentar la imagen digital correspondiente y detectar las bandas (o fragmentos de ADN) presentes en la misma. La segmentación involucra etapas como la determinación de los ejes principales de la imagen (a partir del cálculo de la imagen de autocorrelación), el cálculo de mapas de distancia euclidiana que permiten reducir la complejidad de los cálculos (de dos a una dimensión) y así determinar la ubicación de los carriles y las bandas. Dicho cálculo involucra el suavizado de una señal unidimensional (la proyección) por medio de algoritmos de difusión lineal, así como el cálculo de gradientes para determinar la ubicación de máximos y mínimos locales. Finalmente se describe un algoritmo que permite determinar el tamaño de cada fragmento de ADN presente en el gel, con base en el marcador de peso molecular y en la posición de cada una de las bandas presentes en la imagen.

El método descrito ofrece un desempeño adecuado en la etapa de segmentación (superior, incluso, en ocasiones al ofrecido por algunos sistemas de carácter comercial). La detección de bandas es igualmente un procedimiento completamente nuevo, mucho más robusto y preciso en comparación trabajos anteriores. En este caso, la determinación de los ejes principales de la imagen conduce a mejores resultados en términos de cantidad de bandas detectadas.

* Trabajo de Investigación para optar al título de Magister en Ingeniería (área Ingeniería Electrónica)

** Facultad de Ingenierías Físico-mecánicas. Maestría en Ingeniería (área Ingeniería Electrónica). Director: Dr. Arturo Plata Gómez (Escuela de Física).

ABSTRACT

TÍTULO: ON THE USE OF DIGITAL IMAGE PROCESSING TECHNIQUES FOR IMAGE ACQUIRING AND ANALYSIS OF DNA GELS *.

AUTHOR: MIGUEL ÁNGEL SOTAQUIRÁ GUTIÉRREZ**

KEYWORDS: DNA gels, electrophoresis, euclidean distance maps, autocorrelation, principal components, non-linear diffusion, digital image processing.

DESCRIPTION:

DNA gels analysis plays an important role in genetic of populations and genetic disorders diagnosis studies. The main purpose of this procedure is to determine the size (measured in base pairs) of every fragment, or at least some of them, of DNA present in the gel, process that, until now, is carried without any computational procedure in the Genetics Laboratory of the Universidad Industrial de Santander

This document describes a procedure, based on digital image processing techniques, which leads to image segmentation and bands (DNA fragments) detection. Segmentation procedure involves different stages like estimation of image's main axes (by means of autocorrelation estimation), and also the reduction in complexity of analysis (by using euclidean distance maps, which leads to 1D analysis, instead of 2D analysis). With the estimation of image's main axes is possible to detect lanes and bands exact position. This involves 1D signal filtering, by means of linear diffusion techniques, and also requires gradient calculus to obtain local maxima and minima positions. Finally, an algorithm (based on molecular weight marker and coordinate estimation of every band in the gel) that leads to obtain DNA fragments size is described.

In terms of segmentation, the method described in this work is even better than some systems offered in the market. Bands detection is also a procedure completely new, more robust and precise compared to any other work found until now. In this case, image's main axes estimation leads to better results in terms of number of bands detected.

* Trabajo de Investigación para optar al título de Magister en Ingeniería (área Ingeniería Electrónica)

** Facultad de Ingenierías Físico-mecánicas. Maestría en Ingeniería (área Ingeniería Electrónica). Director: Dr. Arturo Plata Gómez (Escuela de Física).

INTRODUCCION

Sin duda alguna el descubrimiento de la estructura de la molécula de ADN hacia el año 1953^{[19][18]}, y la inmensa cantidad de implicaciones que este hito marca en la historia de la biología, la genética y la misma historia humana (junto con la aparición de nuevas disciplinas del conocimiento, o la fusión de otras, tales como la bioinformática o la biomedicina)^[6], hace que día a día los desarrollos en los campos teórico y práctico en dichas áreas cobren cada vez mayor importancia. Así, tras la obtención del genoma humano, se plantea ahora, con la ingeniería genética, la posibilidad de crear medicinas "a la medida del individuo" donde pequeños segmentos de ADN pueden ser reemplazados o reparados para de esta manera alcanzar un correcto funcionamiento del organismo; igualmente, conceptos aún más complejos (y que por obvias razones han despertado preocupaciones de tipo ético) como la clonación o el papel que pueden jugar las empresas farmacéuticas en la manipulación del genoma humano, están a la orden del día. Paralelamente, se adelantan estudios menos polémicos pero igualmente interesantes: la comprensión de la manera como el código genético hace que el organismo se comporte de una u otra manera frente a un virus (como el VIH), o el origen de algunas enfermedades coronarias, e igualmente lo que ocurre dentro de las células cancerosas desde el punto de vista del análisis del ADN. Son cuestiones que desde hace mucho tiempo han sido una incógnita, y que solo hasta ahora comienzan a develarse.

Se puede pensar que solo aquellos países con los suficientes recursos económicos y tecnológicos pueden tener acceso a aplicaciones de este tipo. Sin embargo, es interesante notar que naciones en desarrollo como Colombia han tenido acceso a buena parte de esa tecnología, y que en la actualidad se adelantan trabajos de carácter investigativo orientados al análisis de moléculas de ADN. El Grupo de Investigación en Genética Humana de la Universidad Industrial de Santander no es la excepción y prueba de ello son los trabajos de investigación llevados a cabo en la actualidad, orientados al análisis de la epidemiología de las enfermedades cardiovasculares junto con investigaciones acerca de la genética del cáncer de colon y gástrico, además de otros trabajos previos relacionados con el área de genética poblacional. Dichos trabajos se llevan a cabo en el Laboratorio de Genética de la Facultad de Salud, donde se cuenta con equipo de última tecnología que permite el procesamiento y análisis de muestras de ADN así como el análisis de cromosomas (citogenética), entre otros.

Entre las técnicas utilizadas para el análisis de moléculas de ADN, se encuentra el proceso de electroforesis, donde una molécula de ADN (previamente fragmentada en trozos de diferentes tamaños) es expuesta a un campo eléctrico que permite la organización de los diferentes fragmentos de acuerdo a su tamaño*. El resultado es almacenado en papel fotográfico.

* En el capítulo 1 se describen todos los conceptos relacionados con el proceso de electroforesis y con el procedimiento llevado a cabo en el Laboratorio de Genética.

El presente trabajo describe el desarrollo de una aplicación que permite automatizar este proceso, llevado a cabo en la actualidad en el Laboratorio de Genética, el cual es completamente manual y presenta desventajas como:

- La exposición del operario a la radiación UV.
- El tiempo requerido para el análisis de la imagen obtenida, el cual debe ser menor a 5 minutos para reducir los efectos de la radiación sobre el personal del Laboratorio. Sumado a ello, la calidad (nitidez) de la imagen obtenida se deteriora en pocos minutos de exposición a la radiación, así como el gel de soporte, todo ello en un lapso de aproximadamente dos horas. Transcurrido este período de tiempo se presenta difusión del material (fragmentos de ADN) separado por medio de la electroforesis, lo cual dificulta su posterior estudio. Se concluye entonces que el análisis de las muestras debe ser inmediato.
- El almacenamiento de los resultados en un medio que no es el adecuado, como lo es el papel fotográfico mencionado con anterioridad. Dicho material se degrada con el tiempo junto con la calidad de la imagen obtenida, impidiendo o dificultando futuros análisis de las muestras obtenidas.

Desde luego, existen sistemas de carácter comercial que permiten el análisis de geles de ADN (algunos de los cuales incluyen la etapa de adquisición). Entre las principales limitantes de dichas aplicaciones se encuentran su costo y la imposibilidad de mejoras a futuro, puesto que se trata de sistemas completamente cerrados.

Este trabajo tiene como objetivo principal la implementación de una aplicación computacional encargada de la adquisición y el análisis semiautomático de las imágenes provenientes de geles de ADN, basada en técnicas de procesamiento digital de imágenes y señales que permitirán segmentar la imagen adquirida y estimar el tamaño de cada una de las bandas presentes en la misma. Para ello, se hará uso del equipo disponible en el Laboratorio de Genética (computador y cámara fotográfica). En la aplicación final se incluirán opciones para la mejora de la imagen adquirida (como manejo del histograma así como algoritmos para la atenuación del ruido).

Los resultados obtenidos tras la implementación del algoritmo son equivalentes y, en ocasiones, superiores a aquellos arrojados por sistemas de tipo comercial encargados del análisis de dichas moléculas de ADN. Se obtiene igualmente un mejor desempeño en comparación con trabajos de carácter académico realizados anteriormente.

El documento está estructurado de la siguiente manera:

En el capítulo 1 se describen los conceptos relacionados con la electroforesis y su utilidad en el análisis de moléculas de ADN. Se describe en detalle el procedimiento llevado a cabo en el Laboratorio de Genética.

El capítulo 2 hace referencia al estado del arte, que comprende los trabajos realizados hasta el momento (revisión bibliográfica) junto con la descripción de algunos sistemas comerciales existentes en la actualidad.

El capítulo 3 es el núcleo central del presente trabajo. En él se describen los algoritmos, basados en técnicas de procesamiento digital de imágenes y señales, tendientes a llevar a cabo la segmentación de la imagen, así como la posterior detección y estimación del tamaño de cada uno de los fragmentos presentes en el gel. En muchos casos se analizan las posibles aproximaciones a la solución del problema en

particular, se justifica la elección de la metodología descrita y se analiza en detalle la solución planteada.

El capítulo 4 describe brevemente la interfaz gráfica implementada, que incluye, además de la etapa de adquisición, los algoritmos diseñados para la segmentación de la imagen y la estimación del tamaño de las bandas, junto con opciones adicionales para la mejora de la imagen.

El capítulo 5 corresponde al análisis de resultados, donde se compara el desempeño de los algoritmos descritos en el capítulo 3 con el ofrecido por sistemas de carácter comercial o mediante el uso de técnicas descritas en trabajos anteriores. Se muestran las bondades del método implementado, en términos de precisión en la segmentación de la imagen así como en la detección de las bandas, o fragmentos, presentes en el gel.

Finalmente, en el capítulo 6 se enumeran los principales logros del trabajo realizado e igualmente se consignan algunas sugerencias o mejoras que a futuro se podrían implementar.

1. EL PROCESO DE ELECTROFORESIS Y LOS GELES DE ADN

En términos generales, la electroforesis es un proceso de separación de partículas de una mezcla con base en su carga eléctrica; el movimiento de dichas partículas se realiza por la acción de un campo eléctrico aplicado a través de un soporte que, para el problema de interés, es un gel. Para el caso de muestras biológicas, las que más comúnmente se separan por medio de esta técnica son las proteínas y el ADN. Estas muestras se ubican en un medio que favorezca el movimiento en presencia de un campo eléctrico. Dicho soporte es precisamente una sustancia coloidal o gelatinosa, que puede ser agarosa o poliacrilamida.

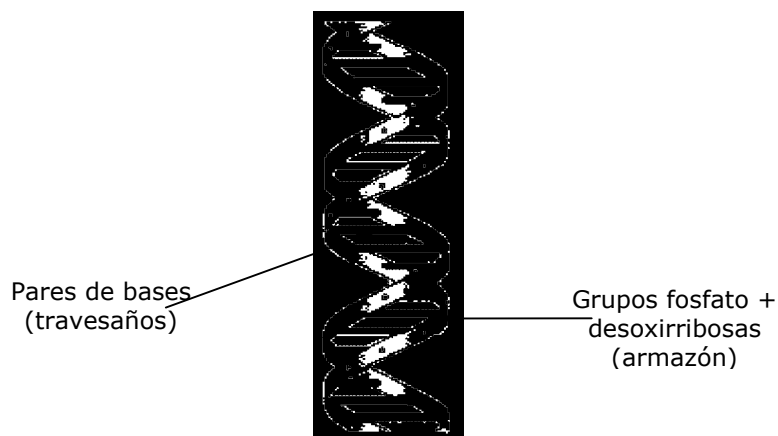
A continuación se describen detalladamente los procesos de electroforesis y la obtención de un gel con muestras de ADN.

1.1 LA ELECTROFORESIS DE MOLÉCULAS DE ADN

Todos los organismos vivos poseen cromosomas en el núcleo de sus células; dentro de esas pequeñas estructuras se encuentra una molécula conocida como ADN (Ácido Desoxirribonucleico), encargada de la síntesis de proteínas (tema del cual se hablará en la sección 1.2) y de la replicación de todas y cada una de las células del organismo. Es, por tanto, la molécula encargada de perpetuar la existencia de una especie determinada.

La estructura de una molécula de ADN se puede asemejar a la de una doble hélice (Figura 1). El armazón de la hélice está compuesto por grupos fosfato y glúcidos del tipo desoxirribosas. En la parte interna se ubican las bases nitrogenadas, agrupadas en pares de bases a manera de travesaño (en últimas se tienen cuatro tipos de bases: Adenina, Guanina, Timina y Citosina)^[7].

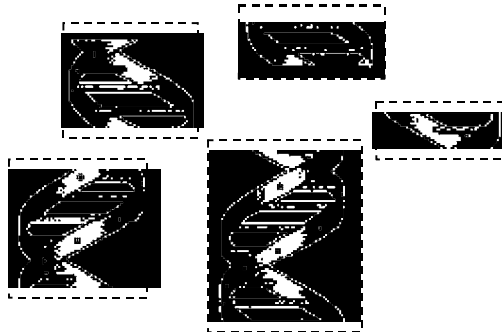
Figura 1. Estructura de la molécula de ADN



Cuando se tiene una molécula de ADN esta puede ser "digerida" o procesada por las enzimas producidas por una bacteria determinada (como la *E. Coli* y la *Hemophilus*

influenzae). Dicha bacteria genera cierto tipo de enzimas (llamadas enzimas de restricción), capaces de cortar la molécula de ADN en fragmentos pequeños y en lugares muy específicos de la cadena de ADN. Tras ese proceso de digestión se tiene entonces la molécula de ADN inicial (similar a la de la Figura 1) fragmentada ahora en segmentos de diferentes tamaños:

Figura 2. Molécula de ADN tras ser digerida por la bacteria (en este caso ha sido fragmentada en cinco partes). El tamaño se mide en cantidad de pares de bases

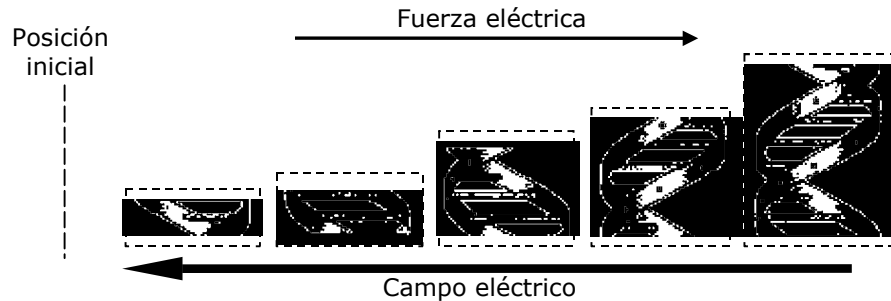


Estos fragmentos originados por la digestión enzimática se deben separar, para lo cual el siguiente paso es el proceso de electroforesis. Se debe tener en cuenta que la molécula de ADN (y por tanto cualquier segmento) posee una carga eléctrica negativa, dada por la presencia de los grupos fosfato. Si entonces a la molécula fragmentada se aplica un campo eléctrico, se observa que cada fragmento experimentará una fuerza eléctrica dada por la ecuación:

$$\vec{F} = q\vec{E} \quad [N] \quad (\text{Ecuación 1})$$

Debido a que la carga q es negativa para cualquiera de los segmentos, el vector fuerza apuntará en dirección contraria al campo eléctrico. La magnitud de dicho vector fuerza será mayor para los fragmentos más pesados (puesto que tendrán mayor masa y por tanto mayor carga eléctrica), en comparación con los elementos más livianos. Se espera entonces que los segmentos más pesados se desplacen una distancia mayor, con respecto a una posición inicial determinada, en comparación con los segmentos más livianos. En una situación ficticia, sin la presencia del soporte o gel, se tendría un resultado como el que se ilustra en la Figura 3:

Figura 3. Ilustración del proceso de electroforesis (considerando la interacción entre el campo eléctrico y la molécula de ADN únicamente)



En el proceso descrito anteriormente, se considera que se tiene una molécula de ADN "pura", por decirlo de alguna manera; esto es, no se encuentra presente ninguna otra sustancia diferente de los fragmentos de dicha molécula.

Se tiene entonces una organización de los fragmentos según su tamaño, de forma ascendente (si se observa la figura 3 de izquierda a derecha). Si bien se ha logrado el objetivo, se trata de una situación completamente ficticia (puesto que es necesario utilizar un soporte que facilite el desplazamiento de las moléculas de ADN) y, en este caso, el procedimiento presenta la desventaja de que segmentos con tamaños similares, mas no idénticos, tendrán cargas eléctricas similares y por tanto se desplazarán la misma distancia a partir de la posición inicial, por lo cual no podrán ser diferenciados en etapas posteriores a la electroforesis. Es esta una de las razones por las cuales los fragmentos se introducen en una sustancia gelatinosa o coloide -el gel-, previo a la aplicación del campo eléctrico.

1.2 GELES DE ADN

El papel que desempeña el gel es sencillo, y su función es similar a la de un filtro o embudo: facilita el paso de moléculas de cierto tamaño (relativamente pequeño, dependiendo de la preparación del gel) y opone resistencia al movimiento de moléculas de mayor tamaño. Ese nivel de resistencia lo dictamina el tamaño del poro (puesto que el gel es precisamente eso: una sustancia porosa). Haciendo más o menos grande el tamaño del poro se puede definir de manera muy precisa el nivel de resistencia al movimiento de una molécula (algo conocido como la "resolución del gel"). El nivel de resistencia al movimiento de un fragmento determinado equivale a una fuerza de fricción, dependiente de la velocidad de la partícula:

$$\bar{F}_r = \beta \bar{v} \quad [N] \quad (\text{Ecuación 2})$$

donde β es el coeficiente de fricción dinámico y v es la velocidad del fragmento. Dicha fuerza de fricción es opuesta a la ejercida por el campo eléctrico. En el equilibrio se alcanza una velocidad constante que depende, además de la carga neta de cada fragmento, del tamaño y forma de cada partícula.

Los geles más usados en procesos de electroforesis de ADN son los de agarosa (que es un coloide natural extraído de las algas marinas) y los de poliacrilamida (un material

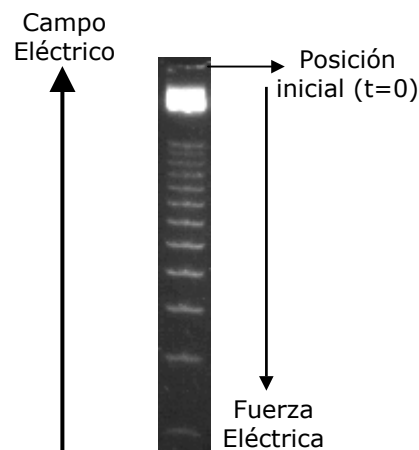
sintético que es usado, por ejemplo, para fabricar lentes de contacto). Los primeros tienen poros bastante grandes (bajo poder de resolución), en comparación con los geles de poliacrilamida; se puede asegurar entonces que las moléculas migran tan sólo en función de su carga eléctrica, puesto que el gel no ejerce ninguna acción de filtrado (prácticamente permite el paso de todas las moléculas de ADN). De esta forma, aquellos fragmentos que posean un tamaño muy cercano entre sí no podrán ser diferenciados tras el proceso de electroforesis (puesto que su carga y tamaño serán muy similares).

En el caso de geles de poliacrilamida, se cuenta con una resolución variable, puesto que el tamaño del poro se puede ajustar al momento de preparar el gel (usando la concentración adecuada de los diferentes componentes que lo conforman). Son, por tanto, mucho más versátiles ya que las moléculas pueden ser separadas de acuerdo con su carga eléctrica y además por su tamaño (moléculas de tamaño similar, pero diferente, podrán ser diferenciadas tras el proceso de electroforesis, siempre y cuando la preparación del gel sea la adecuada).

Cuando se prepara el gel y se introduce en él una muestra de ADN fragmentado, se logrará el efecto opuesto al de la Figura 3: todos los fragmentos tenderán a moverse en dirección contraria a la del campo eléctrico, pero solo aquellos lo suficientemente pequeños (comparados con el tamaño del poro) experimentarán una resistencia despreciable al movimiento. Dicha resistencia al movimiento aumenta a medida que el tamaño del fragmento es mayor. Por tanto, los fragmentos más livianos se desplazarán una distancia mayor, con respecto a una posición inicial determinada, en comparación con los fragmentos más pesados. La ventaja del uso de los geles, específicamente de poliacrilamida, es que la resolución del mismo se puede definir durante su preparación.

Para visualizar la ubicación de los diferentes fragmentos se agrega una sustancia fluorescente (generalmente el Bromuro de Etidio) que tiene la propiedad de intercalarse en la molécula de ADN, permitiendo así su visualización bajo la luz U.V. El resultado es similar al ilustrado en la figura 4:

Figura 4. Imagen obtenida tras aplicar el proceso de electroforesis a un gel con una muestra de ADN



Inicialmente, y con relación a la Figura 4, todos los segmentos están ubicados en la parte superior de la imagen. Tras aplicar el campo eléctrico por un período de tiempo (que puede oscilar entre 30 minutos y 2 horas) los fragmentos más livianos presentan una mayor movilidad que los fragmentos más pesados y por tanto se desplazan una mayor distancia con respecto a la posición inicial (los fragmentos son los pequeños rectángulos con tonalidad gris en la Figura 4). En este caso, los fragmentos más livianos están ubicados en la parte inferior de la imagen, mientras que los más pesados se encuentran en la parte superior. Se ilustra igualmente la dirección del campo eléctrico y de la fuerza eléctrica experimentada por cada fragmento.

1.2.1 Utilidad de los geles de ADN. Cualquiera de los procesos que involucran la identificación de moléculas de ADN con base en su tamaño, requiere el uso de los geles de ADN. Entre dichas aplicaciones se encuentran los procedimientos de secuenciación (cuyo objetivo primordial es conocer, precisamente, la secuencia de pares de bases presentes en uno o varios de los fragmentos de ADN analizados).

Desde el punto de vista clínico, la utilidad de los geles de ADN abarca procesos como:

- Identificación de genes^{*}: en muchas ocasiones se sospecha cuál puede ser la zona de la molécula de ADN en la cuál existe un gen específico. Si se fragmenta dicha molécula y, con base en el proceso de electroforesis aplicado a un gel de ADN, se logra determinar el tamaño (en pares de bases) de los fragmentos presentes en la muestra de ADN, entonces es posible establecer en cuál (o cuales) de los segmentos puede haber presencia de uno o varios genes de interés. Posteriormente se aísla el (los) fragmento (o fragmentos) de interés y se continúa con la etapa de secuenciación, donde precisamente se determina la secuencia de bases (Adenina, Timina, Guanina, Citosina) presentes en el fragmento analizado.
- Diagnóstico de enfermedades genéticas (es decir de carácter hereditario): se han desarrollado actualmente pruebas muy fiables que hacen posible una diagnosis prenatal de varias enfermedades genéticas (como la anemia falciforme, la hemofilia y la distrofia muscular). Para ello, la molécula de ADN es procesada por medio de enzimas de restricción y luego sometida a un proceso de electroforesis sobre un gel. Los segmentos de interés son aislados y posteriormente analizados en busca de secuencias específicas que muestren la existencia de algún tipo de enfermedad.
- La secuenciación del ADN de otros organismos (como virus o bacterias), lo que permitirá comprender los procesos que dan origen a enfermedades y a idear los posibles métodos que permitirán contrarrestar este tipo de afecciones.

El interés en muchas investigaciones (como el diagnóstico de enfermedades o el desarrollo de nuevos tratamientos para las mismas) es la determinación de los genes que están involucrados en el proceso (por ejemplo, es de interés determinar

^{*} Los genes hacen parte de la molécula de ADN y están presentes solo en algunas zonas de dicha molécula. Los genes, que son simplemente secuencias específicas de pares de bases, ejercen sus efectos a través de las moléculas a las cuáles dan origen. Sus productos inmediatos son las moléculas de ácido ribonucleico (ARN), de las cuáles la mayoría está involucrada en la producción de proteínas. Las proteínas, en última instancia, intervienen en diversas funciones vitales como el metabolismo, la contracción muscular o la respuesta inmunológica de un organismo. Las proteínas igualmente pueden ser enzimas (como la insulina y otras hormonas); además son los anticuerpos del sistema inmunológico y la hemoglobina, que transporta oxígeno en la sangre

Se concluye entonces que es la secuencia de bases presente en la molécula de ADN (y por tanto los genes presentes en la misma) quien dictamina el comportamiento de cualquier organismo (sea este animal o vegetal) en términos biológicos.

cuáles de estos genes están involucrados en el desarrollo de enfermedades como el cáncer, o cuáles de ellos se ven inhibidos, o igualmente activados, por virus como el VIH, lo cual conlleva al desarrollo de enfermedades como el SIDA).

1.2.2. Estimación del tamaño de los fragmentos presentes en un gel de ADN.

Las aplicaciones descritas anteriormente implican el uso de un procedimiento que permita determinar el tamaño de cada uno de los fragmentos presentes en el gel de ADN para posteriormente estar en capacidad de aislar aquellos fragmentos de interés.

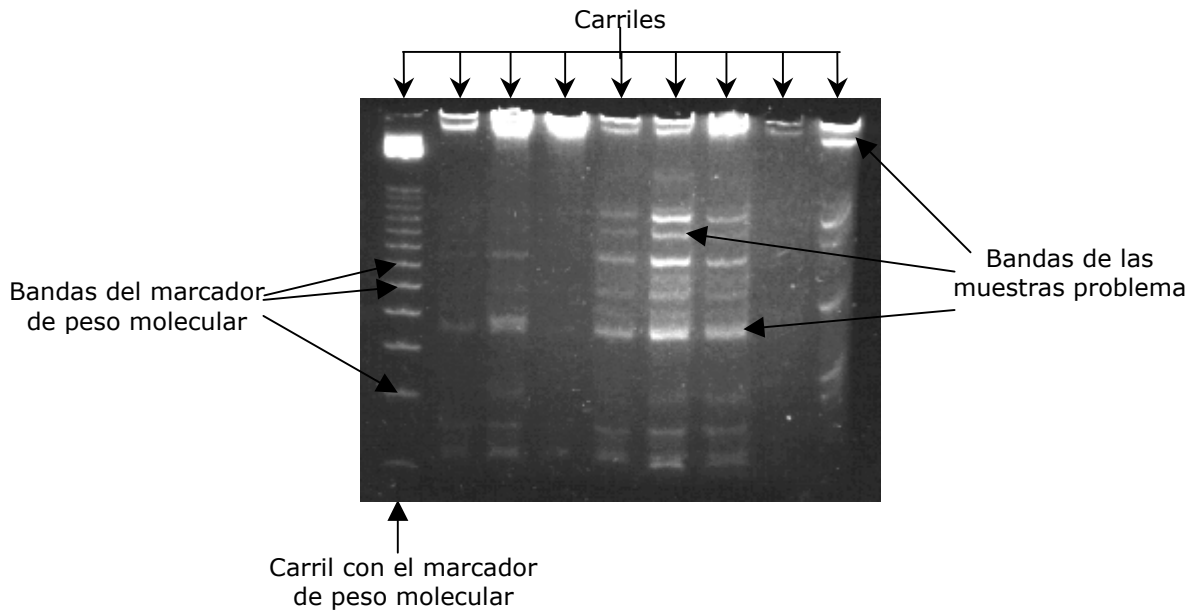
Para ello, junto con la muestra de ADN que se desea analizar, se ubica, en el mismo gel, otra muestra de ADN que ha sido previamente fragmentada por una enzima de restricción específica en la cual se conoce el tamaño de cada uno de sus fragmentos (dicha muestra recibe el nombre de marcador de peso molecular). Algunas enzimas de restricción usadas con este propósito, se ilustran en la Tabla 1:

Tabla 1. Tamaño de algunos fragmentos obtenidos tras la digestión de ADN con algunas enzimas de restricción^[22] (todos los valores están dados en kilopares de bases)

Fragmento	Enzimas				
	<i>EcoRI</i>	<i>HindIII</i>	<i>EcoRI + HindIII</i>	<i>BglII</i>	<i>AvaI</i>
A	21.8	23.7	21.8	22.8	15.9
B	7.52	9.46	5.24	13.6	8.8
C	5.93	6.75	5.05	9.8	6.1
D	4.8	4.26	4.21	2.3	4.6
E	3.41	2.26	3.41	0.46	4.1
F		1.98	1.98		1.8
G		0.58	1.90		1.61
H			1.71		1.55
I			1.32		
J			0.93		
K			0.84		
L			0.58		

Ahora bien, es posible que el gel de ADN no contenga únicamente la muestra de ADN segmentada con una de las enzimas descritas en la Tabla 1, junto con la muestra de ADN que se desea analizar. Más bien, es de uso común ubicar varias muestras de ADN (2 o más) junto con la muestra cuyos segmentos son de tamaño conocido. El resultado es una imagen como la ilustrada en la figura 5:

Figura 5. Un gel de ADN típico



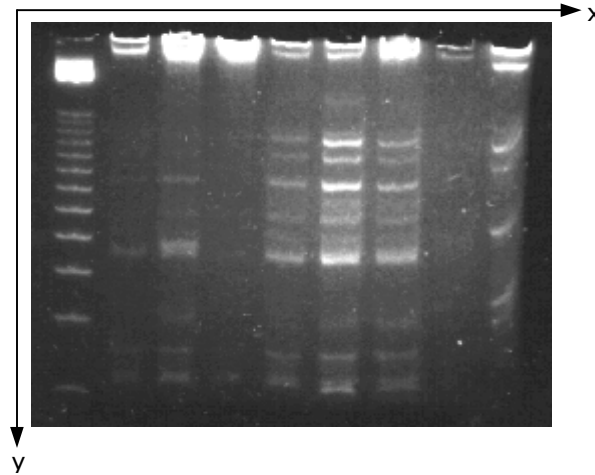
En esta figura se distinguen dos características importantes:

- Los carriles: son cada una de las columnas presentes en la imagen. Cada carril corresponde a una muestra de ADN fragmentada con una enzima de restricción determinada.
- Las bandas: son los pequeños rectángulos de tonalidad gris o brillante presentes en cada uno de los carriles, y que corresponden a un fragmento de la molécula de ADN (el nivel de gris de cada banda, así como su ubicación, dependerá del tamaño del fragmento correspondiente).

Ahora bien, teniendo en un mismo gel el carril con el marcador de peso molecular (que de ahora en adelante se llamará "carril estándar") y las muestras de interés, es posible determinar el tamaño de cada una de las bandas presentes en la imagen a partir, precisamente, de la comparación con dicho carril estándar. Para ello se lleva a cabo el siguiente procedimiento:

1. Se enmarca la imagen del gel en un par de ejes coordenados, como lo muestra la Figura 6:

Figura 6. El gel de ADN y sus correspondientes ejes coordenados

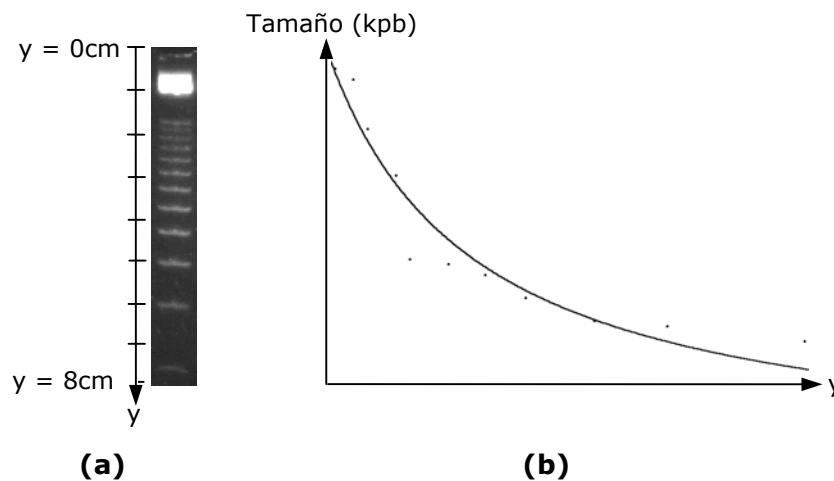


- Se determina la coordenada "y" de cada una de las bandas presentes en el carril estándar. Con base en esta coordenada y en el tamaño de cada uno de los segmentos se construye la curva Tamaño vs. Posición. Por ejemplo, para este caso se asume que el carril estándar es una muestra de ADN digerida con *EcoRI* y *HindIII*, obteniendo así doce segmentos. En la figura 7b se muestran las coordenadas (tamaño, posición sobre el eje "y") correspondientes a la figura 7a. La curva ilustra el ajuste de dichos puntos mediante una función de la forma^[22]:

$$S(y) = \frac{k_1}{y} + k_2 \quad [\text{Pares de bases}] \quad (\text{Ecuación 3})$$

donde S es el tamaño del segmento y k_1, k_2 son constantes que se calculan buscando el mejor ajuste a los datos que se tienen inicialmente.

Figura 7. Ilustración del proceso involucrado en la obtención de la curva Tamaño vs. Posición



3. Se determina la coordenada "y" de cada una de las bandas presentes en la imagen. Teniendo dichas coordenadas y conociendo la curva Tamaño vs. Posición (o, equivalentemente, las constantes k_1 y k_2 en la ecuación 2), se determina el tamaño de cada una de las bandas presentes en la imagen del gel de ADN.

1.3 PROCEDIMIENTO LLEVADO A CABO EN EL LABORATORIO DE GENÉTICA

El Laboratorio de Genética de la Facultad de Salud, cuenta con los reactivos e implementos necesarios para llevar a cabo el proceso de electroforesis. A continuación se describe brevemente el procedimiento involucrado en la obtención del tamaño de los fragmentos presentes en un gel de ADN:

- Se preparan las muestras de ADN y el gel, y posteriormente se exponen a un campo eléctrico. Esta etapa se lleva a cabo en un dispositivo conocido como cubeta de electroforesis.
- Se expone el gel de ADN a radiación UV, gracias a un dispositivo llamado transiluminador. Este proceso permite visualizar claramente las diferentes bandas presentes en el gel, dado que están marcadas con bromuro de etidio.
- Se captura de la imagen por medio de una cámara fotográfica.
- Se determina, de forma completamente manual, el tamaño aproximado de los fragmentos de interés. No es posible enmarcar la imagen dentro de un par de ejes coordenados (puesto que el papel fotográfico no facilita este procedimiento). Se hace necesario entonces calcular el tamaño de los diferentes fragmentos por procedimientos de tanteo, comparando la ubicación del fragmento de interés con el fragmento del carril estándar más cercano.

Se concluye entonces que, el proceso llevado a cabo es completamente manual y altamente impreciso.

2. MÉTODOS EXISTENTES PARA EL PROCESAMIENTO DE GELES DE ADN

En esta sección se presenta un panorama general relacionado con los trabajos, de carácter académico, realizados hasta el momento, así como un análisis de los sistemas comerciales existentes en la actualidad.

2.1 TRABAJOS EN EL CAMPO ACADÉMICO

Se describen a continuación los principales aportes en el campo investigativo, orientados a la determinación, semiautomática, del tamaño de las bandas presentes en una imagen proveniente de un gel de ADN. Se trata de algoritmos implementados con base en técnicas de procesamiento digital de imágenes y señales.

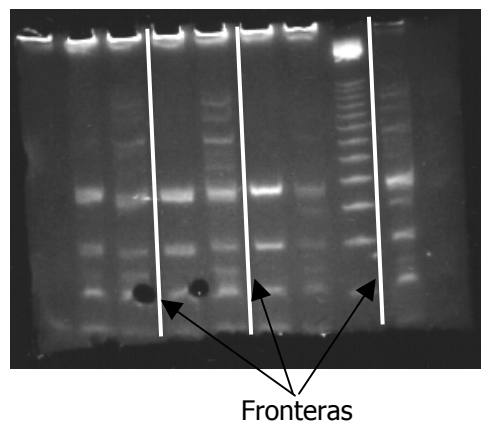
De acuerdo con lo descrito en el capítulo 1, referente al procedimiento involucrado en la determinación del tamaño de los segmentos de ADN, es posible afirmar que las metodologías descritas a continuación están compuestas por dos etapas:

- Segmentación de carriles: con ello se busca aislar cada uno de los carriles del fondo e igualmente diferenciarlos entre sí.
- Obtención de la ubicación de las bandas: se busca determinar la coordenada "y" de las bandas presentes en cada uno de los carriles.

En cada uno de los casos se supone que previamente se cuenta con la imagen almacenada en formato digital.

2.1.1 Segmentación de carriles. Existen dos tipos de aproximación, cada una de las cuáles posee ciertas ventajas pero también algunas desventajas. Dichas aproximaciones son: el análisis bidimensional (basado en técnicas de procesamiento digital de imágenes) y el análisis unidimensional (que en ocasiones implementa técnicas de procesamiento de señales). En ambos casos el objetivo es determinar la zona limítrofe (frontera) entre uno y otro carril, como lo ilustra la figura 8:

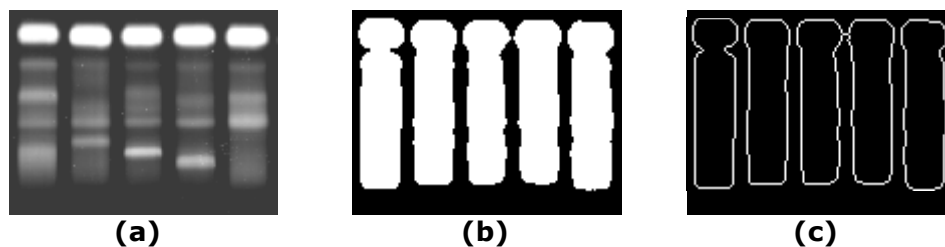
Figura 8. Ilustración del concepto de "frontera" entre uno y otro carril



2.1.1.1 Análisis bidimensional. En los trabajos realizados^[2] se pretende segmentar la imagen del gel de ADN con base en las características de su histograma (específicamente, se busca que este sea bimodal, lo que posteriormente permitirá segmentar la imagen con base en la adecuada estimación del umbral presente en dicho histograma). En primera instancia, este procedimiento impone restricciones como la necesidad de iluminación uniforme sobre el gel, requisito que muchas veces (sino la mayoría) es imposible de cumplir, esto debido principalmente a la variabilidad de la composición entre uno y otro gel (ya que para cada experimento las muestras de ADN cambiarán completamente, junto con la preparación del gel y la cantidad de Bromuro de Etidio aplicado a la muestra).

Si la etapa anterior resulta exitosa, y es posible estimar el umbral adecuado del histograma, se procede a crear una "máscara" binaria, donde los carriles son etiquetados con 1 y el fondo es etiquetado con 0. Posterior a dicho procedimiento se aplica un filtro detector de bordes, se evalúa entonces la transformada de Radón sobre la imagen obtenida con el fin de determinar la orientación de los bordes de interés, para finalmente eliminar aquellos que son perpendiculares a la orientación de los carriles (los bordes perpendiculares a los carriles son el resultado de segmentar las bandas presentes dentro de cada uno de ellos). El procedimiento se ilustra en la figura 9:

Figura 9. Obtención de las zonas límite intercarril tras el procesado de la imagen del gel: a) Imagen original; b) Máscara binaria; c) Resultado de aplicar un filtro detector de bordes a la imagen en (b)



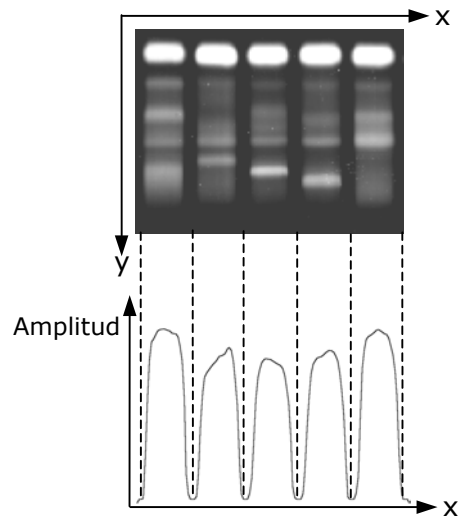
Además de las características inadecuadas de la imagen, existen otras limitantes en este procedimiento:

1. La distribución de bandas debe ser uniforme a lo largo de todos y cada uno de los carriles. La presencia de pocas bandas dentro de un carril hace que en la máscara binaria (figura 9b) cada carril no sea un objeto independiente. Por el contrario, es probable que un carril esté segmentado en varias partes, correspondiente cada una de ellas a una banda. Esto desde luego dificulta el proceso de detección de zonas límite intercarril.
2. La orientación de los carriles debe ser completamente vertical, de esta manera el filtro detector de bordes, cuyo resultado depende en gran parte de la orientación de los carriles^{[9][16]}, permitirá la correcta determinación de las zonas limítrofes.
3. Tras la aplicación del filtro detector de bordes se procede a evaluar la transformada de Radón sobre dicha imagen, y a determinar, mediante un proceso iterativo, la ubicación de las zonas de interés que definen la frontera entre uno y otro carril. La desventaja de esta técnica está en el tiempo requerido para el cálculo de la

transformada Radón y el proceso iterativo para detectar el máximo en dicha transformada (que es una función bidimensional).

2.1.1.2 Análisis unidimensional. Parte del hecho de que no es necesario procesar la imagen completa, ya que basta con conocer correctamente la coordenada "x" (ver Figura 10) en la que se encuentra la frontera entre un carril y otro. Se evalúa entonces la proyección de la imagen sobre dicho eje "x" y posteriormente se aplica un algoritmo detector de mínimos sobre dicha proyección. La ubicación de estos mínimos coincide perfectamente con la ubicación de las zonas limítrofe:

Figura 10. Proyección de la imagen original sobre el eje x. Arriba: imagen original (bidimensional). Abajo: proyección (señal unidimensional)



La proyección se obtiene simplemente como la suma por columnas de la imagen^{[13],[1],[23],[5]}. El procedimiento tiene la ventaja de reducir el costo computacional en comparación con el análisis bidimensional (puesto que ahora se analizará una señal unidimensional). De hecho, la complejidad de los filtros a utilizar se reduce (se usan máscaras unidimensionales para el filtrado por la mediana o por la media) y el algoritmo detector de mínimos hace uso de primeras y segundas derivadas aplicadas a la proyección.

Algunas desventajas de esta aproximación son las siguientes:

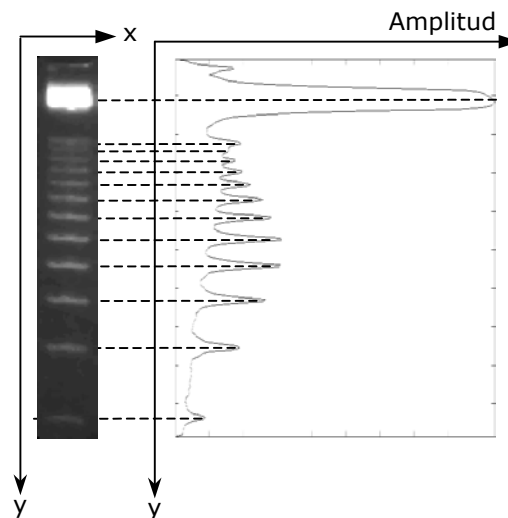
1. Se asume, al igual que en el caso bidimensional, que los carriles son perfectamente verticales. A partir de esta premisa se calcula la proyección como la suma por columnas de la imagen. De nuevo, esta aseveración es equivocada, y conlleva a errores en la segmentación de los carriles, en la posterior detección de las bandas y en la estimación del tamaño de cada fragmento (ver capítulo 5: resultados). La única excepción se encuentra en el trabajo desarrollado por Machado, Siqueira y otros^[13], en donde inicialmente se consideran carriles completamente verticales pero, posterior a esta segmentación inicial se evalúa la correcta orientación de cada uno de ellos con base en el cálculo de momentos de primero y segundo orden. La desventaja que presenta este método es que cada carril debe poseer una distribución uniforme de bandas, las cuales suministran la

información necesaria para estimar de forma correcta la orientación del carril; si el carril contiene solo unas cuantas bandas es probable que la información sea insuficiente, lo cual conlleva a un cálculo erróneo de la orientación del carril. Además, se debe tener en cuenta el hecho de que no es necesario evaluar la orientación para todos y cada uno de los carriles presentes en la imagen, puesto que es posible comprobar que todos poseen exactamente la misma orientación (exceptuando casos en los que la preparación del gel o de las muestras de ADN no son los adecuados).

2. El hecho de evaluar la proyección como la suma por columnas puede llevar a errores en la detección de mínimos locales. Esto se debe a que es posible que el fondo de la imagen contenga ruido, y que además existan carriles con poco contenido de bandas. La combinación de esos dos factores puede hacer que en la proyección sea imposible diferenciar claramente las zonas limítrofe de zonas que pertenecen a un carril determinado.

2.1.2 Detección de bandas. En la bibliografía encontrada, el procedimiento de uso más extendido es similar al implementado en la detección de carriles a partir de una señal bidimensional^{[13],[1],[23],[5],[8],[3]}. Para cada uno de los carriles segmentados se calcula la proyección sobre el eje "y" (en este caso se toma cada uno de los carriles y se evalúa la suma por filas) y se aplica un algoritmo detector de máximos locales a la proyección obtenida. La ubicación de los máximos locales coincide perfectamente con la ubicación de las bandas, como se ilustra en la figura 11:

Figura 11. Un carril (izquierda) y su proyección (derecha)



El procedimiento que permite detectar los máximos locales es muy similar al usado anteriormente en la detección de los mínimos locales, y de nuevo hace uso de derivadas de primero y segundo orden. Por otra parte, las desventajas presentes en este procedimiento son similares a las encontradas en el caso anterior:

1. La evaluación de la proyección como la suma por filas de la imagen correspondiente al carril es inadecuada en el sentido de que algunas bandas que no logran diferenciarse claramente del fondo (como por ejemplo la banda ubicada

- en la parte inferior de la figura 11) pueden no ser detectadas en caso de que haya presencia de ruido en la imagen.
2. Al igual que en el caso anterior, no se puede asumir que las bandas tienen una orientación paralela al eje "x". Es por tanto inadecuado calcular la proyección como la suma por filas de la imagen correspondiente al carril.
 3. En ocasiones^[5] y para evitar interferencia entre un carril y otro, se habla de un "carril efectivo" entendido este como el resultado de tomar una zona (de ancho entre 10 y 15 columnas) alrededor del centro del mismo. Desde luego, esta aproximación es válida siempre y cuando el carril sea completamente vertical (o al menos su orientación sea prácticamente paralela al eje "y"). Sin embargo, existe pérdida de información que puede afectar el proceso de detección de bandas.

Se concluye entonces que los procedimientos de análisis unidimensional presentan ventajas en términos de costo computacional. La principal desventaja es que asumen una perfecta orientación vertical de los carriles, hecho que muy pocas veces es cierto puesto que las muestras son ubicadas manualmente por el usuario y es prácticamente imposible que siempre tengan la orientación ideal. Este hecho determina el comportamiento del algoritmo en las etapas de segmentación y detección de bandas.

2.2 SISTEMAS COMERCIALES

De la gran variedad de sistemas orientados al análisis de geles de ADN, se encontraron los siguientes:

- 1D Image Analysis Software, de Kodak.
- BASys Image Analysis System, de Biotec-Fisher GmbH.
- Doc-It: Adquisition and 1D analysis software, de UVP Inc.
- Labscan Utility v2.0, Nonlinear Dynamics Ltd.
- TotalLab, Nonlinear Dynamics Ltd.
- 1DScan EX v3.1, Scanalytics Inc.

De dichos sistemas se tuvo acceso a una versión de prueba de TotalLab y a una de 1DScan. Más adelante, en el capítulo 5, se analiza el desempeño de esos dos sistemas en comparación con los algoritmos desarrollados para la segmentación de carriles. Sin embargo, se describirán a continuación sus principales características así como su valor comercial.

2.2.1 Principales características. Como es de esperar, los algoritmos implementados en estos sistemas están orientados a la determinación del tamaño de los diferentes segmentos de ADN. Específicamente, implementan etapas como:

- Segmentación de carriles: los procedimientos involucrados incluyen la estimación de la orientación de la imagen (y por tanto de los carriles). Se incluye una interfaz gráfica que permite al usuario la marcación/eliminación manual de uno o varios carriles, en caso de que las características de la imagen no permitan una detección automática de los mismos. No se encuentra descripción acerca del procedimiento implementado para la detección de los carriles (solo en uno de los sistemas evaluados, TotalLab, se pudo comprobar que esa detección se obtiene tras evaluar la proyección de la imagen sobre el eje "x"). Algunos de ellos incluyen, además, la

posibilidad de seleccionar manualmente regiones de interés sobre la imagen, con el fin de analizar solo una parte de ella.

- Detección de bandas: en todos los casos se obtiene la proyección de cada uno de los carriles sobre el eje "y". El objetivo principal en este proceso, como se mencionó anteriormente, consiste en ubicar los máximos locales de dicha proyección, correspondientes precisamente a la coordenada "y" de los diferentes fragmentos (bandas) de ADN. Uno y otro sistema se diferencian en la forma de abordar dicha detección de máximos. Se supone que en todos los casos la etapa final consiste en la evaluación de derivadas de primero y segundo orden. En etapas previas se puede incluir algún tipo de suavizado mediante el uso de filtros por la media, por ejemplo, e incluso haciendo uso de técnicas como la deconvolución*. Dicha técnica es adecuada solo en algunos casos, y se aplica a imágenes ideales (con muy bajo nivel de ruido). Sin embargo, la mayoría de las veces es prácticamente imposible determinar dichas características, y se opta entonces por el uso de filtros suavizantes. De hecho, algunos autores^[3] sugieren que este es un problema bastante complejo, y que la determinación de la forma de la respuesta impulsional del sistema no es un asunto trivial.

Desde luego, se incluyen opciones que permiten al usuario agregar/eliminar manualmente algunas bandas presentes en cada uno de los carriles.

- Estimación del tamaño de los fragmentos: se incluyen diferentes librerías para el uso de variedad de estándares, útiles en la determinación del tamaño de los fragmentos (un estándar es simplemente una enzima específica que fragmenta el ADN en un número determinado de segmentos, cuyo tamaño es conocido – ver Tabla 1). Algunas aplicaciones ofrecen al usuario la posibilidad de elegir la forma que debe tener la función matemática que relaciona tamaños y coordenadas (exponencial, polinómica de diferentes órdenes, lineal, potencias, etc...). La función matemática correspondiente depende precisamente del tipo de estándar utilizado.
- Otras técnicas de procesamiento de imágenes: en varios de los sistemas analizados se ofrecen opciones que permiten corregir la distribución de niveles de gris en la imagen (mejora del contraste, eliminación de ruido de fondo – con base en transformaciones morfológicas como la transformada *Top-Hat*^[16]-, entre otras).
- Funciones adicionales: se ofrece al usuario la opción de almacenar los resultados (esto es el tamaño de cada uno de los segmentos presentes en el gel de ADN) y/o exportarlos a otra aplicación (como Excel).
- Todas las aplicaciones analizadas incluyen, además, módulos para el análisis de geles de ADN bidimensionales, o para el conteo de colonias (opciones que no serán

* En el procedimiento de deconvolución se asume que la proyección del carril sobre el eje "y" (sea esta $f(y)$), es el resultado de la convolución entre una señal "pura" (sin ningún tipo de distorsión, ni ruido), $g(y)$, y una función $h(y)$ que simula precisamente la distorsión introducida en la imagen digital (y que es una combinación de la etapa de adquisición con las características del gel y de las muestras):

$$f(y) = g(y) \otimes h(y)$$

El objetivo del procedimiento consiste entonces en estimar la forma y características de la señal $h(y)$. Conociendo $f(y)$ (la proyección del carril sobre el eje y) junto con $g(y)$ es posible determinar el patrón de proyección "ideal" $g(y)$.

incluidas en el presente trabajo puesto que no son aplicaciones de interés para el Laboratorio de Genética).

- Algunas incluyen el sistema de adquisición de imágenes, otras ofrecen por separado los componentes de hardware y software.

Se observa entonces que son sistemas semiautomáticos, que permiten, en ocasiones de manera interactiva, determinar el tamaño de cada uno de los segmentos presentes en la imagen.

2.2.2 Costo. El costo depende de los componentes incluidos en la aplicación (software y/o hardware). Asumiendo que se cuenta con el sistema de adquisición, se tienen precios que oscilan entre los U\$2300 (si se adquieren por ejemplo módulos útiles únicamente para el análisis de geles unidimensionales, como es el caso) y los U\$5200, para sistemas que no se ofrecen por módulos (y que involucran por tanto opciones como el análisis de geles bidimensionales y el conteo de colonias). Para sistemas que incluyen la etapa de adquisición (cámara CCD e incluso monitor), los precios van desde los U\$ 6000 en adelante, aproximadamente.

Desde luego, es posible actualizar una de estas aplicaciones, pero necesariamente se debe incurrir en costos que (aunque menores) son significativos, dependiendo de la versión del sistema y de las características de la aplicación a la cual se quiere actualizar. Se trata entonces de sistemas completamente cerrados, que no permiten proyección a nuevas investigaciones de carácter algorítmico.

3. ALGORITMO PARA EL ANÁLISIS DE GELES DE ADN, BASADO EN TÉCNICAS DE PROCESAMIENTO DIGITAL DE SEÑALES E IMÁGENES

En este apartado se describen los algoritmos implementados en el presente trabajo, tendientes a resolver el problema de la estimación del tamaño de cada uno de los fragmentos presentes en el gel de ADN, y que mejoran algunas falencias encontradas en trabajos anteriores.

Como se ha mencionado hasta el momento, la solución al problema de la determinación del tamaño de los segmentos de ADN se puede dividir en tres etapas: segmentación de carriles, determinación de la ubicación de las bandas y estimación del tamaño de los fragmentos. A continuación se describe cada una de ellas. Se asume, por supuesto, que la imagen se encuentra almacenada en formato digital.

3.1 SEGMENTACIÓN DE CARRILES

Busca diferenciar los carriles del fondo e igualmente determinar correctamente la frontera entre uno y otro carril. Como se mencionó anteriormente, la segmentación por niveles de gris (análisis del histograma) requiere que las imágenes cumplan requisitos muy específicos (como la iluminación uniforme y el carácter bimodal del histograma), requisitos que en la mayoría de las ocasiones son imposibles de cumplir (debido, principalmente a sustancias como el Bromuro de Etidio, que es agregado a los fragmentos de ADN y encargado de su fluorescencia, y que puede tener diferentes concentraciones entre uno y otro experimento).

Se consideró en primer lugar la segmentación con base en la periodicidad presente en el gel^{[24], [14]} (ver por ejemplo la Figura 6), evaluando inicialmente la transformada de Fourier de la imagen y observando entonces las características del espectro en frecuencia. Se encontró la presencia de algunos picos en dicha transformada, signo de la existencia de algún tipo de periodicidad. El procedimiento presenta una gran desventaja y es que la imagen del gel debe ser prácticamente ideal: su nivel de ruido debe ser mínimo, la distribución de bandas a través de cada carril debe ser uniforme, la iluminación adecuada y un histograma de características bimodales. Solo de esta manera se puede garantizar la existencia de picos en la transformada de Fourier de la imagen, relacionados estos con la periodicidad del gel de ADN. Desde luego, son requisitos que restringen la aplicación de este procedimiento de manera considerable.

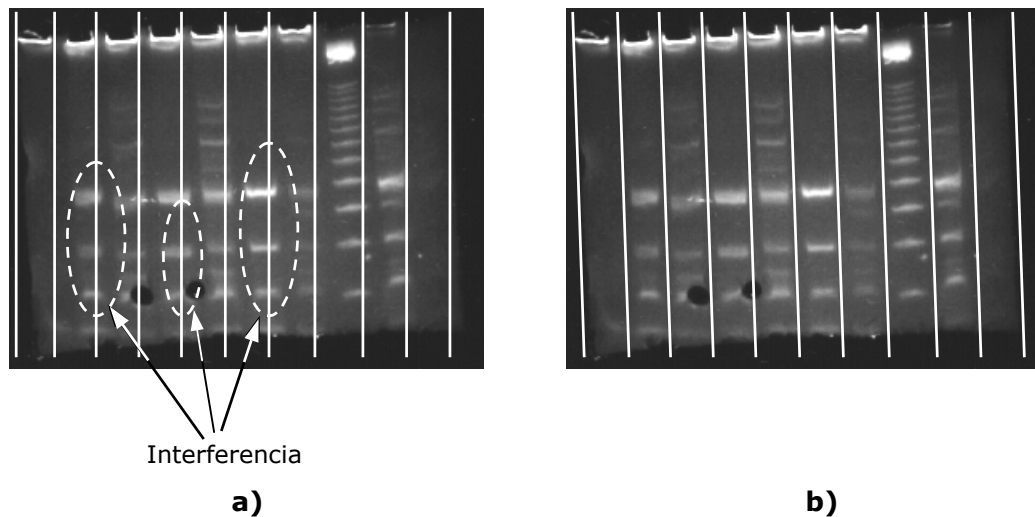
Se optó entonces por el análisis unidimensional, basado en el cálculo de la proyección de la imagen sobre un eje en particular, principalmente debido a las siguientes razones:

- Es apropiado si se tienen en cuenta las características de las imágenes obtenidas (iluminación no uniforme y la inexistencia de un carácter bimodal en su histograma).
- La delimitación entre un carril y otro implica el conocimiento de solo una coordenada de dicha zona limítrofe (por ejemplo, la coordenada "x"). No es necesario analizar la señal bidimensional (la imagen), por el contrario, es posible

llevar a cabo el análisis en una dimensión (la proyección) y luego extender el resultado a dos dimensiones.

Sin embargo, y como se analizó anteriormente, la mayoría de las veces es impreciso asumir que los carriles son completamente verticales. La razón principal es que en el momento de la preparación del gel y la ubicación de las muestras de ADN sobre el mismo por parte del usuario, es muy probable que la orientación de las mismas no sea completamente horizontal, lo cual conlleva a la aparición de carriles que no son perfectamente verticales tras el proceso de electroforesis. Como se ilustrará más adelante (ver capítulo 5), el hecho de asumir una orientación perfectamente vertical redundaría en un aumento de errores al momento de detectar las bandas presentes en el gel y, principalmente, en la posterior estimación del tamaño de cada fragmento. Este efecto se puede apreciar en la Figura 12, donde, como ejemplo comparativo, se han dibujado los carriles presentes en una imagen asumiendo que estos son completamente verticales (Figura 12a) y teniendo en cuenta su orientación (Figura 12b):

Figura 12. Ubicación de los límites intercarril a) Asumiendo carriles verticales; b) Teniendo en cuenta la correcta orientación de la imagen



En la Figura 12a se han señalado además algunas zonas donde existe interferencia entre un carril y otro (los carriles comparten cierto número de bandas). Desde luego, este es un efecto indeseable y se puede suponer desde ya que es necesario estimar inicialmente la orientación de la imagen, como una etapa previa a la obtención de la proyección de la imagen del gel. Este problema se aborda a continuación.

3.1.1 Cálculo de la orientación de la imagen. Existe una gran variedad de aproximaciones que pueden ser consideradas en la determinación de la orientación de una imagen en escala de grises con respecto a los ejes x,y convencionales. Un método útil en la estimación de la orientación de una imagen (que equivale a determinar los ejes principales de la misma), es el uso de la matriz de covarianza. Sin embargo, la complejidad de los cálculos es relativamente alta y depende directamente del tamaño de la imagen. Dicha complejidad se puede reducir cuando se maneja una imagen

binaria, pero esta es una etapa que, según lo descrito anteriormente, es difícil de implementar debido a las características del gel de ADN (iluminación no uniforme y el carácter monomodal del histograma de niveles de gris). Se consideraron entonces algunos métodos espectrales, partiendo del cálculo de la Transformada de Fourier de la imagen del gel y considerando este espectro como una distribución bidimensional en coordenadas polares de la forma $S(r,\theta)$, evaluando posteriormente la función $S(\theta)$, descrita en la ecuación 4, para un radio específico:

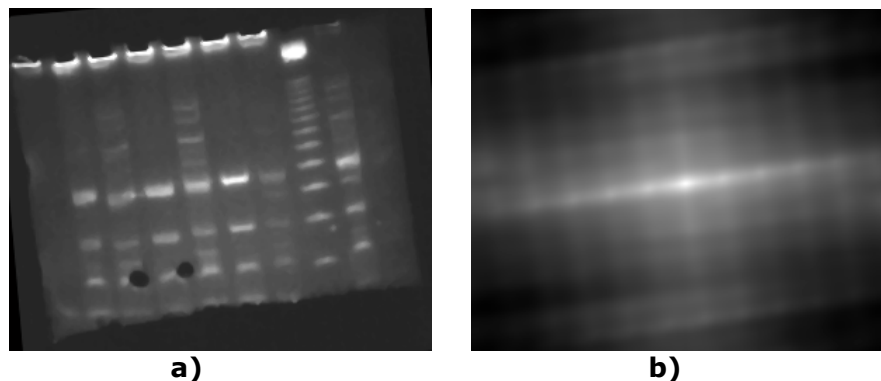
$$S(\theta) = \sum_{r=1}^R S(r,\theta) \quad (\text{Ecuación 4})$$

Si $0 \leq \theta \leq 2\pi$, dicho cálculo equivale a determinar el comportamiento del espectro a lo largo de un círculo centrado en el origen (que coincide en este caso con el centro de la imagen). La presencia de picos en la función $S(\theta)$ manifiesta la existencia de una orientación predominante en la imagen^[9]. El algoritmo descrito funciona de manera adecuada solo para una imagen de prueba (de un total de 18), y se demuestra entonces que no es una solución viable al problema de la determinación de los ejes principales de la imagen.

Se analizó entonces una aproximación morfológica^[11] pero la desventaja principal de este algoritmo es que no arroja resultados globales, únicamente permite determinar la orientación de zonas específicas de la misma (el interés del presente trabajo radica en la obtención de la orientación global).

Se optó entonces por el uso de la imagen de autocorrelación para la obtención de la orientación de la imagen. Dicha autocorrelación tiene la ventaja de que, al ser evaluada sobre una imagen con una orientación determinada, su resultado arroja la aparición de un patrón, lo suficientemente claro, orientado precisamente en la dirección de uno de los ejes principales de la imagen^{[5],[10]}. Este hecho se ilustra en la Figura 13:

Figura 13. Un gel de ADN cuyos carriles no están alineados verticalmente (a) y su correspondiente imagen de autocorrelación (b)



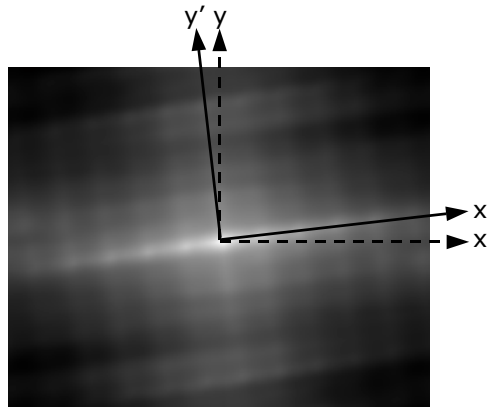
Por su parte, la autocorrelación (específicamente su magnitud) se obtiene a partir del producto entre la transformada de Fourier y su complejo conjugado:

$$\text{Autocorrelación} = \left| TF^{-1} \{ TF(I) \times TF^*(I) \} \right| \quad (\text{Ecuación 5})$$

donde TF^{-1} , TF y TF^* hacen referencia a la transformada de Fourier inversa, la transformada de Fourier y el complejo conjugado de la Transformada de Fourier (todas bidimensionales), e I hace referencia a la imagen del gel.

El siguiente paso consiste en el desarrollo de un método que permita, a partir de la magnitud de la secuencia de autocorrelación, calcular la orientación de la imagen, o, equivalentemente, establecer cuáles son sus ejes principales (su orientación relativa con respecto a los ejes x,y tradicionales). Esos nuevos ejes principales reciben el nombre de x' e y' , de acuerdo con la figura 14:

Figura 14. Los nuevos ejes x',y' y su relación con los ejes x,y tradicionales



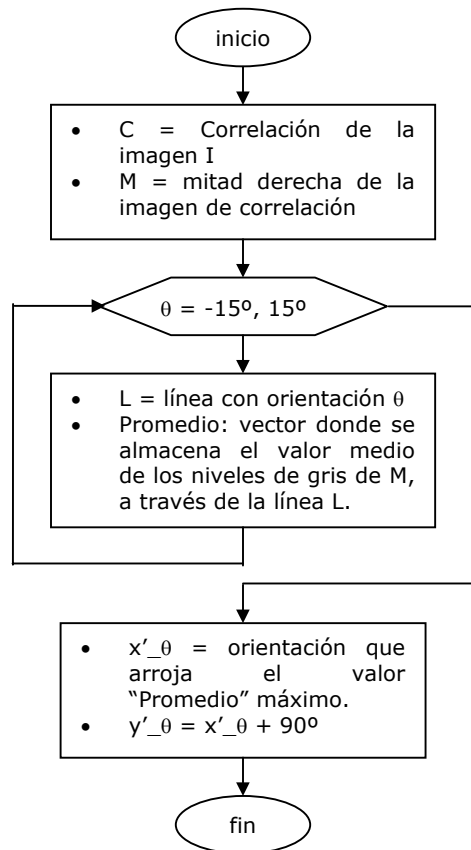
Inicialmente se debe tener en cuenta la simetría de la imagen de la autocorrelación. En la Figura 14 se observa que esta es con respecto al origen de coordenadas y se concluye que en el análisis que se llevará a cabo solo es necesario el uso de una de las dos mitades de la imagen. Se opta por el uso de la mitad derecha de la imagen, correspondiente a los cuadrantes I y IV del sistema de coordenadas cartesiano tradicional (se incluye el cuadrante IV debido a que el eje x' puede formar un ángulo positivo o negativo con respecto al eje " x "). El procedimiento descrito a continuación busca determinar la orientación del eje x' con respecto al eje x . El eje y' estará a 90° del eje x' .

El proceso que permite obtener los ejes principales de la imagen consiste en desplazar una línea recta con uno de sus extremos coincidente con el origen de coordenadas, y que tiene una longitud aproximadamente igual al ancho de la imagen que se está analizando (como en este caso la imagen analizada es la mitad derecha de la imagen original, la longitud de la línea será igual a la mitad del ancho de la imagen original). En cada iteración del algoritmo, la línea rotará 0.2° (en un rango de -15° a 15°), conservando siempre uno de sus extremos en el origen de coordenadas. Se determinan entonces los niveles de gris de la imagen de la autocorrelación que se encuentran bajo esa línea y se estima el valor promedio alcanzado por todos estos

niveles de gris. Al terminar el proceso se determina cuál fue la orientación de la línea que arrojó el mayor promedio, orientación que corresponde (en el 100% de las imágenes analizadas, de un total de 18) precisamente al ángulo existente entre el eje x' y el eje "x". Se procede entonces a calcular la orientación del eje y' , simplemente sumando 90° a la orientación correspondiente al eje x' (como se mencionó anteriormente).

La figura 15 es el diagrama correspondiente al algoritmo descrito con anterioridad:

Figura 15. Algoritmo usado para determinar la orientación de la imagen



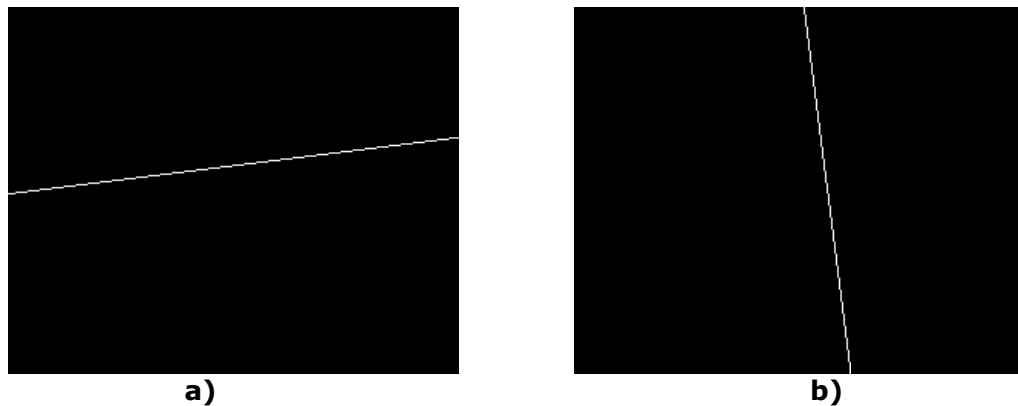
Tras determinar los ejes principales de la imagen se procede a evaluar su proyección sobre el eje x' . La proyección sobre dicho eje, junto con la correcta estimación de la ubicación de sus mínimos, permite finalmente segmentar los carriles presentes en la imagen. Se debe tener en cuenta que el cálculo de dicha proyección no será simplemente la suma por columnas de la imagen (ya que sus ejes coordenados no coinciden, en general, de manera perfecta con los ejes cartesianos tradicionales). Es necesario implementar un procedimiento que permita realizar un "barrido" sobre la imagen teniendo en cuenta la orientación correspondiente. Ese procedimiento se hace con base en el cálculo de mapas de distancia, método que se describe a continuación.

3.1.2 Los mapas de distancia. También reciben el nombre de “Transformación de distancia” y se pueden definir de la siguiente manera^{[17],[12]}: dada una imagen binaria con valores 0/1 la transformada de distancia aplicada sobre dicha imagen binaria, resulta en una imagen en escala de grises, donde cada pixel representa la mínima distancia entre el fondo (pixel con valor 0) y el pixel con valor 1 más cercano. En la determinación de dicha distancia se usan diferentes métricas, entre otras: *City Block*, *Chessboard* y Euclideana^[17] (la implementada en el presente trabajo es precisamente *City Block*, por razones que más adelante serán explicadas). Entre los algoritmos más extendidos se encuentran los iterativos (o paralelos) y los secuenciales (o recursivos), ambos basados en procedimientos de morfología matemática.

Dependiendo de las características de la imagen binaria (forma del objeto y tamaño de la matriz) los algoritmos pueden llegar a requerir un gran tiempo en la obtención de la transformada de distancia. Este es precisamente el caso, ya que se habla de imágenes binarias (con el mismo tamaño de las imágenes provenientes de los geles), mayores a los 640x480 pixeles. Se propone entonces una solución equivalente que permite obtener los mapas de distancia necesarios, pero que es a la vez mucho más simple y computacionalmente eficiente.

El algoritmo planteado parte del uso de dos “marcadores”, cada uno de los cuales es simplemente una imagen binaria, donde el objeto de interés es una línea paralela al eje x' , en un caso, y al eje y' en el otro. A continuación se ilustran estos dos marcadores correspondientes a la imagen de la Figura 13a:

Figura 16. Los dos marcadores utilizados en la obtención de los mapas de distancia: a) Marcador paralelo al eje x' ; b) Marcador paralelo al eje y'



Se puede afirmar que la posición de cada uno de los marcadores dentro de la imagen binaria, es irrelevante. El único requisito es que deben estar ubicados de manera simétrica, en el sentido de que arriba y abajo (o a la derecha y a la izquierda) debe haber dos esquinas a cada uno de los lados (por ejemplo, dos esquinas a la derecha del marcador de la figura 16b y dos a su izquierda, como es el caso).

Se procede entonces a evaluar los dos mapas de distancia correspondientes. Se ilustra el procedimiento para el caso de una imagen de 5x3 (5 filas, 3 columnas); el

resultado se puede extender para el caso de imágenes mucho más grandes (como son las imágenes del gel). La Figura 17 muestra una imagen de ese tamaño donde los pixeles iguales a "1" corresponden precisamente al marcador orientado en una dirección determinada.

Figura 17. Ejemplo para una imagen binaria con un marcador (tamaño 5x3)

$$\text{Marcador} = \begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array}$$

Se procede a crear una nueva matriz (*Grilla*), con cada fila codificada desde 1 hasta el número de filas que contiene la imagen binaria (en este caso 5). El resultado se ilustra en la figura 18:

Figura 18. La imagen binaria y su matriz *Grilla* correspondiente

$$\text{Marcador} = \begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array} \quad \text{Grilla} = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 2 & 2 & 2 \\ \hline 3 & 3 & 3 \\ \hline 4 & 4 & 4 \\ \hline 5 & 5 & 5 \\ \hline \end{array}$$

Ahora bien, se determinan las filas de la matriz en las cuales se encuentran los pixeles que pertenecen al marcador. En este caso (y según la Figura 17 vista de izquierda a derecha) esas filas son: 4, 3 y 2. Este resultado se almacena en un vector (por tanto, en este caso, el vector tendrá componentes [4 3 2]).

Se crea ahora una nueva matriz (*A*) del mismo tamaño de la imagen binaria y cuyas filas son iguales al vector obtenido anteriormente:

Figura 19. Matriz *A* cuyas filas corresponden a las coordenadas de los pixeles pertenecientes al marcador

$$A = \begin{array}{|c|c|c|} \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline \end{array}$$

Finalmente, se calcula la diferencia entre la matriz *Grilla* y la matriz *A* (en este caso el resultado se almacena en la matriz *B*):

Figura 20. Matriz B , resultado de evaluar la diferencia entre la matriz $Grilla$ y la matriz A

$$B = Grilla - A = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 2 & 2 & 2 \\ \hline 3 & 3 & 3 \\ \hline 4 & 4 & 4 \\ \hline 5 & 5 & 5 \\ \hline \end{array} - \begin{array}{|c|c|c|} \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline 4 & 3 & 2 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline -3 & -2 & -1 \\ \hline -2 & -1 & 0 \\ \hline -1 & 0 & 1 \\ \hline 0 & 1 & 2 \\ \hline 1 & 2 & 3 \\ \hline \end{array}$$

Se debe observar que el resultado obtenido representa efectivamente la distancia existente entre cada pixel del marcador y los pixeles etiquetados con "1" (exceptuando los valores negativos). Así por ejemplo, los pixeles iguales a cero corresponden precisamente a los pixeles etiquetados con "1" en el marcador (y por tanto la distancia mínima es cero). De igual manera, por ejemplo, aquellos pixeles ubicados exactamente por encima y por debajo del marcador están a una distancia (entera) igual a 1, y así sucesivamente. Si se toma la magnitud de la matriz resultante, B para el caso de la Figura 20, el resultado coincide con la aplicación formal de la transformada de distancia usando la métrica *City Block* (pero tiene la ventaja de ser mucho menos complejo y adecuado para el desarrollo de la presente aplicación). Igualmente, y más importante aún, se debe notar que los pixeles etiquetados con el mismo valor son paralelos a la línea del marcador (observar por ejemplo el caso de los pixeles etiquetados con 1, -1, 2 o -2: todos ellos son paralelos a dicha línea).

El resultado es normalizado para eliminar los pixeles con valor negativo. Para ello, simplemente se suma a la matriz B una matriz con todos sus componentes iguales al valor absoluto del menor valor presente en B (en este caso, el mínimo de esta matriz es -3). El resultado es el mapa de distancia de interés en este caso:

Figura 21. Mapa de distancias obtenido para el ejemplo de una matriz de 5x3

$$Mapa = \begin{array}{|c|c|c|} \hline -3 & -2 & -1 \\ \hline -2 & -1 & 0 \\ \hline -1 & 0 & 1 \\ \hline 0 & 1 & 2 \\ \hline 1 & 2 & 3 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline 1 & 2 & 3 \\ \hline 2 & 3 & 4 \\ \hline 3 & 4 & 5 \\ \hline 4 & 5 & 6 \\ \hline \end{array}$$

Es de resaltar que la matriz obtenida (*Mapa*) conserva las características deseadas (un grupo de pixeles con igual valor es precisamente una línea paralela al marcador). Si se aplica este algoritmo a las dos imágenes binarias ilustradas en la Figura 16, se obtienen dos mapas: uno con un grupo de líneas paralelo al eje x' y el otro con un grupo de líneas paralelo al eje y' .

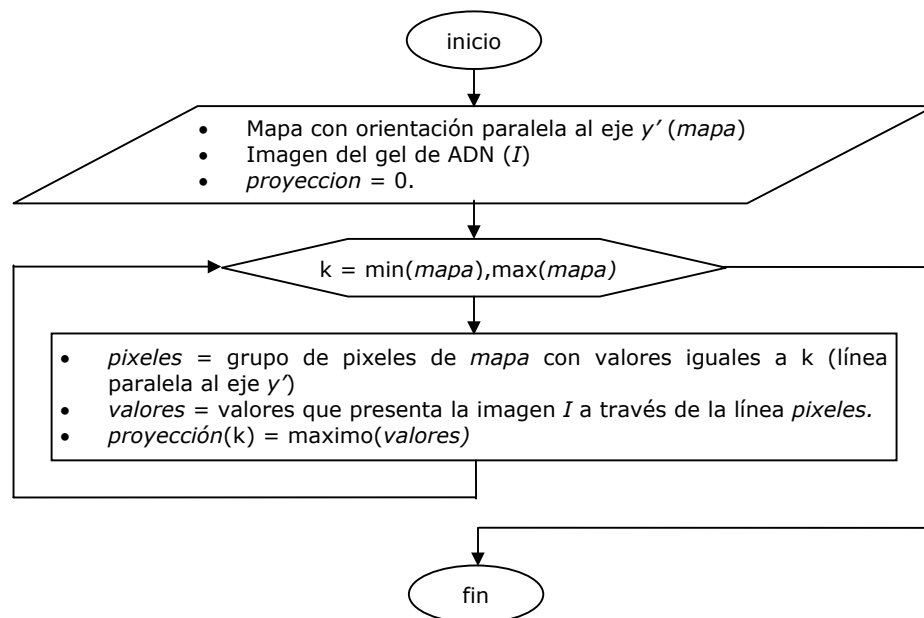
El procedimiento descrito anteriormente es útil únicamente cuando el marcador es una línea recta, y no aplica para otros tipos de marcadores (que en realidad no son de interés en el presente trabajo).

Se describe a continuación el método diseñado para calcular la proyección de la imagen sobre el eje x' .

3.1.3 Cálculo de la proyección de la imagen sobre el eje x' . En la sección 2.1.1.2. se analizó la desventaja de evaluar la proyección como la suma por columnas de la imagen, principalmente debido a que es posible que algunos carriles posean un bajo contenido de bandas y por lo tanto es probable que, mediante la suma por columnas, no sea posible diferenciar claramente dicho carril del fondo. Se propone entonces una modificación a este esquema para el cálculo de la proyección: además de tener en cuenta la orientación de la imagen, se evaluará la proyección como el valor máximo presente en una orientación determinada. Si fuese completamente vertical el carril, esto equivaldría a determinar el valor máximo presente en cada columna y asignar este valor como su correspondiente proyección. El procedimiento tiene la ventaja de funcionar de manera adecuada en carriles donde el contenido de fragmentos de ADN es reducido. Incluso, es probable que solo exista una banda en ese carril, sin embargo, el nivel de gris presente en esa banda (y por tanto el valor máximo a través de una orientación determinada) será mayor en comparación con el nivel de gris del fondo de la imagen.

Se debe recordar que un mapa es simplemente un conjunto de líneas paralelas, en este caso a los ejes x' o y' . Para el cálculo de la proyección de la imagen sobre el eje x' se hará uso del mapa cuyas líneas tienen una orientación es paralela al eje y' . Sin embargo, y con propósitos ilustrativos, se usará el mapa obtenido en la Figura 21. El procedimiento es iterativo y el diagrama correspondiente se ilustra en la Figura 22:

Figura 22. Algoritmo utilizado en el cálculo de la proyección de la imagen sobre el eje x'



La Figura 23 permite observar el funcionamiento de este algoritmo para el caso del mapa de la Figura 21 y asumiendo una imagen I como la ilustrada:

Figura 23. Ilustración del algoritmo usado para determinar la proyección de la imagen sobre el eje x'

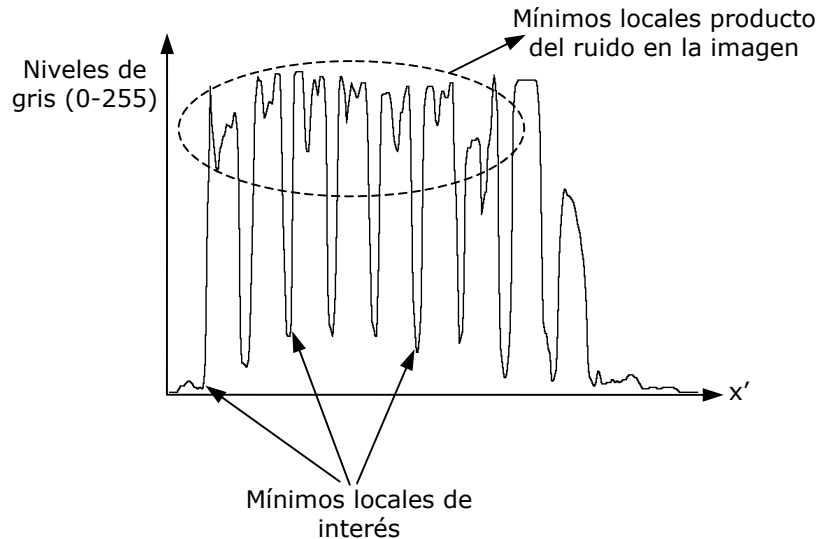
Mapa=			I =		
0	1	2	255	56	32
1	2	3	140	65	25
2	3	4	123	25	12
3	4	5	8	27	18
4	5	6	132	202	252

k	Píxeles en mapa iguales a k			Valores de I a través de la línea píxeles			Proyeccion(k+1)
0	0			255			255
1		1			56		140
	1			140			
2			2			32	123
		2			65		
	2			123			
3			3			25	25
		3			25		
	3			8			
4			4			12	132
		4			27		
	4			132			
5			5			18	202
		5			202		
	5						
6			6			252	252

El número total de iteraciones es siete (puesto que la variable k va desde 0, que es el mínimo de la matriz $mapa$, hasta 6, que es el máximo). Para este caso el vector proyección, también de siete componentes, es [255 140 123 25 132 202 252].

Cuando se aplica el algoritmo propuesto en la Figura 22 a una imagen típica de un gel de ADN, se obtiene la siguiente señal unidimensional (proyección):

Figura 24. Proyección sobre el eje x' de un gel de ADN típico



En la Figura 24 se observan dos características de interés en la proyección sobre el eje x' :

- El ruido presente en la imagen (introducido en la etapa de adquisición e igualmente debido a la misma constitución del gel y de las sustancias que acompañan los fragmentos de ADN), hace que existan mínimos locales que no corresponden en realidad a la zona limítrofe entre un carril y otro. Por el contrario, son zonas que se encuentran dentro de los carriles. Se hace necesario implementar entonces una etapa de suavizado, que elimine estos mínimos locales que no son de interés pero que a la vez conserve aquellos que delimitan un carril y otro.
- Los mínimos locales de interés (ubicados en la parte inferior de la proyección), no comparten el mismo nivel de gris (o al menos dichos niveles no son muy cercanos entre sí). Esto pone de manifiesto la iluminación no uniforme en la imagen, efecto que se ve acentuado en las zonas intermedias de la proyección (y por tanto de la imagen). Es igualmente necesario implementar una etapa de corrección de la tendencia en la iluminación de la imagen, como un procedimiento previo a la detección de los mínimos.

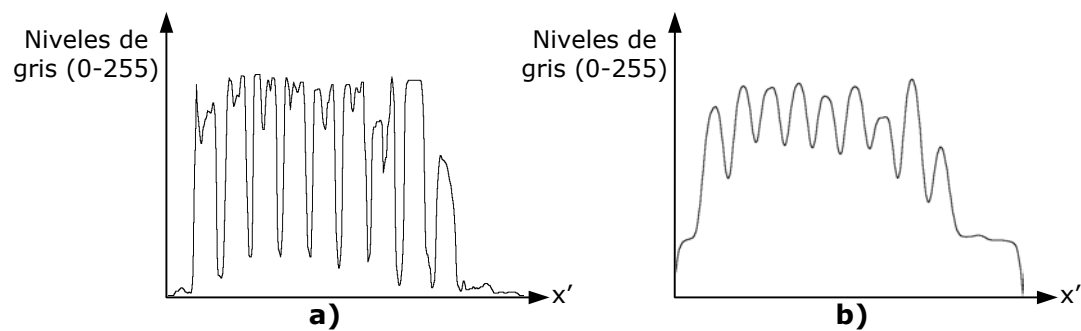
Se describe a continuación la etapa de suavizado.

3.1.4 Suavizado del patrón de proyección. Se pretende usar un procedimiento que permita eliminar los mínimos locales presentes dentro de los carriles, sin que se imposibilite o dificulte la detección de los mínimos locales existentes entre uno y otro

carril. Existe una gran variedad de procedimientos disponibles (como el filtrado por la media o el uso de técnicas de morfología matemática). En última instancia, se debe implementar un procedimiento que funcione de manera adecuada para absolutamente todas las imágenes, y por tanto el problema se reduce a la escogencia de un algoritmo cuyos parámetros (como el tamaño o el tipo de máscara, así como la cantidad de iteraciones) se ajusten a las características de las imágenes, y que no cambien de una a otra.

Se analiza en primer lugar el comportamiento del tipo de filtrado más simple: por la media^[2]. Para ello, se implementa un procedimiento similar al filtrado por la media en señales bidimensionales (imágenes); sin embargo, en este caso la máscara usada es de carácter vectorial y no matricial (específicamente, es de la forma $[1\ 1\ 1]/3$ o alguna similar, de mayor tamaño). El método es un proceso iterativo, donde la señal (proyección) es filtrada varias veces. El resultado del procedimiento, aplicado sobre la Figura 23, es el siguiente:

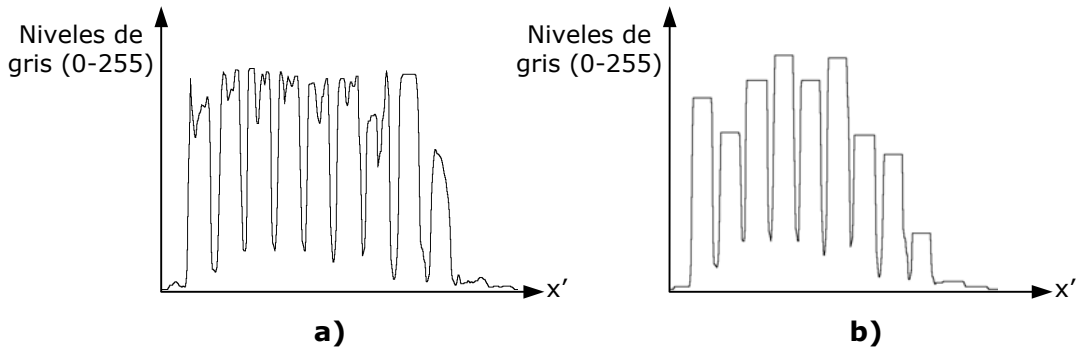
Figura 25. Resultado de aplicar el filtrado por la media a la señal original: a) Señal original; b) Resultado de filtrar la señal en (a)



Se observa que la señal no conserva su forma original; a pesar de ello es interesante notar que los mínimos locales de interés permanecen en su ubicación inicial (coordenada x'), en comparación con la señal original. El resultado ilustrado en la Figura 25b se obtiene utilizando una máscara de tamaño 1×5 y realizando diez iteraciones, o, de manera equivalente, utilizando una de tamaño 1×3 y 20 iteraciones. Desafortunadamente, estos parámetros (el tamaño de la máscara y el número de iteraciones) varían de una imagen a otra y no es posible determinar un par de valores (tamaño y número de iteraciones) que se adecue a las características de todas las imágenes.

Se considera a continuación el uso de un procedimiento de carácter morfológico, específicamente mediante la apertura de la señal original mediante una máscara de tamaño y forma adecuados. El resultado de este algoritmo se ilustra en la Figura 26, donde se ha usado un elemento de estructura (máscara) con forma de semicírculo de radio 10 píxeles:

Figura 26. Resultado de aplicar el filtrado morfológico a la señal original: a) Señal original; b) Resultado de filtrar la señal en (a)



Se observa que, al igual que en el caso anterior, los mínimos locales presentes dentro de los carriles han desaparecido, conservando solo aquellos que indican los límites intercarril. El resultado presenta la desventaja adicional, en comparación con el filtrado por la media, de requerir en ocasiones elementos de estructura con formas diferentes, dependiendo de las características de la señal original. La determinación de la forma que debe tener el elemento de estructura, es un proceso completamente manual (y en ocasiones bastante engorroso).

Se analiza a continuación un procedimiento basado en técnicas de difusión^{[15], [4], [21], [20]}, aplicadas principalmente al procesamiento de imágenes pero fácilmente extendibles a señales unidimensionales (como es el caso). Los algoritmos de difusión permiten remover el ruido presente en una imagen (o, en este caso, en una señal unidimensional) mediante la solución a una ecuación diferencial parcial. Específicamente, dichas técnicas de difusión pueden ser isotrópicas (o lineales) o anisotrópicas (no lineales). El nombre de la técnica (difusión) se basa en la analogía con varios procesos físicos de difusión (como la difusión de temperatura en una barra de metal, o la difusión en dos fluidos en contacto). Matemáticamente, es posible expresar el proceso de difusión como:

$$\frac{\partial I(r,t)}{\partial t} = \nabla \cdot (k \nabla I) \quad (\text{Ecuación 6})$$

donde I es la variable física (temperatura, flujo, etc...) o, en este caso, la señal de interés (la proyección), el símbolo $\nabla \cdot ()$ es el operador divergencia (una derivada parcial con respecto a las coordenadas espaciales – que para el caso de la Ecuación 6 es la variable r), y ∇I es el gradiente (o primera derivada) de la señal. k es una constante.

Se asume entonces que I (la proyección) es función de las coordenadas espaciales (en este caso x' juega el papel de r en la Ecuación 6) y t es un parámetro artificial que indica el número de iteración en el algoritmo (la señal I cambia de una a otra iteración). La ecuación 6 es la ecuación clásica de difusión isotrópica o lineal (puesto que k es una constante). Dicha ecuación se diferencia del proceso de difusión anisotrópica, en que k es una función (no lineal) dependiente de ∇I :

$$\frac{\partial I(r,t)}{\partial t} = \nabla \cdot (f(\nabla I) \nabla I) \quad (\text{Ecuación 7})$$

Dicha función $f(\nabla I)$ recibe el nombre de difusividad, y puede tener variedad de formas, de las cuales, la de uso más extendido es:

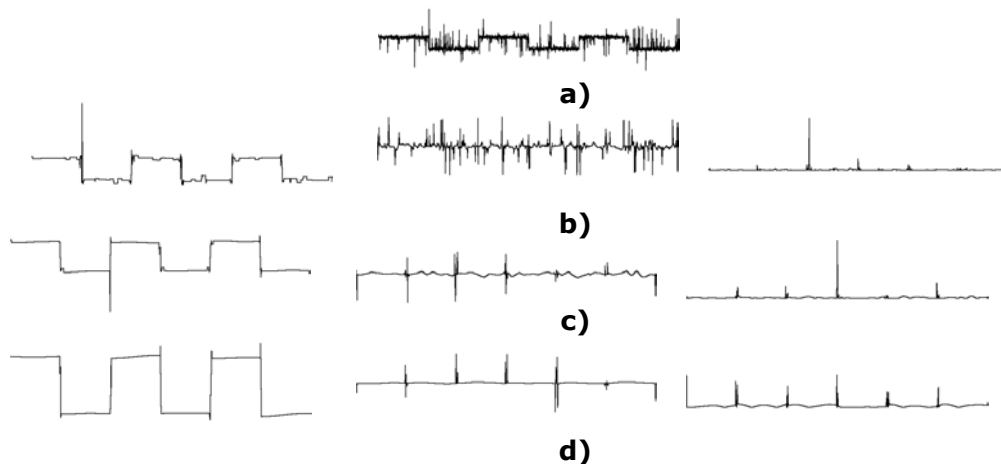
$$k = f(\nabla I) = 1 - e^{\frac{-C}{(\nabla I/\lambda)^m}} \quad (\text{Ecuación 8})$$

Se debe tener en cuenta que el gradiente de la señal (∇I) se aplica sobre una versión suavizada de la misma:

$$\bar{\nabla} I = \bar{\nabla} \left(\frac{1}{(2\pi\sigma^2)^{m/2}} e^{-\frac{|x|^2}{\sigma^2}} \times I \right) \quad (\text{Ecuación 9})$$

Así, la difusión lineal es el simple proceso de suavizado de la señal I mediante una función de forma gaussiana. Por su parte, la difusión no lineal facilita el suavizado de regiones donde la magnitud del gradiente es relativamente pequeña (en comparación con el parámetro λ de la ecuación 8). Sin embargo, en zonas donde el gradiente es mucho mayor (como aquellas en donde se encuentra un mínimo local), el efecto de suavizado es despreciable. Esto se muestra en la Figura 27, donde se ilustra la función original, la función suavizada (producto de resolver la ecuación 7), el flujo (que corresponde al producto $f(\nabla I)\nabla I$) y el nuevo valor correspondiente a $\partial I/\partial t$:

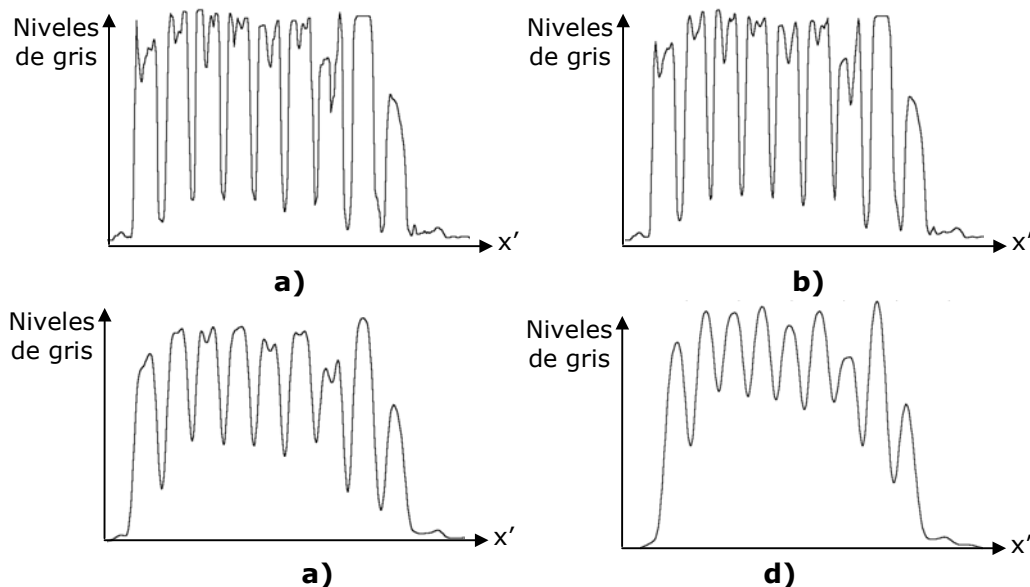
Figura 27. Ilustración del proceso de difusión no lineal (primera columna: señal restaurada; segunda columna: flujo; tercera columna: $\partial I/\partial t$): a) Señal original; b) Tras una iteración; c) Tras tres iteraciones; d) Tras cinco iteraciones



Como se mencionó anteriormente, el efecto neto del proceso de difusión no lineal es restaurar la señal ideal, conservando además los cambios abruptos. Es, por tanto, un procedimiento ideal para el suavizado de la proyección, ya que permite conservar los mínimos locales de interés, eliminando a la vez el ruido presente en las zonas que se encuentran dentro de un carril. Sin embargo, se encuentran algunas desventajas en la

aplicación de este algoritmo. Específicamente, y observando las ecuaciones 8 y 9, se puede afirmar que la complejidad es mayor en comparación con la de algoritmos como el filtrado por la media o el morfológico, en términos del número de variables que se deben manejar. En este caso aparecen nuevas constantes que deben ser determinadas de acuerdo a las características de la señal: σ (la desviación standard de la gaussiana, según la ecuación 9) y los parámetros λ y m de la ecuación 8 (el parámetro C de esta ecuación se calcula previamente con base en un procedimiento descrito en ^[15] y ^[21]). Se tienen entonces cuatro parámetros (si se tiene en cuenta que el proceso es iterativo y por tanto es necesario definir el número de iteraciones), en comparación con los dos de los casos anteriores (el tamaño o forma de la máscara y el número de iteraciones). Se opta entonces por una aproximación equivalente, pero que no involucra tal cantidad de variables. El procedimiento implementado corresponde entonces a un filtrado por difusión lineal modificado, donde no se hace uso del filtro gaussiano como etapa previa a la evaluación del gradiente de la señal. Se hará uso entonces de la ecuación 6, y se asumirá la constante k como 1. Al hacer lineal el proceso, se evita el uso de la función $f(\nabla I)$ y por tanto el cálculo de parámetros como m , λ e igualmente C . De igual manera, no es necesario determinar el valor del parámetro σ (que no es el mismo para todas las señales, sino que varía dependiendo de sus características). De esta manera, el único parámetro de interés en el algoritmo planteado es el número de iteraciones. A continuación se ilustra la aplicación del procedimiento descrito sobre la señal de la Figura 24:

Figura 28. Proceso de difusión lineal (modificado) aplicado a la proyección sobre el eje x' : a) Señal original; b) Tras 7 iteraciones; c) Tras 15 iteraciones; d) Tras 20 iteraciones

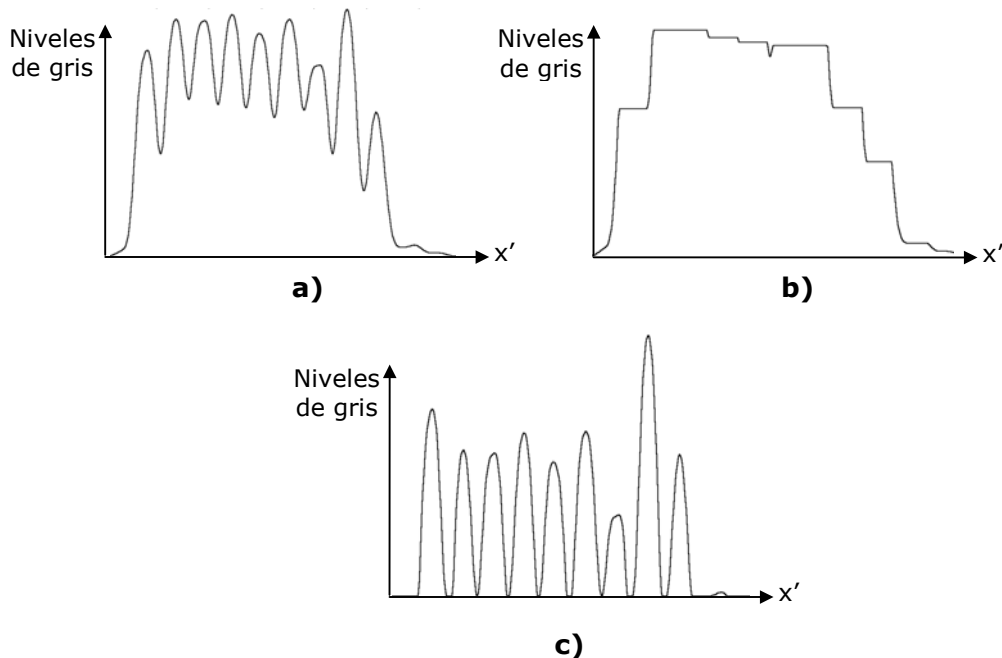


Se observa que el resultado obtenido es similar al caso en el que se usa el filtrado por la media. Sin embargo, la ventaja en este caso es que depende de una sola variable: el número de iteraciones. Tras aplicar la difusión no lineal a cada una de las proyecciones disponibles (provenientes de 18 imágenes), se determinó el número de iteraciones adecuado (20 en total) que funciona en la mayor parte de ellas.

El siguiente paso busca corregir la tendencia de la proyección, de tal manera que todos los mínimos locales se encuentren al mismo nivel (o al menos cercano entre sí). El procedimiento utilizado implementa un algoritmo basado en técnicas de morfología matemática.

3.1.5 Corrección de la tendencia de la iluminación en la proyección. El procedimiento es bastante simple, y parte del hecho de que al calcular la apertura morfológica de la señal suavizada (Figura 28d), es posible obtener precisamente la tendencia que poseen los niveles de gris presentes en dicha proyección. La señal con una tendencia de iluminación uniforme es el resultado de restar a la imagen original (suavizada), el patrón de iluminación obtenido gracias a la apertura morfológica de la misma, procedimiento que se ilustra a continuación:

Figura 29. Ilustración del procedimiento que permite corregir la tendencia en la iluminación de la proyección: a) Señal original (suavizada); b) Tendencia en el patrón de iluminación de a; c) Señal obtenida tras aplicar el procedimiento morfológico

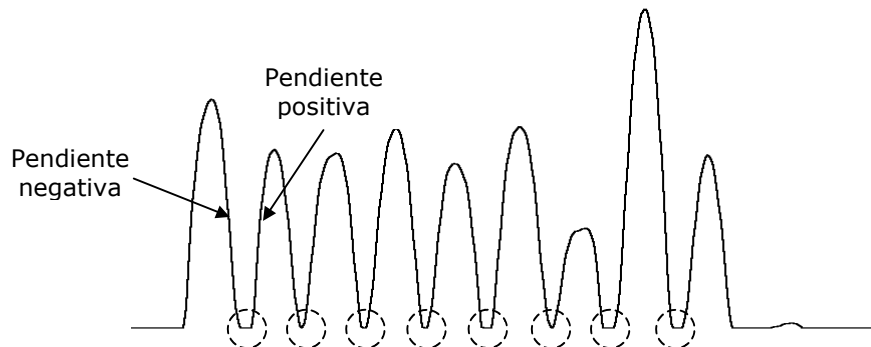


Se debe tener en cuenta que el elemento de estructura, usado en la apertura morfológica, debe tener un tamaño mayor que la distancia existente entre uno y otro pico (y que para el caso de las diferentes imágenes analizadas es un semicírculo de radio 20 píxeles).

Tras obtener una señal como la ilustrada en la Figura 29c, se procede a determinar la ubicación de los mínimos locales, lo cual, en última instancia, permite segmentar los carriles presentes en la imagen.

3.1.6 Estimación de los mínimos locales. Para ello se evalúa el gradiente, primera derivada, de la señal (Figura 29c), teniendo en cuenta que un mínimo local se encuentra en una zona donde dicho gradiente cambia de negativo a positivo:

Figura 30. Los diferentes mínimos locales presentes en la proyección están determinados por el cambio en el signo de la pendiente (negativo/positivo).



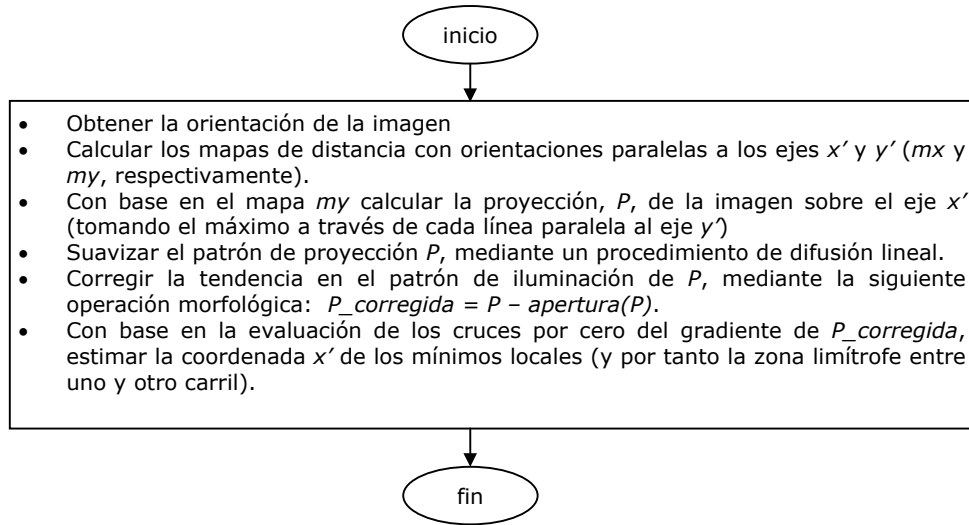
Así, evaluando aquellas zonas del gradiente que tengan las secuencias negativo/positivo en su signo (o, equivalentemente, los cruces por cero en esta primera derivada), es posible determinar por completo la coordenada x' de cada uno de los mínimos presentes en la proyección. Esta etapa permite la segmentación adecuada de los carriles en la mayor parte de las imágenes (característica que se analizará en el capítulo 5). En ocasiones es necesaria la intervención del usuario, debido a alguna de las siguientes razones:

- Es probable que durante el proceso de preparación dos o más carriles se intersecten en una o varias zonas del gel, lo cual conlleva a la eliminación del mínimo local que en principio debería existir en la proyección. Es necesario que el usuario determine la zona que divide un carril de otro.
- La existencia de burbujas o algún tipo de imperfección en el gel o en la muestra, origina la posible aparición de mínimos locales en zonas que se encuentran dentro de un carril.

Por tanto, se habla de una etapa de segmentación de carriles semiautomática, puesto que el proceso de electroforesis es de cierta manera irreplicable, en el sentido de que uno y otro gel, incluso con las mismas muestras de ADN, no son necesariamente idénticos (diferencias que radican, por ejemplo, en factores como la preparación de las muestras de ADN, así como el tiempo de aplicación del campo eléctrico y su intensidad, o la cantidad de Bromuro de Etidio presente en los fragmentos del material genético).

En la Figura 31 se resume el algoritmo que permite segmentar los carriles, partiendo inicialmente de la imagen del gel:

Figura 31. El algoritmo utilizado en la segmentación de carriles

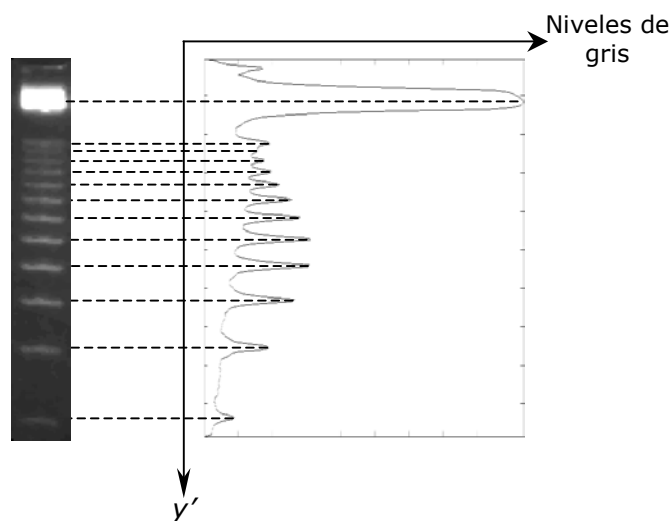


La segmentación correcta de los carriles permite afrontar la siguiente etapa: la detección de bandas y posterior estimación del tamaño de las mismas.

3.2 DETERMINACIÓN DE LA UBICACIÓN DE LAS BANDAS

Se basa en un procedimiento similar a la obtención de la proyección de la imagen sobre el eje x' . La diferencia radica en que para este caso se toma cada uno de los carriles y se evalúa su proyección sobre el eje y' (tomando como base el mapa de distancias cuyas líneas poseen una orientación paralela al eje x'). El resultado de evaluar dicha proyección para un carril determinado, se ilustra en la Figura 32:

Figura 32. Un carril y su proyección sobre el eje y' .



Se observa una perfecta coincidencia entre los máximos locales y la ubicación de las bandas. Se procede entonces a determinar la posición exacta (coordenada y') de dichos máximos locales, con el fin de estimar posteriormente el tamaño de cada fragmento de ADN.

3.2.1 Obtención de la coordenada y' de los máximos locales en la proyección del carril. El procedimiento es idéntico al usado en la determinación de los mínimos locales de la proyección de la imagen. Sin embargo, la diferencia en este caso radica en que la detección está orientada a establecer la coordenada y' de los máximos, y no los mínimos, locales.

Antes de aplicar el algoritmo es necesario suavizar la proyección, de tal manera que los posibles componentes de ruido presentes en la misma sean eliminados o atenuados en su mayor parte. Para ello se usa nuevamente el proceso de difusión lineal, esta vez con un número variable de iteraciones, que puede ser ajustado por el usuario (ver capítulo 4). Esto se debe a que en ocasiones algunas bandas tienen una tonalidad gris imperceptible (debido a la cantidad de Bromuro de Etidio adherida a ellas), que pueden no ser tenidas en cuenta si se implementa un proceso de suavizado automático. Es el personal del Laboratorio quien ajusta la cantidad de iteraciones de tal manera que se obtengan los resultados deseados.

Después de dicho suavizado, se determina la ubicación de los máximos locales, con base en la obtención de la primera derivada y en la determinación de los cruces por cero de la misma, teniendo en cuenta los cambios de signo positivo/negativo, que corresponden precisamente a la ubicación de los máximos locales.

Tras aplicar el procedimiento a todos y cada uno de los carriles presentes en la imagen, se cuenta con la información necesaria para calcular la función (lineal, logarítmica, etc...) que mejor se ajuste a los datos provenientes del carril standard. Finalmente, dicha información (curva de calibración) se aplicará a los demás carriles para así obtener el tamaño de cada fragmento de ADN. Existe gran variedad de algoritmos para la obtención de la curva de calibración y en la sección 4 se dan más detalles al respecto.

Se han descrito hasta el momento los algoritmos implementados útiles en la segmentación de la imagen, la detección de bandas y su localización precisa. Dichos algoritmos se basan en la correcta estimación de los ejes principales de la imagen (en contraste con los trabajos realizados hasta el momento, donde se asume que estos ejes principales coinciden con los ejes x, y convencionales), hecho que se traduce en un sistema mucho más preciso a la hora de estimar el tamaño de los segmentos.

4. IMPLEMENTACIÓN DEL ALGORITMO E INTERFAZ GRÁFICA

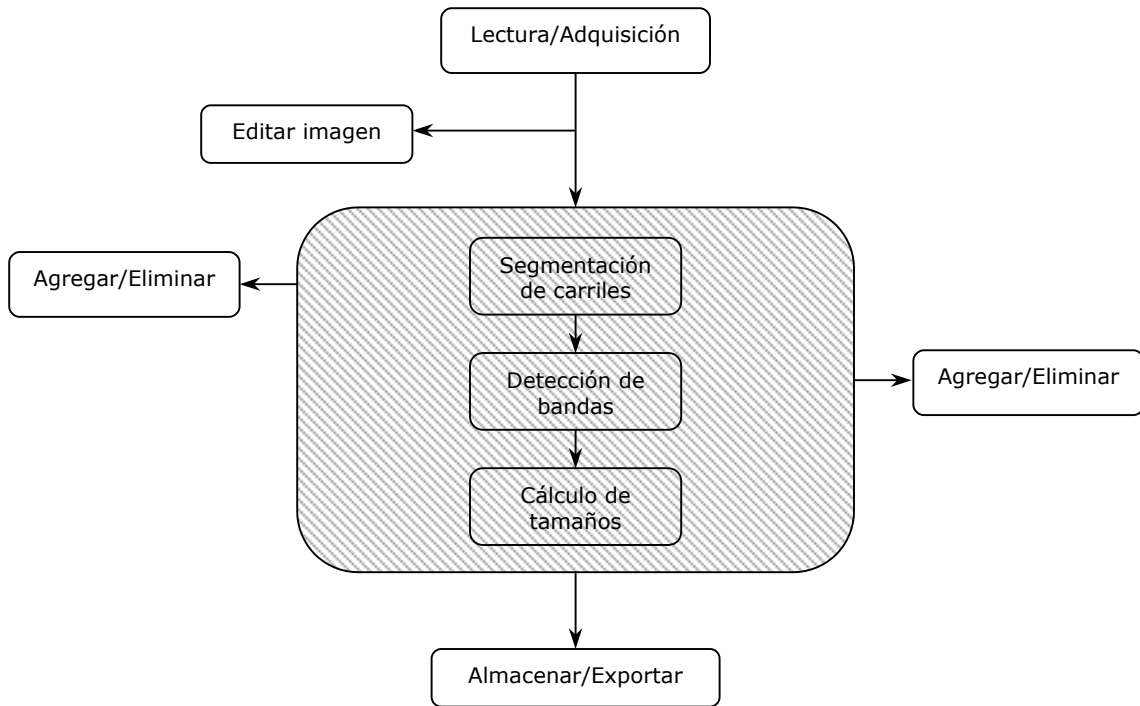
Los procedimientos descritos en la sección 3, que permiten en última instancia determinar el tamaño de cada fragmento de ADN presente en el gel, fueron implementados usando la herramienta computacional Matlab® versión 7.0. De igual forma, la interfaz gráfica, entregada al Laboratorio de Genética, está implementada en este programa. Las razones que sustentan esta decisión son las siguientes:

- Se cuenta con un equipo computacional adecuado para el desarrollo de aplicaciones usando Matlab 7.0.
- El Laboratorio de Genética se encuentra llevando a cabo los trámites necesarios para la consecución de la licencia de Matlab 7.0, con el fin de implementar, además del presente sistema, otra aplicación previamente elaborada con base en esta herramienta computacional (orientadas al análisis de cromosomas) y futuras aplicaciones relacionadas con el análisis de imágenes biológicas.
- Matlab 7.0 ofrece versatilidad en el desarrollo de sistemas de adquisición de imágenes, ya que permite el uso de diferentes dispositivos de adquisición (desde tarjetas Matrox hasta cámaras USB).
- Se cuenta con un sistema abierto a mejoras futuras e inmediatas, puesto que, en términos de programación, es menos complejo comparado con lenguajes de alto nivel como C++. Fácilmente se podrían implementar nuevos módulos (en caso de ser necesarios), como por ejemplo el análisis de geles bidimensionales, así como algunas opciones adicionales de interés para los usuarios de la aplicación.

Desde luego, si se desea desarrollar una aplicación de carácter comercial es necesario implementar los algoritmos en una aplicación como Visual C++, que permita obtener una aplicación independiente del entorno de Matlab. Sin embargo, y por el momento, este no es el interés.

En la Figura 33 se puede observar un diagrama de bloques con los componentes principales del sistema desarrollado:

Figura 33. Diagrama de bloques del sistema implementado en Matlab



Los bloques de color gris son módulos del sistema a los cuales el usuario no tiene acceso. Sin embargo, en ocasiones es necesaria la adición o eliminación bien sea de carriles o de bandas, para lo cual se requiere la selección manual del carril y/o la(s) banda(s) de interés.

A continuación se describen brevemente las principales características de cada módulo:

4.1 MÓDULO "LECTURA/ADQUISICIÓN"

Permite abrir una imagen, o un proyecto (un proyecto contiene la información concerniente a los carriles y las bandas así como el tamaño de cada fragmento, además de la imagen correspondiente), previamente almacenada en el disco duro. Los formatos de imagen aceptados son TIFF y JPEG.

La opción de adquisición permite tener acceso a la información proveniente de una cámara USB o de una tarjeta (el programa está configurado para dispositivos Matrox). Tras adquirir la imagen es posible almacenarla en el disco duro, o ejecutar las etapas siguientes del análisis y, posteriormente, almacenar el proyecto.

4.2 MÓDULO "EDITAR IMAGEN"

Cuenta con opciones adicionales para la manipulación de los niveles de gris de la imagen adquirida o previamente almacenada. Permite aplicar procedimientos para la mejora del contraste, la gamma o el brillo. De igual forma, es posible el uso de

algoritmos que permiten la eliminación del ruido (el suavizado de la imagen) o la mejora del contraste mediante procedimientos morfológicos (transformada *Top-Hat*).

4.3 MÓDULO "SEGMENTAR CARRILES"

En el se encuentran implementados todos los algoritmos descritos en la sección 3.1. Cuenta con la opción adicional de agregar o eliminar, de forma manual, carriles que no sean de interés o que se encuentren segmentados de forma incorrecta.

4.4 MÓDULO "DETECTAR BANDAS"

Al igual que el anterior, implementa los algoritmos descritos en la sección 3.2. Permite la adición o eliminación manual de algunas bandas e incluye, además, una opción para que el usuario establezca el número de iteraciones a utilizar en el algoritmo de suavizado de la proyección.

4.5 MÓDULO "CÁLCULO DE TAMAÑOS"

Tras el uso de los módulos anteriores se cuenta con información correspondiente a la ubicación de las bandas dentro del gel junto con el tamaño de los fragmentos pertenecientes al carril estándar. Con ayuda del *Toolbox curvefitting*, incluido en la versión 7.0 de Matlab, se evalúan diferentes ajustes y su respectivo error cuadrático medio. Específicamente, se lleva a cabo el cálculo para siete tipos de curvas:

- Lineal ($y = ax + b$, con a y b como constantes).
- Cuadrática ($y = ax^2 + b$, con a y b como constantes).
- Cúbica ($y = ax^3 + b$, con a y b como constantes).
- Exponencial ($y = ae^{bx}$, con a y b como constantes).
- Potencias (de dos formas: $y = ax^b$ e $y = ax^b + c$, con a , b y c como constantes).
- Un ajuste que obedece a la ecuación $y = a/x + b$, con a y b como constantes.

Como se mencionó, para cada caso se evalúa el error cuadrático medio y se elige, de manera automática, sin intervención del usuario, aquel ajuste que arroja el parámetro de menor valor (es decir, el mejor ajuste en este sentido). Finalmente, la función obtenida se evalúa para las coordenadas de cada una de las bandas presentes en la imagen, determinando de esta manera el tamaño de cada fragmento de ADN presente en el gel. El resultado obtenido es mostrado al usuario y se incluye la opción de, además de almacenar el análisis como un proyecto, exportar los datos correspondientes al tamaño de los fragmentos a una hoja de cálculo para posteriormente ser visualizados en Excel.

La aplicación implementada permite obtener el tamaño de cada fragmento presente en el gel en un tiempo que oscila entre los 2 y los 4 minutos, lo cual representa una mejora significativa en comparación con un análisis completamente manual (que es más impreciso y no permite analizar cada uno de los fragmentos, puesto que el usuario debe reducir su tiempo de exposición a la radiación UV a un máximo de cinco minutos).

5. RESULTADOS EXPERIMENTALES

El presente capítulo se divide en tres secciones: resultados obtenidos en la segmentación de carriles y en la detección de bandas, así como un análisis de las fuentes de error generadas por algoritmos inadecuados para la segmentación, la detección de bandas y que influyen en la estimación del tamaño de los fragmentos de ADN.

5.1 SEGMENTACIÓN DE CARRILES

En este caso se usó un grupo de 18 imágenes, con diferentes características, adquiridas previamente en el Laboratorio de Genética. Los parámetros medidos en cada una de las imágenes, fueron:

- Porcentaje de aciertos: corresponde a aquellos carriles segmentados correctamente. Está dado por la relación $(\text{total aciertos}/\text{cantidad de carriles}) \times 100\%$.
- Porcentaje de desaciertos: son aquellos carriles que no fueron segmentados. Se obtiene a partir de la relación $(\text{total desaciertos}/\text{cantidad de carriles}) \times 100\%$.
- Cantidad de falsos aciertos: aquellas zonas de la imagen que no corresponden a ningún carril y que sin embargo fueron segmentadas como uno de ellos. Se indica el número de falsos aciertos obtenidos por imagen.
- Cantidad de carriles segmentados parcialmente: es posible que el proceso de segmentación haya dado origen a uno de dos posibles efectos: la combinación de dos carriles en uno solo o la división de un carril en dos partes. Se indica el número de carriles segmentados parcialmente presentes en cada imagen (este parámetro se incluye ya que, como se verá en la sección 7, implementación de la aplicación, la segmentación correcta de carriles segmentados parcialmente es bastante sencilla).

Los resultados se han tabulado en la Tabla 2, donde se ha medido el desempeño de tres algoritmos: el implementado en el presente trabajo, y aquellos en los que se basan dos de las aplicaciones comerciales (versión demostración) descritas en la sección 2.

Tabla 2. Resultados de la segmentación de carriles tras la aplicación del algoritmo planteado

Imagen número	Cantidad de carriles	Porcentaje de aciertos			Porcentaje de desaciertos			Cantidad de falsos positivos			Carriles segmentados parcialmente		
		AP	S1	S2	AP	S1	S2	AP	S1	S2	AP	S1	S2
1	9	100	100	100	0	0	0	0	0	0	0	0	0
2	9	100	100	100	0	0	0	0	0	1	0	0	0
3	10	100	100	50	0	0	50	0	0	0	0	0	3
4	9	100	100	66.7	0	0	33.3	1	0	2	0	0	3
5	8	100	100	0	0	0	100	1	1	0	0	0	7
6	10	10	80	100	0	20	0	0	0	0	2	2	0
7	9	88.9	100	0	11.2	0	100	0	0	2	0	0	9
8	10	100	100	50	0	0	50	0	0	0	0	0	4
9	6	100	83.3	0	0	16.7	100	4	0	0	0	1	0
10	10	100	70	60	0	30	40	4	2	9	0	0	1
11	9	100	100	33.3	0	0	66.7	5	0	0	0	0	1
12	3	100	100	0	0	0	100	6	16	0	0	0	0
13	10	100	20	0	0	80	100	0	0	0	0	7	0
14	8	100	100	0	0	0	100	1	9	0	0	0	0
15	10	90	90	60	10	10	40	0	0	1	0	1	1
16	9	100	100	66.7	0	0	33.3	0	0	0	0	0	3
17	9	100	100	0	0	0	100	0	0	0	0	0	6
18	7	100	0	42.9	0	100	57.1	1	0	0	0	5	2
Promedios		93.83%	85.74%	40.53%	1.17%	14.26%	59.46%	1.27	1.55	0.83	0.11	0.88	2.22

AP = Algoritmo planteado en el presente trabajo
 S1 = Sistema comercial 1
 S2 = Sistema comercial 2

Se observa una ventaja sustancial del algoritmo planteado en este trabajo en comparación con las aplicaciones de tipo comercial, específicamente en aspectos como el porcentaje de aciertos (93.83%, 85.74% y 40.53% en promedio, respectivamente) y el porcentaje de desaciertos (1.17%, 14.26% y 59.46% en promedio, respectivamente). Por otra parte, la cantidad, promedio, de falsos positivos y de carriles segmentados parcialmente es similar, y el comportamiento de los tres sistemas es equivalente en este sentido. Es imposible establecer las razones que originan esta diferencia con respecto a los sistemas de carácter comercial, puesto que no se tiene un conocimiento detallado de los algoritmos que dichos sistemas utilizan en la extracción de los carriles. Se debe tener en cuenta, igualmente, que se está hablando únicamente de dos sistemas de carácter comercial; es probable que otros ofrezcan un mejor desempeño que las aplicaciones evaluadas. Sin embargo, se debe resaltar que los resultados obtenidos con el uso del método implementado en el presente trabajo son bastante aceptables.

Se describen a continuación los resultados obtenidos en la siguiente etapa: la detección de bandas.

5.2 DETECCIÓN DE BANDAS

En el análisis descrito a continuación se asume, por obvias razones, que la segmentación de carriles ha sido llevada a cabo exitosamente (todos los carriles presentes en la imagen están segmentados correctamente); de no ser así es posible corregir el proceso de segmentación de forma manual (tal como se mencionó en el capítulo 4). En este caso se pretende determinar la efectividad del método descrito, basado en la estimación de la correcta orientación de la imagen, en comparación con aquellos algoritmos descritos en la bibliografía, donde se asume que los carriles tienen una orientación vertical. Como se mencionó anteriormente, la correcta detección de las bandas presentes en la imagen redundará desde luego en la estimación precisa del tamaño de cada uno de los fragmentos de ADN presentes en el gel. Cabe anotar que en dicha estimación del tamaño de las bandas influyen igualmente las características propias de la imagen (como por ejemplo la distribución de los carriles y las bandas a través de la misma), tema que será analizado más adelante.

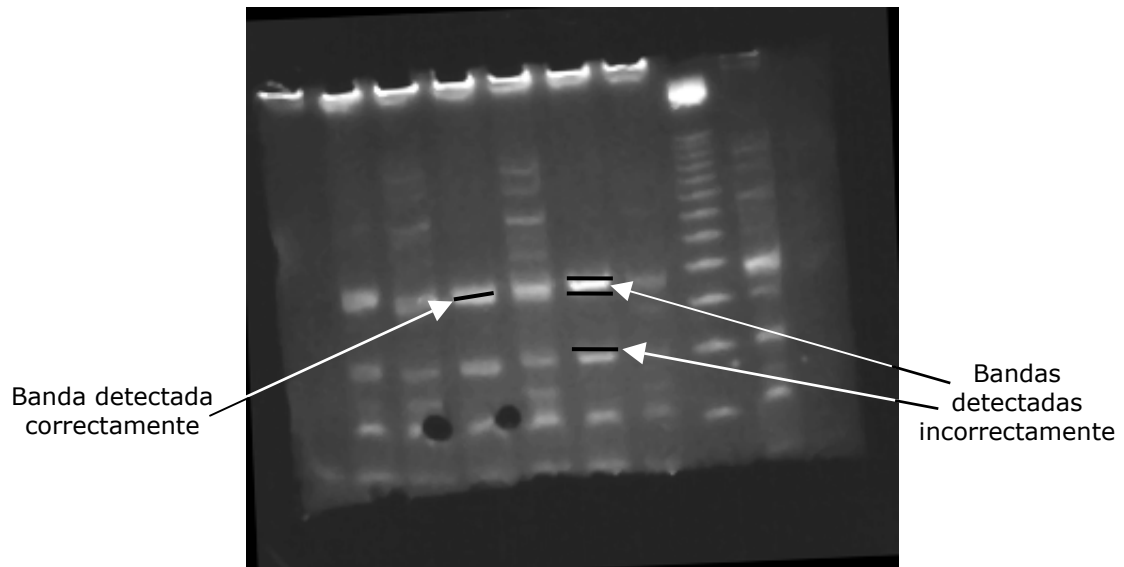
Se llevaron a cabo dos tipos de análisis: uno de ellos partiendo del grupo de imágenes usado en la sección anterior (la obtención de carriles). Sin embargo, se debe aclarar que en este caso el grupo se ve reducido de 18 a 9 imágenes, que son las que presentan algún tipo de orientación diferente de cero (se entiende la orientación de la imagen como el ángulo que forma el eje x' con el eje x convencional) de tal forma que los carriles dejan de ser completamente verticales. Se usa entonces el algoritmo descrito en el presente trabajo y uno que no toma en cuenta la orientación de la imagen. En muchos casos los resultados son similares, en términos de porcentaje de aciertos/errores (esto ocurre específicamente cuando la orientación es muy pequeña, cercana a 0°).

Para el segundo tipo de análisis, se tomó una imagen típica proveniente de un gel de ADN y se estimó la ubicación de las bandas presentes en el mismo para diferentes versiones de la imagen, cada una de las cuales es el resultado de aplicar una rotación a la imagen original. Se comparan, nuevamente, los resultados existentes entre los dos algoritmos: el descrito en el presente trabajo y aquellos que asumen una orientación completamente vertical de los carriles^{[13],[1],[11],[5]}.

5.2.1 Resultados obtenidos sobre un grupo de nueve imágenes. Se analizan tres aspectos en estos resultados:

- Porcentaje de aciertos: se evalúa sobre cada carril. Se calcula como el número de bandas detectadas dividido entre el número total de bandas existentes en el carril. Para que una banda sea catalogada como "correctamente detectada" (un acierto), el algoritmo debe indicar exactamente el lugar de la imagen donde se encuentra dicha banda, hecho que se ilustra en la figura 34:

Figura 34. Una banda detectada correctamente (izquierda) y otras detectadas de forma incorrecta (derecha)



Más adelante se analizarán las fuentes de error que impiden una detección correcta de las bandas (dicho fenómeno se presenta únicamente al usar el algoritmo de detección de carriles asumiendo que estos son completamente verticales).

- Porcentaje de desaciertos: corresponde simplemente al cociente entre el número de bandas que no fueron detectadas y el número total de bandas presentes en cada carril.
- Falsos aciertos: es la cantidad de lugares que fueron marcados como bandas y en donde en realidad no existe ninguna de ellas. Sus orígenes, para el caso del algoritmo planteado en el presente trabajo, están en el ruido de la imagen (debido al proceso de preparación del gel o a la etapa de adquisición). Más adelante se discutirán en detalle las fuentes de error para el caso en el que se asumen carriles completamente verticales.

Todos los parámetros (porcentaje de aciertos, porcentaje de desaciertos y falsos aciertos) son tabulados a continuación. Cabe anotar que, para cada imagen, los porcentajes de aciertos y falsos aciertos han sido promediados, esto es: se evalúa el parámetro para cada una de las bandas de la imagen y luego se promedia y se expresa

como un porcentaje. Los falsos aciertos corresponden a la suma de falsos aciertos de cada una de las bandas. Los datos se han organizado de acuerdo a la orientación de la imagen, de tal manera que se puedan analizar fácilmente las ventajas del método implementado.

Tabla 3. Resultados de la detección de bandas tras la aplicación del algoritmo planteado (método 1) y su comparación con los resultados obtenidos mediante el uso del algoritmo convencional (método 2)

Imagen Número	Orientación* (grados)	% Aciertos		% Desaciertos		Cantidad de falsos aciertos	
		Método 1	Método 2	Método 1	Método 2	Método 1	Método 2
3	0	94.27	94.27	5.73	5.73	6	6
4	0.3745	100	96.29	0	3.71	13	16
9	0.4494	78.78	77.27	21.22	22.73	14	16
6	0.9094	97.5	96.66	2.5	3.34	4	4
2	1.4976	86.74	68.83	13.26	31.17	6	6
1	1.9481	98.33	82.59	1.67	17.41	8	10
7	2.6195	96.11	78.50	3.89	21.50	4	13
5	3.3665	82.27	72.12	17.73	27.88	11	23
8	4.2105	88.08	72.27	11.92	27.73	2	4

- * Orientación: es el ángulo que forma el eje x' de la imagen con respecto al eje x convencional
- La orientación de las imágenes (segunda columna) se determinó de tal forma que se apreciaran diferencias notables entre uno y otro método. Es por ello que dicha orientación no posee una distribución uniforme
- La cantidad de iteraciones utilizada en el suavizado de la proyección fue igual a 3, esperando con esto obtener una gran cantidad de falsos aciertos (a medida que la cantidad de iteraciones aumenta se cuenta con una proyección del carril mucho más suave, reduciendo así la cantidad de falsos aciertos)

Se comprueba inicialmente la hipótesis planteada con anterioridad: el cálculo de la orientación de la imagen, y su posterior uso en la determinación de la ubicación de las bandas, ofrece un mejor desempeño, en términos de porcentaje de aciertos y desaciertos, en comparación con el método convencional que asume una orientación de los carriles perfectamente vertical. Para ello basta comparar las columnas 3, 4 y 5,6 de la tabla 3. Se observa que, a medida que la orientación de los carriles deja de ser completamente vertical (es decir, se aleja de los 0°), el porcentaje de aciertos tras usar el segundo método disminuye paulatinamente, mientras que no ocurre lo mismo cuando se usa el método descrito en el presente trabajo.

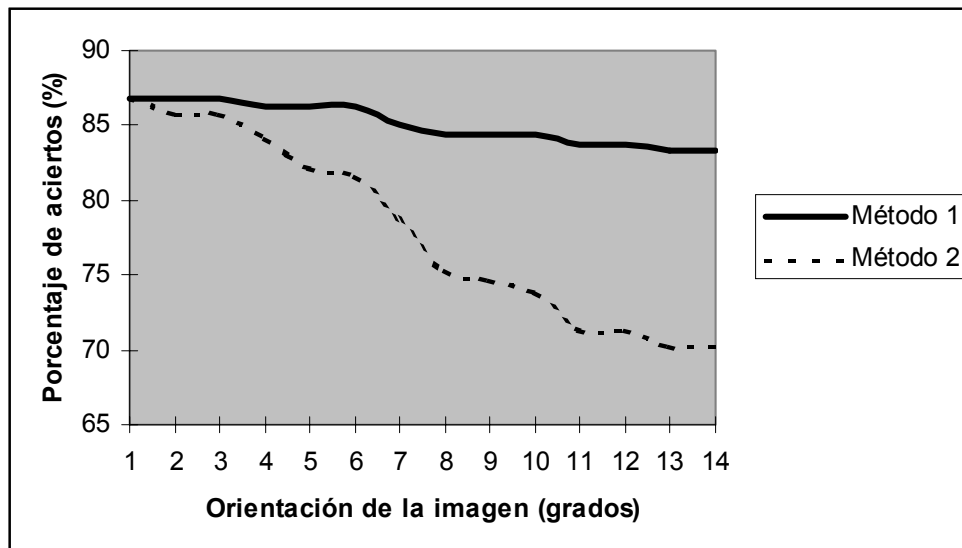
5.2.2 Resultados obtenidos en una imagen específica. Para ilustrar de forma más clara los resultados obtenidos anteriormente, se tomó una imagen de un gel de ADN y, de manera artificial, mediante la aplicación de la función *imrotate* de Matlab, se cambió su orientación dentro de cierto rango de valores (específicamente entre 0° y 11°). Para cada una de las versiones rotadas de la imagen se evaluaron los mismos parámetros ilustrados en la Tabla 3, obteniendo los siguientes resultados:

Tabla 4. Parámetros de desempeño evaluados para una imagen con diferentes orientaciones, usando dos métodos diferentes

Orientación (grados)	% Aciertos		% Desaciertos		Cantidad de falsos aciertos	
	Método 1	Método 2	Método 1	Método 2	Método 1	Método 2
0	86.74	86.74	13.26	13.26	6	6
0.5	86.74	85.71	13.26	14.29	6	6
0.7	86.74	85.71	13.26	14.29	6	6
1.0	86.23	84.12	13.77	15.88	6	6
1.5	86.23	82.05	13.77	17.95	6	8
1.8	86.23	81.53	13.77	18.47	6	8
3.0	85.12	78.66	14.88	21.34	6	9
4.5	84.33	75.23	15.67	24.77	6	11
5.5	84.33	74.66	15.67	25.34	6	11
6.8	84.33	73.86	15.67	26.14	6	11
8.0	83.76	71.33	16.24	28.67	7	14
9.3	83.76	71.33	16.24	28.67	7	14
10.5	83.33	70.16	16.67	29.84	7	14
11.0	83.33	70.16	16.67	29.84	7	14

A continuación se esboza una gráfica Porcentaje de aciertos vs. Orientación, donde se comparan los métodos 1 y 2:

Figura 35. Desempeño de los métodos 1 y 2 para una imagen específica, en términos del porcentaje de aciertos



Se observa claramente la robustez del método 1 (implementado en el presente trabajo), ya que el porcentaje de aciertos permanece prácticamente constante a

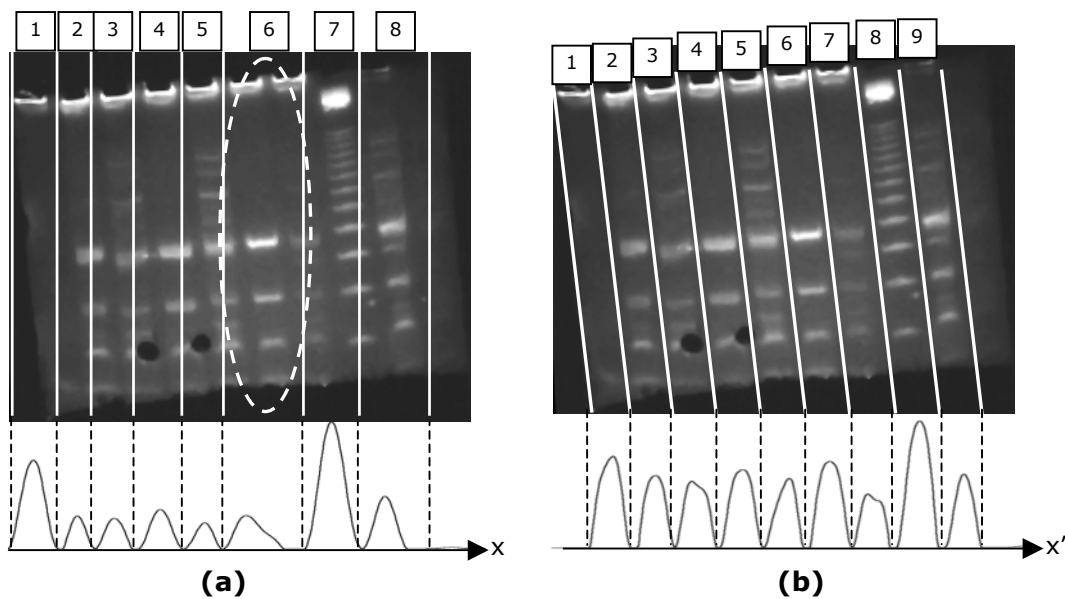
medida que la orientación cambia (lo cual no ocurre con el método 2, cuyo desempeño empeora a medida que los carriles dejan de ser completamente verticales). Si además se observan las dos últimas columnas de la tabla 4, se puede concluir que el desempeño del método 1 es mucho mejor ya que, como en el caso anterior, la cantidad de falsos aciertos permanece prácticamente constante.

Es necesario aclarar que el comportamiento descrito por la figura 35 no necesariamente es el mismo para todas las imágenes, debido precisamente a las características únicas de las mismas (como, por ejemplo, la presencia de burbujas en el gel, o el ruido introducido en la etapa de adquisición).

5.3 FUENTES DE ERROR

La robustez del método planteado radica no solo en la correcta estimación de la orientación de los carriles; se requiere igualmente que la orientación de las bandas (que es, simplemente, perpendicular a la de los carriles) sea tenida en cuenta al momento de detectar las mismas. Esto parte del hecho de que tanto carriles como bandas tendrán una orientación específica que no siempre será vertical y horizontal, respectivamente (como se asume en los trabajos realizados hasta el momento). Es erróneo asumir una orientación vertical de los carriles pero lo es igualmente obtener la proyección de cada carril (útil en la obtención de bandas) como la suma por filas de la imagen (asumiendo así que las bandas son completamente horizontales). Es así como, en principio, y además de las características propias de la imagen o el gel, se pueden atribuir las fuentes de error a estos dos aspectos: la estimación incorrecta de la orientación de los carriles y de las bandas. Es además muy probable que el hecho de asumir carriles completamente verticales dificulte la etapa de detección de los mismos, ya que los mínimos locales, que en principio corresponden a las zonas limítrofe entre uno y otro carril, no serán fácilmente detectables, hecho que se ilustra a continuación:

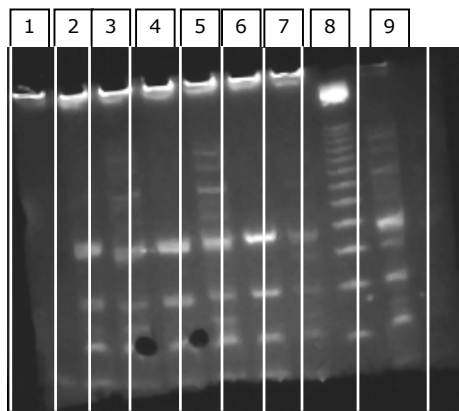
Figura 36. Una imagen cuyos carriles no son completamente verticales, y su correspondiente proyección sobre el eje x (izquierda) y sobre el eje x' (derecha)



Si bien no se puede hablar de una segmentación correcta, es importante resaltar que el problema empeora cuando se presenta una imagen como la ilustrada en la figura 36a, donde, en la zona marcada por la línea punteada, se ha llegado al caso extremo de combinar tres carriles en uno solo, debido a que la proyección sobre el eje x , calculada como la simple suma por columnas - o, de manera alterna, el máximo de cada una de ellas -, no arroja información suficiente, debido a que algunos mínimos locales desaparecen por completo (de hecho, y según la figura 36b, el número total de carriles es 9 y no 8). En contraste, el comportamiento del algoritmo implementado en este trabajo permite la correcta ubicación de todos y cada uno de los carriles presentes en la imagen, mediante el cálculo de la proyección sobre el eje x' . Se debe resaltar igualmente que los niveles de gris alcanzados por la proyección sobre dicho eje - es decir, la altura de los picos - sobrepasan aquellos alcanzados por la proyección sobre el eje x . La ventaja en este caso es que en ocasiones la altura alcanzada por los picos en la proyección sobre el eje x puede no ser suficiente, y dificultar de esta manera el proceso de detección de mínimos (ya que un máximo local puede tener un nivel de gris muy cercano a un mínimo local). Con la evaluación de la proyección sobre el eje x' se minimiza la existencia de este problema.

Tal como lo ilustra la figura 36a, la estimación incorrecta de la orientación de los carriles - es decir, el hecho de asumir que estos son completamente verticales - puede dar origen a dos situaciones - completamente opuestas la una de la otra -: la pérdida de información, que se manifiesta como la no detección de una o más bandas, o el exceso de la misma (la aparición de bandas inexistentes dentro de un carril). Para ilustrar estos casos se asumirá que, de forma manual, se han segmentado los carriles de la figura 36a, de tal manera que se ha eliminado, en parte, el problema descrito en el párrafo anterior (la combinación de tres carriles en uno solo). Desde luego, el principal problema continúa: los carriles han sido segmentados incorrectamente, pero se asumirá algún grado de tolerancia del algoritmo detector de bandas, en el sentido de que, supuestamente, puede soportar que algunos segmentos de los carriles sean eliminados a lado y lado de los mismos (es decir, se tomará solo una parte cada una de las bandas en la evaluación de la proyección). De acuerdo con esto, se tiene una imagen como la ilustrada en la figura 37:

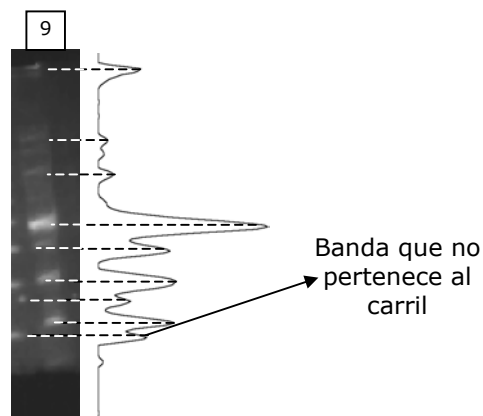
Figura 37. Nueva segmentación de carriles, completamente verticales, para corregir la ausencia de algunos



Se debe observar que el carril número 6 de la figura 36a, ha sido dividido en dos: los carriles 6 y 7 de la figura 37. Se analizarán algunas zonas de estas imágenes para ilustrar los dos efectos descritos anteriormente: la aparición de bandas inexistentes y la desaparición de otras.

5.3.1. Exceso de bandas. Se analizará en detalle el carril número 9 de la figura 37, carril que se encuentra, relativamente, segmentado de manera adecuada (puesto que ninguna banda perteneciente a dicho carril ha sido recortada en la segmentación). La figura 38 muestra el carril de interés junto con su proyección sobre el eje "y" (calculada como el máximo de cada fila de dicha imagen):

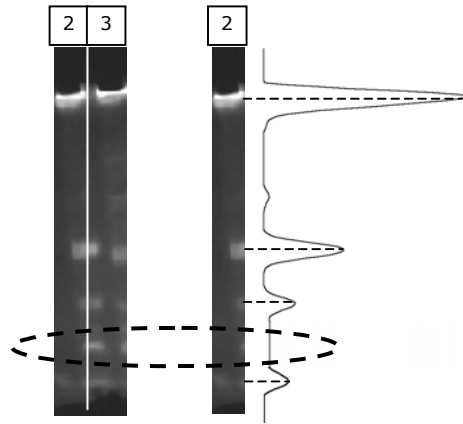
Figura 38. Ilustración del exceso de bandas debido a la incorrecta segmentación de los carriles



Se observa que la última banda, ubicada en la parte inferior, en realidad no pertenece al carril número 9, es procedente del carril número 8 y aparece debido a la incorrecta segmentación del carril. Desde luego este efecto, la aparición de bandas inexistentes, en principio es menor en caso de tener un carril con bajo contenido de bandas (como es el caso en los carriles 1 y 2: el carril 1 posee bajo contenido de bandas y por ello estas no interfieren en el proceso de detección en el carril 2). Se observa entonces que el efecto depende de la imagen y del contenido de fragmentos de ADN presente en cada carril, pero se puede asegurar que casi siempre estará presente. La aparición de "falsas bandas" está directamente relacionada, precisamente, con la aparición de falsos aciertos.

5.3.2. Eliminación de algunas bandas existentes. Corresponde a la no detección de algunas bandas, a pesar de que estas se encuentran presentes dentro del carril. Para ello se puede observar la figura 39, correspondiente al carril 2:

Figura 39. Los carriles 2 y 3 (izquierda) y el carril 2 con su respectiva proyección sobre el eje "y" (derecha)

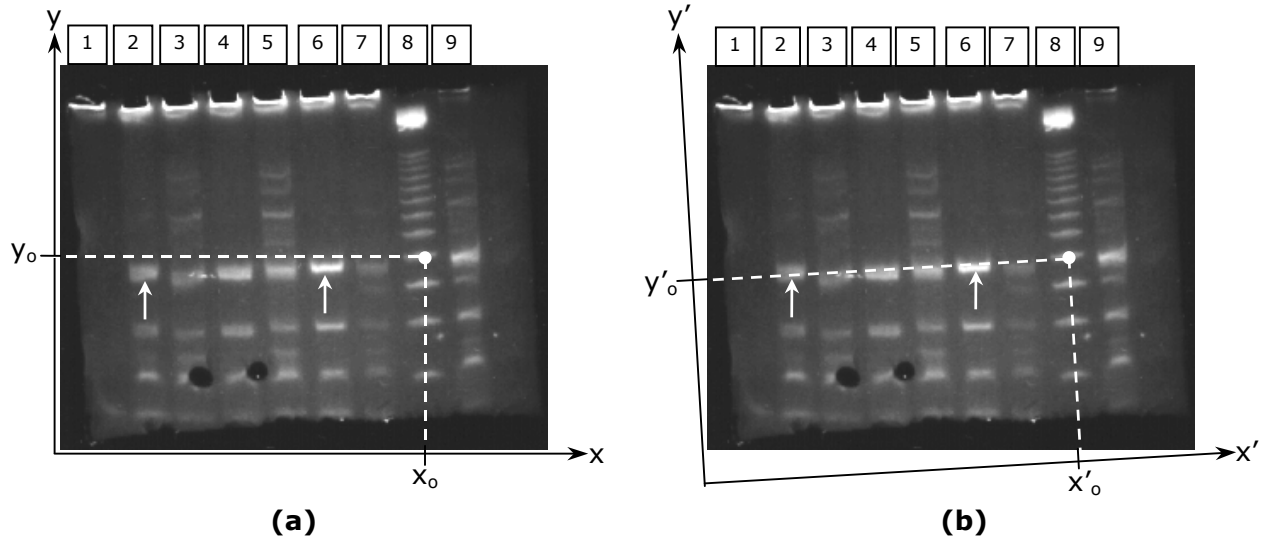


Se debe observar que, en la imagen del lado izquierdo, existe una banda en la zona encerrada por la línea punteada, banda que en realidad pertenece al carril 2. Del lado derecho se observa la proyección sobre el eje "y" convencional, donde no existe ningún pico, máximo local, correspondiente a la banda en mención. Dicho fragmento simplemente no será detectado, debido precisamente a que en la segmentación de los carriles no pudo ser incluido como parte del carril 2.

5.3.3. Errores en la estimación del tamaño de las bandas. Es probable que las características de la imagen o su misma orientación permitan obtener resultados similares al comparar uno y otro método, esto es, en ocasiones las bandas pueden ser detectadas de forma correcta asumiendo carriles completamente verticales, siempre y cuando la orientación de la imagen (definida como el ángulo que forma el eje x' con el eje x convencional) sea pequeña. Sin embargo, es muy probable que la estimación del tamaño de las bandas se vea afectada por asumir una orientación incorrecta de los carriles y las bandas.

El caso anterior se ilustrará con el siguiente ejemplo a partir de la Figura 40, donde se asume que se desea determinar el tamaño que poseen las bandas marcadas con una flecha, ubicadas en los carriles 2 y 6 respectivamente, y donde el carril número 8 es el carril estándar:

Figura 40. Estimación del tamaño de dos fragmentos de ADN asumiendo carriles y bandas completamente verticales y horizontales (a), y teniendo en cuenta la orientación de la imagen (b)



Se debe tener en cuenta que es con base en el carril standard y la coordenada "y" (o y') de las bandas presentes en el mismo, en las cuales se determina el tamaño de cualquier banda (fragmento) presente en el gel (ver sección 1.2.2). Igualmente se debe recordar que las bandas ubicadas en la parte superior de la imagen corresponden a fragmentos de mayor tamaño, en comparación con aquellas ubicadas en la parte inferior de la misma.

Para el primer caso, la figura 40a, se observa que los fragmentos de interés están ubicados por debajo de la línea $y = y_0$, lo cual indica que el tamaño de dichos fragmentos es menor que el tamaño de aquel ubicado en el punto (x_0, y_0) – que es de tamaño conocido por pertenecer al carril estándar -. Se deduce entonces que ambos son de menor tamaño que el fragmento estándar, y específicamente se podría aseverar que aquel ubicado en el carril 2 es el de menor tamaño, puesto que se encuentra más lejos de la línea $y = y_0$; por su parte, el fragmento ubicado en el carril 6 posee un tamaño cercano, pero diferente del fragmento standard.

Ahora bien, observando la figura 40b es posible asegurar que ambos fragmentos tienen en realidad el mismo tamaño del fragmento estándar, puesto que su coordenada y' es prácticamente la misma del fragmento de tamaño conocido (esto es $y' = y'_0$). Se observa entonces que el hecho de asumir carriles y bandas completamente verticales y horizontales, como en el caso de la figura 40a, equivale a introducir errores en la estimación del tamaño de los diferentes fragmentos de ADN presentes en el gel. Se debe observar que el error siempre existe, pero aumenta a medida que el fragmento se encuentra más lejos del carril standard (como es el caso en la figura 40a, donde la diferencia entre la coordenada "y" del fragmento del carril 6 y la coordenada y_0 , es menor en comparación con la misma diferencia pero evaluada para el fragmento del carril 2).

Se resalta entonces la importancia de estimar inicialmente la orientación adecuada de la imagen, hecho que redundará en la correcta segmentación de la misma, así como la adecuada detección de las bandas, evitando además incurrir en errores en la estimación del tamaño de los diferentes fragmentos de ADN presentes en el gel.

6. CONCLUSIONES Y OBSERVACIONES

En términos generales se logró el objetivo principal, como era el desarrollo de una aplicación que permitiese el análisis de geles de ADN. Si bien en principio se esperaba sacar provecho del sistema existente en la actualidad en el Laboratorio de Genética, aquel que permite el análisis de cromosomas, el mismo carácter comercial de la aplicación encargada de esta labor, impide el acceso y manejo del dispositivo de adquisición. Sin embargo, se aprovecharon otros recursos igualmente disponibles en el Laboratorio y de esta forma fue posible implementar la aplicación desarrollada.

A continuación se enumeran en detalle los principales logros de este trabajo:

- Se cumplió el principal objetivo, referente a la implementación de una aplicación tendiente a resolver las necesidades existentes en el Laboratorio de Genética, cuales eran la adquisición y análisis de una imagen proveniente de un gel de ADN. Para ello se sacó provecho del hardware existente en la actualidad en el Laboratorio (cámara digital y computador). La aplicación está equipada además con opciones adicionales para la restauración de la imagen, así como con la posibilidad de exportar los datos obtenidos tras la determinación del tamaño de los fragmentos de ADN.
- Se logró reducir significativamente el tiempo requerido para el análisis de un gel que, dependiendo de la cantidad de fragmentos de ADN que este contenga, puede requerir aproximadamente 15 o 20 minutos al ser realizado de forma manual. Con la aplicación implementada el tiempo requerido oscila entre 2 y 4 minutos.
- Se diseñó un algoritmo completamente nuevo, basado en técnicas de procesamiento digital de imágenes, que suple varias falencias encontradas en otros trabajos realizados hasta el momento en el área de análisis de geles de ADN. Es clara la ventaja que ofrece este método en la obtención de los carriles, comparable incluso con el desempeño de algunos sistemas comerciales. De igual forma, y más importante aún, posee un comportamiento mucho más robusto que aquel ofrecido por algoritmos implementados con anterioridad al presente trabajo, todo ello en términos de precisión en la detección de las bandas y la estimación del tamaño de los fragmentos. La aproximación al análisis de la imagen teniendo en cuenta la orientación de sus carriles es de gran importancia, puesto que este es un parámetro que no puede ser controlado muchas veces, siendo prácticamente imposible obtener en cada experimento una imagen con carriles completamente verticales.
- Tras analizar la posibilidad del manejo de imágenes en color, se concluye que dicha información adicional, el color, no ofrece ventajas significativas al momento de segmentar la imagen y detectar las bandas. Por el contrario, se encuentra el mismo problema existente tras analizar una imagen en niveles de gris: en muchas ocasiones es difícil diferenciar claramente el fondo de las bandas. Por su parte las imágenes en escala de gris tienen la ventaja de ser mucho más simples, en términos de la cantidad de datos e información que poseen. Es por ello que todos los trabajos enfocados al análisis de geles de ADN, obtenidos tras el proceso de electroforesis, utilizan imágenes con tonalidades grises. Sin embargo, no se descarta que en un futuro se desarrollen nuevas técnicas en la preparación de los geles, que le den al color un papel preponderante.

Si bien se tiene un funcionamiento adecuado del sistema, quedan aún aspectos por mejorar, entre ellos:

- Desarrollo en un lenguaje como C++: la principal desventaja del sistema desarrollado es que se ha implementado en un entorno que requiere una aplicación adicional para su funcionamiento: Matlab. Esto se puede tornar en un escollo a la hora de implementar la aplicación en otros Laboratorios (dentro o fuera de la Universidad). Desde luego, los algoritmos podrán ser implementados de manera independiente del equipo computacional que se tenga, sin embargo, la etapa de adquisición dependerá directamente del sistema que se tenga (tarjeta de adquisición, puerto USB, etc...) y será necesario ajustarlos a las necesidades y características de dicho sistema. Se puede considerar dicha implementación en caso de existir algún interés comercial por el sistema desarrollado.
- Optimizar el algoritmo encargado de estimar la orientación de la imagen, el cual es un proceso iterativo que en ocasiones puede tornarse lento dependiendo del tamaño de la imagen (toma entre 1 y 10 s aproximadamente, dependiendo, precisamente, del tamaño de la imagen). Esta es en realidad la etapa que más tiempo requiere durante el análisis de la imagen.

BIBLIOGRAFÍA

- [1] AGARWAL, A.; SRINIVAS MANOHAR, C. Y PUJARI, Arun K. Identification of lanes and bands in DNA autoradiogram images. Proceedings RC-IEEE-EMBS & 14th BMESI, 1995. pp.
- [2] ARIAS GONZÁLEZ, León David y RESTREPO PALACIOS, Alfredo. Desarrollo de una herramienta para la interpretación de placas de electroforesis. Memorias del VII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial. Bucaramanga, noviembre 27-29 de 2002. pp. 67-72.
- [3] BAJLA, I.; HOLLÄNDER, I.; BURG, K. Y FLUCH, S. A novel approach to quantitative analysis of electrophoretic gel images of DNA fragments.
- [4] BLACK, Michael J.; SAPIRO, Guillermo; MARIMONT, David H. Y HEEGER, David. Robust anisotropic diffusion. IEEE transactions on image processing. Vol 7, No. 3, 1998. pp. 421-432.
- [5] CHENG, Wei-Zen; YEN, Kue-Sai; LIN, Chih-Yang; CHING, Yu-Tai y YANG, Yun-Liang. Comparing lanes in the pulsed-field gel electrophoresis (PFGE) images. Proceedings of the 23rd annual EMBS International Conference. Istanbul, Turkey, 2001. pp. 2911-2913.
- [6] COLLINS, Francis S.; GREEN, Eric D.; GUTTMACHER, Alan E. y GUYER, Mark S. A vision for the future of the genomics research: a blueprint for the genomic era. Nature, Vol. 422, abril 24 de 2003. pp. 835-847.
- [7] CURTIS, Helena y BARNES, Sue. Invitación a la Biología. 5^a Ed. Madrid: Editorial Médica Panamericana, 1995. 862 p.
- [8] FUHRMANN, Daniel R.; CHIU, Readman; SCHEIN, Jacqueline E.; WATERSON, Robert; McPHERSON, Jhon D. Y MARRA, Marco A. Automated image analysis for DNA fingerprinting.
- [9] GONZÁLEZ, Rafael y WOODS, Richard. Tratamiento digital de imágenes. Wilmington, Delaware: Addison-Wesley Iberoamericana S.A, 1996. 773p.
- [10] GUY PLANTIER, Anthony Sorice y SAUMET, Jean-Louis. Autocorrelation fitting for texture orientation estimation. Proceedings of the International Conference on Image Processing, 2003.
- [11] HANBURY, Allan y SERRA, Jean. Analysis of oriented textures using mathematical morphology.
- [12] HUANG, C.T. y MITCHELL, O.R. A Euclidean distance transform using grayscale morphology decomposition. IEEE transactions on pattern analysis and machine intelligence, 1994. Vol. 16, pp. 443-448.

- [13] MACHADO, Alexei M; CAMPOS, Mario F.; SIQUEIRA, Ari M. y CARVALHO, Osvaldo S. An iterative algorithm for segmenting lanes in gel electrophoresis images. Proceedings of the XI SIBGRAPI, IEEE Press, 1997.
- [14] OLMOS, Adriana; TRUCCO, Manuel; LEBART, Katia y LANE, David. Detecting ripple patterns in mission videos.
- [15] PERONA, Pietro y MALIK, Jitendra. Scale-space and edge detection using anisotropic diffusion. IEEE transactions on pattern analysis and machine intelligence. Vol. 12, No. 7, 1990. pp. 629-639.
- [16] RUSS, John C. The image processing handbook. 4^a Ed. Boca Raton FL: CRC Press & IEEE Press, 2002. 744p.
- [17] STAROVOITOV, Valery V.; JEONG, Sang-Yong y PARK, Rae-Hong. Texture periodicity detection: features, properties and comparisons. IEEE transactions on systems, man and cybernetics; part A: systems and humans. Vol. 28, No. 6, 1998. pp. 839-849.
- [18] WATSON, James D. The Double Helix: a personal account of the discovery of the structure of DNA. New York: New American Library, 1968. 143 p.
- [19] WATSON, James D. y CRICK, Francis H. Molecular structure of nucleic acids: a structure for Deoxyribose Nucleic Acid. Nature, No. 4356. Abril 25 de 1953. pp. 737-738.
- [20] WEICKERT, J. Applications of nonlinear diffusion in image processing and computer vision. Proceedings of Algorithmy 2000, Vol. LXX, No. 1, 2001. pp. 33-50.
- [21] ----- ROMENY, Bart y VIERGEVER, Max A. Efficient and reliable schemes for nonlinear diffusion filtering. IEEE transactions on image processing, Vol. 7, No. 3, 1998. pp. 398-410.
- [22] WU, Ray; GROSSMAN, Lawrence y MOLDAVE, Kivie. Recombinant DNA methodology. San Diego, California: Academic Press, 1989. 760 p.
- [23] YE, Xiangyun; SUEN, Ching Y.; CHERIET, Mohamed y WANG, Eugenia. A recent development in image analysis of electrophoresis gels. Vision Interface '99, Trois-Rivières, Canada. 1999. pp. 432-438.
- [24] ZAMPIROLI, Francisco de Assis y LOTUFFO, Roberto de Alencar. Classification of the distance transformation algorithms under the mathematical morphology approach.