



Máster en Ingeniería de Sistemas y Servicios para la Sociedad de la Información

Trabajo Fin de Máster		
Título	Segmentación de la glotis en imágenes laríngeas usando snakes	
Autor	Gustavo Xavier Andrade Miranda	VºBº
Tutor	Nicolás Sáenz Lechón	
Ponente	Juana María Gutiérrez Arriola	
Tribunal		
Presidente	Juan Ignacio Godino Llorente	
Secretario	Fernando Pescador del Oso	
Vocal	Danilo Simón Zorita	
Fecha de lectura	13-07-2012	
Calificación		

El Secretario:

DEDICATORIA

A mis padres, que a pesar de la distancia siempre están presentes para trasmitirme todo su amor, brindarme su apoyo incondicional y recordarme que cada día debo esforzarme, trabajar más, hacer camino al andar y paso a paso avanzar en busca de mis más anhelados propósitos.

No podría agradecer a quienes solo quieren mi bienestar y superación, por eso para a ellos simplemente mi inmenso cariño y un eterno abrazo.

AGRADECIMIENTO

Mi imperecedera gratitud a todos los profesores que forman parte del grupo de investigación BYO, quienes con sus conocimientos y orientación han permitido llevar adelante este nuevo reto en mi vida profesional: espero que este trabajo de fin de máster, elaborado con esfuerzo y dedicación, sea de utilidad y permita avanzar en el conocimiento, procesamiento e interpretación de imágenes para fines de investigación y aplicación científica.



Índice

Índice	i
Índice de figuras	iv
Índice de tablas	vii
Índice de ecuaciones	viii
Resumen	x
Summary	xi
1 Preliminares	1
1.1. Planteamiento del problema.	2
1.2. Objetivos.	3
1.2.1. Objetivo General.	3
1.2.2. Objetivos específicos.	3
1.3. Organización del Documento.	4
2 Estado del Arte	5
2.1. Estado del Arte.	6
3 Marco Teórico	11
3.1. Contornos Activos.	12
3.2. Energía Externa.	14
3.3. Energía Interna.	18
3.4. Minimización de las Energías.	20
3.5. Fuerzas externas.	23



3.5.1. Fuerza Externa del tipo Balloon.	24
3.5.2. Fuerza externa del tipo GVF (Gradient Vector Flow).	25
3.6. Inicialización del Contorno Activo.	26
3.6.1. Inicialización naif.	27
3.6.2. Inicialización manual.	27
3.6.3. Inicialización automática.	27
4 Metodología Propuesta	31
4.1. Metodología.	32
4.2. Pre-procesado.	32
4.2.1. Umbralización.	33
4.2.2. Difusión Anisotrópica.	35
4.2.3. Filtro de mediana.	37
4.3. Gradient vector flow.	38
4.4. Inicialización.	39
4.5. Evolución de la snake.	42
5 Marco Experimental	45
5.1. Introducción.	46
5.2. Ajuste de parámetros.	46
5.2.1. Ajuste de umbral.	47
5.2.2. Ajuste del filtro anisotrópico.	49
5.2.3. Ajuste del filtro de mediana.	50



5.2.4. Ajuste en la inicialización.	51
5.3. Pruebas realizadas.	52
5.3.1. Sistema snake vs Sistema experto.	52
5.3.1. Sistema snake vs Sistema watershed.	54
6 Conclusiones y Futuros Trabajos	57
6.1. Conclusiones.	58
6.2. Futuros trabajos.	59
7 Referencias	61

Índice de figuras

Figura 1	Dos ejemplos de videoquimografía (a) cierre normal (b) defecto de cierre (imagen extraída de [10])	6
Figura 2	Representación esquemática de la regla de transformación de los fonovibrogramas para un solo fotograma. (a) segmentación. (b) re-muestreo de las líneas de contorno extraídas. (c) separación del eje de la glotis. (d) asignación de colores a las desviaciones de las cuerdas vocales. (imagen modificada a partir de [9])	7
Figura 3	Fonovibrograma con 21 ciclos de oscilación (imagen modificada a partir de [9])	8
Figura 4	Inicialización de la snake (el contorno cyan representa a la snake en su fase inicial)	13
Figura 5	Evolución de la snake (el contorno cyan representa la inicialización de la snake mientras que el verde muestra la ubicación de la snake al cabo de 15 iteraciones)	13
Figura 6	Resultado de la segmentación final (el contorno verde representa la evolución final de la snake)	14
Figura 7	(a) Imagen con un objeto circular (el contorno verde representa a la snake en $t=0$) (b) Grafico de superficie de (1).	15
Figura 8	Grafico de superficie de la función $E_{ext}(x,y)$	16
Figura 9	Líneas de campo en la frontera del objeto	16
Figura 10	Comportamiento de los vectores en cada píxel de la imagen	17
Figura 11	Gráfico de superficie de la evolución de la snake (la circunferencia de color blanco representa un estado intermedio del proceso de evolución de la snake)	17
Figura 12	Grafico de superficie de la segmentación final	18
Figura 13	Resultado de la segmentación excluyendo la energía interna (el contorno cyan indica la inicialización y el verde representa la segmentación errónea, imagen extraída de [30])	18
Figura 14	Resultado de la segmentación incluyendo la energía interna (el contorno cyan indica la inicialización y el verde representa la correcta segmentación final, imagen extraída de [30])	19
Figura 15	Segmentaciones fallidas usando la fuerza externa KWT (contorno rojo: snake inicial, contorno verde: snake final, imagen extraída de [30])	23

Figura 16	Fuerza externa balloon de inflación. (a) inicialización de la snake realizada dentro del objeto (b) resultado de la segmentación (contorno verde: snake final, líneas rojas representan el movimiento de los snaxels)	24
Figura 17	Fuerza externa balloon de deflación. (a) inicialización de la snake realizada fuera del objeto (b) resultado de la segmentación (contorno verde: snake final, líneas rojas representan el movimiento de los snaxels)	25
Figura 18	Líneas de campo de la fuerza GVF	26
Figura 19	Inicialización de una snake usando el método CoD (en rojo, snake inicial)	28
Figura 20	Líneas de campo que representan nueve regiones disjuntas	28
Figura 21	Inicialización usando FFS	29
Figura 22	Etapas del proceso seguido para la detección de la glotis	32
Figura 23	Imagen laríngea en escala de grises e histograma	33
Figura 24	Imagen laríngea en escala de grises e histograma	34
Figura 25	Umbralización e histograma al 85% de la Figura 23	34
Figura 26	Umbralización e Histograma al 85% de la Figura 24	35
Figura 27	Difusión anisotrópica de la Figura 25	36
Figura 28	Difusión anisotrópica de la Figura 26	36
Figura 29	Imagen a la salida de la etapa de difusión anisotrópica	37
Figura 30	Imagen a la salida del filtro de mediana	37
Figura 31	Líneas de fuerza GVF con $\mu=0.5$	38
Figura 32	Líneas de fuerza GVF con $\mu=0.2$	38
Figura 33	Líneas de fuerza GVF con 20 iteraciones	39
Figura 34	Líneas de fuerza GVF con 200 iteraciones	39
Figura 35	Imagen obtenida tras la inicialización GVF. (a) umbral 0.09 (b) umbral 0.009	40
Figura 36	Bordes obtenidos de las imágenes. (a) umbral 0.09 (b) umbral 0.009	41
Figura 37	Bordes con mayor área. (a) umbral 0.09 (b) umbral 0.009	41
Figura 38	Resultado de la inicialización. (a) umbral 0.09 (b) umbral 0.009	41
Figura 39	Evolución de la snake al cabo de 10 iteraciones	42



Figura 40	Evolución de la snake al cabo de 50 iteraciones	43
Figura 41	Resultado final de la segmentación	43
Figura 42	Ajuste del umbral al 100% del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final	47
Figura 43	Ajuste del umbral al 50% del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final	47
Figura 44	Ajuste del umbral al 85% hacia la derecha del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final	48
Figura 45	Ajuste del umbral al 85% hacia la izquierda del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final	48
Figura 46	Ajuste del filtro anisotrópico: $t=60$, $\lambda=0.25$ (a) $k=20$, (b) $k=1$, (c) $k=6$	49
Figura 47	Ajuste del filtro de mediana. (a) sin máscara (b) máscara 9x9 (c) máscara 3x3	50
Figura 48	Ajuste en la inicialización (a) inicialización de la snake con 0.009 (b) segmentación resultante	51
Figura 49	Ajuste en la inicialización (a) inicialización de la snake con 0.09 (b) segmentación resultante	51
Figura 50	Segmentación con el sistema snake (a) Pratt=0.49 (b) Pratt=0.25	52
Figura 51	Figura de mérito de Pratt sistema snake – sistema experto	53
Figura 52	Figura de mérito de Pratt con un valor de 0.96 (a) sistema snake (b) sistema experto	53
Figura 53	Segmentación con el sistema watershed (a) Pratt=0.57 (b) Pratt=0.64	54
Figura 54	Figura de mérito de Pratt sistema watershed – sistema experto	55
Figura 55	Figura de mérito de Pratt con un valor de 0.95 (a) sistema watershed (b) sistema experto	55



Índice de tablas

Tabla. I. Comparación entre el sistema watershed y el sistema snake.	56
--	----



Índice de ecuaciones

- (1) Energía Externa de la snake.
- (2) Energía Interna de la snake.
- (3) Energía Total de la snake.
- (4) Derivada funcional de la energía total con respecto $X(s)$.
- (5) Derivada funcional de la energía total con respecto $Y(s)$.
- (6) Método del descenso del gradiente.
- (7) Resultado de aplicar el método del descenso del gradiente en (4) y (5).
- (8) Representación discreta con respecto a x .
- (9) Representación discreta con respecto a y .
- (10) Fuerza externa con respecto a x .
- (11) Fuerza externa con respecto a y .
- (12) Representación vectorial de la componente en x de los snaxels en el tiempo τ .
- (13) Representación vectorial de la componente en y de los snaxels en el tiempo τ .
- (14) Representación vectorial de la Fuerza externa con respecto a x en el tiempo τ .
- (15) Representación vectorial de la Fuerza externa con respecto a y en el tiempo τ .
- (16) Representación vectorial de la ecuación (8).
- (17) Representación vectorial de la ecuación (9).
- (18) Matriz de energía interna.
- (19) Evolución de la Snake en la posición x .
- (20) Evolución de la Snake en la posición y .



- (21) Evolución de la snake usando la fuerza externa balloon en la posición x .
- (22) Evolución de la snake usando la fuerza externa balloon en la posición y .
- (23) Función $u(X, Y)$.
- (24) Función $v(X, Y)$.
- (25) Evolución de la snake usando la fuerza externa GVF en la posición x .
- (26) Evolución de la snake usando la fuerza externa GVF en la posición y .
- (27) Ecuación recursiva de la difusión anisotrópica
- (28) Representación vectorial de la Fuerza externa u con respecto a x en el tiempo
- (29) Representación vectorial de la Fuerza externa v con respecto a x en el tiempo



Resumen

El presente trabajo describe una nueva metodología para la detección automática del espacio glotal de imágenes laringeas tomadas a partir de 15 vídeos grabados por el servicio ORL del hospital Gregorio Marañón de Madrid con luz estroboscópica. El sistema desarrollado está basado en el modelo de contornos activos (snake).

El algoritmo combina en el pre-procesado, algunas técnicas tradicionales (umbralización y filtro de mediana) con técnicas más sofisticadas tales como filtrado anisotrópico. De esta forma, se obtiene una imagen apropiada para el uso de las snakes. El valor escogido para el umbral es del 85% del pico máximo del histograma de la imagen; sobre este valor la información de los píxeles no es relevante. El filtro anisotrópico permite distinguir dos niveles de intensidad, uno es el fondo y el otro es la glotis. La inicialización se basa en obtener el módulo del campo GVF; de esta manera se asegura un proceso automático para la selección del contorno inicial. El rendimiento del algoritmo se valida usando los coeficientes de Pratt y se compara contra una segmentación realizada manualmente y otro método automático basado en la transformada de watershed.



Summary

The present work describes a new methodology for the automatic detection of the glottal space from laryngeal images taken from 15 videos recorded by the ENT service of the Gregorio Marañón Hospital in Madrid with videostroboscopic equipment. The system is based on active contour models (snakes).

The algorithm combines for the pre-processing, some traditional techniques (thresholding and median filter) with more sophisticated techniques such as anisotropic filtering. In this way, we obtain an appropriate image for the use of snake. The value selected for the threshold is 85% of the maximum peak of the image histogram; over this point the information of the pixels is not relevant. The anisotropic filter permits to distinguish two intensity levels, one is the background and the other one is the glottis. The initialization is based on the obtained magnitude by GVF field; in this manner an automatic process for the initial contour selection will be assured. The performance of the algorithm is tested using the Pratt coefficient and compared against a manual segmentation and another automatic method based on the watershed transformation.

1

Preliminares

1.1. Planteamiento del problema.

En la medicina actual el uso de imágenes digitales para el diagnóstico de patologías vocales se ha convertido en una práctica cada vez más habitual. La presencia de estos trastornos se debe al aumento de tensión laríngea o hiperfunción, hablar con una frecuencia diferente del tono óptimo, excesiva ingesta de bebidas alcohólicas, uso desmesurado de la voz y, en muchos casos, por factores de agotamiento nervioso, falta de sueño, descontrol emotivo, vida agitada y estrés [1].

El desarrollo de nuevos mecanismos que permitan una eficiente y automática detección del espacio glotal se torna una tarea de relevante importancia para el oportuno descubrimiento de anomalías. Es aquí donde interviene el procesamiento digital de imágenes brindando la información necesaria para que el médico pueda realizar una correcta y rápida valoración de cada uno de sus pacientes.

La detección de la glotis en imágenes laríngeas no es una tarea sencilla; esto se debe a que los dos procedimientos fundamentales que permiten capturar el movimiento vibratorio de las cuerdas vocales (grabaciones de alta velocidad y grabaciones de baja velocidad con luz estroboscópica) [2] poseen fuentes de iluminación diferentes, lo que introduce una variabilidad inter-video, y en el caso de las grabaciones estroboscópicas produce adicionalmente una variación intra-video (debido a los cambios de iluminación entre fotogramas), sin olvidar los problemas de borrosidad y falta de continuidad en la secuencia de las imágenes de las que están exentas las imágenes tomadas con sistemas de alta velocidad. A lo mencionado se suman otros inconvenientes que afectan a ambos tipos de sistemas: sombras en los bordes de las imágenes, rotación de la cámara, movimientos laterales de la cámara o del paciente que producen una cierta deslocalización de las cuerdas vocales. De todas las dificultades presentadas en la adquisición de las imágenes, la segmentación de la glotis resulta una operación fundamental para el cálculo posterior de numerosos parámetros que caracterizan la fonación, ya sea directamente o a través de alguna representación de forma de onda glotal, perfiles de vibración, quimogramas, relación de amplitudes de vibración, relación de periodos de vibración, medidas de las fases de apertura-cierre, etc. [3]. En otras ocasiones, la segmentación de la glotis se utiliza como paso previo para la segmentación de las cuerdas vocales [4-5].

En la actualidad se han propuesto muchas técnicas para la detección del espacio glotal, desde las basadas en procesado clásico (umbralización, filtrado,

operaciones morfológicas, etc.) pasando también por las técnicas de crecimiento de región, y llegando hasta métodos más modernos como el de los contornos activos (snakes). A pesar de las innumerables propuestas desarrolladas para la detección automática de la glotis, la mayoría de estas padecen de sensibilidad a la presencia de ruido y peor aún presentan una alta dependencia del punto de inicialización del proceso de segmentación, lo que las torna soluciones poco prácticas.

En consonancia con lo anterior se plantean los objetivos de esta propuesta, para hacer frente al problema de la detección automática de la glotis, mediante la combinación de técnicas convencionales con técnicas modernas, como las snakes, y de esta manera solucionar las dificultades que aquejan a las metodologías hasta hoy propuestas, brindando una mayor robustez y garantizando una mayor tasa de aciertos.

1.2. Objetivos.

1.2.1. Objetivo General.

Mejorar las metodologías existentes para la identificación automática de la glotis, lo que redundará en algoritmos más precisos que facilitarán el diagnóstico médico.

1.2.2. Objetivos específicos.

- Realizar un estudio exhaustivo de los modelos matemáticos que rigen el comportamiento de las snakes.
- Analizar qué fuerza externa se ajusta mejor a la segmentación laríngea.
- Generar una metodología para el pre-procesado de las imágenes laríngeas que permita adaptar el algoritmo de snakes a la segmentación automática de la glotis.
- Proponer una metodología para la inicialización de las snakes basada en el enfoque del problema inverso.
- Validar los resultados obtenidos.

1.3. Organización del Documento.

El presente documento está compuesto por 6 capítulos: Preliminares, Estado del Arte, Marco teórico, Metodología propuesta, Marco experimental, Conclusiones y futuros trabajos.

En el capítulo 2 se presentan trabajos relativos a la detección de la glotis enfocados a la segmentación basada en regiones, y a la segmentación basada en modelos.

En el capítulo 3 se presentan conceptos relacionados con las snakes como: fuerzas internas, fuerzas externas, minimización y discretización de las energías e inicialización de las snakes con el enfoque del problema inverso. Se incluye también en este apartado, el “*balloon model*” y el “*Gradient vector Flow*”, que son variantes de la fuerza externa propuesta por M.Kass, A. Witkin y D.Terzopoulos en 1988 [6].

En el capítulo 4 se detalla la metodología de trabajo utilizada para segmentar la glotis, en la cual se incluye la explicación detallada de los procedimientos para el pre-procesado (umbralización, filtro anisotrópico y filtro de mediana), implementación del método de las snakes e inicialización del contorno activo.

El capítulo 5 presenta los conjuntos de pruebas realizados y coteja los resultados obtenidos con otro método utilizado para la segmentación de imágenes laríngeas basado en la transformada de watershed [7].

Los resultados de los experimentos son discutidos en el capítulo 6, presentando a su vez, las conclusiones de la tesis y el trabajo futuro.

2 Estado del Arte

2.1. Estado del Arte.

La forma de vida actual ha traído consigo el aumento de las patologías laríngeas, haciendo cada vez más imperioso un buen diagnóstico de las mismas. Uno de los métodos mas usados es la observación directa de la vibración de las cuerdas vocales durante la fonación. En este sentido, se han desarrollado numerosas técnicas de visualización y aplicaciones orientadas a facilitar y hacer más preciso el trabajo del médico. Algunas de las más extendidas actualmente son la videoquimografía [8] y los fonovibrogramas [9], que permiten visualizar y caracterizar con precisión el comportamiento dinámico de las cuerdas vocales.

La videoquimografía se obtiene a partir de vídeos de alta velocidad o de grabaciones estroboscópicas, y es simplemente la quimografía aplicada a la laringología. Se escoge un segmento de los pliegues vocales y se analiza su comportamiento vibratorio, obteniendo una única imagen que es una representación temporal del proceso de apertura y cierre de las cuerdas vocales (Figura 1). La videoquimografía facilita la detección de anomalías en el patrón de vibración y en la onda mucosa.

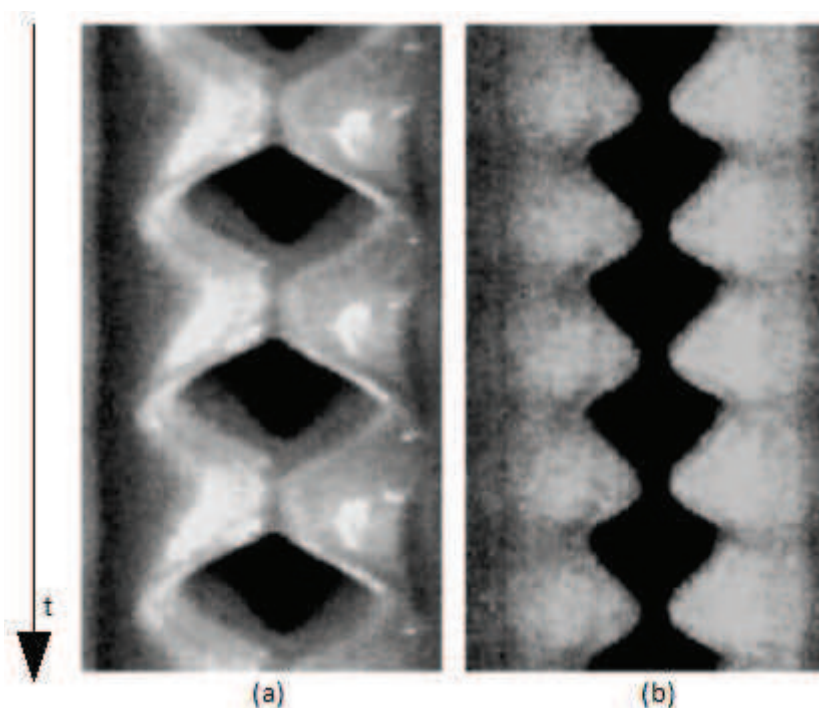


Figura 1. Dos ejemplos de videoquimografía (a) cierre normal (b) defecto de cierre (imagen extraída de [10]).

Los fonovibrogramas tienen como objetivo analizar los patrones de vibración presentes en las cuerdas vocales. Dichas vibraciones son obtenidas a partir de vídeos de alta velocidad, y a partir de estos se localiza los bordes de las cuerdas vocales por medio de algoritmos de segmentación. Posteriormente se genera una matriz de colores de dos dimensiones, que permite distinguir visualmente las oscilaciones realizadas por las cuerdas vocales a lo largo del tiempo (Figura 2 y Figura 3). Los fonovibrogramas son capaces de diferenciar y cuantificar diferentes tipos de vibraciones normales y patológicas de las cuerdas vocales [9].

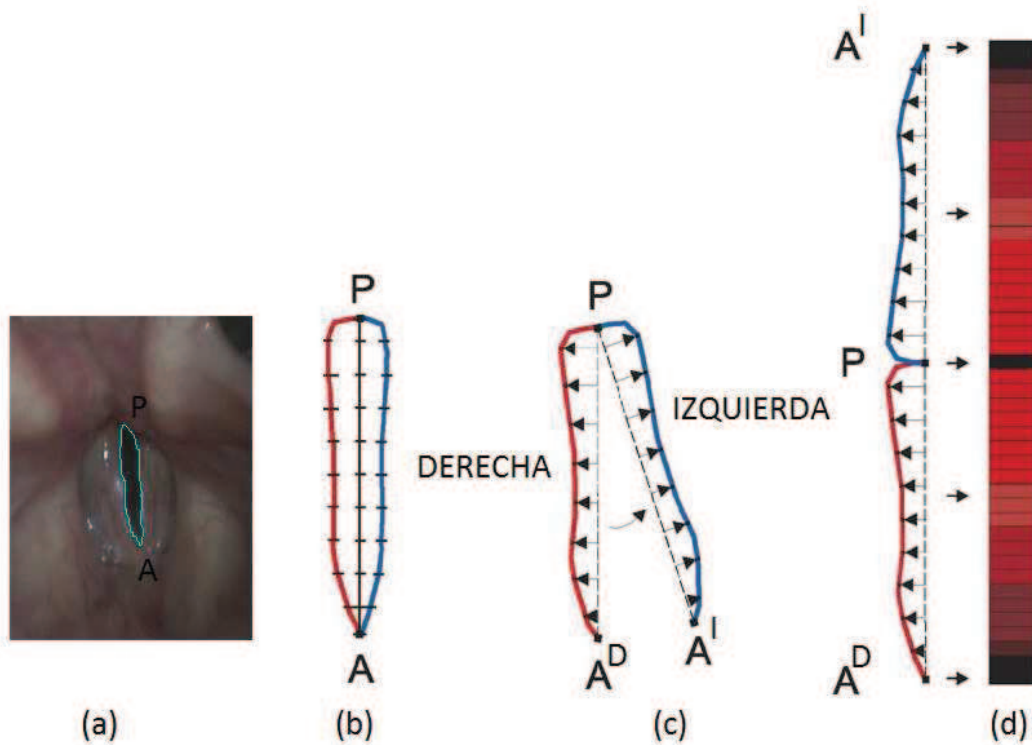


Figura 2. Representación esquemática de la regla de transformación de los fonovibrogramas para un solo fotograma. (a) segmentación. (b) re-muestreo de las líneas de contorno extraídas. (c) separación del eje de la glotis. (d) asignación de colores a las desviaciones de las cuerdas vocales. (imagen modificada a partir de [9]).

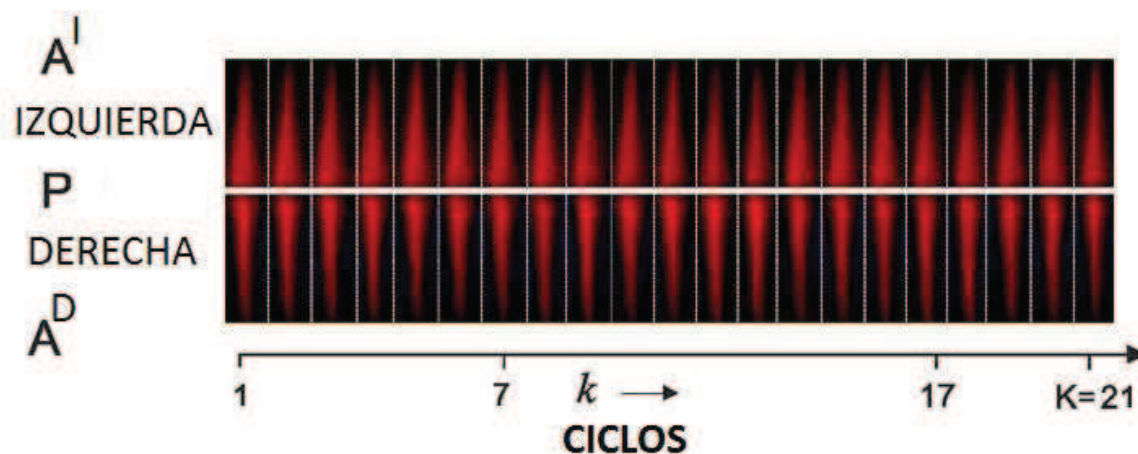


Figura 3. Fonovibrograma con 21 ciclos de oscilación (imagen modificada a partir de [9]).

El principal problema que existe a la hora de utilizar estas técnicas está en los movimientos involuntarios que sufren cámara y paciente durante la exploración, produciendo un falseo de los resultados. La solución viene de la mano del procesado digital de imagen que, aplicado a los vídeos capturados, hace posible atenuar en mayor o menor medida los efectos perniciosos del movimiento. Dentro de este procesado juega un papel fundamental una buena segmentación del espacio glotal.

Actualmente hay muchos trabajos referentes al problema de la detección automática del espacio glotal como paso previo para el análisis de procesos de fonación. Éstos manejan principalmente dos diferentes enfoques para la segmentación: la basada en regiones y la basada en modelos.

En el primer enfoque encontramos los métodos basados en umbralización e histogramas, y los de crecimiento de regiones (*region growing*). Los picos y los valles obtenidos de los histogramas de una imagen representan la distribución de los píxeles a lo largo del rango de colores o de intensidad, y suele usárselos para separar rangos de valores dentro de una imagen. En este sentido, se asume que los histogramas son bimodales (tienen dos picos), lo cual no es el caso en imágenes con bajo contraste y en objetos con perfiles heterogéneos. Metha [11] se vale de este método para la detección del espacio glotal; sin embargo, el procedimiento realizado requiere inspección visual y ajustar el umbral manualmente.

Los métodos de crecimiento de región examinan los píxeles vecinos de un grupo inicial de puntos de semillas y determinan si los vecinos del píxel deberían añadirse a la región. El número final de regiones ha de ser como mucho igual al

número de semillas sembradas. Uno de los inconvenientes que aqueja este método es que requiere de un criterio sólido y bordes relativamente delimitados con el fin de converger hacia la región de interés; el otro es su crítica dependencia de la inicialización, lo que en muchos casos dificulta un procesado completamente automático. Lohscheller [12] por ejemplo permite que el usuario defina a partir de una imagen previamente seleccionada los puntos que serán designados como semillas. Una matriz de dos dimensiones compuesta de umbrales es usada como criterio para detener el crecimiento de las regiones. Wittenberg [13] sugiere usar los píxeles más oscuros de las imágenes como semillas para el algoritmo de crecimiento de regiones, lo que resulta inapropiado en imágenes con sombras y con poco contraste. Yan [14] propone que la inicialización de la semilla en la región de crecimiento se realice mediante técnicas avanzadas de umbralización basadas en los histogramas de las imágenes; el inconveniente es que asevera que las imágenes de alta velocidad tienen una distribución Rayleigh de intensidad, lo cual no se cumple en todas las imágenes, especialmente en los videos estroboscópicos. Demeyer [15] define la semilla como la respuesta máxima del laplaciano de un filtro gaussiano y la intensidad es usada como el único criterio de homogeneidad; este método se aplica a los fotogramas que corresponden a la glotis en su máxima apertura. La región de crecimiento resultante se propaga al resto de la secuencia usando un método *level-set* y los parámetros son elegidos empíricamente.

Dentro de los métodos basados en modelos se encuentran los contornos activos, también conocidos como *snakes*. Reciben este nombre por el movimiento que realizan para delimitar regiones de interés dentro de un objeto. Las snakes son finas bandas elásticas que se acoplan de forma adecuada a contornos no rígidos y amorfos. Para lograrlo se necesita situar el contorno cerca del objeto (inicialización) para que posteriormente sea guiado por las fuerzas externas de la imagen hacia este, y una vez ahí, cualquier evolución adicional no producirá cambios. Sin embargo son muy dependientes de la inicialización de su contorno. Además es bien sabido que los detectores de bordes utilizados para obtener el gradiente, que es la característica en la cual se fundamentan las snakes, son extremadamente sensibles al ruido y en muchos casos no logran una buena localización de los bordes del objeto deseado. Marendic [16] fue el primero en presentar el algoritmo de contornos activos para la extracción de las cuerdas vocales a partir de videos de alta velocidad, adaptando al detector de bordes de Canny [17] un filtro lineal que reduce el ruido y es usado para calcular la energía externa. Cada fotograma del vídeo es inicializado con la snake resultante del cuadro previo, y los parámetros elegidos son seleccionados empíricamente. Palm [18]

desarrolló una variante del método de las snakes para mejorar el comportamiento frente al ruido y para obtener cierto grado de independencia en la inicialización. Sin embargo estima que la glotis siempre va a ser el objeto más oscuro de la imagen y se va a situar en el centro de esta, algo que no ocurre en la mayoría de imágenes estroboscópicas debido a los problemas de iluminación y movimientos. Una de las aplicaciones más recientes basadas en contornos activos es la presentada por Moukalled [19], en la que se emplea un par de curvas abiertas a la derecha y a la izquierda de las cuerdas vocales para segmentar la glotis a partir de una secuencia digital quimográfica. Este método requiere que el usuario ajuste unos parámetros dos veces por vídeo, la primera vez para inicializar las snakes en la izquierda y derecha de las cuerdas vocales y la segunda para verificar la segmentación resultante de un fotograma DKG (*Digital Kymographic*) antes de propagarlo a las etapas de segmentación automatizada [20].

Adicionalmente a los dos enfoques mayormente usados para la segmentación de la glotis, existen otros métodos como [7], que es el resultado de la combinación de algunas técnicas relevantes en el campo del procesamiento digital de imágenes entre las que se incluyen la transformada de *watershed* [21] seguida del uso de la técnica de fusión de regiones (*region merging*) [22] para finalizar con un predictor lineal [23] simple que es el encargado de la toma de decisión. Esta técnica se basa en las características perceptuales del ojo humano para discriminar entre niveles de grises, por tanto desde el punto de vista de un observador humano la glotis siempre debería ser un objeto oscuro rodeado de zonas más claras. Otra ventaja que nos brinda este método es que no necesita inicialización, que es el punto débil que aqueja a la mayoría de las técnicas empleadas para la segmentación de la glotis.

En el presente trabajo de tesis, se pretende desarrollar un método que permita automatizar y optimizar la costosa y compleja tarea de aislar el espacio glotal en imágenes de la laringe. El método propuesto para realizar la detección está basado en varias herramientas avanzadas de procesado de imagen y principalmente en el algoritmo de contornos activos – snakes.

3

Marco Teórico

3.1. Contornos Activos.

El modelo de contornos activos o comúnmente conocido como *snakes* fue introducido en el año de 1988 por M.Kass, A. Witkin y D.Terzopoulos [6]. Desde el momento de su publicación esta técnica ha causado un tremendo impacto en el análisis de imágenes biomédicas, y se ha convertido a su vez en uno de los modelos más populares en lo que a segmentación se refiere. Entre las aplicaciones desarrolladas para la segmentación de imágenes médicas usando contornos activos podemos citar detección de tumores cerebrales [24], segmentación de los huesos en radiografías de la mano [25] o detección de nódulos pulmonares en imágenes obtenidas mediante tomografía computarizada [26]. A las aplicaciones anteriormente mencionadas hay que sumarles las utilizadas para realizar seguimientos, como por ejemplo: [27] para detectar el movimiento de los leucocitos a partir de vídeos en tiempo real, [28] para capturar el movimiento de las estructuras cardiacas a partir de vídeos de ultrasonido.

A continuación se ofrece una introducción a los conceptos y los modelos matemáticos involucrados en las *snakes*, que se basan en su mayoría en [29].

Un contorno activo es una banda elástica fina que se sitúa en una imagen con anticipación para de esta manera poder delimitar un objeto de interés. El nombre de “snake” proviene del comportamiento dinámico que presenta el contorno, ya que este se arrastra sobre la imagen mientras busca las características relevantes y minimiza la energía. Las *snakes* parten del concepto de ecuaciones paramétricas, las cuales se generan cuando las variables x e y se expresan en función de una tercera llamada *parámetro*. Teniendo esto en cuenta, consideremos una curva continua en dos dimensiones donde tanto las coordenadas en x como en y son funciones de un parámetro escalar s ; por tal razón se puede decir que $(X(s), Y(s))$ representa una curva continua en dos dimensiones donde $X(s)$ e $Y(s)$ son respectivamente los valores de las coordenadas x e y en el plano. El parámetro s puede tomar valores en el rango $[0, 1]$, por tanto $(X(0), Y(0))$ representa un extremo de la *snake* y $(X(1), Y(1))$ representa el otro. Por tal motivo para que la curva sea cerrada necesitamos que $(X(0), Y(0))$ y $(X(1), Y(1))$ representen el mismo punto, lo que específicamente significaría $X(0) = X(1)$ e $Y(0) = Y(1)$. El objetivo primordial del modelo de *snakes* es delimitar la frontera de un objeto. Para lograr este cometido el primer paso es colocar la *snake* cerca del objeto: a este paso se lo conoce con el nombre de *inicialización de la snake* (Figura 4).

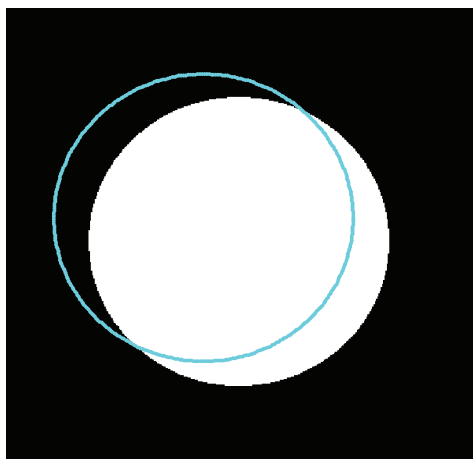


Figura 4. Inicialización de la snake (el contorno cyan representa a la snake en su fase inicial).

Posteriormente, mediante un procedimiento computacional llamado *evolución de la snake*, se realiza un proceso iterativo en el cual la snake se desliza buscando las características sobresalientes de la imagen. Una buena evolución empuja la snake inicial hacia los bordes del objeto deseado y una vez alcanzado el mismo, ninguna evolución adicional producirá cambio alguno. Dicho esto se puede establecer que la evolución de la snake cesa cuando esta converge. La Figura 5 muestra un estado intermedio en la evolución de la snake aplicada a un objeto circular, y la Figura 6 muestra la evolución final de la snake en la que se observa cómo el contorno alcanzó el borde del objeto.

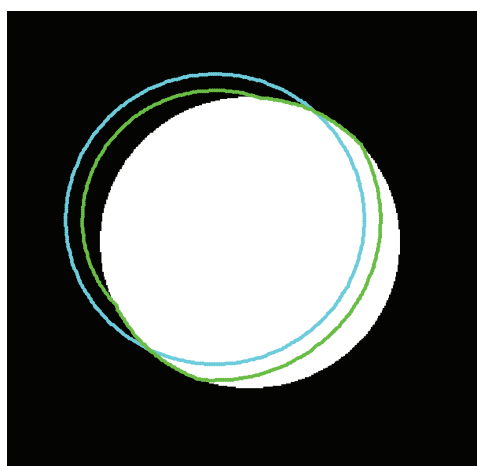


Figura 5. Evolución de la Snake (el contorno cyan representa la inicialización de la snake mientras que el verde muestra la ubicación de la snake al cabo de 15 iteraciones).

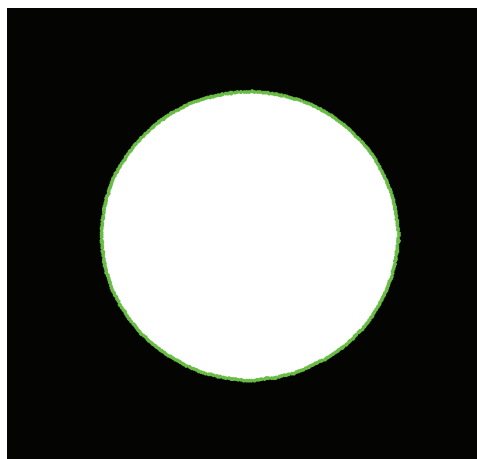


Figura 6. Resultado de la segmentación final (el contorno verde representa la evolución final de la snake).

A pesar de las ventajas que presentan, las snakes no resuelven por sí solas los problemas de la segmentación. Las snakes necesitan que algún otro algoritmo o procedimiento las sitúe cerca del contorno deseado. Sin embargo, en muchas ocasiones, a pesar de que no se cuente con un mecanismo satisfactorio de inicialización automático, se usan para procedimientos de interpretación de imágenes semiautomáticos.

Las snakes son un ejemplo de técnicas en las cuales se hace coincidir un modelo deformable con una imagen por medio de la minimización de energía. El modelo de snake no es más que una curva continua controlada e influenciada por la energía externa (propia de la imagen) y la energía interna (propia de la snake).

3.2. Energía Externa.

Para lograr que las snakes se deslicen a través de la superficie de una imagen, es necesaria una energía funcional que pueda guiarlas hacia las características sobresalientes que se pretenden encontrar. Supongamos la imagen de la Figura 7 con su respectivo diagrama de superficie. En este ejemplo el objeto a segmentar es el círculo blanco y la circunferencia de color verde representa a la snake en $t=0$.

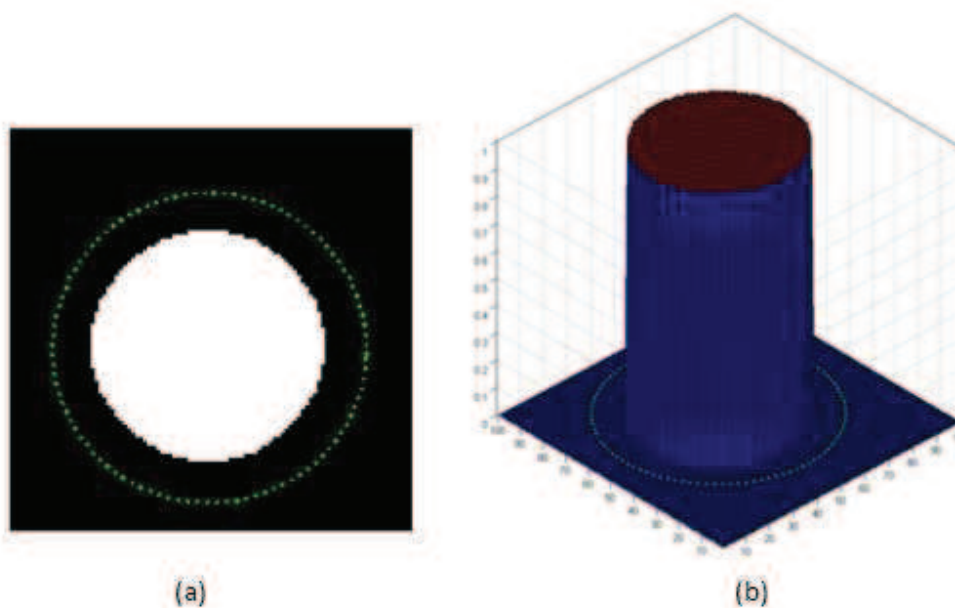


Figura 7. (a) Imagen con un objeto circular (el contorno verde representa a la snake en $t=0$) (b) Grafico de superficie de (1).

Observando la Figura 7 podemos notar que la única característica relevante para obtener el círculo blanco es la variación brusca de intensidad en el contorno del círculo. El gradiente de una imagen representa justamente la razón de cambio de la intensidad con respecto a las coordenadas de la imagen, por tal razón cuanto mayor sea el cambio de los niveles de intensidad, significará que nos encontramos en presencia de un contorno o en la frontera de una región. Por tanto la energía externa dependerá de la ecuación:

$$E_{ext}(X, Y) = - \int_0^1 f(X(s), Y(s)) ds \quad (1)$$

$$f(x, y) = |\nabla I(x, y)|^2 = \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2$$

Donde $I(x, y)$ es la función de intensidad de la imagen y ∇ representa el gradiente. Si observamos cuidadosamente la Figura 8, que representa el gradiente de la imagen anterior, podemos notar que en la frontera del objeto se forma un valle, ya que se concentra la mayor tasa de cambio de intensidad y el gradiente en esos puntos se rige por las líneas de campo mostradas en la Figura 9

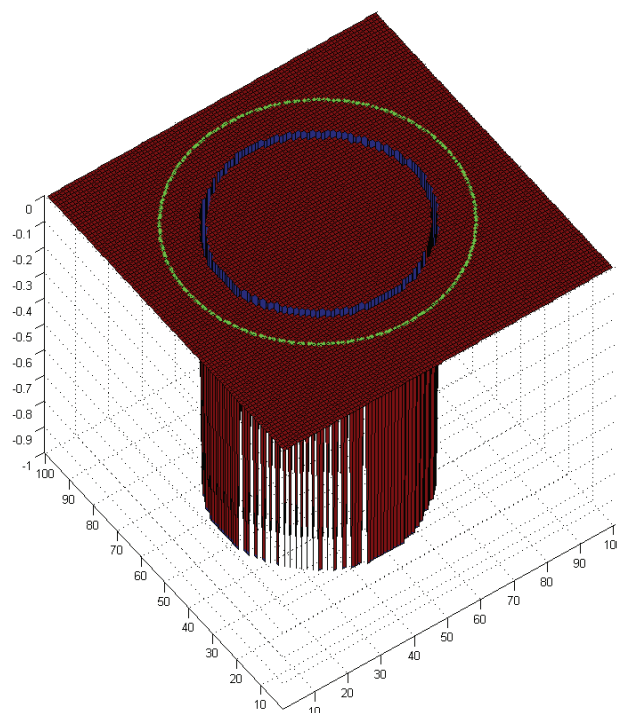


Figura 8. Grafico de superficie de la función $E_{ext}(x, y)$.

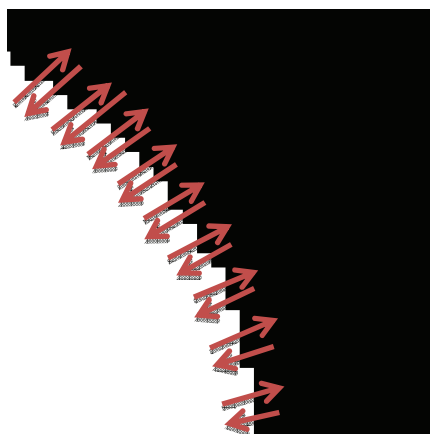


Figura 9. Líneas de campo en la frontera del objeto.

La dirección y la magnitud de los vectores de la Figura 9 corresponden a los de la fuerza externa (este concepto se detallará más adelante). Cada píxel de la imagen tendrá asociado un vector de fuerza externa, la dirección del vector servirá para orientar a la snake, el módulo para saber cuán cerca se encuentra de una variación de intensidad y el sentido para indicar hacia donde debe moverse. La Figura 10 detalla el comportamiento de cada píxel en la frontera del objeto y en sus vecindades. El número 0 en los píxeles representa un vector nulo.

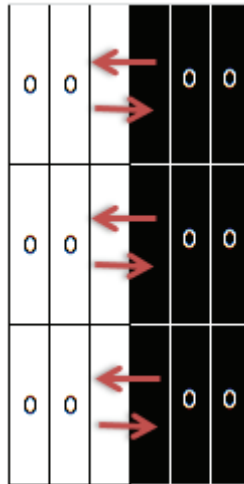


Figura 10. Comportamiento de los vectores en cada píxel de la imagen.

Hasta este punto sólo hemos localizado la región de interés, el siguiente paso es lograr que la snake se mueva hasta situarse en el valle. Para esto es necesario minimizar las energías y evaluarlas en los puntos $(X(s), Y(s))$. La Figura 11 muestra un estado intermedio en la evolución de la snake, mientras que en la Figura 12 se observa cómo logra situarse en el valle formado por el borde del objeto.

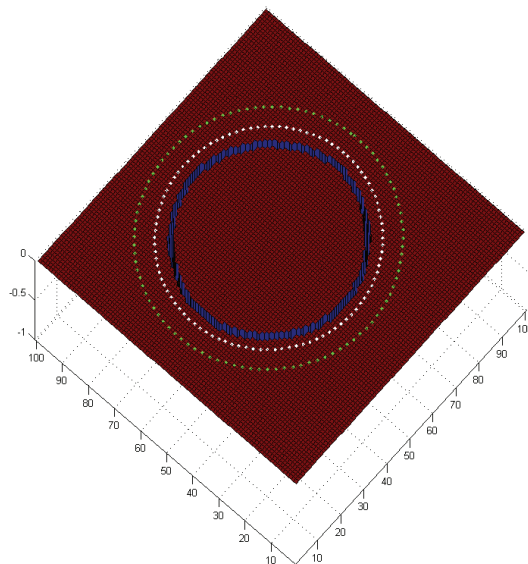


Figura 11. Gráfico de superficie de la evolución de la snake (la circunferencia de color blanco representa un estado intermedio del proceso de evolución de la snake).

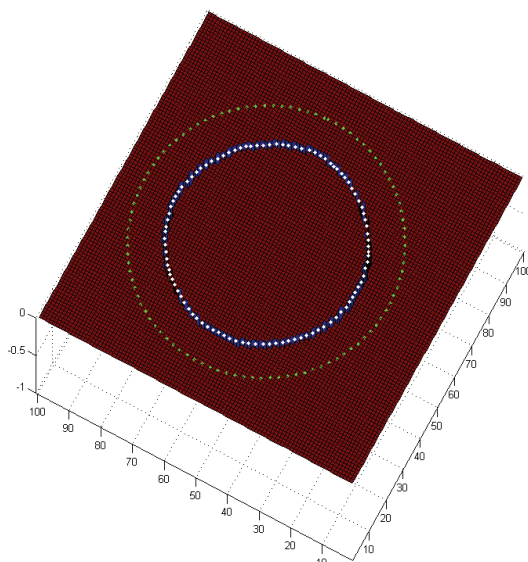


Figura 12. Grafico de superficie de la segmentación final.

3.3. Energía Interna.

La falta de suavidad de las imágenes a causa del ruido generado en la mayoría de equipos utilizados en el momento de la adquisición, provoca que la evolución normal de la snake fracase. Esto se debe a que el módulo del gradiente de la imagen encargada de orientar a la snake hacia el contorno genera valores erróneos, produciendo una incorrecta segmentación. La Figura 13 muestra un proceso fallido de segmentación.



Figura 13. Resultado de la segmentación excluyendo la energía interna (el contorno cyan indica la inicialización y el verde representa la segmentación errónea, imagen extraída de [29]).

El modelo Kass, Witkin y Terzopoulos (KWT) consideró las dificultades presentadas e introdujo un término conocido como *energía interna*. Este término provee a la snake de tolerancia al ruido, para que en el momento de la segmentación pueda situarse sobre el objeto deseado sin conflicto alguno. La energía interna se expresa como:

$$E_{int}(X, Y) = \frac{1}{2} \int_0^1 \alpha \left[\left| \frac{dX}{ds} \right|^2 + \left| \frac{dY}{ds} \right|^2 \right] + \beta \left[\left| \frac{d^2X}{ds^2} \right|^2 + \left| \frac{d^2Y}{ds^2} \right|^2 \right] ds \quad (2)$$

La ecuación descrita esta conformada por un término de primer orden controlado por el parámetro α y uno de segundo orden controlado por el parámetro β . El término de primer orden hace que la snake actúe como una membrana mientras que el término de segundo orden hace que se comporte como una lámina delgada. Ajustando los valores de α y β correctamente podemos lograr que la snake tenga desplazamientos suaves, flexibles pero no irregulares.

Los términos $\left(\frac{dX}{ds}, \frac{dY}{ds} \right)$ y $\left(\frac{d^2X}{ds^2}, \frac{d^2Y}{ds^2} \right)$ indican el cambio de localización de las coordenadas de la snake como resultado de la variación del parámetro s . Si consideramos a la variable s como unidad de tiempo, entonces los términos mencionados anteriormente corresponderían a la velocidad y a la aceleración de la snake respectivamente. Por tal motivo se debe controlar el valor de estas dos magnitudes para garantizar un movimiento apacible. La Figura 14 muestra una correcta segmentación debido a la adición de la energía interna.



Figura 14. Resultado de la segmentación incluyendo la energía interna (el contorno cyan indica la inicialización y el verde representa la correcta segmentación final, imagen extraída de [29]).

3.4. Minimización de las Energías.

El modelo básico de la snake puede ser representado mediante la suma de la energía interna y externa como se muestra a continuación:

$$E_{total}(X, Y) = E_{int}(X, Y) + E_{ext}(X, Y) \quad (3)$$

Para lograr la evolución de la snake es necesario minimizar la energía total. Sin embargo si observamos detenidamente (3) podemos notar que es una *ecuación funcional*, es decir una función de una función. Para resolver una funcional es necesario el uso de herramientas matemáticas avanzadas, conocidas como *cálculo de variaciones* o *cálculo variacional*. En [30] se explica el procedimiento necesario para obtener la derivada de una funcional con respecto a una función. El resultado obtenido se iguala a cero y se procede a encontrar los mínimos. Las derivadas funcionales de la energía total igualada a cero son:

$$\frac{\partial E}{\partial x} = -\alpha \frac{d^2 X}{ds^2} + \beta \frac{d^4 X}{ds^2} - \frac{\partial f}{\partial x} = 0 \quad (4)$$

$$\frac{\partial E}{\partial y} = -\alpha \frac{d^2 Y}{ds^2} + \beta \frac{d^4 Y}{ds^2} - \frac{\partial f}{\partial y} = 0 \quad (5)$$

Las ecuaciones parciales obtenidas en (4) y (5) son conocidas como *ecuaciones de Euler*, donde no se puede obtener una solución única, por consiguiente debemos de valernos de técnicas numéricas como el método del descenso del gradiente para encontrar los mínimos. Para comprender mejor el método del descenso del gradiente consideremos una función arbitraria $h(x)$, tal que x^* es el valor de x en el cual la ecuación $h(x)$ es mínima. Para minimizar $h(x)$ derivamos en función de x e igualamos a cero, $\frac{dh(x)}{dx} = 0$. El principio del método del gradiente es establecer una posible solución inicial $x = x_0$. Luego el valor de la variable x debe ir variando hasta llegar a x^* acorde a que el cambio en x tiene que ser proporcional al negativo de la derivada de $h(x)$. Por tanto la expresión matemática que rige la regla del descenso del gradiente es:

$$\frac{dx}{d\tau} = - \frac{dh(x)}{dx} \quad (6)$$

donde τ denota una variable de pseudo-tiempo. La ecuación (6) guarda una importante relación con el comportamiento que rige la snake cuando el proceso de minimización es alcanzado, ya que cuando la derivada es cero no es necesario ningún cambio adicional en la variable x . Otra alternativa que se puede usar para resolver las ecuaciones (4) y (5) es el método de Newton; el inconveniente de este es su alto costo computacional, lo que lo torna un método poco práctico.

Aplicando el método del descenso del gradiente a las ecuaciones (4) y (5) obtenemos:

$$\frac{\partial X}{\partial \tau} = -\frac{\partial E_{ext}(x)}{\partial x} = \alpha \frac{d^2 X}{ds^2} - \beta \frac{d^4 X}{ds^2} + \frac{\partial f}{\partial x} \quad (7)$$

$$\frac{\partial Y}{\partial \tau} = -\frac{\partial E_{ext}(y)}{\partial y} = \alpha \frac{d^2 Y}{ds^2} - \beta \frac{d^4 Y}{ds^2} + \frac{\partial f}{\partial y}$$

Si tenemos en cuenta que $Fuerza = \nabla Energía$ [31], podemos decir que (7) corresponde a un equilibrio de fuerzas. La fuerza interna (los dos primeros términos) y la externa (el tercero) tratan de balancearse entre ellas. Cuando la fuerza neta entre ambas es cero, la snake detiene su evolución y obtenemos el mínimo local. Ahora necesitamos una representación poligonal del contorno continuo. Esto se logra representando el parámetro continuo $s \in [0,1]$ con los índices $i \in \{0,1, \dots, n-1\}$ donde n representa el número total de *snaxels* (puntos de la snake). Por consiguiente se puede decir que (X_i, Y_i) es la contraparte discreta de $(X(s), Y(s))$. La forma discreta de las ecuaciones (7) es:

$$\frac{X_i^{\tau+1} - X_i^{\tau}}{\zeta} = \alpha(X_{i+1}^{\tau} - 2X_i^{\tau} + X_{i-1}^{\tau}) - \beta(X_{i+2}^{\tau} - 4X_{i+1}^{\tau} + 6X_i^{\tau} - 4X_{i-1}^{\tau} + X_{i-2}^{\tau}) + f_x(X_i^{\tau}, Y_i^{\tau}) \quad (8)$$

$$\frac{Y_i^{\tau+1} - Y_i^{\tau}}{\zeta} = \alpha(Y_{i+1}^{\tau} - 2Y_i^{\tau} + Y_{i-1}^{\tau}) - \beta(Y_{i+2}^{\tau} - 4Y_{i+1}^{\tau} + 6Y_i^{\tau} - 4Y_{i-1}^{\tau} + Y_{i-2}^{\tau}) + f_y(X_i^{\tau}, Y_i^{\tau}) \quad (9)$$

donde τ y $\tau + 1$ representan dos instantes de tiempo sucesivos con una diferencia entre ambos de ζ . Las fuerzas externas se expresan con la notación f_x y f_y donde:

$$f_x(x, y) = \frac{\partial}{\partial x} f(x, y) \quad (10)$$

$$f_y(x, y) = \frac{\partial}{\partial y} f(x, y) \quad (11)$$

La expresión (f_x, f_y) denota los vectores de campo que actúan como fuerza externa para la snake. Es conveniente expresar las ecuaciones (10) y (11) en notación matriz-vector. Para esto debemos emplear la siguiente notación:

$$x^\tau \equiv [X_0^\tau, \dots, X_{n-1}^\tau]^T \quad (12)$$

$$y^\tau \equiv [Y_0^\tau, \dots, Y_{n-1}^\tau]^T \quad (13)$$

$$f_x^\tau \equiv [f_x(X_0^\tau, Y_0^\tau), \dots, f_x(X_{n-1}^\tau, Y_{n-1}^\tau)]^T \quad (14)$$

$$f_y^\tau \equiv [f_y(X_0^\tau, Y_0^\tau), \dots, f_y(X_{n-1}^\tau, Y_{n-1}^\tau)]^T \quad (15)$$

Usando esta notación podemos reescribir las ecuaciones (8) y (9) como:

$$\frac{x^{\tau+1} - x^\tau}{\zeta} = -Ax^\tau + f_x^\tau \quad (16)$$

$$\frac{y^{\tau+1} - y^\tau}{\zeta} = -Ay^\tau + f_y^\tau \quad (17)$$

siendo

$$A = \begin{bmatrix} c & b & a & & & a & b \\ b & c & b & a & & & a \\ a & b & c & b & a & & \\ & & \ddots & \ddots & \ddots & \ddots & \\ & & & a & b & c & b & a \\ a & & & a & b & c & b \\ b & a & & & a & b & c \end{bmatrix} \quad (18)$$

La matriz A es de tamaño $n \times n$ con una estructura pentagonal donde a , b y c son iguales a: $a = \beta$, $b = -(4\beta + \alpha)$, $c = 6\beta + 2\alpha$. De las ecuaciones (16) y (17) se

puede deducir que los valores de $x^{\tau+1}$ e $y^{\tau+1}$ que corresponden a la posición de la snake en el tiempo $\tau + 1$ son:

$$x^{\tau+1} = (I_n + (\gamma)A)^{-1}(x^{\tau} + (\gamma)f_x^{\tau}) \quad (19)$$

$$y^{\tau+1} = (I_n + (\gamma)A)^{-1}(y^{\tau} + (\gamma)f_y^{\tau}) \quad (20)$$

Las ecuaciones (19) y (20) representan la evolución de la snake, y deben ser resueltas de manera iterativa hasta minimizar la energía total. El término I_n corresponde a una matriz identidad de tamaño $n \times n$.

3.5. Fuerzas externas.

Las fuerzas externas suelen clasificarse en fuerzas dinámicas y fuerzas estáticas. Las fuerzas dinámicas son aquellas que dependen del resultado de la deformación de la snake, y las estáticas se calculan únicamente a partir de las imágenes. En términos físicos, las ecuaciones (7) pueden entenderse como el balanceo de las fuerzas internas y externas que actúan sobre el contorno haciendo que este se mueva de acuerdo a la fuerza resultante obtenida. En el modelo KWT la fuerza externa se debe únicamente a la superficie potencial creada a partir del módulo del gradiente de la imagen. La limitación de este modelo se presenta cuando examinamos una región homogénea: en estos puntos el gradiente de la imagen es cero o cercano a cero. Por tal razón, si la snake empieza su camino en estos valores será imposible que pueda converger hacia los bordes del objeto (Figura 15). En otras palabras, las líneas de campo formadas por la fuerza externa (f_x, f_y) , tienen un rango de captura limitado.

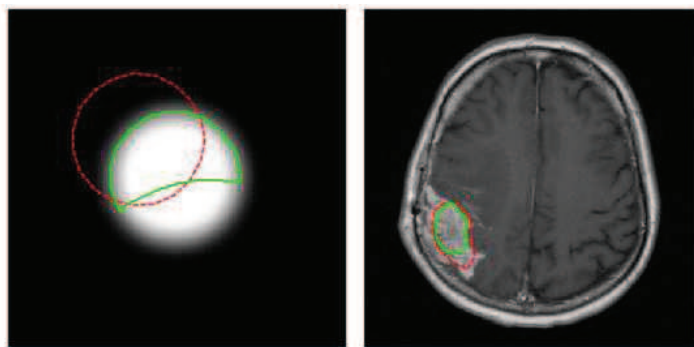


Figura 15. Segmentaciones fallidas usando la fuerza externa KWT (contorno rojo: snake inicial, contorno verde: snake final, imagen extraída de [29]).

A continuación vamos a estudiar dos tipos de fuerzas externas, una dinámica (modelo balloon) y otra estática (modelo GVF).

3.5.1. Fuerza Externa del tipo Balloon.

El modelo del *balloon* (globo) es un ejemplo de una fuerza externa dinámica donde el contorno activo se infla o desinfla (Figura 16 y Figura 17) para alcanzar al objeto deseado. El modelo balloon ejerce fuerzas que son normales al contorno activo. Dichas fuerzas pueden ser a favor del contorno o en contra del contorno [32]. El parámetro s es el encargado de establecer la dirección del contorno: si s aumenta, el contorno se infla, por el contrario si s disminuye, el contorno se desinfla y la dirección normal de la curva parametrizada viene dada por $(-\frac{\partial Y}{\partial s}, \frac{\partial X}{\partial s})$. La evolución de la snake usando la fuerza externa balloon se expresa como:

$$\frac{\partial X}{\partial \tau} = \alpha \frac{d^2 X}{ds^2} - \beta \frac{d^4 X}{ds^2} + f_x(X, Y) - c_\tau \frac{\partial Y}{\partial s} \quad (21)$$

$$\frac{\partial Y}{\partial \tau} = \alpha \frac{d^2 Y}{ds^2} - \beta \frac{d^4 Y}{ds^2} + f_y(X, Y) + c_\tau \frac{\partial X}{\partial s} \quad (22)$$

donde c_τ representa la constante de proporcionalidad que controla el grado de incremento o reducción del contorno. Al observar las ecuaciones (21) y (22) notamos que se sigue incluyendo la fuerza externa KWT. Esto es debido a que no podríamos saber si la snake ha alcanzado o no el borde del objeto sin las líneas de campo que esta fuerza nos proporciona.

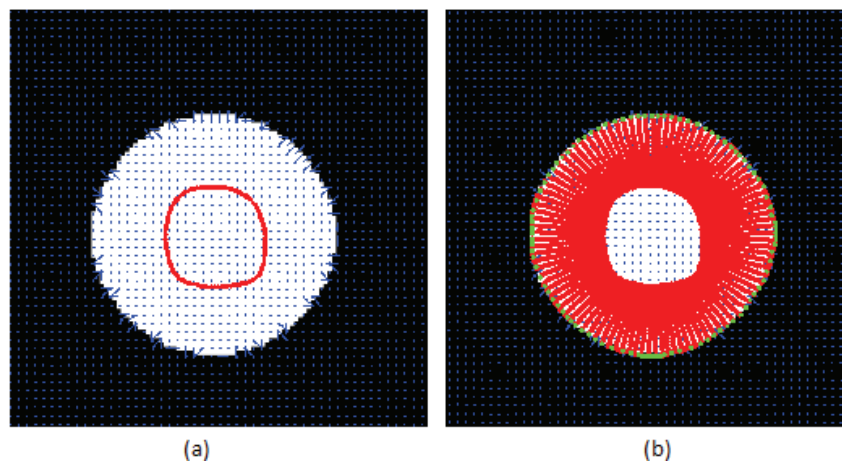


Figura 16. Fuerza externa balloon de inflación. (a) inicialización de la snake realizada dentro del objeto (b) resultado de la segmentación (contorno verde: snake final, líneas rojas representan el movimiento de los snaxels).

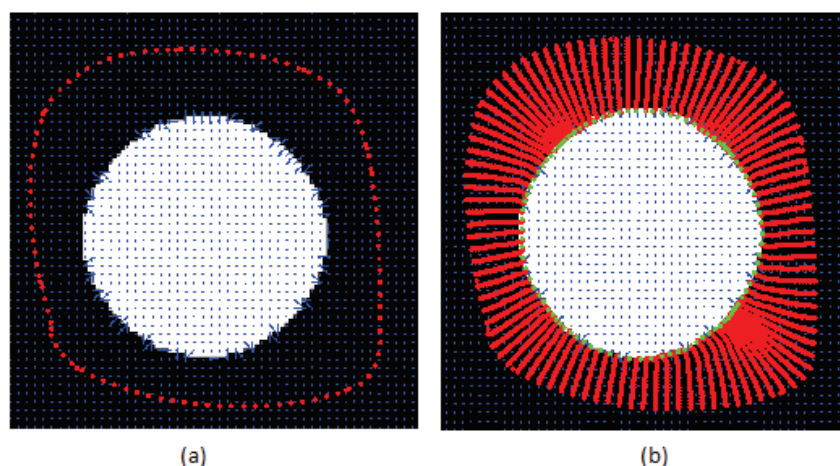


Figura 17. Fuerza externa balloon de deflación. (a) inicialización de la snake realizada fuera del objeto (b) resultado de la segmentación (contorno verde: snake final, líneas rojas representan el movimiento de los snaxels).

La fuerza externa balloon no está exenta de problemas, sobre todo a la hora de la inicialización, ya que debe existir cierto grado de conocimiento relativo de la posición del objeto a segmentar para de esta manera decidir si se inicializa dentro o fuera del objeto.

3.5.2. Fuerza externa del tipo GVF (Gradient Vector Flow).

La limitación que presenta la fuerza externa KWT, se puede subsanar gracias a la implementación propuesta por Xu y Prince [33] conocida como *Gradient Vector Flow* (GVF). La fuerza GVF representa un claro ejemplo de las fuerzas externas estáticas, donde el movimiento de la snake se rige sólo por las líneas de campo. La idea principal del GVF es propagar los vectores generados por la fuerza KWT a sus vecinos y estos a su vez a los suyos. Este proceso se realiza iterativamente a lo largo de cada píxel de la imagen manteniendo la dirección del vecino que lo generó y reduciendo su módulo conforme se aleja (Figura 18). Lo que logra la fuerza GVF es incrementar el rango de movimiento de la snake en la imagen. Las funciones (u, v) correspondientes a las ecuaciones (23) y (24) representan los nuevos vectores de campo, cuyo valor inicial es el de la fuerza KWT (f_x, f_y) .

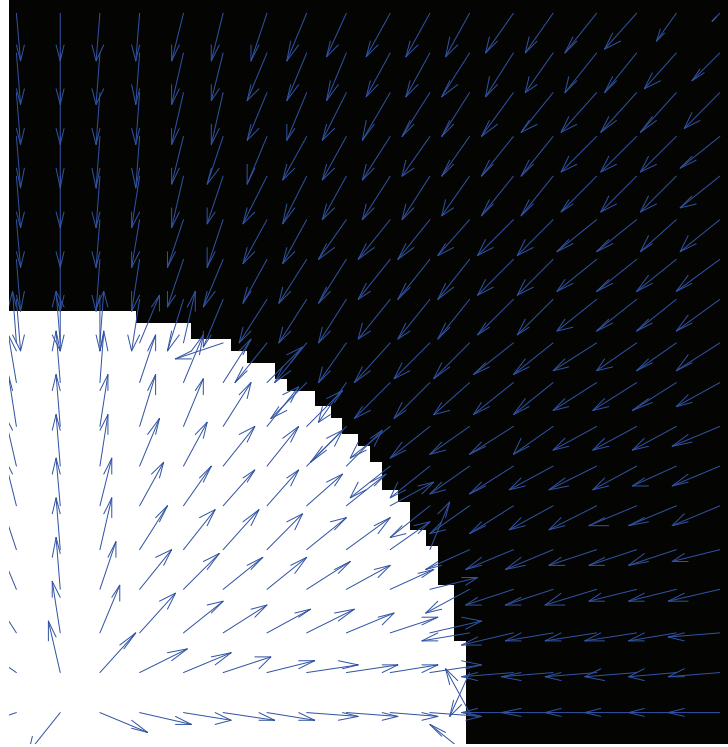


Figura 18. Líneas de campo de la fuerza GVF

$$u = \mu \nabla^2 u - (f_x^2 + f_y^2)(u - f_x) \quad (23)$$

$$v = \mu \nabla^2 v - (f_x^2 + f_y^2)(v - f_y) \quad (24)$$

Donde μ es un parámetro positivo que controla la suavidad del campo (u, v) . El primer término de las ecuaciones (23) y (24) mantiene el campo suave mientras que el segundo hace que el campo formado en los bordes del objeto se propague. Las ecuaciones correspondientes a la evolución de la snake son:

$$\frac{\partial X}{\partial \tau} = \alpha \frac{d^2 X}{ds^2} - \beta \frac{d^4 X}{ds^2} + u(X, Y) \quad (25)$$

$$\frac{\partial Y}{\partial \tau} = \alpha \frac{d^2 Y}{ds^2} - \beta \frac{d^4 Y}{ds^2} + v(X, Y) \quad (26)$$

3.6. Inicialización del Contorno Activo.

En lo relativo a la inicialización se pueden considerar tres opciones fundamentales: naif, manual y automática.

3.6.1. Inicialización naif.

La inicialización *naif* (es decir, ingenua) define el contorno inicial sin conocimiento previo del contenido de la imagen. Por ejemplo, se podría inicializar la snake en la frontera de la imagen o inicializarla con alguna forma geométrica que se asemeje al objeto a segmentar. Generalmente este tipo de inicializaciones se sitúan distantes del objeto de interés y es necesaria una gran cantidad de iteraciones para capturar la región deseada. Además se corre el riesgo de que la snake sea atraída hacia otras características sobresalientes de la imagen, convergiendo hacia objetos no deseados. Este inconveniente se resuelve realizando un pre-procesado previo, de forma que se resalten las características del objeto de interés y se reduzcan en la medida de lo posible las características no relevantes.

3.6.2. Inicialización manual.

En la inicialización manual, el usuario crea un contorno cerca del objeto a segmentar o en su defecto establece un punto en el interior del objeto, de forma que sirva de referencia para crear el contorno. La tarea de establecer un contorno de manera manual suele consumir mucho tiempo en el proceso de interacción usuario – sistema, además de ser tediosa y poco conveniente cuando se quiere realizar un proceso completamente automatizado. En aplicaciones relacionadas con imágenes en tres dimensiones, la inicialización manual es muy compleja y poco práctica.

3.6.3. Inicialización automática.

Para realizar la inicialización automática se pueden considerar tres alternativas: *Centro de divergencia* (CoD), que se define como el punto donde los vectores de todos sus píxeles vecinos empiezan a “emanar”. Este concepto se entiende mejor si observamos en detalle la Figura 19, en la cual la inicialización del contorno se realiza desde el lugar donde nacen las líneas de campo. Uno de los inconvenientes presentados por la técnica CoD son los problemas de sobre-segmentación, por lo que en muchas ocasiones es necesario emplear algoritmos de post-procesado para lograr unir las regiones.

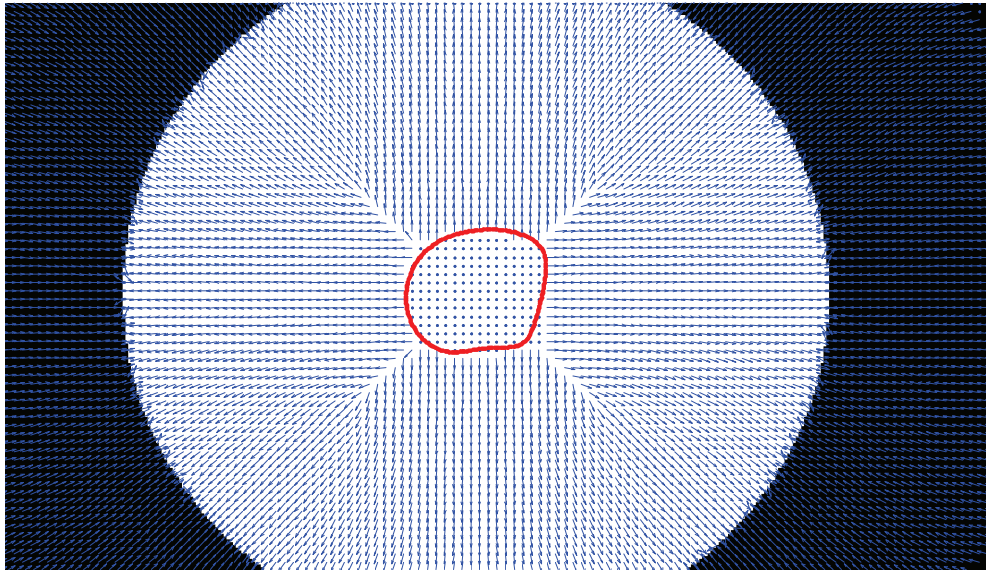


Figura 19. Inicialización de una snake usando el método CoD (en rojo, snake inicial)

Otro método de segmentación automático es el llamado *force field segmentation* (FFS) [34], en el cual las inicializaciones se realizan a partir del campo formado por la fuerza externa, logrando encontrarse N regiones disjuntas pertenecientes a los N objetos a segmentar. La Figura 20 y la Figura 21 ilustran el proceso.

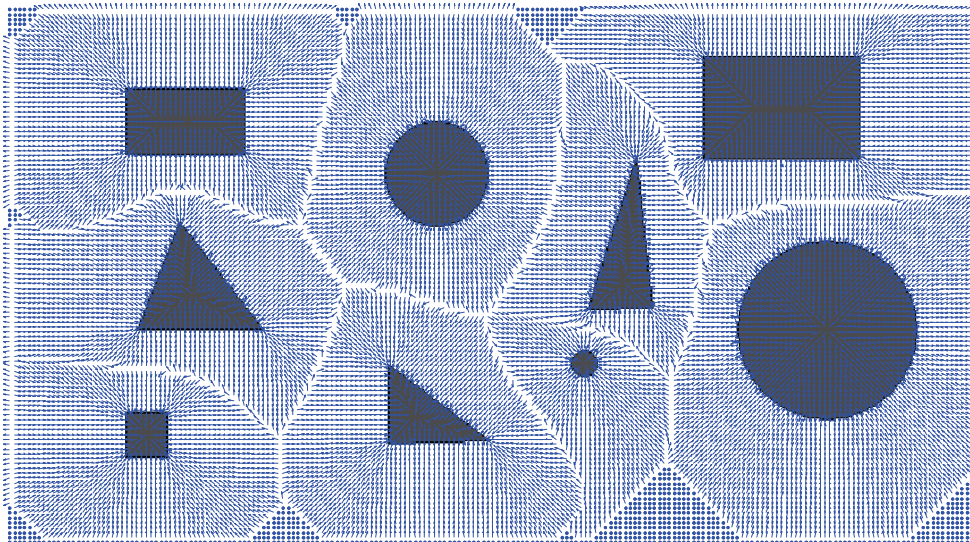


Figura 20. Líneas de campo que representan nueve regiones disjuntas

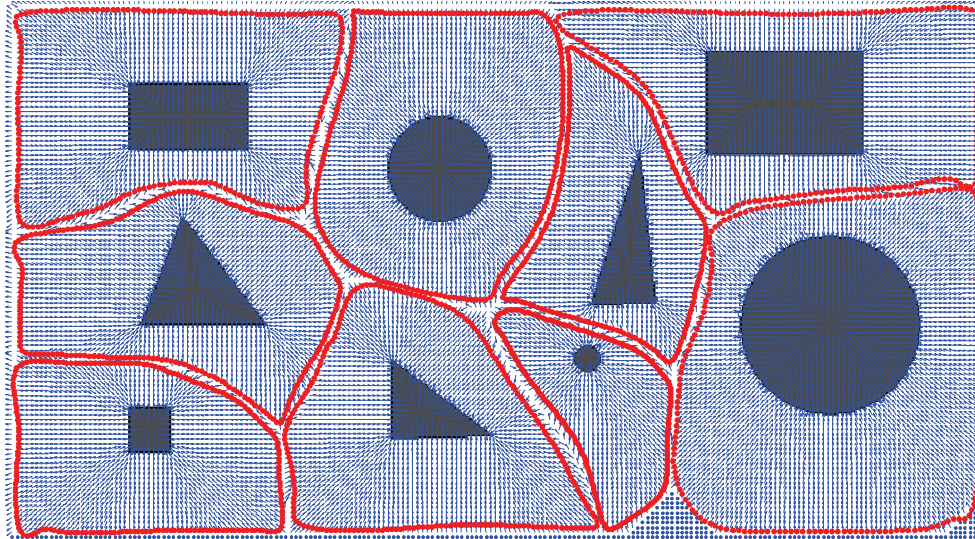


Figura 21. Inicialización usando FFS

El tercer método utilizado para la inicialización automática de la snake es el llamado *Poisson Inverse Gradient* (PIG) [35]. PIG es un claro ejemplo de una solución basada en el enfoque del problema inverso, en el cual se obtiene la energía externa de la imagen $E_{ext}(x, y)$ a partir del campo generado por la fuerza externa.

4 Metodología Propuesta

4.1. Metodología.

El método propuesto permite individualizar la glotis en imágenes laríngeas siguiendo el esquema que se presenta en la Figura 22. El funcionamiento de cada uno de los bloques se detalla a continuación:

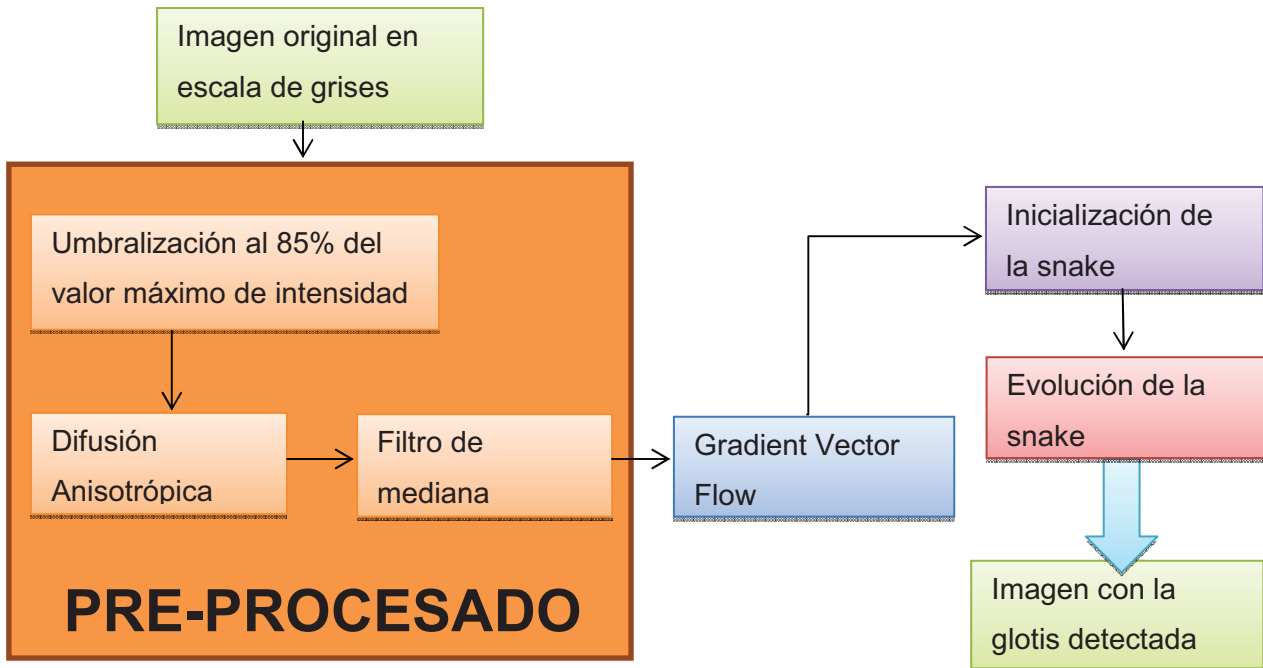


Figura 22. Etapas del proceso seguido para la detección de la glotis

4.2. Pre-procesado.

Antes de empezar el pre-procesado es necesario convertir la imagen original (RGB) a escala de grises por medio de una transformación según el modelo YIQ [17]. Después de dicha conversión se usa la luminancia “Y” para generar la imagen en escala de grises (Figura 23) necesaria para iniciar las etapas de acondicionamiento de la imagen. El pre-procesado tiene como objetivo suavizar la imagen y resaltar los píxeles que correspondan a la glotis para evitar de esta manera que la snake se ajuste a características no deseadas. El bloque de pre-procesado está formado por tres etapas: umbralización, difusión anisotrópica y filtro de mediana.

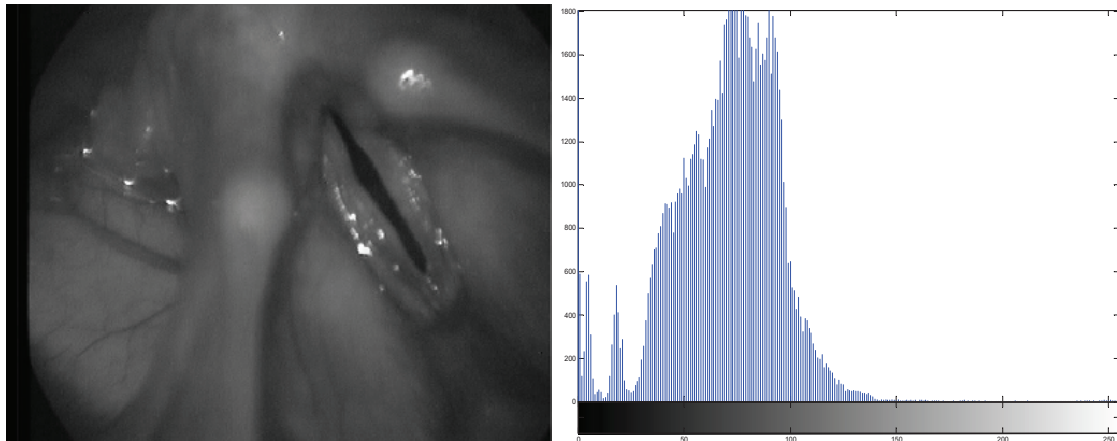


Figura 23. Imagen laríngea en escala de grises e histograma

4.2.1. Umbralización.

La etapa de umbralización tiene como finalidad reducir el contraste presentado en las imágenes para facilitar la labor del filtro anisotrópico. Si observamos la Figura 23 podemos decir a ciencia cierta que la glotis corresponde al objeto negro rodeado de valores de grises. Sin embargo, si pensáramos únicamente en buscar los píxeles más oscuros dentro de la imagen, caeríamos en un error, ya que en los bordes del fotograma se concentra una gran cantidad de píxeles cuyos valores se encuentran cercanos a cero. Para tener una idea más clara de cómo están distribuidos los valores de los píxeles dentro de la imagen observemos el histograma. Del histograma podemos deducir que la glotis no es necesariamente el objeto más oscuro de la imagen. Por tanto, establecer un valor o un rango de valores de umbral es una tarea complicada y más aún sabiendo de la variabilidad inter-vídeo existente. La Figura 24 muestra un fotograma proveniente de un video diferente al de la Figura 23. Claramente observamos la gran diferencia de intensidades presentadas entre ambas, lo que queda aún más patente si vemos el histograma correspondiente a esta imagen y lo comparamos con el de la Figura 23. Con base en el conocimiento de que la glotis es siempre más oscura que el fondo que la rodea, podemos reasignar los valores de los píxeles que superen un cierto valor de intensidad. Al realizar esto lo que se desea es igualar en cierta manera los tonos de grises en el fondo y eliminar información no relevante para la segmentación. El umbral seleccionado inicialmente fue el pico máximo del histograma, ya que este representa a los píxeles localizados en el fondo de la imagen, posteriormente al realizar pruebas con las imágenes de la base de datos, se constató que un umbral más adecuado es el 85% del pico máximo por la izquierda.

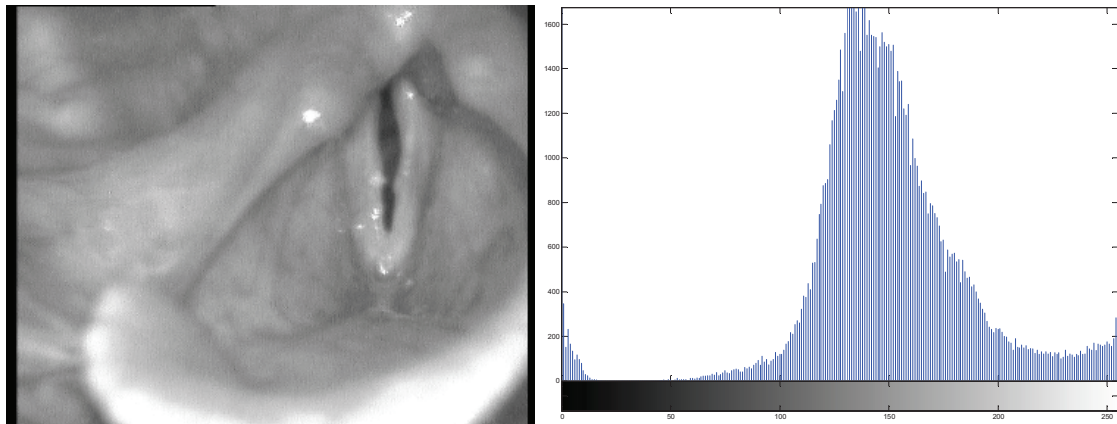


Figura 24. Imagen laríngea en escala de grises e histograma.

Se decidió por este umbral para reducir la mayor cantidad de píxeles oscuros que no pertenezcan a la glotis y emparejar el nivel de la intensidad en el fondo de la imagen. Los valores localizados en los extremos del histograma (ver Figura 24) no son relevantes, ya que la glotis nunca se encontrará en estos rangos. La Figura 25 y la Figura 26 muestran las imágenes obtenidas con sus respectivos histogramas.

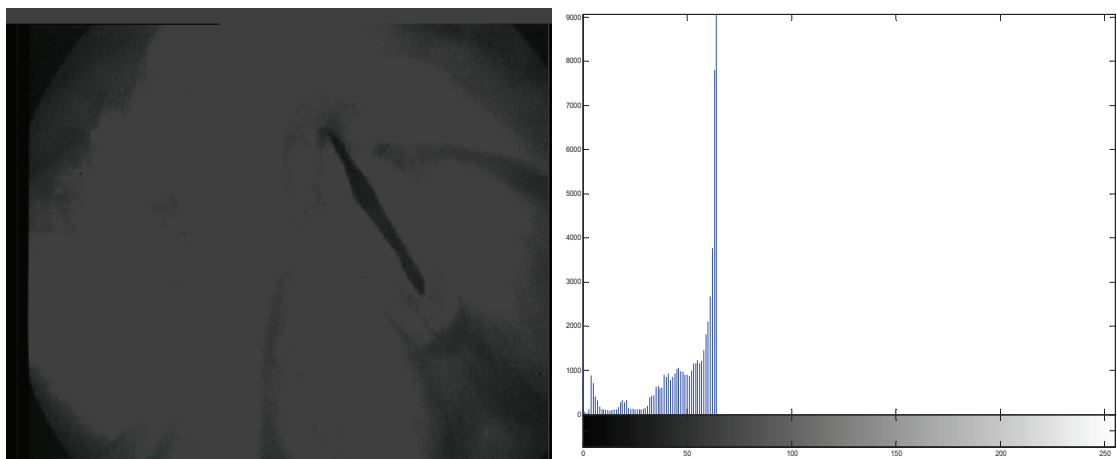


Figura 25. Umbralización e Histograma al 85% de la Figura 23

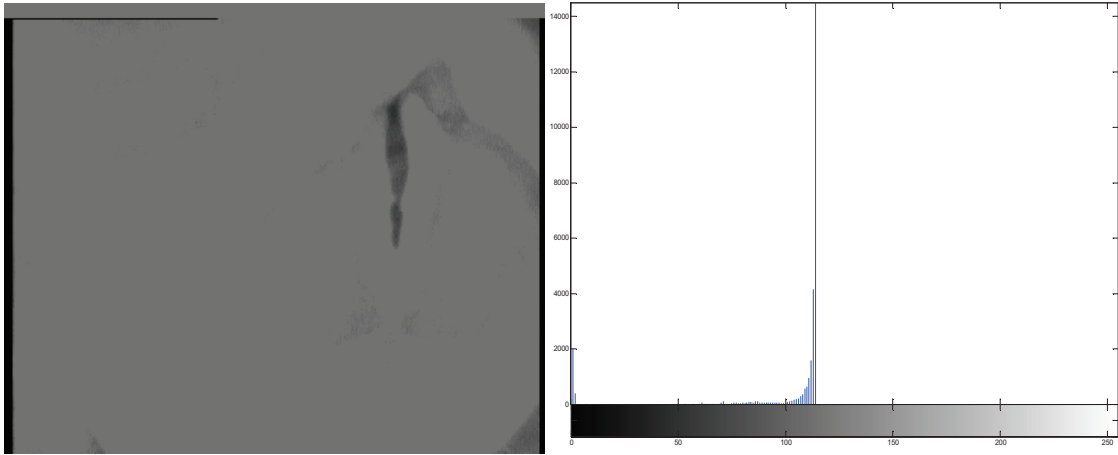


Figura 26. Umbralización e Histograma al 85% de la Figura 24

4.2.2. Difusión Anisotrópica.

El escalado espacial con difusión anisotrópica fue propuesto por P. Perona [36] y tiene como objetivo suavizar las regiones delimitadas por bordes sin afectar a estos. Basándonos en [37] se puede lograr el efecto deseado mediante la aplicación recursiva de la ecuación (27) en todos los píxeles de la imagen obtenida en la etapa de umbralización.

$$I_{i,j}^{t+1} = I_{i,j}^t + \lambda \sum_{l=N,S,E,W} [C_l \cdot \nabla_l I]_{i,j}^t \quad (27)$$

Los valores de i y j indican la posición del píxel dentro de la imagen, donde i es la fila y j la columna; t indica el número de iteración en la que se encuentra el proceso de recursividad; $I_{i,j}^t$ representa el estado en que queda la imagen I una vez realizada la iteración t sobre ella; ∇ introduce la diferencia de intensidad del punto actual con cada uno de los puntos de su vecindad, en una conectividad 4: norte-N, sur-S, este-E y oeste-W; C es el *coeficiente de conducción*, su valor se establece de forma distinta para cada dirección de vecindad, en función de la diferencia de nivel de gris anterior. Cuando ∇ es grande estamos cerca a un borde y C será próximo a cero, mientras que para valores pequeños de ∇ , C tenderá a la unidad. Existen varias formas posibles de determinar la función que implementa C sobre ∇ , siendo la más utilizada una señal con exponencial decreciente y caída controlada por una constante prefijada k . Finalmente el parámetro λ es un valor que regula la velocidad a la que se realiza el proceso y que no debería superar 0.25 para asegurar la estabilidad numérica del sistema. Para cada píxel de la imagen se calculan las diferencias de nivel de gris con sus vecinos en las

cuatro direcciones consideradas. Aquellas diferencias altas (píxel cerca de un borde) se verán multiplicadas por un coeficiente de conducción cercano a cero, no afectando por tanto al resultado de la iteración. Las diferencias pequeñas significarán sin embargo una ligera sustracción o adición al valor del píxel en estudio. De esta forma, tras sucesivas iteraciones, todos los píxeles de la imagen en zonas de nivel de gris similar tenderán a tener el mismo valor, mientras que se mantendrán intactas las diferencias en los bordes. La Figura 27 y Figura 28 muestran el resultado obtenido al aplicar la difusión anisotrópica, cuando se establece el valor de los parámetros t , k y λ en 60, 6 y 0.25 respectivamente.

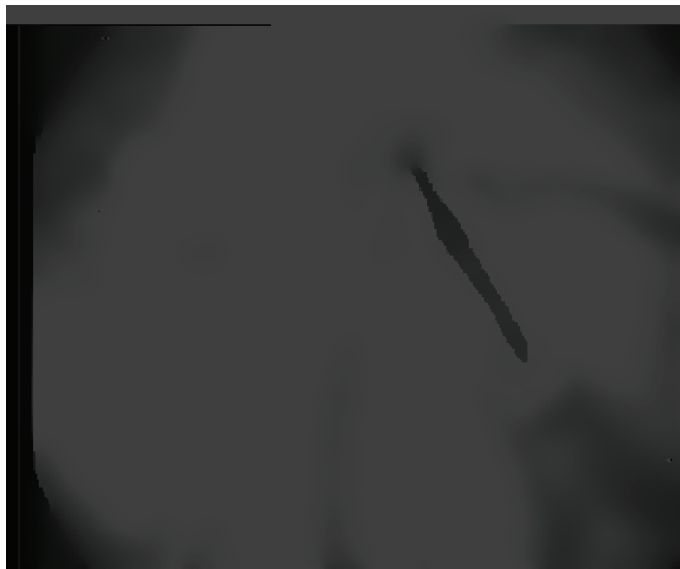


Figura 27. Difusión anisotrópica de la Figura 25



Figura 28. Difusión anisotrópica de la Figura 26

4.2.3. Filtro de mediana.

Un filtro de mediana reemplaza el valor de gris de un punto por la mediana de los niveles de gris de una cierta vecindad [38]. La mediana m de un conjunto de valores es aquel tal que la mitad de los valores del conjunto son menores que m y la otra mitad son mayores que m . La función principal de los filtros de mediana es forzar a los puntos aislados con valores de intensidad muy distintos a sus vecinos (lo que en procesado de imagen se conoce como “ruido sal y pimienta”) a tener valores más próximos a ellos. En la Figura 29 se muestra un ejemplo de una imagen en la que notamos que luego de aplicar la difusión anisotrópica aparece un ruido sal y pimienta que a pesar de ser mínimo podría afectar a la snake si la inicialización se realizara lejos de la glotis. La Figura 30 muestra el resultado obtenido luego de aplicar el filtro de mediana.



Figura 29. Imagen a la salida de la etapa de difusión anisotrópica



Figura 30. Imagen a la salida del filtro de mediana

4.3. Gradient vector flow.

Como ya se explicó en el marco teórico, la fuerza GVF es de tipo estática. Por tal razón no es necesario que se inicialice ningún contorno para generar las líneas de campo y el comportamiento de las líneas dependerá únicamente de las características de la imagen. El primer paso es generar la fuerza KWT que será necesaria para obtener la fuerza GVF. Para esto es necesario recordar la relación que existe entre fuerza y energía (que a su vez responde a la ecuación (2)). Para generar el campo GVF se resuelven las ecuaciones (23) y (24), donde u y v tomarán como valores iniciales los de f_x y f_y . El valor de la constante μ se estableció en 0.2 para todas las imágenes. Con este valor se garantiza que el movimiento de las líneas de campo sea suave y ordenado. La Figura 31 y la Figura 32 muestran el efecto que produce la variación de μ en las líneas de campo.

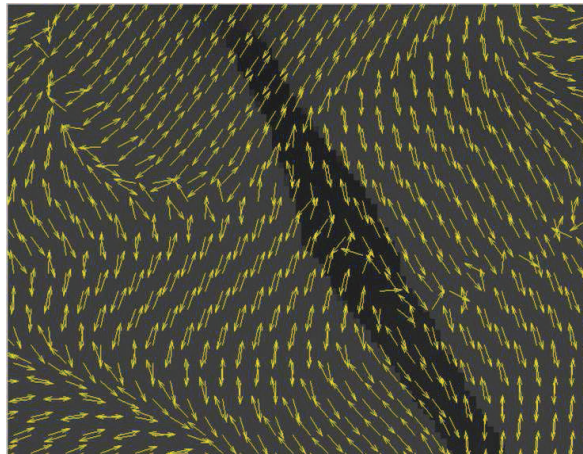


Figura 31. Líneas de fuerza GVF con $\mu = 0.5$

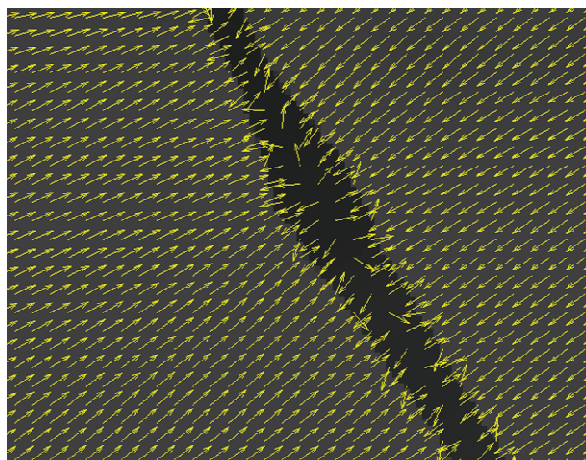


Figura 32. Líneas de fuerza GVF con $\mu = 0.2$

El número de iteraciones representa la capacidad que tiene el campo de atraer a la snake, lo que se ve reflejado en las imágenes de la Figura 33 y Figura 34. Para el caso de este algoritmo se optó por un número de doscientas iteraciones para que de esta manera no sea indispensable inicializar cerca de la glotis.

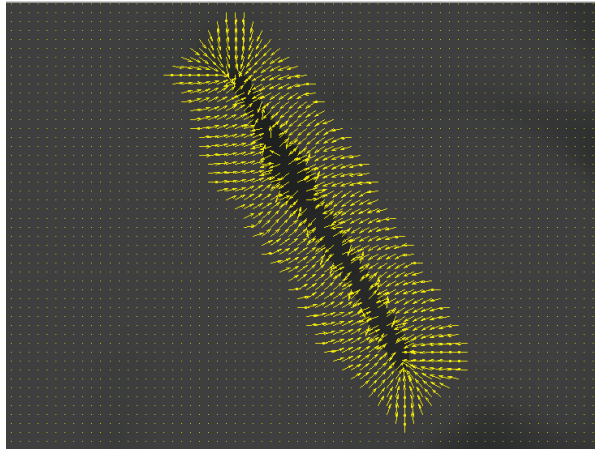


Figura 33. Líneas de fuerza GVF con 20 iteraciones

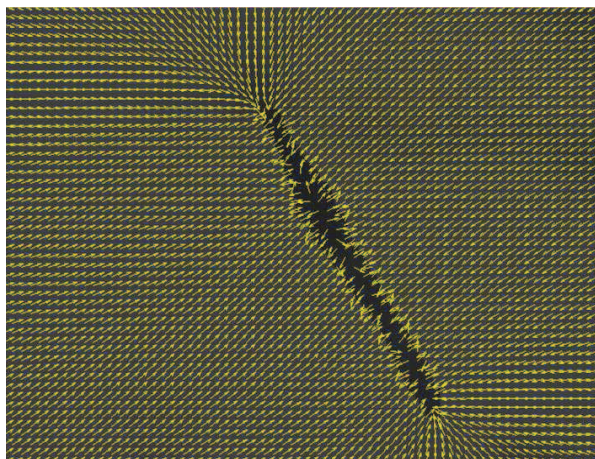
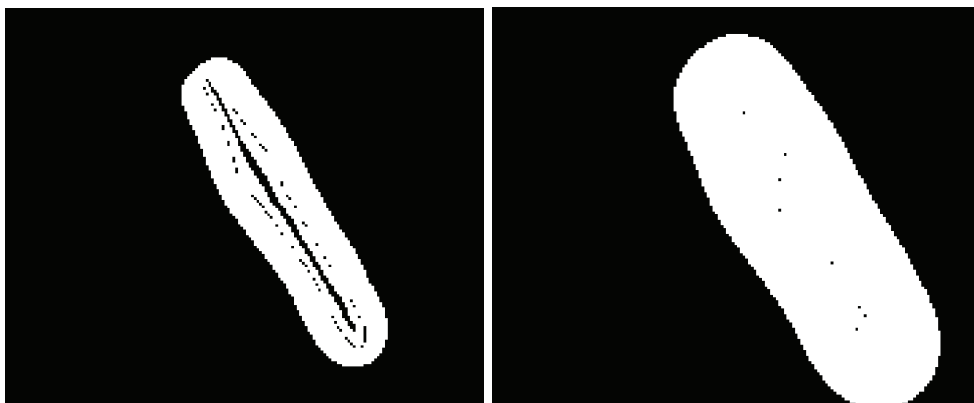


Figura 34. Líneas de fuerza GVF con 200 iteraciones

4.4. Inicialización.

La inicialización está basada en el enfoque del problema inverso. En otras palabras, lo que se pretende es estimar la energía externa a partir de la fuerza externa GVF. Los valores de u y v representan las componentes de los vectores de fuerza externa en los puntos (x, y) . Por ejemplo, el primer vector se define por los componentes de $u(1)$, $v(1)$ y este se muestra en el punto $x(1)$, $y(1)$. El módulo de los vectores de fuerza externa nos indica cuán cerca nos encontramos de una característica sobresaliente de la imagen. Por su parte, si lo que se desea encontrar es

el lugar exacto en el cual se localiza la glotis, el módulo de los vectores no es suficiente. Sin embargo es muy útil en lo que a inicializar la snake se refiere. Seleccionar un valor de módulo depende de cuán cerca se quiera inicializar la snake de la glotis. En todo caso, ambas apreciaciones tienen sus ventajas y desventajas. Para el caso de una inicialización lejos de la glotis se corre el riesgo de que dentro del área de la snake aparezcan otras características relevantes que no fueron eliminadas en la etapa de pre-procesado, lo que traería un error en la segmentación. A esto hay que añadir la mayor cantidad de movimientos que necesita la snake para converger en la glotis. Por otro lado, si inicializamos cerca de la glotis y no se selecciona un valor adecuado podríamos caer dentro de sus píxeles, lo cual conllevaría complicaciones en la normal evolución de la snake. Para establecer la metodología de trabajo se probó con ambos enfoques, decidiéndose finalmente por inicializar la snake cerca de la glotis y así eludir el ruido que queda después de la etapa de pre-procesado. Para la elección del módulo se probó con varios valores hasta encontrar los apropiados. Si se decidiese por una inicialización cerca de la glotis (como así ha sido en este algoritmo) los valores deben ser mayor que 0.09. En caso contrario los valores a tomar tendrían que ser mayores que 0.009. El procedimiento se basa en generar una máscara con un valor de 1 en los píxeles que sean mayores que el umbral definido y cero para los otros casos (Figura 35). Posteriormente se extraen los bordes de las nuevas imágenes obtenidas (Figura 36), se selecciona el borde de mayor área (Figura 37) y finalmente se extrae las coordenadas correspondientes a este borde y se sobrepone en la imagen laríngea (Figura 38). Una consideración adicional es que los puntos del contorno deben estar ordenados en sentido de las manecillas del reloj, ya que si esto no se realiza, la etapa de la evolución de la snake no funcionaría adecuadamente.



(a)

(b)

Figura 35. Imagen obtenida tras la inicialización GVF. (a) umbral 0.09 (b) umbral 0.009

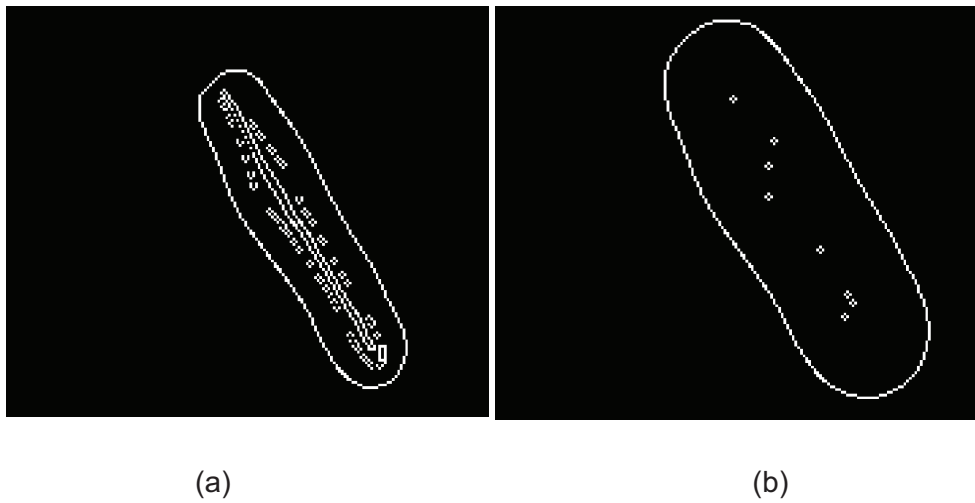


Figura 36. Bordes obtenidos de las imágenes. (a) umbral 0.09 (b) umbral 0.009

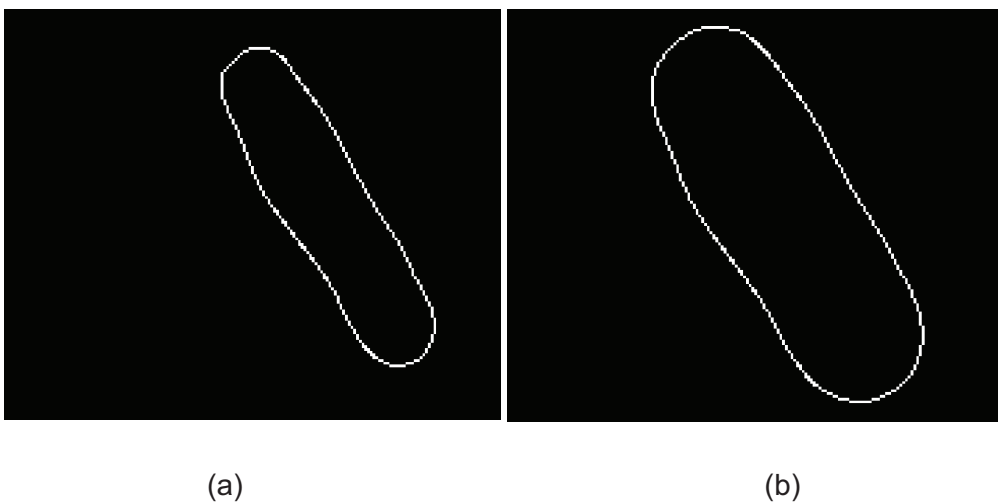
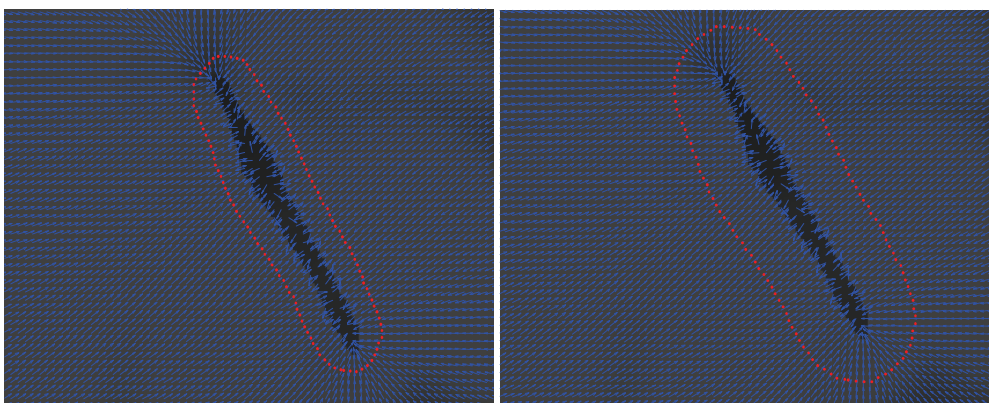


Figura 37. Bordes con mayor área. (a) umbral 0.09 (b) umbral 0.009



(a)

(b)

Figura 38. Resultado de la inicialización. (a)Umbral 0.09 (b)Umbral 0.009

4.5. Evolución de la snake.

Una vez determinado el contorno inicial se procede a representarlo según las ecuaciones (12) y (13). En este caso como nuestra fuerza externa es la GVF, se realiza una modificación en la nomenclatura de las ecuaciones (14) y (15), por la cual las nuevas expresiones quedarían:

$$u_x^t \equiv [u_x(X_0^t, Y_0^t), \dots, u_x(X_{n-1}^t, Y_{n-1}^t)]^T \quad (28)$$

$$v_y^t \equiv [v_y(X_0^t, Y_0^t), \dots, v_y(X_{n-1}^t, Y_{n-1}^t)]^T \quad (29)$$

Como último paso hay que reemplazar los valores de f_x^t y f_y^t de las ecuaciones (19) y (20) por u_x^t y u_y^t . Con esto se logra que la snake se mueva (Figura 39) hasta llegar a la glotis (Figura 40). El número de iteraciones necesarias para que la snake alcance a la glotis es de 50 o menos, por lo que se decidió tomar este valor como punto de finalización de las iteraciones. La Figura 46 ilustra mejor el resultado final de la segmentación.

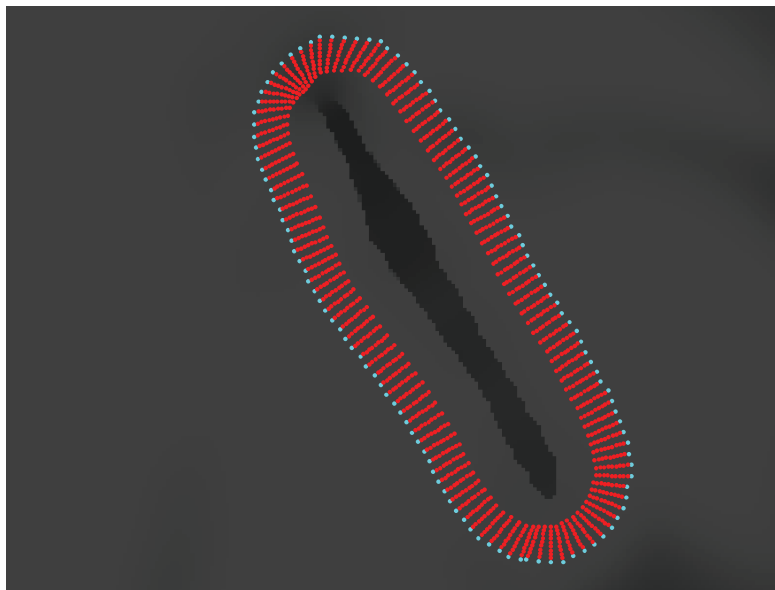


Figura 39. Evolución de la snake al cabo de 10 iteraciones.

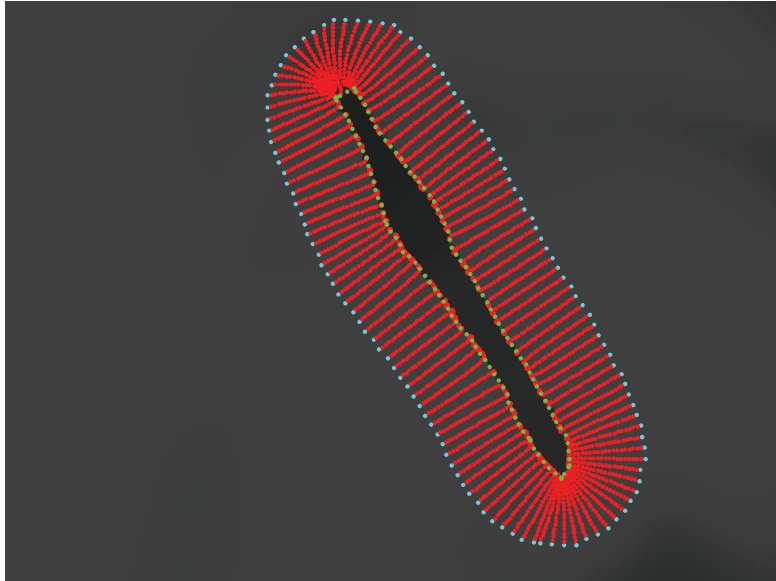


Figura 40. Evolución de la snake al cabo de 50 iteraciones.

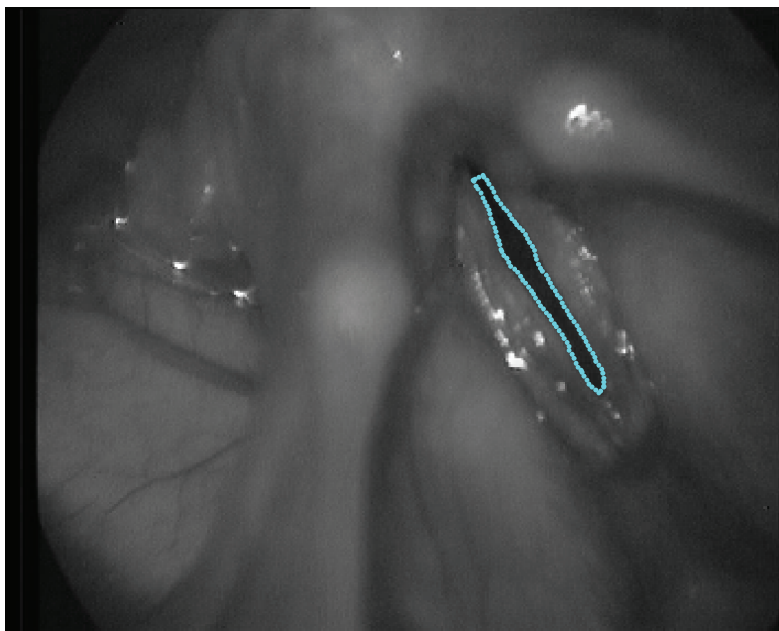


Figura 41. Resultado final de la segmentación.

5

Marco

Experimental

5.1. Introducción.

La metodología descrita en el capítulo anterior ha sido probada con 110 imágenes, extraídas de 15 vídeos grabados en el hospital Gregorio Marañón en Madrid. Los vídeos estroboscópicos se obtuvieron a partir de equipos provenientes de dos fabricantes. Todas las imágenes usadas para este algoritmo presentan las cuerdas vocales abiertas.

El presente capítulo está formado por dos apartados: El apartado 5.2 muestra de qué manera se ve afectada la segmentación cuando varían los parámetros de umbralización, del filtro anisotrópico, del filtro de mediana, y de la inicialización. Cabe mencionar que todos los parámetros se ajustaron según los resultados obtenidos de las pruebas realizadas sobre 15 imágenes, donde cada imagen corresponde a un fotograma de un vídeo distinto. El apartado 5.3 corresponde a las pruebas realizadas una vez fijados todos los parámetros del proceso. Estas consisten en comparar los resultados obtenidos de la metodología propuesta (en adelante lo llamaremos *sistema snake*), con una segmentación realizada manualmente (en adelante *sistema experto*). Para validar aún más la factibilidad del sistema snake se compararon también los resultados obtenidos por el sistema experto con el sistema desarrollado en [7] (en adelante lo llamaremos *sistema watershed*). El algoritmo empleado para comparar las segmentaciones es el algoritmo de Pratt. Este algoritmo calcula una figura de mérito que mide la similitud entre contornos [39]. Para nuestras pruebas se compararon los contornos correspondientes al sistema experto contra el sistema snake y al sistema experto contra el sistema watershed. Por último se compararon los resultados obtenidos por el sistema snake y el sistema watershed.

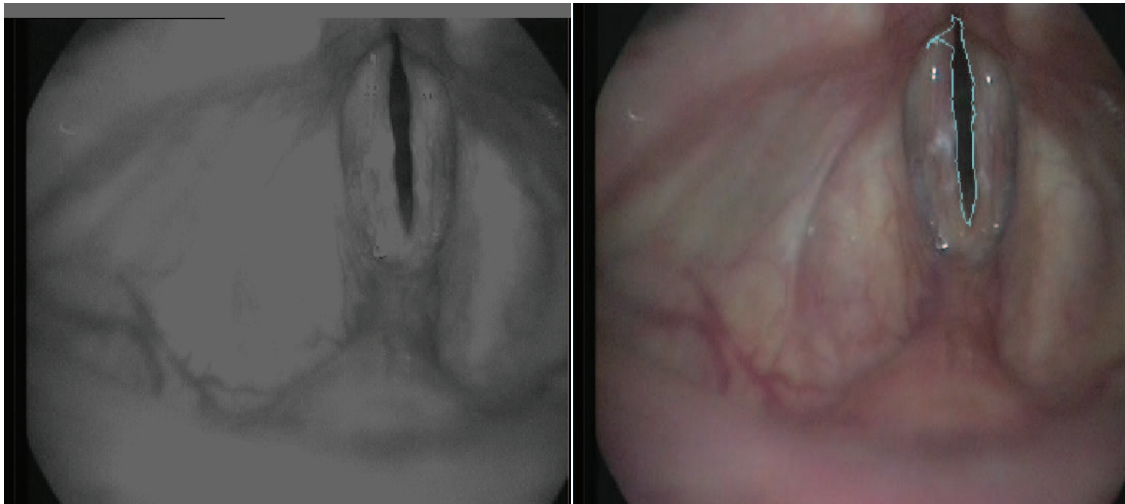
5.2. Ajuste de parámetros.

Los parámetros correspondientes a cada etapa del procesado guardan relación con la etapa siguiente. Los parámetros ajustados en una etapa anterior pueden perjudicar a una posterior, por tal motivo hay que buscar los parámetros que mejor se equilibren entre ellos. No necesariamente los mejores parámetros de una etapa serán los mejores del sistema. Antes de continuar recordemos los valores escogidos en cada una de las etapas de ajuste (ver capítulo 4), que han proporcionado un mejor resultado global con las 15 imágenes de prueba. Umbralización: 85% del pico máximo del histograma, filtro anisotrópico: $t = 60$, $k = 6$ y $\lambda = 0.25$, filtro de mediana: máscara de 3×3 e inicialización: umbral 0.09. Para evaluar la variación de los parámetros en cada

etapa consideraremos que los parámetros de las etapas anteriores y posteriores se mantienen iguales a los de los valores seleccionados.

5.2.1. Ajuste de umbral.

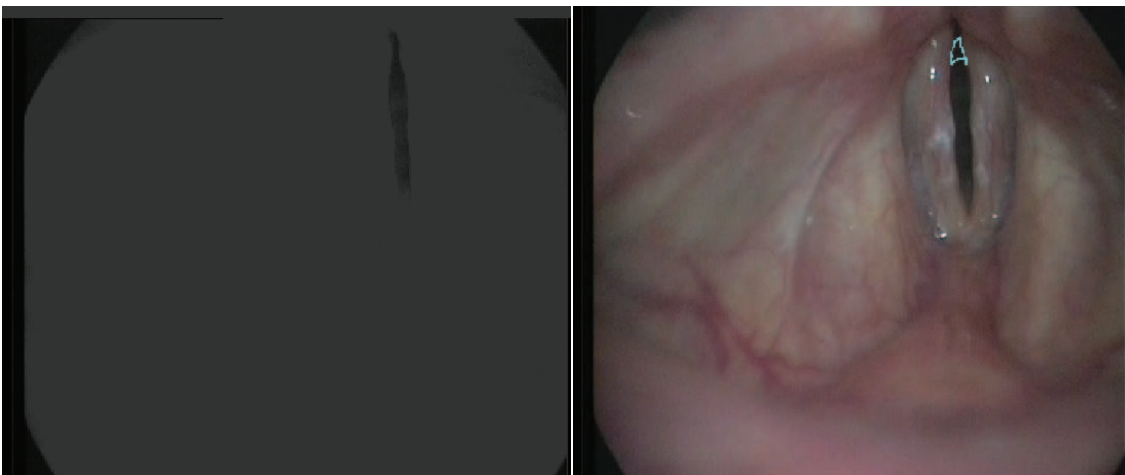
En las imágenes (Figura 42 a 45) podemos observar como la variación de un solo parámetro (umbral del histograma de la imagen) afecta a la segmentación final.



(a)

(b)

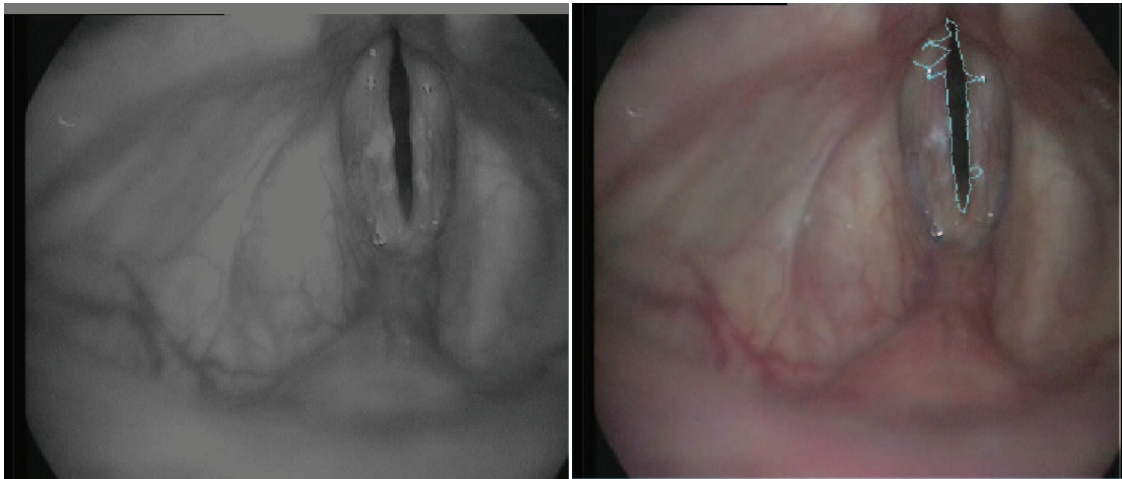
Figura 42. Ajuste del umbral al 100% del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final



(a)

(b)

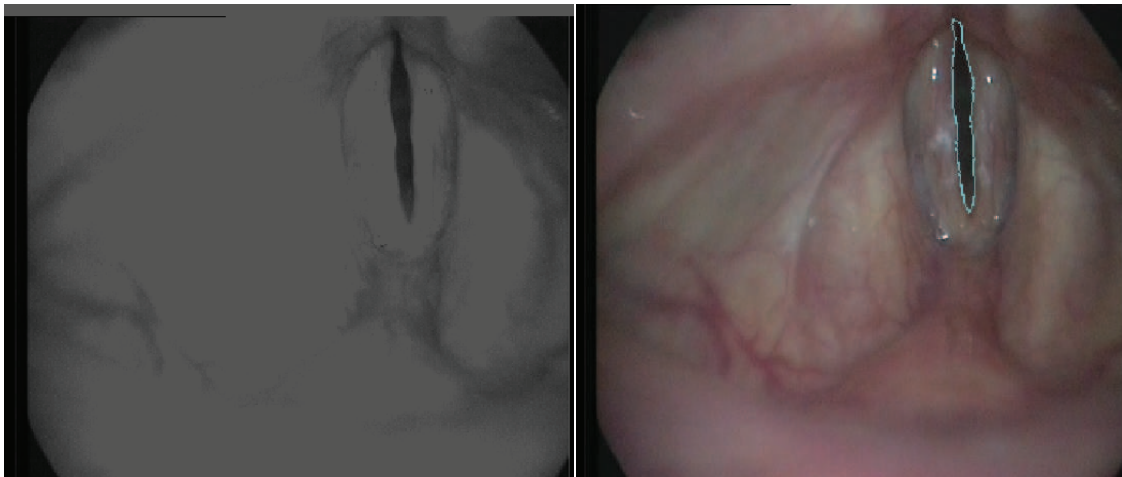
Figura 43. Ajuste del umbral al 50% del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final



(a)

(b)

Figura 44. Ajuste del umbral al 85% hacia la derecha del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final



(a)

(b)

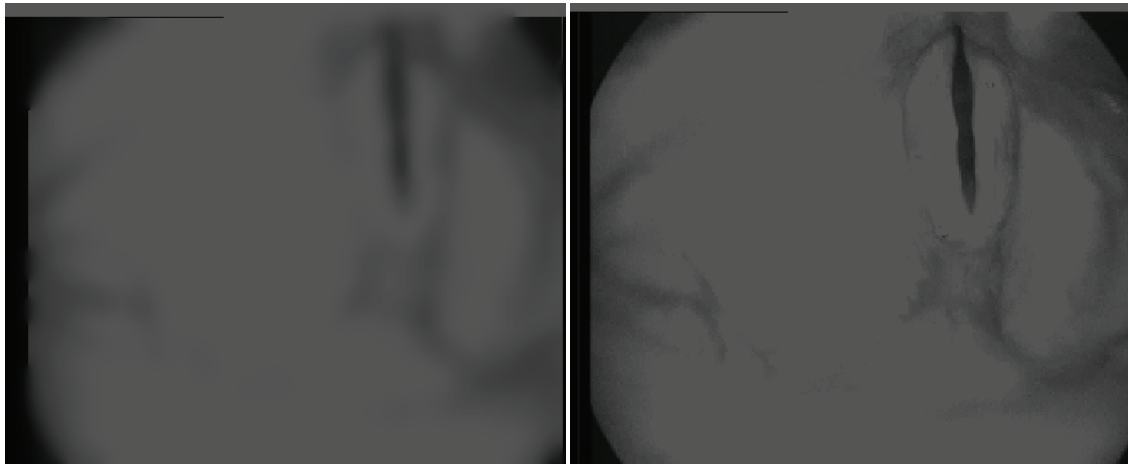
Figura 45. Ajuste del umbral al 85% hacia la izquierda del pico máximo. (a) salida de la etapa de umbralización (b) segmentación final

A simple vista se podría pensar que la Figura 43 es la mejor alternativa para la etapa de umbralización, ya que en esta se distingue fácilmente entre el fondo y la glotis. Sin embargo si observamos dentro de la glotis, podemos notar cómo ciertos píxeles toman valores iguales a los del fondo de la imagen. Dicho inconveniente se agrava cuando se aplica el filtro anisotrópico. Por el contrario, si se selecciona un

umbral mayor o igual al pico máximo (Figura 44 y Figura 42), conservamos algunos detalles oscuros que no serán eliminados tampoco por el filtro anisotrópico.

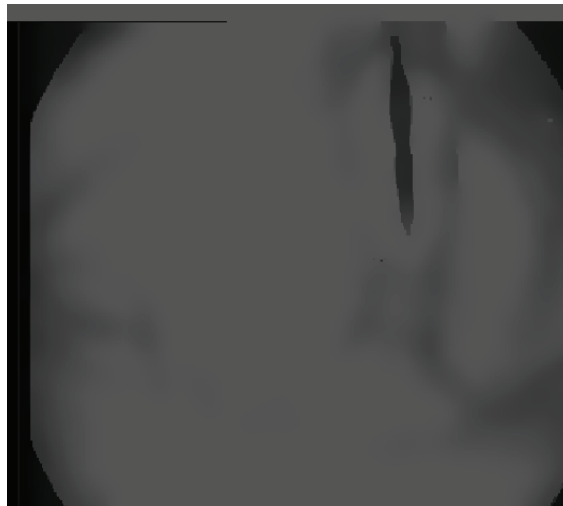
5.2.2. Ajuste del filtro anisotrópico.

Variando el parámetro k que es el encargado de controlar los coeficientes de difusión y manteniendo los otros en sus valores óptimos podemos lograr los efectos mostrados en la Figura 46.



(a)

(b)



(c)

Figura 46. Ajuste del filtro anisotrópico: $t = 60$, $\lambda = 0.25$ (a) $k = 20$, (b) $k = 1$, (c) $k = 6$

También se varió el parámetro t que indica las veces que se realiza la difusión, pero se comprobó que el filtro se ve más afectado por k . El valor de λ se mantuvo fijo en 0.25 en todas los ajuste realizados.

5.2.3. Ajuste del filtro de mediana.

De las 15 imágenes probadas se observó que 10 de ellas presentaban ruido sal y pimienta a la salida del filtro anisotrópico. Por tal motivo se decidió aplicar un filtro de mediana para eliminarlo. Como primera alternativa se intentó con un filtro recursivo de 10 iteraciones y con una máscara de 3x3, pero al comprobar que no se obtenían los resultados deseados, se optó por aplicar el filtro una sola vez, probándolo con 3 tipos de máscaras diferentes (Figura 47). Los mejores resultados se lograron con la máscara de 3x3, donde el ruido fue eliminado en 6 de las 10 imágenes.

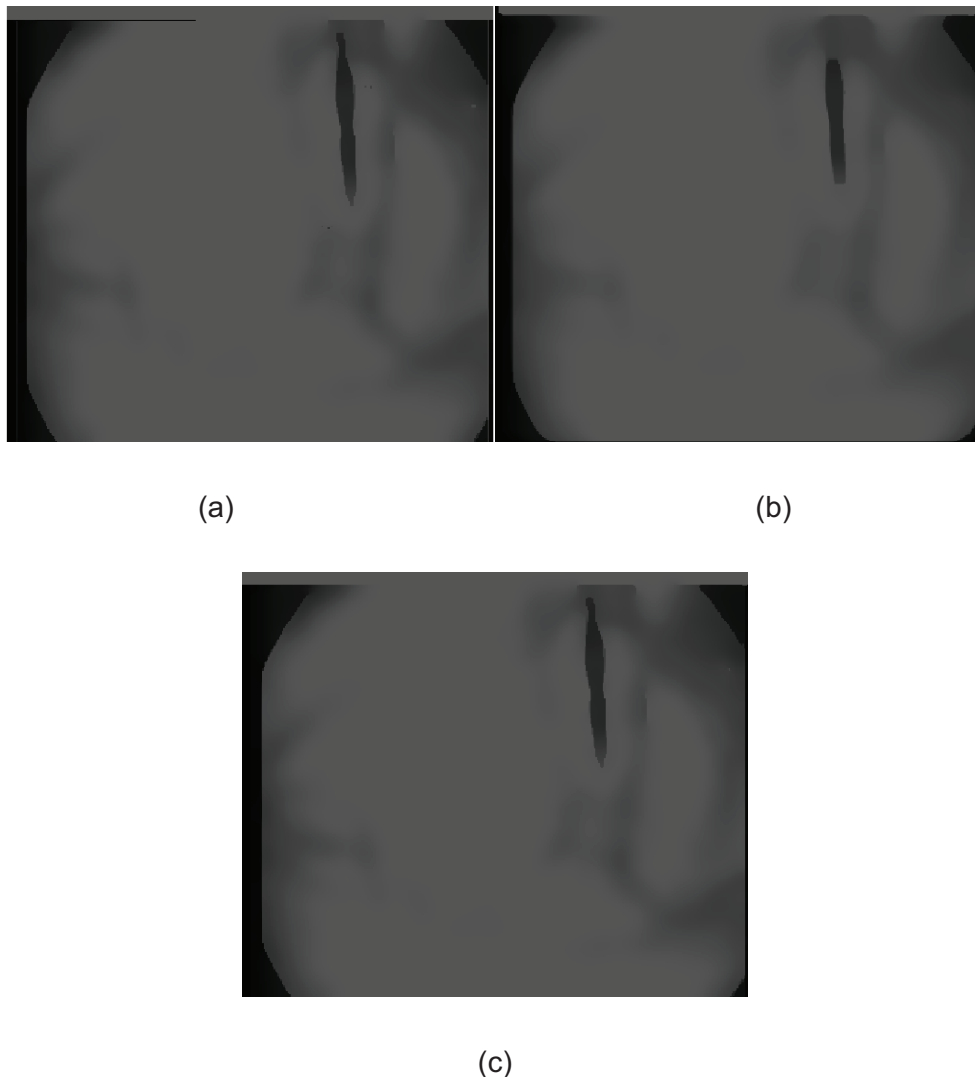
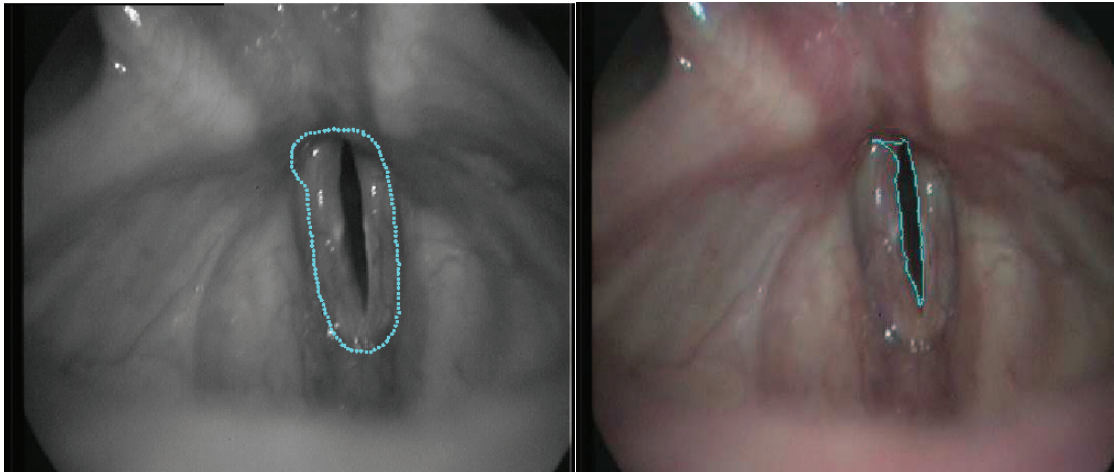


Figura 47. Ajuste del filtro de mediana. (a) sin máscara (b) máscara 9x9 (c) máscara 3x3

5.2.4. Ajuste en la inicialización.

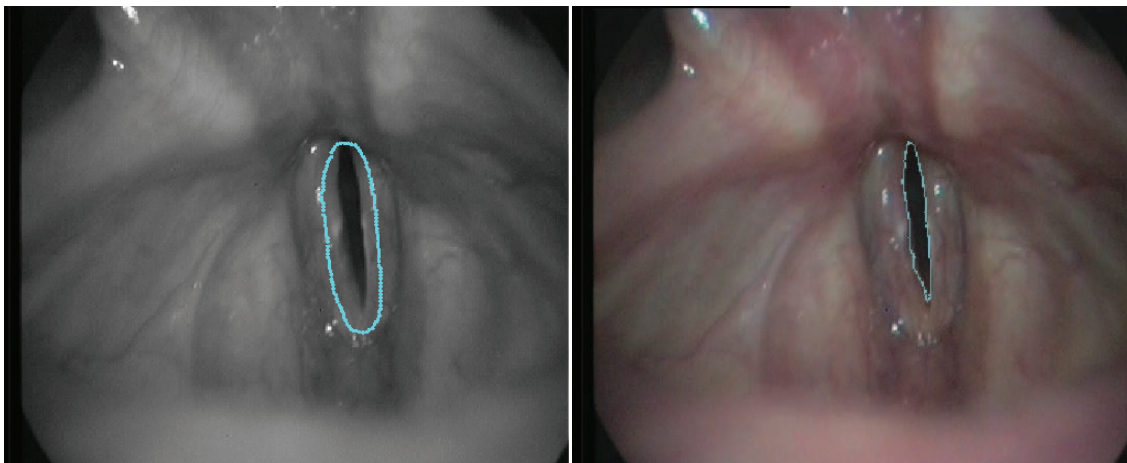
Para la inicialización se disponía de dos posibles alternativas; realizarse cerca de la glotis o lejos de esta. Si recordamos que de las 15 imágenes usadas para ajustar los parámetros, 4 presentaban aún problemas a la salida del filtro de mediana, la necesidad de ajustar los parámetros de la inicialización era imperiosa para no afectar la normal evolución de la snake.



(a)

(b)

Figura 48. Ajuste en la inicialización (a) inicialización de la snake con 0.009 (b) segmentación resultante



(a)

(b)

Figura 49. Ajuste en la inicialización (a) inicialización de la snake con 0.09 (b) segmentación resultante

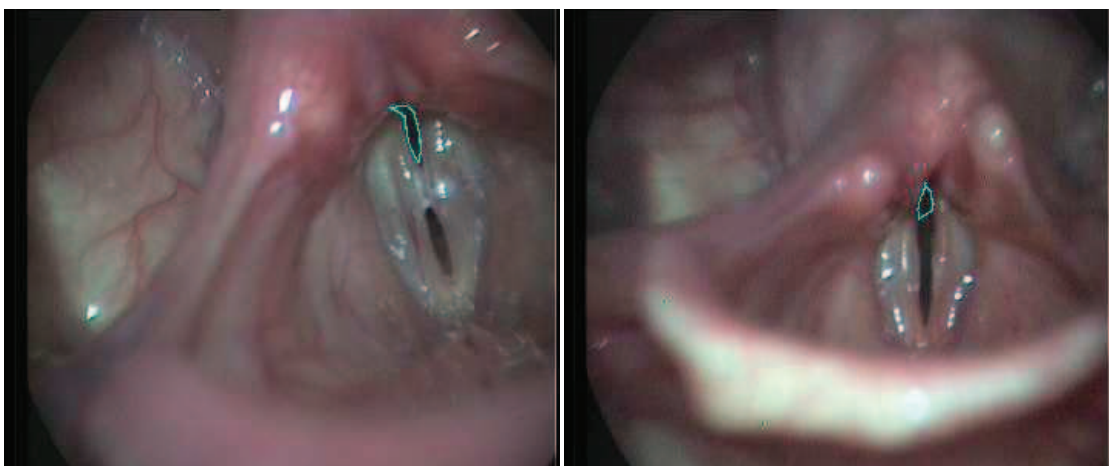
La decisión de optar por una inicialización cerca de la glotis se tomó a partir del hecho de que por más bueno que sea el pre-procesado siempre iban a existir imágenes que mantuvieran ruido y atrajeran a la snake hacia un lugar incorrecto. Por tanto si lo que se desea es estar menos afectado por estos, la solución obvia es mantenerse lejos del ruido y cerca de la glotis. La Figura 48 muestra cómo la inicialización se ve afectada por un cambio de intensidad cercano y la Figura 49 muestra cómo este es ignorado cuando el contorno inicial se encuentra lo suficientemente lejos.

5.3. Pruebas realizadas.

Se realizaron dos tipos de pruebas; la primera compara el sistema snake con el sistema experto y la segunda compara el sistema snake con el sistema watershed. El algoritmo de Pratt utilizado para medir el parecido entre ambos sistemas da como resultado valores entre 0 y 1, donde 1 indica que los dos contornos son iguales y 0 que no hay ninguna similitud. Los resultados de las pruebas se detallan a continuación:

5.3.1. Sistema snake vs Sistema experto.

Al realizar las pruebas en las 110 imágenes se obtuvieron como resultado 14 valores menores a 0.5, en su mayoría estos problemas se originaron en la etapa de pre-procesado. La Figura 50 muestra dos imágenes con errores en la segmentación.



(a)

(b)

Figura 50. Segmentación con el sistema snake (a) Pratt=0.49 (b) Pratt=0.25

Los valores obtenidos se resumen en la Figura 51. Se observa que hay 45 imágenes en el rango comprendido entre los 0.8 - 0.9 y 42 en el rango de 0.9 - 1. Es

decir, el 81% de las imágenes tiene un índice superior a 0.8. El coeficiente de Pratt con el valor más alto se logró en la Figura 52 con un valor de 0.96.

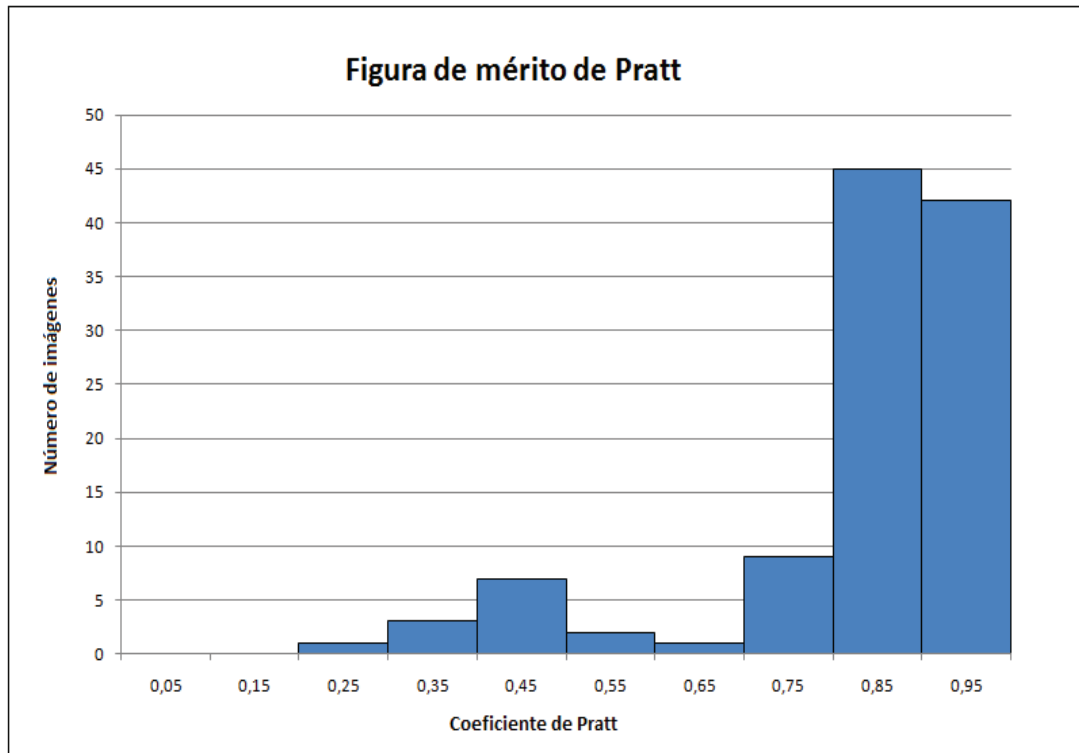


Figura 51. Figura de mérito de Pratt sistema snake – sistema experto

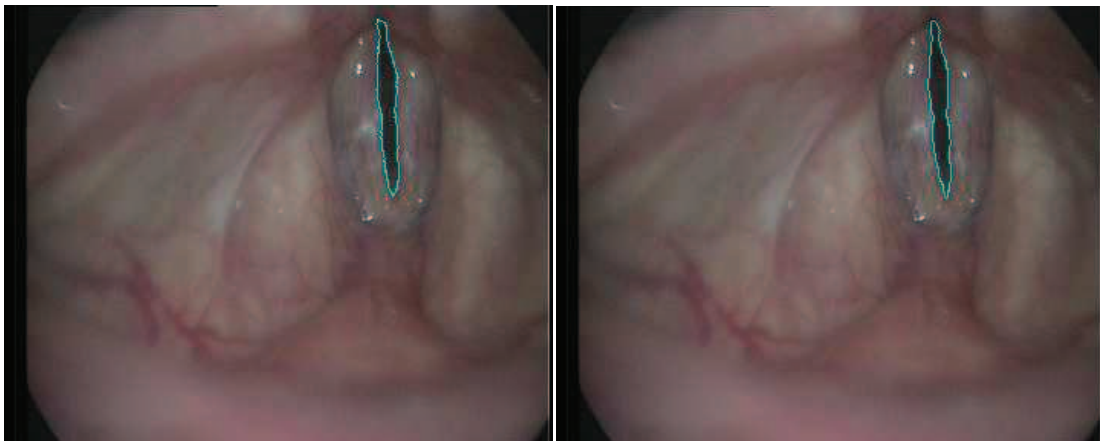
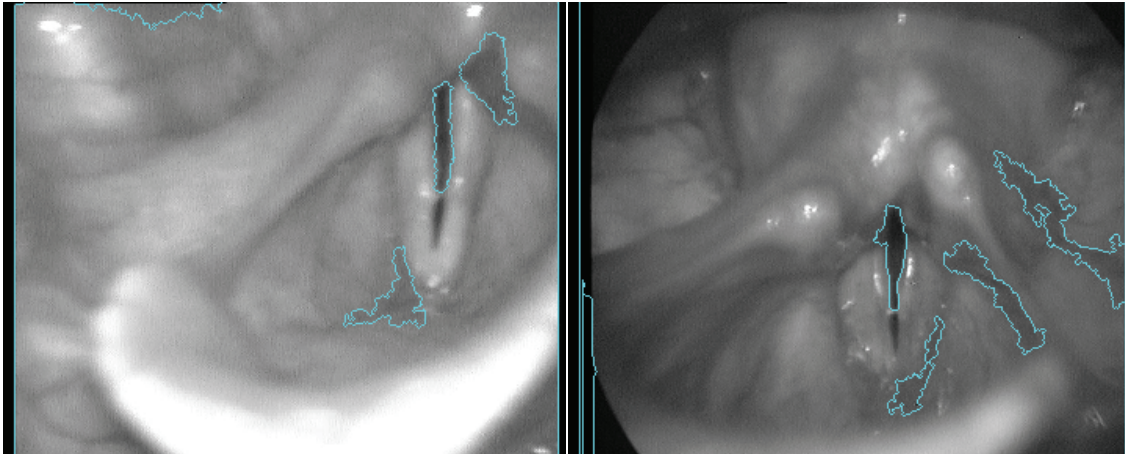


Figura 52. Figura de mérito de Pratt con un valor de 0.96 (a) sistema snake (b) sistema experto

5.3.1. Sistema snake vs Sistema watershed.

El sistema watershed [7] fue evaluado usando los mismos criterios que el sistema snake. De las 110 imágenes sólo una presentó valores entre 0.5 – 0.6. La Figura 53 muestra dos imágenes con los errores más representativos de la segmentación.



(a)

(b)

Figura 53. Segmentación con el sistema watershed (a) Pratt=0.57 (b) Pratt=0.64

Los valores obtenidos se resumen en la Figura 54. Se observa que hay 46 imágenes en el rango comprendido entre los 0.8 - 0.9 y 48 en el rango de 0.9 – 1 (un 85% de las imágenes por encima de 0.8). El coeficiente de Pratt con el valor más alto se logró en la Figura 55 con un valor de 0.95. Para realizar la comparación entre el sistema watershed y el sistema snake se decidió no utilizar el algoritmo de Pratt. Esto se debe a que ambos procedimientos presentan errores y una comparación de sus contornos no brindaría información real. Por lo que se decidió únicamente cotejar los resultados obtenidos cuando se evaluaron ambos sistemas con el experto. Entre los valores a destacar tenemos: el promedio de los coeficientes de Pratt obtenido a partir de la base de datos con las 110 imágenes, el máximo coeficiente, el mínimo coeficiente, número de veces que un sistema supera al otro y el número de coeficientes obtenidos entre 0.9 – 1 (Tabla. I.).

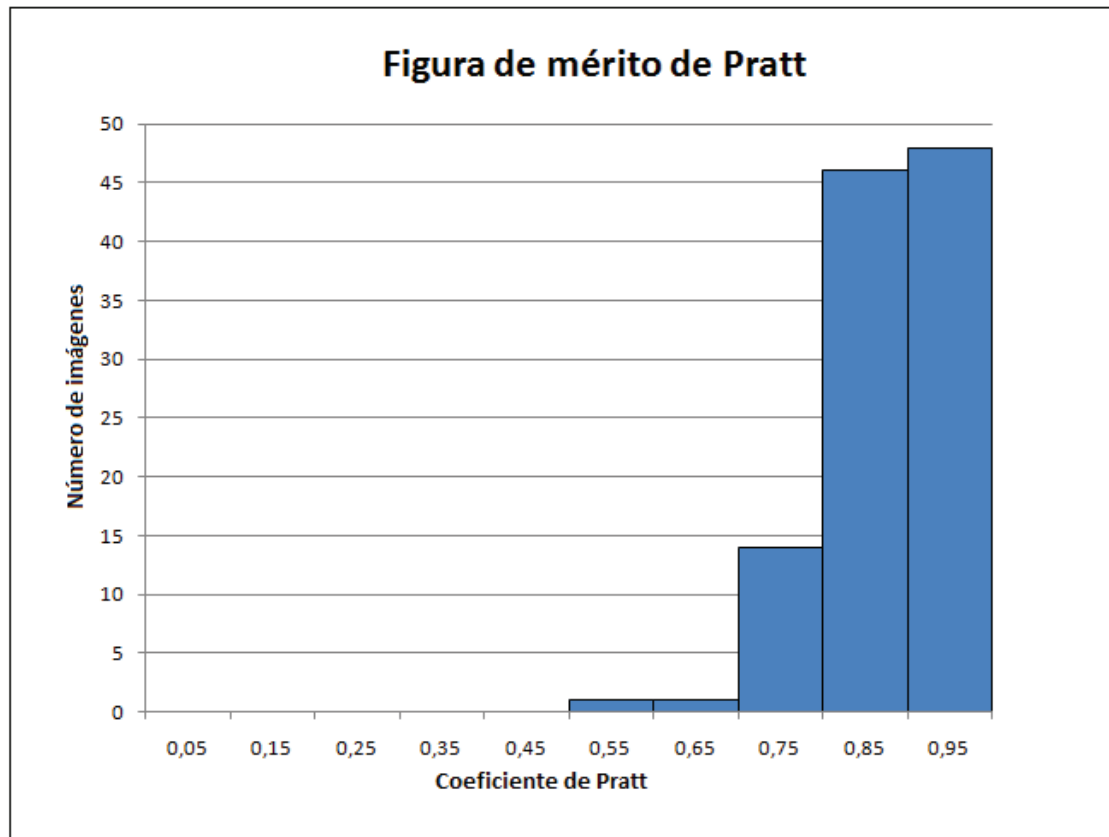
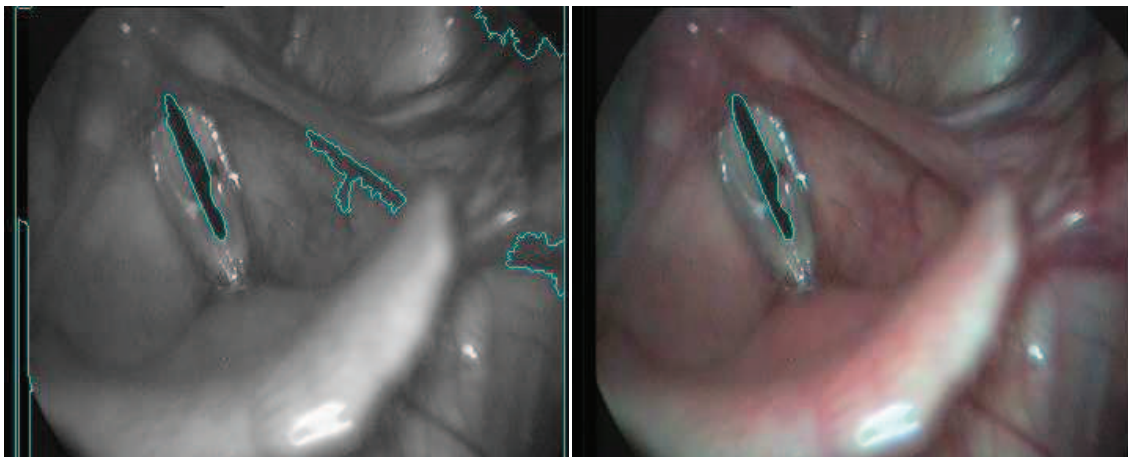


Figura 54. Figura de mérito de Pratt sistema watershed – sistema experto



(a)

(b)

Figura 55. Figura de mérito de Pratt con un valor de 0.95 (a) sistema watershed
(b) sistema experto

	Sistema Watershed	Sistema Snake
Promedio de los 110 coeficientes de Pratt	0,87	0,83
Coficiente máximo de Pratt	0,95	0,96
Coficiente mínimo de Pratt	0,57	0,25
Número de veces que un sistema supera al otro	59	51
Número de coeficientes entre 0,9 - 1	48	42
Número de coeficientes entre 0,9 - 0,8	46	45
Número de coeficientes entre 0,8 - 0,7	14	9
Número de coeficientes entre 0,7- 0,6	1	1
Número de coeficientes entre 0,6 - 0,5	1	2
Número de coeficientes menores a 0,5	0	11

Tabla. I. Comparación entre el sistema watershed y el sistema snake.

6 Conclusiones y Futuros Trabajos

6.1. Conclusiones.

El presente trabajo abordó el problema de la segmentación automática de la glotis en imágenes laríngeas obtenidas a partir de vídeos estroboscópicos. Entre las dificultades en este tipo de segmentaciones podemos mencionar: variación de iluminación entre fotogramas provocados por la luz estroboscópica, movimientos de la cámara y movimientos de las propias estructuras laríngeas.

Los objetivos propuestos han sido cumplidos cabalmente, ya que se ha podido desarrollar un sistema totalmente automático de segmentación con snakes, a diferencia de los métodos mencionados en el estado de arte que no logran una inicialización del todo automática

Este trabajo de fin de máster ha propuesto una alternativa a los métodos ya existentes para la segmentación de la glotis en imágenes laríngeas, el cual ha brindado buenos resultados en la mayoría de imágenes probadas, a pesar de la pobre iluminación que se presenta en muchas de estas.

Se ha podido demostrar también que el uso del algoritmo de snake con fuerza externa GVF necesita de un robusto pre-procesado para converger en la glotis y que la mayoría de los errores fueron provocados por esta etapa.

Otro importante aporte a considerar es el procedimiento usado para la inicialización, que a pesar de ser muy básico resultó ser muy eficiente. Únicamente 2 imágenes de la base de datos presentaron problemas en la inicialización. Esto se debió a que en ambos casos parte de la snake se inicializó dentro de la glotis lo que produjo ciertos errores.

Las pruebas realizadas arrojaron resultados alentadores cuando se lo cotejó con el método basado en la transformada de watershed [7]. De los resultados obtenidos se puede establecer que el basado en watershed realiza las mejores segmentaciones; a pesar de esto, los valores logrados con las snakes se encuentran cercanos a los obtenidos por las watershed, lo cual ya es un gran paso si se tiene en mente que este último es más robusto y más elaborado. La robustez del método de watershed se refleja en su tolerancia a la variabilidad de iluminación, y su mayor elaboración proviene del hecho de usar técnicas sofisticadas como JND que simulan el comportamiento del ojo humano para discriminar niveles de grises, sin olvidar también

la inclusión de un LDA capaz de distinguir la glotis de otros objetos generados por la segmentación.

De los índices mostrados en la Tabla. I, el más representativo es el primero, que representa el promedio de los 110 coeficientes de Pratt correspondientes a las 110 imágenes de la base de datos. En este índice se obtuvo un valor de 0.83, únicamente 0.04 por debajo del obtenido por las watershed. En base a estos resultados se puede concluir que a pesar de no ser el mejor sistema para la segmentación automática de la glotis, es un buen primer paso para la formulación de un nuevo algoritmo basado en snakes.

La causa por la cual 14 imágenes presentaron un coeficiente de Pratt menor a 0.5 se debe a problemas en el pre-procesado que ocasionan que la glotis pierda detalles y se confunda con el fondo. Esto a su vez complica la labor de la snake que únicamente logra segmentar la parte de la glotis que no se vio afectada. Para resolver este tipo de inconvenientes se puede cambiar los parámetros del pre-procesado, sin embargo al realizarse esto no se garantiza que se consiga una mejora significativa. La mejor solución sería proponer algún tipo de variante de fuerza externa en la cual se discrimine entre las regiones delimitadas por el área de la snake durante su movimiento.

Aunque el algoritmo de watershed tuvo mejores resultados, las snakes están mejor preparadas para el seguimiento, lo cual sería provechoso si se implementa en videos estroboscópicos, donde únicamente se usaría para la inicialización el resultado del fotograma anterior.

6.2. Futuros trabajos.

En este apartado se sugieren algunas líneas de investigación que podrían ser de interés para profundizar en este campo.

- Buscar nuevas alternativas para la inicialización de los contornos activos, que sean más robustas frente a pequeñas partículas de ruido o una inicialización realizada dentro de la región glótica buscada.
- Explorar nuevas técnicas basadas en contornos activos como la descrita en [40], la cual usa un enfoque basado en regiones locales para guiar a la snake.

- Estudiar el uso de las snakes para realizar seguimientos en vídeos en tiempo real.
- Estudiar otros tipos de fuerzas externas como *Vector Field Convolution* (VFC) o *Enhanced Gradient Vector Flow* (EGVF) [29].
- Estudiar algún método más sofisticado para determinar si el objeto identificado como la glotis es tal, basados en momentos de formas, LDA o similar [7].
- Estudiar el comportamiento de estos algoritmos con mayor número de imágenes en diferentes condiciones y con imágenes de alta velocidad, que son presuntamente más fáciles de segmentar.
- A partir del conocimiento adquirido se pretende extender la complejidad del problema para ser usado en segmentaciones en tres dimensiones.
- Por último se pretende desarrollar una metodología basada en snake que sea capaz de adaptarse a distintos problemas presentados en la segmentación biomédica.

7

Referencias

-
- [1] Jackson – Menaldi, M.C., La Voz Patológica. Ed. Panamericana 2002.
- [2] Baken, R. J. y Orlikoff, R. F., Clinical measurement of speech and voice, 2 ed., Singular, 2000.
- [3] Osma Ruiz, V.J., Contribución al procesado digital de imágenes para la caracterización de patologías laríngeas, Tesis doctoral, Universidad Politécnica de Madrid, enero 2010
- [4] Manfredi, C., Bocchi, L., Bianchi, S., Migali, N., Cantarella, G. Objective vocal fold vibration assessment from videokymography images. *Biomedical Signal Processing and Control* 2006; 1(2):129-36.
- [5] Friedl, S., Wittenberg, T. Automatic segmentation of vocal folds using active shape models. In: Schade G, editor. *Proceedings of the 6th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research*. Hamburg, IRB Verlag, 2003.
- [6] Kass, M., Witkin, A., y Terzopoulos, D., "Snakes: Active contour models," *International Journal of Computer Vision*, pp.321-331, 1988.
- [7] Osma-Ruiz, V.J., Godino-Llorente, J.I., Saénz-Lechón, N., y Fraile, R., "Segmentation of the glottal space from laryngeal images using the watershed transform," *Computerized Medical Imaging and Graphics*, vol. 32, no 3, pp. 193-201, 2008.
- [8] Wittenberg, T., Tigges, M., Mergell, P., y Eysholdt, U., "Functional imaging of vocal fold vibration: digital multislice high-speed kymography," *Journal of Voice*, vol. 14, no. 3, pp. 422-442, 2000.
- [9] Lohscheller, J., Eysholdt, U., Toy, H., y Dollinger, M., "Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2D-diagrams for visualizing and analyzing the underlying laryngeal dynamics," *IEEE Transactions on Medical Imaging*, vol. 27, no. 3, pp. 300-309, 2008.
- [10] Švec, J.G., Schutte, H.K., Videokymography: High-speed line scanning of vocal fold vibration, *Journal of Voice*, Volume 10, Issue 2, 1996, pp. 201-205, ISSN 0892-1997, 10.1016/S0892-1997(96)80047-6.
- [11] Mehta, D., Deliyski, D., Quatieri, T., Hillman, R., 2011. Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings. *J. Speech Lang. Hear Res* 54, 47–54.
- [12] Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., Dollinger, M., 2007. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med. Image Anal.* 11, 400–413.
- [13] Wittenberg, T., Moser, M., Tigges, M., Eysholdt U, "Recording, processing, and analysis of digital highspeed sequences in glottography," *Mach. Vis. Appl.*, vol.8 (6), pp. 399-404, 1995.
- [14] Yan, Y., Chen, X., Bless, D., "Automatic tracing of vocal-fold motion from high-speed digital images," *IEEE Transactions on Biomedical Engineering*, vol.53, no.7, pp.1394-1400, July 2006 doi: 10.1109/TBME.2006.873751.
- [15] Demeyer, J., Dubuisson, T., Gosselin B., Remacle, M., 2009. Glottis segmentation with a high-speed glottography: a fully automatic method. In: 3rd Adv. Voice Funct. Assess. Int. Workshop.
- [16] Marendic, B., Galatsanos, N., Bless, D., 2001. A new active contour algorithm for tracking vibrating vocal folds. In: *Proc. Int. Conf. ImageProc.*, pp. 397–400.

- [17] Russ, J., *The Image Processing Handbook*, 4th ed, North Carolina State University, Raleigh, USA: July 26, 2002 CRC Press
- [18] Palm, C., Lehmann, T. M., Bredno, J., Neuschaefer-Rube, C., Klajman, S., and Spitzer, K., "Automated analysis of stroboscopic image sequences by vibration profiles," in *Proceedings of the 5th International Workshop on Advances in Quantitative Laryngology, Voice and Speech Research*, Groningen, Netherlands, Apr.2001.
- [19] Moukalled, H., Deliyski, D., Schwarz, R., Wang, S., 2009. Segmentation of laryngeal High-Speed Videoendoscopy in temporal domain using paired active contours. *MAVEBA 1*, 137–140.
- [20] Manfredi, C., Bocchi, L., Cantarella, G., Peretti, G., Videokymographic image processing: Objective parameters and user-friendly interface, *Biomedical Signal Processing and Control*, Volume 7, Issue 2, March 2012, pp. 192-201, ISSN 1746-8094, 10.1016/j.bspc.2011.02.007.
- [21] Bleau, A., Leon, L.J., Watershed-based segmentation and region merging, *Computer Vision and Image Understanding*, 77 (3) (2000), pp. 317–370.
- [22] Shen, D.F., Huang, M.T., A watershed-based image segmentation using JND property, *Proceedings of IEEE ICASSP 2003*, 3 (2003), pp. 377–380
- [23] Duda, R.O., Hart P.E., Stork D.G., *Pattern classification*, (2nd ed.)Wiley-Interscience. (2001)
- [24] Unde, A.S.; Premprakash, V.A.; Sankaran, P.; , "A novel edge detection approach on active contour for tumor segmentation," *Engineering and Systems (SCES), 2012 Students Conference on* , vol., no., pp.1-6, 16-18 March 2012 doi: 10.1109/SCES.2012.6199100
- [25] De Luis-Garcia, R.; Martin-Fernandez, M.; Arribas, J.I.; Alberola-Lopez, C., "A fully automatic algorithm for contour detection of bones in hand radiographs using active contours," *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on* , vol.3, no., pp. III- 421-4 vol.2, 14-17 Sept. 2003 doi: 10.1109/ICIP.2003.1247271
- [26] Keshani, M.; Azimifar, Z.; Boostani, R.; Shakibafar, A., "Lung nodule segmentation using active contour modeling," *Machine Vision and Image Processing (MVIP), 2010 6th Iranian*, pp.1-6, 27-28 Oct. 2010 doi: 10.1109/IranianMVIP.2010.5941138.
- [27] Ray, N.; Acton, S.T.; "Tracking fast-rolling leukocytes in vivo with active contours," *Proceedings of the 2002 International Conference on Image Processing*, vol. 3, pp. 165-168, 2002 doi: 10.1109/ICIP.2002.1038931.
- [28] Mikic, I.; Krucinski, S.; Thomas, J.D.; "Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates," *IEEE Transactions on Medical Imaging*, vol.17, no.2, pp. 274-284, April 1998 doi: 10.1109/42.700739.
- [29] Scott T. Acton y Nilanjay. Ray, *Biomedical Image Analysis: Segmentation*. Morgan & Claypool Publishers series, 2006.
- [30] Troutman, J.L, *Variational calculus with elementary convexity*, New York: Springer-Verlag, 1983.
- [31] R.A. Serway, J.W. Jewett, *Physics for scientists and engineers with Modern physics*, 8 ed, 2010
- [32] Cohen, L.D, On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211–218, Mar. 1991.

-
- [33] Xu, C. y Prince, J.L., "Snakes, Shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, pp. 359, 1998.
- [34] Li, C., Liu J., y Fox, M.D., "Segmentation of external force field for automatic initialization and splitting of snakes," *Pattern Recognition*, vol. 38, pp. 1947, 2005.
- [35] Bing, Li; Acton, S.T.; , "Automatic Active Model Initialization via Poisson Inverse Gradient," *Image Processing, IEEE Transactions on* , vol.17, no.8, pp.1406-1420, Aug. 2008 doi: 10.1109/TIP.2008.925375.
- [36] Perona, P. y Malik, J., "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629-639, 1990.
- [37] Gutiérrez-Arriola J.M., Osma-Ruiz V., Godino-Llorente J.I., Sáenz-Lechón N., Fraile R., Arias-Londoño J.D, "Preprocesado Avanzado de Imágenes Laríngeas para Mejorar la Segmentación del Área Glotal", *Workshop de Tecnologías Multibiométricas para la Identificación de Personas*, Gran Canaria, 5-6 de julio de 2010.
- [38] González, R.C., Wintz, P. (1996), *Procesamiento digital de imágenes*. Addison-Wesley.
- [39] Pratt, W.K., "Digital signal processing, 4 ed.", Wiley-Interscience, 2007, pp. 514-6.
- [40] Lankton, S.; Tannenbaum, A.; , "Localizing Region-Based Active Contours," *Image Processing, IEEE Transactions on* , vol.17, no.11, pp.2029-2039, Nov. 2008 doi: 10.1109/TIP.2008.2004611