Preprocesado Avanzado de Imágenes Laríngeas para Mejorar la Segmentación del Área Glotal

J.M. Gutiérrez-Arriola [a], V. Osma-Ruiz [a], J.I. Godino-Llorente [a], N. Sáenz-Lechón [a], R. Fraile [a], J.D. Arias-Londoño [b]

[a] *E.U.I.T. Telecomunicación-Universidad Politécnica de Madrid, Cra Valencia Km 7, 28031, Madrid.*

[b] *Grupo de Control y Procesamiento Digital de Señales. Universidad Nacional de Colombia. Manizales*

Corresponding autor: V. Osma-Ruiz (vosma@ics.upm.es)

Resumen — **El presente trabajo describe un método avanzado de preprocesado de imagen para mejorar la detección automática del espacio glotal en imagines laríngeas. El sistema puede aplicarse a imágenes obtenidas a partir de exploraciones de alta velocidad o a partir de exploraciones estroboscópicas (baja velocidad), aunque es en estas últimas donde se observan las mayores ventajas, al tratarse de grabaciones de inferior calidad. Con esta nueva técnica de preprocesado se logran resolver ciertos fallos de segmentación producidos por un sistema previo basado en transformada "Watershed" y "Merging". En resumen, se consiguen arreglar o mejorar el 38% de los errores de delineado de la glotis que aparecían en 29 imágenes de un total de 111 segmentadas.**

Index Terms **— Segmentación, preprocesado, difusión anisotró-pica, glotis.**

I. Introducción

Las patologías que pueden afectar a la producción de la voz son muchas y muy variadas. No obstante, su efecto común suele ser la generación de diferentes grados de dificultad para producir una vibración correcta de las cuerdas vocales durante la fonación, asociada a un defecto de cierre que agrava la situación. El análisis de estos dos efectos, particularmente el de vibración de los pliegues, resulta fundamental para el profesional de ORL (Otorrino-laringología) a la hora de diagnosticar disfunciones del sistema fonador. El principal problema con el que se enfrentan los especialistas al llevar a cabo esta labor es la elevada velocidad con que se produce el movimiento de las cuerdas vocales, para que el ojo humano sea capaz de visualizar el proceso con una mínima precisión. Para resolver este inconveniente se han ido desarrollando, a lo largo del último siglo, distintos métodos que de una u otra manera permiten captar el movimiento: técnicas subjetivas como la estroboscopia [1;2] y las grabaciones de alta velocidad [3;4]; u objetivas como la quimografía [5], los diagramas de área glotal [6], y los fonovibrogramas [7]. Estas últimas están cobrando cada vez una mayor importancia ya que facultan al especialista para cuantificar el movimiento [8], además de visualizarlo.

Todas las técnicas objetivas citadas anteriormente adolecen de la necesidad de un procesado de imagen orientado a la segmentación del espacio glotal, bien como parte de su desarrollo, bien para solucionar diversos artificios introducidos durante la exploración, como los movimientos sufridos por el dispositivo de grabación y/o el paciente.

El presente trabajo describe un método avanzado de preprocesado de imagen que permite mejorar la segmentación automática de la glotis obtenida por un sistema desarrollado anteriormente [9] y que combina varias técnicas relevantes en el campo del procesado digital de imagen.

La organización del artículo es como sigue: en el apartado II se describen las herramientas usadas para lograr una detección correcta del espacio glotal (el sistema ya diseñado y el preprocesado mediante difusión anisotrópica); en el apartado III se analizan los resultados obtenidos tras combinar los dos métodos anteriores; y en el apartado IV se destacan las principales conclusiones.

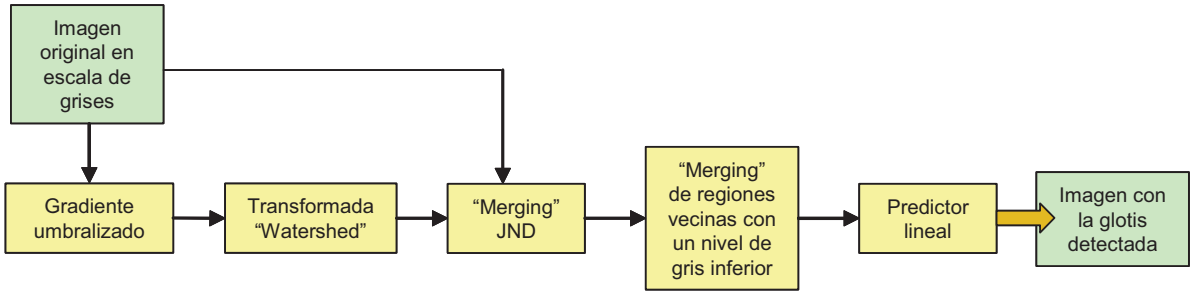

Fig. 1. Etapas del proceso seguido para la detección de la glotis en imágenes laríngeas.

 ISBN: 978-84-693-3389-1

II. MÉTODOS

A. Sistema Previo de Segmentación

El método descrito en [9] permite individualizar la glotis en imágenes laríngeas siguiendo el esquema que se presenta en la Fig. 1. El funcionamiento de cada uno de los bloques es el siguiente:

Transformada "Watershed" [10] del gradiente: el primer paso es convertir la imagen original (RGB) a escala de grises por medio de una transformación según el modelo YIQ. Después de dicha conversión se usa la luminancia "Y" para generar la imagen gradiente, que será umbralizada con nivel 2 para eliminar bordes insignificantes debidos al ruido (aquellos píxeles de la imagen gradiente con un nivel igual o inferior a 2 son colocados a 0). Sobre el gradiente umbralizado se aplica la transformada "Watershed" para conseguir una primera división de la imagen en regiones.

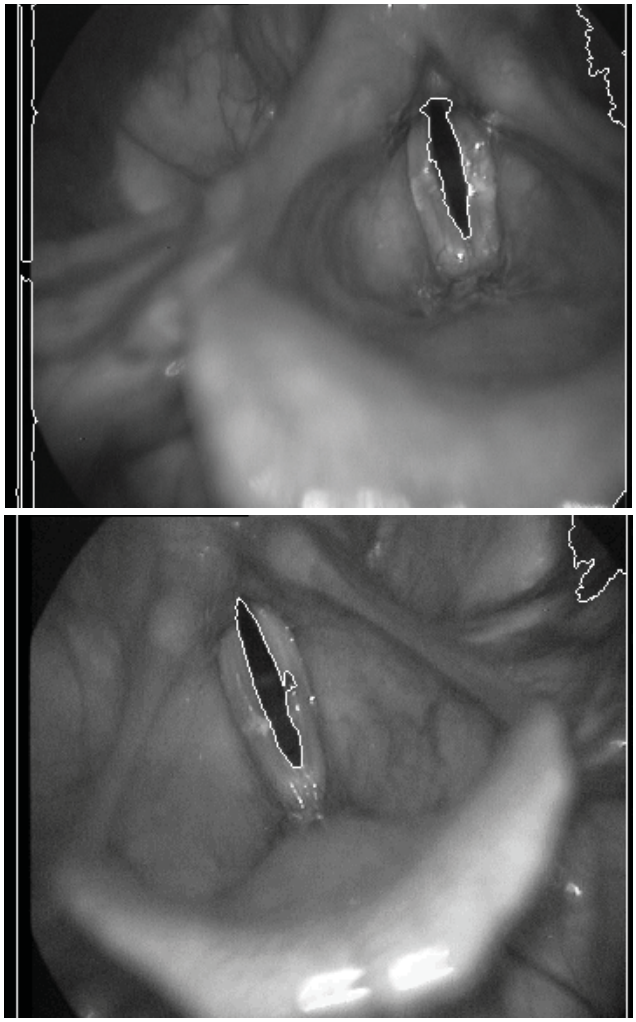

Fig. 2. Errores de segmentación introducidos en ocasiones por el sistema.

"Merging" basado en JND: el principal problema que presenta la transformada "Watershed" es que es muy sensible al ruido, lo que hace que la imagen quede dividida en miles de regiones cuando sólo se esperaban unas cuantas (una por objeto a delimitar). El preprocesado anterior alivia en parte el problema, pero no lo resuelve totalmente y es necesario aplicar un "merging" posterior que garantice la unión de regiones homogéneas. En este sentido, el sistema de la Fig. 1 permite, mediante este bloque, la unión de aquellas regiones que resultan iguales al ojo humano según el criterio de JND (Just Noticeable Difference) [11].

"Merging" de regiones con nivel de gris superior: el tercer paso del sistema consiste en otro proceso de "Merging". En esta ocasión se mantendrán aquellas regiones que no tienen vecinas más oscuras (nivel de gris inferior) y se obligará la fusión de las otras. El objetivo es ahora reducir el número de objetos segmentados uniendo aquellas regiones que no pueden corresponder con el área glotal (nótese que desde el punto de vista de un observador humano la glotis siempre debería ser un objeto oscuro rodeado de zonas más claras).

Discriminador: la última etapa consiste en un proceso de clasificación que permite diferenciar la glotis del resto de objetos presentes en la imagen. Para ello se usa un predictor lineal [12] entrenado en función de los 7 momentos invariantes (completos y binaros) de los distintos objetos.

El sistema presentado permite la detección de la glotis en un 75% de las imágenes analizadas, de manera totalmente automática, y mediante la variación de un único umbral en el 25% restante. No obstante, en 29 imágenes (de un total de 111) aparecen pequeños errores en la delineación del espacio glotal, como los que se presentan en la Fig. 2.

B. Preprocesado. Difusión anisotrópica

Un preprocesado más potente que el presentado en el sistema de la Fig 1. puede reducir el número de regiones resultantes de la división entregada por la transformada "Watershed", haciendo el proceso de "merging" posterior más sencillo y mejorando los resultados.

El preprocesado de la imagen se realizará pues mediante la combinación de dos métodos. En una primera fase se trata de conseguir el suavizado de la imagen original en escala de grises *I*, sin deteriorar los bordes más significativos de esta, mediante el escalado espacial con difusión anisotrópica propuesto por P. Perona [13]. El efecto deseado se logra gracias a la aplicación recursiva de la ecuación 1 en todos los píxeles de *I*. En la segunda etapa se seguirá empleando una umbralización de gradiente que elimine los pocos bordes insignificantes que resten después del filtrado anisotrópico.

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

En la ecuación 1:

1. i y j indican la posición del píxel dentro de la imagen (fila y columna, respectivamente)
2. t indica el nivel de escalado en que se encuentra el proceso (el número de iteración dentro de la recursividad). $I_{i,j}^{t}$ representa pues el estado en que queda la imagen I una vez realizada la iteración t sobre ella.

 ISBN: 978-84-693-3389-1

3. ∇ introduce la diferencia de intensidad del punto actual (*i,j*) con cada uno de los puntos de su vecindad, en una conectividad 4: norte-*N*, sur-*S*, este-*E* y oeste-*W*.
4. *c* es el denominado coeficiente de conducción, su valor se establece también de forma distinta para cada dirección de vecindad, en función de la diferencia de nivel de gris anterior. Cuando ∇ es grande (característica propia de los bordes) *c* será próximo a cero, mientras que para valores pequeños de ∇ *c* tenderá a la unidad. Bajo estas premisas, existen muchas formas posibles para la función que implementa *c* sobre ∇, ya que basta con que, partiendo de 1, sea descendente hacia cero. Una señal muy utilizada es una exponencial decreciente con ritmo de caída controlado por una constante prefijada *K*. La ecuación 2 recoge, como ejemplo, el cálculo del coeficiente de conducción con dirección norte para el píxel (*i,j*).

$$c_{N_{i,j}}^{t} = e^{-\left(\frac{\left|I_{i-1,j}-I_{i,j}\right|}{K}\right)^2} \quad (2)$$

5. λ es un valor que regula la velocidad a la que se realiza el proceso y que no debería superar ¼ para asegurar la estabilidad numérica del sistema.

Para cada píxel (*i,j*) de la imagen se calculan las diferencias de nivel de gris con sus vecinos en las cuatro direcciones consideradas. Aquellas diferencias altas (representativas 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.

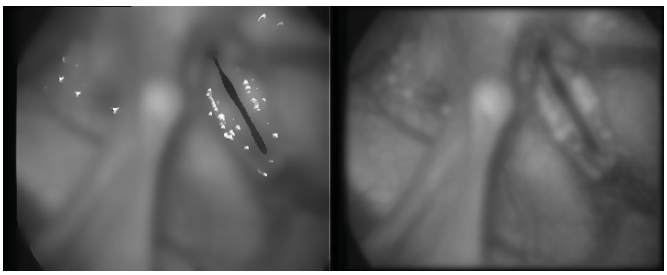

Fig. 3. Ejemplo de escalado con difusión anisotrópica. Izquierda: imagen obtenida por filtrado anisotrópico tras 50 iteraciones. Derecha: imagen filtrada paso bajo.

En la Fig. 3 se presentan como ejemplo dos imágenes de la laringe: la de la izquierda ha sido obtenida mediante un filtrado de difusión anisotrópica con *K*=10, λ=0,2 y 50 iteraciones; la de la derecha es el resultado de aplicar sobre la misma imagen original un típico filtro paso bajo. Es fácil observar como la difusión anisotrópica homogeniza los distintos tejidos de la laringe sin deteriorar los bordes más significativos, mientras que el filtro estándar origina una gran borrosidad en la imagen.

III. Resultados

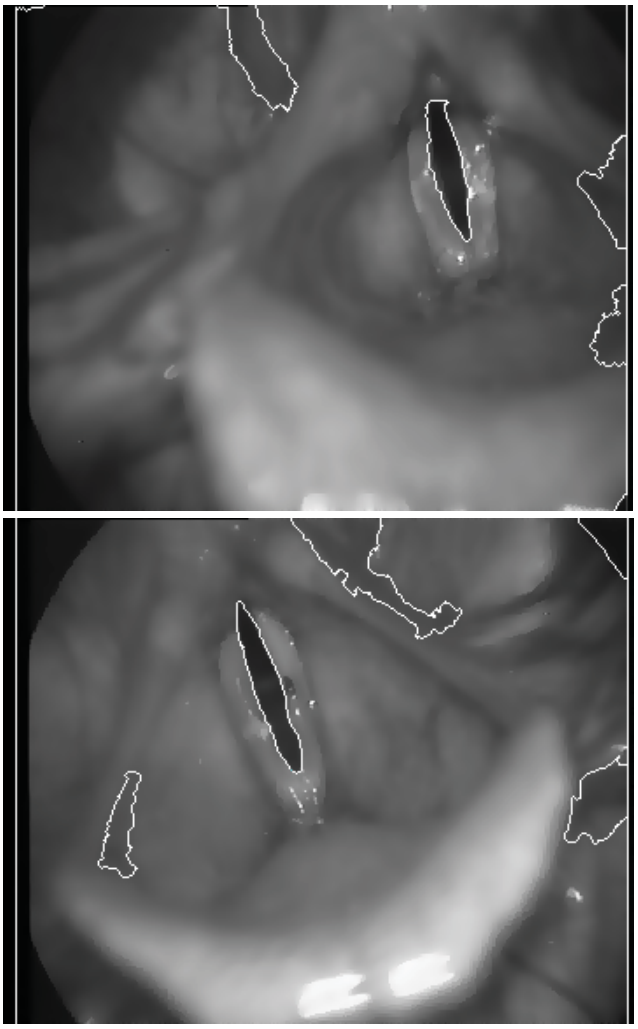

Fig. 4. Ejemplos de imágenes en las que se ha suprimido el error de segmentación. Compárense estas con las imágenes de la Fig. 2.

El proceso de investigación se llevó a cabo en dos fases:

1. Primero se analizaron los resultados que entregaba el sistema descrito en el apartado II-*A* añadiéndole una etapa de difusión anisotrópica como se indicó en el apartado II-*B*. Para ello se seleccionaron 12 imágenes escogidas al azar (6 con error y 6 correctas) y se variaron los parámetros característicos de la difusión. Concretamente se estudiaron 140 variaciones del método, realizando combinaciones de *K* entre 3 y 18 (con razón de 5), λ entre 0,05 y 0,25 (con razón de 0,05) y número de iteraciones entre 5 y 65 (con razón de 10). De todas ellas se escogieron las 4 que entregaban mejores resultados. Considérese que es tan importante analizar las posibles mejoras sobre imágenes con error, como

 ISBN: 978-84-693-3389-1

asegurarse de que las segmentadas correctamente no se empeoran.

2. Con los 4 casos anteriores se volvió a realizar la segmentación, pero esta vez sobre las 111 imágenes disponibles. Los mejores resultados se obtuvieron con los parámetros K=8, λ=0,05 y 55 iteraciones. Con esta configuración se consiguió suprimir el error en 14 de las 29 imágenes citadas y mejorarlo en 4. Sin embargo, como inconveniente, se originó un error, que no existía, en 7 imágenes de las 82 restantes. El balance, en cualquier caso, es positivo pudiéndose considerar una mejora neta de 11 imágenes, lo que supone un porcentaje del 38% de errores de delineado resueltos.

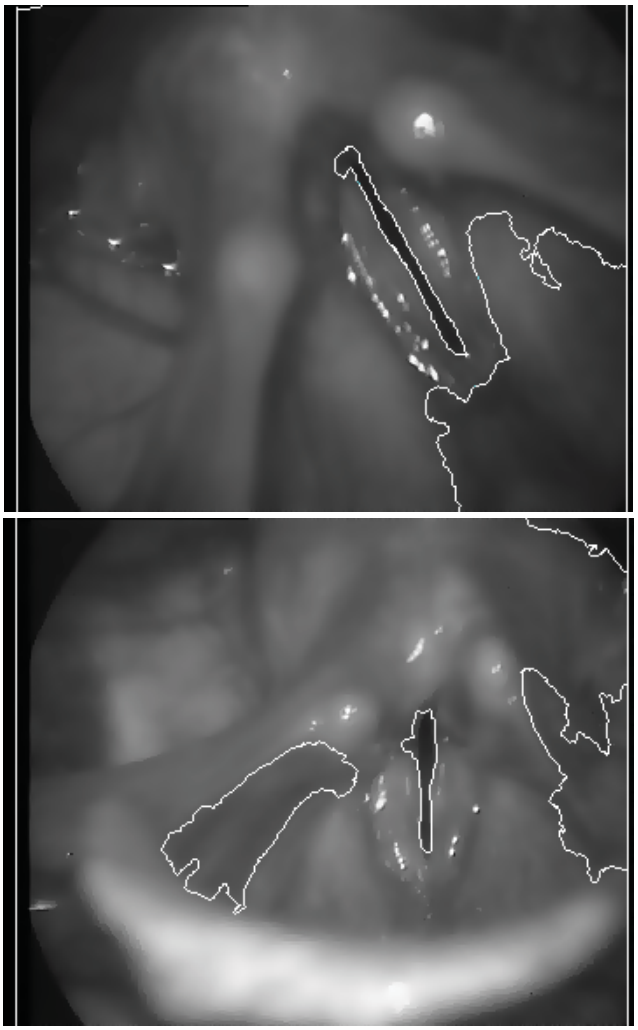

Fig. 5. Ejemplos de imágenes para las que la difusión anisotrópica introduce un error que no existía.

En la **¡Error! No se encuentra el origen de la referencia.**. 4 se recogen los resultados entregados por el nuevo método para las 2 imágenes con error que se mostraron en la Fig. 2.

La Fig. 5 muestra sin embargo otras dos imágenes en las que se genera un error, inexistente cuando son segmentadas únicamente con el método descrito en el apartado II-A, sin preprocesado adicional.

IV. Conclusiones

La difusión anisotrópica es un método de preprocesado de imágenes que permite aumentar la homogeneidad de zonas con niveles de gris similar, mientras que mantiene intactos, e incluso acentúa, los bordes que separan zonas con cambios bruscos.

Se ha demostrado que el filtrado anisotrópico es un buen método para mejorar el funcionamiento de un sistema diseñado previamente para la detección de la glotis en imágenes laríngeas [9]. El objetivo es corregir ciertos errores que aparecen en el delineado de la glotis.

Con un filtro de difusión anisotrópica de parámetros K=8, λ=0,05 y 55 iteraciones, se procesaron 111 imágenes laringoscópicas, 29 con errores de delineación en la glotis. Los resultados tras la segmentación mostraban el problema resuelto en 14 de las 29 imágenes y significativamente mejorado en 4. En 7 imágenes de las 82 restantes apareció un error que no existía. Teniendo en cuenta estos dos aspectos podría decirse que el método de preprocesado presentado en este artículo obtiene buenos resultados al lograr una mejora neta de un 38% sobre las imágenes que presentaban algún error de segmentación.

V. Agradecimientos

Este trabajo ha sido financiado por el proyecto TEC2009-14123-C04.

Referencias

[1] Oertel,M.J., "Über eine neue 'laryngostroboskopische' untersuchungsmethode des kehlkopfes," *Zentralbl.f.d. Mediz. Wissenschaften Heft*, vol. 16, pp. 81-82, 1878.

[2] Rosen, C. A., "Stroboscopy as a research instrument: development of a perceptual evaluation tool," *Laryngoscope*, vol. 115, no. 3, pp. 423-428, 2005.

[3] Schwarz, R., Hoppe, U., Schuster, M., Wurzbacher, T., Eysholdt, U., and Lohscheller, J., "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1099-1108, 2006.

[4] Zhang, Y., Bieging, E., Tsui, H., and Jiang, J. J., "Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging," *Journal of Voice*, 2009, In press.

[5] Wittenberg, T., Tigges, M., Mergell, P., and Eysholdt, U., "Functional imaging of vocal fold vibration: digital multislice high-speed kymography," *Journal of Voice*, vol. 14, no. 3, pp. 422-442, 2000.

[6] Yan, Y., Ahmad, K., Kunduk, M., and Bless, D., "Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology," *Journal of Voice*, vol. 19, no. 2, pp. 161-175, 2005.

[7] Lohscheller, J., Eysholdt, U., Toy, H., and 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.

[8] Manfredi, C., Bocchi, L., Bianchi, S., Migali, N., and Cantarella, G., "Objetive vocal fold vibration assessment from videokymographic images," *Biomedical signal processing and control*, vol. 1, no. 2, pp. 129-136, 2006.

[9] Osma-Ruiz, V. J., Godino-Llorente, J. I., Sáenz-Lechón, N., and 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.

[10] Osma-Ruiz, V. J., Godino-Llorente, J. I., Sáenz-Lechón, N., and Gómez-Vilda, P., "An improved watershed algorithm based on efficient computation of shortest paths," *Pattern Recognition*, vol. 40, no. 3, pp. 1078-1090, 2007.

[11] Shen, D. F. and Huang, M. T., "A watershed-based image segmentation using JND property," in *Proceedings of IEEE ICASSP 2003*, vol. 3, pp. 377-380, Apr.2003.

[12] Duda, R. O., Hart, P. E., and Stork, D. G., *Pattern Classification*, 2 ed., Wiley-Interscience, 2001.

[13] Perona, P. and 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.

 ISBN: 978-84-693-3389-1