# SEGMENTATION OF LARYNGEAL HIGH-SPEED VIDEOENDOSCOPY IN TEMPORAL DOMAIN USING PAIRED ACTIVE CONTOURS

Habib J. Moukalled[1,2], Dimitar D. Deliyski[1,2], Raphael R. Schwarz[1,3], Song Wang[2]
[1] Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC, USA
[2] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA
[3] Siemens AG, Erlangen, Germany

***This paper introduces a method for segmention of the vocal-fold edges in temporal domain from laryngeal high-speed videoendoscopy (HSV). The method employs a pair of active contours (snakes), which deform within a series of kymographic images derived from the HSV data. By following a set of deformation rules, this pair of active contours converges to the desired boundaries of the glottis. The proposed method was tested on a dataset of 98 HSV samples, of which 96 were successfully segmented. The new method substantially outperforms existing methods. However, more precise analysis revealed that of the 96 successfully segmented HSV samples, 18 exhibited a fine error up to ±1 pixel, and 78 samples exhibited errors exceeding a pixel. The large majority of the gross errors (76%) were due to problems near the posterior and anterior commissures, which warrants further investigation for improving the accuracy and reliability of the method.***

***Keywords:*** **high-speed videoendoscopy, active contour segmentation, snakes, glottis, digital kymography**

## I. INTRODUCTION

Laryngeal high-speed videoendoscopy (HSV) contains unprecedented amount of information about the vibration of the vocal folds that is potentially clinically useful. However, navigating through the enormous amount of HSV data is difficult and impractical. In order for HSV to gain widespread clinical use, there is a need for image-processing algorithms for automatic extraction of the relevant vocal-fold vibratory features. That is the long-term purpose of this project.

This problem has been investigated in recent years. Yan *et al.* developed an algorithm to segment the glottis from HSV data by globally thresholding pixel intensities on a per-frame basis [1]. Lohscheller *et al.* developed an algorithm that takes advantage of HSV's 3D structure by performing a modified 3D seeded region growing for segmentation of the glottis and post-processing for reconstruction of the vocal-fold boundaries [2]. However, such local image thresholding or region growing algorithms are usually sensitive to image homogeneity and noise.

Active contours, or snakes, are deformable models that can dynamically converge towards the desired image features [3]. The deformation of a snake follows certain specified rules on the whole contour, which may make it more robust to image noise. A closed-loop snake has been used to analyze the PE-segment within HSV data [4]. A pair of open-curve snakes has been applied to the right and left vocal folds to segment the glottis from videokymography [5]. A HSV movie can be represented in temporal domain as a digital kymography (DKG) playback [6]. Therefore, an attractive approach for HSV-segmentation can be achieved by segmenting the glottis from all spatial-temporal kymographic images of HSV.

In this study, we employed a pair of open-curve snakes (right and left) on DKG images to segment the glottis, for which deformation rules enforce the temporal resolution of HSV. Fig. 1 illustrates two open-curve snakes, right and left, attracted to pixels with large gradient magnitude (aligned with the glottal boundaries), which is derived from DKG. In Fig. 1 (not drawn to scale) the white squares are the vertices, termed snaxels, which make up the right and left snakes. The white lines connecting the snaxels are spline segments. And the space between the vertical white lines denotes time.


***Fig. 1:*** Snakes are attracted to the pixels with large magnitude of gradient within a kymographic image.

The proposed method exhibits the following merits over previous methods: (a) the snake convergence is facilitated due to the absence of complex geometries in kymographic images; (b) the deformation of the snakes can be optimized by using time-delayed discrete dynamic programming; (c) the temporal resolution of HSV helps constrain snake deformation since DKG images exhibit continuity along the time axis; (d) the method is robust to the disappearing glottis during the closing phase; (e) the initialization procedure is simple and scalable; and (f) the method segments the right and left vocal-fold edges concurrently, while maintaining separate left and right segmentation results.

## II. METHOD

### *A. Snake -energy Minimization.*

***Energy Minimizing Splines.*** A snake is a spline deformed in the spatial domain of a digital image in order

to minimize an energy functional comprised of internal forces derived from the snake's shape, and external forces derived from image features [3]. A snake is parameterized by the vector $v(s) = [x(s),y(s)]$, where $s \in [0,1]$, and seeks to minimize the following energy functional [3,7]:

$$E = \int_0^1 E_{\text{int}}(v(s)) + E_{image}(v(s))ds \qquad (1)$$

The internal force $E_{\text{int}}$ acting on snake $v(s)$ is a soft constraint used to make the snake's shape smooth and is given by:

$$E_{\text{int}}(v(s)) = \frac{1}{2}[\alpha(s) \mid v'(s) \mid^2 + \beta(s) \mid v''(s) \mid^2] \qquad (2)$$

Where $v'(s)$ and $v''(s)$ are the first and second derivatives, respectively; $\alpha$ and $\beta$ are two weights used to adjust the snakes elasticity and rigidity, respectively, which in turn influences the snake's shape. The image forces $E_{image}$ acting on the snake $v(s)$, is a force that counter-balances the internal force $E_{\text{int}}$, and makes the snake align with desirable image features. For example $E_{image}$ can be:

$$E_{image}(v(s)) = -\mid \nabla I(x,y) \mid^2, \qquad (3)$$

where $\nabla I$ is the image gradient. By combining Eqs. (1) and (2) we obtain:

$$E = \int_0^1 \frac{1}{2}[\alpha(s)|v'(s)|^2 + \beta(s)|v''(s)|^2] + E_{image}(v(s))ds. \qquad (4)$$

Using the calculus of variations, Eq. (4) has a numerical solution that can be obtained in $O(n)$ time [3]. By using specialized external force fields the convergence of snakes using the variational calculus framework can be significantly improved.

***Snake Deformation Rules**:* In order to enhance convergence of the paired temporal snakes, three snake-deformation rules are applied: (1) no closed loops are permitted in the right and left snakes (i.e. snaxels of right and left snakes are defined by the time axis of kymographic images and can only move up or down within a kymographic slice during deformation); (2) in the absence of glottal-edge information right and left snakes are attracted to each other (i.e. regions in the kymograms where the vocal folds are in contact); and (3) right and left snakes are not allowed to pass each other in the deformation.

***Time-delayed Discrete Dynamic Programming.*** The variational calculus framework for snake-energy minimization uses higher-order derivatives in order to approximate an energy minimizing spline from discrete data. Hard constraints are typically non-differentiable; as a consequence numerical instability occurs. In order to overcome the instability of variational approaches, snake energy is minimized using discrete dynamic programming [7].

The discretization of the internal energy term of a snake given in Eq. (2), yields:

$$E_{\text{int}}(v_i) = \frac{1}{2}[\alpha_i \mid v_i - v_{i-1} \mid^2 + \beta_i \mid v_{i+1} - 2v_i + v_{i+1} \mid^2], \qquad (5)$$

where $v_i$ corresponds to the $i^{th}$ snaxel. By discretizing Eq. (4) we obtain:

$$E_{total} = \sum_{i=1}^{n} \left(E_{\text{int}}(v_i) + E_{image}(v_i)\right), \qquad (6)$$

which can be viewed as a discrete multistage decision-making process, or better yet, a dynamic-programming problem [7].

Before dynamic programming can be applied, we must make the observation of a correspondence between minimizing the total energy measure of a snake and the problem of minimizing a function of the form [7]:

$$\begin{aligned} &E_{total}(v_1, v_2, ..., v_n) = \\ &E_1(v_1, v_2, v_3) + E_2(v_2, v_3, v_4) + ... + E_{n-2}(v_{n-2}, v_{n-1}, v_n), \end{aligned} \qquad (7)$$

where each $v$ is a state variable that can take $m$ possible values. In the general case,

$$E_{i-1}(v_{i-1}, v_i, v_{i+1}) = E_{\text{int}}(v_{i-1}, v_i, v_{i+1}) + E_{image}(v_i). \qquad (8)$$

Now, the dynamic programming solution involves generating a sequence of functions of one variable, $\{S_i\}_{i=1}^{n-1}$ (the optimal value function), where for obtaining each $S_i$ a minimization is performed over a single variable. For example, given the energy function shown in Eq. (8), with $n = 4$, we have:

$$\begin{aligned} &S_1(v_2, v_1) = \min_{v_0}[S_0(v_1, v_0) + E_1(v_1, v_2, v_3)], \\ &S_2(v_3, v_2) = \min_{v_1}[S_1(v_2, v_1) + E_2(v_2, v_3, v_4)], \\ &S_3(v_4, v_3) = \min_{v_2}[S_2(v_3, v_2) + E_3(v_3, v_4, v_5)], \\ &\min_{v_0,...,v_4} E(v_0, v_1, v_2, v_3, v_4) = \min_{v_3}[S_3(v_4, v_3) + E_4(v_4, v_5, v_6)]. \end{aligned} \qquad (9)$$

And in the general case [7],

$$\begin{aligned} &S_i(v_{i+1}, v_i) = \\ &\min_{v_{i-1}} [S_{i-1}(v_i, v_{i-1}) + E_i(v_i, v_{i+1}, v_{i+2})]. \end{aligned} \qquad (10)$$

The discrete dynamic-programming solution for snake-energy minimization has a $O(nm^2)$ memory requirement and $O(nm^3)$ theoretical complexity, where $n$ is the total number of stages (number of snaxels) and $m$ is the total number of decisions at a given stage (neighborhood size).

Fig. 2 gives insight to the dynamic programming solution for snake-energy minimization as a pair of temporal snakes deform within a glottal opening in a DKG image. The gray tiles of Fig. 2 represent the magnitude of the gradient, black tiles correspond to the snaxels of the right snake, black tiles with a dot correspond to snaxels from the left snake, and black tiles with a square correspond to snaxels where the right and left snake are overlapping. During the snake-deformation procedure snaxel movement is limited to a column-wise neighborhood, which prohibits the occurrence of closed loops (self intersections) in the right and left snakes and significantly reduces the search space needed for snake-energy minimization.

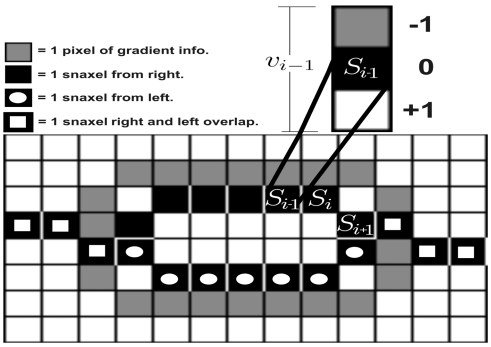


***Fig. 2*:** Snake energy is minimized by finding the optimal state variable $v_{i-1}$ along the direction of the column.

*B. Experimental Design.*

***DKG Snake Toolbox.*** In order to test the new method, a custom software, DKG Snake Toolbox, was built. It allows a user to scroll through the HSV and DKG frames, which are dynamically linked. After the user manually marks the anterior and posterior commissures, an initial DKG image is selected at the 50% anterior-posterior level. Then, a snake-initialization line is placed in the middle of the glottis, spanning through the time axis of the kymogram. The right and left snakes are deformed in order to segment the glottis. After the result is verified, the remaining DKG images are automatically segmented.

***Preprocessing the HSVs.*** The laryngeal tissues being observed are covered with a superficial layer, the lamina propria, which is highly reflective due to hydration and/or mucus presence. In general, light reflections represent a significant problem for snakes, because they introduce spurious noise into the gradient maps governing the snake deformations. Through the duration of a HSV recording, glottal openings exhibit distinctly dark intensities. Pixels having intensity values higher than the median intensity value of the entire recording more than likely correspond to light reflections, which can be easily suppressed.

Once light reflections have been suppressed, specialized gradient maps are computed. The gradient in the spatial domain is calculated for every frame of the HSV. Since the snaxels of the right and left snakes are restricted to column-wise neighborhoods of movement, calculation of the gradient is performed using only the rows of a given frame in order to accent the horizontal edge information. A custom gradient map with gradients normal to the right vocal-fold edge is computed for the right snake, and a gradient map with gradients normal to the left vocal-fold edge is computed for the left snake.

***Contour Embedding.*** In order to keep the right and left snakes attracted to each other in regions of the kymogram where the vocal folds are in contact, a new parameter, snake intensity (*snakeInt*), is devised. After each iteration of the dynamic programming, the right and left snakes are embedded in the opposing snake's edge map as a salient edge with intensity values between 0 and 255. This effectively bounds the right snake between the right vocal-fold edge and the left snake, and the left snake is bounded by the right vocal-fold edge and the right snake. This can prevent the right and left snakes from moving across one another during deformation.

***Human Data.*** Fourteen vocally-normal speakers (7 men and 7 women between 22 and 29 years of age) have been recorded using with a Phantom V.7.1 (Vision Research, Inc., Wayne, NJ) monochromatic camera (16,000 fps, 320x320 pixels, 12-bit depth) connected to a 70° rigid laryngeal endoscope and a 300-W xenon light source. Each speaker produced the vowel /i/ in seven phonatory conditions, varying in register, pitch and loudness. Thousand-frame tokens of sustained phonation have been extracted from each recording to yield a total of 98 HSV samples.

## III. Results and Discussion

***Snake Parameter Adjustment.*** In early works on snakes, the parameters $\alpha$ and $\beta$ were shown to be sensitive parameters used to weight the snake model's continuity and rigidity, respectively. In the time-delayed discrete dynamic programming algorithm, $\alpha$ and $\beta$ are not as sensitive as their classical counterparts [7]. For all results obtained in this paper, we have set $\alpha = 10$ and $\beta = 3$. The only parameters that have been adjusted were the *snakeInt* and the column-wise neighborhood size (*colSize*). *colSize* and *snakeInt* are adjusted twice per recording, once in order to initialize the right and left snakes, and one additional time for the automated segmentation stage.

Figs. 3 and 4 show the values of *colSize* and *snakeInt* used for the initialization and segmentation stages for female and male subjects, respectively. Fig. 5 provides an example of (A) the initial positions of the right and left snakes in the toolbox, (B) the deformation results for the initial kymogram, and (C) phases of the opening cycle with the deformation results (the white contours along the

glottis) presented in the spatial domain of the HSV for a female subject.

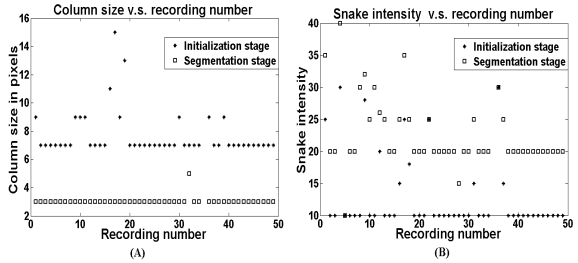


***Fig. 3:*** *colSize* and *snakeInt* for 49 female subjects.

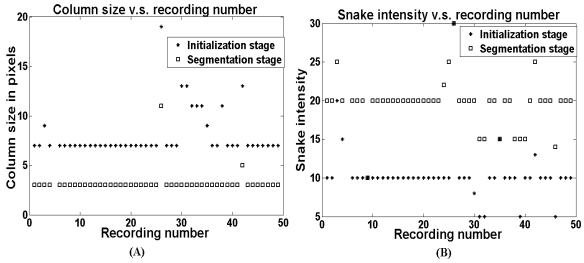


***Fig. 4:*** *colSize* and *snakeInt* for 49 male subjects.

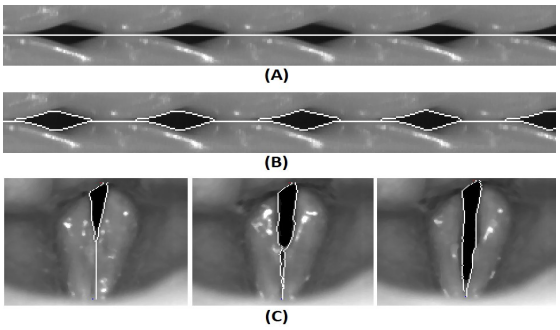


***Fig. 5:*** (A) Initialization of the right and left snakes, (B) deformation results of right and left snakes for the initial kymogram, (C) three phases of the opening cycle with right and left snakes presented in spatial domain.

***Validity and Reliability.*** From the 98 samples in the dataset, 96 samples were successfully segmented using the temporal paired snakes, and 2 presented difficulties due to poor lighting. That is an overall reliability of 98%, which is a highly-encouraging result. In all HSV samples, most DKG images were analyzed without gross errors, i.e. divergence of the snake from the correct edge by more than one pixel, usually due to attraction to the wrong nearby edge. Of the 96 successfully-segmented samples, 78 exhibited at least one DKG image with at least one gross error, 59 of which (76%) were due to a failure of the right and left snakes to attract to each other near the commissures, mainly the posterior commissure. Those instances can be easily corrected by introducing an adaptively sized column-wise neighborhood and appropriate pre-processing when automating the method.

***Accuracy.*** In all HSV samples, most DKG images exhibited sub-pixel accuracy of segmentation. Of the 18 samples free of gross errors, 1 had no single DKG image with a snake differing from the target edge, and 17 exhibited at least one DKG image containing an instance of an error up to ±1 pixel.

## IV. Conclusion

The proposed paired temporal snake algorithm exploits the HSV temporal resolution for obtaining a segmentation of the glottis by following a set of snake-deformation rules. The snake deformation strategy employs a dynamic programming algorithm, in which the optimization of the snake-energy function decreases monotonically with respect to the asymptotic rate of growth of the algorithm, and thus the global convergence is guaranteed. The development of the algorithm is still in progress, to be extended to a fully-automatic method for segmentation of the glottis from HSV. This algorithm is reliable and fast, yet highly scalable in terms of the degrees of parallelism that can be exploited in the future.

## Acknowledgments

This project is funded by NIH R01 grant DC007640: “Efficacy of Laryngeal High-Speed Videoendoscopy”.

## References

[1] Yan Y, Chen X, Bless D. Automatic tracing of vocal-fold motion from high-speed digital images. *IEEE Transactions on Biomedical Engineering*, 53(7):1394-1400, 2006.

[2] Lohscheller J, Toy H, Rosanowaski F, Eysholdt U, Dollinger M. Clinically evaluated procedure for reconstruction of vocal-fold vibrations from endoscopic digital high-speed videos. *Medical Image Analysis*, 11(1):400-413, 2007.

[3] Kass M, Witkin A, Terzopoulos D. Active contour models. *International Journal of Computer Vision*, 1(1):321-331, 1987.

[4] Lohscheller J, Dollinger M, Schuster M, Schwarz R, Eysholdt U, Hoppe U. Quantitative investigation of the vibration pattern of the substitute voice generator. *IEEE Transactions on Biomedical Engineering*, 51(8):1394-1400, 2004.

[5] Manfredi C, Bocchi L, Bianchi S, Migali N, Cantarella G. Objective vocal fold vibration assessment from videokymographic images. *Biomedical Signal Processing and Control*, 1:129–136, 2006.

[6] Deliyski D, Petrushev P, Bonilha H, Gerlach T, Martin-Harris B, Hillman R. Clinical Implementation of Laryngeal High-Speed Videoendoscopy: Challenges and Evolution. *Folia Phoniatrica et Logopaedica*, 60(1):33-44, 2008.

[7] Amini A, Weymouth T, Jain R. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):855–867, 1990.