Automatic contour tracking in ultrasound images

Min Li†, Chandra Kambhamettu†, Maureen Stone‡
†University of Delaware, ‡University of Maryland Dental School

(Received March 11, 2004; accepted September 1, 2004)

Contact Address
Min Li
Video/Image Modelling and Synthesis Lab
Department of Computer and Information Sciences
University of Delaware, Newark, DE 19716 USA
Fax: 302-831-8458 /Phone: 302-831-0556
E-mail: mli@cis.udel.edu

Abstract

In this paper, a new automatic contour tracking system, EdgeTrak, for the ultrasound image sequences of human tongue is presented. The images are produced by a Head and Transducer Support System (HATS). The noise and unrelated high-contrast edges in ultrasound images make it very difficult to automatically detect the correct tongue surfaces. In our tracking system, a novel active contour model is developed. Unlike the classical active contour models which only use gradient of the image as the image force, the proposed model incorporates the edge gradient and intensity information in local regions around each snake element. Different from other active contour models that use homogeneity of intensity in a region as the constraint and thus are only applied to closed contours, the proposed model applies local region information to open contours and can be used to track partial tongue surfaces in ultrasound images. The contour orientation is also taken into account so that any unnecessary edges in ultrasound images will be discarded. Dynamic programming is used as the optimisation method in our implementation. The proposed active contour model has been applied to human tongue tracking and its robustness and accuracy have been verified by quantitative comparison analysis to the tracking by speech scientists.

Keywords: Snake; Tracking; Tongue; Ultrasound images

1 Introduction

Ultrasound imaging is one of most attractive ways of acquiring image sequences of the tongue during speech. It does not expose the subject to radiation and can capture time-varying features in real-time. With the Head and Transducer Support System (HATS) (Stone and Davis, 1995), the head of the subject is fixed and the transducer is placed below the chin in a known position. In this way, accurate and reliable ultrasound images can thus be obtained during natural speech.

To reconstruct the tongue shapes from ultrasound images, automatic extraction and tracking of the tongue surface is necessary to avoid manual extraction which is time consuming. We developed a system, EdgeTrak, that can track the tongue surfaces through a sequence of two-dimensional ultrasound images. The user input is just several points along the tongue surface in a single frame. An approximated contour is obtained by B-spline interpolation. This contour is then attracted to the tongue surface by an automatic optimisation process. The optimised contour in a current frame can be used to approximate the tongue surface in the temporally immediate adjacent frame and the automatic optimisation process is applied in this adjacent image again. The steps are repeated through all images to produce tongue surfaces for a sequence of images.

In ultrasound images, there are always high-contrast edges unrelated to the structure of interest, and the tongue surface may be interrupted in several places (Unser and Stone, 1992). These noise characteristics make it difficult to automatically track tongue contours in ultrasound images. Our system uses snake (Kass, Witkin and Terzopoulos, 1988) as the tool for detecting the tongue surface. Snake is an active contour defined within an image that can move closer and closer to the edge while its associated energy is minimised. The energy terms of the snake are classified as internal and external energies. The internal energy is related to the contour shape and the minimisation goal for internal energy is to get smooth and continuous curves. This makes it possible to estimate the edge positions even in places where the surface is interrupted. The external energy is computed from the image data and it is the only term that attaches the active contour to the image. Cohen proposed the balloon model (Cohen, 1991; Cohen and Cohen, 1993),

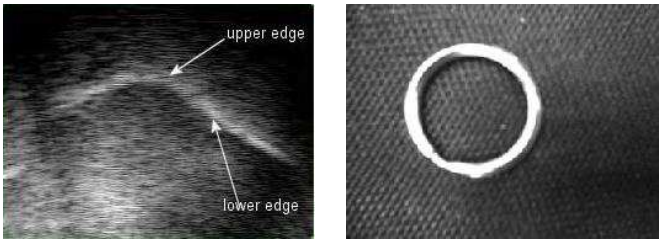
Gunn et al. introduced the dual active contour model (Gunn and Nixon, 1997) in order to prevent the active contour from stopping at local minima. Chalana et al. and Akgul et al. applied temporal smoothness (Chalana and Linker, 1996; Akgul, Kambhamettu and Stone, 2000) in addition to the spatial constraints in a single frame. Chan et al. introduced a region-based external energy (Chan and Vese, 2001) instead of the gradient of the edge for closed contour. Amini et al. developed dynamic programming (Amini, Weymouth and Jain, 1990) as the optimisation process for the snake model.

In the EdgeTrak system, temporal smoothness is not added to the internal energy component in order to give more flexibility to tracking during large tongue motion; however, this can be easily added. We use the contour in the previous frame as the initialisation of the current contour and use dynamic programming to optimise the location of the contour. To deal with the noise and unrelated edges in images, region information is applied to open contours and intensity in local regions is incorporated with edge gradient as the external energy. To the best of our knowledge, the active contour model in our system is the first model that applies region information to open contours.

Akgul et al. also presented a tongue contour tracking system (Akgul, Kambhamettu and Stone, 1999; Akgul et al. 2000). In their external energy definition, only gradient information is used; this would cause some tracking problems since the tongue surface cannot be distinguished from other high-contrast edges in the images. We solve this problem by introducing an intensity related constraint. See details in Section 2.

2 Novelty of the proposed active contour model

Among the different energies in a snake model, the external energy is usually related to gradient of the image. In reality, images are generally noisy and there are always high-contrast unrelated edges which make the gradient information insufficient to extract edges of interest. By constraining the homogeneity of intensity (the image brightness) in a region, the edge of a region in a noisy image



(a) (b)

Figure 1: (a) Example of ultrasound image of the tongue. (b) Example of closed contour.

can be successfully extracted (Chalana, Costa and Kim, 1995; Chan and Vese, 2001), but this constraint has some limitations:

First, it can only be applied to closed contours, and can not be used in some applications where open contours need to be tracked. One example of such applications is the human tongue tracking in ultrasound images. The ultrasound images are formed by propagating ultrasound waves through the subject's tongue so that part of its surface is obtained in the image (Stone and Davis, 1995). An ultrasound tongue image is shown in figure 1(a). The bright white band is the air reflection at the upper surface of the tongue. The lower edge of the band is the upper surface of the tongue, and the upper edge of the band has no physical interpretation. Thus, only lower edge is of interest to speech scientists though both edges have high gradient. It is hard to distinguish them by only using gradient information and there is no enclosed region where the constraint of homogeneity of intensity can be applied.

The second limitation of the constraint of intensity homogeneity can be seen from the example image in figure 1(b). In this image there is a key-chain ring which has the shape of a band. If the outer edge of the key-chain ring is of interest, the constraint of intensity homogeneity will fail since the region enclosed by the inner edge is more homogeneous than the region enclosed by the outer edge.

The proposed snake model in this paper combines the edge gradient and intensity in local regions. The local regions are not enclosed by the objective contour, they are in fact associated with each snake element. With the proposed snake model, the upper edge and lower edge of the

air reflection in the ultrasound images, or the inner edge and the outer edge of the key-chain ring, can be distinguished. The proposed snake model has been applied to human tongue tracking and its robustness and accuracy has been verified by the speech scientists via a quantitative analysis in this paper. The related software, EdgeTrak, is in use at several institutions for studying various aspects of tongue with applications in otolaryngology, linguistics, etc. (li, Kambhamettu and Stone, 2004).

3 The active contour model

The active contour model, or snake (Kass et.al, 1987), is an energy minimisation method to extract edges in images. The energy definition for snakes is:

$$E_{Total} = \alpha E_{int} + \beta E_{ext} \tag{1}$$

where E_{int} is the internal energy, E_{ext} is the external energy, α and β are the weighting parameters. E_{int} controls the contour shape and it is only related to the geometry property of the contour. E_{ext} attaches the contour to the image and defines the image features that are of interest.

Given a contour which is a set of points $[v_0, v_1, ..., v_{n-1}]$, the internal energy controls the smoothness and continuity of the contour and is defined as (Akgul and Kambhamettu 1999):

$$E_{int}(v_i) = \alpha_1(1 - \frac{\vec{v_{i-1}v_i}.\vec{v_iv_{i+1}}}{|\vec{v_{i-1}v_i}|\,.\,|\vec{v_iv_{i+1}}|}) + \beta_1||v_i - v_{i-1}| - d| \tag{2}$$

where v_i is the i^{th} snake element, α_1 and β_1 are the weighting parameters. d is the average length between two continuous snake elements.

The external energy is usually defined as the negative of the image gradient (Gunn et al., 1997; Akgul et al., 2000) and we use the normalised external energy as:

$$E_{ext}(v_i) = 1 - |\bigtriangledown I(v_i)|\,/M \tag{3}$$

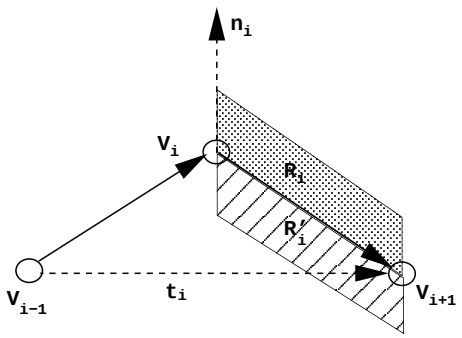


Figure 2: The definitions for t_i, n_i, R_i and $R_i^{'}$.

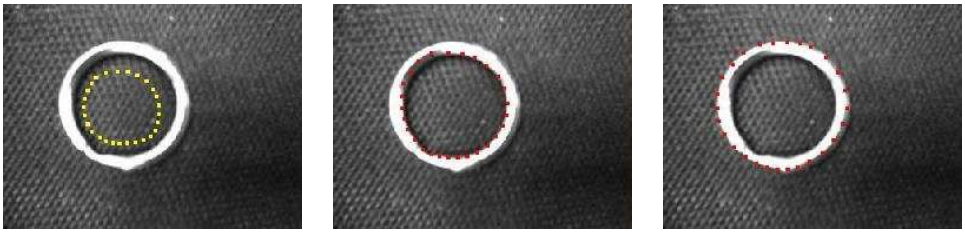
where M is the normalisation constant, I is the image intensity. At each pixel (x, y) of the image, the image gradient is defined as $\bigtriangledown I(x,y) = (\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y})$.

In reality, using only gradient information as the external energy is not enough due to the image noise and the high-contrast edges unrelated to the structure of interest. The constraint of homogeneity of intensity in a region is also not appropriate in case of open contours, or closed contours of band-shape objects. A region based band energy is presented below to solve these problems and it is also the main contribution of this paper.

In our active contour model, the contour is a set of snake elements $[v_0, v_1, ..., v_{n-1}]$ and the order of these elements are kept in the whole optimisation process. For snake element v_i, we define its tangent t_i as the direction of the line connecting its two neighbour elements:

$$t_i = \frac{v_{i+1} - v_{i-1}}{|v_{i+1} - v_{i-1}|}. \tag{4}$$

The normal vector n_i of element v_i can be obtained by rotating t_i 90 degrees in the counter-clockwise direction. Then we can define two regions R_i and $R_i^{'}$ for v_i. R_i is a quadrilateral with one edge connecting v_i and v_{i+1} while another edge is in the normal direction. $R_i^{'}$ is same as R_i except that it is in the opposite direction of the normal vector. For a band-shape object, R_i should be inside the band and $R_i^{'}$ should be outside the band, or vice versa. The difficulty in defining R_i and $R_i^{'}$ is that we cannot easily decide the edge length of the quadrilateral in the



(a) (b) (c)

Figure 3: Extraction of the outer edge of a key-chain ring. (a) Snake initialisation. (b) Edge extracted without band energy. (c) Edge extracted with band energy.

normal direction. This should depend on the application and the length should be smaller than the depth of the band. In our current system which is designed for tongue contour tracking from ultrasound images produced by HATS (Stone and Davis, 1995), there are 33 snake elements for each snake. We simply approximate the edge length as the average length between adjacent snake elements. The edge length is several pixels with this approximation. This definition guarantees that the edge length is smaller than the depth of the bright white band as shown in figure 1(a). The definitions for t_i, n_i, R_i and $R_i^{'}$ are shown in figure 2.

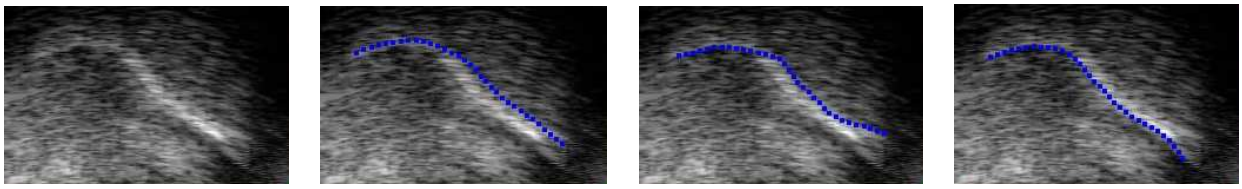
Suppose R_i is inside the band and the band-shape object of interest has a high intensity value than the background of the image, then the difference between the mean intensity of region R_i and the mean intensity of region $R_i^{'}$ should be large. The mean intensity difference between R_i and $R_i^{'}$ is:

$$dif(v_i) = \frac{1}{n \cdot N} \cdot \left(\sum_{p_j \in R_i} I(p_j) - \sum_{p_j^{'} \in R_i^{'}} I(p_j^{'}) \right) \tag{5}$$

where p_j is the pixel in region R_i, $p_j^{'}$ is the pixel in region $R_i^{'}$, n is the number of pixels in region R_i or $R_i^{'}$ and N is the intensity normalisation constant. In our application, N is 255.

The region based band energy is then defined as:

$$E_{band}(v_i) = \begin{cases} pen & dif(v_i) < 0 \\ 1 - dif(v_i) & otherwise \end{cases} \tag{6}$$



(a) (b) (c) (d)

Figure 4: Extraction of tongue contour. (a) Ultrasound tongue image. (b) Snake initialisation. (c) Edge extracted without band energy; some snake elements are attracted to the unrelated high-gradient upper air reflection edge. (c) Edge extracted with band energy.

where *pen* is a penalty constant applied to v_i when the mean intensity difference between R_i and $R_i^{'}$ is less than zero. In our application, we let $pen = 2$ and obtain good results for the tongue edge extraction. For snake elements which are located at the two ends of the snake and have no normal definitions, we use E_{band} of their neighbours to approximate their band energies.

Note that we are interested in finding the best location of each snake element v_i, where the intensity difference between R_i and R_i' is the maximum (in case $E_{band}(v_i)$ is the only constraint which needs to be minimised) among several possible positions of v_i (see section4). We do not care about the absolute value of this difference, which means the depth of R or R' is not critical. Since v_i is dynamically re-positioned, R_i and R_i' need to be re-calculated according to the orientation of v_i at each iteration of the optimisation process. For fast implementation, we weight the intensity values of all pixels inside R and R' uniformly.

Now we have both intensity and gradient information for a snake element and we define a new external energy:

$$E_{ext}^{'}(v_i) = E_{band}(v_i) \cdot E_{ext}(v_i). \tag{7}$$

$E_{ext}^{'}(v_i)$ uses both intensity and gradient instead of only using gradient. As we explained above, $E_{band}(v_i)$ is related to the intensity difference between two regions R_i and R_i' around each snake element v_i. One should see that R_i and R_i' are related to the orientation of the snake at v_i, thus $E_{band}(v_i)$ is the intensity difference measure along the normal direction of v_i while the gradient measure $E_{ext}(v_i)$ has nothing to do with the orientation of the snake. Most importantly,

the gradient is just for the snake element while the intensity information comes from neighbour regions around the snake element. This is very helpful in the tracking problem when the speckle noise is presented in the image since speckles are not favored by $E'_{ext}(v_i)$ where the intensity value is calculated over regions. Also the unrelated edges in the images such as the upper edge of the air reflection in the ultrasound image and the inner edge (or the outer edge if the orientation of the contour is reversed) of the key-chain ring will get a penalty from $E_{band}(v_i)$ and can not attract the active contour any more.

The performance of band energy is shown in figure 3 where the outer edge of the key-chain ring is the interest. Figure 3(a) is the initialisation of the snake. Without the band energy, the snake is attracted to the high-contrast inner edge as shown in figure 3(b). With the band energy and appropriate contour orientation definition(counter-clockwise), the outer edge of the key-chain ring is correctly extracted as shown in figure 3(c).

Band energy is important in order to correctly detect the human tongue surface in ultrasound images (see figure 4 for example). Without the band energy, some snake elements are attracted toward unrelated high-gradient edges (the upper edge of the air reflection) while with band energy, the tongue surface is correctly extracted.

The band energy definition depends on the normal direction of snake element. In the above key-chain ring example, one can reverse the contour orientation to extract the inner edge of the key-chain ring easily since region R_i and R'_i can be exchanged. In case the object of interest has lower intensity than the background of the image, the band energy can still work in the same way with appropriate contour orientation definition.

4 Optimisation process

In our tracking system, the optimisation method is based on dynamic programming (Amini et al., 1990). The contour of each frame is initialised by copying the contour from the previous frame. The normal of snake element v_i is recalculated in each optimisation step. From the definitions

of $E_{int}(v_i)$ and $E'_{ext}(v_i)$ in Equations (2) and (7) respectively, one can see that the energy of the snake element v_i only depends on two neighbours of this element and itself. The optimisation for one contour can be processed in multiple steps. Each step is decomposed into n independent stages. In stage i only the energy of v_i is minimised and the elements under consideration are only v_{i-1}, v_i and v_{i+1}. After n stages, energies of all snake elements are minimised and the energy of each element are summed up as the current E_{Total}. This process continues iteratively until the E_{Total} does not decrease any more. Compared with the exhaustive search method, the search cost is dropped from $O(l^n)$ to $O(n * l^3)$ with dynamic programming (n is the number of snake elements and l is the size of the search space respectively).

An efficient way to define the search space for the snake element v_i is to restrict the search along the normal direction of this point. In our application, search is in the normal direction and the position of each snake element is rearranged along the tangent direction of this point after every step of the optimisation process. The purpose of the rearrangement is to keep all snake elements evenly located along the contour while the current contour shape is kept unchanged. In EdgeTrak system, the size of search space is $l = 5$ by default. It has been found in practice that this search space works for tracking most tongue contours. In case of large tongue motion which means that the snake initialisation copied from the previous frame is far away from the true tongue surface, the search space needs to be increased by user.

$E'_{ext}(v_i)$ depends on regions R_i and R'_i. These two regions are decided by the normal of the snake element. In each step of the optimisation process the normal is calculated to decide the search direction and at the same time R_i and R'_i can be obtained according to the normal.

5 Experiment results

EdgeTrak has been used to track the tongue surfaces in ultrasound images. In EdgeTrak system, the user input is just several points along the tongue surface in the first frame. An approximated contour is obtained by B-spline interpolation. This contour is then attracted towards the tongue

	'yaya'	'golly'	'he sought'
expert 1 vs. expert 2	3.77	2.47	2.50
automatic vs. expert 1	2.64	1.83	2.39
automatic vs. expert 2	3.59	2.20	3.02

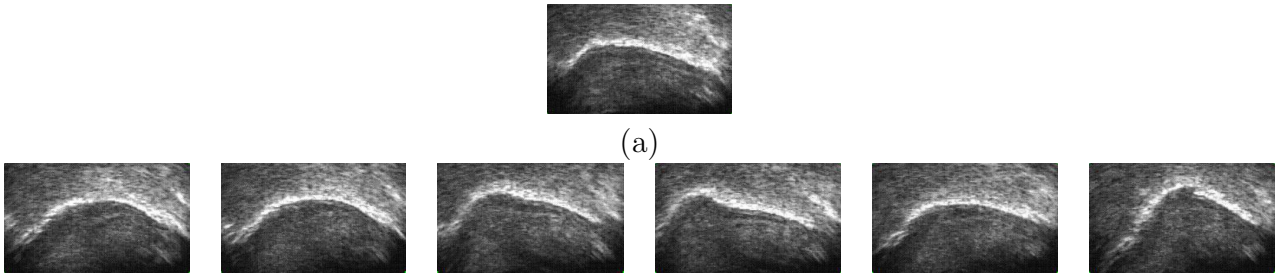
Table 1: Mean distance errors in pixels. 1 pixel=0.295mm.

surface by automatic dynamic programming optimisation process. Every frame in the sequence gets its snake initialisation from its previous frame and the snake is optimised in the same way as in the first frame. The tracking of an example sequence shown in figure 5 is performed and the results are shown in figure 6. Another example sequence is shown in figure 7 and its tracking result is shown in figure 8. The visual inspection of the tracked contours shows that our snake model works pretty well.

One more sequence is shown in figure 9. In this sequence, more speckle noise is present. Tracking result by EdgeTrak is shown in figure 10. The tongue surface, which is the lower edge of the air reflection is successfully tracked. Unrelated high-contrast edges, e.g. the upper edge of the air reflection is discarded due to the introduced band energy in EdgeTrak. Without the band energy, it is difficult to distinguish the upper edge of the air reflection from the tongue surface, as shown in figure 4.

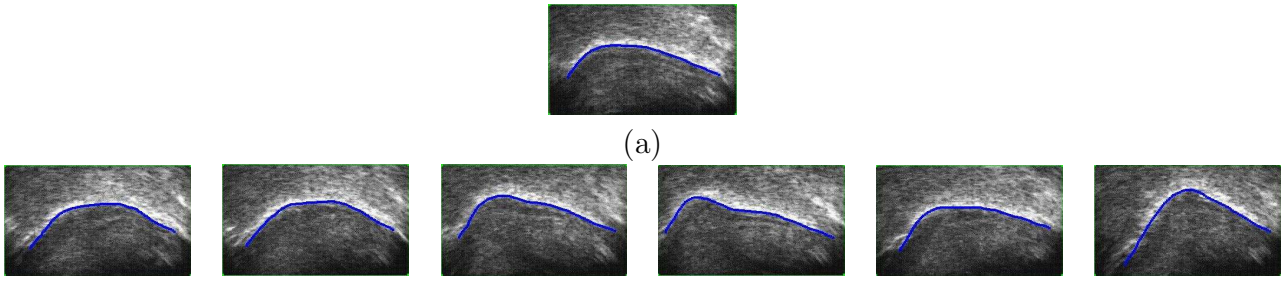
Note that we only use B-spline interpolation to get the snake initialisation from the user input in the first frame. In the snake optimisation process, the contour smoothness is controlled by the internal energy defined in Equation (2). Alternative approach is where snake is modeled via B-spline and the internal energy is not required since the smoothness of the snake is encoded in the spline formulation (Cipolla and Blake, 1990). In our current tracking system, we use explicit snake smoothness definition so that the user can control the contour smoothness interactively.

In order to verify the tracking results quantitatively, we compare the difference between the automatic tracking results and the manual contours drawn by the speech scientists, and compare the difference between the manual contours drawn by different speech scientists. The difference between any two contours was calculated using a Mean Sum of Distances (MSD) by measuring



(a) (b) (c) (d) (e) (f) (g)

Figure 5: Image sequence of example 1. Every 10th frame from 67 frames is shown. Image (a) is the first frame.



(a) (b) (c) (d) (e) (f) (g)

Figure 6: Tracked contours for the sequence in figure 5. User input is only seven points in the first frame.

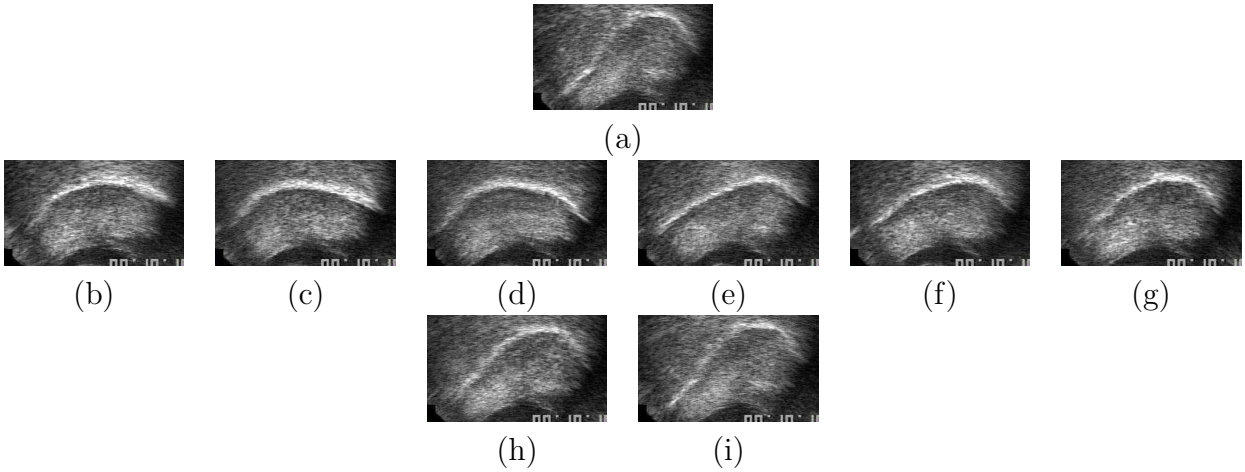


Figure 7: Image sequence of Example 2. Every 4th frame from 33 frames is shown. Image (a) is the first frame.

the distances between closest snake elements of each contour. The MSD between two contours $U = [u_1, u_2, ..., u_n]$ and $V = [v_1, v_2, ..., v_n]$ is defined as:

$$MSD(U, V) = \frac{1}{2n}(\sum_{i=1}^{n} \min_j |v_i - u_j| + \sum_{i=1}^{n} \min_j |u_i - v_j|). \quad (8)$$

Contours tracked by the automatic tracking system, EdgeTrak, and manual tracking by two speech scientists for three speech sequences were compared. The speeches for these three sequences are 'yaya', 'golly' and 'he sought' respectively. The comparison is listed in table 1. As the numbers indicate, the automatic contours are not isolated from the expert detected contours and the pixel errors between the automatic contours and manually drawn contours by scientists are quite low.

6 Conclusion

An automatic contour tracking system, EdgeTrak, for the ultrasound image sequences of human tongue is presented. This tracking system is based on a novel snake model. In this snake model, region information around each snake element is incorporated with the image gradient and the contour orientation is taken into account. Compared with the traditional snake model and other

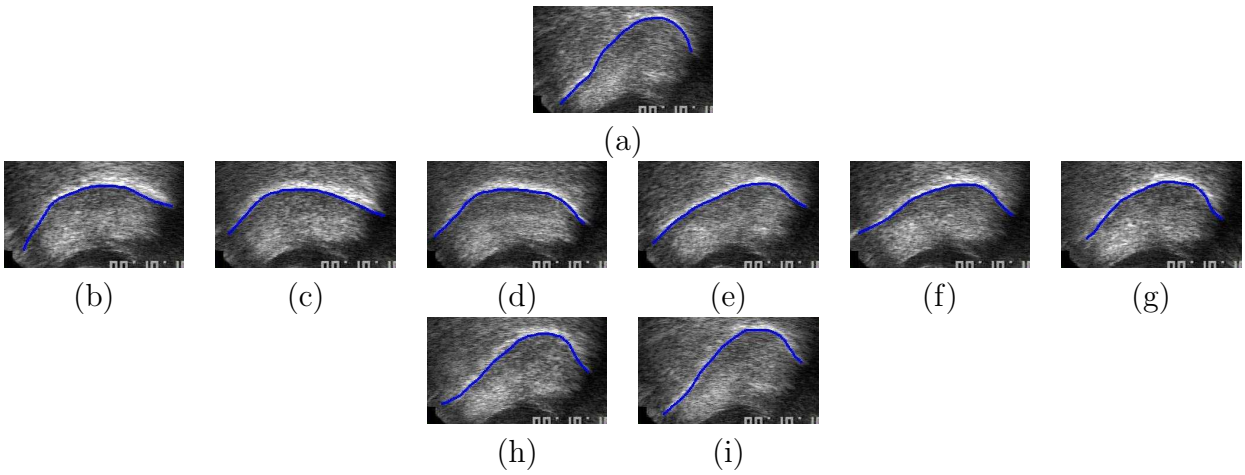


Figure 8: Tracked contours for the sequence in figure 7. The user input is only seven points in the first frame.

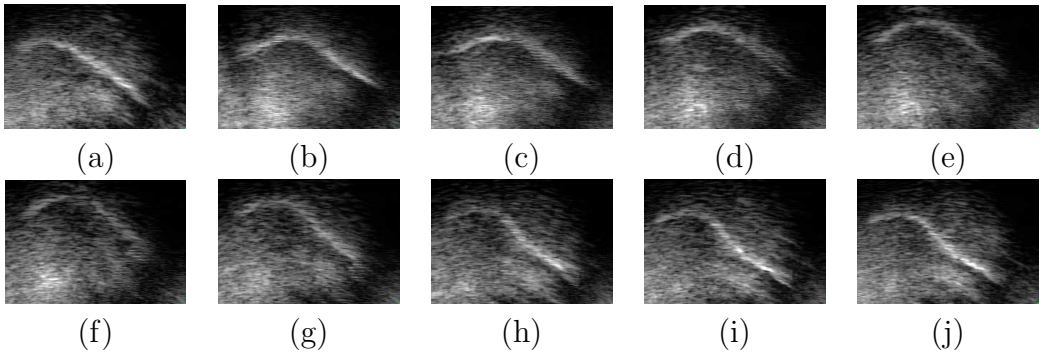


Figure 9: A difficult sequence with more noise.

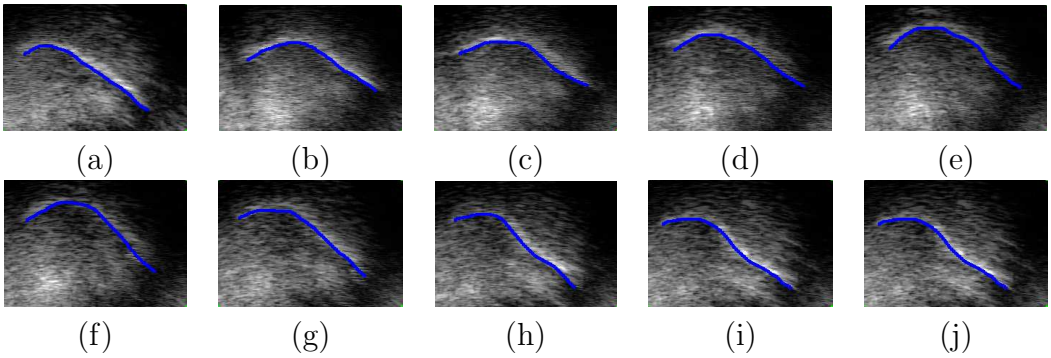


Figure 10: Tracked contours for the sequence in figure 9.

models which use homogeneity of intensity in a closed region as the image constraint, our snake model is robust to the speckle noise and can be applied to open contour tracking problems where region information is involved.

The robustness of the proposed model has been verified by comparing the automatic tracking results and the manual contours drawn by the speech scientists. EdgeTrak is currently being used by scientists at several institutions. Feedbacks from them indicate that the system is efficient and accurate for speech research and related applications.

Acknowledgment

This research was funded in part by NIDCD/NIH grant number R01 DC01758.

References

Akgul, Y.S., & Kambhamettu, C. (1999). A new multi-level framework for deformable contour optimization. *CVPR99* (pp. II:465-470).

Akgul, Y.S., Kambhamettu, C., & Stone, M. (1999). Automatic extraction and tracking of the tongue contours. *IEEE Transactions on Medical Imaging*, 18(10), 1035-1045.

Akgul, Y.S., Kambhamettu, C., & Stone, M. (2000). A task-specific contour tracker for ultrasound. *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis* (pp.135-142). Hilton Head Island, South Carolina.

Amini, A.A., Weymouth, T. E., & Jain, R. C. (1990). Using dynamic programming for solving variational problems in vision. *PAMI*, 12(10), 855-867.

Chalana, V., Costa, W. S., & Kim, Y. (1995). Integrating region growing and edge detection using regularization. *Proc. SPIE. Medical Imaging 1995: Image Processing, Murray H. Loew; Ed.*, 2434, 262-271.

Chalana, V., & Linker, D.T. (1996). A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Transactions on Medical Imaging*, 15(3), 290-298.

Chan, T.F., & Vese, L.A. (2001). Active contours without edges. *IP*, 10(2), 266-277.

Cipolla, R., & Blake, A. (1990). The dynamic analysis of apparent contours. *ICCV90* (pp. 616-623).

Cohen, L.D. (1991). On active contour models and balloons. *CVGIP*, 53(2), 211-218.

Cohen, L.D., & Cohen, I. (1993). Finite-element methods for active contour models and balloons for 2-D and 3-D images. *PAMI*, 15(11), 1131-1147.

Gunn, S.R., & Nixon, M.S. (1997). Robust snake implementation: a dual active contour. *PAMI*, 19(1), 63-68.

Kass, M., Witkin, A.P., & Terzopoulos, D. (1988). Snakes: active contour models. *IJCV*, 1(4), 321-331.

Li, M., Kambhamettu, C., & Stone, M. (2004). http://vims.cis.udel.edu/EdgeTrak.

Stone, M., & Davis, E.P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *The Journal of The Acoustical Society of America*, 6, 3107-3112.

Unser, M., & Stone, M. (1992). Automatic detection of the tongue surface in sequences of ultrasound images. *The Journal of The Acoustical Society of America*, 91(5), 3001-3007.