SNAKE FOR BAND EDGE EXTRACTION AND ITS APPLICATIONS

Min Li, Chandra Kambhamettu
Video/Image Modeling and Synthesis Lab
Department of Computer and Information Sciences
University of Delaware
Newark, DE,19716 USA
{mli,chandra}@cis.udel.edu

Maureen Stone
Vocal Tract Visualization Lab
Department of Oral and Craniofacial Biological Sciences
University of Maryland Dental School
Baltimore, MD, 21201 USA
mstone@umaryland.edu

ABSTRACT

A novel snake model suitable for edge extraction of band-shape objects is presented in this paper. Based on the proposed model, an edge tracking system, EdgeTrak, has been developed which is being used by speech scientists in speech research and other related applications. Unlike the classical active contour models which only use gradient of the image as the image force, the proposed snake model incorporates the edge gradient and intensity information in specific regions around each snake element. It can be used to extract edges that are open or closed contours, which makes it 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 snake model also takes into account the contour orientation so that any unrelated edges in the image will be discarded even if these edges have high gradient, or enclose a homogeneous region. Dynamic programming is used as the optimization method in our implementation and the image information update is naturally incorporated in the optimization process. Experiment results on face edge and human tongue tracking are also presented in this paper and the robustness and accuracy of the proposed model is verified by quantitative and qualitative analysis.

KEY WORDS

Image Processing, Snakes, Tracking, Edge Extraction.

1 Introduction

Snake [1], or active contour, has attracted a considerable amount of attention and is popularly used for automatic extraction and tracking of object edges. Snake is an energy minimization model whose energy terms are classified as internal and external. The internal energy is related to the contour shape and the minimization 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 usually is the negative of the image gradient and is the term that attaches the active contour to the image. Cohen [2] [3] proposed the balloon model, Gunn [4] introduced the dual active contour model to prevent the active contour from stopping at local minima. Wang [5] introduced the B-Spline represen-
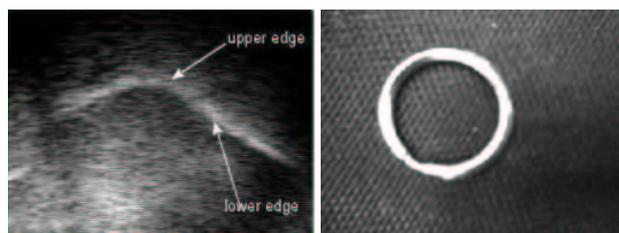
Figure 1. Left: An example of ultrasound images of the tongue. Right: An example of closed contour

tation of snake, which is a multistage active contour model. Chalana [6] and Akgul [7] applied temporal smoothness in addition to the spatial constraint in a single frame. Chan [8] introduced a region-based external energy instead of the gradient of the edge of a closed contour. Amini [9] developed dynamic programming as the optimization process for the snake model to guarantee the global optimization.

Although different energy types have been proposed in these active contour models, the external energy is usually related to the 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 in a region, the edge of a region in a noisy image can be successfully extracted [8], but this constraint has some limitations:

First, it can only be applied to closed contours. It can not be used in applications where open contours need to be tracked, such as tracking the surface of the human tongue in ultrasound images. The ultrasound images are formed by propagating ultrasound waves through a section of the subject's tongue, and the surface of the upper tongue part is obtained in the image [10]. An ultrasound tongue image is shown in the left of Figure 1. 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 is useless. 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 the right of Figure 1. 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 both edge gradient and intensity in specific regions. The specific regions are not enclosed by the object contour. They are in fact associated with each snake element, and are split into two parts: one part is inside the band and the other part is outside the band. Whether one part is inside or outside the band is defined by the orientation of the contour. By considering the intensity difference of these two regions, 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 track the human tongue from ultrasound images, and also the human face boundary from video images. A developed system, EdgeTrak, is being used in speech and swallowing research by speech scientists, and its robustness and accuracy is verified by quantitative and qualitative analysis in this paper.

2   The Active Contour Model

The active contour model, or snake [1], is an energy minimization method to extract edges in images. The energy definition for snakes is:

$$E_{Total} = \alpha E_{int} + \beta E_{ext} \qquad (1)$$

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.

The internal energy controls the smoothness and continuity of the contour and is defined as [11]:

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

where v_i is the ith 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 [4] [12] [11] and we use the normalized external energy as:

$$E_{ext}(v_i) = 1 - |\nabla I(v_i)|/M \qquad (3)$$

where M is the normalization constant.

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 interest. The constraint
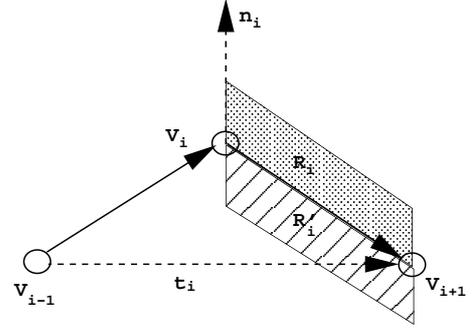
Figure 2. The definitions for t_i, n_i, R_i and R_i'.

of homogeneity of intensity in a region is also not appropriate in case of open contours or closed contours for a band-shape object. A region based band energy is presented below to solve these problems.

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 is kept throughout the optimization process. For snake element v_i, we define its tangent t_i as the direction of the line connecting its two neighbor elements:

$$t_i = \frac{v_{i+1} - v_{i-1}}{|v_{i+1} - v_{i-1}|}. \qquad (4)$$

The normal vector n_i of element v_i can be obtained by rotating t_i 90 degrees in the count-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. 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 can not easily decide the edge length of the quadrilateral in the normal direction. This should depend on the application and the length should approximately be the depth of the band. In our tracking system, we simply approximate this length as the average length between adjacent snake elements. 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:

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

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 normalization constant. In our application, N is 255;
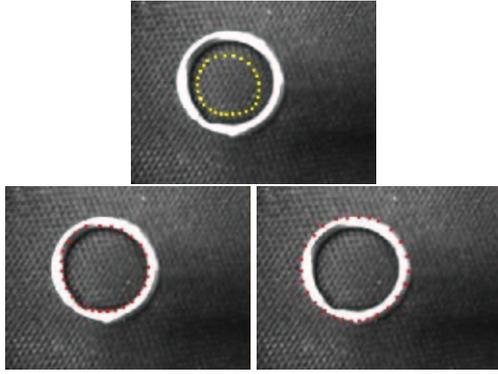
Figure 3. Extraction of the outer edge of a key-chain ring. Snake elements are shown with different colors for visualization purpose. Top: Snake initialization. Bottom left: edge extracted without band energy. Bottom Right: edge extracted with band energy

The region based band energy is then defined as:

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

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 get good results for the edge extraction.

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

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

$E_{ext}'(v_i)$ uses both intensity and gradient instead of only using gradient. Most importantly, the gradient is just for the snake element while the intensity information comes from neighbor 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 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 will 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. The top image is the initialization of the snake. Without the band energy, the snake is attracted to the high-contrast inner edge as shown in the bottom left image. With the band energy and appropriate contour orientation definition(counter-clockwise), the outer edge of the key-chain ring is correctly extracted in the bottom right image.

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

The band energy definition depends on the normal direction of the 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' are interchanged. 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.

3 Optimization Process

The original snakes [1] uses a variational approach as the optimization method. Variational approaches can not guarantee global optimality of the solution. Dynamic programming [9] ensures global optimality of the solution and the contour information can be dynamically updated during the optimization process.

In EdgeTrak, the optimization method is based on dynamic programming [9]. The normal of the snake element v_i is recalculated in each optimization 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 neighbors of the element and itself. The optimization 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 minimized and the elements under consideration are only v_{i-1}, v_i and v_{i+1}. After n stages, the energies of all snake elements are minimized and the energy of each element is 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 with dynamic programming is dropped from $O(l^n)$ to $O(n * l^3)$ (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 the point. In fact, due to the aperture problem, only the deformation along the normal direction can be detected. 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 optimization 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.

$E_{ext}'(v_i)$ 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 optimization 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.
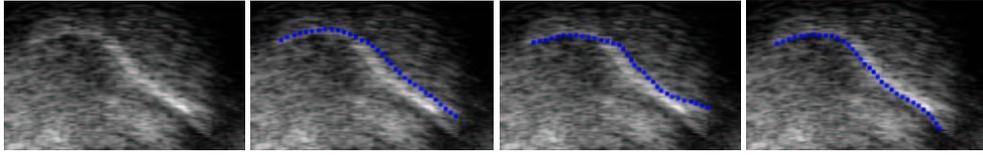
Figure 4. Extraction of the lower edge of human tongue. Left: Ultrasound tongue image. Middle left: Snake initialization. Middle right: edge extracted without band energy; some snake elements are attracted to the uninteresting high-gradient upper edge of the air reflection. Right: edge extracted correctly with band energy.
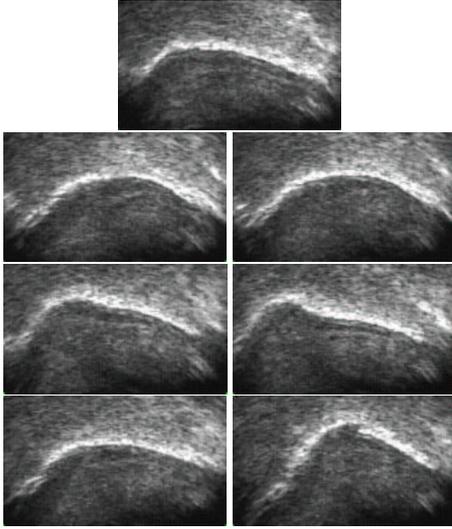
Figure 5. Image sequence of human tongue motion. Every 10th frame from 67 frames is shown. Images are ordered from top to bottom, left to right.
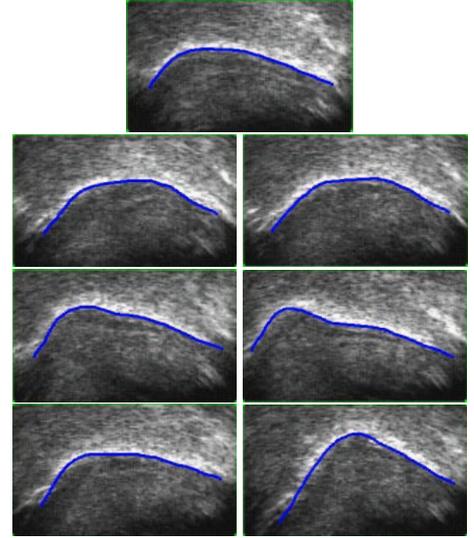
Figure 6. Tracked contours for the sequence in Figure 5. User input is only seven points in the first frame. All contours are tracked automatically.

4 Experiment Results

4.1 Validation of Human Tongue Tracking in Ultrasound Images

EdgeTrak has been applied to ultrasound image sequences of human tongue motion for tracking. In this 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 surface by the automatic dynamic programming optimization process. Every frame in the sequence gets its snake initialization from the previous frame and the snake is optimized in the same way as in the first frame. The tracking result for Figure 5 is shown in Figure 6. Another ultrasound image sequence is shown in Figure 7 and its tracking result is shown in Figure 8. The visual inspection of the tracked contour shows that our snake model works pretty well.

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

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

Contours tracked by EdgeTrak and manual tracking by two speech scientists for three speech sequences were compared. The speech materials 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.

EdgeTrak is currently being used by speech scientists.
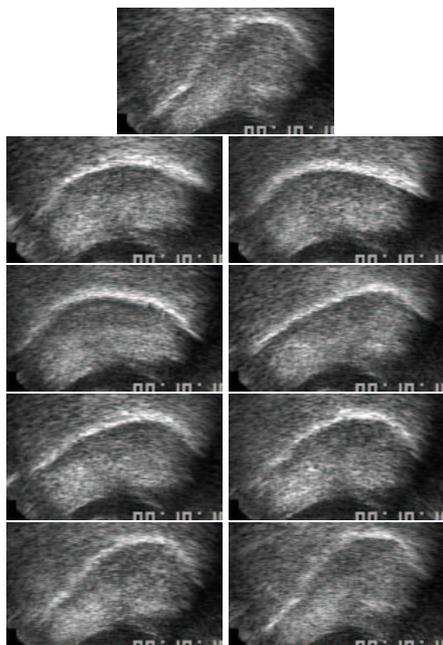
Figure 7. Image sequence of human tongue with different motion. Every 4th frame from 33 frames is shown. Images are ordered from top to bottom, left to right.

	"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=0.295 mm.

Our feedback from them indicates that the system is efficient and robust for speech research and related applications.

4.2 Tracking the Human Face

Although there is no obvious band shape presented in the edge of the face, the proposed snake model can still be applied to face tracking by considering that there is a virtual band along the face edge. The actual face boundary should be the outer edge of this virtual band. The difference between the virtual band intensity and the intensity of the non face area will help us to correctly locate the face boundary.

There is no ground truth for the face boundary to evaluate our tracking results. By visually comparing the face boundary tracked without band energy (Figure 9) and the tracking results with band energy (Figure 10), one can see that the face boundary is correctly tracked by the proposed snake model. Unrelated edges such as the high-gradient part below the lip are successfully discarded by using the band energy. Note that we have defined our contour to be

Figure 8. Tracked contours for the sequence in Figure 7. The user input is only seven points in the first frame. All contours are tracked automatically.

"open" on the face boundary, in order to only track the important details during facial expressions.

5 Conclusion

A snake model which is suitable for edge extraction of band-shape objects is presented in this paper. Region information around each snake element is incorporated with the image gradient and the contour orientation is taken into account in our snake model. Compared with the traditional snake model and other 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 the edge tracking system that has been developed based on the proposed snake model and is being used by speech scientists. The feedbacks from them indicates that the system is efficient and accurate for speech research and related applications. The proposed snake model can also be used to extract edges of non band-shape objects by considering a virtual band along the object edge. Its application to human face tracking has been shown in this paper and the tracking result has been verified by qualitative evaluations.
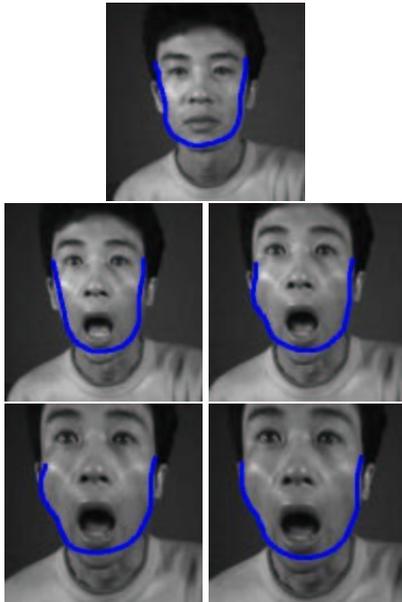
Figure 9. Face tracking without the band energy for a 'surprise' sequence. Note that the boundary is not correctly located at the chin. There are 20 frames in this sequence and every 4th frame is shown. The person in the image is moving toward the camera at the same time with a 'surprised' expression.

Figure 10. Face tracking of the sequence in Figure 9 with band energy. Note that the face boundary is correctly tracked compared with Figure 9.

6 Acknowledgment

References

[1] M. Kass, A.P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, vol. 1, no. 4, pp. 321–331, January 1988.

[2] L.D. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-d and 3-d images," *PAMI*, vol. 15, no. 11, pp. 1131–1147, November 1993.

[3] L.D. Cohen, "On active contour models and balloons," *CVGIP*, vol. 53, no. 2, pp. 211–218, March 1991.

[4] S.R. Gunn and M.S. Nixon, "Robust snake implementation: A dual active contour," *PAMI*, vol. 19, no. 1, pp. 63–68, January 1997.

[5] M. Wang, J. Evans, L. Hassebrook, and C. Knapp, "A multistage, optimal active contour model," *PAMI*, vol. 5, no. 11, pp. 1586–1591, November 1996.

[6] V. Chalana and D.T. Linker, "A Multiple Active Contour Model for Cardiac Boundary Detection on Echocardiographic Sequences," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 290–298, June. 1996.

[7] Y.S. Akgul, C. Kambhamettu, and M. Stone, "A Task-Specific Contour Tracker for Ultrasound.," *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, Hilton Head Island, South CarolinaI*, pp. 135–142, 2000.

[8] T.F. Chan and L.A. Vese, "Active contours without edges," *IP*, vol. 10, no. 2, pp. 266–277, February 2001.

[9] A.A. Amini, T.E. Weymouth, and R.C. Jain, "Using dynamic programming for solving variational problems in vision," *PAMI*, vol. 12, no. 9, pp. 855–867, September 1990.

[10] M. Stone and E.P. Davis, "A head and transducer support system for making ultrasound images of tongue/jaw movement.," *The Journal of The Acoustical Society of America*, vol. 6, pp. 3107–3112, December. 1995.

[11] Y.S. Akgul and C. Kambhamettu, "A coarse-to-fine deformable contour optimization framework," *PAMI*, vol. 25, no. 2, pp. 174–186, February 2003.

[12] Y.S. Akgul, C. Kambhamettu, and M. Stone, "Automatic Extraction and Tracking of The Tongue Contours.," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 1035–1045, October. 1999.