Group Average Difference: A Termination Criterion for Active Contour

Tong Kuan Chuah · Jun Hong Lim · Chueh Loo Poh

Published online: 20 July 2011
© Society for Imaging Informatics in Medicine 2011

Abstract This paper presents a termination criterion for active contour that does not involve alteration of the energy functional. The criterion is based on the area difference of the contour during evolution. In this criterion, the evolution of the contour terminates when the area difference fluctuates around a constant. The termination criterion is tested using parametric gradient vector flow active contour with contour resampling and normal force selection. The usefulness of the criterion is shown through its trend, speed, accuracy, shape insensitivity, and insensitivity to contour resampling. The metric used in the proposed criterion demonstrated a steadily decreasing trend. For automatic implementation in which different shapes need to be segmented, the proposed criterion demonstrated almost 50% and 60% total time reduction while achieving similar accuracy as compared with the pixel movement-based method in the segmentation of synthetic and real medical images, respectively. Our results also show that the proposed termination criterion is insensitive to shape variation and contour resampling. The criterion also possesses potential to be used for other kinds of snakes.

Keywords Termination criterion · Active contour · Gradient vector flow · Area difference · Group average

Introduction

Active contours or snakes [1] are moving curves used for detecting image boundaries. It is mostly applied in segmen-

tation tasks: these include segmentations applied on radar imagery [2], computed tomography [3], magnetic resonance images of the muscles [4], and ultrasonic breast images [5]. Active contour can be broadly categorized into two types: parametric and geodesic. Parametric active contour employs an energy functional minimization formulation to search for the correct location for a parameterized contour [1]. Hence, it is more suitable for segmentation of a single structure in an image. Geodesic active contour uses level set methods to find the locations of final contours. It is more suitable for segmentation of multiple structures in an image due to its robustness in merging and splitting the contours during evolution [6]. For both types of active contours, terminating the contour evolution in a fast and accurate manner has been a challenge that requires further research attention.

Although theoretically the formulation of active contour expects that there will be no movement of the contour when it reaches the desired location (such as reaching a force balance of internal and external forces), the ideal force balance and zero contour movement is practically difficult to obtain because it may involve any or all of these: very small time step, long evolution time, and extensive tuning of force parameters. In view of these difficulties, several termination criteria were proposed [7–12]. For parametric active contour, existing methods are mostly based on pixel movements of the contour. They include sum of norms [7], sum of distance, maximum distance, minimum distance, median distance, and average distance. Ray et al. [10] has also proposed a definite termination criterion that is based on the direction of pixel movements. The criterion terminates the evolution when the percentage of contour pixels moving in opposing directions in subsequent iterations is above the certain predetermined threshold. Generally, the termination criteria used in parametric active contour are not applicable to geodesic active contour because there is no pixel correspondence between the curves

T. K. Chuah · J. H. Lim · C. L. Poh (✉)
Division of Bioengineering, School of Chemical & Biomedical Engineering, Nanyang Technological University,
N1.3-B2-09, 70 Nanyang Drive,
Singapore 637457, Singapore
e-mail: CLPoh@ntu.edu.sg

of subsequent iterations in geodesic active contour. For geodesic active contour, termination can be based on the energy functional [9, 11], entropy and simulated annealing [8], and the change of length of the contour [12]. The termination criteria used for geodesic active contour, such as that used in [9, 11], have not been demonstrated in the use of parametric active contour. Hence, their applicability is still unclear. Despite much effort in proposing new termination criteria, studies that assess termination criteria by systematically comparing them are sparse.

To assess how well a termination criterion performs for both types of active contours, three important factors should be considered. They include accuracy, segmentation speed, and shape insensitivity of the criterion. Shape insensitivity measures the extent that the shape change of the target shape can have (while holding the termination parameters constant) before the criterion fails. A criterion is judged to be a failure if it does not return an accurate result within reasonable time or a number of iterations. Although shape insensitivity is a qualitative measure, it is important in practice because the shape of the target structure can vary significantly across different samples (of the same anatomical location) and/or different slices (in the same volumetric dataset). Different slice locations may need to be segmented at different instances or for different patients. An active contour termination criterion is useful in practice if its performance (such as segmentation time and accuracy) is insensitive to this shape variation [13].

One more factor should be considered for the assessment of termination criteria for parametric active contour: insensitivity to contour resampling. Contour resampling is the addition and removal of contour vertices in subsequent iterations, as suggested in [14]. The contour resampling step is necessary for avoiding vertex clustering and vertex dispersion. A termination criterion insensitive to contour resampling is applicable for both types of parametric active contour: with and without resampling. Contour resampling results in the loss of pixel correspondence of contours in subsequent iterations. As a result, all pixel movement-based termination criteria are inapplicable to contour with resampling. They are considered to be sensitive to contour resampling.

In this paper, we present a termination criterion called group average difference (GAD) which is designed for parametric gradient vector flow (GVF) active contour [15] with contour resampling and unit normal force selection. This termination criterion is based on the image difference between subsequent iterations, as opposed to most of the existing criteria that use pixel movement [7, 10]. This paper presents studies that evaluate the proposed termination criterion in terms of segmentation speed, accuracy, shape insensitivity, and insensitivity to contour resampling by comparing with existing termination criteria. The merits and limitations of the proposed criterion are also discussed.

Method

Overview

This paper presents the GAD termination criterion by first studying its trend line and comparing with that of existing termination criteria (*trend line study*). Secondly, the termination criterion is evaluated with synthetic images of different target shapes (*time accuracy study*). To demonstrate that the proposed criterion is applicable to real medical images, we applied the GAD for the segmentations of three different anatomical structures of varying shapes using medical images. Sample anonymized medical images were used with ethics approval. The segmentations in medical images include femur, tibia, and humerus using magnetic resonance (MR) images. The anatomical structures chosen have varying shapes across different slices within a volumetric dataset. Segmentation of these structures is important for visualization in diagnosis and surgical planning [16–18]. All medical images used underwent Otsu's thresholding [19] before the application of the active contour so that the active contour essentially evolved on an edge map of a binary image.

The purpose of the *trend line study* was to compare the trend of the proposed criterion with the trends of existing criteria. This comparison enables one to select or design a steadily decreasing metric for termination without having to change the energy functional of the active contour. The comparison of trend lines also allows visualization of the robustness or weakness of a termination metric.

The *time accuracy study* was carried out for studying the termination time, accuracy, and shape sensitivity of existing criteria and the proposed criterion. In this study, we compare the average pixel movement (a.k.a. average distance) with GAD across six different test shapes of varied complexity using synthetic images. Average pixel movement (APM) is chosen amongst other existing termination criteria because it is representative of most other pixel movement-based criteria as shown from the results of the *trend line study*. The shapes tested include various curvatures, corners, and inward and outward concavities.

Both studies in this paper employ the GVF active contour with contour resampling (no contour resampling for tests of APM termination method) and unit normal force selection.

GVF Snake with Contour Resampling and Unit Normal Force Selection

The parametric GVF snake used in this paper follows the force balance equation [15]

$$\mathbf{x}(s, t) = \alpha \mathbf{x}''(s, t) - \beta \mathbf{x}'''(s, t) + \mathbf{v} \quad (1)$$

where s is the parameter of the contour $\mathbf{x}=[x(s), y(s)]$; x, y are the horizontal and vertical spatial coordinates, respec-

tively; vector field $\mathbf{v}=[u(x, y), v(x, y)]$ is the GVF field; and α and β are weights for the first and second derivative terms of internal forces in the energy functional of traditional snake [1]. Solution of Eq. 1 provides the critical point (minimum in this case) of the energy functional. The left-hand side of the equation is the time derivative of the contour (subscript here indicates derivative); the right-hand side of the equation is the balanced force. Both the balanced force and the time derivative of the contour equal zero at the critical point.

The balanced force comprises both internal force and external force (GVF field) $\mathbf{v}$. The internal force is made up of the second and fourth partial derivatives of the contour with respect to the parameter s . For all the images processed in this paper, we use $\alpha=0.85$ and $\beta=0.01$. Snake evolution time step, ζ , is set to 0.075. These values (α , β , and ζ) were chosen based on results from a separate optimization study of the active contour. Forward time and central space numerical schemes were applied for implementation.

To derive $\mathbf{v}$, the edge map $f(x, y)$ used is

$$f(x, y) = |\nabla[G_\sigma(x, y)*I(x, y)]|^2 \quad (2)$$

$I(x, y)$ is the thresholded binary image of interest; $G_\sigma(x, y)$ is a 5×5 two-dimensional Gaussian kernel with a standard deviation $\sigma=2$; ∇ is the gradient operator, and $*$ is the convolution operator. The edge map is employed in the numerical solution for the GVF components u and v as follows:

$$u_t(x, y, t) = \mu \nabla^2 u(x, y, t) - [u(x, y, t) - f_x(x, y)] \cdot [f_x(x, y)^2 + f_y(x, y)^2] \quad (3a)$$

$$v_t(x, y, t) = \mu \nabla^2 v(x, y, t) - [v(x, y, t) - f_y(x, y)] \cdot [f_x(x, y)^2 + f_y(x, y)^2] \quad (3b)$$

The parameter μ is the regularization term associated with smoothing relative to image gradient; u_t and v_t are time derivatives of the horizontal and vertical GVF components, respectively; f_x and f_y are horizontal and vertical spatial derivative of the edge map, respectively; and ∇^2 is the Laplacian operator.

The regularization term μ is set to 0.2. While solving Eqs. 3a and 3b, u and v were initialized using the directions (having magnitudes of unity) of the external force field of the traditional snake. Rounded value of $\sqrt{N}$ iterations were applied (N is the maximum of the image size along either the horizontal or the vertical direction). The resulting GVF from this solution is a static force field which does not change in the process of snake evolution. The iterations use a time step (ζ_{GVF}) value of 0.82 for the image used in the *trend line study*

and medical images in the *time accuracy study*, and a value of 0.01 for the synthetic shapes in the *time accuracy study*. This time step value is chosen based on the closeness between the initial contour and the target boundary. Whenever termination criteria are compared in this paper, these evolution parameters mentioned are set to be the same in the criteria (note: not to confuse ζ_{GVF} , the time step for obtaining GVF, with ζ , which is the time step for snake evolution).

During contour evolution, all the external forces encountered by the contour are resolved into the tangential and normal directions with respect to the contour. The tangential component of each force is discarded because it does not contribute to the shape of the contour, but causes the problem of vertex dispersion or vertex clustering of the contour. If a normal component is present on the contour, only the direction (magnitude of unity) is used to step the contour forward in time. The process of incorporating normal force onto the contour for evolution is called *normal force selection*. Taking unity magnitude for normal force makes the contour move in a more consistent way which is dominated by the external forces relative to the shape of the contour. To further avoid vertex dispersion and clustering problem, contour resampling operation is carried out (but not for evolutions terminated based on pixel movement) for each time step of the evolving contour, as suggested by [14].

Pixel Movement-Based Termination Criteria

This section defines existing [7] (Eq. 4) and potential (Eqs. 5 to 9) pixel movement-based termination metrics for parametric active contour. A contour on a digital 2D image is defined by a finite number of pixels that represent the vertices of the contour. Let this finite number be n ; the vector $\vec{x}$ is a column vector that lists the horizontal coordinates, and the vector $\vec{y}$ is the column vector that lists the vertical coordinates of the n points. We denote $\vec{x}$ and $\vec{y}$ at the i th iteration as $\vec{x}_i$ and $\vec{y}_i$. For contour without resampling during evolution, n is a fixed number throughout evolutions. Let the distance traveled by each of n contour pixels at the i th time step be

$$d_i = \left[\left(\vec{x}_i - \vec{x}_{i-1} \right)^2 + \left(\vec{y}_i - \vec{y}_{i-1} \right)^2 \right]^{\frac{1}{2}}$$

The termination metrics at the i th time step are

$$\text{Sum of norm} = \left\| \vec{x}_i - \vec{x}_{i-1} \right\| + \left\| \vec{y}_i - \vec{y}_{i-1} \right\| \quad (4)$$

$$\text{Sum of distance} = \sum_1^n d_i \quad (5)$$

$$\text{Average distance (APM)} = \frac{\sum_1^n d_i}{n} \quad (6)$$

$$\text{Maximum distance} = \max(d_i) \quad (7)$$

$$\text{Minimum distance} = \min(d_i) \quad (8)$$

$$\text{Median distance} = \text{median}(d_i). \quad (9)$$

The symbol $\|\dots\|$ denotes Euclidean norm. The sum of distance is obtained by first calculating the distances moved by each of n contour pixels, and then all the n distances are added together to form the metric.

Group Average Difference Termination Criterion

Once the contour evolves through one time step, a list of coordinates that represent the vertices of the contour in double precision is obtained. The coordinates are rounded when they are drawn onto the pixelized image. This rounding causes the contour to be broken at some instances. Fortunately, due to the resampling operation, the breakage of the contour is no larger than 1 pixel at every instance. This breakage is repaired on the binary image describing the contour with the background using the following method:

1. Apply Gaussian low-pass filter (3×3 kernel size, $\sigma=0.5$) after changing the binary image to double precision.
2. Convert the double precision image back to binary image, with all non-zero pixels eventually carrying the value of 1.
3. Thin the contour to 1-pixel width using the morphological algorithm described in [20].

Once the contour has been repaired, the binary image is filled with a morphological hole-filling operation. This operation fills any background pixels that cannot be reached by filling the background from the edge of the image. The filling is based on four-connectivity neighborhood. A filled area is obtained for every iteration step. This filled area forms the segmented mask of target structure and is the basis for setting the termination criterion.

Two images of filled areas in subsequent iterations are compared using XOR logical operator. The more the two images share the same filled area, the less logical ones will appear in the resulting image. Denoting the filled image of the i th iteration as I_i , for an $M \times N$ binary image (M is the size at the vertical direction and N is the size at the horizontal direction), we have

$$I_i \text{ XOR } I_{i-1} = I_d(i), \quad 2 \leq i < \infty \quad (10)$$

$$[D]_k = \left[\sum_{y=1}^M \sum_{x=1}^N I_d(x, y) \right]_i \quad k = i - 1 \quad (11)$$

By summing up the number of logical ones in I_d , we would expect a decreasing metric as the snake approaches equilibrium position. The quantity D is denoted as the image difference. The index k is the iteration index of D , which is always one integer less than i .

However, due to the use of only normal force of unity magnitude on the contour for evolution, D will not decrease to zero. This is because when the contour is near the equilibrium position, the contour location will fluctuate within 3 pixels around the exact edge. This is a pseudo-equilibrium state. The fluctuation will only be of practical concern if the image is small (e.g., smaller than 100×100 pixels) and/or the target structure is thin (e.g., having thickness smaller than 10 pixels). Recognizing that when all the vertices of the contour reach their pseudo-equilibrium position the value D will fluctuate around a constant, we define the quantity called the *group average*, G .

$$[G]_j = \frac{\sum_{k=(1+g(j-1))}^{jg} D_k}{g}, \quad 1 \leq j < \infty, g \in \mathbb{Z} \quad (12)$$

In Eq. 12, g is an integer representing the number of iterations to form a group. The group average has iteration index j . The calculation of group average is non-overlapping (i.e., the data used for computing the first group average are not used for computing the second group average, etc.). Since D fluctuates around a constant, we can expect G to be approaching a constant when g is reasonably large. The value of g is selected to be 200 in our application. For a quantity that approaches zero while the contour progresses and settles, we define the *group average difference* (E) as

$$[E]_p = [G_{j+1} - G_j], p = j, \quad 1 \leq p < \infty, \quad (13)$$

The index p is the iteration index of the group average difference. The evolving contour is expected to terminate when p is increasing and E is approaching zero.

Analysis

For the *trend line study*, we study the trend line of each metric versus the number of iterations, up to 1,200 iterations (equivalent to 1,200 time steps). The decrease of trend and the steadiness of decrease are examined. A steadily decreasing metric is deemed to be a metric more suitable for practical applications. It is because a steadily decreasing metric provides the ease of determining a threshold for termination, i.e., a value near to zero, whereas it is difficult to determine a threshold for a non-decreasing metric.

For both the *trend line study* and the *time accuracy study*, the maximum number of iterations is set at 10,000.

If a termination metric threshold cannot be reached within this maximum number of iterations, the evolution is forced-terminated.

In the *time accuracy study*, the time to termination is used as an indicator of speed of segmentation. All contour evolutions were implemented using MATLAB (The Math-Works, Inc.) on IBM workstation with Intel Xeon CPU 5160 at 3.00 GHz. The number of iterations to termination is used as an indicator to the shape sensitivity of a termination method. Failure to terminate within 10,000 iterations is deemed as the failure of a particular termination criterion to terminate within a reasonable time. Shape sensitivity is measured by the frequency of failure of a termination metric to terminate within a reasonable time. By having unified threshold values for APM and GAD, the shape sensitivity of the termination criteria can be studied.

For all the segmentations carried out using active contour, accuracy is assessed using the dice similarity coefficient (DSC) [21], verified with the ground truth. The ground truths for medical images are obtained from manual segmentation carried out by an expert. DSC is defined as $2TP/(2TP+FP+FN)$, where TP is true positive, TN is true negative, FP is true positive, and FN is false negative [21]. This coefficient is a measure of overlap between the ground truth and the active contour segmentation. It takes into account both positive overlap and negative overlap because it decreases as the area of false positive and/or false negative increase(s).

Experimental Results

Trend Line Study

In this study, trend lines of termination metrics of the active contour on 16 images (six synthetic images and ten real medical images) were analyzed. The six synthetic images comprise binary images of different shapes (“[Time Accuracy Study](#)”) and the real medical images are MR images of the femur, tibia, and humerus (“[Evaluation of Termination Methods on Medical Images](#)”). In general, the results show that it is difficult to find a good threshold for all shapes using pixel movement-based metrics, whereas GAD shows a steadily decreasing trend approaching zero for all shapes, which enables the threshold to be better determined. For simplicity, trend lines of one representative image are presented. The representative image is a knee magnetic resonance imaging (MRI) slice with a size of 171×171 pixels.

The initial contour for this study is shown in Fig. 1a. The final contour for the pixel movement-based evolution (without contour resampling each time step) is shown in Fig. 1b, and the final contour for evolution with contour resampling (that is used in obtaining GAD) is shown in

Fig. 1c. All of the contours are shown to overlap with the MRI slice (histogram equalized for display) used. To show an example of how the pixel movement-based criteria vary over evolution time steps, we have plotted the six termination metrics in Fig. 2 using the representative knee MR image.

From the evolution of the contour in Fig. 1a to the final contour shown in Fig. 1b, we obtained the trend line for each termination criterion as shown in Fig. 2a–f. In this evolution process, a visual inspection of the evolving contour suggested that an accurate and efficient termination metric should terminate the evolution between iterations 800 and 1,000.

From the evolution of the contour in Fig. 1a to the final contour shown in Fig. 1c, the trend lines for image difference, group average, and GAD are shown in Fig. 3a–c, respectively.

Referring to Fig. 2, the minimum distance and maximum distance (Fig. 2d, e) have shown to be non-decreasing metrics which are difficult to be used directly for automatic termination. Although the median distance (Fig. 2f) is a decreasing metric, it shows a great extent of fluctuation. The sum of norm, sum of distance, and APM (Fig. 2a–c) show similar decreasing trends, although their ranges are different. APM can be a representative of these three metrics because its valleys are nearest to zero. It is useful to highlight that these three metrics hit their first low around iteration 600. On one hand, if a threshold were to be set at the level of the metric at iteration 600, a premature termination could occur (because accurate termination for this segmentation should occur between iterations 800 and 1,000 based on observations). On the other hand, if a threshold were to be set at a very low value of a metric not covered in these plots, the time taken to termination could be unacceptably long.

The problem of pixel movement-based termination criterion is not only its inability to be applied when contour resampling is present but also its difficulty to find an accurate and efficient termination threshold. From our study, the value of the metric corresponding to the first lowest valley of a pixel movement-based metric is highly dependent on the shape of the target structure. The unpredictability of pixel movement-based metric is high with respect to target shape change. The problem of shape dependence of the pixel movement-based criterion will be evident when a change in target shape causes the failure of the criterion to terminate (see “[Time Accuracy Study](#)”).

In contrast to the pixel movement-based termination metrics, although image difference is a non-decreasing metric (Fig. 3a), its trend is predictable based on observations. Therefore, using image difference, we are able to derive a predictably steadily decreasing metric, GAD. It is evident from Fig. 3c that a threshold value can easily be chosen by a value near to zero so that the evolution can terminate automatically at iterations 800 to 1,000.

Fig. 1 **a** Initial contour on the MRI slice used. **b** Final contour from pixel movement-based evolution. **c** Final contour for evolution with contour resampling

Time Accuracy Study

This study examines the speed and accuracy of each termination method. The segmentation speed and the accuracy of the methods are measured by the time of termination and DSC, respectively. To obtain objective measures of the time of termination and DSC across different target shapes, the values of time and DSC are averaged across different test shapes and only the average values are used for comparing different termination methods. The test images consist of six synthetic binary images having test shapes of varied complexities (Fig. 4). The images are of moderate size, 325×273 pixels.

Initial contours for all the synthetic images are all circles of radius 45 pixels centered at the center of the image. This study consists of two parts: the first part examines the termination methods when all associated parameters (including smoothness parameters of contour, time step, APM threshold, GAD threshold, and group size) are optimized to the specific shape *and* the specific method. The aim of this part of the study was to compare APM with GAD in their best achievable performance. The second part of this study examines the feasibility of the methods for automatic implementation in which different shapes need to be segmented. This part of the study compares APM with GAD while each parameter (including smoothness parameters of contour, time step, APM threshold, GAD threshold, and group size) is set to a unified value for all shapes in an anatomical structure. In this study, the best combination of parameters is chosen for each method for comparison.

Results from the first part of the study are shown in Table 1. APM termination method gives an accuracy, measured by DSC, of $97.55 \pm 1.59\%$. GAD termination gives an accuracy of $97.55 \pm 1.65\%$, which is approximately the same as that of APM. The average time taken using GAD method to terminate is 69 s, and the average time for termination using APM is 24 s. In this case, in which all the parameters are optimized for the specific shape and specific method, the APM method is three times faster than the GAD method.

To test the time and accuracy of the termination methods considering its robustness with variation of target shape in automatic implementation, the second part of this *time accuracy study* tests each method with fixed values of parameters across all test shapes. For the APM method, the combination of parameters including smoothness parameters of contour, time step, and the APM threshold that gives the best average accuracy of all shapes are chosen for comparison. Similarly for the GAD method, the combination of parameters including smoothness parameters of contour, time step, GAD threshold, and the group size that gives the best average accuracy of all shapes are chosen for comparison. Table 2 gives the comparison results of APM with GAD in their best variation studied while keeping parameters unified across all shapes.

In this second part of the *time accuracy study*, the GAD method gives a slightly lower accuracy ($97.70 \pm 1.42\%$) compared with the DSC accuracy of APM ($97.75 \pm 1.49\%$). However, the accuracy difference is neither large nor significant statistically (Fig. 4). The average time for termination using GAD is 102 s, almost half of the average time using APM, which is 179 s.

In this automatic implementation in which all termination parameters were set constant for all images, the results reveal the *shape sensitivity* nature of the APM termination method because in three out of six test images, the termination metric did not reach the threshold set before reaching the maximum number of iterations allowed. The GAD method is relatively shape-insensitive because the contours are able to terminate before hitting the maximum number of iterations. Because of this shape sensitivity nature of the pixel movement-based termination metric, its average time for termination becomes longer compared with the GAD method in automatic implementation.

Evaluation of Termination Methods on Medical Images

For the purpose of testing the termination criterion in situations where anatomical structures have different shapes

Fig. 2 Trend lines of sum of norm (a), sum of distance (b), average distance (APM) (c), maximum distance (d), minimum distance (e), and median distance versus number of iterations (f)

in different subjects or at different slice locations, three exemplar anatomical structures (sagittal femur, sagittal

tibia, and axial humerus) were segmented using sample anonymized medical images. The sizes of all medical

Fig. 3 Image difference (a), group average (b), and group average difference (c), as defined in Eqs. 11, 12, and 13, respectively

images used in this section are the same, 320×320 pixels. All medical images were preprocessed to binary images before the contour evolutions. The preprocessing uses optimal threshold for each image for best delineation of the desired structure. The unified termination parameters (i.e., APM threshold, GAD threshold, and group size) for automatic implementation were obtained by optimizing between the accuracy and the time for termination, for each termination method. APM threshold was optimized for each anatomical structure.

Because of the large extent in which the anatomical target shape can vary across different slices in a volumetric datasets, and that the purpose of this study was to compare termination methods but not initialization methods, the initial contours are not fixed across different target shapes, but are consistent across different termination methods for the same target shape. Being consistent in the usage of

initial contour ensures that the comparison of termination methods is unaffected by variations of the initial contour. The initial contour for each medical image is defined using an interpolated cubic spline curve by receiving a number of user-specified points. Arbitrary number of points is used to define the initial contour for the medical images so that the initial contour for each image is fairly smooth and near to but not at the intended final location (where a contour should converge). This way of initialization is to simulate the contour evolution at and only at the termination stage to observe the termination pattern using the APM and GAD termination criteria, respectively.

Similarly, for each shape in the medical images, smoothness parameters and time steps are optimized (in another study) for each target anatomical structure and used during the contour evolution of the medical images of that structure, consistently across different termination methods.

Fig. 4 *Leftmost column* Initial contours. *Middle column* Final contours of APM terminated evolutions. *Rightmost column* Final contours of GAD terminated evolutions. All final contours are from automatic implementation

Table 1 Comparison of time taken, number of iterations to termination, and accuracy for six synthetic test shapes using the conventional method and the proposed method

	Conventional method (optimized)				Proposed method (optimized)				
	APM				GAD				
	Threshold	Time (s)	Iterations	DSC	Threshold	Grp size	Time (s)	Iterations	DSC
Indented circle	0.0466	13	363	0.9890	0.25	15	22	390	0.9881
Heart	0.0533	15	436	0.9896	0.25	50	20	350	0.9904
Quadrilateral	0.0699	39	1,110	0.9849	0.25	50	66	1,200	0.9838
Triangle	0.0597	11	323	0.9717	0.25	100	27	500	0.9457
Fish	0.0469	47	1,361	0.9700	0.25	50	93	1,550	0.9714
Mean		24	679	0.9755			69	1,282	0.9755
SD		15	442	0.0159			65	1,281	0.0165

All parameters optimized for each shape and each method

These parameters are implemented in the settings for active contour runs in examining and comparing the conventional termination method (APM) and the proposed method (GAD).

Table 3 shows the comparison of the automatic implementation for the APM and GAD termination criteria. Both methods of termination achieved very good accuracy measured by DSC. GAD has an average accuracy of $96.77 \pm 0.92\%$ and APM has an average accuracy of $96.75 \pm 0.94\%$. There is no significant difference in accuracy using these two methods. However, the shape sensitivity of APM becomes apparent when six out of nine images were unable to terminate before reaching the limit for the number of iterations; in contrast, in the case of the GAD termination criterion, none of the images fails to terminate before hitting the iteration limit. Shape sensitivity indicates a need for shape-specific tuning of the termination metric. Without shape-specific tuning, a

shape-sensitive termination method can result in a long segmentation time (which is not the optimal segmentation time the method can possibly achieve). In this automatic implementation, the total segmentation time of all images tested using APM is 2,121 s, whereas the total time using GAD is 607 s. In this case, there is about 60% total time reduction using GAD compared with using APM. The average segmentation time of all images tested using APM is 236 ± 217 s, whereas the average time using GAD is 67 ± 50 s. Both total segmentation time and average segmentation time of GAD represent about 60% time reduction compared with APM. The coefficient of variation (COV) of the segmentation time is 92% using APM and 75% using GAD. The COV can be regarded as a measure of relative shape sensitivity for comparing the methods because the segmentation time would have a large variation from its mean if the termination criterion cannot terminate the contour within the maximum allowed number of

Table 2 Comparison of time taken, number of iterations to termination, and accuracy for six synthetic test shapes using the conventional method and the proposed method

	Conventional method (optimized)				Proposed method (optimized)				
	APM				GAD				
	Threshold	Time (s)	Iterations	DSC	Threshold	Grp size	Time (s)	Iterations	DSC
Indented circle	0.0466	12	363	0.9891	0.25	100	39	700	0.9902
Heart	0.0466	29	850	0.9912	0.25	100	56	1,000	0.9896
Quadrilateral	0.0466	336	10,000	0.9858	0.25	100	105	1,900	0.9847
Triangle	0.0466	338	10,000	0.9781	0.25	100	101	2,000	0.9742
Star	0.0466	335	10,000	0.9521	0.25	100	268	5,100	0.9536
Fish	0.0469	47	1,361	0.9700	0.25	100	268	1,550	0.9714
Mean		179	5,313	0.9775			102	1,900	0.9770
SD		173	5,137	0.0149			65	1,670	0.0142

All parameters unified for all shapes; best accuracy combination of parameters for each method

Table 3 Comparison of time taken, number of iterations to termination, and accuracy for nine medical images of three anatomical locations using the conventional method and the proposed method in automatic implementation

	Conventional method (optimized)				Proposed method (optimized)				
	APM				GAD				
	Threshold	Time (s)	Iterations	DSC	Threshold	Grp size	Time (s)	Iterations	DSC
Femur lateral	0.0468	418	10,000	0.9731	0.25	50	76	1,000	0.9710
Femur middle	0.0468	496	10,000	0.9655	0.25	50	42	500	0.9660
Femur medial	0.0468	6	194	0.9733	0.25	50	14	200	0.9730
Tibia lateral	0.0431	187	10,000	0.9802	0.25	50	133	2,500	0.9805
Tibia middle	0.0431	544	10,000	0.9742	0.25	50	155	1,700	0.9536
Tibia medial	0.0431	6	270	0.9726	0.25	50	61	1,000	0.9741
Humerus proximal	0.0460	332	10,000	0.9533	0.25	50	40	600	0.9536
Humerus intermediate	0.0460	13	345	0.9584	0.25	50	77	1,100	0.9580
Humerus distal	0.0460	119	10,000	0.9567	0.25	50	9	200	0.9586
Total		2,121					607		
Mean		236	6,757	0.9675			67	978	0.9677
SD		217	4,865	0.0094			50	745	0.0092

Middle refers to the location of the intercondylar slice in knee MRI

iterations in some cases. The larger COV of segmentation time resulted from using APM suggests that APM is more shape-sensitive than GAD.

The final contours for four representative results using real medical images are shown in Fig. 5. Their corresponding trend lines for both the APM and GAD evolutions are shown in Fig. 6. There is no visible difference on the image between the segmentation results terminated by APM and GAD.

The tests of GAD termination criterion using medical images ascertain the decreasing trend of the proposed metric (Fig. 6). These tests further illustrate the difficulty in selecting a specific APM threshold because the APM threshold is highly shape-sensitive, i.e., requiring different thresholds to terminate optimally. For example, the optimum APM termination threshold used for femur is optimal only for the medial slice of femur. For the middle slice of femur in which the target shape is different (not to confuse medial with middle), the threshold causes failure in terminating within the maximum number of iterations. On the other hand, using a threshold optimum for the middle slice for the medial slice may in turn result in a premature termination in the medial slice.

In contrast, without laborious manual search for a specific threshold for specific shape, an evolution is able to terminate within a reasonable number of iterations (in this paper, we consider 10,000 iterations as maximum for reasonable number of iterations) for all nine shapes tested using GAD (Table 3).

Discussion

Segmentation Time

Significant variations in the shape of a target structure can be encountered in practical situations such as: (1) same anatomical locations of different patients; (2) different pathological severity at the same anatomical location of the same or different patient(s); and (3) different slices of the same volumetric dataset. Hence, in an automatic implementation to segment an anatomical structure, it is important that the termination criterion in use is insensitive to this shape variation and terminates the snake evolution in a reasonable time and accuracy.

Because the time for termination of the proposed method is longer than that of the optimized-threshold APM termination (Table 1), it is probable that one favors the use of the optimized-threshold APM to terminate a contour. However, in practical application of automatic segmentation task which involves extracting structures of different shapes, it is unlikely that the user would search for the optimized parameters for each shape because it involves work as laborious as manually segmenting the target structure. In an automatic segmentation task, due to variations of shapes across different images, a termination method that does not need user intervention in searching for the correct threshold value is usually required. Table 2 provides the results of the methods run without user intervention across different shapes. The results suggest that due to the shape-sensitive

Fig. 5 Results of three representative segmentations of anatomical structures. *Leftmost column* Initial contours. *Middle column* Final contours of APM terminated evolutions. *Rightmost column* Final contours of GAD terminated evolutions. *Top row* Segmentation of

femur middle slice using sagittal knee MRI. *Second row* Segmentation of femur medial slice from the same dataset. *Third row* Segmentation of humerus intermediate slice from axial shoulder MRI. *Bottom row* Segmentation of humerus distal slice from the same dataset

nature of APM termination, its segmentation speed becomes disadvantageous in automatic implementation. Although GAD has one more parameter involved (the group size), its relative insensitivity to the change of target shape suggests that the tuning required is considerably less than that of the

Fig. 6 *Left column* Trend lines for APM (fixed threshold). *Right column* Trend lines for GAD (unified group size and threshold). *Top row* Trend lines for segmentation of femur middle slice. *Second row* Trend lines for segmentation of femur medial slice. *Third row* Trend lines for segmentation of humerus intermediate slice. *Bottom row* Trend lines for segmentation of humerus distal slice. The coordinates of first valleys of APM lines are labeled for over-evolved cases

application of APM, which only has one parameter to tune but very sensitive to the change of target shape. We have shown that the average time for termination using GAD in automatic implementation can be as short as half that using APM in the synthetic images. Tests in medical images ascertain the notion of the shape sensitivity of APM and that its time for automatic termination is longer than that using GAD.

Contour Resampling

All tests of APM termination criterion in this paper were carried out without contour resampling at every time step because contour resampling will cause the loss of vertex correspondence between the contours of subsequent time steps. This is an inherent restriction of using the pixel movement-based termination criterion. On the other hand, for stable evolution without redundant and/or insufficient vertices to characterize the contour, all tests of the GAD termination criterion were carried out with contour resampling at every time step. Since the GAD termination criterion can work with contour resampling and the APM criterion cannot, the pixel movement-based termination criteria such as APM are considered to be sensitive to the use of contour resampling.

Number of Iterations

The number of iterations for a parametric active contour depends on several factors such as the image size, the object size, use of normal force selection, use of contour resampling, and the shape of the object relative to the initial contour. In this study, the large numbers of iteration are mainly due to the *normal force selection* formulation that keeps the contour fluctuating around the exact boundary sought and, in medical images, partially caused by the relatively large size of images. Normal force selection is a formulation important to avoid vertex clustering and vertex dispersion while evolving the contour. The formulation also provides a more predictable movement pattern for the evolving snake since only unit normal force is applied onto the contour at any time. Because the contour is formulated using normal force selection, the movement of contour will not decrease to zero, but will settle at a pseudo-equilibrium state in which the contour fluctuates around the exact solution within approximately ± 3 pixels. While gaining the advantage that the contour vertices do not cluster/disperse because they do not move in the direction tangential to the contour, the associated effect which may be disadvantageous is the large number of iterations before the contour can be terminated by any criterion set objectively.

Image Size

A relatively small image is used for the *trend line study* because the use of a small image will result in the contour

reaching an equilibrium position at smaller number of iterations. Therefore, a relatively low limit of iteration (1,200) is sufficient to facilitate the analysis of the entire trend of termination metric. For the *time accuracy study*, this paper examined the methods of interest with images of moderate size. The limitation (in terms of time and accuracy) of the proposed metric will be evident on small (or thin) structure on a large image, or any structures on a small image. For these instances, the pixel movement-based metrics could perform much better in speed (including segmentation speed and evolution speed) and accuracy compared with GAD, although problems of shape sensitivity and sensitivity to contour resampling remain.

A Similar Termination Metric

A metric similar in concept to our proposed metric based on the change of contour length was recently proposed by Wang et al. [12]. Our metric is based on the area difference of the contour. We use the group size, g , to remove the fluctuation of the base metric (image difference) because the fluctuation of it is large and the number of iterations involved in parametric active contour is large; Wang et al. used a threshold of iteration, $T_{\text{iteration}}$, similar in concept to our group size. Since the number of iterations to the target shape for their geodesic active contour is small and the fluctuation of their metric is also small, they did not use $T_{\text{iteration}}$ to average the base metric but used it as the number of iterations if termination occurs due to the base metric (change in contour length) being below the threshold for $T_{\text{iteration}}$ iterations. Both our proposed GAD and Wang's termination criterion are insensitive to the type of contour.

Limitations of the Study

This paper examines the proposed termination criterion from the point of view of application to GVF active contour with normal force selection. Our focus in this paper is primarily to illustrate the advantages of the proposed metric compared with pixel movement-based metrics especially in situations where contour resampling is necessary.

The proposed method itself is applicable to all kinds of active contour regardless of its force balance formulation, but the group size is a parameter that needs to be adjusted when the formulation or numerical implementation of active contour is changed because the extent of fluctuation of the contour at the desired boundary is changed. Although the proposed metric shows good potential to be applied to other variations of active contours (e.g., geodesic active contours [6], the balloons [22], and region-based active contours [23]), its comparative performance (compared with other metrics) in those applications remains to be studied. This paper only compares the termination methods using 2D

active contour, not active surface in 3D. An extension of the proposed method would be to measure GAD based on volume difference (2D is based on area difference).

All images used in this paper possess defined boundaries. We compare the termination criteria of interest in cases where defined boundaries are present because even with a defined boundary, the contour can fluctuate when it reaches the boundary, as shown in the results. Consequently, terminating the evolving contour automatically is not trivial. The paper does not intend to address the challenge in accurately terminating an evolving contour in cases of ambiguous boundary, which is the subject of active contour formulation (e.g., introduction of a region-based term in the energy functional).

Conclusion

We have presented a new termination criterion called GAD designed for parametric active contour. We assessed the robustness of the criterion by considering its trend, speed, accuracy, shape insensitivity, and insensitivity to contour resampling in comparison to a pixel movement-based criterion. The proposed termination criterion showed high predictability of decreasing trend in a typical evolving active contour. In automatic applications, it has considerably higher average speed to termination than the pixel movement-based termination criteria with comparable accuracy. Given that a fixed threshold of termination metric was used across different target shapes, the failure (to terminate) rate of the proposed termination criterion is very low compared with the highly shape-sensitive pixel movement-based termination metrics. We have also demonstrated the insensitivity of the proposed termination criterion to contour resampling. Since the proposed criterion only depends on the segmentation outcomes of the contour and not the formulation of the contour, it should be applicable to different types of active contour/surfaces. Our future plan is to apply this new criterion on other variations of active contours.

Acknowledgments T. K. Chuah would like to express his gratitude to the university for providing financial support through Nanyang President Graduate Scholarship. The authors thank the Ministry of Education (MOE), Singapore, for supporting the project through SUG09/07 and AcRF Tier 1 (RG34/07) grant.

References

- Kass M, Witkin A, Terzopoulos D: Snakes: Active contour models. *International Journal of Computer Vision* 1:321–331, 1988
- Horritt MS, Mason DC, Luckman AJ: Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *International Journal of Remote Sensing* 22:2489–2507, 2001
- Jiang H, Cheng Q: Automatic 3D segmentation of CT images based on active contour models. *Proceedings of the 11th IEEE International Conference on Computer-Aided Design and Computer Graphics: City*
- Jurcak V, et al: Automated segmentation of the quadratus lumborum muscle from magnetic resonance images using a hybrid atlas based-geodesic active contour scheme. *Conference Proceedings: Annual International Conference of the IEEE EMBS 2008*:867–870, 2008
- Liu B, Cheng HD, Huang J, Tian J, Liu J, Tang X: Automated segmentation of ultrasonic breast lesions using statistical texture classification and active contour based on probability distance. *Ultrasound in Medicine and Biology* 35:1309–1324, 2009
- Chan TF, Vese LA: Active contours without edges. *IEEE Transactions on Image Processing* 10:266–277, 2001
- Acton ST, Ray N: *Algorithm KWT*: Morgan & Claypool Publishers, 2009
- Chen Y, Zhao W, Wang Z: Level set segmentation algorithm based on image entropy and simulated annealing. *Proc. ICBBE: City*
- Lee SH, Seo JK: Level set-based bimodal segmentation with stationary global minimum. *IEEE Transactions on Image Processing* 15:2843–2852, 2006
- Ray N, Chanda B, Das J: A fast and flexible multiresolution snake with a definite termination criterion. *Pattern Recognition* 34:1483–1490, 2001
- Shuai Y, Sun H, Xu G: SAR image segmentation based on level set with stationary global minimum. *IEEE Geoscience and Remote Sensing Letters* 5:644–648, 2008
- Wang XF, Huang DS, Xu H: An efficient local Chan–Vese model for image segmentation. *Pattern Recognition* 43:603–618, 2010
- Liu J, Gao W, Huang S, Nowinski WL: A model-based, semi-global segmentation approach for automatic 3-D point landmark localization in neuroimages. *IEEE Transactions on Medical Imaging* 27:1034–1044, 2008
- Lobregt S, Viergever MA: Discrete dynamic contour model. *IEEE Transactions on Medical Imaging* 14:12–24, 1995
- Xu C, Prince JL: Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* 7:359–369, 1998
- Dong X, Zheng G: Automatic extraction of proximal femur contours from calibrated X-ray images using 3D statistical models: An in vitro study. *International Journal of Medical Robotics and Computer Assisted Surgery* 5:213–222, 2009
- Baldwin MA, Langenderfer JE, Rullkoetter PJ, Laz PJ: Development of subject-specific and statistical shape models of the knee using an efficient segmentation and mesh-morphing approach. *Computer Methods and Programs in Biomedicine* 97:232–240, 2010
- Von Jan U, Sandkühler D, Kirsch L, Rühmann U, Overhoff HM: Surgical planning for the humeral part of a shoulder endoprosthesis based on landmarks determined from 3D ultrasound volumes. *International Journal of Computer Assisted Radiology and Surgery* 1:229–231, 2006
- Otsu N: Threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern SMC* 9:62–66, 1979
- Lam L, Lee S-W, Suen CY: Thinning methodologies—A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14:page 879, bottom of first column through top of second column, 1992
- Fripp J, Crozier S, Warfield SK, Ourselin S: Automatic segmentation of the bone and extraction of the bone–cartilage interface from magnetic resonance images of the knee. *Physics in Medicine and Biology* 52:1617–1631, 2007
- Cohen LD: On active contour models and balloons. *CVGIP: Image Understanding* 53:211–218, 1991
- Jehan-Besson S, Gastaud M, Barlaud M, Aubert G: Region-based active contours using geometrical and statistical features for image segmentation. *Proceedings of the IEEE International Conference on Image Processing: City*