Enhanced Active Contour Models for Tracking Shape-Changing Hand Gestures in Taiwanese Sign Language

Ming-Kai Chien, Pi-Fuei Hsieh, Ming-Jer Tsai, and Chung-Hsien Wu
Department of Computer Science and Information Engineering
National Cheng Kung University
Tainan 701, Taiwan, R.O.C.
{p7693136, pfhsieh, p7694163}@mail.ncku.edu.tw, and chwu@csie.ncku.edu.tw

ABSTRACT

This paper presents a vision-based system for tracking and recognizing the hand gestures that change from one shape to another for conveying a single sign gesture. The widely-used active contour model (ACM) often fails to track the objects with long knobs, such as hands with fingers because of the difficulty progressing into boundary concavities. The enhanced ACM is proposed to overcome the difficulty by combining the gradient vector flow (GVF) and the inflation pressure force as the external forces. The tracked hand shape contours were then characterized by affine-invariant Fourier descriptors for recognition. Transition probabilities between shape changes were included to improve the recognition. The gestures of Taiwanese sign language (TSL) were chosen for test due mainly to the finite set of well-defined hand shapes and trajectories. Experiments showed that the enhanced ACM was reliable in tracking changes in hand shapes and that the system was feasible for recognizing the shape-changing hand gestures.

1: INTRODUCTIONS

Hand gesture recognition has gained a lot of attentions in the fields of human-computer interaction. An extensive review of the existing techniques for interpretation of hand gestures has been presented in [1]. The interfaces of hand gesture recognition can be classified into the glove-based and vision-based interfaces. Data Glove has been used to develop gestures recognition of TSL to capture the flexion of 10 finger joints, the rolling of palm, and other 3-D motion information [2]. The glove-based gesture interfaces require the user to wear a cumbersome glove, which is not convenient for signers. Furthermore, the expensive mechanical gloves are limited on finger movements. These drawbacks make the mechanical gloves less than popular in spite of the high accuracy and the real-time capability. A vision-based system [3] that recognizes continuous hand gestures in a stationary background uses an extraction algorithm to trace the moving hand and to extract the hand region, uses the Fourier descriptors to characterize spatial features and the motion analysis to characterize the temporal features. They combine the two features as input for hidden Markov models to recognize the gestures. In [4], single-handed gestures that are sequences of distinct hand shapes have been considered. A given hand shape can undergo motion and discrete changes.

The contour information, which can be used to represent a hand shape, plays a crucial role in recognition of shape-changing hands. We adopted the snake model [5] to detect and track the contours of shape-changing hand gestures. Snakes are defined as curves that move under the influence of the internal forces coming from the curve itself and the external forces computed from the image data. The internal and the external forces are defined so that the snake will conform to an object boundary or other desired features within an image. There are two key difficulties with snakes. First, the initial contour must be close to the true boundary otherwise it will likely converge to a wrong result. Second, active contours can hardly converge into the deep concavities of the boundary. For solving the problems, we selected the GVF snakes [6][7] and an inflating pressure force [8][9] as our external forces to enhance the active contour model. When the contour is tracked, the series of points along the contour are used as inputs to the Fourier descriptors [10] to generate features for further recognition. The Fourier descriptors are suitable for hand contour representation because they are invariant with translation, rotation, and scaling transformation of the input shapes. The sign words in the category of shape changing hand gestures take a large proportion in TSL. Thanks to the finite set of well-defined hand shapes and trajectories in TSL, those sign words are ideal test data for the experiment about shape-changing gestures.

This paper is organized as follows. Section 2 introduces active contour models, including the traditional snakes, the GVF snakes, and the enhanced active contour model. Section 3 describes the overall recognition system. Section 4 presents the experimental results. Finally, Section 5 draws conclusions.

2: ACTIVE CONTOUR MODELS

2.1: Active Contour Model

A traditional snake is a curve $w(s) = [x(s), y(s)]$, $s \in [0,1]$, which moves through the spatial domain of an image to minimize the following energy functional,

$$E = \int_0^1 \frac{1}{2}[\alpha |w'(s)|^2 + \beta |w''(s)|^2] + E_{ext}(w(s))ds , \quad (1)$$

where α and β are weighting parameters that control the snake's tension and rigidity, respectively, and $w'(s)$ and $w''(s)$ denote the first and second derivatives of $w(s)$ with respect to *s*. The external energy function E_{ext} is derived from the image so that it takes on its smallest values at the features of interest, such as boundaries. Given a gray-level image *I*(*x*, *y*), which is viewed as a function of continuous position variables (*x*, *y*), the typical external energies that are designed to lead an active contour toward step edges [5] are

$$E_{ext}^{(1)}(x, y) = -|\nabla I(x, y)|^2 , \quad (2)$$

$$E_{ext}^{(2)}(x, y) = -|\nabla[G_\sigma(x, y) * I(x, y)]|^2 , \quad (3)$$

where ∇ is the gradient operator and $G_\sigma(x, y)$ is a two-dimensional Gaussian function with standard deviation σ. From the calculus of variations [11], the minimization of the snake function *E* can be found by solving the following Euler equations

$$\alpha w''(s) - \beta w''''(s) - \nabla E_{ext} = 0 , \quad (4)$$

where $w''''(s)$ is the fourth derivatives of $w(s)$. To find a solution to (4), the snake is made dynamic by treating $w(s)$ as a function of time *t*. The partial derivative of w with respect to *t* is then set equal to the left hand side of (4) as follows,

$$w_t(s,t) = \alpha w''(s,t) - \beta w''''(s,t) - \nabla E_{ext} . \quad (5)$$

When the solution $w(s, t)$ stabilizes, the term $w_t(s, t)$ vanishes and a solution of (4) is achieved.

2.2: Gradient Vector Flow (GVF) Snake

The GVF snake [6] is proposed to achieve better contour detection. The basic idea of the GVF snake is to extend the influence range of the image force to a larger area by generating a GVF field. The GVF field is computed from the image. In detail, a GVF field is defined as a vector field $\mathbf{v}(x, y) = [u(x, y), v(x, y)]$ for any pixel (*x*,*y*). The vector field **v**(*x*,*y*) is computed by minimizing the energy functional

$$\varepsilon = \iint \mu(u_x{}^2 + u_y{}^2 + v_x{}^2 + v_y{}^2) + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 \, dxdy , \quad (6)$$

where ∇f is the gradient of the edge map *f* derived from the input image *I*(*x*, *y*), μ is a regularization parameter. Using the calculus of variations [11], the GVF field can be obtained by solving the corresponding Euler equations.

Similar to (4), the force balance equation of the GVF snake can be expressed as

$$\alpha w''(s) - \beta w''''(s) + \mathbf{v} = 0 . \quad (7)$$

GVF snake's larger capture range and concavity tracking ability are attributed to the diffusion operation shown in (6).

By replacing both μ and $|\nabla f|^2$ in (6), the generalized GVF (GGVF) [7] is defined as the equilibrium solution of the following vector partial differential equation

$$\mathbf{v}_t = g(|\nabla f|)\nabla^2 \mathbf{v} - h(|\nabla f|)(\mathbf{v} - \nabla f) . \quad (8)$$

The weighting functions can be selected such that $g(\cdot)$ decreases as $h(\cdot)$ increases. The following weighting functions are used for GGVF.

$$g(|\nabla f|) = e^{-(|\nabla f|/K)} , \quad h(|\nabla f|) = 1 - g(|\nabla f|) . \quad (9)$$

Using this pair of weighting functions, the GGVF field conforms to the edge map gradient at strong edges but varies smoothly away from the boundaries.

We found that the GVF snake often failed to track intact hand shapes, especially those associated with fingers, in an image sequence.

2.3: Enhanced Active Contour Model

To solve the problem of tracking intact hand shapes with fingers, we included an outward pressure force as another external force in addition to the GVF snake. The pressure force is introduced from the balloon model [8][9]. In a balloon model, a pressure force $f_p(s)$ is added with the snake force as a second external force to push the curve outward or inward. The pressure force is represented as follow:

$$f_p(s) = k\mathbf{n}(s) , \quad (10)$$

where $\mathbf{n}(s)$ is the normal unit vector to the curve at pixel $w(s)$ and *k* is the amplitude of this force.

Combining an outward pressure force with the GVF snake, the enhanced force balance equation is thus given by

$$\alpha w''(s) - \beta w''''(s) + \phi \mathbf{v} + f_p(s) = 0 , \quad (11)$$

where **v** is the GVF external force, and f_p is the inflating force and ϕ is the weight.

3: RECOGNITION SYSTEM

The developed recognition system consists of four phases: skin detection, hand shape tracking, hand shape representation, and gestures recognition. Flow chart of the recognition system is shown in Fig. 1.

3.1: Skin Detection

Color is one of the most special clues in tracking objects. In this study, skin color was used to detect hands. Color is generally represented as RGB as captured from a video camera. Averaging RGB for each pixel in a color image produced a gray-level image. Besides, the color image was transformed from the RGB color space to the YIQ [12] color space. The I-component of YIQ was checked if it was equal to or greater than a preset threshold. Finally, skin pixels were detected by combining the gray-level image and the I-component image, as shown in Fig. 2.

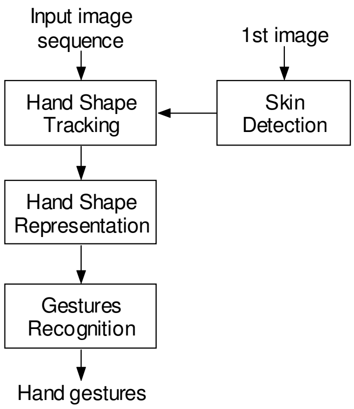

Fig. 1 Flow chart of the shape-changing hand gestures recognition system.

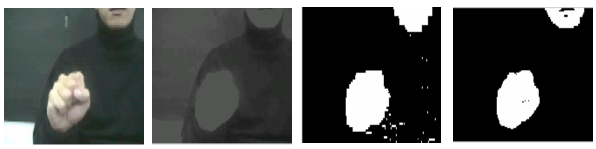

Fig. 2 (a) Original image (b) Gray-level image (c) I-component image (d) Final result image.

3.2: Hand Shape Tracking

The objective of the hand shape tracking phase is to extract from a video sequence the contours of deformable hand shapes in a two-dimensional image space. Each hand shape contour can be represented by a set V of point-tokens (x_i, y_i) as

$$V = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}, \quad (12)$$

where N is the number of the sampled points on the contour. Since different hand shape contours may have different numbers of the contour points and the coordinates may have overlaps or separation, we must keep the distance between the points constant by removing or inserting contour points, as shown in Fig. 3.

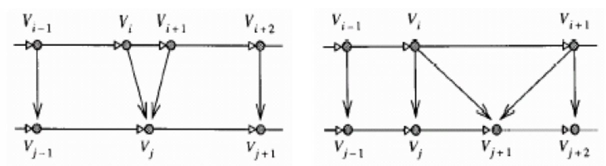

Fig. 3 (a) Remove and (b) insert contour points.

We proposed the enhanced ACM to achieve this goal. Fig. 4 shows that the enhanced ACM tracked fingers better than the GVF snake.

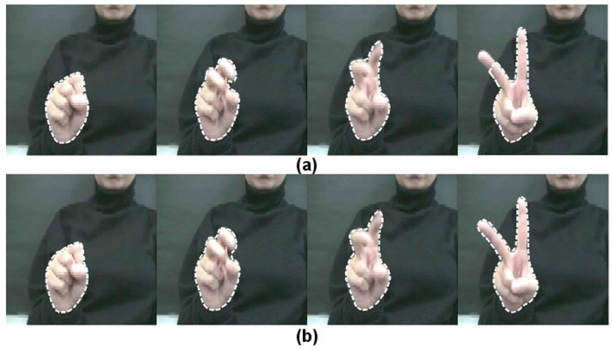

Fig. 4 Comparison of hand shape tracking using (a) the GVF snake and (b) the enhanced ACM. The middle finger was tracked by the enhanced ACM but missed by the GVF snake.

3.3: Hand Shape Representation

After tracking the hand shape contour in the image sequences, the hand shape representation was characterized by the hand shape contour nodes. Fourier descriptors provide a means for representing the boundary of a two-dimensional shape. In this phase, the Fourier descriptors transform the contour point-token set V that represent a hand shape to a set of normalized coefficients that are invariant under any affine transformation (translation, rotation, scaling, and shearing) [10].

3.4: Gestures Recognition

A shape-changing hand gesture is defined herein as a gesture that changes from a hand shape to another hand shape for a meaningful sign word. In the recognition phase, whether the hand shape changes is first determined by analyzing the variation in the video sequence. If a change occurs, the beginning shape and the final shape are taken into recognition. Transition probabilities between shape changes are incorporated into the process of recognition to increase the accuracy of recognition.

4: EXPERIMENTAL RESULTS

Our experiments consisted of two parts: tracking of hand shapes and recognition of shape-changing hand gestures.

4.1: Data Set

In experiments, we tested 20 TSL sign words [13] whose hand shapes were changed from one to another for a word. The selected sign words (shape-changing hand gestures) were related to 21 frequently used hand shapes, as shown in Fig. 5. To collect test samples, each sign word was performed three times by nine individuals, yielding a total of 27 video sequences of shape-changing hand gestures for each word. The video sequences were acquired from a digital video camera of 24-bit RGB colors and 160×120 resolution for the image segment with 30 frames per seconds (FPS).

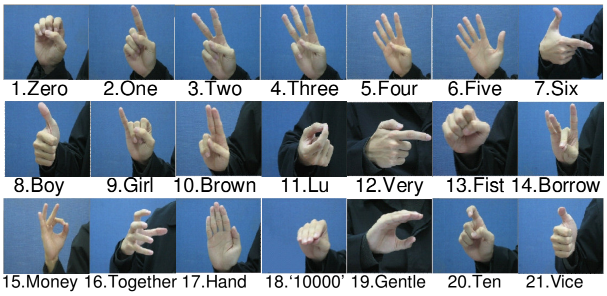

Fig. 5 Hand shapes related to the selected TSL sign words.

4.2: Result of Hand Shape Tracking

A video sequence that recorded the shape-changing hand gestures in TSL was an input to the developed system. The tracking performance was evaluated in terms of the Hausdorff distance [14] for the 21 hand shapes associated with selected shape-changing gestures. Fig. 6 shows that the enhanced ACM gave higher accuracy than the GVF snake especially for hand shapes 3-6 and 15. Note that those hand shapes are composed of extended fingers, making concavities along the contours. Fig. 7 demonstrates several examples of tracking shape-changing hand gestures in TSL using the enhanced ACM.

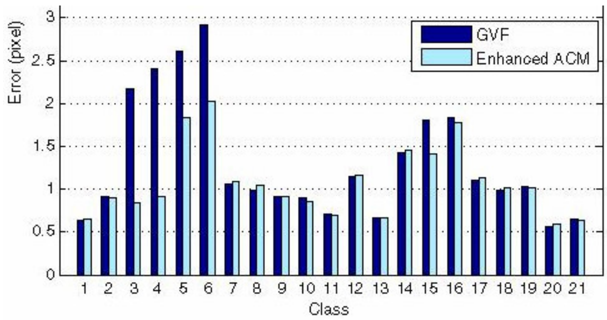

Fig. 6 Tracking performance comparison between the GVF snake and the enhanced ACM.

4.3: Result of Shape-Changing Hand Gestures Recognition

There were 594 training samples acquired from the static images for the 21 hand shapes in TSL and 540 testing image sequences in our experiment. Table 1 shows the recognition result for the shape-changing hand gestures using the GVF snake and the enhanced Active Contour Model, respectively. In the process of recognition, we included transition probabilities to improve the accuracy of gesture recognition.

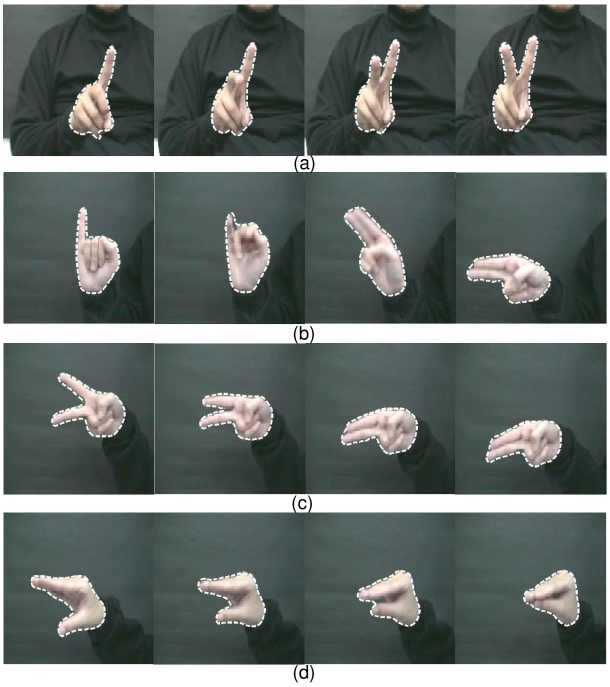

Fig. 7 Tracking results of TSL sign gestures (a) 'and', (b) 'feminine', (c) 'cut', and (d) 'message'.

Table 1 Result of shape-changing hand gestures recognition.

No.	Sign	Hand shape1→	Hand shape2	Acc (%) GVF	Acc (%) Enh. ACM
1	And	One	Two	56	74
2	Die	One	Ten	96	96
3	Nice	Six	Male	100	100
4	Masculine	Male	Brown	93	93
5	Thanks	Male	Vice	100	100
6	Feminine	Female	Brown	89	89
7	Increase	Lu	Very	100	100
8	Decrease	Very	Lu	100	100
9	Cut	Two	Brown	100	100
10	Kick	Borrow	Five	100	100
11	Have?	Hand	'10000'	100	100
12	Free	Money	Five	85	100
13	100	Fist	One	81	81
14	200	Fist	Two	63	85
15	300	Fist	Three	67	78
16	400	Fist	Four	63	93
17	Tell	Zero	Five	52	100
18	Catch	Five	Fist	96	96
19	Pain	Five	Together	100	100
20	Message	Gentle	'10000'	100	100
			Average	87.05	94.25

5: CONCLUSIONS

This study develops a vision-based system for recognizing shape-changing hand gestures in TSL. The enhanced active contour model that is proposed herein improved the performance of hand shape tracking. The system uses the Fourier descriptors to effectively represent hand shapes regardless of the variance in rotation, translation, and scaling. The system also takes into account the transition probabilities between shape changes for recognizing the shape-changing hand gestures more accurately. The experimental results demonstrated the feasibility of the system for recognizing shape-changing signs of TSL.

In the future, the difficulty of detecting hand shapes in the complex background needs to be overcome. The transition probabilities of changing shapes also need to be estimated and established for practical use. The investigation can be advanced to 3-D tracking and recognition to handle a larger lexicon of TSL.

Acknowledgments

The authors would like to thank the participators in gesture recording for their devotion to setting up the TSL lexicon regarding shape-changing sign gestures. Also, the authors would like to thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC94-2614-E-006-087.

REFERENCES

[1] V. I. Pavlovic, R.Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 677-695, July 1997.

[2] R. H. Liang and M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," *Proc. of IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 558-567, Apr. 1998.

[3] F. S. Chen, C. M. Fu, and C. L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Journal of Image and Vision Computing,* vol. 1, pp. 745-758, Mar. 2003.

[4] A. Ramamoorthy, N. Vaswani, S. Chaudhury, and S. Banerjee, "Recognition of dynamic hand gestures," *Pattern Recognition,* vol. 36, pp. 2069-2081, 2003.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int'l J. Computer Vision*, vol. 1, pp. 321-331, 1988.

[6] C. Xu and J. L. Prince, "Snakes, shapes and gradient vector flow," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 359-369, Mar. 1998.

[7] C. Xu and J. L. Prince, "Generalized gradient vector flow external forces for active contours," *Signal Processing*, vol. 71, pp. 131-139, 1998.

[8] L. D. Cohen, "On active contour models and balloons," *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 221-228, Mar. 1991.

[9] L. D. Cohen, and I. Cohen, "Finite-element methods for active contour models and balloons for 2D and 3D images," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 15, no. 11, pp. 1131-1147, 1993.

[10] K. Arbter, W. E. Snyder, H. Burkhardt and G. Hirzinger, "Application of affine-invariant Fourier descriptors to recognition of 3-D objects," *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 12, no. 7, July 1990.

[11] P. V. O'Neil, *Advanced Engineering Mathematics*, 3rd ed., Wadsworth publishing, pp. 510~535, 1991.

[12] J. D. Foley et al., *Computer Graphics*, Addition Wesley, 1992.

[13] 史文漢暨丁立芬, *手能生橋*, 第一冊~第二冊, 中華民國聾人協會發行, 2004.

[14] D.P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 9, pp. 850-863, Sept. 1993.