

# A Comparative Study of Region Matching Based on Shape Descriptors for Coloring Hand-drawn Animation

Yoshihiro Kanamori  
University of Tsukuba  
Email: kanamori@cs.tsukuba.ac.jp

**Abstract**—The work of coloring hand-drawn animation is done by manually specifying and painting each closed region in line drawings. To make this process more efficient, this research creates associations between closed regions in line drawings of adjacent frames, which need to be colored with the same color. Feature values are first computed from the shape of each closed region, and a cost of associating pairs of closed regions is computed from these feature values. The combination of associations that minimizes the total cost is then computed based on these costs. Three shape descriptors for computing feature values for each closed region are examined: ellipses, Fourier descriptors, and shape context; and the accuracy of making associations using each of them is studied.

## I. INTRODUCTION

Since the advent of hand-drawn 2D animation, people have been fascinated with it. However, even now with the appearance of production-support software, production of hand-drawn animation is very labor-intensive and requires many hours of work. This production includes the following processes: drawing line drawings on paper and scanning them for each frame, cleaning up the scanned line drawings, and manually selecting and coloring each of the closed regions in the line drawings. The work of coloring regions holds promise for automation because it is simpler than other processes.

To automate the coloring process, we find associations between closed regions in adjacent frames of the animation, and then can propagate colors to corresponding regions in adjacent frames. Most existing such methods use feature values derived from characteristic points on the outlines of closed regions to find associations between closed regions [1], [2]. Examples of the characteristic points include intersection points and points where the curvature changes significantly. However, such characteristic points on outlines are quite susceptible to changes in shape due to motion in the animation, and this reduces the accuracy of finding associations. In this research, the following three shape descriptors, which are less susceptible to changes in outline shape, are compared.

- 1) Ellipses roughly approximating the shapes of closed regions [3],
- 2) Fourier descriptors expressing low-frequency components of closed-region shape [4], and
- 3) Shape context, representing closed region shape in polar coordinates and then as a 2D histogram of log-distance and angle [5].

Of these, the method in [3] was originally proposed for associating closed regions in line drawings of hand-drawn animation. To associate closed regions, we should consider the position and scale of each region to differentiate the region from others within the same frame. On the other hand, Fourier descriptors [4] and shape context [5] were developed to find shapes similar to a single input shape, such as a logo or road marker, so they provide for invariance under transformations such as translation, scaling and rotation. We thus modify these descriptors to take into account geometric transformations in finding associations for the purposes of this research. The cost of associating two closed regions is derived using these shape descriptors, and the best matching of closed regions between two adjacent frames is computed using bipartite graph matching.

This paper reports on the results of experiments comparing the accuracy of these methods using real, hand-drawn animation image data.

## II. RELATED WORK

A simple method for painting closed regions on multiple frames all at once is to use an “Onion fill tool”. With this function multiple frames are stacked and all of the closed regions that include a coordinate specified by the user are colored with the same color. Such tools are not often used in practice, because the multiple closed regions to be painted must be overlapping, and it is not easy to specifying how to paint correctly in one try.

Existing methods make links between closed regions using outline shapes [1], [2]. Examples of outline shape descriptors used include coding the outline shape as a text string, or using an array of so-called *dominant points*. However, these methods are not effective when the shape of the outline changes significantly.

Another method uses the skeletal structure of objects in the line drawing to find associations between closed regions [6], but skeletal structure can only be applied to articulated characters that have joints.

Sýkora et al. [7] proposed a method of applying color and texture to hand-drawn animation by matching the positions of images between successive frames. They also proposed *LazyBrush* [8], which they used for coloration. However, their method uses rigid shape matching [9], so it requires several seconds to match positions for each successive frame and has

difficulty find associations when shapes change in a non-rigid manner.

Generally, there has been much research in the computer vision field on feature values used for finding such associations in natural images. However, the images used in this research are line drawings, which do not contain the kind of shading information found in natural images, and only shape feature values obtainable from line drawings are considered applicable. Details on shape feature values can be found in references such as [10]. This research compares the accuracy of finding associations using shape feature values computed from ellipse-based shape descriptors proposed recently [3], from Fourier descriptors [4], and from shape context [5].

### III. BASIC PROCESS OF ATTACHING CORRESPONDENCES

Regarding the input, line drawings are of course required, but with only the line drawings, there is no way to check if the coloring was done correctly. Thus, we take colored images as input and perform a region partitioning by color to obtain both the correct coloration data and the closed-region data, assuming that each closed region is colored with a single color. Line drawings are extracted as the lines marking the outlines of the regions (mainly drawn in black) in the input images.

This research compares three shape descriptors within the same framework. After region partitioning, associations linking pairs of closed regions in successive frames are calculated. Once these associations have been made for all frames, they are used to create chains of closed regions between frames. When a color is assigned to a single closed region, the color is propagated to adjacent frames by following these chains of associations.

The following describes in detail how these associations are made. Let  $N_f$  and  $N_{f+1}$  denote the number of closed regions in each of two successive frames,  $f$  and  $f+1$ . A cost matrix,  $A = \{a_{ij}\}$  is computed. Each element,  $a_{ij}$ , of this cost matrix is the cost of associating the closed region  $i$  in frame  $f$  with the closed region  $j$  in the frame  $f+1$  (Note  $i = 1, 2, \dots, N_f$  and  $j = 1, 2, \dots, N_{f+1}$ ). For  $A$  obtained in this way, the Hungarian algorithm [11] is used to solve for a bipartite graph matching. If the numbers of closed regions in the two frames are not the same (i.e.,  $N_f \neq N_{f+1}$ ), then one or more closed regions are left without associations. This is understood to mean that a closed region has newly appeared or disappeared in one of the frames.

This research compares methods that compute the elements,  $a_{ij}$ , in the cost matrix,  $A = \{a_{ij}\}$ , using three different shape descriptors. These shape descriptors are described below.

### IV. ELLIPTICAL DESCRIPTOR

The method in reference [3] approximates each closed region using an ellipse (Fig. 1). The ellipse  $i$  approximating closed region  $i$  is computed from the centroid,  $\mathbf{t}_i$ , and covariance matrix,  $C_i$  of the positions of each pixel in closed region  $i$ . Here, the largest and smallest eigenvalues of the covariance matrix,  $C_i$ , are designated  $\lambda_i^{max}$  and  $\lambda_i^{min}$ , and their associated eigenvectors are  $\mathbf{e}_i^{max}$  and  $\mathbf{e}_i^{min}$  (assumed to be unit vectors), respectively. Then the major and minor axes of ellipse  $i$  are given by  $\sqrt{\lambda_i^{max}}\mathbf{e}_i^{max}$  and  $\sqrt{\lambda_i^{min}}\mathbf{e}_i^{min}$  respectively.

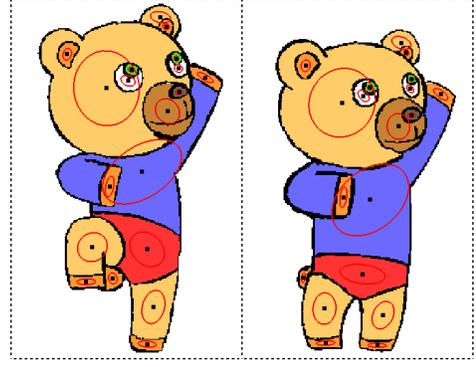


Fig. 1. Elliptic shape descriptors. In the method from reference [3], associations between closed regions in consecutive frames are made based on ellipses (red) approximating each closed region.

The association cost,  $a_{ij}$ , is derived from the orientation, size and position of the ellipses approximating the two closed regions,  $i$  and  $j$ , as follows.

$$a_{ij} = w_{angle} a_{ij}^{angle} + w_{scale} a_{ij}^{scale} + w_{pos} a_{ij}^{pos}, \quad (1)$$

$$a_{ij}^{angle} = \cos^{-1}(\mathbf{e}_i^{max} \cdot \mathbf{e}_j^{max}), \quad (2)$$

$$a_{ij}^{scale} = \sqrt{\frac{\max\{\lambda_i^{max}, \lambda_j^{max}\}}{\min\{\lambda_i^{max}, \lambda_j^{max}\}}} + \sqrt{\frac{\max\{\lambda_i^{min}, \lambda_j^{min}\}}{\min\{\lambda_i^{min}, \lambda_j^{min}\}}} - 2, \quad (3)$$

$$a_{ij}^{pos} = \|\mathbf{t}_i - \mathbf{t}_j\|^2, \quad (4)$$

where  $w_{angle}$ ,  $w_{scale}$ , and  $w_{pos}$  are weightings. Coordinates  $\mathbf{t}_i$  and  $\mathbf{t}_j$  are normalized by the image size. The values of  $a_{ij}^{angle}$ ,  $a_{ij}^{scale}$ , and  $a_{ij}^{pos}$  each approach zero if the orientation, size, and position respectively, of the ellipses approximating closed regions  $i$  and  $j$  approach each other. To increase the accuracy of associations further, information regarding regions adjacent to regions  $i$  and  $j$  is also added to cost  $a_{ij}$ . This is done by computing the centroid of the regions adjacent to each of regions  $i$  and  $j$  and adding the distances they move the association cost. The centroid,  $\mathbf{n}_i$ , of the regions adjacent to closed region  $i$  is computed as follows.

$$\mathbf{n}_i = \frac{\sum_{k \in \{O_i, i\}} \alpha_k \beta_k \exp\left(-\frac{\|\mathbf{t}_k - \mathbf{t}_i\|^2}{\sigma_i^2}\right) \mathbf{t}_k}{\sum_{k \in \{O_i, i\}} \alpha_k \beta_k \exp\left(-\frac{\|\mathbf{t}_k - \mathbf{t}_i\|^2}{\sigma_i^2}\right)}, \quad (5)$$

$$\sigma_i = \sigma_0 \max_{k \in O_i} \|\mathbf{t}_k - \mathbf{t}_i\|, \quad (6)$$

where  $O_i$  is the set of closed regions adjacent to closed region  $i$ .  $\alpha_k$  is the area of closed region  $k$ . If the area of closed region  $k$  is large, then region  $k$  has a greater effect on the position of the centroid. However, if region  $k$  has only small contact with region  $i$ , it is desirable to give region  $k$  a smaller weighting so that the centroid position is not overly affected by movement of region  $k$ . For this reason, we multiply the weighting by a contact rate,  $\beta_k \in [0, 1]$ . The contact ration is a value expressing what proportion of the total length of the outline of region  $i$  is in contact with region  $k$ . Further, background regions tend to be large, so if region  $k$  is part of the background, the computation is done setting  $\alpha_k = \alpha_i$  and  $\mathbf{t}_k = \mathbf{t}_i$  to reduce its effect. The effects of closed region  $k$  also must be reduced if the position of its centroid is far from that of closed region  $i$ , so the distance between the centroids of the two regions is also used as a weighting, with  $\sigma_0$  as

the coefficient. Finally, the association cost  $a_{ij}$  is computed by adding the term  $a_{ij}^{neighbor}$  multiplied by the weighting  $w_{neighbor}$ , as in the following equations.

$$a_{ij}^{neighbor} = \|\mathbf{n}_i - \mathbf{n}_j\|^2, \quad (7)$$

$$a_{ij} = w_{angle} a_{ij}^{angle} + w_{scale} a_{ij}^{scale} + w_{pos} a_{ij}^{pos} + w_{neighbor} a_{ij}^{neighbor}. \quad (8)$$

## V. FOURIER DESCRIPTORS

There are many variations of Fourier descriptors. Reference [4] reports on experiments using the following three Fourier descriptors to search for similar shapes.

- Centroid distance: computes the distance between the shape centroid and a point moving along the outline.
- Changes in curvature: computes changes in the curvature of the outline.
- Cumulative angle: accumulates the angle between the shape outline and a tangent.

It finds that the centroid distance produces the best results, so this research uses a Fourier descriptor based on centroid distance.

The Fourier descriptor based on centroid distance from reference [4] is described as follows. The shape outline is first converted to a polygonal curve line with  $N$  segments of equal length. The distance from each vertex,  $t$  ( $t = 0, 1, \dots, N-1$ ), to the centroid is  $r_t$ . Then the  $n$ -th Fourier coefficient,  $u_n$  ( $n = 0, 1, \dots, N-1$ ) is computed as follows.

$$u_n = \frac{1}{N} \sum_{t=0}^{N-1} r_t \exp\left(\frac{-j2\pi nt}{N}\right), \quad (9)$$

where  $j$  is the imaginary unit. The DC component,  $u_0$ , has only a real part, and depends on the scale of the shape. In reference [4], the following shape feature vector,  $\mathbf{f}$ , is used to provide invariance over translation, scaling, and rotation.

$$\mathbf{f} = \left( \frac{|u_1|}{|u_0|}, \frac{|u_2|}{|u_0|}, \dots, \frac{|u_{N-1}|}{|u_0|} \right). \quad (10)$$

When comparing shapes, the initial  $K$  ( $0 \leq K \leq N$ ) elements of the feature vector  $\mathbf{f}$  are selected, and the Euclidean distance between them is calculated as a degree of similarity.

*Computation used in this research:* This research also computes the distance between centroids of two closed regions to take their relative positions into consideration. The terms in Equation (10) are divided by  $|u_0|$  to achieve scaling invariance, but in this research, relative size is also a consideration, so this division is omitted. As a result, the feature vector,  $\mathbf{f}_i$ , for closed region  $i$  is computed as follows.

$$\mathbf{f}_i = (|u_0|, |u_1|, \dots, |u_{K-1}|), \quad (11)$$

This is the vector of absolute values of  $K$  Fourier coefficients. The association cost,  $a_{ij}$  for closed regions  $i$  and  $j$  is computed as follows.

$$a_{ij} = w_{trans}^{fourier} \|\mathbf{t}_i - \mathbf{t}_j\| + \|\mathbf{f}_i - \mathbf{f}_j\|, \quad (12)$$

where  $\mathbf{t}_i$  and  $\mathbf{t}_j$  are the centroids of closed regions  $i$  and  $j$  (normalized by the image size), and  $w_{trans}^{fourier}$  is a weighting

value. The parameter values used were  $N = 64$ ,  $K = 4$ , the same as in reference [4], and  $w_{trans}^{fourier} = 0.5$ .

## VI. SHAPE CONTEXT

Shape context [5] is a two-dimensional histogram of the positions of points relative to one point, in terms of log-distance and angle. For invariance to scaling, each distance is divided by the average distance before computing the log distance. Using log distance captures the distribution in more detail close to the origin, and in less detail farther away. Log distances and angles are divided into  $N_{dist}$  and  $N_{rad}$  partitions respectively, producing a histogram with a total of  $N_{bin} = N_{dist}N_{rad}$  bins. To compute the similarity,  $C_{ij}$ , between two histograms,  $i$  and  $j$ , a  $\chi^2$  is used.

$$C_{ij} = \frac{1}{2} \sum_{b=1}^{N_{bin}} \frac{|h_i(b) - h_j(b)|^2}{h_i(b) + h_j(b)}, \quad (13)$$

where  $h_i(b)$  and  $h_j(b)$  are the values in the  $b$ -th bins of the normalized histograms for  $i$  and  $j$ .

*Computation used in this research:* As with the Fourier descriptor case, this research also takes into consideration the distances between centroids of closed regions and changes in scale, so distances from the centroid for each region are not divided by the average distance. The association cost,  $a_{ij}$ , between closed regions  $i$  and  $j$  is calculated by the following equation.

$$a_{ij} = w_{trans}^{shape} \|\mathbf{t}_i - \mathbf{t}_j\| + C_{ij}, \quad (14)$$

where  $w_{trans}^{shape}$  is a weighting factor.

Parameter values used were number of log-distance partitions,  $N_{dist} = 5$ , number of angle partitions,  $N_{rad} = 12$ , the same as in reference [5], and  $w_{trans}^{shape} = 4$ .

## VII. EXPERIMENTAL RESULTS

The comparison program was implemented in C++ using OpenGL, GLUT and GLUI and was run on a notebook PC with a 2.3 GHz Intel Core i7-2820QM CPU and 3.2 GB of memory. The coloring for each region obtained from already colored images was used as correct data. In experiments, associations were first created using each of the shape descriptors. Color was then applied to a single frame, propagated automatically to adjacent frames, and the extent to which color was applied correctly was evaluated quantitatively. The frame with the most closed regions was selected as the frame to begin coloring. When new closed regions appear in a frame, those regions remain without associations, so a frame with as many regions as possible was selected to minimize the number of regions left without color.

Experimental results are shown in Figs. 3 and 4. The images are trimmed to accommodate available space. Associations were applied using the ellipse, Fourier, and shape context descriptors. Closed regions that were assigned the wrong color are shown in red, and regions that had no association and were not assigned a color are shown in green. The numbers of errors and unassociated regions is shown. For these results, the correctness rate is defined as the proportion of the number closed regions that are colored correctly, and this is graphed in Fig. 2. For each of Figs. 3 and 4, the Fourier and shape

TABLE I. COMPUTATION TIMES (S) FOR MAKING ASSOCIATIONS FOR DIFFERENT IMAGE SIZES AND NUMBERS OF FRAMES. COMPUTATION TIMES INDICATE TIME REQUIRED TO MAKE ASSOCIATIONS FOR ALL FRAMES.

Fig. (Image size, no. of frames)	Elliptic	Fourier descriptor	Shape context
Fig. 3 (3614×2053, 14 images)	0.11	0.056	0.032
Fig. 4 (3614×2053, 57 images)	0.51	0.32	0.19

context descriptors produced slightly higher accuracy than the ellipse descriptor.

Computation times are shown in Table I. Region partitioning was implemented with the same process for all three methods and required 2.3 s and 8.1 s respectively for each of the figures. Although the time required for region partitioning increased with the image size, making the associations was very fast in all cases. Results also showed that the Fourier descriptor and shape context were both faster than the ellipse method .

Considering the graphs in Fig. 2, all of the results showed a rapid drop in correctness in the frames before and after the frame to which color was applied, which had the most closed regions. The frames with the most regions tend to be cut up by objects like hair, hands, or arms, which increase the number of closed regions. Such multiple closed regions caused by screening should really be handled by making the association with a single closed region before partitioning. In other words, one-to-many associations are needed for closed regions. However, the current framework with bipartite graph matching cannot handle one-to-many associations and leaves closed regions without associations. This results in a drop in colorization accuracy. This result shows that making one-to-one associations between closed regions a limitation in the approach itself.

## VIII. CONCLUSION AND FUTURE ISSUES

This research has studied a method toward automating the work of coloring frames in the current production of 2D hand-drawn animations. Three shape descriptors for closed regions were used, based on ellipses, Fourier descriptors, and shape context. The cost of making associations between regions based on each of these was computed, and the associations were made using bipartite graph matching. Experimental results showed that both the Fourier and shape context descriptors produced better results than the ellipse descriptor, both for accuracy and computation speed. However, the current

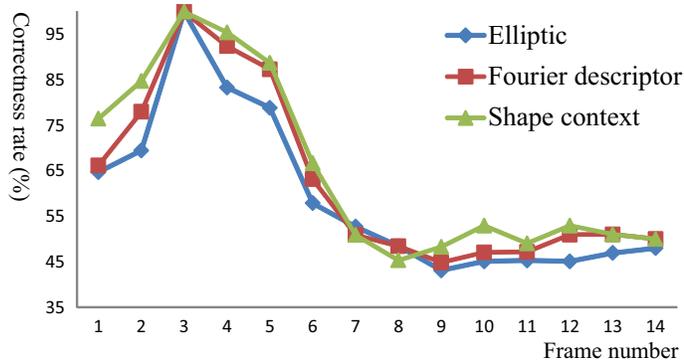
methods only support one-to-one associations between closed regions, so cases requiring one-to-many associations, in which new closed regions appear due to screening, cannot be handled. Accuracy drops quickly for cases when such screening occurs often. In the future, methods that handle the effects of such screening well need to be developed.

## ACKNOWLEDGMENT

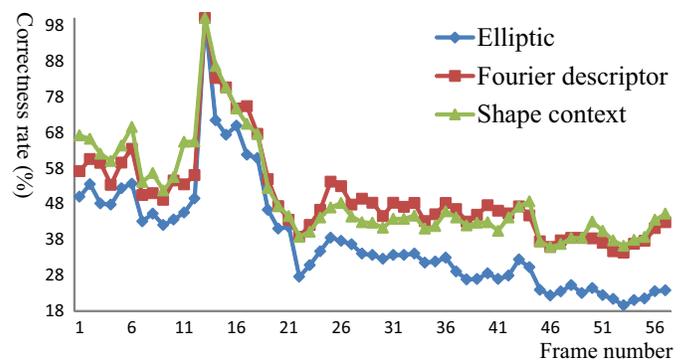
The author would like to express sincere gratitude to the animation company that provided the real animation data used in this paper. The copyright of the image data is reserved by the animation company.

## REFERENCES

- [1] J. S. Madeira, A. Stork, and M. H. Gross, "An approach to computer-supported cartooning," *The Visual Computer*, vol. 12, no. 1, pp. 1–17, 1996.
- [2] P. Garcia Trigo, H. Johan, T. Imagire, and T. Nishita, "Interactive region matching for 2D animation coloring based on feature's variation," *IEICE Transactions (E92-D)*, no. 6, pp. 1289–1295, 2009.
- [3] Y. Kanamori, "Region matching with proxy ellipses for coloring hand-drawn animations," in *SIGGRAPH Asia 2012 Technical Briefs*, ser. SA '12, 2012, pp. 4:1–4:4. [Online]. Available: <http://doi.acm.org/10.1145/2407746.2407750>
- [4] D. Zhang and G. Lu, "A comparative study of fourier descriptors for shape representation and retrieval," in *Proc. of 5th Asian Conference on Computer Vision (ACCV)*. Springer, 2002, pp. 646–651.
- [5] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
- [6] J. Qiu, H. S. Seah, and F. Tian, "Auto coloring with character registration," in *Proceedings of the 2006 international conference on Game research and development*, ser. CyberGames '06, 2006, pp. 25–32.
- [7] D. Sýkora, M. Ben-Chen, M. Čadík, B. Whited, and M. Simmons, "Textoons: practical texture mapping for hand-drawn cartoon animations," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, ser. NPAR '11, 2011, pp. 75–84. [Online]. Available: <http://doi.acm.org/10.1145/2024676.2024689>
- [8] D. Sýkora, J. Dingliana, and S. Collins, "LazyBrush: Flexible painting tool for hand-drawn cartoons," *Computer Graphics Forum*, vol. 28, no. 2, pp. 599–608, 2009.
- [9] D. Sýkora, J. Dingliana, and S. Collins, "As-rigid-as-possible image registration for hand-drawn cartoon animations," in *Proceedings of the 7th International Symposium on Non-Photorealistic Animation and Rendering*, ser. NPAR '09, 2009, pp. 25–33. [Online]. Available: <http://doi.acm.org/10.1145/1572614.1572619>
- [10] M. Yang, K. Kpalma, and J. Ronsin, "A Survey of Shape Feature Extraction Techniques," in *Pattern Recognition*, 2008, pp. 43–90.
- [11] C. Papadimitriou and K. Stieglitz, *Combinatorial optimization : algorithms and complexity*. Prentice-Hall, 1982.



(a) Correctness rates for Fig. 3



(b) Correctness rates for Fig. 4

Fig. 2. Graphs of correctness rates (%) for the examples of automatic coloring in Figs. 3 and 4. The frames with the most closed regions in each were the 3-rd, and 13-th, respectively, so the correctness rate for those frames is 100%.

	1-st frame	2-nd frame	4-th frame	14-th frame
Input images				
	68 regions	59 regions	66 regions	50 regions
Elliptic				
	23 errors (13 unassociated)	15 errors (4 unassociated)	13 errors (4 unassociated)	28 errors (3 unassociated)
Fourier descriptor				
	24 errors (4 unassociated)	13 errors (0 unassociated)	5 errors (0 unassociated)	26 errors (0 unassociated)
Shape context				
	17 errors (8 unassociated)	9 errors (0 unassociated)	3 errors (0 unassociated)	26 errors (2 unassociated)

Fig. 3. Results of experiments using a 14-frame sequence. Red indicates closed regions where an incorrect color was assigned, and green indicates closed regions where no association was made, so no color was assigned. Color was assigned to the 3-rd frame, and the results from the 1-st, 2-nd, 4-th, and 14-th frames only are shown here. ©2006 Nibariki/GNDHDDT.

	1-st frame	11-th frame	12-th frame	14-th frame	15-th frame	57-th frame
Input images						
	70 regions	101 regions	107 regions	119 regions	98 regions	84 regions
Elliptic						
	40 errors (0 unassociated)	11 errors (0 unassociated)	56 errors (0 unassociated)	24 errors (0 unassociated)	29 errors (0 unassociated)	57 errors (11 unassociated)
Fourier descriptor						
	30 errors (0 unassociated)	47 errors (0 unassociated)	47 errors (0 unassociated)	20 errors (0 unassociated)	19 errors (0 unassociated)	48 errors (5 unassociated)
Shape context						
	22 errors (2 unassociated)	32 errors (0 unassociated)	35 errors (0 unassociated)	16 errors (0 unassociated)	19 errors (0 unassociated)	49 errors (27 unassociated)

Fig. 4. Results of experiments using a 57-frame sequence. Red indicates closed regions where an incorrect color was assigned, and green indicates closed regions where no association was made, so no color was assigned. Color was assigned to the 13-th frame, and the results from the 1-st, 11-th, 12-th, 14-th, 15-th, and 57-th frames only are shown here. ©2010 GNDHDDTW.