Proceedings of the 8th International Symposium on Mathematical Morphology,
Rio de Janeiro, Brazil, Oct. 10–13, 2007, MCT/INPE, v. 2, p. 61–62.
http://urlib.net/dpi.inpe.br/ismm@80/2007/06.15.14.12

Rotation, scale and translation-invariant segmentation-free grayscale shape recognition using mathematical morphology

Sidnei Alves de Araujo and Hae Yong Kim

Universidade de São Paulo (USP), Brazil
{saraujo,hae}@lps.usp.br

1. Introduction

Rotation, scale and translation-invariant grayscale shape recognition is an important problem in computer vision. However, to our knowledge, all practical techniques in the literature first "simplifies" the image using operations like segmentation/binarization and detection of edges and corner points. Some approaches that achieve RST-invariance by detecting interest points and edges include: curvature scale space [1], orientation code histograms [2], geometric hashing [3], and generalized Hough Transform [4]. Other techniques, like [5], first binarizes the image, isolates the shapes, normalizes their area and extracts some RST-invariant features. The most commonly used RST-invariant features are Hu's moments [6] and Zernike's moments [7]. However, these "image simplifying operations" throw away the rich grayscale information, are noise-sensitive and prone to errors. Thus, a segmentation-free approach is attractive. Segmentation-free approaches were proposed to recognize printed character [8] and handwritten numeral string [9], but they are not RST-invariant. Mathematical morphology has been used successfully in many works related to shape recognition, for example [10]. In this paper, we present a RST-invariant, segmentation-free gray-level shape recognition method using mathematical morphology approach. It is composed of three steps. In the first and second steps, filters based on dilations and erosions prune out the pixels that have no chance of matching the query shape. The third step makes use of the conventional template matching to recognize the query shape.

2. Algorithm Description

Let A be the grayscale image to be analyzed and Q the query grayscale template. The problem is to find all occurrences of Q in A. However, the instances of Q in A may be rotated, scaled, translated and with diverse brightnesses and contrasts. Our approach consists of three cascaded steps of filtering. Each filtering successively excludes pixels that have no chance of matching the query pattern. In order to measure the similarity between two grayscale images independent of brightness/contrast, we use the correlation coefficient with some modification to force that shapes with very low contrast yield no correlation. Let x and y be vectors of same size obtained from the query shape Q and any subimage of A, respectively. Then contrast-aware correlation coefficient is defined:

$$Corr(x,y) = \begin{cases} 0, \text{if } |\beta| \leq t_\beta \text{ or } 1/t_\beta \leq |\beta| \\ \frac{\tilde{x}\tilde{y}}{\|\tilde{x}\|\|\tilde{y}\|} = \frac{\beta\tilde{x}^2}{\|\tilde{x}\|\|\tilde{y}\|}, \text{otherwise} \end{cases}$$

where $\beta = \tilde{x}\tilde{y}/\tilde{x}^2$ is the contrast correction factor, $0 < t_\beta \leq 1$ is a contrast threshold, $\tilde{x} = x - \overline{x}$ is the mean-corrected vector, and $\overline{x}$ is the mean of x. Similar definitions are applicable to y.

In this paper, we define the dilation and erosion of a grayscale image I by a flat structuring element B with domain $D(B)$ as:

$$(I \oplus B)(s,t) = max_{(x,y)\in D(B)}\{I(s+x, t+y)\}$$
$$(I \ominus B)(s,t) = min_{(x,y)\in D(B)}\{I(s+x, t+y)\}$$

Note that, differently from the usual, our dilation definition does not include the reflection of the structuring element.

In the first step of the filtering, gray-level morphological dilations and erosions by flat annular structuring elements (Figure 1a) are used to determine a set of pixels of A (called "first grade candidate pixels") that have chance of matching Q. For each candidate pixel, the probable scale factor is also computed. Given A and a set of l radii $\{r_0, r_1, ..., r_{l-1}\}$, we build a 3D image $C_A[x,y,k]$, $0 \leq k < 2l$:

$$C_A[x,y,k] = \begin{cases} (A \oplus B_{r_{k/2}})(x,y), \text{if } k \text{ is even} \\ (A \ominus B_{r_{(k-1)/2}})(x,y), \text{if } k \text{ is odd} \end{cases}$$

where B_r is the flat annular structuring element with radius r. In our experiments we have used $l = 13$, and the set of radii $\{0, 2, ..., 24\}$ pixels. Given the query shape Q and a set of n scales $\{s_0, s_1, ..., s_{n-1}\}$, Q is resized to each scale factor, yielding the resized queries $Q_0, Q_1, ..., Q_{n-1}$. We build a matrix C_Q with n rows (scales) and $2l$ columns (radii):

$$C_Q[i,k] = \begin{cases} (Q_i \oplus B_{r_{k/2}})(x_0,y_0), \text{if } k \text{ is even} \\ (Q_i \ominus B_{r_{(k-1)/2}})(x_0,y_0), \text{if } k \text{ is odd} \end{cases}$$

where (x_0, y_0) is the central pixel of Q and $0 \leq i < n$, $0 \leq k < 2l$.

The matrices C_Q and C_A and a contrast threshold t_β are used to detect the correlation $CisCorr$ at the best matching scale for each pixel (x,y):

Proceedings of the 8th International Symposium on Mathematical Morphology,
Rio de Janeiro, Brazil, Oct. 10–13, 2007, MCT/INPE, v. 2, p. 61–62.
http://urlib.net/dpi.inpe.br/ismm@80/2007/06.15.14.12

$$CisCorr_{A,Q}(x,y) = MAX_{i=0}^{n-1}[|Corr(C_Q[i], C_A[x,y])|]$$

A pixel (x,y) is classified as a first grade candidate pixel if $CisCorr_{A,Q}(x,y) \geq t_1$, for some threshold t_1. The probable scale of (x,y) is s_i, where i is the argument that maximizes $CisCorr$.

The second step upgrades some of the first grade candidate pixels to second grade, while discards those that are not upgraded. For each second grade candidate pixel, the probable rotation angle is estimated. This filtering uses flat structuring elements disposed on radial lines (Figure 1a). Given Q and the set of m angle inclinations (in our example, $\alpha_0 = 0, \alpha_1 = 10, ..., \alpha_{m-1} = 350$) a vector R_Q with $2m$ elements is built:

$$R_Q[j] = \begin{cases} (Q \oplus B_{\alpha_{j/2}}^{r_{l-1}})(x_0, y_0), \text{if } j \text{ is even} \\ (Q \ominus B_{\alpha_{(j-1)/2}}^{r_{l-1}})(x_0, y_0), \text{if } j \text{ is odd} \end{cases}$$

where (x_0, y_0) is the central pixel of Q, $B_\alpha^{r_{l-1}}$ is the flat radial structuring element with inclination α and length r_{l-1} (the largest sampling circle radius) and $0 \leq j < 2m$. For each first grade candidate pixel (x,y), the vector R_A is computed:

$$R_A[x,y,j] = \begin{cases} (A \oplus B_{\alpha_{j/2}}^{\lambda})(x,y), \text{if } j \text{ is even} \\ (A \ominus B_{\alpha_{(j-1)/2}}^{\lambda})(x,y), \text{if } j \text{ is odd} \end{cases}$$

where $\lambda = s_i r_{l-1}$ is the largest radius resized to the probable scale s_i of pixel (x,y), and $0 \leq j < 2m$. The vectors $R_A[x,y]$, R_Q and a contrast threshold t_β are used to detect the correlation $RasCorr$ at the best matching rotation angle:

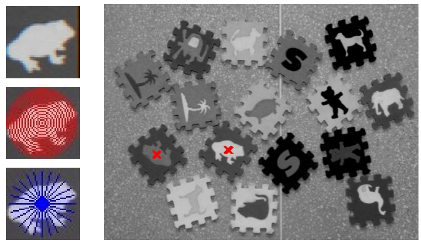
$$RasCorr_{A,Q}(x,y) = MAX_{j=0}^{2m-1}[|Corr(R_A[x,y], cshift_j(R_Q))|]$$

where "$cshift_j$" denotes circular shifting j positions of the argument vector. A first grade pixel (x,y) is upgraded to second grade if $RasCorr_{A,Q}(x,y) \geq t_2$. The probable rotation angle of each candidate pixel (x,y) is α_j where the argument j maximizes $RasCorr$.

In the third step, the second grade candidate pixels are filtered using a conventional template matching. This task is easy because the probable scale and angle for each candidate pixel are known. If the absolute value of contrast-aware correlation coefficient between the query shape Q and the image A at position (x,y) is larger than a threshold t_3, the query shape is considered to be found.

3. Experimental Results

Some experiments were executed to evaluate the performance of the proposed method. Figure 1 depicts one of them. We have applied the method to detect three query shapes (frog, bear and mouse) in three images where the query shapes appear in different rotations and scales. All shapes were correctly



(a) (b)

Figure 1. Detection of the frog shape. (a) Query shape Q (51×51 pixels), Annular structuring elements and Radial structuring elements. (b) Analyzed image A (465×338 pixels) where the red "x" indicate shape matchings.

detected. Our algorithm takes 69s while the brute force algorithm takes 4 hours (this algorithm performs template matchings between each pixel of A and Q rotated by every angle and scaled by every scale). We also executed successfully some other experiments using remote sensing images.

References

[1] F. Mokhtarian and A. K. Mackworth, *A Theory of Multiscale, curvature-based shape representation for planar curves*, IEEE T. Pattern Anal. Mach. Intell. **14** (1992), no. 8, 789–805.

[2] F. Ullah and S. Kaneko, *Using orientation codes for rotation-invariant template matching*, Pattern Recognition **37** (2004), no. 2, 201-209.

[3] H. J. Wolfson and I. Rigoutsos, *Geometric hashing: An overview*, IEEE Comput. Sci. Eng. **4** (1997), no. 4, 10–21.

[4] D. H. Ballard, *Generalizing the Hough transform to detect arbitrary shapes*, Pattern Recognition **13** (1981), no. 2, 111–122.

[5] L. A. Torres-Mendez and J. C. Ruiz-Suarez and L. E. Sucar and G. Gomez, *Translation, rotation, and scale-invariant object recognition*, IEEE T. Systems, Man and Cybernetics **30** (2000), no. 1, 125–130.

[6] M. K. Hu, *Visual pattern recognition by moment invariants*, IEEE T. Inf. Theory **8** (1962), no. 2, 179–187.

[7] J. Flusser and T. Suk, *Rotation Moment Invariants for Recognition of Symmetric Objects*, IEEE T. Image Processing **15** (2006), no. 12, 3784–3790.

[8] H. Y. Kim, *Segmentation-Free Printed Character Recognition by Relaxed Nearest Neighbor Learning of Windowed Operator*, Proc. of the XII SIBGRAPI, 1999, pp. 195–204.

[9] T. M. Ha and M. Zimmermann and H. Bunke, *Off-line handwritten numeral string recognition by combining segmentation-based and segmentation-free methods*, Pattern Recognition **31** (1998), no. 3, 257–272.

[10] L. Yu and R. Wang, *Shape representation based on mathematical morphology*, Pattern Recognition Letters **26** (2005), no. 9, 1354–1362.