Symbol Classification Using Dynamic Aligned Shape Descriptor

Alicia Fornés
Computer Vision Center, UAB
Ed. O, 08193, Bellaterra
Dept. Computer Sciences, UAB
Ed. Q, 08193, Bellaterra
Barcelona, Spain
e-mail: afornes@cvc.uab.es

Sergio Escalera
Computer Vision Center, UAB
Ed. O, 08193, Bellaterra
Dept. Mat. Aplicada i Anàlisi, UB
Gran Via Corts Catalanes 585, 08007
Barcelona, Spain
e-mail: sergio@maia.ub.es

Josep Lladós, Ernest Valveny
Computer Vision Center, UAB
Ed. O, 08193, Bellaterra
Dept. Computer Sciences, UAB
Ed. Q, 08193, Bellaterra
Barcelona, Spain
e-mail: {josep,ernest}@cvc.uab.es

***Abstract*—Shape representation is a difficult task because of several symbol distortions, such as occlusions, elastic deformations, gaps or noise. In this paper, we propose a new descriptor and distance computation for coping with the problem of symbol recognition in the domain of Graphical Document Image Analysis. The proposed D-Shape descriptor encodes the arrangement information of object parts in a circular structure, allowing different levels of distortion. The classification is performed using a cyclic Dynamic Time Warping based method, allowing distortions and rotation. The methodology has been validated on different data sets, showing very high recognition rates.**

***Keywords*-Graphics Recognition, Symbol Recognition, Symbol Description**

I. Introduction

Symbol recognition is a particular case of object recognition and one of the main topics of Graphics Recognition. Symbols are synthetic visual entities made by humans to be understood by humans. They can appear in scanned document images or in natural scenes captured by a camera.

From the point of view of symbols in documents, the descriptor should ideally guarantee intra-class compactness and inter-class separability. It should be tolerant to noise, degradation, occlusions and distortion (including shear). And due to isolated symbols which are present in graphical documents, we must also take into account the variations in rotation, scaling and translation. On the contrary, in the camera-based symbol recognition domain, the system should cope with a totally different problematic: uncontrolled environments, illumination changes, and changes in the point of view (perspective).

According to Zhang and Lu [1], numerous shape descriptors, tolerant to such distortions, have been proposed. They can be classified in continuous and structural approaches. Continuous approaches use a feature vector derived from the image photometry to describe the shape. Structural approaches tend to represent the shape using structures like string, tree, graph or grammar, where the similarity measure is done by string, tree, graph matching or parsing, respectively. These approaches capture the spatial arrangement of symbol parts, which usually suffer from complex distortions.

In this paper, we introduce a novel symbol descriptor, the Dynamic Aligned Shape Descriptor (D-Shape). It encodes the spatial probability of appearance of the shape pixels and their context information. As a result, a robust technique to deal with noise and elastic deformations is obtained. The circular descriptor is stretched and aligned using a variation of the cyclic Dynamic Time Warping (DTW) algorithm, which makes the description rotation invariant. Moreover, the alignment cost is used as a measure of similarity, and thus, the description is useful for symbol retrieval and classification without the need of a learning stage.

The paper is organized as follows: Section 2 presents the D-Shape descriptor. Section 3 describes the DTW-based classification. Section 4 presents the experimental results, and finally, Section 5 concludes the paper.

II. Dynamic Aligned Shape Descriptor

In this Section we present the definition of the descriptor. Firstly, we define the location of some concentric circles, and for each one, we compute the location of the voting points. Secondly, these voting points will receive votes from the pixels of the shape, depending on their distance to each voting point.

A. Computation of the concentric circles

By defining a circular structure from the center of the object region, spatial arrangement of object parts is shared among voting points located in concentric circles. These points will be used for describing the neighbouring region of the shape. The goal is to locate isotropic equidistant points, so that the inner and external part of the symbol could be described using the same number of voting points. Thus, the descriptor defines in the same way all the regions of the symbol. For this purpose, the points are distributed in concentric circles, with the following constraints: each concentric circle is located at the same distance from its neighbouring ones, and all the points located in a circle are equidistant each other (see Fig.1(a)).

Given the number of concentric circles T and the radius of the most external circle R, we define C_i as the concentric

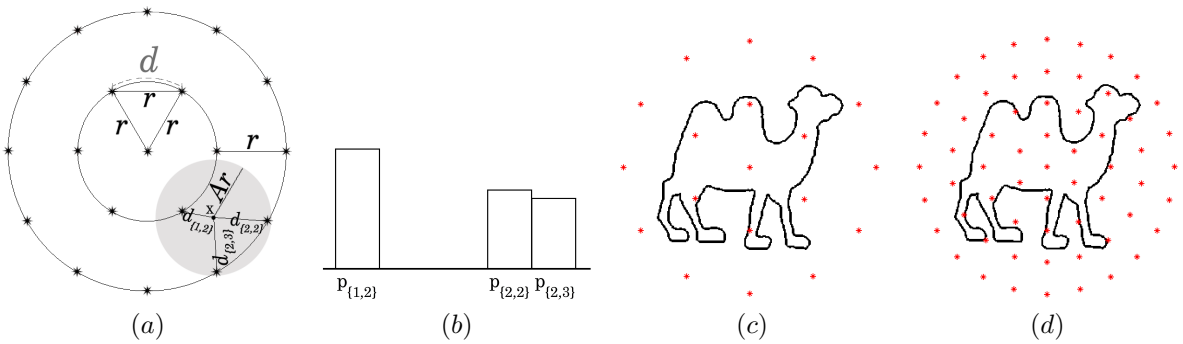


Figure 1. D-Shape description.(a) Location of the voting points. (b) Descriptor vector update after the analysis of **x**. (c) and (d) voting points for different number of circles.

circle i, R_i as the radius of C_i, and N_i as the number of points in C_i. We define $R_0 = 0$ as the center of the circles, $R_1 = r$ as the radius of the most internal circle, and $R_T = R$ as the radius of the most external one.

For satisfying the first constraint, each circle C_i must be located at the same distance r from C_{i+1} and C_{i-1}. For this purpose, $R_i = i \cdot R_1 = R_{i-1} + R_1$, where $R_1 = \frac{R}{T} = r$, ensuring:

$$R_i - R_{i-1} = R_1 = r, \forall i \in [2, .., T] \quad (1)$$

For satisfying the second constraint, we must determine the number of voting points that can be located in each circle, so that the distance between them is also r. For the first circle C_1 with $R_1 = r$, only 6 voting points can be located with a distance r between them, and a distance r to the center of the circle (see Fig.1(a)). Then, we can compute d as the distance in the perimeter of the circle between them as:

$$d = \frac{2\pi R_1}{6} = \frac{\pi}{3} R_1 \quad (2)$$

Consequently, the perimeter $Perim_i$ of each circle C_i is divided in N_i parts of length d:

$$\frac{Perim_i}{N_i} = d, \text{ then } \frac{2\pi R_i}{N_i} = \frac{\pi R_1}{3}, \text{ then } N_i = 2 \cdot 3i \quad (3)$$

Finally, the N_i voting points P will be located in the circle at an angular distance of $\frac{2\pi}{N_i}$. This procedure is detailed in Algorithm 1.

B. Computation of the histogram

The D-Shape codifies the spatial arrangement of object characteristics based on a defined blurring degree, which determines the *shape* deformation allowed to the object. First of all, the input binary or grey-level segmented image I is preprocessed for obtaining the important shape features. In our case, we base the relevant shape features as those points in the image with high gradient magnitude. In particular, for each foreground pixel x, we save its corresponding gradient magnitude in $I'(x)$ if its gradient magnitude is higher to the 20% of the highest gradient magnitude in the symbol region. Secondly, the voting points P are determined so that the center of the shape (the centroid) coincides with the center of the circles, without any image resizing. Once we have the voting points P, the descriptor vector ν is computed. For this purpose, each point x from I' is taken into account in the description process. Firstly, the distances from the relevant point x to its neighbouring voting points are computed (see Fig.1(a)). Notice that the neighbouring voting points will be determined depending on the voting influence area A, which determines the blurring degree. Then, the inverse of these distances is computed and normalized by the sum of total distances. These values are weighted by the magnitude of its gradient $I'(x)$, and added to the corresponding positions of the descriptor vector ν (see Fig.1(b)). Finally, the vector is normalized ($\sum \nu = 1$).

Notice that the length of ν is defined by parameter T, which defines the degree of spatial information taken into account in the description process (see Fig.1(c),(d)). In the way that we increase the number of circles T or decrease the influence voting area A, the description becomes more local. Contrary, if T decreases or A increases, the blurring degree is higher, making the description more tolerant to irregular deformations. Thus, an optimal parameters of T and A should be obtained for each particular problem. This procedure is detailed in Algorithm 1.

III. Distance computation

Once we have computed the D-Shape descriptor vector of two shapes, a matching algorithm is required for determining the distance between them. Since the matching method must be tolerant to distortions, elastic deformations, occlusions, gaps, and rotation, we propose a cyclic version of the Dynamic Time Warping algorithm for bi-dimensional shapes.

The Dynamic Time Warping algorithm [2] is used for comparing signals by matching two one-dimensional vectors. It is a much more robust distance measure for time series than Euclidean distance, allowing similar samples to match even if they are out of phase in the time axis.

Algorithm 1 D-Shape Description Algorithm.

Require: An image of high gradient magnitudes I' with dimensions $W \times Z$, the number of circles T, $T \geqslant 1$, and the voting influence area A, $A \geqslant 1$

Ensure: Descriptor vector ν

Computation of the location of the voting points P

Define the radius of the most external circle $R_T = \max\{\frac{W}{2}, \frac{Z}{2}\}$, and the radius of the most internal circle $R_1 = \frac{R_T}{T} = \frac{R}{T}$.

Define $P = \{p_{\{0,1\}}, p_{\{1,1\}}, ..., p_{\{1,6\}}, ..., p_{\{T,1\}}..., p_{\{T,2\cdot 3T\}}\}$ as the set of voting points. Thus, $p_{\{i,j\}}$ corresponds to the voting point j of the circle i.

Define the centre of the circle $p_{\{0,1\}} = (\frac{W}{2}, \frac{Z}{2})$ as the center of I'.

for each concentric circle $C_i, i = 1, .., T$ **do**

$N_{\{i\}} = 2 \cdot 3 \cdot i = 6 \cdot i$, as the number of points in the circle i

$R_{\{i\}} = R_i \cdot i$, as the radius of the circle i

$\alpha = \frac{2\pi}{N_i}$, as the angle increment for each point

$p_{\{i,j\},j=1,..,N_i} = (R_i \sin(j\alpha), R_i \cos(j\alpha))$, as the voting points in the circle i

end for

Computation of the descriptor vector ν.

Initialize $\nu_i = 0$, $i \in [1, .., |P|]$, where $|P|$ corresponds to the number of points in P (the cardinal of P). The order of indexes in ν are:

$\nu = \{p_{\{0,1\}}, p_{\{1,1\}}, ..., p_{\{1,6\}}, ..., p_{\{T,1\}}..., p_{\{T,2\cdot 3\cdot T\}}\}$

for each point $\mathbf{x} \in I'$, $I'(\mathbf{x}) > 0$ **do**

Initialize $D_x = 0$

for each $p_{\{i,j\}}$ satisfying $||\mathbf{x} - p_{\{i,j\}}||^2 < A \cdot R_1$ **do**

$d_{\{i,j\}} = d(\mathbf{x}, p_{\{i,j\}}) = ||\mathbf{x} - p_{\{i,j\}}||^2$

$D_x = D_x + \frac{1}{d_{\{i,j\}}}$

end for

Update the probabilities vector ν positions as follows:

$\nu(p_{\{i,j\}}) = \nu(p_{\{i,j\}}) + I'(\mathbf{x}) \cdot \frac{1/d_{\{i,j\}}}{D_x}$

end for

Normalize the vector ν as follows:

$d' = \sum_{i=1}^{|P|} \nu_i$, $\nu_i = \frac{\nu_i}{d'}, \forall i \in [1, .., |P|]$

In case of bidimensional data, instead of applying a 2-DTW algorithm (with very high time complexity), the 2D representation is typically reduced by encoding them in 1D signals. Consequently, the time complexity is significantly reduced. In this cases, the distance between two points is substituted by the distance between two vectors.

The distance computation method proposed in this paper is a variation on the DTW algorithm, that no only adapts the algorithm to bidimensional data, but also is tolerant to rotation by the definition of a cyclic version of the DTW. The first step consists in resampling the descriptor vector ν for obtaining a matrix F of features. For this purpose, each row i in the matrix will consist in the resampling of the voting points of the circle C_i, so that each row i will contain the same number of columns as the most external circle ($2 \cdot 3 \cdot N_T = 6 \cdot N_T$). Due to this resampling, the matching must take into account that some voting points in the descriptor ν are repeated several times, which is particularly true in the small concentric circles. For this reason, the weight of this points Q in the computation of the

Algorithm 2 D-Shape Matching Algorithm.

Require: Two D-Shape descriptors ν_1 and ν_2 with T concentric circles

Ensure: Matching Distance $Cost$

Define the feature matrix $F1$ from ν_1 as the resampling of the voting points of each circle to the length of $2 \cdot 3 \cdot T = 6T$.

$f_1(k+1, 1..6T) = \text{resample}(\nu_1(p_{\{k,1\}}), .., \nu_1(p_{\{k,6i\}}))$, where $k = 0..T$

$$F_1 = \begin{pmatrix} f_1(1,1) & \cdots & f_1(1,6T) \\ \cdots & \cdots & \cdots \\ f_1(T+1,1) & \cdots & f_1(T+1,6T) \end{pmatrix}$$

Define the feature matrix $F2$ from ν_2, as the resampling of the voting points of each circle to the length of $2 \cdot 3T = 6T$, and duplicate the matrix.

$f_2(k+1, 1..6T) = \text{resample}(\nu_2(p_{\{k,1\}}), .., \nu_2(p_{\{k,6i\}}))$, where $k = 0..T$

$$F_2 = \begin{pmatrix} f_1(1,1) & \cdots & f_1(1,6T), & f_1(1,1) & \cdots & f_1(1,6T) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_1(T+1,1) & \cdots & f_1(T+1,6T), & f_1(T+1,1) & \cdots & f_1(T+1,6T) \end{pmatrix}$$

Define Q as the weight of each concentric circle:

$Q(0) = \frac{1}{6T}$, $\quad Q(k) = \frac{1}{6T/6k} = \frac{6k}{6T} = \frac{k}{T}, \forall k \in [1, .., T]$

Initialize the distance matrix MD as follows:

$MD(0,0) = 0$

$MD(0,j) = 0; \forall j \in [1, .., 6T]$

$MD(0,j) = \infty; \forall j \in [6T+1, .., 12T]$

$MD(i,0) = MD(i-1,0) + \text{dist}(f_1(i), f_2(1)); \forall i \in [1..T+1]$

where $\text{dist}(f_1(i), f_2(j)) = \sum_{k=1}^{T} Q(k) \cdot (f_1(k,i) - f_2(k,j))^2$

for each i, $i = 1, .., T+1$ **do**

for each j, $j = 1, .., 12T$ **do**

$$MD(i,j) = \min \begin{Bmatrix} MD(i,j-1) \\ MD(i-1,j) \\ MD(i-1,j-1) \end{Bmatrix} + \text{dist}(f_1(i), f_2(j))$$

$$\text{dist}(f_1(i), f_2(j)) = \sum_{k=1}^{T} Q(k) \cdot (f_1(k,i) - f_2(k,j))^2$$

end for

end for

Compute the length of the warping path Z and normalize the matching cost:

$$Cost = \frac{\min\{MD(M, N+1), .., MD(M, 2N)\}}{Z}$$

distance between two columns must be decreased depending on the degree of resampling performed. The most external circle has no resampling, and for this reason, $Q(C_T) = 1$. The most internal circle has been repeated $N_T = 6T$ times due to the resampling, and consequently, $Q(C_0) = \frac{1}{6 \cdot T}$. The weight for the circle i is the following:

$$Q(C_i) = \frac{1}{N_T/N_i} = \frac{6i}{6T} = \frac{i}{T} \tag{4}$$

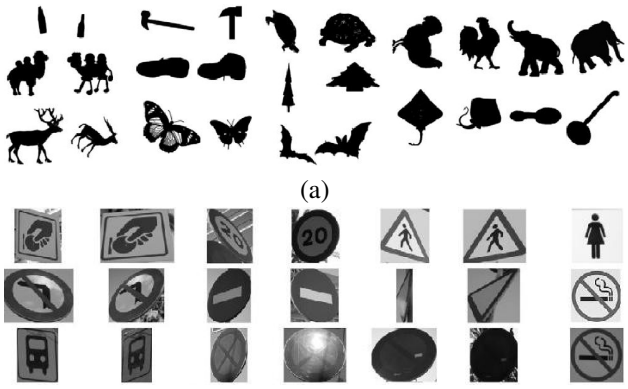
Thus, when computing the distance between two vectors, the weight Q will determine the influence of each rest in the vector. This algorithm is fully detailed in Algorithm 2.

IV. Experimental Results

The proposed methodology has been compared with SIFT [3], Zoning, BSM [4] and Zernike Moments [5]. Seven moments are used in Zernike moments. The optimum number

of circles for the D-Shape descriptor (11 circles) and the optimum grid size for the BSM and Zoning (16× 16 regions) have been computed via cross-validation using a 10% of the samples for validation. These descriptors are trained using 50 runs of Gentle Adaboost with decision stumps, and the one-versus-one ECOC design with the Euclidean distance decoding [6]. We also compare with a 3-Nearest Neighbour classifier. Contrary, for the D-Shape, only the DTW-based algorithm has been used, avoiding a training step. The classification score is computed by means of stratified 10-fold cross-validation, testing for the 95% of the confidence interval with a two-tailed t-test.

The public 70-class MPEG7 repository data set contains 20 instances of each class, obtaining a total of 1400 instances. It has been chosen because it contains samples with high intra-class variability in terms of scale, rotation, rigid and elastic deformations, as well as a low inter-class variability. A pair of samples for some categories of the data set are shown in Fig 2(a).



(a) (b)

Figure 2. Data sets. (a) MPEG-7, (b) Grey Traffic signs

For the BSM and Zoning descriptors, the Hotelling alignment is applied for rotating the image. Then, each feature set is classified using the one-versus-one scheme with Gentle Adaboost and a 3-Nearest Neighbor classifier. The performance and confidence interval obtained by each descriptor and classifier are shown in Table I. We can observe that the recognition rate of the D-Shape (83.7%) clearly outperforms the others (less than 78%).

The second data set is composed by 17 classes of grey-level traffic sign symbols, with a total of 550 samples acquired with a digital camera from real environments. It contains the common distortions from real environments, such as illumination changes, partial occlusions, or changes in the point of view (see Fig 2(b)). In grey-level data sets, the Zernike moments and Zoning are not suitable descriptors, so the comparison is performed with SIFT and BSM descriptors. In Table II the results show that the D-Shape (87%) significantly outperforms the others.

Table I
MPEG-7 CLASSIFICATION

Descriptor	3NN	ECOC G.Adaboost
BSM	65.79 (8.03)	**77.93 (7.25)**
Zernike	43.64 (7.66)	51.29 (5.48)
Zoning	58.64 (10.97)	65.50 (6.64)
SIFT	29.14 (5.68)	32.57 (4.04)
D-Shape	**83.7 (1.69)**	-

Table II
TRAFFIC SIGNS CLASSIFICATION

SIFT (E.G.Adaboost)	BSM (E.G.Adaboost)	D-Shape (3NN)
62.12 (9.08)	75.23 (7.18)	**87.04 (2.91)**

V. CONCLUSION

In this paper we have presented the D-Shape descriptor, which encodes the spatial arrangement of the shape using voting points located in concentric circles. It copes with distortions, occlusions, scale and noise, and allows different degradation levels. Thanks to the use of a cyclic DTW-based method, it becomes robust to elastic deformations and rotation. The results show that it outperforms the state of the art methods.

ACKNOWLEDGMENT

This work has been partially supported by the Spanish projects TIN2008-04998, TIN2009-14633-C03-03, TIN2009-14404-C02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018).

REFERENCES

[1] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, pp. 1–19, 2004.

[2] J. B. Kruskal and M. Liberman, "The symmetric time-warping problem: From continuous to discrete," in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, D. Sankoff and J. B. Kruskal, Eds. Addison-Wesley Publishing Co., September 1983, pp. 125–161.

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] S. Escalera, A. Fornés, O. Pujol, P. Radeva, G. Sánchez, and J. Lladós, "Blurred Shape Model for binary and grey-level symbol recognition," *Pattern Recognition Letters*, vol. 30, no. 15, pp. 1424–1433, 2009.

[5] A. Khotanzad and Y. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[6] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, no. 2, pp. 337–374, 2000.