Partial Shape Matching and Retrieval under Occlusion and Noise

Leonardo Chang[1,2], Miguel Arias-Estrada[2],
José Hernández-Palancar[1], and L. Enrique Sucar[2]

[1] Advanced Technologies Application Center (CENATAV)
7A # 21406, Siboney, Playa, C.P. 12200, Havana, Cuba
{lchang,jpalancar}@cenatav.co.cu
[2] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)
Luis Enrique Erro # 1, Tonantzintla, C.P. 72840, Puebla, Mexico
{lchang,ariasmo,esucar}@ccc.inaoep.mx

Abstract. Shape information have proven to be useful in many computer vision applications. In this work, a self-containing shape descriptor for open and closed contours is proposed. Also, a partial shape matching method robust to partial occlusion and noise in the contour is proposed. Both the shape descriptor and the matching method are invariant to rotation and translation. Experiments were carried out in the Shapes99 and Shapes216 datasets, where contour segments of different lengths were removed to obtain partial occlusion as high as 70%. For the highest occlusion levels the proposed method outperformed other popular shape description methods, with up to 50% higher bull's eye score.

Keywords: partial occlusion, open contours, shape representation, partial shape matching.

1 Introduction

Shape information have proven to be useful in many image processing and computer vision applications such as object detection, image retrieval and 3D curve reconstruction. However, shape representation and matching remains as one of the most challenging topics in computer vision, partly because of partial occlusion and noise in the shape information extracted in real images. Most previous works on shape description and matching assume that the whole shape is always visible in images and that the complete shape contour can be extracted, so they work only for closed contours and assume a one-to-one correspondence between shape points [4,9]. Dealing with partial occlusion and noise implies that the shape descriptor should be able to represent both open and closed contours, and that part of the contour fragments should match with part of the shape model, which makes the shape matching problem more difficult than that for closed shapes. Other considerations are the robustness with respect to the image scale, rotation and translation.

In this work, we propose a shape descriptor that is particularly suitable for partial shape matching of open/closed contours. Our descriptor, named OCTAR

(Open/Closed contours Triangle Area Representation), measures the convexity or concavity of contour segments using the signed areas of triangles formed by every pair of contour points and their middle point. Based on this descriptor, we also propose a partial shape matching method robust to partial occlusion and noise in the extracted contour. The matching method finds for every contour fragment in the query shape their best match in the shape model. We extend the OCTAR descriptor to represent the spatial configuration of two contour fragments. Individual matches with coherent spatial configurations wrt. the model are joined to form object hypotheses in an agglomerative hierarchical process. Later, hypotheses are evaluated based on the coverage of the model contour and measuring the global shape similarity. The conducted experiments on the Shapes99 and Shapes216 datasets showed that while increasing the occlusion level in shape contour, the difference in terms of bull's eye score gets larger in favor of OCTAR compared to other state of the art methods.

The rest of the paper is organized as follows. Section 2 discusses shape description and matching approaches based on the area of triangles formed by contour points, which are more closely related to our work. Section 3 presents the proposed OCTAR shape descriptor. Sections 4 and 5 describe the proposed partial matching method. The performed experiments and discussion are presented in Section 6. Finally, Section 7 concludes the paper with a summary of our proposed methods, main contributions, and future work.

2 Related Work

Several methods have used the area of triangles formed by contour points as the basis for shape representations. In [6], Roh and Kweon have proposed the use of shape features based on triangle area using five equally spaced contour points $p_1(t), p_2(t), p_3(t), p_4(t)$ and $p_5(t)$ from a closed boundary of N points. For each selection $t = \{1, 2, ..., N\}$ they defined the shape invariant as: $I(t) = \frac{A(p_5(t)p_1(t)p_4(t)) \cdot A(p_5(t)p_2(t)p_3(t))}{A(p_5(t)p_1(t)p_3(t)) \cdot A(p_5(t)p_2(t)p_4(t))}$, where $A(p_a p_b p_c)$ is the area of the triangle formed by points p_a, p_b and p_c. Finally, the shape signature of a boundary is obtained by plotting the value $I(t)$ versus t for the different values of $t = \{1, 2, 3, ..., N\}$.

Rube *et al.* [7] proposed a method named Multi-scale Triangle-Area Representation (MTAR). This representation uses the area of the triangles formed by each three consecutive and equally spaced points on a closed boundary. A MTAR image is obtained by thresholding the area function at zero and taking the locations of the negative values. To reduce noise effect, they apply a Dynamic Wavelet Transform to each contour sequence at various scale levels. At each wavelet scale level a TAR image is computed. In order to match two MTAR image sets of two shapes, several global features are used to discard very dissimilar shapes. Then, a similarity measure D_s between each two MTAR images at certain scale is computed. D_s is based on finding a number of initial correspondences between two sets of maxima in the MTAR images using only two maxima in each image. After that, the lowest cost node is extended to include all other maxima and its cost is considered as D_s.

   More recently, the triangle-area representation signature (TAR) proposed by
Alajlan *et al.* [1] have shown very good results in shape retrieval. TAR is com-
puted from the area of the triangles formed by the points on the shape boundary
at different scales. For the matching, the optimal correspondence between the
points of two shapes is searched using a Dynamic Space Warping algorithm.
Based on the established correspondence, a distance is derived, and global fea-
tures are incorporated in the distance to increase the discrimination ability and
to facilitate the indexing in large shape databases.
   The aforementioned approaches can only deal with shapes of closed boundary.
Also, the contour of each object is represented by a fixed number of sample points
and no partial matches of the shape are allowed, hence, how these approaches
work under occluded, noisy or uncompleted contours is not well-defined. In this
paper, we propose a self-containing, triangle area-based shape descriptor able to
represent both open and closed contours, and a partial matching method that
takes advantage of the properties of this descriptor to provide robustness to
partial occlusion and noise in the contour.

3   Shape Descriptor

In order to find partial shape correspondences between contour fragments and
a model, a shape descriptor must be able to represent both open and closed
contours, must be self-contained, and invariant to rotation and translation.
In this paper, we propose a shape descriptor based on the triangle area rep-
resentation, that meets these properties. The proposed descriptor is named
OCTAR, from Open/Closed contour Triangle Area Representation. The use of
triangle areas provides discriminative data about shape features such as the con-
vexity/concavity at each curve segment. For contour sequences represented in
counter clockwise direction, positive, negative and zero values of OCTAR indi-
cate convex, concave and straight-line points, respectively.
   Given a sequence of N ordered points, $\mathcal{P} = \{p_1, p_2, ..., p_N\}, p_i \in \mathbb{R}^2$, repre-
senting a contour fragment, for each pair of points $\langle p_i, p_j \rangle$ in $\mathcal{P}$ we compute the
area of the triangle formed by these two points and their middle point $p_* \in \mathcal{P}$
(see Fig. 1(a)). The signed area of the triangle formed by these points is given
by

$$TAR(i, j, *) = \frac{1}{2} \det \begin{pmatrix} x_i & y_i & 1 \\ x_* & y_* & 1 \\ x_j & y_j & 1 \end{pmatrix}, \tag{1}$$

where $\det()$ is the matrix determinant. If the middle point between $\langle p_i, p_j \rangle$ does
not exist, i.e., $(i - j)$ is even, p_* is interpolated from $p_{\lfloor (i+j)/2 \rfloor}$ and $p_{\lceil (i+j)/2 \rceil}$.
   To obtain the OCTAR descriptor of contour fragment $\mathcal{P}$, denoted as $\Theta^{\mathcal{P}}$, the
triangle area is normalized by the area of the equilateral triangle inscribed in the
minimum enclosing circle of the sub-contour $\{p_i, ..., p_j\} \subseteq \mathcal{P}$ (it can be proved
that this is the maximum area triangle of all possible triangles inside a circle),
and represented in a log space to make the descriptor more sensitive to the area
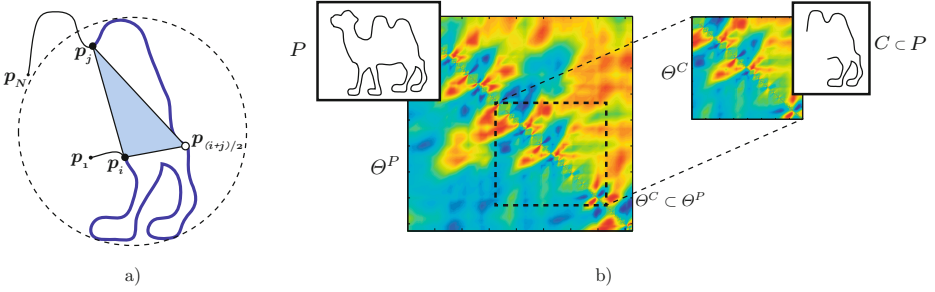of nearby contour points than to those of points farther away:

Fig. 1. *(best seen in color.)* (a) The OCTAR descriptor is computed from the area of the triangles formed by every pair of points in the shape and their middle point. In (b) the OCTAR descriptor matrices, Θ^P and Θ^C, of two shapes, P and C, are shown. The self-containing property of the proposed descriptor can also be appreciated.

$$\Theta^{\mathcal{P}}(i,j) = \log\left(1 + \frac{TAR(i,j,*)}{\mathcal{A}(\{p_i,...,p_j\})}\right), \tag{2}$$

where $\mathcal{A}(\{p_i,...,p_j\})$ is the area of the equilateral triangle inscribed in the minimum enclosing circle of the sub-contour $\{p_i,...,p_j\} \subseteq \mathcal{P}$. We add one to the normalized triangle area to make $\Theta(i,j)$ positive.

The similarity of two OCTAR descriptors Θ^P and Θ^Q of the same size is given by

$$\Phi(P,Q) = 1 - \left(\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|\Theta^P(i,j) - \Theta^Q(i,j)\right|\right), \tag{3}$$

where $M \times N$ is the size of the descriptor matrices.

The proposed shape descriptor has three important properties. i) OCTAR is able to represent both open and closed contours, since it does not make any assumption over the contour closeness. ii) The triangle area based representation makes it invariant to rotation and translation. iii) OCTAR is self-contained as for any $C \subset \mathcal{P}$ it holds that $\Theta^C \subset \Theta^{\mathcal{P}}$, this property implicitly allows to retrieve partial matches from the contour description. Fig. 1 (b) shows the OCTAR signature of two shapes, being the second shape a subset of the first. As it can be appreciated, the OCTAR descriptor of the second shape is contained in that of the first shape.

4   Partial Shape Matching

Given the set of contour fragments $\mathcal{F} = \{f_1, f_2, ..., f_K\}$ that represent the query shape image, where each fragment $f_k = \{p_1, p_2, ..., p_{N_k}\}, p_i \in \mathbb{R}^2$ is a sequence of N_k points, and the shape model $\mathcal{Q} = \{q_1, q_2, ..., q_M\}, q_i \in \mathbb{R}^2$, a sequence of M points. What we want to find is the best correspondence between the part $f_k(a,l) = \{p_a, ..., p_{a+l-1}\}, f_k(a,l) \subseteq f_k$ of a contour fragment and the part $\mathcal{Q}(b,l) = \{q_b, ..., q_{b+l-1}\}, \mathcal{Q}(b,l) \subseteq \mathcal{Q}$ of the model, where a and b are the initial points in f_k and $\mathcal{Q}$, respectively, and l is the part length.

Based on the proposed OCTAR descriptor, we introduce a method for finding partial matches between contour fragments and a model. In order to find partial matches of arbitrary sizes we have to compare all possible sub-blocks of the descriptor matrices to find the corresponding sub-blocks with the maximum similarity. With this aim, we build a 4D tensor $\mathbf{T}(k, a, b, l) = \Phi(f_k(a, l), \mathcal{Q}(b, l))$, where Φ is the similarity measure between descriptor matrices defined in Equation 3. To efficiently build $\mathbf{T}$, we use the integral image optimization to access the partial descriptor differences in constant time, as suggested in [5]. This optimization is possible thanks to the self-containing property of the OCTAR descriptor.

In order to select the best match between part $f_k(a, l) \subseteq f_k$ and $\mathcal{Q}(b, l) \subseteq \mathcal{Q}$, the simplest criterion could be to select those fragments with the maximum similarity in $\mathbf{T}$. However, given the observation that when very short fragments are involved in a matching, it is neither discriminative nor reliable, even when having the highest similarity values. To overcome this limitation, we propose a more robust alternative.

For each length value l, we select the best matching part of length l in each contour fragment k,

$$\mathcal{L}(k, l) = \max_{a,b} \mathbf{T}(k, a, b, l). \tag{4}$$

Given the exhaustive character of $\mathbf{T}$, in a neighborhood of the best matching part there will be a large amount of strong matches in $\mathcal{L}$. In order to detect the best match, we build a histogram of the similarity values in $\mathcal{L}$, i.e., $h(r) = |\{\mathcal{L}(k, l) : \mathcal{L}(k, l) \in \mathbf{bin}(r)\}|$. From this histogram, we select as the best match for each contour fragment, the correspondence with the largest l in the highest frequency bin.

5   Hypotheses Evaluation

In presence of partial occlusion and noise, not every contour fragment $f_k \in \mathcal{F}$ has to be part of the object. Therefore, we have to select among the set of candidate matches those that really belong to the object. The number of possible combinations of contour fragments that can be joined to form the object is exponential with respect to the number of contour fragments. In order to reduce the number of possible combinations, spatial information is used. Each matched contour fragment is mapped to its corresponding part in the model and the object centroid estimated. Only matches with neighboring object centroid estimations can be later joined as an object hypothesis. Beside reducing computational time, this step avoids false positives from an early stage.

In order to use further the spatial information in an object hypothesis, we extended the OCTAR shape descriptor, defined in Equation 2, such that it could express the spatial configuration between two contour fragments $\mathcal{P} = \{p_1, p_2, ..., p_N\}$ and $R = \{r_1, r_2, ..., r_M\}$, where each point $p_i \in \mathcal{P}$ is related with each point $r_j \in R$ through the area of the triangle formed by p_i, r_j, and the first point in R (see Fig. 2). The spatial configuration descriptor of two contour fragments $\mathcal{P}$ and R is defined as

Fig. 2. (*best seen in color*). Extended OCTAR descriptor for contours spatial configuration. The same two contours under different configurations have different descriptor.

$$\Theta^{\mathcal{P},R}(p_i, r_j) = \log\left(1 + \frac{TAR(p_i, r_j, r_1)}{\mathcal{A}(\{p_i, r_j, r_1\})}\right). \tag{5}$$

In order to form object hypotheses, we use an agglomerative hierarchical approach in the spatial configuration space. Initially, each match is a hypothesis; later, in each iteration, the two hypotheses which contour fragments configuration is more similar to its corresponding configuration of model parts are joined, restricted to those fragments with neighboring object centroid estimations.

Once the hypotheses hierarchy is obtained, each object hypothesis is evaluated according to two criteria. The first is based on the coverage of the model contour by the hypothesis, and is defined as $\mathcal{E}_{COV} = \frac{1}{M}\sum_i^M w_i$, where w_i takes value of one or zero to indicate whether the ith point of the model contour has been matched or not, and M is the number of points in the model. The second hypothesis evaluation measure assesses the shape similarity in a global manner. With this aim, we compared the OCTAR descriptor of the concatenation of all the contour fragments in a hypothesis with the OCTAR descriptor of the corresponding model parts concatenated, $\mathcal{E}_{GLB} = \Phi(\hat{f}, \hat{\mathcal{Q}})$, where $\hat{f}$ and $\hat{\mathcal{Q}}$ are the concatenated sequences of contour fragments and model parts points, respectively, in the hypothesis. Finally, we select as final match the hypothesis with the biggest linear combination of $\mathcal{E}_{COV}$ and $\mathcal{E}_{\Theta}$, defined as $\mathcal{M}^* = \arg\max\left[\alpha\mathcal{E}_{COV} + (1-\alpha)\mathcal{E}_{GLB}\right]$, where α and $(1-\alpha)$ are the weights associated to $\mathcal{E}_{COV}$ and $\mathcal{E}_{GLB}$, respectively. In all our experiments we used $\alpha = 0.5$.

6   Experimental Results

The performance of the proposed method has been evaluated on two different well-known datasets. The first dataset is the Kimia Shapes99 dataset [8], which includes 9 categories and 11 shapes in each category with variations in form, occlusion, articulation and missing parts. The second dataset is the Kimia Shapes216 dataset [8]. The database consists of 18 categories with 12 shapes in each category. In the two datasets, in each image there is only one object, defined by its silhouette, and at different scales, rotations and positions.

In order to show the robustness of the proposed method to partial occlusion, we generated another 14 datasets by artificially introducing occlusion of different magnitudes (10%, 20%,...,70%) to the Shapes99 and Shapes216 datasets. Occlusion was added by randomly removing 2 to 5 fragments of the entire contour,
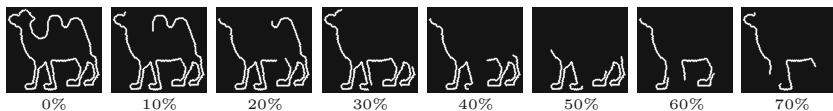
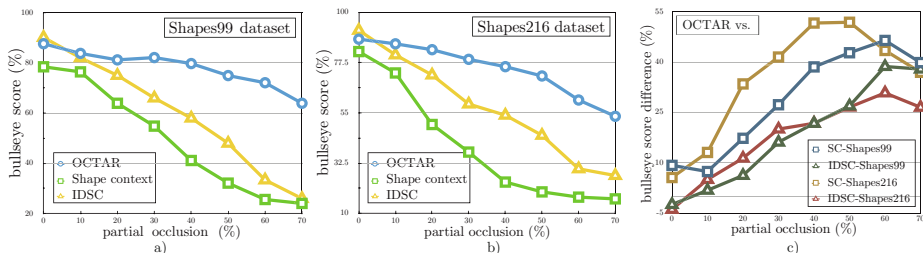**Fig. 3.** Example image from the Shapes216 dataset with different levels of occlusion

Fig. 4. (*best seen in color*). Bull's eye score comparison between OCTAR, shape context and IDSC in the a) Shapes99 and b) Shapes216 datasets with different partial occlusions. In c) the OCTAR improvement compared to shape context and IDSC.

which total length represents the desired partial occlusion. A sample image from the Shapes216 dataset at different occlusion levels is shown in Fig. 3.

As a measure to evaluate and compare the performance of the proposed shape matching schema in a shape retrieval scenario we use the so-called bull's eye score. Each shape in the database is compared with every other shape model, and the number of shapes of the same class that are among the $2N_c$ most similar is reported, where N_c is the number of instances per class. The bull's eye score is the ratio between the total number of shapes of the same class and the largest possible value. The results obtained by our proposed method were compared with those of the popular shape context [2] and IDSC [3] descriptors. Fig. 4 shows the behavior of the bull's eye score of each method.

As expected, our proposal outperforms the shape context and IDSC methods. Moreover, while increasing the occlusion level, the difference in terms of bull's eye score gets bigger, with about 30 - 50% higher bull's eye score across highly occluded images; Fig. 4(c) shows the advantages of OCTAR over the other two, in particular for highly occluded contours. The computation time of our proposed method has been also evaluated and compared to other methods. OCTAR, shape context and IDSC took as average 45, 1710, and 6 milliseconds for extracting and matching features of two shapes, respectively. These results were obtained on a single thread of a 2.2 GHz CPU and 8GB RAM PC. As it can be seen, OCTAR is faster than shape context and comparable to IDSC.

7 Conclusion and Future Work

As a result of this work, a shape descriptor for open and closed contours, and a partial shape matching method have been proposed. The proposed descriptor

and matching method allow us to find the best matching parts of a query object with a model in presence of partial occlusion and noise. Also, the proposed method is invariant to rotation and translation. The conducted experiments supported the mentioned contributions, showing larger robustness to partial occlusion than other methods in the state of the art.

As one of the main contributions of our proposal, it does not assume that the whole shape is visible in the image, nor that the complete shape contour can be extracted or that there are no more objects in the background. Therefore, the method has great potential for use in object detection, recognition and categorization from "real" images (RGB or grayscale images), using edge information, e.g., extracted using Canny. Future work will focus on this subject.

References

1. Alajlan, N., El Rube, I., Kamel, M.S., Freeman, G.: Shape retrieval using triangle-area representation and dynamic space warping. Pattern Recognition 40(7), 1911–1920 (2007)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE TPAMI 24(4), 509–522 (2002)
3. Ling, H., Jacobs, D.W.: Shape classification using the inner-distance. IEEE TPAMI 29(2), 286–299 (2007)
4. Yang, M., Kpalma, K., Ronsin, J.: A survey of shape feature extraction techniques. In: Pattern Recognition, pp. 43–90 (November 2008)
5. Riemenschneider, H., Donoser, M., Bischof, H.: Using partial edge contour matches for efficient object category localization. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 29–42. Springer, Heidelberg (2010)
6. Roh, K.-S., Kweon, I.-S.: 2-d object recognition using invariant contour descriptor and projective refinement. Pattern Recognition 31(4), 441–455 (1998)
7. Rube, I.E., Alajlan, N., Kamel, M., Ahmed, M., Freeman, G.H.: Robust multiscale triangle-area representation for 2d shapes. In: ICIP (1), pp. 545–548. IEEE (2005)
8. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of shapes by editing their shock graphs. IEEE TPAMI 26(5), 550–571 (2004)
9. Zhang, D.: Review of shape representation and description techniques. Pattern Recognition 37(1), 1–19 (2004)