WHAT PARTS OF A SHAPE ARE DISCRIMINATIVE?

Vittal Premachandran and Ramakrishna Kakarala

School of Computer Engineering, Nanyang Technological University, Singapore-639798

ABSTRACT

Shape is a distinguishing feature of objects. But when objects have similar global shapes, discriminating them becomes difficult. This calls for the identification of important parts of the shapes. In this paper, we introduce the concept of discriminative parts and propose a method to identify the same. We show how levels of importance can be assigned to different regions of contours, based on their discriminative capability. Our experiments show that the method is promising and can identify semantically meaningful segments as being important segments.

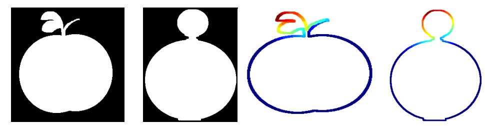
***Index Terms*—** Shape, Image Retrieval, Object Recognition

1. INTRODUCTION

While looking at images of objects, we perceive a wide variety of information such as colour, shape, contrast, etc. Among these, shape seems to have a big effect on object memorability. Understanding object shapes is a key towards recognising objects, and this has been acknowledged by the computer vision community. The object recognition/image retrieval literature has many works related to shape matching, and the development of shape descriptors [1, 2]. However, to the best of our knowledge, there is no work that looks at importance of object parts. In this paper, we would like to ask the following questions. What parts of an object's shape make it distinctive? Are the distinctive parts consistent among all shape instances of a particular class?

Consider the example images shown in Figure 1. The silhouette in Figure 1a is that of an apple, while the silhouette in Figure 1b is that of a pocket watch. We have no difficulty in distinguishing between the two even though a large portion of the object is quite similar to each other. Given such shapes that have strong overall visual similarity, we tend to give more importance to the differences in the two shapes while distinguishing between them. Ex: The stem of the apple is considered to be the distinguishing part between the two objects shown. Similarly, we would like the shape-matching algorithms to automatically identify the discriminative regions. Figures 1c and 1d shows the results from our algorithm, which is able to successfully identify discriminative regions.

Many state-of-the-art shape matching techniques cannot make such fine-scale distinction. There is an implicit assumption that all parts of a shape are equally important. This assumption might be detrimental towards the development of better shape matching techniques in the future. Widely-used shape descriptors such as Shape Context (SC) [1], and Inner Distance Shape Context (IDSC) [2], get confused between the two classes shown in Figure 1. Hence, there is a need to develop techniques that can match different parts of the shape with different levels of importance.



(a) (b) (c) (d)

Fig. 1: [Best viewed in color] (a) and (b) show two example images that are globally similar to each other. (c) and (d) show the heat maps on the contour, which indicate the level of importance. Most important regions are depicted by red, and the least important regions are depicted by blue.

While previous work has aimed at developing shape descriptors, and better shape matching techniques, in this work, we try to identify the "important" regions of a shape. The term "important" is quite subjective. So, we define important parts of the shape to be the discriminative parts of the shape, which helps enhance intra-class shape similarity, and reduce the inter-class shape similarity.[1]

Thus, the goal of this paper is to introduce the concept of discriminative shape parts. To this end, we propose techniques for decomposing a shape into parts, show how to identify the discriminative parts among them. We also develop a measure of consistency of discriminative parts among various class instances.

2. RELATED WORK

Shape-matching, being an active area of research, has numerous publications attributed to it. In this section, we refer some of the relevant and influential work from this area. Belongie et al. [1] proposed the Shape Context (SC) shape descriptor, which is a 2-D histogram of distances and angles. They also

[1]We use the terms important, discriminative, and distinctive, interchangeably throughout the paper.

showed that good-quality matching can be done using uniformly sampled points from the shape's contour as opposed to using points of extreme curvature as landmark points. Ling et al. [2] extended this work and proposed the Inner Distance Shape Context (IDSC) to perform articulation invariant shape matching. Both these descriptors are global shape descriptors and try to match the global visual similarity between two shapes.

Bronstein et al. [3] linked the concept of Pareto-optimality to quantify matching of partial shapes. This was later used by Donoser et al. in [4], where they propose a new shape descriptor that can perform partial matching of shapes. While there has been a progress from matching shapes globally to matching partial segments of the shapes, there is still no work that looks at the importance of various parts of the shape.

Most current-day shape descriptors, whether they are global shape descriptors [1, 2, 5], or partial shape descriptors [4, 6], are obtained with the help of uniformly spaced points on the object's contour. Hence, there is an implicit assumption about the uniform importance of all parts of the shape. However, as we have seen from the example in Figure 1, there are some parts of the shape that need to be given more importance than others.

Xu et al. [7] measure the contour flexibility at various points on the contour. They calculate the bendable potential at each point on the contour and show that flexibility is a characteristic of the shape, which is not uniformly distributed across the complete contour. This work shows that certain parts of the contour have different characteristics compared to the rest of the contour. The bendable potential, in the work of Xu et al., was not calculated using the knowledge about the classes that the shape belongs to. We, on the other hand, exploit the information present in the class labels to identify regions of the shape that are highly distinctive when compared to others.

The closest work related to the identification of distinctiveness/uniqueness of images that we could find was not in the area of shape-matching, but in the area of image memorability. The authors of [8] and [9] try to identify "memorable" regions of an image by looking at some of the intuitive features extracted from the same. They try to understand what is it about an image region that makes it memorable. In similar vein, we try to identify what is it about a shape that catches our eye, or makes it distinctive.

3. DISCRIMINATIVE PARTS

3.1. Extracting Contour Segments

Following a learning paradigm, let us consider the set of all positive shapes to be $\mathcal{S}^+ = \{S_1^+, ..., S_{|\mathcal{S}^+|}^+\}$, and the set of all negative shapes to be $\mathcal{S}^- = \{S_1^-, ..., S_{|\mathcal{S}^-|}^-\}$. By our definition of important parts, we need to find those parts in a given shape S_i^+, which maximally discriminates it, among all other negative shapes, $\mathcal{S}^-$. In the classification literature, good decision boundaries are usually obtained by comparing positive class instances with "hard" negative examples. Hard negatives are those instances that usually turn up as false positives. We take a similar approach while comparing the shape S_i^+ to the set $\mathcal{S}^-$.

We identify the "hardest" example as,

$$S_h^- = \min_{S_j^- \in \mathcal{S}^-} \Psi_{IDSC}(S_i^+, S_j^-), \tag{1}$$

where S_h^- is the hardest example, $\Psi_{IDSC}(S_i^+, S_j^-)$ is the cost obtained while comparing the positive example S_i^+ with a negative example S_j^- using IDSC [2]. Considering that there are noisy instances in the negative class, we not only compare the shape S_i^+ to its hardest counterpart, but to its k hardest counterparts.

Given a shape S_i, we extract its contour $\mathcal{C}_i$. For identifying distinctive parts, we first split the contour into multiple contour segments

$$\mathcal{C}_i^{seg} = \{C_i^{seg_1}, C_i^{seg_2}, ..., C_i^{seg_p}\}. \tag{2}$$

Dividing a shape into multiple semantic regions is a difficult task. The usual tendency is to make cuts on the contour at regions of extreme curvature. However, one then has to find good thresholds to define "extreme curvature". In our approach, we do not make the cuts at corner points, or points of extreme curvature. We randomly split the contour into p segments, each of equal length. We repeat this procedure T times. So, we end up with T random contour cuts, with each cut dividing the contour into p parts. Randomly cutting the contour into different parts does not need any manual supervision in terms of threshold selection, or identification of "meaningful" parts. In Section 4, we show from examples that randomly cutting the contour into parts does not affect the identification of semantically meaningful segments.

To compare contour segments of the i-th positive shape, $\mathcal{C}_i^{seg^+}$, against the contour segments obtained from the k hard negative examples, $\mathcal{C}_{h_{(1,...,k)}}^{seg^-}$, we need a descriptor to describe the segments. It is important that the shape descriptor be a partial shape descriptor and not a global shape descriptor, as we are trying to compare parts of different shapes.

We choose the shape descriptor proposed by Donoser et al. [4], as it has many good properties. Firstly, it is simple to compute the descriptor. The descriptor of a contour segment with n landmark points is a matrix of size $n \times n$, with the (u, v)th entry given as,

$$\alpha_{uv} = \angle(L_{u,v}, L_{v,v-\Delta}), \tag{3}$$

where u and v are two points on the contour, $L_{u,v}$ is the line segment joining the two points, and Δ is an indicator of the number of points before point v with respect to which the angle α_{uv} is calculated. Secondly, the descriptor is invariant to rotations. And, finally, this shape descriptor can be used for quick computation of distances between any two contour segments of any length. The authors of [4] make use of the integral image data structure to perform this efficient computation. More details about their descriptor can be found in their paper [4].

3.2. Distinctiveness of a Segment

The most distinctive positive segment is the one that is least likely to be found among the segments obtained from the negative classes. Therefore, we define the distinctiveness of a particular contour fragment to be inversely proportional to the likelihood of finding it, given the negative classes. The likelihood can be calculated in many ways. We can either train a max-margin classifier and find the distance of each segment to the margin, or we can model the distribution of the positive and negative segments using some prior knowledge, and calculate the likelihood of a new contour fragment belonging to either of the two classes. In this paper, we compute the distinctiveness of a segment using the distances obtained by matching it to other negative contour segments.

Let $d(s,t)$ denote the cost of matching the s-th segment from a positive shape S_i^+, to the t-th segment from a negative shape S_j^-. $d(s,t)$ is the normalized Frobenius Norm between the respective shape descriptors. The probability of a segment being distinctive is then directly proportional to the distance between that segment and other negative segments. We define the log-likelihood of a segment being distinctive as

$$\log P(C_i^{seg_s^+} | \mathcal{C}_{h_{(1,\ldots k)}}^{seg^-}) = \frac{1}{|\mathcal{C}_{h_{(1,\ldots k)}}^{seg^-}|} \sum_j d(seg_s^+, seg_j^-), \tag{4}$$

where $j \in \{1, ..., |\mathcal{C}_{h_{(1,\ldots k)}}^{seg^-}|\}$, and $|.|$ denotes the size of the set. To reduce notational clutter, from now on, we denote $C_i^{seg_s}$ as seg_s, $\log P(C_i^{seg_s^+} | \mathcal{C}_{h_{(1,\ldots k)}}^{seg^-})$ simply as $\mathcal{L}(seg_s^+)$, and $P(C_i^{seg_s^+} | \mathcal{C}_{h_{(1,\ldots k)}}^{seg^-})$ as $P(seg_s^+)$. The probability of a particular segment, seg_s^+, being distinctive can now be written as

$$P(seg_s^+) = \frac{1}{Z} \exp(\mathcal{L}(seg_s^+)). \tag{5}$$

Z is the normalization constant, which is calculated as

$$Z = \sum_{seg_s^+ \in \mathcal{C}_i^{seg^+}} \exp(\mathcal{L}(seg_s^+)). \tag{6}$$

From our experiments, we found that the log-likelihood Eq. 4 need not be calculated as the average distance over all negative segments. It can be obtained using the average distance over just the k closest segments to the segment under consideration.

We use the method just described to calculate the distinctiveness probability for all segments. Taking the average probability over all random cuts gives us the heat map that was shown in Figures 1c and 1d. More examples are shown in Figure 3. As can be seen from these figures, our method is able to consistently identify discriminative regions that are semantically meaningful.

3.3. Part Consistency

The method described in the previous two subsections helps us to get the heat maps of a particular shape, which indicates the importance of a particular part. If there are many object instances in a particular class, we would like that the heat maps of different objects indicate similar importance to similar parts. In this subsection, we propose a way to perform this check.

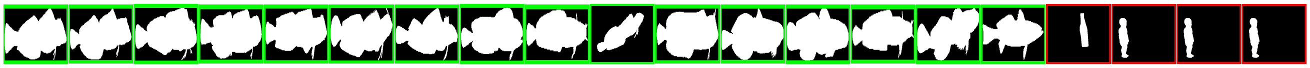
We start off by finding correspondences between points on pairs of shapes. This can be achieved using IDSC. IDSC not only finds the cost of matching two shapes, but it also gives a correspondence between points on the shapes using dynamic programming. We randomly choose a particular shape as the reference shape, S_{ref}, and calculate the point correspondences between the reference shape and every other shape in the class. For each point on the shape, we can also obtain the distinctiveness value using the method described above. Thus, for M objects in a class, each with N points on them, we can generate an $N \times M$, matrix of distinctiveness values, which we denote as Q. The entry $Q(r,c)$ corresponds to the distinctiveness value of a point in shape S_c, whose corresponding point in the reference shape, S_{ref}, is r.

If the importance values of all the corresponding points agree with each other, then the singular values of the matrix, Q, arranged in descending order, should drop off steeply. We quantify this steep drop using the area under the curve (AUC) of the plot of the top M singular values. The larger the AUC, the more disagreement there is between the corresponding points of shapes in a class. In the next section, we show plots of the singular values and their corresponding AUCs for some example classes.

4. EXPERIMENTS

We conduct our experiments on the well-known MPEG7 CE-Shape-1 Part B dataset. The dataset consists of 1400 images grouped into 70 different classes, with each class containing 20 example images. Most state-of-the-art shape matching techniques get confused while matching shapes that have overall visual similarity. So, we try to identify the distinguishing parts for each object in the database using the method described in Section 3. The top row of Figure 3 shows the output of our algorithm for 5 instances of a particular class, 'Apples'. We can see that even though our method uses random cuts to segment out parts, we are still able to localise on the semantically meaningful segments. Also worth noting is that our method is able to consistently identify the same segments, across all instances, as the discriminative segment of

(a) Uniform Sampling



(b) Importance Sampling

Fig. 2: Figure shows the top 20 best matching objects to the fish shown in first column. The green boxes indicate correct retrievals, while the red boxes indicate incorrect retrieval. (a) Retrieval results from uniformly sampled IDSC. (b) Retrieval results using IDSC generated using importance sampling. Importance values were obtained using the method described in Section 3.

apples.

Figure 4 shows the plots of the singular values for some example classes. We have also shown the corresponding AUC on the graph. The agreement among what is considered as distinctive among class ‘Apples’ (Figure 3) is confirmed by the small AUC. For classes such as camels, which are articulating objects, the articulation causes different parts to be identified as distinctive. However, we found that the humps of the camel did stand out to be more distinctive than other parts (final object in row 3 of Figure 3). Figure 3 also shows more examples outputs from our algorithm.

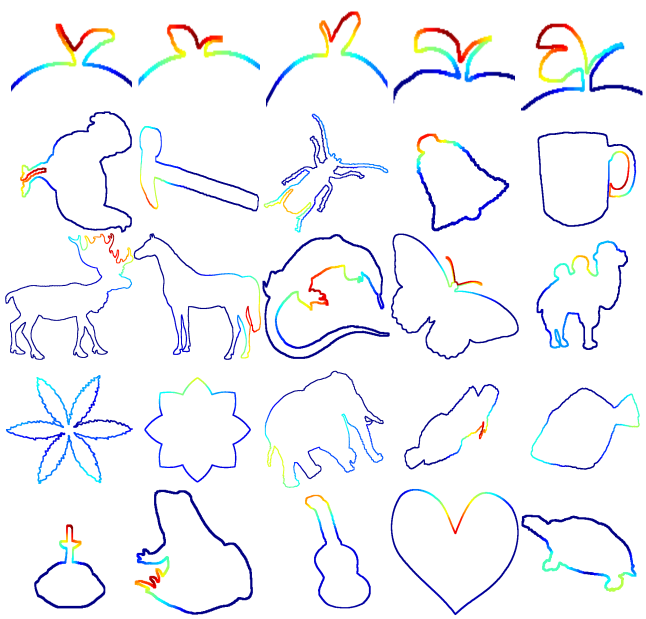


Fig. 3: [Best viewed in colour] Figure shows the importance heat maps of some example objects generated using the method described in Section 3. Top row shows the cropped-out heat maps of 5 apples. Notice that our method is able to consistently identify the semantically meaningful regions.

Finally, we implemented a variant of IDSC that used non-uniform sampling as opposed to the regularly used uniform sampling of the object contours. We sampled points based on the heat values at different locations of the object. More points were sampled from regions that were considered to be more important, and less points from regions that were less

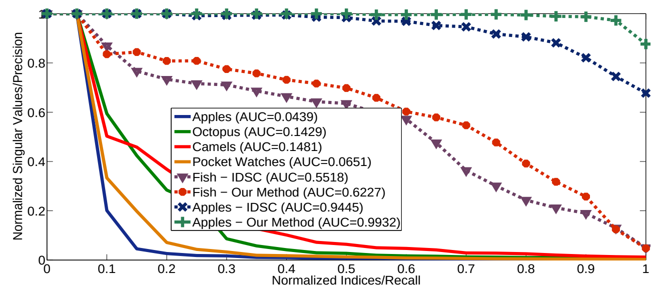


Fig. 4: Bold lines show the singular value plots. The dashed lines are the Precision/Recall plots.

important. We then generated the IDSC shape descriptor at each of the original uniformly sampled landmark points. The descriptor is now affected by the extra “important” points. We can consider the new descriptor as a weighted generalization of the uniformly weighted shape context. Figure 2 shows a comparison of the retrieval results obtained using the original descriptor and our importance sampling based descriptor. We can see that our descriptor is able to retrieve more correct examples than the original IDSC, which gets confused by globally similar objects.

Figure 4 also shows the precision/recall plots for two example classes. The P/R plots are calculated as in [10]. Ideally, we would like the P/R plot to be at 100% precision for all levels of recall. So, greater the area under the curve, better is the retrieval method. As can be seen from the plots, our method of generating the descriptor (using importance sampling of the contour) helps improve the precision of the retrievals.

5. CONCLUSION

In this paper, we have introduced the concept of distinctive parts. We have explained why it is important to look out for important regions in contours. We provide a simple way to extract the distinctive parts of any given object. Results from our experiments shows that the method is promising. As a part of future work, we would like to explore how we can make better use of the important discriminative regions to come up with more discriminative shape-matching techniques.

6. REFERENCES

[1] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 509–522, 2002.

[2] H. Ling and D.W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 286–299, 2007.

[3] A.M. Bronstein, M.M. Bronstein, A.M. Bruckstein, and R. Kimmel, "Partial similarity of objects, or how to compare a centaur to a horse," *Int. J. of Computer Vision*, vol. 84, no. 2, pp. 163–183, 2009.

[4] M. Donoser, H. Riemenschneider, and H. Bischof, "Efficient partial shape matching of outer contours," in *ACCV*. 2009, pp. 281–292, Springer.

[5] V. Premachandran and R. Kakarala, "Perceptually motivated shape context which uses shape interiors," *arXiv preprint arXiv:1212.4608*, 2012.

[6] T. Ma and L.J. Latecki, "From partial shape matching through local deformation to robust global shape similarity for object detection," in *CVPR*, 2011, pp. 1441–1448.

[7] C. Xu, J. Liu, and X. Tang, "2d shape matching by contour flexibility," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 1, pp. 180–186, 2009.

[8] D. Parikh, P. Isola, A. Torralba, and A. Oliva, "Understanding the intrinsic memorability of images," *Journal of Vision*, vol. 12, no. 9, pp. 1082–1082, 2012.

[9] A. Khosla, J. Xiao, A. Torralba, and A. Oliva, "Memorability of image regions," in *NIPS*, Lake Tahoe, USA, December 2012.

[10] E. Baseski, A. Erdem, and S. Tari, "Dissimilarity between two skeletal trees in a context," *Pattern Recognition*, vol. 42, no. 3, pp. 370–385, 2009.