AGGREGATING CONTOUR FRAGMENTS FOR SHAPE CLASSIFICATION

Song Bai, Xinggang Wang, Xiang Bai

Dept. of Electronics and Information Engineering, Huazhong Univ. of Science and Technology

ABSTRACT

In this paper, we address the problem of building a compact representation for shape. We first decompose shape into meaningful contour fragments, and each fragment is described by a certain descriptor, *e.g.*, Shape Context. Then inspired by the coding scheme Vector of Locally Aggregated Descriptors widely used in image representation, we try to aggregate the contour fragments into a very compact vector of limited dimension to stand for a shape, and we name the new designed shape descriptor as Vector of Aggregated Contour Fragments (VACF). We apply VACF to shape classification task on the well-known MPEG-7 shape benchmark, and the experimental results show that the accuracy of our proposed method outperforms other state-of-the-art algorithms with much smaller memory usage.

Index Terms— VACF, Compact Representation, Shape Classification, Contour Fragments

1. INTRODUCTION

Shape is an informative and distinctive feature of an image. Human beings usually perceive an object by its shape first. Shape classification is a fundamental problem in computer vision. Given a set of training shapes with their labels known, shape classification is to determine the category of the shapes in the testing set. Traditional shape classification algorithms usually start with designing a robust shape descriptor [1, 2, 3]. Then for the training shapes, correspondences are found between each other using matching method such as dynamic programming. The final decision is made by a simple nearest neighbor (NN) classifier. The limitation of the traditional classification methods is obvious. It is time consuming to perform the pairwise matching. When applied to large database, traditional matching-based methods is impractical to be conducted due to its high time complexity.

Inspired by the Vector of Locally Aggregated Descriptors (VLAD) [4] that is used for image representation, we try to build a distinctive and compact shape representation to fulfill the task of shape classification efficiently and precisely. However, VLAD are based on many mature local descriptors designed specifically for image, such as SIFT [5], and it is difficult to apply VLAD to shape analysis due to the lack of local descriptors for shape. Our solution is to decompose the outer contour of shape into fragments [6, 7], which is described by a certain shape descriptor. The contour fragments are regarded as the raw basic shape descriptions, and they are aggregated again to form the final descriptor for shape with a coding scheme similar to VLAD. The new designed descriptor for shape is named as Vector of Aggregated Contour Fragments (VACF).

Our proposed method discards the pairwise matching algorithm, and propose a compact representation for shape. The "compact" here not only means that our descriptor is merely a single vector in place of a set of vector conventionally, but also indicates that the vector requires small memory usage with limited vector dimension. The property of compactness is really important to the training stage and the testing stage later, since it can reduce the running time and decrease the memory usage largely.

Moreover, the distinctive classifier such as SVM can be used to select the representative contour fragments to handle the problem brought by the large shape deformation. The experimental results show that our proposed method outperforms other state-of-the-art algorithms in terms of shape classification precision.

The remainder of this paper is organized as follows: We review some related shape classification methods in Section 2. In Section 3, we introduce the detailed information of VACF. Section 4 propose some methods to improve the performance of VACF. The experimental results are presented in Section 5 to show the advantage of VACF. Conclusions are given in Section 6.

2. RELATED WORK

In the past years, many algorithms have been proposed for shape classification in the literature. Bai *et al*. propose a skeleton-based approach to match skeleton graphs using geodesic paths in [8]. In [9], a directed and acyclic shock graph is defined to conduct shape matching. Meanwhile, some contour-based methods are also proposed. Belongie *et al*. [1] propose a popular shape descriptor called Shape Context, and Ling and Jacobs [2] replace Euclidean distance

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61222308; by the Program for New Century Excellent Talents in University in China under Grant NCET-12-0217; by the Fundamental Research Funds for the Central Universities under Grant HUST.

used in Shape Context by geodesic distance to handle the articulated shapes. Height functions are defined to describe the contour of the shape in [3]. In [10], a three-level framework that consists of models for contour segments is used to shape classification with Bayesian classifier. The contour of shape is mapped into a string of symbols in [11] and a modified edit distance is used to compute the similarity between strings of symbols in [12, 13]. The combination of contour and skeleton for shape classification is well studied in [14] with a gaussian mixture model.

VACF is partly inspired by Vector of Locally Aggregated Descriptors (VLAD) that proposed in [4] to solve the task of natural image representation. VLAD can be regraded as a non probabilistic Fisher Kernel [15] that is designed for large scale image classification, and VALD emphasizes on the very large scale image retrieval. VLAD accumulates the local image descriptors to form a compact representation for image. As is discussed above, it is difficult to directly apply VLAD to shape analysis, so the outer contour of the shape is decomposed into fragments by Discrete Contour Evolution (DCE) [6] to obtain the local shape descriptors. DCE obtains a novel rule called hierarchical convexity rule to identify convex parts at different stages of a proposed contour evolution method. Closest to our work is Wang *et al*. [7]. They propose Bag of Contour Fragments (BCF) to form the mid-level representation of shape, however VACF is more compact than BCF, and is more suitable for large scale shape classification.

VACF aggregates the contour fragments in a similar way to VLAD, however several differences also exist. VACF focuses directly on the analysis of shape based on the contour fragments, while VLAD is mainly about the image representation. We also provide several methods to improve the performance of VACF in shape classification, such as principal component analysis.

3. THE VECTOR OF AGGREGATED CONTOUR FRAGMENTS

In this section, we show in detail the algorithm of decomposing a shape into contour fragments, and the coding scheme of aggregating the fragments.

3.1. Extracting Contour Fragments

Contour fragment is an informative feature for shape recognition, since they can reveal both local and global information. Our proposed VACF also adopts the contour fragments as the raw descriptions for shape. In order to get meaningful contour fragments, we utilize a robust algorithm called Discrete Contour Evolution (DCE) [6].

Given the outer contour $\mathcal{S}$ of a certain shape, DCE is applied to obtain a simplified polygon $\nu = \{v_1, v_2, \ldots, v_{n_v}\}$ on $\mathcal{S}$. The vertex of the polygon is denoted by v_i $(i = 1, 2, ..., n_v)$, and n_v is the number of vertex. By tracing the path along the contour from v_i to v_j, we can obtain a fragment f_{ij}. All the contour fragments $\mathcal{F}(\mathcal{S})$ of the contour $\mathcal{S}$ can be obtained after enumerating all the pair of vertices

$$\mathcal{F}(\mathcal{S}) = \{f_{ij} = (v_i, v_j), i \neq j, i, j \in [1, 2, \ldots, n_v]\} \quad (1)$$

Notice that f_{ij} and f_{ji} are not the same, since their directions are opposite. An illustration for the extraction of fragments is showed in Fig. 1.

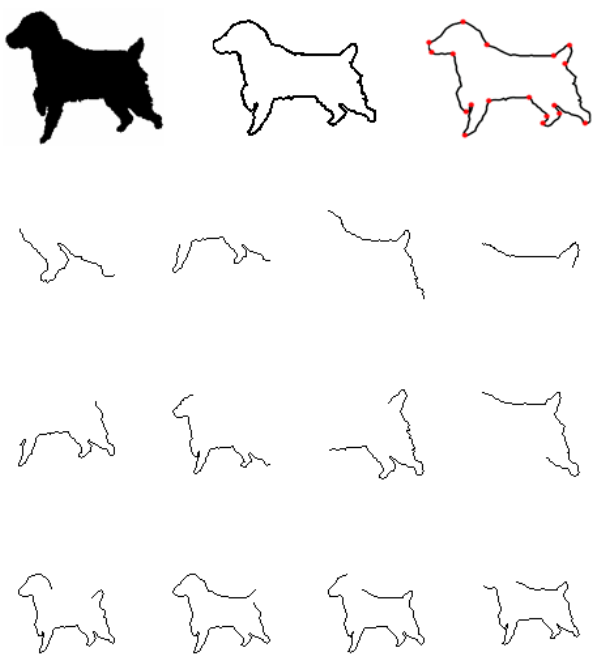

Fig. 1. The illustration of decomposing the contour into fragments. The first row shows the original shape and the corresponding outer contour. The red points represent the vertexes of the polygon obtained by DCE. Some typical contour fragments of different lengths are presented in the second to the fourth row.

Finally, each fragment is described by Shape Context to obtain a vector representation $x_{ij} \in \mathbf{R}^{d \times 1}$. We equidistantly get 5 reference points on the fragment f_{ij}, and for each point we compute the shape context histogram, which are cascaded later to form the fragment descriptor x_{ij}. Note that the dimension of the shape context histogram is $d = 60$ (12 for dividing angle space and 5 for dividing radius space), so the feature vector dimension of a contour fragment is $5 \times 60 = 300$.

3.2. Aggregating Contour Fragments

After these contour fragments are obtained that can be assumed as the local descriptors for shape, our proposed method pays much attention to aggregating the contour fragments to

generate a much compact representation for shape with limited codebook size and limited vector dimension. A natural solution is VLAD proposed for image analysis in [4].

We first learn a small codebook $\mathcal{B} = \{b_1, b_2, ..., b_K\} \in \mathbf{R}^{d\times K}$ off-line with k-means (where the size of codebook K is typically 10 or 20). Then each contour Fragment x_{ij} in a certain shape is assigned to the closest cluster in the codebook. For each cluster b_k, the residual $x_{ij} - b_k$ of the contour fragment x_{ij} assigned to b_k is recorded. In order to accumulate the residuals, we sum up all the residuals for the k-th visual word as

$$\mathcal{R}_k = \sum_{x_{ij}:NN(x_{ij})=b_k} x_{ij} - b_k \tag{2}$$

At last, Let $\mathcal{R}$ be the concatenation of the aggregated residuals: $\mathcal{R} = [\mathcal{R}_1\mathcal{R}_2 \dots \mathcal{R}_K] \in R^{D\times 1}$, where $D = d \times K$. R is subsequently normalized as $\mathcal{R} := \mathcal{R}/||\mathcal{R}||_2$, and the new designed descriptor for shape is called Vector of Aggregated Contour Fragments (VACF).

4. THREE METHODS TO IMPROVE VACF

The burstiness effect [16], which is observed in text retrieval initially, is a phenomena that some visual words appear more times in an article than a statistically independent model would predict. We assume that the same phenomenon also exist in shape analysis when we decompose a shape into contour fragments. It is natural to find that some bursty fragments in a shape dominate the visual similarity computed between VACFs. In this section, we offer three methods to improve the performance of VACF in view of burstiness.

4.1. Power Normalization

The Power Normalization (PN), which is introduced in [15], can increase the small feature values, and also restrain the high feature values with a simple algebraic manipulation, which means that it can reduce the burstiness effect efficiently.

As is proved later, Power Normalization can significantly improve the performance of VACF. Before VACF is finally L_2-normalized, a element-wise nonlinearity operation is defined for each component $r_i (i = 1, 2, ...D)$

$$r_i := |r_i|^\alpha \times sign(r_i) \tag{3}$$

where the value of α is between 0 and 1. The VACF vector is subsequently L_2 normalized. In practise, α is set to 0.2 throughout our experiments.

4.2. Residual Average Pooling

Note that standard VACF, defined in Eq. 2, sums up all the residuals, which results in that the contour fragments contribute unequally to describe the shape. In [17], Jegou *et al*. propose Residual Normalization (RN) to solve the same problem in image analysis. However, as for the case of shape analysis, it does not make much difference probably due to the fact that property of contours fragments is different to the local image descriptors, such as SIFT.

We propose to shorten the big gap for describing the shape between different clusters, while keep their distinctive abilities through averaging the residual by the number of fragments assigned to them. We name it Residual Average Pooling (RAP).

$$\tilde{\mathcal{R}_k} = \frac{1}{N_k} \sum_{x_{ij}:NN(x_{ij})=b_k} x_{ij} - b_k \tag{4}$$

where N_k denotes the number of contour fragments assigned to the k-th cluster for a certain shape.

4.3. Principal Component Analysis

Principal Compone Analysis (PCA) is a linear dimension reduction algorithm, which is adopted in a widely variety of applications in computer vision.

In our experiment, PCA is applied directly to VACF to reduce the the dimension of VACF significantly (From D to $\bar{D}$ as is showed in Table 1). It makes much sense especially when VACF is applied to big data. As is showed in the experiment later, the discriminative ability of VACF is not impaired too much, and on the contrary it is sometimes beneficial because of the feature selection ability of PCA. We will discuss the effect of PCA in the experiment.

5. EXPERIMENTS

We evaluate our proposed method for shape classification on the widely-used MPEG-7 shape dataset [18]. It consists of 1400 silhouette images grouped into 70 classes, and each class has 20 shapes. Some typical shapes are presented in Fig. 2. Considering that our descriptor for shape is a simple and compact vector, we directly adopt SVM as the classifier with the fast linear SVM toolbox [19]. Two strategies are used to obtain the classification accuracy: (1) Half Training. We randomly select half of shapes per class for training the classification model, and the rest is used for testing. This process is repeated for 20 times to decrease the uncertainty of

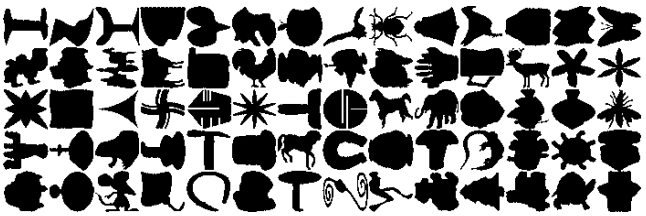

Fig. 2. Some typical shapes from MPEG-7 dataset

Methods	K	D	$\bar{D}$=D	$\bar{D}$=450	$\bar{D}$=300	$\bar{D}$ = 150
VACF	20	6000	$94.01 \pm 0.85\%$	$94.19 \pm 0.88\%$	$93.98 \pm 0.79\%$	$91.35 \pm 0.80\%$
VACF+RAP	20	6000	$95.18 \pm 0.79\%$	$94.82 \pm 0.65\%$	$94.44 \pm 0.97\%$	$92.03 \pm 0.85\%$
VACF+PN	20	6000	$96.88 \pm 0.68\%$	$97.32 \pm 0.58\%$	$97.10 \pm 0.71\%$	$96.39 \pm 0.81\%$
VACF+PN+RAP	20	6000	$97.10 \pm 0.83\%$	$97.03 \pm 0.67\%$	$97.09 \pm 0.68\%$	$96.38 \pm 0.73\%$

Table 1. The classification accuracy (Half training) of VACF with different improving methods before and after dimension reduction with PCA. K denotes the codebook size, and D is the full feature dimension of VACF. $\bar{D}$ is the dimension of VACF after PCA is applied.

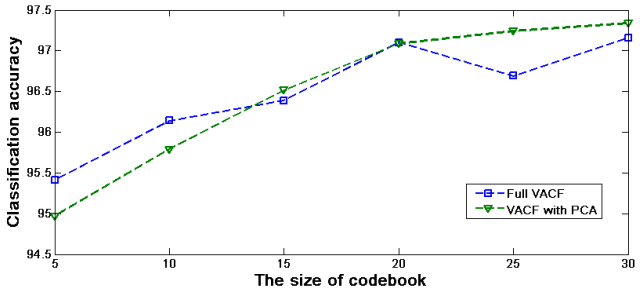

Fig. 3. The effect of codebook size on the classification accuracies with half training. The size of codebook ranges from 5 to 30, which is rather small. The blue line and the green line denote classification accuracy before and after the dimension of VACF is reduced to 300.

performance brought by the random selection of training set. (2) Leave-one-out. Training with all shapes leaving one out for testing and repeating this procedure for all shapes in the dataset.

We discuss the effect of PN, RAP and PCA in Table 1. It can be drawn from the table that the standard VACF already performs better than many other shape classification algorithms. Power Normalization improves the classification accuracy significantly without additional memory and time usage. RAP can also enhance the discriminative ability of VACF, and produce a more faithful result when it interplays with Power Normalization. PCA is applied to reduce the dimension of VACF in an unsupervised way. PCA can preserve the local structure of the data manifold largely, so the performance of VACF with dimension reduction is also comparable. An interesting phenomena is that sometimes PCA improves the performance on the contrary, and our interpretation is that the ability of feature selection in PCA is amplified by SVM.

The influence of the codebook size is depicted in Fig. 3. We can find that the performance increases with the size of coodbook increasing in general. Notice that the division of contour fragments in VACF is rather coarse, and we get $20k$ fragments per cluster on average when the size of codebook is fixed to 20. The property of the small codebook size makes it fast to conduct the procedure of codebook learning.

The comparison between VACF and some related shape

Algorithm	Half training	Leave-one-out
Skeleton paths [14]	86.7%	-
Class segment set [10]	90.9%	-
Contour segments [14]	91.1%	-
ICS [14]	96.6%	-
String of symbols [11]	-	97.36%
Robust symbolic [12]	-	98.57%
Kernel-edit distance [13]	-	**98.93%**
VACF	**$97.32 \pm 0.58\%$**	98.64%

Table 2. Classification accuracy of different algorithms on MEPG-7 dataset.

classification methods is presented in Table 2. The proposed VACF, in combination with PN and PCA, obtains the highest classification accuracy of 97.32% when half training is applied. The experimental results show that VACF is much more powerful to solve the task of shape classification. What is more important is the memory usage of VACF. As is discussed above, the dimension of VACF is rather small, we can get a competitive classification accuracy when the dimension of VACF is only 300.

6. CONCLUSION

In this paper, we develop a compact representation for shape based on the contour fragments. Our proposed method Vector of Aggregated Contour Fragment (VACF) outperforms other state-of-the-art shape classification algorithms with much smaller feature dimension. VACF saves time and memory usage both on-line and off-line, which is suitable for big data and necessary for mobile device.

In the future, we will study how to apply the proposed VACF to web scale hand-written and sketch image retrieval.

7. REFERENCES

[1] Serge Belongie, Jitendra Malik, and Jan Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[2] Haibin Ling and David W Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

[3] Junwei Wang, Xiang Bai, Xinge You, Wenyu Liu, and Longin Jan Latecki, "Shape matching and classification using height functions," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 134–143, 2012.

[4] Hervé Jégou, Florent Perronnin, Matthijs Douze, Cordelia Schmid, et al., "Aggregating local image descriptors into compact codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1704–1716, 2012.

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

[6] Longin Jan Latecki and Rolf Lakämper, "Convexity rule for shape decomposition based on discrete contour evolution," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 441–454, 1999.

[7] Xinggang Wang, Bin Feng, Xiang Bai, Wenyu Liu, and Longin Jan Latecki, "Bag of contour fragments for robust shape classification," *Pattern Recogination*, 2014.

[8] Xiang Bai and Longin Jan Latecki, "Path similarity skeleton graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1282–1292, 2008.

[9] Kaleem Siddiqi, Ali Shokoufandeh, Sven J Dickinson, and Steven W Zucker, "Shock graphs and shape matching," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 13–32, 1999.

[10] Kang B Sun and Boaz J Super, "Classification of contour shapes using class segment sets," in *Computer Vision and Pattern Recognition*. IEEE, 2005, vol. 2, pp. 727–733.

[11] Mohammad Reza Daliri and Vincent Torre, "Shape recognition and retrieval using string of symbols," in *International Conference on Machine Learning and Applications*. IEEE, 2006, pp. 101–108.

[12] Mohammad Reza Daliri and Vincent Torre, "Robust symbolic representation for shape recognition and retrieval," *Pattern Recognition*, vol. 41, no. 5, pp. 1782–1798, 2008.

[13] Mohammad Reza Daliri and Vincent Torre, "Shape recognition based on kernel-edit distance," *Computer Vision and Image Understanding*, vol. 114, no. 10, pp. 1097–1103, 2010.

[14] Xiang Bai, Wenyu Liu, and Zhuowen Tu, "Integrating contour and skeleton for shape classification," in *Computer Vision Workshops (ICCV Workshops)*. IEEE, 2009, pp. 360–367.

[15] Florent Perronnin, Jorge Sánchez, and Thomas Mensink, "Improving the fisher kernel for large-scale image classification," in *European Conference on Computer Vision*. 2010, pp. 143–156, Springer.

[16] Hervé Jégou, Matthijs Douze, and Cordelia Schmid, "On the burstiness of visual elements," in *Computer Vision and Pattern Recognition*. IEEE.

[17] Jonathan Delhumeau, Philippe-Henri Gosselin, Hervé Jégou, and Patrick Pérez, "Revisiting the vlad image representation," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 653–656.

[18] Longin Jan Latecki, Rolf Lakamper, and T Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Computer Vision and Pattern Recognition*. IEEE, 2000, vol. 1, pp. 424–429.

[19] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.