

## Research Article

# Improving Shape Retrieval by Integrating AIR and Modified Mutual $k$ NN Graph

Nouman Qadeer,<sup>1</sup> Dongting Hu,<sup>1</sup> Xiabi Liu,<sup>1</sup> Shahzad Anwar,<sup>1</sup> and Malik Saad Sultan<sup>2</sup>

<sup>1</sup>Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup>Department of Computer Science, Faculty of Science, University of Porto, 4169-007 Porto, Portugal

Correspondence should be addressed to Nouman Qadeer; noumansoomro@gmail.com

Received 3 October 2014; Accepted 5 January 2015

Academic Editor: Jianping Fan

Copyright © 2015 Nouman Qadeer et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In computer vision, image retrieval remained a significant problem and recent resurgent of image retrieval also relies on other postprocessing methods to improve the accuracy instead of solely relying on good feature representation. Our method addressed the shape retrieval of binary images. This paper proposes a new integration scheme to best utilize feature representation along with contextual information. For feature representation we used articulation invariant representation; dynamic programming is then utilized for better shape matching followed by manifold learning based postprocessing modified mutual  $k$ NN graph to further improve the similarity score. We conducted extensive experiments on widely used MPEG-7 database of shape images by so-called bulls-eye score with and without normalization of modified mutual  $k$ NN graph which clearly indicates the importance of normalization. Finally, our method demonstrated better results compared to other methods. We also computed the computational time with another graph transduction method which clearly shows that our method is computationally very fast. Furthermore, to show consistency of postprocessing method, we also performed experiments on challenging ORL and YALE face datasets and improved baseline results.

## 1. Introduction

Content based image retrieval (CBIR) remained an important and challenging problem in computer vision from many decades [1]. CBIR processes take a query such as query by image, region, or sketch and return most similar images from database based on similarity of low level features instead of textual annotations or media metadata [2–4]. Generally, image retrieval is divided into two steps. The first step consists of shape, texture, and color feature extraction of image or object and the second step involves similarity measure of query image with all database images and then images are ranked according to similarity measure. Detailed discussions can be found in [1–4]. Our method focused on shape based feature extraction and retrieval of binary images. Although flood of color images are on its peak, importance of binary images could not be neglected as they are widely used in trademark images, patent images, technical drawings, or some applications such as medical images [5], botanical

collections [6], and road signs [7] and in many different areas. Shapes are a primary source of information for feature extraction of binary as well as color images. Shapes images are taken directly from binary images (i.e., black and white) or threshold in gray images where one side of threshold will be white and the other side will be black as there is little or no information of texture and there is no color information. Traditionally, feature extraction or feature representation remained to be a crucial step for improving retrieval rate. Lately, some similarity measures methods also improved retrieval accuracy rate. Recently, some postprocessing methods contributed significantly in image retrieval accuracy, such as graph transduction, cotransduction, spectral matching, and metasimilarity [8–11].

In this paper we propose integration of articulation invariant representation (AIR) feature extraction method with postprocessing, modified mutual  $k$ NN graph, manifold learning method to improve the shape retrieval score with respect to context information. At first, we extracted

features of each shape using AIR method [12]. After feature extraction, dynamic programming (DP) was used to measure pairwise shape similarity. After achieving distances by using DP, we applied modified mutual  $k$ NN graph [13]. For proving the efficiency and effectiveness of our method we conducted experiments on popular MPEG-7 database. We calculate accuracy using so-called bulls-eye score. Our method improved the baseline results of AIR method. We compared our method with other rival methods and our method got good results compared to other methods. We, also, did experiments with respect to computational time as computational time is another challenging area of content based image retrieval. To show the computational efficiency of our method we integrated AIR representation method with graph transduction method [14] and we got almost the same retrieval accuracy results but on the cost of huge computational time which showed our method is promising and efficient.

To the best of our knowledge, AIR feature representation is a baseline method, till now, for shape feature representation on planar shapes, especially on challenging MPEG-7 database. Although histogram based shape context (SC) [15] feature representation remained to be a popular method for shape images, it lacks articulation invariance and the accuracy level of SC is lower than AIR. For shape matching or similarity measure, DP is more robust to outlier and noise, and it is also invariant to scale, rotation, and translation. Also, it proved to be an efficient matching algorithm in perspective of accuracy. Since similarity measure methods are not ideal in calculating interclass and extraclass distances, thus we used postprocessing method, modified mutual  $k$ NN graph, to achieve similarity with respect to context information. It gives the affinity-wise similarity and in our method these similarities yield very good results of retrieval compared to other methods. Moreover, modified mutual  $k$ NN graph is computationally inexpensive.

While we reproduced the results of inner distance shape context (IDSC) with modified mutual  $k$ NN graph [13] which clearly showed deterioration of results compared to our method, to further prove the consistency of modified mutual  $k$ NN graph we integrate it with different feature representations on ORL and YALE face databases and improved the baseline results, respectively.

## 2. Related Work

There are three subproblems in shape based image retrieval. The first problem is feature extraction for good representation of shape. The second problem is similarity measure or shape matching which calculates the distance of query image with the rest of images in the database and reranks those images with respect to query image. The third problem addressed the postprocessing methods to obtain similarity with respect to context information. Our discussion will be around these three problems.

Feature extraction and shape matching are very active research area in literature [16]. Similarity of shape has also been very well analyzed in the field of psychology [17]. In literature, many descriptors and similarity methods have

been reported but we focused more on the methods of silhouette images [18]. Based on shape contours, curvature scale space (CSS) and some varying methods are used for simple binary images [19–21]. CSS method removes the irrelevant inflections. After removing inflections, it checks the characteristics of object. Recently, many efforts have been dedicated to overcome the drawbacks of CSS such as computational cost [21] and inaccurate similarity measure [20, 21].

Shape context (SC) proposed by Belongie et al. described the relative distribution such as distance and orientation [15]. SC combining with Thin Plate Spline (TPS) is considered to be discriminate. Generally, SC achieves good accuracy on the cost of high computational time in feature extraction and image matching. SC was extended by adding statistics of tangent vectors at landmark points [22]. Ling and Jacobs [23] extended the SC by using the geodesic distance between the contour points instead of Euclidean distance to measure the spatial relations between points on shapes. From one point of view, deformation, pose, and self-occlusion cause large variation in a shape class, while, on the other hand, different shape examples may have more identical components and variation can be addressed by employing different configuration of these components [24, 25]. Geometrical, noise, and occlusion invariant shape descriptor, Height Function (HF), is used for shape feature extraction [26]. In HF, predefined sample points of contour of each object are represented and on each sample point height function is calculated which follows the smoothing of height function for robustness. Contour flexibility feature extraction method was developed for feature extraction of both local and global features [27]. Bag of contour fragments (BCF) is an extended form of Bag-of-Words (BoW) for shape images that were used in [28]. In BCF, shape is divided into set of contour fragments and each contour is calculated by SC which follow the encoded into shape code. BCF worked well on trainable classifier but it performed poorly in shape matching. For shape matching Latecki and Lakämper [29] utilized visual parts made up by simplified polygons of contours. Feature-driven generative model for probabilistic shape matching was proposed by Tu and Yuille [30]. Siddiqi et al. [31] used the shape matching using graphical method. Dynamic and hierarchical curve matching method was introduced by Felzenszwalb and Schwartz [24] for avoiding the problems related with local or global methods. Multiscale representation of triangle areas for shape matching was introduced by Alajlan et al. [32], which consist of partial and global shape information. Based on SC, symbolic descriptor was defined by Daliri and Torre, and to overcome the problems caused by occlusions and deformation, edit distance for final matching was used in [33]. On the format of BoW, shape vocabulary was used as shape descriptor [34]. It addressed the issue of time consuming shape matching and speeds up the matching process as distance metric of global descriptors is compared instead of time consuming local features. Edge orientation autocorrelogram (EOAC) [35] is another binary image method. In EOAC, image edge orientations are used. These edge orientations are quantized and were given input to two-dimensional histogram. EOAC was used in PATSEEK

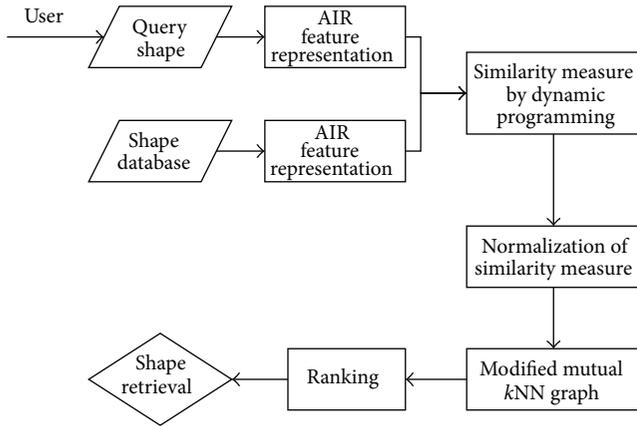


FIGURE 1: Proposed method of shape retrieval.

search engine, developed by US patent office, in an effort to efficiently analyze the US patent image database [36].

As content based image retrieval is resurging more actively, different graphical methods are developing along with feature extraction methods and similarity measures. After obtaining good feature representation and applying matching algorithm to get similarity matrix, manifold learning as a postprocessing method has been used in the context of image retrieval. Improving the ranking of retrieved shapes by employing data manifold structure was proposed by Zhou et al. [8]. Otherwise stated, manifold learning improves the ranking results by keeping the context information of closest objects of database. Earlier developed for semisupervised learning, context information with label propagation of graph transduction for silhouette retrieval was proposed by Bai et al. [9]. It adds together various similarity as well as dissimilarity measures to better establish the relationship among many objects. Furthermore, in the work of Kotschieder et al. [13], the performance was improved by keeping the context information on inner distance shape context representation. Our work is similar to that of Kotschieder et al., but our work used articulation invariant representation instead of inner distance shape context, and our method also outperforms the results of the aforementioned method.

### 3. Proposed Work

Our proposed work consists of two parts. First, we need to extract the features or get a feature representation by using articulation invariant representation (AIR) method [1]. After getting  $N \times N$  pairwise distance matrix  $A$  by using dynamic programming (DP), we applied the modified mutual  $k$ NN graph to improve the similarity. Our method is depicted in Figure 1.

**3.1. Feature Representation.** For feature representation we used AIR, which is geometrical invariant, such as scale, translation, and rotation. Along with these invariances, it is also invariant to articulation. Another reason for using AIR is that it outperformed other feature representation methods on shapes, especially on large intraclass variation MPEG-7

database. As we have shape  $S \subset \mathbb{R}^2$ , such that  $S = \{\cup_i^n p_i\} \cup \{\cup_{i \neq j, 1 \leq i, j \leq n} q_{ij}\}$ , here,  $p_i$  is the part of shape  $S$  and  $q_{ij}$  is the junction of shape  $S$ .  $M$  is the numbers of set of conditions of 2D shapes are given  $\{S_k\}_{k=1}^M$ ; feature representation  $\tilde{R}$  can be found which meets the criteria of (1). Complete glossaries are as follows:

$S$ : 2D projection

$M$ : Fixed number of conditions

$p_i$ : Parts of 2D shape  $S$

$q_{ij}$ : Junction of shape  $S$

$S_k = \{\cup_i^n p_{ik}\} \cup \{\cup_{i \neq j, 1 \leq i, j \leq n} q_{ijk}\}$ , for all  $k = 1$  to  $M$

$\{g_{lk}\}_{l=1}^t$ , set of  $t$  points that constitute  $S_k$

$\tilde{R}$ : The shape representation of (1)

$D$ : Distance

ID: Inner distance

$A_{ij}$ : Shape distance

$W_{ij}$ : Affinity distance

$\epsilon_k$ : Total error

$\epsilon_{P_k}$ : Weak perspective of real world

$\epsilon_{D_k}$ : Distance error

$\epsilon_{S_k}$ : Projection error

$T$ : Part-wise affine normalization to perform transformation

$h_{lk}$ : Histogram of point

$z$ : Total number of bins in histogram  $h_{lk}$

$v$ : Vertex of graph

$e$ : Edge of graph

$G_{mmkNN}(N, k)$ : Modified mutual  $k$ NN graph with  $N$  vertices and  $k$  edges.

Consider

$$\tilde{R}(S_1) = \tilde{R}(S_k), \quad \forall k = 1 \text{ to } M. \quad (1)$$

If  $g_{1k}, g_{2k} \in S_k$  are the two points, then we can measure a respective distance  $D(g_{1k}, g_{2k})$  by

$$D(g_{1k}, g_{2k}) = c, \quad \forall k = 1 \text{ to } M. \quad (2)$$

Here,  $c$  represents a constant. Using above distance we obtained feature representation  $\tilde{R}$  that satisfies (1). Because of many variations in viewpoint or varying effect, part by part affine normalization will be performed, to cope with such changes in shape  $S_k$ . This normalization process will essentially find a transformation  $T$ . So every  $g_{lk} \in S_k$  will be transformed as  $T(g_{lk}) \rightarrow g'_{lk}$ . To find the articulation changes in  $S_k$  we compute the distance of two points  $g'_{1k}, g'_{2k} \in S_k$  using the inner distance (ID), which is articulation invariant:

$$D(g_{1k}, g_{2k}) = \text{ID}(g'_{1k}, g'_{2k}), \quad \forall k = 1 \text{ to } M. \quad (3)$$

Ideally, (3) should construct  $\tilde{R}$  by satisfying (1), but practically there is an error as in

$$D(g_{1k}, g_{2k}) = c + \varepsilon_k, \quad \forall k = 1 \text{ to } M. \quad (4)$$

In (4),  $\varepsilon_k$  is an error of different type which can be described as below:

$$\varepsilon_k = \varepsilon_{p_k} + \varepsilon_{D_k} + \varepsilon_{S_k}, \quad \forall k = 1 \text{ to } M. \quad (5)$$

In (5), error  $\varepsilon_{p_k}$  occurs due to weak perspective of complex real world. Inner distance gave error  $\varepsilon_{D_k}$  when pathway connecting two points  $g_{1k}, g_{2k}$  passes over the junction.  $\varepsilon_{S_k}$  denotes the error of projection of 3D image on 2D image. Due to varying effects and various view points, affine normalization will be performed:

$$T(p_{ik}) \rightarrow p'_{ik}, \quad (6)$$

where  $T$  shows a minimum enfolding parallelogram which transforms the part  $p_{ik}$  of shape  $S_k$  toward unit square. Thus, it is used for the affine transformation that consists of scale, translation, rotation, and shear transformation. After achieving transformation  $T(g_{ik}) \rightarrow g'_{ik}$  we will compute two points  $g_{1k}, g_{2k}$  using inner distance (ID) and inner angle (IA). At last, shape context descriptor will be built on each point  $g'_{ik}$  and it is called histogram  $h_{ik}$ :

$$h_{ik}(z) = \# \left\{ g'_{mk}, m \neq 1 \leq m \leq t : \text{ID}(g'_{ik}, g'_{mk}) \times \text{IA}(g'_{ik}, g'_{mk}) \in \text{bin}(z) \right\}. \quad (7)$$

Here,  $z$  denotes the number of bins. Total 60 bins were used. Distance bins were 12 and angular bins were 5. The result of these two types of bins is 60. Now, feature representation satisfying (1) can be constructed as  $\tilde{R}(S_k) = (h_{1k}, h_{2k}, \dots, h_{tk})$ .

**3.2. Similarity Measure.** For shape retrieval, usually, similarity of shape or dissimilarity of shape also known as pairwise shape distance is calculated for which best possible correspondence of contour points needs to be found. These optimal points are used for ranking of shapes in the database for shape retrieval. Similarity measure or shape distance was calculated by DP and we obtained  $N \times N$  matrix for further postprocessing method. This DP method is invariant to translation, rotation, scale, and also robust against outliers and noise. This and other similarity measures are nonmetric and breach the triangle in equality. DP can be calculated as follows: consider there are two shapes  $S_1$  and  $S_2$  and they have contour point sequences, such as  $p_1, p_2, \dots, p_n$  for  $S_1$  with  $n$  points and  $x_1, x_2, \dots, x_m$  for  $S_2$  with  $m$  points. The matching  $\pi$  between shape  $S_1$  and shape  $S_2$  is mapping from  $1, 2, \dots, n$  to  $0, 1, 2, \dots, m$ , where  $p_i$  can be matched to  $x\pi(i)$  if  $\pi(i) \neq 0$  and will be left unmatched otherwise.  $\pi$  is used for minimization of the cost of matching  $H(\pi)$  as follows:

$$C(\pi) = \sum_{1 \leq i \leq n} c(i, \pi(i)). \quad (8)$$

Here,  $c$  is the matching cost and can be calculated as follows:

$$c(i, j) \equiv \frac{1}{2} \sum_{1 \leq z \leq Z} \frac{[h_{S_1,i}(z) - h_{S_2,j}(z)]^2}{h_{S_1,i}(z) + h_{S_2,j}(z)}. \quad (9)$$

Here,  $h_{S_1,i}$  and  $h_{S_2,j}$  are the shape context histograms of  $p_i$  and  $x_j$  and  $Z$  represents the numbers of histograms.

According to true analysis of database, dissimilarities should be high between extraclass and low for intraclass. Similarity measure did not follow this rule fully. That is why modified mutual  $k$ NN graph improves the accuracy after thoroughly analysis.

**3.3. Modified Mutual  $k$ NN-Graph.** For improving the retrieval rate and to obtain the affinity matrix  $W$ , we used modified mutual  $k$ NN graph manifold learning method as a postprocessing method. It is more consistent on shapes as well as on images and it is computationally very cheap. Such postprocessing manifold learning methods are used for improving the  $N \times N$  similarity matrix  $A$  whose similarities are calculated among all images of database. There are two steps which improve these precalculated matrices. The first step is the normalization of the  $N \times N$  distance matrix  $A$  which transforms the  $A$  into affinity matrix  $W$  and the second step is more thorough analysis of pattern of objects such as finding the shortest paths by construction of nearest neighbor graphs.

If  $N \times N$  distance matrix  $A$  is given, we can do part by part normalization to achieve  $N \times N$  affinity matrix  $W$  as follows:

$$W_{ij} = \exp\left(-\frac{A_{i,j}^2}{\sigma_{i,j}^2}\right). \quad (10)$$

Here,  $\sigma$  is an element-wise normalization parameter, while  $A$  is a  $N \times N$  distance matrix and  $W$  is an affinity matrix. The normalization parameter can be defined as below:

$$\sigma_{ij} = \sigma_i \sigma_j, \quad (11)$$

$$\sigma_i = A(i, i_K). \quad (12)$$

In (12),  $i_K$  shows the  $K$ th neighbor of object  $i$ . Distances between features of images are not obvious metric and cannot be measured only the scale of 0 to 1. Importance of normalization for improving retrieval has been given in literature [37]. After normalization, distance matrix can be converted into affinity matrix and underlying structure of data can be defined by measuring a neighborhood in graph. Graph will try to match similarity using underlying structure.

Ideally, the normalization should give clear and embedded submanifolds of matching shapes but normalization methods are not optimal; thus we used analysis step after the normalization. Graph based methods are then used for performing such analysis which find the locally underlying structures. Interdependence between objects can be found using connected regions in these graphs such as detecting shorter routes within similar shapes rather than within nonmatching shapes, to define the neighborhood graph with modified mutual  $k$ NN-graph. Graph can be constructed in



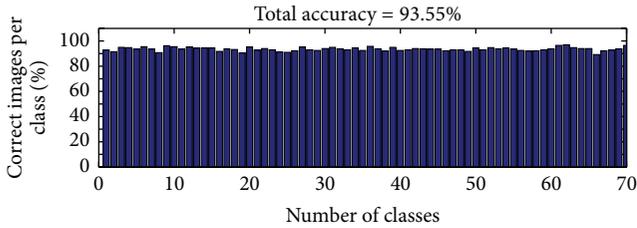


FIGURE 3: Classwise accuracy of AIR.

TABLE 1: Retrieval rate on MPEG-7 database (bulls-eye top 40).

Method	Bulls-eye score
CSS [39]	75.44%
SC + TPS [15]	76.51%
Generative models [10]	80.03%
IDSC + DP [23]	85.40%
IDSC + modified mutual $k$ NN [13]	93.40%
AIR [12]	93.55%
AIR + graph transduction [11]	99.90%
Our method (without normalization)	99.47%
Our method	<b>99.89%</b>

Results are also reported before and after normalization of modified mutual  $k$ NN graph. We achieved bulls-eye score 99.47% without normalization and 99.89% after normalization. Results clearly indicate the significance of normalization. IDSC is more robust method and it is articulation invariant but its accuracy is 85.40% which is much below than our method. AIR achieved 99.55. Our method improved results to 99.89%. AIR with graph transduction performed little well but its computational time is too high. Figure 4 shows top 10 retrieved images of MPEG-7 database. Most left column images are query images while remaining 19 column images are retrieved images from database. Retrieved images are rotation translation, scale invariant, and deformation invariant. We also showed the accuracy on each class in Figure 5. Figure 5 showed the improvement in each category and accuracy of each category is more consistent compared to Figure 3.

We also measured computational time with another post-processing graph transduction method. Computation was performed on Core i5 2.40 GHz processor with 4 GB RAM installed using MATLAB. Comparison of computational time is given in Figure 6.

Graph transduction is another manifold learning post-processing method. We applied graph transduction method on  $N \times N$  matrix of pairwise distance. Although results are promising, major drawback of this method is computational time. Compared to graph transduction method, modified mutual  $k$ NN graph showed huge boost with respect to computational time. On AIR calculated shape distances, time of modified mutual  $k$ NN graph was calculated as 10.83 seconds (without normalization) and 12.07 seconds (after normalization). Conversely, 7468.89 seconds were calculated by graph transduction on AIR shape distances.

TABLE 2: Retrieval rate on ORL database (bulls-eye top 15).

Method	Bulls-eye score
Baseline [38]	62.35%
Our method	<b>78%</b>

TABLE 3: Retrieval rate on YALE face B database (bulls-eye Top 15).

Method	Bulls-eye score
Baseline [38]	69.48%
Our method	<b>75.04%</b>

4.2. *Retrieval on ORL Face Database.* ORL has 400 grayscale faces of different 40 peoples and each person has 10 different images with different illumination, pose, and expression. After getting distance matrix of  $400 \times 400$  from feature representation [38], we applied the postprocessing modified mutual  $k$ NN graph. This integration improved the baseline results which were measured by bulls-eye score from top 15 closest neighbors. Results are shown in Table 2.

4.3. *Retrieval on Yale Face Database B.* Yale face database B is the database of faces with different poses and illumination. In this database 165 faces are collected of 15 subjects and each subject has 11 different images under different conditions. Feature representation [38] was used for obtaining subset. After calculating distance we applied modified mutual  $k$ NN graph. Performance was computed using bulls-eye score from top 15 closest matches. Results are shown in Table 3.

## 5. Conclusion

Our work tackled the problem of shape retrieval. Articulation invariant representation was used to extract features of binary images. After obtaining similarity matrix by dynamic programming, we integrate the postprocessing modified mutual  $k$ NN graph for obtaining the context information and get similarity among different images. We conducted shape retrieval experiments on widely used MPEG-7 database for shape images and measured accuracy by so-called bulls-eye score. We also reported results before and after normalization of modified mutual  $k$ NN graph which clearly indicates that normalization improved accuracy results, significantly. Compared to other methods our method got better results. We also demonstrated that computational performance of our method is much higher than other competitive methods. To show consistency of postprocessing method, we also performed experiments on ORL and YALE face databases and baselines results were improved.

In the future we would like to extend this shape retrieval on cluttered images using deformation invariant shape matching and diffusion process.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

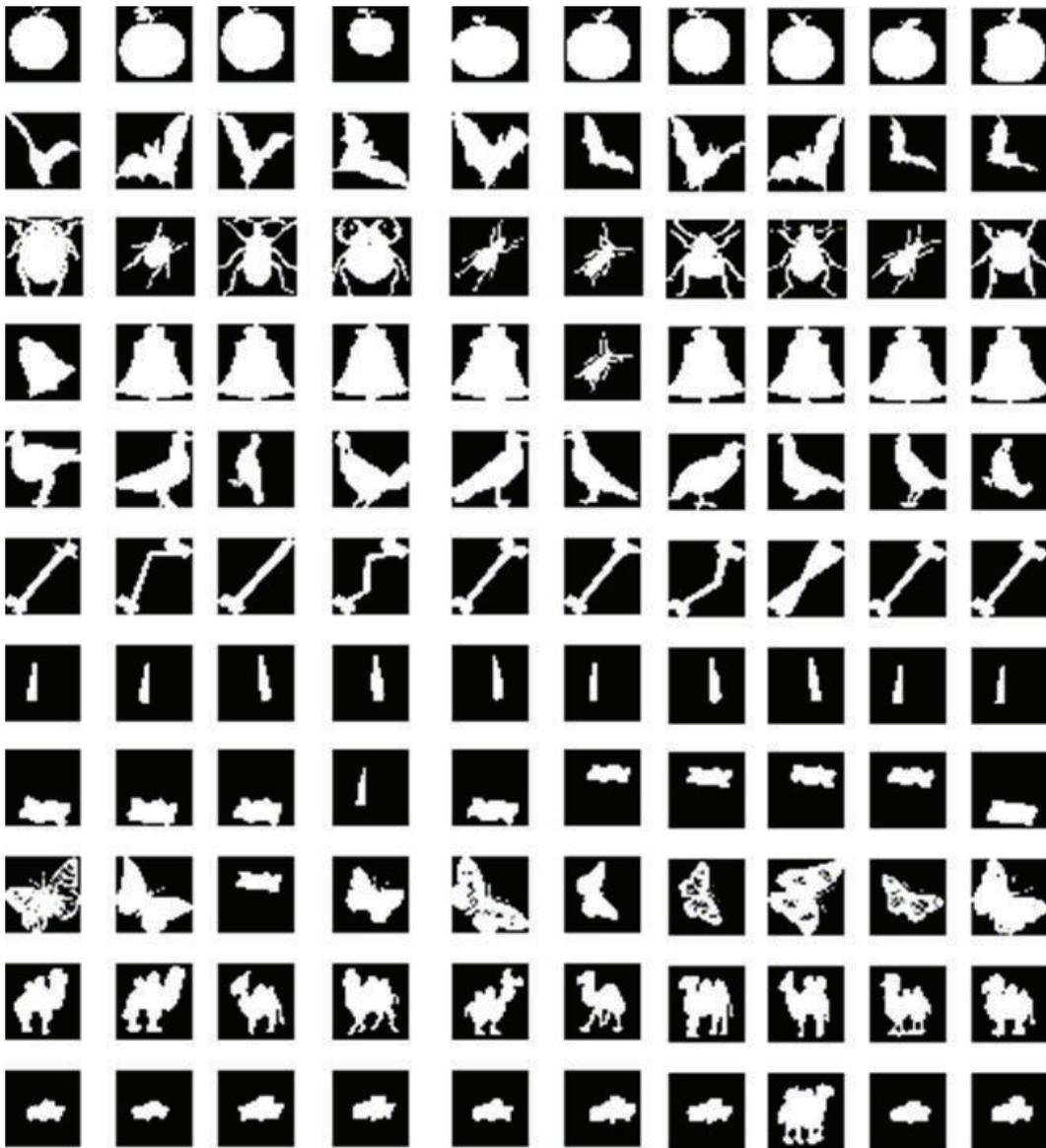


FIGURE 4: Left most column images show the query images while other column images are retrieved images.

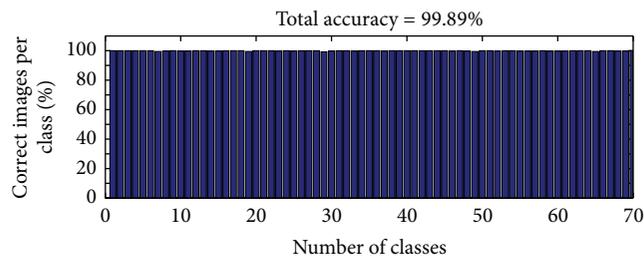


FIGURE 5: Classwise accuracy of our method.

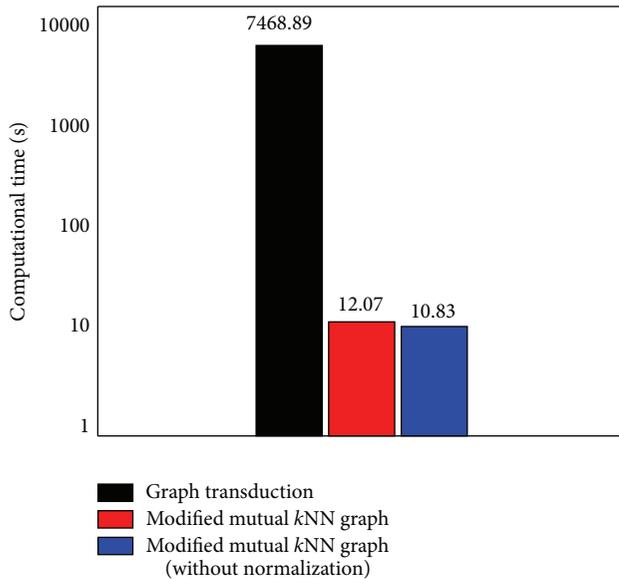


FIGURE 6: Computational time in seconds.

## Acknowledgments

This research was supported by National Natural Science Foundation of China (Grant nos. 60973059 and 81171407) and Program for New Century Excellent Talents in University of China (Grant no. NCET-10-0044).

## References

- [1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, article 5, 2008.
- [3] R. Veltkamp, H. Burkhardt, and H.-P. Kriegel, Eds., *State-of-The-Art in Content-Based Image and Video Retrieval*, vol. 22, Springer, 2001.
- [4] F. Long, H. Zhang, and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management*, Signals and Communication Technology, pp. 1–26, Springer, Berlin, Germany, 2003.
- [5] S. Antani, L. R. Long, and G. R. Thoma, "Content-based image retrieval for large biomedical image archives," in *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO '04)*, San Francisco, Calif, USA, September 2004.
- [6] I. Yahiaoui, N. Hervé, and N. Boujemaa, "Shape-based image retrieval in botanical collections," in *Advances in Multimedia Information Processing—PCM 2006*, vol. 4261 of *Lecture Notes in Computer Science*, pp. 357–364, Springer, Berlin, Germany, 2006.
- [7] P. Dukkipati and L. Brown, "Improving the recognition of geometrical shapes in road signs by augmenting the database," in *Proceedings of the 3rd International Conference on Computer Science and Its Applications*, pp. 8–13, 2005.
- [8] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Advances in Neural Information Processing Systems*, vol. 16, pp. 169–176, Max Planck Institute for Biological Cybernetics, Cambridge, Mass, USA, 2004.
- [9] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu, "Learning context-sensitive shape similarity by graph transduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 861–874, 2010.
- [10] A. Egozi, Y. Keller, and H. Guterman, "Improving shape retrieval by spectral matching and meta similarity," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1319–1327, 2010.
- [11] X. Yang, X. Bai, L. J. Latecki, and Z. Tu, "Improving shape retrieval by learning graph transduction," in *Computer Vision—ECCV 2008*, pp. 788–801, Springer, Berlin, Germany, 2008.
- [12] R. Gopalan, P. Turaga, and R. Chellappa, "Articulation-invariant representation of non-planar shapes," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, vol. 6313 of *Lecture Notes in Computer Science*, pp. 286–299, Springer, Berlin, Germany, 2010.
- [13] P. Kotschieder, M. Donoser, and H. Bischof, "Beyond pairwise shape similarity analysis," in *Computer Vision—ACCV 2009*, pp. 655–666, Springer, Berlin, Germany, 2010.
- [14] X. Yang, X. Bai, L. J. Latecki, and Z. Tu, "Improving shape retrieval by learning graph transduction," in *Computer Vision—ECCV 2008*, vol. 5305 of *Lecture Notes in Computer Science*, pp. 788–801, Springer, Berlin Heidelberg, 2008.
- [15] S. Belongie, J. Malik, and J. 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.
- [16] R. C. Veltkamp and M. Hagedoorn, "State of the art in shape matching," in *Principles of Visual Information Retrieval*, Advances in Pattern Recognition, pp. 87–119, Springer, London, UK, 2001.
- [17] E. Goldmeier, "Similarity in visually perceived forms," *Psychological Issues*, vol. 8, no. 1, pp. 1–136, 1972.
- [18] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.
- [19] F. Mokhtarian, "Silhouette-based object recognition with occlusion through curvature scale space," in *Computer Vision—ECCV '96*, vol. 1064 of *Lecture Notes in Computer Science*, pp. 566–578, Springer, Berlin, Germany, 1996.
- [20] J. van der Poel, C. W. D. de Almeida, and L. V. Batista, "A new multiscale, curvature-based shape representation technique for image retrieval based on DSP techniques," in *Proceedings of the 5th International Conference on Hybrid Intelligent Systems (HIS '05)*, pp. 373–378, November 2005.
- [21] C. W. D. de Almeida, R. M. C. R. de Souza, and N. L. Cavalcanti Jr., "A shape-based image retrieval system using the curvature scale space (CSS) technique and the self-organizing map (SOM) model," in *Proceedings of the 6th International Conference on Hybrid Intelligent Systems (HIS '06)*, IEEE, December 2006.
- [22] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: breaking a visual CAPTCHA," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 134–144, June 2003.
- [23] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.
- [24] P. F. Felzenszwalb and J. D. Schwartz, "Hierarchical matching of deformable shapes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, June 2007.

- [25] G. McNeill and S. Vijayakumar, "Hierarchical procrustes matching for shape retrieval," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 1, pp. 885–892, New York, NY, USA, June 2006.
- [26] J. Wang, X. Bai, X. You, W. Liu, and L. J. Latecki, "Shape matching and classification using height functions," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 134–143, 2012.
- [27] C. Xu, J. Liu, and X. Tang, "2D shape matching by contour flexibility," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 180–186, 2009.
- [28] X. Wang, X. Bai, W. Liu, and L. J. Latecki, "Feature context for image classification and object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 961–968, 2011.
- [29] L. J. Latecki and R. Lakämper, "Shape similarity measure based on correspondence of visual parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1186–1190, 2000.
- [30] Z. Tu and A. L. Yuille, "Shape matching and recognition—using generative models and informative features," in *Computer Vision—ECCV 2004*, vol. 3023, pp. 195–209, Springer, Berlin, Germany, 2004.
- [31] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, "Shock graphs and shape matching," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 13–32, 1999.
- [32] N. Alajlan, M. S. Kamel, and G. H. Freeman, "Geometry-based image retrieval in binary image databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1003–1013, 2008.
- [33] M. R. Daliri and V. Torre, "Robust symbolic representation for shape recognition and retrieval," *Pattern Recognition*, vol. 41, no. 5, pp. 1799–1815, 2008.
- [34] X. Bai, C. Rao, and X. Wang, "Shape vocabulary: a robust and efficient shape representation for shape matching," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3935–3949, 2014.
- [35] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, and H. Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern Recognition*, vol. 36, no. 8, pp. 1725–1736, 2003.
- [36] A. Tiwari and V. Bansal, "PATSEEK: content based image retrieval system for patent database," in *Proceedings of the 4th International Conference on Electronic Business (ICEB '04)*, Beijing, China, December 2004.
- [37] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Proceedings of the 18th Annual Conference on Neural Information Processing Systems (NIPS '04)*, December 2004.
- [38] J. Jiang, B. Wang, and Z. Tu, "Unsupervised metric learning by self-smoothing operator," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 794–801, November 2011.
- [39] F. Mokhtarian, S. Abbasi, and J. Kittler, "Efficient and robust retrieval by shape content through curvature scale space," in *Proceedings of International Workshop on Image Databases and Multimedia Search*, pp. 35–42, 1996.

