A NEW MULTI-SCALE FUZZY MODEL FOR HISTOGRAM-BASED DESCRIPTORS

Lunshao Chai[a], Zhen Qin[b], Honggang Zhang[a], Jun Guo[a], Bir Bhanu[b]

[a]Beijing University of Posts and Telecomuunictions, Beijing, 100876, China
[b]University of California at Riverside, Riverside, CA 92521, USA

ABSTRACT

In this paper, we first propose a general *Multi-Scale Fuzzy Model (MSFM)* which handles distortions at different scales in *Histogram-Based Descriptors(HBDs)*. This model can be applied both on one-dimensional HBDs and multi-dimensional HBDs. We then focus on applying MSFM on the widely used *Shape Context* for a *Simplified Multi-scale Fuzzy Shape Context (SMFSC)* descriptor. Fuzzy models are barely used in multi-dimensional HBDs due to the significant increase of computational complexity. We show that by introducing an intra-bin point location approximation and an approximate iterative fuzzification approach, the algorithm can be simplified and thus SMFSC hardly increases computational complexity. Experiments on standard shape dataset show that SMFSC improves upon the Inner Distance Shape Context. We also applied SMFSC on *Content-Based Product Image Retrieval* and the experimental results further validate the effectiveness of our model.

Index Terms— histogram-based descriptors, multi-scale fuzzy model, simplified multi-scale fuzzy shape context

1. INTRODUCTION

Histogram-based descriptors (HBDs) have played an important role in various multimedia and vision tasks such as image retrieval [1, 2] and shape analysis [3, 4]. HBDs have several advantages. Firstly, the histograms are well-defined and have clear physical interpretations. Secondly, the similarity between two HBDs is easy to measure by using existing distances such as Euclidean distance, χ^2 distance, Histogram Intersection distance and *Earth Mover's Distance (EMD)* [5]. Thirdly, most HBDs have achieved basic necessary built-in invariances such as translation, scale and rotation invariance. Nevertheless, there is one crucial drawback of HBDs. HBDs can only precisely measure the similarity of two images based on the assumption that the two images are perfectly aligned, which is hardly the case in most real-world applications. That's because to capture rich and precise information, the image blocks corresponding to histogram bins in HBDs are usually small and absolute, and small distortion or noise will lead to misalignment.

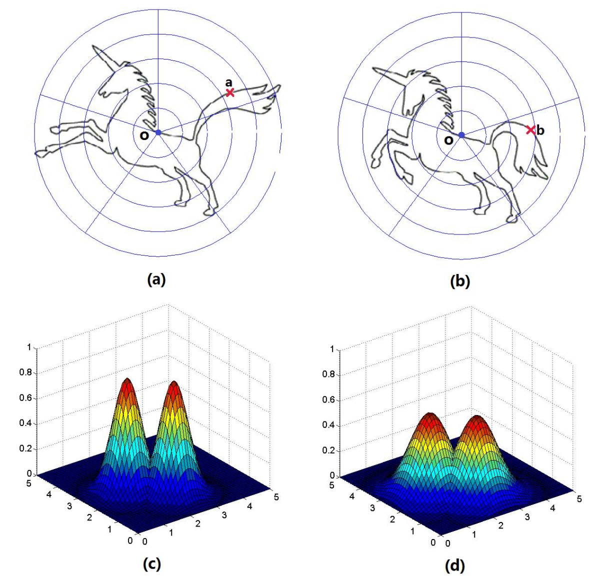

Fig. 1. An illustration of effect of different fuzzification scales. Point a and point b fall into different grids of shape context of point o due to deformation. (c) and (d) show how the two points are assigned to different histogram bins of different fuzzification scales.

Targeting to address this problem, many solutions followed the idea of fuzzy model have been proposed. The basic idea is to use one-to-many correspondences instead of using the one-to-one patterns in crisp models when matching the bins in histograms. These solutions fall into two categories: feature domain methods [6, 7, 8, 9, 10, 11, 12] and measure domain methods [5, 13, 14, 15, 16] (see section 2 for a detailed discussion). Generally, well-designed measure domain methods can be used to measure the similarity of different descriptors. But feature domain methods are usually more effective and flexible because they can be optimized based on in-depth study of the specific descriptor. More importantly, for many tasks including image retrieval and image recognition, increased computations of applying feature domain fuzzy methods is in the off-line stage, while that of measure domain methods is in the on-line stage. Hence in these circumstances, feature domain methods are favorable options.

An important problem with existing feature domain methods is that they are all designed to handle histogram distortions of one certain scale, while the scales of histogram distortions vary a lot from image to image. The preset fuzzification scale may be of little or no help for other deformation scales. Fig. 1 illustrates the effect of different fuzzification scales. We can see that fuzzification scale (d) can better mitigate the deformation between shape (a) and shape (b). Besides, the histogram distortion problem in multi-dimensional HBDs have been realized and discussed but few rigorous multi-dimensional fuzzy models are proposed.

To overcome these problems, we propose a general feature domain multi-scale fuzzy model termed *Multi Scale Fuzzy Model (MSFM)* which handles multi-scale distortions. It can be applied on any HBDs, both one-dimensional ones and multi-dimensional ones. To verify the effectiveness of our model, we apply it on the widely used two dimensional *Inner Distance Shape Context* descriptor. We introduce two approximations in the generation of the descriptor to reduce the computational complexity, and name this relaxed version *Simplified Multi scale Fuzzy Shape Context (SMFSC)*. The simplified fuzzification process hardly increases computational complexity but still results in better performance.

2. RELATED WORK

2.1. Feature domain fuzzy models

Fuzzy logic [17] was introduced to the image processing society to mitigate the distortions of ordinary crisp HBDs. The first and most successful feature domain methods are applied in color descriptors. In [6], Konstantinidis et al. proposed a new fuzzy linking method of color histogram creation based on the L*a*b* color space and provided a histogram which contains only 10 bins. In [7], a fuzzy system was proposed to produce a fuzzy linking histogram, which takes the three channels of HSV as inputs, and forms a 10-bin histogram as the output based on the color selection method in [6]. Wang et al. proposed a robust kernel tracking method using fuzzy color histogram in [8]. Chuang et al. [9] presented a ratio histogram based on fuzzy color histogram to detect suspicious objects in an abnormal event.

There are many widely used multi-dimensional HBDs such as *SIFT* [18], *HOG* [19] and *Shape Context* [3]. And multi-dimensional histogram distortion problem has been discussed and some solutions have been proposed, but they are not rigorous or well-defined fuzzy models.Liu et al. proposed a soft shape context descriptor in which one point is assigned to angular neighboring bins [10]. Wang et al. tried to address histogram distortion by angular blur [11]. They enlarged angular span, letting bins be overlapped in angular directions. These two solutions only deal with one dimension in shape context, namely the angular dimension. In this sense, Ayed et al. went much further, they defined fuzzy rulers on both angular and radial dimension and got a fuzzy descriptor [12]. However, in their algorithm points are treated equal, no matter they are in the center of a bin or near the boundary. This is not reasonable because points near the boundary are more possible to deform to another bin.

2.2. Measure domain fuzzy models

In general, measure domain methods use cross-bin distances to handle histogram distortions in HBDs and many distance measures have been proposed. The *Earth Mover's Distance (EMD)* proposed by Rubneret et al. [5] defines the distance computation between distributions as a transportation problem. EMD is very effective for distributions with sparse structures. However, the time complexity of EMD is larger than $O(N^3)$ where N is the number of histogram bins. This prohibits its application from multi-dimensional HBDs. Indyk and Thaper [13] proposed a fast (approximative) EMD algorithm by embedding the EMD metric into an Euclidean space. The time complexity of the embedding is $O(N d log \triangle)$, where N is the size of feature sets, d is the dimension of the feature space and $\triangle$ is the diameter of the union of the two feature sets to be compared. Grauman and Darrell [14] used the pyramid matching kernel for feature set matching. A pyramid of histograms of a feature set is extracted as a description of an object. Then the similarity between two objects is defined by a weighted sum of histogram intersections at each scale.

Ling et al. also proposed a fast EMD algorithm, namely EMD-L_1 [15]. EMD-L_1 utilizes the special structure of the L_1 ground distance on histograms for a fast implementation of EMD. Later Ling et al. proposed another cross-bin distance, diffusion distance, for histogram comparison [16]. They defined the difference between two histograms to be a temperature field, and treated the histogram similarity as a diffusion process. The diffusion distance is much faster than EMD-L_1 and has comparable performance with EMD-L_1.

2.3. Shape context

The original *shape context* was introduced by Belongie[3] for shape description. It describes the relative spatial distribution (distance and orientation) of landmark points around feature points. Given N sample points $p_1, p_2, \ldots, p_N$ on a shape, the shape context at point x_i is defined as a histogram h_i of the relative coordinates of the remaining $N-1$ points. The distance between two shape context histograms is defined using the χ^2 statistic.

Based on shape context, Ling et al. proposed a new *Inner Distance Shape Context (IDSC)* descriptor in [20]. They extended the shape context by replacing the Euclidean distance with the inner-distance. They also designed a dynamic programming based method for shape matching and comparison. Their algorithm demonstrated much better performance in comparison with the original shape context.

3. MULTI-SCALE FUZZY MODEL

We now describe our proposed general multi-scale fuzzy model (MSFM) in detail.

3.1. Fuzzy Histogram-based Descriptors

Let $\boldsymbol{x}$ represents a datapoint p. The possible value range $\boldsymbol{\Omega}$ of $\boldsymbol{x}$ is divided to L subranges $\{\boldsymbol{X_i}\}$, each of which corresponds to one bin in the feature histogram $\{b_i\}$. In traditional crisp histograms, the assignment to bin b_i from p is determined as:

$$\Delta b_i = \begin{cases} 1, & if\ \ \boldsymbol{x} \in \boldsymbol{X_i} \\ 0, & otherwise \end{cases} \quad , \quad i = 1, 2, \ldots, L. \tag{1}$$

To handle the possible distortion of image features, in fuzzy logic the datapoint p is represented by a distribution, denoted by $d(\boldsymbol{x})$, which satisfies:

$$d(\boldsymbol{x}) \geq 0 \ \ , \ \ and \tag{2}$$

$$\int\limits_{\Omega} d(\boldsymbol{x}) = 1. \tag{3}$$

Then the assignment to each bin from p is determined as:

$$\Delta b_i = \int\limits_{X_i} d(\boldsymbol{x}). \tag{4}$$

Since $d(\boldsymbol{x})$ determines the component the histogram draws from each data point, it is called the membership function. Many kinds of membership functions have been used. Fig. 2 shows some of them.

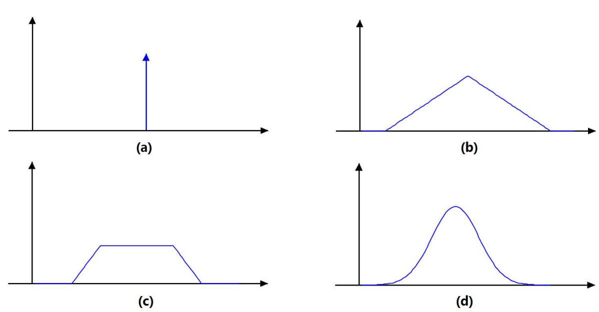

Fig. 2. Examples of membership functions. (a) is the impulse function, which can be used to analyze crisp histograms in fuzzy models. (b) is a triangular function. (c) is a trapezoidal function. (d) is a normal distribution function.

3.2. Multi-Scale Fuzzy Model

Generally fuzzy HBDs have better performance than crisp HBDs because fuzzy models have a better representation of image features with distortions and noises. However, how to build an appropriate fuzzy model is still an open question: what scale of fuzzification should the feature have? What makes the problem even more difficult is that the distortion may be different in each pairwise comparison, so it is impossible to learn a global distortion scale. Another way is to check the distortion in each fuzzification scale, which is the basic idea of our MSFM. We check K levels of distortion scales and the final fuzzy HBDs $\boldsymbol{H}$ is the concatenation of each fuzzy histograms of each fuzzification scale:

$$\boldsymbol{H} = [\boldsymbol{H_1}, \boldsymbol{H_2}, \ldots, \boldsymbol{H_K}]. \tag{5}$$

The distance between two MSFM fuzzy HBDs $\boldsymbol{H}$ and $\boldsymbol{G}$ can be measured as:

$$dis_{MSFM}(\boldsymbol{H}, \boldsymbol{G}) = \sum_{i=1}^{K} dis_{SC}(\boldsymbol{H_i}, \boldsymbol{G_i}), \tag{6}$$

where $dis_{SC}(\boldsymbol{X}, \boldsymbol{Y})$ can be any traditional distance measures of the corresponding crisp HBD. Note that there is a trade-off between the number of scale levels and the computational complexity of the descriptor. The appropriate number of scale levels is determined by the distortions of the histogram, which is further determined by the distortions and noises of the images.

4. SIMPLIFIED MULTI-SCALE FUZZY SHAPE CONTEXT

We apply our proposed MSFM model on the Shape Context-based shape descriptor IDSC introduced in [20] (see Section 2(3)). In each fuzzy scale, every datapoint sampled from the contour is represented by a two dimensional normal distribution with covariance matrix

$$\boldsymbol{Cov} = \alpha \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{7}$$

where α is a parameter. However, this will lead to huge computational cost. Considering the shape context of one sampled point, in traditional IDSC, we just need to check which bins the other $N-1$ points belong to, where n is the number of sampled points. But in fuzzy histogram with MSFM applied, we must calculate $K \times M \times (N-1)$ two dimensional integrals, where K is the number of fuzzification scales and M is the number of bins each point belongs to.

In order to reduce the high computational complexity, We use two approximation techniques. The first one is an intra-bin point location approximation. The idea is to pretend that all points falling in each image block are at the center of that block. In this way all these points have the same assignment to their correspondence bins. Then the assignment can be pre-calculated and saved, and the time-consuming integral of $d(\boldsymbol{x})$ is no longer needed.

The second approximation is using an approximate iterative fuzzifying approach. According to the characterization of normal distribution, about 95% of the integral value is drawn within two standard deviations from the mean, so we can assign the shape context value only within this range. Also, for K scales of fuzzification, we need K assignments and calculate them respectively. To further reduce the time complexity, we approximate the fuzzication in an iterative manner. In the kth level we fuzzify the output histogram from the $(k-1)th$ level.

With these two approximations, the computational cost is greatly reduced. The fuzzy inner distance shape context is simplified, and we name this relaxed fuzzy descriptor *Simplified Multi-scale Fuzzy Shape Context (SMFSC)*. A description of our fuzzification process is illustrated in **Algorithm 1**.

Algorithm 1: SMFSC extraction of an image

Input: sampled points $P = \{p_1, \ldots, p_N\}$; $f(\cdot)$ is datapoint-to-bins assignment derived from the two dimensional normal distribution.

for $n = 1$ to N **do**
  Step 1: Calculate a IDSC histogram $\boldsymbol{h_n^0}= \{b_j^0\}$ for p_n.
    for $i = 1$ to N without n **do**
      if p_i in image block corresponding to b_j^0 **then**
        $b_j^0 \leftarrow b_j^0 + 1$;
      end if
    end for
  Step 2: Iteratively calculate a fuzzy histogram $\boldsymbol{h_n^k}= \{b_j^k\}$ for p_n at the kth fuzzification scale, and append it to the fuzzy histogram of the whole image $\boldsymbol{H_k}$ at the kth scale.
    for $k = 1$ to K **do**
      for $j = 1$ to J **do**
        $b_j^k = f(\boldsymbol{h_n^{k-1}})$;
      end for
      $\boldsymbol{H_k(n)}$=$\boldsymbol{h_n^k}$;
    end for
end for

Output: $\boldsymbol{H}$= [$\boldsymbol{H_1}$, $\boldsymbol{H_2}$, ..., $\boldsymbol{H_K}$]

---

5. EXPERIMENTS

In this section the proposed SMFSC is tested for two tasks. The first experiment is performed on the widely tested shape data sets, MPEG7 CE-Shape-1 shapes. The result shows that SMFSC outperforms traditional methods. We then evaluate SMFSC on a real-world application, *Content-based Product Image Retrieval (CBPIR)*, which further demonstrates the effectiveness of SMFSC.[1]

[1] Matlab code implementation of SMFSC can be downloaded for research usage at https://dl.dropboxusercontent.com/u/57435211/icme2013w.zip.

5.1. MPEG shape database

The widely tested MPEG7 CE-Shape-1 database consists of 1400 silhouette images from 70 classes. Each class has 20 different shapes. The recognition rate is measured by the so-called Bullseye test: For every image in the database, it is matched with all the images and the top 40 most similar candidates are counted. At most 20 of the 40 candidates are correct hits so the possible best result is 20 for one image. The Bullseye score of the test is the ratio of the number of correct hits for all images to the number of possible best result for all the images.

We use the same setting as in [20]. The number of sampled points is 100 (300 were used in [3]); the number of radial bins is 8; the number of angular bins is 12, and the number of fuzzification scales is 2. In Table 1 SMFSC is compared with other algorithms. It shows that our algorithms outperform all the alternatives. The speed of our is almost the same with ID-SC, and in the same range as as those of shape contexts [3], curve edit distance [21] and generative model [22].

To further analyze the effectiveness of our algorithm, we did two other experiments in the same setting with different scale levels of fuzzification. In the first experiment, we only use one scale level, which means the output of first fuzzification level is directly used as descriptor. The Bullseye score of this experiment is 85.63%. The result demonstrates that both the fuzzification and the extension to multi-scale contribute to the improvement of retrieval performance. In the second expriment we use three scale levels, and the bullsey score is 85.74%, only a slight improvement for the additional scale. This shows that for the MPEG7 CE-Shape-1 database, a two scales fuzzification is enough to handle the distortions.

5.2. Content-based Product Image Retrieval

Content-based Product Image Retrieval(CBPIR) is an emerging application-oriented field of Content-based Image Retrieval with the prevalence of E-Commerce sites such as Amazon and eBay. Image segmentation is the bottle neck of using shape feature in many tasks such as image classification and general image retrieval. But in CBPIR, the background is mostly very simple and the shapes of the products can be easily extracted. So Shape feature can be an important cue in the retrieval task.

In this part, we build a retrieval system utilizing SMFSC. The retrieval results are compared with the retrieval system in [27] with *Radial Harmonic Fourier Moments (RHFM)* [28] and *Elliptical Fourier Shape Descriptors (EFSD)* [29]. We conduct the experiment on two databases, the *Product Image Categorization Data Set (PI100)* [30] and the CPImage10 [31]. We choose 10 categories from PI100 (see examples from the 10 categories in Fig. 3). There are 100 images in each category's samples gallery and 20 query images for each category, which are not included in the sample gallery. CPImage10 is a subset of the CPImage dataset (see examples

Table 1. Bullseye scores of different methods for MPEG7 CE-SHAPE-1

Alg.	CSS [23]	Vis. Parts [24]	SC+TPS [3]	Curve Edit [21]	Dis. Set [25]
Score	75.44%	76.45%	76.51%	78.17%	78.38%
Alg.	MCSS [26]	Gen. Mod. [22]	MDS+SC+DP [20]	IDSC+DP [20]	**SMFSC**(2 scales)
Score	78.8%	80.03%	84.35%	85.40%	**85.73%**

in Fig. 4). There are also 10 categories with 100 images in each category, and we randomly choose 20 of them as query images and perform retrieval in the whole dataset.

Fig. 3. Example images of the selected PI100 images

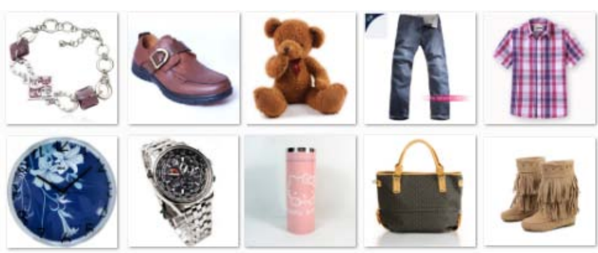

Fig. 4. Example images of the CPImage10

Precision ratio and recall ratio criterion are used to evaluate the retrieval result. Fig. 5 shows the precision ratio and recall ratio at the first R retrieval result. From these results, we can see that SMFSC performs consistently better than RHFM and shape histogram at all R values. In addition, the feature extraction time of SMFSC is in the same range with moment.

6. CONCLUSION AND FUTURE WORK

In this paper we first proposed a new fuzzy model, MSFM. The model is novel in two respects. Firstly, it is a general model which can be used on both single dimensional HBDs and multi-dimensional HBDs. Secondly, it formulates different scales of histogram distortions, which makes it more robust and effective. Then we apply this model on the widely used two dimensional shape descriptor IDSC, and propose SMFSC. Experiment on the MEPG shape dataset shows that SMFSC outperforms traditional algorithms. Experiments on a real-world application , CBPIR, also demonstrate its effectiveness. We make our code and dataset with experiment setting publicly available for easy future references.

We are interested in generalizing our model to other widely used multi-dimensional HBDs, such as SIFT and ORB. We also plan to go further with SMFSC, including finding ways to conduct fuzzification with no or less approximation, and exploring the possibility of using Bag-of-Feature-like model to speed up the matching.

7. ACKNOWLEDGEMENT

This work was partially supported by National Natural Science Foundation of China under Grant No.61273217, 61175011 and 61171193, the 111 project under Grant No.B08004, and the Fundamental Research Funds for the Central Universities.

8. REFERENCES

[1] R. Arandjelovic and A. Zisserman, “Smooth object retrieval using a bag of boundaries,” in *ICCV*, 2011.

[2] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan, “Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set,” in *CVPR*, 2012.

[3] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape context,” in *TPAMI*, 2002.

[4] I. Kokkinos, M. Bronstein, R. Litman, and A. Bronstein, “Intrinsic shape context descriptors for deformable shapes,” in *CVPR*, 2012.

[5] Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval,” in *IJCV*, 2000.

[6] K. Konstantinidis, A. Gasteratos, and I. Andreadis, “Image retrieval based on fuzzy color histogram processing,” in *Optics Communications*, 2005.

[7] S. Chatzichristofis and Y. Boutalis, “A hybrid scheme for fast and accurate image retrieval based on color descriptors,” in *ICAISC*, 2007.

[8] Y. Wang, Q. Pan, C. Zhao, and Y. Cheng, “Fuzzy color histogram based kernel tracking under varying illumination,” in *FSKD*, 2007.

[9] C. Chuang, J. Hsieh, L. Tsai, P. Ju, and K. Fan, “Suspicious object detection using fuzzy-color histogram,” in *ISCAS*, 2008.

[10] D. Liu and T. Chen, “Soft shape context for iterative closest point registration,” in *ICIP*, 2004.

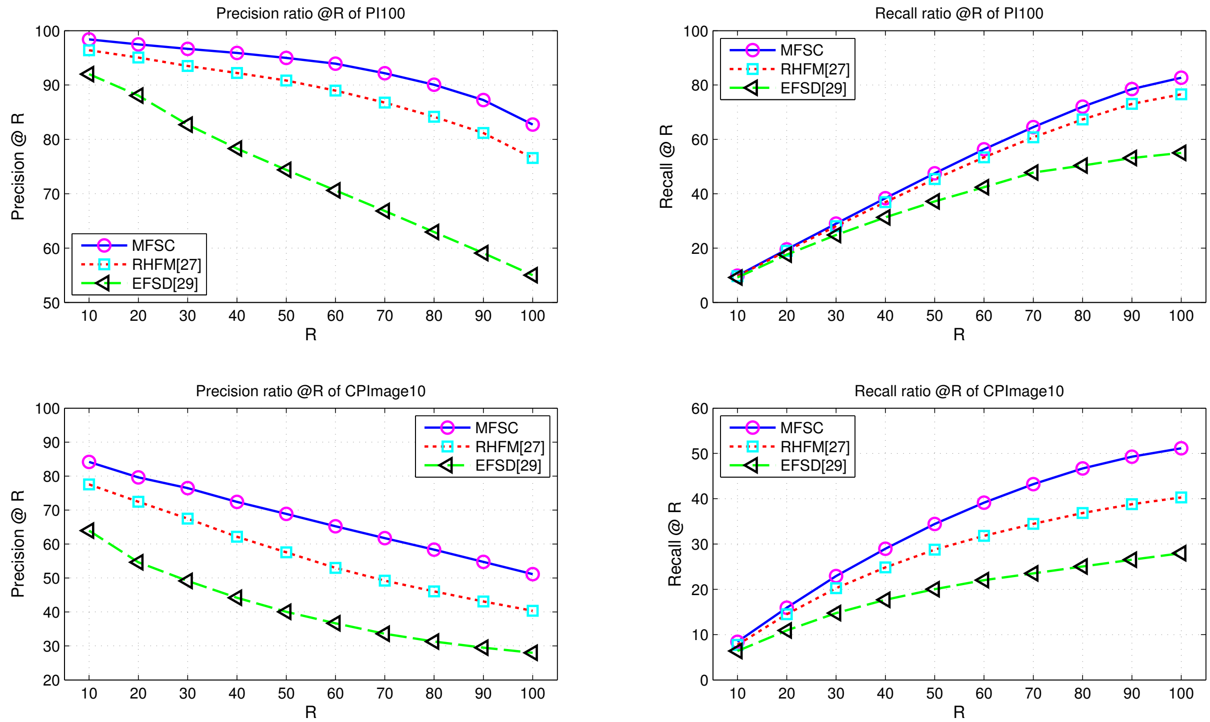

Fig. 5. Retrieval results on PI100 and CPImage10 datasets

[11] L. Wang, J. Shi, G. Song, and I. Shen, "Object detection combining recognition and segmentation," in *ACCV*, 2007.

[12] A. B. Ayed, M. Kardouchi, and S. Selouani, "Rotation invariant fuzzy shape contexts based on eigenshapes and fourier transforms for efficient radiological image retrieval," in *ICMCS*, 2012.

[13] P. Indyk and N. Thaper, "Fast image retrieval via embeddings," in *Workshop on Statistical and computational Theories of Vision*, 2003.

[14] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *ICCV*, 2005.

[15] H. Ling and K. Okad, "Emd-l_1: An efficient and robust algorithm for comparing histogram-based descriptors," in *ECCV*, 2006.

[16] H. Ling and K. Okad, "Diffusion distance for histogram comparison," in *CVPR*, 2006.

[17] L. A. Zadeh, "Fuzzy sets," in *Information and Control*, 1965.

[18] D. Lowe, "Distictive image features from scale-invariant keypoints," in *IJCV*, 2004.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005.

[20] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," in *TPAMI*, 2007.

[21] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "On aligning curves," in *TPAMI*, 2003.

[22] Z. Tu and A. L. Yuille, "Shape matching and recognition-using generative models and informative features," in *ECCV*, 2004.

[23] F. Mokhtarian, S. Abbasi, and J. Kittler, *Efficient and Robust Retrieval by Shape Content through Curvature Scale Space*, 1997.

[24] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using affine-invariant regions," in *TPAMI*, 2005.

[25] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," in *TIP*, 2003.

[26] A. C. Jalba, M. H. F. Wilkinson, and J. B. T. M. Roerdink, "Shape representation and recognition through morphological curvature scale spaces," in *TIP*, 2006.

[27] L. Chai, H. Zhang, Z. Qin, J. Yu, , and Y. Qi, "Multi-feature content-based product image retrieval based on region of main object," in *ICICS*, 2011.

[28] H. Ren, Z. Ping, W. Bo, W. Wu, and Y. Sheng, "Multidistortion-invariant image recognitin with radial harmonic fourier moments," in *Optical Society of America*, 2003.

[29] F. P. Kuhl and C. R. Giardina, "Elliptic fourier features of a closed contour," in *Computer Graphics and Image Processing*, 1982.

[30] X. Xie, L. Lu, M. Jia, H. Li, F. Seide, and W.g Ma, "Mobile search with multimodal queries," in *Proc. of the IEEE*, 2008.

[31] L. Chai, Z. Qin, H. Zhang, J. Guo, and C. Shelton, "Re-ranking using compression-based distance measure for content-based commercial product image retrieval," in *ICIP*, 2012.