International Journal of Soft Computing 12 (1): 79-85, 2017
ISSN: 1816-9503

Sketching Method Based on Earth Mover's Distance for Image Contour Matching

F. Nayyeri and M.F. Nasrudin
Faculty of Information Science and Technology, Center for Artificial Intelligence Technology,
University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract: Finding similar images to a given query image can be computed by different distance measures. One of general distance measures is the Earth Mover's Distance (EMD). Although, EMD has proven its ability to retrieve similar images in >95% true, high execution time is its major drawback. Therefore, previous algorithms of EMD could not run efficiently when performing retrievals from large databases. A contour-matching algorithm has been presented that quickly estimate the minimum weight matching using an embedding of the EMD into L_1. This low-distortion algorithm somehow solves the time problem by sacrificing the performance due to generating heavily tailed image feature vector.

Key words: Earth Mover's Distance, EMD, sketching, dimension reduction, statistics, feature vector

INTRODUCTION

Image retrieval in large databases is a problem of interest in vision and database communities. The central question is how to design a similarity measure that quantifies the perceptual notion of two images being similar. Suppose, there is a large image database and a query image which can be identified manually by a user or created automatically by an application. Then, the similarity between each image in the data base and the query image can be determined by an appropriate similarity model which computes the distance between their corresponding feature representations. Retrieving the most similar images is possible using computed similarity values via distance values (Beecks *et al*., 2010). The appropriate distance measure should redirect human perception. In other words, there should be smaller distance value between similar images based on human perception and larger distance value between perceptually different images (Zhang and Lu, 2003). To know how good the retrieval is a performance measurement is needed to measure accuracy of the retrieval.

Similarity measurements: To calculate similarity between 2 images $A = \{a_i\}$ and $B = \{b_i\}$ numerous similarity measures have been proposed. They are divided into 2 categories.The cell-by-cell similarity measures such as Minkowski, Cosine distance, Kullback-Liebler divergence and Jeffrey divergence. They only compare contents of corresponding image cells, means comparing a_i and b_i for all i but not a_i and b_j for $i \neq j$. The second category that is cross-cell measures, like quadratic distance, Mahalanob is Kolmogorov-Smirnov distance, match distance and earth mover's distance compares non-corresponding cells using the ground distance d_{ij} which is the distance value between corresponding feature representation for cell i and cell j. Unsurprisingly, the cross-cell measures are less sensitive than cell-by-cel,l measures to the position of cell boundaries (Rubner *et al*., 2000).

Distance measurements: In arithmetic, the Euclidean distance which is also known as cartesian distance is the normal distance calculated between 2 points with ruler. Suppose, there are 2 points $p = (p_1, p_2, ..., p_n)$ and $q = (q_1, q_2, ..., q_n)$ in an n-dimensional space. The Euclidean distance computed between p and q is given by Eq. 1:

$$\text{Euclidean distance } (p,q) = \sqrt{\begin{matrix}(p_1 - q_1)^2 + (p_2 - q_2)^2 + \\ ... + (p_n - q_n)^2\end{matrix}} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (1)$$

Using Eq. 1, a vector in the metric space is described as a line between 2 points in Euclidean space that we call them vector tip and vector tail. Considering the length of this line is the calculated distance between its tip and its tail, clearly the Euclidean norm of this vector is the Euclidean distance between its tip and tail (Nayyeri and Nasrudin, 2015). To reveal how similar 2 image are the minimum cost matching between their features makes a set of correspondences. Computing the optimal similarity image retrieval has a complexity which directly depends on the size of database. That is to say, the complexity of comparison between images in a large databaseis magnified.

Corresponding Author: F. Nayyeri, Faculty of Information Science and Technology, Center for Artificial Intelligence Technology, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

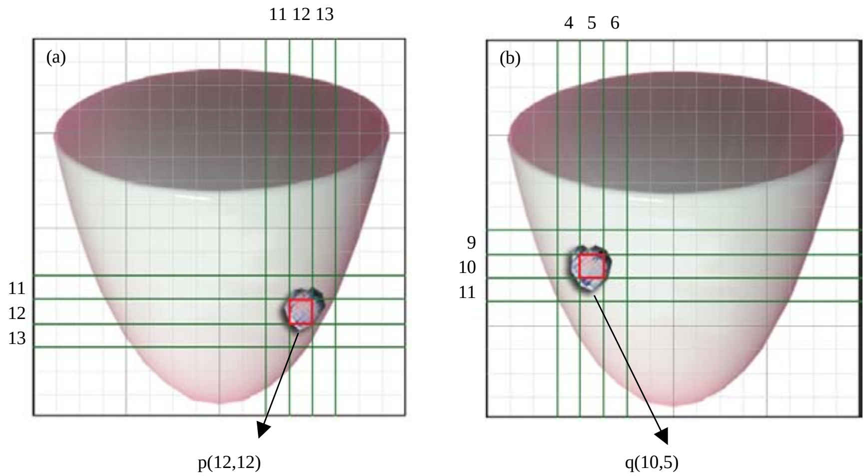


Fig. 1: Two bowls with designs in different positions

Early work on image retrieval typically solved this issue by representing each image by a point in a multidimensional space and using norms (e.g., Euclidean norm) to define the distance between 2 such points. For example, a natural method for extracting colour information of an image is to compute its colour histogram. That means splitting the colour space into cells and counting the number of pixels that fall into that cell for each cell and finally computing the Euclidean distance between each pair of cells. In Fig. 1, two images with a little dissimilarities are illustrated. This dissimilarity is in the form of different positions in the design of 2 bowls. Euclidean distance (EU) of two corresponding designs in these images is computed as following:

$$\begin{aligned}\text{Euclidean distance}(p,q) &= \sqrt{(p_{1x}-q_{1x})^2+(p_{2x}-q_{2x})^2}\\ &= \sqrt{(12-10)^2+(12-5)^2}\\ &= \sqrt{53}=7.28\end{aligned}$$

The minimum cost matching with enhancing a path algorithm is achieved by Belongie *et al*. (2002) that estimates the transform between two shapes. Using this method an impressive results of shape matching is obtained; however, due to cubic complexity of the proposed shape correspondence method high running time is noted. Afterwards, a new approach was proposed to detect the common patterns in two images using EMD as the framework (Tan and Ngo, 2005). This method provides the best estimation in terms of the scale and location of the common patterns.

MATERIALS AND METHODS

The Earth Mover's Distance (EMD) is a similarity measure for weighted point sets. For the first time EMD was introduced by Rubner *et al*. (2000) to improve the irregularities regarding the perceptual similarity experimented in other similarity measures such as Kullback-Leibler divergence, Minkowski-form distance and Jeffrey divergence. The emd provides a mechanism to compute the similarity of 2 multi-dimensional probability distributions in some feature space. As a very simple description, consider we have two distributions: one is a collection of holes in space and the other is a mass of the earth spread out in the same space. The EMD measures the minimum amount of works required to fill the holes with earth. A formal definition of the EMD will be given shortly. Suppose:

$$W=\sum_{i=1}^{m} w_i \text{ and } U=\sum_{j=1}^{n} u_j$$

are the total weighted point sets of A and B, respectively. To find a rigid motion of A to B with the minimum EMD, bisconsidered to be fixed and A is transformed to B. The transformation from A to B only change the position of points in terms of translations and rotations not their weights. All possible rigid motions in the plane is denoted by I that is the rotation about the origin by angle θ shown by R_θ and the translation by t in the plane shown by T_t. In general, transformed version of A is denoted in Eq. 2. Translated version of A to B:

$$A(t)=\{a_1(t),(a_2)(t),\ldots,a_m(t)\}$$

Rotated version of A to B:

$$A(t,\theta)=\{a_1(\theta),(a_2)(\theta),...,a_m(\theta)\} \tag{2}$$

Transformed version of A to B:

$$A(t,\theta)=\{a_1(t,\theta),(a_2)(t,\theta),...,a_m(t,\theta)\}$$

The EMD between A and B is a function EMD defined in Eq. 3:

$$EMD(A,B)=\min_F \frac{\sum_{i=1}^{m}\sum_{j=1}^{i} d_{ij}(t,\theta)}{\min\{Q,U\}} \tag{3}$$

where, d_{ij} is the Euclidean distance between A and B with F being a set of all feasible flows between A and B.We deal with the Euclidean EMD where d_{ij} is given by the L_2-norm. Informally, the minimum amount of work is measured by EMD to transform one set into the other one by weight transportation (Cabello *et al*., 2008). Potential use of EMD for measuring shape similarity was first proposed. After that the EMD became a popular similarity measure in image retrieval applications and computer vision (Rubner *et al*., 2000; Lv *et al*., 2004) shape matching (Cohen and Guibasm, 1999, Giannopoulos and Veltkamp, 2002, Grauman and Darrell, 2004) and music score matching (Cabello *et al*., 2008). For these applications, having a fast estimation to calculate the minimum distance between 2 weighted point sets is useful. In the above example to count EMD, Euclidean distances between all weighted point sets should be computed and then minimum distance between each pair of point sets will be found. Finding the minimum distances between each pair of points in two images with η points is in (η^3) arithmetic operations. Therefore, the typical EMD is very time-consuming and it is the most important drawback for EMD.

Embedding of earth mover distance into normed spaces: To exemplify the concept of EMD, assume an image representing by a set of pixels and each pixel is a point in 2-dimensional plane. The distance between 2 point sets of two different images is formally defined as the least weight matching between those point sets which is the minimum amount of work required to transform one set to researchers.

Embedding the metric into a normed space means mapping each point in the metric space (2-dimentional) in to a point in a normed space (1-dimensional) so that the distance between the images of every 2 points is comparable to the distance between the points themselves. The least weight matching of contour features is calculated via the EMD embedding into (Nayyeri and Nasrudin, 2013).

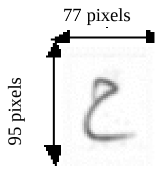
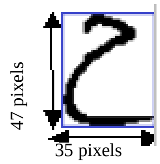
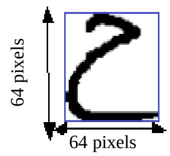
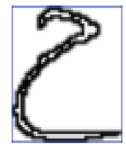
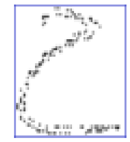
Dimension reduction in L_1: The exact EMD has the complexity of (η^3) and the embedding EMD to L_1 which computes approximation instead of exact EMD, reduces the complexity to $0(\eta^2)$. Indyk and Thaper in their work provide a low-distortion embedding of EMD into L_1. Their mapping has a provable distortion. However, their experimental results show that the empirical distortion is much smaller. This allows them to reduce the execution time of search problem in EMD to the equivalent problem in L_1.

Sketching: In recent year a novel approach for similarity search in un-normed metrics has been proposed. Sketching technique is proposed to reduce the complexity of EMD to $0(\eta)$ by using dimension reduction in L_1 (Nayyeri and Nasrudin, 2015). In this method, after scanning the data, the original data of image matrix Ais multipliedby the data of image matrix B which both are mapped to L_1 by a random vector R. Each element of the vector R is either 0 or 1 and the subtraction of their results is considered as the corresponding element of sketch vector. This step is repeated k times and finallysum of all elements of sketch vector is what is called sketching EMD. Using the mapping method as explained, the similarity search data structure is implemented by constructing a nearest neighbour data structure; then in order to answer query, the data structure is simply queried. The properties of the embedding imply that this approach guarantees that the answer is an approximate nearest neighbour where the approximation factor depends on the distance distortion incurred by the mapping.

Applying both methods of embedded EMD and sketching needs some pre-processing on query and test images. As explained in detail by Nayyeri and Nasrudin (2013), first of all image have to be converted to bitmap with only 0 and 1 bits. Then the image is resized to 64×64 pixels. In the next step, the contour of image is extracted. Then the number of black spots is equalized in order to have the same length of L_1 vector (in this work the black spots are equalized to 150). Finally for exact EMD, equal number of black spots of images are selected randomly and for sketching method the processed image is mapped to L_1 vector. The pre-processing steps for one of the query images is illustrated in Table 1.

In order to map a 2-dimensional image matrix to 1-dimensional vector, each 4 neighbouring pixels are added and the results is assigned to the corresponding place of another matrix. This process will repeat till only one element is assigned in last mapping. As an example, the steps of mapping elements of a 16 times 16 pixel image is shown in Fig 2. Afterwards, all the elements will be assigned to a 1-dimenstional vector. This process is shown in Fig. 3.

Table 1: Pre-processing steps onone sample image

Original images	First process	Second process	Third process (No. of pixels: 384)	Fourth process (No. of pixels: 150)
				

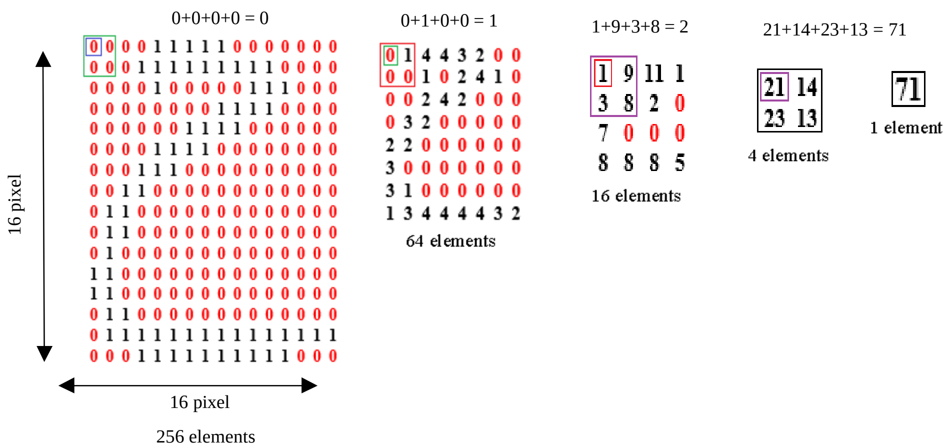


Fig. 2: Mapping process of a 16×16 image

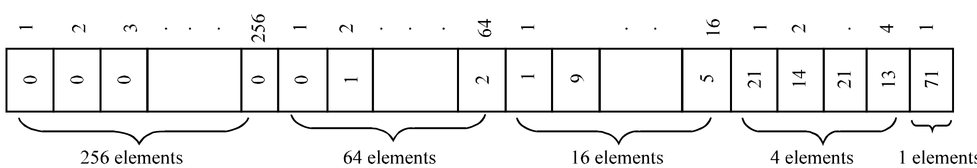


Fig. 3: L_1 embedding vector (length for 16×16 pixel image)

RESULTS AND DISCUSSION

Three methods of EMD, embedded EMD and sketching are implemented in this research. This method is applied on the Persian letter dataset (Zhang and Lu, 2003). Although, only classes of the dataset with non-dot letters are chosen. The selected dataset is divided in to 15 classes including 11,790 images. In Table 2 some of the samples from this dataset are shown randomly.

The results of pre-processing on some of randomly selected query images are presented in Table 3. Table 4 shows the results of performance and execution time of these methods for 100 randomly selected images in the dataset. The performance is judged based on Mean Average Precision (MAP). Firstly, the Average Precision (AP) of relevant retrieved images among the 10 top-ranked retrieved images of the test image sets is calculated. MAP is the average across multiple evaluations of AP. As it can be cleared from Table 4, the performance of exact EMD is 97% with the highest execution time (37 h and 54 min). The embedded EMD improved the execution time (by 58%) by sacrificing the performance.

In the sketching method applied on Persian letter dataset, the performance is improved by 94% and the running time is increased by 12% (Table 3 and 4). Sketching is a dimensionally reduced method designed to approximate the specific summary statistics. This algorithmmay be regarded as special-purpose data compressions and often outperforms embedding EMD by sacrificing the execution time. In this reserach this dimension reduction method is applied on non-dot letters of Persian letter dataset. Since, the letters with dots needs some improvements in method to have better results. Some of the samples of Persian letters with dots are illustrated in Table 5.

Table 2: Some samples of handwritten persian letter images

Classes	Letters	Correct shapes	Samples 1	2	3	4	5
1	A						
2	A (2)						
3	Da						
4	Ha						
5	Ra						
6	Se						
7	Sa						
8	Ta						
9	Aa						
10	Ka						
11	La						
12	Me						
13	Ve						
14	He						
15	Ye						

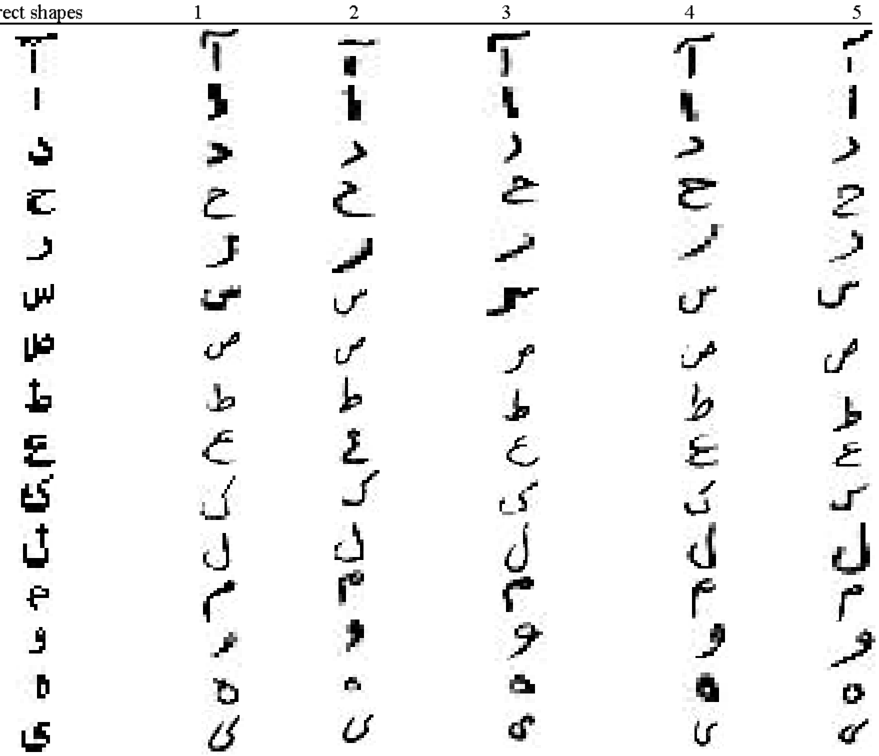
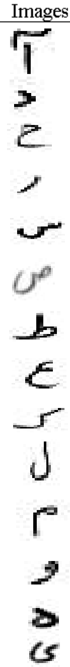


Table 3: Pre-processing results onchosen classes of handwritten persian letter images

Classes	Images	Dimension	Pre-process steps (Width: 64, Height: 64) 1	2	3	4
1		Width: 46, Height: 70				
2		Width: 19, Height: 25				
3		Width: 35, Height: 47				
4		Width: 25, Height: 37				
5		Width: 53, Height: 36				
6		Width: 51, Height: 36				
7		Width: 52, Height: 53				
8		Width: 37, Height: 56				
9		Width: 61, Height: 45				
10		Width: 33, Height: 59				
11		Width: 29, Height: 46				
12		Width: 33, Height: 33				
13		Width: 16, Height: 24				
14		Width: 35, Height: 35				



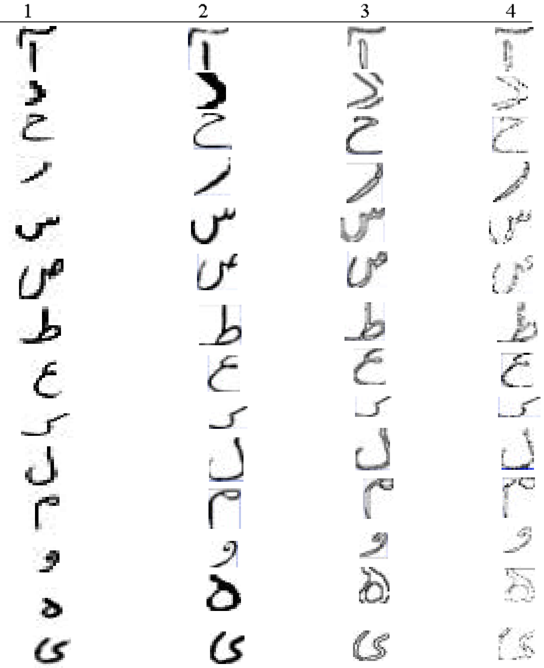
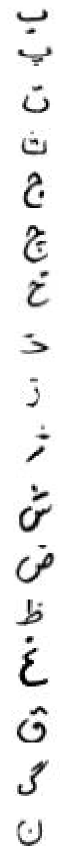


Table 4: Results of 3 methods

Methods	Complexity	Performance MAP (%)	Execution time hh:mm
Exact EMD	$0\ (\eta^3)$	97	37:54
Embedded EMD	$0\ (\eta^2)$	84	15:42
Sketching	$0\ (\eta)$	89	17:38

Table 5: Some samples of Persian letters with dots

Correct shapes	Samples
ب	
پ	
ت	
ث	
ج	
چ	
خ	
ذ	
ز	
ژ	
ش	
ض	
ظ	
غ	
ق	
گ	
ن	



CONCLUSION

This study aims to improve the performance of embedded EMD by a method of dimension reduction called sketching which is a distance estimation method based on specific summary statistics. Finally, this methodis applied on Persian letter matching using 11,790 images and the results show an improvement in the performance of embedded EMD at the cost of 12% more dilation than embedded EMD.

RECOMMENDATIONS

As a future research, these improvements will be applied on sketching method in order to increase the performance.

REFERENCES

Beecks, C., M.S. Uysal and T. Seidl, 2010. A comparative study of similarity measures for content-based multimedia retrieval. Proceedings of the 2010 IEEE International Conference on Multimedia and Expo (ICME), July 19-23, 2010, IEEE, Germany, Europe, ISBN:978-1-4244-7491-2, pp: 1552-1557.

Belongie, S., J. Malik and J. Puzicha, 2002. Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell., 24: 509-522.

Cabello, S., P. Giannopoulos, C. Knauer and G. Rote, 2008. Matching point sets with respect to the Earth movers distance. Comput. Geom., 39: 118-133.

Cohen, S. and L. Guibasm, 1999. The earth mover's distance under transformation sets. Proceedings of the 7th IEEE International Conference on Computer Vision 1999, Vol. 2, September 20-27, 1999, IEEE, California, USA., ISBN: 0-7695-0164-8, pp: 1076-1083.

Giannopoulos, P. and R.C. Veltkamp, 2002. A pseudo-metric for weighted point sets. Proceedings of the European Conference on Computer Vision, May 28-31, 2002, Springer, Berlin, Germany, pp: 715-730.

Grauman, K. and T. Darrell, 2004. June. Fast contour matching using approximate earth mover's distance. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), Vol. 1, June 2, 2004, IEEE, Cambridge, Massachusetts, ISBN: 0-7695-2158-4, pp: 1-27.

Lv, Q., M. Charikar and K. Li, 2004. Image similarity search with compact data structures. Proceedings of the 13th ACM International Conference on Information and Knowledge Management, November 08-13, 2004, ACM, Washington, D.C., USA., ISBN: 1-58113-874-1, pp: 208-217.

Nayyeri, F. and M.F. Nasrudin, 2013. Image matching using dimensionally reduced embedded earth movers distance. J. Appl. Math., 2013: 1-11.

Nayyeri, F. and M.F. Nasrudin, 2015. Similarity Comparison of Images Based on Earth Movers Distance. Lambert Academic Publishing, Saarbrucken, Germany, Pages: 108.

Rubner, Y., C. Tomasi and L.J. Guibas, 2000. The earth mover's distance as a metric for image retrieval. Intl. J. Comput. Vision, 40: 99-121.

Tan, H.K. and C.W. Ngo, 2005. Common pattern discovery using earth mover's distance and local flow maximization. Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV 2005), Vol. 2, October 17-21, 2005, IEEE, Hong Kong, China, ISBN:0-7695-2334-X, pp: 1222-1229.

Zhang, D. and G. Lu, 2003. Evaluation of similarity measurement for image retrieval. Proceedings of the IEEE International Conference Neural Networks and Signal Processing, December 14-17, 2003, Nanjing, China, pp: 928-931.