Efficient Word Image Retrieval Using Earth Movers Distance Embedded to Wavelets Coefficients Domain

Raid Saabni
Department of Computer Science
Triangle R&D Center, Kafr Qarea
Tel Aviv-Yaffo Academic College, Tel-Aviv, Israel
Email: raidsa@mta.ac.il

***Abstract*—In this paper we use the Earth Movers Distance (EMD) algorithm to measure similarity between shapes for recognizing and searching Arabic words. We have used the Shape Context and the Angular Radial Partitioning descriptors to evaluate matching and recognizing with EMD. Based on the encouraging results of high accuracy and recall, we follow the low-distortion embedding of the Earth Mover's Distance to map the shapes in the database under the EMD distance, into a normed space of wavelet coefficients as differences of coefficients histograms. The approximate k-nearest neighbors in the database of the embedded shapes are retrieved in sub linear time using a Locality-Sensitive Hashing (LSH) and generate a short list of candidates. This short list of candidates is used in a filter and refine strategy and the exact results are achieved using the original EMD on this short list. We demonstrate our method on the MNIST dataset and the freely available Arabic Printed Text Image (APTI) database. Our method achieves a speedup of 4 orders of magnitude over the exact method, at the cost of only a 2.4% reduction in accuracy.**

I. Introduction

Off-line cursive text recognition has achieved a great attention for many years due to its important contribution in the digital libraries evolution in tasks such as key word searching, handwriting recognition and others. There are two general approaches in the literature: the $Segmentation - Based$ approach, where words have to be segmented into characters to be recognized, and the $Holistic$ $(Segmentation - Free)$ approach which entails the recognition of the whole word without segmenting it to characters. In a holistic approach, recognizing a given shape is a very expensive task when the size of the lexicon is large. In keyword searching and script recognition tasks, shapes have to be matched to sets of multiple appearances of different words which may lead to matching to few millions of shapes. For example, computing the similarity of a given word to multiple shapes of each word in the English lexicon based on Dynamic Time Warping (DTW) similarity measurement may take more than few hours on an average personal computer. Therefore, to reduce time needed for searching (matching) a shape within a large lexicon, it is desirable to make as few as possible distance calculations to candidates, where, sub-linear size is preferable. This option is not feasible with non-metric spaces, such as many existing similarity functions, including DTW and others. A known and familiar technique to achieve such improvement will include mapping, or embedding, the set of shapes represented as vectors(points) in some feature space into a low-dimensional space and then conducting the search there with the help of multidimensional indexing methods. Intuitively, such embedding is required to preserve distances which means that distances in the embedding space have to approximate distances of the objects in the original space. At the same time, searching time in the embedding space is expected to be significantly reduced. The earth movers distance (EMD) is a natural and intuitive metric to measure similarity between histograms. If we think about the first histogram as a set of piles of sand sitting on the ground, one pile for each coordinate. The second histogram can be seen as a set of holes. To quantify the difference between the two histograms, we can measure the cost (flow) of moving the grains of sand in order to fill the holes. In this case, EMD is the minimal total ground distance traveled weighted by the amount of sand moved (flow). EMD makes sure that shifts in sample values are not penalized excessively. EMD has been successfully used for image retrieval [9, 21], image registration [5] and pattern matching in medical images [12].

In this paper, we use the EMD metric to measure similarity between two shapes using two different features extracted from their boundary. We also use the mapping presented in [25] for embedding of different shapes represented using their contours into a normed space. Embedding is performed using a linear time algorithm for approximating the EMD for low dimensional histograms using the sum of absolute values of the weighted wavelet coefficients of the difference histogram. The embedding of our shapes as feature vectors to a normed space enables the use of state-of-the-art methods for fast computation of the approximate k-nearest neighbors. Finding the approximate k-nearest neighbors produces a short list which enables applying expensive matching methods, yet keeps sub linear searching time. The high performance of the presented approach enables matching shapes from handwritten documents to large lexicons of words for indexing, keyword searching, or reading of hand printed documents.

The rest of this paper is organized as follows: in section II, we briefly overview the state-of-the-art in handwriting word matching and searching. In section III, we describe our approach in details. Experimental results and some directions for future work are presented in sections IV and V.

II. Related Work

Two general approaches for word matching have been presented in the literature, the segmentation and the holistic or segmentation-free approaches. In the segmentation approach, the crucial step is to split a scanned bitmap image of a document into individual characters. In segmentation-free approach line and word segmentation are used for creating an index based on word matching. Rachev and Ruschendorf[19] discussed different approaches to word matching. Word image matching was applied in [17] using the weighted Hausdorff distance. The authors in [14] performed word matching using global and local features based on profile signatures and morphological cavities. A voting system is used in [28] to fuse multiple handwritten word recognition techniques based on ranks and confidence values.

Many systems for word spotting and searching presented in previous work are based on variants of DTW. Different sets of features were used and gave good results comparing to the competing techniques [18]. Manmatha *et al.* [18] were among the first to introduce DTW for word spotting. They examined several matching techniques and showed that DTW, in general, provides better results. Saabni and El-Sana [22, 23] used DTW with features extracted from contours [22] or sliding windows [23] to compare shapes for keyword searching tasks where templates of the searched keywords were synthetically generated. Shrihari *et al.* [27] presented a design of a search engine for handwritten documents. They indexed documents using global image features, such as stroke width, slant, word gaps, as well as local features that describe the shapes of characters and words. Lavrenko *et al.* [15], presented a segmentation-free approach using the upper word and projection profile to spot word images without segmenting into individual characters and showed feasibility of their approach even for noisy documents. Gatos *et al.* [7], propose a segmentation-free approach which combines image preprocessing, synthetic data creation, word spotting and user feedback techniques.

Much work has been done on embedding finite metric spaces into low-dimensional normed spaces in order to enable efficient and fast nearest neighbor extraction. Such embeddings have been extensively studied in pure mathematics [3], and have found application in a variety of settings [6], usually using one of the l_p norms. In domains with a computationally expensive distance measure, significant speed-ups can be obtained by embedding objects into another space with a more efficient distance measure. Several methods have been proposed for embedding arbitrary spaces into a Euclidean or pseudo-Euclidean space. Some of these methods, in particular Multidimensional Scaling (MDS)[30], Bourgain embeddings [11, 3], Locally Linear Embedding (LLE) [20], need to evaluate exact distances between the query and most or all database objects, and thus are not designed for efficient nearest neighbor retrieval. Methods that can be used for efficient retrieval include Lipschitz embeddings, FastMap [6], MetricMap [29] and BoostMap [1, 24].

Efficient embedding of the EMD metric to a normed space [13] have been used by Grauman and Darrell [9] for contour based matching. Indyk and Thaper[13] use a randomized multi-scale embedding of histograms into a space equipped with the l_1 norm. The multi-scale hierarchy is obtained by a series of random shifting and dyadic merging of bins. They show that the l_1 norm computed in this space, averaged over all random shifts, is equivalent to the EMD. Additional efficient embedding of the EMD metric has been presented by Shridonkar and Jacobs[25] using a novel method for approximating the EMD distance of two histograms using a new metric on the weighted wavelet coefficients of the difference histogram using the l_1 distance. This new metric was experimentally shown to follow EMD closely without any significant performance difference. The wavelet EMD metric can be computed in $O(n)$ time.

III. Our Approach

In the presented work, we address the problem of recognizing an individual Arabic word seen as a multi-component shape. The MNIST dataset of handwritten digits, Follows the constraint restricting each shape to be one connected component, therefore the Shape Context(SC) [2] descriptor is used with it. The Angular Radial Partitioning (ARP) [4] descriptor works directly on multicomponent shapes, therefore, it is used to generate feature vectors based on shapes of Arabic words with multiple components. The presented system uses the Earth movers distance (EMD) to match shapes using the histograms generated using the SC and ARP descriptors. All histogram in that feature space, are mapped into a normed space by performing an embedding process for approximating the EMD for low dimensional histograms using the sum of absolute values of the weighted wavelet coefficients of the difference histogram. Using Local Sensitivity Hashing (LSH), we find the approximate k-nearest neighbors to produce a short list on which we apply the original expensive EMD algorithm to present final results. In the next subsections, we explain the feature extraction process, matching with EMD and the optimization process for fast word retrieval using embedding and Local Sensitivity Hashing (LSH).

A. Feature Extraction

The first feature set we have used to match shapes as one connected component was the 'Shape Contexts' descriptor which was presented by Belongie and Malik *et al.* [2]. It is a boundary based descriptor which describes a distribution of all boundary points with respect to each point on the boundary. The Shape Context descriptor of each single boundary point computes the histogram of relative polar coordinates and have been proved to be one of the efficient features for matching binary images. There is no direct extension of the Shape Context to multi component shapes considering its dependency with a clear and ordered boundary of the matched component. The Angular Radial Partitioning [4], transforms an image data which may contain multi components into a new structure that supports measurement of the similarity between images. In the first step the image is converted to gray intensity followed by applying an edge detection algorithm. The resulting edge map image is normalized to WXW pixels in order to achieve the scale invariance property. The algorithm uses the surrounding circle of the resulted image for partitioning it into MXN sectors, where M is the number of radial partitions and N is the number of angular partitions. The angle between adjacent angular partitions is $\theta = 2\pi/N$ and the radius of successive concentric circles is $\rho = R/M$, where R is the radius of the surrounding circle of the image. The number of edge points in each sector of I is chosen to represent the sector feature, See Figure Fig. 1.

B. Earth Mover's Distance

Computing the similarity distance between two sequences is an easy task when the ground distance between two single

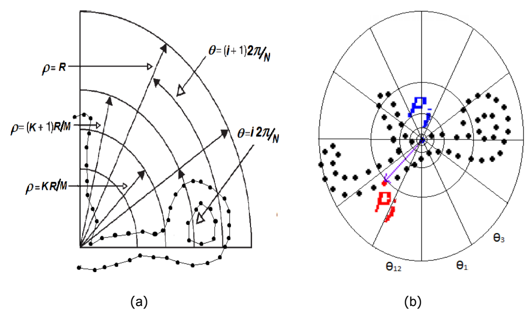

Fig. 1. (a) The Angular Radial Portioning Descriptor, and (b) the Shape Context Descriptor from one contour point view.

perceptual features is given and the sequences are totally aligned. This measurement becomes more complicated when the sequences of features are not aligned and no obvious correspondences between different bins in the two sequences can be found. The Earth Mover's Distance (EMD) is a method to evaluate dissimilarity between two multi-dimensional distributions in some feature space where a ground distance between single features is given. Intuitively, given two distributions, one can be seen as piles of sand and the other as a collection of holes. The EMD measures the least amount of work needed to fill the holes with the sand. Here, a unit of work corresponds to transporting a unit of sand from one pile to a hole depending on their distance.

Computing the EMD is based on a solution to the well-known transportation problem [10]. Suppose that several suppliers, each with a given amount of goods, are required to supply several consumers, each with a given limited capacity. For each supplier-consumer pair, the cost of transporting a single unit of goods is given. The transportation problem is then to find a least-expensive flow of goods from the suppliers to the consumers that satisfies the consumers' demand. Matching histograms (in the presented case, feature descriptors of contour points are treated as histograms) can be naturally cast as a transportation problem by defining one histogram as the supplier and the other as the consumer, and by setting the cost for a supplier-consumer pair to equal the ground distance between an element in the first histogram and an element in the second. Intuitively, the solution is then the minimum amount of work required to transform one signature into the other.

Formally, let us define the first signature to be $\{(q_i, w_i)\}_{i=1}^{m}$ with m entries and the second as $\{(p_i, w_i)\}_{i=1}^{n}$ with n entries(p_i) having the weights(w_i). Let the flow between p_i to q_j be f_{ij} and d_{ij} be the ground distance between the entries p_i and q_j. We can solve this problem using the following linear programming problem: Find the flow $F = [f_{ij}]$ that minimizes the following work for the signatures P and Q:

$$Work(P, Q, F) = \sum_{j=1}^{n} \sum_{i=1}^{m} f_{ij} d_{ij} \tag{1}$$

Subject to constraints to allows moving supplies from P to Q and not vice versa and limiting the amount of supplies that can be sent by the clusters in P to their weights, and the clusters in Q to receive no more supplies than their weights, and force to move the maximum amount of supplies possible. The normalization factor is the total weight of the smaller signature which is needed when the two signatures have different total weight, in order to avoid favoring smaller signatures.

C. Embedding to the Wavelet Coefficients Domain

Embedding the set of shapes into a normed space, will enable fast approximate search with sub-linear time and improve the efficiency. We start the process by converting all shapes to the feature space using Shapes extracted contour in clockwise direction for handwriting digits and the ARP descriptor for the Arabic words. The embedding process presented by Shirdhonkar and Jacobs [25], requires histograms and not signatures which means that SC sequences need to be normalized to the same size. After embedding is done a serious of Locality sensitive hashing function are generated to enable fast search of the presented word part. Shirdhonkar and Jacobs[25] presented a novel method for approximating the EMD for histograms using a new metric on the weighted wavelet coefficients of the difference histogram and show that this is equivalent to EMD. Experimentally they show that this metric follows EMD closely and can be used instead without any significant performance difference. The wavelet EMD metric can be computed in $O(n)$ time. For applications that need to store computed histogram descriptors for future use, they split the wavelet EMD computation into two parts. First, the histogram descriptor is converted into the wavelet domain and its coefficients are scaled according to equation (2). Computing the EMD distance in the next step is done as the l_1 (Manhattan) distance between these coefficients. In order to choose the wavelet domain for embedding, the authors have tested few, and showed that the $Coiflets$ of order 3 and the $Symmetric\ Daubechies$ wavelets of order 5 had low error rates. We follow their results and use order 3 $Coiflets$ in our system. Intuitively speaking, the wavelet transform splits up the difference histogram according to scale and location. Each wavelet coefficient represents an EMD subproblem that is solved separately. For a single wavelet, the mass to be moved is proportional to the volume of $|\psi_j(x)|$, i.e. to $2^{jn/2}$. The distance traveled is proportional to the span of the wavelet 2^{-j} (According to Meyers [14] convention, a wavelet at scale j is the mother wavelet squeezed 2^j times.) The sum of all distances is an approximation to EMD and called the wavelet EMD between two histograms, as seen from equation (2).

Approximation by scale and location separation is similar to the way packages are shipped over large distances. The total journey is broken into several hops, short and long. Short hops connect the source and destination to shipping hubs, while long hops connect the shipping hubs themselves. Packages from nearby towns merge at shipping hubs to travel together. Thus, the package journey is split into multiple scales, and the sum of the distances traveled is an approximation to the actual distance.

$$d(p)_{wemd} = \sum_{\lambda} 2^{-j(1+n)/2} |p\lambda| \tag{2}$$

where, p is the n dimensional difference histogram and p_λ are its wavelet coefficients. The index λ includes shifts and the scale j.

Following Shirdhonkar and Jacobs[25] approach, we also split the wavelet EMD computation into two parts. First, the histogram descriptors are converted into the wavelet domain using $Coiflets$ of order 3 resulting a $2500-coordinate$ vector of coefficients for each shape. In the next step, computing the EMD distance is done as the l_1 (Manhattan) distance between these vectors coefficients. In our case, the set of shapes was converted to a set of histograms and mapped to the normed space of weighted wavelet coefficients to approximate the EMD metric.

D. Fast k-Nearest Neighbor approximation

Computing exact nearest neighbors in high dimensions is a very difficult task. Few methods seem to be significantly better than a brute-force computation of all distances. However, it has been shown that by computing nearest neighbors approximately, it is possible to achieve significantly faster running times with a relatively small actual errors. Methods and structures for both exact and approximate nearest neighbor searching such as kd-tree and box decomposition trees can not be used due to the high dimensionality of the given feature space. In the presented approach, we use the Locality sensitive hashing (LSH) which manages high dimensional points and guaranties better performance. Locality sensitive hashing (LSH) is a technique for grouping points in space into 'buckets' based on some distance metric operating on the points. Points that are close to each other under the chosen metric are mapped to the same bucket with high probability. This is based on the simple idea that, if two points are close together, then after a projection operation these two points will remain close together. The basic idea is to hash the input items so that similar items are mapped to the same buckets with high probability (the number of buckets being much smaller than the universe of possible input items). LSH [8], uses several hash functions of the same type to create a hash value for each point of the dataset. Each function reduces the dimensionality of the data by projection onto random vectors. The data is then partitioned into bins by a uniform grid. Since the number of bins is still too high, a second hashing step is performed to obtain a smaller hash value. At query time, the query point is mapped using the hash functions and all the data points that are in the same bin as the query point are returned as candidates. The final nearest neighbors are selected by a linear search through candidate data points.

IV. Experimental results

We have used two recognition systems to test and evaluate the proposed approach. The first system is for handwriting digit recognition and was tested using the MNIST dataset with six different configurations. The six configurations use three methods of matching, (the Hungarian method (HUM), the EMD and Embedded EMD(Embedded EMD)) and two shape descriptors, (SC and the ARP). The second system is a key word searching system using eight configurations. Four matching algorithms with the same two sets of features, the Shape Context and The ARP. The four matching algorithm are the previous three and the well known Dynamic Time Warping algorithm. The Hungarian method was used in [2] to find the correspondence between the two sets of points in the matched shape. The shape context descriptor of each point was used to measure there similarity between each pair of points. While the Hungarian method is used to find the correspondence, the DTW and EMD algorithms are used to measure the similarity between the two shapes without finding the correspondence between points of the contours. Using the Shape Context with complete Arabic words was done by neglecting additional strokes and joining results of recognizing each connected component (Word-Part) in the word to one value indicating the similarity. With the second system, we have used the APTI [26] printed Arabic database. The APTI database was created in low-resolution with a lexicon of $113,284$ different Arabic words as testing samples with 10 different fonts in 10 sizes. These fonts have been selected to cover different complexity of shapes of Arabic printed characters. We have used the evaluation protocols 13 and 14 to report results in terms of word recognition rates. Each one of these protocols uses a set of three fonts of multiple sizes for training and a set of additional two fonts with the same sizes for testing. The well known MNIST database have been used for digit recognition. The MNIST dataset, is derived from the NIST dataset, and has been created by Yann LeCun [16]. The MNIST dataset consists of handwritten digit images. The examples for training include $60,000$ examples for all digits and $10,000$ examples for testing. All digit images in this dataset have been size-normalized and centered in a fixed size image.

TABLE I. $Precision$ AND $Recall$ OF THE DIFFERENT CONFIGURATIONS OF THE TWO SYSTEMS IN TERMS OF COUNTING FALSE POSITIVES AND FALSE NEGATIVES. THE LAST COLUMN SHOWS TIME SPEEDUP IN ORDERS OF MAGNITUDE USING THE PRESENTED APPROACH.

Handwriting Digit Recognizer using MNIST Dataset					
	Shape Context Descriptor		ARP Descriptor		speedup in orders
	precision	Recall	precision	Recall	of magnitude
HUM	96.15%	95.17%	89.81%	94.6%	0
EMD	93.25%	93.8%	93.8%	93.4%	0
APX_EMD	92.1%	92.8%	92.75%	92.5%	4.5
Key Word searching using Private Dataset					
	Shape Context Descriptor		ARP Descriptor		speedup in orders
	precision	Recall	precision	Recall	of magnitude
DTW	82.6%	84.8%	86.1%	85.4%	0
HUM	81.1%	82.2%	82.9%	82.6%	0
EMD	84.1%	84.8%	85.5%	85.3%	0
APX_EMD	83.15%	84.1%	84.15%	84.4%	4.1

As we can see in Table I, results in terms of precision and recall are slightly better using the EMD for key word searching and very close to the Hungarian method (HUM) in digit recognition. Using the fact that a distance preserving embedding of the EMD can be performed without real decrease in recognition rates, a real speedup (last column in the table) of four to five orders of magnitude is achieved using the LSH. It is important to notice that the main argue of the presented approach is to enable fast and efficient search of key words while keeping high recognition rates, as we can see in Table I, therefore, we have not perform any improvement to the recognition system by using basic versions of DTW and EMD on a simple version of the database.

V. Conclusion

We have presented a novel approach for fast searching of Arabic word-parts using the Earth Movers Distance via embedding and k-nearest neighbor approximation using Locality Sensitive Hashing(LSH) to produces a short list which enables applying expensive matching methods, yet keeps sub linear searching time. Experimental results show that searching handwritten word-parts within large lexicon can be done $20,000$ times faster than comparing to all samples, without real decrease in accuracy.

VI. Acknowledgments

We would like to thank Prof' D. Ron and Dr' Alexander Bronstein from Tel-Aviv University for their insightful comments which led to several improvements in the presentation of this paper. This work has been supported in part by the German Research Foundation under grant no. FI 1494/3-2.

References

[1] V. Athitsos, J. Alon, S. Sclaroff, and G. Kollios. Boostmap: A method for efficient approximate similarity rankings. In *in IEEE Conference on Computer Vision and Pattern Recognition*, pages 268–275, 2004.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:509–522, 2002.

[3] J. Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel J. Math*, 25:46–52, 1985.

[4] Abdolah Chalechale, Alfred Mertins, and G Naghdy. Edge image description using angular radial partitioning. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 151, pages 93–101. IET, 2004.

[5] C. Chefdhotel and G. Bousquet. Intensity-based image registration using emd. In *Medical Imaging, proc of the SPIE*, 2007.

[6] C. Faloutsos and K. Lin. fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *Proceedings ACM SIGMOD Conference*, pages 163–174, 1995.

[7] B. Gatos, T. Konidaris, K. Ntzios, I. Pratikakis, and S. Perantonis. A segmentation-free approach for keyword search in historical typewritten documents. In *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, volume 1, pages 54–58, 29 Aug.-1 Sept. 2005.

[8] A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimensions via hashing. In *In Proceedings of the 25th Intl Conf on Very Large Data Bases*, 1999.

[9] K. Grauman and T. Darrell. Fast contour matching using approximate earth movers distance. In *In IEEE Conference on CVPR*, volume 01, page 220227, 2004.

[10] F. L. Hitchcock. The distribution of a product from several sources to numerous localities. *Journal of Math and Phys*, 20:224–230, 1941.

[11] G. Hjaltason and H. Samet. .properties of embedding methods for similarity searching in metric spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:530–549, 2003.

[12] A. Holmes, C. Rose, and C. Taylor. Transforming pixel signatures into an improved metric space. *Image and Vision Computing*, 20(9):701–707, 2002.

[13] P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *In 3rd International Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.

[14] P. Keaton, H. Greenspan, and R. Goodman. Keyword spotting for cursive document retrieval. In *Document Image Analysis, 1997. (DIA '97) Proceedings., Workshop on*, pages 74–81, 20 June 1997.

[15] V. Lavrenko, T. Rath, and R. Manmatha. Holistic word recognition for handwritten historical documents. In *DIAL '04: Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL'04)*, page 278, Washington, DC, USA, 2004. IEEE Computer Society.

[16] Yann LeCun. The mnist database of handwritten digits. http://www.research.att.com/ yann/exdb/mnist, AT&T Labs, 1994.

[17] Y. Lu, C. Tan, H. Weihua, and L. Fan. An approach to word image matching based on weighted hausdorff distance. *Sixth International Conference on Document Analysis and Recognition (ICDAR01)*, pages 10–13, 2001.

[18] R. Manmatha and T. Rath. Indexing handwritten historical documents - recent progress. *the Proc. of the Symposium on Document Image Understanding (SDIUT-03)*, pages 77–85, 2003.

[19] S. T. Rachev and L. Ruschendorf. Mass transportation problems. *Theory. Springer*, 1, 1998.

[20] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science 22 December*, 290:2323–2326, 2000.

[21] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth movers distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121, 2000.

[22] R. Saabni and J. El-Sana. Keyword searching for arabic handwritten documents. In *The 11'th International Conference on Frontiers in Handwriting recognition (ICFHR2008), Montreal*, pages 716–722, 2008.

[23] R. Saabni and J. El-Sana. Word spotting for handwritten documents using chamfer distance and dynamic time warping. In *Document Recognition and Retrieval XVIII*, 2011.

[24] Raid Saabni and Alexander Bronstein. Fast keyword searching using'boostmap'based embedding. In *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, pages 734–739. IEEE, 2012.

[25] S. Shirdhonkar and D.W. Jacobs. Approximate earth movers distance in linear time. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference.*, pages 1–8, 2008.

[26] F. Slimane, R. Ingold ND S. Kanoun, A. M. Alimi, and J. Hennebert. A new arabic printed text image database and evaluation protocols. In *the 10th International Conference on Document Analysis and Recognition (ICDAR)*, page 946950, 2009.

[27] S. Srihari, C. Huang, and H. Srinivasan. A search engine for handwritten documents. *Document Recognition and Retrieval XII, San Jose, CA,Society of Photo Instrumentation Engineers (SPIE)*, pages pp. 66–75, January 2005.

[28] B. Verma, P. Gader, and W. Chen. Fusion of multiple handwritten word recognition techniques. *Pattern Recognition Letters*, (22) 9:991–998, 2001.

[29] X. Wang, J. T. L. Wang, K. I. Lin, D. Shasha, B. A. Shapiro, and K. Zhang. An index structure for data mining and clustering. *Knowledge and Information Systems*, vol. 2:161–184, 2000.

[30] F. Young and R. Hamer. Multidimensional scaling: History, theory and applications. *Hillsdale, New Jersey: Lawrence Erlbaum Associates*, 1987.