Contextual Distance Refining for Image Retrieval

Islam Almasri*

Abstract: Recently, a number of methods have been proposed to improve image retrieval accuracy by capturing context information. These methods try to compensate for the fact that a visually less similar image might be more relevant because it depicts the same object. We propose a new quick method for refining any pairwise distance metric, it works by iteratively discovering the object in the image from the most similar images, and then refine the distance metric accordingly. Test show that our technique improves over the state of art in terms of accuracy over the MPEG7 dataset.

Keywords: Contextual distance, image distance, image retrieval.

1. INTRODUCTION

The problem of image matching have been previously investigated by studying the features of a pair of images, this goes by extracting features from both images and calculating a numerical similarity measure accordingly, examples of these are the bag-of-features method [1, 2], non-negative matrix factorization-based approaches [3-6], and the inner distance shape context (IDSC) [7].

However, pairwise image matching only considers the image of the object and not the contextual data [8-10]. The idea is that for an image (A) containing an object, an image that has the same object (B), might look less similar than another image for another object (C), because of the different perspective and orientation. A better pairwise similarity measure in this case will (and should) recognize image B as more similar to A than C is.

Recently, the importance of the context of the shapes has been recognized to improve the performance of shape matching. Jegou et al. [11] have proposed the Contextual Dissimilarity Measure (CDM) by assuming that a good image ranking is asymmetrical. Bei et al. [12] proposed a new method for context-sensitive shape similarity using Graph Transduction (GT), which is further improved by Wang et al. [13, 14] by using Shortest Path Propagation (SSP) on the graph [13], and using multiple graph instead of one single graph [14]. Lately, Wang et al. [15] developed the first supervised contextual similarity learning algorithm.

All the previous methods, besides the one proposed in [15], have the advantage of being unsupervised methods. The GT method performs well for being customized for each query object, however, as we will show in this paper, their method is costly, and we proposed a more efficient way of context-aware shape retrieval. Our method Contextual Distance Refining (CDR) takes a pairwise distance criterion and alters the distances by injecting the context information. Our method is still unsupervised and the context will be based on an iterative and dynamic consideration of the neighbors.

The rest of the paper is organized as follows: we lay down the basis for our method and describe the implementation algorithm in Section 2. Experimental results on are presented in Section 3. Finally, conclusive remarks are presented in Section 4.

2. CONTEXTUAL DISTANCE REFINING

In this section our method for contextual image retrieval is described. First a description of the basis of idea is provided. Then, a mathematical and algorithmic approach is proposed. And we finish with a discussion of the approach in terms of complexity, parameter dependence and how it differs to other proposed methods.

2.1. Context Deduction

The pairwise similarity measures fail to take into consideration the contextual information [16, 17], for example in (Fig. 1), the constellation of points represent the similarity measures of the images, the closer the points the more similar are the images. We can see that a point is more relevant to point than point despite being further (less similar) from.

The figure is a much more simplified representation of the distances, showing the points in a 2-dimensional plan which might suggest that distances between some points are dependent thus, can be projected on a plan. While this is not the case, the figure still shows how we can derive the context of a shape from the neighboring shapes. This extends the definition of "similar" images by considering two
parameters, in the context of this paper, an image \( x_i \) is similar image to an image \( x_j \) if:

- Image \( B \) depicts the same object in \( A \), and
- Image \( B \) is visually similar to image \( A \).

The idea behind our approach is this: assume a query image \( x_i \) for a certain object, and the top \( N \) images depicting the same object are to be retrieved. We denote the image database as a collection of image features \( X = \{x_i\}_{i=1}^{I} \), while query image feature as \( x_o \). A distance measure \( d_{ij} \) is defined between a pair of images \( x_i \) and \( x_j \), \( i, j = 1, \ldots, I \). The distance is smaller for similar images. Now a naive context-unaware approach of retrieving similar images is to rank them in ascending order according to \( d_{ij} \), and take the first \( N \). Now assume a different image \( x_k \) depicting the same object as \( x_o \), then images \( x_i \) similar to \( x_i \) should also be considered similar to \( x_o \) even if \( d(x_i, x_o) \) of these images is high. So now all the values of \( d_{ik} \) can be considered as contextual information that can be embedded in \( d_{ij} \). The new refined distances will provide a better distance measure.

### 2.2. Distance Refining

Now a method has to be defined to include the determined contextual information within the distance measure, let’s define \( d'_{ij} \) as the new distance metric and can be determined by (1). Since both distances are based on the same metric, they can be added without scaling, the \( \alpha \) factor represents our confidence on whether image \( x_k \) depicts the same object as \( x_i \) or not.

\[
d'_{ij} = d_{ij} + \alpha d_{ik} \tag{1}
\]

To guarantee the symmetry of the new distance metric, we also used another image \( x_l \) from the same class as \( x_j \) to regularize the distance as

\[
d''_{ij} = d'_{ij} + \alpha (d_{ik} + d_{il}) \tag{2}
\]

Now we can generalize the equation given a set of images \( N_l \) similar to image \( i \):

\[
d'_{ij} = d_{ij} + \alpha \left[ \frac{1}{|N_l|} \sum_{k \in N_l} d_{ik} + \frac{1}{|N_j|} \sum_{l \in N_j} d_{il} \right] \tag{3}
\]

We call \( N_i \) as context of \( x_i \) here. Now the problem is how to find such an context \( N_i \)? We define the context of \( x_i \) as its \( N \) nearest neighbors from \( X \):

\[
N_i = N - \operatorname{argmin}_{x_k \in X} d_{ik} \tag{4}
\]

However, the context size \( N \) is very difficult to find. At one hand, if \( N \) is too large, some irrelevant images will be included within \( N_i \) since \( d_{ij} \) is not a credible distance metric itself, thus degrade the quality of \( d'_{ij} \). At the other hand, if \( N \) is too small, the effect of \( N_i \) will be limited. To solve this problem, instead of finding the context one time using \( d_{ij} \), we propose a new iterative approach to refine the distance metric \( d_{ij} \) while approximating the contexts \( N_i \) alternately.

### 2.3. Iterative Refining Algorithm

We propose an iterative approach for refining. Define a distance function \( \hat{d}_{ij} \), where \( t \) is the iteration number. In each iteration, the image excluded by \( N_i \) while minimizing \( \hat{d}_{ij} \) is added to the context \( N_i \) of \( x_i \), and then distances are updated according to (3). The algorithm is summarized in Algorithm 2.3.

Contextual Distance Refining (CDR) Algorithm.

Original distance matrix \( D = [d^0_{ij}]_{i, j=0}^{N} \); Balance parameter \( \alpha \); Iteration number \( T \);

\( t = 1, \ldots, T \)

Update the distance matrix

\[
d^t_{ij} = d^{t-1}_{ij} + \alpha \left[ \frac{1}{|N_i|} \sum_{k \in N^{t-1}_i} d^{t-1}_{ik} + \frac{1}{|N_j|} \sum_{l \in N^{t-1}_j} d^{t-1}_{il} \right] \tag{5}
\]

Find the new context item \( R_i^t \)

\[
R_i^t = \operatorname{argmin}_{x_k \in N^{t-1}_i, x_k \in X} d_{ik} \quad N_i^t = N_i^{t-1} \cup R_i^t \tag{6}
\]

Output distance matrix \( D^t = [d^t_{ij}] \).

An important feature of our algorithm is that the size of context is not fixed like CDM [11] or TG [12]. The context and distance metric are refined step by step coherently.

### 2.4. Discussion

It can be noticed from the algorithm that it can be applied on any pairwise distance measure to improve it. The algorithm will return a new set of distances that are improved based on the locality and context of a certain query. Then the top entries can be picked according to the ascending order of the elements.

One major improvement over the Graph Transduction method [12] is the method by which context is extrapolated from the distance matrix. In their paper they first they have to select a certain number of candidate neighbors, these are the top \( n \) images according to the original distance metric,
then their algorithm is applied on these n candidates. It can be shown that their algorithm is $O(TI^2)$, where $T$ is the number of iterations and $I$ is the number of images in the database. While the algorithm we propose is $O(T^2I)$, since $T$ is usually chosen a small fixed number (as we will show in the experiment section) and $T = I$, the algorithms end up being $O(T^2)$ and $O(I)$ respectively, which allows our method to work on the square number of initial candidates preserving the same performance. And, as will be shown in the Experimental Results Section below, there is no sacrifice in retrieval rates too.

3. EXPERIMENTAL RESULTS

In this section we examine the accuracy improvement of our algorithm applied on a pairwise distance metric, this is done against other context-aware algorithms. Then, we show experiments on the parameter values, with a brief description of what they affect.

3.1. MPEG7 Dataset

The MPEG7 Core Experiment CE-Shape-1 data set [18], the dataset contains 1400 images arranged as 70 classes 20 images each. To compare our results with the state of art, we use the IDSC distance [7], which is used as the base distance for Bai et al. [12]. The results are shown in the (Fig. 2) below.

![Fig. (2). Precision/recall values for: a) CDR: our proposed method b) GT: Graph transduction c) CDM: Contextual dissimilarity measure and d) IDSC: original distance metric.](image)

The compared distances are: the original distance (IDSC), the distance after applying Graph Transduction, the distance after applying CDM [11] and our distance CDR. CDR gave more improvement for most of the precision/recall values and a significant improvement over Graph Transduction overall. CDM did not perform well in this case because their basis assumption of CDM is that the distance between a data sample and its neighbors should be consistent. However, due to the variance of local distribution of shape database, this assumption will degrade the retrieval performance.

3.2. Choice of Parameters

Next we show results of changing the algorithm parameters $T$ and $\alpha$. (Fig. 3) shows the effect of changing $\alpha$ on the average accuracy of the algorithm (area below precision/recall curve). Note that a smaller $\alpha$ is used when the images of different classes are close in terms of original distance; a smaller $\alpha$ means that the contextual information derived from neighbors should not be trusted with high probability. A higher $\alpha$ is ideal for more clustered data, which results in even more clustering and dispersing non relevant images.

![Fig. (3). Effect of varying $\alpha$ on precision.](image)

Experiments on different values of $T$ showed similar results in (Fig. 4); at first, more improvement is gained by including more neighbors and adjusting the distances accordingly. But as we include more neighbors the performance starts to degrade because less relevant neighbors are included. $T$ is usually picked according to how many images we expect to belong to the same cluster.

![Fig. (4). Effect of varying $T$ on precision.](image)

4. CONCLUSION

Our method can be integrated with any pairwise distance metric to improve accuracy; it iteratively captures contextual information from a set of close images. The tests show a
Contextual Distance Refining for Image Retrieval

significant improvement of accuracy for image retrieval. The effect of the choice of parameters is investigated. Our future work will focus more on automating and dynamically changing the parameters, depending on the Image Distance Matrix to improve accuracy and enhance automation, which is important in unsupervised search.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by Chongqing Key Laboratory of Computational Intelligence (Grant No. CQ-LCI-2013-02).

REFERENCES

[1] J. van Gemert, C. Veenman, A. Smeulders, and J.-M. Geusebroek, “Visual word ambiguity”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 7, 2010, pp. 1271-1283.

[2] J. Wang, X. Gao, Q. Wang, and Y. Li, “Prodis-contrs: learning protein dissimilarity measures and hierarchical context coherently for protein-protein comparison in protein database retrieval”, BMC Bioinformatics, vol. 13, 2012, p. 7.

[3] F. A. González, J. C. Caicedo, O. Nasraoui, and J. Ben-Abdallah, “Nmf-based multimodal image indexing for querying by visual example”, In: Proceedings of the ACM International Conference on Image and Video Retrieval, CIVR ’10. ACM, New York, NY, USA, 2010, pp. 366-373. doi:10.1145/1816041.1816094.

[4] J. J.-Y. Wang, H. Bensmail, and X. Gao, “Joint learning and weighting of visual vocabulary for bag-of-feature based tissue classification”, Pattern Recogn., vol. 46, no. 12, 2013, pp. 3249-3255.

[5] J. J.-Y. Wang, H. Bensmail, and X. Gao, “Multiple graph regularized nonnegative matrix factorization”, Pattern Recogn., vol. 46, no. 10, 2013, pp. 2840-2847.

[6] J. J.-Y. Wang, X. Wang, and X. Gao, “Non-negative matrix factorization by maximizing correntropy for cancer clustering”, BMC Bioinformatics, vol. 14, 2013, p. 107.

[7] L. Latecki, R. Lakamper, and T. Eckhardt, “Shape descriptors for non-rigid shapes with a single closed contour”, In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000. vol. 1, 2000, pp. 424-429.

[8] A. Mans, M. Bennamoun, and R. Owens, “A novel representation and feature matching algorithm for automatic pairwise registration of range images”, Int. J. Comput. Vis., vol. 66 no. 1, 2006 pp. 19-40.

[9] M. M. B. Contshc: Learning context sensitive shape similarity by graph transduction”, Lect. Notes Comput. Sci. (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5164 LNCS PART 2, 2008, pp. 81-90.

[10] URL http://doi.acm.org/10.1145/1816041.1816094

[11] Y.-S. Hwang, H. Park, and Y. Chung, “Pairwise matching of spots in 2-de images using hopfield network”, Lect. Notes Comput. Sci., 2005, pp. 264-271.

[12] X. Bai, X. Yang, L. Latecki, W. Liu, and Z. Tu, “Learning context-sensitive shape similarity by graph transduction”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32 no. 5, 2010, pp. 861-874.

[13] K. Srijan, and C. Jawahar, “Towards exhaustive pairwise matching in large image collections”, Lect. Notes Comput. Sci. (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7583 LNCS (PART 1) (2012) pp. 219-228.

[14] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, and N. Tang, “Learning context-sensitive similarity by shortest path propagation”, Pattern Recogn. vol. 44, no. 10-11, 2011, pp. 2367-2374.

[15] J. J.-Y. Wang, H. Bensmail, and X. Gao, “Multiple graph regularized protein domain ranking”, BMC Bioinformatics vol. 13, 2012, p. 1.

[16] H. Ling, and D. Jacobs, “Shape classification using the inner-distance”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 2, 2007, pp. 286-299.

[17] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek, “Accurate image search using the contextual dissimilarity measure”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 1, 2010 pp. 2-11.

[18] S. Kropf, H. Heuer, M. Grünig and K. Smalla, “Significance test for comparing complex microbial community fingerprints using pairwise similarity measures”, J. Microbiol. Meth., vol. 57, no. 2, 2004, pp. 187-195.