

Semi-Automatic Photo Clustering with Distance Metric Learning

Dinghuang Ji^a, Meng Wang^b, Qi Tian^c, Xian-Sheng Hua^b

^a Institute of Computing Technology, Beijing 100190, P. R. China

^b Microsoft Research Asia Beijing 100080, P. R. China

^c University of Texas at San Antonio, USA.

ABSTRACT

Photo clustering has been widely explored in many applications such as album management. But automatic clustering can hardly achieve satisfying performance due to the large variety of photos' content. This paper proposes a semi-automatic photo clustering scheme that attempts to improve clustering performance with users' interactions. Users can adjust the results of automatic clustering, and a set of constraints among photos are generated accordingly. A distance metric is then learned with these constraints and we can re-implement clustering with this metric. We conduct experiments on different photo albums, and experimental results have demonstrated that our approach is able to improve automatic photo clustering results, and it is better than pure manual adjustment approach by exploring distance metric learning.

Keywords: photo clustering, distance metric learning

1. INTRODUCTION

With the popularity of digital cameras, recent years have witnessed a rapid growth of personal photos. People capture photos to record their lives and share them on the web. Clustering is an effective approach to helping users manage, browse and annotate photos by grouping them into visually and semantically consistent clusters. A typical application is batch annotation. Given an image set, manually annotating each image will be a labor-intensive process and there is also a waste of efforts as many images are close to some others, and meanwhile learning-based automatic annotation technology is still not mature enough despite there are extensive research efforts [21][22][23]. Batch annotation is an effective approach to reducing the cost by directly assigning a set of tags to a batch of images. Therefore, if the images in a set can be effectively clustered, then users can easily adopt batch annotation and tags only need to be assigned once for each cluster.

Extensive efforts have been dedicated to automatic image clustering [1][2][3][4]. However, although encouraging performance has been shown, automatic photo clustering can hardly achieve satisfying results in many cases due to the large variety of photos' content. For example, the effective features and distance metric for photo clustering may vary across different albums. Without supervision, we can hardly verify the effectiveness of different features and distance metrics in the clustering process.

In this paper, we introduce a novel photo clustering approach. Instead of automatic clustering, users can iteratively manually adjust clustering results. Then a set of equivalence and inequivalence constraints are generated based on the adjustments, and distance metric learning is performed to construct a better distance measure for photo pairs. Consequently, we implement clustering with the updated metric. This process repeats until satisfactory clustering results are obtained. Figure 1 illustrates the main scheme of our approach. We will show our approach can obtain better results than automatic clustering or pure manual adjustments of clustering results.

The organization of the rest of this paper is as follows. In Section 2, we provide a review on the related work. In Section 3, we introduce our semi-automatic photo clustering approach. Experiments are presented in Section 4. Finally, we conclude the paper in Section 5.

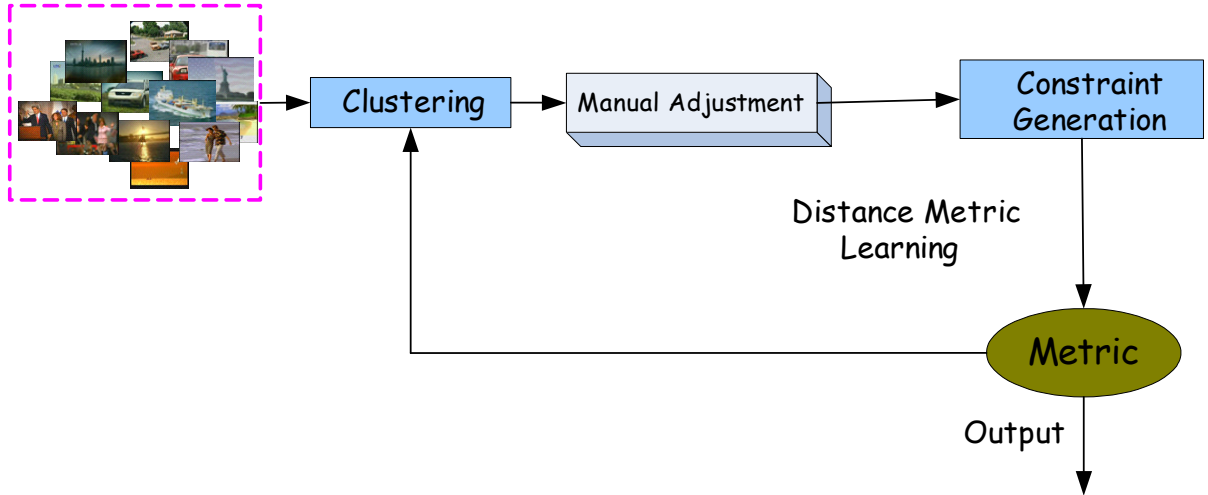


Figure 1. An schematic illustration of semi-automatic image clustering.

2. RELATED WORK

There is extensive research on image clustering. Some of them explore textual information that is associated with the images, such as surrounding text or tags [1][2][3]. Goldberger et al. [4] proposed an image clustering approach based on information theoretic scheme. Kennedy and Naaman [5] and Simon et al. [6] proposed two clustering methods to extract representative images from an image set. In [7], an AutoAlbum scheme is proposed, which adopts a photo clustering approach with the help of temporal information and the order of photo creation. John et al. [8] then further proposed PhotoTOC [8], which supplies a new interface to assist users efficiently get the photos they want.

Distance metric learning is intended to construct an optimal distance metric for the given learning task based on the pairwise relationships among samples. A number of algorithms have been proposed for distance metric learning. Bar-Hillel proposed Relevant Components Analysis (RCA) method to learn a linear transformation from the equivalence constraints, which can be used directly to compute the distance between two examples [9]. Xing et al. [10] formulated distance metric learning as a constrained convex programming problem by minimizing the distance between the data points in the same classes under the constraint that the data points from different classes are well separated. Neighborhood Component Analysis (NCA) [11] learned a distance metric by extending the nearest neighbor classifier. Weinberger et al. [12] proposed the maximum-margin nearest neighbors (LMNN) method that extends NCA through a maximum margin framework. Alipanahi et al. [13] show a strong relationship between distance metric learning methods and Fisher Discriminant Analysis (FDA). Hoi et al. [14] proposed a semi-supervised distance metric learning method that integrates both labeled and unlabeled examples. In our work, we apply a distance metric learning algorithm based on the constraints among photos generated from users' manual adjustments. We will show its effectiveness despite that the generated constraints are noisy.

3. SEMI-AUTOMATIC PHOTO CLUSTERING

As introduced in Section 1, our semi-automatic photo clustering works as follows. First, photos are grouped into a certain number of clusters (the number can be specified by users). Then users can view the clustering results and make manual adjustments. Based on the adjustments, equivalence and inequivalence constraints among photos are generated and a distance metric learning algorithm is performed. Consequently, we perform clustering again with the learned distance metric and the process can repeat until a satisfactory clustering performance is achieved. Table 1 illustrates the implementation process.

Input:Photos $\mathcal{F} = x_1, x_2, \dots, x_n$;The number of clusters m ; //the number can be specified by users**Output:**Clusters C_1, C_2, \dots, C_m ;

1. Initialize distance metric \mathbf{M} to be Euclidean distance;
2. Cluster \mathcal{F} into m clusters with the distance metric \mathbf{M} ;
3. Manually adjust clustering results;
4. Generate constraints based on the manual adjustments;
5. Perform distance metric learning with the constraints to obtain an updated \mathbf{M} ;
6. Go to 1, unless user is satisfied with the clustering performance.

Table 1. The implementation process of the semi-automatic photo clustering approach

3.1 Photo Clustering with a Distance Metric

We adopt spectral clustering approach in our scheme. It is a technique that explores the eigenstructure of a similarity matrix to partition samples into disjoint clusters with samples in the same cluster having high similarity and points in different clusters having low similarity [15][16][17]. The spectral clustering can be viewed as a graph partition task. The simplest and most straightforward way to construct a partition of the similarity graph is to solve the Min-cut problem, and we can also use Normalized-cut or Ratio-cut to replace Min-cut [18][19] in order to obtain more stable clusters. Existing studies have shown that spectral clustering outperforms many conventional clustering algorithms such as k-means algorithm. Here we adopt the method proposed in [18]. In our scheme we have learned a distance metric from users' manipulation of clustering results, thus we adopt the distance metric in the computation of the similarity matrix, i.e.,

$$W_{ij} = \exp \left(-\frac{(x_i - x_j)^T \mathbf{M} (x_i - x_j)}{\sigma^2} \right) \quad (1)$$

The clustering process is illustrated in Table 2.

Input:Photos $\mathcal{F} = x_1, x_2, \dots, x_n$;The number of clusters m ; //the number can be specified by users;Distance metric \mathbf{M} ;**Output:**Clusters C_1, C_2, \dots, C_m ;

1. Construct a similarity graph based on Eq. (1);
2. Compute the unnormalized Laplacian \mathbf{L} ;
3. Compute the first m generalized eigenvectors u_1, \dots, u_m of the generalized eigenproblem $\mathbf{L}u = \lambda \mathbf{D}u$;
4. Let $\mathbf{U} \in \mathcal{L}^{n \times m}$ be the matrix containing the vectors u_1, \dots, u_m as columns.
5. For $i = 1, \dots, n$, let $y_i \in \mathcal{R}^m$ be the vector corresponding to the i -th row of \mathbf{U} ;
6. Cluster the points $(y_i)_{(i=1, \dots, n)}$ in \mathcal{R}^k with the k-means algorithm into clusters C_1, \dots, C_m .

Table 2. The implementation process of the spectral clustering algorithm with distance metric \mathbf{M}

3.2 Manual Adjustment and Constraint Generation

Clustering results can be manipulated in different ways, such as moving one photo from one cluster to another, merging two clustering or splitting one cluster into multiple clusters. Here we mainly consider the moving manipulation and leave the exploration of other manipulations to our future work (see Section 5). We denote the manipulation as $Move(x, C_i, C_j)$, which means removing a sample x from C_i and adding it to C_j , as shown in Figure 2. For this manipulation, we can assume that x forms dissimilar constraint with each remained samples in C_i as it has been moved out from C_i . On the contrary, we can generate a similar constraint for x and each sample in C_j . Therefore, if the sizes of C_i and C_j are u and v respectively, we obtain $u - 1$ dissimilar and v similar constraints

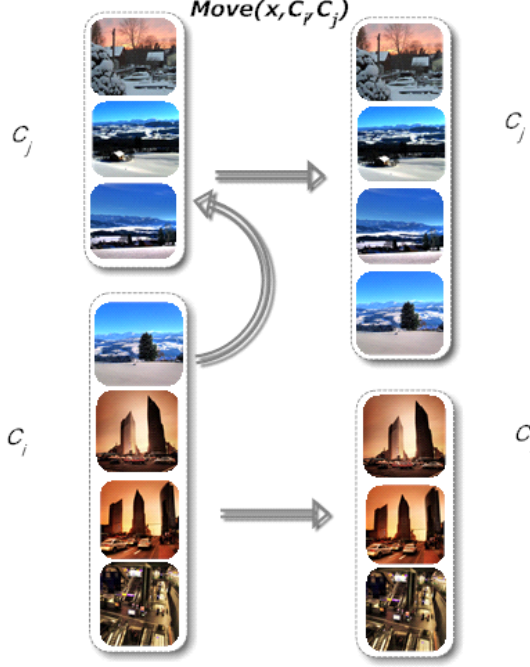


Figure 2. An example of the manipulation of clustering results.

3.3 Distance Metric Learning

We adopt the Global Distance Metric Learning by Convex Programming (GDMLCP) method proposed by Xing et al. [10] to learn the distance metric based on the constraints generated from users' manipulations. Denote by \mathcal{S} and \mathcal{D} the sets of equivalence constraints and inequivalence constraints. Then the distance metric \mathbf{M} is learned with the following formulation

$$\begin{aligned} \text{minimize}_{\mathbf{M}} \quad & g(\mathbf{M}) = \sum_{(x_i, x_j) \in \mathcal{D}} \|x_i - x_j\|_{\mathbf{M}}, \\ \text{s.t.} \quad & \sum_{(x_i, x_j) \in \mathcal{S}} \|x_i - x_j\|_{\mathbf{M}}^2 \geq 1, \quad \mathbf{M} \succeq 0 \end{aligned} \quad (2)$$

The above optimization problem can be solved with an iterative process, and in each iteration we first implement a gradient ascent step on $g(\mathbf{M})$, and then apply an iterative projection method to ensure that the constraints hold.

4. EXPERIMENT

4.1 Experimental Settings

We conduct experiments with 6 different personal albums that are collected from Flickr [20]. These photos are captured at different locations around the world and contain diverse content, including the records of cityscape, landscape, wide life, etc. Table 1 illustrates the number of photos in these albums.

Many of the photos are of high resolution. To speed up feature extraction, we resize each photo such that its width is 240 pixels and then we extract the following features: (1) 64-dimensional color histogram; (2) 75-dimensional edge histogram; (3) 225-dimensional block-wise color moment features generated from 5-by-5 partition of the image; (4) 128-dimensional wavelet texture features.

4.2 Experiment Result

We compare the following three methods:

Album	Photo Num
Germany	163
HongKong	136
London	204
LongExposure	185
Manasquan	333
Pairs	265

Table 3. The numbers of photos of the albums used in our experiments

1. Automatic clustering, i.e., we perform spectral clustering with Euclidean distance and do not adjust clustering results;
2. Semi-automatic clustering without distance metric learning, i.e., we perform spectral clustering with Euclidean distance and then adopt pure manual adjustments on the clustering results;
3. Semi-automatic clustering with distance metric learning, i.e., we learn a new distance metric after users adjust clustering results and then re-implement spectral clustering.

In our experiments, the number of clusters is specified by the user. In the semi-automatic methods, in each iteration we allow the user to make 10 manual adjusts (i.e., implement 10 *Move* manipulations). For each album, we evaluate the result using the F-score, which is defined as $2 \times (precision \times recall) / (precision + recall)$. It is worth mentioning that we need to obtain the ground truth of clustering for object evaluation. In our experiments, we ask a user to manually adjust clustering results until he is satisfied, and then we regard the obtained results as our ground truths. We average the F-scores of all clusters as our performance evaluation metric in our experiments. The parameter σ in spectral clustering is set to the median value of the pairwise distances of all photos.

Figure 2 illustrates the performance comparison. We can see that semi-automatic methods can effectively improve clustering performance in comparison with automatic clustering. By adopting distance metric learning, the clustering performance can be significantly improved in comparison with pure manual adjustments. In our experiments, it takes an average of 5.6s, 3.7s, 6.2s, 6.8s, 12.1s and 9.7s respectively in each iteration for the albums Germany, HongKong, London, LongExposure, Manasquan and Pairs, respectively. All these time costs are recorded on a PC with Pentium 3.40G CPU and 2G memory.

5. CONCLUSION

This paper introduces a semi-automatic photo clustering approach. Different from automatic method that directly applies a clustering algorithm, our approach allows humans to manually adjust clustering results, and then a distance metric is learned accordingly. With this new metric, clustering is re-implemented to obtain better results, and this process can repeat until satisfying performance is obtained.

Clustering images with user interaction is a novel scheme and we will continue research this direction. There are several future works to further improve our approach:

1. Online distance metric learning. In our current approach, we learn a distance metric based on the accumulated constraints from users' manipulations, but this method will introduce intensive computation when the number of constraints becomes great. An online distance metric learning method that directly updates previously learned metric with newly generated constraints can significantly reduce the computational cost.
2. Constraint refinement. The constraints generated from users' manipulations may contain significant noises due to the impurity of clusters. For example, when we perform the *Move*(x, C_i, C_j) manipulation, C_j may not be a satisfactory cluster, i.e., users do not consider that all the samples in C_j should be grouped into a cluster and they will further manipulate it. Therefore, a constraint refinement that is able to remove noises can help learn better distance metric and consequently improve clustering performance.

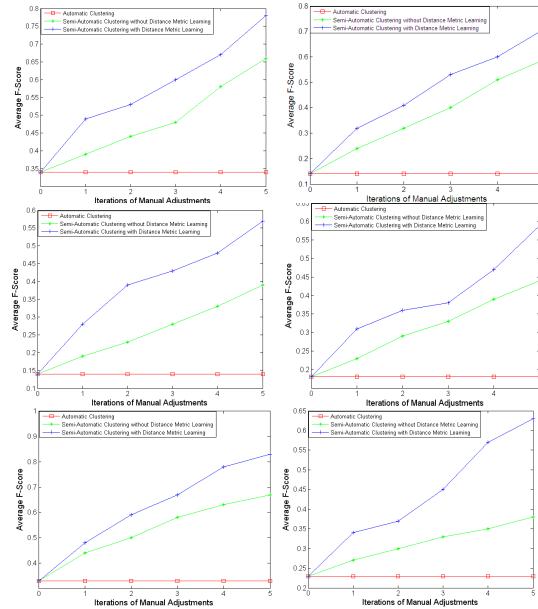


Figure 3. Clustering performance comparison of different methods.

3. More manual adjustment patterns. In this work we just consider *Move* manipulation. In our future work we will introduce more manipulations such that users can adjust clustering results in a more efficient way, such as Merge and Split, which merge two clusters into one or split a cluster into two subsets respectively.

REFERENCES

1. P. A. Moellic, J. E. Haugeard, and G. Pitel, *Image clustering based on a shared Nearest Neighbors Approach for Tagged Collections*, ACM International Conference on Image and Video Retrieval, 2008.
2. D. Cai, X. He, Z. Li, W. Y. Ma, and J. R. Wen, *Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information*, ACM International Conference on Multimedia, 2004.
3. S. Wang, F. Jing, J. He, Q. Du, and L. Zhang, *IGroup: Presenting Web Image Search Results in Semantic Clusters*, ACM International Conference on Human Factors in Computing Systems, 2007.
4. J. Goldberger, S. Gordon, and H. Greenspan, *Unsupervised image-set clustering using information theoretic Framework*, IEEE Transactions on Image Processing, 15(2):449-458, 2006.
5. L. Kennedy and M. Naaman, *Generating Diverse and Representative Image search Results for Landmarks*, International World Wide Web Conference, 2008.
6. I. Simon, N. Snavely, S. M. Seitz, *Scene Summarization for Online Image Collections*, International Conference on Computer Vision, 2007.
7. J.C. Platt, *Proc. AutoAlbum: Clustering digital photographs using probabilistic model merging*. In Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, pages 96-100, 2000.
8. J.C. Platt, M. Czerwinski, B. Field, *PhotoTOC: Automatic Clustering for Browsing Personal Photographs*. Fourth IEEE Pacific Rim Conference on Multimedia (2003)
9. A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. *Learning a Mahalanobis metric from equivalence constraints*. Journal of Machine Learning Research (JMLR), 6,2005.
10. E. P. Xing, A. Y. Ng, and S. Russell, *Distance metric learning with application to clustering with side information*, in Proceedings of NIPS, 2003.
11. J. Goldberger, S. T. Roweis, G. E. Hinton, and R. Salakhutdinov. *Neighbourhood components analysis*. In Advances in Neural Information Processing Systems (NIPS), 2004.
12. K. Weinberger, J. Blitzer, and L. Saul. *Distance metric learning for large margin nearest neighbor classification*. In Advances in Neural Information Processing Systems (NIPS). 2006.

13. B. Alipanahi, M. Biggs, and A. Ghodsi. *Distance metric learning versus fisher discriminant analysis*. In Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI), 2008.
14. S. C.H. Hoi, W. Liu, and S.-F. Chang. *Semisupervised distance metric learning for collaborative image retrieval*. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
15. U. von Luxburg. *A tutorial on spectral clustering*. In Statistics and Computing, vol.17, no.4, pp.395-416, December 2007.
16. Bach, F. and Jordan, M. *Learning spectral clustering*. Advances in Neural Information Processing Systems 16 (NIPS) (pp. 305 - 312). Cambridge, MA: MIT Press.
17. Ding, C. *A tutorial on spectral clustering*. Talk presented at ICML.
18. J. Shi and J. Malik. *Normalized cuts and image segmentation*. IEEE Trans. PAMI, 22(8):888-905, 2000.
19. P. K. Chan, M. D. F. Schlag, and J. Y. Zien. *Spectral K-way ratio-cut partitioning and clustering*. IEEE Trans. CAD, 13(9):1088-1096, 1994.
20. Flickr. <http://www.flickr.com>.
21. M. Wang, X. S. Hua, R. Hong, J. Tang, G. -J. Qi, and Y. Song, *Unified Video Annotation Via Multi-Graph Learning*, IEEE Transactions on Circuits and Systems for Video Technology, 19(5), 2009.
22. M. Wang, X. S. Hua, J. Tang, and R. Hong, *Beyond Distance Measurement: Constructing Neighborhood Similarity for Video Annotation*, IEEE Transactions on Multimedia, 11(3), 2009.
23. M. Wang, X. S. Hua, Y. Song, J. Tang, and L. Dai, *Multi-Concept Multi-Modality Active Learning for Interactive Video Annotation*, IEEE International Conference on Semantic Computing, 2007.