ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

A Novel Approach for Automatic Face Reorganization

Bhargav.K.M[1], Faraz Ahmed.K[2], Manish Jain[3], Neha Kumari[4]

8th Semester, Dept. of ISE, SJB Institute of Technology, Bengaluru, India

ABSTRACT: Each image contains several faces and is associated with a few names in the corresponding caption; the goal of face naming is to infer the correct name for each face. We first propose a new method called regularized low-rank representation by effectively utilizing weakly supervised information to learn a low-rank reconstruction coefficient matrix while exploring multiple subspace structures of the data. Specifically, by introducing a specially designed regularizer to the low-rank representation method. With the inferred reconstruction coefficient matrix, a discriminative affinity matrix can be obtained. Moreover, we also develop a new distance metric learning method called ambiguously supervised structural metric learning by using weakly supervised information to seek a discriminative distance metric. Hence, another discriminative affinity matrix can be obtained using the similarity matrix (i.e., the kernel matrix) based on the Mahalanobis distances of the data. Observing that these two affinity matrices contain complementary information, we further combine them to obtain a fused affinity matrix, based on which we develop a new iterative scheme to infer the name of each face.

KEY TERMS: Matrix, Low-Rank, Automatic face Naming

I. INTRODUCTION

To tag faces in news photos, Berg *et al.* proposed to cluster the faces in the news images. Ozkan and Duygulu developed a graph-based method by constructing the similarity graph of faces and finding the densest component. Guillaumin *et al.* proposed the multiple-instance logistic discriminant metric learning (MildML) method. Luo and Orabona proposed a structural support vector machine (SVM)-like algorithm called maximum margin set (MMS) to solve the face naming problem. Recently, Zeng *et al.* proposed the low-rank SVM (LR-SVM) approach to deal with this problem, based on the assumption that the feature matrix formed by faces from the same subject is low rank. In the following, we compare our proposed approaches with several related existing methods. Our rLRR method is related to LRR and LR-SVM. LRR is an unsupervised approach for exploring multiple subspace structures of data. In contrast to LRR, our rLRR utilizes the weak supervision from image captions and also considers the image-level constraints when solving the weakly supervised face naming problem. Moreover, our rLRR differs from LR-SVM [9] in the following two aspects. 1) To utilize the weak supervision, LR-SVM considers weak supervision information in the partial permutation matrices, while rLRR uses our proposed regularizer to penalize the corresponding reconstruction coefficients. 2) LR-SVM is based on robust principal component analysis (RPCA) . Similarly to , LR-SVM does not reconstruct the data by using itself as the dictionary. In contrast, our rLRR is related to the reconstruction based approach LRR. Moreover, our ASML is related to the traditional metric learning works, such as large-margin nearest neighbors(LMNN) , Frobmetric , and metric learning to rank (MLR). LMNN and Frobmetric are based on accurate supervision without ambiguity (i.e., the triplets of training samples are explicitly given), and they both use the hinge loss in their formulation. In contrast, our ASML is based on the ambiguous supervision, and we use a max margin loss to handle the ambiguity of the structural output, by enforcing the distance based on the best label assignment matrix in the feasible label set to be larger than the distance based on the best label assignment matrix in the infeasible label set by a margin. Although a similar loss that deals with structural output is also used in MLR, it is used to model the ranking orders of training samples, and there is no uncertainty regarding supervision information in MLR (i.e., the groundtruth ordering for each query is given).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

Figure 1: Illustration of face naming task in which we aim to interfere which name matches which face

II.APPROACH

In this paper, we propose a new scheme for automatic face naming with caption-based supervision. Specifically, we develop two methods to respectively obtain two discriminative affinity matrices by learning from weakly labeled images. The two affinity matrices are further fused to generate one fused affinity matrix, based on which an iterative scheme is developed for automatic face naming. To effectively infer the correspondences between the faces based on visual features and the names in the candidate name sets, we exploit the *subspace structures among faces* based on the following assumption: the faces from the same subject/name lie in the same subspace and the subspaces are linearly independent. Liu *et al.* showed that such subspace structures can be effectively recovered using LRR, when the subspaces are independent and the data sampling rate is sufficient. They also showed that the mined subspace information is encoded in the reconstruction coefficient matrix that is block-diagonal in the ideal case. As an intuitive motivation, we implement LRR on a synthetic dataset.This near block-diagonal matrix validates our assumption on the *subspace structures among faces*. Specifically, the reconstruction coefficients between one face and faces from the same subject are generally larger than others, indicating that the faces from the same subject tend to lie in the same subspace. However, due to the significant variances of inthe- wild faces in poses, illuminations, and expressions, the appearances of faces from different subjects may be even more similar when compared with those from the same subject. Our main contributions are summarized as follows.

1) Based on the caption-based weak supervision, we propose a new method rLRR by introducing a new regularizer into LRR and we can calculate the first affinity matrix using the resultant reconstruction coefficient matrix.

2) We also propose a new distance metric learning approach ASML to learn a discriminative distance metric by effectively coping with the ambiguous labels of faces. The similarity matrix (i.e., the kernel matrix) based on the Mahalanobis distances between all faces is used as the second affinity matrix.

3) With the fused affinity matrix by combining the two affinity matrices from rLRR and ASML, we propose an efficient scheme to infer the names of faces.

4) Comprehensive experiments are conducted on one synthetic dataset and two real-world datasets, and the results demonstrate the effectiveness of our approaches.

III. RESULTS

Affinity Matrix

Since the principles of proximity and smooth-continuation arise from local properties of the configuration of the edges, we can model them using only local information. Both of these local properties are modeled by the distribution of smooth curves that pass through two given edges. The distribution of curves is modeled by a smooth, stochastic motion of a particle. Given two edges, we determine the probability that a particle starts with the position and direction of the first edge and ends with the position and direction of the second edge. The *affinity* from the first to the second edge is

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

the sum of the probabilities of all paths that a particle can take between the two edges. The change in direction of the particle over time is normally distributed with zero mean. Smaller the variance of the distribution, the smoother are the more probable curves that pass between two edges. Thus the variance of the normal distribution models the principle of smooth-continuation. In addition each particle has a non-zero probability for decaying at any time. Hence, edges that are farther apart are likely to have fewer curves that pass through both of them. Thus the decay of the particles models the principle of proximity. The affinities between all pairs of edges form the *affinity matrix* .

Learning discriminative affinity matrices For automatic face naming

In this section, we propose a new approach for automatic face naming with caption-based supervision. We formally introduce the problem and definitions, followed by the introduction of our proposed approach. Specifically, we learn two discriminative affinity matrices by effectively utilizing the ambiguous labels, and perform face naming based on the fused affinity matrix. We introduce our proposed approaches rLRR and ASML for obtaining the two affinity matrices respectively. In the remainder of this paper, we use lowercase/uppercase letters in boldface to denote a vector/matrix (e.g., **a** denotes a vector and **A** denotes a matrix). The corresponding nonbold letter with a subscript denotes the entry in a vector/matrix (e.g., *ai* denotes the *i* th entry of the vector **a**, and *Ai, j* denotes an entry at the *i* th row and *j* th column of the matrix **A**). The superscript _ denotes the transpose of a vector or a matrix. We define **I***n* as the *n* ×*n* identity matrix, and **0***n*, **1***n* ∈ R*n* as the *n*×1 column vectors of all zeros and all ones, respectively. For simplicity, we also use **I**, **0** and **1** instead of **I***n*, **0***n*, and **1***n* when the dimensionality is obvious. Moreover, we use **A** ◦ **B** (resp., **a** ◦ **b**) to denote the element-wise product between two matrices **A** and **B** (resp., two vectors **a** and **b**). tr*(***A***)* denotes the trace of **A** (i.e., tr*(***A***)* = _*i Ai,i*), and _**A**,**B**_ denotes the inner product of two matrices (i.e., _**A**,**B**_ = tr*(***A_B***)*). The inequality **a** ≤ **b** means that *ai* ≤ *bi* ∀*i* = 1, . . . , *n* and **A** 0 means that **A** is a positive semidefinite (PSD) matrix. _**A**_*F* = *(_i, j A*2*i , j)*1*/*2 denotes the Frobenious norm of a matrix **A**. _**A**_∞ denotes the largest absolute value of all elements in **A**.

Learning Discriminative Affinity Matrix With Regularized Low-Rank Representation (rLRR)

We first give a brief review of LRR, and then present the proposed method that introduces a discriminative regularizer into the objective of LRR. *1) Brief Review of LRR:* LRR [2] was originally proposed to solve the *subspace clustering* problem, which aims to explore the subspace structure in the given data **X** = [**x**1, . . . , **x***n*] ∈ R*d*×*n*. Based on the assumption that the subspaces are linearly independent, LRR [2] seeks a reconstruction matrix **W** = [**w**1, . . . ,**w***n*] ∈ R*n*×*n*, where each **w***i* denotes the representation of **x***i* using **X** (i.e., the data matrix itself) as the dictionary. Since **X** is used as the dictionary to reconstruct itself, the optimal solution **W**∗ of LRR encodes the pairwise affinities between the data samples. As discussed in [2, Th. 3.1], in the noise-free case, **W**∗ should be ideally block diagonal, where *W*∗ *i, j* _= 0 if the *i* th sample and the *j* th sample are in the same subspace.

Learning Discriminative Affinity Matrix by Ambiguously Supervised Structural Metric Learning (ASML)

Besides obtaining the affinity matrix from the coefficient matrix **W**∗ from rLRR (or LRR), we believe the similarity matrix (i.e., the kernel matrix) among the faces is also an appropriate choice for the affinity matrix. Instead of straightforwardly using the Euclidean distances, we seek a discriminative Mahalanobis distance metric **M** so that Mahalanobis distances can be calculated based on the learnt metric, and the similarity matrix can be obtained based on the Mahalanobis distances. In the following, we first briefly review the LMNN method, which deals with fully-supervised problems with the groung-truth labels of samples provided, and then introduce our proposed ASML method that extends LMNN for face naming from weakly labeled images.

Inferring names of faces

With the coefficient matrix **W**∗ learned from rLRR, we can calculate the first affinity matrix as **A***W* = 12 *(***W**1+W2) and normalize **A***W* to the range [0, 1]. Furthermore, with the learnt distance metric **M** from ASML, we can calculate the second affinity matrix as **A***K* = **K**, where **K** is a kernel matrix based on the Mahalanobis distances between the faces. Since the two affinity matrices explore weak supervision information in different ways, they contain complementary information and both of them are beneficial for face naming. For better face naming performance, we combine these two affinity matrices and perform face naming based on the fused affinity matrix. Specifically, we obtain a fused affinity matrix **A** as the linear combination of the two affinity matrices, i.e., **A** = *(*1 − *α)***A***W* + *α***A***K*, where *α* is a parameter in the range [0, 1]. Finally, we perform face naming based on **A**. Since the fused affinity matrix is obtained based on rLRR and ASML, we name our proposed method as rLRRml.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

IV. CONCLUSION

In this paper, we have proposed a new scheme for face naming with caption-based supervision, in which one image that may contain multiple faces is associated with a caption specifying only who is in the image. To effectively utilize the caption-based weak supervision, we propose an LRR based method, called rLRR by introducing a new regularizer to utilize such weak supervision information. We also develop a new distance metric learning method ASML using weak supervision information to seek a discriminant Mahalanobis distance metric. Two affinity matrices can be obtained from rLRR and ASML, respectively. Moreover, we further fuse the two affinity matrices and additionally propose an iterative scheme for face naming based on the fused affinity matrix. The experiments conducted on a synthetic dataset clearly demonstrate the effectiveness of the new regularizer in rLRR. In the experiments on two challenging real-world datasets (i.e., the Soccer player dataset and the Labeled Yahoo! News dataset), our rLRR outperforms LRR, and our ASML is better than the existing distance metric learning method MildML. Moreover, our proposed rLRRml outperforms rLRR and ASML, as well as several state-of-the-art baseline algorithms.

V. FUTURE ENHANCEMENT

To further improve the face naming performances, we plan to extend our rLRR in the future by additionally incorporating the _1-norm-based regularizer and using other losses when designing new regularizers. We will also study how to automatically determine the optimal parameters for our methods in the future.

VI. ACKNOWLEDGEMENT

I would thank to Mrs. Usha N, Lecturer, Dept of ISE and staff of Information Science and Engineering for their support.

REFERENCES

[1] P. Viola and M. J. Jones, “Robust real-time face detection,” *Int. J. Comput. Vis*., vol. 57, no. 2, pp. 137–154, 2004.

[2] G. Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 663–670.

[3] T. L. Berg *et al.*, “Names and faces in the news,” in *Proc. 17th IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, USA, Jun./Jul. 2004, pp. II-848–II-854.

[4] D. Ozkan and P. Duygulu, “A graph based approach for naming faces in news photos,” in *Proc. 19th IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 1477–1482.

[5] P. T. Pham, M. Moens, and T. Tuytelaars, “Cross-media alignment of names and faces,” *IEEE Trans. Multimedia*, vol. 12, no. 1, pp. 13–27, Jan. 2010.

[6] M. Guillaumin, J. Verbeek, and C. Schmid, “Multiple instance metric learning from automatically labeled bags of faces,” in *Proc. 11th Eur. Conf. Comput. Vis.*, Heraklion, Crete, Sep. 2010, pp. 634–647.

[7] J. Luo and F. Orabona, “Learning from candidate labeling sets,” in *Proc. 23rd Annu. Conf. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2010, pp. 1504–1512.

[8] X. Zhang, L. Zhang, X.-J. Wang, and H.-Y. Shum, “Finding celebrities in billions of web images,” *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 995–1007, Aug. 2012.

[9] Z. Zeng *et al.*, “Learning by associating ambiguously labeled images,” in *Proc. 26th IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 708–715.