Automatic Face Naming by Learning Discriminative Affinity Matrices from Weakly Labeled Images

Ajay kumar[1], Ritika Mehra[2], Er.Sandeep Garg[3]
Dept of CSE
R. P. Inderprastha Institute of Technology, India

Abstract- Given a collection of images, where 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. In this paper, we propose two new methods to effectively solve this problem by learning two discriminative affinity matrices from these weakly labeled images. 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, we penalize the corresponding reconstruction coefficients related to the situations where a face is reconstructed by using face images from other subjects or by using itself. 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. Comprehensive experiments demonstrate the effectiveness of our approach.

Keywords- Affinity matrix, caption-based face naming, distance metric learning, low-rank representation (LRR).

I. INTRODUCTION

In Social networking websites (e.g., Facebook), photo sharing websites (e.g., Flickr) and news websites(e.g., BBC), an image that contains multiple faces can be associated with a caption specifying who is in the picture. For instance, multiple faces may appear in a news photo with a caption that briefly describes the news. Moreover, in TV serials, movies, and news videos, the faces may also appear in a video clip with scripts. In the literature, a few methods were developed for the face naming problem (see Section II for more details).

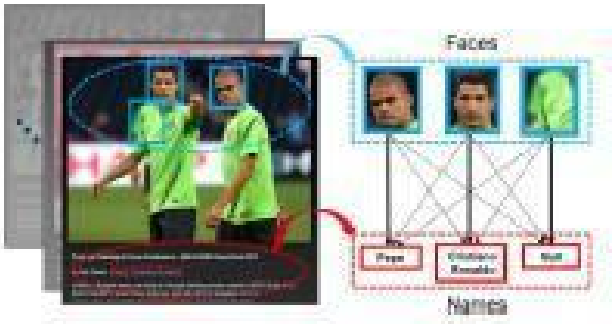


Fig. 1. Illustration of the face-naming task, in which we aim to infer which name matches which face, based on the images and the corresponding captions. The solid arrows between faces and names indicate the ground truth face-name pairs and the dashed ones represent the incorrect face-name pairs, where null means the ground truth name of a face does not appear in the candidate name set.

In this paper, we focus on automatically annotating faces in images based on the ambiguous supervision from the associated captions. Fig. 1 gives an illustration of the face naming problem. Some preprocessing steps need to be conducted before performing face naming. Specifically, faces in the images are automatically detected using face detectors [1], and names in the captions are automatically extracted using a name entity detector. Here, the list of names appearing in a caption is denoted as the candidate name set. Even after successfully performing these preprocessing steps, automatic face naming is still a challenging task. The faces from the same subject may have different appearances because of the variations in poses, illuminations, and expressions. Moreover, the candidate name set may be noisy and incomplete, so a name may be mentioned in the caption, but the corresponding face may not appear in the image, and the correct name for a face in the image may not appear in the corresponding caption.

Each detected face (including falsely detected ones) in an image can only be annotated using one of the names in the candidate name set or as null, which indicates that the ground-truth name does not appear in the caption. 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.

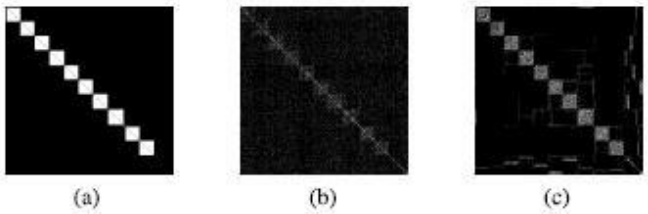


Fig.2. Coefficient matrix W* according to the ground truth and the ones obtained from LRR and rLRR. (a) W* according to the groundtruth. (b) W* from LRR. (c) W* from our rLRR.

To obtain the first affinity matrix, we propose a new method called regularized low-rank representation (rLRR) by incorporating weakly supervised information into the low-rank representation (LRR) method, so that the affinity matrix can be obtained from the resultant reconstruction coefficient matrix. 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. [2] 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 and the resultant reconstruction coefficient matrix is shown in Fig. 2(b) (More details can be found in Sections V-A and V-C). 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 large than others, indicating that the faces from the same subject tend to lie in the same subspace [2]. However, due to the Significant variances of in the-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.

Consequently, as shown in Fig. 2(b), the faces may also be reconstructed using faces from other subjects. In this paper, we show that the candidate names from the captions can provide important supervision information to better propose a method called rLRR by introducing a new discover the subspace structures. In Section III-C2, we first regularizer that incorporates caption-based weak supervision into the objective of LRR, in which we penalize the reconstruction coefficients when reconstructing the faces using those from different subjects. Based on the inferred reconstruction coefficient matrix, we can compute an affinity matrix that measures the similarity values between every pair of faces. Compared with the one in Fig. 2(b), the reconstruction coefficient matrix from our rLRR exhibits more obvious block-diagonal structure in Fig. 2(c), which indicates that a better reconstruction matrix can be obtained using the proposed regularizer. Moreover, we use the similarity matrix (i.e., the kernel matrix) based on the Mahalanobis distances between the faces as another affinity matrix. Specifically, in Section III-D, we develop a new distance metric learning method called ambiguously supervised structural metric learning (ASML) to learn a discriminative Mahalanobis distance metric based on weak supervision information. In ASML, we consider the constraints for the label matrix of the faces in each image by using the feasible label set, and we further define the image to assignment (I2A) distance that measures the incompatibility between a label matrix and the faces from each image based on the distance metric. Hence, ASML learns a Mahalanobis distance metric that encourages the I2A distance based on a selected feasible label matrix, which approximates the groundtruth one, to be smaller than the I2A distances based on infeasible label matrices to some extent.

II. RELATED WORK

Recently, there is an increasing research interest in developing automatic techniques for face naming in images [3]–[9] as well as in videos [10]–[13]. To tag faces in news photos, Berg et al. [3] proposed to cluster the faces in the news images. Ozkan and Duygulu [4] developed a graph-based method by constructing the similarity graph of faces and finding the densest component. Guillaumin et al. [6] proposed the multiple-instance logistic discriminant metric learning (MildML) method. Luo and Orabona [7] proposed a structural support vector machine (SVM)-like algorithm called maximum margin set (MMS) to solve the face naming problem. Recently, Zeng et al. [9] 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 [2] and LR-SVM [9]. 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) [14]. Similarly to [15], 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) [16], Frobmetric [17], and metric learning to rank (MLR) [18]. 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 groundtruth ordering for each query is given). Our ASML is also related to two recently proposed approaches for the face naming problem using weak supervision, MildML [6], and MMS [7]. MildML follows the multi-instance learning (MIL) assumption, which assumes that each image should contain a face corresponding to each name in the caption. However, it may not hold for our face naming problem as images, these facial pictures square measure employs a maximum margin loss to handle the structural output without using such an assumption. While MMS also uses a maximum margin loss to handle the structural output, MMS aims to learn the classifiers and it was designed for the classification problem.

Our ASML learns a distance metric that can be readily used to generate an affinity matrix and can be combined with the affinity matrix from our rLRR method to further improve the face naming performance. Finally, we compare our face naming problem with MIL [19], multi-instance multilabel learning (MIML) [20], and the face naming problem in [21]. In the existing MIL and MIML works, a few instances are grouped into bags, in which the bag labels are assumed to be correct. Moreover, the common assumption in MIL is that one positive bag contains at least one positive instance. A face straightforward way to apply MIL and MIML methods for we solving the face naming problem is to treat each image as a bag, the faces in the image as the instances, and the names in the caption as the bag labels. However, the bag labels (based on candidate name sets) may be even incorrect in our problem because the faces corresponding to the mentioned names in the caption may be absent in the image. Besides, one common assumption in face naming is that any two faces in the same image cannot be annotated using the same name, which indicates that each positive bag contains no more than one positive instance rather than at least one positive instance. Moreover, in [21], each image only the contains one face. In contrast, we may have multiple faces in a one image, which are related to a set of candidate names in the annotation,

III. LEARNING DISCRIMINATIVE AFFINITY MATRICES FOR AUTOMATIC FACE NAMING

Illustrates the system flow of the planned framework of Search-Based Face Annotation (SBFA), that consists of the subsequent steps: (1) facial image information collection; (2)face detection and facial feature extraction; (3) high-dimensional facial feature indexing; (4) learning to refine frail labeled data; (5) similar face retrieval; (6) face annotation by majority pick on the similar faces with the refined labels. the primary four steps square measure sometimes conducted before the take a look at part of a face annotation task, whereas the last 2 steps square measure conducted throughout the take a look at part of a face annotation task, that sometimes ought to be done terribly with efficiency. we have a tendency to in brief describe every step below. The primary step is that the information assortment of facial pictures as shown in Fig. 1(a), within which we have a tendency to crawled a set of facial pictures from the computer network by associate existing internet computer programmed (i.e., Google) per a reputation list that contains the names of persons to be collected. because the output of this creeping method, we have a tendency to shall acquire a set of facial pictures, every of them is related to some human names. Given the character of one. These 2 works were planned and printed when the conference version of this study [13].

Web images, these facial pictures square measure typically strident, that don't continuously correspond to the proper human name. Thus, we have a tendency to decision such reasonably internet facial pictures with strident names as frail labeled facial image information. The second step is to pre-process internet facial pictures to extract face-related data, as well as face detection and alignment, facial region extraction, and facial feature illustration. For face detection

and alignment, we have a tendency to adopt the unsupervised face alignment technique planned. For facial feature illustration, we have a tendency to extract the GIST texture options to represent the extracted faces. As a result, every face may be pictured by a d-dimensional feature vector. The third step is to index the extracted options of the faces by applying some economical high-dimensional categorization technique to facilitate the task of comparable retrieval within the subsequent step. In our approach, have a tendency to adopt the neck of the woods Sensitive Hashing (LSH), a really fashionable and effective high-dimensional categorization technique. Besides the categorization step, associate other key step of the framework is to have interaction an unsupervised learning theme to boost the label quality of the frail labeled facial pictures.

This method is extremely necessary to the whole search- based annotation framework since the label quality plays a vital think about the ultimate annotation performance. All the on top of square measure the processes before expansion a question facial image. Next we have a tendency to describe the method of face annotation throughout the take a look at part, above all, given a question facial image for annotation, we have a tendency to 1st conduct an analogous face retrieval method to go looking for a set of most similar faces (typically high K similar face examples) from the antecedently indexed facial information. With the set of high.

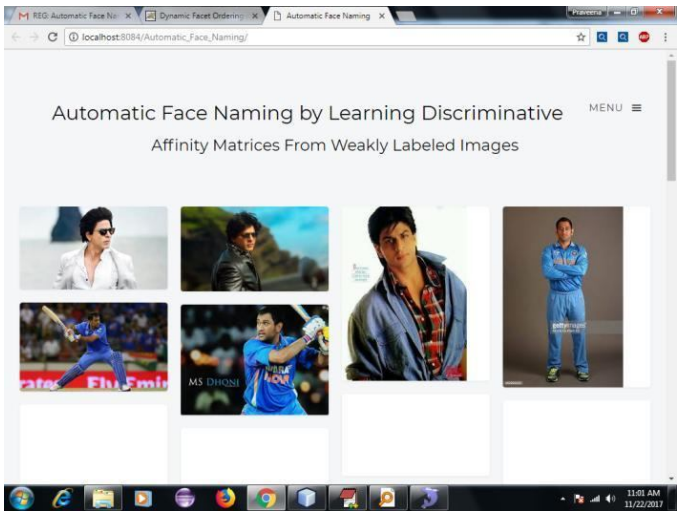
K similar face examples retrieved from the information, succeeding step is to annotate the facial image with a label (or a set of labels) by using a majority pick approach that mixes the set of labels related to these high K similar face examples. during this paper, we have a tendency to focus our attention on one key step of the on top of framework, i.e., the unsupervised learning method to refine labels of the frail labeled facial pictures. A. To identify the face of persons in the image: Since the principles of proximity supported assumption that subspaces area unit linearly freelance ,LRR seeks a reconstruction matrix W=[w1,….,wn] Rd*n. wherever every Badger State denotes the illustration of xi victimization X because the wordbook. since X is employed as wordbook to reconstruct itself, best resolution W* of LRR encodes the pair-wise affinities between knowledge samples.

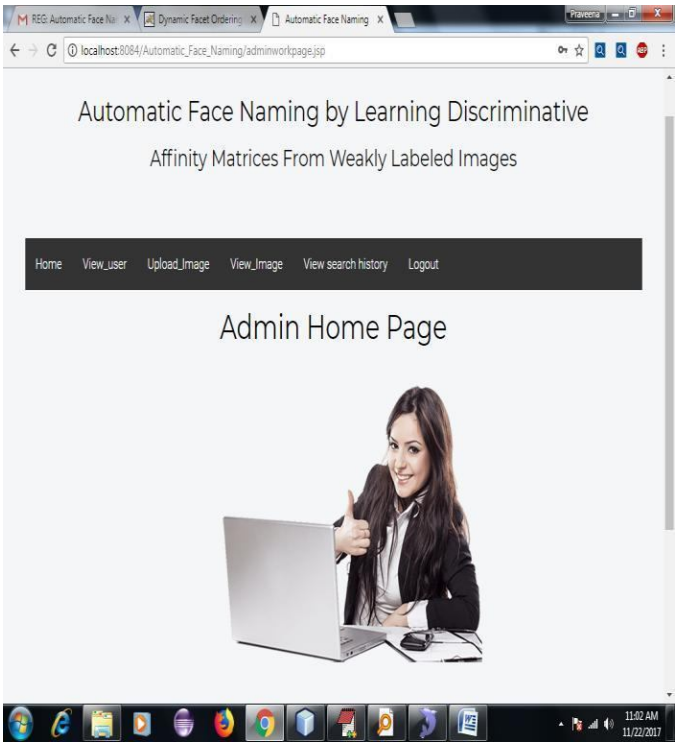
In the noise-free case W* ought to be ideally block diagonal wherever W*i,j isn't up to zero if the ith sample and jth sample area unit in same mathematical space. LRR learns the constant matrix W in Associate in Nursing unattended means. Based on the motivation we have a tendency to introduce new regularization ||WfH||2F by incorporating weak supervised data wherever H n*n is outlined supported candidate name set. we have a tendency to penalise the nonzero entries in W, wherever corresponding try of faces don't share any common name in candidate name set, and meantime we have a tendency to penalise entries appreciate state of affairs wherever face is reconstructed by itself. Once we have a tendency to get optimum resolution W*, affinity matrix Aw will be computed as Aw= 1/2(W*+W*') and Aw is any normalized to be at intervals the vary of [0,1]

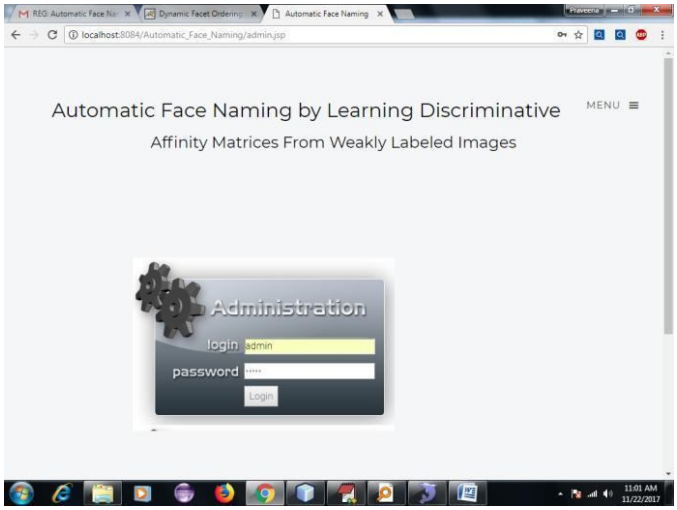
B. To improve the face naming performances:

Algo: Input: The feasible lebel sets {yi|mi=1},the affinity matrix A,the initial lebel matrix Y(1) and the parameters Niter,Θ. 1: for t=1:Niter do 2: update B by victimization B=[b1,….,bp+1]', wherever bc=(Ayc/1'yc), λc=1,….,p with yc being the c-th column of Y(t)',and bp+1= Θ1 3: update Y(t+1) by solving m subproblems in (19) 4: break if Y(t+1)=Y(t); 5:end for Output: the label matrix Y(t+1) 3.3 To implement new scheme for face naming with caption based supervision: With the constant matrix learned from rLRR,we will calculate the primary affinity matrix and as Aw and normalize it to the vary [0, 1].Furthermore, with the learnt distance metric M from ASML, we will calculate the second affinity matrix as AK = K, wherever K may be a kernel matrix supported the Mahalanob is distances between the faces. Since the 2 affinity matrices explore weak management data in numerous ways in which, they contain complementary data and each of them square measure helpful for face naming. For higher face naming performance, we tend to mix these 2 affinity matrices and perform face naming supported the amalgamate affinity matrix. Specifically, we tend to acquire a amalgamate affinity matrix A because the linear combination of the 2. affinity matrices. i.e. A=(1- d)Aw+dAk. Finally, we tend to perform face naming supported A. Since the amalgamate affinity matrix is obtained supported rLRR and ASML, we tend to name our projected technique as rLRRml.

IV. RESULT

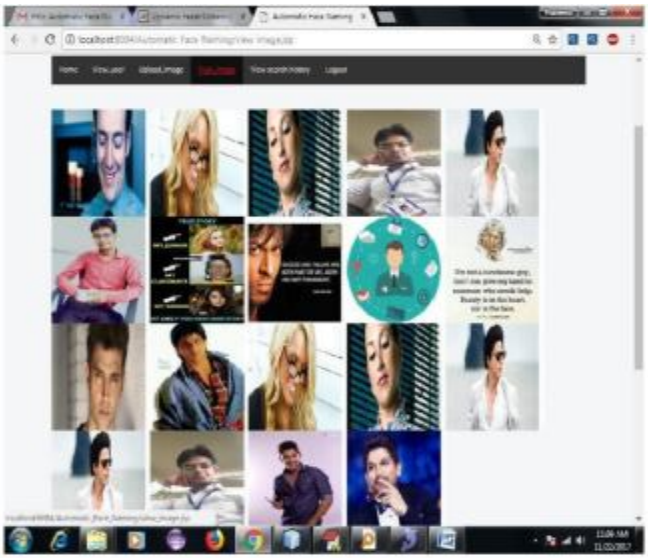

Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
MENU


Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
Home
View_user
Upload_Image
View_Image
View search history
Logout
Admin Home Page

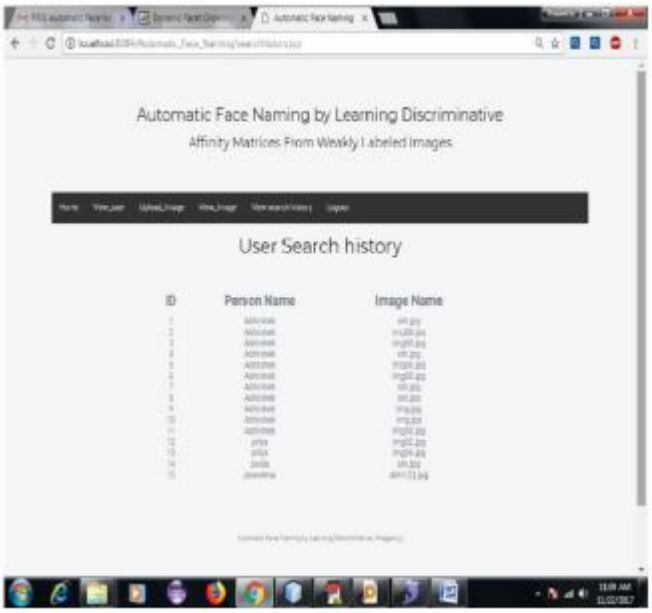

Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
MENU
Administration
login
admin
password
Login

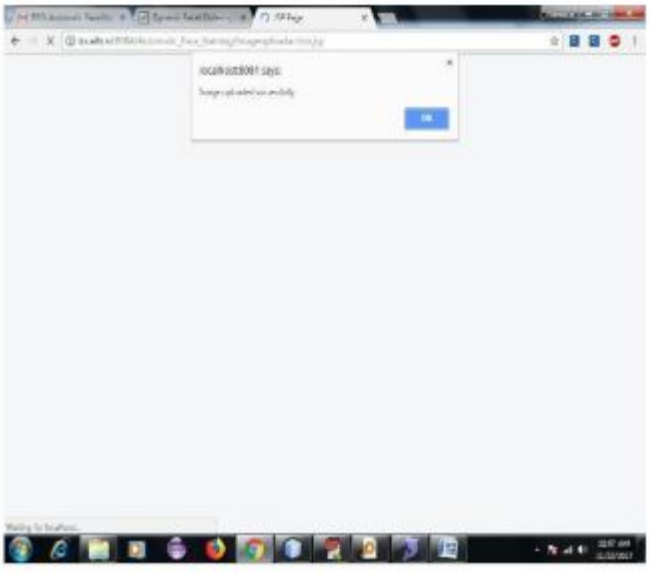

Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
Admin Home Page
ID Name
email
Mobile
Dob
Address
gender
Status
Permission

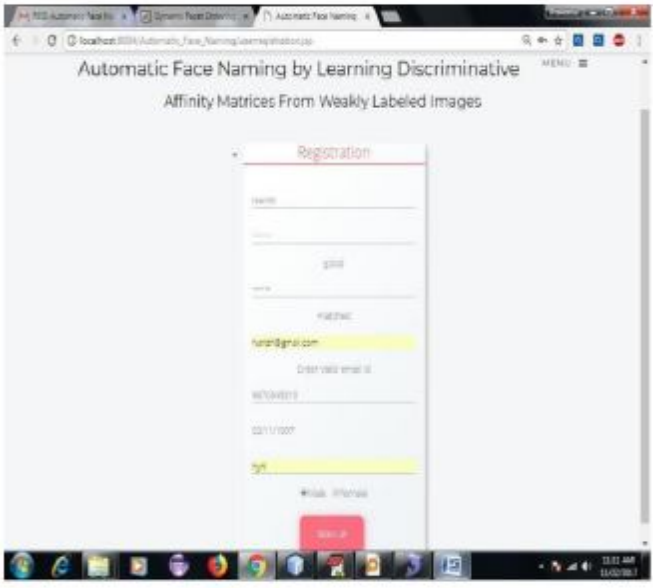

Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
Choose image file




Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
Choose image file

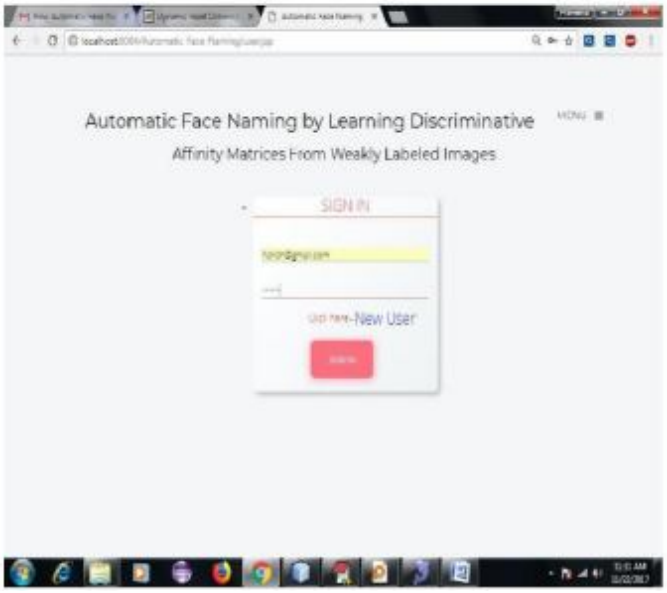

Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
User Search history



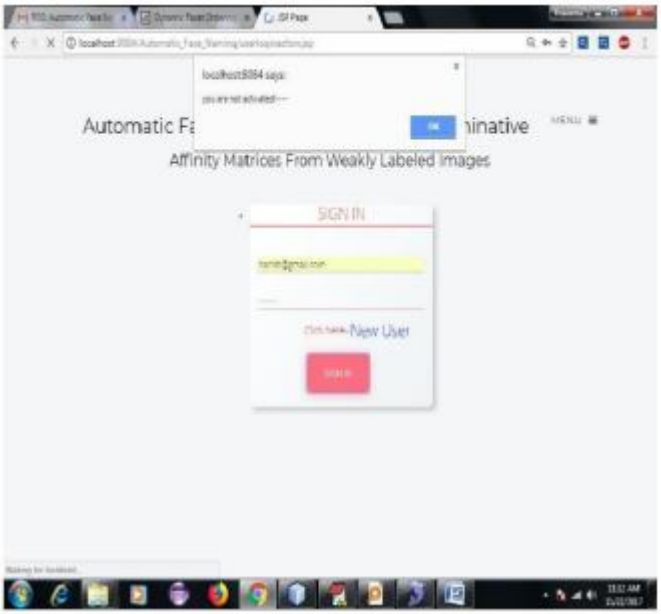

Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images

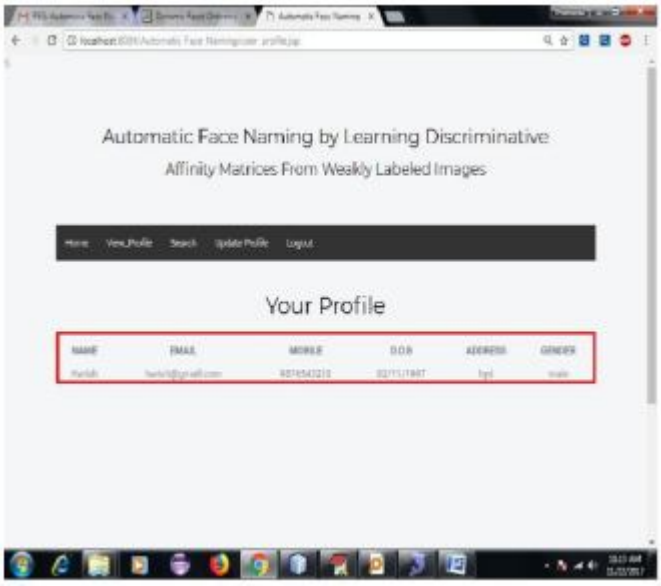

Registration

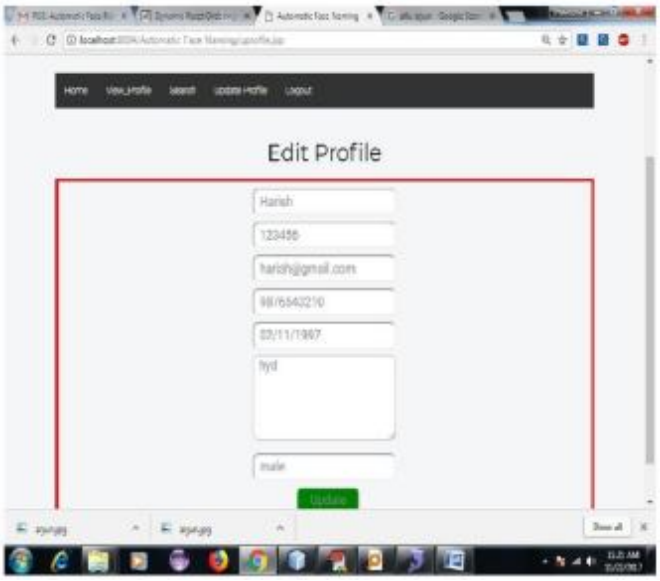

Admin Home Page
ID Name
email
Mobile
Dob
Address
gender
Status
Permission


Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
SIGN IN
New User


Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
User Home Page


Affinity Matrices From Weakly Labeled Images
SIGN IN
New User


Automatic Face Naming by Learning Discriminative
Affinity Matrices From Weakly Labeled Images
Your Profile



V. CONCLUSION

In this paper the experiments on 2 difficult real-world datasets (i.e., the participant dataset and therefore the labelled Yahoo! News dataset), our rLRR outperforms LRR, and our ASML is best than the present distance metric learning technique MildML. Moreover, our planned rLRRml outperforms rLRR and ASML, in addition as many progressive baseline algorithms. To more improve the face naming performances, we have a tendency to conceive to extend our rLRR within the future by in addition incorporating the 1-normbased regularizer and victimization different losses once planning new regularizers. we have a tendency to planned new theme during this paper for determination drawback of automatic face naming, that detects name or caption of the face located in image of multiple faces containing victimization higher than technique. Algorithms for this method we have a tendency to used LRR based mostly rLRR with introduction of latest regularizer to utilize weak oversight data. we have a tendency to develop ASML for brand new distance metric. rLRR ANd ASML obtained 2 affinity matrices by fusing this 2 affinity matrices we have a tendency to planned an repetitious theme. we are going to conjointly study a way to mechanically verify the best parameters for our ways within the future.

REFERENCES

[1] P. Viola and M. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.

[2] G.Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in Proceedings of the 27th

International Conference on Machine Learning, Haifa, Israel, Jun. 2010, pp. 663–670.

[3] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y. W. Teh, E. G. Learned-Miller, and D. A. Forsyth, "Names and faces in the news," in Proceedings of the 17th IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, Jun. 2004, pp. 848–854.

[4] Chunhua Shen, Junai kim"scalable large margin mahalanobis distance metric learning" IEEE transaction on neural networks.vol 21.no 9.september 2010

[5] Yue deng,Quionghai dai, "Low rank structure learning via non convexheuristic recovery" IEEE Tranaction on networks and learning systems, International Journal For Technological Research In Engineering Volume 4, Issue 2, October-2016 ISSN (Online): 2347 - 4718 www.ijtre.com Copyright 2016.All rights reserved. 235 vol. 24, no. 5, may 2013

[6] Xinxing Xu,Ivor W.tsang,and Dong Xu, "soft margin multiple kernel learning" IEEE Tranaction on networks and learning systems, vol. 24, no. 3, mar 2013

[7] Ishwarya, madhu B,veena potdar, "face and name matching in a movie by graphical methods in a dynamic way", International journal of scientific & technology research volume 2,issue 7,july 2013

[8] Peng wang, Qiang ji,"Robust face tracking via collaboration of generic and specific models", IEEE transaction on image processing vol. 17,no.7,july 2008

[9] Lei pand and chang wah Ngo "Unsupervised celebrity face naming in web videos" IEEE transaction on multimedia vol. 17,no.6,june 2015.

[10] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma, "Annosearch: Image autoannotation by search," in CVPR, 2006, pp. 1483–1490. 1, 2

[11] L. Wu, S. C. H. Hoi, R. Jin, J. Zhu, and N. Yu, "Distance metric learning from uncertain side information for automated photo tagging," ACM TIST, vol. 2, no. 2, p. 13, 2011.

AUTHOR'S PROFILE:

Ajay Kumar, PG Scholar, Dept of CSE, R.P. Inderprastha Institute of Technology, India,
Email: ajaykushwahacse@gmail.com.

Er. Ritika Mehra, Received the Master of Technology degree in Embedded Systems from the N.C.C.E.ISRANA Panipat, She received the Bachelor Of Technology from the J.M.I.T. College, Radaur, Yamuna Nagar. She is currently working as Assistant Professor and a Head of the Department of CSE with R. P. Inderprastha Institute Of Technology, Karnal. Her interest subjects are Data Base Management System, Software Engineering, Operating System, Digital Image Processing and etc,
Email:- er.ritika2410@gmail.com.

Er.Sandeep Garg, Received the Master of Technology degree in Network Security and Management from the I.T.M. Gurgaon, he received the Master of Science degree from Guru Nanak Khalsa College, Karnal. He is currently working as Assistant Professor of CSE with R.P.Inderprastha Institute of Technology, Karnal. His interest subjects are Data Structure,Computer Network, Artificial Intelligence, Image Processing and etc,
Email: Sandeep.1091@gmail.com.