Distinctive Analysis of The Facial Matrices For Automatic Face Annotation

Kavitha G L[1], D V Pranathi Suhasini[2]
[1, 2] Dept of Information Science and Engineering,
[1, 2] Atria Institute of Technology
Bangalore - 560024 , Karnataka, India.

Abstract- *While trading with images which comprises of several faces captioned with respective names, it may so happen that it can be incorrectly annotated. The face naming procedure that we propose, exploits the weakly labeled image dataset, and aims at labeling a face in the image precisely. We propose this effective face naming technique which is self regulated and aims at appropriately labeling a face in an image.*

This is a perplexing task because of the very large appearance variation in the images, as well as the potential discrepancy between images and their captions.

We propose a method called Regularized Low-Rank Depiction (RLRD) which productively employs the weakly named image information to decide a low-rank matrix which is obtained by inspecting many subspace structures of the recreated data. From the recreation method used we infer a discriminatory matrix. We also organize the Large Margin Nearest Neighbor (LMNN) method to label an image, which further leads to another kernel matrix. This is based on the Mahalanobis distances of the data. The two corresponding facial matrices can be fused in such a way as to enhance the quality of each other and it is used as a new reiterative plan to deduce the names of each facial image. Experimental results on synthetic and real world data sets proves the efficiency of the proposed method.

Keywords- Weakly Labeled Image Dataset, Regularized Low Rank Depiction (RLRD), Large Margin Nearest Neighbor (LMNN).

I. INTRODUCTION

The extensive growth of Internet based photo sharing has become beneficial to many real world applications.

Many of the facial images shared over different social media are wrongly annotated. A few methods were proposed in the literature for this image annotation problem. We aim at automatic image naming which stand on the uncertain affiliated captions.

Initial steps include using automated face detectors [1] and label entity detectors. The series of labels are expressed as the candidate label set. Despite these initial steps self-regulated face labeling is very challenging because of the large appearance variation in the images, and mismatch between images and their captions. Besides this, the candidate label set may sometimes be disturbed and incomplete and so a labeled image may not have the right labeled caption.

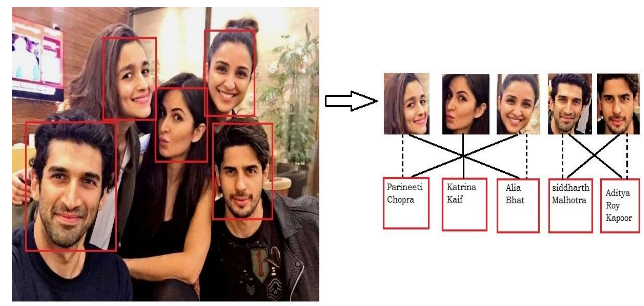

Fig. 1. Gives an inset to the facial image annotation problem. The solid lines represent the correctly named faces in the image and the dotted lines represent the weakly annotated faces.

Every face recognized would use only one label from the candidate label set or it may be set to null, indicating that the unidentified entity does not appear in the caption.

We introduce a new system of self-regulated face naming with label-based control. We obtain two correspondingfacial matrices by determining the wrongly named images. These two matrices which are discriminated and merged into a single merged matrix based on which a reiterative plan is advanced for the self-regulated face naming.

We propose a new method called Unsupervised Regularized Low Rank Depiction to obtain the first facial matrix by consolidating wrongly labeled image information from the Unsupervised Label Refinement (ULR) method, so that the recreated matrix can be eventually obtained. To productively interpret the likeliness between the faces based on the visual appearance of the faces and the labels in the candidate label set, we accomplish the subspace structures [2] among faces based on the following inference that the faces of the same subject are present in the same subspace and the subspaces are linearly absolute.

Universal Label Refinement (ULR) [3] is devised to amplify the naming quality by using graph based and low-rank learning scheme. It is a scheme to refine the labels of the facial images by exploring machine learning techniques.

Introducing our proposed method, the RLRD is a new regularized approach which consolidates with the caption based weak supervision into the unbiased ULR in which we rebuke the regeneration of the faces using different subjects; based on the interpreted recreated matrix we can encrypt the similarity between each pair of faces.

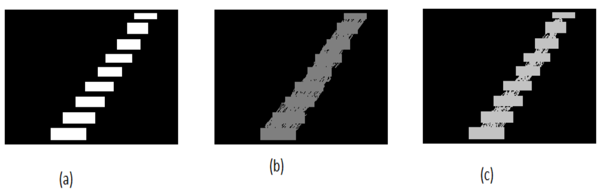

Fig. 2.
(a) Original image W* according to the ground truth
(b) W* from ULR Algorithm
(c) W* from RLRD Algorithm (proposed system).

Besides, the kernel matrix is based on the Mahalanobis distances between the faces as another consistent facial matrix.

Large Margin Nearest Neighbor (LMNN) [4] scheme uses the Mahalanobis distances steadily improving the kNN (k nearest neighbor) grouping through Euclidean distances. LMNN classification works well with PCA Principle Component Analysis than Linear Discriminant Analysis when some form of dimensionality reduction is mandatory for preprocessing.

In contemplation of RLRD and LMNN we examine the weak supervision in discrete and creative way. The two corresponding facial matrices are combined to obtain a merged facial matrix that is utilized for face labeling.

II. RELATED WORK

Programmed face labeling is one of the key area of interests these days. Most of the research are focused on developing techniques for automatic image naming. Berg et al. [5] presented face clumping method to annotate the faces in news pictures. M Guillaumin [6] introduced the multiple-instance metric learning from automatically labeled bags of faces (MildML). Ozkan and Duygulu [7] developed a graph-based method by constructing the similarity graph of face.

Zeng et al. [8] developed the low-rank SVM (LR-SVM) method which makes use of an assumption that the feature matrix of faces from the same subject is low rank. Luo and Orabona [9] developed learning from candidate labeling sets method for face naming.

Following is the comparison between our proposed method and existing systems:

Our proposed method RLRD is recounted to LR-SVM [8] and ULR [3]. In case of LR-SVM approach, LR-SVM considers distant supervision data in the permutation matrices, whereas RLRD utilizes regularizer that we have proposed, to deal with the recreation coefficients. In LR-SVM, data is not recreated by using itself as the base. In case of RLRD, it is related to the recreation-based approach of ULR. ULR is an unsupervised method that evaluates multiple subspace structures of data. Whereas, RLRD considers the image-level constraints to solve the face labeling problem in images..

Large-margin nearest neighbors (LMNN), is a traditional metric learning system. LMNN is constructed on appropriate supervision without any uncertainty. LMNN utilizes the hinge loss function. LMNN was proposed to learn distance metric M that supports the squared Mahalanobis distance between each training sample and its target neighbors to be smaller than those between this training sample and samples from other classes. The LMNN algorithm is built on the remark that the kNN will correctly classify an example if its k-nearest neighbors share the same label. The algorithm attempts to increase the number of training examples with this property by learning a linear transformation of the input space that precedes kNN classification using Euclidean distances

LMNN learns a distance metric that can be used to produce a facial matrix and can be fused with the facial matrix obtained from RLRD approach for the betterment of image labeling performance.

In the existing systems, such as MIL and MIML, data objects are represented as bags of instances. The distance between the data objects (bags) is a set-to-set distance. MIL makes use of class-to-bag distance, which assesses the relationships between the classes and the bags. The face labeling problem is solved by applying MIL and MIML method, in which each image is treated as a bag, faces in the image as the instances and names in the candidate name set as bag labels.

In some cases, the bag labels may be incorrect due to absence of names in the caption to which a face corresponds.

III. DIFFERENTIATION OF FACIAL MATRICES FOR AUTOMATIC FACE ANNOTATION

In this paper, we introduce a new system of self-regulated face naming with label-based control. This is exciting because of the essential mismatch between numerous facial images and their captions. We work on two facial matrices by making use of the ambiguous labels, to achieve image annotation based on the facial matrix attained by fusing the two facial matrices. Further in the paper, we brief our approach called Regularized Low-Rank Depiction (RLRD). The facial matrix obtained from this method is fused with the facial matrix obtained from the LMNN [4] method.

I_n is defined as the n×n similarity matrix, and 0_n , $1_n \in R_n$ as the n×1 column vectors of all zeros and ones, in the corresponding order. Also, we use I, 0 and 1 instead of I_n, 0_n, and 1_n in the case where the magnitudes are evident. tr(A) represents the trace of A and <A, B> means the dot product of two matrices. A◦B represents the element-wise multiplication of two matrices A and B (a◦b in case of vectors a and b). $\|A\|_\infty$ denotes the greatest absolute value of all the elements contained in matrix A. $\|A\|_F = (\sum_{i,j} A^2{}_{i,j})^{1/2}$ represents the Frobenious norm of the matrix A. $a \leq b$ implies that $a_i \leq b_i$ ∀ i = 1,...,n . $A \geq 0$ denotes that A is a positive semidefinite matrix (PSD matrix).

A. Problem Statement

While dealing with facial images which is captioned with corresponding names, it may so happen that it may be wrongly annotated. The face naming technique that we propose, is self regulated and aims at correctly labeling a face in an image.

This wrong annotation may happen due to the variation in the images and mismatch between the images and their captions. In this paper, we present methods for face naming using collection of images with captions. This is carried out in two steps:

1) Retrieve all faces of a particular person from the data set.
2) Establish the correct association between the names in the captions and faces in the image.

Let us assume that we have m images, each of which consists of r_i names and n_i faces, ∀ i = 1…..m. Let q ∈ {1,...,p} denote a name and x ∈ R^d denote a face, where p is the total number of names in all the captions and *d* is the feature dimension. Thereafter, each image can be represented as (X^i , N^i), where X^i =[$x^i{}_1$,...,$x^i{}_{ni}$] ∈ $R^{d\times ni}$ is the data matrix for faces, that are in the *i*th image with each x^i_f being the *f*th face in the image (f = 1,...,N_i), and N^i={ q^i_1,..., $q^i_{r_i}$} is the corresponding set of candidate names with each q^i_j ∈ { 1,...,p} being the *j*th name (j = 1,...,r_i). Further, let X =[X^1 …..X^m] ∈ $\mathbb{R}^{d\times n}$ represent the data matrix of the faces from all m images, where n $=\sum_{i=0}^{m} n_i$. After defining a binary label matrix Y =[Y^1,...,Y^m] ∈ $\{0,1\}^{(p+1)}$with each Y^i ∈ $\{0,1\}^{(p+1)\times n_i}$ being the label matrix for each image X^i, the next step is to infer the facial label matrix Y based on the candidate name sets $\{N^i|_{i=1}^m\}$. When the ground-truth name of a face does not appear in the associated candidate name set N^i, we make use of the (p+1)th name to denote null class, so that the face can be assigned to the (p+1)th name. The label matrix Y^i for each image should satisfy the following image-level constraints [8].

1) Distinctiveness: In the same image, two faces cannot be annotated with the same name except the (p+1)th name, i.e., $\sum_{f=1}^{n_i} y^i_{j,f} \leq 1, \forall j = 1, \dots, p$.
2) Expediency: the faces in the *i*th image should be tagged
using the names from the set:
$\tilde{N}^i = N^i \cup \{(p+1)\}$, i.e., $Y^i_{j,f} = 0, \forall f = 1, \dots, and$
j ∉ $\tilde{N}^i$.
3) Non-Pleonastic: In the ith image, each face should be tagged exactly one name from the set $\tilde{N}^i$, i.e., $\sum_j Y^i_{j,f} = 1, \forall f = 1, \dots, n_i$.

B. Face Naming Using Facial Matrix

The feasible set of $\mathbf{Y}^i$ for the *i*th image, based on image-level constraints can be defined as follows:

$$y^i = \left\{ Y^i \in \{0,1\}^{(p+1)\times n_i} \;\middle|\; \begin{array}{c} 1'_{(p+1)}(Y^i \circ T^i)1_{n_i} = 0, \\ 1'_{(p+1)}Y^i = 1'_{n_i} \\ Y^i 1_{n_i} \leq [1'_p, n_i]' \end{array} \right\} \quad (1)$$

The matrix $T^i \in \{0,1\}^{(p+1)\times n_i}$ has rows related to the indices of the names in $\tilde{N}^i$ are all zeros and rest of rows are all ones.

The feasible set for the label matrix can be represented as $y = \{Y = [Y^1, \ldots, Y^m] \mid Y^i \in y^i \; \forall_i = 1, \ldots, m\}$.

Let $A \in \mathbb{R}^{n \times n}$ be a facial matrix, which meets the condition $A = A'$ and $A_{i,j} \geq 0, \forall i, j$. Each $A_{i,j}$ expresses the pair-wise similarity between the *i*th face and the *j*th face. Our goal is to learn a proper A such that $A_{i,j}$ is large if and only if the *i*th face and the *j*th face share the same ground-truth name. Then, the face naming problem can be solved based on the facial matrix A obtained. We solve the following, to annotate the faces in an image:

$$\max_{Y \in y} \sum_{c=1}^{p} \frac{y_c' A y_c}{1' y_c} \quad s.t \quad Y = [y_1, y_2, \ldots, y_{(p+1)}]' \qquad (2)$$

$y_c \in \{0,1\}^n$ correlates to the *c*th row in Y. The faces with the same label are clustered as one group, and the sum of the average similarities for each group is maximized.

We propose RLRD method to learn the Unsupervised Regularized Low-Rank recreation matrix. We obtain our first facial matrix from our RLRD method. Likewise, we make use of LMNN method to find alternative facial matrix. Lastly, we fuse these two facial matrices into one single facial matrix in order to accomplish image tagging.

C. Learning Refinement of Facial Matrix with Regularized-Low-Rank Dipiction (RLRD)

We will initially examine ULR (unsupervised label refinement) and then present our proposed method, RLRD.

1) Description of ULR

ULR was planned to improve the face labelling quality through a graph-based and low-rank learning (LRR) method. ULR makes use of content-based image search face annotation, face annotation presentation on database. LRR is intended to solve the subspace clustering problem. The objective of LRR is to gauge the structure of subspace in the specified data $X = [x_1, \ldots, x_n] \in R^{d \times n}$. LRR efforts to obtain a recreation matrix W, which is based on an postulation that the subspaces have linearly independent vectors . This recreation matrix W is given by $W = [w_1, \ldots, w_n] \in R^{n \times n}$, where each w_i denotes the representation of x_i using X as the base. Because X is used as the base to recreate itself, the ideal solution W∗ of LRR encodes the pair-wise similarity between the data matrices. The efficiency problem of LRR is given as:

$$\min_{W,E} \|W\|_* + \lambda \|E\|_{2,1} \quad \text{s.t } X = XW + E \qquad (3)$$

where $E \in R^{d \times n}$ is the recreation error, $\lambda > 0$ is a tradeoff parameter, $\|W\|_*$ which is the nuclear form, is used to replace rank(W) as commonly used in the rank minimi- zation problems, and $\|E\|_{2,1} = \sum_{j=1}^{n} (\sum_{i=1}^{d} (E_{i,j})_2)^{1/2}$ is a regularizer that supports the recreation error E to be column-wise sparse. LRR performs better than sparse subspace clustering method, and hence produces better results in most of the real world applications that includes Faceprints.

Graph based method is proposed to determine the most relevant subset among the set of possible faces related to the query name, where the most relevant subset is likely to match with the faces of the queried person. Graph based method is implemented to rectify the correct faces of a queried person using both text and visual appearances. This approach eliminates the wrong tags, by applying geometrical constraint. The geometrical distance corresponding to the *i*th assignment refers to

$$\sqrt{X^2 + Y^2} \text{ where,}$$

$$X = \frac{\text{locX(i)}}{\text{sizeY(image1)}} - \frac{\text{locX(match(i))}}{\text{sizeX(image2)}}$$

$$Y = \frac{\text{locY(i)}}{\text{sizeY(image1)}} - \frac{\text{locY(match(i))}}{\text{sizeY(image2)}} \qquad (4)$$

And locX is the X coordinate and locY is Y coordinate of the feature points in the images, sizeX and sizeY hold X and Y sizes of the images and match(i) resembles the matched keypoint in the second image of the ith feature point in the first image.

Unsupervised Label Refinement (ULR) mission is to learn a refined label matrix $F^* \in R^{n \times m}$ to progress the initial raw label matrix Y . ULR makes use of an notion called “label smoothness”. i.e., the more alike the visual contents of two facial images, the more likely they share the same labels. The label smoothness principle is framed as an idealization problem of reducing the following loss function $E_s(F,W)$:

$$E_s(F,W) = \frac{1}{2} \sum_{i,j=1}^{n} W_{i,j} \| F_{i*} - F_{j*} \|_F^2 = tr(F^T L F) \qquad (5)$$

Where W is a weight matrix of a sparse graph, $\| \cdot \|_F$ denotes the Frobenius norm, L = D−W signifies the Laplacian matrix where D is a diagonal matrix with diagonal elements as $D_{ii} = \sum_{j=1}^{n} W_{i,j}$ and tr denotes a trace function.

We implement a new word $\|W \circ H\|_F^2$, which is called regularization term that comprises of the weak supervised information. Definition of $H \in \{0,1\}^{n\times n}$ depends on the candidate name sets $\{N^i|^m{}_{i=1}\}$. $H_{i,j} = 0$ if the following two conditions fulfil:

1) the i^{th} face and the j^{th} face has at least one name in common, in the corelated candidate name sets and
2) i = j. If not, $H_{i,j} = 1$.

And so forth, non-zero entries in W, where the corelated pair of faces have no names in mutual in their candidate name sets, and the entries that corelate to the circumstances where a face is recreated by itself, are corrected. Therefore, the subsequent facial matrix W is expected to be more distinguishable, with information related to weak supervision encoded in H.

By implementing the new regularizer $\|W\circ H\|^2{}_F$ into ULR, the new optimization problem is achieved as follows:

$$\min_{W,E,} \|W\|_*+\lambda\|E\|_{2,1}+\frac{\gamma}{2}\|W\circ H\|^2_F \text{ s.t. } X = XW+E \quad (6)$$

where $\gamma \geq 0$ is a used to balance the new regularizer with the other term. This problem is referred to as RLRD. By setting the parameter γ to zero, the RLRD problem in Eq(5) can be reduced to the ULR problem .

Once we obtain the ideal solution W* after solving Eq(6), the facial matrix A_W can be computed as $A_W = \frac{1}{2}(W^*+W^{*'})$.

3) Optimization: To obtain equivalent optimization problem , an intermediate variable J is introduced in Eq(6):

$$\min_{W,E,J} \|J\|_* + \lambda\|E\|_{2,1} + \frac{\gamma}{2}\|W\circ H\|^2_F \text{ s.t. } X = XW+E,\ W = J. \quad (7)$$

We Consider the following augmented Lagrangian function from Augmented Lagrangian Method (AML):

$$L= \|J\|_* + \lambda\|E\|_{2,1} + \frac{\gamma}{2}\|W\circ H\|^2_F + \langle U, X-XW-E\rangle + \langle V,W-J\rangle + \frac{\rho}{2}(\|X-XW-E\|^2_F + \|W-J\|^2_F) \quad (8)$$

where ρ is a positive penalty parameter and $U \in R^{d\times n}$ and $V \in R^{n\times n}$ are the Lagrange multipliers. Notably, we set the following parameters as follows:

$E_0 = X - XW_0$, $W_0 = (1/n)(1_n1'_n - H)$, $J_0 = W_0$ and U_0,V_0 as zero matrices. The following steps are performed recursively at the *t*th iteration, until convergence is achieved.

1) Fix the others and update J_{t+1} by

$$\min_{J_{t+1}} \|J_{t+1}\|_* + \frac{\rho_t}{2}\left\|J_{t+1} - \left(W_t + \frac{V_t}{\rho_t}\right)\right\|_F^2$$

which can be solved in closed form using the singular value thresholding method.

2) Fix the others and update W_{t+1} by

$$\min_{W_{t+1}} \frac{\gamma}{2}\|W_{t+1}\circ H\|_F^2 + \langle U_t, X - XW_{t+1} - E_t\rangle + \langle V_t, W_{t+1} - J_{t+1}\rangle + \frac{\rho_t}{2}\|X - XW_{t+1} - E_t\|_F^2 + \frac{\rho_t}{2}\|W_{t+1} - J_{t+1}\|_F^2 \quad (9)$$

Due to the new regularizer $\|W\circ H\|_F^2$ this problem cannot be solved as in [2] by using pre-computed SVD. We use the gradient descent method to efficiently solve (7), where the gradient

with respect to W_{t+1} is $(H\circ H)\circ W_{t+1}+$ $\rho_t(X'X+I)W_{t+1} + V_t - \rho_t J_{t+1} -$ $X'(\rho_t(X-E_t)+U_t)$

3) Fix the others and update E_{t+1} by

$$\min_{E_{t+1}} \frac{\lambda}{\rho_t}\|E_{t+1}\|_{2,1} + \frac{1}{2}\left\|E_{t+1} - \left(X - XW_{t+1} + \frac{U_t}{\rho_t}\right)\right\|_F^2$$

4) Update U_{t+1} and V_{t+1} by respectively using

$$U_{T+1} = U_t + \rho_t(X - XW_{t+1} - E_{t+1})$$
$$V_{t+1} = V_t + \rho_t(W_{t+1} - J_{t+1}).$$

5) Update ρ_{t+1}using

$\rho_{t+1} = min(\rho_t(1+\Delta\rho), \rho_{max})$ where Δ_ρ and ρ_{max} are the constant parameters.

6) The iterative algorithm stops if the two convergence conditions are both satisfied

$$\|X - XW_{t+1} - E_{t+1}\|_\infty \leq \epsilon$$
$$\|W_{t+1} - J_{t+1}\|_\infty \leq \epsilon$$

where ϵ is a constant parameter.

D. Large Margin Nearest Neighbor Classification (LMNN)

Weinberger and Saul [4] proposed the LMNN method to learn a distance metric M that promotes the squared Mahalonobis distances between each training sample and its target neighbours to be smallers than the distance between this training sample and samples from other classes. In LMNN, the metric is trained with the goal that the k-nearest neighbors always belong to the same class and the examples from various classes are separated by a large margin. The algorithm is based on an observation that an example will be classified correctly by KNN decision rule, if its K-nearest neighbors share the same label. Large Margin Nearest Neighbor (LMNN) metric learning algorithm has been used widely in many applications and has produced promising results.

LMNN optimizes matrix M with the help of semidefinite programming. The objective is twofold: For every data point $\vec{x}_i$, the target neighbours should be close and imposters (differntly labelled) should be far away. The learned metric causes the input vector $\vec{x}_i$ to be surrounded by training instances of the same class. This optimization is illustrated in figure 3.

Let $\{(x_i, y_i)|^n_{i=1}\}$ be the n labeled samples: $x_i \in R^d$ denotes the i^{th} sample, with d being the feature dimension, and $y_i \in \{1,...,z\}$ denotes the label of this sample, with z being the total number of classes. $\eta_{i,j} \in \{0,1\}$ indicates whether x_j is a target neighbor of x_i. i.e, $\eta_{i,j} = 1$ if x_j is a target neighbour of x_i, and $\eta_{i,j} = 0$ if x_j is a target neighbor of x_i ,$\forall i,j \in \{1..n\}$. $\nu_{i,l} \in \{0,1\}$ indicates whether x_l and x_i are from different classes. i.e, $\nu_{i,l} = 1$ if $y_l \neq \mathbf{y_i}$, and $\nu_{i,l} = 0$ if $y_l = \mathbf{y_i}$, $\forall$ i,l $\in\{1,...,n\}$. The squared Mahalanobis distance between t-wo samples x_i and x_j can be defined as:

$$d^2_M(x_i,x_j)=(x_i-x_j)'M(x_i-x_j).$$

LMNN minimizes the following idealization problem:

$$\min_{M\geq 0} \sum_{(i,j):\eta i,j=1} d^2_M(x_i,x_j) + \mu \sum_{(i,j,l)\in S} \xi_{i,j,l}$$
$$\text{s.t } d^2_M(x_i,x_l) - d^2_M(x_i,x_j) \geq 1-\xi_{i,j,l}\ , \forall (i, j,l) \in S,\ \xi_{i,j,l} \geq 0\ , \forall (i, j,l) \in S \quad (10)$$

where $\xi_{i,j,l}$ is a slack variable, μ is a tradeoff parameter and $S =\{ (i, j,l)|\eta_{i,j} = 1, \nu_{i,l} = 1, \forall i, j,l \in\{ 1,...,n\}\}$. Therefore, $d^2_M(x_i,x_j)$ is the squared Mahalanobis distance between x_i and its target neighbor x_j, and $d^2_M(x_i,x_l)$ is the squared Mahalanobis distance between x_i and x_j that belong to different classes. The slack variable can condone the cases when $d^2_M(x_i,x_l) - d^2_M(x_i,x_j)$ is smaller than one. The LMNN problem in Eq. (10) can be equivalently reformulated as the idealization problem as follows:

$$\min_{M\geq 0} \sum_{(i,j):\eta i,j=1} d^2_M(x_i,x_j) + \mu \sum_{(i,j,l)\in S} |1 - d^2_M(x_i,x_l) + d^2_M(x_i,x_j)|_+$$

Where $|\cdot|_+$ is the truncation function.

Algorithm 1 summarizes the entire learning process.

Algorithm 1: LMNN

Input: Data samples $\mathbf{\{x_i,y_i\}^N_{i=1}}$, number of target neighbors K, output dimension m, maximum number of optimization iterations T.

Result: matrix $L \in R^{d\times m}$

Initialize L with the first m leading eigen vectors of the covariance matrix of the data samples $\mathbf{\{x_i\}^N_{i=1}}$;

For t=1 to T ***do***

Randomly generate subsamples S;

Calculate the descending direction d;

Use line search algorithm to find the step length λ;

Update $L \leftarrow L + \lambda d$;

if the termination condition satisfies ***then*** ***break***;

IV. ANNOTATION OF FACIAL IMAGES

The first facial matrix A_w can be calculated as, $A_W = \frac{1}{2}(W^* + W^{*'})$, using coefficient matrix W^* learned from RLRD, and regularize A_W to the range [0,1]. The second facial matrix can be calculated from learnt distance metric M of LMNN as A_K = K, where K is a kernel matrix depending upon the Mahalanobis distance. These two facial matrices use weak supervision material in diverse ways. Thus, the two facial matrices comprise of interdependent data which is beneficial for face annotation. We associate the two facial matrices obtained from our RLRD and LMNN to attain better accuracy, and we call this fused facial matrix as RLRD, which is our proposed method. This fused facial matrix A is the linear combination of the two facial matrices derived from RLRD and LMNN, where A is given by, $A=(1-\alpha)A_W +\alpha A_K$, where α is a parameter in the range [0, 1]. Finally, the image face naming tagging is carried out based on A. We work on image face annotation by resolving the following idealization problem:

$$\max_{Y\in y} \sum_{c=1}^{p} \frac{Y'_c A Y_c}{1'Y_c} \quad \text{S.T, } Y = [\, y_1 \ldots\ldots\ldots y_{(p+1)}]'. \qquad (11)$$

But, the above problem is computationally costly to solve. To solve this problem, we propose an iterative method.

At each iteration, an objective function is estimated using $\tilde{y}'_c Ay_c / 1'\tilde{y}_c$ that can substitute $y'_c Ay_c/1'y_c$, where $\tilde{y}_c$ is the solution for y_c inferred from the previous iteration. Therefore, we solve the linear programming problem at each iteration, as follows:

$$\max_{y \in Y} \sum_{c=1}^{p} b'_c y_c \text{ , s.t. } Y = [y_1,...,y_{(p+1)}]' \qquad (12)$$

where $b_c = A\tilde{y}_c / 1'\tilde{y}_c, \forall c = 1,...,p$. f the faces may not annotated with their correct name.

The problem in Eq. (12) can be reformed by defining $B \in R^{(p+1)\times n}$ as $B=[b_1,...,b_{p+1}]$. The reformulated form is as follows:

$$\max_{y \in Y} \langle B,Y \rangle \qquad (13)$$

The viable set for Y is defined as $\mathcal{Y} = \{Y = [Y^1,...,Y^m] | Y^i \in \mathcal{Y}^i, \forall i = 1,...,m\}$. Matrix B can be expressed as $B=[B^1,...,B^m]$, where each $B^i \in R^{(p+1)\times ni}$ correlates to Y^i. Then, the objective function in Eq. (13) can be conveyed as $\langle B,Y \rangle = \sum_{i=1}^{m} \langle B^i,Y^i \rangle$. Therefore, Eq. (13) can be optimized by solving m sub-problems, with each sub-problem related to one image in the following form:

$$\max_{y^i \in Y^i} \langle B^i,Y^i \rangle \quad \forall i = 1,...,m \qquad (14)$$

The ith problem in Eq. (14) can be reformulated as a minimization problem as follows:

$$\min_{Y^i_{q,f} \in \{0,1\}} \sum_{q \in \mathcal{N}^i} \sum_{f=1}^{n_i} -B^i_{q,f} Y^i_{q,f}$$
$$\text{S.T } \sum_{q \in \mathcal{N}^i} Y^i_{q,f} = 1 \quad \forall f = 1,...,n_i$$
$$\sum_{f=1}^{n_i} Y^i_{q,f} \leq 1 \quad \forall q \in N^i$$
$$\sum_{f=1}^{n_i} Y^i_{(p+1),f} \leq n_i \qquad (15)$$

in which we have left out the elements $\{Y^i_{q,f} |_{q \notin \mathcal{N}^i}\}$, because these elements are zeros according to the feasibility constraint in Eq. (1). In this paper, we adopt the Hungarian algorithm to efficiently solve the problem in Eq. (15). Certainly, for an i^{th} image, the cost c(f, p +1) for assigning a face X^i_f to the corresponding null name is set to $-B^i_{(p+1),f}$ and the cost c(f , q) for assigning a face X^i_f to a real name q is set to $-B^i_{q,f}$.

The iterative face naming algorithm is as follows:

Algorithm 2: Face Naming Algorithm

Input: The feasible label sets $\{\mathcal{Y}^i|_{i=1}^m\}$, the affinity matrix A, the initial label matrix Y(1) and the parameters $\tilde{N}_{iter}, \theta$.

1: ***for*** $t = 1 : \tilde{N}_{iter}$ ***do***

2: Update B by using $B = [b_1,..., b_{p+1}]'$, where $b_c = \frac{A\tilde{y}_c}{1'\tilde{y}_c}, \forall c = 1,...,p$ with $\tilde{y}_c$ being the c-th column of $Y^{(t)'}$, and $b_{p+1} = \theta 1$.

3: Update Y(t +1) by solving m sub problems in Eq (14).

4: break if Y(t +1)=Y(t).

5: ***end for***

Output: the label matrix Y(t +1)

V. EXPERIMENTS

We examine our proposed schemes RLRD, LMNN algorithms for face labeling by means of two real-world datasets.

A. Real-world Datasets

1) Movie Face Database (MFD) - MFD is made from frames extracted from movies of dissimilar languages. MFD database consists of 4512 facial images corresponding to 430 actors collected from roughly 103 movies. MFD consists of 67 male and 33 female actors with at least 200 images for each actor.

2) SCface – Surveillance Cameras face database
SCface is a database of static images of human faces. Images were taken in unrestrained indoor environment using five video surveillance cameras of various qualities. Database contains 4160 static images (in visible and infrared spectrum) of 130 subjects.

VII. CONCLUSIONS

This paper explored a promising search-based face annotation framework, in which we fixated on tackling the critical problem of enhancing the label quality and proposed a RLRD algorithm.

TABLE 1:
Two document examples with their naming results for LRR, ULR and RLRD, shows the maximum number of accurately named faces in an image.

#images	LRR	ULR	RLRD
	Alia Bhat, Priyanka Chopra.	Alia Bhat, Priyanka Chopra, Parineeti Chopra, Tabbu	Tabbu, Parineeti Chopra, Priyanka Chopra, Alia Bhat,
	Dennis, Mohsen, Hamed, Yi,Sam	Deva,Carl, Kuang, Julian, Bailey, Ragib	Nancy, Dennis, Deva, Mohsen, Hamed, Yi,Sam, Carl, Kuang, Julian, Bailey, Ragib, Xian

We presented a method for face detection and naming which minimizes computation time while accomplishing high detection accuracy. To productively employ the face naming of the facial images we introduce RLRD and by means of this scheme we increase the evaluation of auto face annotation performance. We also strengthen the LMNN algorithm which explores on discriminating Mahalanobis distance metric. Our proposed methods focus on tackling the critical problem of enhancing the label quality and accurately naming the facial images. We examine the two challenging and exciting real-world datasets from which we can certify that our RLRD and LMNN outperforms ULR and kNN respectively and several other baseline algorithms. Our future work will address the issues of duplicate names and explore and explore other techniques to further improve the label refinement task.

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] Dayong WANG, Chu Hong HOI, Ying HE, Jianke ZHU ‘Mining Weakly-Labeled Web Facial Images for Search-Based Face Annotation’,2014.

[4] K. Q. Weinberger and L. K. Saul ‘Distance metric learning for large margin nearest neighbor classification’ J. Mach. Learn. Res., vol. 10, pp. 207–244, Feb. 2009.

[5] T. L. Berget 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.

[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] 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

[8] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, “Face recognition from caption-based supervision,” Int. J. Comput. Vis., vol. 96, no. 1, pp. 64–82, 2012

Kavitha G L is an MTech (CSE) graduate, pursuing PhD in cloud computing. She is currently working as Asst. Professor in Atria Institute of Technology, Bangalore. Her area of interests includes machine learning and image processing.

D V Pranathi Suhasini is in 1st year M.E in Atria Institute of Technology, Bangalore. Her area of interests includes image processing.