Retrieval-based Face Annotation by Weak Label Regularized Local Coordinate Coding

Dayong Wang*, Steven C.H. Hoi*, Ying He*, Jianke Zhu†, Tao Mei‡, Jiebo Luo§
*School of Computer Engineering, Nanyang Technological University, 639798, Singapore
†College of Computer Science, Zhejiang University, Hangzhou, 310027, P.R. China
‡Media Computing Group, Microsoft Research Asia, Beijing, 100080 P.R. China
§Department of Computer Science, University of Rochester, Rochester, NY 14627, USA.
E-mail:{s090023, chhoi, yhe}@ntu.edu.sg, jkzhu@zju.edu.cn, tmei@microsoft.com, jluo@cs.rochester.edu

Abstract—Auto face annotation, which aims to detect human faces from a facial image and assign them proper human names, is a fundamental research problem and beneficial to many real-world applications. In this work, we address this problem by investigating a retrieval-based annotation scheme of mining massive web facial images that are freely available over the Internet. In particular, given a facial image, we first retrieve the top n similar instances from a large-scale web facial image database using content-based image retrieval techniques, and then use their labels for auto annotation. Such a scheme has two major challenges: (i) how to retrieve the similar facial images that truly match the query; and (ii) how to exploit the noisy labels of the top similar facial images, which may be incorrect or incomplete due to the nature of web images. In this paper, we propose an effective Weak Label Regularized Local Coordinate Coding (WLRLCC) technique, which exploits the principle of local coordinate coding by learning sparse features, and employs the idea of graph-based weak label regularization to enhance the weak labels of the similar facial images. An efficient optimization algorithm is proposed to solve the WLRLCC problem. Moreover, an effective sparse reconstruction scheme is developed to perform the face annotation task. We conduct extensive empirical studies on several web facial image databases to evaluate the proposed WLRLCC algorithm from different aspects. The experimental results validate its efficacy. We share the two constructed databases "WDB" ($714,454$ images of $6,025$ people) and "ADB" ($126,070$ images of $1,200$ people) to the public. To further improve the efficiency and scalability, we also propose an offline approximation scheme (AWLRLCC), which generally maintains comparable results but significantly reduces the annotation time.

Index Terms—face annotation, content-based image retrieval, machine learning, label refinement, web facial images, weak label

1 Introduction

Auto face annotation, which aims to detect human faces from a photo image and to tag the facial image with the human names, is a fundamental research problem and beneficial to many real-world applications. Typically, face annotation is formulated as an extended face recognition problem, in which face classification models are trained from a collection of well-controlled labeled facial images using supervised machine learning techniques [1], [2]. Recent studies [3] have attempted to explore a promising retrieval-based face annotation (RBFA) paradigm for facial image annotation by mining the world wide web (WWW), where a massive number of weakly labeled facial images are freely available.

Generally, it is easy to construct a large scale web facial image database by exploring web search engine with names as queries. All the returned images can be directly tagged with the query names. Due to the noisy nature of web images, the initial name labels of such a facial image database may be incorrect or incomplete without extra manual refinement effort. For example, for Google search engine the precision of the top 200 returned images is around 0.4 with "Nick Lachey" as a query [4]. It is also extremely time-consuming to manually label (tag) all the images with the correct names. In this work, we refer to this kind of web facial images with noisy/incomplete names as weakly labeled facial images. In the RBFA framework, given a query (test) facial image, first, we retrieve its top n similar instances from weakly labeled facial image database by exploiting content-based image retrieval (CBIR) techniques [5]–[7]. Second, we tag the query image with human names based on the weak (noisy) name labels of the top-ranking similar images. Such a paradigm is inspired by search-based general image annotation techniques since face annotation can be viewed as a special case of generic image annotation [8], [9], which has been extensively studied yet remains a challenging problem.

In general, there are two key problems for the retrieval-based face annotation (RBFA) technique: The first problem is how to efficiently retrieve a short list of the most similar facial images from a large facial image database for a query facial image. It typically relies on an effective content-based facial image retrieval (CBIR) solution. The recent work [6] mainly addresses this problem, in which an effective image representation technique is proposed for facial image retrieval by employing both the local and global features. The second problem is how to effectively exploit the short list of candidate facial images and their weak label information for the face annotation task. This is critical because the associated labels of web facial images are noisy. To address this critical problem, we propose a novel Weak Label Regularized Local Coordinate Coding (WLRLCC) algorithm to boost the annotation performance by a unified learning scheme, which exploits the local

coordinate coding principle for learning more discriminative features and makes use of the graph-based regularization for enhancing the weak labels simultaneously. The main contributions of this paper are as follows:

- We propose a novel Weak Label Regularized Local Coordinate Coding (WLRLCC) algorithm for the retrieval-based face annotation paradigm.
- We conduct extensive experiments to evaluate the proposed algorithm for automated face annotation on two large-scale web facial image databases, which are also shared to facilitate related research for other researchers.
- We propose an offline approximation scheme AWLRLCC to further reduce the running time without introducing much degradation of annotation performance.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 briefly introduces the proposed retrieval-based face annotation (RBFA) framework. Section 4 presents the proposed Weak Label Regularized Local Coordinate Coding (WLRLCC) scheme together with an effective face name annotation solution based on sparse reconstruction. Section 5 shows the offline approximation algorithm for WLRLCC. Section 6 introduces the construction of our weakly labeled retrieval database. Section 7 shows the experimental results of our empirical studies. Section 8 discusses the limitations, and Section 9 concludes this paper.

2 Related Work

Our work is closely related to several groups of research work. The first group of related work is face recognition and verification, which are classic research problems in computer vision and pattern recognition and have been extensively studied for many years [11], [12]. Although traditional face recognition methods can be extended for automatic face annotation [2], they usually suffer from a few common drawbacks. For example, they usually require high-quality facial image databases collected in well-controlled environments. This drawback has been partially addressed in recent benchmark studies of unconstrained face detection and verification techniques on the facial images collected from the web, such as the LFW benchmark [13], [14].

The second group is related to generic image annotation [15]–[17]. The common techniques usually apply existing object recognition techniques to train classification models from human-labeled training images, or attempt to infer the correlation or joint probabilities between query images and annotation keywords [15], [18], [19]. Given limited training data, semi-supervised learning methods have been widely used for image annotation [8], [20]. For example, Wang et al. proposed to refine the model-based annotation results with a label similarity graph by following a random walk approach [20], [21]. Although semi-supervised learning approaches can leverage both labeled and unlabeled data, its performance still quite depends on the amount of labeled data. It is usually fairly time-consuming and expensive to collect enough well-labeled training data in order to achieve satisfying performance in large-scale scenarios. Recently, the retrieval-based image annotation paradigm by mining web images has attracted more and more attention [20], [22], [23]. A few studies in this area have attempted to develop efficient content-based indexing and search techniques to facilitate annotation/recognition tasks. For example, Russell et al. developed a large collection of web images with ground truth labels to facilitate object recognition tasks [22]. There are also several studies that aim to address the final annotation process by exploring effective label propagation. For example, Wright et al. proposed a classification algorithm based on sparse representation, which predicts the label information based on the class-based feature reconstruction [24]. Tang et al. presented a sparse graph-based semi-supervised learning (SGSSL) approach to annotate web images [8]. Wu et al. proposed to select heterogeneous features with structural grouping sparsity and suggested a Multi-label Boosting scheme (denoted as "MtBGS" for short) for feature regression, where a group sparse coefficient vector is obtained for each class (category) and further used for predicting new instances [9]. Wu et al. proposed a multi-reference re-ranking scheme (denoted as "MRR" for short) for improving the retrieval process [6].

The third group is face annotation on the collections of personal/family photos. Several studies have mainly focused on the annotation task on collections of personal/family photos [25]–[27], which often contain rich context clues, such as personal/family names, social context, GPS tags, timestamps, etc. In addition, the number of persons/classes is usually quite small, making such annotation tasks less challenging. These techniques usually achieve fairly impressive annotation results. Some techniques have been successfully deployed in commercial applications, e.g., Apple iPhoto,[1] Google Picasa,[2] Microsoft easyAlbum [26], and Facebook face auto-tagging solution.[3]

The fourth group deals with face annotation by mining weakly labeled facial images on the web. A few studies consider a human name as an input query, and mainly aim to refine the text-based search results by exploiting visual consistency of facial images [4], [28], which is closely related to automatic image re-ranking problems. For example, Ozkan and Duygulu proposed a graph-based model for finding the densest sub-graph as the most related result [28]. Following the graph-based approach, Le and Satoh proposed a new local density score to represent the importance of each returned image [29]. The generative approach such as the Gaussian mixture model had also been adopted to the name-based search scheme and achieved comparable results [1]. Recently, a discriminant approach was proposed in [30] to improve the generative approach and avoid to explicitly compute the pairwise similarities in a graph-based approach. Inspired by query expansion [31], the performance of name-based scheme can be further improved by introducing the images of "friends" of the query name. Unlike these studies of filtering the text-based retrieval results, some studies have attempted to directly annotate each facial image with the names extracted from its caption information. For example, Berg et al. proposed a

1. http://www.apple.com/ilife/iphoto/
2. http://picasa.google.com/
3. http://www.facebook.com/

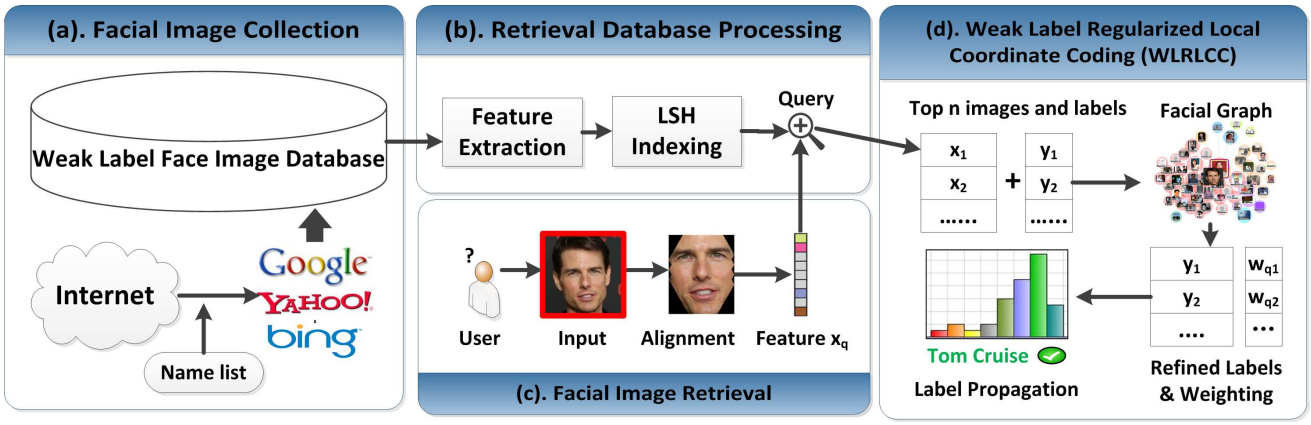


Fig. 1. The system diagram of the Retrieval-based Face Annotation (RBFA) scheme. (a) We collect weakly-labeled facial images from WWW using a web search engine; (b) We perform face detection and alignment, then extract GIST features from the detected faces, and finally apply LSH [10] to index the high-dimensional facial features; (c) A query facial image uploaded by the user is transformed into a feature vector with the same preprocessing step; using our content-based facial image retrieval engine, a short list of the top-n most similar facial images and their associated names are retrieved and passed to the subsequent learning and annotation stage; (d) The proposed WLRLCC scheme is applied to return the final list of (ranked) annotated face names.

probability model which is combined with a clustering algorithm to estimate the relationship between the facial images and the names in their captions [32]. For the facial images and the detected names in the same document, Guillaumin et al. proposed to iteratively update the assignment based on a minimum cost matching algorithm [30]. To further improve the annotation performance, they adopted supervised distance metric learning techniques to grasp the important discriminative features in low dimensional spaces.

Our work is fundamentally different from the previous studies of "*text-based face annotation*" and "*caption-based face annotation*." The key difference lies in two major aspects. First, our work aims to solve the generic content-based face annotation problem, where the facial images are directly used as the input query images, and the task is to return the corresponding names in the query images. Second, for the top-ranking similar candidate images, the proposed WLRLCC algorithm aims to achieve a new discriminative feature representation and refined label information in a unified learning scheme. However, the caption-based annotation scheme only considers the assignment between the facial images and the names appearing in their corresponding surrounding text. Therefore the caption-based annotation scheme is only suitable for the scenario where both images and their captions are available, and cannot be employed in our RBFA framework due to the shortage of caption information. In addition, this work is related to our previous work in [3], [33], which proposed an unsupervised label refinement (ULR) technique to enhance the label information over the entire facial image database. This paper is different from the ULR algorithm in several aspects. First, it is difficult for the ULR to handle huge databases due to its high computational cost, while, the WLRLCC algorithm is only applied to a short list of the most similar images for each query image and therefore is independent of the entire retrieval database size. Second, ULR focuses on refining the label information over the whole database, and its simple majority voting scheme may not be effective enough in exploiting the short list of the most similar images. In contrast, the proposed WLRLCC algorithm comprehensively resolves this problem by fully exploiting the short list of top similar images via a unified optimization scheme, which learns both sparse features and enhanced labels. In addition, according to recent research work [34], [35], the face annotation performance could be further improved by combining the proposed WLRLCC algorithm with supervised inductive learning techniques in a unified framework.

The proposed learning methodology for WLRLCC is partially inspired by several groups of existing work in machine learning, including local coordinate coding [36], [37], graph-based semi-supervised learning [38], and multi-label learning [39].

3 Retrieval-based Face Annotation Framework

In this section, we briefly introduce the proposed Retrieval-based Face Annotation (RBFA) paradigm. Figure 1 illustrates the proposed framework, which consists of the following four major stages: (i) data collection of facial images from WWW; (ii) facial image preprocessing and high-dimensional facial feature indexing; (iii) content-based facial image retrieval for a query facial image; (iv) face annotation by the proposed Weak Label Regularized Local Coordinate Coding (WLRLCC) algorithm. The details of each stage are described as follows.

The first stage, as shown in Figure 1(a), collects a database of weakly labeled facial images, which can be crawled from the web. Specifically, we can choose a list of desired human names and submit them to existing web search engines (e.g., Google) to crawl their related web facial images. As the output of this crawling process, we obtain a collection of web facial images, each of them is associated with some human names.

The collected web images are useful for many applications, e.g. object classification [40] and animal classification [41]. In our framework, we use them as the retrieval database in a data-driven scheme. Since the labels of these web images are usually noisy, we refer to such web facial images with noisy names as weakly labeled facial images.

The second stage, as shown in Figure 1(b), pre-processes the weakly label facial image database, including face detection, face alignment, facial feature representation, and high dimensional feature indexing. For facial region detection and alignment, we adopt OpenCV and the unsupervised face alignment technique DLK proposed in [42]. Generally, any facial feature which is represented in the vector format could be used. In our system, we extract the GIST features [43] to represent the aligned facial regions. Finally, we apply the Locality-Sensitive Hashing (LSH) to index the facial features in our solution [10].

The first two stages must be done before annotating a query facial image. The next two stages are related to online processes of annotating a query facial image. As shown in Figure 1(c), given a query facial image, we employ a similar face retrieval process (kNN with L2 distance) to find a short list of the most similar faces (e.g., top-n similar faces) from the indexed face databases using the LSH technique.

After obtaining a set of the similar faces for the query image, the last stage applies the proposed WLRLCC algorithm for name annotation, as shown in Figure 1(d). Specifically, WLRLCC learns local coordinate coding for each of the similar facial images and enhances the weak label matrix via an iterative optimization process. Based on the learning results, a sparse reconstruction algorithm is applied to perform the final face name annotation. Next, we present the details of WLRLCC.

4 Weak Label Regularized Local Coordinate Coding (WLRLCC)

In this section, we present the proposed WLRLCC for the face annotation task based on a list of the similar facial images.

4.1 Preliminaries

Throughout the paper, we denote the matrices by upper case letters, e.g. X, D; we denote the vectors by bold lower case letters, e.g. $\mathbf{x}, \mathbf{x}_i$; we denote the scalars by the normal letters, e.g. x_i, x_{ij}, X_{ij}, where x_i is the i-th element of the vector $\mathbf{x}$, x_{ij} is the j-th element of the vector $\mathbf{x}_i$, and X_{ij} is the element in the i-row and j-column of the matrix X.

Consider a query facial image $\mathbf{x}_q \in \mathbb{R}^d$ in a d-dimensional feature space, which is associated with an unknown class denoted by a class label vector $\mathbf{y}_q$. Assume the n retrieval results of the query image $\mathbf{x}_q$ are $\{(\mathbf{x}_i, \mathbf{y}_i)|i = 1, 2, \ldots, n\}$, where $\mathbf{y}_i \in \{0, 1\}^m$ is the name label vector of its corresponding facial image $\mathbf{x}_i$ and $\|\mathbf{y}_i\|_0 = 1$, and m is the total number of classes (names) among all the similar facial images. Let $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ denote the feature matrix of the retrieval results. We represent the initial name information with a label matrix $\tilde{Y} \in \mathbb{R}^{n\times m}$, where $\tilde{Y}_{i\star} = \mathbf{y}_i$, the i-th row of matrix, denotes the class label vector for $\mathbf{x}_i$.

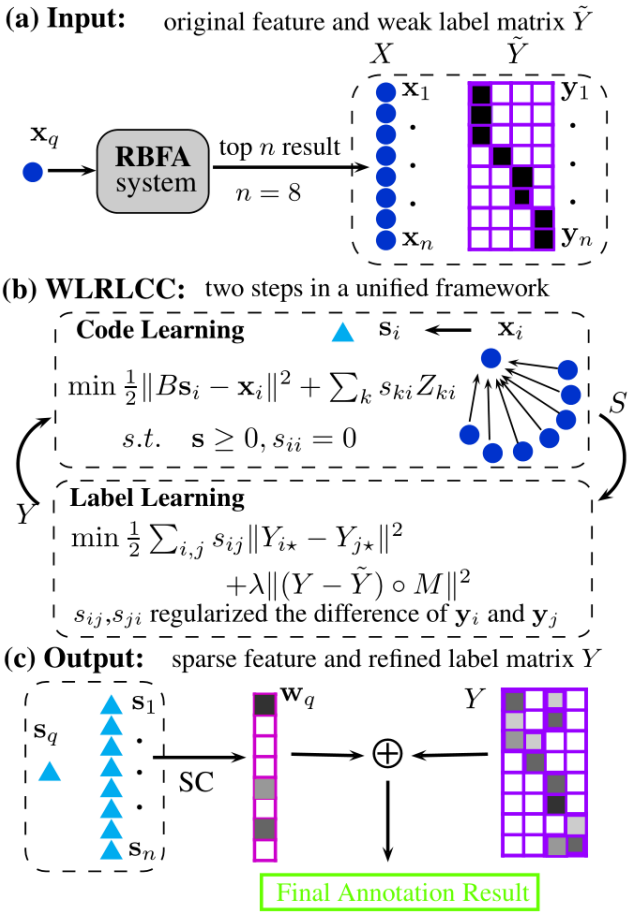


Fig. 2. Illustration of the proposed WLRLCC process. In this example, we show the top $n = 8$ retrieved images with $m = 4$ candidate names for query image $\mathbf{x}_q$: (a) gives the input data; (b) illustrates the two key steps of the WLRLCC process; (c) shows the output coding result and final annotation scheme.

4.2 Problem Formulation

4.2.1 Sparse Features via Local Coordinate Coding

Sparse representation has been successfully applied in many applications. Recently, it has been argued from both theoretical and empirical perspectives that it is important to exploit the locality information for the linear embedding of high dimensional data on the manifold in Local Coordinate Coding (LCC) [36]. In detail, the goal of LCC is to develop a new representation of feature vectors in which each feature vector is described as a linear combination of several nearby items among the basis vectors. The sparse representation of each vector is locally adapted, hence the name of "Local Coordinate Coding".

In our framework, we adopted the LCC algorithm for the coding step of WLRLCC. Specifically, we reconstruct the locality coding $\mathbf{s}_i$ of the i-th facial image $\mathbf{x}_i$ in the retrieval results with the dictionary $B = [X, I] \in \mathbb{R}^{d\times(n+d)}$ as follows:

$$\min_{\hat{\mathbf{s}}_i} \frac{1}{2}\|\mathbf{x}_i - B\hat{\mathbf{s}}_i\|^2 + \lambda \sum_{k=1}^{n+d} |\hat{s}_{ik}| \|B_{\star k} - \mathbf{x}_i\|^2 \text{ s.t. } \hat{s}_{ii} = 0; \quad (1)$$

where $\mathbf{s}_i$ is a sub-vector of $\hat{\mathbf{s}}_i$ with its first n elements: $\hat{\mathbf{s}}_i = [\mathbf{s}_i, \xi_i]$, ξ_i is related to the noise information in our framework, I is the identity matrix, and $B_{\star k}$ is the k-th column of dictionary B. A similar leave-one-out representation model is adopted in sparse subspace clustering [44]. For simplicity, we also denote the coding problem for $\mathbf{x}_i$ by $e(\hat{\mathbf{s}}_i; \mathbf{x}_i)$.

In order to make sure the neighbor samples of each $\mathbf{x}_i$ are included in the retrieval results X, we apply query expansion for the query image $\mathbf{x}_q$ by including the top n' nearest samples of each query result into the final retrieval database. In our experiment we simply fix n' to 3 for all cases. One difference between our coding algorithm and the original LCC algorithm is that we directly construct the dictionary B instead of learning it by optimization, which is proposed in [45] and achieves excellent experiment results on the face recognition task. This method is also well known as "*cross-and-bouquet*" model for dense error correction [46].

According to our formulation, the j-th element of $\mathbf{s}_i$ tries to measure the similarity between the facial images $\mathbf{x}_i$ and $\mathbf{x}_j$, so a negative value is contrary to this intuitive notion and doesn't make sense in our scenario. As a result, we introduce an extra non-negative constraint to the previous formulation following Non-Negative Sparse Coding [37], in which all the elements in $\hat{\mathbf{s}}_i$ are forced to be non-negative: $\hat{s}_{ij} \geq 0, j = 1, 2, \ldots, n$.

Finally, we can give the formulation for all the n facial images in the retrieval results:

$$E_1(\hat{S}; X) = \sum_{i=1}^{n} e(\hat{\mathbf{s}}_i; \mathbf{x}_i). \tag{2}$$

where $\hat{S} \in \mathbb{R}^{(n+d)\times n} = [S; \Xi]$, $S \in \mathbb{R}^{n\times n}$ is the non-negative local coordinate coding of X, and $\Xi \in \mathbb{R}^{d\times n}$ is the noise matrix.

4.2.2 Weak Label Enhancement

The previous formulation shows that the j-th local coefficient s_{ij} of facial image $\mathbf{x}_i$ essentially encodes the locality information between $\mathbf{x}_i$ and $\mathbf{x}_j$, where $j \neq i$. Specifically, a larger value of s_{ij} indicates that $\mathbf{x}_j$ is more representative of $\mathbf{x}_i$. In addition, from a view of graph-based semi-supervised learning, for any two facial images, the smaller their local distance, the more likely they should belong to the same person. As a result, a larger value of s_{ij} implies that the name labels of $\mathbf{x}_i$ and $\mathbf{x}_j$ are more likely to be the same.

Based on the above motivation, we can give the following formulation to enhance the initial weak label matrix $\tilde{Y}$ as follows:

$$\min_{Y\geq 0} \frac{1}{2} \sum_{i,j} s_{ij} \|Y_{i\star} - Y_{j\star}\|^2 + \lambda \|(Y - \tilde{Y}) \circ M\|_{\mathrm{F}}^2 \tag{3}$$

where $M = [h(\tilde{Y}_{ij})]$ is an indicator matrix: $h(x) = 1$ if $x > 0$ and otherwise $h(x) = 0$, and $\circ$ denotes the Hadamard product of two matrices. In the above objective function, the first term enforces that the class labels of two facial images $\mathbf{x}_i$ and $\mathbf{x}_j$ to be similar if the local sparse coefficient s_{ij} is large, and the second term is a regularization term that prevents the refined label matrix being deviated too much from the initial weak matrix. Since the initial label matrix is noisy and incomplete, we apply the regularization of the second term on only these nonzero elements in $\tilde{Y}$.

In general, the optimal solution to the problem in Eq. (3) is dense; however, the ideal true label matrix is often very sparse. We introduce some convex sparsity constraints, i.e., $\|Y_{i\star}\|_1 \leq \varepsilon, \varepsilon \geq 1$, where $i = 1, 2, \ldots, n$ (we choose $\varepsilon = 1$ in this work). These constraints are included to limit the number of name labels assigned to each facial image.

4.2.3 Weak Label Regularized Local Coordinate Coding

The above two optimization tasks of "sparse feature learning" and "label enhancement" are performed separately. Specifically, the sparse features S are first learned from the optimization in Eq. (1), and then used by the optimization in Eq. (3) to refine the label matrix Y. To better exploit the potential of the two learning approaches, we propose the Weak Label Regularized Local Coordinate Coding (WLRLCC) scheme, which aims to reinforce the two learning tasks via a unified optimization framework. Specifically, the optimization of WLRLCC can be formulated as follows:

$$\begin{aligned} &\min_{\hat{S},Y} E_1(\hat{S}; X) + E_2(Y, S) = \min_{\hat{S},Y} \frac{1}{2}\|B\hat{S} - X\|_{\mathrm{F}}^2 + \\ &\lambda_1 tr(\mathbf{1} \cdot (\hat{S} \circ V)) + \lambda_2 tr(Y^\top L Y) + \lambda_3 \|(Y - \tilde{Y}) \circ M\|_{\mathrm{F}}^2 \\ &\text{s.t. } \hat{S}_{ii} = 0, \|Y_{i\star}\|_1 \leq 1, i = 1, 2, \ldots, n, \ \hat{S} \geq 0, Y \geq 0 \end{aligned} \tag{4}$$

where $V \in \mathbb{R}^{(n+d)\times n}$, $V_{ij} = \|B_{\star i} - X_{\star j}\|^2$, $L = D - S$, D is a diagonal matrix, with $D_{ii} = \frac{\sum S_{i\star} + \sum S_{\star i}}{2}$, $Y \in \mathbb{R}^{n\times m}$, $\mathbf{1}$ is all-one-element matrix with dimension $n \times (n + d)$, and $tr(\cdot)$ denotes a trace function. In the above, $\lambda_2 tr(Y^\top L Y)$ is a label smoothness regularizer which connects the label matrix and the sparse features.

In Figure 2, we show the whole process of the proposed WLRLCC algorithm with a simple example by using the top $n = 8$ retrieval results and $m = 4$ candidate names.

4.3 Optimization

The optimization problem in Eq. (4) is generally non-convex. To solve this challenging optimization, we propose to solve $\hat{S}$ and Y alternatively by iteratively solving two optimization steps: (1) *Code Learning*, and (2) *Label Learning*. The step of updating $\hat{S}$ can be transformed into a weighted non-negative sparse coding problem, and the step of updating Y is a quadratic programming problem.

4.3.1 Code Learning

By first fixing the label matrix Y and ignoring the constant terms, the optimization problem in Eq. (4) can be reformulated as follows:

$$\begin{aligned} Q_Y(\hat{S}) = \min_{\hat{S}} \frac{1}{2}\|B\hat{S} - X\|_{\mathrm{F}}^2 + tr(\mathbf{1} \cdot (\hat{S} \circ Z)) \\ \text{s.t. } \hat{S}_{ii} = 0, i = 1, 2, \ldots, n, \text{and } \hat{S} \geq 0 \end{aligned} \tag{5}$$

where $Z = \lambda_1 V + \lambda_2 U$, $U \in \mathbb{R}^{(n+d)\times n}$, for all $j = 1, 2, \ldots, n$, if $i \leq n$, $U_{ij} = \frac{1}{2}\|Y_{i\star} - Y_{j\star}\|^2$; otherwise $U_{ij} = 0$. The other variables follow the same definitions in Eq.(4).

The optimization problem in Eq. (5) can be further separated into a series of sub-problems for each coding coefficient $\hat{S}_{\star i}$ of facial image $X_{\star i}$. Each sub-problem is a weighted non-negative sparse coding problem, which can be written into the following optimization:

$$\min_{\hat{\mathbf{s}}\geq 0} \frac{1}{2}\|B\hat{\mathbf{s}} - X_{\star i}\|^2 + \sum_{k=1}^{n+d} \hat{s}_k Z_{ki} \quad \text{s.t.} \quad \hat{s}_i = 0. \tag{6}$$

In our approach, we adopt the Fast Iterative Shrinkage and Thresholding Algorithm (FISTA) [47], a popular and efficient algorithm for the linear inverse problem that has been already implemented for sparse learning in [48]. Since we aim to solve the problem related to only the top-n images in the retrieval results, where n is usually small, FISTA is efficient enough for our application.

4.3.2 Label Learning

Similarly, by fixing $\hat{S}$ and ignoring the constant terms, the optimization problem in Eq.(4) can be reformulated as follows:

$$\begin{aligned} Q_{\hat{S}}(Y) = \min_{Y} tr(Y^\top LY) + \lambda\|(Y-\hat{Y})\circ M\|_{\mathrm{F}}^2 \\ \text{s.t. } Y \geq 0, \|Y_{i\star}\|_1 \leq 1, i = 1,2,\ldots,n. \end{aligned} \tag{7}$$

where $\lambda = \frac{\lambda_3}{\lambda_2}$, and the other variables follow the same definitions in Eq.(4). The optimization problem is a quadratic program with a series of closed and convex constraints (ℓ_1 ball). In order to efficiently solve this problem, we first vectorize the label matrix $Y \in \mathbb{R}^{n\times m}$ into a column vector $\bar{\mathbf{y}} = vec(Y) \in \mathbb{R}^{(n\cdot m)\times 1}$, and then rewrite the previous optimization problem as follows:

$$Q(\bar{\mathbf{y}}) = \min_{\bar{\mathbf{y}}\geq 0} \bar{\mathbf{y}}^\top \Psi \bar{\mathbf{y}} + \mathbf{c}^\top \bar{\mathbf{y}}, \text{ s.t. } \forall i, \sum_{k=0}^{m-1} \bar{y}_{k\cdot n+i} \leq 1. \tag{8}$$

where $\Psi = I_m \otimes L^\top + \lambda R$, I_m is an identity matrix with dimension $m\times m$, $R = diag(vec(M))$, $\mathbf{c} = -2\lambda R^\top \cdot vec(\tilde{Y})$, and $\bar{y}_j$ is the j-th element in $\bar{\mathbf{y}}$. In order to solve this problem, we employ the multi-step gradient scheme [47] and the efficient Euclidean projection algorithm [49].

Specifically, in order to achieve the optimal solution $\bar{\mathbf{y}}^\star$, we recursively update two sequences $\{\bar{\mathbf{y}}^{(k)}\}$ and $\{\mathbf{z}^{(k)}\}$ in the multi-step gradient scheme, where k is the iteration step. Commonly at each iteration k, the variance $\mathbf{z}^{(k)}$ is named as the search point and used to construct the combination of the two previous approximate solutions $\bar{\mathbf{y}}^{(k-1)}$ and $\bar{\mathbf{y}}^{(k-2)}$. In our problem, the sub-block problem is defined as:

$$\bar{\mathbf{y}}^{(k+1)} = \arg\min_{\bar{\mathbf{y}}\geq 0} \frac{t}{2}\|\bar{\mathbf{y}} - \mathbf{v}\|^2 \text{ s.t. } \forall i, \sum_{k=0}^{m-1} \bar{y}_{k\cdot n+i} \leq 1. \tag{9}$$

where $\mathbf{v} = \mathbf{z}^{(k)} - \frac{1}{t}\mathbf{g}$ and $\mathbf{g} = 2\Psi\mathbf{z}^{(k)} + \mathbf{c}$, t can be a fixed positive constant or searched in a backtracking step [47]. The key problem is to solve Eq. (9) efficiently. In our work, we employ the Euclidian projection algorithm [49], which has been shown as an efficient solution to such kind of ℓ_1 ball constrained problem with linear time complexity of $O(n)$ as compared with other algorithms of $O(n\log(n))$.

Further, in order to apply the Euclidian projection algorithm, we split $\bar{\mathbf{y}}$ into a series of m-dimensional sub-vectors with $\bar{\mathbf{y}}^i = [\bar{y}_{k\cdot n+i})]_{k=0}^{m-1}$, where $i = 1,2,\ldots,n$, and similarly we can split the vector $\mathbf{v}$. Specifically, each optimal solution $\bar{\mathbf{y}}^{i\star}$ for sub-vector $\bar{\mathbf{y}}^i$ can be achieved by solving the following problem:

$$\bar{\mathbf{y}}^{i\star} = \arg\min_{\bar{\mathbf{y}}^i\geq 0} \frac{t}{2}\|\bar{\mathbf{y}}^i - \mathbf{v}^i\|^2 \quad \text{s.t.} \quad \|\bar{\mathbf{y}}^i\|_1 \leq 1. \tag{10}$$

The problem in Eq. (10) has a non-negative constraint, which is different from the one in [49]. The optimal solution to this problem is given as:

$$\bar{y}_j^{i\star} = h(v_j^i)\max(|v_j^i| - \lambda^\star, 0), j = 1,2,\ldots,m. \tag{11}$$

where $h(\cdot)$ is the indicator function in Eq. (3), $\bar{y}_j^i$ and v_j^i are the j-th elements in $\bar{\mathbf{y}}^i$ and $\mathbf{v}^i$, and $\lambda^\star$ is the optimal solution for the dual form of Eq. (10). Suppose $S = \{j|v_j^i \geq 0\}$, the value of $\lambda^\star$ can be computed as follows:

$$\lambda^\star = \begin{cases} 0 & \sum_{k\in S} v_k^i \leq 1, \\ \bar{\lambda} & \sum_{k\in S} v_k^i > 1. \end{cases} \tag{12}$$

where $\bar{\lambda}$ is the unique root of function $f(\lambda) = \sum_{k\in S}\max(|v_k^i| - \lambda, 0) - 1$, which is continuous and monotonically decreasing in $(-\infty, \infty)$. The root $\bar{\lambda}$ can be achieved by a bisection search in linear time complexity of $O(dim(\bar{\mathbf{y}}^i))$. It is well-known that the multi-step gradient scheme converges at $O(\frac{1}{k^2})$, where k is the iteration step, indicating our optimization can be solved efficiently.

4.4 Face Name Annotation

After obtaining the sparse coding matrix S and the enhanced label matrix Y, the last key step of our framework is to perform the face name annotation. First, the query facial image $\mathbf{x}_q$ is also projected into the local coordinate coding space. Specifically, the new coding $\mathbf{s}_q$ w.r.t. $\mathbf{x}_q$ can be achieved by solving the following optimization problem:

$$[\mathbf{s}_q; \xi_q] = \arg\min_{\hat{\mathbf{s}}\geq 0} \frac{1}{2}\|B\hat{\mathbf{s}} - \mathbf{x}_q\|^2 + \lambda\sum_k \hat{s}_k\|B_{\star k} - \mathbf{x}_q\|^2 \tag{13}$$

where the parameter setting is the same as the previous WLRLCC algorithm.

Although the manifold structure is supposed to be well-fitted in the new local coordinate space, it is still unsuitable to directly use a one-vs-one similarity measure for face name annotation. For the second step, we employ a sparse reconstruction scheme in the local coordinate coding space to recover the potential weighting vector $\mathbf{w}_q$ for name annotation in the label space. The optimization problem is given as follows:

$$\min \|\mathbf{w}_q\|_1 \quad \text{s.t.} \quad \mathbf{s}_q = S_{I_n}\cdot \mathbf{w}_q \tag{14}$$

where $S_{I_n} = S + I_n$, S is the local coordinate codes obtained from the previous step and I_n is an $n\times n$ identity matrix.

Finally, the label vector $\mathbf{y}_q$ can be directly computed by:

$$\mathbf{y}_q = Y^\top \cdot \mathbf{w}_q$$

The value $y_{qk}, k \in \{1, 2, \ldots, m\}$ measures the confidence of the k-th name being assigned to the query facial image $\mathbf{x}_q$. Thus, a name with a large confidence value will be ranked on the top position for the final annotation result.

5 Offline Approximation For WLRLCC

To reduce the computational cost time of the proposed WLRLCC algorithm, one straight-forward solution is to adopt the PCA dimension reduction techniques over the original high dimensional feature space. The smaller the new dimension space is, the less time it takes for WLRLCC algorithm. The key limitation for the PCA-based approximation is the information loss during dimension reduction may affect the final annotation performance.

In this section, we propose an offline approximation scheme for WLRLCC (AWLRLCC for short), which can both significantly reduce the running-time and achieve comparable results. In detail, we first pre-compute the local coordinate coding for each facial image in the retrieval database with its own n neighborhoods by Eq. (13) and save all the coding results. Then, in the annotation step, for each query image we can directly reconstruct the sparse features of its n nearest instances based on the offline coding results without extra computational costs.

We denote by $\mathfrak{D} = \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N\}$ the whole retrieval database, where N is the total number of facial images. For each facial image $\mathbf{d}_i$, we denote by $\mathfrak{N}(\mathbf{d}_i)$ its n similar images in the retrieval database except itself. Then, we can achieve its encoding result $\mathbf{c}_i$ with Eq. (13), where the dictionary is constructed as $[\mathfrak{N}(\mathbf{d}_i), I]$. For the query image $\mathbf{x}_q$, its n most similar facial images $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] = [\mathbf{d}_{\mathfrak{I}_1}, \mathbf{d}_{\mathfrak{I}_2}, \ldots, \mathbf{d}_{\mathfrak{I}_n}]$ is also a subset of the whole retrieval database $\mathfrak{D}$, where $\mathfrak{I}$ is an index vector for the whole retrieval database. In the offline computation step, the sparse coding of $\mathbf{x}_i$ (or $\mathbf{d}_{\mathfrak{I}_i}$) is $\mathbf{c}_{\mathfrak{I}_i}$ and its j-th element $c_{\mathfrak{I}_i j}$ measures the weight of the j-th item in the nearest neighbor set $\mathfrak{N}(\mathbf{d}_{\mathfrak{I}_i})$. As a result, the key problem in our approximation scheme is to construct $\mathbf{x}_i$'s (or $\mathbf{d}_{\mathfrak{I}_i}$'s) new coding result $\mathbf{s}_i$ for WLRLCC based on the previous coding results. The new coding results should share the same dictionary $\bar{X}$, which is constructed with all the neighbor sets $\mathfrak{N}(\mathbf{d}_{\mathfrak{I}_i})$ of $\mathbf{d}_{\mathfrak{I}_i}, i = 1, 2, \ldots, n$: $\bar{X} = \bigcup_{i=1}^{n} \mathfrak{N}(\mathbf{d}_{\mathfrak{I}_i}) \bigcup X$ where the number of dictionary items in $\bar{X}$ is $\bar{n}, \bar{n} \geq n$. Suppose the function $\Upsilon(\bar{X}, \mathbf{d})$ returns the index value of the vector $\mathbf{d}$ in the dictionary $\bar{X}$, the new coding result $\mathbf{s}_i \in \mathbb{R}^{\bar{n}}$ of $\mathbf{x}_i$ can be constructed based on $\bar{X}$ with the following rules: i) For $j = 1, 2, \ldots, n$, $s_{i\Upsilon(\bar{X}, \mathfrak{N}(\mathbf{d}_{\mathfrak{I}_i})_j)} = c_{\mathfrak{I}_i j}$. ii) The left items in $\mathbf{s}_i$ are set as zero. Similar to the general WLRLCC algorithm, the local coordinate coding result $\mathbf{s}_q$ of $\mathbf{x}_q$ can be achieved by solving the problem in Eq. (13), where $B = [\bar{X}, I]$.

Figure 3 shows an example for AWLRLCC, where $\otimes$ denotes the query image $\mathbf{x}_q$, $\bullet$ denotes the database images, and $n = 3$. As shown in Figure 3(b), the nearest neighbor set of $\mathbf{x}_q$ is $\mathfrak{N}(\mathbf{x}_q) = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} = \{\mathbf{d}_4, \mathbf{d}_3, \mathbf{d}_5\}$, and the index vector is $\mathfrak{I} = [4, 3, 5]$. Based on the reconstructed dictionary $\bar{X} = \mathfrak{N}(\mathbf{d}_4) \bigcup \mathfrak{N}(\mathbf{d}_3) \bigcup \mathfrak{N}(\mathbf{d}_5) \bigcup \mathfrak{N}(\mathbf{x}_q) = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_9]$, the value mapping rules between the pre-computed coding $\mathbf{c}_i$ and the reconstructed coding $\mathbf{s}_i$ are illustrated in Figure 3(c). E.g., for the facial image $\mathbf{d}_4$ with $j = 1$, the corresponding dictionary item is $\mathbf{d}_5$ and $\Upsilon(\bar{X}, \mathbf{d}_5) = 5$, so that $s_{45} = c_{41}$.

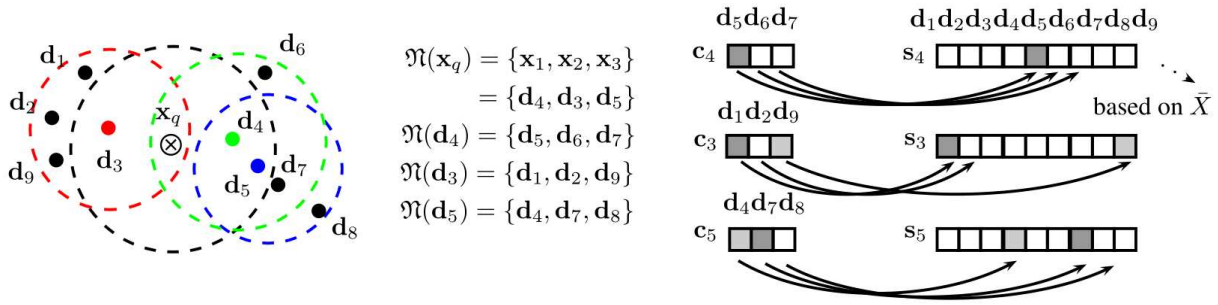
6 Retrieval Database Construction

Constructing a proper retrieval database is a key step for a retrieval-based face annotation system. In literature, some web facial image databases are available from the previous research work, e.g., LFW [13],[4] Yahoo!News [30],[5] and FAN-Large [50][6]. One issue with these three databases is that although the number of persons is quite large as compared to regular face databases, the number of images for each person is quite small, making them inappropriate for retrieval-based face annotation tasks. For example, the LFW database has a total of $13,233$ images for $5,749$ people, in which each person on average has no more than 3 facial images. Unlike these database, another database in literature, PubFig [51][7], is more appropriate. It has 200 persons and $58,797$ images, as constructed by collecting online news sources. Due to the copyright issue, only image URL addresses were released from the previous work. Since some URL links are no longer available, only a total of $41,609$ images were collected by our crawler. For each downloaded image, we crop the face image according the given face position rectangle and resize all the face images into the same size (128×128).

In order to evaluate the retrieval-based face annotation scheme in even larger web facial image databases, we construct two new databases: (i) "WLF: Weakly Labeled Faces on the web", which contains $6,025$ famous western celebrity and $714,454$ web facial images in total; (ii) "WLF-cn", an Asian web facial image database, which consists of $1,200$ persons and $126,070$ images in total. Since there is almost no overlap between "WLF" and "WLF-cn", we adopt both to evaluate the generalization of the system on different types of databases. In general, there are two main steps to build a weakly-labeled web facial image database: (i) Construct a name list of popular persons; and (ii) Query the existing search engine with the names, and then crawl the web images according to the retrieval results. To facilitate future research by other researchers, we have made our databases and their related information publicly available[8].

In the first step, for "WLF" database, we collect a name list consisting of $6,025$ names downloaded from the website: **IMDb**,[9] which covers the actors and actresses who were born between 1950 and 1990. Specifically, we collect these names with the billboard: "Most Popular People Born In yyyy", where yyyy is the year of birth. e.g. the webpage[10] presents all the actor and actresses who were born in 1975 in the popularity order. For the "WLF-cn" database, we collected a name list consisting of 600 Asian male celebrities and 600 Asian female celebrities from two web pages.[11] In the second

4. http://vis-www.cs.umass.edu/lfw/
5. http://lear.inrialpes.fr/people/guillaumin/data.php
6. http://www.vision.ee.ethz.ch/~calvin/fan-large/
7. http://www.cs.columbia.edu/CAVE/databases/PubFig/
8. http://stevenhoi.org/WLF/
9. http://www.imdb.com
10. http://www.imdb.com/search/name?birth_year=1975
11. http://goo.gl/XsqtV and http://goo.gl/5uCS5



(a) Database and Query (b) Point Relationship (c) Original Features → New Features

Fig. 3. An example for AWLRLCC. (a) shows the database images • ($\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_9\}$) and query image ⊗ ($\mathbf{x}_q$); (b) shows the nearest neighbor sets of $\mathbf{x}_q$, $\mathbf{d}_4$, $\mathbf{d}_3$, and $\mathbf{d}_5$; (c) shows the value mapping rules between the pre-computed coding $\mathbf{c}_i$ and the reconstructed coding $\mathbf{s}_i$ based on the new dictionary $\bar{X}$. As an example, for the facial image $\mathbf{d}_4$ with $j = 1$, the corresponding dictionary item is $\mathbf{d}_5$ and $\Upsilon(\bar{X}, \mathbf{d}_5) = 5$, so that $s_{45} = c_{41}$

step, we submit each name in the aforementioned lists as a query to search for the related web images by Google image search engine. The top 200 retrieved web images are crawled automatically. After that we use the Viola-Jones algorithm to detect the faces and use the Deformable Lucas-Kanade (DLK) algorithm [42] to align facial images into the same well-defined position. The web images in which no faces are detected are ignored directly. As a result, we collected $714,454$ facial images of $6,025$ persons in the "WLF" database, and $126,070$ images of $1,200$ persons in the "WLF-cn" database.

For evaluation, we built two "evaluation(test) datasets" by randomly choosing 119 names from the name list of "WLF" and 110 names from the name list of "WLF-cn." We issued each of these names as a text query to Google image search and crawled the web images with the top 200-th to 400-th search results. Note that we did not consider the top 200 retrieval results since they had already been used in the retrieval database. This aims to examine the generalization performance of the proposed technique for the unseen facial images. We requested our staff to manually examine the retrieved facial images and remove those irrelevant facial images for each name. As a result, the evaluation dataset for "WLF" contains $1,600$ facial images, and the evaluation dataset for "WLF-cn" contains $1,300$ facial images.

In the following experiments, we denote the "WLF" database as "WDB-600K" and denote the "WLF-cn" database s "ADB-P100." In order to evaluate the effect of the number of persons, we build four subsets of the whole database "WDB-600K": "WDB-040K", "WDB-100K", "WDB-200K", and "WDB-400K." For example, WDB-040K includes only all the facial images of 400 celebrities, and there are $53,448$ images in WDB-040K. In order to evaluate the effect of the number of images per person, we build a subset "ADB-P050" for the whole database ADB-P100, which means that only half of the facial images per person are collected into ADB-P050 and there are about 50 facial images per person on average.

7 Experimental Results

7.1 Experimental Testbed

We compare the proposed WLRLCC against a baseline algorithm based on Soft-Max Weighted majority voting ("SMW") and five existing algorithms for face/image annotation, including "S-Recon" [24], "SGSSL" [8], "LCC" [36], "MtBGS" [9], and "MRR" [6]. Most of these algorithms were proposed in recent years, as briefly introduced in Section 2. For facial feature representation, we adopt the 512-dimensional GIST features [43] for both "WDB" and "ADB" databases. For the "PubFig" database, we adopt the 2891-dimensional LBP [52] features by dividing the face images into 7×7 blocks, which is further projected into a lower 500-dimensional feature space using Principal Component Analysis (PCA). More details on feature representation are discussed in Section 7.7. To evaluate the annotation performances, we adopt the *hit rate* at the top-t annotated results as the performance metric, which measures the likelihood of having the true label among the top-t annotated names for a query facial image.

Furthermore, we discuss the cross validation issue. For all the algorithms, we only conducted it in the experiments on the WDB-040K; as a result, the parameters with the best average hit rate results are directly used in the following experiments based on the other databases. Specifically, we randomly divide the query set into two parts of equal size, in which one part is used as the validation set to find the optimal parameters by a grid search, and the other part is used for performance evaluation. Such a procedure is repeated 10 times and the average performance is computed over the 10 trials.

7.2 Impact of the Top-n and Top-t Settings

This experiment aims to evaluate the impact on the final annotation performance under varied settings by varying the values of n and t for Top-n retrieved facial images and Top-t annotated names.

7.2.1 *Impact of the Top-t Settings*

TABLE 1 shows the hit rate performance of the baseline SMW algorithm ($n = 10$) and the WLRLCC algorithm ($n = 40$)

under different settings of t values based on the WDB-040K database.

TABLE 1
The hit rate performance of SMW ($n = 10$) and WLRLCC ($n = 40$) with different Top-t

	t = 1	t = 2	t =3	t=4	t = 5
SMW	0.6090	0.7200	0.7631	0.7881	0.8006
WLRLCC	0.7649	0.8006	0.8254	0.8394	0.8514

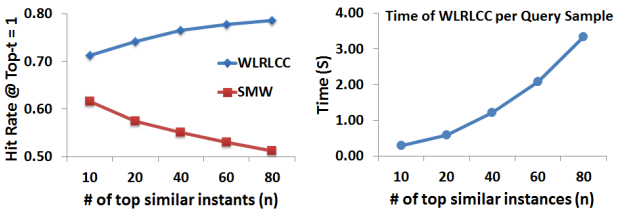
We observe that given a fixed n value, increasing the value of t generally leads to a better annotation performance and the improvement becomes marginal when t is large. This observation is not surprising as assigning more names certainly gives a better chance to hit the relevant name. Specially, for $t = 1$ the hit rate measures the precision of the annotation system, and for a large t value, the hit rate is very similar to the recall metric. In practice, we mainly focus on a small value of t since users usually would not be interested in a long name list.

7.2.2 *Impact of the Top-n Settings*

Figure 4 (a) and TABLE 2 show the hit rate performance of the baseline SMW algorithm and the WLRLCC algorithm under different settings of n values based on the WDB-040K database. For both algorithms, we fix t to 1. The average running time of WLRLCC algorithm for one query sample is presented in Figure 4 (b) and the last row of TABLE 2.

TABLE 2
The hit rate performance of SMW and WLRLCC with different Top-n, and the running time of WLRLCC.

	$n = 10$	$n = 20$	$n = 40$	$n = 60$	$n = 80$
SMW	0.6090	0.5738	0.5513	0.5300	0.5119
WLRLCC	0.7226	0.7441	0.7649	0.7694	0.7760
Time(s)	0.295	0.579	1.205	2.077	3.337



(a) (b)

Fig. 4. (a) The hit rate performance of SMW and WLRLCC with different Top-n,(b) The average running time of WLRLCC for one query sample.

Several observations can be drawn from the results. First of all, for the baseline SMW algorithm, when increasing the value of n, the annotation performance generally decreases. However, for the proposed WLRLCC algorithm, increasing n leads to improve the annotation performance. It is not difficult to explain this observation. In particular, for any two facial images, the smaller their local distance is, the more likely they belong to the same person; as a result, for the top n similar instances, a small n value results in a high precision and a low recall, while a large n value often produces a high recall and a low precision. Because SMW is a simple weighted majority voting algorithm without any re-ranking, it favors the retrieval result of a high precision; this is why a small parameter n gets a better performance. On the other hand, the WLRLCC algorithm adopts a re-ranking scheme and typically prefers a high recall of the retrieval results such that a potentially larger pool of relevant facial images can be exploited for re-ranking. Second, we observe that the running time of the WLRLCC algorithm increases when increasing value of n due to the cost of encoding more instances and a larger dictionary.

The above observations are beneficial to determining the appropriate parameters in our rest experiments. For SWM, the parameter n for Top-n retrieved images is set to 10. For the other algorithms that usually carry a re-ranking or sample selection step, we thus set the parameter n to 40, as a trade-off between annotation performance and running time.

7.3 Evaluation of Auto Face Annotation on "WDB", "ADB", and "PubFig"

In this experiment, we compare the proposed WLRLCC algorithm against the other algorithms on the three databases: WDB-040K, ADB-P100, and PubFig.

For the two weakly labeled databases WDB-040K and ADB-P100, the average annotation performance with Top-t=1 are shown in TABLE 3. As the Asian celebrity database (ADB-P100) is independent of the Western celebrity database (WDB-040K), we hope the two sets of results are useful to validate the generalization of retrieval-based face annotation system and the proposed WLRLCC algorithms.

TABLE 3
The annotation performance of WLRLCC and the other algorithms on database WDB-040K and WDB-P100.

	Hit Rate @ WDB-040K	Hit Rate @ ADB-P100
WLRLCC	0.7649 $\pm$ 0.010	0.7129 $\pm$ 0.017
S-Recon	0.7369 $\pm$ 0.011	0.6914 $\pm$ 0.014
SGSSL	0.7310 $\pm$ 0.011	0.6889 $\pm$ 0.016
LCC	0.7430 $\pm$ 0.012	0.6915 $\pm$ 0.015
MtBGS	0.7223 $\pm$ 0.015	0.6571 $\pm$ 0.016
MRR	0.6640 $\pm$ 0.010	0.5708 $\pm$ 0.013
SMW	0.6090 $\pm$ 0.019	0.5043 $\pm$ 0.017

Several observations can be drawn from the results. First, it is clear that the proposed WLRLCC algorithm consistently outperforms the other algorithms on both databases. Considering the database WDB-040K, the performance of the baseline SMW algorithm is about 61.0%, while the performance of the proposed WLRLCC algorithm is 76.5%. The better annotation performance validates the effectiveness of the proposed WLRLCC algorithm for retrieval-based face annotation. Second, for a small t value ($t = 1$), compared with SMW, WLRLCC achieves a significant improvement of about $+15.5$ points. Further, by examining a larger value of t, e.g. $t = 5$, the

hit rate performances of WLRLCC and SMW are 85.1% and 80.1% respectively, leading to an improvement of about $+5.0$ points. This result shows that a greater improvement can be achieved by the WLRLCC algorithm for a smaller value of t, which is critical to real-world applications where only top annotated results are concerned. Third, the similar result based on the database ADB-P100 shows that WLRLCC is always helpful to improve the annotation performance given different databases.

In addition to the above two weakly labeled databases, we also test on the PubFig database. Note that the initial label information of this database is of quite high quality as it has been manually purified and all the facial regions are just for the selected person even if there are multiple faces in the same image. In our experiments, we manually remove the cropped non-face images and some incorrect images for each person. In our experiment, PubFig is adopted as a well-labeled database in order to help us examine the impact of annotation performance by varied label noise settings. Specifically, regarding the construction of the query set, we randomly sample 10 images per person for all the 200 persons in the testbed. For PubFig, there are a total of $39,609$ images in the retrieval database, and $2,000$ images in the query database. To simulate different noise levels for the retrieval database, we randomly sample $p\%$ (10%, 20%, 40%) of images from each person and change their labels to the other randomly generated names. In order to fully observe the impact of the noise, one would prefer a higher recall of the face retrieval result and annotation result; we thus set the values of n and t to 40 and 5 respectively, for both WLRLCC and SMW algorithms. The experimental results are shown in TABLE 4.

TABLE 4
The evaluation of different noise percentage based on the database "PubFig".

Noise Percentage	p = 0%	p = 10%	p = 20%	p = 40%
WLRLCC	0.6760	0.6505	0.6115	0.5280
Performance Decline	—	-4%	-10%	-22%
SWM	0.5715	0.4540	0.4120	0.3250
Performance Decline	—	-21%	-28%	-43%

Several observations can be drawn from the results. First of all, for all the cases, the proposed WLRLCC algorithm consistently outperforms the baseline SMW algorithm. Second, when extra noise labels are added into the retrieval database, the task becomes more difficult and thus the annotation performance decreases. For example, by randomly mislabeling 10% images per person, the performance of WLRLCC will drop to 65.1%, which is about 96% of the result without noise. For comparison, the impact caused by noise labels is more serious for the SMW algorithm, which drops to 79% of the result without noise under the same noise level. This indicates that the proposed WLRLCC algorithm is more robust for noisy labels and is able to effectively improve the annotation performance for the weakly labeled databases. Finally, it is interesting to observe the overall annotation performance on the PubFig database is worse than the results on both WDB and ADB. We think three possible reasons may explain such observation: (i) The number of images per person (n_{ipp}) in the PubFig varies a lot. As shown in Figure 5, we randomly distribute n_{ipp} of PubFig and WDB-040K along the x-axis. The n_{ipp} of WDB-040K is stable and around 100, however, some values of n_{ipp} of PubFig are very small (e.g. 23, 27), which are insufficient for data-driven scheme and reduce the annotation performance; (ii) It may be because the facial images were cropped from the selected regions on this database without further alignment (instead of adopting the DLK algorithms on the previous databases which would remove some partial face images and reduce the size of retrieval database); (iii) The ratio of duplicate images on PubFig is smaller than that of WDB and ADB. To show this, for each query set of the three databases, we try to estimate the ratio of queries that have duplicate images in their retrieval databases. Note that we count a retrieved image is duplicate to the query whenever the distance is small enough (e.g., 10^{-4} in our setting). According to statistics, the duplicate ratios are about 8.0% (128 out of $1,600$), 5.0% (65 out of $1,300$), and 1.1% (22 out of $2,000$) for WDB-040K, ADB-P100, and PubFig, respectively. This result validates that duplicate images also may affect the face annotation performance.

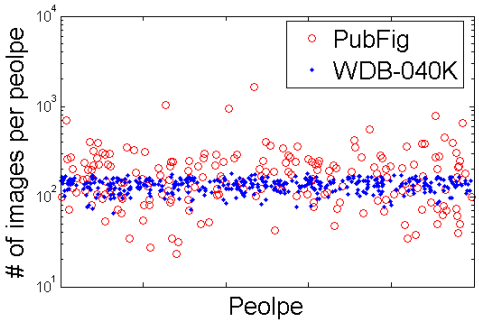


Fig. 5. The distribution of the number of images per person for the database WDB-040K and PubFig.

7.4 Evaluation on the Database Size

This experiment aims to examine the impact of database size on the annotation performance. There are two ways to vary the database size: one is to increase the number of facial images for each person, and the other is to fix the number of images per person but increase the number of unique persons. We will evaluate the impact under both settings in the following.

7.4.1 Varied Numbers of Persons

We examine the impact of the number of persons on the series of four databases: WDB-040K, WDB-100K, WDB-200K, WDB-400K, and WDB-600K, where the numbers of persons are increased from 400 to $6,025$. The experimental results are presented in Figure 6, where for clarity only the annotation result with Top-t=1 are shown.

We can draw some observations from the results. First, similar to the previous observations, the WLRLCC algorithm consistently obtains the best annotation performance among

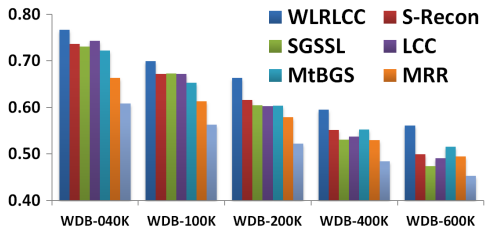


Fig. 6. Comparison between WLRLCC and the other algorithms on WDB-040K, WDB-100K, WDB-200K, WDB-400K, and WDB-600K

all the compared algorithms under different database sizes. Second, it is clear that the annotation performance decreases when increasing the number of persons, similar to the observation of some existing work on facial image retrieval in [6]. This is because increasing the number of persons leads to a larger database of more images for the same query, making the retrieval task more challenging.

7.4.2 *Varied Numbers of Images Per Person*

We further examine the impact of the number of images per person collected into the retrieval database based on two database ADB-P050 and ADB-P100, where the number of images per person is increased from about 50 to 100. The experimental results are shown in Figure 7 with $t = 1$. From the results, it is obvious to observe that the larger the number of facial images per person, the better the average annotation performance achieved. This is also similar to the observation found in the previous work [3], where more potential images in the retrieval database are beneficial to the annotation task.

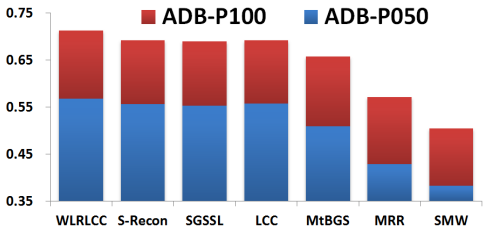


Fig. 7. Comparison between "ADB-P050" and "ADB-P100", with Top-$t = 1$.

7.5 Evaluation of Approximation for WLRLCC

In this section, we aim to examine the efficacy of the approximation scheme to trade off between annotation accuracy and time efficiency. To this purpose, we evaluate the running time and annotation performance of the three algorithms: the original WLRLCC algorithm, the PCA-based approximation scheme for WLRLCC, and the offline approximation scheme (AWLRLCC). For the PCA-based approximation, we randomly select 1000 images from the retrieval database to learn the eigenvectors, from which both the query image $\mathbf{x}_q$ and the top-n nearest neighbors X would the be projected into a new k-dimensional feature space, which can be further applied by the original WLRLCC algorithm. For the offline approximation, all the facial images in the retrieval database are first encoded based on its top-n neighbors, then the query image x_q is annotated with its top-ranked candidate images based on the proposed offline approximation scheme for WLRLCC (AWLRLCC). All the experimental results are presented in TABLE 5.

TABLE 5
The running time of different approximation schemes, based on database WDB-040K .

	WLRLCC	PCA-300	PCA-200	PCA-100	AWLRLCC
Time (s)	1.205 ±0.041	0.773 ±0.028	0.630 ±0.020	0.524 ±0.030	0.216 ±0.017
Hit Rate	0.7649 ±0.010	0.7482 ±0.013	0.7434 ±0.013	0.7240 ±0.015	0.7560 ±0.020

Several observations can be drawn from the results. First, by using PCA, the running time can be significantly reduced. For example, for the new feature "PCA-100", the PCA-based approximation scheme needs only about 0.52 second; however, this is paid by a drop of the annotation performance to about 72.4%. Second, we found that the proposed offline approximation algorithm (AWLRLCC) not only takes less time than the PCA-based scheme, but also achieves better annotation results. For example, it achieves the best approximation result 75.6% with only about 0.22 second. Although the proposed "AWLRLCC" algorithm achieves a good approximation result with less time cost, we note that its disadvantage is the cost of extra storage (memory) space because all the pre-computed results should be saved for the on-the-fly reconstruction.

To further evaluate the approximation result of AWLRLCC, we compare it with the original WLRLCC algorithm on all the databases, as shown in Figure 8. From the results, it is obvious that AWLRLCC always achieves good approximation results on different databases.

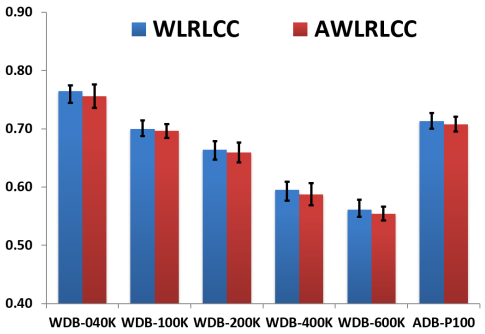


Fig. 8. Comparison between WLRLCC and AWLRLCC over all database.

7.6 Evaluation on Other Face Annotation Schemes

As we mentioned previously, the existing algorithms proposed for "text-based face annotation" or "caption-based face annotation" cannot be directly applied to our "retrieval-based face annotation" framework due to the shortage of caption information. However, the proposed WLRLCC algorithm can

be generally applied to the previous two schemes. In this section, we aim to apply the proposed WLRLCC technique for the existing two face annotation tasks, and examine if the proposed WLRLCC algorithm is comparable to the stat-of-the-art algorithms [30].

Following the same setup and settings as the previous studies, we conduct this experiment on the same database "*Labeled Yahoo! News* as used in [30], which has over $20,071$ documents, $31,147$ faces, and $5,873$ names. For the feature representation, we adopt PCA to project the high dimensional facial features into a low dimensional space (100). For the proposed WLRLCC algorithm, we use the whole database as the retrieval database, and initialize the name vector of each face with the detected names in its caption information.

7.6.1 Text-based Face Annotation

For the "text-based face annotation" task ("*Single-person Retrieval*"), the goal is to retrieve all facial images of one person for a given name. We adopt the same 23 query names as used in [30]. For each query name, we collect all the facial images which contain the query name in their corresponding captions, and re-rank the face sets according to the annotation result produced by WLRLCC. The final mAP scores are shown in TABLE 6. We observe that the proposed WLRLCC algorithm can achieve better results than the *"Graph-based"* algorithms and the *"Generative Model"*. It is generally competitive to the state-of-the-art *"SMLR model."* which adopts the original face descriptor and automatically finds which dimensions to use.

TABLE 6
The mAP scores over 23 queries for the proposed WLRLCC and three other algorithms in [30], where the results of the other algorithms were taken from [30]

GB-eps	GB-kNN	GM-QS	SMLR-2304D	WLRLCC
73.6	77.1	85.0	89.1	89.0

7.6.2 Caption-based Face Annotation

For the "caption-based face annotation" task ("*Multi-person Naming*"), the goal is to associate a face to one name of its corresponding caption. We firstly annotated all the $14,827$ facial images in the test set, then restrict the annotation results with names detected in the caption information, at last a threshold value is used to decide whether the top-ranked name should be assigned. The summary of names and faces association performance and the final naming precision are shown in TABLE 7, where the results of the compared algorithms are taken from [30]. Similar to the previous experiment, we observe that the proposed WLRLCC algorithm achieves better result than the *"Graph-based"* algorithms and highly competitive result as the *"Generative Model."*

7.6.3 Search-based Face Annotation

In this section, we aim to compare with the existing ULR approach [3] for search-based face annotation. As discussed in Section 2, the ULR algorithm [3] aims to refine the weak labels over the entire database and may be limited for its simple majority voting based annotation. in contrast, WLRLCC adopts a re-ranking scheme by fully exploiting the short list of top similar images via a unified optimization scheme. Following the same setting, we first apply WLRLCC on the same data set used in [3]. The result of WLRLCC is shown in the fourth column of TABLE 8, in which it is clear to see that WLRLCC achieves a better result.

Moreover, we note that the task addressed by WLRLCC is very different from that of ULR. In fact, the WLRLCC algorithm can be benefited from the refined label matrix learned by the ULR algorithm. In particular, we can exploit the refined the label matrix by ULR to construct the initial matrix $\tilde{Y}$ for each query image in Eq. 4. To show this, we implement such a combined cascade framework and show the result of this framework in the fifth column of TABLE 8.

TABLE 7
The summary of names and faces association performance obtained by *Graph-based method*, *Generative Model* [30], and the proposed WLRLCC

	Graph-based	Generative Model	WLRLCC
Correct: name assigned	6585	8327	7801
Correct: no assignment	3485	2600	3164
Incorrect: no assignment	1007	765	2820
Incorrect: wrong name	3750	3135	1024
Precision	0.68	0.74	0.74

TABLE 8
The performance of WLRLCC and the cascade framework combined with ULR.

	ORI	ULR	WLRLCC	ULR+WLRLCC
Hit Rate	0.548 $\pm$0.013	0.715 $\pm$0.008	0.766 $\pm$0.016	0.784 $\pm$0.013

As a summary, for the retrieval-based face annotation task, the proposed WLRLCC algorithm achieves promising results based on different database sets. It also can be combined with the previous ULR algorithm to further improve the annotation performance. Finally, WLRLCC is not restricted to retrieval-based face annotation tasks, but also can be easily applied to tackle the other existing face annotation tasks in literature with encouraging results highly competitive to the state-of-the-art algorithms on the same public web facial image databases.

7.7 Evaluation of Facial Feature Representation

In this experiment, we evaluate the impact of different types of facial feature representations for the proposed WLRLCC algorithm based on the database WDB-040K. TABLE 9 shows the annotation performance of different features. All of these features are extracted from the aligned facial images by the DLK algorithm [42]. The features of "GIST", "Edge", "Color", and "Gabor" are generated by the toolbox in http://goo.gl/BzPPx. For "LBP", each aligned facial image is divided into 7×7 blocks [52], leading to 2891-dimensional features.

TABLE 9
The performance of WLRLCC with different facial features based on database WDB-040K.

	GIST	Edge	Color	Gabor	LBP_{2891}	LBP_{512}
Hit Rate	0.7649	0.3724	0.4265	0.4805	0.7928	0.7584
Time(s)	1.205	0.681	0.457	0.658	45.830	1.239

From the results, we observe that the original LBP_{2891} feature achieves the best performance, which is consistent with the existing face recognition studies that show LBP is one of the best facial features. However, the high dimensionality of LBP feature leads to an extremely high running time of WLRLCC, which takes about 46 seconds on average for each query sample. By reducing the dimensionality of the original LBP feature to the same dimensionality of GIST (512D) via PCA, denoted as "LBP_{512}", the new feature LBP_{512} performs slightly worse than GIST. Based on the result, for the databases aligned by the DLK algorithm (e.g., WDB and ADB), we adopted the GIST features to represent the facial images in our experiments. For the other database (e.g., PubFig), we adopted the LBP features to represent the facial images.

8 Limitations

Despite the encouraging results achieved, our work is limited in some aspects. First, we assume the top ranked web facial images are related to the query name, which is certainly true for celebrities. However, when the query person is not well-known, the relevant facial images on the internet may be very limited. This is a common issue of all the existing data-driven face/image annotation techniques. It could be partially resolved by exploiting social contextual information from social networks or exploring a combination with model-based annotation techniques. Second, WLRLCC relies on good facial feature representation techniques and an effective facial image retrieval scheme, which is still technically challenging, despite being studied for many years.

9 Conclusions

We investigate the retrieval-based face annotation problem and present a promising framework to address the problems in mining massive weakly labeled facial images freely available on the WWW. We proposed a Weak Label Regularized Local Coordinate Coding (WLRLCC) algorithm, which effectively exploits the principles of both local coordinate coding and graph-based weak label regularization. Moreover, a sparse reconstruction scheme is developed to perform face annotation task. We have conducted extensive empirical studies on several web facial image databases, which validate the efficacy of WLRLCC. To further improve the efficiency and scalability, we then proposed an offline approximation scheme (AWLRLCC) to speed up the original WLRLCC algorithm, saving a significant amount of computational time while maintaining comparable performance.

Acknowledgements

This work was supported in part by Singapore MOE Tier-1 research grant (RG33/11), Microsoft Research grant, and IDM National Research Foundation (NRF) Research grant (MDA/IDM/2012/8/8-2 VOL 01). Jianke Zhu was supported by Fundamental Research Funds for the Central Universities.

References

[1] 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 *CVPR*, 2004, pp. 848–854. 1, 2

[2] J. Zhu, S. C. Hoi, and M. R. Lyu, "Face annotation by transductive kernel fisher discriminant," *IEEE Trans. Multimedia*, vol. 10, no. 1, pp. 86–96, 2008. 1, 2

[3] D. Wang, S. Hoi, Y. He, and J. Zhu, "Mining weakly-labeled web facial images for search-based face annotation," *IEEE T KNOWL DATA EN*, vol. 99, no. PrePrints, pp. 1–14, 2012. 1, 3, 11, 12

[4] A. Holub, P. Moreels, and P. Perona, "Unsupervised clustering for google searches of celebrity images," in *IEEE FG'08*, 2008, pp. 1–8. 1, 2

[5] S. C. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Semi-supervised svm batch mode active learning with applications to image retrieval," *ACM TOIS*, vol. 27, no. 3, pp. 1–29, July 2009. 1

[6] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable face image retrieval with identity-based quantization and multi-reference re-ranking," in *CVPR*, 2010, pp. 3469–3476. 1, 2, 8, 11

[7] S. C. Hoi, W. Liu, and S.-F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval and clustering," *ACM T MULTIM COMPUT*, vol. 6, no. 3, pp. 18:1–18:26, Aug. 2010. 1

[8] J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, and R. Jain, "Image annotation by knn-sparse graph-based label propagation over noisily tagged web images," *ACM TIST*, vol. 2, pp. 14:1–14:15, Feb. 2011. 1, 2, 8

[9] F. Wu, Y. Han, Q. Tian, and Y. Zhuang, "Multi-label boosting for image annotation by structural grouping sparsity," in *ACM Multimedia*, 2010, pp. 15–24. 1, 2, 8

[10] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li, "Modeling lsh for performance tuning," in *CIKM*, 2008, pp. 669–678. 3, 4

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003. 2

[12] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition, 2nd Edition.* Springer, 2011. 2

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," UMASS, Tech. Rep. pp. 07–49, October 2007. 2, 7

[14] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *CVPR*, 2010, pp. 2707–2714. 2

[15] A. Hanbury, "A survey of methods for image annotation," *J. Vis. Lang. Comput.*, vol. 19, pp. 617–627, Oct. 2008. 2

[16] P. Wu, S. C.-H. Hoi, P. Zhao, and Y. He, "Mining social images with distance metric learning for automated image tagging," in *WSDM*, 2011, pp. 197–206. 2

[17] H. Xia, P. Wu, S. C. Hoi, and R. Jin, "Boosting multi-kernel locality-sensitive hashing for scalable image retrieval, in *SIGIR*, 2012, pp. 55–64. 2

[18] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *ECCV*, 2002, pp. 97–112. 2

[19] G. Carneiro, A. B. Chan, P. Moreno, and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Trans Pattern Anal Mach Intell*, vol. 29, no. 3, pp. 394–410, March 2007. 2

[20] C. Wang, F. Jing, L. Zhang, and H.-J. Zhang, "Image annotation refinement using random walk with restarts," in *ACM MM*, 2006, pp. 647–650. 2

[21] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, Nov 1999. 2

[22] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: A database and web-based tool for image annotation," *Int. J. Comput. Vision*, vol. 77, no. 1-3, pp. 157–173, 2008. 2

[23] X. Rui, M. Li, Z. Li, W.-Y. Ma, and N. Yu, "Bipartite graph reinforcement model for web image annotation," in *ACM MM*, 2007, pp. 585–594. 2

[24] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 2, pp. 210–227, Feb. 2009. 2, 8

[25] G. Wang, A. Gallagher, J. Luo, and D. Forsyth, "Seeing people in social context: recognizing people and social relationships," in *ECCV*, 2010, pp. 169–182. 2

[26] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang, "Easyalbum: an interactive photo annotation system based on face clustering and re-ranking," in *CHI*, 2007, pp. 367–376. 2

[27] J. Y. Choi, W. D. Neve, K. N. Plataniotis, and Y. M. Ro, "Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks," *IEEE Transactions on Multimedia*, vol. 13, no. 1, pp. 14–28, Feb. 2011. 2

[28] D. Ozkan and P. Duygulu, "A graph based approach for naming faces in news photos," in *CVPR*, 2006, pp. 1477–1482. 2

[29] D.-D. Le and S. Satoh, "Unsupervised face annotation by mining the web," in *ICDM*, 2008, pp. 383–392. 2

[30] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Face recognition from caption-based supervision," in *Int. J. Comput. Vision*, vol. 96, no. 1, pp. 64–82, Jan. 2011. 2, 3, 7, 12

[31] T. Mensink and J. J. Verbeek, "Improving people search using query expansions," in *ECCV*, 2008, pp. 86–99. 2

[32] T. L. Berg, A. C. Berg, J. Edwards, and D. Forsyth, "Who's in the picture," in *NIPS*, 2006, pp. 264–271. 3

[33] D. Wang, S. C. Hoi, and Y. He, "Mining weakly labeled web facial images for search-based face annotation," in *SIGIR*, 2011, pp. 535–544. 3

[34] D. Wang, S. C. H. Hoi, and Y. He, "A unified learning framework for auto face annotation by mining web facial images," in *CIKM '12*, 2012, pp. 1392–1401. 3

[35] D. Wang, S. C. Hoi, P. Wu, J. Zhu, Y. He, and C. Miao, "Learning to name faces: A multimodal learning scheme for search-based face annotation," in *SIGIR*, 2013. 3

[36] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *NIPS*, 2009, pp. 2259–2267. 3, 4, 8

[37] P. O. Hoyer, "Non-negative sparse coding," *CoRR*, vol. cs.NE/0202009, 2002. 3, 5

[38] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *ICML*, 2003, pp. 912–919. 3

[39] Y.-Y. Sun, Y. Zhang, and Z.-H. Zhou, "Multi-label learning with weak label," in *AAAI*, 2010, pp. 593–598. 3

[40] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from google's image search," in *ICCV'05*, vol. 2, oct. 2005, pp. 1816–1823 Vol. 2. 4

[41] T. Berg and D. Forsyth, "Animals on the web," in *CVPR*, vol. 2, 2006, pp. 1463–1470. 4

[42] J. Zhu, S. C. Hoi, and L. V. Gool, "Unsupervised face alignment by robust nonrigid mapping," in *ICCV*, 2009, pp. 1265–1272. 4, 8, 12

[43] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans Pattern Anal Mach Intell*, vol. 29, pp. 300–312, Feb. 2007. 4, 8

[44] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *CoRR*, vol. abs/1203.1005, 2012. 5

[45] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 2, pp. 210–227, April 2008. 5

[46] J. Wright and Y. Ma, "Dense error correction via l1-minimization," *IEEE Trans. Inf. Theor.*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010. 5

[47] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Img. Sci.*, vol. 2, pp. 183–202, March 2009. 6

[48] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009. 6

[49] J. Liu and J. Ye, "Efficient euclidean projections in linear time," in *ICML*, 2009, pp. 657–664. 6

[50] M. Ozcan, J. Luo, V. Ferrari, and B. Caputo, "A large-scale database of images and captions for automatic face naming," in *BMVC*, 2011, pp. 29.1–29.11. 7

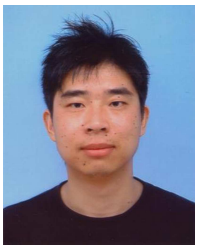
[51] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and Simile Classifiers for Face Verification," in *ICCV*, 2009, pp. 365–372. 7

[52] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *ECCV*, vol. 1, 2004, pp. 469–481. 8, 12

Dayong Wang is currently a PhD candidate in the School of Computer Engineering, Nanyang Technological University, Singapore. He received his Bachelor degree in Computer Science from Tsinghua University, Beijing, P.R. China. His research interests include machine learning, computer vision, pattern recognition, and multimedia information retrieval.

Steven C. H. Hoi is currently an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. He received his Bachelor degree in Computer Science from Tsinghua University, Beijing, P.R. China, and his Master and Ph.D degrees in Computer Science and Engineering from Chinese University of Hong Kong. His research interests include machine learning, multimedia information retrieval, web search and data mining. He is a member of IEEE and ACM.

Ying He received the BS and MS degrees in Electrical Engineering from Tsinghua University, and the PhD degree in Computer Science from Stony Brook University. He is currently an Associate Professor at the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests fall in the broad areas of visual computing. He is particularly interested in the problems that require geometric computation and analysis.



Jianke Zhu is currently an Associate Professor at Zhejiang University. He obtained Bachelor degree from Beijing University of Chemical Technology in 2001, his Master degree from University of Macau in 2005, and his PhD degree in the Computer Science and Engineering department at the Chinese University of Hong Kong. His research interests are in pattern recognition, computer vision, and statistical machine learning.

Tao Mei is a Researcher with the Media Computing Group in Microsoft Research Asia. His current research interests include multimedia information retrieval and computer vision. He received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, in 2001 and 2006, respectively.

Jiebo Luo joined the University of Rochester in Fall 2011 after over fifteen years at Kodak Research Laboratories, where he was a Senior Principal Scientist leading research and advanced development. He is the Editor-in-Chief of the Journal of Multimedia, and has served on the editorial boards of the IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Multimedia, etc. He has authored over 200 papers and 70 US patents. Dr. Luo is a Fellow of the SPIE, IEEE, and IAPR.