IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 4, Ver. I (July – Aug. 2015), PP 29-35
www.iosrjournals.org

Mining Weakly Labeled Web Facial Images For Search-Based Face Annotation

Mr. Rohit D. Mane
Bharati Vidyapeeth's collage of Engineering Kolhapur, India.

Abstract: *This paper investigates a framework of search-based face annotation (SBFA) by mining weakly labeled facial images that are freely available on the World Wide Web (WWW). One challenging problem for search-based face annotation scheme is how to effectively perform annotation by exploiting the list of most similar facial images and their weak labels that are often noisy and incomplete. To tackle this problem, we propose an effective unsupervised label refinement (ULR) approach for refining the labels of web facial images using machine learning techniques. We formulate the learning problem as a convex optimization and develop effective optimization algorithms to solve the large-scale learning task efficiently. To further speed up the proposed scheme, we also propose a clustering-based approximation algorithm which can improve the scalability considerably. We have conducted an extensive set of empirical studies on a large-scale web facial image testbed, in which encouraging results showed that the proposed ULR algorithms can significantly boost the performance of the promising SBFA scheme.*

I. Introduction

Due to the popularity of various digital cameras and the rapid growth of social media tools for internet-based photo sharing, recent years have witnessed an explosion of the number of digital photos captured and stored by consumers. A large portion of photos shared by users on the Internet are human facial images. Some of these facial images are tagged with names, but many of them are not tagged properly. This has motivated the study of auto face annotation, an important technique that aims to annotate facial images automatically.

Auto face annotation can be beneficial to many real world applications. For example, with auto face annotation techniques, online photo-sharing sites (e.g., Facebook) can automatically annotate users' uploaded photos to facilitate online photo search and management. Besides, face annotation can also be applied in news video domain to detect important persons appeared in the videos to facilitate news video retrieval and summarization tasks.

Classical face annotation approaches are often treated as an extended face recognition problem, where different classification models are trained from a collection of well labeled facial images by employing the supervised or semi-supervised machine learning techniques. However, the "model-based face annotation techniques are limited in several aspects.

First, it is usually time-consuming and expensive to collect a large amount of human labelled training facial images. Second, it is usually difficult to generalize the models when new training data or new persons are added, in which an intensive retraining process is usually required. Last but not least, the annotation/recognition performance often scales poorly when the number of persons/classes is very large.

Some emerging studies have attempted to explore a promising search-based annotation paradigm for facial image annotation by mining the World Wide Web (WWW), where a massive number of weakly labeled facial images are freely available. Instead of training explicit classification models by the regular model-based face annotation approaches, the search-based face annotation (SBFA) paradigm aims to tackle the automated face annotation task by exploiting content-based image retrieval (CBIR) techniques, in mining massive weakly labeled facial images on the web. The SBFA framework is data-driven and model-free, which to some extent is inspired by the search-based image annotation techniques for generic image annotations. The main objective of SBFA is to assign correct name labels to a given query facial image. In particular, given a novel facial image for annotation, we first retrieve a short list of top K most similar facial images from a weakly labeled facial image database, and then annotate the facial image by performing voting on the labels associated with the top K similar facial images.

One challenge faced by such SBFA paradigm is how to effectively exploit the short list of candidate facial images and their weak labels for the face name annotation task. To tackle the above problem, we investigate and develop a search-based face annotation scheme. In particular, we propose a novel unsupervised label refinement (URL) scheme by exploring machine learning techniques to enhance the labels purely from the weakly labelled data without human manual efforts. We also propose a clustering-based approximation (CBA) algorithm to improve the efficiency and scalability. As a summary, the main contributions of this paper include the following:

We investigate and implement a promising search based face annotation scheme by mining large amount of weakly labelled facial images freely available on the WWW. We propose a novel ULR scheme for enhancing label quality via a graph-based and low-rank learning approach. We propose an efficient clustering-based approximation algorithm for large-scale label refinement problem. We conducted an extensive set of experiments, in which encouraging results were obtained.

We note that a short version of this work had appeared in SIGIR2011. This journal article has been significantly extended by including a substantial amount of new content. The remainder of this paper is organized as follows: Section 2 reviews the related work. Section 3 gives an overview of the proposed search-based face annotation framework. Hare we present the proposed unsupervised label refinement scheme. Our experimental results of performance evaluation, the limitation of our work and conclusion finally concludes this paper.

II. Related Work

Our work is closely related to several groups of research work. The first group of related work is on the topics of face recognition and verification, which are classical research problems in computer vision and pattern recognition and have been extensively studied for many years. Recent years have observed some emerging benchmark studies of unconstrained face detection and verification techniques on facial images that are collected from the web, such as the LFW benchmark studies. Some recent study had also attempted to extend classical face recognition techniques for face annotation tasks. Comprehensive reviews on face recognition and verification topics can be found in some survey papers and books.

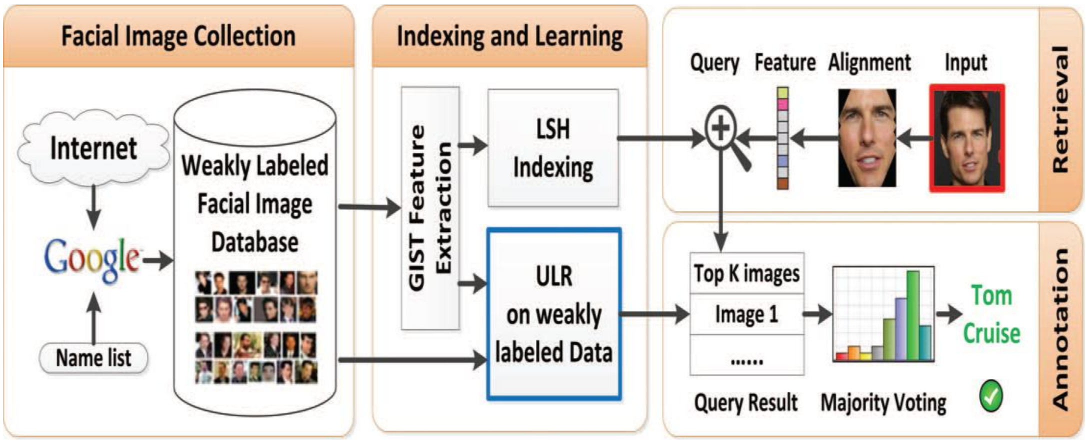
The second group is about the studies of generic image annotation. The classical image annotation approaches usually apply some existing object recognition techniques to train classification models from human-labeled training images or attempt to infer the correlation/probabilities between images and annotated keywords. Given limited training data, semi-supervised learning methods have also been used for image annotation. For example, Wang et al. proposed to refine the model-based annotation results with a label similarity graph by following random walk principle. Similarly, Pham et al. proposed to annotate unlabelled facial images in video frames with an iterative label propagation scheme. Although semi-supervised learning approaches could leverage both labeled and unlabeled data, it remains fairly time-consuming and expensive to collect enough well-labeled training data to achieve good performance in large-scale scenarios. Recently, the search-based image annotation paradigm has attracted more and more attention. For example, Russell et al. built a large collection of web images with ground truth labels to facilitate object recognition research. However, most of these works were focused on the indexing, search, and feature extraction techniques. Unlike these existing works, we propose a novel unsupervised label refinement scheme that is focused on optimizing the label quality of facial images towards the search-based face annotation task.

The third group is about face annotation on personal/family/social photos. Several studies have mainly focused on the annotation task on personal photos, which often contain rich contextual clues, such as personal/family names, social context, geotags, and timestamps and so on. The number of persons/classes is usually quite small, making such annotation tasks less challenging. These techniques usually achieve fairly accurate annotation results, in which some techniques have been successfully deployed in commercial applications, for example, Apple iPhoto, Google Picasa, Microsoft easy Album, and Facebook face auto tagging solution.

The fourth group is about the studies of face annotation in mining weakly labeled facial images on the web. Some studies consider a human name as the input query, and mainly aim to refine the text-based search results by exploiting visual consistency of facial images. For example, Ozkan and Duygulu proposed a graph-based model for finding the densest sub-graph as the most related result. Following the graph-based approach, Le and Satoh proposed a new local density score to represent the importance of each returned images, and Guillaum in et al. introduced a modification to incorporate the constraint that a face is only depicted once in an image. On the other hand, the generative approach like the Gaussian mixture model was also been adopted to the name-based search scheme and achieved comparable results. Recently, a discriminant approach was proposed in [44] to improve over the generative approach and avoid the explicit computation in graph-based approach. By using ideas from query expansion [45], the performance of name-based scheme can be further improved with introducing the images of the “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 possibility model combined with a clustering algorithm to estimate the relationship between the facial images and the names in their captions. For the facial images and the detected names in the same document (a web image and its corresponding caption), Guillemin et al. proposed to iteratively update the assignment based on a minimum cost matching algorithm. In their follow-up work, they further improve the annotation performance by using distance metric learning techniques to achieve more discriminative feature in low-dimension space.

Our work is different from the above previous works in two main aspects. First of all, our work aims to solve the general content-based face annotation problem using the search-based paradigm, where facial images are directly used as query images and the task is to return the corresponding names of the query images. Very limited research progress has been reported on this topic. Some recent work mainly addressed the face retrieval problem, in which an effective image representation has been proposed using both local and global features. Second, based on initial weak labels, the proposed unsupervised label refinement algorithm learns an enhanced new label matrix for all the facial images in the whole name space; however, the caption-based annotation scheme only considers the assignment between the facial images and the names appeared in their corresponding surrounding-text. As a result, the caption-based annotation scheme is only applicable to the scenario where both images and their captions are available, and cannot be applied to our SBFA framework due to the lack of complete caption information.

The fifth group is about the studies of purifying web facial images, which aims to leverage noisy web facial images for face recognition applications. Usually these works are proposed as a simple pre-processing step in the whole system without adopting sophisticated techniques. For example, the work in applied a modified k means clustering approach for cleaning up the noisy web facial images. Zhao et al. [48] proposed a consistency learning method to train face models for the celebrity by mining the text-image co-occurrence on the web as a weak signal of relevance toward supervised face learning task from a large and noisy training set. Unlike the above existing works, we employ the unsupervised machine learning techniques and propose a graph-based label refinement algorithm to optimize the label quality over the whole retrieval database in the SBFA task.



(a) (b) (c)

Fig.1. the system flow of the proposed search-based face annotation scheme.

(a) We collect weakly labelled facial images from WWW using web search engines.
(b) We pre-process the crawled web facial images, including face detection, face alignment, and feature extraction for the detected faces; after that, we apply LSH to index the extracted high dimensional facial features. We apply the proposed ULR method to refine the raw weak labels together with the proposed clustering-based approximation algorithms for improving the scalability.
(c) We search for the query facial image to retrieve the top K similar images and use their associated names for voting toward auto annotation.

Search-Based Face Annotation

Fig. illustrates the system flow of the proposed framework of search-based face annotation, which consists of the following steps:
1. Facial image data collection;
2. Face detection and facial feature extraction;
3. High-dimensional facial feature indexing;
4. Learning to refine weakly labeled data;
5. Similar face retrieval; and
6. Face annotation by majority voting on the similar faces with the refined labels.

The first four steps are usually conducted before the test.

The first step is the data collection of facial images as shown in Fig. 1a, in which we crawled a collection of facial images from the WWW by an existing web search engine (i.e., Google) according to a name list that contains the names of persons to be collected. phase of a face annotation task, while the last two steps are conducted during the test phase of a face annotation task, which usually should be done very efficiently. We briefly describe each step below. The second step is to pre-process web facial images to extract face-related information, including face detection and alignment, facial region extraction, and facial feature representation. For face detection and alignment, we adopt the unsupervised face alignment technique proposed in

For facial feature representation, we extract the GIS Texture features to represent the extracted faces. As a result, each face can be represented by a d-dimensional feature vector.

The third step is to index the extracted features of the faces by applying some efficient high-dimensional indexing technique to facilitate the task of similar face retrieval in the subsequent step. In our approach, we adopt the locality sensitive hashing (LSH), a very popular and effective high-dimensional indexing technique.

All the above are the processes before annotating a query facial image. Next, we describe the process of face annotation during the test phase. In particular, given a query facial image for annotation, we first conduct a similar face retrieval process to search for a subset of most similar faces (typically top K similar face examples) from the previously indexed facial database. With the set of top K similar face examples retrieved from the database, the next step is to annotate the facial image with a label (or a subset of labels) by employing a majority voting approach that combines the set of labels associated with these top K similar face examples.

Algorithms: The above optimization tasks belong to convex optimization or more exactly quadratic programming (QP) problems. It seems to be possible to solve them directly by applying generic QP solvers. However, this would be computationally highly intensive since matrix F can be potentially very large, for example, for a large 400-person database of totally 40,000 facial images, F is a 40,000 _ 400 matrix that consists of 16 million variables, which is almost infeasible to be solved by any existing generic QP solver.

Clustering-Based Approximation

The number of variables in the previous problem is n _ m, where n is the number of facial images in the retrieval database and m is the number of distinct names (classes). For a small problem, we can solve it efficiently by the proposed MGA-based algorithms (SRF-MGA or CCFMGA).

For a large problem, we can adopt the proposed CDA-based algorithms (SRF-CDA or CCF-CDA), where the number of variables in each sub problem is n. However, when n is extremely large, the CDA-based algorithms still can be computationally intensive. One straightforward solution for acceleration is to adopt parallel computation, which can be easily exploited by the proposed SRF-CDA or CCF-CDA algorithms since each of the involved sub optimization tasks can be solved independently. However, the speedup of the parallel computation approach quite depends on the hardware capability. To further enhance the scalability and efficiency in algorithms, in this paper, we propose a clustering-based approximation solution to speed up the solutions for large-scale problems.

In particular, the clustering strategy could be applied in two different levels: 1) one is on “image-level,” which can be used to directly separate all the n facial images into a set of clusters, and 2) the other is on “name-level,” which can be used to First separate the m names into a set of clusters, then to further split the retrieval database into different subsets according to the name-label clusters. Typically, the number of facial images n is much larger than the number of names m, which means that the clustering on “imagelevel” would be much more time-consuming than that on “name-level.” Thus, in our approach, we adopt the “namelevel” clustering scheme for the sake of scalability and efficiency. After the clustering step, we solve the proposed ULR problem in each subset, and then merge all the learning results into the final enhanced label matrix F.

Clustering-Based Approximation

The number of variables in the previous problem is n*m, where n is the number of facial images in the retrieval database and m is the number of distinct names (classes).

For a small problem, we can solve it efficiently by the proposed MGA-based algorithms. For a large problem, we can adopt the proposed CDA-based algorithms, where the number of variables in each subproblem is n. However, when n is extremely large, the CDA-based algorithms still can be computationally intensive. One straightforward solution for acceleration is to adopt parallel computation, which can be easily exploited by the proposed SRF-CDA or CCF-CDA algorithms since each of the involved sub optimization tasks can be solved independently. However, the speedup of the parallel computation approach quite depends on the hardware capability. To further enhance the scalability and efficiency in algorithms, in this paper, we propose a clustering-based approximation solution to speed up the solutions for large-scale problems.

In particular, the clustering strategy could be applied in two different levels: 1) one is on “image-level,” which can be used to directly separate all the n facial images into a set of clusters, and 2) the other is on “name-level,” which can be used to First separate the m names into a set of clusters, then to further split the retrieval database into different subsets according to the name-label clusters. Typically, the number of facial images n is much larger than the number of names m, which means that the clustering on “imagelevel” would be much more time-consuming than that on “name-level.” Thus, in our approach, we adopt the “namelevel” clustering scheme for the sake of scalability and efficiency. After the clustering step, we solve the proposed ULR problem in each subset, and then merge all the learning results into the final enhanced label matrix F.

Comparison Schemes and Setup

In our experiments, we know the algorithms described for solving the proposed ULR task. We finally adopted the soft-regularization formulation of the proposed ULR technique in our evaluation since it is empirically faster than the convex-constraint formulation according to our implementations. To better examine the efficacy of our technique, we also implemented some baseline annotation method and existing algorithms for comparisons.

Evaluation of Facial Feature Representation

In this experiment, we evaluate the face annotation performance of five types of facial features for the baseline ORI algorithm. As a result, each small subproblem could be solved efficiently. Besides, as the subproblems on different subsets are independent, a parallel computation framework could also be adopted for further acceleration.

Experiment Testbed

In our experiments, we collected a human name list consisting of popular actor and actress names from the IMDb website: http://www.imdb.com. In particular, we collected these names with the billboard: “Most Popular People Born In yyyy” of IMDb, where yyyy is the born year. For example, the webpage2 presents all the actor and actresses who were born in 1975 in the popularity order. Our name list covers the actors and actresses who were born between 1950 and 1990.

Comparison Schemes and Setup

In our experiments, we implemented all the algorithms described previously for solving the proposed ULR task. We finally adopted the soft-regularization formulation of the proposed ULR technique in our evaluation since it is empirically faster than the convex-constraint formulation according to our implementations. To better examine the efficacy of our technique, we also implemented some baseline annotation method and existing algorithms for comparisons. “ULR”: the proposed unsupervised label refinement method, denoted as “ULR” for short. To evaluate their annotation performances, we adopted the hit rate at top t annotated results as the performance metric, which measures the likelihood of having the true label among the top t annotated names. For each query facial image, we retrieved a set of top K similar facial images from the database set, and return a set of top T names for annotation by performing a majority voting on the labels associated with the set of top K images.

Evaluation of Facial Feature Representation: In this experiment, we evaluate the face annotation performance of five types of facial features for the baseline ORI algorithm. Table 1 shows the annotation performance.

I. Evaluation of Auto Face Annotation

We aim to evaluate the auto face
Annotation performance based on the search-based face annotation scheme. Without loss of generality, we first evaluated the proposed ULR algorithm from different aspects on database “DB0400,” and then verified its performance on large-scale database “DB1000.” Several observations can be drawn from various results.

First of all, it is clear that ULR which employs unsupervised learning to refine labels consistently performs better than the ORI baseline using the original weak label, the existing CL algorithm, MKM algorithm, and the LPSN algorithm. The promising result validates that the proposed ULR algorithm can effectively exploit the underlying data distribution of all data examples to refine the label matrix and improve the performance of the search-based face annotation approach. Second, we note that the ULR algorithm outperforms its special case “URL□ ¼0” without the sparsely regularize in the SRF formulation, which validates the importance of the sparsity regularize. Finally, when T is small, the hit rate gap, i.e., the hit rate difference between ORI and ULR is more significant, and the annotation performance increases slowly when T is large. In practice, we usually focused on the small T value since users typically would not be interested in a long list of annotated names.

I. Evaluation on Varied Numbers of Images per Person in Database

This experiment aims to further examine the relationship between the annotation performance and the number of facial images per person in building the facial image database. Unlike the previous experiment with top 100 retrieval facial images per person in the database, we created three variables of varied-size databases, which consist of top 50, 75, and 100 retrieval facial images per person, respectively. We denote these three databases as P050, P075, and P100, respectively.

II. Evaluation on a Larger Database

This experiment aims to verify the annotation performance of the proposed SBFA framework over a larger retrieval database: “DB1000.” As the test database is unchanged, the extra facial images in the retrieval database are definitely harmful to the nearest facial retrieval result for each query image. A similar result could also been observed in, where the mean average precision became smaller for a larger retrieval database.

Some observations can be drawn from the experimental results. First of all, the proposed ULR algorithm also could efficiently enhance the initial noisy label and achieve the best performance over the other algorithms. Second, all the algorithms perform slightly worse on the larger retrieval database.

To further improve the system performance, we adopt a simple weighted majority voting scheme in the third step of Fig. 1. Specifically, we assign a weighting value to each facial image in the short list of similar faces according to its ranking position.

III. Evaluation of Optimization Efficiency

This section aims to conduct extensive evaluations on the running time cost by the four different algorithms. We refer the four algorithms with the following abbreviations:

- SRF-MGA: Soft-regularization formulation solved by the multistep gradient algorithm.
- SRF-CDA: soft-regularization formulation solved by the coordinate decent algorithm.
- CCF-MGA: Convex-constraint formulation solved by the multistep gradient algorithm.

1. CCF-CDA: Convex-constraint formulation solved by the coordinate decent algorithm.

IV. Evaluation of Clustering-Based Approximation

In this experiment, we aim to evaluate the acceleration performance of the two proposed clustering-based approximation schemes (BCBA and DCBA) on the large database DB1000. A good approximation is expected to achieve a high reduction in running time with a small loss in annotation performance. Thus, this experiment evaluates both running time and annotation performance. The running time of CBA scheme mainly consists of three parts: 1) the time of constructing the similarity matrix C; 2) the time of clustering; and 3) the total time of running ULR algorithm in each subset.

Limitations

Despite the encouraging results, our work is limited in several aspects. First, we assume each name corresponds to a unique single person. Duplicate name can be a practical issue in real-life scenarios. One future direction is to extend our method to address this practical problem. For example, we can learn the similarity between two different names according to the web pages so as to determine how likely the two different names belong to the same person. Second, we assume the top retrieved web facial images are related to a query human name. This is clearly true for celebrities. However, when the query facial image is not a well-known person, there may not exist many relevant facial images on the WWW, which thus could affect the performance of the proposed annotation solution. This is a common limitation of all existing data-driven annotation techniques. This might be partially solved by exploiting social contextual information.

III. Conclusions

This paper investigated a promising search-based face annotation framework, in which we focused on tackling the critical problem of enhancing the label quality and proposed a ULR algorithm. To further improve the scalability, we also proposed a clustering-based approximation solution, which successfully accelerated the optimization task without introducing much performance degradation. From an extensive set of experiments, we found that the proposed technique achieved promising results under a variety of settings. Our experimental results also indicated that the proposed ULR technique significantly surpassed the other regular approaches in literature. Future work will address the issues of duplicate human names and explore supervised/semi-supervised learning techniques to further enhance the label quality with affordable human manual refinement efforts.

Acknowledgment

This work was supported by Prof. S.B. Takmare, Bharati Vidyapit's collage of engineering, Kolhapur.

References

[1]. Social Media Modeling and
[2]. Computing, S.C.H. Hoi, J. Luo, S. Boll, D. Xu, and R. Jin, eds. Springer, 2011.
[3]. D. Wang, S.C.H. Hoi, and Y. He,
[4]. "Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2011.
[5]. P. Belhumeur, J. Hespanha, and D.
[6]. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Pattern Analysis and Machine
[7]. Intelligence, vol. 19, no. 7, pp. 711-720, July 1997.