A Survey on an Efficient Bayesian Similarity Based Face Image Annotation Model for Weakly Labeled Web Facial Images

Tini Vasanth[#1], A.Nirmal Kumar[*2]

[#]PG Scholar, Department of Computer Science and Engineering, Christian College of Engineering and Technology, Dindigul, Tamilnadu - 624619, India.
[1]tini.vasanth@gmail.com

* Assistant Professor,Department of Computer Science and Engineering, Christian College of Engineering and Technology, Dindigul, Tamilnadu - 624619, India.
[2]sa.nirmalkumar@gmail.com

***Abstract*-This paper proposes a novel effective approach of Unsupervised label refinement (URL) for exploiting the top list of most similar web facial images and their weak labels retrieved from World Wide Web (WWW).Duplicate names can be a practical issue in real-life scenarios.To tackle this problem I propose Bayesian framework which measures the similarity of different names by defined it as maximum a posteriori (MAP) estimation for single person annotation.To further improve the performance of further annotation contextual features are extracted with GIST feature to improve the identification of normal persons other than celebrities.**

Keywords—**Bayesian framework, contextual,face annotation, label refinement,machine learning,web facial images,weak label.**

I. INTRODUCTION

Digital photo albums for internet-based photo sharing are gaining much popularity in both number and size due to the rapid grow of digital cameras and mobile phone cameras.These large collections require the annotation of some semantic information to facilitate browsing, manipulation and sharing of photos.This has led to auto face annotation which annotates facial images automatically.

Auto face annotation, is used to detect faces from a facial image and then assign each of them with proper human names, is a fundamental research area and is finding use in many real-world applications. For example, online photo-sharing sites (e.g., Facebook) can automatically annotate users' uploaded photos to facilitate online photo search and management which is useful in many areas of research. 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 [2], [3].

Classical face annotation ,where different classification models are trained from a collection of well labeled facial images by employing the supervised or semi-supervised machine learning techniques [2], [4], [5],[6], [7].Model-based face annotation techniques are limited in several ways,They are time consuming and expensive to collect a large amount of human-labeled training facial images as well as an intensive retraining process is required to generalize the models when new training data or new persons are added.

Recently, some studies have attempted toexplore a promising search-based annotation paradigm for facial image annotation by mining the World Wide Web (WWW), where a massive number of weakly labelled facialimages are freely available. Instead of training classification models by the regular model-based face annotation approaches, the search-based face annotation focuses on how to tackle the automated face annotation task by exploiting content-based image retrieval(CBIR) techniques[8][9] in mining weakly labeled web facial images.The main goal of SBFA framework 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 effectively exploit the short list of candidate facial images and their weak labels for the face name annotation task. To handle the above problem, I investigate and develop a search-based face annotation scheme. In particular, I propose a novel unsupervised label refinement (URL)scheme by using machine learning techniques to enhance the labels purely from the weakly labelled datawithout human manual efforts. I also propose a clustering-based approximation (CBA) algorithm to improve the efficiency and scalability of the annotation.

Yet another challenge faced in this paper is due to duplicate names which can be a practical issue in real-life scenarios.This problem is handled by an additional Bayesian framework with which can learn the similarity between two different names according to the web pages so as to find out how likely the two different names belong to the same person.

When the query facial image is not a well-known person,there may not exist many relevant images on the WWW,which could affect the performance.In this paper I

tackle this problem by extracting the contextual information in addition to the GIST features.

II SEARCH-BASED FACE ANNOTATION

The system flow of the proposed frameworksearch-based face annotation, 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 conducted before the test 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.The steps are briefly described below.

The first step is the data collection of facial images in which 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 are collected. As the output of this crawling process, a collection of facial images is obtained, each of them is associated with some human names. Given the nature of web images, these facial images are often noisy, which do not always correspond to the right human name.Such kind of web facial images with noisy names are called as weakly labeled facial image data.

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, the unsupervised face alignment technique is proposed. For facial feature representation, the GIST texture features are extractedalong 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 this approach, the locality sensitive hashing (LSH), a very popular and effective high-dimensional indexing technique is proposed. Besides the indexing step, another key step of the framework is to engage an unsupervised learning scheme to enhance the label quality of the weakly labeled facial images. This process is very important to the entire search based annotation framework since the label quality plays a critical factor in the final annotation performance.All the above are the processes before annotating a query facial image. Next, the process of face annotation during the test phase is described. In particular, given a query facial image for annotation, a similar face retrieval process to search for a subset of most similar faces is conducted (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 thatcombines the set of labels associated with these top K similar face examples.

In this paper, attention is on one key step of the above framework, i.e., the unsupervised learning process to refine labels of the weakly labeled facial images.

III UNSUPERVISED LEARNING

In machine learning, the unsupervised learning can be defined as that of finding hidden patterns from unlabeled or weakly labeled data. Hence in case of unsupervised learning the examples given to the learner are unlabeled, therefore there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning. The problem of density estimation in statistics lies in close contact withunsupervised learning.However unsupervised learning also includes many other techniques that are used to explain the key characteristics of the data. The methods used to pre-process the data in data mining is employed in many unsupervised learning techniques.

The few approaches that are commonly used in unsupervised learning includes:1)clustering (e.g k-means,mixture models, hierarchical clustering)2)Hidden Markov models,3)blind signal separation using feature extraction techniques(e.g., principal component analysis, independent component analysis, non-negative matrix factorization, singular value decomposition).

The main objective of SBFA framework is to assign correct name labels to a given query facial image. In particular, given a novel facial image for annotation, firsta short list of top K most similar facial images is retrieved 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.To correctly perform this retrieval an unsupervised machine learning technique called unsupervised label refinement is used.In unsupervised label refinement (ULR) no labels are given to the learning algorithm it learns on its own by considering groups of similar input.

The algorithms used for unsupervised label refinement in this paper are Multi-step Gradient algorithm for ULR and Coordinate Descent algorithm for ULR.

IV CLUSTERING-BASED APPROXIMATION

To improve the efficiency and scalability of the annotated facial image clustering-based approximation (CBA) algorithm is proposed.

Clustering items into different groups is a fundamental problem in theinformation sciences. Many typical clustering optimization problems are NP-hard and therefore cannot be expected that the problem can be solved optimally in a reasonable amount of time. Although the use of heuristics is common, in this dissertation it is better to use approximation algorithms, whose performance ratio in relation to the optimal

solution can be guaranteed and whose running time is a polynomial function of the problem instance size.

The bulk of the dissertation concerns mainly withcorrelation clustering. Correlation clustering isclustering a collection ofelements into small clusters based on similar and dissimilar features. The problem instance does not include a distance relation between the elements.The elements are partitioned into clusters so that the number of pairs either correctly or incorrectly classified withrespect to the input judgment labeling is either maximized or minimized respectively. It is worthstudying both complete instances, in which every pair is labeled, and general instances, in which some input pairs might not havelabels.

Suppose there are n facial images and m names then 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 "image-level" would be much more time-consuming than that on "name-level." Thus, in this approach, the "name-level" clustering scheme is adopted for the sake of scalability and efficiency. After the clustering step, the proposed ULR problem in each subset is solved, and then all the learning results are merged into the final enhanced label matrix.

V.BAYESIAN FRAMEWORK

A practical issue faced in annotation in the real life scenarios is due to duplicate names. This problem is rectified by an additional Bayesian framework with which the similarity between two different names according to the web pages so as to find out how likely the two different names belong to the same person is learned.Bayesian inference is a method of inferencewhich is used to update the probability estimate for a hypothesis as additional evidence is acquired by using the Bayes rule. Bayesian updating is an important technique used throughout statistics, and especially in mathematical statistics.

Face annotation is achieved with a Bayesian framework, in which the similarity of faces is measured in terms of maximum a posteriori (MAP) estimation. Bayesian framework which measures the similarity of different names by defined it as maximum a posteriori (MAP) estimation for single person annotation. Themaximum a posteriori probability (MAP) estimatein Bayesian statistics is considered asmodeof posterior distribution.On the basis of empirical data, the MAP is used to obtain a point estimate of an unobserved quantity. MAP has close relation to Fisher's method of maximum likelihood (ML), but unlike fisher's method it employs an augmented optimization objective to incorporate a prior distribution over the quantity one wants to estimate. MAP estimation can be considered as a regularization of ML estimation. MAP works on the principle of majority voting or majority ranking which name gets the highest vote or highest rank is retrieved to assign to the query image in case of duplicate names. The proposed system learns the similarity between two different names according to the web pages so as to determine howlikely the two different names belong to the same person.

In addition contextual features arealso extractedalong with the GIST feature to improve the identification of normal persons other than celebrities. This identifies the person when the query facial image is not a well-known person, and retrieves more relevant facial images on the WWW, and thus improves the annotation performance to a great extent.

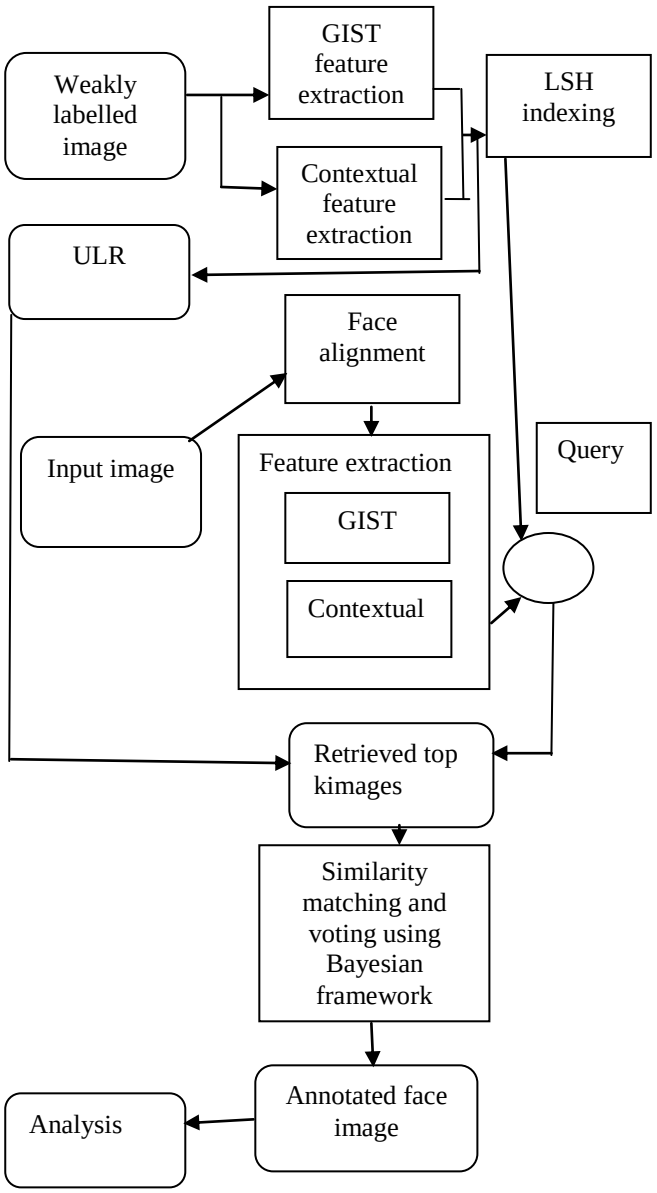

Fig 1: Architecture diagram

VI. RELATED WORK

My work is closely related to several groups of research works.

The first group is about a work that [4] presents a new framework for semi-automated face annotation in family photo album applications. This have reformulated the face annotation from a pure recognition problem to a problem of similar face search and annotation propagation and have developed a solution to this problem by seamlessly integrating content-based image retrieval and face recognition algorithms in a Bayesian framework .Difficulty arises in accurate face alignment.

The second group is about naming people in videos [2]whichPeople play an important role in most video content, and tools for identifying the individuals appearing in a video would be especially helpful. Here, we label news video with textual information extracted from the transcripts.To make the quality of the results acceptable, this work presents a semi-supervised scheme, where a few faces of each person in the database are labeled manually, and this information is then propagated to similar unlabeled faces. This is achieved using a graph based algorithm for learning the name-face alignments jointly from labeled and unlabeled examples. The problem that arises here is computational complexity.

The third group is about the studies of automated photo tagging which automatically tagsmassive unlabeled photos searchable by text search engines [7]. This work presents a retrieval based approach for automated photo tagging. To tag a test image, the proposed approach first retrieves k similarsocial images which shares the largest visual similarity with the test image. Based on the tagging of the visually similar images the tags of the test image are derived. Improved photo tagging performance is not obtained.

The fourth 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 [5] applied a modified k means clustering approach for cleaning up the noisy web facial images. Zhao et al. [10] 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, the proposed paper employs the unsupervised machine learning techniques and proposes a graph-based label refinement algorithm to optimize the label quality over the whole retrieval database in the SBFA task.

The fifth group is about the studies of Coding.In this a weak label regularized localcoordinate coding (WLRLCC) technique [9], which uses the local coordinate codingprinciple in learning sparsefeatures,and mean while employs the graph-based weak label regularization principle to enhance the weak labels of the short list of similar facial images.Limited by discriminative ability of features.

The sixth group employed supervised learning [10]in whicha framework for semantic image annotation and retrieval is proposed. Annotation and retrieval are definedas classification problems in which each class has a group of database images labeled with a common semantic labelwhich improves the robustness but difficulty arises in tagging.

The seventh group proposes a framework in which it extracts a non-rigid mapping between facial images[5]. The Lucas-Kanade image registration approach is used which employs the unsupervised face alignment technique. A robust optimization scheme is used to handle appearance variations. As unsupervised technique is used the method is fully automatic and handle the pose variations and expressions. It improvesefficiency but works on faces with similar appearances and does not focus on lighting issues in joint face alignment scheme.

The eight group proposed a framework for face annotation [15].Transductive Kernel Fisher Discriminant (TKFD) scheme for face annotation, which compared to traditional supervised annotation methods with few training data is considered to be more efficient. The main idea of this approach is to solve the Fisher's discriminant by incorporating the information of both labeled and unlabeled datausing deformed kernelsto improve the efficiency of results obtained but critical result obtained with limited amount of labeled data for large-scale face annotation tasks.

CONCLUSION

This paper investigates a promising search-based face annotation framework, in which focus is on tackling the critical problem of enhancing the label quality and proposed a ULR algorithm. To further improve the scalability,a clustering-based approximation solution is proposed, which successfully accelerated the optimization task withoutintroducing much performance degradation. In order to learn about the similarities between two different names the Bayesian framework is proposed as well as to identify not a well-known person, the contextual features are also extracted in addition to GIST features.

REFERENCES

[1]. D. Wang, S.C.H. Hoi, and Y. He, "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.

[2]. S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming andDetecting Faces in News Videos," IEEE Multimedia, vol. 6, no. 1, pp. 22-35, Jan.-Mar. 1999.

[3]. P.T. Pham, T. Tuytelaars, and M.-F. Moens, "Naming People inNews Videos with Label Propagation," IEEE Multimedia, vol. 18, no. 3, pp. 44-55, Mar. 2011.

[4]. L. Zhang, L. Chen, M. Li, and H. Zhang, "Automated Annotationof Human Faces in Family Albums," Proc. 11th ACM Int'l Conf. Multimedia (Multimedia), 2003.

[5]. 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

theNews," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 848-854, 2004.

[6]. J. Yang and A.G. Hauptmann, "Naming Every Individual in NewsVideo Monologues," Proc. 12th Ann.ACM Int'l Conf. Multimedia(Multimedia), pp. 580-587. 2004.

[7]. L. Wu, S.C.H. Hoi, R. Jin, J. Zhu, and N. Yu, "Distance Metric Learning from Uncertain Side Information for AutomatedPhoto Tagging," ACM Trans. Intelligent Systems and Technology,vol. 2, no. 2, p. 13, 2011.

[8]. S.C.H. Hoi, R. Jin, J. Zhu, and M.R. Lyu, "Semi-Supervised SVMBatch Mode Active Learning with Applications to ImageRetrieval," ACM Trans. Information Systems, vol. 27, pp. 1-29, 2009.

[9]. Dayong Wang, Steven C.H.Hoi, Ying He and Jianke Zhu "Retrieval-based Face Annotation by Weak Label Regularized Local Cordinate Coding".

[10]. Gustavo Carneiro, Antonio B.Chan, Pedro .J.Moreno and Nuno Vasconcelos "Supervised Learning of Semanitic Classes for Image Annotation and Retrieval" vol 29 No: 3 march-2007.

[11]. P. Wu, S.C.H. Hoi, P. Zhao, and Y. He, "Mining Social Images with Distance Metric Learning for Automated Image Tagging, "Proc. Fourth ACM Int'l Conf. Web Search and Data Mining (WSDM '11), pp. 197-206, 2011.

[12]. D. Wang, S.C.H. Hoi, and Y. He, "Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation," Proc. 34th Int'lACM SIGIR Conf. Research and Development in Information Retrieval(SIGIR), 2011.

[13]. P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces versusFisherfaces: Recognition Using Class Specific Linear Projection,"IEEE Pattern Analysis and Machine Intelligence, vol. 19, no. 7,pp. 711-720, July 1997.

[14]. W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "FaceRecognition: A Literature Survey," ACM Computing Survey, vol. 35, pp. 399-458, 2003.

[15]. J. Zhu, S.C.H. Hoi, and M.R. Lyu, "Face Annotation Using Transductive Kernel Fisher Discriminant," IEEE Trans. Multimedia, vol. 10, no. 1, pp. 86-96, Jan. 2008.

[16]. Y. Tian, W. Liu, R. Xiao, F. Wen, and X. Tang, "A Face Annotation Framework with Partial Clustering and Interactive Labeling," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR),2007.

[17]. J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang, "EasyAlbum: An Interactive Photo Annotation System Based on Face Clustering and Re-Ranking," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), pp. 367-376, 2007.

[18]. D. Anguelov, K. Chih Lee, S.B. Go¨ktu¨ rk, and B. Sumengen, "Contextual Identity Recognition in Personal Photo Albums," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '07), 2007.

[19]. 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 Trans. Multimedia, vol. 13, no. 1, pp. 14-28, Feb. 2011.

[20]. D. Ozkan and P. Duygulu, "A Graph Based Approach for Naming Faces in News Photos," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1477-1482, 2006.

