TAG ME: An Accurate Name Tagging System For Web Facial Images using Search-Based Face Annotation

Mr. Ansari Mohammed Abdul Qadir[1], Prof. R.P. Dahake[2]

[1] *PG Student, Department of Computer Engineering, MET'S Institute of Engineering, BKC Adgaon, Nashik, Maharashtra, India*
[2] *Associate Professor, Department of Computer Engineering, MET'S Institute of Engineering, BKC Adgao, Nashik, Maharashtra, India*

ABSTRACT

Search based face annotation (SBFA) is an effective technique to annotate the weakly labeled facial images that are freely available on World Wide Web. The main objective of search based face annotation is to assign correct name labels to given query facial image. One drawback for search based face annotation theme is how to effectively perform annotation by exploiting the list of most similar facial pictures and their weak labels that are incomplete. To tackle this drawback, this system introduced a good unsupervised label refinement (URL) approach for purification the labels of web facial pictures exploitation machine learning technique. To additional speed up the projected theme, the clustering based approximation is introduced which may improve quantify ability significantly.

Keyword : - *Face annotation, SBFA, machine learning, unsupervised label refinement, web facial images, weak label.*

---

1. INTRODUCTION

Due to the recognition of varied digital cameras and also the rapid growth of social media tools for internet-based photo sharing, recent years have witnessed associate explosion of the amount of digital photos captured and hold on by consumers. An outsized portion of photos shared by users on the Internet area unit human facial pictures. A number of these facial images area unit labeled with names, however several of them don't seem to be tagged properly. This has actuated the study of automatic face annotation, a vital technique that aims to annotate facial pictures. Auto face annotation may be useful to several real world sites (e.g., Facebook) will automatically annotate users' Uploaded photos to facilitate online icon search and management. Besides, face annotation may also be applied in news video domain to sight important persons appeared within the videos to facilitate news video retrieval and account tasks. Classical face annotation approaches area unit usually treated as an extended face recognition downside, wherever completely different classification models area unit trained from a group of well labeled facial pictures by using the supervised or semi-supervised machine learning techniques. However the "model-based face annotation" techniques area unit restricted in many aspects. First, it's typically time-consuming and high priced to gather an outsized quantity of human-labeled trained facial pictures. Second, it's typically difficult to generalize the models once a new trained knowledge or new person's area unit are formed, within which associate intensive training process is typically needed. Last however not least, the annotation/recognition performance usually scales poorly when the amount of persons/classes is extremely massive.

Recently, some rising studies have tried to explore a promising search-based annotation paradigm for facial image annotation by mining the globe Wide net (WWW), wherever a vast variety of weakly labeled facial images area unit freely accessible. Rather than coaching express classification models by the regular model-based face

annotation approaches, the search-based face annotation (SBFA) paradigm aims to tackle the machine-controlled face annotation task by exploiting content-based image retrieval (CBIR) techniques in mining huge weakly labeled facial pictures on the net. The search-based face annotation (SBFA) framework is data-driven and model-free, that to some extent is inspired by the search-based image annotation technique for generic image annotations. The main objective of search-based face annotation (SBFA) is to assign correct name labels to a given question facial image. Particularly, given a unique facial image for annotation, we tend to initial retrieve a brief list of prime K most similar facial pictures from weakly labeled facial image information, and then annotate the facial image by performing ballot on the labels related to the top k similar facial pictures. One challenge Janus-faced by such search-based face annotation (SBFA) paradigm is a way to effectively exploit the listing of candidate facial pictures and their weak labels for the face name annotation task. To tackle the higher than downside, we tend to investigate and develop a search -based face annotation theme. Particularly, we propose a unique unattended label refinement (URL) scheme by exploring machine learning techniques to enhance the labels strictly from the decrepit labeled knowledge without human manual efforts.

The system is broadly classified into 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.
6. Face annotation by majority voting on the similar faces with the refined labels.

The first four steps are usually 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. Let us see each step one by one.

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

The second step is to preprocess 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 system adopts the unsupervised face alignment technique proposed in [23]. For facial feature representation, the system extracts the GIST 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 the system approach, system adopt the locality sensitive hashing (LSH), a very popular and effective high-dimensional indexing technique. 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, now the system describes the process of face annotation during the test phase. In particular, given a query facial image for annotation, the system 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. In this topic, the system focuses our attention on one key step of the above framework, i.e., the unsupervised learning process to refine labels of the weakly labeled facial images.

2. RELATED WORK

This System is closely related to several groups of research work now let us see these groups one by one in details:

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 [03], [04]. 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 [05], [06], [07].

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. [09] have proposed to refine the model based annotation results with a label similarity graph by following random walk principle. Similarly, Pham et al. [10] have proposed to annotate unlabeled 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. . For example, Russell et al. [11] 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. Un-like these existing works, this system shows 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 in [12], [13], [14] have mainly focused on the annotation task on personal photos, which often contain rich contextual clues, such as personal/family names, social context, geotags, 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 easyAlbum, 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 [15] have explained the concept of a graph-based model for finding the densest subgraph as the most related result. Following the graph-based approach, Le and Satoh [16] have worked on a new local density score to represent the importance of each returned images, and Guillaumin et al. [17] have 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 [18] to improve over the generative approach and avoid the explicit computation in graph-based approach. By using ideas from query expansion [19], 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. [20] have thrown lights on 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), Guillaumin 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. The work here 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 [21] 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 searched based face annotation (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 preprocessing step in the whole system without adopting sophisticated techniques. For example, the work in [02] applied a modified k-means clustering approach for cleaning up the noisy web facial images. Zhao et al. [22] 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 system 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 searched based face annotation (SBFA) task.

3. PROCEEDING METHODOLOGY

3.1 System Overview

TAGME system flow architecture is shown in following figure the overall system is divided into three modules:

a) Facial Image Collection.

b) Indexing and Learning.

c) Annotation and Retrieval.

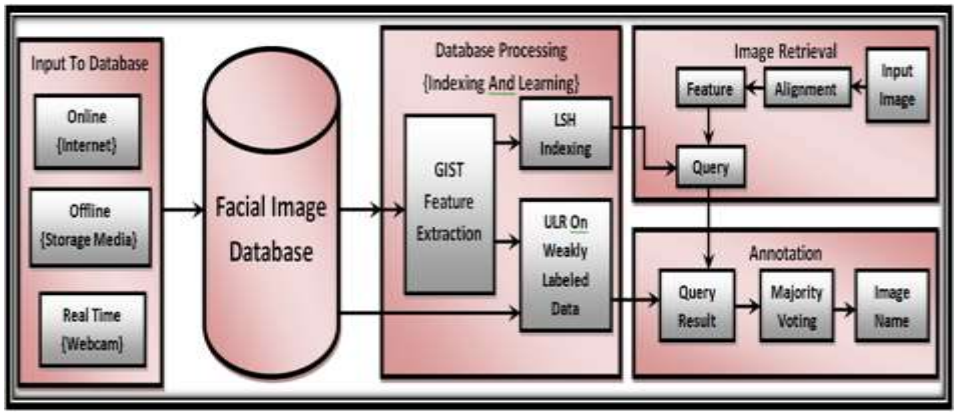

Fig. 1.TAGME System Flow Architecture

The short descriptions of these modules are as follows:

(a) Weakly labeled facial images collection from WWW using web search engines.

(b) Preprocess the facial images crawled from the web, including face detection, face alignment, and feature extraction for the detected faces, then we apply LSH indexing to index the draw out high-dimensional facial features. We employ ULR method to purify the raw weak labels composed with the clustering-based approximation algorithms for bettering the scalability.

(c) Search for the query image to save the best K identical images and use their associated names for voting toward auto annotation.

The above steps are broadly classified as:

1. Facial image database (weakly labeled images).
2. Face alignment, detection and feature extraction.
3. Indexing of facial feature.
4. Purifying weakly labeled data by learning.
5. Similar face retrieval.
6. Face annotation using majority voting on the similar faces with the refined labels.

From the above steps first to four steps are generally administered earlier the evaluation test of a face annotation task, as long as the final two steps are administered along the test phase of a face annotation task, which generally should be done very efficiently. We briefly describe each step below.

The starting step is the facial image database as shown in Fig. 1, in which we crawled facial images from the web by using web search engine (i.e., Google) this is the collection of images used for image database and these facial images is associated with some human names. As we know these images are usually noisy, which do not

constantly agree with the right human name, hence these types of the facial images are called weakly labeled facial images.

The next step is to preprocessing the web facial images to obtain face related details or information, including face alignment, detection, facial region extraction and facial feature representation. For face alignment and detection, the system adopts the unsupervised face alignment scheme discussed in [23]. For facial feature representation, we used the GIST texture features to produce the extracted faces. By using the result of GIST each face can be depicted by a d-dimensional feature vector.

The next step is to perform indexing on the depicted features of the faces by employing few adept high-dimensional indexing techniques to promote the task of very much alike face retrieval in the consequent step. According to the system approach, system used the locality sensitive hashing (LSH), it is most famous and effective high-dimensional indexing technique. In addition to the indexing step, another key step of the framework is to employ an unsupervised learning scheme to improve the label quality of the facial images which are weakly labeled. This action is very essential to the whole search based annotation framework after all the label quality plays a detracting factor in the final annotation performance.

All of the above are the processes before annotating a query facial image. Now we define the face annotation process during the test phase. In specific, first we gives a query image i.e. facial image for annotation, firstly the system administer a very much alike face retrieval process to search for a subset of best similar faces (usually best K similar face examples) from the earlier indexed facial database. With the collection of best K similar face examples reclaim from the database, the next step is the annotation of the facial image with a label (or a subset of labels) by using a voting approach based on the majority concept which is consist of the combination of the set of labels related with these best K similar face examples. In this topic, the system focuses our attention to one of the key step for the framework discuss above, i.e., the unsupervised learning process to purify labels of the weakly labeled facial images.

3.2 Algorithms

1) Multistep Gradient Algorithm for ULR

Multi-Step Gradient Algorithm is an efficient algorithm to solve the problem in SRF i.e. here sparsity is not taken into consideration but it is efficient to solve the optimization tasks and also this algorithm efficiently solve the quadratic programming problem. The pseudo code or Multi-step Gradient Algorithm is summarized in Algorithm1.

Parameters	Meaning
$Q \in R^{(n.m)x(n.m)}$	Vectorizing Matrix
x^*	Optimal Solution
K	No. of Iteration
$Z^{(k)}$	Search Point
T	Lipshitz Constant
α_0	Regularization Parameter, always $\alpha>0$.

Table 1. Parameters Used in Algorithm 1.

Algorithm 1 : Multi-step Gradient Algorithm for ULR

Input: $Q \in \mathbb{R}^{(n.m)\times(n.m)}, \mathbf{c} \in \mathbb{R}^{n.m}, t \in \mathbb{R}$

Output: $\mathbf{x}^*$

begin

$\alpha_0 = 1; k = 1; \mathbf{z}^{(0)} = \mathbf{x}^{(0)} = \mathbf{x}^{-1} = 0;$

repeat

Case SRF : Achieve $\mathbf{x}^{(k)}$;

Case CCF : Achieve $\mathbf{x}^{(k)}$;

$\alpha_k = \frac{1+\sqrt{4\alpha_{k-1}^2+1}}{2}$;

$\mathbf{z}^{(k)} = \mathbf{x}^k + \frac{\alpha_{k-1}-1}{\alpha_k}(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$;

$k = k + 1$;

until *CONVERGENCE*;

2) Coordinate Descent Algorithm for URL

Co-ordinate Descendent Algorithm is an efficient algorithm to solve the problem in SRF as well as in CCF i.e. here sparsity is taken into consideration and it is efficient to solve the optimization tasks also this algorithm efficiently solve the quadratic programming problem, the difference between SRF and CCF is that here the scalability is further improved. This algorithm can take the advantages of the power of parallel computation when solving a very large-scale problem. The pseudo code for Multi-step Gradient Algorithm is summarized in Algorithm 2.

Parameters	Meaning
$X \in R^{nxd}$	Extracted Facial Image Features Matrix.
N	No. of Facial Images
D	No. of Feature Dimensions
$Y \in [0,1]^{nxm}$	Initial Raw Label Matrix
$F^* \in R^{nxm}$	Refined Label Matrix

Table 2. Parameters Used in Algorithm 2.

Algorithm 2 : Co-ordinate Descent Algorithm for ULR

Input: $X \in \mathbb{R}^{n\times d}, Y \in [0,1]^{n\times m}$

Output: $F^* \in \mathbb{R}^{n\times m}$

begin
  $t = 0$ and $F^{(t)} = Y$;
repeat
for $i = 1$ **to** n **do**
    Case SRF : Achieve $F_{i*}^{(t+1)}$;
    Case CCF : Achieve $F_{i*}^{(t+1)}$;
end for
  $t = t + 1$;
until *CONVERGENCE*;

3.3 Expected Result

According to the input image face detection and face alignment can be done and successful implementation of content based searched. Other users can search the image using content based search. Other users can provide rating to the searched images which can used to re-rank images in future searches. Face annotation using percentage of matching or using user rating are possible. Statistical reports will be generated. User friendly environment help will be there for each module in the system.

4. DISSCUSSION & ANALYSIS

As per the above discuss literature survey or reference papers, and base paper, the above system can be implemented using following steps:

(a) Collect weakly labeled facial images from WWW, real-time and offline sources.

(b) Preprocess the crawled web facial images, including face detection, face alignment for the detected faces, after that apply the proposed ULR method to refine the raw weak labels together with the proposed clustering-based approximation algorithms for improving the scalability.

(c) Search for the query facial image to retrieve the top K similar images and use their associated

5. CONCLUSIONS

Accurate name tagging system for Web facial images using search-based face annotation mainly focused on real life problem of name tagging over social media, over the internet and also on tackling the critical problem of enhancing the label quality. Second advantage of the system is the improvement in the scalability and it also uses a clustering-based approximation solution, which successfully accelerated the optimization task without any performance degradation.

6. REFERENCES

[1] Dayong Wang, Steven C.H. Hoi,Ying He,Jianke Zhu, Mining Weakly Labeled Web Facial Images For Searched-Based Face Annotation, IEEE Transaction on Knowledge and Data Engineering, Vol-26 2014.

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

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

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

[5] 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, technical report 07-49, 2007.

[6] H.V. Nguyen and L. Bai, Cosine Similarity Metric Learning for Face Verification, Proc. 10th Asian Conf. Computer Vision (ACCV 10), 2008.

[7] M. Guillaumin, J. Verbeek, and C. Schmid, Is that You? Metric Learning Approaches for Face Identification, Proc. IEEE 12th Intl Conf. Computer Vision (ICCV), 2009.

[8] Z. Cao, Q. Yin, X. Tang, and J. Sun, Face Recognition with Learning Based Descriptor, IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 2707-2714, 2010.

[9] C. Wang, F. Jing, L. Zhang, and H.-J. Zhang, Image Annotation Refinement Using Random Walk with Restarts, 14th Ann. ACM Intl Conf. Multimedia, pp. 647-650, 2006.

[10] P. Pham, M.-F. Moens, and T. Tuytelaars, Naming Persons in News Video with Label Propagation, Proc. VCIDS, pp. 1528-1533, 2010.

[11] B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman, LabelMe: A Database and Web-Based Tool for Image Annotation, Intl J. Computer Vision, vol. 77, nos. 1-3, pp. 157-173, 2008.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] D.-D. Le and S. Satoh, Unsupervised Face Annotation by Mining the Web, Proc. IEEE Eighth Intl Conf. Data Mining (ICDM), pp. 383-392, 2008.

[17] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, Automatic Face Naming with Caption-Based Supervision, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2008.

[18] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, Face Recognition from Caption-Based Supervision, Intl J. Computer Vision, vol. 96, pp. 64-82, 2011.

[19] T. Mensink and J.J. Verbeek, Improving People Search Using Query Expansions, Proc. 10th European Conf. Computer Vision (ECCV), vol. 2, pp. 86-99, 2008.

[20] T.L. Berg, A.C. Berg, J. Edwards, and D. Forsyth, Who's in the Picture, Proc. Neural Information Processing Systems Conf. (NIPS), 2005.

[21] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, Scalable Face Image Retrieval with Identity Based Quantization and Multi-Reference Re-Ranking, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 3469-3476, 2010.

[22] M. Zhao, J. Yagnik, H. Adam, and D. Bau, Large Scale Learning and Recognition of Faces in Web Videos, Proc. IEEE Eighth Intl Conf. Automatic Face and Gesture Recognition (FG), pp. 1-7, 2008.

[23] J. Zhu, S.C.H. Hoi, and L.V. Gool, Unsupervised Face Alignment by Robust Non rigid Mapping, Proc. 12th Intl Conf. Computer Vision (ICCV), 2009.