ISSN:0975-9646

Search Based Face Annotation Using Weakly Labeled Facial Images

Shital Shinde[1], Archana Chaugule[2]

[1]*Computer Department, Savitribai Phule Pune University*
D.Y.Patil Institute of Engineering and Technology, Pimpri, Pune-18
Mahatma Phulenagar, 120/2 Mahaganpati soc, Chinchwad, Pune-19, MH, India

[2] *D.Y.Patil Institute of Engineering and Technology, Pimpri, Pune-18, Savitribai Phule Pune University*
DYPIET, Pimpri, Pune-18, MH, India

***Abstract*— Automated face annotation aims to automatically detect human faces from a photo and further name the faces with the corresponding human names. Most of the images are obtained from World Wide Web (WWW), so there is possibility of getting noisy and incomplete images. This study tackle problem by investigating a search-based face annotation (SBFA) paradigm for mining large amounts of facial images. Given a query facial image for annotation, the idea of SBFA is to first search for top-n similar facial images from facial image database and then exploits these top-ranked similar facial images and their weak labels for naming the query facial image. Major challenging task of search based annotation scheme is how to perform effective annotation by exploiting the list of most similar facial images and their weak labels. Unsupervised label refinement (ULR) approach is used for refining the labels of facial images. In this proposed work unsupervised label refinement approach effectively solves the problem of weak labelling. A clustering-based approximation algorithm is proposed to improve the scalability of naming and images. In future it will evaluate a larger database to develop more efficient solutions.**

***Keywords*— Search based face annotation, unsupervised label refinement, clustering based approximation, weak label, auto annotation, face annotation.**

I. INTRODUCTION

Day by the digital accessories for capturing images is increasing and sources for sharing that images e.g., social media tools/sites are also increasing [1, 13]. Large numbers of images are shared over these social sites, but many times images shared by person are not having any label so it becomes problematic in understanding name of person from image if any random person sees it. The main aim of image annotation process is to automatically assign associate label to images, so image retrieving users are able to query images by labels and automatically detect human faces from a photo image and further name the faces with the corresponding human names.

Tagging facial images are known as face annotation and now a day many techniques are introducing for annotation. Auto face annotation is used for automatic face image annotation without any human intervention [2, 7, 8, 17]. Facial annotation is also applying for videos, such as annotation of facial images from news video is done and then it showed on television so that peoples can recognize person in TV [3, 10].

The model base annotation has more limitations i.e. it is more time consuming and more costly to collect large amount of human labeled training facial image. It is more difficult to generalize the models when new persons are added, in which retraining process is required and last the annotation performance is become poor when the number of persons is very more. To address the challenges "Auto face annotation" is important technique which automatically gives name to relevant person images [2, 7]. This technique is more beneficial to different real world application of search based face annotation. The main objective of search-based face annotation is to assign correct name labels to a given query facial image. There are several groups of research work, research is carried for facial images recognition, feature exaction, annotation etc. Various techniques are applied for face recognition, face annotation e.g. cluster based possibility model, graph based approach [16], query expansion [19], retrieval based face annotation [23], and search based face annotation [1, 7, 25] schemes.

II. LITERATURE SURVEY

Various techniques are present for face annotation in mining weakly labelled facial images from www. This study shows that most of the techniques those accepts name of person as input and process text-based search for achieving face images. Many researchers are trying to design system, which accepts image-based input data and produces text based output. Various research groups are working for successfully fulfilment of this objective.

These are classical research problems in computer vision and pattern recognition and have been studied from many years. G.B. Huang et al. [5] designed Labeled Faces in the Wild. Z. Cao et al. [6] presented a novel approach to address the representation and the matching issue in face recognition. In proposed work they firstly worked on approach that encodes the micro-structures of the face by a new learning-based encoding technique. They used unsupervised learning schemes to learn an encoder from the training sets. In next step they applied PCA technique to get a compact face descriptor. Results obtained shows that study done by them suggested that the discriminative ability of the descriptor can be improved by a simple normalization mechanism after PCA. To handle the large pose variation in real-life scenarios case, they proposed a

pose-adaptive matching method which uses pose-specific classifiers to deal with various pose combinations of the matching facial image pair. This studied approach is comparable with the state-of the-art methods on the Labeled Face in Wild.

X.-J. Wang et al. presented AnnoSearch scheme, a novel scheme to annotate images using search and data mining technologies. They solved this problem in two-stages; first is searching for semantically and visually similar images on the internet, and mining annotations from annotation. One accurate keyword is necessary for enabling text-based search for a set of semantically similar images. After that content related search task is performed on this retrieved set to obtain visually similar images. As final processing step of their work, annotations are mined from the descriptions. These proposed works has benefit that is no supervised training process is adopted in this process, and as a result, it handles large amount of vocabulary. It also ensures a highly scalable image database. In their work they mentioned that in follow-up work they will work on reinforcing the labels of images from large scale database and they are interested to resolve the problem of how to annotate query images without associated keywords [7].

J. Tang et al. [8] proposed a novel kNN-sparse graph based semi-supervised learning method with regularization on number of training labels, which is used to annotate various noisily-tagged web images by label propagation. Here the graph is constructed to handle the semantically different links. It is generated by reconstructing each and every sample from its k nearest neighbours to improve the efficiency, and in the same study the approximate method is applied to accelerate the kNN search. And the regularization is proposed to handle the noise in the training labels. Experimental results of this study showed a key factor, which affects the performance of image annotation process with the tags as trained labels. Actually, in image annotation scheme, there is no need to correct all the noisy tags; they collected the correct image label pairs as much as possible for training. They also decided to focus on how to construct an effective training set from the community-contributed images and tags in future work.

Guillaumin et al. [14] introduced a modification to incorporate the constraint that a face is only depicted once in an image. This work has two scenarios of naming persons in database for finding face of person and assigning name to all faces. The text based result is not more accurate. Graph based approach is improved by introducing the constraint, objective function generative models have previously been proposed to solve the multi-person naming task by comparing generative and graph based methods. Graph based method is extended in future to multi-person naming. M. Guillaumin et al. [15] proposed a method to iteratively update the assignment based on a minimum cost matching algorithms. In follow-up work of this, they further uses metric learning technique to enhance the annotation performance to gain lots of distinguish features in low dimension space [17, 18].

Z. Wu et al. mainly addressed the face retrieval problem; by using local and global features it proposed an effective image representation from study. Future Work visual word vocabulary for face is improved by designing a supervised learning algorithm. They proposed highly scalable system, and they planned it by using a computer cluster to apply on a web-scale image database [21]. Retrieval-Based Annotation Approach, Shows that retrieval based approach are applied with distance metric learning also various different techniques are implemented with these retrieval based or search based face annotation [23].

M. Zhao et al. [22] proposed a system that can learn and recognize faces by combining signals from large scale weakly labeled text, image, and video. First, consistency learning is proposed to create face models for popular persons. It uses the text-image co-occurrence on the web as a weak signal of relevance and learns the set of consistent face models from this very large and noisy training set. It recognizes peoples in videos; they applied face detection and tracking to extract faces from various videos. And then, key faces are selected for each track for fast and robust recognition. Face tracks are further clustered to get more compact and robust representation. The face tracks are clustered to get more representative key faces and remove duplicate key faces. Majority voting and probabilistic voting algorithms are combined to recognize each cluster of face tracks. They studied various active learning possibilities in case of improving the recognizer to grow across age variations. Proposed work provides another direction which would be shows that how to combine high precision face-based retrieval scheme and high-recall text based retrieval scheme. D. Wang, S.C.H. Hoi, and Y. He [24] this proposed work adopted a unified framework of Unifying Transductive and Inductive Learning.

There are two key points that differentiate the proposed work from existing work. First key point is that it is used to solve general content based face annotation problem using search based face annotation scheme where face image is used as query image. Second point is that unsupervised label refinement algorithm is used as suitable algorithm for enhancement of new label matrix for weakly labeled facial images [1]. This work related to recent work of the WLRLCC method [23].

III. SEARCH BASED FACE ANNOTATION

System Architecture:

In this section, it briefly introduce the framework of SBFA system. Figure 1 illustrates the proposed framework that consists of the following four major modules:

(1) The database construction module by crawling facial images from the World Wide Web;
(2) The database indexing module towards fast retrieval of high-dimensional facial features;
(3) The content-based facial image retrieval module for searching a query facial image;
(4) The automated face annotation module for naming the query by mining the top-k retrieved similar facial images and their corresponding weakly 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. It can

annotate the query facial image by exploiting both weakly label information and visual contents of top-ranked facial images to maximize the annotation efficiency.

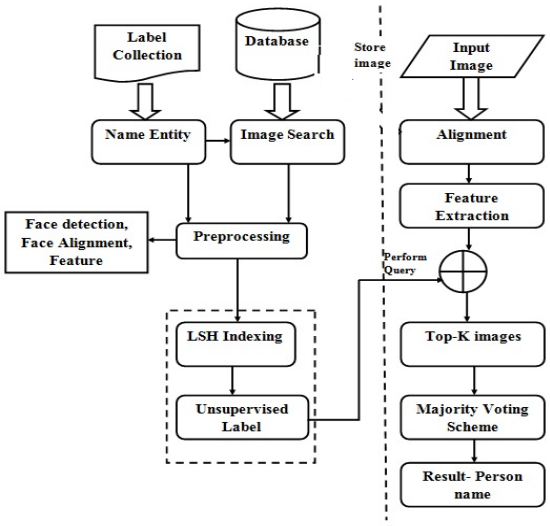


Figure1: Search Based Face Annotation System Flow

System Algorithms:

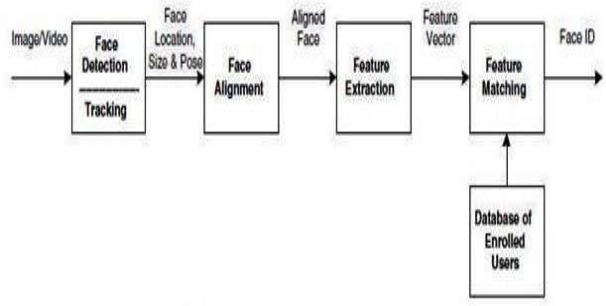


Figure 2: Face recognition processing flow

Face recognition: The face recognition problem can be divided into two main stages: face verification (or authentication), and face identification (or recognition). The detection stage is the first stage; it includes identifying and locating a face in an image. The recognition stage is the second stage; it includes feature extraction, where important information for discrimination is saved, and the matching, where the recognition result is given with the aid of a face database. Several face recognition methods have been proposed. In the vast literature on the topic there are different classifications of the existing techniques. The following is one possible high-level classification:

Holistic Methods- The whole face image is used as the raw input to the recognition system. An example is the well-known PCA-based technique.

Local Feature-based Methods*:* Local features are extracted, such as eyes, nose and mouth. Their locations and local statistics (appearance) are the input to the recognition stage.

Pre-processing: It pre-process facial images to extract face-related information, including face detection and alignment, facial region extraction, and facial feature representation. As a result, each face can be represented by a d-dimensional feature vector. An object recognizer using PCA (Principle Components Analysis):-

1. Get the Eigen vectors that form the Eigen space
2. Get or set the labels for the corresponding training image
3. Get or set the Eigen distance threshold.
4. Get the average Image.
5. Get the Eigen values of each of the training image
6. Create an object recognizer using the specific training data and parameters, it will always return the most similar object.
7. Create an object recognizer using the specific training data and parameters
8. Given the Eigen value, reconstruct the projected image.
9. Get the Euclidean eigen-distance between input image and every other image in the database
10. Given the image to be examined, find in the database the most similar object, return the index and the Eigen distance
11. Try to recognize the image and return its label.

Unsupervised Face Refinement: It is 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. It denoted by X IRnxd the extracted facial image features, where n and d represent the number of facial images and the number of feature dimensions, respectively. Further it denoted by Ω= {n1, n2, . . .,; nm} the list of human names for annotation, where m is the total number of human names. It denotes Y [0; 1] nxm as the initial raw label matrix to describe the weak label information, in which the ith row Y represents the label vector of the ith facial image X IRd. In this application, Y is often noisy and incomplete.

In particular, for each weak label value Yij, Yij ≠ 0 indicates that the ith facial image Xi has the label name n, while Yij ≠ 0 indicates that the relationship between ith facial image Xi and jth name is unknown. Note that it usually have Yi* = 1 since each facial image in database was uniquely collected by a single query. Following the terminology of graph-based learning methodology, build a sparse graph by computing the weight matrix W = [Wij] IRnxn, where Wij represents the similarity between Xi and Xj.

Clustering base Approximation: To further enhance the scalability and efficiency in algorithms, in this paper, proposed 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 “image level” would be much more time-consuming than that on

"name-level." Thus, this approach, adopt the "name level" clustering scheme for the sake of scalability and efficiency. After the clustering step, solve the proposed ULR problem in each subset, and then merge all the learning results into the final enhanced label matrix F.

Algorithm:

Input: c, qc, Lloop

Output: clustering highest order result list Llist

Add M0 to Llist;

Repeat

Remove the largest cluster M_l from L_{list};

For i=1 to t do

Bisect M_l to $M_1^{(i)}$ and $M_2^{(i)}$;

Compute sum of squared error (SSE_i);

Select the result with the lowest SSE_i value;

Add $m_1^{(i)}$,$m_2^{(i)}$ to L_{list};

Until $|L_{list}|=q_c$;

In the Clustering scheme, the ith row C_{i*} is used as the feature vector for class Xi. In each step, the largest cluster is bisected for I_{loop} times and the clustering result with the lowest sum-of-square-error (SSE) value is used to update the clustering lists. In our framework, we set I_{loop} to 10. The details of the Bisect Clustering Based Approximation (BCBA) scheme are illustrated in above algorithm, where qc is the cluster number.

Weighted Majority Voting (WMV): It is used to combines the set of labels associated with these top K similar face examples. It is based on the top-n retrieval images. The confidence weight depends on the Euclidean distance between the query image and the similarity image. In particular, for the i-th nearest similar face, it assigns a weight coefficient wi to the corresponding label vector yi by $w_i = \frac{\phi(X_q, X_i)}{\sum_{j=1}^{n} \phi(X_q, X_j)}$

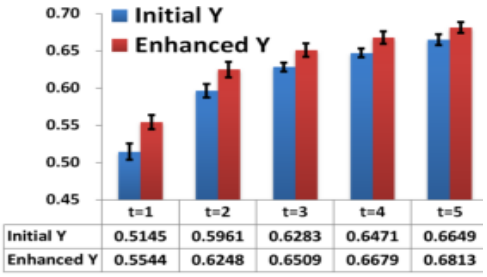
where $\phi(\cdot, \cdot)$ is related to the distance between the query image x_q and its i-th nearest sample x_i .

$$\phi(X_q, X_t) = \frac{1}{1+\exp(-\|X_q - X_t\|^2)}$$

IV. Results

Evaluation of Label Enhancement:

This experiment aims to evaluate the performance of the refined label matrix Y. It presents the results of the proposed SBFA algorithm on the most noisy DB-100K database. Similar observations can also been observed on the other databases by employing different algorithms. The ULR, CBA algorithm is applied to learn and enhance the label matrix. After that, both the refined label matrix Y and the initial weak label matrix are used in the face name annotation step. The comparison results are presented in Figure 3, it represents evaluation of the enhanced label matrix Y on the DB-100K database.



	t=1	t=2	t=3	t=4	t=5
Initial Y	0.5145	0.5961	0.6283	0.6471	0.6649
Enhanced Y	0.5544	0.6248	0.6509	0.6679	0.6813

Figure 3: Evaluation of the enhanced label matrix Y on the DB-100K database.

Next experiment is to demonstrate the qualitative tagging performance achieved by different SBFA and ULR methods for automatic image tagging tasks. To achieve this purpose, it randomly select several images from the test set, and applied a number of different ULR methods to annotate them using the proposed search based annotation approach.

Evaluation of Varied k Values:

Figure 4 shows the performance of SBFA at top t tags by varying k, the number of top retrieved similar images from 10 to 60. From the results, observed that k affects the annotation performance. In particular, when k is about 40 to 50, the proposed method achieved the best average precision. This is reasonable because if k is too small, some relevant images may not be retrieved, while if k is too large, lots of irrelevant images could be retrieved, leading to engage many noisy tags in the list of candidate tags. Both of the above situations could degrade the annotation performance.

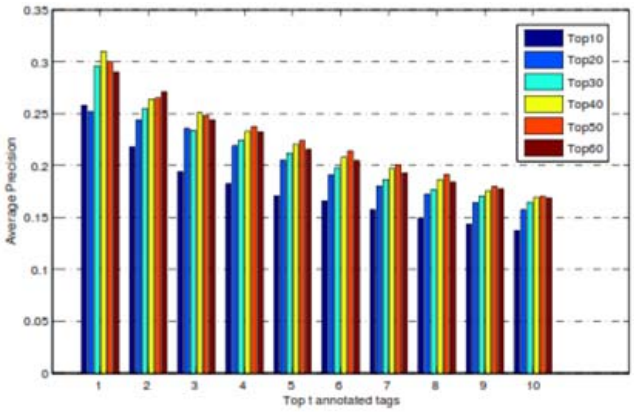


Figure 4: Comparisons of average precision under different top k similar images used

V. Conclusions

It investigate a promising search-based face annotation framework, in which it focused on tackling the critical problem of enhancing the label quality and proposed a ULR algorithm. To further improve the scalability, it also proposed a clustering-based approximation solution, which successfully accelerated the optimization task without introducing much performance degradation. It uses unsupervised labeled refinement (ULR) approach for improving efficiency and scalability of proposed system. Results obtained shows that the proposed ULR technique can significantly boost the performance of the promising search-based face annotation scheme.

REFERENCES

[1] Dayong Wang, Steven C. H. Hoi, Ying He, and Jianke Zhu, "Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation", IEEE Transactions On Knowledge And Data Engineering, vol. 26, no. 1, January 2014.

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

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

[4] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[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] 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.

[7] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma, "AnnoSearch: Image Auto Annotation by Search," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 14831490, 2006.

[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 Trans. Intelligent Systems and Technology, vol. 2, pp. 14:1-14:15, 2011.

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

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[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," Int'l 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] D. Wang, S.C.H. Hoi, Y. He, and J. Zhu, "Retrieval-Based Face Annotation by Weak Label Regularized Local Coordinate Coding," Proc. 19th ACM Int'l Conf. Multimedia (Multimedia), pp. 353362, 2011.

[24] D. Wang, S.C.H. Hoi, and Y. He, "A Unified Learning Framework for Auto Face Annotation by Mining Web Facial Images," Proc. 21st ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 1392-1401, 2012.

[25] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li, "Modeling LSH for Performance Tuning," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 669-678, 2008.

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