International Journal of Innovative Technologies
WWW.IJITECH.ORG
ISSN 2321-8665
Vol.03,Issue.01,
May-2015,
Pages:0022-0025

A Framework for Search-Based Face Annotation by Mining Weakly Labeled Facial Images

K. E. NARESH KUMAR[1], R. SUDESHNA[2]
[1]Assistant Professor, Dept of CSE, RGMCET, Nandyala, AP, India, E-mail: nareshforgovt@gmail.com.
[2]PG Scholar, Dept of CSE, RGMCET, Nandyala, AP, India, E-mail: chinna.suna@gmail.com.

Abstract: This paper will investigates a framework of search-based face annotation (SBFA) by mining weakly labeled facial images which are freely available present on the World Wide Web (WWW). For search-based face annotation scheme there is one challenging problem that is how to effectively perform annotation by exploiting the list of most similar facial images and their weak labels which are often noisy and incomplete. In order to tackle this problem, an effective unsupervised label refinement (ULR) approach for refining the labels of web facial images using machine learning techniques were proposed. The learning problem is formulated as a convex optimization and develops effective optimization algorithms that are used to solve the large-scale learning task more efficiently. In order to speed up the proposed scheme, a clustering-based approximation algorithm is proposed that can improve the scalability considerably. An extensive set of empirical studies have been conducted on a large-scale web facial image tested, in which the encouraging results will show that the proposed ULR algorithms will significantly boost the performance of the promising SBFA scheme.

Keywords: Face Annotation, Machine Learning, Content-Based Image Retrieval, Label Refinement, Weak Label, Web Facial Images.

I. INTRODUCTION

In Recent years, an explosion of the number of digital photos captured and stored by consumers have witnessed due to the popularity of various digital cameras and the rapid growth of social media tools for internet-based photo sharing. A large portion of photos were shared by users in the Internet are usually human facial images. Some of these facial images are tagged along with their names, but many of them are not tagged their names properly. This has motivated for the study of auto face annotation, which is an important technique that usually aims to annotate facial images automatically. To many real world applications, the auto face annotation can be beneficial. For example, with the help of auto face annotation techniques, online photo-sharing sites (e.g., Face book) can automatically annotate users' uploaded photos in order to facilitate online photo search and management. Face annotation can also be applied in news video domain in order to detect important persons which are appeared in the videos for facilitating news video retrieval and summarization tasks. Classical face annotation approaches are often considered as an extended face recognition problem, in which different classification models are trained from a collection of well labeled facial images by applying the supervised or semi-supervised machine learning techniques. However, in different aspects the "model-based face annotation" Techniques are limited.

Initially, it is time-consuming and expensive for collecting the large amount of human-labeled training facial images. Secondly, it is difficult for generalizing the models an intensive retraining process is usually required when new training data or new persons are added. Lastly, when the number of persons or classes is very large the annotation or recognition performance often scales poorly. To explore a promising search-based annotation paradigm for facial image annotation by mining the World Wide Web (WWW),some emerging studies have attempted where a massive number of weakly labeled facial images are freely available. Instead of training explicit classification models which are obtained by the regular model-based face annotation approaches, the search-based face annotation (SBFA) paradigm will aim to tackle the automated face annotation task for exploiting content-based image retrieval (CBIR) techniques, in case of mining massive weakly labeled facial images present on the web. The SBFA framework is model-free and data-driven, which is inspired by the search-based image annotation techniques for generic image annotations. Assigning correct name labels to a given query facial image is the main objective of SBFA.

In particular, given a novel facial image for annotation process, a short list of top K most similar facial images are retrieved from a weakly labeled facial image database, and then it will annotate the facial image by performing voting based on the labels which are associated with the top K similar facial images. One challenge is faced by SBFA paradigm that how to effectively exploit the short list of candidate facial images and their weak labels are used for the face name annotation task. In order to tackle the above problem, a search-based face annotation scheme is investigated and developed. In particular, a novel

unsupervised label refinement (URL) scheme is proposed by exploring machine learning techniques in order to enhance the labels purely obtained from the weakly labeled data without manual efforts. A clustering-based approximation (CBA) algorithm is also proposed for improving the efficiency and scalability. The main contributions of this paper will include the following:

- A promising search based face annotation scheme is investigated and implemented by mining large amount of weakly labelled facial images that are freely available on the Web.
- A novel ULR scheme is proposed for enhancing label quality via low-rank learning approach and a graph-based.
- An efficient clustering-based approximation algorithm is proposed for large-scale label refinement problem.
- An extensive set of experiments were conducted, in which encouraging results were obtained.

A short version of this work had appeared in SIGIR2011. By including a substantial amount of new content, this article has been significantly extended. This paper is organized as follows: SectionII will review the related work. SectionIII will give an overview of the proposed search-based face annotation framework. SectionIV will represent the applications of proposed unsupervised label refinement scheme. SectionV will conclude this paper.

II. SAMPLE WORK

This work is closely related to several groups of research work. The first group of related work is based on the topics of verification and face recognition, which are classical research problems that arise in computer vision and pattern recognition and have been extensively studied for several years. In recent years, some emerging benchmark studies of unconstrained face detection and verification techniques on facial images have been observed which are collected from the web, known as the LFW benchmark studies. Some recent studies had also attempted for extending classical face recognition techniques used for face annotation tasks. Comprehensive reviews present on face recognition and verification topics can be found in some survey papers. The second group is about the description of generic image annotation. The classical image annotation approaches will usually apply some existing object recognition techniques in order to train classification models that are obtained from human-labeled training images or attempt to infer the correlation or probabilities among the images and annotated keywords. For given limited training data, semi-supervised learning methods have also been used for image annotation. For example, Wang et al. had proposed to refine the model-based annotation will result with a label similarity graph which is followed by random walk principle.

Similarly, Pham et al. proposed for annotating unlabeled facial images in case of video frames along with an iterative label propagation scheme. Although semi-supervised learning approaches would leverage both labeled and unlabeled data, it will remain fairly time-consuming and expensive in order to collect enough well-labeled training data for achieving good performance in large-scale scenarios. Recently, the search-based image annotation paradigm has attracted more and more attention. For example, Russell et al. had built a large collection of web images along with ground truth labels for facilitating the object recognition research. However, most of these works is focused on the feature extraction, indexing, and search techniques. Unlike these existing works, a novel unsupervised label refinement scheme is proposed which is focused on optimizing the label quality of facial images towards the search-based face annotation task. The third group describes face annotation present on personal or family or social photos. Several studies have mainly focused on the annotation task present on personal photos, which will often contain rich contextual clues, known as personal or family names, geotags ,social context, timestamps and so on.

The number of persons or classes is usually quite small, making such annotation tasks less challenging. These techniques will usually achieve fairly accurate annotation results, in which some techniques have been successfully deployed in case of commercial applications, for example, Microsoft easy Album, Apple iPhoto, Google Picasa, and Face book face auto tagging solution. The fourth group will describe about the studies of face annotation in mining weakly labeled Present on the web. Some studies will consider a human name as the input query, and mainly it will aim to refine the text-based search results used for exploiting visual consistency of facial images. For example, Ozkan and Duyguluhad proposed a graph-based model for finding the densest sub-graph considered as the most related result. By following the graph-based approach, Le and Satoh had proposed a new local density score for representing the importance of each returned images. Guillaumin et al. had introduced a modification in order to incorporate the constraint that a face is only obtained once in an image. On the other side, the generative approach like the gaussian mixture model is also adapted to the name-based search scheme, and comparable results are achieved. Recently, a discriminate approach is proposed in order to improve over the generative approach and avoid the explicit computation present in graph-based approach.

The performance of name-based scheme can be further improved by using ideas from query expansion, by introducing the images of the “friends” of the query name. Unlike these studies regarding facial images by filtering the text-based retrieval results, some studies have attempted in order to directly annotate each facial image with the names that are extracted from its caption information. For example, Berg et al. had proposed a possibility model combined with a clustering algorithm for estimating the relationship among the facial images and the names in their captions. For the facial images and the detected names present in the same document (a web image along with its caption), Guillaumin et al. had proposed in order to iteratively update the assignment which is based on a minimum cost matching algorithm. In their follow-up work, they further improve the annotation performance by using distance metric learning techniques for achieving more discriminative feature in low-dimension space. The work performed in this is different from the above previous works mainly in two main aspects. Initially, this work will aim to solve the general content-based face

annotation problem by using the search-based paradigm, where the facial images are directly used as query images and here the task is to return the corresponding names of the query images. Based on this topic, very limited research progress has been reported.

Some recent work, had mainly addressed the face retrieval problem, in which an effective image representation has been proposed by using both local and global features. Based on initial weak labels, the proposed unsupervised label refinement algorithm will learn an enhanced new label matrix for all the facial images present in the whole name space; however, the caption-based annotation scheme will only consider the assignment between the facial images and the names that are appeared in their corresponding surrounding-text. With this result, the caption-based annotation scheme is only applicable to the scenario where both the images and their captions are available, and due to the lack of complete caption information they cannot be applied to our SBFA framework. The fifth group describes about the studies of purifying web facial images that aims to leverage noisy web facial images used for face recognition applications. In the whole system, these works are proposed as a simple preprocessing step without adopting sophisticated techniques. For example, a modified k means clustering approach is applied for cleaning up the noisy web facial images. Zhao et al. had proposed a consistency learning method in order to train face models for the celebrity by mining the text-image co-occurrence present on the web as a weak signal of relevance towards supervised face learning task obtained from a large and noisy training set.

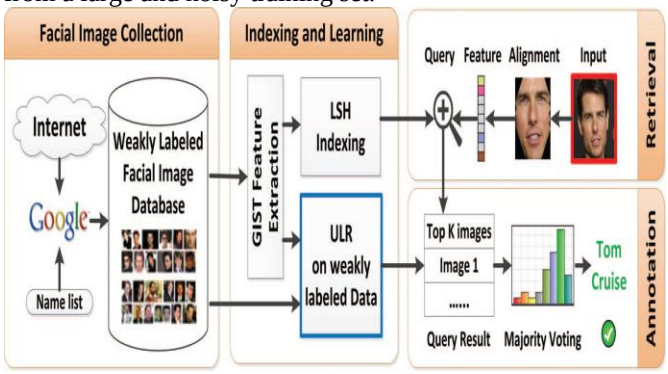

Fig.1.System Architecture.

Apart from the above existing works, the unsupervised machine learning techniques were employed and proposed a graph-based label refinement algorithm for optimizing the label quality over the whole retrieval database which is present in the SBFA task. Finally, it is noted that this work is also related to recent work of the WLRLCC method and latest work is present on the unified learning scheme. Instead of enhancing the label matrix across the entire facial image database, for each individual query, the WLRLCC algorithm is focused on learning more discriminative features for the top retrieved facial images which is very different from the ULR task in this paper. Last but not least, it is noted that the learning methodology for solving the unsupervised label refinement task are partially inspired by some existing studies in case of machine learning, which includes graph-based semi-supervised learning and multi label learning techniques.

III. PROPOSED MECHANISM

A. Unsupervised Label Refinement Obtained By Learning on Weakly Labeled Data

Preliminaries: It is denote by X IRnd the extracted facial image features, where n and d will represent the number of facial images and also the number of feature dimensions, respectively. Further it is denoted by n1; n2; . . . ; n_m the list of human names for annotation, where m denotes the total number of human names. It is also denoted by Y; 1nm the initial raw label matrix for describing the weak label information, in which the ith row Yi will represent the label vector of the ith facial image xi IRd. In this application, Y is often referred to as noisy and incomplete. In particular, for each weak label value Yij, Yij will indicate that the ith facial image xi has the label name as nj, while Yij will indicates that the relationship among ith facial. Image xi and jth name is described as unknown. It is noted that it usually have kYi k0 since each facial image present in the database was uniquely collected from a single query. By following the terminology of graph-based learning methodology, a sparse graph is built by computing the weight matrix W IRnn, where Wij will represent the similarity among xi and xj.

Problem Formulation: The final goal of the unsupervised label refinement problem is to learn a refined label matrix F2 IRnm, which is expected to be more accurate, compared to the initial raw label matrix Y. since nothing else is present but the raw label matrix Y and the data examples X themselves is described, this is a challenging task. In order to tackle this problem, a graph-based learning solution is proposed based on a key assumption of "label smoothness," i.e., the more similar that the visual contents of two facial images, the more likely they will share the same labels. The label smoothness principle will be formally formulated as an optimization problem for minimizing the following loss function, where k kF will denote the Frobenius norm, W denotes the weight matrix of a sparse graph which is constructed from the n facial images, L will denote the Laplacian matrix where D is denoted as a diagonal matrix with the diagonal elements such as Di Pn and tr will denote the trace function.

It is problematic by directly optimizing the above loss functionas it yields a trivial solution. In order to overcome this issue, it is noticed that the initial raw label matrix usually, though it is being noisy, it still contains some correct and useful label information. Thus, when it is optimized to search for F, it shall avoid the solution F being deviated too much from Y. To this end, it is formulated the following optimization task for the unsupervised label refinement which includes a regularization term for reflecting this concern. where it is a regularization parameter and F 0 enforces F is represented as nonnegative. Next, defining an appropriate function fo is studied. One possible choice of Þ is to simply set the Y k2 F. However, this is not appropriate as Y is often very sparse, i.e., due to the incomplete nature of Y many elements of Y are zeros. Thus, the above choice is problematic as it may simply force many elements of F to zeros without the consideration of the label smoothness.

A more appropriate choice of the regularization will be applied only to the nonzero elements of Y. To this end, the following are proposed, where S is described as a "sign" matrix where if x > 0 and 0 otherwise, and denotes the Hadamard product (i.e., the entry wise product) among two matrices. Finally, it is noticed that the solution of the optimization is generally dense, which is again not desired since the true label matrix is often described as sparse. In order to take the sparsity into consideration, a sparsity regularize is proposed by following the "exclusive lasso" technique where it introduces an '1 norm in order to combine the label weights for the same person corresponding to different names, and an '2 norm in order to combine the label weights ofdifferent persons together. By combining this regularizer and the previous formulation, the final formulation. Where are two regularization parameters. The above formulation will combine all the terms which are present in the objective function that refers to as "soft-regularization formulation" or "SRF".

IV. APPLICATIONS

Face annotation will find its application in the field of:

- It can achieve relatively high performance without the user interaction.
- When the user interaction is considered, it should be reduced to an acceptable level.
- In case of Face annotation at macro scale and micro scale.
- In case of Wild landmark face annotation.
- Used in online photo album management and also in video domain.

V. CONCLUSION

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

VI. REFERENCES

[1]Social Media Modeling and Computing, S.C.H. Hoi, J. Luo, S. Boll, D. Xu, and R. Jin, eds. Springer, 2011.

[2] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and Detecting 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 the News," 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 News Video Monologues," Proc. 12th Ann. ACM Int'l Conf. Multimedia (Multimedia), pp. 580-587. 2004.

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

[8]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.

[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]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. 1483-1490, 2006.

Author's Profile:

R.Sudeshna recived the B.TechDegree in Information Technologyfrom Jawaharlal Nehru Technological University, Anantapur (College-G. Pullaiah college of Engineering & Technology, Kurnool.) Pursing M.Tech CSE in RGMCET, Nandyala, Andhra Pradesh, India.

K.E.Naresh Kumar, Recived the B.Tech Degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Anantapur, India, in 2006 and Masters in Computer Science from Visweswaraya Technological University, Belgaum, India, in 2010. Working as Assistant professor in CSEDept at R.G.M.C.E.T, Nandyal, India.