A Survey on Search Based Web Facial Image Annotation Methods

Ms. Shalaka S. Dixit
Student , M.Tech. C.S.E. 3rd Sem
GHRIETW
Nagpur

Ms. Antara Bhattacharya
Asst. Professor
GHRIETW
Nagpur

Abstract— **With the advances in multimedia technologies collection of digital images is growing rapidly. Due to the popularity of various digital cameras and the rapid growth of social media tools, internet-based photo sharing have increased in daily life. A large portion of photos shared by users on the Internet are human facial images. Some of these facial images are tagged with names, but many of them are not tagged properly. Content Based Image retrieval is very important area of research in the field of image retrieval. The search-based face annotation (SBFA) paradigm aims to tackle the automated face annotation task by exploiting content-based image retrieval (CBIR) techniques. Many researchers develop and use lots of approaches towards image annotation. Automatic image annotation is the process of automatically assigning semantic labels to images. This paper presents the survey of different approaches for automatic annotation and image based annotation retrieval. This paper aims to cover the latent space and generative approaches for automatic image annotation.**

Keywords— Automatic Image Annotation, Content Based Image Retrieval, search-based face annotation

I. INTRODUCTION

Now a days due to increase in digital media like camera, mobile phones collection of digital images is growing rapidly. Many of photos the shared by users on the internet are human facial images. Some of these facial images are tagged with names, but many of them are not tagged with proper names. So, this gives the motivation in the field of annotation of images or tagging of facial images with proper names. The image annotation technique can be performed either manually or automatically. Manual image annotation can be performed by libraries using indexing and then later retrieving their image collection. But manual image annotation is quite expensive, time consuming and labour intensive procedure. On the other side, automatic image annotation approach to annotate and retrieve images based on a training set of images. In this technique images can be described using a small vocabulary blob. Blobs in images can be generated from image features using clustering. With the help of training set of images with proper annotations probabilistic models allow to predict the probability of generating a word given the blobs in an image. This may be used to automatically annotate and retrieve images given a word as a query or even an image as a query .The face annotation can be important in many real world application. For example, Labelling photos on online social network, labelling facial images in video frames using transcript. There are main two approaches in the research work: Classical face annotation and Search based face annotation. Classical face annotation can be helpful in the creation of different classification models, which are trained from a collection of well labelled faces. Classification model can be can be created by employing the supervised and semi-supervised machine learning technique. Classical face annotation is also known as "model- based face annotation" as different models can be created by this approach. But for this approach a large number of training faces will be needed and which should be well labelled also. So, due to this reason Model-based approach is basically limited in the aspect that, it is time-consuming and expensive to collect such kind of images. Also, it is difficult to modify the existing model or to create new model when a new training data or facial image is added. Also classical face annotation or model based face annotation gives poor scalability where numbers of classes are very large. Second approach is Search based annotation which can basically overcome limitation occurred in classical annotation approach. This approach uses the concept of content based image retrieval (CBIR) technique. Content-based image retrieval is a technique which uses visual contents to search images from large scale image databases according to users' interests. Basically it uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. The visual contents in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. A framework called search based face annotation framework (SBFA) is derived with the help of search based annotation approach. SBFA paradigm aims to tackle the automated face annotation task by exploiting content-based image retrieval (CIBR) techniques. So, SBFA framework is data driven and model –free. SBFA can be supervised or unsupervised. In this research work an unsupervised technique is used. With the help of unsupervised technique in SBFA, a novel unsupervised label refinement (ULR) is developed. ULR is used for refining the labels which are associated with particular image. Labels are often noisy and do not necessarily give correct names for the image. So, ULR is such a method which can purify the

noisy labels and helps to find out accurate labels related with the image. This unsupervised label refinement scheme used in search based face annotation framework is focused on optimizing the label quality of facial images towards the search-based face annotation task.

This paper presents a survey of the research related to the search based web facial image annotation. The rest of the paper is organized as follows: Section II reviews research work related to classical face annotation and Content based image retrieval. (CBIR).Section III reviews framework for Search based face annotation (SBFA). Table I provide comparative analysis of methods discussed for face annotation. Section IV concludes the paper.

II. RELATED WORK

A. Classical Face Annotation

Classical face annotation approach is a very old and basic approach for face annotation problems. This technique is applied where the database has collection of well labeled images. Then classification models will be generated by employing supervised or semi-supervised machine learning technique on those images. When a new image is or training data of a person is added into databases then again a new classification models has to be generated by using the new added image. Classification in the classical approach can be particularly "pattern classification", where each pixel is in an image considered as a co-ordinate in high dimensional space. This particular patter classification work can be found in P. Belhumeur [2]. This paper uses the advantage of observation that the images of a particular face, under varying illumination but fixed pose, lies in a 3D linear subspace of the high dimensional image space . The problem area in this paper can be simply stated as given a facial images labelled with person's identity which will be treated as learning set, and an unlabeled set of facial images from the same group of people forms the test set, indentify each person in the test set. So, this In this paper two classical techniques were used: Eigen faces and Fisher faces. In Eigen faces method is based on linearly projecting the image space into a low dimensional feature space. Eigen faces method uses principal component analysis for dimensionality reduction which provides the projection which maximizes total scatter across all classes. Fisher faces is a classical technique in pattern recognition. It has been applied in different fields of computer vision depending on which features are being used.

As the learning set is labelled, it makes sense to use this information to build a more reliable method for reducing the dimensionality of the feature space. By using class specific linear methods for dimensionality reduction and simple classifiers in the reduced feature space, it is easy to get better recognition rates than with either the Linear Subspace method or the Eigenface method. Fisher's Linear Discriminant (FLD) [3] is an example of a *class specific method*, in the sense that it tries to "shape" the scatter in order to make it more reliable for classification. But still there is a drawback, when the data base contains large number of images. The classical face annotation does not scale well when image database is large. Also, when a new image is added to the dataset, intensive retraining of

process is usually required. So, its tedious and time consuming process to make classes when new data is added to database by classical face annotation approach. So, many of the drawbacks in the classical approach for face annotation can be effectively overcome by using content based image retrieval technique.

B. Content Based Image retrieval (CBIR)

CBIR is a technique where content such as , shape, texture, and spatial layout to represent image. The visual contents in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. Some images comes with definite labels and some of them are unlabelled. Some research work in [4] proposed by Wang et al., gives refine the model-based annotation results with a label similarity graph by following random walk principle. Also by using some semi supervised techniques, research work in [5] proposed by Pham et al., gives the annotation of unlabelled facial images in video frame with an interactive 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. So, recently the search base face annotation paradigm has gain the attention in the research work of face annotation, [6], [7], [8]. For example, in research work [9], proposed by Russell et al. created a large collection of web images with genuine labels to facilitate object recognition research. However, most of these works were focused on the indexing, search, and feature extraction techniques that is there was no great involvement of labels annotated to those images. So, in above research work contents of an image was the core area for the image annotation. But using only content of images and use of indexing, searching and feature extraction technique on image for annotation was not sufficient. To again improve the performance in annotation of images, contextual information which comes along with images can also be used. So, for gathering all contextual images related with an images, personal, social or family photos can be used. Several studies [10], [11], [12], [13] have mainly focused on the annotation task on personal photos, which often contain rich contextual clues or information, such as personal/family names, social context, geotags, timestamps and so on. This was an effective technique but with one limitation that number of classes generated was too small which makes annotation task less effective. To overcome this, and increase scalability in sense of number of classes.

Although semi-supervised learning approaches could leverage both labeled and unlabeled data, it remains fairly time-consuming and expensive to collect enough well-labelled training data to achieve good performance in large-scale scenarios. So, recently the search base face annotation paradigm has gain the attention in the research work of face annotation, [6], [7], [8]. For example, in research work [9],

proposed by Russell et al. created a large collection of web images with genuine labels to facilitate object recognition research. However, most of these works were focused on the indexing, search, and feature extraction techniques that is there was no great involvement of labels annotated to those images. So, in above research work contents of an image was the core area for the image annotation. But using only content of images and use of indexing, searching and feature extraction technique on image for annotation was not sufficient. To again improve the performance in annotation of images, contextual information which comes along with images can also be used. So, for gathering all contextual images related with an images, personal, social or family photos can be used. Several studies [10], [11], [12], [13] have mainly focused on the annotation task on personal photos, which often contain rich contextual clues or information, such as personal/family names, social context, geotags, timestamps and so on. This was an effective technique but with one limitation that number of classes generated was too small which makes annotation task less effective. To overcome this, and increase scalability in sense of number of classes.

So, another new approach in face annotation is using weakly labelled facial images by mining on web. It consider a human name as the text input query, and aims to refine the text-based search results by exploiting visual consistency of facial images. A graph-based model is generated in for finding the densest sub-graph as the most related result of input query is proposed in the research by Ozkan and Duygulu [14]. On the other hand, the generative approach like the gaussian mixture model was also been adopted for the name-based search scheme [15], [16] and achieved good comparable results. Also, in [17] a discriminant approach was proposed to improve over the generative approach and avoid the explicit computation in graph-based approach. By using ideas from query expansion [18]. This name based scheme can further be improved by introducing 'friends' of query name. These studies particularly focused on filtering the text based retrieval result. Also, a research work is proposed [19] which have used partial clustering and interactive labeling for face annotation. An unsupervised stage is used for partial clustering which is used to find out the most evident clusters instead of grouping all instances into clusters, which gives a good initial labelling for later user interaction purpose. In the Interactive stage an efficient labelling procedure is proposed which is based on minimization of both global system uncertainty and estimated number of user operations. Some, studied have attempted to improve annotation result, by directly annotating each facial image with names extracted from caption information of the image. a possibility model is developed in combination with a clustering algorithm to estimate the relationship between the facial images and the names in their captions which is proposed in research work [20] proposed by Berg et al. This proposed work is basically used in an framework called Search based face annotation (SBFA), where images with weakly labeled names are collected and then annotation task is performed with the help of SBFA.

III. SEARCH BASED FACE ANNOTATION(SBFA)

The recent approach in research work of face annotation is SBFA which exploits the characteristics of CBIR technique. In research work [21] proposed by Dayong Wang SBFA framework is developed. This framework comprises of several steps. The first step is to collect all the images with labeled or weakly labeled facial images. These facial images are extracted by mining world wide web. These images are then stores in database. In the second step an efficient algorithm is applied for face alignment and facial feature extraction. In the next step indexing of facial feature is done with help of LSH technique. After the indexing is completed, the next step is to learn and refine weakly labelled data of all images. In the above research work a Unsupervised label refinement (ULR) technique is used to perform this work. This step is the most important and innovative technique which is responsible for efficient annotation task. Next step is to extract similar faces. This is performed with the help of comparing and matching indexed features of images and their refined labels. Last step will to annotate extracted facial images with proper label. Here, in the research work [21] face annotation task is performed by majority voting on the similar faces with the refined label. So, in[21] annotation is performed effectively than other previously discussed. It has improved its efficiency and scalability. But the proposed framework used in [21] is only used for famous personality or well known faces. The method used with this framework in [21] is completely unsupervised so accuracy in annotation is hard to maintain.

IV. CONCLUSION

A wide variety of researches have been made on image annotation on multimedia databases. Each work has its own technique, contribution and limitations. In Classical approach techniques, it becomes challenging when number of persons increases in database. It becomes very time consuming and expensive when a new person data is added in the database. In CBIR, it removes drawbacks of classical approach but still number of persons/classes formed with this techniques is very small. Also, if CBIR method is combined with contextual information of images it improves its performance that previous but again scalability issue remains. So, this limitation is removed in recent approach which is SBFA framework, which effectively improved its performance in annotation and scalability than previous. But this approach is only applied to famous personality photos so this is one of the limitation. In this paper, we attempted to provide a comprehensive survey on search based web facial image annotation. As a survey paper, we might not include each and every aspect of individual works; however we have focused on the wide variety of approaches used for face annotation research work.

TABLE I. COMPARATIVE ANALYSIS OF IMAGE ANNOTATION METHOS

Methods	Comparative Analysis		
	Problem Addressing	*Advantages*	*Limitations*
Pattern classification method : Eigenfaces versus Fisherfaces	Face recognition for images with varying luminance	Significantly reduce the face annotation workload and good accuracy	Number of persons/classes usually quite small.
Graph based model (Text based searching)	A face naming method that learns from labelled and unlabeled examples using iterative label propagation in a graph of connected faces or name-face pairs	The method does not require training for any specific person and thus it can be applied to any number of people	Face recognition and retrieval is a difficult and an error-prone problem due to large variations in pose, illumination and expressions
Partial clustering and interactive labelling	Creation of clusters of facial images and annotation to those clusters.	Partial clustering creates only evident clusters and discards irrelevant clusters	All clusters have to be labelled by user initially
Search based face annotation (SBFA)	Face annotation by mining facial images of web, and refinement of weak labels of images	Applicable for large data sets, Enhancement in the label quality , Image is directly used as input, Efficient face recognition technique is used	Applied to only queried images i.e. famous personalities

REFERENCES

1. Dayong Wang, Steven C.H. Hoi, Member, IEEE, Ying He, and Jianke Zhu, "Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation", IEEE Transaction On Knowledge and Data Engineering, Vol. 26, No. 1, January 2014.
2. 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.
3. H.V. Nguyen and L. Bai, "Cosine Similarity Metric Learning for Face Verification," Proc. 10th Asian Conf. Computer Vision (ACCV '10), 2008.
4. C. Wang, F. Jing, L. Zhang, and H.-J. Zhang, "Image Annotation Refinement Using Random Walk with Restarts," 14th Ann. ACM Int'l Conf. Multimedia, pp. 647-650, 2006.
5. P. Pham, M.-F. Moens, and T. Tuytelaars, "Naming Persons in News Video with Label Propagation," Proc. VCIDS, pp. 1528-1533, 2010 .
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. 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.
8. X. Rui, M. Li, Z. Li, W.-Y. Ma, and N. Yu, "Bipartite Graph Reinforcement Model for Web Image Annotation," Proc. 15th ACM Int'l Conf. Multimedia, pp. 585-594, 2007.
9. B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," Int'l J. Computer Vision, vol. 77, nos. 1-3, pp. 157-173, 2008.
10. 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.
11. 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.
12. 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.
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. 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.
15. 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
16. 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.
17. 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.
18. 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.
19. T.L. Berg, A.C. Berg, J. Edwards, and D. Forsyth, "Who's in the Picture," Proc. Neural Information Processing Systems Conf. (NIPS), 2005.
20. Y. Tian, W. Liu, R. Xiao, F. Wen, and X. Tang, "A Face Annotation Framework with Partial Clustering and Interactive Labelling", Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2007.
21. A.W.M. Smeulders, M. Worring, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1349–1380.