

Enhanced Web Facial Image Annotation with Image Based Data Retrieval Technique

Ms. Shalaka S. Dixit
RTMNU: Computer Science & engineering
GHRIETW
Nagpur, India
shalaka.dixit09@gmail.com

Ms. Antara Bhattacharya
RTMNU: Computer Science & engineering
GHRIETW
Nagpur, India
antara.bhattacharya@raisoni.net

Abstract—With the advancement in multimedia technology there is continuous increase in sharing of photos videos, and sharing is done by posting online photos over the internet. We see such digital high quality photos over social networking sites. The photos shared on internet in daily life contain 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. In CBIR image is retrieved based on the content provided. The search-based face annotation (SBFA) paradigm aims to tackle the automated face annotation task by using content-based image retrieval (CBIR) techniques. Many researchers have been done for image annotation. Automatic image annotation is the process of automatically annotating the labels related to that image. This paper presents the idea of automated image annotation technique.

Keywords— *Automatic Image Annotation, Content Based Image Retrieval (CBIR), Search-based face annotation (SBFA), Image content retrieval*

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 colour, 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 novel approach for the face annotation and how efficiently additional data related to an image can be retrieved. 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) and reviews framework for Search based face annotation (SBFA). Section III presents the proposed system which is the combination of methods of SBFA and Partial clustering-Interactive Labeling method. 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 patten 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 labeled 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 are 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 Discriminate (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 labelled 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 are 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 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 byOzkan 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 discriminate 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 labelling 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 labeling 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 labelled names are collected and then annotation task is performed with the help of SBFA.

C. 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 labelled or weakly labelled 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.

III. PROPOSED SYSTEM

Proposed system is automated face annotation. Annotation can not be only name of the facial image but it can also be related information of that image. In the proposed system , first standard database `imm_face_db` is used which contains images for face annotation. It also contains binary format of those images. First step to achieve face annotation is identification of facial image. Face identification can be

achieved by applying edge detection technique. The best edge detection technique is canny edge detector. In proposed system, Derivative edge Detector algorithm, is used, in which Canny edge detection technique is used. Next step is to apply initial labeling to images. Database can contain more than one image of same person in different poses. So, first task is to form clusters of similar faces by partial clustering and then apply labels to each cluster by interactive labeling stage. Algorithm for edge detection used in proposed system increases the accuracy in detection of facial image because it calculates inner as well as outer edges in the image. Also in second step partial clustering stage improve the scalability even if there is very large number of images present in the database. Interactive labeling helps to provide needed labels and information to image cluster, so it avoids naming images individually which is time consuming. Following section will expand over both steps in proposed system.

3.1 Derivative Edge Detector Algorithm

This algorithm is used for finding outer as well as inner edges of the facial image. In this algorithm Canny edge detection technique is used. An edge can be defined as points in a digital image at which the image brightness changes sharply or has discontinuities. Technically, it is a discrete differentiation operation, computing an approximation of the gradient of the image intensity function. There are many methods for edge detection, but most of them can be grouped into two categories: search-based and zero-crossing based. The search-based methods detect edges by first computing a measure of edge strength, usually a first-order derivative expression such as the gradient magnitude, and then searching for local directional maxima of the gradient magnitude using a computed estimate of the local orientation of the edge, usually the gradient direction. Here, we are using such a method implemented by Sobel known as Sobel operator. The operator calculates the *gradient* of the image intensity at each point, giving the direction of the largest possible increase from light to dark and the rate of change in that direction. We are using these gradients or derivatives in X direction and Y direction for matching.. We first create a data set or template model from the edges of the template image that will be used for finding the pose of that object in the search image. Here we are using a variation of Canny’s edge detection method to find the edges. You can read more on Canny’s edge detection. For edge extraction, Canny uses the following steps:

Step 1: Find the intensity gradient of the image

Use the Sobel filter on the template image which returns the gradients in the X (G_x) and Y (G_y) direction. From this gradient, we will compute the edge magnitude and direction. Once the edge direction is found, the next step is to relate the edge direction that can be traced in the image. There are four possible directions describing the surrounding pixels:

0 degrees, 45 degrees, 90 degrees, and 135 degrees. We assign all the directions to any of these angles.

Step 2: Apply non-maximum suppression

After finding the edge direction, we will do a non-maximum suppression algorithm. Non-maximum suppression traces the left and right pixel in the edge direction and suppresses the current pixel magnitude if it is less than the left and right pixel magnitudes. This will result in a thin image.

Step 3: Hysteresis threshold

Doing the threshold with hysteresis requires two thresholds: high and low. We apply a high threshold to mark out those edges we can be fairly sure are genuine. Starting from these, using the directional information derived earlier, other edges can be traced through the image. While tracing an edge, we apply the lower threshold, allowing us to trace faint sections of edges as long as we find a starting point.

Step 4: Save the data set

After extracting the edges, we save the X and Y derivatives of the selected edges along with the coordinate information as the template model. These coordinates will be rearranged to reflect the start point as the center of gravity.

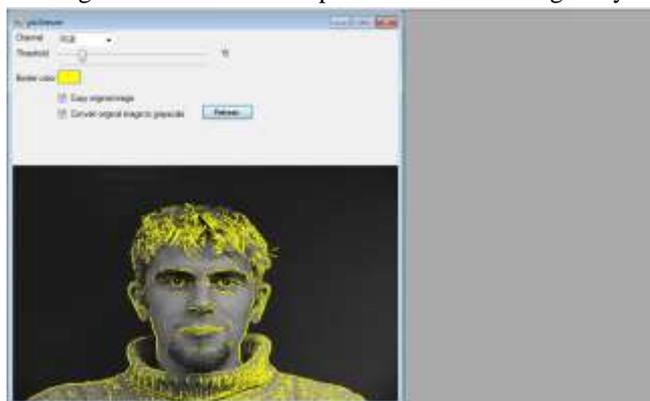


Fig. 3.1 Edge Detection with Derivative Edge Detector algorithm

3.2 Facial feature indexing with label

In this process the matrix of values of calculated edges of images are assigned with label as well as additional information. That is not only name of image is assigned as a label, but a structured label containing Information such as first name, last name, email id, age etc are also assigned to the image. But as the database contains multiple images of the same person. The images or matrices of the similar images can be clustered into one, with the help of values got from edge detection and a cluster can be assigned with such structured label. This can be achieved by partial clustering and Interactive labelling.

3.2.1 Partial Clustering

With the help of partial clustering, only “evident” clusters are kept, while other faces, that cannot be grouped tightly enough, remain in the litter bin. These evident clusters usually contain only one entity that is different facial image of same person, so a user can do group labelling with only one click. It is different from traditional clustering algorithms, in the partial clustering algorithm, it will not group all samples into clusters. The basic assumption in this algorithm is that the noisy samples which are difficult for clustering will be distributed uniformly after the spectral embedding.

3.2.2 Interactive Labelling

The partial clustering algorithm will automatically group similar faces into several evident clusters, and groups unmatched faces into a background cluster, called the litter-bin. After the partial clustering stage, we use an “Initial labelling” procedure to annotate these evident clusters. Since faces in an evident cluster most likely belong to a single individual, user annotation interactions on these clusters can be significantly reduced. However, the workload of face annotation in the litter-bin is still huge. In this section, we propose a parameter-free, iterative labelling procedure to address this problem. In each step, the system uses the information from the labelled faces to automatically infer an optimal subset of unlabeled faces for user annotation. This annotation step will be iteratively used until all faces are labelled. Using this strategy, the overall user interactions can be reduced by finding an optimal subset of unlabeled faces in each annotation step.

Here labels to be provide to the image cluster is not only name of that person in the image but also we can give extra contextual information related with that image.

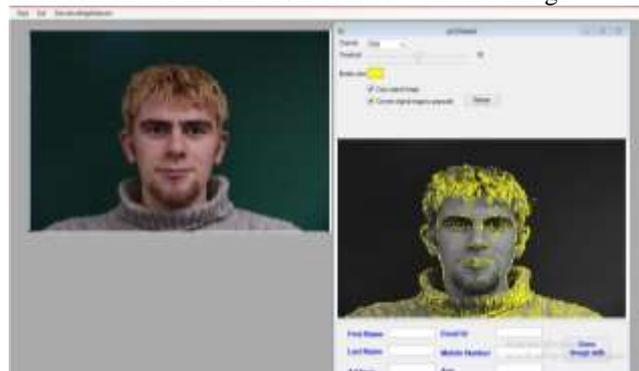


Fig. 3.2 Facial Feature indexing with label

IV. CONCLUSION

A wide variety of researches have been made on image annotation on multimedia databases. Every work has its own technique, some contribution and have limitations also. As in the Classical approach techniques, it becomes challenging when number of persons increases in database. It is very time consuming and expensive process when a new person data is added in the database. In CBIR, it can remove the drawbacks of classical approach but it is not efficient as

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 than previous technique but again scalability issue persists. So, this limitation is again 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. But in SBFA, Face recognition is not performed with a perfect face recognition algorithm. Solving Face annotation can be difficult for real life photographs which can be in various poses, with different expressions. So in such problem it is most important to use a good face recognition or detection algorithm in the situation of face annotation. As face detection is the most important step in the face annotation. So In this paper the research work is to use an efficient algorithm for face detection in problem of face annotation. In this research work algorithm used is “Derivative Edge Detector” which uses Sobel operator and Canny Edge detection technique. After Detection of face it important to provide initial labels to faces. But for improving efficiency and scalability here partial clustering and interactive labeling is performed. Which can effectively reduce the work of manually labeling all the images initially in the database.

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] 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. 1 477-1482, 2006.
- [3] H.V. Nguyen and L. Bai, “Cosine Similarity Metric Learning for Face Verification,” Proc. 10th Asian Conf. Computer Vision (ACCV '10), 2008.
- [4] 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.
- [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.