Face Annotation on Weakly Labeled Images using Nearest Neighbor Calculation

Ann Veena Paul
Department of Computer Science & Engineering
M.A College of Engineering, Kothamangalam, Kerala, India

Merin K Kurian
Department of Computer Science & Engineering
M.A College of Engineering, Kothamangalam, Kerala, India

Eldo P Ealias
Department of Computer Science & Engineering
M.A College of Engineering, Kothamangalam, Kerala, India

Abstract

Social networks such as face book, twitter etc. are widely used in our day to day life. Face image retrieval using content based method is an emerging technology in many real world applications for automatic face annotation. Our main goal is to retrieve the similar images from large scale database using content based. One challenging problem for search-based face annotation scheme is how to effectively perform annotation by exploiting the list of most similar facial images and their weak labels that are often noisy and incomplete. To tackle this problem, we propose an effective label refinement approach based on the semantic closeness of the image with name-list with which we had crawled the web search engines. Locality-Sensitive Hashing (LSH) to index the facial features in our solution and a Clustering of images on name-level is proposed to improve the efficiency of the annotation. We conduct extensive empirical studies on several web facial image databases to evaluate the proposed classification algorithm from different aspects.

Keywords: Locality Sensitive Hashing, Label Refinement, Face Annotation, Clustering Based Approximation, Performance

I. INTRODUCTION

With the rapid growth of web photo sharing portals and social networks, massive amounts of images and photos have been uploaded and shared on the internet. Most of the images uploaded to internet are facial images. Efficient face annotation scheme can recognize the faces and annotate them properly. The images uploaded to internet are actually a treasure to future generations. But there is a problem that the efficient methods to retrieve all the images properly are not available. The main reason behind this is that most of the images are not properly tagged.

Mining web facial images on the internet has emerged as a promising paradigm towards auto face annotation. Content-based image retrieval systems require users to query images by their low-level visual content. This not only makes it hard for users to formulate queries, but also can lead to unsatisfied retrieval results. Because of this, Image annotation is introduced. The aim of image annotation is to automatically assign keywords to images, so image retrieval users are able to query images by Keywords and automatically detect human faces from a photo image and further name the faces with the corresponding human names.

Face annotation can applied for online photo sharing applications and video domains. The goal of an automated image tagging task is assigning with some pre-trained image models. One of the challenges is the need for tools that automatically analyze the visual content with semantically meaningful annotations. There are two problems in classic models of facial annotation. First, it is time consuming and expensive to create a training set which contains the images and exact labels. Second, it is difficult to generalize the models when new images are added. Different techniques are proposed to overcome these issues and thereby making facial annotation in an efficient way up to an extent.

II. RELATED WORKS

Different techniques are used in retrieving facial images based on search query. Most of the users use person's name as the search query. So it is effective to label the images with their exact names. The automatic face recognition techniques can annotate the faces with exact labels and it also help to improve the search more efficiently.

A. Classic Model of Face Annotation

This model of face recognition can be done either by comparing the features of two input images or by comparing an input image with the training data set. This is a straight forward method that reduces the workload of user when searching for same person's image. The face recognition is a challenging part of all time since the actual face recognition efficiency is affected by many

factors such as illumination, lighting, camera quality, pose of photo taken etc. So most of the face recognition algorithms can perform well in controlled conditions [1].

B. Retrieval based Face Annotation

Dayong Wang, Steven C.H. Hoi, Ying He, Jianke Zhu [3] proposed the retrieval based face annotation. The paper introduces an effective Weak Label Regularized Local Coordinate Coding (WLRLCC) technique, which exploits the local coordinate coding principle in learning sparse features. It employs graph-based weak label regularization principle to enhance the weak labels of short listed similar facial images. This method overcomes two major challenges that are being faced in labelling problem: how efficiently retrieve short list of similar images and how to annotate them. This is an optimization algorithm, which boosts the performance of retrieval based face annotation. They also develop an effective sparse reconstruction scheme to perform the final face name annotation.

C. Content Based Image Retrieval Content-based image retrieval (CBIR)

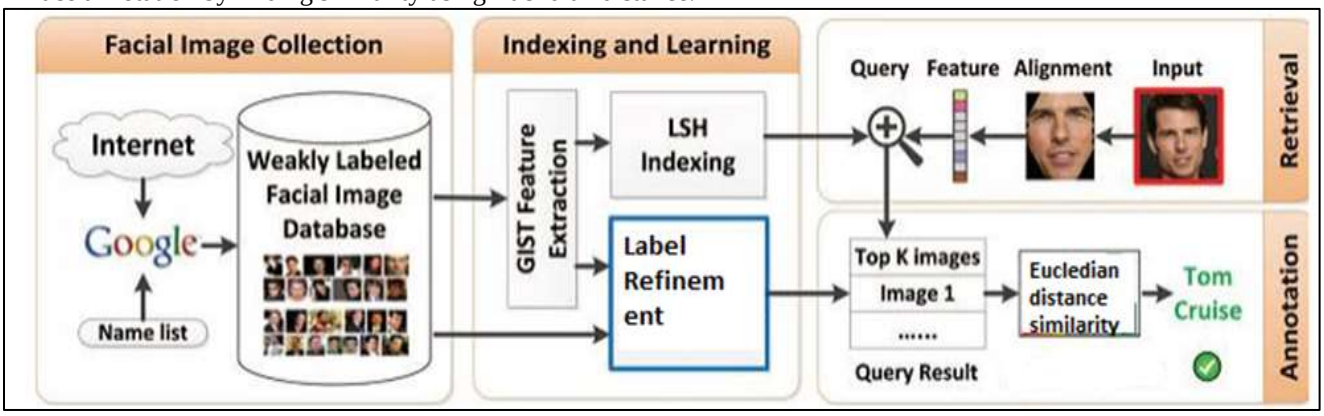
It is opposed to traditional concept-based approaches. Content Based Image Retrieval is an efficient technique for improving the performance of image retrieval. Various methods are used for this purpose and Support Vector Machine (SVM) is very important one in this field. This provides a supervised learning technique which analyses data and learning patterns. This has high importance in collecting relevance feedback. This approach has many drawbacks, sometimes the SVM offer small number of label examples. Another problem is that, this method does not consider the redundancy of results and therefore system selects multiple examples in relevance feedback, that may be similar (or even identical) to each other [5].

III. PROPOSED SYSTEM

The main feature of this work is that SBFA is data driven and it is model free, so it can provide large scalability than the other existing techniques can provide. As per this method, whenever an image is uploaded, K similar images are retrieved and the annotation is performed by finding the Euclidian similarity.

It consists of the following steps:

1) Facial images data collection,
2) Label refinement
3) Face detection and facial feature extraction,
4) Locality sensitive hashing
5) Clustering based approximation
6) Face annotation by finding similarity using Euclidian distance.



(a) (a) (a)

Fig. 1: The Framework of the proposed system.(a) The database construction by crawling facial images from the World Wide Web; (b) The database indexing for fast facial feature retrieval in high-dimensional space; (c) The content-based facial image retrieval for a query facial image and the automatic face annotation by mining the retrievable facial images and the corresponding labels.

The first four steps are 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 thus should be done very efficiently. The first step is the collection of facial image data as shown in Figure 1(a), in which we crawled facial images from the WWW by web search engines (e.g., Google) based on a name list that stores the names of persons to be collected. This crawling process produces a collection of facial images, each of them is associated with some human name. Given the nature of web images, these facial images are often noisy, which do not always correspond to the right human name. Thus, we call such kind of web facial images with noisy names as weakly labelled facial image data. So we have to perform a label refinement. The third step is to pre-process web facial images to extract face-related information, including face region detection and alignment, face region extraction, and facial feature representation. For facial region detection and alignment, we adopt the unsupervised face alignment technique in [14]. For facial feature

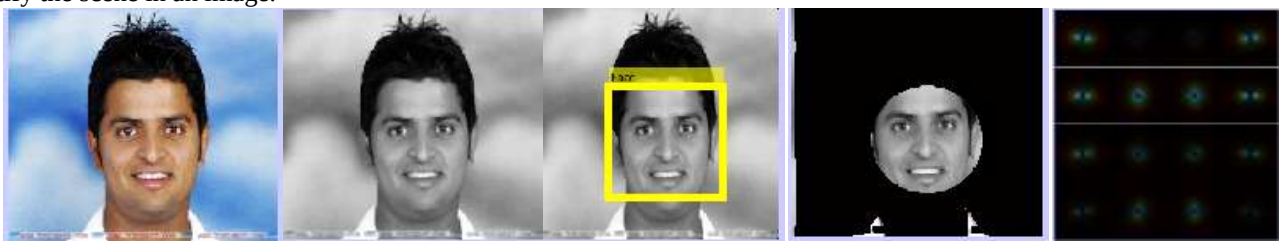
representation, we extract the GIST features [13] to represent the extracted faces. As a result, each face can be represented as a d-dimensional vector. The fourth step of the framework is to index the extracted features of the faces by applying some efficient high-dimensional indexing technique to facilitate the task of similar face retrieval in the subsequent step. In our approach, we adopt the Locality-Sensitive Hashing (LSH) [10], a popular and effective high-dimensional indexing technique for approximate nearest neighbour search. Next we describe the process of face annotation during the test phase. In particular, given a query facial image for an- notation, we first conduct a similar face retrieval process to search for a subset of most similar faces (typically top k similar face examples) from the previously indexed facial database. With the set of top k similar face examples retrieved from the database, the next step is to annotate the facial image with a label (or a subset of labels) by employing a Euclidean based distance approach that combines the set of labels associated with these top k similar face examples.

D. Pre-Processing

The input image and the database images are apply to the pre-processing method. The noises in the frames reduces the quality of the frames. Each frames are considered as images. In order to improve the quality of the images we normally employ some filtering operations. Median filter is used for filtering. The median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings. Instead of simply replacing the pixel value with the median of neighboring pixel values. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value.

E. Gist Feature

In computer vision, GIST descriptors are a representation of a low-dimensional image that contains enough information to identify the scene in an image.



(a) (b) (c) (d) (e)

Fig. 2: (a) the original input image (b) image is converted to grayscale (c) face recognition (d) face alignment (e) Gist feature representation of input image

The GIST descriptors allow a very small size representation of an image. They can represent the dominant spatial structure of the scene from a set of perceptual dimensions. The authors have tried to capture the gist descriptor of the image by analyzing the spatial frequency and orientation. The global descriptor is constructed by combining the amplitudes obtained at the output of the K Gabor filters at different scales E and orientations O. To reduce the feature vector size, each filtered output image is scaled and divided into N * N blocks (N between 2 and 16), which gives a vector of dimension N * N * K * E * O. This dimension can be further reduced by a principal component analysis (PCA). After the pre-processing step of the input character image, the next step consists on changing the character image into different scales and orientations. Finally, the features vectors are calculated for each scale, orientation and frequency. Those features vectors are combined to form a global feature descriptor which is reduced by a principal component analysis (PCA).

F. Label Refinement

In this section, we refine the labels of web facial image data by comparing weakly labelled data with name list we used to crawl the images.

We denote by $X \in Rn\times d$ the extracted facial image features, where n and d represent the number of facial images and the number of feature dimensions, respectively. Further we denote by = {n1, n2. . . nm} the list of human names for annotation, where m is the total number of human names. We also denote by $Y \in [0, 1]$ n×m the initial raw label matrix to describe the weak label information, in which the i-th row $Yi*$ represents the label vector of the i-th facial image $xi \in Rd$. In our application, Y is often noisy and incomplete. In particular, for each weak label value Yij, Yij = 0 indicates that the i-th facial image xi has the label name nj , while Yij = 0 indicates that the relationship between i-th facial image xi and j-th name is unknown. Note that we usually have $\|Yi*\|0 = 1$ since each facial image in our database was uniquely collected by a single query.

We refine this weak label by comparing the weak label name yj with nj and the similarity between yj and nj is computed and if similar weak label is replaced with the corresponding name list label.

G. *Locality Sensitive Hashing*

After a linear projection and then assignment of points to a bucket via quantization, points that are nearby are more likely to fall in the same bucket than points that are farther away. We use the following notation: D is the input data set; its points are denoted by p with various subscripts; q is the query point. LSH has three parameters: the quantization width w, the number of projections k, and the number of repetitions L. A Gaussian random vector v is a vector with every coordinate independently taken from the normal distribution N(0;1). We pick a random shift value b is taken uniformly from the interval [0;w].

Using LSH consists of two steps:

1) *Step 1: Indexing*

- Randomization: -Select k random vectors v with the same dimensionality as the data, where each coordinate is a Gaussian random variable N (0, 1), and a scalar bias term b from a uniform random distribution between 0 and w.
- One-line Projection: - Take a point p€D, compute its dot product with the Gaussian random vector v, and quantize the result with step w=p.v+b/w. The bias term b does not affect our performance, but simplifies the analysis to follow because it ensures that the quantization noise is uncorrelated with the original data.
- Multiline projection: Obtain an array of k integers by doing k one-line projections. All points that project to the same k values are members of the same (k-dimensional) bin. At this stage, a conventional hash is used to reduce the k-dimensional bin identification vector to a location in memory. With a suitable design, this hash produces few collisions and does not affect our analysis.
- Repeat by hashing the data set to k-dimensional bins into a total of L times. Thus, we place every point in the dataset into L tables.

2) *Step 2: Search*

- Compute the L (k-dimensional) bins for the query point using the same random vectors and shift values as in the indexing stage.
- Retrieve all points that belong to all identified bins (we call them candidates), measure their distance to the query point, and return the one that is closest to query point.

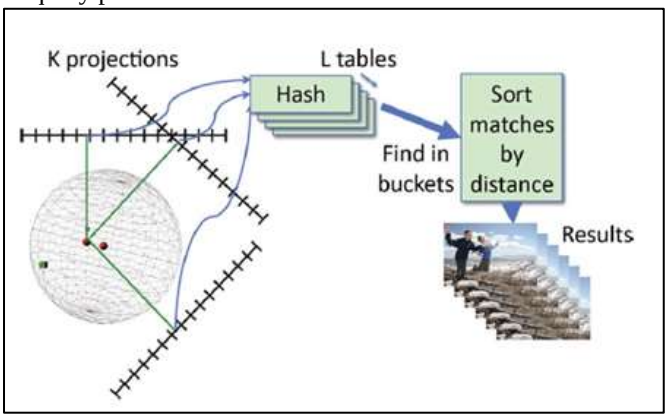


Fig. 3: Indexing with LSH

H. *Name Level Clustering*

Clustering-Based Approximation: - The number of variables in the previous problem is n * m, where n is the number of facial images in the retrieval database and m is the number of distinct names (classes). In particular, the clustering strategy could be applied as "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

I. *Euclidean Distance*

The Euclidean distance or Euclidean metric is the "ordinary" distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm. Older literature refers to the metric as Pythagorean metric. The Euclidean distance between point's p and q is the length of the line segment connecting them.

In Cartesian coordinates, if p = (p1, p2,..., pn) and q = (q1, q2,..., qn) are two points in Euclidean n-space, then the distance (d) from p to q, or from q to p is given by the Pythagorean formula:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

The position of a point in a Euclidean n-space is a Euclidean vector. So, p and q are Euclidean vectors, starting from the origin of the space, and their tips indicate two points.

J. Annotation and Similar Image Retrieval

Given a query facial image, we employ a similar face retrieval process to find the most similar face from the indexed face databases using the LSH technique. After obtaining the most similar faces for the query image, retrieve the label associated with this closest image. Now search for the cluster which has the label, we obtained before and show the similar images.

K. Evaluation of Performance of Proposed Algorithm

To evaluate the annotation performances, we used Precision and Recall as the performance metric. Performance is measured in terms of Precision and Recall. Precision P is defined as the ratio of the number of retrieved relevant images r to the total number of retrieved images n, i.e., P =r/n [1]. Precision measures the accuracy of the retrieval.

Recall is defined by R and is defined as the ratio of the number of retrieved relevant images r to the total number m of relevant images in the whole database, i.e., R=r/m [1]. Recall measures the robustness of the retrieval.

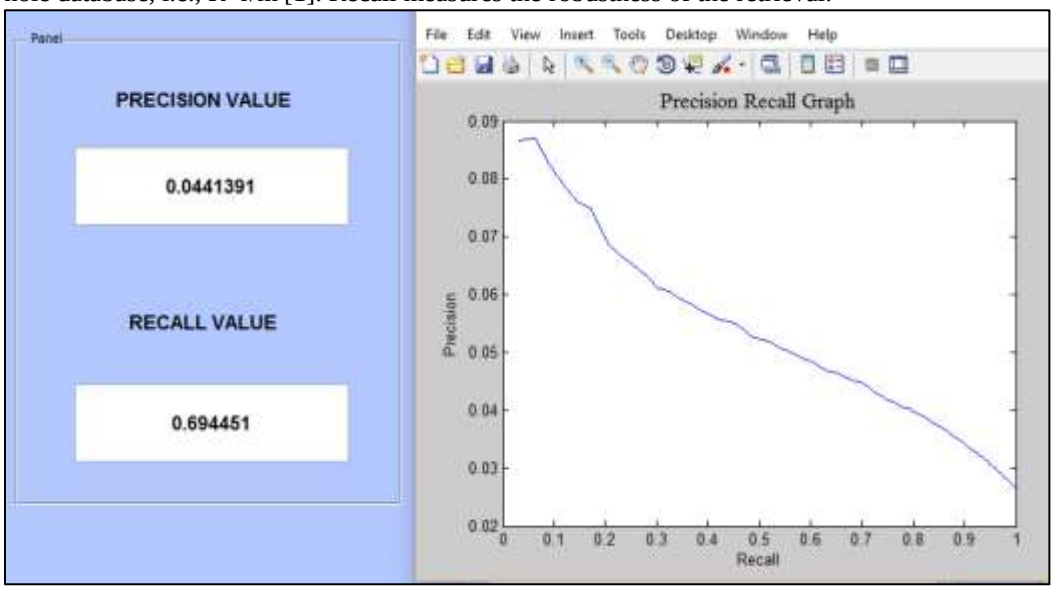


Fig. 4: Performance evaluation of proposed algorithm

IV. CONCLUSION AND FUTURE WORK

This paper investigated a promising search-based face annotation framework, in which we focused on tackling the critical problem of enhancing the label quality. To further improve the scalability, we also proposed a Clustering-based Approximation (CBA) solution, which successfully accelerated the optimization task without introducing much performance degradation. From an extensive set of experiments, we found that the proposed technique achieved promising results under a variety of settings. Our experimental results also indicated that the proposed distance based approach improved the scalability. As future work a more robust label refinement approach based on semantic similarity can be explored to further enhance the accuracy of face recognition problem.

REFERENCES

[1] 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 Conf. Computer Vision and Pattern Recognition (CVPR), pp. 848-854, 2004.
[2] J. Zhu, S.C. Hoi, and M.R. Lyu, "Face Annotation by Transductive Kernel Fisher Discriminant," IEEE Trans. Multimedia, vol. 10, no. 1, pp. 86-96, Jan. 2008.
[3] D. Wang, S. Hoi, Y. He, and J. Zhu, "Mining Weakly-Labeled Web Facial Images for Search-Based Face Annotation," IEEE Trans. Knowledge and Data Eng., vol. 99, no. PrePrints, pp. 1-14, 2012.
[4] A. Holub, P. Moreels, and P. Perona, "Unsupervised Clustering for Google Searches of Celebrity Images," Proc. Eighth IEEE Int'l Conf. Automatic Face & Gesture Recognition (FG '08), pp. 1-8, 2008.
[5] S.C. 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, no. 3, pp. 1-29, July 2009.
[6] 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.
[7] S.C. Hoi, W. Liu, and S.-F. Chang, "Semi-Supervised Distance Metric Learning for Collaborative Image Retrieval and Clustering," ACM Trans. Multimedia Computing, Comm., and Applications, vol. 6, no. 3, pp. 18:1-18:26, Aug. 2010.
[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, Feb. 2011.

[9] F. Wu, Y. Han, Q. Tian, and Y. Zhuang, "Multi-Label Boosting for Image Annotation by Structural Grouping Sparsity," Proc. ACM Int'l Conf. Multimedia, pp. 15-24, 2010.
[10] 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.
[11] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, Dec. 2003.
[12] Handbook of Face Recognition, second ed., S.Z. Li and A.K. Jain, eds. Springer, 2011.
[13] Matthijs Douze,Hervé Jégou "Evaluation of GIST descriptors for web-scale image search"
[14] J. Zhu, S. C. Hoi, and L. V. Gool, "Unsupervised face alignment by robust nonrigid mapping," in ICCV, 2009, pp. 1265–1272