An enhanced approach for Mining of Facial Images and Annotation based on Ensemble learning method

Tarang Boharupi, Pranjali Joshi

Abstract— **In this paper, we present an improved method for the mining of facial images and annotation using an ensemble learning method. Machine learning techniques such as k-means clustering, support vector machine and neural network are widely used in the research field of Bioinformatics, data mining, capturing knowledge in expert systems and so on. The extracted features can be deducted from the images and classified using ensemble learning method. The proposed system addresses the problem of annotating names to the facial images. The supervised ensemble learning method Random forest classifier is used for the classification of facial images and to perform annotation based on majority voting. In this proposed work, the face is detected using the global threshold technique and features of faces are extracted using SIFT-PCA algorithm. The main motivation behind this hybrid approach is to provide simple, effective and less complex model for mining of facial images and annotation for input query.**

Index Terms— **Annotation, Face recognition, Image mining, Ensemble learning method, Web facial images**.

I. INTRODUCTION

On social networking sites, a large number of human facial images are shared. Some of these images are tagged correctly, but most of them are wrongly labeled. Such images we called as weak labeled images. So there is a strong need to annotate the name correctly to these facial images so we get correctly labeled images for input query. Face recognition and annotation can be beneficial to many real world applications for example face recognition used for identifying the particular person securing the cybercrime. Besides, Face annotation can be applied to detect the important person in the video domain. The report addresses the problem of mining of weakly labeled images and annotation. We explore the issue such as how to effectively retrieve the facial images from a huge database of images for a particular given query.

In this proposed work, we collect the facial images from the database and Global threshold technique is used to detect faces in an image by converting RGB into YcBcR. We calculate the time required to train the images for input query. SIFT and PCA technique will be used to process the image by extracting the features and to reduce the dimensionality of data. The Proposed methodology is used for mining of weakly labeled images from database based on preprocessing facial images. The Proposed system uses an ensemble learning method for classifying the images which will further matched for mining of images. So, this combined method increases the face identification rate. On the basis of majority voting the name is annotated to the facial images correctly.

Manuscript received Dec, 2015.

Tarang Boharupi *received the B.E. in Computer science and engineering, in 2013 from BMCOE, Nagpur affiliated to RTMNU University. Perusing M.E. degree in Computer Engineering from Pune Institute of Computer Technology, Pune affiliated to Savitribai Phule Pune University.*

II. LITERATURE SURVEY

Mining of weak labeled web facial images over the internet has been studied with different methods. All these methods are performed on it such that human names are used as input query and expect to refine results based on text by attaining exist face images.

D. Wang et.al [1] proposes a model of Search-based face annotation which plays an important role in face annotation. The scheme firstly retrieves a short rundown of top-K most comparative facial pictures from a vast scale web facial image database, and then gives the name to the query facial image by mining the tags related to the top-K similar facial images. They have applied bisecting k-means clustering and divisive based clustering algorithm for enhancing the performance on large scale facial images. To take care of the issue of uncertainty between a few names and one face, a modified *k*-means clustering method was utilized in which faces are assigned to the closest group (each cluster corresponding to one name) after a number of iterations. Nearest neighbor classifier is one of the unsupervised machines learning algorithm, which cluster the objects having similar properties in one cluster and objects having different properties in different clusters. The similarity between two objects is measured using metrics the Euclidean distance and Manhattan distance. Search based face annotation takes less running time, but final annotation performance degrades.

Challenges:-

a) There is a test in proficiently recovering the top-K most comparable facial images from a huge facial image database for giving an input query facial image.

ISSN: 2278 – 7798

b) There is a test in viably exploit the short list of Candidate facial pictures and their weak labels for naming the human Faces automatically.

The SBFA consists of following steps.
Step 1: Facial image information collection;
Step 2: Face detection and facial feature extraction;
Step 3: High-dimensional facial feature indexing;
Step 4: Learning to refine weakly labeled data;
Step 5: Similar face recovery.
Step 6: Face annotation by greater part voting on the similar appearances with the refined labels.

S. Satoh et.al [2] developed a Name-It, a system that knowledge source, the system is given news and videos which include image grouping and transcripts got from sound tracks or closed caption texts. The framework can then either surmise the name of a given face or yield the name candidates, or can find the countenances in news videos by a name. To fulfill this task, the framework extracts face from image sequences and names from transcripts, each of which could correspond to important persons in news topics. The projected framework takes full benefit of advanced image and natural language process. The image processing contributes to the mining of face sequences which provide rich information for face name coordination. The processing helps to pick the best frontal view of a human face in a face sequence to boost the face identification that is required for the processing. It is more time consuming and more costly to gather a large amount of human labeled training facial images. Here, using the video caption the face detected and extract names from transcripts. It achieves good face association, but annotation performance is poor.

D. Wang, S.C.H. Hoi, Y. He. And J. Zhu et.al [3] proposed WLRLCC algorithm which is used as a part of mining gigantic web facial images for automatic human face annotation. It has two difficulties first is the mean by which adequately retrieve most of comparative facial images while second test is how to effectively perform the annotation approach to images. Weak label regularized local coordinate coding (WLRLCC) is optimization algorithm. This algorithm boosts the performance of face annotation on huge scale web facial pictures. This algorithm uses with a neural network, which boosts the performance of face annotation on huge scale web facial pictures. It gives better performance as well as improves retrieval accuracy.

Y. He et.al [4] introduced a framework by consolidating both transductive and inductive learning methods to mine web facial images for face annotation. They have proposed Weak label Laplacian support vector machine (WL-LapSVM) algorithm by receiving WLRLCC algorithm that helps the execution of annotation. Support vector machine is a supervised learning method that analyzes data and recognizes patterns for classification. Given a set of training, each marked as belonging to one of class, SVM build model from training data and assigns the new example into one class or another. SVM performs both linear and non-linear classification. In non-linear classification, SVM uses kernel trick. Support Vector Machine work very well in a practice, giving a good result for classification, but gives the problem of over fitting and has a high semantic gap. The user must choose the kernel function and its parameters, but the rest is automatic. Kernel trick helps to boost the performance.

L. Zhang et.al [5] presented a learning system to automate the face annotation in family photo collection. Firstly, methodologies of image retrieval based on contents and face recognition area are consistently coordinated to accomplish automated annotation. Secondly, face annotation is developed in a Bayesian structure, in which the face likeness measure is characterized because the greatest a posteriori (MAP) estimation. Thirdly, to manage with the missing feature options, marginal probability is employed in order that samples which have missing values are compared with those having the complete feature set to validate a non-biased decision. The experimental evaluation has been conducted among a family Photoshop's of few thousands of photographs and also the results show that the proposed methodology is successfully and advanced in automated face annotation in family album management. They have done effective and efficient face recognition on the basis of content based image retrieval. For final annotation Naïve-Bayes classifier has used which is simple using names and faces.

Mensink et.al [6] were intrigued to discover images of peoples on the web sites and more plainly named the new facial images. Starting result focused around the text, but they are not perfect because execution is focused around pre-assumption. To enhance over poor performance, query expansion strategy is utilized. They have connected this thought right on early proposed method on which filter the initial result set. They have given a straight forward preprocessing step without utilizing sophisticated techniques. Mensink et.ai [6] proposed modified k means clustering methodology for cleaning up the noisy websites facial images. Their aims to correct noisy web facial images for face recognition. They have got good results, but take more computation time.

Tamara L. Berg et al. [7] shows great face association is possible for a dataset of mistakenly and vaguely labeled face images. Their dataset is 44,773 face images, got by applying a face discoverer to approximately half a million subtitled news images. This dataset is more practical than usual face recognition datasets, as a result of it contains faces captured "in the wild" in a variety of configurations with related to the photo camera, taking a variety of features, and under the light of widely varied color. Every face image is connected with a group of names, automatically terminated from the associated caption. However, not all such sets contain the correct name. They bunch face images in appropriate discriminant coordinates. They utilize a clustering procedure to break ambiguities in naming and identify incorrectly labeled faces. A merging procedure then identifies variation of names that refer to the same individual. The subsequent representation can be used to mark faces in news images or to arrange news pictures of individual persont. They report

ISSN: 2278 – 7798

results for (a) the original dataset (b) the datasets resulting from our clustering, merging and cleaning process, without using cluster information (c) the datasets resulting from our clustering, merging and cleaning process, including their cluster structure.

Tamara L. Berg et al. [8] Developed the graph based strategy in which a graph is formed by considering faces as nodes, and the weights of edges connected between nodes are the comparability of faces, is very closely related to their problem. They developed the method to discover the similarities subset with a possible similar set of faces with query person names where SIFT describers speak to the likeness in the picture. They have found that content based image retrieval performance is improved for face recognition. In text based approach first image were annotated with content and further search using this approach from the user database management system. This methodology is used to managing images. Through text and content description, images can be organized by semantic or topical hierarchies which provide facility to easily explore and browse base on standard Boolean queries.

Peter N. Belhumeur et al. [9] built up face recognition algorithm techniques which is harsh to extensive variety in lighting direction and facial expression. Taking an example of pattern classification approach, they have considered every pixel in an image as a direction in a high-dimensional space. They have exploited an advantage of the perception that the picture of a selected face, under varying illumination, however fixed expression; lie in a 3Dimention linear subspace of the high dimensional image space if the face is a Lambertian surface expect shadowing. On the other hand, since faces are not really Lambertian surfaces and do so turn out self-shadowing; pictures will veer off from this linear subspace. Instead of explicitly modeling this deviation, we sequentially project the picture in a way which subspace in a manner which discounts those areas of the human face with expansive deviation. Our projection system is predicated on Fisher's Linear Discriminant and creates well distinguished classes in a low-dimensional subspace, even in the extreme variety of lighting and facial expressions.

Jun Yang et al. [10] approaches problem of Naming Every Individual in News Video Monologues with a statistical learning method. Naming each distinct individual showing up in broadcast videos with names detected from the news video caption leads to better access of the news video content. Two classes of data separated from multiple video modalities have been investigated, namely features, which helps to recognize the right name of each and every individual, as well as constraints, which reveal the connection between the names of diverse persons. The individual naming issue is planned into a learning framework which calculates the most probably name for each individual based on the expression, and refines the predictions using the constraints. Experiments performed on ABC World New Tonight and CNN Headline News videos exhibit that this methodology outperforms a non-learning option by a large amount. They have described a machine learning model for naming every individual person in broadcast news videos. Trained based on some manually named people, this model predicts the most likely name of each individual who is giving a monologue speech based on features extracted from multiple modalities including speech, overlaid text, transcript, etc. Equivalence constraints relating the names of different persons are used to refine the predictions. Experiments have shown that this approach significantly outperforms a non-learning alternative in terms of the accuracy of the names assigned, particularly when the constraints are applied.

Timo Ahonen et.al. [11] Studied face recognition with local binary patterns**.** In this work, they introduced a new approach for face identification, which considers both shape and texture data to represent the facial pictures. Rather than the EBGM approach, a clear extraction of the face feature vector (histogram) is received in their calculation. The face picture is initially divided into little regions from that the Local Binary Pattern (LBP) components are separated and concatenated into a one feature histogram efficiently representing the face picture. The surfaces of the facial areas are generally encoded by the LBP patterns whereas the complete dimensions of the face are recovered by the development of the face feature histogram. The idea behind utilizing the LBP options is that the face images can be seen as composed of micro-patterns which are invariant with regard to monotonic gray scale transformations. In their methodology, a face picture is initially divided into many pieces (facial regions) from that we extract local binary patterns and build a global feature histogram that speaks both the statistics of the facial micro-patterns and their spatial abstraction. Then, face identification is performed using a nearest neighbor (k-means) classifier in the computed feature space with $\chi 2$ as a difference measure. The proposed face representation can be easily mined in a single scan through the human image, without doing any complex analysis as in the EBGM calculation. Face pictures can be seen as a composition of micro-patterns which can be well described by LBP. They exploited this observation and proposed a simple and efficient presentation for face recognition. In their approach, a face image is firstly divided into several regions (facial regions) from which authors extract local binary patterns and construct a global component histogram that represents both the measurements of the facial micro-patterns and their spatial areas.

III. PROPOSED SYSTEM

Face recognition is a system by detecting the face in an image, with the effect of estimating and normalizing for translation, scale and in-plane rotation. Given a normalized image, the features, either global or local, are extracted and consolidated in a compact face representation which can then be put in a database and compared with face representations derived at later times. Before face annotation, we have to go through face recognition.

ISSN: 2278 – 7798

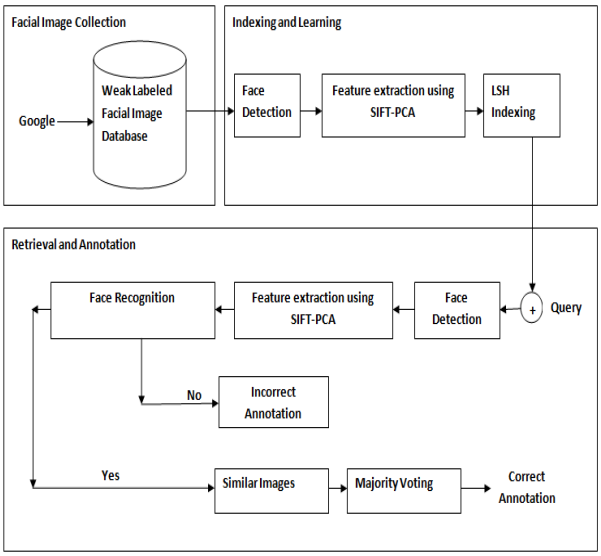

Figure 1: Proposed System

A. Collection of Facial Images:

The first step is the data collection of facial images as shown in Fig. 2, in which we collects a facial images from the WWW by an existing web search engine (i.e., Google) according to a name list that contains the names of persons to be collected. As the output of this crawling process, we shall obtain a collection of facial images; each of them is associated with some human names. Given the nature of facial 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 labeled facial image data.

Figure 2: Collection of facial images

B. Face Detection and Feature Extraction:

We have an input image of particular persons; on that system will detect face first by converting the RGB into YcBcR as shown in fig. 3. After face detection we have to apply SIFT-PCA algorithm for feature extraction. The Scale invariant feature transform assign and build the key points by calculating the difference of Gaussian extreme. PCA calculates eigenvalues and eigenvectors from the set of key points. Dimension reduction occurs by ignoring the directions in which the covariance is small. SIFT is a scale invariant feature transform which is used for Invariant to scale change, Invariant to rotation change, Invariant to illumination change, Robust to the addition of noise, Robust to a substantial range of fine transformation, Robust to 3D viewpoint, Highly distinctive for discrimination. Principle component analysis (PCA) is a multivariate technique that analyzes a face data in which observation are described by several inter-correlated dependent variables. With the strong robustness of SIFT-PCA, a good performance of feature extraction is performed. Features can be extracted by the following algorithm.

Algorithm 1 SIFT algorithm

Input : Set of facial images
Output: Extracted features

1. While not at correct annotation **do**
2. Read input images
3. Perform Face detection and feature extraction using SIFT-PCA
4. If Face of person detects
5. Get all trained images
6. Use recognize method
7. Calculate Eigen distance between input and trained image
8. Return image with minimum Eigen distance.
10. Return recognized face
11. **Else**
12. Go to next input image
13.End.

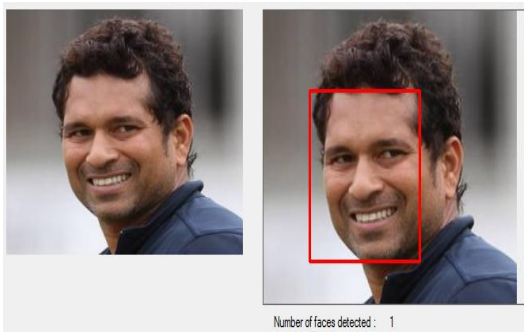

Figure 3: Face detection

C. Face recognition:

The features extracted during training of images now classified using ensemble learning method. These features of trained images now matched with features of input image for face recognition. The following fig. 4, the extracted key-point using SIFT-PCA is matched for face recognition. Random forest classifier is used for face recognition.

Algorithm 2 Random Forest classifier

Input : I ϵ D, Input images
Output: Face recognition.

ISSN: 2278 – 7798

1. While not at face recognition **do**
2. Read input query images
3. Perform Face detection and feature extraction using SIFT-PCA.
4. If Face of person Extracted
5. Calculate most similar object .
6. **for** each of sample
7. Get Eigen distances.
8. Choose best split among all predictors at each node.
9. **Return** recognized face
10. **Else**
11. Go to next input query image
12.**End**

Figure 4: Key point Matching

D. Mining of Images:

"Random forest classifier" is applied to classify the images based on Sift features. In a randomized tree, the split at each node happens by using only a randomly selected subset of all features. By using "Random Forest classifier", the face can be recognized and mining of images can be done. The following fig. 5, shows the mining of facial images.

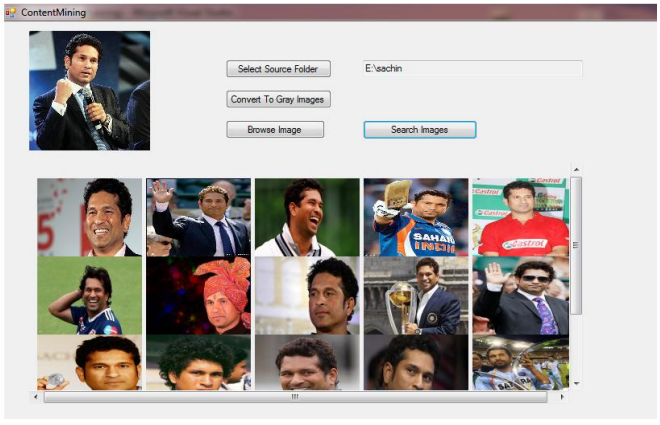

Figure 5: mining of facial images

E. Annotation:

Every Image will vote for a label. After voting on all images, the one with a maximum vote will be considered as a winner and the image will be annotated with the label of winner as shown in fig.6

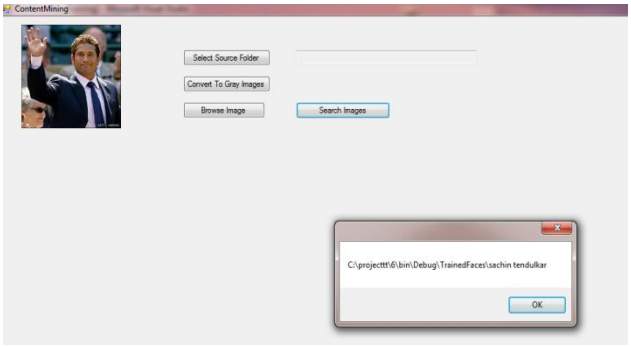

Figure 6: Annotation of images

The implemented application uses the collection of facial images for mining and annotation of images. The algorithm looks very simple and less complex in structure and thus the question arises whether these algorithms are suitable for a large database of images. The question can be answered only by pure experimentation and testing the model to predict the same.

IV. RESULTS ANALYSIS

For the comparison of supervised algorithm's accuracy, precision, recall and error rate these parameters are considered.

Precision (P):

Precision is the percentage of correctly mined search images among all the search queries that were assigned to the category by the classifier.

$$P = \frac{\text{Mining of images by the ensemble learning algorithm}}{\text{All Facial images retrieved by the Enssemble learning algorithm}}$$

Precision for this P = 17/18

=94.44%

Recall (R):

It is the percentage of correctly classified search queries among all the queries belonging to that category.

$$R = \frac{\text{Mining of images by the ensemble learning algorithm}}{\text{All Facial images retrieved by the human}}$$

Recall for this R = 17/20

= 85%

ISSN: 2278 – 7798

3. Accuracy (A):

Accuracy is calculated by the following formula

$$\mathbf{A} = \frac{P*R}{\left(\frac{1}{2}\right)*(P+R)}$$

Accuracy for this,

A = (94.44*85) / (0.5*(94.44+85))

= 89.471 %

Methods	Total Number of faces	Relevant faces	Retrieved faces	Precision (%)	Recall (%)	Accuracy (%)
UEL	20	17	19	89.47	85	87.17
SVM	20	16	19	84.21	80	82.05
Proposed system	20	17	18	94.44	85	89.47

Fig 7: Data table for Precision, Recall and Accuracy

Ensemble learning method is trained on a same training data set. They have tested on four different data sets and results are calculated as follows. For comparison of the two algorithms, precision, recall and accuracy these parameters are considered. The recognition rate is dependent on the number of trained images. More the trained images more will be the accuracy.

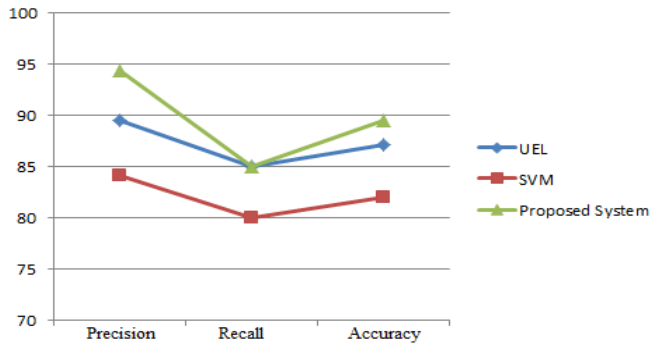

Fig 8: Bar chart

V. DISCUSSION AND FUTURE WORK

The current system aims to classify the images using ensemble learning method. SIFT and PCA technique will be used to process the image by extracting the features and to reduce the dimensionality of data. The proposed ensemble learning method gives better results for face recognition and annotation. Future work includes increasing the size of the database and the range of the pose and illumination conditions which the system can handle. We can use the combination of different classifier with feature extraction techniques for mining of images and annotation.

VI. CONCLIUSION

In this paper, the survey on mining weakly labeled database approaches is done. Proposed system provides a handy way to input the image into the system which supports Mining and annotation. SIFT is a scale invariant feature transform which is used for Invariant to scale change, Invariant to rotation change, Invariant to illumination change ,Robust to addition of noise, Robust to substantial range of affine transformation, Robust to 3D view point, Highly distinctive for discrimination. The accuracy of the system is a better than other existing works. Ensemble learning method algorithm provides accuracy 89.47%. The proposed technique dealing with designing fast, secure and reliable face recognition system. SIFT and PCA technique will be used to process the image by extracting the features and to reduce the dimensionality of data. Finally those extracted SIFT features are used to train the "random forest classifier" for image classification and recognition. So, this combined method increases the face identification rate.

REFERENCES

[1]Dayong Wang, Ying He,and Jianke Zhu, " Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation" IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 2, January 2014.

[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]D. Wang, S.C.H. Hoi, Y. He1, and J. Zhu, "Retrieval-Based Face Annotation by Weak Label Regularized Local Coordinate Coding," Proc. 18th ACM Int'l Conference. Multimedia (Multimedia), pp. 354-362, year 2014.

[4] D. Wang, Y. He, and S.C.H. Hoi, "A Unified Learning Framework for Auto Face Annotation by Mining Web Facial Images," Proc.21st ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 1382-1401, 2012.

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

[6] T. Mensink and J.J. Verbeek, "Improving People Search Using Query Expansions," Proc. 10th European Conference. Computer Vision (ECCV), vol. 2, pp. 86-99,year 2008.

[7] T.L. Berg, A.C. Berg, M. Maire, J. Edwards, R. White, E.G. Learned-Miller, Y.W. Teh, and D.A. Forsyth, "Names and Faces in the News," Proc. IEEE computer science Conf. Computer Vision and Pattern Recognition (CVPR), pp. 848-854, year 2004.

[8] D. Ozkan and P. Duygulu, "A Graph Based Approach for Naming Faces in News Photos," Proc. IEEE computer science Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1477-1482, year 2006.

[9] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection,"

ISSN: 2278 – 7798

IEEE Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July 1997.

[10] 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.

[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Recognition with Local Binary Patterns," Proc. European Conf. Computer Vision (ECCV), vol. 1, pp. 469-481, 2004.

[12] Stone, T., Darrell, Z., Zickler, T: Autotagging facebook: Social network context improves photo annotation. In: CVPR Workshop, year 2008.

[13]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.

[14]Patrik Kamencay, Martin Breznan, Dominik Jelsovka, and Martina Zachariasova, "Improved Face Recognition Method based on segmentation Algorithm using SIFT-PCA", 978-1-4673-1118-2/12/$31.00 ©2012 IEEE.

[15]Tong Liu, Sung-Hoon Kim, Hyon-Soo Lee, Hyung-Ho Kim, " Face Recognition base on a New Design of Classifier with SIFT keypoints",978-1-4244-4738-1/09/$25.00 ©2009 IEEE.

[16]Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", IEEE Conf. on Computer Vision and Pattern Recognition, pp. 506-513, 2004.

[17] L. C. Zhang, J. W. Chen, Y. Lu and P. Wang, "Face Recogntion Using Scale Invariant Feature Transform and Support Vector Machine", the 9th International Conference for Young Computer Scientists , pp.1766-1770, 2008.

[18] Gautam Narang, Soumya Singh Arjun Narang, "Robust Face Recognition Method Based on SIFT Features Using Levenberg-Marquardt Backpropagation Neural Networks", 978-1-4799-2764-7/13/$31.00 ©2013 IEEE.

[19] J. Zhu, S.C.H. Hoi, and L.V. Gool, "Unsupervised Face Alignment by Robust Nonrigid Mapping," Proc. 12th Int'l Conf. Computer Vision (ICCV), 2009.

[20] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli, "On the use of SIFT features for face authentication", Proc. Of IEEE Int Workshop on Biometrics, in association with CVPR, pp. 35-41, NY, 2006.

[21] X.-J. Wang, L. Zhang, F. Jing, and W.-Y.Ma.Annosearch: Image auto-annotation by search. In CVPR'06, pages 1483–1490, 2006.

[22]Verbeek, J., Triggs, B.: Region classification with Markov field aspectmodels. In: CVPR (2007).

[23] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," Pattern Recognit., vol. 33, no. 11, pp. 17711782, 2000.

[24] L. Torres, L. Lorente, and J. Vilà, "Face recognition using self-eigenfaces," in International Symposium on Image/Video Communications Over Fixed and Mobile Networks. Rabat, Morocco, 2000, pp.44-47.

Tarang Boharupi received the B.E. in Computer science and engineering, in 2013 from BMCOE, Nagpur affiliated to RTMNU University. Perusing M.E. degree in Computer Engineering from Pune Institute of Computer Technology, Pune affiliated to Savitribai Phule Pune University.

ISSN: 2278 – 7798