ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Unsupervised Celebrity Face Naming with HOG Scheme

Anugraha Raj.S, Sreenimol K.R

Post Graduate Student, Dept. of CS, Mangalam College of Engineering, Ettumanoor, Kottayam, Kerala, India

Associate Professor, Dept. of CS, Mangalam College of Engineering, Ettumanoor, Kottayam, Kerala, India

ABSTRACT: Nowadays character identification from popular web videos is very challenging task due to huge variation in the approach of each and every person or celebrity in the web videos .In this paper investigating the problem of missing tag or label detection in unconstrained videos with user-created Metadata. Instead of relying on supervised learning, a better relationship made from image domain and value content. Those relationships mainly include spatial-temporal context and visual similarities. And the knowledge base includes weakly tagged images along with set of names and celebrity social networks. Merging of suitable relationship with knowledge base is done through conditional random field. The proposed system gives three kinds of relationship sets, Face to Face, Name to Name and Face to Name. The new approach introduce here, which can encounter the closest relationship with right feature or faces in web videos, thereby reduce missing tag problem with celebrity face identification to an ideal extent.

KEYWORDS: Celebrity face naming, social network, unconstrained web videos, unsupervised learning, Graph cutting, Histogram of Oriented Gradient (HOG), Speed Up Oriented Feature (SURF).

I. INTRODUCTION

Global video sharing sites like YouTube, Netflix have great important in today's modern lifestyle. Among them YouTube got more popularity. Most of the web videos are uploaded by individuals, in that 80% are people related. In those 70% percentage are celebrity related videos. But unfortunately, majority of celeb-videos suffered with face identification problem. Technically called missing tag or labeling problem which means that the user description along with every uploaded video is insufficient because of several incomplete data. It is not unusual that a pointed celebrity does not appear in the video, and vice versa. One reason behind of this is description is appearing in a video is not mentioned. This will result unsatisfactory video sharing experience.

Fig.1. Example of Web video illustrating the challenge of associating the names (italic) in metadata with the detected faces (with bounding boxes) in the video

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Ideal solution is to find a alternative mechanism for right face naming according to metadata information. Merging face and name or any other features within a relationship and followed by this establishment of corresponding metadata can resolve this missing tag problem to good extent. User video experience can be also improved by reducing noisy problems.

Fig. 1 illustrates the problem with a real example of Web video. Out of the fourteen faces (of four celebrities) detected in the video, only four of them have names mentioned in the metadata. Furthermore, among the three celebrities who are mentioned, only two of them appear in the video. In other words, there are missing faces and names in the video and text respectively. Additionally, a common characteristic of Web videos, as shown in Fig. 1, is that faces appear wildly different as a result of motion blur, lighting and resolution changes. In brief, the challenge of name-face association can be attributed to incomplete text labels, noisy text and visual cues.

Here leveraging on rich relationships rather than rich texts [1]–[4] based Web video domain, a method based on histogram oriented gradient (hog) with conditional random field (CRF) [9], [10] is proposed to address the problem of face naming. Typically 3 kinds of relationships are formed in the work. Namely;

i) Face to Face (F2F)
ii) Face to Name (F2N)
iii) Name to Name (N2N)

First two relationships (F2F, F2N) exploits particular relationship with in a single video called within-video. The function is to assign the names mentioned in the metadata with exact face detected in video and it is notated as “null assignment” or “uncertenity”.N2N extend the naming system in a single video to ‘between video’ concept, by performing its task on group of celebrity videos. One benefits of later one is allow the rectification of names incorrectly tagged and the filling in of missing names not found in metadata.

The main contribution of this paper is find a better alternative for face tagging problem in domain unrestricted web videos for celebrity face naming.

II. RELATED WORK

Now currently available research efforts on face labeling mostly concentrate on domain web images [16] – [18] and constrained videos [3]- [9], such as TV serial, news bulletins and movies .All those existing works can be categorized broadly in to three classes: model-based, search-based and constrained clustering-based face labelling.

Name-It [2] one early existing face-name associating system, processes information from the videos and can infer possible name for a given face or locate a face in news videos by name. To accomplish this task, the system takes a multimodal video analysis approach i.e. face sequence extraction and similarity evaluation from videos, name extraction from transcripts, and video-caption recognition. Name-It system can associate faces in news videos with their right names without using a priori face-name association set. In other words, Name-It system extracts face-name correspondences only from news videos. Two categories of information extracted from multiple video modalities have been explored, namely features, which helps to distinguish the true name of every person, as well as constraints, which reveal the relationships among the names of different persons. Multiple instance of learning [20] also another method for face labeling

A very modern form of face recognition scheme is introduced in DeepFace [21]. The network architecture is based on the assumption that once the alignment is completed, the location of each facial region is fixed at the pixel level. Therefore it is possible to learn from the raw pixel values, without any need to apply several layers of convolutions as is done in many other networks.

Search-based approaches investigate and implement a promising search based face annotation scheme. Here mining large amount of poor labeled web images freely available on the WWW. For better understanding, suitable example is mining weakly labeled web facial images for search-based face annotation [22] .Also formulate the learning problem as a convex optimization and develop effective optimization algorithm to solve the major learning task efficiently. To further speed up the proposed scheme also propose a clustering based approximation algorithm which can improve the availability considerably.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

The most related works to this paper are cluster-based approaches. The fundamental assumptions behind of this are that that faces belonging to a person can be densely clustered and hence be exploited for face naming. Existing approaches are Gaussian mixture models (CGMM) [17], [18] graph-based clustering (GC) [17] and face-name association by commute distance (FACD) [23].

III.PROPOSED SYSTEM

A. *Problem definition and Notation: Relationship Modeling*

Given a video in which the inputs to problem of face tagging consisting of a set of observed or detected faces from a video and celebrity names occupied from metadata. Celebrity faces represented as a set $N = \{C_1, C_2, C_3, \cdots\cdots C_M\}$ and celebrity names as sequence $S = (X_1, X_2, X_3, \ldots\ldots X_P)$ where M and P represent count of faces and names respectively. Then the problem can be defined as assigning at most one $C_i \in N$ to $X_i \in S$, from the assignment it is understable that every face in from a video given a name or no name (null). Also the output to the problem represented as the $Y = \{Y_1, Y_2, \ldots\ldots, Y_N\}$ gives the indexed variables which indicate the correct face assignment with exact name.

Conditional random field is used to model the graph for name interference. Inference is accomplished by drawing upon available ‘‘features’’ that correspond to each node and each edge. These features include both image data and context from the embedding social network. Mathematical representation is $G = (V, E)$, vertices V = (S, Y) represents set of faces and edges E represent the defined relationship between faces and between face and names. Fundamentally the problem is to trace out possible and suitable label assignments and then periodically pick out the best one as the solution to maximise the probability assignment. As a part of this initially estimate the conditional probability $p(X|Y)$. Following the local Markov property in CRF [12], we assume that two indexed variables $y_i, y_j \in Y$ are independent of each other if there are no edges between them. This can be illustrates by example in Fig.2. The variable y_1 is dependent on variable y_4, but not dependent on variabley_2.The dependent variable is termed as “factor”. Here $\{y_1, x_1\}$ is a factor and $\{y_1, y_4, x_1, x_4\}$ is also a factor. The inference of names can be solved with off-the-shelf algorithms such as Markov Chain Monte Carlo (MCMC) [13] or Loopy Belief Propagation (LBP) [14], [15]

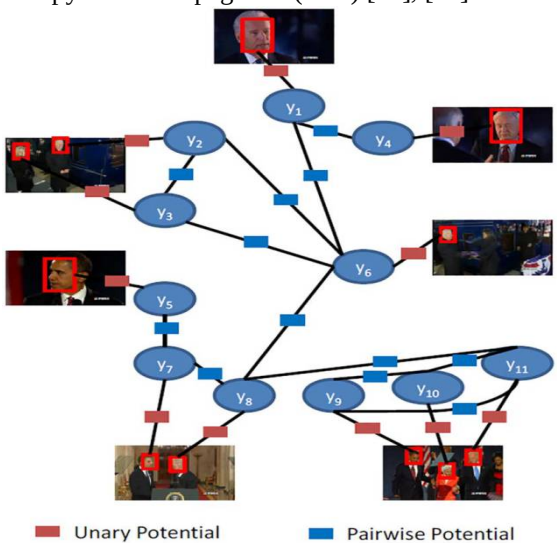

Fig.2 Example of graph depicting the modelling of relationships for face naming as an optimization

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

B. *Two types of potentials:*

With respect to the relationship modelling ultimately considering two kinds of [11] potentials, namely unary potential and pair-wise potential. Unary potential includes Face-Name relationship modelling, whereas pair–wise potential considers multivariate relationships.

Unary Potential: The unary potential [15] reads the likelihood of a face x_i being labelled with a name or "null" category. For this purpose, model the name as Multivariate Gaussian Distribution. Uncertainty is the exact term used to indicate the null category. Suppose the uncertainty is higher value then probability is uniformly distributed. Reversely when the probability of labelling name is very high then uncertainty becomes lower.

Pair-Wise Potential: Pairwise potential consist of linear combination of three relations, namely spatial, temporal and visual relations. The pairwise represents noticeable relationship between two faces. In spatial relationship, two frames of different shots, the spatial locations of faces, as well as their overlapping area, give clue to the identity of face. Similarly in temporal relationship, appearance of face at different time period gives clues whether the names assigned to the faces should be unique to each other. But the visual relation represents the background changes and colour changes.

C. *Architecture of Face labeling:*

The whole frame work for the unsupervised celebrity face naming is based on architecture shown in Fig. 3. Input consists of set of frames in a video and each frames includes number of different faces. Each faces have corresponding feature set and find matching among them. If there is a matching then a value is returned that means the corresponding name of the face. Now in (F2F) metadata , the data corresponds to each video that resides. If no matching occurs system searches in to the metadata. Meta data includes video name, frame number and celebrity names etc. By searching, corresponding faces are extracted from the video. As an output from metadata the feature set is generated and matching process restarted with input frames. Suppose matching again results as negative, then searches continuing with second meta data(N2N metadata) which includes images and names, and then find matches the celebrity data .

D. *Celebrity Face Naming with HOG scheme:*

According to the basic concept of face labelling, called Histogram Oriented Gradient (HOG) can be used. HOG is a combination of a series of steps. Before applying HOG scheme initially create the metadata also adds new test images in to the meta data. The whole system following F2F and N2N relationships. The basic objective of HOG system is object recognition. The basic idea behind of HOG system is Local shape information often well described by the distribution of intensity gradients without precise information about the location of the edges themselves. According to celebrity naming problem in HOG based object recognition, initial step is to divide image into small sub-images called "cells". Cells can be rectangular (R-HOG) or circular (C-HOG) . After this accumulate a histogram of edge orientations within that cell. In next stage, the combined histogram entries are used as the feature vector for describing the object.

Orientation binning and block normalization are further steps in here. In orientation binning is generating cell histograms. Each cells contains number of pixel values and these each pixel castes a vote for histogram channel. The basic consideration for this voting will be the values found in gradient computation. The cells themselves can change to rectangular or radial in shape. Also the histogram channels are spread over 0 to 180 degrees or 0 to 360 degrees and that depends on whether the gradient is "signed" or "unsigned". Block normalization is next subsequent step in which , gradient affected by the illumination changes are normalized.

Why HOG in unsupervised celebrity face tagging system? Because it can capture edge or gradient structure that is very characteristic of local shape. But surf based method used in early study not good in recognizing exact shape of the object. Capturing edge or gradient structure that closely relates the characteristic of local shape, within cell rotations and translations do not make changes in HOG values and the illumination invariance achieved through normalization. The method is similar to edge oriented scheme, scale-invariant feature transformation and shape contexts.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

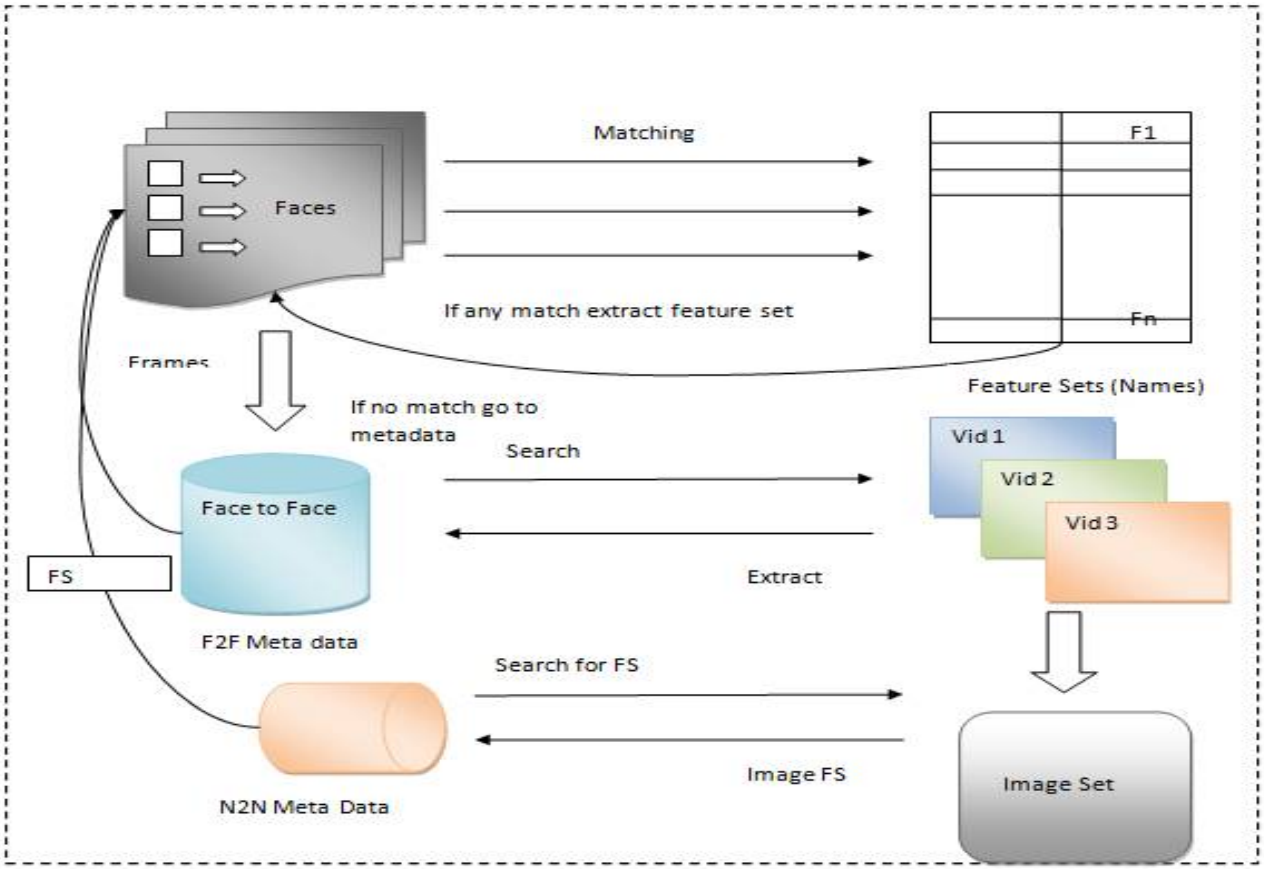

Fig.3 Architecture for unsupervised celebrity face naming framework.

IV. EVALUATION RESULTS

Similar to [16], [17], [19], [23], the performance is measured by accuracy and precision. Both measures count the number of faces correctly labelled, but differ where accuracy also includes the counting of faces without labels.

Note that accuracy and precision are calculated across all the faces in a test collection, rather than averaged over videos. Recalling of matching is not used here because we do not consider the problem of "retrieving all faces given a name", rather we are dealing with the problem of whether a face is labelled with a correct name (precision), otherwise labeled as "null" if the name is missing from metadata (accuracy).

Here evaluation study done on the basis of SURF based face recognition in existing approach [1]. Speeded Up Robust Features (SURF) [24] is a local feature detector and descriptor that can be used for tasks such as object recognition , registration , classification and 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The proposed method which takes Histogram of oriented gradients as feature set is compared against Surf features. The evaluation is done on the accuracy measurements i.e. no of faces identified on different web videos

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

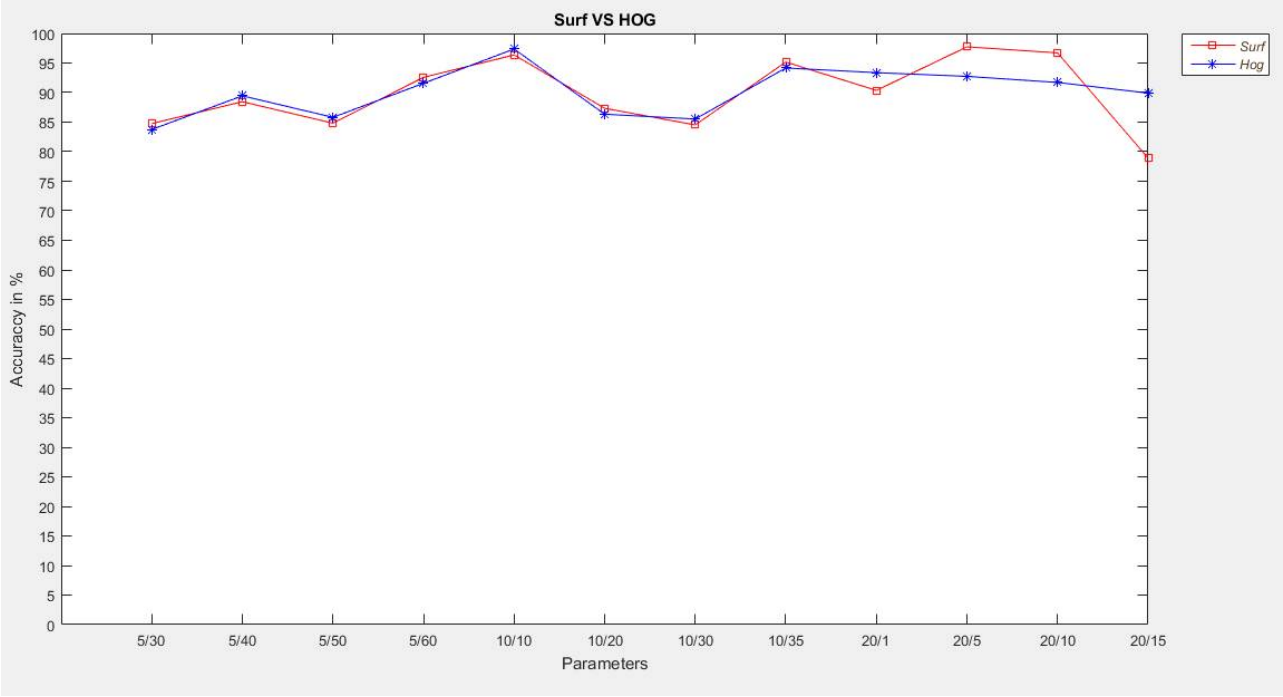

Fig.4 Comparison between surf based face recognition and HOG based recognition method

The proposed face naming system with HOG scheme implemented with MATLAB. Basic idea used for comparison is F2N relationship. For feature matching several parameters are trained. For example in surf based naming, for particular feature point extraction it needs to consider number of parameters like gradient, gaussian parameters etc. Those parameters are here represented with values 5/30, 5/40, 5/50 etc in x-coordinates. The graph shown in the Fig.4 representing the accuracy changes occurs with each parameters used in the experiment also graph shows a generalized result. Both cases include a window size and sigma value, sigma gives the actual Gaussian value. Here in 5/30, 5 is Gaussian value and 30 is the sigma value respectively. To create masking programme wavelength and sigma values are very essentials. According to the graph for an example, if out of 10 images 9 images are detected with exact identity then result is 99 % of accuracy. The good accuracy is resulted if the ratio between parameter is 10:10.So the evaluation results shows that the accuracy level of the new method is vary with ratio between each parameters.

V. CONCLUSION

The paper represented modelling the solution for celebrity face naming problem .Problem is experimented with a new method in face naming called HOG scheme. Dealing with the incomplete and noisy metadata, CRF smoothly encodes F2F and F2N relationships also permitting null category by considering uncertainty labelling. HOG scheme results a good effect than the previous face labelling method. Shape oriented feature detection based on HOG scheme shows a stable accuracy in almost level of parameter ratio. Therefore experiments results shows that parameter property leads to a good performance superiority over current methods. The price of improvement, nevertheless, also comes along with increase in processing time and the number of false positives.

REFERENCES

1. Lei Pang and Chong-Wah Ngo ,'Unsupervised Celebrity Face Naming in Web Videos, IEEE Transactions on multimedia, vol. 17, no. 6, june 2015
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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

3. Y. F. Zhang, C. S. Xu, H. Q. Lu, and Y. M. Huang, "Character identification in feature-length films using global face-name matching," IEEE Trans. Multimedia, vol. 11, no. 7, pp. 1276–1288, Nov. 2009
4. M. R. Everingham, J. Sivic, and A. Zisserman, ""Hello! My name is Buffy"–automatic naming of characters in TV video," in Proc. Brit. Mach. Vis. Conf., 2006, pp. 92.1–92.10.
5. Z. Stone, T. Zickler, and T. Darrell, "Toward large-scale face recognition using social network context," Proc. IEEE, vol. 98, no. 8, pp. 1408–1415, Aug. 2010.
6. L. Y. Zhang, D. V. Kalashnikov, and S. Mehrotra, "A unified framework for context assisted face clustering," in Proc. Int. Conf. Multimedia Retrieval, 2013, pp. 9–16
7. Y. Y. Chen, W. H. Hsu, and H. Y. M. Liao, "Discovering informative social subgraphs and predicting pairwise relationships from group photos," in Proc. ACM Int. Conf. Multimedia, 2012, pp. 669–678
8. J. Choi, W. De 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.
9. J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in Proc. Int. Conf. Mach. Learn., 2001, pp. 282–289
10. C. Sutton and A. McCallum, "An introduction to conditional random fields," Found. Trends Mach. Learn., vol. 4, no. 4, pp. 267–373, 2012.
11. W. Li and M. S. Sun, "Semi-supervised learning for image annotation based on conditional random fields," in Proc. Conf. Image Video Retrieval, 2006, vol. 4071, pp. 463–472
12. G. Paul, K. Elie, M. Sylvain, O. Marc, and D. Paul, "A conditional random field approach for face identification in broadcast news using overlaid text," in Proc. IEEE Int. Conf. Image Process., Oct. 2014, pp. 318–322.
13. C. P. Robert and G. Casella,Monte Carlo StatisticalMethods (Springer Texts in Statistics). New York, NY, USA: Springer-Verlag, 2005.
14. M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," Found. Trends Mach. Learn., pp. 1–305, 2008.
15. J. S. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," IEEE Trans. Inf. Theory, vol. 51, no. 7, pp. 2282–2312, Jul. 2005
16. V. F. Mert Özcan, L. Jie, and B. Caputo, "A large-scale database of images and captions for automatic face naming," in Proc. Brit. Mach. Vis. Conf., 2011, pp. 29.1–29.11.
17. M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Automatic face naming with caption-based supervision," in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2008, pp. 1–8
18. T. Berg, A. Berg, J. Edwards, and D. Forsyth, "Who's in the picture?," in Proc. Neural Inf. Process. Syst., 2005, pp. 137–144.
19. M. Tapaswi, M. Bäuml, and R. Stiefelhagen, ""Knock! Knock! Who is it?" Probabilistic person identification in TV-series," in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2012, pp. 2658–2665.
20. J. Yang, R. Yan, and A. G. Hauptmann, "Multiple instance learning for labeling faces in broadcasting news video," in Proc. ACM Int. Conf.Multimedia, 2005, pp. 31–40.
21. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2014, pp. 1701–1708.
22. D. Y. Wang, S. Hoi, Y. He, and J. K. Zhu, "Mining weakly labeled web facial images for search-based face annotation," IEEE Trans. Knowl.Data Eng., vol. 26, no. 1, pp. 166–179, Jan. 2014
23. J. Bu et al., "Unsupervised face-name association via commute distance,"in Proc. ACM Int. Conf. Multimedia, 2012, pp. 219–228.
24. S. W. Smoliar and H. Zhang, "Content-based video indexing and retrieval,"IEEE Multimedia, vol. 1, no. 2, pp. 62–72, Jun. 1994.

BIOGRAPHY

Anugraha Raj.S received the B-Tech from M.G University in 2014 and doing M.Tech in Computer Science from 2014 in Mangalam College of Engineering. Her area of interests includes data mining, security in computing and artificial intelligence.

Sreenimol K.R received B.Tech from Govt RIT ,Kottayam and M.tech from CUSAT, Cochin.She is currently an associate professor in Mangalam College of Engineering, Kottayam. Her area of interests includes data mining, computer architecture, social networks.