Vol. 7(24), Apr. 2017, PP. 3344-3353

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

Marking Faces Utilizing Annotations as a Part of Features for Videos

Pankaj Agarkar[1] and S.D. Joshi[2*]

Department of Computer Engineering, Research Scholar at JJTU, Zhunzhuno, Pune, India

Department of Computer Engineering, Research Guide at JJTU, Zhunzhuno, Pune, India

***Corresponding Author's E-mail:** sdj@live.com*

Abstract

Face Annotation is a note or description added to the image for better understanding. Also it can help to improve better search due to detailed description. If this annotation technique is used in video that can help in better searching of videos. The goal is to annotate unseen faces in videos with the words that best describe the image. Initially the database containing images and description mapping of that image will be gathered. Later videos that need to be processed will be considered. These videos will be converted to frames. This frame will act as images. These images will be processed with the existing database. If the faces are matched then it will be considered with the matching annotation. The matching results will produce thee matching annotation or null (the images that are not matched). Further training can be provided by the later result.The problem of naming can be traced back to name face association, where the goal is to align the observed faces with a given set of names in videos. Our proposed system give the Face candidate retrieval by name Automated video indexing by the person's name Automated creation of face-name correspondences database from thousands of hours of news videos. Use of Annotations has increased in images by adding Videos can also use this approach for associating face-name for videos can be a approach for better video searching. It will help for users to search desired videos, eg. News videos. Also systems with manual caption exist. If such system gets implemented then captions can get added automatically. Automatic tagging of people in videos will improve the search results. It can be further enhanced by considering different parameters like image background and other parameters for providing better description.

***Keywords**: Face Annotations, social network, Face recognition, unconstrained web videos mining, unsupervised.*

1. INTRODUCTION

Due to the popularity of various digital cameras and the rapid growth of social media tools for internet-based photo-video sharing, recent years have witnessed an explosion of the number of digital photos captured and stored by consumers. A large portion of photos/videos shared by users on the Internet are human facial images. Some of these facial images are tagged with names, but many of them are not tagged properly. This has motivated the study of auto face annotation, an important technique that aims to annotate facial images automatically. Auto face annotation can be beneficial to many realworld applications. For example, with auto face annotation techniques, online photo-sharing sites (e.g., Facebook) can automatically annotate users' uploaded photos to facilitate online photo search and management. Besides, face annotation can also be applied in news video

Article History:
Received Date: Oct. 12, 2016
Accepted Date: Feb. 05, 2017
Available Online: Mar. 02, 2017

domain to detect important persons appeared in the videos to facilitate newsvideo retrieval and summarization tasks.Classical face annotation approaches are often treated as an extended face recognition problem, where differentclassification models are trained from a collection of well-labeled facial images by employing the supervised or semi-supervised machine learning techniques. However, the "model-based face annotation" techniques are limited in several aspects. First, it is usually time-consuming and expensive to collect a large amount of human-labeled training facial images. Second, it is usually difficult to generalize the models when new training data or new persons are added, in which an intensive retraining process is usually required. Last but not least, the annotation/recognition performance often scales poorly when the number of persons/classes is very large.

Recently, some emerging studies have attempted to explore a promising search-based annotation paradigm for facial image annotation by mining the World Wide Web (WWW), where a massive number of weakly labeled facial images are freely available. Instead of training explicit classification models by the regular model-based face annotation approaches, the search-based face annotation(SBFA) paradigm aims to tackle the automated face annotation task by exploiting content-based image retrieval(CBIR) techniques [8], [9] in mining massive weakly labeled facial images on the web. The SBFA frameworkis data-driven and model-free, which to some extent is inspired by the search-based image annotation techniques [10], [11], [12] for generic image annotations. The main objective of SBFA is to assign correct name labels to a given query facial image. In particular, given a novel facial image for annotation, we first retrieve a short list of top K most similar facial images from a weakly labeled facial image database, and then annotate the facial image by performing voting on the labels associated with the top K similar facial images.

One challenge faced by such SBFA paradigm is how to effectively exploit the short list of candidate facial images and their weak labels for the face name annotation task. Totackle the above problem, we investigate and develop a search-based face annotation scheme. In particular, we propose a novel unsupervised label refinement (URL) scheme by exploring machine learning techniques to enhance the labels purely from the weakly labeled data without human manual efforts. We also propose a clustering- based approximation (CBA) algorithm to improve the efficiency and scalability. As a summary, the main contributionsof this paper includes the following:

- We investigate and implement a promising search based face annotation scheme by mining large amount of weakly labeled facial images freely available on the WWW.
- We propose a novel ULR scheme for enhancing label quality via a graph-based and low-rank learning approach.
- We propose an efficient clustering-based approximation algorithm for large-scale label refinement problem.
- We conducted an extensive set of experiments, in which encouraging results were obtained.

2. RELATED WORK

The Name-It system associates names and faces in news videos. Assume that we're watching a TV news program. When persons we don't know appear in the news video, we can eventually identify most of them by watching only the video. To do this, we detect faces from a news video, locate names in the sound track, and then associate each face to the correct name. For face-name association, we use as many hints as possible based on structure, context, and meaning of the news

video. We don't need any additional knowledge such as newspapers containing descriptions of the persons or biographical dictionaries with pictures.Similarly, Name-It can associate faces in news videos with their right names without using an a priori face-name association set. In other words, Name-It extracts face-name correspondences only from news videos. Name-It takes a multimodal approach to accomplish this task. For example, it uses several information sources available from news videos-image sequences, transcripts, and video captions. Name-It detects face sequences from image sequences and extracts name candidates from transcripts. It's possible to obtain transcripts from audio tracks by using the proper speech recognition technique with an allowance for recognition errors. However, most news broadcasts in the US already have closed captions. (In the near future, the worldwide trend will be for broadcasts to feature closed captions.) Thus we use closed-caption texts as transcripts for news videos. In addition, we employ video-caption detection and recognition. We used "CNN Headline News" as our primary source of news for our experiments [2].

Identification of characters in films, although very intuitive to humans, still poses a significant challenge to computer methods. In this paper, we investigate the problem of identifying characters in feature-length films using video and film script. Different from the state-of-the-art methods on naming faces in the videos, most of which used the local matching between a visible face and one of the names extracted from the temporally local video transcript, we attempt to do a global matching between names and clustered face tracks under the circumstances that there are not enough local name cues that can be found. The contributions of our work include: 1) A graph matching method is utilized to build.

Face-name association between a face affinity network and a name affinity network which are, respectively, derived from their own domains (video and script). 2) An effective measure of face track distance is presented for face track clustering. 3) As an application, the relationship between characters is mined using social network analysis. The proposed framework is able to create a new experience on character-centered film browsing. Experiments are conducted on ten feature-length films and give encouraging results [3].

Personal photographs are being captured in digital form at an accelerating rate, and our computational tools for searching, browsing, and sharing these photos are struggling to keep pace. One promising approach is automatic face recognition, which would allow photos to be organized by the identities of the individuals they contain. However, achieving accurate recognition at the scale of the Web requires discriminating among hundreds of millions of individuals and would seem to be a daunting task. This paper argues that social network context may be the key for large-scale face recognition to succeed. Many personal photographs are shared on the Web through online social network sites, and we can leverage the resources and structure of such social networks to improve face recognition rates on the images shared. Drawing upon real photo collections from volunteers who are members of a popular online social network, we asses the availability of resources to improve face recognition and discuss techniques for applying these resources [5].

Important inference problems in statistical physics, computer vision, error-correcting coding theory, and artificial intelligence can all be reformulated as the computation of marginal probabilities on factor graphs. The belief propagation (BP) algorithm is an efficient way to solve these problems that are exact when the factor graph is a tree, but only approximate when the factor graph has cycles. We show that BP fixed points correspond to the stationary pointsof the Bethe approximation of the free energy for a factor graph. We explain how to obtain region-based free energy

approximations that improve the Bethe approximation, and corresponding generalized belief propagation (GBP) algorithms. We emphasize the conditions a free energy approximation must satisfy in order to be a "valid" or "maxent-normal" approximation. We describe the relationship between four different methods that can be used to generate valid approximations: the "Bethe method," the "junction graph method," the "cluster variation method," and the "region graph method." Finally, we explain how to tell whether a region- based approximation, and its corresponding GBP algorithm, is likely to be accurate, and describe empirical results showing that GBP can significantly outperform BP.[15]

Current video management tools and techniques are based on pixels rather than perceived content. Thus, state-of-theart video editing systems can easily manipulate such things as time codes and image frames, but they cannot "know," for example, what a basketball is. Our research addresses four areas of content based video management.[16]

3. PROPOSED WORK

A. System Architecture

We proposed framework, which is by formulating the problem of within video face labeling as an optimization problem under conditional random field (CRF). Multiple relationships are then defined to characterize the sets of faces and names in the CRF. The scope of project includes the proposed technique should recognize faces to match with the database.This may be used in applications like video searching, for better video indexing. Here scope of project is related to video search.

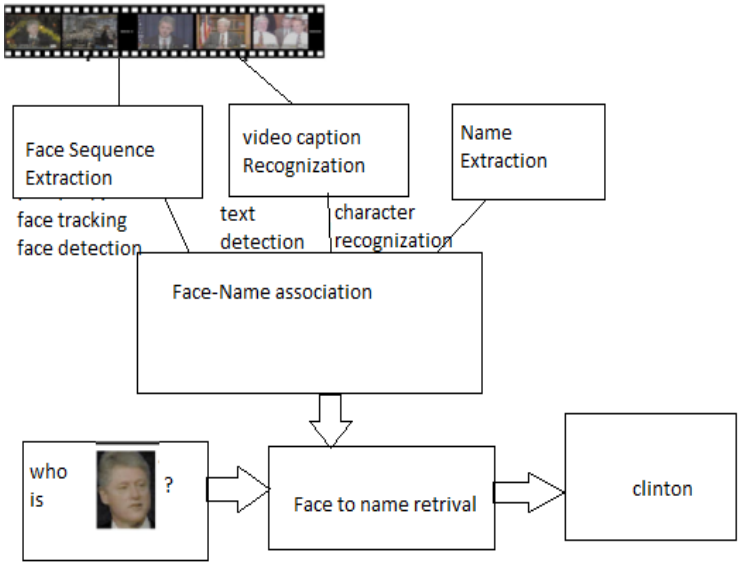

Fig.1. System Architecture

1) Face sequence extraction:

A video is given as input.The input video is made up of frames. Approx 20-25 frames/sec are there in a video. These frames are extracted from the video.Total frames in a video=Frame rate * total seconds in video. The frames are extracted from video. Give the input as video then features are extracted from it in the form of frames then we match them according to time and space.

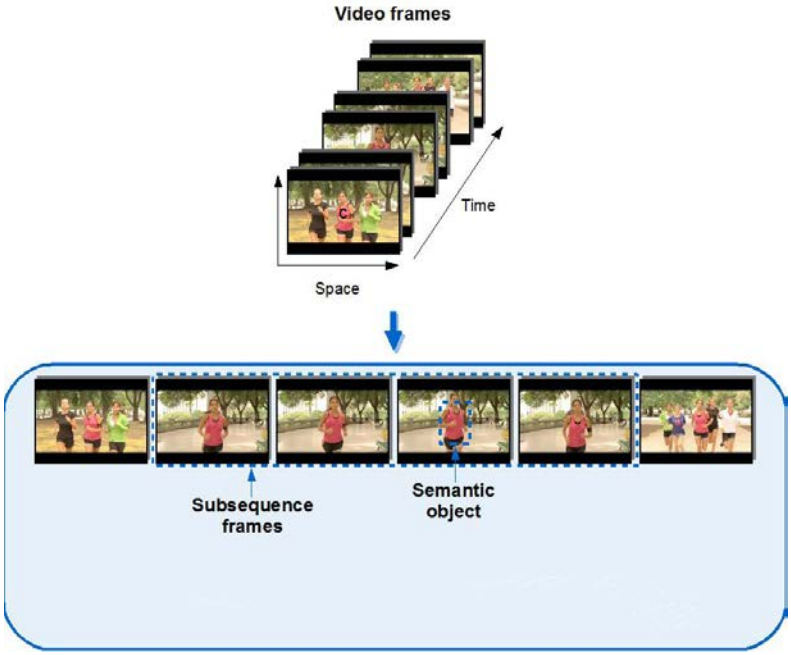

Fig. 2. Frame Extraction from video input

2) *Video caption recognizing:*

video name is also considered as a part of face identification. Although it does not play an integral part but yet it can help in sometimes.

3) Face recognition:

(a) *Verification (one-to-one matching):* When presented with a face image of an unknown individual along with a claim of identity, ascertaining whether the individual is who he/she claims to be.

(b) Identification (one-to-many matching): Given an image of an unknown individual, determining that person's identity by comparing (possibly after encoding) that image with a database of (possibly encoded) images of known individuals.

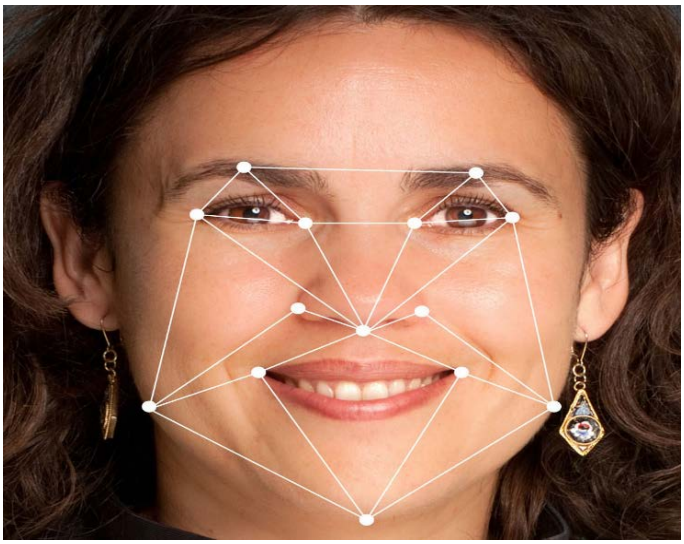

Fig. 3. Face Recognition

4) Face-Name Association:

To index and retrieve personal photos based on an understanding of "who" is in the photos, annotation (or tagging) of faces is essential. However, manual face annotation by users is a time-consuming and inconsistent task that often imposes significant restrictions on exact browsing through personal photos containing their interesting persons.

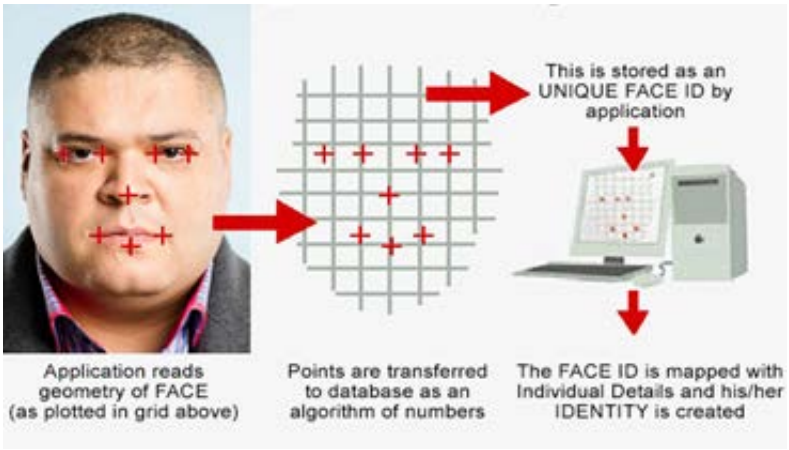

Fig. 4. Face-Name Association

5) Face to name retrieval:

a) Skin color extraction: After getting frames skin-tone color is extracted from the input image as the most important information of human face.

b) Face judgement: After lines-of-face detection, there may be some remaining noises because the lines-of-face template can only detect skin-tone contour.

c) Template matching: The matched template will be used compares with face name association. And the corresponding name will be considered.

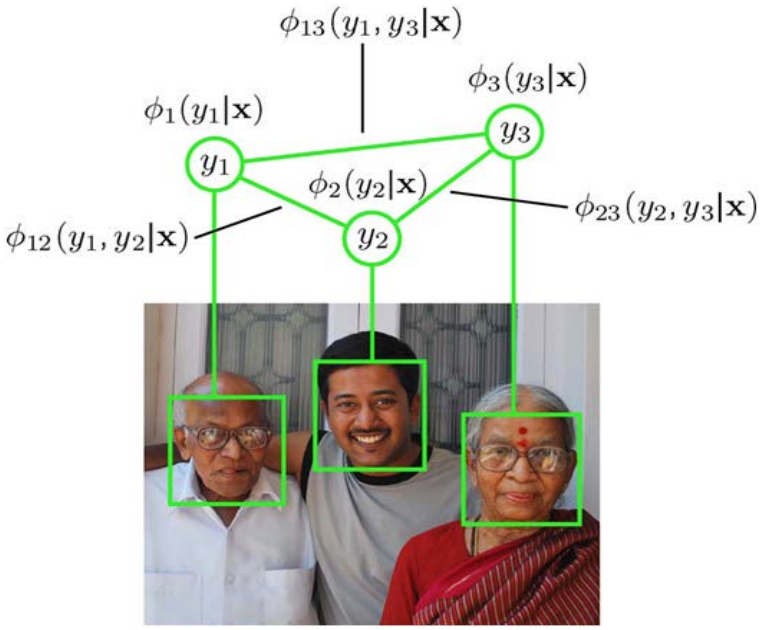

Fig. 5. Face to name retrival

B. Algorithm: Within-video face labelling Algorithm

Input: The sets of faces S and names N in a video V
Output: Face Labels Y that maximizes p(y/x) in

1) Constructing a graph G by modeling the unary potential for each face $x_i \in S$ where an edge between x_i and $y_i \in Y$ is weighted with unary potential.
2) Establishing edges for any pairs of $y_i \in Y$ and $y_j \in Y$ in G that satisfy the condition in two frames with identical area temporal relationship with their edgeweights set respectively based on spatial visual relationship.
3) Performing loopy belief propogation it basically calculates the co_occrances statistics of celebrities as proportion of videos where both names are tagged on G for face labelling.

C. *Mathematical Model*

1. Celebrity names as N=c1, c2...cM
2. Detected face sequence as S=x1, x2 …xN,
   Here M and N are number of names and faces respectively.
3. Output face annotations are given as Y=y1, y2... yN
4. Probability is given as

$$p(y|x) = \frac{1}{Z(x)} + \prod_{c \epsilon C} \Phi(yc, xc)$$

ϕ is potential function.

$$Z(\mathbf{x}) = \sum_y \prod_{c \in C} \Phi(\mathbf{y}_c, \mathbf{x}_c)$$

5. Is a partition function served for normalizing the probability score. $\emptyset$ Has unary potential μ(yi,xi) and pairwise potential y(yi,yj,xi,xj). New equation will be

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{c \in C_\mu} \mu(y_i, \mathbf{x}_i) \prod_{c \in C_\psi} \psi(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j)$$

D. Partial Implementation Module

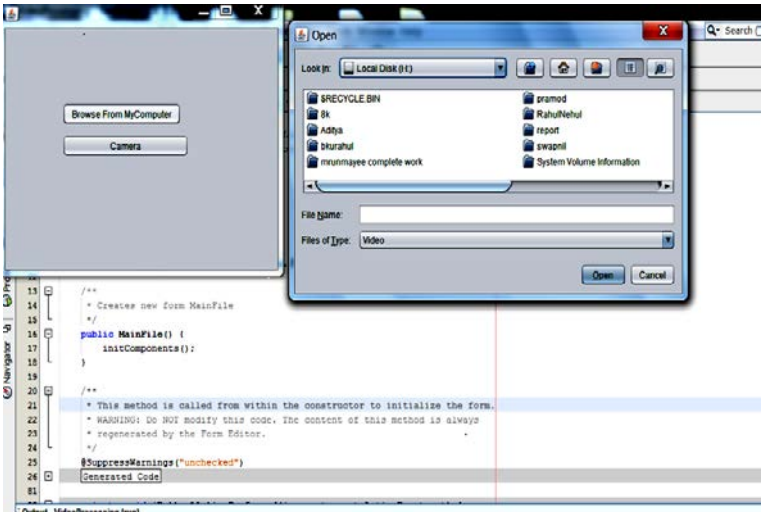

Fig. 6. Snapshort of video Input taken from computer

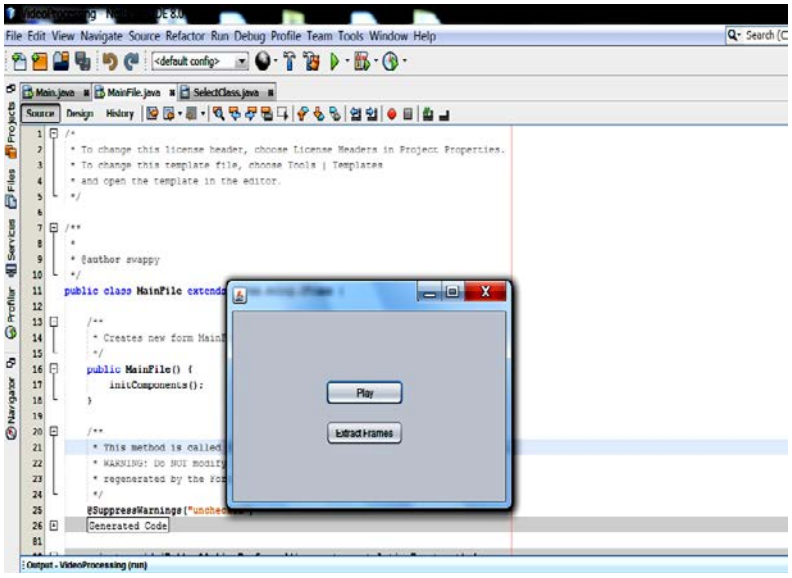

Fig. 7. Extraction of video into frames after taking input

In given above two snap short we done with 1st module partial implementation in that we taken video as input browse from computer or either from camera two options provided their according to take input.After that the second snapshort shows the paly button that extracts all images from videos and stored in terms of frames in some dadabase records according to fraction of second the frames appeared. the extraction button provided for the extract the features from videos that nothing but number of frames images of that person belong that respective videos it further used for face to name retrivals.

E. Expected results

The below fig.8 shows the Excepted result of the perfeormance of calculating precision and recall for better annotaion of faces into videos improvement of processing time.It will help to search engine for faster search of videos.

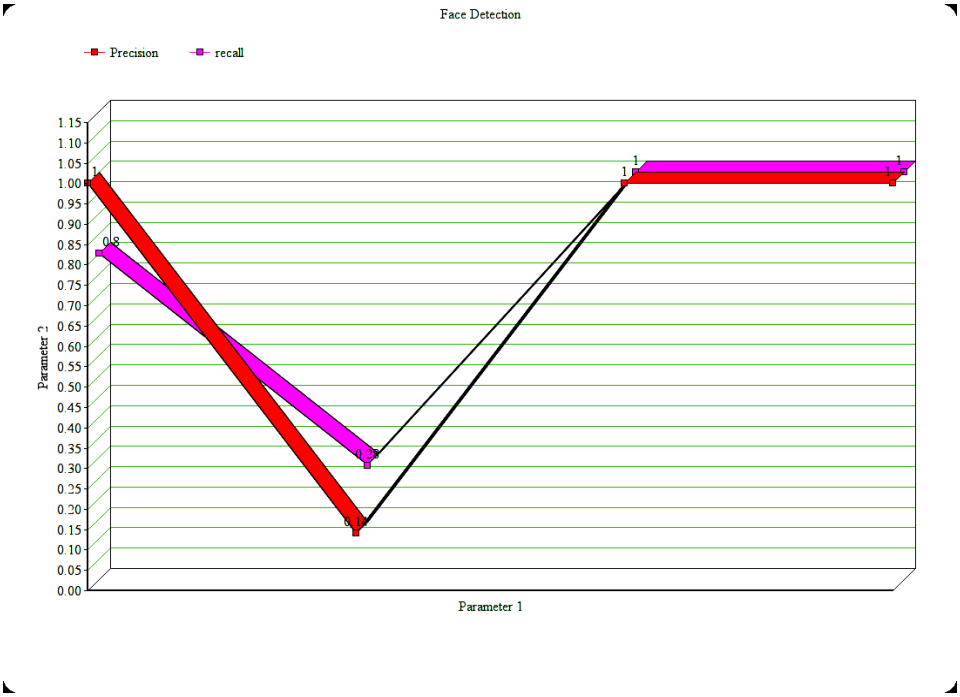

Fig. 8. Expected Result of Performance(precision & Recall)

Example:

1) total faces in a frame 5

Total faces detected 4.

Precision=4/4=1

Recall=4/5

2) Total faces in a frame 7

Total faces detected 2.

Precision: 1/7

Recall: 2/7

CONCLUSION

We have presented an approach for celebrity naming in the Web video domain. Our system associates faces and names in news videos by integrating face-sequence extraction and similarity evaluation, name extraction, and video-caption recognition into a unified factor: co occurrence.We considered. Graph-based and generative approaches to solving two tasks: finding faces of a single person, and naming all the faces in a data set. We have shown that we can obtain significant improvements over existing methods by improving and extending an existing graph-based method. Adding a language model, and enhancing face detection and facial feature localization can bring further improvements, as this will lead to cleaner data sets from which to construct the similarity graphs. The potential applications of the methods proposed in this paper include web-based photo retrieval by name, automatic photo annotation, and news digest applications.

REFERENCES

[1] J. Yang and A. G. Hauptmann, "Naming every individual in news video monologues," in Proc. ACM Int. Conf.Multimedia, 2004, pp. 580-587.

[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] 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 socialsubgraphs 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] A. 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] A. P. Robert and G. Casella, Monte Carlo StatisticalMethods (Springer Texts in Statistics). New York, NY, USA: Springer-Verlag, 2005.

[14] J. S. Yedidia, W. Freeman, and Y.Weiss, "Constructing free-energy approximations and generalized belief propagation

[15] S. W. Smoliar and H. Zhang, "Content-based video indexing and retrieval," *IEEE Multimedia*, vol. 1, no. 2, pp. 62–72, Jun. 1994.