MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 - 6953

NAMING FACES USING ANNOTAIONS BASED ON EXTERNAL KNOWLEDGE FROM VIDEOS

Navale Ashwini M., Prof.Pankaj Agarkar

Department of Computer Engineering
Dr.D.Y.Patil School of Engineering Pune, Maharashtra, India

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 if used in videos 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. This 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 exists. 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 naming, social network, unconstrained web videos mining, unsupervised.*

1. INTRODUCTION

Labeling celebrities in Web videos is a challenging problem due to large variations in face appearance. The problem becomes increasingly important due to the massive growth of videos in Internet. According to surveys, 1 about 80% of popular videos are people-related and among the people-related videos, about 75% are about celebrities. To date, most search engines index these videos with user-provided text descriptions (e.g., title, tag), which are often noisy and incomplete. The descriptions are given globally, and hence the correspondences between celebrity names and faces are not explicit. It is not unusual that a mentioned celebrity does not appear in the video, and vice versa, a celebrity actually appearing in a video is not mentioned. [4] For these reasons, searching people-related videos may yield unsatisfactory retrieval performance, either because of low recall or low precision. Ideally, finding the direct correspondences between names and faces could help rectify the potential errors in text descriptions and thus serve as a preprocessing step for video indexing. Furthermore, user search experience could be improved if the name-face correspondence is visualized, for example, by showing the name of a celebrity when a cursor moves over a face.[2] The problem of celebrity naming can be traced back to nameface association, where the goal is to align the observed faces with a given set of names. This problem has been attempted in the domains of news videos movies and TV series, capitalizing on the rich set of time-coded information including speech transcripts and subtitles. Nevertheless, these approaches often assume the ideal situation where the text cue is “rich” such that the given name set is free-of-noise and can perfectly match the observed faces. As a consequence, directly extending these approaches to Web video domain is not straightforward. Utilization of rich context information for face naming is also studied in the domain of personal album collection, by using timestamps, geotags, personal contact lists and social networks.[5] Nevertheless, these approaches cannot be directly applied for domain unrestricted videos, because of the absence of context cues and prior knowledge such as family relationships for problem formulation.

2. LITERATURE SURVEY

- We investigate the problem of face identification in broadcast programs where people names are obtained from text overlays automatically processed with Optical Character Recognition (OCR) and further linked to the faces throughout the video. To solve the face-name association and propagation, we propose a novel approach that combines the positive effects of two Conditional Random Field (CRF) models: a CRFfor person diarization (joint temporal segmentation and association of voices and faces) that benefit from the combination of multiple cues including as main contributions the use of identification sources (OCR appearances) and recurrent local face visual dentification of the person clusters that improves identification performance thanks to the ubackground (LFB) playing the role of a namedness feature a second CRF for the joint ise of further diarization statistics.[7]

- We describe a probabilistic method for identifying characters in TV series or movies. We aim at labeling every character appearance and not only those where a face can be detected. Consequently, our basic unit of appearance is a person track (as opposed to a face track). We evaluate our approach on the first 6 episodes of The Big Bang Theory and achieve an absolute improvement of 20% for person identification and 12% for face recognition.[8]

- In modern face recognition, the conventional pipeline consists of four stages: detect⇒align⇒represent⇒classify. We revisit both the alignment step and the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and derive a face representation from a nine-layer deep neural network. This deep network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Our method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.[9]

3. MODULE DESCRIPTION

1. **Face sequence extraction:** This modules consists of 2 parts

   a) ***Frames extraction:*** 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.

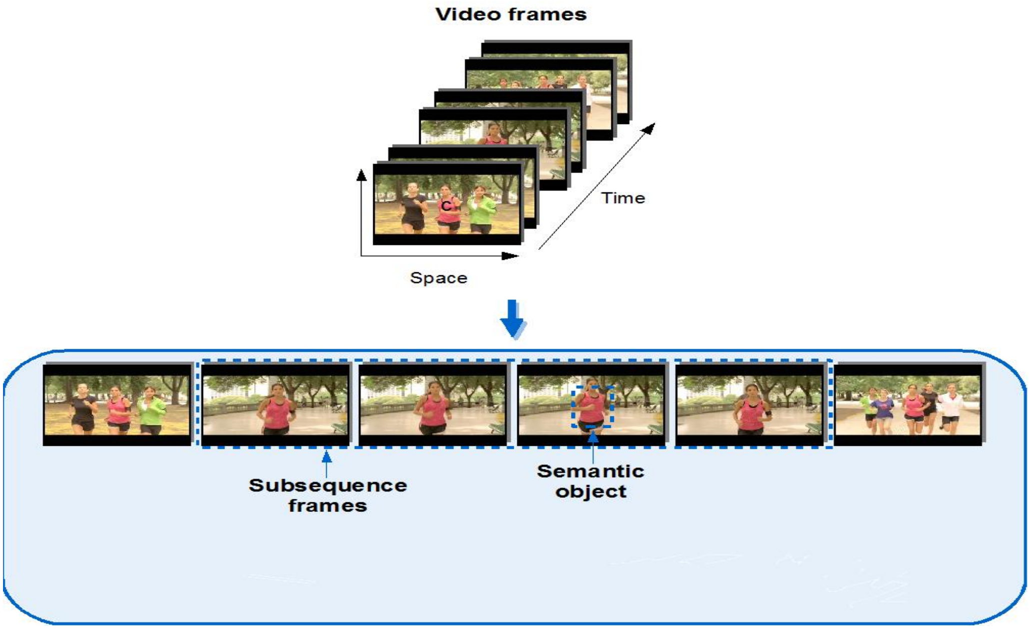

Fig.1: frames Extraction

b) Face sequence association: Frames that contains images are considered. From that frames faces are extracted.The recognition of human faces is not so much about face recognition at all – it is much more about face detection! It has been proven that the first step in automatic facial recognition – the accurate detection of human faces in arbitrary scenes, is the most important process involved.

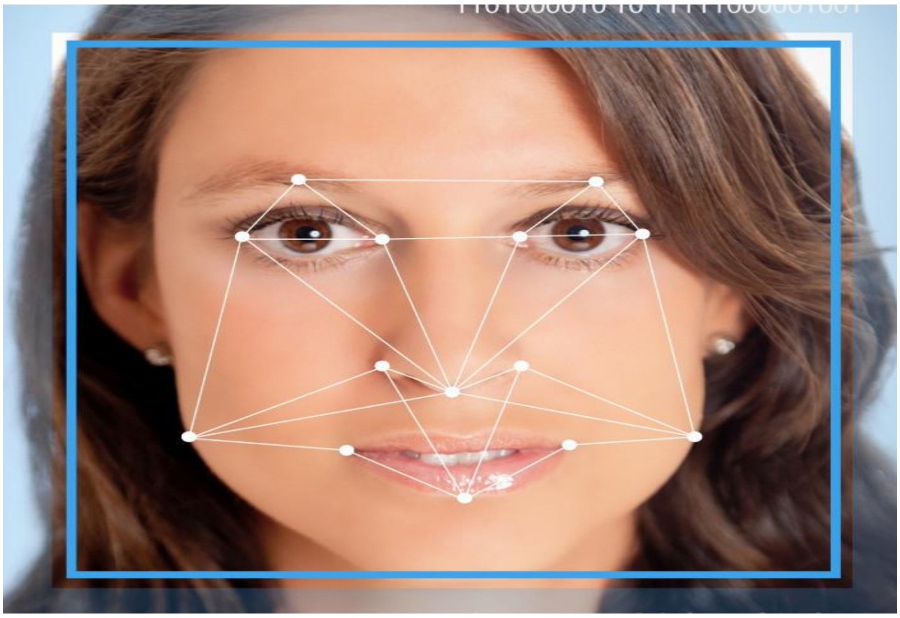

Fig.2: Face sequence association

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

3. Face name association: This part is the database part in the project. Here all the faces and its corresponding geometry features are stored.

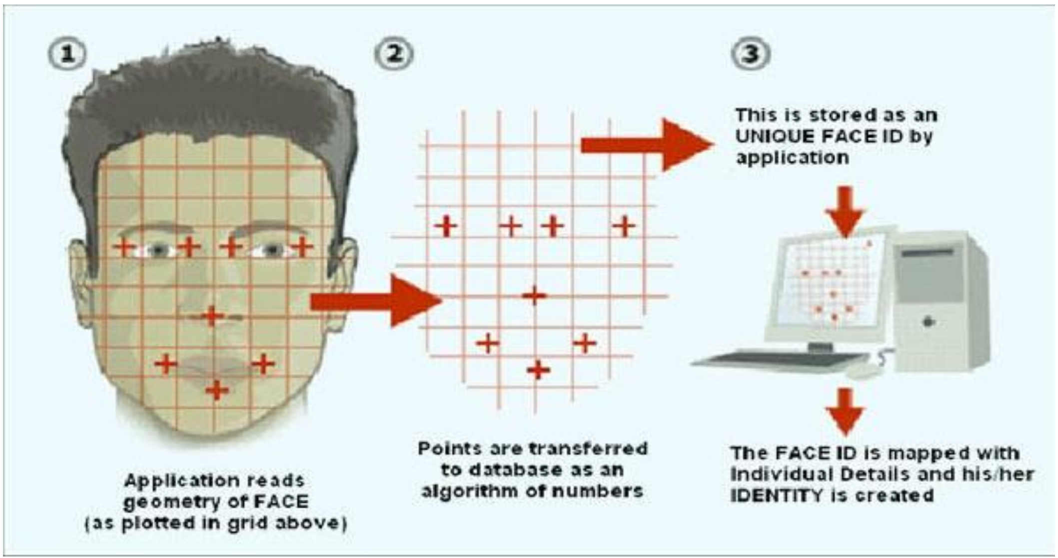

Fig.3: Face name association

While under going the face- name association, following factors are considered
a) Lighting compensation: This will adjust contrast of image.
b) Eigenfaces: Face Images are projected into a feature space ("Face Space") that best encodes the variation among known face images. The face space is defined by "Eigenfaces" which are the eigenvectors of the set of faces.

4. Face to name retrival:

a) Skin color extraction: After getting frames skin-tone color is extracted from the input image as the most important information of human face. Further coarse detection, edge extraction and blurring are executed for skin-tone flag map. This emphasizes the edges of skin-tone for face candidate detection.

b) Face judement: 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.

4. CONCLUSION AND FUTURE WORK

We have presented the modeling of multiple relationshipsusing CRF for celebrity naming in the Web video domain. In view of the incomplete and noisy metadata, CRF softly encodes these relationships while allowing null assignments by considering the uncertainty in labeling. Experimental results basically show that these nice properties lead to performance superiority over several existing approaches. The consideration of between-video relationships also results in further performance boost, mostly attributed to the capability of rectifying the errors due to missing names and persons. The price of improvement, nevertheless, also comes along with increase in processing time and the number of false positives. Fortunately, the proposals of leveraging social relation and joint labeling by sequential video processing still make CRF scalable in terms of speed and memory efficiency. While the overall performance of the proposed approach is encouraging, the effectiveness is still limited by facial feature similarity, which is used in the unary energy term and pairwise visual relationship. It can be further enhanced by considering different parameters like image background and other parameters for providing better description.

ACKNOWLEGEMENT

We thank our guide Prof. Pankaj Agarkar ME coordinator of CSE for his inspiration and guidance.

REFERENCES

[1]. 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.

[2]. 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.

[3]. 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.

[4]. J. Choi, W. De Neve, K. N. Plataniotis, and Y. M. Ro, "Collaborativeface 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.

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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.
[9]. 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.
[10]. 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.
[11]. D. Y. Wang, S. C. Hoi, Y. He, J. K. Zhu, T. Mei, and J. B. Luo, "Re- trieval-based face annotation by weak label regularized local coordi- nate coding," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 3, pp. 550–563, Mar. 2014.