



International Journal of Engineering, Business and Enterprise Applications (IJEBA)

www.iasir.net

Automatic Video Annotation using ECGM

S. Divya Meena¹, S. Vidhya Meena²
ME CSE¹, BE CSE²

Kingston Engineering College, Vellore– 632 059, Tamil Nadu, INDIA^{1,2}

Abstract: Identifying faces of characters in movies, though a challenging problem has drawn a significant research interest and led to many interesting applications. Automatic Video Annotation can be used to match the Face and Name of a character in movies. While there are many a methods which produces hopeful results in clean environment, the performance of these methods fail in complex movie scenes due to the inevitable noises generated during face tracking, detecting and clustering. The paper uses two schemes namely; (i) Face Name Matching with Cluster (ii) Face Name Matching without Cluster. Both the schemes use Error Correcting Graph Matching (ECGM) Algorithm in common. The complex character change in movies is handled by Simultaneous Graph Partition and Graph Matching. Two simulated noises are used to achieve an In-Depth Sensitivity Analysis. Finally Global Face-Name Graph Matching based framework for robust movie character identification is used. In brief, the core of the problem is to exploit the relations between videos and the associated texts in order to label them with the corresponding names in the cast.

Keywords: ECGM; Sensitivity Analysis; Video Annotation.

I. Introduction

The explosion of movie and TV has resulted in large amount of digital video data. The need of the hour is for an efficient and effective technique with which the videos can be understood and organised. Automatic video annotation can be applied for this purpose. This project focuses on annotating characters in the movie and TVs, which is called “Movie character identification”.

Character identification problem develops a relation between videos and the texts associated to it, with which we can label the faces of characters with its appropriate name. Video Annotation is similar to identifying faces in news videos. But, in news videos, names of the faces are available from the local texts that appear simultaneously.

And in TVs and Movies, names of the characters are mentioned in the script and not in the screenplay. But the script doesn't contain any timestamp to align to the video.

Movies content and its structure can be decided by the characters involved, which make the focus point of the audience. Automatic character identification is essential for semantic movie index and retrieval scene segmentation, summarization and other applications.

II. System architecture

The system architecture explains the various processes in video annotation. This process altogether forms the video annotation technique. Following are the process involved in the video annotation technique;

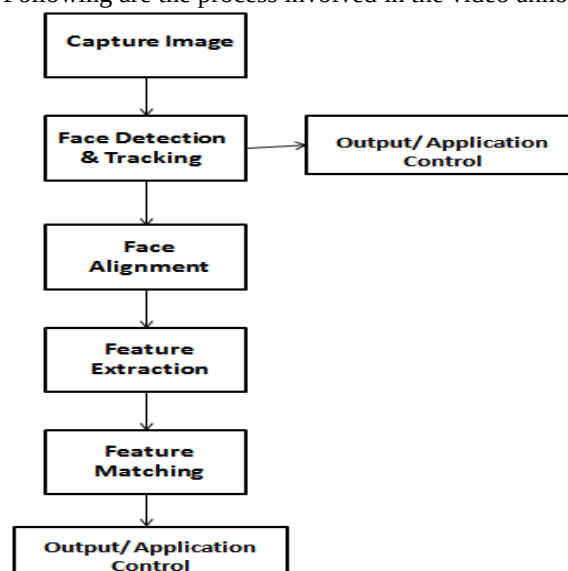


Figure 1: Architecture Diagram for Face detection and Recognition

Error Correcting Graph Matching (ECGM):

The two Schemes that are considered in robust face name graph matching uses ECGM algorithm, a powerful tool for graph matching with distorted inputs i.e. to measure the similarity of two graphs.

In ECGM, the difference between the two groups is measured by editing distance which is a sequence of graph edits operations. The optimal match is achieved with the least edit distance.

Method 1:

Method 1 Face tracking framework is clustered using constrained K-means, where the number of clusters is set as the number of distinct speakers. Co-occurrence of names in script and face in videos represents the corresponding face graph and name graph. We use ordinal graphs for robust representation and an additional ECGM-based graph matching method is introduced. The face and name graph construction will be represented in rank ordinal level, which gets the strength of the relationships in a rank order from the weakest to strongest. Rank order data carry no numerical meaning and thus are less sensitive to the noises.

Method 2:

The proposed framework for Method 2 has two differences from Method 1.

1. Face tracks clustering step requires no cluster number.
2. A graph partition component is added before numbered graph representation, since the face graph and name graph may have different number of vertexes.

Method 2 groups face tracks from different characters, instead of from same character. For this reason, we use Similarity propagation for clustering the face tracks. For each cluster, we take a sample and the face tracks are recursively clustered through the appearance based similarity broadcast and propagation. With this, we expect a high purity cluster with large number of clusters.

Graph partition is introduced before graph matching, since one character name may correspond to several graph matching. If the partitioned graph achieves an optimal graph matching with the name graph, then the face cluster can be grouped further. In brief, we perform face clustering in two steps as (i) Clustering by appearance i.e., Face Clustering (ii) Modifying by script. Simultaneously, we optimize face clustering and graph matching to improve the robustness against errors and noises.

III. System analysis

A Global face-name graph matching based framework for robust movie character identification is used. Two methods are considered. There are connections as well as differences between them. Regarding the connections, firstly, the two methods both belong to the global matching based category, where external script resources are utilized. Secondly, to improve the robustness, the ordinal graph is employed for face and name graph representation and a novel graph matching algorithm called Error Correcting Graph Matching (ECGM) is introduced.

Regarding the differences, method (1) sets the number of clusters when performing face clustering.

The face graph is restricted to have an identical number of vertices in the name graph. While, in the method (2), no cluster number is required and face tracks are clustered based on their intrinsic data structure. Moreover method (2) has an additional module of graph partition compared with a method (1). From this perspective, the method (2) can be seen as an extension to the method (1).

IV. Sensitive analysis

Unavoidable noises are generated during the process of face detection, face tracking and face clustering. This may be due to the pose, expression, illumination variation as well as low resolution problem. All this implicitly mean that the derived face graph and the name graph do not exactly match. So, the face graph can be viewed as a noisy version of the name graph. Therefore, sensitivity analysis step has become a must to evaluate the performance of the methods.

In this section, we first use two types of simulated noises namely, coverage noise and intensity noise. The sensitivity score for measuring ordinal graph downgrading is defined to evaluate the robustness of the ordinal similarity graph to noises.

V. Coverage noise and intensity noise

In the graph construction process for character identification, it is found that several noises such as vertex substitution, edge substitution and edge destruction or creation are involved. According to that, we use two types of noises, coverage noise and intensity noise for simulation.

The graph connectivity structure is very important in graph matching. We use the graph to edit operations of edges creation and destruction to simulate the changes to the topology of the graph. We call this as coverage noise. The creation or destruction probability for each existing or potential edge denotes the coverage noise level. Intensity noise refers to the changes in the edge's weight. It has involvement with the quantitative variation of the edges, but with no care for the graph structure. A random value distributed uniformly in the range denotes the intensity noise level.

VI. System design

A. Detection

In detection module we are going to detect the face of the movie characters using Error Correcting Graph Matching (ECGM), which is a powerful tool for graph matching with distorted Inputs.

The Microsoft Face Detection system provides a solution that can automatically detect faces in still images and real time video feeds. The system can detect an arbitrary number of faces at any scale and location. Input to the system is either a photographic image or a video. The output of face detection is then aligned. If it detects no faces, it will return an empty array.

The system attempts to eliminate false positives by skipping non-faces. It applies a color filter and an edge filter to improve the precision of the detection. It uses a lighting correction to further eliminate false positives. The detected faces along with confidence are outputted in the form of rectangle.

The following graphic illustrates the Face Detection system:

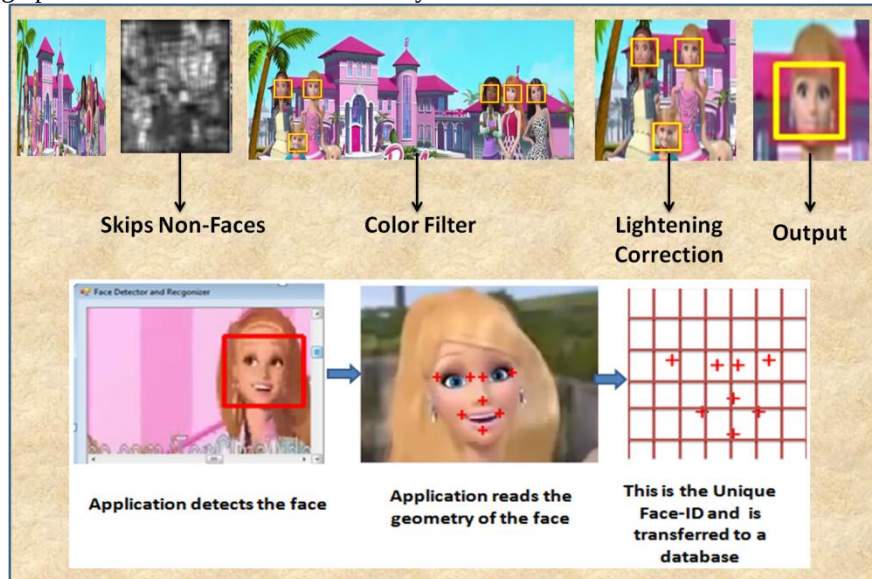


Figure 2: Face detection

B. Recognition

Face recognition is the task of recognizing a person using digital face images. FRS is used to measure the similarity between two faces or images and produce the output. Automated FRSs typically involve finding key facial landmarks (such as the centre of the eyes) for alignment, normalizing the face's appearance, choosing a suitable feature representation, learning discriminative feature combinations, and developing accurate and scalable matching. To recognize the face of the movie characters which is we previously stored on the face database, we just need to give the real name of it. This will be done with the help of the Eigen Object Recognizer we use to recognize the face. Four key factors that significantly compromise recognition accuracy are posing, illumination, expression, and aging it shows the impact of facial aging on face-recognition performance. Thus, deployments of fully automated FRSs are mostly limited to scenarios in which we can largely constrain these factors.

VII. Conclusion

Face recognition in video is an active topic in computer society because of the many potential applications it has. Two different methods and a recent work on face retrieval are introduced. Finally, some challenges and research directions are discussed. Although we mainly focus on video-based approaches, recent years have witnessed some interesting still-image based approaches, which could be helpful for face recognition in video.

References

- [1] J. Sang, C. Liang, C. Xu, and J. Cheng, "Robust movie character identification and the sensitivity analysis," in ICME, 2011, pp. 1-6.
- [2] J. Sang and C. Xu, "Character-based movie summarization," in ACM MM, 2010, pp 71-89.
- [3] R. G. Cinbis, J. Verbeek, and C. Schmid, "Unsupervised Metric Learning for Face Identification in TV Video," in International Conference on Computer Vision, 2011.
- [4] M. Xu, X. Yuan, J. Shen, and S. Yan, "Cast2face: character identification in movie with actor-character correspondence," in ACM Multimedia, 2010, pp. 831-834.
- [5] L. Lin, X. Liu, and S. C. Zhu, "Layered graph matching with composite cluster sampling," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 8, pp. 1426-1442, 2010.
- [6] H. Cevikalp and B. Triggs, "Face recognition based on image sets" in CVPR, 2011, vol. 42, no. 18, pp. 1526-1542.

- [7] N. Cherniavsky, I. Laptev, J. Sivic, and A. Zisserman “Semisupervised learning of facial attributes in video”, in The first international workshop on parts and attributes (in conjunction with ECCV 2010), 2010.
- [8] M. Guillaumin, J. Verbeek, and C. Schmid, “Multiple instance metric learning from automatically labeled bags of faces”, in ECCV, 2010.
- [9] P. Pham, M. Moens, and T. Tuytelaars. Cross-media alignment of names and faces. *IEEE Transactions on Multimedia*, 12(1):pp.13–27, 2010.
- [10] R. Hong, M. Wang, M. Xu, S. Yan, and T.-S. Chua, “Dynamic captioning: video accessibility enhancement for hearing impairment”, in *ACM Multimedia*, 2010, pp. 421–430.
- [11] Y. Zhang, C. Xu, J. Cheng, and H. Lu, “Naming faces in films using hypergraph matching,” in *ICME*, 2009, pp. 278–281.
- [12] T. Cour, B. Sapp, A. Nagle, and B. Taskar, “Talking pictures: Temporal grouping and dialog-supervised person recognition,” in *CVPR*, 2010, pp.1014–1021.
- [13] R. G. Cinbis, J. Verbeek, and C. Schmid, “Unsupervised Metric Learning for Face Identification in TV Video,” in *International Conference on Computer Vision*, 2011, pp. 278–281 .
- [14]. Arandjelović, O., Cipolla, R.: “A pose-wise linear illumination manifold model for facerecognition using video” in *Computer Vision and Image Understanding*, pp. 113–125, (2009)
- [15] J. Sang, C. Liang, C. Xu, and J. Cheng, “Robust movie character identification and the sensitivity analysis,” in *ICME*, 2011, pp. 1–6
- [16] J. Yang and A. Hauptmann, “Multiple instance learning for labelling faces in broadcasting news video,” in *ACM Int. Conf. Multimedia*, 2005, pp. 31–40.
- [17] A. W. Fitzgibbon and A. Zisserman, “On affine invariant clustering and automatic cast listing in movies,” in *ECCV* (3), 2002, pp. 304–320.
- [18] O. Arandjelović and R. Cipolla, “Automatic cast listing in feature length films with anisotropic manifold space,” in *CVPR* (2), 2006, pp. 1513–1520.
- [19] D. Ramanan, S. Baker, and S. Kakade, “Leveraging archival video for building face datasets,” in *ICCV*, 2007, pp. 1–8.
- [20] M. Everingham and A. Zisserman, “Identifying individuals in video by combining generative and discriminative head models,” in *ICCV*, 2005, pp. 1103–1110.
- [21] R. G. Cinbis, J. Verbeek, and C. Schmid, “Unsupervised Metric Learning for Face Identification in TV Video”, in *International Conference on Computer Vision*, 2011.
- [22] M. Xu, X. Yuan, J. Shen, and S. Yan, “Cast2face: character identification in movie with actor-character correspondence,” in *ACM Multimedia*, 2010, pp. 831–834.
- [23] M. Everingham, J. Sivic, and A. Zisserman, “Hello! my name is... buffy automatic naming of characters in tv video,” in *Proceedings of BMVC*, 2006, pp. 889–908.
- [24] J. Sivic, M. Everingham, and A. Zisserman, “Who are you? – learning person specific classifiers from video,” in *Proceedings of CVPR*, 2009, pp. 711–734.
- [25] T. Cour, C. Jordan, E. Mitsakaki, and B. Taskar, “Movie/script: Alignment and parsing of video and text transcription,” in *ECCV* (4), 2008, pp. 158–171.