Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 4, April 2013, pg.232 – 237

RESEARCH ARTICLE

Character Identification Using Graph Matching Algorithm

S. Bharathi [1], P. Manirathnam [2], K. Ramya[3], R. Ranjith Kumar [4], Mrs. M. Tamilarasi [5]

[1,2,3,4]Department of Computer Science and Engineering, Anna University Chennai, India

[5]Assistant Professor, Department Of Computer Science and Engineering,

K.S.R. College Of Engineering, Tiruchengode, India

[1] rathiranya@gmail.com ; [2] manirathnam60@gmail.com ; [3] ramya1736@yahoo.com ; [4] ranjith.rhl@gmail.com

Abstract— In this paper we are going to present a new efficient technique to identify the faces of the characters automatically in movies that drawn a significant research that led to many interesting applications. It is a huge problem due to the variations of each character in it. The existing applications will provide a good result for clean environment, but for complex environments the performance is limited due to the external noises generated while face tracking and clustering processes. In this paper we present a scheme for Global face-name matching for robust character identification. This include: 1) A noise insensitive character relationship representation is incorporated.2) we introduce an edit operation based graph matching algorithm.3) Complex character changes are handled by simultaneously graph portioning and matching.4) Beyond existing character identification approach, we further perform an in-depth sensitivity analysis by in introducing two types of simulated noises. The above proposed schemes will demonstrate a state-of-art-performance on movie character identification in various complex movies.

Key Terms: - Character identification; graph matching and graph portioning algorithm; graph; edit; sensitivity analysis

I. INTRODUCTION

The explosion of movie and TV provides a huge amount of digital video data. It led to the need of more efficient techniques of video content analysing and organization. Automatic video annotation is one of the key techniques to the proposed schemas. Automatic video explanation is one of the key techniques. In this paper our focus is on explaining characters in the digital video's and TV's which is termed as movie character identification. The main goal is to identify the faces of the characters in the digital video and label them with their corresponding names in the cast. The textual cues like lists, scripts, sub-titles and closed captions are usually demoralized. In the Fig. 1 shows an example in our experiment. In a movie, characters are the focus point of interests for the audience. Their occurrences provide as a lots of clues about the structure of the movie and its contents. Automatic character identification is essential for semantic movie index and retrieval, scene segmentation, summarization and also for other applications

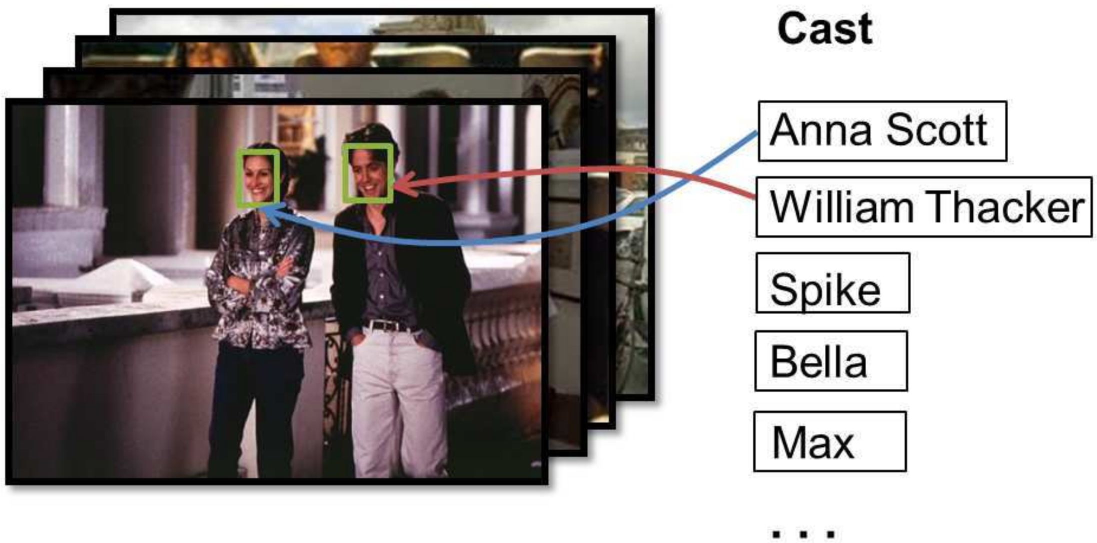

Fig.1 Examples of character identification from movie "Notting Hill"

Character identification, still very intuitive to humans, is a massively challenging task in a computer vision. The reason is four-fold: 1) weakly supervised textual cues. There are indistinctness problem in establishing the correspondence between the names and faces. The indistinctness can arise from a reaction shot where the person speaking may not be shown in the frames, and indistinctness can arise in partially labelled frames when multiple speakers are available in the same scene.2) Face identification is in videos is not easier as in images. Low resolution, occlusion, non-rigid deformations, large motion, complex background and other uncontrolled conditions may lead the face identification and tracking more unreliable. In movies the scenario is worse. This brings unexpected noises to the character identification. 3) The same character will appear as a different one during the movie. There may be a huge pose, expression and illumination variation, wearing, clothing, even makeup and hair-style changes. Moreover the characters in the movie will go throw different ages e.g. from young age to old age. Sometimes may be different actors playing different ages of the same character.4) determination for the number of identical faces is not trivial due to the remarkable intra-class variance, the same character name will correspond to faces of huge variant appearances. It will be unreasonable to set the number of identical faces. Our study is motivated by these challenges and aims to find solutions for a robust framework for movie character identification.

II. TECHNIQUES

Category 1: Cast list based:

These above methods will only make use of the case list textual source. In the "cast list discovery" problem faces are clustered by appearance and faces of a particular character are expected to be collected in a few pure clusters. Names for the clusters are then manually selected from the cast list. The authors have addressed the problem of finding particular characters by constructing a model/classifier of the character's appearance from user-provided guidance data. An interesting work combining character identification with web image repossession is proposed. The character names in the cast are used as queries to search face images and compose gallery set. The survey face tracks in the movie are then identified as one of the characters by multi-task joint scrubby illustration and categorization. Cast-specific metrics are adapted to the people appearing in a particular video in an unsubstantiated manner. The clustering as well as identification performance is demonstrated to be improved. These cast list based methods are easy for understanding and implementation. Though, without other textual cues, they either need physical labelling or assure that no robust clustering and classification performance due to the large intra-class variances.

Category 2: Subtitle or Closed caption, local matching based:

Subtitle and closed caption provide time-stamped dialogue, which can be demoralized for coalition to the video frames. They further extended their work in by replacing the nearest neighbour classifier by multiple kernel learning for features combination. In the new skeleton, non-frontal faces are handled and the coverage is extended. Researchers from University of Pennsylvania utilized the readily available time-stamped resource, the closed captions, which are demonstrated more reliable than OCR-based subtitles. They investigate on the faintness issues in the local alignment between video, screenplay and closed captions. A partially-supervised multiclass classification problem is formulated. Recently, they attempted to address the character identification problem without the use of screenplay. The reference cues in the closed captions are employed as multiple instance constraints and face tracks grouping as well as face-name association are solved in a convex formulation. The local matching based methods require the time-stamped information, which is either extracted by OCR or unavailable for the majority of movies and TV series. Besides, the uncertain and partial gloss makes local matching based methods more sensitive to the face detection and tracking noises.

Category 3: Script/Screenplay, Global matching based:

Global matching based methods open the possibility of character identification without OCR-based subtitle or closed caption. Since it is not an easy task to get local name cues, the task of character identification is formulate as a global matching problem. The method we are proposing belongs to this category and can be considered as an extension. In movies, the names of characters hardly ever directly appear in the subtitle, while the movie script which contains character names has no time information. Without the local time information, the task of character identification is formulated as a worldwide identical problem between the faces detected from the video and the names take out from the movie script. Compared with local matching, global figures are used for name-face organization, which enhance the forcefulness of the algorithms.

III. SCHEMES

In this paper two schemes are considered .There is both similarities and dissimilarities between them. Regarding similarities, the proposed both schemes belong to the global matching based category, where external script resources are used. For improving robustness the ordinal graph is employed for face and name graph representation. The novel graph matching algorithm called Error Correcting Graph Matching (ECGM) is introduced. Regarding the dissimilarities, scheme 1 sets the number of clusters when performing face clustering (e.g., K-means, spectral clustering). The face graph having same number of vertexes with the name graph. No cluster number is required and face tracks are clustered based on their intrinsic data structure in the scheme 2. Scheme 2 is said to be as the extension of scheme 1. Because scheme 2 has an additional module of graph partition compared with scheme 1.

SCHEME 1:

The proposed framework for scheme 1 is shown in Fig.2. It is similar to the framework of [2]. Face tracks are 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 clusters in video constitutes the corresponding face graph and name graph. We modify the traditional global matching framework by using ordinal graphs for robust representation and introducing an ECGM-based graph matching method. For face and name graph construction, we propose to represent the character co-occurrence in rank ordinal level, which scores 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. The affinity graph used in the traditional global matching is interval measures of the co-occurrence relationship between characters. While continuous measures of the strength of relationship holds complete information, it is highly sensitive to noises.

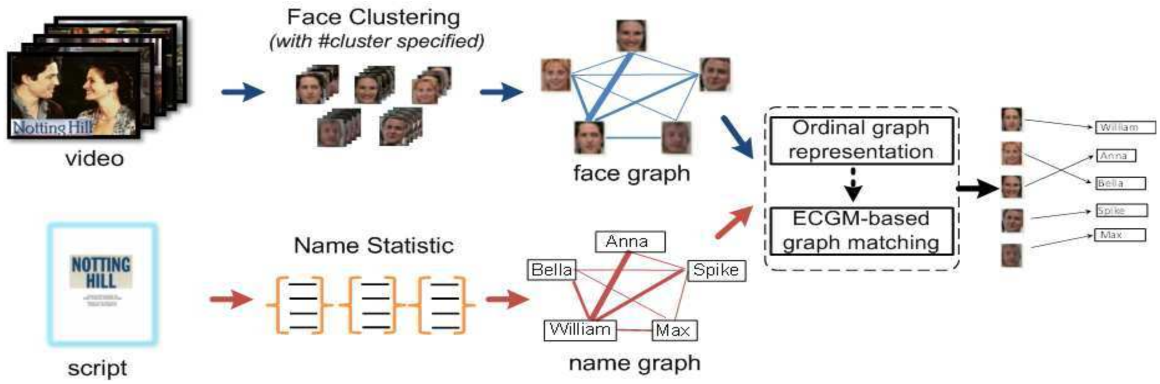

Fig. 2. Framework of Category 1: Face-name graph matching with #cluster prespecified

SHCEME 2:

Scheme 2 has two differences from scheme 1 first, no cluster number is required for the face tracks clustering step. Second, since the face graph and name graph may have different number of vertexes, a graph partition component is added before ordinal graph representation. The basic premise behind the scheme 2 is that appearances of the same character vary significantly and it is difficult to group them in a unique cluster. Take the movie "The Curious Case of Benjamin Button" for example. The hero and heroine go through a long time period from their childhood, youth, middle-age to the old-age. The intra-class variance is even larger than the inter-class variance. In this case, simply enforcing the number of face clusters as the number of characters will disturb the clustering process. Instead of grouping face tracks of the same character into one cluster, face tracks from different characters may be grouped together. In scheme 2, we utilize affinity propagation for the face tracks clustering. With each sample as the potential center of clusters, the face tracks are recursively clustered through appearance-based similarity transmit and propagation. High cluster purity with large number of clusters is expected. Since one character name may correspond to several face clusters, graph partition is introduced before graph matching. Which face clusters should be further grouped (i.e., divided into the same subgraph) is determined by whether the partitioned face graph achieves an optimal graph matching with the name graph. Actually, face clustering is divided into two steps: coarse clustering by appearance and further modification by script. Moreover, face clustering and graph matching are optimized simultaneously, which improve the robustness against errors and noises.

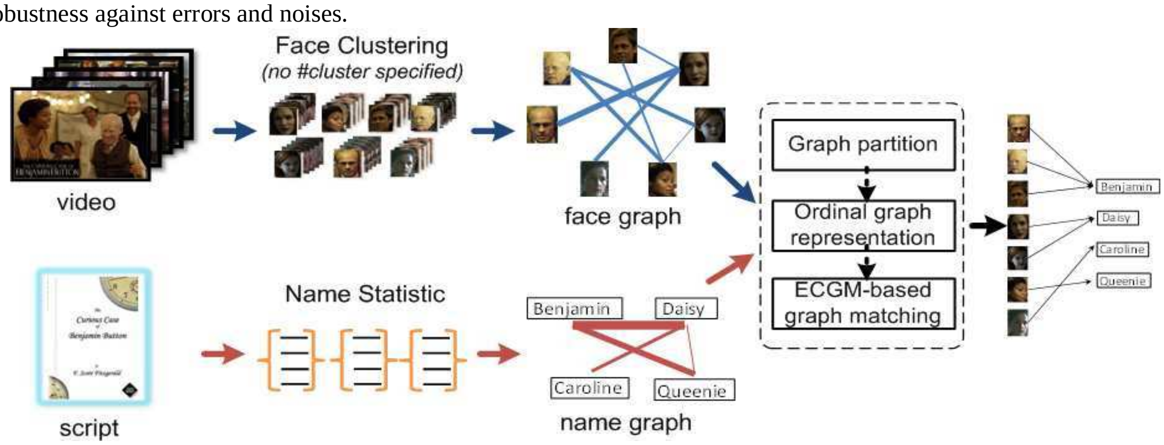

Fig. 3. Framework of scheme 2: Face-name graph matching without #cluster pre-specified

IV. ECGM-BASE GRAPH MATCHING

ECGM is a powerful tool for graph matching with deformed inputs. It has various applications in prototype recognition and computer vision. In order to calculate the resemblance of two graphs, graph edit operations are defined, such as the deletion, insertion and substitution of vertexes and edges. Each of these operations is auxiliary assign a certain cost. The costs are application dependent and usually reflect the possibility of graph distortion. The more likely certain distortion is to occur, the slighter is its cost. Through error correcting graph matching, we can define proper graph edit operations according to the noise exploration and design the edit cost function to advance the concert. For explanation expediency, we provide some notations and definitions taken from. Let L be a finite alphabet of labels for vertexes and edges.

Notation: A graph is a triple g = (V, α, β), where V is the finite set of vertexes, α: V → L is vertex labelling function, and β : E → L is edge labelling function. The set of edges E is implicitly given by assuming that graphs are fully connected, i.e., E = V × V. For the notational convenience, node and edge labels come from the same alphabet.

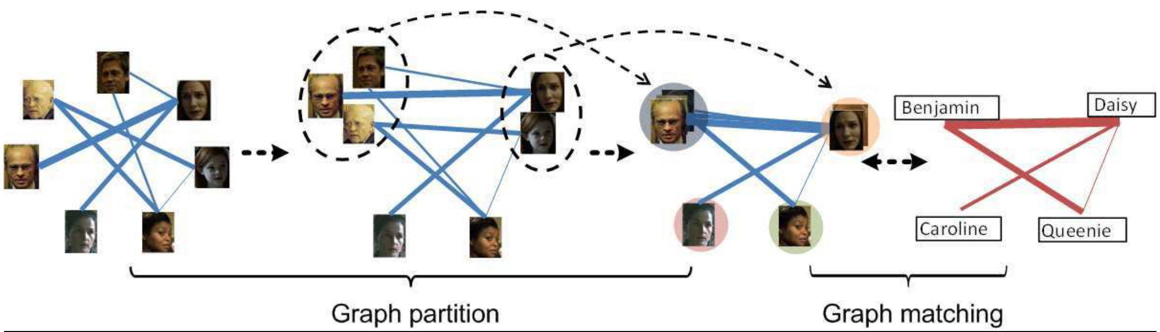

Fig.3. Simultaneously graph partition and matching for scheme 2

V. CONCLUSION

We have shown that the proposed two schemes for face identification are useful to improve the results for clustering and identification of the face tracks extract from unrestrained movie videos. From the kindliness scrutiny, we have also shown that to some degree, such schemes have better robustness to the noises in constructing similarity graphs than the traditional methods. A third conclusion is a principle for developing robust character identification method: intensity alike noises must be emphasizing more than the coverage alike noises. In the future, we will extend our work to investigate the optimal functions for different movie genre. Another goal of future work is to make the most of more character relationships, e.g., the sequential statistics for the speakers, to build affinity graphs and improve the robustness.

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] Y. Zhang, C. Xu, H. Lu, and Y. Huang, “Character identification in feature-length films using global face-name matching,” IEEE Trans.Multimedia, vol. 11, no. 7, pp. 1276–1288, November 2009.
[3] M. Everingham, J. Sivic, and A. Zissserman, “Taking the bite out of automated naming of characters in tv video,” in Jounal of Image and Vision Computing, 2009, pp. 545–559.
[4] C. Liang, C. Xu, J. Cheng, and H. Lu, “Tvparser: An automatic tv video parsing method,” in CVPR, 2011, pp. 3377–3384.
[5] J. Sang and C. Xu, “Character-based movie summarization,” in ACM MM, 2010.
[6] 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.
[7] T. Cour, B. Sapp, C. Jordan, and B. Taskar, “Learning from ambiguously labeled images,” in CVPR, 2009, pp. 919–926.
[8] J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, “Video-based face recognition on real-world data.” in ICCV, 2007, pp. 1–8.
[9] S. Satoh and T. Kanade, “Name-it: Association of face and name in video,” in Proceedings of CVPR, 1997, pp. 368–373.
[10] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y. W. Teh, E. G. Learned-Miller, and D. A. Forsyth, “Names and faces in the news,” in CVPR, 2004, pp. 848–854.
[11] J. Yang and A. Hauptmann, “Multiple instance learning for labelling faces in broadcasting news video,” in ACM Int. Conf. Multimedia, 2005, pp. 31–40.

[12] A. W. Fitzgibbon and A. Zisserman, “On affine invariant clustering and automatic cast listing in movies,” in ECCV (3), 2002, pp. 304–320.
[13] O. Arandjelovic and R. Cipolla, “Automatic cast listing in feature-length films with anisotropic manifold space,” in CVPR (2), 2006, pp. 1513– 1520.
[14] D. Ramanan, S. Baker, and S. Kakade, “Leveraging archival video for building face datasets,” in ICCV, 2007, pp. 1–8.
[15] M. Everingham and A. Zisserman, “Identifying individuals in video by combining ”generative” and discriminative head models,” in ICCV, 2005, pp. 1103–1110.

