ISSN:0975-9646

Implementation on Eccentric Identified from Picture using Graph Methodology

Pradnyesh J. Bhisikar, Amit sahu

Computer Science & Engg.
G.H.RaisoniCollege of Engg. And Mgmt.

ABSTARCT: **Rapidly face identification of eccentric in picture has drawn significant research interests and led to various applications. We investigate the problem of rapidly labelling appearances of eccentric in TV or picture material with their names. The contribution of this paper is two-fold: (1) we propose a generative model, named eccentric picture, to depict the temporal character correspondence between picture and script, from which face-name relationship can be automatically learned as a model parameter, and meanwhile, video scene structure can be effectively inferred as a hidden state sequence; (2) we find fast algorithms to accelerate both model parameter learning and state inference, resulting in an efficient and global optimal alignment. We first segment scenes in the movie by analysis and alignment of script and movie. Then we conduct sub story discovery and content attention analysis based on the scene analysis and character interaction features. Different from the existing methods that are categorized as static captioning, dynamic captioning puts scripts at suitable positions to help hearing impaired audience better recognize the speaking characters.**

Keywords: eccentric;TV;character;scripts;discovery.

I. INTRODUCTION

The development of video and TV provides large amount of digital video data. This has led to the requirement of efficient and effective techniques for video content understanding and organization. Automatic video annotation is one of such key techniques. In this paper our focus is on annotating characters in the video and TVs, which is called video eccentric identification [1]. The objective is to identify the faces of the characters in the video and label them with the corresponding names in the cast. The textual cues, like cast lists, scripts, subtitles and closed captions are usually exploited. In a video, person are the focus center of interests for the audience. Automatic character identification in videos is essential for semantic movie analysis such as movie indexing, summarization and retrieval. Character identification, though very intuitive to humans, is a tremendously challenging task in computer vision. In the paper second part is that video – the subtitles record what is said, but not by whom, whereas the script records who says what, but not when. However, by automatic matching of the two sources, it is possible to extract who says what and when. Knowledge that a person is speaking then gives a very weak cue that the person may be visible in the video. The flourishing TV industries have contributed to an explosive growth of video content. However, such increasing quantity has not yet been accompanied by an improvement in its accessibility. Huge amount of TV videos have become a burden of storage and management. Therefore, the automatic parsing or indexing approach is of great importance. Unfortunately, low-level visual information is still difficult to solve this problem due to the well-known semantic gap. In addition to effective exploitation of cues from textual annotation, success depends on robust computer vision methods for face processing in video. We propose extensions to our method for connecting faces in video [4], which provides robust face tracks, and a novel extension of the ‘‘pictorial structure” method [5] which gives reliable localization of facial features in presence of significant pose variations. This paper is an extended version of [1]. Character identification, though very intuitive to humans, is a tremendously challenging task in computer vision. The reason is four-fold: 1) Weakly supervised textual cues [7]. There are ambiguity problem in establishing the correspondence between names and faces: ambiguity can arise from a reaction shot where the person speaking may not be shown in the frames1; ambiguity can also arise in partially labelled frames when there are multiple speakers in the same scene2. 2) Face identification in videos is more difficult than that in images [8]. Low resolution, occlusion, nonrigid deformations, large motion, complex background and other uncontrolled conditions make the results of face detection and tracking unreliable. In movies, the situation is even worse. This brings inevitable noises to the character identification. 3) The same character appears quite differently during the movie [3]. There may be huge pose, expression and illumination variation, wearing, clothing, even makeup and hairstyle changes. Moreover, characters in some movies go through different age stages, e.g., from youth to the old age. Sometimes, there will even be different actors playing different ages of the same character. 4) The determination for the number of identical faces is not trivial [2]. Due to the remarkable intraclass variance, the same character name will correspond to faces of huge variant appearances. It will be unreasonable to set the number of identical faces just according to the number of characters in the cast. Our study is motivated by these challenges and aims to find solutions for a robust framework for movie character identification.

II. RELATED WORK

The crux of the character identification problem is to exploit the relations between videos and the associated texts in order to label the faces of characters with names.

Name-it [2] is the first name-face association system proposed for news video, which is based on the co-occurrence between the detected faces and names extracted from the transcript. A face is labeled with a name which frequently co-occurs with it. Yang et al. [3] employed the closed caption and speech transcript to build models predicting the probability that a name in the text matches to a face on the video frame. To cope with this problem, Everingham et al. [5] [6] proposed to align the film script and the subtitle to generate time stamped name annotation. Based on that, they learned character specific classifiers from video and extended the coverage of the method in [7]. Previous work on the recognition of characters in TV or movies has often ignored the availability of textual annotation. In the ‘‘cast list discovery” problem [3,6], faces are clustered by appearance, aiming to collect all faces of a particular character into a few pure clusters (ideally one), which must then be assigned a name manually. It remains a challenging task to obtain a small number of clusters per character without merging multiple characters into a single cluster. The combination of face detection and text has also been applied previously to face recognition in video. In [8], transcripts (spoken text without the identity of the speaker) and video of news footage were combined to recognize faces. Much attention was directed at how to predict from a name appearing in the transcript (typically spoken by a news anchor-person) when (relatively) the person referred to might appear in the video; addition of a standard face recognition method to this information gave small improvements in accuracy.

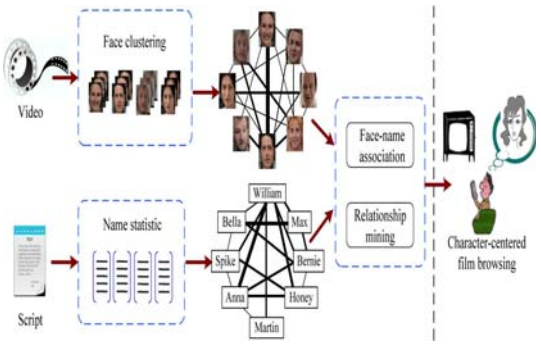

Fig. 1. Framework of strategy 2: Face-name graph matching without #cluster pre-specified

1. Overview of our approach

In an eccentric video, the interacting among the characters resembles them into a relationship network, which makes a video be treated as a small society [2]. In this paper, we propose a global face-name graph matching based framework for robust movie character identification. Two strategies are considered. There are connections as well as differences between them every character has his/her social position and keeps a certain relationship with others. In the video, faces can stand for characters and the co-occurrence of the faces in a scene can represent an interaction between characters. Hence, the statistical properties of faces can preserve the mutual relationship in the character network. Regarding the connections, firstly, the proposed two strategies 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. As the same way in the film script, the spoken lines of different characters appearing in the same scene also represents an interaction. Thus, the names in front of the spoken lines can also build a name affinity network. Both the statistical properties of the faces and the names motivate us to seek a correspondence between the face affinity network and the name affinity network. The name affinity network can be straightforwardly built from the script. For the face affinity network, we first detect face tracks in the video and cluster them into groups corresponding to the characters.

2. Robust Character Identification from Eccentric Video

In this section first review on previous work on character identification by global name-face graph matching. Based on investigations of the noises generated during the affinity graph construction process, we construct the name and face affinity graph in rank ordinal level and employ *ECGM* with specially designed edit cost function for name face match. Then in this model discuss on following techniques.

2.1. *Proficiencies*

2.1.1. Category 1: Cast list discovery based method:

These methods will only make use of the cast 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. 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.

2.1.2. 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 neighbor 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.

2.1.3. 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.

2.2. *Strategies*

In this paper two strategies are considered. There are both similarities and dissimilarities between them. Regarding similarities, the proposed both strategies 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, strategies 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 strategies 2. Strategies 2 is said to be as the extension of strategies 1. Because strategies 2 have an additional module of graph partition compared with strategy 1.

2.2.1. Strategy 1:

The proposed framework for strategy 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.

2.2.2. Strategy 2:

Strategy 2 has two differences from strategy 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 strategy 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 strategy 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 sub-graph) 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.

3. ECGM-BASE METHOD

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 [2].

Notation: A graph is a triple g = (V, α, β), where V is the finite set of vertexes, α: V → L is vertex labeling function, and β: E → L is edge labeling 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.

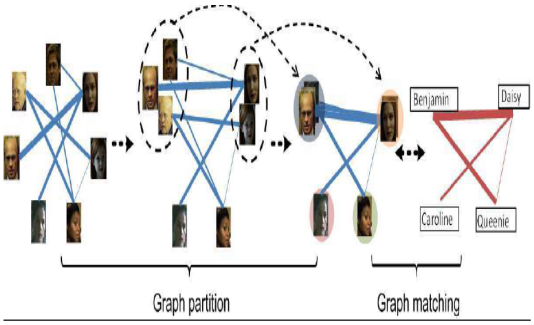

Fig.2. Simultaneously graph partition and matching for strategy 2.

4. 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 eccentric 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 eccentric identification method: intensity a like noise must be emphasizing more than the coverage a like noise. 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 multiple character identification and if there is one person play different role then identified that person i.e. using makeup identified person.

REFERENCES

[1] J. Sang, C. Liang, C. Xu, and J. Cheng, "Robust movie character identification and the sensitivity analysis," in *Proc. ICME*, 2011, pp. 1–6.

[2] C. Liang,C.Xu, J. Cheng, andH.Lu, "Tvparser: An automatic tv video parsing method," in *Proc. Comput. Vis. Pattern Recognit.*, 2011, pp.3377–3384.

[3] R. G. Cinbis, J. Verbeek, and C. Schmid, "Unsupervised metric learning for face identification in TV video," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 1559–1566.

[4] J. Sang and C. Xu, "Character-based movie summarization," in *Proc. ACM Int. Conf. Multimedia*, 2010, pp. 855–858.

[5] R. Hong, M. Wang, M. Xu, S. Yan, and T.-S. Chua, "Dynamic captioning: Video accessibility enhancement for hearing impairment,"*ACM Trans. Multimedia*, pp. 421–430, 2010.

[6] M. Xu, X. Yuan, J. Shen, and S. Yan, "Cast2face: Character identification in movie with actor-character correspondence," *ACM Multimedia*, pp. 831–834, 2010.

[7] 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, Nov. 2009.

[8] T. Cour, B. Sapp, C. Jordan, and B. Taskar, "Learning from ambiguously labeled images," in *Proc. Comput. Vis. Pattern. Recognit.*, 2009,pp. 919–926.

[9] D. Ramanan, S. Baker, and S. Kakade, "Leveraging archival video for building face datasets," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[10] J. Sivic, M. Everingham, and A. Zissserman, "Who are you?—Learning person specific classifiers from video," in *Proc. Comput. Vis. Pattern Recognit.*, 2009, pp. 1145–1152.