

## Character Identification in Movie Using Graph Matching Algorithm

T. R. Mageswari, M.Tech CSE Iyr, Mrs.Murunya M.C.A., M.E.,  
PRIST University, Kumbakonam Campus.  
HodOf CSE Dept, PRIST University, Kumbakonam Campus.

**Abstract:** The character Identification in films are a challenging problem due to the huge variation in the appearance of each character. It drawn significant research interests and had a many interesting applications. In this paper, we investigate the problem of identifying characters using graph matching algorithm to get the most accurate identification. The contributions of our work include: 1) The character relationship representation with a noise insensitive 2) Use an edit operation based graph matching algorithm. 3) Graph partition and graph matching for complex character changes by handle simultaneously. 4) The existing character identification approaches, we can perform an in-depth sensitivity analysis by introducing two types of simulated noises. The proposed framework is able to create a new experience on character identification in movie to demonstrate state-of-the-art performance in various genres of movies.

### 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 analyzing and organization. Automatic video annotation is one of such key techniques. We focus is on annotating characters in the movie and TVs, which is called movie character identification. The objective is to identify the faces of the characters in the video and label them with their names in the cast. The cast lists, scripts, subtitles and closed captions are usually exploited are known as textual cues. Fig.1 shows an example in our identification.. The characters are the focus center of interests for the audience in movies. These provide lots of clues about the movie structure and content. The automatic character identification is essential for semantic movie index and retrieval , scene segmentation , summarization and other applications .

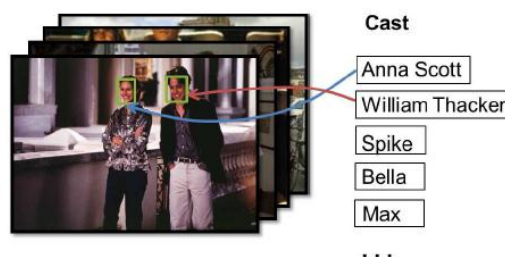


Fig. 1. Examples of character identification from movie "Notting Hill".

The character identification is a tremendously challenging task in computer vision. The reasons are following: 1) The textual cues are weakly supervised. The ambiguity problem in establishing the correspondence between names and faces: the ambiguity can arise from a reaction shot where the person speaking may not be shown in the frames and in partially labeled frames when there are multiple speakers in the same scene. 2) The face identification in videos is more difficult than that in images. Low resolution, occlusion, non rigid deformations, large motion, complex background and other uncontrolled conditions make the results of face detection and tracking unreliable. This brings inevitable noises to the character identification. 3) The characters appears quite differently during the movie. There may be huge pose, expression and illumination variation, wearing, clothing, even makeup and hairstyle changes. The characters in some movies go through different age stages, e.g., from youth to the old age. 4) The 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 just according to the number of characters in the cast. This paper motivated by these challenges and aims to find solutions for accurate movie character identification.

The usage of the utilized textual cues, we can divide the existing movie character identification methods into three categories.

**1.Cast list based:** In these methods we utilize the cast list textual resource. In this category, faces are clustered by appearance and faces of a particular character are expected to be collected in a few pure clusters. The character names in the cast are used as queries to search face images and constitute gallery set. The probe face tracks in the movie are then identified as one of the characters by multi-task joint sparse representation and classification. The metric learning is introduced into character identification in uncontrolled videos recently. The clustering as well as identification performance are demonstrated to be improved. These cast list based methods are easy for using the understanding and implementation of the identification.

**2.Subtitle or Closed caption, Local matching based:** Subtitle and closed caption provide time-stamped dialogues, which can be exploited for alignment to the video frames. The rest of the faces were then classified into these exemplars for identification. Time-stamped name annotation and face exemplars are generated. They further extended their work by replacing the nearest neighbor classifier by multiple kernel learning for features combination. In the new framework, non-frontal faces are handled and the coverage is extended. The local matching based methods require the time-stamped information, which is either extracted by OCR (i.e., subtitle) or unavailable for the majority of movies and TV series.

**3.Script/Screenplay, Global matching based:** Global matching based methods open the possibility of character identification without OCR-based subtitle or closed caption. This is not easy to get local name cues, the task of character identification is formulated as a global matching problem. In the movies, the names of characters directly appear in the subtitle, while the movie script which contains character names has no time information. If the local time information is not present, the task of character identification is formulated as a global matching problem between the faces detected from the video and the names extracted from the movie script. The comparison of the local matching, global statistics are used for name-face association, which enhances the robustness of the algorithms.

**The proposed work differs from the existing research in following fold:**

- The characters may show various appearances, the representation of character is often affected by the noise introduced by face tracking, face clustering and scene segmentation. The extensive research efforts have been concentrated on character identification and many applications have been proposed, little work has focused on improving the robustness. Here we observed in our investigations that some statistic properties are preserved in spite of these noises.
- The face track clustering serves as an important step in movie character identification. In the proposed work we lose of one face cluster corresponding to one character name. Face track clustering and face-name matching are jointly optimized and conducted in a unique framework.
- In the movie character identification, sensitivity analysis offers valid tools for characterizing the robustness to noises for a model. In this work, our aim to fill this gap by introducing two types of simulated noises.

**Overview of Our Approach**

Our paper approach is to propose a global face-name graph matching based framework for movie character identification. Here we considered two schemes. One is connections and another is differences between them. In the regarding the connections, the proposed two schemes both belong to the global matching based category, where external script resources are utilized. In the improvement of the robustness, the ordinal graph is employed for face and name graph representation and we introduce a graph matching algorithm called Error Correcting Graph Matching (ECGM) is introduced. With the regarding the differences, scheme 1 sets the number of clusters when performing face clustering (e.g., K-means, spectral clustering). The face graph is restricted to have identical number of vertices with the name graph. While, in scheme 2, no cluster number is required and face tracks are clustered based on their intrinsic data structure (e.g., mean shift, affinity propagation). Moreover, as shown in Fig.2 and Fig.3, scheme 2 has an additional module of graph partition compared with scheme 1.

**1) Scheme 1:** 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. In ECGM, the difference between two graphs is measured by edit distance which is a sequence of graph edit operations. The optimal match is achieved with the least edit distance.

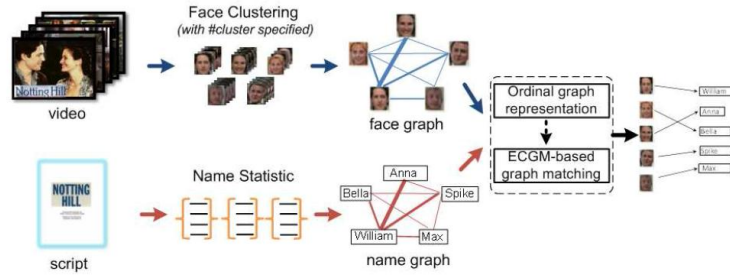


Fig. 2. Framework of scheme 1: Face-name graph matching with #cluster pre-specified.

2) **Scheme 2:** This is shown in the following figure 3. It has two differences from scheme 1 in Fig. 2. One is no cluster number is required for the face tracks clustering step and another is the face graph and name graph may have different number of vertices, a graph partition component is added before ordinal graph representation.

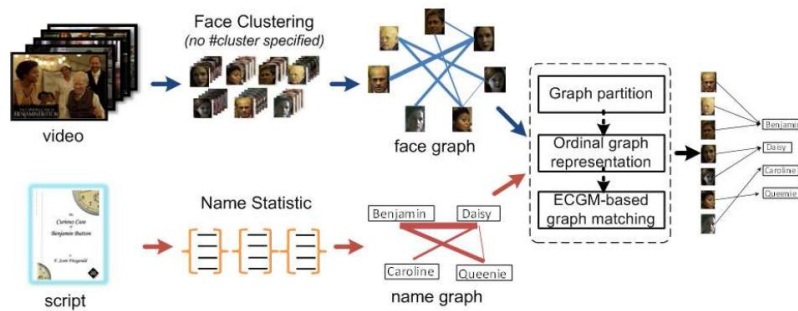


Fig. 3. Framework of scheme 2: Face-name graph matching without #cluster pre-specified.

**ECGM-based Graph Matching**

ECGM is a powerful tool for graph matching. The measurement of the similarity of two graphs, graph edit operations are defined, such as the deletion, insertion and substitution of vertices and edges. Each of these operations is further assigned a certain cost. The costs are application dependent and usually reflect the likelihood of graph distortions. The more likely a certain distortion is to occur, the smaller is its cost. Through error correcting graph matching, we can define appropriate graph edit operations according to the noise investigation and design the edit cost function to improve the performance. For explanation convenience, we provide some notations and definitions taken from Let  $L$  be a finite alphabet of labels for vertices and edges.

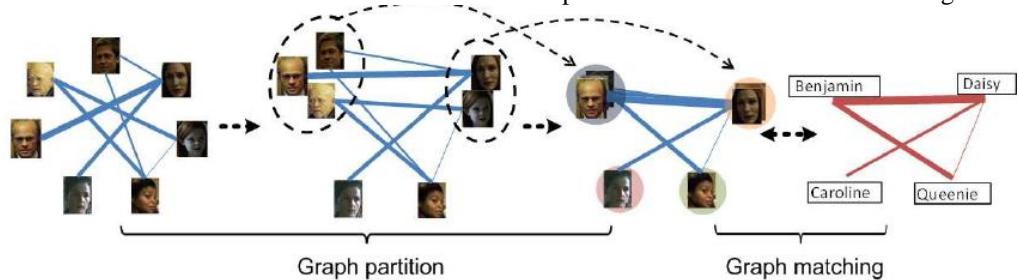
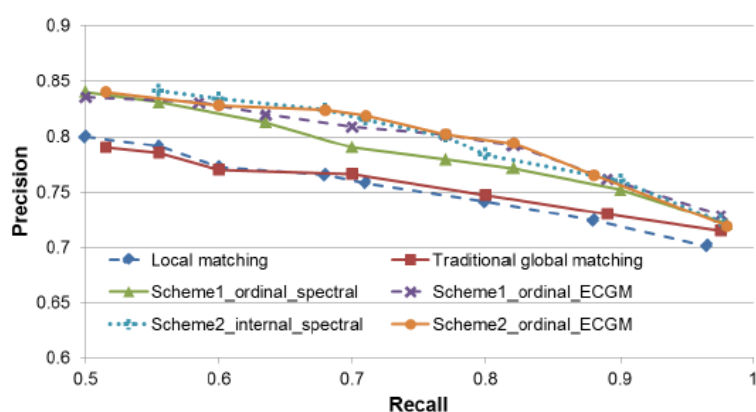


Fig. 4. Simultaneously graph partition and matching for scheme 2.

**Notation:** A graph is a triple  $g = (V, \_1, \_2)$ , where  $V$  is the finite set of vertices,  $\_1 : V \rightarrow L$  is vertex labeling function, and  $\_2 : E \rightarrow L$  is edge labeling function. The set of edges  $E$  is implicitly given by assuming that graphs are fully connected, i.e.,  $E = V \times V$ . For the notational convenience, node and edge labels come from the same alphabet  $L$ .

**Definition 1.** Let  $g_1 = (V_1, \_1, \_2)$  and  $g_2 = (V_2, \_3, \_4)$  be two graphs. An ECGM from  $g_1$  to  $g_2$  is a bijective function  $f : \hat{V}_1 \rightarrow \hat{V}_2$ , where  $\hat{V}_1 \subseteq V_1$  and  $\hat{V}_2 \subseteq V_2$ . We say that vertex  $x \in \hat{V}_1$  is substituted by vertex  $y \in \hat{V}_2$

if  $f(x) = y$ . If  $\_1(x) = \_3(f(x))$ , the substitution is called an identical substitution. The cost of identical vertex or edge substitution is usually assumed to be zero, while the cost of any other edit operation is greater than zero.



## II. Conclusion

The proposed schemes are useful to improve results for clustering and identification of the face tracks extracted from uncontrolled movie videos. With the usage of the sensitivity analysis, we have also shown that to some degree, such schemes have better robustness to the noises in constructing affinity graphs than the traditional methods. The another conclusion is a principle for developing robust character identification method: intensity alike noises must be emphasized more than the coverage alike noises. In the future, we will extend our work to investigate the optimal functions for different movie genres. Another goal of future work is to exploit more character relationships, e.g., thesequential 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. Zisserman, "Taking the bite out of automated naming of characters in tv video," in *Journal of Image and Vision Computing*, 2009, pp. 545–559.
- [4] C. Liang, C. Xu, J. Cheng, and H. Lu, "Tvparsr: 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 labeling 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.