

# The Process of Face Detection and Recognition in Multi-videos

**G.Bakkiyaraj**

Department of CSE  
Roever Engineering College  
Perambalur, India  
bakkiya10@gmail.com

**V.Dinesh**

Department of CSE  
Roever Engineering College  
Perambalur, India  
veldinesh3909@gmail.com

## Abstract

*The future of the world depends upon latest technologies and innovative ideas. The innovative concepts are helps to upgrade our knowledge. In face detection of videos captured in unconstrained real-world scenarios is an interesting problem with many potential applications. This paper presents an efficient approach to face detection, in which the intensity differences between pixels in the grayscale face images are used as features. In this paper we present two schemes of global face-name matching based framework for robust character identification. The contributions of this work include: A noise insensitive character relationship representation is incorporated. We introduce an edit operation based graph matching algorithm. Complex character changes are handled by simultaneously graph partition and graph matching. Beyond existing character identification approaches, we further perform an in depth sensitivity analysis by introducing two types of simulated noises. The proposed schemes demonstrate state-of-the-art performance on movie character identification in various genres of movies.*

## Keywords:

*Image identification; graph matching; graph partition; gray Scale; graph edit.*

## I. INTRODUCTION

### A. Objective and Motivation

The proliferation of movie 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 movie and TVs, which is called movie character. The objective is to identify the faces of the label marks in the technical faces to compare to multiple detection of images in videos. In analysis to train the images in system database can be stored and retrieved.

In Characters in the video and label them with the names in the cast faced in image. The textual cues, like cast lists, scripts, subtitles and closed captions are usually exploited. Fig.1 shows an example in our experiments. In a movie, characters are the focus center of interests for the audience. Their occurrences provide Lots of clues about the movie structure and content. Automatic character identification is essential for semantic movie index and retrieval, scene segmentation, summarization and other applications.

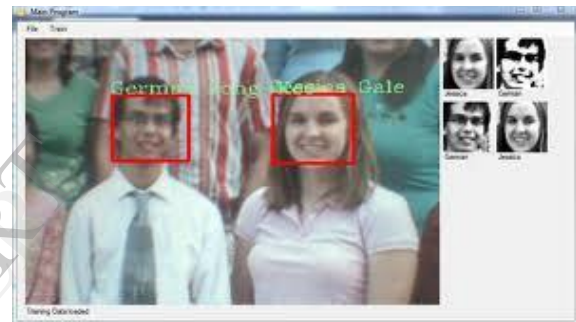


Fig. 1. Examples of character identification from "Video".

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.
- 2) Face identification in videos is more difficult than that in images.
- 3) The same character appears quite differently during the movie.
- 4) The determination for the number of identical faces is not trivial.

Our study is motivated by these challenges and aims to find solutions for a robust framework for movie character identification the faces detected from the video and the names extracted from the movie script.

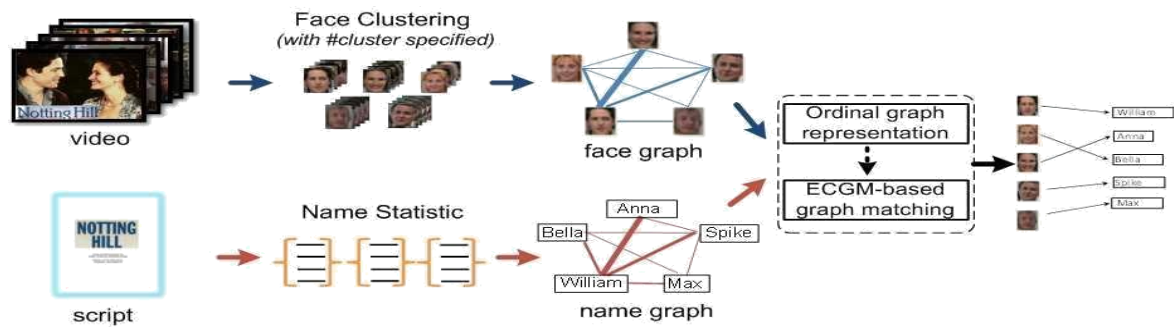


Fig. 2 Framework of scheme 1: Face-name graph matching with #cluster pre specified.

## B. Related Work

The crux of the character identification problem is to exploit the relations between videos and the associated texts in order

- 1) The name in the subtitle/closed caption finds no Corresponding faces in the video.
- 2) Multiple names in the subtitle/closed caption correspond to multiple faces in the video.

### 1) Category 1: Cast list based:

These methods only utilize the case list textual resource. 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 character names in the cast are used as queries to search face image.

The probe face tracks in the movie are then identified as one of the characters by multi-task joint sparse representation and classification. Recently, metric learning is introduced into character identification in uncontrolled videos. Cast-specific metrics are adapted to the people appearing in a particular video in an unsupervised manner. The clustering as well as identification performance is demonstrated to be improved. These cast list based methods are easy for understanding and implementation. However, in this scene is created and detected. To the best of our knowledge, there have been no efforts directed at the sensitivity analysis for movie character identification. the subtitle, while the movie script which contains character names has no time information. Without the local time information.

### 2) Category 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. 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 (i.e., subtitle) or unavailable for the majority of movies and TV series (i.e., closed caption). Besides, the ambiguous and partial annotation makes local matching based methods more sensitive to the face detection and tracking noises.

### 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 easy to get local name cues, the task of character identification is formulated as a global matching problem in. In movies, the names of characters seldom 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 global matching problem. For name-face graph matching, we utilize the ECGM algorithm. 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 the faces detected from the video. and tracking unreliable. In movies, the situation is even worse. This brings inevitable noises to the character identification. The same character appears quite differently during the movie.

In global statistics are used for name-face association, which enhances the robustness of the algorithms. Our work differs from the existing research in three-fold:

- Regarding the fact that characters may show various appearances, the representation of character is often affected by the noise introduced by face tracking, face clustering and scene segmentation. Although extensive research efforts have been concentrated on character identification and many applications have been proposed, little work has focused on improving the robustness. We have observed in our investigations that some statistic properties are preserved in spite of these noises. Based on that, we propose a novel representation for character relationship and introduce a name-face matching method which can accommodate a certain noise.
- Face track clustering serves as an important step in movie character identification. In most of the existing methods, some cues are utilized to determine the number of target clusters prior to face clustering, e.g., in, the number of clusters is the same as the number of distinct speakers appearing in the script. While this seems convinced at first glance, it is rigid and even deteriorating the clustering results sometimes. In this paper, we lose the restriction of one face cluster corresponding to one character name.
- Sensitivity analysis is common in financial applications, risk analysis, signal processing and any area where models are developed. Good modeling practice requires that the modeler provides an evaluation of the confidence in the model, for example, assessing the uncertainties associated with the modeling process and with the outcome of the model itself. For movie character identification, sensitivity analysis offers valid tool for Characterizing the robustness has been no efforts directed character identification.

### C. Overview of Our Approach

In this paper, we propose a global face-name graph match-in based framework for robust movie character identification. Two schemes are considered. There are connections as well as differences between them. Regarding the connections, firstly, the proposed two schemes 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.

**Scheme 1:** The proposed framework for scheme 1 is shown in Fig.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. For name-face graph matching, we utilize the ECGM algorithm. In ECGM, the difference between two graphs is measured by edit distance which is a sequence of graph edit operations.

**Scheme 2:** The proposed framework for scheme 2 is shown in Fig.3. It has two differences from scheme 1 in Fig.2. 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. In this paper, we aim to fill this gap by introducing two types of simulated noises. 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. There may be huge pose, expression and illumination variation, wearing, clothing, even makeup and hairstyle changes 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. The determination for the number of identical faces is not trivial. Due to the remarkable infraclass variance, the same character name will correspond to faces of huge variant appearances. Closer they are, and the larger the edge weights between them are. In this sense, a name affinity graph from script analysis and a face affinity graph from video analysis uncertainties associated with the modeling process and with the outcome of the model itself. and illumination variation, wearing, clothing, even makeup and hairstyle changes characters in some movies.

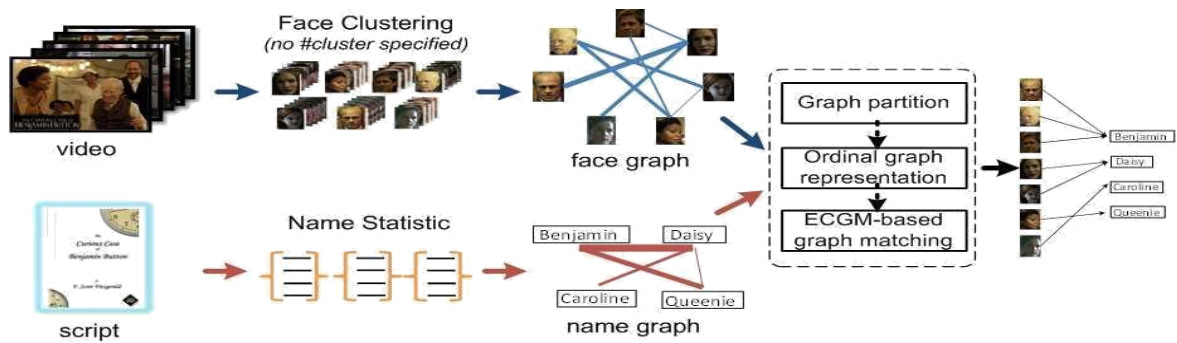


Fig.3 Framework of scheme 2: Face-name graph matching without #cluster pre-specified.

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.

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.

In general, the scheme 2 has two advantages over the scheme 1.

- For scheme 2, no cluster number is required in advance and face tracks are clustered based on their intrinsic data structure.
- Regarding that movie cast cannot include pedestrians whose face is detected and added into the face track, restricting the number of face tracks clusters the same as that of name from movie cast will deteriorate the clustering process.

## II. SCHEME 1: FACE-NAME GRAPH MATCHING WITH NUMBER OF CLUSTER SPECIFIED

In this section we first briefly review the framework of traditional global graph matching based character identification. 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 face-name matching. In this sense, a name affinity graph from script analysis and evaluation of the confidence in the model, for example, assessing the uncertainties faces and images are identified the movie character. The more scenes where two characters appear together, the closer they are, and the larger the edge weights outcome of the model itself for movie character identification.

### A. Review of Global Face-Name Matching Framework

In a movie, the interactions among characters resemble them into a relationship network. Co-occurrence of names in script and faces in videos can represent such interactions. Affinity graph is built according to the co-occurrence status among characters, which can be represented as a weighted graph  $G = \{V, E\}$  where vertex  $V$  denotes the characters and edge  $E$  denotes relationships among them. The more scenes where two characters appear together, the closer they are, and the larger the edge weights between them are. In this sense, a name affinity graph from script analysis and a face affinity graph from video analysis can be constructed. Fig.4 demonstrates the adjacency matrices corresponding to the name and face affinity graphs from the movie "Noting Hill"<sup>3</sup>. All the affinity values are normalized into the interval  $[0, 1]$ . We can see that some of the face affinity values differ much from the corresponding name affinity values (e.g.  $\{WIL, SPI\}$  and  $\{Face1, Face2\}$ ,  $\{WIL, BEL\}$  and  $\{Face1, Face5\}$ ) due to the introduced noises. Subsequently, character identification is formulated as the problem of finding optimal vertex to vertex matching between two graphs. A spectral graph matching algorithm is applied to find the optimal name-face correspondence.

### B. Ordinal Graph Representation

The name affinity graph and face affinity graph are built based on the co-occurrence relationship. Due to the imperfect face detection and tracking results, the face affinity graph can be seen as a transform from the name affinity graph by affixing noises. We have observed in our investigations that, in the generated affinity matrix some statistic proper ties of the characters are relatively stable and insensitive to the noises, such as character A has more affinities with character B than C, character D has never changed.



we assume that while the absolute quantitative affinity values are changeable, the relative affinity quantitative affinity values are changeable, the relative affinity Relationships between characters (e.g. A is closer to B than to C) and the qualitative affinity values (e.g. whether D has co-occurred with A) usually remain unchanged. In this paper, we utilize the preserved statistic properties and propose

To represent the character co-occurrence in rank order. We denote the original affinity matrix as  $R = \{r_{ij}\} N \times N$ , where  $N$  is the number of characters. First we look at the cells along the main diagonal (e.g. A co-occur with A, B co-occur with B). We rank the diagonal values  $r_{ii}$  in ascending order, then the ordinal affinity matrix  $R$ :

$$\tilde{r}_{ii} = I_{r_{ii}} \quad (1)$$

Where  $I_{r_{ii}}$  is the rank index of original diagonal affinity value  $r_{ii}$ . Zero-cell represents that no co-occurrence relationship is specially considered, which a qualitative measure is. From the perspective of graph analysis, there is no edge between the vertexes of row and column for the zero-cell. Therefore, change of zero-cell involves with changing the graph structure or topology. To distinguish the zero-cell change, for each row in the original affinity matrix, we remain the zero-cell unchanged. The number of zero-cells in the  $i^{th}$  row is recorded as  $null_i$ . Other than the diagonal cell and zero-cell, we sort the rest affinity values in ascending order, i.e., for the  $i^{th}$  row, the corresponding cells  $\tilde{r}_{ij}$  in the  $i^{th}$  row of ordinal affinity matrix:

$$\tilde{r}_{ij} = I_{r_{ij}} + null_i \quad (2)$$

Where  $I_{r_{ij}}$  denotes the order of  $r_{ij}$ . Note that the zero-cells are not considered in sorting, but the number of zero-cells will be set as the initial rank order<sup>4</sup>. The ordinal matrix is not necessarily symmetric. The scales reflect variances in degree of intensity, but not necessarily equal differer. We illustrate in Fig.5 corresponding to the affinity matrices in Fig. 4. It is shown that although there are major differences between original name and face affinity matrices, the derived ordinal affinity matrices are basically the same. The differences are generated due to the changes of zero-cell. A rough conclusion is that the ordinal affinity matrix is less sensitive to the noises than the original affinity matrix. We will further validate the advantage of original graph representation in the experiment section.

### C. ECGM-based Graph Matching

ECGM is a powerful tool for graph matching with distorted inputs. It has various applications in pattern recognition and computer vision. In order to measure the similarity of two graphs, graph edit operations

are defined, such as the deletion, insertion and substitution of vertexes and edges. Each of these operations is further assigned a certain cost. The costs are application dependent and usually reflect the livelihood 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 vertexes and edges.

**Notation:** A graph is a triple  $g = (V, \_, \_)$ , where  $V$  is the finite set of vertexes,  $\alpha : V \rightarrow L$  is vertex labeling function, and  $\beta : E \rightarrow L$  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.

### III. SCHEME 2: FACE-NAME GRAPH MATCHING WITHOUT NUMBER OF CLUSTER SPECIFIED

In Scheme 2 requires no specification for the face cluster number. Standard affinity propagation [29] is utilized for face tracks clustering. The similarity input  $s(i, k)$  is set as the Earth Mover's Distance (EMD) [30] between face tracks. All face tracks are equally suitable as exemplars and the preferences  $s(k, k)$ . There are two kinds of messages, "availability" and "responsibility", changed between face tracks. With "availability"  $a(i, k)$  initialized to be zero, the "responsibilities"  $r(i, k)$  are computed and updated using the rule. The message-passing procedure converges when the local decisions remain constant for certain number of iterations. In our case, high cluster purity with large number of clusters is encouraged. Since no restriction is set on the one-to-one face-name correspondence, the graph matching method expected to cope with the situations where several face clusters correspond to the character name. In view of this, a graph partition step is conducted before graph matching. Traditional graph partition aims at dividing a graph into disjoint sub graphs of the same size. Therefore, the scheme 2 provides certain robustness to the infraclass variance, which is very common in movies where characters change appearance significantly or go through a long time period. (b) Regarding that movie cast cannot include pedestrians whose face is detected and added into the face track, restricting the number of face tracks clusters the same as that of name from movie cast will deteriorate the clustering process.

In this paper, graph partition is only used to denote the process of dividing original face graphs. Instead of separately performing graph partition and graph matching, and using the partitioned face graph as input for graph matching graph partition and graph matching are optimized in a unique framework.

#### IV. CONCLUSIONS

We have shown that the proposed two schemes are useful to improve results for clustering and identification of the face tracks extracted from uncontrolled movie videos. From 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. A third conclusion is a principle for developing robust character identification method: intensity like noises must be emphasized more than the coverage alike noises.

#### V. FUTURE WORKS

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., 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 2013.
- [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, "TV parser: 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*, 2013.
- [6] R. Hong, M. Wang, M. Xu, S. Yan, and T.-S. Chua, "Dynamic captioning: video accessibility enhancement for hearing impairment," in *ACM Multimedia*, 2012, pp. 421–430.
- [7] T. Cour, B. Sapp, C. Jordan, and B. Taskar, "Learning from ambiguously labeled images," in *CVPR*, 2012, 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] S. Satoh and T. Kanade, "Name-it: Association of face and name in video," in *Proceedings of CVPR*, 2007, pp. 368–373.
- [11] 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*, 2012, pp. 848–854.
- [12] J. Yang and A. Hauptmann, "Multiple instances learning for labeling faces in broadcasting newsvideo," in *ACM Int. Conf. Multimedia*, 2013, pp. 31–40.