# ROBUST FACE-NAME GRAPH MATCHING FOR MOVIE CHARACTER IDENTIFICATION

Mamta Tukaram Gavhale
M.Tech , Student,
PRRM Engineering College.

Ashraf Shaik
Ph.D , HOD Of CSE Dept,
PRRM Engineering College .

## Abstract:

Automatic face identification of characters in movies has drawn significant research interests and led to many interesting applications. It is a challenging problem due to the huge variation in the appearance of each character. Although existing methods demonstrate promising results in clean environment, the performances are limited in complex movie scenes due to the noises generated during the face tracking and face clustering process. In this paper we present two schemes of global face-name matching based framework for robust character identification.

The contributions of this work 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 partition andgraph matching. 4) 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.

## Index Terms:

Character identification, graph matching, graph partition, graph edit, sensitivity analysis.

## 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 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. 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 [2], [3], scene segmentation [4], summarization [5] and other applications [6].

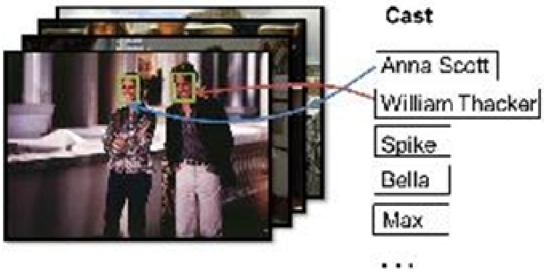


Fig. 1. Examples of character identification from movie "Not ting Hill".

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 frames 1; ambiguity can also arise in partially labeled frames when there are multiple speakers in the same scene 2. 2) Face identification in videos is more difficult than that in images [8]. Low resolution, occlusion, non-rigid 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 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. Our study is motivated by these challenges and aims to find solutions for a robust framework for movie character identification.

### B. Related Work:

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

1I.e., the name in the subtitle/closed caption finds no corresponding faces in the video.

2I.e., multiple names in the subtitle/closed caption correspond to multiple faces in the video.

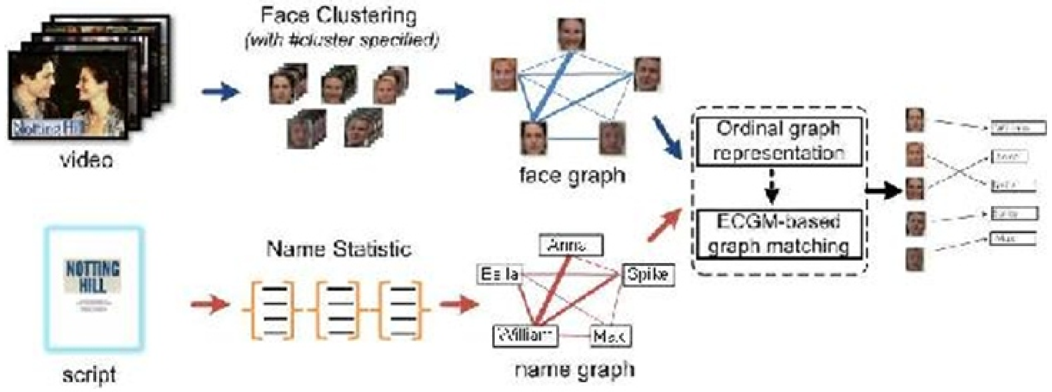


Fig. 2. Framework of scheme 1: Face-name graph matching with #cluster pre-specified.

to label the faces of characters with names. It has similarities to identifying faces in news videos [9], [10], [11]. However, in news videos, candidate names for the faces are available from the simultaneously appearing captions or local transcripts. While in TV and movies, the names of characters are seldom directly shown in the subtitle or closed caption, and scrip- t/screenplay containing character names has no time stamps to align to the video. According to the utilized textual cues, we roughly divide the existing movie character identification methods into three categories.

**1) Category 1:** Cast list based: These methods only utilize the case list textual resource. In the “cast list discovery” problem [2], [3], 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. Ramananet al. proposed to manually label an initial set of face clusters and further cluster the rest face instances based on clothing within scenes [4]. In [15], the authors have addressed the problem of finding particular characters by building a model/classifier of the character’s appearance from user-provided training data. An interesting work combining character identification with web image retrieval is proposed in [7].

**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. Everingham et al. [8], [3] proposed to combine the film script with the subtitle for local face-name matching. Time-stamped name annotation and face exemplars are generated. (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 [2], [22], [4]. Our method belongs to this category and can be considered as an extension to Zhang’s work [2]. 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 between the faces detected from the video and the names extracted from the movie script. Compared with local matching, global statistics are used for name-face association, which enhances the robustness of the algorithms.

Our work differs from the existing research in threefold:

- Regarding the fact that characters may show various appearances, the representation of character is often affected

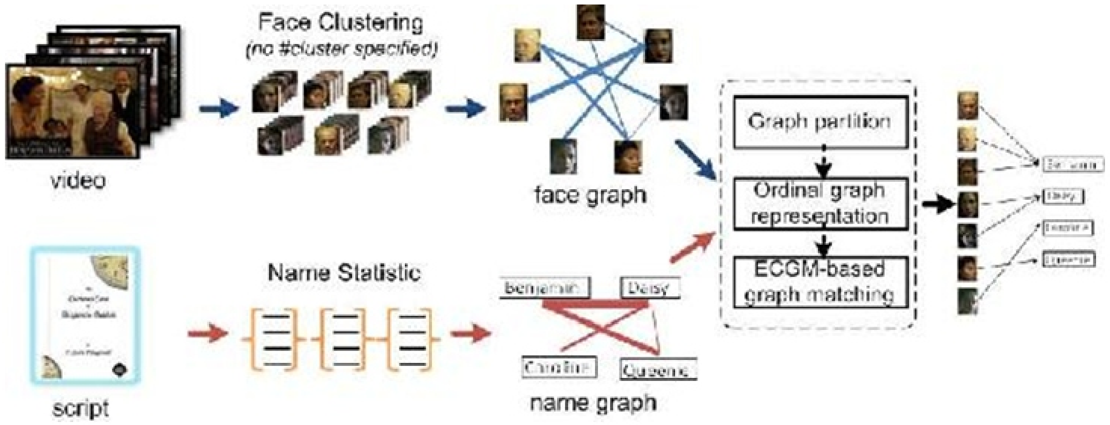


Fig. 3. Framework of scheme 2: Face-name graph matching without #cluster pre-specified.

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.

### C. Overview of Our Approach:

In this paper, we propose a global face-name graph matching 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. 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 vertexes 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. From this perspective, scheme 2 can be seen as an extension to scheme 1.

**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 [25], 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. 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 with the least edit distance. According to the noise analysis, we define appropriate graph edit operations and adapt the distance functions to obtain improved name-face matching performance.

## 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.

### 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.

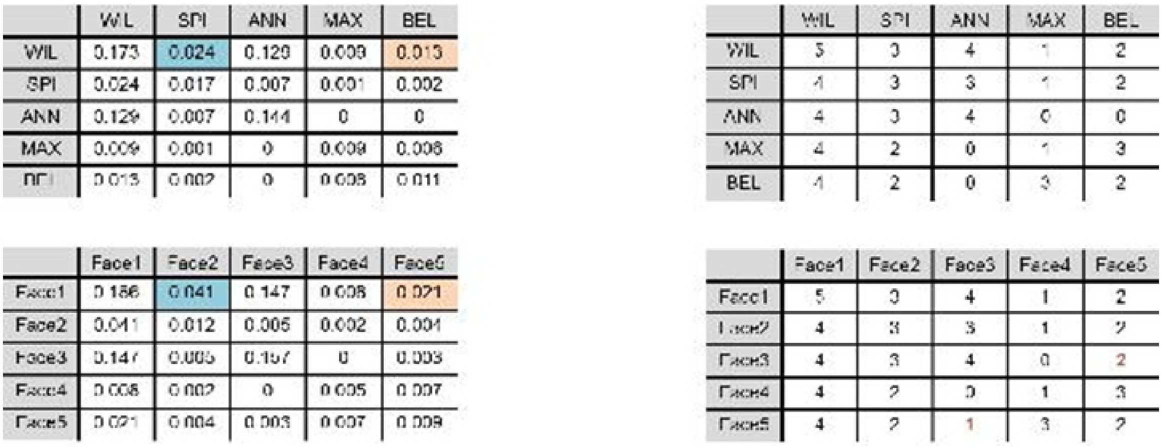
The more scenes where two characters appear together, the closer they are, and the larger the edge weights between them are. In this sense, a name affinity graph from script analysis and a face affinity graph from video analysis can be constructed. Fig.4 demonstrates the adjacency matrices corresponding to the name and face affinity graphs from the movie ”Noting Hill” 3.

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. More technical details can be referred to [2].

### 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 co-occurred with character A, etc. Delighted from this, we assume that while the absolute.

3 The ground-truth mapping is WIL-Face1, SPI-Face2, ANN-Face3, MAX-Face4, BEL-Face5



|  | WIL | SPI | ANN | MAX | BEL |
|---|---|---|---|---|---|
| WIL | 0.173 | 0.024 | 0.129 | 0.009 | 0.013 |
| SPI | 0.024 | 0.017 | 0.007 | 0.001 | 0.002 |
| ANN | 0.129 | 0.007 | 0.144 | 0 | 0 |
| MAX | 0.009 | 0.001 | 0 | 0.009 | 0.008 |
| BEL | 0.013 | 0.002 | 0 | 0.008 | 0.011 |

|  | Face1 | Face2 | Face3 | Face4 | Face5 |
|---|---|---|---|---|---|
| Face1 | 0.186 | 0.041 | 0.147 | 0.008 | 0.021 |
| Face2 | 0.041 | 0.012 | 0.005 | 0.002 | 0.004 |
| Face3 | 0.147 | 0.005 | 0.157 | 0 | 0.003 |
| Face4 | 0.008 | 0.002 | 0 | 0.005 | 0.007 |
| Face5 | 0.021 | 0.004 | 0.003 | 0.007 | 0.009 |

|  | WIL | SPI | ANN | MAX | BEL |
|---|---|---|---|---|---|
| WIL | 5 | 3 | 4 | 1 | 2 |
| SPI | 4 | 3 | 3 | 1 | 2 |
| ANN | 4 | 3 | 4 | 0 | 0 |
| MAX | 4 | 2 | 0 | 1 | 3 |
| BEL | 4 | 2 | 0 | 3 | 2 |

|  | Face1 | Face2 | Face3 | Face4 | Face5 |
|---|---|---|---|---|---|
| Face1 | 5 | 3 | 4 | 1 | 2 |
| Face2 | 4 | 3 | 3 | 1 | 2 |
| Face3 | 4 | 3 | 4 | 0 | 2 |
| Face4 | 4 | 2 | 0 | 1 | 3 |
| Face5 | 4 | 2 | 1 | 3 | 2 |

Fig. 4. Example of affinity matrices from movie "NottingHill ": (a) Name affinity matrix RNAME (b) Face affinity matrix RFACE

quantitative affinity values are changeable, the relative affinity relationships between characters (e.g. A is more 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.

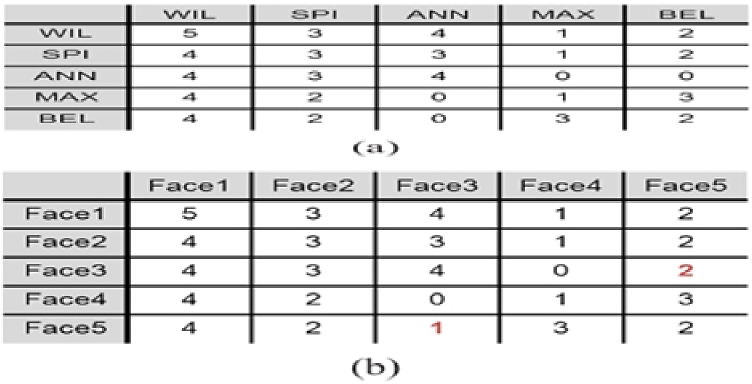
We denote the original affinity matrix as $R = \{r_{ij}\}_{N\times N}$ , where N is the number of characters. First we look at the cells along the main diagonal (e.g. A co-occur with A, B co-occur with B). We rank the diagonal affinity values $r_{ii}$ in ascending order, then the corresponding diagonal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

R
$\tilde{r}_{ii} = Ir_{ii}$ (1)
where$Ir_{ii}$ is the rank index of original diagonal affinity value $r_{ii}$. Zero-cell represents that no co-occurrence relationship is specially considered, which is a qualitative measure. 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 ith row is recorded as nulli. Other than the diagonal cell and zero-cell, we sort the rest affinity values in ascending order, i.e., for the ith row, the corresponding cells $\tilde{r}_{ij}$ in the ith row of ordinal affinity matrix:

$\tilde{r}_{ij} = Ir_{ij} + null_i$ (2)
where$Ir_{ij}$ 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 4. The ordinal matrix is not necessarily symmetric. The scales reflect variances in degree of intensity, but not necessarily equal differences. We illustrate in Fig.5 an example of ordinal affinity matrices corresponding to the affinity matrices in Fig. 4. It is shown that although there are major differences between original name and face affinity

**4 It can be considered that all the zero-cells rank first and the rest cells rank from NULLI + 1.**



|  | WIL | SPI | ANN | MAX | BEL |
|---|---|---|---|---|---|
| WIL | 5 | 3 | 4 | 1 | 2 |
| SPI | 4 | 3 | 3 | 1 | 2 |
| ANN | 4 | 3 | 4 | 0 | 0 |
| MAX | 4 | 2 | 0 | 1 | 3 |
| BEL | 4 | 2 | 0 | 3 | 2 |

(a)

|  | Face1 | Face2 | Face3 | Face4 | Face5 |
|---|---|---|---|---|---|
| Face1 | 5 | 3 | 4 | 1 | 2 |
| Face2 | 4 | 3 | 3 | 1 | 2 |
| Face3 | 4 | 3 | 4 | 0 | 2 |
| Face4 | 4 | 2 | 0 | 1 | 3 |
| Face5 | 4 | 2 | 1 | 3 | 2 |

(b)

Fig. 5. Example of ordinal affinity matrices corresponding t o figure 4:

(a)Name ordinal affinity matrix R(b) Face ordinal affinity matrix R 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 ordinal 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 [26]. 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 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 [28]. 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 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 5.

Definition 1. Let g1 = (V1, α1, β1) and g2 = (V2, α2, β2)

Be two graphs. An ECGM from g1 to g2 is a bijective function

$\hat{f}$ : $\hat{V}$1→ $\hat{V}$2, where V1⊆ V1 and $\hat{V}$2 ⊆ V2.

We say that vertex x ∈ $\hat{V}$1 is substituted by vertex y ∈ V2 if f(x) = y. If α1(x) = α2(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.

## 5 For weighted graphs, edge label is the weight of the edge.

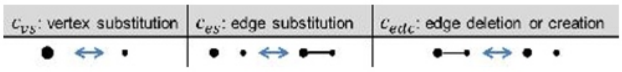

$c_{vs}$: vertex substitution
$c_{es}$: edge substitution
$c_{edc}$: edge deletion or creation

Fig. Three basic graph operators for editing graphs.

## Sensitivity Analysis:

Random coverage and intensity noise of different noise levels are generated for the sensitivity analysis. We note that for scheme 2, the simulated noises are added on the original affinity graphs. We first present the curve of ordinal graph sensitivity score change in Fig. 13. It demonstrates how sensitive the average precision of name- face matching is with respect to different coverage noise levels and intensity noise levels, respectively. It is shown that the ordinal graph is more stable towards the intensity noise. When intensity noise νC ≤ 0.08, the sensitivity score remains stable.Fig.14 shows the curve of face track classification precision@recall= 0.8 versus the simulated noises. From Fig.14(a), we can see that the proposed scheme 1 and scheme 2 basically remain stable when the intensity noise νI ≤ 0.08, while Traditional global matching-tends to decline even when νI = 0.02. From Fig.14(b), coverage noise deteriorates the

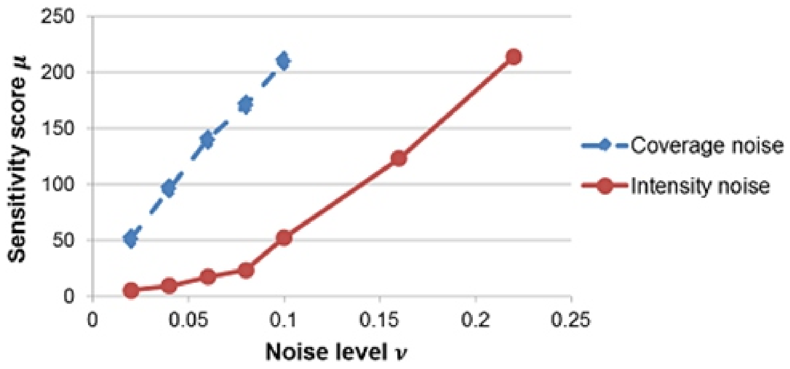
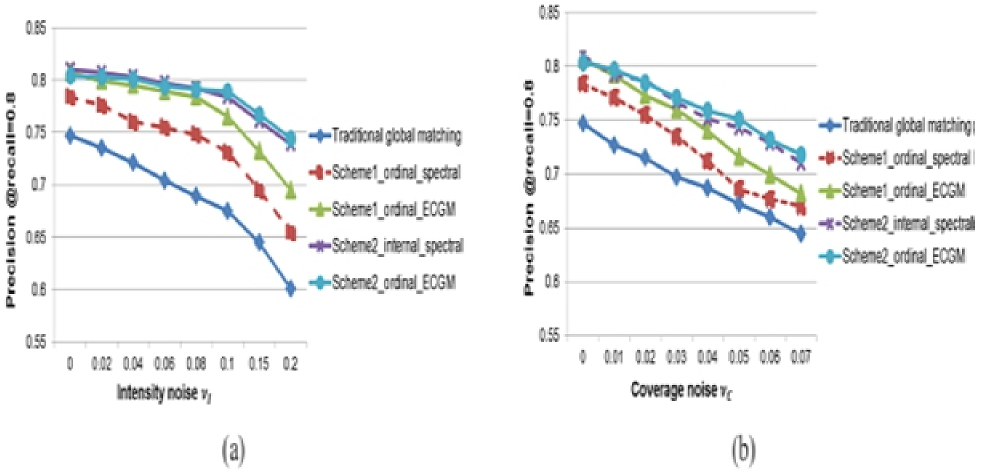


Fig. The sensitivity score μ v.s. noise level for coverage noise and intensity noise.



(a) (b)

Fig.. The precision@recall=0.8 v.s. and the simulated noise level. (a) Intensity noise (b) coverage noise classification precision of the proposed schemes as well as Traditional global matching. However, scheme 2 has a better tolerance to the coverage noises than scheme 1. This is because the simulated noises are added before the graph partition step, optimizing the graph partition helps reduce the sensitivity to noises.

We reach the conclusion that the proposed methods are more robust to the intensity noises than to the coverage noises, i.e., ordinal representation and simultaneous graph partition with graph matching have the ability to tolerate the random variations to the values of weighted edges and manage to match graphs correctly as long as the topological structure is preserved. This finding is of great importance.

According to our observation and experimental results, though the weights of face affinity relations are imprecise, basically the gene rated name and face affinity graph have the same topology. On one hand, this serves as one of the validations that we treat zero-cell different from the nonzero-cell in constructing ordinal graph. On the other hand, the design of robust character identification method needs focusing more on handling the intensity noises.

## 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 alike noises must be .

## 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 b ite 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 automat ictv 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.