Face And Name Matching In A Movie By Grapical Methods In Dynamic Way

Ishwarya, Madhu B, Veena Potdar

Abstract: With the flourishing development of the movie industry, a huge amount of movie data is being generated every day. It becomes very important for a media creator or distributor to provide better media content description, indexing and organization, so that users can easily browsing, skimming and retrieving the content of interest. Our goal is to automatically determine the cast of a feature-length film and match it with the character name. This is challenging because the cast size is not known, with appearance changes of faces caused by extrinsic imaging factors like illumination, pose, and expression often greater than due to differing identities. Although in the existing system the performances are limited due to the noises generated during the face tracking and face clustering process. The contributions of this work include: A noise insensitive character relationship. An error correcting graph matching algorithm is introduced. Complex character changes are handled simultaneously by graph partition and graph matching.

Key Words*:* Graph matching, Graph Partition, Face Tracking, Face clustering, error correction graph matching, indexing, skimming.

INTRODUCTION

In a film, characters are the focus center of interests for the audience. With the flourishing development of the movie industry, a huge amount of movie data is being generated every day. It becomes very important for a media creator or distributor to provide better media content description, indexing and organization, so that users can easily browsing, skimming and retrieving the content of interest .Their occurrences provide meaningful presentation of the video content. Character identification in feature-length films, although very intuitive to humans, still poses a significant challenge to computer methods. This is due to the fact that characters may show variation of their appearance including scale, pose, illumination, expression and wearing in a film. This can be due to:

1) Weakly supervised textual cues [6]. 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, ambiguity can also arise in partially labeled frames when there are multiple speakers in the same scene.

2) Face identification in videos is more difficult than that in images [4]. 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 [1]. 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. The work is motivated by these challenges and aims to find solutions for a robust framework for movie character identification. The objective of this work is to identify the faces of characters in the film and label them with their names. Based on our work, users can easily use the name as a query to select the characters of interest and view the related video clips.

1. EXISTING SYSTEM

In the previous work the use of face tracking and speaker detection shows the benefits of exploiting the specific properties of video. In contrast, one aspect of TV and movie footage which has been neglected is the audio. Although some of the methods showed promising face identification results in the video, most of them are used in news videos which can easily get candidate names for the faces from the simultaneous appearing captions or temporally local transcripts. Unlike the news videos which are presented by the third person such as the anchor or the reporter, in the films, the names of characters are seldom directly appearing in the subtitle, which makes it difficult to get the local name cues. Hence, many efforts on film analysis were devoted to major characters detection or automatic cast listing but not assigning real names to them. While the availability of script and subtitle makes the audio track seemingly redundant, since the script specifies who is speaking, and the subtitles specify when, there might be more information to be extracted from the audio. One area where the audio might usefully be applied is resolving the ambiguity in the subtitle/script timing. Another interesting

- Ishwarya 4th sem Mtech,Dr.AIT ishwarya27.06@gmail.com, Madhu B Assistant Professor Department of CSE
- Dr AIT bmadhu.cs@gmail.com , Veena Potdar Assistant Professor Department of CSE Dr AIT veenapotdar@gmail.com

possibility is to attempt to localize the speaker in the frame based on the audio, augmenting the visual speaker detection. In the particular domain of TV and movies, there is also ‘‘grammar” of editing in cinematography, for example alternating close-up shots during a dialogue, which could be exploited [2]. And later on the video based approach for the face recognition is presented which is able to process real world data, comprising large variations of a subject’s visual appearance. The results show that the combination of video based face recognition with a local appearance based approach is able to handle a large number of variations of facial appearance [5]. The different variations of facial appearance entail that some frames are more ambiguous than others. Therefore, two main observations are exploited to derive two different schemes to weight the contribution of each individual frame to the overall classification result. The first, distance-to model (DTM), takes into account how similar a test sample is to the representatives of the training set. Test samples that are very different from the training data will generally produce larger distances than more similar data. Consequently, they are more likely to cause a misclassification. DTM is used to reduce the impact of these samples on the final score. The second weighting scheme, distance-to-second closest (DT2ND), reduces the impact of frames which deliver ambiguous classification results. We do not discard “bad” matches (in terms of distance or score) but only reduce their contribution. Affine invariant image to image matching was used to achieve robustness to pose and a simple band pass filter to illumination changes. Faces are first affine registered and then classified in a Kernel PCA space using combined image and contextual text based features. 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 [1], [3]. 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.

2. PROPOSED SYSTEM

The cast clustering algorithm is based on pair wise comparisons of face that correspond to sequences of moving faces. Hence, the first stage of the proposed method is automatic acquisition of face data from a continuous feature-length film. The steps are as follows

(i) Temporally segment the video into *shots*.

(ii) Detect faces in each and, finally.

(iii) Collect detections through time by tracking in the (*X, Y,* scale) space. The detection of faces in cluttered scenes on an independent, frame by frame basis. 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. The traditional global matching framework is modified by using ordinal graphs for robust representation and introducing an ECGM-based graph matching method. For face and name graph construction, x, 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. 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. The observations from investigations are, in the generated affinity matrix some statistic properties 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. In this paper, the utilization of the preserved statistic properties and propose to represent the character co-occurrence in rank order. Representation of the original affinity matrix as

$$R=\{rij\}N\times N, \quad (1)$$

where N is the number of characters. First look at the cells along the main diagonal (e.g. A co-occur with A, B co occur with B). Then rank the diagonal affinity values rii in ascending order, followed by the corresponding diagonal cells r’ii in the rank ordinal affinity matrix R’:

$$r'ii=Irii \quad (2)$$

Where Irii is the rank index of original diagonal affinity value rii. 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 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 r’ij in the ith row of ordinal affinity matrix:

$$r'ij=Irijs+nulli \quad (3)$$

Where Irij denotes the order of rij. Note that the zero-cells are not considered in sorting, but the number of zero-cells will be set as the initial rank order. The ordinal matrix is not necessarily symmetric. The scales reflect variances in degree of intensity, but not necessarily equal differences.

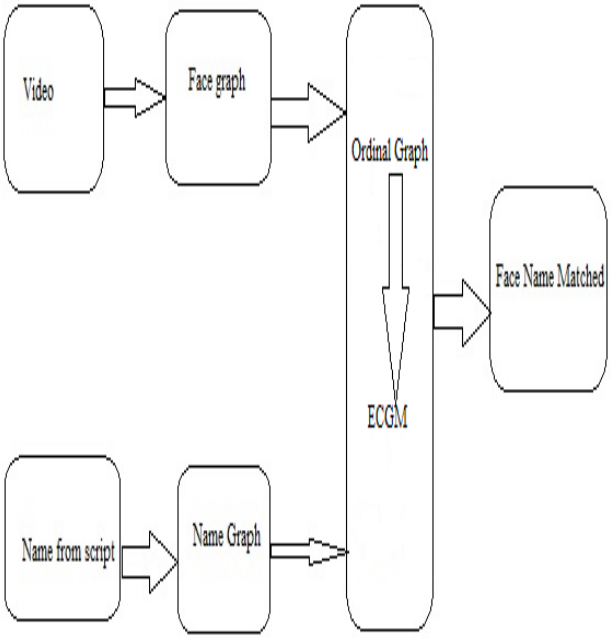

Figure1.1: Face Name matching with clusters specified.

For name face graph matching, The ECGM algorithm is utilized. 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. 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 likelihood of graph distortions. The more likely a certain distortion is to occur, the smaller is its cost. Through error correcting graph matching, appropriate graph edit operations can be defined according to the noise investigation and design the edit cost function to improve the performance. This sequence of flow is show in the Figure 1.1.Another method is used, it is different from the clustering in the following ways: 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. 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.

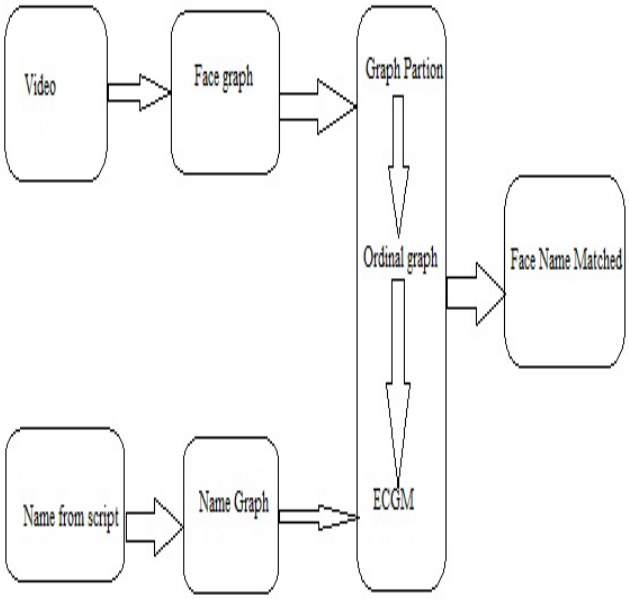

Figure1.2: Face Name matching with no clusters specified.

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. Affinity propagation is utilised for the face tracks clustering. With each sample as the potential centre of clusters, the face tracks are recursively clustered through appearance based similarity transmit and propagation. The sequence of the face name matching with no clusters specified is shown in Figure 1.2.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. 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. In general, the second method has two advantages over the clustering. (a) No cluster number is required in advance and face tracks are clustered based on their intrinsic data structure. Therefore, the second method provides certain robustness to the intra class variance, which is very common in movies where characters change appearance significantly. (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 addition, there is some chance that movie cast does not cover all the characters. In this case, pre specification for the face clusters is risky, face tracks from different characters will be mixed together and graph matching tends to fail. Sensitivity analysis plays an important role in characterizing the uncertainties associated with a model. To explicitly analyze the algorithm’s sensitivity to noises, two types of noises, coverage noise and intensity noise, are introduced. Based on that, sensitivity analysis is performed by investigating the performance of name face matching with respect to the simulated noises. 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. Due to the pose, expression, illumination variation as well as occlusion and low resolution problem, inevitable noise is generated during the process of face detection, face tracking and face tracks clustering, which mean the derived face graph does not precisely match the name graph. In this sense, the face graph can be seen as a noisy version of the name graph. Therefore, the sensitivity of the graph matching algorithms to the mixed noise is an important metric to evaluate the performance of the method and a sensitivity analysis step has become a must. For movie character identification, sensitivity analysis offers valid tools for characterizing the robustness of the algorithms to the noises from subtitle extraction, speaker detection, face detection and tracking. The sensitivity is defined to evaluate the robustness of the proposed ordinal affinity graph to noises.

3 SENSITIVITY ANALYSIS

Due to the pose, expression, illumination variation as well as occlusion and low resolution problem, inevitable noise is generated during the process of face detection, face tracking and face tracks clustering which mean the derived face graph does not precisely match the name graph. In this sense, the face graph can be seen as a noisy version of the name graph. Therefore, the sensitivity of the graph matching algorithms to the mixed noise is an important metric to evaluate the performance of the method and a sensitivity analysis step has become a must. For movie character identification, sensitivity analysis offers valid tools for characterizing the robustness of the algorithms to the noises from subtitle extraction, speaker detection, face detection and tracking.

Cost Function for ECGM

The costs for different graph edit operations are designed by automatic inference based on the training set. Parameters λ1 and λ2 embody the likelihood of different graph distortions. Set a threshold Thscore to discard noise before matching. The face tracks with function scores lower than Thscore are refused to classify to any of the clusters and will be left unlabeled. The sensitivity score function should be consistent with the likelihood of the graph distortion i.e., noises. Therefore, the ordinal graph sensitivity score μ are defined in accordance to the definition of the cost function for ECGM.

$$\mu = \sum_{x \epsilon v 1} \lambda 1 |\alpha 1(x) - \alpha 2(x)| + \sum_{e \epsilon \varepsilon 1} |\beta 1(e) - \beta 2(e)| + \sum_{\substack{\beta 1(e).\beta 2(e) \neq 0 \\ \beta 1(e) \neq \beta 2(e)}} \lambda 2 \qquad (4)$$

Where g1 means the ordinal graph before demoted by the noises and g2 is the corresponding demoted ordinal graph.

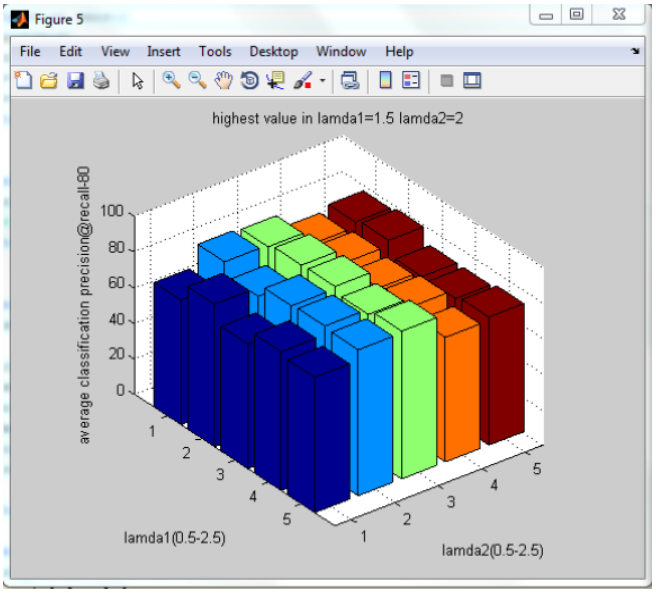

Figure 3.1: The average face track classification precision as lamda 1 and lamda 2 changes for the clustered option

The average face track classification precision as lamda 1 and lamda 2 changes for the clustered option is shown in figure 3.1 in which the highest value for lamda 1=1.5 and for lamda 2=2. The average face track classification precision as lamda 1 and lamda 2 changes for the clustered option is shown in figure 8.7 in which the highest value for lamda 1=2 and for lamda 2=1.5.

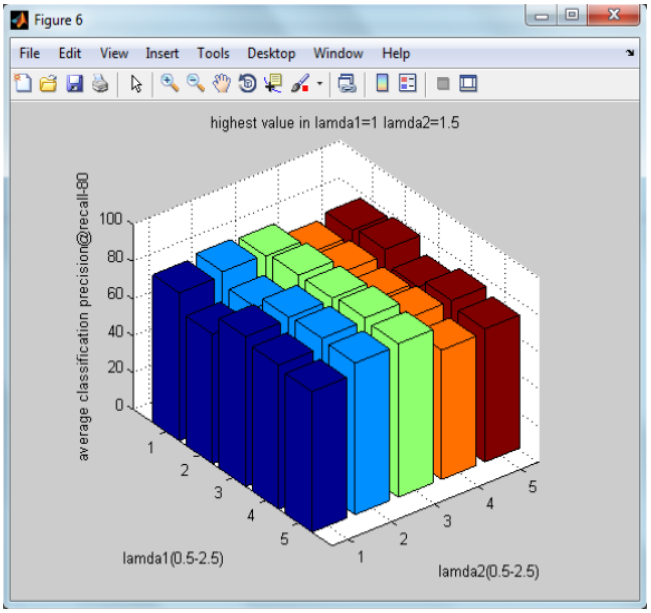

Figure 3.2: The average face track classification precision as lamda 1 and lamda 2 changes in the cluster not being specified

CONCLUSION

In this paper, a novel framework is proposed for character identification in feature length films. Different from the previous work on naming faces in the videos, most of which relied on local matching, a global matching method is

proposed. A graph matching method has been utilized to build name face association between the name affinity network and the face affinity network which are, respectively, derived from their own domains (script and video). As an application, the relationship between characters are mined and provided a platform for character-centered film browsing. In the future, improvement of our current work can be done along three directions. 1) In face name association, some useful information such as gender and context information will be integrated to refine the matching result. 2) Currently film content search and browsing is on the scene level to keep the integrity of the story. Later extend video annotation and organization on the shot level to achieve better accuracy and completeness in responding 3) will explore to generate a movie trailer related to a certain character or a group of characters.

REFERENCES

[1]. Y. Zhang, C. Xu, H. Lu, and Y. Huang, “Character identification in feature-length films using global face-name matching,”

[2]. M. Everingham, J. Sivic, and A. Zissserman, “Taking the bite out of automated naming of characters in tv video,” .

[3]. C. Liang, C. Xu, J. Cheng, and H. Lu, “Tvparser: An automatic tv video parsing method,” .

[4]. T. Cour, B. Sapp, C. Jordan, and B. Taskar, “Learning from ambiguously labelled images,” .

[5]. J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, “Video-based face recognition on real-world data.”

[6]. O. Arandjelovic and R. Cipolla, “Automatic cast listing in feature-length films with anisotropic manifold space,”