ISSN No. 0976-5697
Volume 6, No. 6, July - August 2015
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

ROBUST FACE-NAME GRAPH MATCHING FOR DIFFERENT MOVIE FRAMES

T.P.Senthilkumar
Department of computer science
Gobi Arts & Science College
Gobichettipalayam

R.Aravind
Department of computer science
Gobi Arts & Science College
Gobichettipalayam

Abstract: Auto face identification of characters in films has drawn most research interests and led to many interesting applications. Since huge variation in the appearance of each character is found, it is a challenging problem. Existing methods evaluate 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. This study presents two scheme of global face-name matching based frame work for robust character identification. This contribution of this study include: 1) A noise insensitive character relationship representation is incorporated.2) The study introduces an edit operation based graph matching algorithm.3) Complex character changes are handled by simultaneously graph partition and graph matching. algorithm4) 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 movies.

Keywords: Character identification, graph matching, graphpartition, graph edit, sensitivity analysis.

I.INTRODUCTION

Image Processing is a technique to enhance raw images received from cameras/sensors placed on satellites, space probes and aircrafts or pictures taken in normal day-to-day life for various applications. Various techniques have been developed in Image Processing during the last four to five decades. Most of the techniques are developed for enhancing images obtained from unmanned spacecraft, space probes and military reconnaissance flights.

Image Processing systems are becoming popular due to easy availability of powerful personnel computers, large size memory devices, graphics software, etc. Image Processing is used in various applications such as: Remote Sensing, Medical Imaging, Non-destructive Evaluation, Forensic Studies and Textiles Material Science, Military, Film industry, Document processing, Graphic arts and Printing Industry. The common steps in image processing are image scanning, storing, enhancing and interpretation.

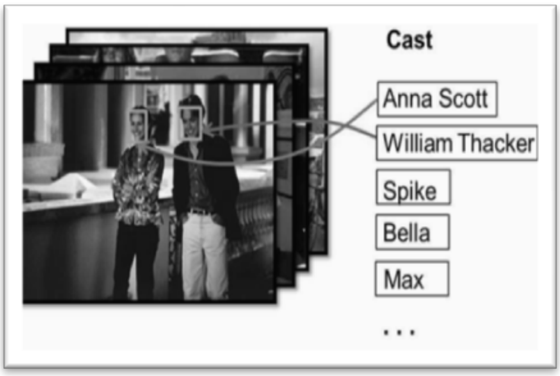

Fig.1. Examples of character identification from movie "Notting Hill".

A. Objective and Motivation

This paper proposes 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, the proposed two schemes both belong to the global matching based category, where external script resources are utilized. 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 .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

B.Related Work

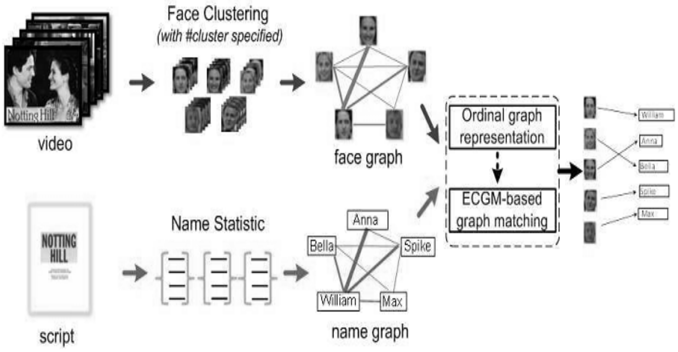

Fig.2. Framework of scheme 1: Face-name graph matching with #cluster pre-specified

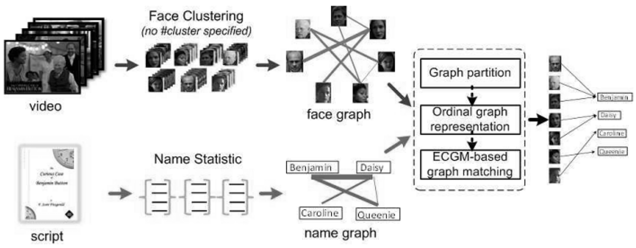

Fig.3. Framework of scheme 2: Face-name graph matching without #cluster pre-specified.

II. ROBUST CHARACTER IDENTIFICATION

In this section primarily we briefly review our 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.

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 [1, 12].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 the frames 1; 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 [2]. Low resolution, occlusion, nonrigid deformations, large motion, complex background and other uncontrolled conditions make the results of face detection and tracking unreliable [3]. 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 [4]. 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 be even different actors playing different ages of the same character.4) The determination for the number of identical faces is not trivial [5]. Due to the remarkable intra-class variance, the same character name will correspond to faces huge variant appearances. It will be unreasonable to set the number of identical faces just according to the number of characters in the cast [11].

A. Review of Global Name-Face Matching Framework

In movies, the names of characters seldom directly appear in the subtitle, while the movie script which contains character names has no time information. Hence, the task of character identification can be formulated as a global matching problem between the faces detected from the video and the names extracted from the movie script[13].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 [14]. In this sense, a name affinity graph from script analysis and a face affinity graph from video analysis can be constructed. 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.[6] Subsequently, a spectral graph matching algorithm is applied between the name affinity graph and the face affinity graph 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 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.Delighted from this, we assume that while the absolute 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.

	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

Table 1. Example of Face Affinity Matrix

We denote the original affinity matrix as $R = \{rij\}$ $n\times n$, 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-occurs with B). We rank the diagonal affinity values rii in ascending order, then the corresponding diagonal cells $\tilde{r}ii$ in the rank ordinal affinity matrix $\tilde{R}$ is:

$$\tilde{r}ii = Irii \quad (1)$$

where $Irii$ is the rank index of original diagonal affinity value rii. For each row in the original affinity matrix, we first remain the 0-cell unchanged. Then, other than the diagonal cell and 0-cell, we rank the rest affinity values in ascending order the rank order of rij is denoted as $Irij$). The rank ordinal matrix is not necessarily symmetric and the
scale reflects differences in degree of intensity, but not necessarily equal differences. There are major differences between original name and face affinity matrices, the derived rank ordinal affinity matrices are basically the same. A rough conclusion is that the affinity rank ordinal matrix is less sensitive to the noises than the original affinity matrix, which coincides with our assumption.

.

III. ERROR CORRECTING GRAPH MATCHING

Error correcting graph matching (*ECGM*) is a powerful concept that has various applications in pattern recognition and computer vision. Its application is focused on distorted inputs [7]. In order to measure the similarity of two graphs, graph edit operations are defined, such as the deletion, insertion and substitution of nodes 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 [8]. 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 [9, 10]. Let $\mathcal{L}$ be a finite alphabet of labels for nodes and edges.

Notation: A graph is a triple $g = (\mathcal{V}, \alpha, \beta)$, where $\mathcal{V}$ is the finite set of nodes. $\mathcal{V} \rightarrow \mathcal{L}$ *is* node labeling function, and $\beta : \mathcal{E} \rightarrow \mathcal{L}$ *is* edge labeling function. The set of edges $\mathcal{E}$ is implicitly given by assuming that graphs are fully connected, i.e., $\mathcal{E} = \mathcal{V} \times \mathcal{V}$. Node and edge labels
(for weighted graphs, edge label is the weight of the edge)come from the same alphabet for notational convenience.

Definition 1. Let $g1 = (\mathcal{V}1, \alpha1, \beta1)$ and $g2 = (\mathcal{V}2, \alpha2, \beta2)$ be two graphs. An error correcting graph matching (*ECGM*) from $g1$ to $g2$ is an objective function: $\hat{}\mathcal{V}1 \rightarrow \hat{}\mathcal{V}2$, where$\hat{}\mathcal{V}1 \subseteq \mathcal{V}1$ and $\hat{}\mathcal{V}2 \subseteq \mathcal{V}2$.We say that node $x \in \mathcal{V}1$ is substituted by node $y \in \mathcal{V}2$ if $(x) = y$. If $\alpha1\ (x) = \alpha2\ (f(x))$, the substitution is called an identical substitution. The cost of identical node or edge substitution is usually assumed to be zero, while the cost of any other edit operation is greater than zero.

IV. PERFORMANCES ANALYSIS

The following **Table & Fig 1** describes experimental result for proposed system over all experimental result analysis. The table contains Number of Videos cluster, number of frame images and average face detection details are shown

Number of Videos	No. of. Frame Images	AVG % Face Detection
Cluster Videos A	580	72.5
Cluster Videos B	597	74.62
Cluster Videos C	578	72.25
Cluster Videos D	557	69.62
Cluster Videos E	579	72.37
Cluster Videos F	569	71.12

Table2: Number of videos cluster, number of frame images and average face detection

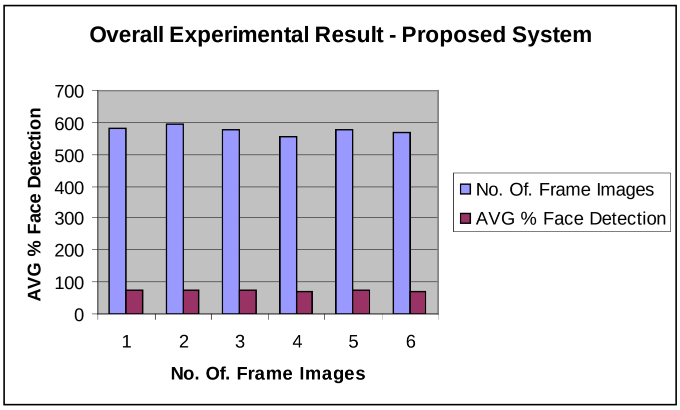

Fig4. Overall Experimental Results - Proposed System

V.CONCLUSION

The proposed schemes are useful to improve results for clustering and identification of the face tracks in different movie frames. 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.

VI.REFERENCES

[1] T. Cour, B. Sapp, C. Jordan, and B. Taskar, "Learning from ambiguously labeled images," in *CVPR*, 2009, pp. 919–926.

[2] T. Cour, C. Jordan, E. Miltsakaki, and B. Taskar, "Movie/script: Alignment and parsing of video and text transcription," in ECCV (4), 2008, pp. 158–171.

[3] J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, "Video-based face recognition on real-world data." in *ICCV*, 2007, pp. 1–8.

[4] 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.

[5] 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.

[6] 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.

[7] E. Bengoetxea, "Inexact graph matching using estimation of distribution algorithms," *PhD thesis, Ecole Nationale Superieure des Telecommunications*, 2002

[8] H. Bunke, "On a relation between graph edit distance and maximum common sub graph," *Pattern Recognition Letters*, vol. 18, pp. 689–694, 1997.

[9] M. Everingham, J. Sivic, and A. Zissserman, "Taking the bite out of automated naming of characters in tv video," in Journal of Image and Vision Computing, 2009, pp. 545–559.

[10] M. Everingham and A. Zisserman, "Identifying individuals in video by combining "generative" and discriminative head models," in ICCV, 2005, pp. 1103–1110.

[11] J.Sang and C. Xu, "Character-based movie summarization," in ACM MM, 2010.

[12] T. Cour, B. Sapp, C. Jordan, and B. Taskar, "Learning from ambiguously labeled images," in CVPR, 2009, pp. 919–926.

[13] J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, "Video-based face recognition on real-world data." in ICCV, 2007, pp. 1–8.

[14] J. Yang and A. Hauptmann, "Multiple instance learning for labeling faces in broadcasting news video," in ACM Int. Conf. Multimedia, 2005, pp. 31–40.