IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.
www.ijiset.com
ISSN 2348 – 7968

A Novel Method for Movie Character Identification based on Graph Matching

Mr. Bhausaheb Salve[1], Prof. S.A Shinde[2]

[1] M.E. Computer (II), VPCOE Baramati, Pune University,
Pune, Maharashtra, India
E-mail: salvebs1486@gmail.com

[2] Computer Department, Assistant Professor, VPCOE Baramati, Pune University,
Pune, Maharashtra, India
E-mail: meetsan_shinde@yahoo.com

Abstract

The face identification of character in images is very easy, but in videos is difficult due to the huge variation in the appearance of each character such as extrinsic imagining factor that is illumination, pose, expressions and wearing, various hair-styles, and make-up of characters.

In a good situation existing methods gives better results, however in complex movie scenes performance is limited because face tracking and clustering processes generate noise.

So, Novel method for movie character identifications based on graph matching is proposed. This method uses global face name graph matching algorithm for robust character identifications in movies. Also edit operations are used for graph matching algorithm. For complex character changes are handled by simultaneously graph matching and graph partitioning. There are two types of simulated noises introduced to improve sensitivity analysis performances.

Keywords: *Character identification, Graph edit, graph matching, graph partition, Multimedia database, Optical character recognition (OCR), clustering, Earth mover distance (EMD), Error correcting graph matching (ECGM).*

1. Introduction

Automatic character identification in movies is essential for semantic movie analysis such as movie indexing, summarization and retrieval. Character identification, though very intuitive to humans, is a tremendously challenging task in computer vision.

When you watch a movie or TV, mostly you don't know all the character's names in the movies or video. In a movie, the audience focused on characters and real name of the characters. Mostly character can be identified based on occurrences of various persons references and apply these occurrences in the movies based on many visual appearances and different speeches of characters. Suppose the woman with colored hair or the man with a different voice tones. A problem may occur of weaker annotation i.e. a screenplay or given direct supervision and a cast list with various head shots positions. So, in this situation character identification is impossible. If you may rely on movie structure to aid in resolution character occurrences on screen often alternate during dialogues, and when a new scene begins, a new set of characters is usually present.

System is interested in automatic labeling of people in TV or film material with their identity. A very big challenging problem due to the large variation in intra-person appearance variations because photometric factors such as various lighting colors, view point, and scale, or due to changes in expressions, hair styles, or occlusions, so appearances of each characters and the weakness ambiguity of available annotation. But with the increasing development of the movie industry, a large amount of digital movie data is being generated every day, for organization and understanding the video contents the effective and efficient techniques are required. Such techniques are called as automatic video annotation. Now a day's various research efforts have been examined on character identification and a lot of applications and techniques have been proposed, there have been rather few efforts directed at the robustness analysis of character identification.

This system focused on annotation of characters in the movies and TVs, which is called movie character identification [2].The textual method such as cast lists, scripts, subtitles and closed captions are used to and the faces of the persons in the movie or video and label them with the corresponding real names. Other applications are essential for automatic character identification such as movie index and retrieval (MIR) [3], [4], scene segmentation, summarization [5]. Character identification in the movies or TVs, its challenging task in computer field due to weakly supervised textual parameter [7].For making the relation between faces and names of character, ambiguity can come from the video shot, if person speaking but not present in the scene so problem occurred for face-name relation, and if number of speaker are more in same frame. Face tracking and face clustering process may be fail due to occlusion, low resolution video, large-motion, changed background and various conditions. Face identification of image is more easy than video [8]. The situation is very worse in video. This brings noises to the character identification in video. Some character playing role like from youth to the old age in movies. Also various actors playing role for different ages of the same character name. There may be various poses, a lot of expression and illumination effect, dressing, wearing, even make-up and hairstyle changes. The remarkable intra-class variance, the same person name will corresponds to the faces of large variant appearances. The determination for the number of identical faces is not trivial [3]. According to the number of persons in the cast list is equal to number of identical faces from face cluster is difficult.

So this system designed for these challenges and aims to d solutions for a novel method for movie character identification based on face name graph matching.

2. Related Work

Mostly in movie or TV, the names of characters are directly shown in the closed caption and captions, and script/screenplay containing character names has no time stamps to align to the video.

In the character identification problem, the relations between videos and the texts in order to label the faces of characters with names[2].It has similarities to identifying faces in news videos [09], [11].

According to the utilized textual cues, we divide the existing movie character identification methods into three classes.

2.1 Class 1: Based on Cast List:

This class uses only the case list textual resource. In [12],[13]problem in the 'cast list discovery', making less pure clusters which contains faces clustered by using appearances and expected faces of particular character, then by using cast list select the names for name clustering. Ramanan et al. proposed to face clustering based on clothing and wearing within scenes [14]. In [15], using model or classier of the person's appearance from training data set for problem in finding a character. In [16] system proposed for character identification with web image retrieval system.by using the characters names from the cast list as queries to find face image. By using multitasking joint sparse representation and classification technique, identified the faces in the movies. Now for character identification in uncontrolled videos metric learning method was implemented also cast metrics is used to the peoples are in video, but not in supervised manner. The clustering as well as identification performances are demon-started to be improved. These cast list based methods are easy for implementation and understanding purpose.

Sometime without other textual cues, they either need manual labeling or guarantee no perfect clustering and classification performance due to the large intra-class variances.

2.2 Class 2: Subtitle or Closed Caption, Based on Local Matching:

This class Subtitle or closed caption used for time-stamped speeches and dialogues, which can be used for video frames alignment. Everingham et al. [17], [4] for local face-name matching make a combination of the film script with the subtitle. Time stamped name annotation and face exemplars are generated. The rest of the faces were then classified into these formats for face identification. They further their work in [18], for features combination by taking the near neighbor classifier/method by multiple kernel learning method. In the new method, non-frontal faces are handled and the coverage is extended. Now a days researcher are uses available time stamped sources and closed captions, which are more reliable than Object Character Recognition(OCR) method based [19],[7]. In this method covered ambiguity issues in the alignment of screenplay, closed captions and movie videos. A partially-supervised multiclass classification problem is formulated. Now, they attempted to address the character identification problem without the use of screenplay [20]. The reference cues in the closed captions are used as multiple instance constraints, the face tracks grouping as well as face-name matching are solved in a con-vex formulation method. The time-stamped information is used for local matching, which is either extracted by OCR (i.e., subtitle) or unavailable for some of the movie videos and TV shows (i.e., closed caption).The ambiguous and partial annotation makes local matching based methods more sensitive to the face detection and face tracking noises.

2.3 Class 3: Script/Screenplay, Based on Global Matching:

Without OCR based method subtitle and closed caption is used in global matching of character identification, so it is difficult to get local name cues, the method of character identification is represented as a global matching problem in [3], [21], [4]. This method is an extension to Zhang's method [3]. In movies, the names of characters seldom directly appear in the sub-title, the script of movie contains a character's name which has no time information. Without the local time information, the method 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

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.
www.ijiset.com
ISSN 2348 – 7968

matching, global statistics are used for name-face association, which enhances the robustness of the algorithms

3. Programmer's Design

3.1 System Architecture

The proposed system [1] is shown in Fig. 1.Face tracks are clustered using classification of Haar cascade feature, then set the number of character is equal to number of names. Calculating co-occurrences of names from script analysis and number of face track from video frames. Constructing face graph and name graph. Applying traditional global matching methods for face-name graph matching. Generate face affinity matrix and name affinity matrix based on number of co-occurrences of character faces and character names. In this method vertex to vertex matching of face-name graph is constructed. In the ordinal graph representation, affixing the noises which generate in incorrect face classification and tracking process. Generate ordinal face affinity matrix and ordinal name affinity matrix and represent character co-occurrence in rank order level.

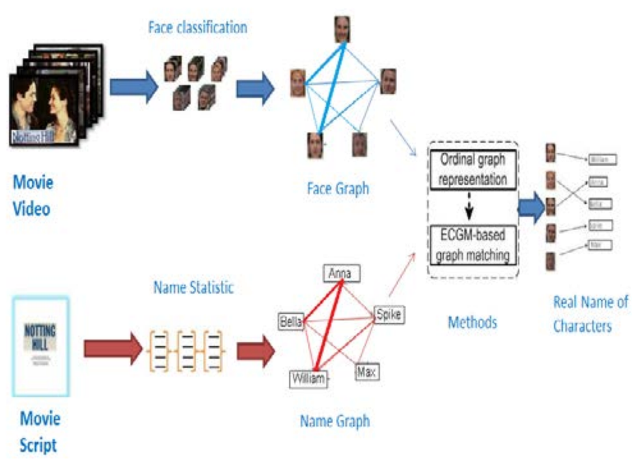


Fig.1 System Architecture [01].

The correct representation and introducing an ECGM based graph matching method. For face and name graph construction, system 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. 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, implement 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 least edit distance. According to the noise analysis, define appropriate graph edit operations and adapt the distance functions to obtain improved name-face matching performance.

3.2 Mathematical Model

3.2.1 Global Face-Name Matching

Graphs built according to the co-occurrences status among character, which is denoted by weighted graph .i.e.

$$G = (V, E)$$

Where, V= character,

E= Relationship among characters.

3.2.2 Ordinal Graph Representation

The original affinity matrix is $R = (r_{ij})_{N*N}$,
Where, N= no. of characters, the cell along the main diagonal.

Rank the diagonal affinity values r_{ii} in ascending order, then corresponding diagonal cells $\tilde{r}_{ii}$ in rank order affinity matrix R,

$$\tilde{r}_{ii} = I_{rii} \qquad (1)$$

Where, I_{rii} is the rank index of original diagonal affinity value r_{ii}.

Zero-cell represents that no co-occurrence relationship but there is any 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, so 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, 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,

$$\tilde{r}_{ij} = I_{rii} + null_i \qquad (2)$$

Where, $\tilde{r}_{ij}$ denotes the order of r_{ij}.

3.2.3 ECGM-Based Graph Matching

In this system set the number of face track clusters as the same with the number of character names, the name and face affinity graph have the same number of vertices.

Let λ be a finite alphabet of label for vertexes and edges.

Notation: A graph is a triplet $g = (v, \alpha, \beta)$, where v is the finite set of vertexes. $\alpha: v \rightarrow \lambda$ is vertex labeling function.

Definition 1: Let $g_1 = (v_1, \alpha_1, \beta_1)$ and $g_2 = (v_2, \alpha_2, \beta_2)$ be two graphs. An ECGM from g_1 to g_2 is a bijective function f: $\widehat{v1} \rightarrow \widehat{v2}$, where $\widehat{v1} \subseteq v1$ and $\widehat{v2} \subseteq v2$.

Definition 2: The cost function f: $\widehat{v1} \rightarrow \widehat{v2}$, from graph $g_1 = (v_1, \alpha_1, \beta_1)$ to $g_2 = (v_2, \alpha_2, \beta_2)$ is shown by,

$$\gamma(f, g_1, g_2) = \sum_{x \in v1-\widehat{v1}} cvd(x) + \sum_{x \in v2-\widehat{v2}} cvi(x) + \sum_{x \in \widehat{v1}} cvs(x) + \sum_{x \in \widehat{v1}} ces(e) \quad (3)$$

Definition 3: The cost function an ECGM in our face name ordinal affinity graph matching,

$$\gamma(f, g_1, g_2) = \sum_{x \in v1-\widehat{v1}} cvd(x) + \sum_{x \in v2-\widehat{v2}} cvi(s) + \sum_{x \in \widehat{v1}} |\alpha1(x) - \alpha2(x)| cvx(x) + \sum_{\beta1(e),\beta2(e) \neq 0\, \beta1(e) \neq \beta2(e)} cedc(e) + \sum_{x \in \hat{e}} |\beta1(e) - \beta2(e)| ces(e) \quad (4)$$

Where, $|\alpha1(x) - \alpha2(x)|$ and $|\beta1(e) - \beta2(e)|$ measure the degree of vertex substitution and edge substitution respectively.

According to the likelihood of graph distortions during the graph construction process, assign different costs to the edit operation of vertex substitution, edge substitution and edge creation/destruction. The cost function C is shown as,

$C = (c_{vd}, c_{vi}, c_{vs}, c_{es}, c_{edc},) = (\infty, \infty, \lambda1, 1, \lambda2)$ (5)

Where λ1 and λ2 are embody the likelihood of different graph distortions.

Consider N face cluster and character names, the number of possible states for solution space is the permutation of N, i.e. N! = N X (N-1) X….X 2 X 1.

4. Experimental Result

4.1 Face Detection Result:

4.1.1 Face Track Extracted-

Table shows the total number of true as well as false face track detected in the movie video.

Table 1: Experimental Result of face track extracted

Movie Name(ID)	***Total No. of Face extracted***
Twelve Angry Men (F1)	7727
Inception (F2)	3871
Notting Hill (F3)	16811
The Curious Case (F4)	2210
Bolt (F5)	2186
Dr.Seuss. The.Lorax (F6)	528
Inception II (F7)	1378
Seven (F8)	2430
Forrest Gump (F9)	1890
Leon: The prof. (F10)	2970

4.1.2 Face Track Detection Result-

Face track detection calculated on the accuracy of true face detected with the total number of face extracted from movie video. Following table contain true face detection and false face detection.

Table 2: Experimental Result of Face Detection.

Movie Name(ID)	***True Face Detection***	***False Face Detection***
Twelve Angry Men (F1)	6417	1310
Inception (F2)	2436	1435
Notting Hill (F3)	11711	5100
The Curious Case of Benjamin Button (F4)	1773	437
Bolt (F5)	1806	380
Dr.Seuss. The.Lorax (F6)	391	137
Inception II (F7)	993	385
Seven (F8)	1752	678
Forrest Gump (F9)	1394	496
Leon: The professional (F10)	1918	1052

4.1.3 Face-Name Matching Accuracy-

The annotation of real name of the character is a main part of the system. In this section calculate the how many face correctly annoted in the movie video. Following table shows the face-name matching result of the system.

Table 3: Experimental Result of Face Name Matching.

Movie Name(ID)	***True Face Track Detected***	***Face Correct Annoted***	***Accuracy (%)***
Twelve Angry Men (F1)	6417	6210	96.8
Inception (F2)	2436	2310	94.9
Notting Hill (F3)	11711	10954	93.6
The Curious Case (F4)	1773	1580	89.2
Bolt (F5)	1806	1750	96.5

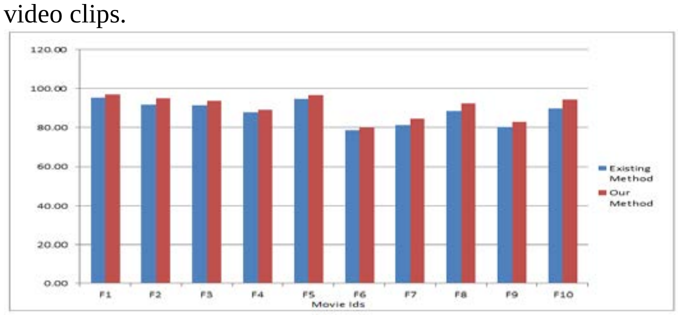
Dr.Seuss. The.Lorax (F6)	391	310	80.1
Inception II (F7)	993	840	84.6
Seven (F8)	1752	1620	92.5
Forrest Gump (F9)	1394	1170	83.1
Leon: The professional (F10)	1918	1804	94.2

4.2 Comparisons- Following table shows the comparison between existing method and our method of face track detection accuracy.

Table 4: Comparisons of Existing method and Our method.

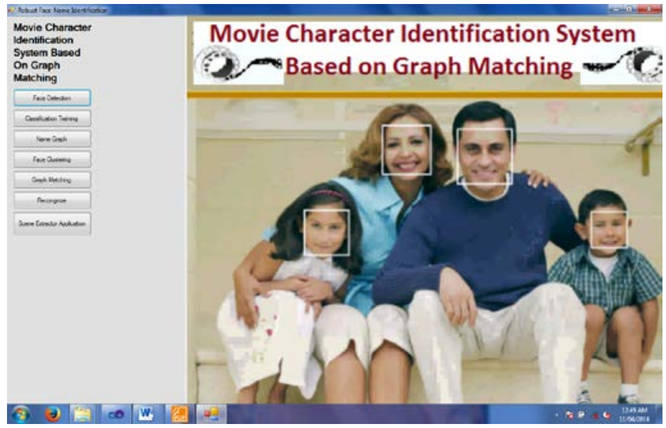
Movie Name(ID)	***Existing Method***	***Our Method***
Twelve Angry Men (F1)	95.2	96.8
Inception (F2)	92.1	94.9
Notting Hill (F3)	91.7	93.6
The Curious Case (F4)	88.1	89.2
Bolt (F5)	94.5	96.5
Dr.Seuss. The.Lorax (F6)	78.6	80.1
Inception II (F7)	81.3	84.6
Seven (F8)	88.7	92.5
Forrest Gump (F9)	80.4	83.1
Leon: The professional (F10)	89.8	94.2

Following graph shows the comparison result of movies video clips.

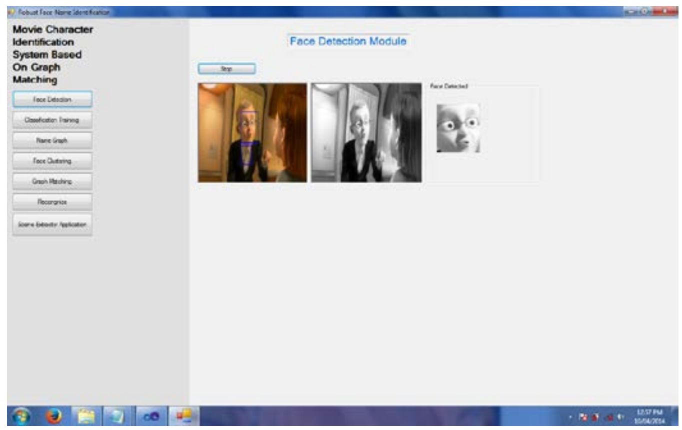


5. Experimental Result Snapshot

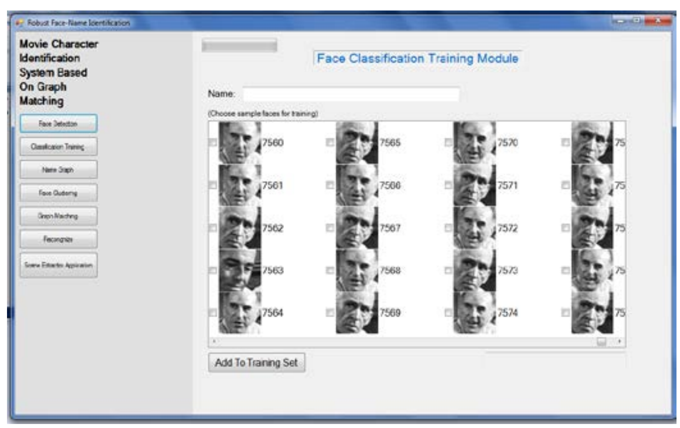
5.1 GUI of movie character identification system-



5.2 Snapshot showing Face Detection module-

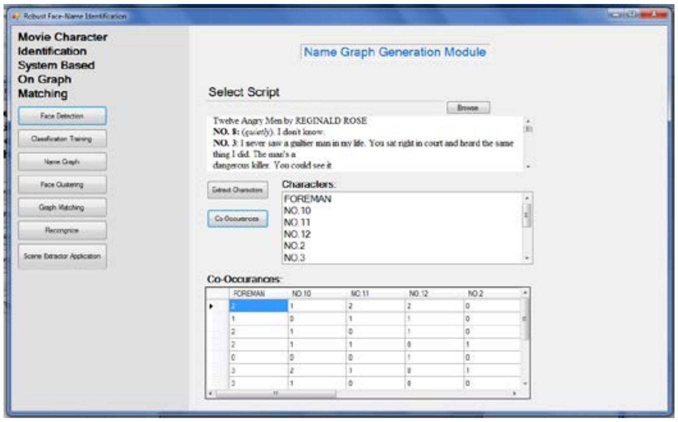


5.3 Snapshot Showing face classification for training set. -

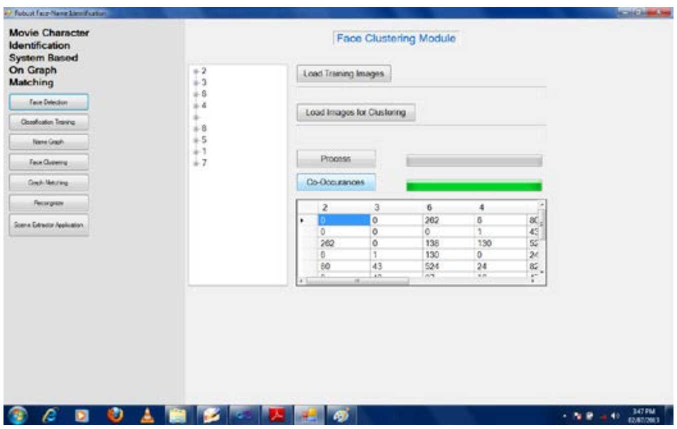


5.4 Snapshot showing Name Graph Generation-

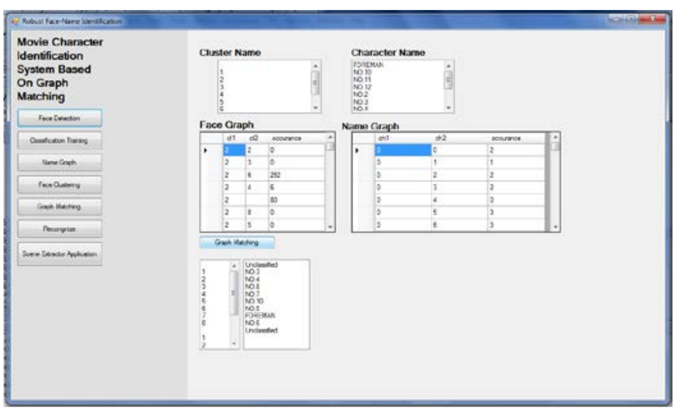
ISSN 2348 – 7968



5.5 Snapshot showing Face extracted from video and calculate co-occurrences.-

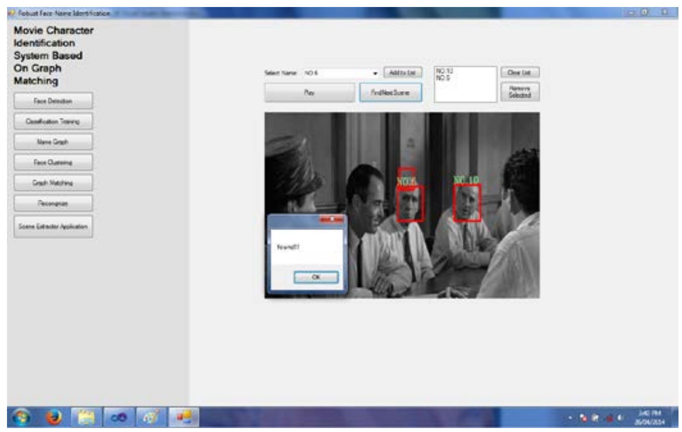


5.6 Snapshot showing ECGM based graph matching.-

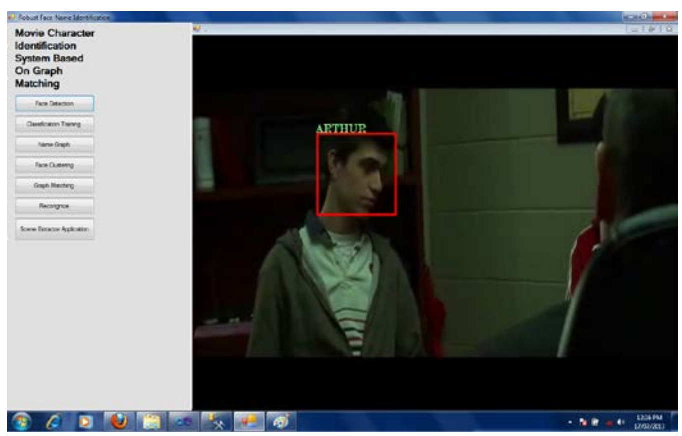


5.7 Snapshot showing generated output of system (Real Name of character).-

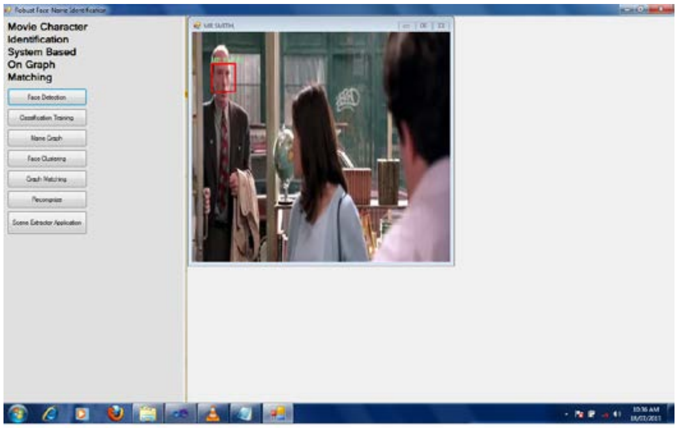
5.8 Snapshot showing characters from “Twelve angry man" movie video with scene extracted application.



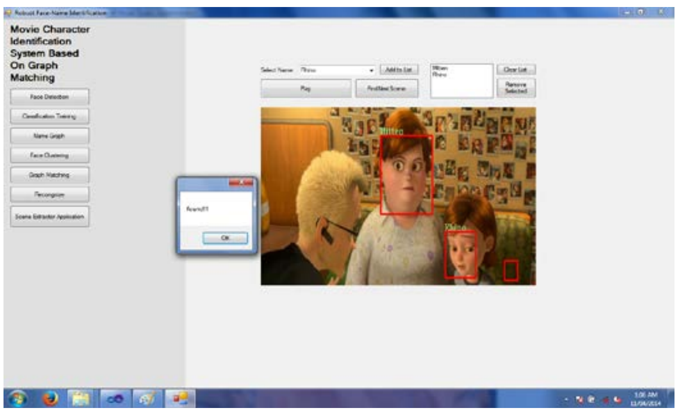
5.9 Snapshot showing characters from “Inception" movie video.



5.10 Snapshot showing characters from “Notting Hill" movie video.



5.11 Snapshot showing characters from "Bolt "movie video with scene extract application.



6. Applications

Character identification, there are many applications, such as

1. character-based video retrieval,
2. personalized video summarization,
3. Intelligent playback and video semantic mining,
4. Character Relationship mining
5. To identify characters in movies and label faces with their name.
6. To provide security at confidential area of banking sector.
7. To maintain employee attendance records and to restrict unauthorized person entry in MNC.
8. Webcam applications- Face detection in Facebook, Security at ATM machine Authentication for PC and Other security applications.

7. Conclusions

In this system, a novel method for movie character identification based on graph matching is proposed. The scheme proposed in this system improves result for classification and identification of the face tacks extracted from uncontrolled movie videos. These schemes have better robustness to the noises in comparison with traditional methods. This system is scalable in terms of datasets of movies and videos which can also contain different movie types.

Future work will be focused on improvement in following:

1. To investigate the optimal functions for different movie generes.
2. In face name matching, some useful information such as gender and context will be integrated to the matching result.
3. To exploit more character relationship, e.g., the sequential statistics for the speaker, to build affinity group and improve the robustness.
4. To conduct a more study on a larger datasets of videos.

Acknowledgments

I express great many thanks to Prof. Santosh A. Shinde, for his great effort of supervising and leading me, to accomplish this fine work. To college and department staff, they were a great source of support and encouragement. To my friends and family, for their warm, kind encourages and loves. To every person gave us something too light my pathway, I thanks for believing in me

References

[1] Jitao Sang and Changsheng Xu, "Robust Face-Name Graph Matching for Movie Character Identification," IEEE Transaction. On multimedia, Vol. 14, No. 3, June 2012.

[2] J. Sang, C. Liang, C. Xu, and J. Cheng, "Robust movie character identification and the sensitivity analysis," in Proc. ICME, 2011, pp. 1–6.

[3] 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, Nov. 2009.

[4] M. Everingham, J. Sivic, and A. Zissserman, "Taking the bite out of automated naming of characters in TV video," J. Image Vis. Comput., pp. 545–559, 2009.

[5] J. Sang and C. Xu, "Character-based movie summarization," in Proc. ACM Int. Conf. Multimedia, 2010, pp. 855–858.

[6] R. Hong, M. Wang, M. Xu, S. Yan, and T.-S. Chua, "Dynamic captioning: Video accessibility enhancement for hearing impairment," ACM Trans. Multimedia, pp. 421–430, 2010.

[7] T. Cour, B. Sapp, C. Jordan, and B. Taskar, “Learning from ambiguously labeled images,” in Proc. Comput. Vis. Pattern. Recognit., 2009, pp. 919–926.

[8] J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, “Video-based face recognition on real-world data,” in Proc. Int. Conf. Comput. Vis., 2007, pp. 1–8.

[9] S. Satoh and T. Kanade, “Name-it: Association of face and name in video,” in Proc. Comput. Vis. Pattern Recognit., 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 Proc. Comput. Vis. Pattern. Recognit., 2004, pp. 848–854.

[11] J. Yang and A. Hauptmann, “Multiple instance learning for labeling faces in broadcasting news video,” in Proc. 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 Proc. ECCV, 2002, pp. 304–320.

[13] O. Arandjelovic and R. Cipolla, “Automatic cast listing in featurelength films with anisotropic manifold space,” in Proc. Comput. Vis. Pattern Recognit., 2006, pp. 1513–1520.

[14] D. Ramanan, S. Baker, and S. Kakade, “Leveraging archival video for building face datasets,” in Proc. Int. Conf. Comput. Vis., 2007, pp. 1–8.

[15] M. Everingham and A. Zisserman, “Identifying individuals in video by combining “generative” and discriminative head models,” in Proc. Int. Conf. Comput. Vis., 2005, pp. 1103–1110.

[16] R. G. Cinbis, J. Verbeek, and C. Schmid, “Unsupervised metric learning for face identification in TV video,” in Proc. Int. Conf. Comput. Vis., 2011, pp. 1559–1566.

[17] M. Everingham, J. Sivic, and A. Zissserman, “Hello! my name is buffy automatic naming of characters in tv video,” in Proc. BMVC, 2006, pp. 889–908.

[18] J. Sivic, M. Everingham, and A. Zissserman, “Who are you?—Learning person specific classifiers from video,” in Proc. Comput. Vis. Pattern Recognit., 2009, pp. 1145–1152.

[19] T. Cour, C. Jordan, E. Miltsakaki, and B. Taskar, “Movie/script: Alignment and parsing of video and text transcription,” in Proc. ECCV, 2008, pp. 158–171.

[20] T. Cour, B. Sapp, A. Nagle, and B. Taskar, “Talking pictures: Temporal grouping and dialog-supervised person recognition,” in Proc. Comput. Vis. Pattern Recognit., 2010, pp. 1014–1021.

[21] Y. Zhang, C. Xu, J. Cheng, and H. Lu, “Naming faces in films using hypergraph matching,” in Proc. ICME, 2009, pp. 278–281.

[22] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, “Sensitivity analysis for chemical models,” Chem. Rev., vol. 105, no. 7, pp. 2811–2828, 2005.

[23] E. Bini, M. D. Natale, and G. Buttazzo, “Sensitivity analysis for fixedpriority real-time systems,” Real-Time Systems, vol. 39, no. 1, pp. 5–30, 2008.

[24] R. E. Hanneman, Introduction to Social Network Methods. Riverside, CA: Univ. California, 2000, Online Textbook Supporting Sociology 157.

[25] E. Bengoetxea, “Inexact graph matching using estimation of distribution algorithms,” Ph.D. dissertation, Ecole Nationale Supérieure des Télécommunications, Paris, France, 2003.

[26] A. Sanfeliu and K. Fu, “A distance measure between attributed relational graphs for pattern recognition,” IEEE Trans. Syst. Man Cybern., vol. 13, no. 3, 1983.

[27] H. Bunke, “On a relation between graph edit distance and maximum common subgraph,” Pattern Recognit. Lett., vol. 18, pp. 689–694, 1997.

[28] B. J. Frey and D.Dueck, “Clustering by passing messages between data points,” Science, vol. 315, pp. 972–977, 2007.

[29] Y. Rubner, C. Tomasi, and L. J. Guibas, “A metric for distributions with applications to image databases,” in Proc. Int. Conf. Comput. Vis., 1998, pp. 59–66.L. Lin, X. Liu, and S. C. Zhu, “Layered graph matching with composite cluster sampling,” IEEE Trans. Pattern Anal.Mach. Intell., vol. 32, no. 8, pp. 1426–1442, Aug. 2010.

[30] H. Chui and A. Rangarajan, “A new point matching algorithm for nonrigid registration,” Comput. Vis. Image Understand., vol. 89, no. 2–3, pp. 114–141, 2003.

Authors

Salve Bhausaheb S. received his B.E. degree in Information Technology engineering from the Pune University, Pune, in 2010. He is currently working toward the M.E. degree in Computer engineering from the University of Pune, Pune. He has 04 years of teaching experience at undergraduate level. His research interests lies in Digital Image Processing.

Shinde Santosh A. received his B.E. degree in computer engineering (First Class with Distinction) in the year 2003 from Pune University and M. E. Degree (First Class with Distinction) in Computer Engineering in 2010 from Pune University. He has 11 years of teaching experience at undergraduate and postgraduate level. Currently he is working as Assistant Professor in Department of Computer Engineering of VPCOE, Baramati, Pune University. His autobiography has been published in Marquie’s Who’ Who, an International Magazine of Prominent Personalities of the World in the year 2012. He is also a Life Member of IACSIT and ISTE professional bodies. His research interests are Digital Image processing and web services.