# A Review on: Automatic Movie Character Annotation by Robust Face-Name Graph Matching

Bhandare P.S.
Research Scholar
Sinhgad College of Engineering, korti, Pandharpur, Solapur University, INDIA

Gadekar P.R.
Assistant Professor
Sinhgad College of Engineering, korti, Pandharpur, Solapur University, INDIA

Bhise Avdhut S.
HOD, Department of Information Technology, College of Engineering (Poly), Pandharpur, Solapur, INDIA

Bandgar Vishal V.
Assistant Professor
College of Engineering (Poly), Pandharpur, Solapur, INDIA

## ABSTRACT
Now a day's character Identification from films is a very challenging task due to the huge variation in the appearance of each & every character. It will lead significant research interests and may have many interesting applications in today's life. In this paper, we investigate the problem of identifying characters & annotating them with respective name using graph matching algorithm to get the most accurate identification result. The contribution of our work include: 1) the character-character relationship representation including a noise insensitive 2) Use an edit operation based error correcting graph matching algorithm.3) Graph partitioning and graph matching for more complex character changes to handle simultaneously. 4) The existing character identification approaches, also we are going to perform an in-depth sensitivity analysis which will introduce two types of simulated noises.

## General Terms
Image Processing, Pattern Recognition, and Movie Character annotation from video.

## Keywords
Character identification, Graph edits, graph matching, graph partition, Video Optical character recognition (VOCR), Error correcting graph matching (ECGM). Movie index and retrieval (MIR), Optical character recognition (OCR).

## 1. INTRODUCTION
The detonation of movie and TV videos content large amount of video data. Which will led to the need of more efficient techniques for analyzing and organization of video content. Automatic video character annotation is one of such key techniques. Our focus is on identifying & labeling characters with particular name in the movie and TV videos, where this technique is called movie character identification [1]. The objective for our project is to identify the faces of the characters from the video and label them with their names in the name cast. The cast lists, scripts, subtitles and closed captions are usually overworked are known as textual prompts. Fig.1 shows an example in our character identification. The characters are the focus for center of interests for the audience in movies & TV videos. These provide lots of clues about the movie structure and content. This technique of automatic character identification is very essential for semantic movie index and content retrieval, scene segmentation [7], summarization[2] and other applications such as movie index and retrieval (MIR) [3], [4].

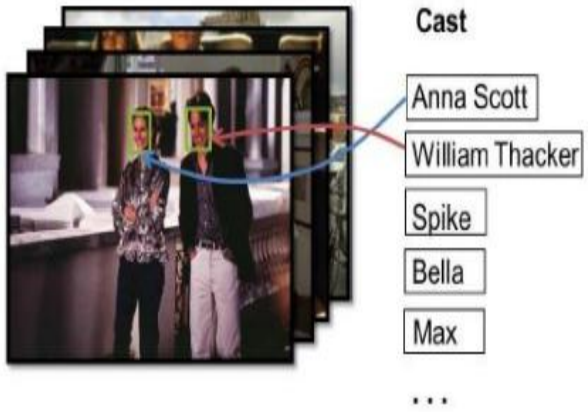


**Fig1: Movie character Identification example**

The movie character identification is a extremely challenging task in computer sight. The reasons are following :1) The textual cues[8] are weakly monitored .The ambiguity problem in establishing the relationship between names and faces: the ambiguity can arise from a reaction shot where the person speaking may not be shown in the frames 1 and in partially labeled frames when there are multiple speakers in the same scene. 2) The face identification in videos is more complex than that in images as video is sequence of images which are continuously moving [9][24]. The problems that may arise in the video data are: a) Low resolution, b) Occlusion, c) non rigid deformations, d) large motion, e) complex background and f) Other uncontrolled conditions make the results of face detection and that of tracking undependable. This brings predictable noises to the movie character identification. 3) The characters appear quite differently during the movie [3]. The character in the movie may have huge poses; expressions and illumination variations, different wearing at different time in the video, different clothing, even change in makeup and hairstyle. One movie may contain characters which will go through different age stages, e.g., from child age to youth age & youth age to the

old age. 4) For this reason the determination for the number of similar faces from the video is big. Due to the significant intra-class variance, the character having names will correspond to different faces of huge variant appearances. It will be illogical to set the number of similar faces just according to the number of characters in the cast. This paper motivated by all these difficult challenges and which will lead aims to find solutions to accurate movie character identification & annotation.

## 2. LITERATURE REVIEW

The main problem in linking faces from face database with names from name script is to form relations between images or videos and texts for labeling the faces with their corresponding names. The very first proposal for face-name association in news videos is Name-it on which is based on the co-occurrence between the detected faces and names extracted from the video transcript [5]. The system which built a database of named faces by video optical character recognition (VOCR) is called Named Faces system [6]. Yang et al. have developed a closed caption and speech transcript, based on that he built model. Thus which is improved their methods doing dynamic captioning which usages multiple instance of learning for partially labeled faces to reduce collected data by users. Speech transcript was also used for finding people those are frequently appearing in the news videos. In news videos, candidate names are made available from local matching while in TV and movie videos, the characters names are rarely directly produced in the subtitle or closed caption which contains names and no time stamps for aligning the video. The proposed system for the same frame work has shown in Fig. 2 which will use clustering method?

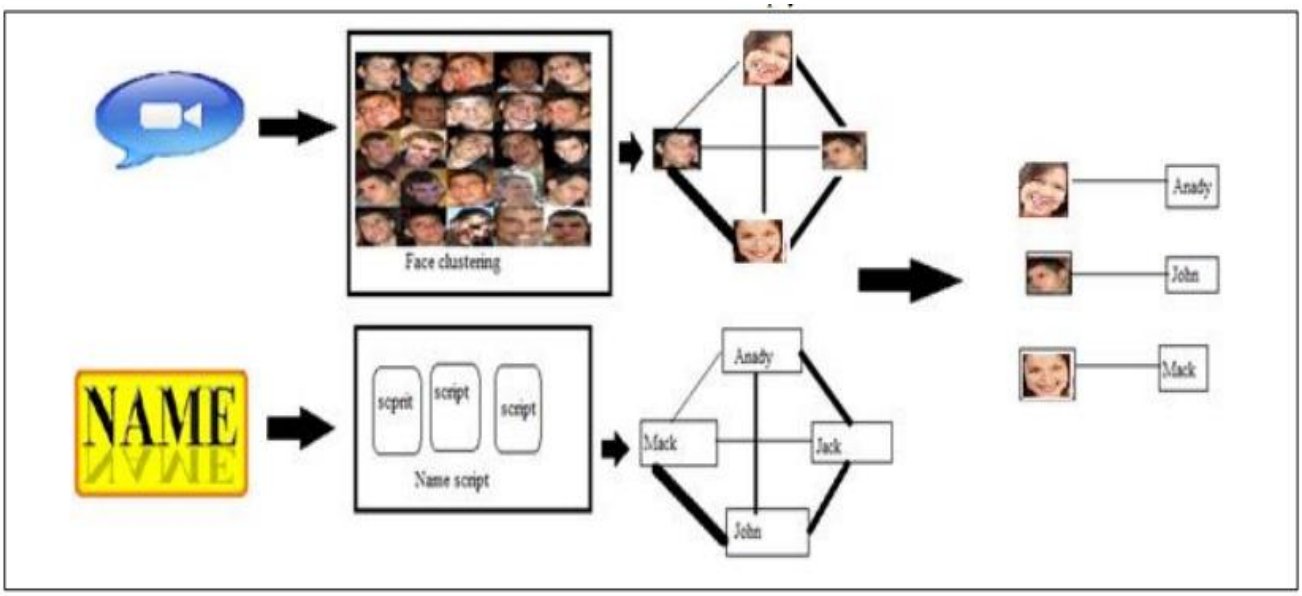


**Fig 2: Framework for face name relationship.**

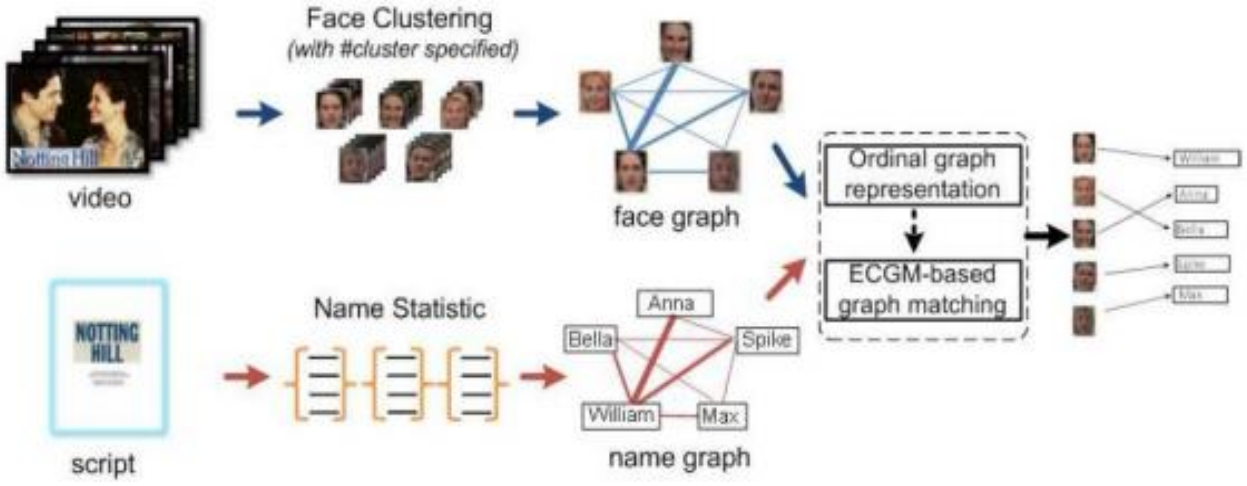


**Figure 3: Frame work of Scheme 1: Face name graph matching with clusters specified**

## 3. RELATED WORK

The crux of the character identification problem is to overwork the relations between videos and the related texts in order to label the faces of characters from the video with names from the name script. It has similarities for identifying faces in news videos [10][11][12]. However, in news videos, candidate names for the faces are available from the simultaneously appearing captions or local transcripts. According to the utilized textual cues, we divide the existing movie character identification methods into three basic categories:

### 3.1. Cast list based

This method gives utilization of the cast list textual resource. Where in this "Cast list discovery" problem [13][14], character faces are clustered by their appearance and faces of a particular character are expected to be collected in a some pure clusters. The names of characters in the cast are used as queries to search face images of characters and establish gallery set. The investigated face tracks from the movie are then recognized as one of the characters by multitask joint thin representation and classification. The metric learning is to be introduced into identification of character from uncontrolled videos [15]. The clustering and identification operation is must be amended. These methods based on cast list are easy for understanding and implementation of the identification.

### 3.2 Subtitle or Closed caption, local matching based:

The method Subtitle and closed caption provides time-stamped dialogues, which can be overworked for aligning the video frames. The remaining faces from the video are then classified for identification into these exemplars. By using this method they are going to generate time-stamped name annotation and face exemplars. They further extended their work which is going to replace the nearest neighbor classifier by multiple kernels learning to combine the various different features. Non-frontal faces are handled in the new framework and the reporting is extended. The methods based on local matching require the information in time-stamped format, which is either unavailable for the majority of movies and TV videos or may be extracted by OCR (i.e., subtitle)[16]. Recently, they attempted to address the movie character identification problem without the using screenplay [17].

### 3.3 Script/Screenplay, Global matching based:

The methods based on Global matching open the possibility of character identification without using OCR-based subtitle or closed caption. This is more difficult to get local name cues and the task of identification of characters is derived as a global matching problem [18]. In the movies, the characters names directly appear in the subtitle, where as the movie script that contains character names has no time information. If the information about local time is not present, the task of identifying character is formulated as a global matching problem between the detected faces from the movie video or TVs and the names extracted from the movie script or name script. The comparison of the local matching & global statistics are used for face-name association, which helps to enhance the robustness of the algorithms. Sensitivity analysis is common applications in financial, risk analysis, signal processing and any other areas where models are developed [19], [20].

## 4. THE PROPOSED WORK

• In movie video various appearances may be shown by one character, the noise introduced by face tracking is often affect representation of character, scene segmentation and face clustering. The extensive research efforts have been concentrated on character identification & automatic annotation and many applications have been proposed by this research, little work has focused on improving the robustness. Here our observation in investigation is that some statistic properties are preserved to inspire of these noises.

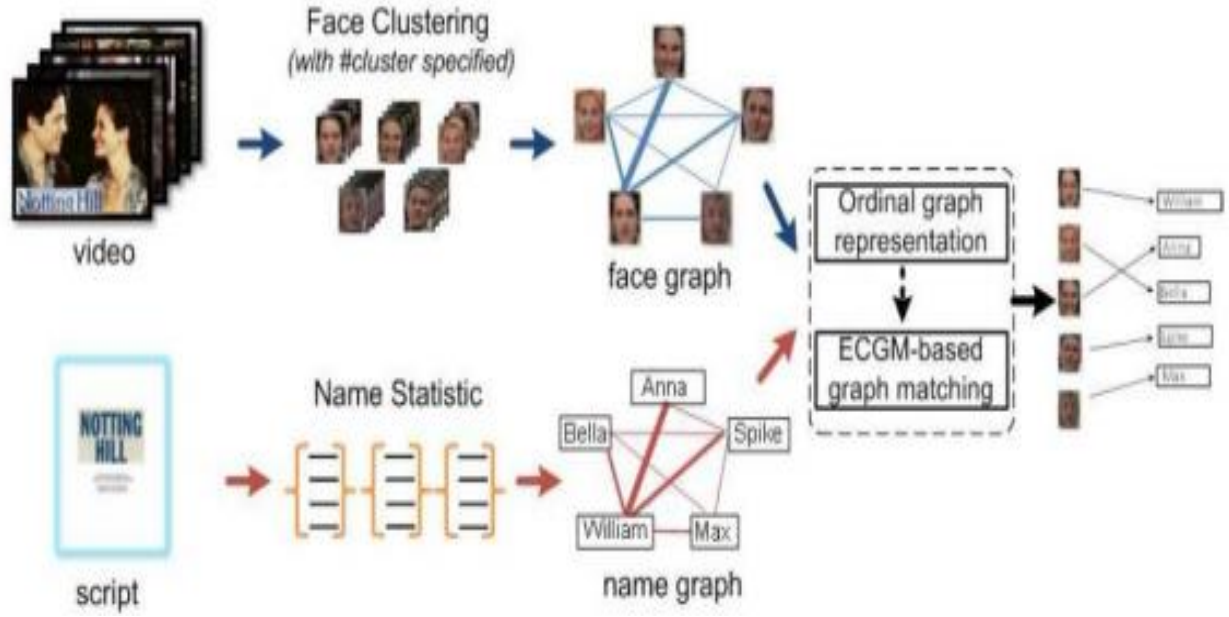


**Fig.4: Frame work of Scheme 2: Face name graph matching with number of clusters specified**

• The most important step in movie character identification is face track clustering. In this proposed work we choose any of one face clusters corresponding to one character name. Face-name matching and face track clustering and are jointly optimized and conducted in a unique framework.

• Sensitivity analysis in the field of movie character identification, offers valid tools for characterizing the robustness to reduce noises for a model. In this work, we are going to fill up such gap by introducing two types of simulated noises.

## 5. OVERVIEW OF OUR APPROACH

Our paper approach is to propose framework for movie character identification which is based on global face-name graph matching. Here in this we are going to considered two schemes .One is connections and another is differences between them. In first scheme regarding the connections, it proposed two schemes which belong to the category based global matching; script resources which are external are utilized. To improve the robustness of work, the ordinal graph is employed both graphs face graph and name graph representation and after representing ordinal graph we are going to introduce a graph matching algorithm called Error Correcting Graph Matching (ECGM). Regarding the differences, scheme 1 contains sets of the number of clusters when performing face clustering (e.g., K-means, spectral clustering). The face graphs are 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 built-in data structure (e.g., mean shift, affinity propagation). Moreover, as shown inFig.3 and Fig.4, scheme 2 has an one additional module of graph partition compared with scheme 1.

### 5.1 Scheme 1:

Face tracks are clustered using constrained K-means algorithm, where the number of clusters is set equals to the number of distinct speakers. Co-occurrence of names in script and face clusters in video form two graphs one of them is face graph and the other is name graph. The difference between two graphs is measured in ECGM by using edit distance method which is a sequence of graph edit operations. The optimal match is achieved with least edit distance. For construction of face and name graph, we propose to represent the movie character co-occurrence in rank ordinal level [21],which scores the strength in the relationships in a rank order from the weakest to strongest. This scheme is described in a Figure 3.

### 5.2 Scheme 2:

This is shown in the following figure 4. It has two differences from scheme 1 in Figure 3. One of them no cluster number is required for the face tracks clustering step and another are both graphs face graph and name graph may have different number of vertexes, a graph partition component is added before ordinal graph representation.

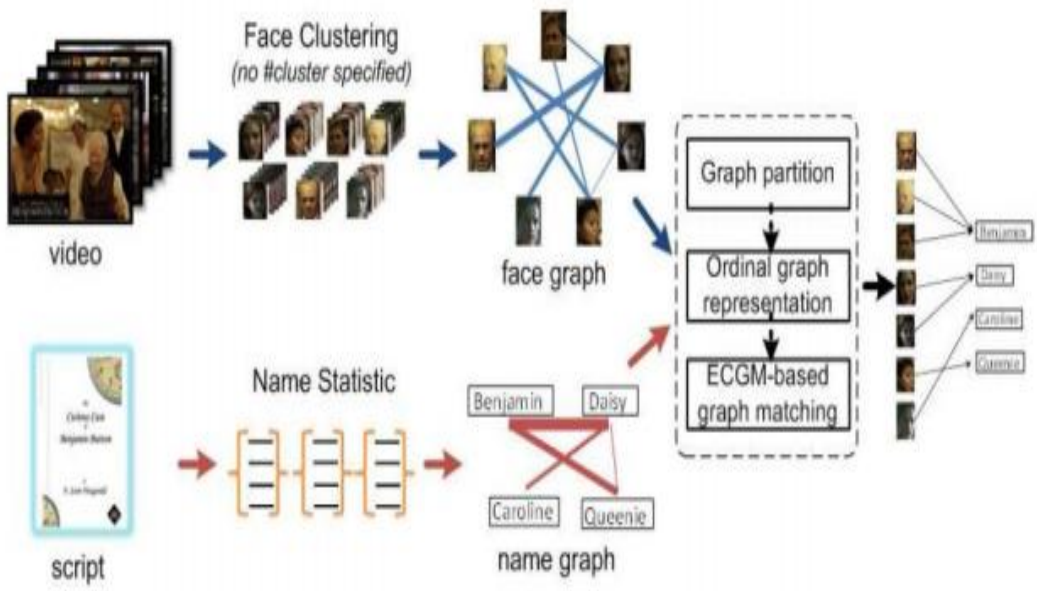


**Fig 4 : Frame work of Scheme 1: Face name graph matching with #clusters specified**

## 6. REVIEW OF GLOBAL FACE-NAME MATCHING FRAMEWORK

In a movie, the interactions among different characters corresponds them into a new network relationship network. Co-occurrence of names from script and faces from videos can represent such interactions. Affinity graph is built by considering the co-occurrence status among them denotes relationships between them. In the movie video scenes, where two or more characters appear together at same time, if they closer top each other, larger the value of edge weights between them. In this video sense, two types of graphs name affinity graph from name script analysis and a face affinity graph from video analysis can be constructed. All the affinity values from both graphs are normalized into the interval of [0, 1]. From this conclusion we can say that some of the face affinity values are different much from the corresponding name affinity values due to the noises that are introduced. Subsequently, character identification is formulated as the problem where are going to find optimal vertex to vertex matching between name graphs & face graph. A spectral graph matching algorithm is applied for finding the optimal name-face correspondence.

### 6.1 ECGM-based Graph Matching

Error correcting Graph Matching (ECGM) algorithm is a powerful tool for graph matching. It has various applications the field of pattern recognition and computer vision [22]. To measure the similarity between two graphs, operations for editing graph are defined, such as the deletion of node, insertion of node and substitution of vertexes and edges by another vertex or edge. Each of these operations is assigned a certain cost. The costs are depends on application and are usually reflect the likeness of graph straining. The more likely a certain distortion is occurred, and then its cost is smaller. Through error correcting graph matching tool, we can define appropriate operations for editing graph according to the noise investigation and the edit cost function is designed to improve the performance. For explanation, here we provide some notations with definitions from [25]. Let L be a finite alphabet of labels for vertexes and edges.

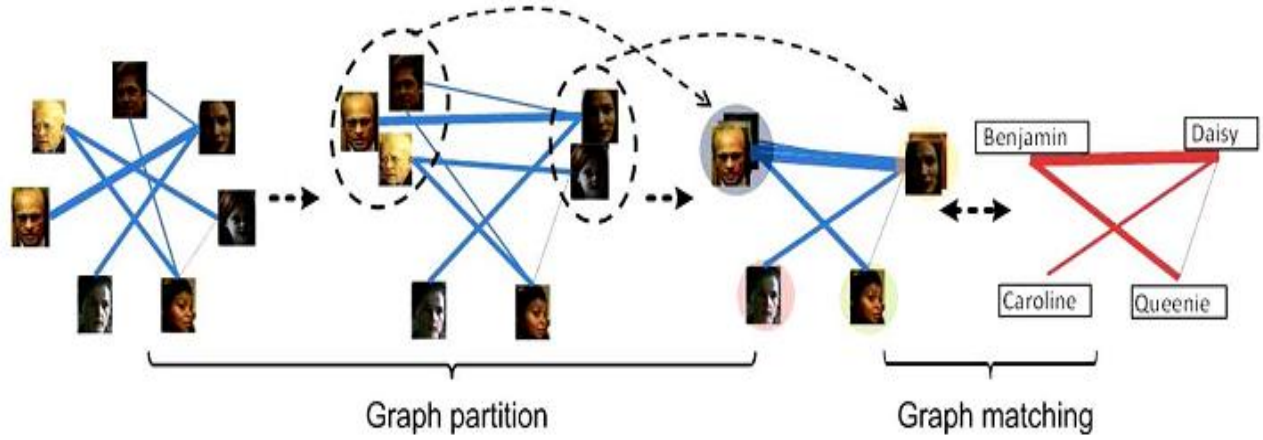


**Fig. 5: Simultaneously graph partition and matching for scheme 2**

**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. Set of edges ε is implicitly given by assuming that graphs are fully connected,

i.e., ε = ν × ν. For the notational convenience, node and edge labels come from the same alphabet 5.

**Definition 1**.Let g1 = (V1, $\alpha_1$, $\beta_1$) and g2 = (V2, $\alpha_2$, $\beta_2$ ) be two graphs. An ECGM from g1 to g2 is a objective function f : v1 → v2, where v1 ⊆ V1 and v2 ⊆ V2. We say that vertex x ∈v1 is substituted by vertex y ∈v2 if f(x) = y. If $\alpha_1$ (x) = $\alpha_2$ (f(x)), such type of substitution is called an identical substitution. The cost for identical vertex or the cost edge substitution is assumed to zero, while the cost of any other graph edit operation is always greater than zero.

## 7. CONCLUSION

The proposed schemes are useful to improve results for clustering and identification of the face tracks extracted from uncontrolled movie videos. 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 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. 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. We can implement this technique for online face annotations[23].

## 8. REFERENCES

[1] J. Sang, C. Liang, C. Xu, and J. Cheng, "Robust movie character identification and the sensitivity analysis," in Proc. 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, Nov. 2009.

[3] 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.

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

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

[7] C.Liang,C.Xu,J.Cheng,andH.Lu,"Tvparser: An automatic tv video parsing method," in Proc. Comput. Vis. Pattern Recognit., 2011, pp.3377–3384.

[8] T. Cour, B. Sapp, C. Jordan, and B. Taskar, "Learning from ambigu-ously labeled images," in Proc. Comput. Vis. Pattern. Recognit., 2009,pp. 919–926.

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

[10] S. Satoh and T. Kanade, "Name-it: Association of face and name invideo," in Proc. Comput. Vis. Pattern Recognit., 1997, 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 Proc. Comput. Vis. Pattern. Recognit., 2004, pp. 848–854.

[12] J. Yang and A. Hauptmann, "Multiple instance learning for labelingfaces in broadcasting news video," in Proc. ACM Int. Conf. Multi-media, 2005, pp. 31–40

[13] A.W. Fitzgibbon and A. Zisserman, "On affine invariant clustering and automatic cast listing in movies,"in Proc. ECCV, 2002, pp. 304–320.

[14] O. Arandjelovic and R. Cipolla, "Automatic cast listing in feature-length films with anisotropic manifold space," in Proc. Comput. Vis.Pattern Recognit., 2006, pp. 1513–1520.

[15] 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.

[16] J. Sivic, M. Everingham, and A. Zissserman, "Who are you?—Learning person specificclassifiers from video," in Proc. Comput. Vis. Pattern Recognit., 2009, pp. 1145–1152

[17] 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.

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

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

[20] E. Bini, M. D. Natale, and G. Buttazzo, "Sensitivity analysis for fixed-priority real-time systems," Real-Time Systems, vol. 39, no. 1, pp. 5–30,2008.

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

[22] E. Bengoetxea, "Inexact graph matching using estimation of distribu-tion algorithms," Ph.D. dissertation, Ecole Nationale Supérieure des Télécommunications, Paris, France, 2003.

[23] Bokefode J.D, Ubale S. A, Modani D. G, Bhandare P.S. "Enhancing the web site structure to provide easy traversal on a website with minimum changes to its structure ", International Journal of Computer Engineering & Technology (IJCET), Volume 5, Issue 1, January (2014),ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375.

[24] Bhandare P.S, Bokefode J.D, Bhise A. S, More P. B, "Analysis of Electrocardiograph using Perceptron Feed Forward Neural Network" International Journal of Computer Applications (0975 – 8887)Volume 90 – No 1, March 2014.

[25] H. Bunke, "On a relation between graph edit distance and maximum common subgraph," Pattern Recognit. Lett., vol. 18,