Vol.3 Issue 9, Pg.: 23-29
September 2015

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

ROBUST FACE-NAME GRAPH MATCHING FOR MOVIE CHARACTER IDENTIFICATION

P.Raja[1], R. Gayathiri[2]

[1]M.phil. Research scholar, Department of computer science, Kaamadhenu arts and Science College,
[2] Assistant Professor, Department of computer science, Kaamadhenu arts and Science College.

Abstract: - In this thesis proposes a global face-name graph matching based framework for robust movie character identification. The proposed system 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.

This research study 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. The main contributions of this study includes, first noise insensitive character relationship representation is incorporate, next study introduces an edit operation based graph matching algorithm, next complex character changes are handled by simultaneously graph partition and graph matching and beyond existing character identification approaches. The proposed schemes demonstrate state-of-the-art performance on movie character identification in various movies.

1. Introduction

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems.

This thesis 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.

2. Literature Review

Mark Everingham, Josef Sivic and Andrew Zisserman described, , the authors investigated the problem of automatically labelling appearances of characters in TV or film material with their names. This is tremendously challenging due to the huge variation in imaged appearance of each character and the weakness and ambiguity of available annotation. However, they demonstrated that high precision can be achieved by combining multiple sources of information, both visual and textual. The principal novelties that we introduce are: (i) automatic generation of time stamped character annotation by aligning subtitles and transcripts; (ii) strengthening the supervisory information by identifying when characters are speaking. In addition, they incorporate complementary cues of face matching and clothing matching to propose common annotations for face tracks, and consider choices of classifier which can potentially correct errors made in the automatic extraction of training data from the weak textual annotation. Results are presented on episodes of the TV series ‘‘Buffy the Vampire Slayer”.

Jitao Sang And Changsheng Xu explained about a movie summary which is helpful for movie producer to promote the movie as well as audience to capture the theme of the movie before watching the whole movie. Most exiting automatic movie summarization approaches heavily rely on video content only, which may not deliver ideal result due to the semantic gap between computers calculated low-level features and human used high-level understanding. They incorporate script into movie analysis and propose a novel character-based movie summarization approach, which is validated by modern film theory that what actually catches audiences' attention is the character [9]. They first segment scenes in the movie by analysis and alignment of script and movie. Then they conduct sub story discovery and content attention analysis based on the scene analysis and character interaction features. Given obtained movie structure and content attention value, we calculate movie attraction scores at both shot and scene levels and adopt this as criterion to generate movie summary. The promising experimental results demonstrate that character analysis is effective for movie summarization and movie content understanding.

Timothee Cour and Benjamin Sapp, the authors have accessed only to partially labeled image and video collections data. For example, personal photo collections often contain several faces per image and a caption that only specifies who is in the picture, but not which name matches which face. Similarly, movie screenplays can tell us who is in the scene, but not when and where they are on the screen. They formulated the learning problem in this setting as partially-supervised multiclass classification where each instance is labeled ambiguously with more than one label. They theoretically showed that effective learning is possible under reasonable assumptions even when all the data is weakly labeled. Motivated by the analysis, they propose a general convex learning formulation based on minimization of a surrogate loss appropriate for the ambiguous label setting. They applied their framework to identify faces culled from web news sources and to naming characters in TV series and movies. They experiment on a very large dataset consisting of 100 hours of video, and in particular achieve 6% error for character naming.

Johannes Stallkamp, Hazim K. Ekenel and Rainer Stiefelhagen, the authors presented the classification sub-system of a real-time video-based face identification system which recognizes people entering through the door of a laboratory. Since the subjects are not asked to cooperate with the system but are allowed to behave naturally, this application scenario poses many challenges. Continuous, uncontrolled variations of facial appearance due to illumination, pose, expression, and occlusion need to be handled to allow for successful recognition. Faces are classified by a local appearance-based face recognition algorithm. The obtained confidence scores from each classification are progressively combined to provide the identity estimate of the entire sequence. They introduced three different measures to weight the contribution of each individual frame to the overall classification decision. They are distance to-model (DTM), distance-to-second-closest (DT2ND), and their combination. Both a k-nearest neighbor approach and a set of Gaussian mixtures are evaluated to produce individual frame scores.

Andrew Fitzgibbon and Andrew Zisserman, the authors developed a distance metric for clustering and classification algorithms which is invariant to affine transformations and includes priors on the transformation parameters. Such clustering requirements are generic to a number of problems in computer vision. They extended existing techniques for affine-invariant clustering, and show that the new distance metric outperforms existing approximations to affine invariant distance computation, particularly under large transformations. In addition, they incorporate prior probabilities on the transformation parameters. This further regularizes the solution, mitigating a rare but serious tendency of the existing solutions to diverge. For the particular special case of corresponding point sets they demonstrate that the affine invariant measure they introduced may be obtained in closed form.

3. Overview of Project

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 evaluates 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 schemes of global face-name matching based framework for robust character identification.

The contributions of this study include:

- A noise insensitive character relationship representation is incorporated.
- The study introduces an edit operation based graph matching algorithm.
- Complex character changes are handled by simultaneously graph partition and graph matching.
- 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. The experimental study of this system has been developed using Visual Studio .Net 2005 as front end and SQL Server 2000 as back end. C# is used as the coding language.

4. Architecture Diagram

For face and name graph construction, the system represents the character co-occurrence in rank ordinal level, which scores the strength of the relationships in a rank order from the weakest to strongest. The affinity graph used in the traditional global matching is interval measures of the co-occurrence relationship between characters.

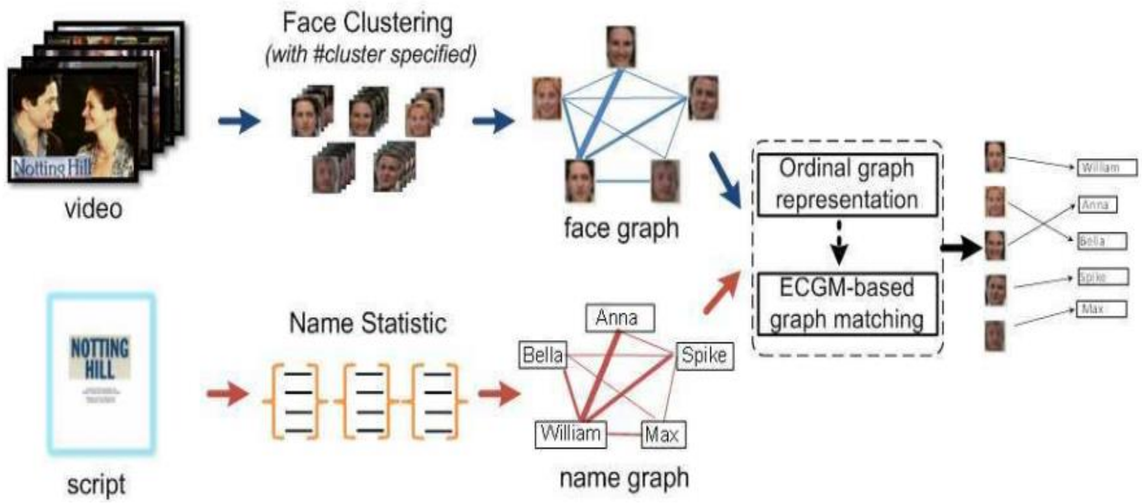

Figure 1: Architecture diagram for face name graph matching

5. Techniques

5.1 Data Input

I) Movie file selection using open file dialog control. AVI File is selected as input and saved in table.

II) Movie file is selected from table, and split into individual frames using AVI extractor "avifil32.dll" methods and saved as bitmaps. The bitmaps folder in the project is used to save all the frames. The record is saved into 'Bitmaps' folder with movie id and frame id.

III) Movie file is selected from table, frame id is selected from the retrieved bitmap frame id is selected and title sentence is added.

IV) Character names found in title is added into 'FaceNames' table with movie id, frame id and name.

5.2 Face Recognition

I) After the movie id selection, from the bitmap frames face area is found out and the details are saved in 'Faces' table with movie id, frame id and face data.

II) From the frames, each one is selected.

III) Converted into gray scale image

IV) Morphological filter is applied with erosion property (3X3) matrix is given as input for erosion process.

V) Then Contour (Border) is found out. Then based on the given width/height ratio, places where images can be found out.

VI) Select face regions' Image data are saved into database with X and Y location along with width and height of the area.

5.3 Face Clustering

I) after the movie id selection, faces are clustered such that K Means clustering is applied with 'N' clusters is given as input.

II) Based on the color difference in the bitmap pixels, the face similarity is calculated.

5.4 Forming Face Graph

I) Based on the faces (multiple characters) appeared in the bitmap frames, relationship between faces is formed.

II) For example, one frame contains Face A, B and C other contains A and C. So A is more related with C and less related B.

III) The edge weight is fixed based on the relationship/ common occurrence between faces.

5.5 Forming Name Graph

I) Based on the names (multiple characters named) appeared in the bitmap frames title, relationship between names is formed.

II) For example, one frame contains name A, B and C other contains A and C. So A is more related with C and less related B.

III) The edge weight is fixed based on the relationship/ common occurrence between names.

5.6 Graph Matching

I) Matching is done based on common occurrences of faces and names.

II) For example If frame 1 contains Face A, B and C with name X, Y and Z.

III) Then frame 2 contains Face A and C with name X and Z, then it is sure that A and B have the names X and Z.

IV) After intersecting all the frames with Face/Name occurrence, we try to match the names with faces.

6. Proposed System Algorithms

I) Frames from multiple movie files are consolidated (grouped) as if they are taken from single movie.

II) For the given face (in the selected frames), the names appeared are grouped. For example if Name Jack from Frame 1 of movie 1 and Name George from frame 1 of Movie 2, occurred for the same face, then both are considered as same actor name.

III) Likewise all the faces in all the frames are checked for combined name appearances in both movies taken for frame selection.

IV) Common names for same face data in two movie frames are compared for occurrences in multiple places and treated as same character.

7. Experiments

7.1Module Description

1. Add Movie

In this module, the movie file is selected and added into table. AVI File is selected as input. Media Player control is provided to check that the AVI file is running properly.

2. Split into Frames

In this module, the movie file is selected from table, and split into individual frames and saved as bitmaps. The bitmaps folder in the project is used to save all the frames. The record is saved into 'Bitmaps' folder with movie id and frame id.

3. Add Title

In this module, the movie file is selected from table, frame id is selected from the retrieved bitmap frame id is selected and title sentence is added. The character names found in title is added into 'FaceNames' table with movie id, frame id and name.

4. Face Recognition

In this module, after the movie id selection, from the bitmap frames face area is found out and the details are saved in 'Faces' table with movie id, frame id and face data.

5. Face Clustering

In this module, after the movie id selection, faces are clustered such that K Means clustering is applied with 'N' clusters is given as input. Based on the color difference in the bitmap pixels, the face similarity is calculated.

6. Forming Face Graph

In this module, based on the faces (multiple characters) appeared in the bitmap frames, relationship between faces is formed. For example, one frame contains Face A, B and C other contains A and C. So A is more related with C and less related B. The edge weight is fixed based on the relationship/ common occurrence between faces.

7. Forming Name Graph

In this module, based on the names (multiple characters named) appeared in the bitmap frames title, relationship between names is formed. For example, one frame contains name A, B and C other contains A and C. So A is more related with C and less related B. The edge weight is fixed based on the relationship/ common occurrence between names.

8. Graph Matching

In this module, matching is done based on common occurrences of faces and names. For example If frame 1 contains Face A, B and C with name X, Y and Z. Then frame 2 contains Face A and C with name X and Z, then it is sure that A and B have the names X and Z. After intersecting all the frames with Face/Name occurrence, we try to match the names with faces.

8. Conclusion and Future Enhancements

At present, the faces are clustered, graphs formed for name and faces, then graph matching is carried out. The noise removal process is also carried out and faces are effectively clustered. Different movies are taken and so list of names for single face (occurred in both movies) can also found out.

In the future, we may 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, characters speaking in the scene, pose recognition and improve the robustness.

REFERENCES

[1] 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.

[2] O. Arandjelovic, A. Zisserman, Automatic face recognition for film character retrieval in feature-length films, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, 2005, pp. 860–867.

[3] A.W. Fitzgibbon, A. Zisserman, on affine invariant clustering and automatic cast listing in movies, Proceedings of the 7th European Conference on Computer Vision, vol. 3, Copenhagen, Denmark, 2002, pp. 304–320.

[4] J. Sivic, M. Everingham, A. Zisserman, and Person spotting: video shot retrieval for face sets, in: Proceedings of the International Conference on Image and Video Retrieval, Singapore, 2005, pp. 226–236.

[5] P. Felzenszwalb, D. Huttenlocher, Pictorial structures for object recognition, International Journal of Computer Vision 61 (1) (2005) 55–79. pp. 860–867.

[6] M. Everingham, A. Zisserman, Identifying individuals in video by combining 'generative' and discriminative head models, in: Proceedings of the 10th International Conference on Computer Vision, Beijing, China, 2005, pp. 1103– 1110.

[7] T. Berg, A. Berg, J. Edwards, M. Maire, R. White, Y.W. Teh, E. Learned-Miller, D.Forsyth, Names and faces in the news, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, 2004, pp. 848– 854.

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

[9] J. Monaco. How to Read a Film: The Art, Technology, Language, History and Theory of Film and Media. Oxford Univ. Press, New York, 1982.

[10] G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi, and Y. Avrithis. Movie summarization based on audiovisual saliency detection. In ICIP 2008, pages 2528–2531, October 2008.

[11] Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. J. Kuo. Techniques for movie content analysis and skimming. IEEE Signal Processing Magazine, 23:79–89, March 2006.

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