ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

A framework for automatic Script-to-Movie production

C.Nithya[1] , M.Priyadharsini[2]

II.M.Tech(Multimedia Technology), KSR College of Engineering, Erode, India[1]

Assistant Professor, Department of Computer Science, KSR College of engineering, Erode, India[2]

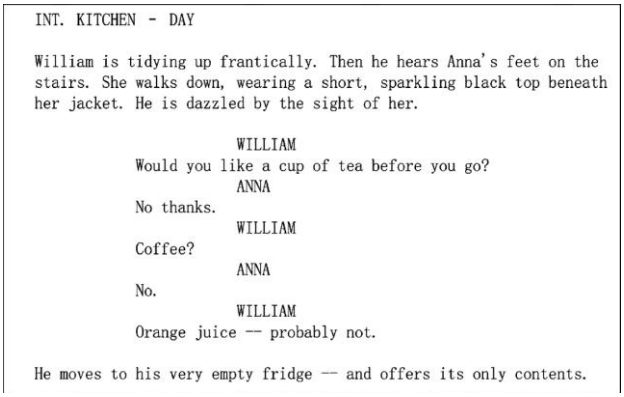
Abstract: In past, a production of movie was a highly professional work. It needs some more team work, advanced devices, techniques, time and money investment. A new technique is introduced called script – to – movie composition, for the automatic production of movie according to the user's designed script. This makes easy editing and filming for the producers. To make a new movie, a video database is maintained with the semantic description, about the different characters in various time and places. By using this S2M composition, the semantic story plot and syntactic visual content are identified as a video segment to narrate the user designed script.

Index Terms: movie composition, semantic story plot, syntactic visual content, video database.

I. INTRODUCTION

An important art and entertainment form is movie, which is popular with people all over the world. According to the reports of Motion Picture Association of America, the global movie industry released 560 films in 2010. This is not achieved by every one of the movie industry. Because, movie making is a specialized work, it needs group of skilled labors, highly developed plans and procedures. In this movie production, more time and efforts spent on video filming and editing.

The producers have to make the film with proper audio-visual materials and organize them in a particular sequential order. For all these, we need massive manual operations, both labor-intensive and skill-required. To overcome this, the two methods are proposed; they are mash up [2] and remix [3] to collect materials from the existing movies.



INT. KITCHEN - DAY

William is tidying up frantically. Then he hears Anna's feet on the stairs. She walks down, wearing a short, sparkling black top beneath her jacket. He is dazzled by the sight of her.

WILLIAM
Would you like a cup of tea before you go?
ANNA
No thanks.
WILLIAM
Coffee?
ANNA
No.
WILLIAM
Orange juice -- probably not.

He moves to his very empty fridge -- and offers its only contents.

Fig.1. a movie script with relevant scene time, Place, characters, dialogs and actions

The fig.1. Shows the movie script with the detailed story plot about time, place and characters. So it is qualified to be a semantic reference for the movie production. For this new form of production, the user designed script is given as input and the story video with the suitable place, time and characters are produced as a output. This type of movie production is known as S2M composition.

This S2M composition is a general- purpose application; in this the users can feed the story plot and get a relevant video output. A professional movie not only narrates the story plot, but also provides the audience with the high quality video outputs.

II. MOVIE COMPOSITIONS

A. Manual Movie Production

In the manual movie production more effort and time is spent on video filming and editing, where producers need to concentrate more on proper filming on audio-visual materials and organize them in a sequential order to narrate the story plot. These steps usually contain massive manual operations, which are both labor-intensive and skill-required. This movie production always needs proficient work with more time and money investment.

B. Script to Movie Composition

In Script to Movie Composition, there will be an automatic production of movie. In this movie production the two different ideas are used. They are mash up and remix, to collect video materials from existing videos and also other approaches [4], [5] to simplify the video content organization. For the immediate selection and arrangement of these video content, they are encapsulated as a black box this is not visible to end producer. It not only decreases the burden of filming and editing but also it helps in quick script design and video preview.

Here we use new form of production as S2M composition, whose input is the user designed script and output is a relevant story video of matched time, place and characters composed from massive existing movies.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

Several technical challenges are needed to be addressed to support the movie composition application: a) S2M is a general-purpose application that is not designed for a particular movie or TV drama, its users in principle can submit arbitrary story plot and have cause for a reasonable video output. b) Given user designed script and abundant audio-visual materials, the core problem of S2M composition is the video segment selection and organization. It considers both semantic and syntactic requirements and finds an optimal solution to well balance them synthetically. c) Movie production is not just a visual work, but instead an integrated multimedia authoring containing video, audio and text.

III. RELATED WORK

Video authoring and video annotation are two most related areas of S2M composition.

A. Video Authoring

It concerns raw material selection and their chronological arrangement and a lot of research efforts have been devoted on various video types including news video [6], home video [7], sports video [8], concert recording [2], documentary films [2] and movies [9], [4], [10], [11], [5]. The ideas of showing dynamic images while writing a story is first originated by textable movie project [9]. Textable movie was designed to automatically select and edit movies in real time based on textual input from the teller.

B. Video Annotation

Video annotation bridges high-level semantic description with low-level video features. Semantic scene extraction outlines the basic unit for advanced video content understanding. Compared with low – level shot detection, scene partition not only consider visual feature smooth but also require high-level semantic correlation. The proposed scene transition graph (STG) [23] to model the visual similarity and temporal adjacency between any two shots, then applied a hierarchical agglomerative-clustering to split the STG into a group of sub graphs, each of which corresponded to an individual scene segment.

Normalized cut (NCut) [14], [15] is used to segment the above shot similarity graph and a global optimal, solution was indicated by the Eigen vectors of the graph laplacian. Besides semantic scenes, another important story element is the character. Character appearance and interaction provide a meaningful representation of the story plot.

IV. OUR METHOD

The system frame work of our proposed S2M composition is shown in fig.2, where the whole flow can be functionally divided into two parts: off-line annotation and on-line composition.

The Off-line annotation adopts a hierarchical alignment idea to explain the video content with its original script description from which important plot clues, such as time, places and character identities, can be extracted and indexed to form a semantically annotated video material database. The on-line composition is responsible for organizing proper video material in accordance with user designed story.

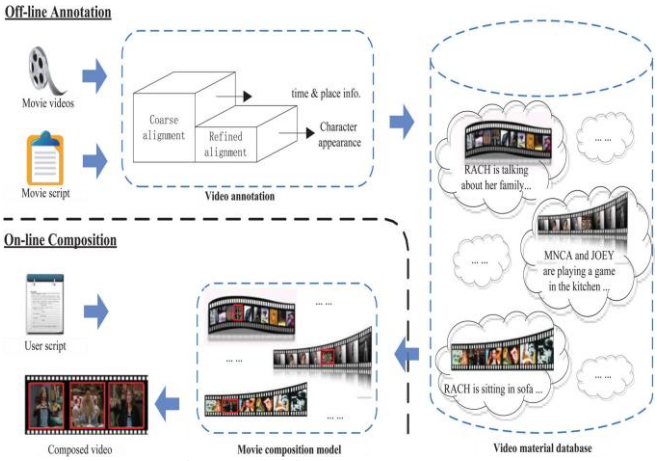


Fig. 2. The framework of the proposed S2M composition.

V. CONTRIBUTION OF OUR WORK

We raise a novel and interesting S2M application for story movie production, through which people's story idea can be automatically converted into relevant video content. Our work focuses on story plot and characters interaction.

We propose an integrated computational frame work to support the S2M composition. It includes a hierarchal alignment method to automatically analyze and annotate existing movie video with their script descriptions. On the other hand, it formulates the S2M composition as a constrained optimization problem, where semantic story plot and syntactic viewing aesthetic are jointly considered in the video content generation process.

VI. EXPERIMENTAL RESULTS

For material annotation, we investigate the performance of coarse video/script alignment from video face naming and semantic scene segmentation, and validate the effectiveness of refined shot/ dialog alignment by comparing multiple pattern alignment methods.

For movie composition, both quantitative and qualitative experiments are conducted to study the performance of the proposed S2M composition model, where modules of semantic matching and syntactic smoothing are numerically tested under various fusion parameters and the quality of composed video is subjectively evaluated through a well designed user study experiments.

A. Data Set

For the experiment, the popular two drama series are stored in the database. Both the series are featured with a set of main characters and also other characters, which involves in other episodes. There are number of interactions

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

in each drama series, which involves in the qualitative evaluation.

B. Off – line Annotation

Off – line annotation involves the hierarchal alignment, to match script description with the video content. It also aligns the face naming and scene segmentation.

Implementation

The result of an evaluation depends on the pattern translation method, in which the shot/dialog is aligned to video/script. Particularly it extracts the name and face appearance pattern.fig.4. The pattern appearance is retrieved in an alternative optimization manner. For this first set the initial scaling factor $\theta = 1$ and iterate the pattern alignment. To verify the effectiveness of the sequence matching algorithm, we compare the random match and full match appearance pattern, respectively and then their summation ratios are used to compute θ. (sum (n) / (sum (f)).

The manually labeled shot-dialog alignment matrix G^* and relevant scaling factor θ^* are used to evaluate the scaling factor by exp $(-|\theta - \theta^*|)$.

To calculate the pattern matching result, the following terms have to be calculate,

The standard precision,

- $\mathbf{G_{pre}}$ = (correctly matched pairs) / (all matched pairs)

The recall,

- $\mathbf{G_{rec}}$ = (correctly matched pairs) / (len(f))

The F – measure,

- $\mathbf{G_{Fmeans}} = (2 * G_{pre} * G_{rec}) / (G_{pre} + G_{rec})$

Where,

G is the matching matrix,

f is the face appearance pattern,

C. On – line composition

The main objective of the on-line composition is to measure the performance of the proposed S2M composition model, by combining two sides of semantic matching and syntactic smoothing. Due to the lack of evaluation criteria, the direct evaluation of generated video quality is not possible.

Implementation

The two main components of movie composition model is semantic matching and syntactic smoothing. For semantic matching the pattern matching matrix S is calculated and for syntactic smoothing the CENTRIST descriptor is adopted to depict the scene feature E. The three experiments are designed to evaluate the performance of the video composition model.

- *Experiment 1*

The effect of various similarity functions on the semantic matching is investigated and its performance is measured by script – based video retrieval.

- *Experiment 2*

This second experiment changes the study focus from the individual semantic matching module to the holistic composition model.

- *Experiment 3*

It examines the dynamic characteristics of the proposed movie composition model under various fusion parameters from its two main components.

Semantic matching

$\mathbf{V_{sem}} = - t_r (B^T S)$

Where,

$\mathbf{B} = [B_{ij}] \in R^{M*N}$,a binary indication matrix.

$\mathbf{S} = [S_{ij}] \in R^{M*N}$,a Semantic matching matrix.

Syntactic matching

$\mathbf{V_{syn}} = \rho_1 t_r (KK^T)$

Where,

ρ is the fusion parameter,

$t_r (KK^T)$, where $K = (I – 1/M\ 1_{M*M})$ BE

and

I is a unit matrix,

1_{M*M} is a full 1 square matrix,

$\mathbf{B} = [B_{ij}] \in R^{M*N}$,a binary indication matrix.

VII. CONCLUSIONS

S2M composition provides a new way for story movie production and people's can easily convert their story into continuous video content in an automatic manner. It also helps the producers to focus on storytelling rather than technical details. The off-line annotation brings an abundant video material database with detailed semantic description and on-line composition selects and organizes proper video segments to visualize user designed story plot. So it does not need any training data and it can easily deploy in large-scale applications. In the future, the user can narrate their story plot in a secured manner for automatic video output.

REFERENCES

[1] Theatrical Market Statistics 2010, Motion Picture Association ofAmerica, 2011, Report. [Online]. Available: http://www.mpaa.org/policy/industry.

[2] Francesco Cricri1, Kostadin Dabov1, Igor D.D. Curcio2, Sujeet Mate2, Moncef Gabbouj1, "Multimodal Event Detection in User Generated Videos", IEEE International Symposium on Multimedia,2011.

[3] N. Nitta and N. Babaguchi, "Example-based video remixing," *MultimediaTools Appl.*, vol. 51, no. 2, pp. 649–673, 2011.

[4] Pablo Cesar, Dick C. A. Bulterman, Luizfernando Gomes Soares, "Introduction to Special Issue: Human-Centered Television—Directions in Interactive Digital Television Research", ACM, 2008.

[5] Edirlei Soares de Lima, Bruno Feijó, Antonio L. Furtado," Automatic Video Editing For Video-Based Interactive Storytelling", UNIRIO – Departamento de Informática Aplicada,

[6] Gonzalo Ramos, Ravin Balakrishnan,"Fluid Interaction Techniques for the Control and Annotation of Digital Video", ACM, 2003.

[7] Wei-Ting Peng1, Yueh-Hsuan Chiang2, Wei-Ta Chu2, Wei-Jia Huang2, Wei-Lun Chang2, Po-Chung Huang2, and Yi-Ping Hung, Aesthetics-Based Automatic Home Video Skimming System, Springer-Verlag Berlin Heidelberg 2007.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

[8] Changsheng Xu, Jian Cheng, Yi Zhang, Yifan Zhang, Hanqing Lu,"Sports Video Analysis: Semantics Extraction,Editorial Content Creation and Adaptation", Journal of Multimedia,2009.
[9] C.Vaucelle,G.Davenport, and T. Jehan, "Textablemovie: Improvising with a personal movie database," in *Proc. SIGGRAPH*, 2003.
[10] Pablo Cesar,"Interactivity and User Participation in the Television Lifecycle: Creating, Sharing, and Controlling Content", ACM, 2008.
[11] E. Y.-T. Shen, H. Lieberman, and G. Davenport, "What's next?: Emergent storytelling from video collection," in *Proc. CHI*, , pp. 809–818.
[12] Timur R. Almaev, Michel F. Valstar,"Local Gabor Binary Patterns from Three Orthogonal Planes for Automatic Facial Expression Recognition".
[13] Olivier Duchenne, Ivan Laptev, Josef Sivic, Francis Bach and Jean Ponce," Automatic Annotation of Human Actions in Video", INRIA.
[14] Wenbing Tao, Hai Jin, Yimin Zhang,"Color Image Segmentation Based on Mean Shift and Normalized Cuts", IEEE Transactions on Systems, 2007.
[15] S. Foucher, L. Gagnon,"Automatic Detection and Clustering of Actor Faces based on Spectral Clustering Techniques", 4th Canadian Conference on Computer and Robot Vision, Montreal, 2007.
[16] K.Mahesh and Dr.K.Kuppusamy,"A New Hybrid Video Segmentation Algorithm using Fuzzy C Means Clustering", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.
[17] Chao Liang, Changsheng Xu, Jian Cheng, Hanqing Lu," TVParser: An Automatic TV Video Parsing Method".
[18] Matthew B. Blaschko ,Andrea Vedaldi, Andrew Zisserman, "Simultaneous Object Detection and Ranking with Weak Supervision".
[19] Mark Everingham, Josef Sivic and Andrew Zisserman, "Hello! My name is... Buffy" – Automatic Naming of Characters in TV Video".
[20] Deva Ramanan, Simon Baker, Sham Kakade, "Leveraging archival video for building face datasets".
[21] O. Arandjelovic and R. Cipolla, "Automatic cast listing in featurelength films with anisotropic manifold space," in *Proc. CVPR*, , pp. 1513–1520.
[22] Chao Liang, Yifan Zhang, Jian Cheng, Changsheng Xu, and Hanqing Lu," A Novel Role-Based Movie Scene Segmentation Method", Springer-Verlag Berlin Heidelberg 2009.
[23] C. Liang, C. S. Xu, J. Cheng, and H. Q. Lu, "Tvparser: An automatic drama video parsing method," in *Proc. CVPR*, , pp. 3377–3384.