Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision, Malta, September 15-17, 2005 (pp82-87)

Video Summarization Based Handout Generation from Video Lectures: A Gesture Recognition Framework

HAFIZ ADNAN HABIB, MUID MUFTI
Computer Engineering Department
University of Engineering & Technology
Taxila, Punjab
PAKISTAN

Abstract: handouts from video lectures are one useful summarization for student's reference. Variety of summarization approaches are proposed in literature for specific applications. A hand gesture recognition based video summarization framework is proposed in this research paper to summarize video lectures in form of handouts. Three types of hand gestures are considered: writing hand gesture, idle/explaining hand gesture and erasing hand gesture. Video summarization for handout generation considers transition among these three types of gestures.

Keywords: video summarization, hand gesture, handout generation, video lectures, content based video processing.

Introduction to Problem:

Deriving compact representation from video sequences that are intuitive for users and that let them easily and quickly browse large collections of video data is fast becoming one of most important topics in content based video processing. Such representation which collectively refer to as video summaries, rapidly provide user with information about content of particular sequence being examined while preserving the essential message [1]. Video capturing systems produce massive data and automatic summarization of this massive data becomes necessary and challenging in certain circumstances. Video summarization is being incorporated to summarize videos in different circumstances: surveillance videos, sports videos, interview sessions, documentaries and entertainment production are to name a few.

Distance learning is evolving as common practice in today's world and massive video data is produced from recording of lecture proceedings and scientific presentations. Students consult these video lectures and gain knowledge. These video lectures are accompanied by power point presentations which give clue about contents covered in video lecture. Sometimes such power point presentations not included. For such cases, teachers make explanations on white board by drawing sketches or deriving equations or writing comments. These explanations on white board become part of video being recorded but these are not available as the case with power point presentation of lectures. In this research paper, these proceedings on white board are collectively called handouts for certain video lecture and are of great convenience for students if available for student accompanied with video lectures.

This problem is traditionally solved by placing a transparent slide over light projector illuminating some wall. Teacher make all explanation work upon

transparent slide. Once that transparent slide is filled, then new transparent slide is utilized. So handouts are easily developed by scanning those presentation slides and utilized by students as handouts of lecture. This traditional approach lacks due to certain reasons: separate light projector required for such operation and inconvenience for teacher as being rather unusual style of giving explanation to students. A scanner system is required for further scanning of written transparent slides.

Video Summarization System:
A good quality video capturing system focusing at white board and streaming video to connected PC or storing in system memory was required. Either offline or online processing can be incorporated for video lectures. Sony's CCD handy cam was used in this experimentation.

General Approaches to Video Summarization:
Most video content can be broadly categorized into two classes [1]: event based content and uniformly informative content. Event based content videos contain easily identifiable story units that form either a sequence of different events or a sequence of events and non-events. Best example of programs in which a sequence of events and non-events occur are sports programs. Here, the events may correspond to highlights such as touchdowns, home runs and goals. Uniformly content programs can't be broken down to a series of events as easily as event based content. For this type of content are sitcoms, presentation videos, documentaries and home movies.

For event based content, since type of events of interest are well defined, one can use knowledge based event detection techniques. In this case, processing will be domain specific and new set of events and event detection rules must be derived from each application domain. Summarization of soccer video is best example of this kind.

Gesture Recognition Based Video Summarization Framework:
Gesture recognition is involved in number of activities: augmented interfaces, virtual desk, human computer interaction, sign language recognition etc. Both static and dynamic gestures are analyzed by systems. By static gesture, particular shape is considered for particular meaning or action. American sign language recognition is one such example. Similarly there are dynamic gestures. Particular motion of gesture is meaningful for observing system.

In this research, a novel framework based upon gesture recognition is presented for summarization of video lectures in form of handouts. This framework involves three different type of gestures and is designed around three main stages. Three gestures are: Hand writing gesture, White board erasing gesture and idle gesture. Three main stages are: white board registration, gesture recognition and handout saving. Hand writing gesture is defined as gesture involved in writing activity. White board erasing gesture is involved in erasing contents upon white board. Idle gesture exists at transition b/w hand writing and white board erasing gesture. It is also considered for no activity situations.

Fig I. Writing Gesture.

Fig. II. Erasing Gesture

White Board Registration:
This is first stage for video summarization. White board is actually the area where writing activity has to be done. A shape based approach is considered for registration of white board. Shape extraction is performed by initially applying gradient operator of Sobel and then adaptive threshold is considered by moving small window over whole gradient image. Once white board is recognized, then it is considered as region of interest (ROI) for rest of analysis. It is restricted that writing gesture has to operate in this ROI. This restriction reduces computational requirements and produces stable results for gesture recognition.

Gesture Recognition:
Gesture recognition [20], [21] is main step in this framework. Two types of gestures are particularly analyzed and extracted. Hand writing gesture is defined as writing gesture upon white board. While erasing gesture erases contents upon white board. Both writing gesture and erasing gesture are defined as static postures. Gesture recognition is performed by technique mentioned in [19]. The mentioned algorithm is particularly specified for state based summarization and recognition of gesture

Generating Handout:
Refer to Fig. II. Handout generation process is accomplished in state based fashion. Final aim is to extract handout before white board contents are completely erased. This is main assumption for handout extraction that erase of complete white board means handout. Three states are defined: hand writing state, idle state and erasing state. Temporary handouts are taken while gesture is performing transition b/w writing state and idle state by continuously analyzing ROI in video stream. Whenever an erasing gesture is found, then system considers the amount of erase. Complete erase is considered as command to consider latest temporary handout as finalized handout.

Conclusion & Future Work:
Gesture recognition based approach proved very successful for automatic extraction of handouts as summarization of video lectures. This functionality is functioning at the institute.

In future directions, it is aimed to include power point transition information extraction directly from video sequence and handout extraction simultaneously. Both these processes should be carried in real time by computationally efficient design of algorithms.

Reference:

[1] Belle L. Tseng, Ching-Yung Lin, John R. Smith. "Video Summarization and Personalization for Pervasive Mobile Devices" In proceedings of SPIE Conference on Storage and Retrieval for Media Databases 2002, volume 4676, pages 359-370, San Jose CA, 23-25 January 2002.

[2] Nuno Vasconcelos, Andrew Lippman. "Baysian modeling of video editing and structure: semantic features for video summarization and browsing". In proceedings of IEEE international conference on Image Processing, Chicago, IL, 4-7 October 1998.

[3] Daniel DeMenthon, Vikrant Kobla, David Doerman. "Video Summarization by Curve Simplification". In proceedings of ACM Multimedia Conference, pages 211-218, Bristol, England, 12-16 September 1998.

[4] Shingo Uchihashi, Jonathan Foote, Andreas Girgenson, John Boreczky. "Video Manga: Generating semantically meaningful video summaries". In proceedings of ACM Multimedia Conference, pages 283-392, Orlando, October 5, 1999.

[5] A. Stefanidis, P. Partsinevelos, P. Agouris, P. Doucette. "summarizing video datasets in spatiotemporal domain" In proceedings of International workshop on Advanced Spatial Data Management, Greenwich, England, 6-7 September 2000.

[6] Jung Hwan Oh, Kien A. Hua.

[7] S.M. lacob, R. L. Lagendijk, and M. E. lacob. Video abstraction based on asymmetric similarity values. In Proceedings of SPIE Conference on Multimedia Storage and Archiving Systems IV, volume 3846, pages 181-191, Boston, MA, September 1999.

[8] Krishna Ratakonda, Ibrahim M. Sezan, and Regis J. Crinon. Hierarchical video summarization. In Proceedings of SPIE Conference Visual Communications and Image Processing, Volume 3653, pages 1531-1541, San Jose, CA, January 1999.

[9] MA. Ferman and A. M. Tekalp. Two-Stage hierarchical video summary extraction to match low-level user browsing preferences. IEEE Transactions on Multimedia, 5(2): 244-256, 2003.

[10] Dirk Farin, Wolfgang Effelsnberg, and Peter H. N. de With. Robust clustering based video summarization with integration of domain knowledge. In proceeding of the IEEE International Conference on Multimedia and Expo 2002, pages 89-92, Lausanne, Switzerland, 26-29 August 2002.

[11] Yossi Rubner, Leonidas Guibas, and Carl Tomasi, the earth mover's distance, multi dimensional scaling, and color based image retrieval. In Proceeding of the ARPA Image Understanding Workshop, pages 661-668, New Orleans, LA, May 1997.

[12] Yahiaoui, B. Merialdo, and B. Huet. Comparison of multi episode video summarization algorithms. EURASIP Journal on Applied Signal Processing, 3(1): 48-55, 2003.

[13] Yihong Gong and Xin Liu. Video summarization using singular value decomposition. In Proceedings of the IEEE Conference on Computer Vision and Pattern position. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Hilton Head Island, SC, pages 174-180, 13-15 June 2000.

[14] Matthew Cooper and Jonathan Foote. Summarizing video using non-negative similarity matrix factorization. In International Workshop on Multimedia Signal Prcessing, ST. Thomas, Us Virgin Islands, 9-11 December 2002.

[15] Ahmet Ekin and A. Murat Tekalp. Automatic soccer video analysis and summarization. In Proceedings of SPIE Conference on Storage and Retrieval for Media Databases 2003, volume 5021, pages 339-350, Santa Clara, Ca, 20-24 January 2003.

[16] Baoxin Li, hao Pan, and Ibrahim Sezan. A general framework for sports video summarization with its application to soccer. In Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, pages 169-172, Hong Kong, 6-10 April 2003.

[17] Romain Cabasson and Ajay Divakaran. Automatic extraction of soccer video highlights using a combination of motion and audio features. In Proceedings of SPIE Conference on Storage and Retrieval for Media Datanbases 2003, Volume 5021, PAGES 272-276, Santa Clara, CA, 20-24 January 2003.

[18] Howard D. Wactlar. Informedia search and summarization in the video medium. In Proceedings of the IMAGINA 2000 Conference, Monaco, 31 January-5 February 2000.

[19] Aaron F. Bobick, Andrew D. Wilson, “A state-based Technique for the summarization and recognition of gesture” 1995 IEEE Conference Proceeding.

[20] Vladmir I. Pavlovic, Rajeev Sharma, Thomas S. Haung. “Visual interpretation of hand gestures for human computer interaction: a review” IEEE transactions on Pattern Analysis and Machine Intelligence Vol. 19. No. 7. July 1997.

[21] Beifang Yi, Frederick C. Harris, Ling Wang, Yusan. “Real time natural hand gestures” IEEE CS and AIP 2005.

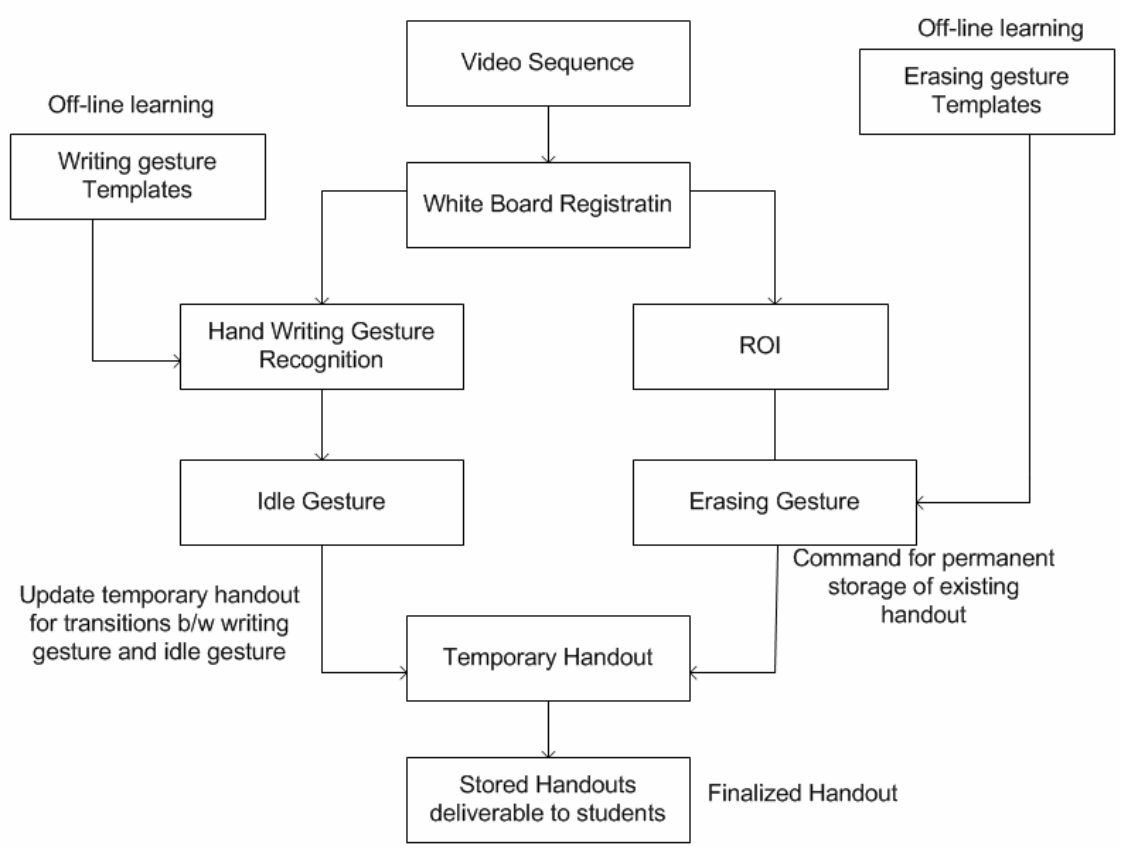

Fig III. Gesture Recognition framework for Summarization of Video Lectures for Handout generation.