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.
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:


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