The Automatic Video Editor

Sam Yip
Entertainment Technology Center
Carnegie Mellon University
Pittsburgh, PA 15213
+1 412 268 6528
syip@andrew.cmu.edu

Eugenia Leu
Entertainment Technology Center
Carnegie Mellon University
Pittsburgh, PA 15213
+1 412 268 6528
eyj@andrew.cmu.edu

Hunter Howe
Entertainment Technology Center
Carnegie Mellon University
Pittsburgh, PA 15213
+1 412 268 6528
hhowe@andrew.cmu.edu

ABSTRACT
More and more home videos are being produced with the increasing popularity of digital video camcorders. Yet the resulting home videos tend to be very long and boring to watch. The precious memories within those videos are ultimately lost. The problem is that the average home videographer does not have the time, or the editing skills to edit their home videos. It is a shame to let all those precious moments go to waste.

There are editing software in the market that allow the user to edit their own home videos. For example, Apple's iMovie [1], Microsoft's MovieMaker [2], and the Hitchcock editing system [3]. But they still demand time, skill and effort from the user.

With the above issue in mind, we developed the Automatic Video Editor, an application that analyzes a home video and edit it automatically into a condensed and interesting mini-movie.

In our video, we will first show a clip of an unedited home video. Afterwards, we will show the output video created by our Automatic Video Editor.

1. SYSTEM OVERVIEW
The Automatic Video Editor takes in two inputs: an original home video and a music track. First, using our video processing system, we analyze the video to discover and mark a set of desirable visual characteristics within it. Then, using our editing rules, the application will automatically construct a new output video. Let us discuss the editing rules in more detail.

2. EDITING RULES

2.1 Contrast
Filmmakers have always manipulated the visual elements on screen to intensify the viewing experience. In *The Visual Story* [4], filmmaker Bruce Block dissected the structure in video into the visual components. The visual components are space, line, shape, tone, color, movement, and rhythm. By contrasting these visual components, the resulting video will be more intense and interesting. Figure 1 illustrates a contrast in color. A video clip that is predominantly red will contrast with a video clip that is predominantly blue. Detailed descriptions of these visual components are described further in the book [4].

MM'03, November 2-8, 2003, Berkeley, California, USA.
ACM 1-58113-722-2/03/0011.

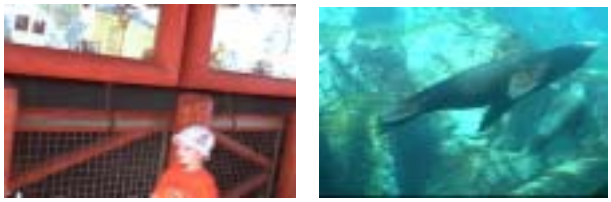

Figure 1. Scenes that Contrast in Color

Using this concept of contrasting visual components as one of our editing rules, the Automatic Video Editor will pick out segments from the source video that contrast with each other the most.

2.2 Desirable Visuals
There are other characteristics that are desirable in a video. For example, we prefer to see videos that are stable and not jerky. Therefore, our system will disregard all the shots with jerky camera movements. We also allow the user to tweak the desirability levels of three characteristics in a video. The three characteristics are 1) the amount of faces in a video; 2) the amount of close ups in the video; 3) the amount indoor/outdoor scenes in the video. For example, if you want to see a lot of outdoor shots in your output video, then you can set the outdoor desirability level higher, and the Automatic Video Editor will insert more shots containing outdoor scenes into the output video. We also provide a pre-set level for all these characteristics.

2.3 Video Processing System
Before we can apply our editing rules to the input home video, the application needs to detect the visual characteristics in the input video. The Informedia processing system [5] at Carnegie Mellon University provides us with the ability to automatically identify various visual characteristics within a source video. Currently, the Informedia system can detect camera zooms, camera pans, hard cuts, faces, indoor and outdoor scenes. For the other visual characteristics that we want to identify in the video, such as color, space and tone, we detect them manually and record them using an application that we have developed. We predict that in the near future, detection of all the visual characteristics that we need can be done automatically and the entire video processing step in our system will not require human intervention.

2.4 Special Effects
Learning from expert human editors and observing movie trailers of the action genre, we found that certain special effects make the video more fun to watch. Therefore we added the Picture in Picture (PiP), the slow-motion, and the fast-motion effects into

our output video. The effects cannot occur too often because it will be irritating and jarring. Therefore, they are triggered by the frequency of the music beats. As the frequency of music reaches a high point, which is usually the climax of the song, the special effects are triggered to start.

We also learned that the contents within the Picture in Picture must be carefully chosen. As illustrated in Figure 2, we do not want the PiP window to block a person's face because that is irritating for the viewer. Also, the PiP window should not look too convoluted. As Figure 3 illustrates, a crowd of people will create a messy, undesirable composition, because the viewer is confused as to what they are looking at.

Figure 2. Undesirable Pip composition

Figure 3. Undesirable Pip composition

Therefore, when we construct the PiP effect, we chose clips which only contain one face to be in the foreground, and we chose clips which contains outdoor scenes, without many people, to be in the background.

2.5 Sound

The music track provides us with the editing rhythm as well as sets the final length of the output video. As mentioned above, the frequency of the beats also trigger the special effect to appear. The marking of when the beats occur is done manually using an application that we developed. There are commercial applications that can detect song beats automatically. In the future, we plan to incorporate the automatic beat detection into our system.

3. VIDEO SCENARIO

In the video, we show a sample output video from our Automatic Video Editor within a family vacation scenario. The family returns from a vacation with a lot of unedited home videos. We will see a short sample of those unedited home videos and then we will show the output video created by the Automatic Video Editor. We conducted evaluations with dozens of Carnegie Mellon students on similar trip videos, learning from those evaluations the importance of composition rules, music, and faces in producing trailers.

The evaluations also pointed out the need to vary the settings of the number of on screen faces based on the source material and the audience. A video of a lunch party full of people was found to be interesting with lots of face shots, but an aquarium video with many face shots was found to be interesting only by those who were in the video. Others who were not in the video preferred to see more fish and tank shots and less people shots.

4. FUTURE WORKS

This project spawned a lot of new ideas that would be interesting to pursue in the future. One idea is a video editing game similar to the game Frequency [6] which will allow the player to edit a video and play a video game at the same time. Another idea is to create a home video that is also a karaoke.

5. SUMMARY

Our video shows an example of our Automatic Video Editor turning an unedited home video into a more condensed and interesting video.

6. ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation under Cooperative Agreement No. IRI-9817496.

Our thanks to the Informedia Project and Entertainment Technology Center for giving guidance and advice to this project.

7. REFERENCES

[1] Apple. iMovie. http://www.apple.com/imovie/ (2003)

[2] Microsoft. Movie Maker. http://www.microsoft.com/windowsxp/moviemaker/ (2003)

[3] John Doherty, Lynn Wilcox, Andreas Girgensohn. A Hitchcock Assisted Video Edited Night at the Opera. In ACM Multimedia Proceedings(Juan-les-Pins, France, 2002), ACM Press, pp. 660-661

[4] Block, Bruce, A. The Visual Story: Seeing the Structure of Film, TV and New Media. Focal Press, (April 2001).

[5] Informedia. Informedia processing system. http://www.informedia.cs.cmu.edu/ (2003)

[6] Harmonix. FreQuency. http://www.harmonixmusic.com/freqoverview.html (2001)