5-2006

A Study of Video Motion and Scene Complexity

Huahui Wu

A Study of Video Motion and Scene Complexity

Huahui Wu, Mark Claypool and Robert Kinicki
(flashine, claypool, rek)@cs.wpi.edu
Computer Science Department
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609

1. Introduction

Ph.D. research [1] shows that the motion and scene complexity characteristics of video content have a significant impact on the quality of streaming video. For example, for a video that has low motion and low scene complexity, its requirement of streaming bandwidth is low and the streaming application does not need to reduce its bitrate as much as the bitrate of a video that has high motion and high scene complexity. Besides, the motion and scene complexity also have impact on the choice of the right scaling methods and right repair amount.

Video motion and scene complexity characteristics are studied in this report. In particular, 9 different video clips are encoded to MPEG files and the MPEG files are analyzed with statistics measurements. The results of different measurements are compared with a 3-person preliminary user study.

The rest of this report is organized as follows. Section 2 describes experiment settings, Section 3 and Section 4 study video motion and scene complexity characteristics, respectively, and Section 5 summarizes the report.

2. Experiment Settings

Video	Description
Container	A working container ship
Hall	A hallway
News	Two news reporters
Foreman	A talking foreman
Paris	Two people talking with high-motion gestures
Silent	A person demonstrating sign language
Coastguard	Panning of a moving coastguard ship
[1]Mobile	Panning of moving toys
Vectra	Panning of a moving car

Table 1: Video Clips

Nine video clips are used for the experiments. Each video clip has 300 raw images and the size of each frame is 352x288 pixels (CIF). Table 1 provides an identifying

[1] http://bmrc.berkeley.edu/frame/research/mpeg/

name and a short description of the video content. All the experiments use the Berkeley MPEG encoder and decoder[1] on Linux. However, the results should hold for other MPEG encoders since the choice of encoder has little impact on compression relative to the impact on compression due to choice of quantization level and GOP pattern. The quantization values for I, P and B frames are all 3 to yield a high picture quality in every frame. A commonly-used MPEG GOP pattern, 'IBBPBBPBBPBBPBB', and a typical full motion frame rate of 30 frames per second (fps) are used.

The encoded MPEG files are then analyzed with the Berkeley mpeg_stat tools. With the option "-block_info file", all the macroblock information, including block number, frame type (I/P/B), quantization value, block size, block type (skip / intra / 0_motion / motion), and motion vector (if applicable) will be output into an ASCII file. This block information file is then analyzed with customized C++ programs or Linux text tools such as "grep" and "awk".

3. Video Motion

3.1 Preliminary User Scores

Video	User1	User2	User3	Average	
Coastguard	1.00	0.88	1.00	0.96	High
Container	0.48	0.25	0.78	0.51	Low
Foreman	0.97	0.72	0.86	0.85	Medium
Hall	0.38	0.28	0.23	0.30	Low
Mobile	1.00	1.00	1.00	1.00	High
News	0.39	0.41	0.25	0.35	Low
Paris	0.67	0.50	0.63	0.60	Medium
Silent	0.66	0.69	0.67	0.67	Medium
Vectra	1.00	1.00	1.00	1.00	High

Table 2: Preliminary User Scores for Motion

For each clip, each of three users divides the frame into 16 equal blocks and counts the motion characteristic of each block during the clip. For each block, the user can rate the motion in that block with 5 scales: 0, ¼, ½, ¾ to 1, where 0 means no motion and 1 means lots of motion, considering both temporal and spatial domain. A simple example would be, if half of a block changes during half of the playout period, this block's motion should be rated as ¼. Later, the scores of 16 blocks are added to get the motion score of the video clip. Table 2 shows the normalized (divided by 16) user scores and the average scores of three users.

3.2 Perceived Motion Energy Spectrum (PMES)

Ma et al. [2] proposed a shot based motion energy representation, namely, perceived motion energy spectrum (PMES) image, obtained by a temporal energy filter and a global motion filter. In this method, a temporal energy filter is designed to eliminate disregarded object motion in a scene, and a global motion filter is designed to shield object motions from the camera motions. Figure 1 shows the comparison of PMES score and user score. Visually, the correlation is low.

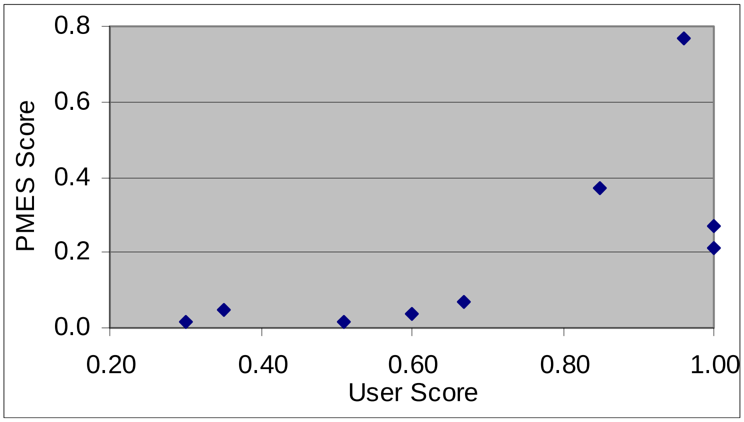

Figure 1 PMES Score versus User Score

3.3 Percentage of Interpolated Macroblocks in B Frames (PIMB)

Tripathi *et al.* [3] uses the percentage of interpolated Macroblock in B frames (PIMB) as a measure of motion. The author suggests, a high number of interpolated macroblocks implies that a greater portion of the frame is similar to frames that are already existing in the stream (i.e. less motion) and a low number of interpolated macroblocks implies that there are a greater number of changes between frames (i.e. more motion). Figure 2 shows the comparison between PIMB scores and user scores. Visually the correlation is not high.

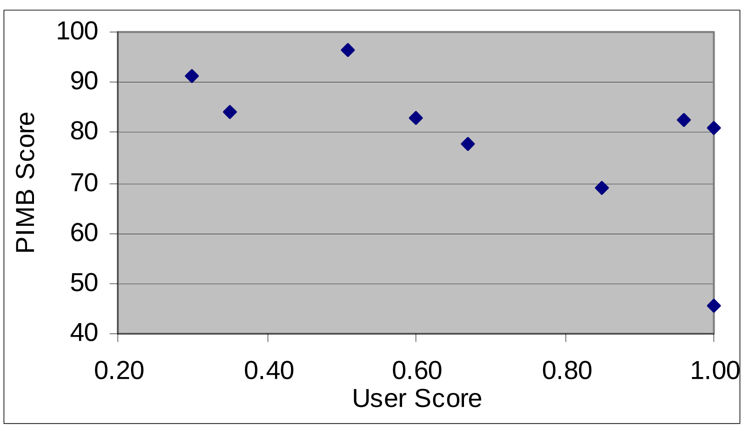

Figure 2: PIMB Score versus User Score

3.4 Percentage of Forward or Intra-coded Macroblocks (PFIM)

We believe a better measure of motion is the percentage of encoded macroblocks, i.e. the percentage of Forward or Intra-coded Macroblocks (PFIM) in all the frames.

Figure 3 shows the comparison of PFIM score and user score. Visually the correlation is significant.

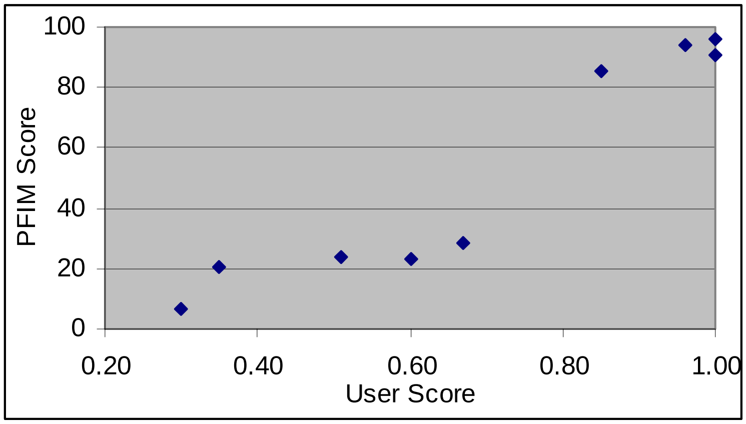

Figure 3: PFIM Score versus User Score

4. Video Scene Complexity

4.1 Preliminary User Scores

Video	User 1	User 2	User 3	Average	
Coastguard	0.56	0.81	0.66	0.63	Medium
Container	0.47	0.59	0.63	0.61	Medium
Foreman	0.72	0.69	1.00	0.96	High
Hall	0.69	0.53	0.39	0.48	Low
Mobile	1.00	1.00	1.00	0.91	High
News	0.63	0.59	0.53	0.65	Medium
Paris	0.94	0.94	0.75	0.86	High
Silent	0.81	0.91	0.77	0.71	Medium
Vectra	0.94	0.91	0.84	0.95	High

Table 3: Preliminary User Scores for Scene Complexity

For each clip, each of three users divides the frame into 16 equal blocks and counts the scene complexity of each block during the clip. For each block, the user can rate the motion in that block with 5 scales: 0, ¼, ½, ¾ to 1, where 0 means blank screen and 1 means extremely complicate scene. A simple example would be, if half of a block is complicate, this block's scene complexity should be rated as ½. Later, the scores of 16 blocks are added to get the scene complexity score of the video clip. Table 2 shows the normalized (divided by 16) user scores and the average scores of three users.

4.2 Average Intra-coded Block Size (IBS)

We use the average of intra-coded block size as a measure of scene complexity. If the scene is simple, there is not much information to be encoded, so the intra-coded block size will be small. If the scene is complicated, the intra-coded block size should be large to contain all the information. Figure 4 shows the results with some correlation, but not as much as expected.

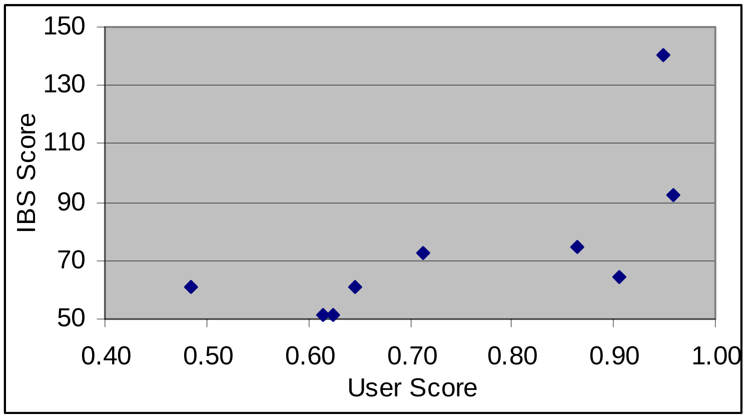

Figure 4: IBS Score versus User Score

5. Summary

This report describes a study of video motion and scene complexity characteristics. Preliminary user studies were conducted to rate the subjective scores. Different objective metrics are measured and compared with the subjective scores. The results show that the proposed metric *Percentage of Forward or Intra-coded Macroblocks (PFIM)* is highly correlated with the user's score of motion characteristics while the proposed metric *Average Intra-coded Block Size (IBS)* has a more modest correlation with user's score of scene complexity.

Reference:

[1] H. Wu. ARMOR – Adjusting Repair and Media Scaling with Operations Research for Streaming Video. *Ph.D. Thesis*, Worcester Polytechnic Institute, May 2006.

[2] Y.F. Ma and H.J. Zhang. A New Perceived Motion based Shot Content Representation. In *Proceedings of International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, October 2001.

[3] A. Tripathi and M. Claypool. Improving Multimedia Streaming with Content-Aware Video Scaling. In *Proceedings of the Second International Workshop on Intelligent Multimedia Computing and Networking (IMMCN)*, Durham, North Carolina, USA, March 2002.