SUBJECTIVE EVALUATION OF MULTI-USER RATE ALLOCATION FOR STREAMING HETEROGENEOUS VIDEO CONTENTS OVER WIRELESS NETWORKS

Xiaoqing Zhu and Bernd Girod

Information Systems Laboratory, Stanford University, Stanford, CA 94305, USA
{zhuxq,bgirod}@stanford.edu

ABSTRACT

Multi-user rate allocation is essential for preventing multiple simultaneous video streaming sessions from congesting a time-shared wireless network. Although the problem fits naturally within an optimization framework, it remains unclear which objective function to optimize, especially for video streams with heterogeneous resolutions, frame rates and content complexity. To address this issue, we have conducted subjective tests and have collected viewers’ opinion scores for various quality combinations of high-definition (HD) and standard-definition (SD) video snapshots. The test results yield a simple subjective quality model, useful for comparing performance of several rate allocation schemes. It is shown that the subjectively preferred allocation can be closely approximated by minimizing weighted or equal-weight sum of mean-squared-error (MSE) distortion of all streams. Significant performance gains can be achieved over media-unaware TCP-friendly allocation, up to 1.5 in MOS, when network resources are limited.

Index Terms— subjective test, rate allocation, multi-user video streaming, wireless networks

1. INTRODUCTION

Video streaming over wireless networks is compelling for many applications, ranging from surveillance camera networking to interconnecting home entertainment devices. Rate allocation becomes an important task in the presence of multiple simultaneous streaming sessions. Since the video streams time-share the common wireless radio channel, their rates need to be jointly allocated even when there is no shared link in their routes. The rate allocation problem is further complicated by heterogeneity in both the video contents and wireless link capacities, as pointed out in [1] and [2].

Many research efforts have been dedicated to joint rate adaptation of multiple video streams based on various quality metrics such as mean-squared-error (MSE) video distortion [3] and mean-opinion-score (MOS) [4]. In general, it remains to be explored which objective function to optimize, to better reflect human preference in a realistic application scenario. The work in [1] has shown that minimizing MSE distortion is the subjectively preferred criteria when video streams of the same resolution time-share a wireless LAN. However, this may not necessarily carry over to video streams with heterogeneous resolutions and frame rates.

Consider, for instance, one high-definition (HD) video stream competing with one standard-definition (SD) video stream over two wireless links, as in Fig. 1. Should the rate allocation scheme maximize total MSE distortion of both streams? Or, rather, should one account for their resolution differences and assign different weight to the distortion of each stream? In the latter case, how much weight should be assigned to the HD and SD stream, respectively?

This paper addresses the above questions. We conduct subjective viewing tests of pairs of image snapshots from HD and SD video streams in various quality combinations. Analysis of the viewers’ opinion scores, collected from 20 participants for each of the 4 sequence pairs, leads to a simple subjective quality model. It is shown that mean-opinion-score (MOS) of an image pair can be approximated as bilinear function of individual image MOS, which, in turn, fits linearly to MSE distortion of that image. This allows subjective evaluation of several rate allocation schemes, including media-aware schemes that minimize weighted or equal-weight sum of MSE distortions, and media-unaware scheme with TCP-friendly allocation.

The rest of the paper is organized as follows. The next section describes the subjective viewing test. In Section 3, we analyze the test results, present the subjective quality model and compare subjective evaluation of various rate allocation schemes.

2. SUBJECTIVE VIEWING TEST

Subjective viewing tests are conducted to collect viewers’ opinion scores of various quality combinations of two video sequences. For faster testing time, different video quality levels are presented using still image snapshots instead of continuously playing video segments. Each pair of test images are displayed side-by-side, on two 24-inch Samsung Syncmaster 245BW LCD monitors at a distance...
of 1.5 m from the viewer. Potential inconsistencies between the two monitors are mitigated by randomizing which monitor displays which image. The testing room is kept dark, as recommended by [5]. Figure 2 illustrates the test setup.

Each test involves one HD video sequence with a resolution of 1280 × 720 pixels per frame and 60 frames per second (fps), and one SD video sequence at 704 × 576 pixels per frame and 30 frames per second. Four test data sets are chosen to cover a range of content complexity. The rate-PSNR tradeoff curves of the HD/SD video pairs are plotted in Fig. 3.

Each video is represented in 5 quality levels, corresponding to the uncompressed version, and compressed versions with quantization parameters (QP) at 24, 34, 40 and 44, using x264, a fast implementation of the H.264/AVC codec [6][7]. In addition to the 25 HD/SD image pairs, each test data set also contains pairs of identical HD or SD images on both displays, for all 5 quality levels. Scores collected from these 10 image pairs are used to calculate MOS of individual HD or SD images.

Each test comprises of one training session and two scoring sessions. The training session contains pairs of HD and SD images at the highest, intermediate and lowest quality levels, as highlighted in Fig. 4. The purpose of the training session is to familiarize the viewer with the expected range of perceived image qualities. No scores are collected at this stage.

In the scoring session, the viewer provides an opinion score for each presented image pair, based on the perceived overall quality of both images. The opinion scores range from 1 - 5, where a score of 1 means “bad” and a score of 5 means “excellent”. Fractional scores are allowed between adjacent description levels. All 35 image pairs are presented in random orders, to mitigate the contextual effect [5]. The viewers are allowed sufficient time to observe and score each image pair. Each data set is presented to 20 participants.

3. TEST RESULTS

3.1. Subjective Quality Model

Figure 5 shows the mean opinion score (MOS) of each HD or SD image, averaged over 40 readings from 20 participants in each test, as a function of its mean-squared-error (MSE) distortion. The standard deviations of the MOS values are also plotted in the form of error bars. It can be observed that MOS of each image can be fitted as a linear function of its MSE distortion:

\[
X = X_0 - \alpha D_{HD}
\]
\[
Y = Y_0 - \beta D_{SD}.
\]

The parameters \(X_0\) and \(Y_0\) correspond to the highest MOS for uncompressed images, and are around 4.5 for all images. The slope of each curve, \(\alpha\) or \(\beta\), is content-dependent, and can be calculated by looking up MSE distortion of images with the lowest MOS, typically around 1.5. The fitting curves are shown as dotted lines in the same figure. Interestingly, MOS values of images containing many complex details tend to be less sensitive to changes in MSE distortion, as in Fountain and Harbor, while MOS values of images containing simple scenes and smooth regions tend to vary more rapidly with MSE distortion, as in Ice and Raven. This can be explained by the spatial masking effect in the human vision system [8].
In addition, MOS of an image pair can be fitted to a bilinear function of individual image MOS:

\[ Z = w_0 + w_1 X + w_2 Y + w_3 XY \]  

In (3), \( Z \) denotes MOS of the image pair; \( X \) and \( Y \) are MOS values of the HD and SD image, respectively. Surprisingly, a common set of parameters \( w_0 - w_3 \) can be used to fit all 4 data sets despite their content differences. The maximum fitting error is 0.24 in MOS, comparable to the standard deviation of the collected scores. Figure 6 shows the contour plot of this fitted function. As expected, MOS of the image pair increases with MOS of either image, and is slightly higher with more balanced qualities of both images. We note that the weights for the HD and SD images are similar even though their areas on the display differ by a factor of 2.3.

Combining (1) - (3) and the video distortion-rate tradeoff described by the following parametric model [9]:  

\[ D(R) = D_0 + \frac{\theta}{R - R_0} \]  

we can derive the subjective opinion score as a function of the given rate allocation. Figure 7 shows the contour plots of such functions, which are content-dependent.

### 3.2. Comparison of Rate Allocation Schemes

We now consider rate allocation for two competing streams over parallel wireless links, as in Fig. 1. Given the video sequence pairs, wireless link capacities and total utilization limit \( \gamma \), different rate allocation schemes would result in different rate pairs \( (R_{HD}, R_{SD}) \) along the tradeoff line:

\[ \frac{R_{HD}}{C_1} + \frac{R_{SD}}{C_2} = \gamma \]  

Figure 8 shows the MOS derived from the subjective quality model along the tradeoff line, for the sequence pair Dijana vs. Crew. Different lines indicate different values of \( \gamma \). Note that the MOS values become more sensitive to rate allocation results with lower values of \( C_2 \), i.e., more stringent network resources. Results from four allocation schemes are also marked on the graph:

- **MAX-MOS**: allocation by maximizing the MOS value derived from the subjective quality model, marked in red "o" signs. This corresponds to the subjectively preferred allocation, and serves as an upper bound of performance.
- **MIN-WMSE**: allocation by minimizing weighted sum of MSE distortion of both streams, marked in purple "x" signs. The weights for the HD and SD streams are chosen as \( \alpha \) and \( \beta \) from (1) and (2).
- **MIN-MSE**: allocation by minimizing total MSE distortion of both sequences, marked in blue "*" signs.
- **TCP-Friendly**: allocation based on TCP-Friendly Rate Control (TFRC) [10], marked in green "+" signs. Both streams are assigned equal rates, since they observe same round-trip-time and packet loss ratios over the common wireless network.

From Fig. 8, it can be noted that the MIN-WMSE scheme closely approximates the subjectively preferred allocation. The MIN-MSE scheme tends to allocate slightly higher rate for the HD stream due to mismatch in the weights, yet still achieves close-to-optimal MOS. In contrast, the media-unaware TCP-friendly allocation results in significantly lower MOS values when network resources are limited.
This paper presents a subjective viewing test designed to evaluate rate allocation results for streaming heterogeneous video content over wireless networks. We collect opinion scores from 20 participants for each of the 4 HD/SD video pairs, presented in the form of snapshots in various quality combinations. A simple subjective quality model is derived from the test results. It is shown that MOS of an image pair can be fitted as a bilinear function of individual image MOS, with a common set of parameters irrespective of image content. In addition, the subjective MOS value relates linearly with the objective MSE distortion of individual HD or SD images. The model is then used to compare the performance of several rate allocation schemes. The subjectively preferred allocation can be closely approximated by both MIN-WMSE and MIN-MSE schemes, whereas media-unaware TCP-friendly allocation may incur significantly suboptimal allocation results when network resources are limited. Despite simplicity of the test scenario, we believe that the insights obtained from this work may apply to more general conditions, in designing rate allocation schemes for streaming heterogeneous video contents over wireless.

5. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Joyce Farrell for many insightful discussions on the test design. We thank Dr. Peter van Beek at Sharp Labs of America, Dr. Erwin Bellers at NXP Semiconductors and Dr. Jianhua Zheng at HiSilicon Technologies for providing some of the test video sequences.

6. REFERENCES