Evaluation of Filtering Mechanisms for MPEG Video Communications

S. Sedigh, J. Joshi, A. Bashandy, A. Ghafoor

School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907

Abstract

In this paper, we evaluate two video filtering mechanisms for MPEG-1 video, namely, low pass filtering and selective frame dropping. The evaluation provides tradeoffs between the reduction achieved in bandwidth requirements and the perceptual quality of the video sequences delivered to the client. Extensive experiments revealed that these filtering mechanisms result in a significant reduction in bandwidth requirements, while maintaining acceptable perceptual quality.

1. Introduction

With the growth of the Internet and advances in video coding techniques, applications such as video conferencing, Internet TV, medical applications, and Internet games have become increasingly popular. Most of these applications are intended for use in a heterogeneous networking environment such as the Internet, which is built on links, routers, and workstations with vastly different capabilities. The quality requirements of these applications are generally specified by a set of *Quality of Service* (QoS) parameters 0. Applications requiring video data typically involve massive amounts of data communication, leading to the need for efficient video compression and transmission techniques 0. Several compression techniques are currently in use, including JPEG for full color image compression, Px64 or H.261 [3] for video-based communications, and MPEG for full-motion video applications. The MPEG standard is the most widely used video compression technique, which allows high compression ratios. However, data intensive applications such as telemedicine demand high bandwidth and computational power at the client side to guarantee QoS. A number of filtering mechanisms can be employed to reduce the data volume at the time of communication, resulting in lower transmission and processing requirements. On the other hand, filtering can result in degradation of the perceptual quality of the video delivered to the client side. To objectively evaluate the quality of filtered video sequences, quality metrics based on models of the human visual system can be used.

In this paper, we evaluate two video filtering mechanisms for compressed video sequences, namely, low pass filtering and selective frame dropping [4]. Low pass filtering is based on dropping high frequency components of the *Discrete Cosine Transform* (DCT) of the video sequence. In selective frame dropping, certain frames are dropped during the course of transmission. As MPEG-1 is an established video compression technique, this study is focused on MPEG-1 coded video data. The evaluation of the filtering schemes provides a tradeoff between the reduction achieved in end-to-end bandwidth requirements and the effect of filtering on the perceptual quality of video sequences delivered to the client. To evaluate the impact of filtering on the perceptual quality of filtered video sequences, the *Institute for Telecommunication Science* (ITS) video quality metric [5] is used. It is demonstrated that these filtering mechanisms result in a significant reduction in network requirements, while maintaining acceptable perceptual quality.

2. Filtering mechanisms

Various filtering mechanisms have been proposed for MPEG streams, including hierarchical filtering, mixing, and dropping [10]. Hierarchical filters can be used for processing substreams in a video data stream, each of which can be associated with a particular QoS and handled independently. These substreams may take different Internet routes, each of which may have distinct network QoS parameters. Internet links of different capacity can utilize this feature to increase the end-to-end transmission rate. The splitting filter splits a single stream into multiple substreams. The splitting can be based on frame type. For example, an MPEG stream can be partitioned into substreams containing only I-, P- or B-frames. Video stream splitting can also be implemented in the frequency domain. In this case, a stream is split into its low and high frequency coefficients. Mixing filters are used to multiplex substreams, e.g., mixing an n-person audio conference into one flow. These filters are suitable for video conferencing. Other filtering mechanisms employ a variety of schemes, including selective frame dropping and low pass filtering. The frame dropping scheme is the so-called "importance-oriented" approach, where frames with minute impact on the quality are dropped to achieve a reduction in frame rate and the Internet bandwidth. Dropping filters can be used to serve clients connected over lower capacity links. Low pass filters discard the higher frequency coefficients of each block in a frame, thus reducing the resolution and bitrate of the frames while maintaining the frame rate. As processing of the compressed bitstream is a computationally intensive task, low pass filters may be more effectively used by decoding the MPEG video stream prior to processing, and encoding the processed stream prior to transmission. Codec filters perform all video processing steps, including compression, decompression, encoding and color translation, in one phase.

In our experiment, we have focused on two filtering mechanisms, low pass filtering, and selective frame dropping. For the purpose of evaluation, two different low pass filtering schemes have been implemented for video sequences that have a mixture of I-, B-, and P-frames. Various levels of selective frame dropping are also evaluated.

Low pass filtering is one of the simplest approaches to filtering video data. In this mechanism, higher frequency components of every macro-block of selected frames in the video sequence are discarded. Low pass filters drop all the DCT coefficients above a particular cutoff value, say p, where p = 1,2,3,4,6,8,12,16,20,24,28,32,36,40, 44, 48. The choice of p is governed by the arrangement of the coefficients in a zigzag order, where lower frequency components lie closer to the DC component.

Two different filtering schemes can be employed for low pass filtering. In the first scheme, DCT coefficients are dropped from all of the three frame types. In the second scheme, coefficients are dropped from only P and B frames.

For the experiment, four sample videos, including *Fly*, *Space*, *Hula* and *Pirates*, were used. Two videos, namely, *Fly* and *Space* contained all three frame types, thus both of the low pass filtering schemes described above were applied to these two videos. Both videos have approximately the same resolution, although Space has less motion artifacts. As the other two videos, *Hula* and *Pirates*, have only I-frames, only the first low pass filtering scheme is applied to them.

In selective dropping filtering mechanism, selected frames are dropped from of the bit stream according to a prespecified dropping rule. In general, this mechanism is implemented by dropping one frame out of every fixed number of frames. The number of frames between two dropped frames is generally constant, and referred to as the "*skip distance.*" As some frames are more important than others, two different dropping schemes have been considered in the experiment. One scheme drops only B-frames, and the other drops both I and B-frames.

The frame dropping mechanism is characterized by its simplicity and low computational complexity, which facilitates software implementation on existing servers and workstations [4].

3. Video Quality Assessment

Video quality assessment plays an important role in the development of new and existing filtering schemes. Image quality assessment has been a research topic for several decades. Various image quality metrics have been developed, such as the Strehl measure [11], or measures of image similarity based on the human visual system (HVS) [8]. In assessing video quality the temporal effects of degradation, as well as the spatial effects, need to be evaluated. International committees, such as CCIR and MPEG have been active in the formulation of standards for the assessment of video quality.

The video quality metric used in an experiment should be selected carefully to yield accurate results with a reasonable amount of computation. One of the most commonly used quality measures is the quantitative video quality metric developed at the *Institute for Telecommunication Science* (ITS) [5]. This metric was chosen for our analysis, due to its relative accuracy and ease of computation. An overview of the ITS system is presented in the following section.

The ITS video quality assessment system is designed to emulate human perception, and provides results that agree closely with quality judgments made by a large panel of viewers. In this system, the quantitative measure is a linear combination of three quality impairment measures, m_1, m_2, and m_3. m_1 is a measure of spatial distortion, and represents the spatial differences between the original and received video frames. m_2 and m_3 are measures of temporal distortion. m_2 considers the global temporal distortion of the video sequence at the receiving site, whereas m_3 considers the frame with the most temporal distortion, or in other words, "jerkiness", in the entire video sequence.

These quantitative measures are based on two features, namely the *spatial information (SI)* feature and the *temporal information (TI)* feature. The spatial information feature of a frame is given by:

$$SI[F_n] = STD_{space}\{Sobel[F_n]\}$$

where STD_{space} is the standard deviation operator over the horizontal and vertical spatial dimensions of a frame. F_n is the nth frame in the video sequence, and *Sobel* is a high pass filtering operator used for edge detection. In essence, we obtain a sequence of numbers, where each number corresponds to the spatial distortion associated with an individual frame received at the client side in comparison with the original video frame.

The overall quality measure s is given as

$$s = 4.77 - 0.992m_1 - 0.27m_2 - 0.356m_3$$

For the development of this measure, a set of subjective tests were conducted by ITS in accordance with CCIR Recommendation 500-3 [12]. These recommendations specify viewing conditions, rating scales, etc. For this experiment, viewers were shown a number of original and degraded video pairs, and asked to rate the difference between the original video and degraded video on a scale of 1 to 5, with the following interpretations: (1) very annoying, (2) slightly annoying, (3) annoying, (4) perceptible but not annoying, (5) imperceptible.

The objective results obtained by the quantitative measure were subsequently compared to the subjective assessment done by the panel of viewers. The standard deviation of the error between the subjective and estimated scores was 0.4 units of the impairment scale of 1 to 5, thus differences in perceptual quality below 0.4 should be considered insignificant.

4. Experiments and Results

For our experiment, four MPEG streams of varying sizes and frame composition were used. In the *Fly* video, only a portion of the image continuously changes. In this respect, *Fly* is similar to sequences used in typical videoconferencing or medical sonogram data. *Space* is similar to *Fly* in resolution, but has less change of scene between successive frames. Both videos are composed of I, P, and B frames. The *Hula* and *Pirates* videos have only I frames, and are similar to sequences used in commercials or sports close-ups, where the presence of each frame is critical in the overall quality of the sequence. In essence, the four videos selected represent most video sequences intended for Internet applications.

Table 1: MPEG sequences used in experiment

Video	Typical applications
Fly	Video conferencing, medical applications, distance learning, (a portion of the picture continuously changes)
Space	(similar to Fly)
Hula	Commercials, sports close-ups, etc. (where each frame is critical)
Pirates	(similar to Hula)

The main goal of the experiments is to study the tradeoff between Internet bandwidth needed to transmit video data and the perceptual quality observed at the client side. The common factors that affect these parameters are the cutoff frequency and the frame dropping rate.

For the low pass filtering scheme, it was noticed that reducing the cutoff frequency reduces the number of DCT coefficients in a single frame. This leads to a reduction in the average frame size, then resulting in a reduction of Internet bandwidth requirements for transmitting the data. Fig. 1 depicts the relation between cutoff frequency and the reduction achieved in bandwidth requirements. It can be seen that increasing the cutoff frequency, which corresponds to dropping fewer DCT coefficients, yields a monotonic decrease in bandwidth reduction.

Several experiments were performed with selective frame dropping. For the video sequences *Space* and *Fly*, only B-Frames were dropped. The resulting reduction in bandwidth was not significant, as shown in Fig 2. In this figure, "*skip distance*" denotes the number of video frames skipped before dropping a frame. For instance, a skip distance of two denotes a selective frame dropping scheme where every third frame of the original video sequence is dropped.

The ITS video quality metric has been used in our experiments to compare the original unfiltered videos with videos received after using the filtering schemes. In interpreting the results, differences in perceptual quality of less than 0.4 should be disregarded, as explained in Section 3.

The results of the low pass filtering are presented in Fig. 3. Disregarding changes in perceptual quality of less than 0.4, the perceptual quality of the filtered video sequence increases with the cutoff coefficient. This is intuitive, because an increase in the cutoff coefficient must result in fewer droppings of DCT coefficients, and hence less degradation in video quality. For cutoff coefficients above 16, all four videos evaluated, *Fly, Space, Pirates, and Hula*, yield perceptual qualities of above 4, corresponding to a "perceptible, but not annoying" rating on the subjective scale. For even higher values of the cutoff coefficient, the perceptual quality is almost identical to that of the original video sequence for all four videos.

The results of applying the two different low pass filtering schemes to *Fly* and *Space* are illustrated in Fig. 4. In the first scheme, as explained in Section 3.1, DCT coefficients are dropped from only P and B frames. In the second scheme, DCT coefficients are dropped from all three frame types. The latter scheme yields lower perceptual quality, which is intuitive, as preserving I-frames results in less degradation.

The results of applying the selective frame dropping scheme to all four test videos are presented in Fig. 5. Intuitively, the perceptual quality should increase as less frames are dropped. However, as noted from the figure, the perceptual quality of the *Space* and *Pirates* videos does not increase with cutoff coefficient. Both videos have a decrease in perceptual quality. This decrease occurs at the fourth skip distance for *Space*, and at the fifth skip distance for *Pirates*. Observing the values of m_1, m_2, and m_3 for these videos revealed that m_3 changes unexpectedly at these points. As explained in Section 4, m_3 represents the maximum jerkiness in the received video.

In order to get insight into these phenomena, we need to consider the process of "frame filling" at the client side. In this process, while using the selective frame dropping scheme, the dropped frame is replaced at the client side by the frame immediately preceding it. If the original video sequence has a sudden scene change at the dropped frame, replacing this frame with the preceding frame can cause significant jerkiness. If the video sequence is relatively long (over a few minutes), jerkiness in one frame can have only an instantaneous impact on perceptual quality, and generally does not significantly affect the overall quality perceived by the viewer. In computing m_3 in the ITS measure, the length of the video is not taken into consideration, hence significant jerkiness in one frame has the same impact on the overall quality of short and long video sequences. This suggests that the ITS measure, in its current form, may not be suitable for evaluating the perceptual quality of videos filtered with the selective frame dropping scheme. An appropriate modification to the scheme, is such that m_3 represents the "normalized" maximum jerkiness. This normalization can be done by incorporating the length of the video sequence into the computation of m_3. Alternatively, instead of taking a single maximum, the mean of values corresponding to a number of frames with maximum jerkiness could be considered.

To incorporate the effect of change in m_3, we modified the ITS metric for the *Pirates* and *Space* videos. As mentioned above, for both videos, the perceptual quality obtained by the original ITS metric did not increase with cutoff coefficient, due to a change in m_3. To modify m_3, instead of considering the point of maximum jerkiness, we considered the frame with the second largest amount of jerkiness. The modified equation for m_3 is:

$m_3 = MAX2_{time}\{4.23LOG_{10}(TI[D_n]/TI[O_n])\}$,

where $MAX2_{time}$ returns the second largest value for the time history of each test sequence.

Accordingly, the perceptual quality based on the modified value of m_3 is calculated.

As illustrated in Fig. 6, this modification yields a significant change in perceptual quality, which now increases with cutoff coefficient, as expected.

Figs. 7, 8, and 9 depict the tradeoff between perceptual quality and reduction in Internet bandwidth for the filtering schemes evaluated. As seen in Fig. 7, the low pass filtering scheme results in reductions of up to eighty percent in Internet bandwidth, while maintaining acceptable perceptual quality, as high as 4.6 on the ITS scale. The disadvantage of this filtering scheme is its high computational complexity, as mentioned earlier, as it requires decoding of the video stream prior to processing.

Figs. 8 and 9 depict the tradeoff between perceptual quality and reduction in Internet bandwidth for the selective frame dropping scheme. The perceptual quality has been evaluated with the original ITS metric for Fig. 8, and with the modified ITS metric for Fig. 9. These figures demonstrate that a greater reduction in Internet bandwidth is achieved at the cost of degradation in perceptual quality. The reduction in bandwidth achieved by selective frame dropping is less than that achieved by the low pass filtering scheme. Moreover, for comparable reductions in bandwidth, the perceptual quality attained is lower than that of the low pass filtering scheme. The advantage of this method is its low computational complexity.

5. Conclusion

In this paper, we have evaluated two filtering mechanisms for compressed video data, namely low pass filtering and selective frame dropping. The evaluation was based on the reduction achieved in the Internet bandwidth required for transmission of the video, and the perceptual quality of the filtered video sequence. It was observed that the ITS video quality metric is inappropriate for the evaluation of video sequences filtered by selective frame dropping, and a modification to the ITS metric was proposed.

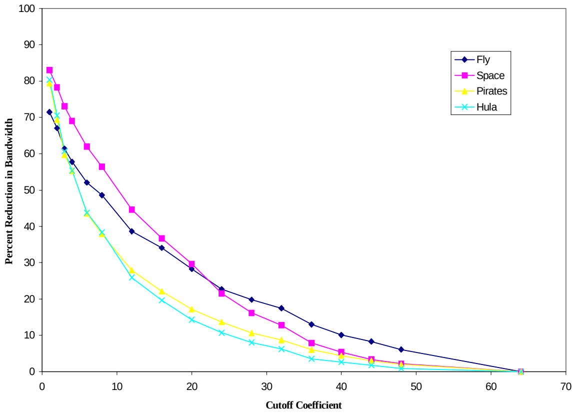

Fig. 1. Reduction in Bandwidth vs. Cutoff Coefficient

References

[1] M. Baentsch, L. Baum, G. Molter, S. Rothkugel,

and P. Sturm, “World Wide Web Caching: The Application Level View of the Internet,” *IEEE Comm.*, vol. 35, no. 6, June 1997.

[2] K. Calvert, M. Doar, E. Zegura, “Modeling Internet Topology,” *IEEE Comm.*, vol. 35, no. 6, June 1997.

[3] B. Furht, W. Smoliar, H. and Zhang, “Video and Image Processing in Multimedia Systems,” Kluwer Academic Publishers, 1995.

[4] H. Fahmi, S. Baqai, A. Bashandy, and A. Ghafoor, "Dynamic Resource Allocation for Multimedia Document Retrieval over High Speed LANs," to appear in *Multimedia Tools and Applications.*

[5] Webster, C. Jones, M. Pinson, S. Voran, and S. Wolf, “An Objective Video Quality Assessment System Based on Human Perception,” in *Proc. SPIE’93*, San Jose, California, Feb. 1993, vol. 1913, pp. 15-26.

[6] R. Storn, “Modeling and Optimization of PET-Redundancy Assignment for MPEG Sequences,” Technical Report TR-95-018.

[7] JTC 1/SC 29 N, *“MPEG I: Coded representation of Picture, Audio and Multimedia/Hypermedia Information”*, International Standards Organization, May 1993.

[8] L. Peterson, and B. Davie, “Computer Networks - A Systems Approach,” Morgan Kaufmann Publishers, Inc., 1996.

[9] T. Frese, C. Bouman, and J. Allebach, “A Methodology for Designing Image Similarity Metrics Based on Human Visual System Models,” *Proc. SPIE’97*, San Jose, California, Feb. 1997, vol. 3016, pp. 8-14.

[10] N. Yeadon, F. Garcia, D. Hutchinson and D. Shepherd, “Filters: QoS Support Mechanisms for Multipeer Communications,” *IEEE J. Select. Areas Commun.*, vol. 14, no. 7, Sept. 1996.

[11] W. Pratt, “Digital Image Processing,” Wiley and Sons, New York, 1974.

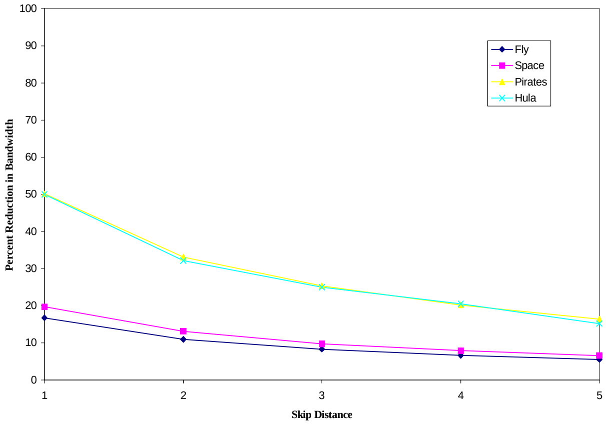

Fig. 2. Reduction in Bandwidth vs. Frame Dropping Skip Distance

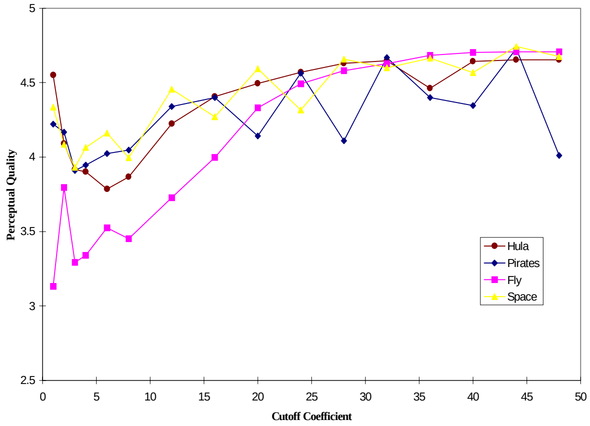

Fig. 3. Perceptual Quality vs. Cutoff Coefficient

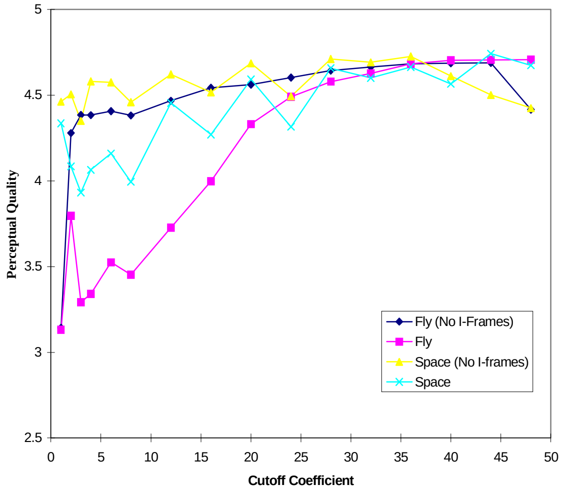

Fig. 4. Perceptual Quality vs. Cutoff Coefficient

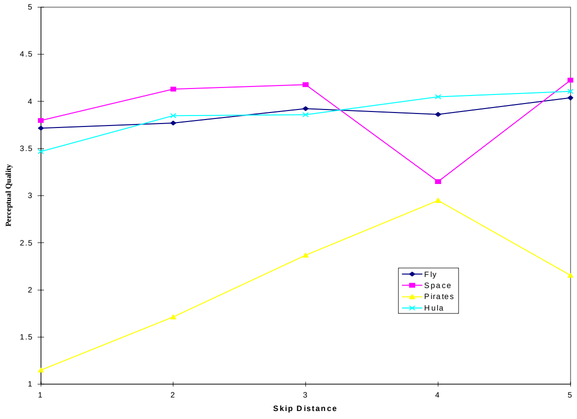

Fig. 5 Skip Distance vs. Cut off frequecy

[12] CCIR Recomm. 500-3, *“Method for the Subjective Assessment of the Quality of Television Pictures”*.

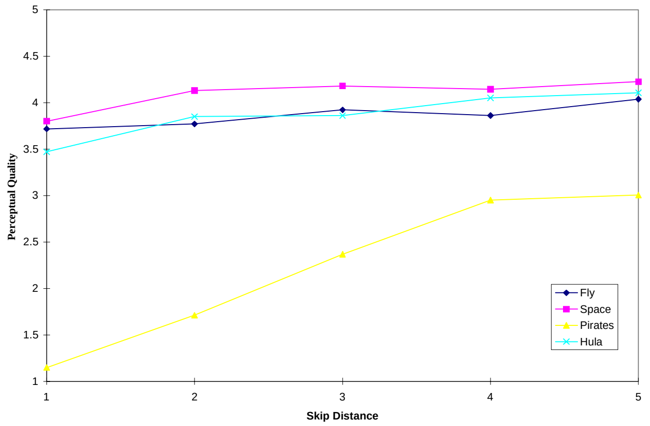

Fig. 6. Perceptual Quality vs. Frame Dropping Skip Distance, Modified ITS Metric

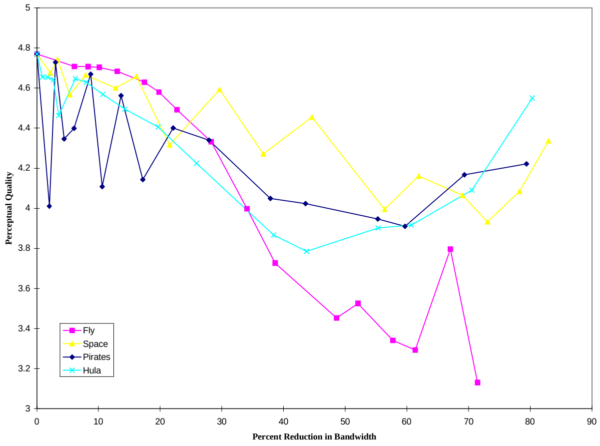

Fig. 7. Perceptual Quality vs. Reduction in Bandwidth
Lowpass Filtering

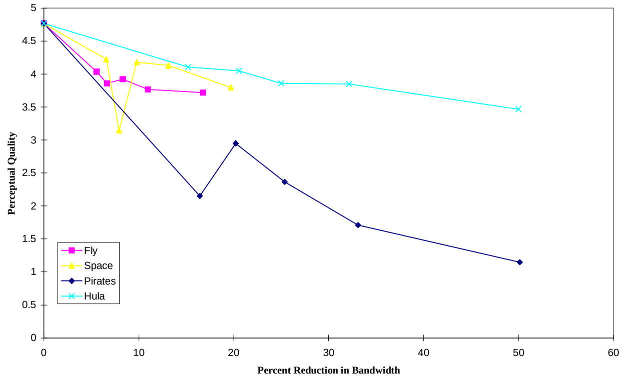

Fig. 8. Perceptual Quality vs. Reduction in Bandwidth
Selective Frame Dropping, Original ITS Metric

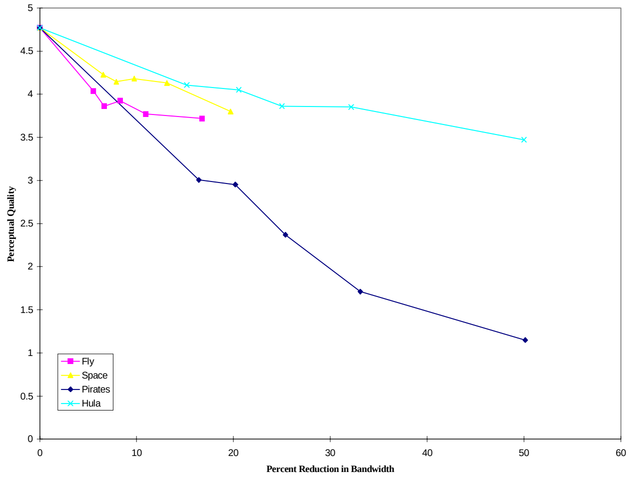

Fig. 9. Perceptual Quality vs. Reduction in Bandwidth
Selective Frame Dropping, Modified ITS Metric