# Video Quality Assessment as Impacted by Video Content over Wireless Networks

Asiya Khan, Lingfen Sun and Emmanuel Ifeachor
Centre for Signal Processing and Multimedia Communication
School of Computing, Communications and Electronics
University of Plymouth, Plymouth PL4 8AA, UK.
e-mail: (asiya.khan, l.sun, e.ifeachor)@plymouth.ac.uk

***Abstract*—The primary aim of this paper is to assess video quality for all content types as affected by Quality of Service (QoS) parameters both in the application and network level. Video streaming is a promising multimedia application and is gaining popularity over wireless/mobile communications. The quality of the video depends heavily on the type of content. The contributions of this paper are threefold. First, video sequences are classified into groups representing different content types using cluster analysis based on the spatial (edges) and temporal (movement) feature extraction. Second, we conducted experiments to investigate the impact of packet loss on video contents and hence find the threshold in terms of upper, medium and lower quality boundary at which users' perception of service quality is acceptable. Finally, to identify the minimum send bitrate to meet Quality of Service (QoS) requirements (e.g. to reach communication quality with Mean Opinion Score (MOS) greater than 3.5) for the different content types over wireless networks. We tested 12 different video clips reflecting different content types. We chose Peak-Signal-to-Noise-Ratio (PSNR) and decodable frame rate (Q) as end-to-end video quality metrics and MPEG4 as the video codec. From our experiments we found that video contents with high Spatio-Temporal (ST) activity are very sensitive to packet loss compared to those with low ST-activity. Further, content providers usually send video at highest bitrate resulting in over provisioning. Through our experiments we have established that sending video beyond a certain bitrate does not add any value to improving the quality. The work should help optimizing bandwidth allocation for specific content in content delivery networks.**

***Keywords-QoS, MPEG4, video content classification, video quality evaluation, wireless communication***

## I. INTRODUCTION

The current trends in the development and convergence of wireless internet IEEE802.11 applications and mobile systems are seen as the next step in mobile/wireless broadband evolution. Multimedia services are becoming commonplace across different transmission platforms such as Wi-Max, 802.11 standards, 3G mobile, etc. Users' demand of the quality of streaming service is very much content dependent. Streaming video quality is dependent on the intrinsic attribute of the content. For example, users request high video quality for fast moving contents like sports, movies, etc. compared to slow moving like news broadcasts, etc. where to understand the content is of more importance. The future internet architecture will need to support various applications with different QoS (Quality of service) requirements [1]. QoS of multimedia communication is affected both by the network level and application level parameters [2]. In the application level QoS is driven by factors such as resolution, frame rate, colour, video codec type, audio codec type, etc. The network level introduces impairments such as delay, cumulative inter-frame jitter, burstiness, latency, packet loss, etc.

Video quality can be evaluated either subjectively or based on objective parameters. Subjective quality is the users' perception of service quality (ITU-T P.800) [3]. The most widely used metric is the Mean Opinion Score (MOS). Subjective quality is the most reliable method however, it is time consuming and expensive and hence, the need for an objective method that produces results comparable with those of subjective testing. Objective measurements can be performed in an intrusive or non-intrusive way. Intrusive measurements require access to the source. They compare the impaired videos to the original ones. Full reference and reduced reference video quality measurements are both intrusive [4]. Quality metrics such as Peak-Signal-to-Noise-Ratio (PSNR), VQM [5] and PEVQ [6] are full reference metrics. VQM and PEVQ are commercially used and are not publicly available. Non-intrusive methods (reference-free), on the other hand do not require access to the source video. Non-intrusive methods are either signal or parameter based. More recently the Q value [7] is a non-intrusive reference free metric. Non-intrusive methods are preferred to intrusive analysis as they are more suitable for on-line quality prediction/control.

Recent work has focused on the wireless network (IEEE 802.11) performance of multimedia applications [8],[9]. In [10],[11],[12] the authors have looked at the impact of transmission errors and packet loss on video quality. In [13] authors have proposed a parametric model for estimating the quality of videophone services that can be used for application and/or network planning and monitoring, but their work is limited to videophone. Similarly, in [14] authors have taken into consideration a combination of content and network adaptation techniques to propose a fuzzy-based video transmission approach. In [15] the authors have proposed content based perceptual quality metrics for different content types, whereas, in [16],[17] video content is divided into several groups using cluster analysis [18]. In [19],[20] authors have looked at video quality assessment of low bitrate videos in multiple dimensions, e.g. frame rate, content type, etc. They have only considered parameters in the application level.

However, very little work has been done on the impact of different types of content on end-to-end video quality e.g. from slow moving (head and shoulders) to fast moving (sports) for streaming video applications under similar network conditions considering both network level and application level parameters. We have looked at the two main research questions in the network level and application level as:

(1) What is the acceptable packet error rate for all content types for streaming MPEG4 video and hence, find the threshold in terms of upper, medium and lower quality boundary at which the users' perception of quality is acceptable?

(2) What is the minimum send bitrate for all content types to meet communication quality for acceptable QoS (PSNR >27 dB) as it translates to a MOS of greater than 3.5 [21]?

To address these two questions, we first classified the video contents based on the spatial and temporal feature extraction into similar groups using cluster analysis [18]. We then carried out experiments to investigate the impact of Packet Error Rate (PER) and hence, find the threshold in terms of upper, medium and lower quality boundary above which the users' perception of quality is acceptable and identified the minimum acceptable Send Bitrate (SBR) for the content types. We chose Peak-Signal-to-Noise-Ratio (PSNR) and decodable frame rate (Q) [5] as end-to-end video quality metrics and MPEG4 as the video codec. In the presence of packet loss video quality becomes highly time-variant [20],[21]. One of the significant problems that video streaming face is the unpredictable nature of the Internet in terms of the send bitrate, and packet loss. We further investigated the impact of video quality over the entire duration of the sequence and hence observe the type of errors using objective video quality metrics such as PSNR. These could help in resource optimization and the development of QoS control mechanisms over wireless networks in the future. Our focus ranges from low resolution and low send bitrate video streaming for 3G applications to higher video send bitrate for WLAN applications depending on type of content and network conditions The proposed test bed is based on simulated network scenarios using a network simulator (NS2) [22] with an integrated tool Evalvid [23]. It gives a lot of flexibility for evaluating different topologies and parameter settings used in this study.

The paper is organized as follows. The video quality assessment problem is formulated in section II. Section III classifies the contents. In section IV the experimental set-up is given. Section V presents the experiments conducted and analysis of results. Conclusions and areas of future work are given in section VI.

## II. PROBLEM STATEMENT

In multimedia streaming services, there are several parameters that affect the visual quality as perceived by the end users of the multimedia content. These QoS parameters can be grouped under application level QoS and network level QoS parameters. Therefore, in the application level perceptual QoS of the video bitstream can be characterized as:

*Perceptual QoS = f (Content type, SBR, frame rate, codec type, resolution, ….)*

whereas, in the network level it is given by:

*Perceptual QoS = f (PER, delay, latency,jitter, ….)*

It should be noted that the encoder and content dimensions are highly conceptual. In this research we chose MPEG4 as the encoder type. We further extracted spatial and temporal features of the video and classified video content accordingly. In the application level we chose send bitrate and in the network level we chose packet error rate as QoS parameters. Hence the main contributions of the paper are three-fold.

(1) Most frequent content types are classified into three main groups by extracting temporal (movement) and spatial (blockiness, blurriness and brightness) feature using a well known tool called cluster analysis.
(2) We define the threshold at which packet loss is acceptable for all content types and
(3) We identify the minimum send bitrate for all content types for acceptable quality.

## III. CONTENT CLASSIFICATION

The chosen video sequences ranged from very little movement, i.e. small moving region of interest on static background to fast moving sports clips. The choice of video sequences was to reflect the varying spatio-temporal activity of the content representative of typical content offered by content providers e.g. news type of content or fast moving sports content. In future, we will consider movie clips and carry out segment by segment analysis of the content features extracted. The content classification was done based on the temporal and spatial feature extraction using well known tool called cluster analysis [18].

The design of our content classification method is given in Fig. 1.

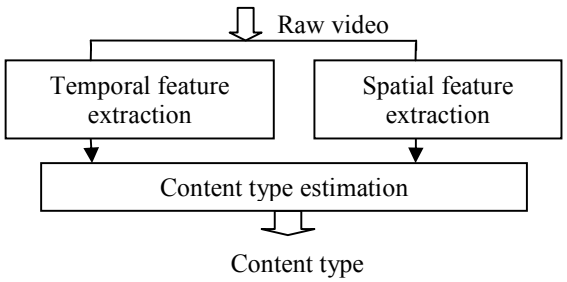


Figure 1. Content classification design

### A. Temporal Feature Extraction

The motion of the temporal sequence can be captured by removing temporal-domain redundancies. The motion can be accumulated into one image that represents the activity of

the whole temporal sequence. Temporal-domain redundancy reduction techniques are well established in the video compression literature. Hybrid video compression standards employ backward and bidirectional prediction as specified by the ISO/IEC MPEG coders such as MPEG-4 part 10 [24]. On the other hand, wavelet-based video coders employ sophisticated motion-compensated temporal filtering techniques as reported in [25] and [26]. To reduce the energy of prediction error, video coders employ motion estimation and motion compensation prediction on blocks of pixels referred to as macroblocks. The outcome of the motion estimation process is a 2-D motion vector representing the relative displacement of a macroblock relative to a reference video clip. The motion compensation prediction subtracts the macroblocks of the current video clip from the best matched location of the reference video clip as indicated by the relevant motion vector. The movement in a video clip can be captured by the SAD value (Sum of Absolute Difference). In this paper, we have used the SAD values as temporal features and are computed as the pixel wise sum of the absolute differences between the two frames being compared and is given by eq. (1).

$$SAD_{n,m} = \sum_{i=1}^{N}\sum_{j=1}^{M}|B_n(i,j) - B_m(i,j)| \quad (1)$$

where $B_n$ and $B_m$ are the two frames of size N X M, and i and j denote pixel coordinates.

### B. Spatial Feature Extraction

The spatial features extracted were the blockiness, blurriness and the brightness between current and previous frames [27].

Blockiness measures the blocking effect in video sequence. For example, in contrast areas of the frame blocking is not appreciable, but in smooth areas these edges are conspicuous. The blockiness measure is calculated the visibility of a block edge determined by the contrast between the local gradient and the average gradient of the adjacent pixels [28] and is given by eq. (2).

$$Blockiness = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[x_m^n(i,j) - \bar{x}_m^n]^2\} \quad (2)$$

where $x_m^n(i,j)$ denotes the pixel value in location (i,j) of the mth block in the nth frame, $\bar{x}_m^n$ denotes the mean of the pixel values of the mth block in the nth frame, M denotes the number of blocks per frame, and N denotes the number of frames under investigation from the video sequence.

Blurring measurement is based on the measure of local edge expansions. The vertical binary edge map is first computed with the Sobel filter. Then, the local extrema in the horizontal neighbourhood of each edge point are detected, and the distance between these extrema (xp) is computed. Blurring is computed as the average of the edge expansions for all edge points and is given be eq. (3).

$$\text{Bluriness} = \frac{1}{N_e}\sum_{m=1}^{M}\sum_{n=1}^{N}|xp_1 - xp_2| \quad (3)$$

where $N_e$ is the number of edge points. $xp_1$ and $xp_2$ are the local extrema in the horizontal neighborhood of each edge point.

Brightness (Br) is calculated as the modulus of difference between average brightness values of previous and current frames and is given by eq. (4).

$$Br_{av\{n\}} = \sum_{i=1}^{N}\sum_{j=1}^{M}\left|Br_{av(n)}(i,j) - Br_{av(n-1)}(i,j)\right| \quad (4)$$

where $Br_{av\{n\}}$ is the average brightness of *n*-th frame of size *N X M*, and *i* and *j* denote pixel coordinates.

### C. Cluster Analysis

We chose 12 video sequences reflecting very low spatial and temporal to very high spatial and temporal activity. Based on the table of mutual Euclidean norm in the joint temporal and spatial sense between pair of sequences, we created dendrogram on the basis of a nearest distance in a 4-dimensional Euclid-space. The dendrogram or tree diagram constructed in this way classifies the content. The features (i.e. SAD, blockiness, blurriness and brightness measurements) extracted are given in normalized form. Fig. 2 shows the obtained dendrogram (tree diagram) where the video sequences are grouped together on the basis of their mutual distances (nearest Euclid distance).

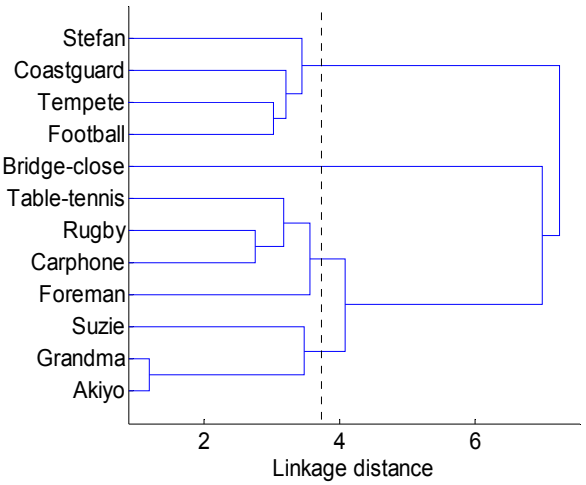


Figure 2. Tree diagram based on cluster analysis

According to Sturge's rule (k = 1 + 3.3logN), which for our data will be 5 groups. However because of the problems identified with this rule [29] we split the data (test sequences) at 38% from the maximum Euclid distance into three groups. (see the dotted line on Fig. 2) as the data contains a clear 'structure' in terms of clusters that are similar to each other at that point. Group 1 (sequences Grandma, Suzie and Akiyo) are classified as 'Slight Movement', Group 2 (sequences Carphone, Foreman, Table-tennis and Rugby) are classified as 'Gentle Walking' and Group3 (sequences Stefan and Football) are classified

as 'Rapid Movement'. We found that the 'news' type of video clips were clustered in one group, however, the sports clips were put in two different categories i.e. clips of 'stefan' and 'football' were clustered together, whereas, 'rugby' and table-tennis' were clustered along with 'foreman' and 'carphone' which are both wide angle clips in which both the content and background are moving. Also 'bridge-close' can be classified on its own creating four groups instead of three. But as it is closely linked with the first group of SM we decided to put it in SM. In future, we will create more groups and compare it to our existing classification.

The cophenetic correlation coefficient, c, is used to measure the distortion of classification of data given by cluster analysis. It indicates how readily the data fits into the structure suggested by the classification. The value of c for our classification was 79.6% indicating a good classification result. The magnitude of c should be very close to 100% for a high-quality solution.

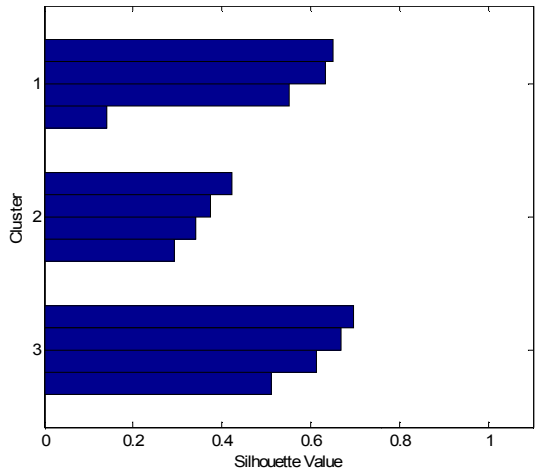


Figure 3. k-means of all contents types

To further verify the content classification from the tree diagram obtained (Fig. 2) we carried out K-means cluster analysis in which the data (video clips) is partitioned into *k* mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. K-means computes cluster centroids differently for each measured distance, to minimize the sum with respect to the specified measure. We specified k to be three to define three distinct clusters. In Fig. 3 K-means cluster analysis is used to partition the data for the twelve content types. The result set of three clusters are as compact and well-separated as possible giving very different means for each cluster. Cluster 3 in Fig. 3 is very compact for the four video clips, whereas cluster 2 is reasonable compact. However, cluster 1 can be further divided into more groups. For example the video clip of bridge-close can be in a separate group. This will be looked in much detail in future work. All results were obtained using MATLAB™ 2008 functions.

The three content types are defined for the most frequent contents for mobile video streaming as follows:

1. Content type 1 – Slight Movement (SM): includes sequences with a small moving region of interest (face) on a static background. See Fig. 4.

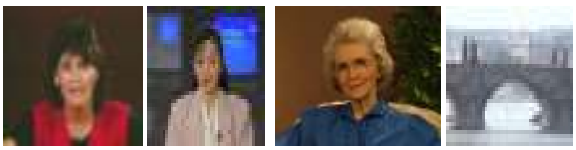

Figure 4. Snapshots of typical 'SM' content

2. Content type 2 – Gentle Walking (GW): includes sequences with a contiguous scene change at the end. They are typical of a video call scenario. See Fig. 5.

Figure 5. Snapshots of typical 'GW' content

3. Content type 3 – Rapid Movement (RM): includes a professional wide angled sequence where the entire picture is moving uniformly e.g sports type. See Fig. 6.

Figure 6. Snapshots of typical 'RM' content

### *D. Comparison with the spatio-temporal dynamics*

Video sequences are most commonly classified based on their spatio-temporal features. In order to classify video clip according to the spatial and temporal complexity of its content, a spatio-temporal grid [30] is considered and is depicted in Fig. 7.

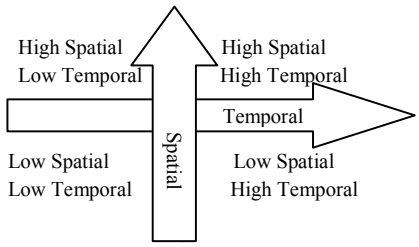


Figure 7. The spatio-temporal grid used for classifying a video sequence according to its content dynamics

From Fig. 7 the spatio-temporal grid divides each video into four categories based on its spatio-temporal features as follows:

- Low spatial – Low temporal activity: defined in the bottom left quarter in the grid.

- Low spatial – High temporal activity: defined in the bottom right quarter in the grid.
- High spatial – High temporal activity: defined in the top right quarter in the grid.
- High spatial – Low temporal activity: defined in the top left quarter in the grid.

Figure 8 shows the principal co-ordinates analysis also known as multidimensional scaling of the twelve content types. The function cmdscale in MATLAB™ is used to perform the principal co-ordinates analysis. cmdscale takes as an input a matrix of inter-point distances and creates a configuration of points. Ideally, those points are in two or three dimensions, and the Euclidean distances between them reproduce the original distance matrix. Thus, a scatter plot of the points created by cmdscale provides a visual representation of the original distances and produces representation of data in a small number of dimensions. In Fig. 8 the distance between each video sequence indicates the characteristics of the content, e.g. the closer they are the more similar they are in attributes.

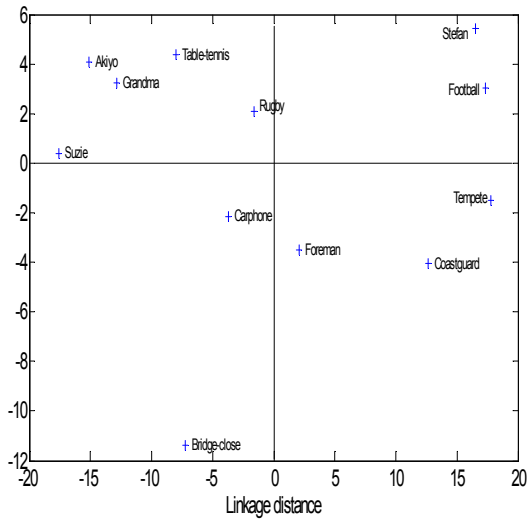


Figure 8. Principal co-ordinate analysis of all contents

Comparing Fig.7 to Fig. 8 we can see that classifying contents using feature extraction, contents of Football and Stefan are high spatial and high temporal and fit in the top right hand side, similarly contents of Bridge-close would fit in the bottom left hand side as they have low spatio-temporal features. Whereas, contents like Grandma and Suzie are in top left hand side indicating high spatial and low temporal features. Similarly, Foreman, Coastguard and Tempete are in the bottom right hand side with high temporal and low spatial features as expected. Only the video sequence of Carphone has been put in the bottom left hand side and will be investigated further.

## IV. EXPERIMENTAL SET-UP

For the tests we selected twelve different video sequences of qcif resolution (176x144) as it is recommended for low bitrate videos especially over mobile environments and encoded in MPEG4 format with an open source ffmpeg [31] encoder/decoder with a Group of Pictures (GOP) pattern of IBBPBBPBB. In future we will choose H.264 as it the recommended codec for low bitrates. The frame rate was fixed at 10fps. Each GOP encodes three types of frames - Intra (I) frames are encoded independently of any other type of frames, Predicted (P) frames are encoded using predictions from preceding I or P frames and Bi-directionally (B) frames are encoded using predictions from the preceding and succeeding I or P frames.

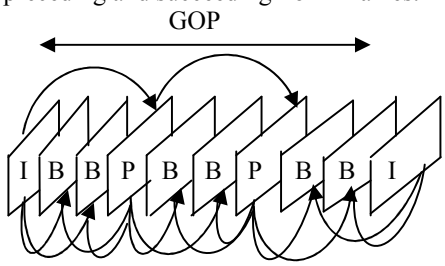


Figure 9. A sample of MPEG4 GOP (N=9, M=3)

A GOP pattern is characterized by two parameters, GOP(N,M) – where N is the I-to-I frame distance and M is the I-to-P frame distance. For example, as shown in Fig.9, G(9,3) means that the GOP includes one I frame two P frames, and six B frames. The second I frame marks the beginning of the next GOP. Also the arrows in Fig. 9 indicate that the B frames and P frames decoded are dependent on the preceding or succeeding I or P frames [32].

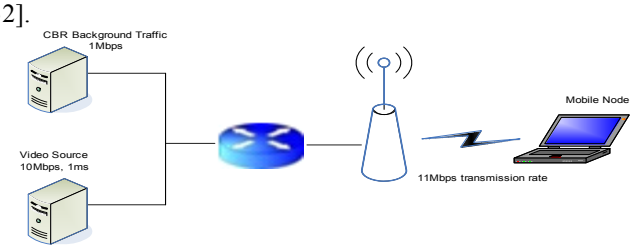


Figure 10. Simulation setup

The experimental set up is given in Fig 10. There are two sender nodes as CBR background traffic and MPEG4 video source. Both the links pass traffic at 10Mbps, 1ms over the internet. The router is connected to a wireless access point at 10Mbps, 1ms and further transmits this traffic to a mobile node at a transmission rate of 11Mbps 802.11b WLAN. No packet loss occurs in the wired segment of the video delivered path. The maximum transmission packet size is 1024 bytes. The video packets are delivered with the random uniform error model. The CBR rate is fixed to 1Mbps to give a more realistic scenario. The packet error rate is set in the range of 0.01 to 0.2 with 0.05 intervals. To account for different packet loss patterns, 10 different initial seeds for random number generation were chosen for each packet error rate. All results generated in the paper were obtained by averaging over these 10 runs.

## V. EXPERIMENT AND ANALYSIS OF RESULTS

We considered both network level and application level factors and used performance metrics to evaluate video quality affected by both factors. The performance metrics used were average PSNR and decodable frame rate Q [7]. PSNR given by (1) computes the maximum possible signal energy to noise energy. PSNR measures the difference between the reconstructed video file and the original video file.

$$PSNR(s,d) = 20 \log \frac{Max}{\sqrt{MSE(s,d)}} \quad (5)$$

Max is the maximum pixel value of the image, which is 255 for 8 bit samples. Mean Square Error (MSE) is the cumulative square between compressed and the original image.

Decodable frame rate (Q) [7] is defined as the number of decodable frames over the total number of frames sent by a video source. Therefore, the larger the Q value, the better the video quality perceived by the end user. The decodable frame number is the number of decodable I/P/B frames. Considering in a GOP I frame is decodable only if all the packets that belong to the I frame are received. Similarly P frame is decodable only if preceding I or P frames are decodable and all the packets that belong to the current P frame are received well. The B frame is decodable only if the preceding and succeeding I or P frame are both decodable and all the packets that belong to the current B frame are all received.

We chose 4 different experiments as outlined in sub-sections A-D below. The motivation of these experiments was to address the two research questions raised in the Introduction section. Experiments 1-3 (sub-sections A-C) address the first question by looking at the impact of packet error rate on end-to-end quality. Whereas, experiment 4 (sub-section D) addresses the second question to identify the minimum acceptable bitrate to meet acceptable QoS.

### A. Experiment 1 – Average PSNR Vs PER

Video quality is measured by taking the average PSNR over all the decoded frames across network PER from 0.01 to 0.2 (20%). All videos were encoded at a send bitrate of 256kb/s. This experiment is conducted to answer the first research question: What is the acceptable PER for maintaining the minimum QoS requirement of 27dB for the different content types ?

Fig. 11 show the average PSNR vs the PER for all 12 video clips. It shows that the average PSNR is better for slight movement compared to gentle walking which in turn is better than rapid movement which shows the dependence on content type. From our results, we found that for slight movement the video quality stays above the threshold of PSNR > 27dB (MOS >3.5) for upto 20% packet loss. However, for gentle walking and rapid movement that value drops to 10% and 6% respectively.

We observe from Fig. 11 that the drop in video quality is much higher for fast moving contents compared to that of slow moving contents. E.g. for 'Akiyo' at 0.01 PER the PSNR is 44dB and at 0.2 (20%) PER it is 27.67dB. However, for 'Football' it is 33dB for a PER of 0.01 and 20dB for PER of 0.2. Even though the percentage drop in quality is more or less the same, 20dB is unacceptable for communication standards. This can be furthered explained by the fact that the bitrate was fixed at 256kb/s. If the bitrate is varied then the impact of packet error rate is much greater on fast moving contents.

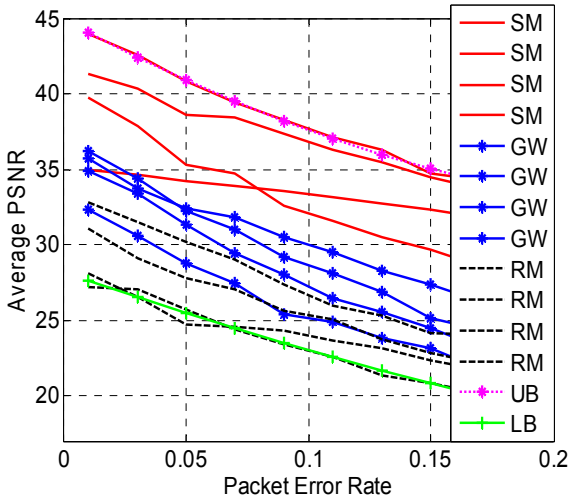


Figure 11. Packet Error Rate vs Average PSNR

Further, we derive an upper, medium and lower boundary for PSNR as a function of PER for the three content types of SM, GW and RM and hence know the threshold for acceptable quality in terms of the PSNR for the three content types with 95% confidence level and goodness of fit of 99.71% and Root Mean squared Error (RMSE) of 0.3235 is given by equations (6), (7) and (8):

$$\text{SM: } PSNR = 122.3(PER)^2 - 88.36(PER) + 42.6; \quad PER \leq 20\% \quad (6)$$

$$\text{GW: } PSNR = 64.9(PER)^2 - 73.75(PER) + 34.43; \quad PER \leq 10\% \quad (7)$$

$$\text{RM: } PSNR = 76.8(PER)^2 - 68.87(PER) + 31.43; \quad PER \leq 6\% \quad (8)$$

### B. Experiment 2 – Q Vs PER

The experimental set up is the same as in A but we measured Q value [7] instead of PSNR vs PER and addressed the above research question in terms of Q [7] instead of PSNR.

Fig. 12 shows the decodable frame rate (Q) of all 12 contents and shows that Q is higher when the PSNR is higher for all the video clips. In comparison to Fig 3 the decodable frame rate does not directly compare to the PSNR. However, from our results we found higher values for the average PSNR for 'slight movement' and it did not correspond to a higher value of Q. This is because the Q value is derived from the number of decodable frames over the total number of frames sent by a video source [5] i.e. it is sensitive to the number of frames and packets lost.

Therefore, as the content becomes more complex we would expect the video quality to degrade more for less I-frames lost compared to that of simpler contents. Hence, we conclude that for slight movement 20%, for gentle walking 10% and for rapid movement 6% packet loss is acceptable.

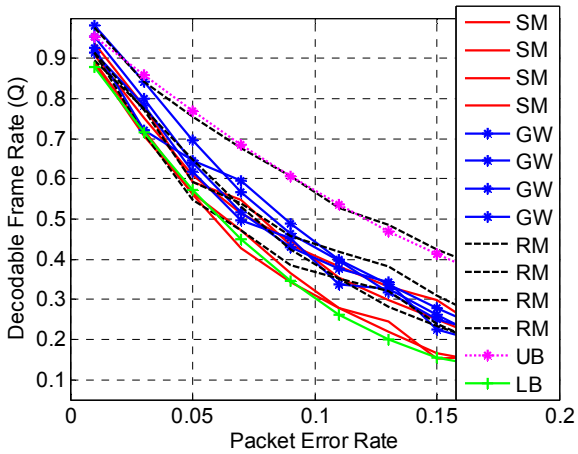


Figure 12. PER vs Q for all content types

Further, we derive an upper, medium and lower boundary for Q value as a function of PER for the three content types of SM, GW and RM and hence know the threshold for acceptable quality in terms of the Q value for the three content types with 95% confidence level and goodness of fit of 99.71% and RMSE of 0.0117 is given by the equations (9), (10) and (11):

$$SM: Q=19.89(PER)^2 - 8.03(PER) + 0.967; \quad PER \leq 20\% \quad (9)$$

$$GW: Q=18.09(PER)^2 - 7.88(PER) + 1.02; \quad PER \leq 10\% \quad (10)$$

$$RM: Q=13.84(PER)^2 - 6.5(PER) + 0.975; \quad PER \leq 6\% \quad (11)$$

Table I summarizes the findings of Figs. 11 and 12 and outlines the PSNR and Q values for acceptable quality at 20%, 10% and 6% PER for all three content types in terms of the I, P and B frames lost. We observe from Table I that for content type of SM the Q value is much lower compared to that of the PSNR. It shows that visually the quality is much lower at 20% packet loss rendering PSNR to be not a very good predictor of visual quality. For SM, Q-value out-performs the PSNR.

TABLE I
PSNR AND Q VALUES FOR THREE CONTENT TYPES @ 20%, 10% AND 6% PACKET LOSS

| | I-frames lost | P-frames lost | B-frames lost | PSNR | Q-value |
|---|---|---|---|---|---|
| SM | 8 | 14 | 43 | 27.67 | 0.458 |
| GW | 8 | 7 | 22 | 28.103 | 0.602 |
| RM | 8 | 11 | 12 | 25.57 | 0.615 |

### C. *Experiment3 – PSNR Vs Time*

We further looked at the relationship between the PSNR over the entire duration of the sequence for all three content types.

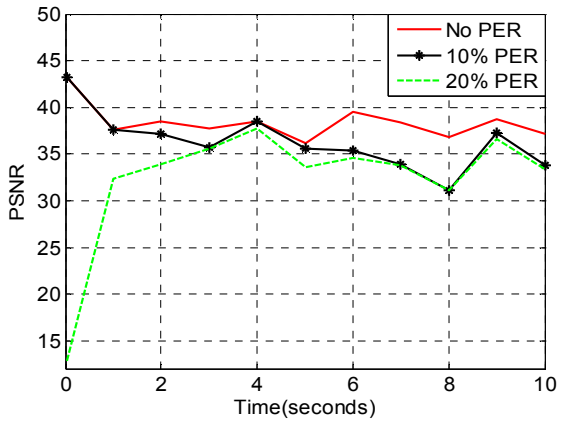


Figure 13. PER effects for SM for 32kb/s SBR

In Fig. 13 we investigate the source of effects caused by packet errors over the entire duration of the sequence. For 'slight movement' we compare the PSNR values for no transmission errors to 10% and 20% packet loss. The PSNR values are the same for a new I-frame over the duration of the sequence. The error occurs in the B-frames and propagates to the P-frames as expected. We observe two effects, the PSNR decreases over the entire duration and the second a more ragged response curve when packet errors of 10% and 20% are introduced. We also observe that for a send bitrate of 32kb/s the video quality is still acceptable for 20% packet loss.

Fig. 14 shows the effects of no packet loss, 10% and 20% packet loss for 'Gentle walking' at a send bitrate of 80kb/s. Again as previously mentioned the video quality reduces over the time duration and we observe a much bigger loss in quality as the packet loss increases to 20%.

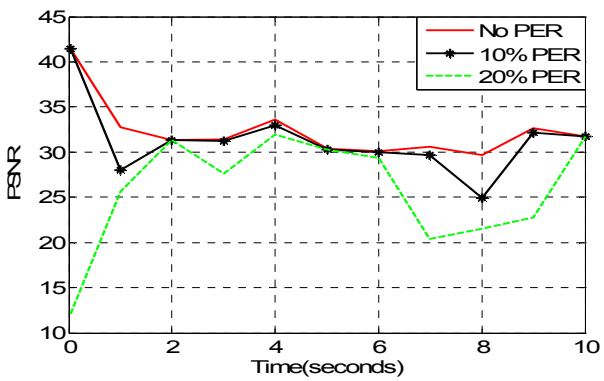


Figure 14. PER effects for GW for 80kb/s SBR

Whereas, from Fig. 15 in 'rapid movement' the video quality degrades fairly quickly with the increase of packet

error rate i.e. for 10% packet loss the video quality is completely unacceptable.

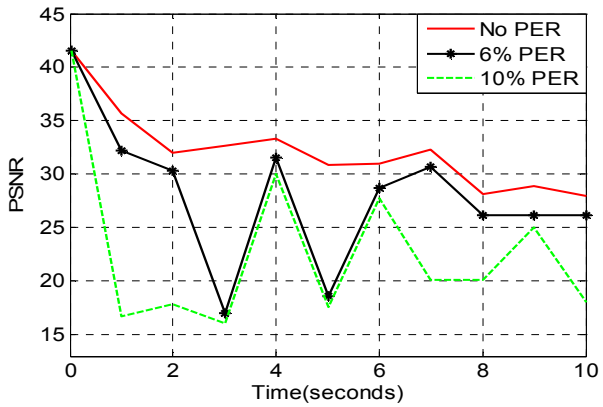


Figure 15. PER effects for RM for 256kb/s SBR

While PSNR is not a good predictor of the visual quality, it can serve as a detector of clearly visible distortions. It can be observed, however that the perceived quality degradation increases in the duration of the sequence. Due to the auto-correlation of the time series (each sample is dependent on the previous and following sample) the values are not independent. We also observed that as the scene activity in the video sequence becomes more complicated e.g. for 'rapid movement' at 20% packet loss the quality is completely unacceptable deteriorating at a much faster speed. All degraded video clips can be found in [33].

Fig. 16 shows that visually the quality of SM, GW and RM is unacceptable at 20%, 10% and 6% packet loss for some frames. Also from Table I we observe that even though PSNR value is acceptable (MOS>3.5) for all three content types, however, the end-to-end perceptual quality is unacceptable. From Fig. 13, the PSNR at 3.4 seconds for SM shows a value of 35dB, whereas the frames (101-103) from Fig. 16a show that the perceptual quality does not follow for those frames. Similarly, for GW at 5.2s (Fig. 14) the PSNR is 30dB and for RM at 3.2s it is 17dB. The PSNR values of GW and RM reflect the perceptual quality better compared to SM. Further from Table I it can be seen that for SM, more B-frames are lost compared to GW and RM. B-frames affect the quality least in MPEG4 GOP. I-frames take priority, then P-frames and finally B-frames. Also the values of Q correlate well with PSNR for GW and RM. However, for SM it does not. Q-value for SM actually shows that at 20% the quality is less than acceptable compared to that of PSNR. This is an area of future work to carry out substantive subjective tests to verify the results of this paper. Also it confirms previous studies [34] that PSNR is not a good indicator of perceptual quality.

### *D. Experiment 4 – Average PSNR Vs PER Vs SBR*

The send bitrate versus PSNR curve is shown in Fig. 17 for all contents. From Fig. 17 we observe that there is a minimum send bitrate for acceptable quality (PSNR > 27dB) for all content types. For high definition IPTV applications PSNR of 32dB is recommended. Therefore, in Fig. 17 we have chosen 32dB as minimum acceptable PSNR as compared to 27dB to illustrate the point of optimizing bandwidth. A PSNR value of 35db is considered "good" for streaming applications [35]. Also there is a maximum send bitrate for the three content types that gives maximum quality (PSNR > 38db). For example for the content category of SM, send bitrate of 30kbps or more gives a maximum PSNR of 38dB. However, in RM higher send bitrates are required for maximum quality i.e. > 370kb/s. From Fig. 17 it can be derived that when the send bitrate drops below a certain threshold, which is depended on the video content, then the quality practically collapses. Moreover, the quality improvement is not significant for send bitrates higher than a specific threshold, which is also dependent on the spatial and temporal activity of the clip.

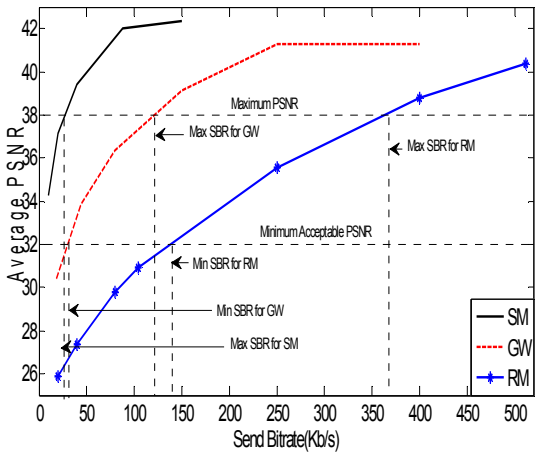


Figure 17. MOS Vs Send Bitrate for the three contents

The experimental set up is the same as in section IV, but we changed the video send bitrate to achieve the minimum send bitrate for QoS requirements and to address the research question: What is the minimum SBR for the different video content types with time variant quality acceptable for communication quality (>27dB)?

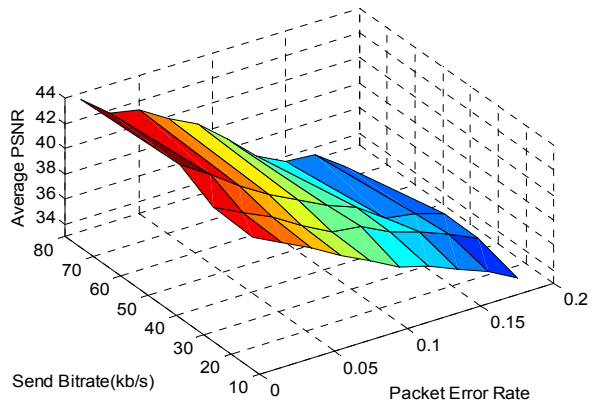


Figure 18. Average PSNR Vs PER and SBR for 'SM'

The send bitrates ranged from 18kb/s to 384kb/s. We chose one video clip from each category. We suggest a minimum send bitrate for all three categories that achieve an average PSNR values of higher than 27dB for the video content types as it translates to a MOS of greater than 3.5 [23] which is an acceptable score for the telecommunication industry.

Fig. 18 shows the average PSNR over the video send bitrates of 18kb/s, 32kb/s, 44kb/s and 80kb/s. We found that for slow movement low bitrate of 18kb/s is acceptable as it yields an average PSNR of 30dB without any packet loss. As the send bit rate is increased to 80kb/s, average PSNR is greater than 40dB indicating that the bandwidth should be re-allocated to optimize it.

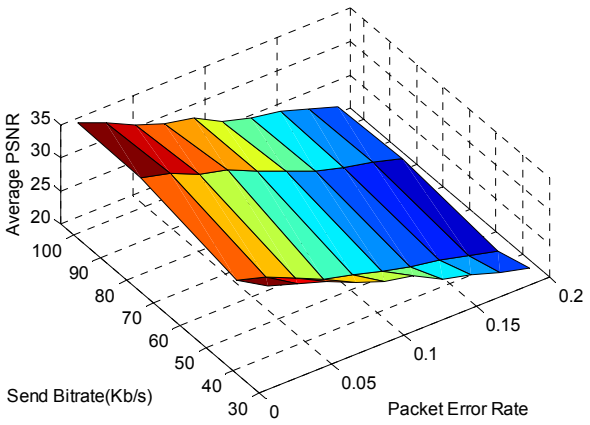


Figure 19. Average PSNR Vs PER and SBR for 'GW'

In Fig. 19 we chose send bitrates of 32kb/s, 44kb/s, 80kb/s and 104kb/s, as bitrates less than 18kb/s will give poor video quality rendering them meaningless. We suggest a send bitrate of 32kb/s for gentle walking as it gives an average PSNR value of approximately 29dB. However, with higher packet loss the quality falls below the acceptable level.

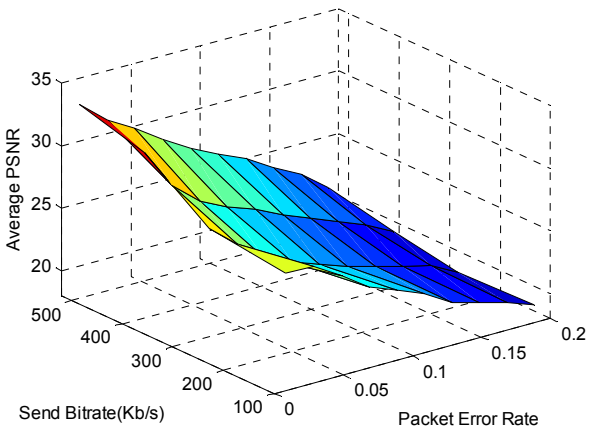


Figure 20. Average PSNR Vs PER and SBR for 'RM'

In Fig. 20 we chose bitrates of 80kb/s, 104kb/s, 256kb/s, 384kb/s and 512kb/s as bitrates less than 80kb/s gave meaningless results in terms of very low PSNR. From our results we suggest a minimum send bitrate of 256kb/s as it yields a PSNR of 30dB. Increasing the send bitrate improves the quality with no packet loss. However, increasing the send bitrate does not compensate for the higher packet loss effect of streaming video quality for fast moving content due to network congestion issues. In fact, when the network is congested the bitrate should be reduced to release congestion. However, the quality of fast moving videos reduces if the bitrate is reduced beyond a certain threshold.

Therefore, the quality of video in 'rapid movement' degrades much more rapidly with an increase in packet loss compared to that of 'slight movement' and 'gentle walking'.

## VI. CONCLUSIONS

In this paper we classified the most significant content types and have established guidelines for the transmission of MPEG4 streaming video over wireless networks in terms of acceptable packet loss and minimum send bitrate. The contents were first classified using cluster analysis into three groups with good prediction accuracy. The video quality is evaluated in terms of average PSNR and decodable frame rate, Q. The acceptable PER was found to be 20%, 10% and 6% for the three content categories of SM, GW and RM respectively. We found that for content category of SM the Q value was more sensitive compared to PSNR as it gave a lower value for 20% packet loss which was more representative visually. However, for GW and RM very little difference was found between PSNR and Q.

Through the first three experiments, we established that as the ST-activity in the content increases it becomes more sensitive to network impairments such as packet loss. Although for low ST-activity videos the acceptable PER was found to be 20% in terms of the PSNR, however, visually looking at the videos, we found that quality was not acceptable at such high packet losses due to blocking and blurring effects. This also confirms previous studies that PSNR is not a good reflector of visual quality - thus addressing the first question raised in the Introduction section.

To address the second question raised in the Introduction section, through our fourth experiment we identified the minimum SBR for acceptable QoS for the three content types as 18, 32 and 256kb/s for SM, GW and RM respectively. Hence, we have established that sending video beyond a certain bitrate does not add any value to improving the end user quality.

We believe that the results would help in optimizing resource allocation for specific content in content delivery networks and the development of QoS control methods for video over mobile/wireless networks. Future direction of our work is to further investigate the more perceptual-based quality metric and adapt the video send bitrate depending on network conditions over wireless networks. Also subjective tests will be carried out to verify our results.

(a) Frames 101-103, PER @ 20% for SM encoded at 32kb/s

(b) Frames 156-158, PER @ 10% for GW encoded at 80kb/s

(a) Frames 96-98 , PER @ 6% for RM encoded at 256kb/s

Figure 16. Perceptual quality comparison for the 3 content types at PER 20%, 10% and 6%

## ACKNOWLEDGMENT

This paper is an invited extended version of the conference paper A.Khan, L. Sun and E. Ifeachor, "Impact of video content on video quality for video over wireless networks", published in Proc. of 5th ICAS, Valencia, Spain, 20-25 April 2009.

The work reported here is supported in part by the EU FP7 ADAMANTIUM project (contract No. 214751).

## REFERENCES

[1] G. Ghinea and J. P. Thomas, "QoS impact on user perception and understanding of multimedia video clips", *Proc. Of ACM Multimedia '98*, Bristol, UK, pp. 49-54, 1998.

[2] A. Khan, Z. Li, L. Sun and E. Ifeachor, "Audiovisual quality assessment for 3G networks in support of E-healthcare", *Proc. of CIMED* , Plymouth, UK, 25-27 July 2007.

[3] ITU-T. Rec P.800, Methods for subjective determination of transmission quality, 1996.

[4] Video quality experts group, multimedia group test plan, Draft version 1.8, Dec 2005, www.vqeq.org.

[5] http://compression.ru/video/index.htm

[6] http://www.pevq.org

[7] C. Ke, C. Lin and C. Shieh, "Evaluation of delivered MPEG4 video over wireless channels", Journal of mobile multimedia, Vol. 3, No. 1, pp. 047-064, 2007.

[8] P. Chondros, A. Prayati, C. Koulamas and G. Papadopoulos, "802.11 performance evaluation for multimedia streaming", *Fifth*

*International Symposium on Communication Systems, Networks and Digital Signal Processing*, Patras, Greece, 19-21 July, 2006.

[9] B. Munir, N. K. Chilamkurti and B. Soh, "A comparative study of voice over wireless networks using NS-2 simulation with an integrated error model", *International Conf. on WiCOM,* 22-24 Sept. 2006.

[10] Y. Koucheryavy, D. Moltchanov and J. Harju, "Performance evaluation of live video streaming service in 802.11b WLAN environment under different load conditions", MIPS, Napoli, Italy, November 2003.

[11] Z. He, H. Xiong, "Transmission distortion analysis for real-time video encoding and streaming over wireless networks", IEEE transactions on Circuits and Systems for Video Technology, Vol. 16, No. 9, Sept. 2006.

[12] S. Kanumuri, P. C. Cosman, A. R. Reibman and V. A. Vaishampayan, "Modelling packet-loss visibility in MPEG2 video", *IEEE Transactions on Multimedia*, Vol. 8, No. 2, April 2006.

[13] K. Yamagishi, T Tominaga, T. Hayashi and A. Takahasi, "Objective quality estimation model for videophone services", *NTT Technical Review*, Vol.5, No. 6, June 2007.

[14] V. Vassilou, P. Antoniou, I. Giannakou and A. Pitsillides " Delivering adaptive scalable video over wireless internet", *International Conference on Artificial Neural Networks (ICANN),* Athens, Greece, September 10-14, 2006.

[15] M. Ries, C. Crespi, O. Nemethova and M. Rupp, "Content based video quality estimation for H.264/AVC video streaming", *Proc. Of the IEEE Wiress Communication & Networking Conference*, Hong Kong, China, Mar. 2007.

[16] Y. Suda, K. Yamori and Y. Tanaka, "Content clustering based on users' subjective evaluation", Information and Telecommunication Technologies, 2005. APSITT 2005 Proceedings. 6th Asia-Pacific Symposium, Pages 177-182, Nov. 2005.

[17] A. Khan, L. Sun and E. Ifeachor, "Impact of video content on video quality for video over wireless networks", 5th ICAS, Valencia, Spain, 20-25 April 2009.

[18] W. J. Krzanowski, "Principles of Multivariate Analysis", Clarendon press, Oxford, 1998.

[19] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang and M. Etoh, "Cross-dimensional perceptual quality assessment for low bitrate videos", IEEE Transactions on Multimedia, Vol. 10, No.7, pp. 1316-1324, Nov. 2008.

[20] S. Winkler and F. Dufaux, "Video quality evaluation for mobile applications", *Proceedings of SPIE Visual Communications and Image Processing*, Lugano, Switzerland, 2003.

[21] O. Nemethova, M. Ries, M. zavodsky and M. Rupp, "PSNR-based estimation of subjective time-variant video quality for mobiles", *Proc. of MESAQIN 2006*, Prag, Tschechien, June, 2006.

[22] NS2, http://www.isi.edu/nsnam/ns/

[23] J. Klaue, B. Tathke and A. Wolisz, "Evalvid – A framework for video transmission and quality evaluation", *In Proc. Of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pp. 255-272, Urbana, Illinois, USA, September 2003.

[24] MPEG-4, Information Technology, Coding of Audio-Visual Objects, Part 10: Advanced Video Coding, ISO/IEC 14496-10, 2005.

[25] P. Chen, "Fully scalable subband/wavelet coding," Ph.D. dissertation, Rensselaer Polytechnic Inst., Troy, NY, May 2003.

[26] R. Ohm, M. Schaar, and J. Woods, "Interframe wavelet coding—Motion picture representation for universal scalability," *Signal Process.Image Commun.*, vol. 19, no. 9, pp. 877–908, Oct. 2004.

[27] M. Al-Mualla, C. Canagarajah and D. Bull, "Video coding for mobile communications", Academy Press – An imprint of Elsevier Science, 2002.

[28] S. Minami and A. Zakhor, "optimization approach for removing blocking effects in transform coding, "IEEE Trans. On Circuits and Systems for Video Technology, Vol. 5, No. 7, pp. 74-82, 1995.

[29] R. J. Hyndman, "The problem with Sturges' rule for constructing histograms", Monash University 1995.

[30] N. Cranley and L.Murphy, "Incorporating user perception in adaptive video streaming services", in Digital Multimedia Perception and Design (Eds. G. Ghinea and S. Chen), published by Idea Group, Inc., May 2006. ISBN: 1-59140-860-1/1-59140-861-X.

[31] Ffmpeg, http://sourceforge.net/projects/ffmpeg

[32] J. Mitchell and W. Pennebaker, "MPEG Video: Compression Standard, Chapman and Hall, 1996, ISBN 0412087715.

[33] www.tech.plymouth.ac.uk/spmc/staff/akhan/video_sequences/videoclips.html

[34] R. Feghali, F. Speranza, D. Wang and A. Vincent, "video quality metric for bit rate control via joint adjustment of quantization and frame rate", IEEE Transactions on Broadcasting,, Vol. 53, No.1, March 2007 pp 441-446.

[35] Telchemy application notes, "Understanding of video quality metrics", Telchemy, Feb. 2008. http://www.telchemy.com.