Toward Deployable Methods for Assessment of Quality for Scalable IPTV Services

Patrick McDonagh, Amit Pande, *Member, IEEE,* Liam Murphy, *Member, IEEE,* and Prasant Mohapatra, *Fellow, IEEE*

***Abstract*—Scalable video coding (H.264 SVC) is an attractive option for video service providers due to its ability to adapt a video's bitstream at the server to suit different network conditions and device characteristics. Lowering a video's bitrate can be achieved through reductions in frame rate, spatial resolution, and/or by increasing the quantization levels applied to the video sequence. In this paper, we first evaluate the effects of such scalability using some full-reference and no-reference video quality metrics, namely PSNR, SSIM, blocking, and blurring. No-reference metrics have the ability to capture the degradation in video quality caused by employing scalability in one or more dimensions. We study if conclusions drawn in previous works, which are based on well-known test video content, hold true for real-world broadcast content. We then discuss how, using these results for a particular content type, the use of no-reference metrics can be enabled in place of, or to supplement, existing widely used full-reference quality assessment metrics. We conduct an experimental analysis by transmitting video encoded at different scalability points over a lossy network to ascertain the effect of loss when scalability is employed in one or more dimensions. We analyze these results using a reduced reference metric called delta-blocking, which can detect visual damage of frames that causes a decrease in a user's quality of experience when perceived by the user. If the levels of packet loss are excessively high, this can lead the decoder to drop some video frames. To combat this type of frame loss, we propose a simple windowing algorithm that can automatically re-align the corresponding values for reduced-reference comparison, allowing for video quality monitoring to continue.**

***Index Terms*—SVC, digital multimedia broadcasting, quality of service, no-reference metrics, reduced-reference metrics.**

I. Introduction

Due to the increases in bandwidth efficiency provided by H.264 Advanced Video Coding (AVC), it has quickly become popular in networked video service applications. H.264 allows network and service operators to increase capacity on their networks while still maintaining the same standard of video quality. H.264/AVC provides gains in compression efficiency of up to 50% over a wide range of bit rates and video resolutions compared to previous standards such as MPEG-2 [1].

Manuscript received June 6, 2012; revised October 17, 2012; accepted March 21, 2013. Date of publication April 29, 2013; date of current version May 20, 2013. This paper was recommended by Associate Editor G.-M. Muntean.

P. McDonagh and L. Murphy are with the School of Computer Science and Informatics, University College Dublin, Dublin 4, Ireland (e-mail: patrick.mcdonagh@ucd.ie).

A. Pande and P. Mohapatra are with the Department of Computer Science, University of California, Davis, CA 95616 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2013.2255776

Networked video services, such as video-on-demand (VoD) are constantly increasing in popularity and are set to become a major feature in the Future Internet. As these services increase both the range and quality of the content they will provide, customers will expect high visual quality, where content is delivered with minimal latency and fewer interruptions due to loss. All of these constraints require that the networks providing these services are able to meet minimal delay constraints, while providing adequate bandwidth to deliver high quality content.

However, due to the variability of network conditions (particularly in the wireless domain), it may not always be possible to meet the constraints required by some networked services. This may be due to lack of available bandwidth in the delivery network or errors as a result of a low signal-to-noise ratio between the source and receiving node (in wireless domain). In this case, the Quality of Experience (QoE) of the customer will decrease unless remedial action is taken.

One possible solution to these problems is the use of scalable video coding. This can be used to decrease the bandwidth required for the video stream (at some expense of quality). Scalability can be achieved in 3 dimensions, firstly by increasing the levels quantization applied to the sequence, secondly by reducing the spatial resolution and finally, by decreasing the temporal resolution (framerate).

Before proceeding we first clarify the different categories of objective metrics which can be used for video quality assessment, namely; Full-reference, reduced reference and no-reference. Full-reference metrics require the original source sequence in order to assess the quality of the encoded or received video. Reduced-reference metrics do not require the original sequence but do require some information regarding the original sequence to assess quality. No-reference metrics do not require any information regarding the source video and use information from the encoded/received video only or other parameters (network measurements) when assessing quality.

A. Key Contributions

The key contributions of this paper are as follows:

1) Previous work in the area [2]–[7], which investigated the impact of scalability in one or more dimensions typically used video sequences from similar test sets (*foreman, park joy, blue sky*, etc.). In this work, *we investigate if their conclusions remain true for more complex*

(multi-angle, longer duration) broadcast content. This is carried out using full-reference metrics, namely Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) [8]. We have shown, even when using complex broadcast content, how video service providers can maximize quality for a given target bitrate by taking the nature of the content into account when choosing the parameters for the scalable layers. The outcome of this work provides a much more deployable and flexible solution than previous works.

2) We *motivate the use of no-reference [9] metrics* which are specifically designed to assess the impact of compression used by modern video codecs. We combine these no-reference metrics with some conclusions (both our own and those from previous works) drawn about impact of scalability on a particular content type to highlight their use when used independently from full-reference metrics.
3) The effect of network losses is perceived most easily in terms of blocking errors / artifacts. *We investigate the impact of the loss perceived in terms of increased levels of blocking for different scalability dimensions with bandwidth variations.* We show how no-reference metrics can be used at both the source and destination node and then compared in a *reduced reference fashion to detect errors in the visual output due to packet loss*. We name this reduced-reference metric "delta-blocking". We observe that, in both the quantization and spatial dimensions, the impact of loss is larger in higher layers than at lower layers.
4) Finally, when using this "delta-blocking", the direct comparison between original and received frame values is only possible when the metric is aware (and can take account) of any frames that could not be decoded so that offsets between the source and destination values can be computed. In order to do this, *we propose a simple windowing algorithm which can detect lost frames and compute the required increase in offset*.

Section II below discusses some related work in the area of scalable video coding and video quality assessment. Section III discusses our content selection, encoding setup and lossy transmission test. Section IV discusses the results, in terms of quality, for the evaluation of the encoding phase and the motivation of the use of no-reference metrics. Section V discusses the use of no-reference metrics in a reduced-reference fashion, as well as the results of the lossy transmission scenario. Section VI provides some conclusions which can be drawn and directions for future work.

II. Related Work

A. H.264 Scalable Video Coding

An extension to the existing H.264 AVC standard, known as H.264 Scalable Video Coding (SVC) allows for the encoding of a video sequence into multiple representations (or layers) at varying degrees of quality, with incremental increases in quality being achieved as layers are combined in a hierarchical fashion. The coding efficiency of scalable coding is superior to that of "simulcasting" the supported spatio–temporal resolutions and bitrates in separate bit streams [10].

A H.264 SVC bitstream is composed of a collection of layers with the lowest quality representation of the video being referred to as the "base" layer along with incremental increases in quality being achieved through the addition of subsequent "enhancement" layers. The combination of the base layer and all the enhancement layers provides the highest quality representation of the video.

This scalability of the video stream can be achieved through the degradation of video quality in 3 different dimensions: spatial, temporal and quantization. A single degradation step between one layer and the previous lower layer can involve the reduction in quality in one or more of these dimensions. Spatial degradation refers to the reduction of the resolution of the source video, in our case we use dyadic downsampling. Using this mode, the resulting downsampled version (at the lower resolution) must now be re-sized or scaled to fit the screen area previously occupied by original, higher resolution sequence.

Temporal scalability is the reduction of the frame rate to a lower frame rate, thus saving bandwidth at the expense of poor representation of any motion present in the video. The effects on quality due to temporal scalability may vary widely depending on the level of motion present within a particular scene. For the case of quantization, as we transition from the highest enhancement toward the base layer, the frames of each layer are subject to higher levels of compression or quantization, this is primarily controlled by the quantization parameter (QP). For a fixed resolution and frame rate, the relationship between lowering QP and output video quality can be approximated as an exponential growth function [11].

B. Network Transmission of H.264 SVC

Content transmission problems such as loss or, high amounts of jitter or delay will result in service degradation due to the required video data not being available for decoding and rendering. This will lead to playback stoppage or visible errors with decoded frames. The IPTV service provider may be able to adapt their video service (in terms of modifying the video content or how the content is transported) when network conditions are poor in order to minimise the impact on their service [12] [13]. However, in order to do this information must first be fed-back to the service provider to make them aware of service delivery issues. The authors of [4] provide a full-reference metric which can be used to provide this information when scalable video encoding is employed.

In [14] the authors propose an architecture when Intra-coded frames are prioritized in order to improve stream robustness in the case of loss. This will ensure that any errors that occur as a result of loss will be constrained to within a single GOP. Their architecture makes use of 2 priority classes for wireless (802.11) transmission, which is not sufficient for the SVC case, where multiple layers may be employed and the decision as to which layers to prioritize requires a more fine-grained approach. Similarly, [15] performs traffic and quality characterization of SVC video but uses a single layer. [16]

proposes a SVC evaluation on IP networks, but use only full-reference PSNR metric.

As mentioned above, existing works don't compare the effect of different scalability options on degradation of perceived visual quality. [17] considers the use of a rate control mechanism with SVC for video multicast over wireless networks. However, they consider six temporal resolutions/ layers. The reasoning behind this choice is not justified. Similarly, Fallah *et al.* [18] use SVC in conjunction with link adaptation without any emphasis on which dimension to degrade.

C. Video Quality Estimation

Assessing the quality of encoded or transmitted video, requires the assessment of network parameters regarding the transmission process (e.g. RTP header inspection) or from the decoded video itself. This data (from either or both of the sources) can then be used to infer the impact of the transmission process on the received video.

The goal of any proposed objective metric is to provide a rating which would be closely correlated with subjective ratings from a collection of viewers. PSNR and SSIM are two widely used objective full-reference metrics. These require that both the decoded and the originally transmitted video be available to enable computation of a quality score. We chose these metrics for a number of reasons. These metrics are widely used in the literature, therefore the use of these metrics will allow for easier comparison with existing works. While it may be argued that PSNR does not correlate closely with perceptual quality, it is however widely used in industry to assess encoding quality/codec comparison; for example in the x264 encoder,[1] Xiph.org,[2] which is what this work aims to achieve. SSIM closely relates to perceptual video quality [19] and has been widely used in research community.

The EvalVid framework [20] allows evaluation of H.264 videos using subjective metrics (such as MOS) and objective metrics (such as PSNR). As such, any derivations of EvalVid, such as EvalSVC [21] also restrict video quality evaluation to PSNR based metrics, which has the disadvantages of being a full-reference metric and thus, not practical in consumer delivery scenario. Although the use of full-reference metrics such as PSNR and SSIM is beneficial for experimental analysis, it is not feasible for deployment purposes. In [22] the authors undertake a subjective evaluation of SVC however, the effect of the transmission process (and the subsequent requirement for no-reference evaluation) is not taken into account. Seeling et al. [23] present an encoding quality comparison between H.264 and VP8, another recently developed codec. [24] propose SVC evaluation using a neural-network based mapping from objective network measurements to subjective user ratings, but it requires extensive off-line user-trials. A performance analysis of SVC is undertaken in prior work, however, they user PSNR (full-reference) as the only metric [25].

The visual features based models employ measurements of blurring and blocking in video [8]. Such no- or reduced- reference metrics can be used for quality-adaptation in practical scenarios by having an application level feedback mechanism. We use these metrics to evaluate SVC video in our experiments.

[1]http://x264.nl
[2]http://xiph.org

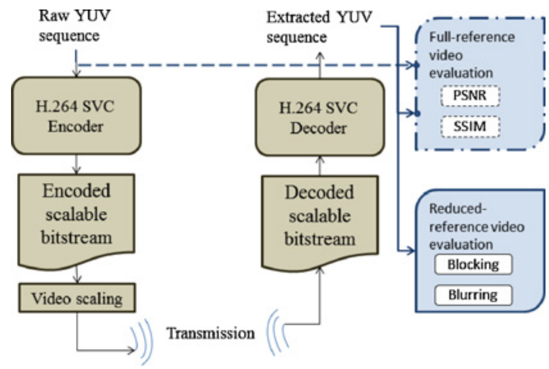

Fig. 1. Block Diagram outlining Experimental Analysis.

III. Experimental Outline

The main objective behind this work is to quantify the effect of scalability options and network transmissions on the end-user experience of watching a video. The scalable codec provides three dimensions of scalability namely; temporal, spatial and quality scalability. Therefore, any video can be scaled in any of these dimensions to counter the fluctuations in network bandwidth or degradation in user quality. A specific trajectory (i.e. a specific choice such as first scaling in quality and then in spatial resolution) is referred to as 'degradation path' in our experiments. Our objective is to investigate the quality degradation along these degradation paths, which can instead be used to choose suitable scalability options to satisfy bandwidth or quality constraints.

A. Video Quality Analysis and Bandwidth Savings

The block diagram of Figure 1 provides an outline of our experimental analysis. The raw YUV sequence is encoded using H.264 SVC encoder into three different streams each stream scaling down in one dimension only – temporal, spatial or quantization. It is at this point where we measure quality using both full and no-reference metrics by producing YUV stream from the encoded video.

Upon completion of the encoding step and extraction of the required streams for analysis, these streams are then converted into raw YUV sequences where they are used for comparison against the original sequence to assess the impact on video quality due to the degradation in the spatial, temporal or quantization dimensions. For the case of spatial degradation, the lower resolution sequences are first upscaled to the original sequence resolution in order to allow comparison to take place. An equivalent procedure takes place for temporally degraded sequences, in order to construct a video for comparison. This is achieved by repeating the frames found in the temporally degraded sequence the requisite number of times to produce the same frame-rate as original sequence. The video quality analysis using the original and degraded sequences is carried out using the MSU Video Quality Measurement Tool [26], which provides implementations of the metrics stated above.

When discussing bitrate savings as content is compressed, typically rate-distortion curves are used, however, if we wish to apply our results to a more general case i.e. categories of videos (videos of a similar content type), explicit references to bitrate only apply to one video sequence. In our work we express bandwidth savings as a percentage of bandwidth saved to provide a more general result.

In order to calculate the bandwidth savings achieved for a particular sequence w.r.t the bandwidth required by the original full quality sequence, we use a tool provided with JSVM to calculate the average bitrate required for the layer that we wish to analyze and express this as a percentage of the bandwidth required by the full quality video.

Section 4 below presents the results of our analysis, where we investigate the impact on quality for all 3 scalable dimensions for the 3 different content types.

In the next step, we introduce a reduced-reference metric named "delta-blocking". This metric uses decoder behavior (when faced with loss) as a trigger for detecting visible errors present in playback. This metric also presents a simple windowing algorithm which can be used to detect and correct frame loss which would lead to a mis-alignment of reduced-reference data, causing spurious errors in the results. This metric is validated by transmitting video content over a wireless network with a range of loss rates. This shows, firstly how errors in playback can be detected (delta-blocking) and secondly, demonstrating how the windowing algorithm operates and can be used to re-align reduced reference data if full-frame loss occurs.

Finally, we present an analysis of the impact of loss for a series of points in each scalable dimension. The video streams are transmitted over a wired network with finite bandwidth and with some loss introduced. We then decode the video to obtain an extracted YUV stream and then evaluate the impact of this loss using no-reference metrics (used in a reduced-reference fashion). In case of network losses, full-reference video evaluation tools are not usable because the original video and received videos are not aligned together (have different number of frames because of frame drops). These results are presented in Section 5.

B. Video Database

For the purpose of our experimental analysis, the decision was made to use clips from broadcast television content and not the standard video quality assessment sequences, (such as *blue sky*, *foreman* and *park joy*) which are typically used for experiments of this type. The reason for this is that, these clips are generally quite short in length and only contain a single angle / camera shot, with varying degrees of motion. Real-world broadcast content is typically composed of a number of different angles, each with varying degrees of motion and complexity. The nature of this broadcast content is also further characterized by the switching between these different angles/shots. However, the standard assessment sequences typically do not contain any switching between shots and thus lack the complexity of composed broadcast content.

TABLE I
YUV SEQUENCE PARAMETERS

Title	Soccer	News	Trailer
No. Of Frames	791	1504	1402
Frames Per Second	30	25	24
Seq. Length (sec)	26.4	60.2	58.4
Motion	High Motion	Low Motion	Medium Motion
Scene Cuts	Frequent	Infrequent	Frequent
Texture Similarity	High	Medium	Low

As a result, any conclusions or recommendations about degradation when using the standard sequences would only hold true for that particular video (or video of a very similar nature). Supposing we were to follow this principle, if for example, we were to compose a broadcast video made up from a collection of these different standard sequences (since we assume people do not watch the same ten seconds of video repeatedly), we would only be able to achieve an optimal degradation path for one of the sub-clips, thus making the choice of an optimal degradation path for the other clips redundant. Therefore, the use of the prepared sequences for selecting the optimal degradation path is only accurate when all of the content in the video sequence is of a very similar form.

In our approach, we use broadcast content which among the 3 videos have varying levels of motion, scene cuts and coding complexity. The reason for doing this is to investigate if we can infer some knowledge regarding the degradation at a sequence level, which is what would be utilized in a real-world broadcast service, as opposed to the per shot/angle level which the standard test sequences can only provide us with. Thus, since our approach uses the concept of whole sequences and the knowledge about degradations for that type of sequence, this provides us with more flexible and deployable concepts than those previously presented in the area.

C. Content Parameters and Descriptions

As detailed above, 3 different videos were used; *Soccer, News* and *Trailer*. Prior to encoding, all the videos were raw YUV sequences in the 4:2:0 Chroma subsampling format. At full resolution, the videos were all 1024 pixels wide and 576 pixels in height. The video specific parameters such as frame rate, sequence length, etc. are summarized in Table I.

In order to distinguish between the different levels of motion, angle changes and sequence complexity a brief description of each sequence is provided below:

1) *Soccer:* This sequence contains a number of different shots from a variety of angles, all but one containing a large amount of movement. with players moving around throughout the frame. All of these shots are zoomed, close–up shots. A single, wide angle, low motion sequence is present at the end. Angle changes are frequent and this sequence is a representative of typical sports videos.
2) *News:* Two different shots are present in this sequence. Both are of a newscaster in the center of the frame with a large static background to the side and above. Both contain very low amounts of motion with the

Fig. 2. Screenshots for (a) Soccer, (b) News & (c) Trailer.

movement being predominantly concentrated in the face and sometimes the body of the newscaster. There is very infrequent (15 seconds or greater) between the switching of these shots. This video is typical representative of news content in broadcast television.

3) *Trailer:* This sequence contains a large number of different shots due to its nature. The content contained within the individual shots is typically, of low to medium levels of motion, with the exception of a small number of high motion shots, as is typical with movie trailers/ advertisements.

These sequences were obtained from online sources such as YouTube in the MP4 container format before being converted to raw YUV format. Screenshots from each of the 3 sequences are presented in Figure 2.

D. Encoding using JSVM

The 3 different videos in YUV format were encoded using the JSVM H.264 SVC encoder [27] at a variety of different resolutions, frame-rates and quantization levels. For the spatial dimension, the resolutions used for encoding were 1024x576 at full quality, 512x288 and 256x144 at the base layer. The choice was taken not to resize or crop the video content to match standard broadcast content so that no video content was altered or lost, which could affect the results.

For investigating the effect of increased quantization for each of the content types, the following quantization parameters (QPs) used were: 20, 26, 32, 38, and 44 (note that the higher the quantization parameter, the greater compression achieved, at the expense of video quality).

Due to the differences in source frame-rates, the values for temporal degradations were different for each sequence. However, it can be summarized as, from the top layer, the next highest layer is achieved by halving the previous layer's frame rate. This step is repeated a further 2 times until a frame rate for the base layer is obtained.

Note, in all of the above videos a fixed Group of Pictures (GoP) size of 16 was used; 1 Intra (I-) frame followed by 15 Predicted (P-) frames, Bi-Predicted (B-) frames were disabled for these experiments to simulate the worst-case scenario for encoding complexity, providing an upper bound on bandwidth requirements. For example, we may have a sequence of the same content type, that is slightly more complex to encode. For sequences with framerates between 24 and 30fps an I-frame (which can be used for resynchronization in the case of loss) is transmitted approximately every 0.5 – 0.7 seconds. In the case of data loss, subsequent inter-coded frames will be decoded with errors. The minimum time required to halt the error propagation (through successful decoding of an I-frame) will be 0.5 – 0.7s, however if a subsequent I-frame is dropped, or if two I-frames from 2 different GoPs are dropped, the error propagation will have a duration greater than one GoP. A shorter GoP may be employed to improve the robustness of the stream in the face of loss, however, this will increase the bitrate of the stream. The video service provider will have to consider the tradeoff between loss robustness and stream bitrate. JSVM's de-blocking filter was enabled for these experiments for all blocks. Motion estimation was carried out using pictures of the highest enhancement layers and using the fast search algorithm with a range of 32. Signal-to-Noise ratio enhancements were encoded using Medium Grained Scalability. The base layer was encoded as an AVC compatible bitstream.

The output of each encoded sequence is a raw H.264 scalable bitstream containing all the layers specified prior to the encoding step. This H.264 bitstream can then be used to extract a "sub-bitstream" containing the layers that are required for analysis. In the case of the analysis of the effect of temporal scalability, in order to produce the lower frame-rate versions the QP 20 version of the sequence was used, since this allows for direct comparison between these two dimensions. JSVM provides a tool *DownConvertStatic* in order to produce a temporally-downsampled bitstream which is then converted to YUV for comparison using the metrics. In our evaluation we specify the full video, containing all the layers as having the "original bitrate", against which all other sub-bitstreams are compared against with regards to video quality and bandwidth savings.

E. Characterizing the Effects of Network Losses

The purpose of this experiment is twofold: to investigate the degradation in one particular scalable dimension in the

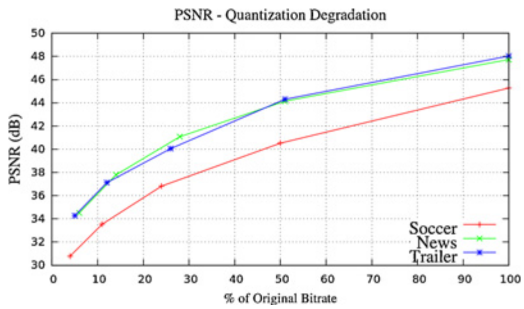

Fig. 3. PSNR for Increased Quantization.

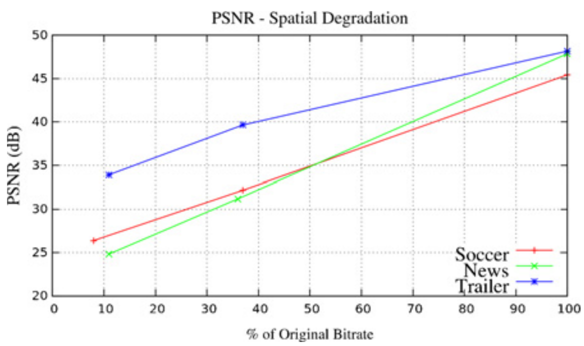

Fig. 4. PSNR for Spatial Degradation.

presence of network loss, and secondly, for a particular content type, determine the scalable dimension which provides better perceptual performance than others in the presence of loss. The introduction of errors in the decoded video due to network losses requires that a no-reference metric be used, because full reference metrics such as PSNR, SSIM cannot be measured in any accurate manner. The misalignment of frames due to frame losses makes these metrics unsuitable for such scenario. Furthermore, it is not possible to have the original video at the receiver side in a practical deployment/ live service environment. Such tests can only be used for static evaluation of video quality but not for dynamic modification of the service when loss is experienced.

In order to do this, we use the no-reference blocking metrics, which give a perceptual experience and can be practically deployed in a wireless/ cellular network. The MSU Blocking and Blurring [28] metrics are examples of no-reference metrics.

The MSU Blocking metric is used to measure the degree of blocking/ macro-blocking present in a source sequence. In a codec such as H.264, where visual data is encoded as a series of macroblocks, the primary effect of loss will be visible in terms of lost or damaged macroblocks, thus we argue this metric is well suited for assessing the impact of loss in this scenario. In our results we calculate a "delta" blocking value, that is a difference in the maximum or average blocking value between the original and received (with loss) sequence.

In the case of quantization and spatial degradation, we use the difference between the maximum blocking value for the original and received sequence as it indicates how the blockiness is increased due to losses in the received sequence compared to the highest level of blocking in the original sequence. Thus, we are comparing the worst-case in both

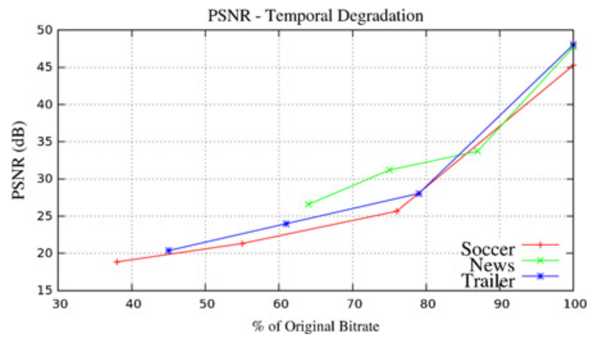

Fig. 5. PSNR for Temporal Degradation.

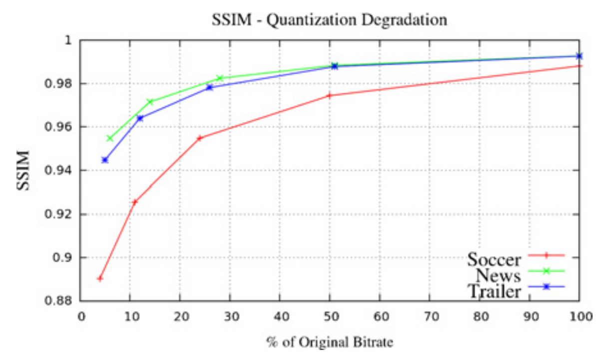

Fig. 6. SSIM for Increased Quantization.

sequences to show how loss and scalability in that particular dimension are related. In the case of temporal degradation, while the maximum value is important, the increase in the difference between two sequences average blocking values provides a better indication of the level of video quality degradation. This is due to the fact that once a frame is damaged due to losses, it must be repeated the requisite number of times to achieve the original framerate thus increasing the average blocking value for the entire sequence.

IV. Results

A. Full Reference Metrics: PSNR, SSIM.

Figures 3 – 8 present the results of the impact on video quality for each of the 3 degradation dimensions with respect to the full-reference metrics Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM). We should observe that due the different nature of each of video content types, there will be some variance around the PSNR value for the full quality video, despite the same encoding parameters being used.

As can be seen in Figure 3, we can observe the effect on PSNR for increased levels of quantization. We can see that for all 3 content types, the relationship between increased quantization and PSNR is approximately logarithmic. When compared with the case for spatial degradation in Figure 4 and temporal degradation in Figure 5, we can see that for an equivalent saving in bandwidth, quantization outperforms both in terms of maintaining perceptual quality for all 3 videos. In Figure 4 we can see that the relationship between spatial degradation and PSNR can be approximated in all 3 cases as a linear relationship.

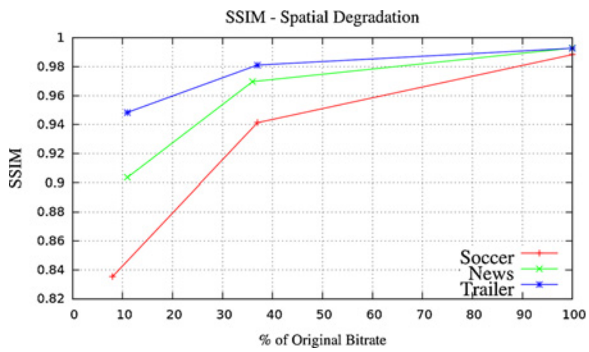

Fig. 7. SSIM for Spatial Degradation.

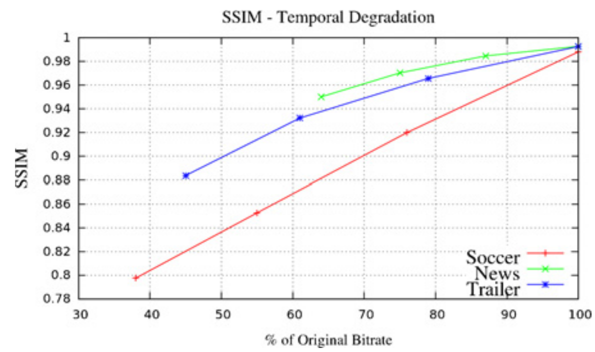

Fig. 8. SSIM for Temporal Degradation.

Using spatial degradation, we can observe that each of the video content types offer an approximately equivalent bandwidth saving for a similar degradation in PSNR from each of their respective full-quality versions. However, we can also observe that the effect of spatial degradation is more pronounced in the case of *news*. This can be partially attributed to the reduction in quality of static items such as; logos and titles, which are present throughout the sequence, thus lowering the per-frame value throughout the sequence.

Figure 5 presents the results for the loss of video quality, in terms of PSNR under temporal degradation. Here we can observe that the low motion of content of *news* performs best at maintaining quality as the frame rates decrease. However, we can see that while *news* is able to maintain a better quality that the other 2 sequences, it does not perform as well in terms of bandwidth reduction. This indicates that when the reduction in the number of frames per second is the same for each video (halving each time), in the case of *news*, although having the greater quality, due to the relatively low amount of motion data in the full frame rate sequence, as the frame rate decreases, this does not lead to a large reduction in bandwidth.

However, we can observe that for *soccer* and *trailer*, which both contain more motion data, the reduction in frame rate leads to the discarding of greater amounts of motion data (at the cost of quality) leading to greater bandwidth savings.

Presented in Figures 6, 7 and 8 are the results for impact of visual quality, in terms of SSIM for each of the three degradation dimensions. Again we can see that quantization outperforms both spatial and temporal degradations for each content type in terms of the trade-off between video quality and bandwidth savings.

In Figure 6 we can see that once again the relationship between increased quantization and SSIM can be approximated as a logarithmic relationship. We can also see that the *soccer* sequence suffers a greater decrease in SSIM than the other sequences. However, this would still be classed as "fair" under the quality of experience mapping presented in [3].

Under the use of spatial degradation as presented in Figure 7, we can observe a similar behavior for the case of *soccer*, this would indicate that the higher motion content of this sequence is particularly sensitive in terms of structural similarity to any degradation in visual quality, either due to increased quantization or due to spatial degradation. In this case however, for the lowest spatial resolution, the resulting SSIM value would be classed as "poor" under the scale presented in [3]. We can also observe that for the *news* sequence, under spatial degradation, the effect on SSIM is much more pronounced than the case for increased quantization for an equivalent bandwidth. Again, this can be attributed to the greater impact on the quality for static items such as, logos and titles under spatial degradation.

For the case of temporal degradations, as shown in Figure 8, we can observe that, again, while *news* is able to maintain the best quality in terms of SSIM, it does not provide the same bandwidth reduction as the other two sequences.

Since we are measuring the structural similarity, for the case of low motion sequences, the structural difference between two (or more) sequential frames will be lower than the difference between frames of a sequence with higher levels of motion. For example, in the case of the higher motion *soccer* sequence, we can see for the level of reduction in frame rate the decrease in SSIM is much greater and would be classed as "poor" while at its lowest frame rate *news* would be classed as "good". For all 3 content types, the relationship between decreased frame rate and SSIM can be approximated as a linear.

B. Comparing Scalable Dimensions by Video

Figure 9 (a) – (f) presents the results of the impact of each scalable dimension, in terms of PSNR and SSIM, for each video. These are the same results as in Figures 3 – 8 but have been grouped by video to allow for the comparison between the effects of each scalable dimension for each video.

Again we can see that, for both PSNR and SSIM, scalability in the quantization dimension provides the best video quality for a given bandwidth saving for all 3 videos. Furthermore, we can observe that both quantization and spatial scalability provide a roughly equal saving in bandwidth at the lowest level of scalability (highest quantization/lowest resolution).

Temporal scalability, in all 3 videos performs the worst in terms of bandwidth efficiency versus video quality. In the case of the *soccer* and *trailer* sequences at its lowest frame rate, temporal scalability achieves a bitrate which is approximately 40%–45% of the original, however in the case of *news*, the bandwidth required is approximately 62% of the original. As stated before, this is due to the fact that *news* contains low levels of motion, relative to the other two sequences, therefore there are lower amounts of encoded motion vector data to remove at each temporal layer, thus the achieved bandwidth savings for the *news* sequence are not as great as

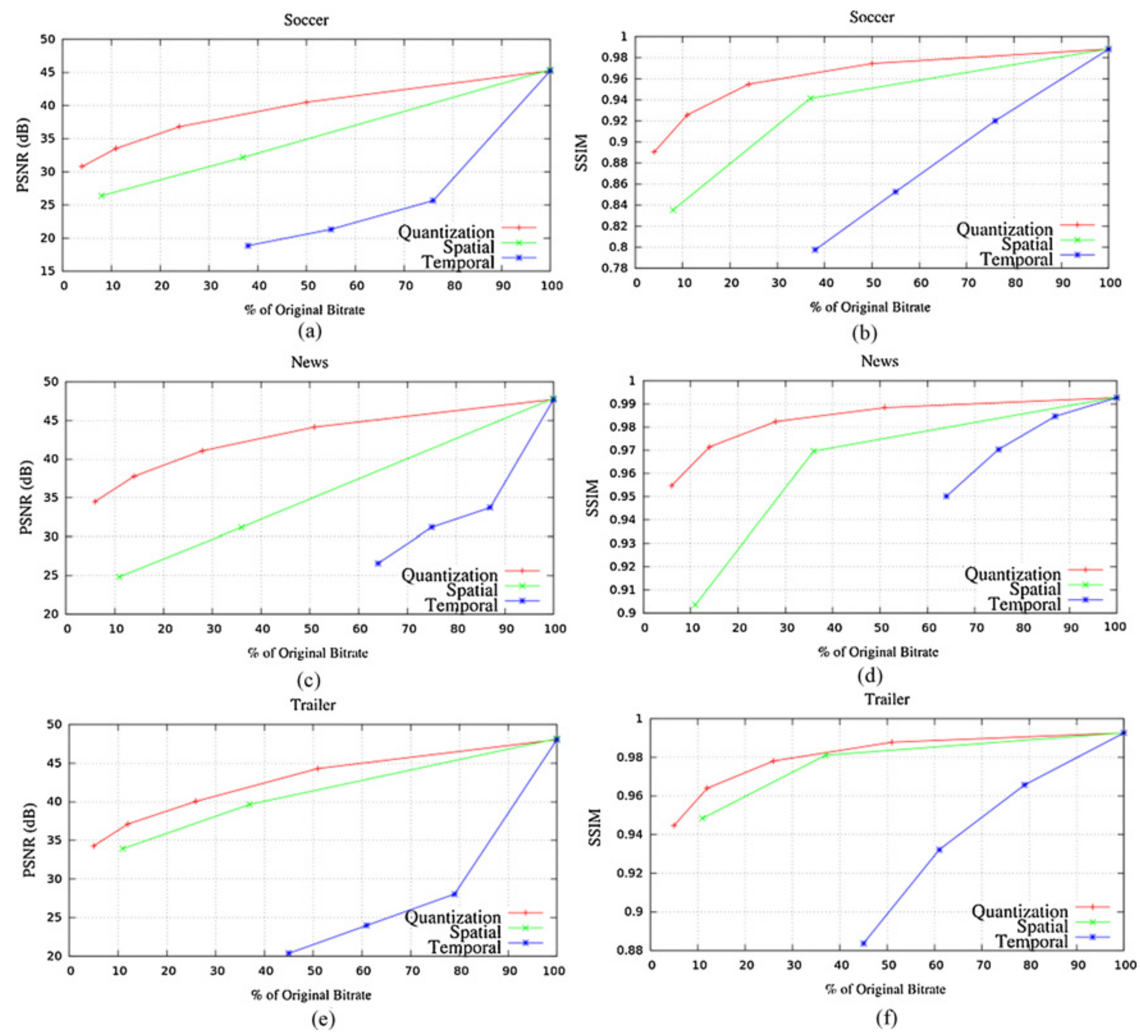

Fig. 9. Comparing impact on PSNR and SSIM for each scalable dimension.

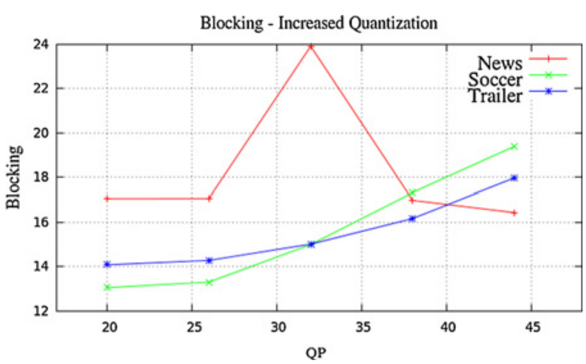

Fig. 10. MSU Blocking - Increased Quantization.

for the *trailer* or *soccer* sequences. It can be noted that our results and conclusions, which are made about full complex sequences, found in broadcast content correspond closely to the results presented in [3], [29], [30], which use the standard test sequences.

C. Content Based Degradation Path

As can be seen in Figures 3 – 8, the overall trend, in terms of degradation of (PSNR or SSIM) for a given dimension is approximately the same across all 3 content types. There are small degrees of variation, e.g. the lower bandwidth savings provided by *news* under temporal degradation. This variation would be expected due to the heterogeneity of sequences which could be assigned to the same content "category". However, none of this influences the fact that a selected degradation path would have a similar impact on quality across all 3 content types in terms of PSNR or SSIM.

D. No Reference Metrics - MSU Blocking and Blurring

Figures 10, 11, 12 and 13 present the results of increased levels of quantization and decreased resolution for each of the 3 videos with respect to MSU Blocking and Blurring.

In Figure 10 we can see that in the case of *soccer* and *trailer* that, as larger amounts of quantization is applied to the source video, the levels of appearance of blocking in the output video is increased. This is due to the fact that as we increase quantization, less and less (spatial) detail is preserved and smoothness of transition between two consecutive parts of the image is reduced, leading to sharp transitions between the two and causing increased levels of blocking. As a consequence of this, as can be seen in Figure 11, this leads to lower levels of blurring at macroblock boundaries as the smoothness

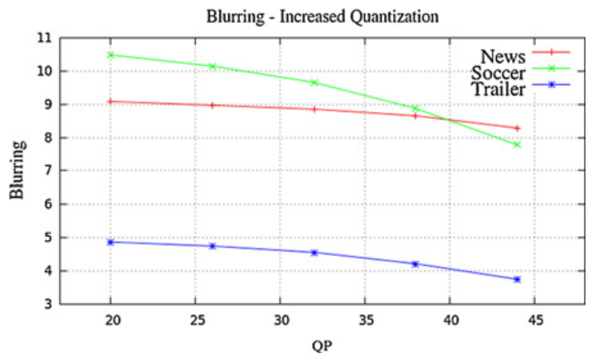

Fig. 11. MSU Blurring - Increased Quantization.

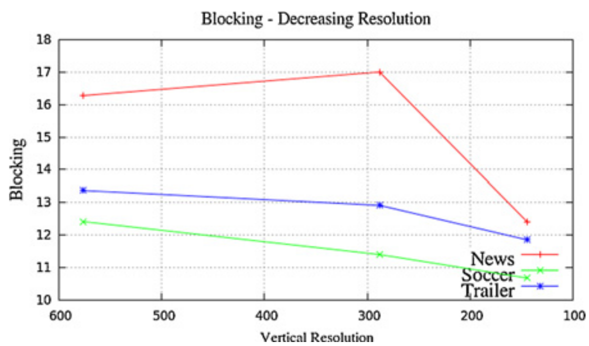

Fig. 12. MSU Blocking - Decreasing Spatial Resolution.

of the image is replaced by sharper transitions. Conversely, within each macroblock, the compression process results in increased blurring, as there is a reduction in contrast between neighboring pixels.

In Figure 10, for the case of *news* we can observe a large increase in the level of blocking when a quantization parameter of 32 is used. Upon further investigation it was found that the blocking metric is particularly sensitive to long stretches of horizontal (or vertical lines) which are present throughout the *news* sequence. It considers these, mistakenly to be artifacts of the compression during the encoding step and therefore gives them a higher value. An example of this is the horizontal border between the subtitles/headlines (at the bottom of the frame) and the camera shot of the news presenter, as these are present throughout the sequence our average is increasing at each frame, giving us our inconsistent value.

We believe that the presence of this horizontal border and its particular interaction with the compression algorithm generates this anomalous results with quantization parameter 32. However, the metric that we introduce later in this paper, calculates the change in this blocking value caused by network impairments. As a result, this metric is not sensitive to such anomalous results.

Figure 12 presents the results for MSU Blocking for decreasing spatial resolution. Here, we can observe that as the spatial resolution is reduced, the overall trend is that the levels of blocking are decreased by varying amounts. This would indicate that as the spatial resolution is decreased the overall blockiness of the sequence decreases as the spatial downsampling process is combining multiple pixels leading

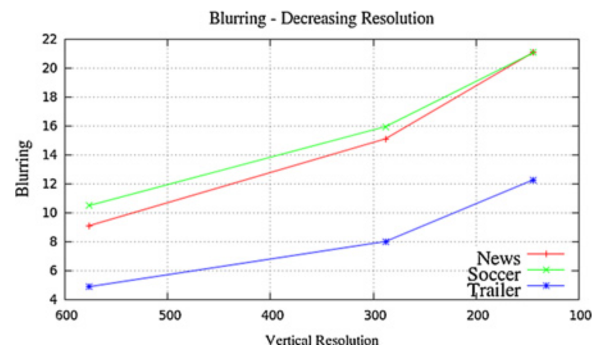

Fig. 13. MSU Blurring - Decreasing Spatial Resolution.

to smoother transitions between two consecutive parts of the image. Furthermore, when the spatially reduced frame is displayed at its original size, the spatially reduced frame appears blurred, this phenomenon is confirmed in Figure 13.

This trend is most pronounced in the case of the *news* sequence. This is most likely due to the fact that the background of the *news* sequence, for the most part, contains content which is shot out of focus (and appears smooth at the full resolution), which, when spatially down-sampled, leads to an even greater smoothening of the transitions between these areas in the image, leading to the greater observed decrease in blocking between the medium and low resolutions.

While the no-reference metrics do not provide as much information as full-reference metrics, they do capture the increased levels of artifacts introduced as part of the compression process, which affects customer Quality of Experience The behavior across all 3 videos is varied due to the differences in content type but is what would be expected as greater levels of compression are applied, eliminating the need for full-reference assessment. Therefore, these no-reference can be used to assess quality degradations in terms of compression artifacts, and when combined with the conclusions drawn about the impact of scalability for a particular type of video content (motion levels, complexity), supplementing or replacing full-reference metrics. Since we are making our measurements at a sequence level, there is no panacea which will satisfy *all* sequences of a certain type. However, for a deployed system some tradeoffs must be made and performance might have to be compromised at times to enable a realistic solution. We argue that our work shows how a deployed system, which is able to make judgements about scalability at a sequence level, might be employed.

V. Reduced-Reference Assessment and Automated Re-Alignment of Sequence Data in the Presence of Frame Loss

In this section we demonstrate how no-reference measurements made at the video server and at the customer's terminal device can be used to detect if any visual damage or frame loss has occurred to the video content due to packet loss. This information may then be used to trigger a request to degrade the video quality by shedding SVC video layers until such a point where the bandwidth of the downscaled video can be supported. We also present an algorithm which can be used to realign the no-reference values for comparison in the presence of total video frame loss. This is performed as a separate experiment from the assessment of loss on different scalable dimensions as to allow to a clearer explanation of the algorithm.

We argue that video *service* quality from the perspective of a paying subscriber is essentially a binary decision; it is either acceptable or unacceptable with regards to their expectations of the service (*video quality* is subject to a more finer grained scale, typically MOS) . After the content has been encoded, from the location from which its transmitted to the point where it is viewed, the only factor influencing quality is loss [3]. Furthermore, with the multitude of sources being currently used to generate video content, ranging from professional high-definition camcorders to cellphones with varying degrees of capture quality, the notion of video *service* quality being associated purely with clarity or fidelity of representation no longer holds true.

With the emergence of residential broadband services such as Digital Subscriber Line (DSL) and Hybrid Fiber Coax (HFC) it is not unusual for these network operators to provide customers with network equipment such as residential gateways and set-top boxes to access their broadband connection and networked services. A set-top box performs two major functions, firstly as stated above, it provides a connection to the operators IPTV network and secondly, it provides the decoding functionality so that streamed video may be displayed to the viewer. Since the network operator will generally configure the set-top box with the necessary parameters required to connect to their network, there is an opportunity for the network operator to install additional software to manage the connection. This software may take the form of firewalls, authentication handling and in our case, monitoring software in the form of the blocking metric.

1) *Our Approach:* As stated previously, for macroblock based codecs such as H.264, the most obvious manifestation of these errors due to packet loss are likely to take the form of visibly damaged regions corresponding to missing macroblock data. We propose a reduced-reference measure to quantify the end-user perception of video quality (in terms of blockiness) and report to the network service provider with a feedback of the perceptual loss introduced by the wireless network, without the need to unnecessarily drill down to a more complex assessment strategy. For a deployed service such as IPTV [31], [32], a more complex no-reference metric may require for example, information regarding the content within the video to perform its evaluation. However, for any large-scale IPTV deployment and in particular for broadcast television, this method will prove difficult to implement. Again, for any carrier grade service, customers expect video content to be delivered free of errors and any disruption to the viewing experience, due to visible artifacts will lead to an unacceptable level of satisfaction with their video service. However, in some cases the levels of loss can lead to a situation where a frame cannot be reconstructed at all, leading to the frame being dropped. In the case of a reduced reference metric, such as the one presented below, where the decoder compares the blocking value between the original source frame and the one computed for the frame received, the loss of a frame can lead to a misalignment between the correct values used for comparison, leading to spurious results. In this case, it is necessary for the monitoring system, when faced with the next "good" frame to be able to quickly re-align itself with the correct value so that the ability to monitor quality in the presence of frame drops can be maintained.

Let B_S(i) denote the blocking measure at source for frame *i* and B_D(i) measures the blocking observed by the end user while viewing frame *i*. Many metrics have been proposed in research literature for measuring blocking values in a no-reference manner [33], and [34]. This work differs in so far that ADB is 1) performed for a video sequence and 2) used to detect the occurrence of macro block errors and not the result of the compression process. In this work, we use the MSU Blocking metric for computation of these values. Intuitively, we take a difference of the source and destination blocking measurements to get a reduced-reference measure of blocking, referring to it as ADB (Absolute Difference of Blocking measures). Since B_S(i) is computed at the video source after the encoding process, for the same frame, the value of B_D(i), should be the exact same at the receiver, unless something has affected the visual content of the frame. This difference in blocking values between the source and receiver values, or "delta-blocking" enables the decoder to detect frame damage and to quantify the level of damage caused.

$$ADB(i) = |B_S(i) - B_D(i)| \tag{1}$$

To test the accuracy of this simple solution, we conducted an evaluation using an IEEE 802.11 network testbed delivering video content; using a H.264 encoded video stream, and MSU Blocking metrics values to get ADB. We next correlate ADB with packet loss events. The packet loss event was forced by specifying a packet loss rate (PLR) on the wireless device (in this case running the Madwifi driver) the video was then decoded (with losses) the resultant blocking values were calculated. However, the ADB values show many spurious measurements which are due to misalignment of frames due to frame loss event. The result is not accurate because some packet loss events cause frame losses which are unnoticed by the decoder. Hence, the i^{th} frame at decoder may correspond to the $(i+1)^{th}$ frame in source video after first frame loss event. A simple absolute difference (ADB) will not account for these events, leading to erroneous values.

However, there is an interesting observation: ADB(i) does capture the effect of packet losses in blocking introduced in the viewed video. The effect of different packet losses is different depending on motion/ content of exact frame in video and type/ importance of lost packet. A packet loss can lead to two errors: loss of information required to reconstruct next frame or incorrect reconstruction of next frame because of lost partial information regarding the frame. The former leads to a temporal loss in video quality, while the latter leads to blocking artifacts (unexpected visible errors). Lack of time-stamping information in received video makes it difficult to match the frame to each other. There exists a lot of relevant

Algorithm 1 : ADB_MF(i)=MatchFrameAlgorithm(B_S(i), B_D(i), W)

// W is size of window used.
NumWindow = Length(B_D(i) / W)
NumLost = 0 //Initialize to 0
for j=0 : (NumWindow - 1)
for i=1 : Window
ADB (i + j * W) = abs(B_S(i+j+NumLost) - B_D(i+j*W))
endfor
for last M frames compute the ADB values with and without 1 frame loss and find minimum (min)
if (min) == $ADB_{withloss}$
NumLost = NumLost + 1

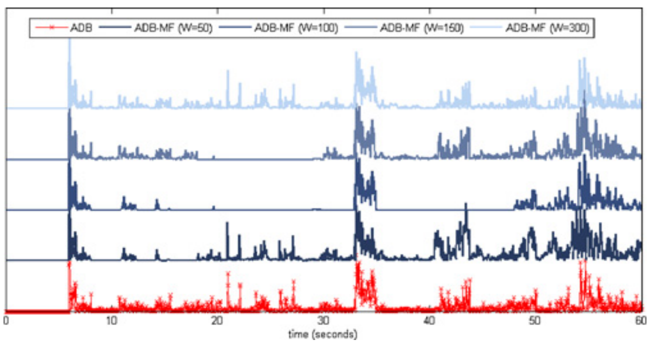

Fig. 14. The smoothing of ADB values obtained with different window sizes (indicated in parenthesis). Packet loss rate is 0.10%.

work w.r.t. sequence matching in other domains that we believe might be relevant to the our problem here of frame matching. In particular, the application of some well-known algorithms in bioinformatics that deal with sequence alignment with gaps [35] and statistical significance estimation of such alignments [36]–[38]. However, these sophisticated algorithms have high computational complexity, thereby we next propose a simple algorithm.

We use a windowing approach to remove the errors introduced in spatial (blocking) measurements due to frame losses. The main idea of the windowing algorithm is to compute *ADB(i)* values over a small window. Before moving to next window, it estimates whether a packet loss event occurred in the past window or not this is done by measuring the *ADB(i)* values of last M (say M=10) frames with the *ADB(i)* values computed with source window shifted by 1 element. If the minimum of these two values is the ADB value with shifted element, we increase a counter *NumLost* which is used as an offset in computation of *ADB(i)* for future windows. This process is repeated for each window. These values are called *ADB_MF(i)*. Algorithm 1 explains this process.

Figure 14 shows the smoothing obtained for an experiment conducted with a prepared source video with different window sizes (decreasing window size from top to bottom) . The red (bottom) plot shows *ADB(i)* values. It can be seen that different windows show different regions of smoothening with the best performance obtained by W=100. We can use this value, or alternatively, we can take the minimum value obtained by different window sizes.

Figure 15 gives more results illustrating the improvements gained by MF algorithm over direct ADB computations.

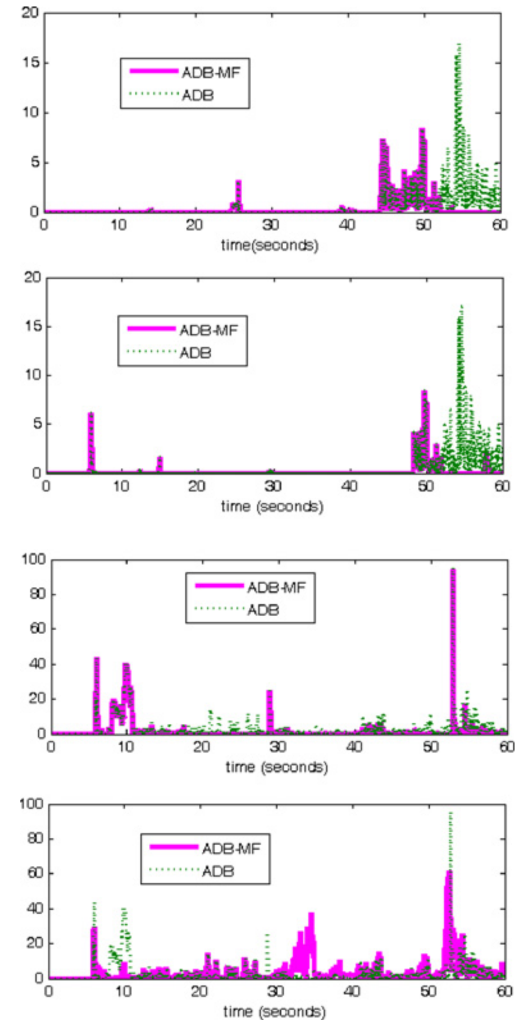

Fig. 15. The smoothing of ADB values obtained with different packet loss rates (0.02%, 0.05%, 0.25% and 0.5% respectively) using the Match Frame (MF) algorithm.

ADB_MF(i) values retain the blocking errors introduced by the wireless network yet reduces most of the spurious noise introduced by the wireless network at very low computational complexity. The results are reported over packet loss rates of 0.02%, 0.05%, 0.25% and 0.5%.

It can be observed that the *ADB_MF(i)* algorithm doesn't discard the reported blocking of frames caused by packet losses, rather it only removes the noise introduced in observation due to frame losses. For each different loss case we can observe periods where errors are mistakenly reported by the *ADB* algorithm, when in fact the ADB value would be zero had the case for missing frames been taken into account. It can be seen that in the face of packet loss, using this windowing algorithm spurious frame error reports (computed using ADB) can be minimized. In the 0.02% loss case there were 298 damaged frames, *ADB* reported 862 lost frames, while *ADB_MF(i)* reported that there were 298 damaged frames, matching the manually calculated version.

For higher loss rates, further investigation is required to fine-tune the optimal window size, however we have included results for higher loss rates to demonstrate the smoothing

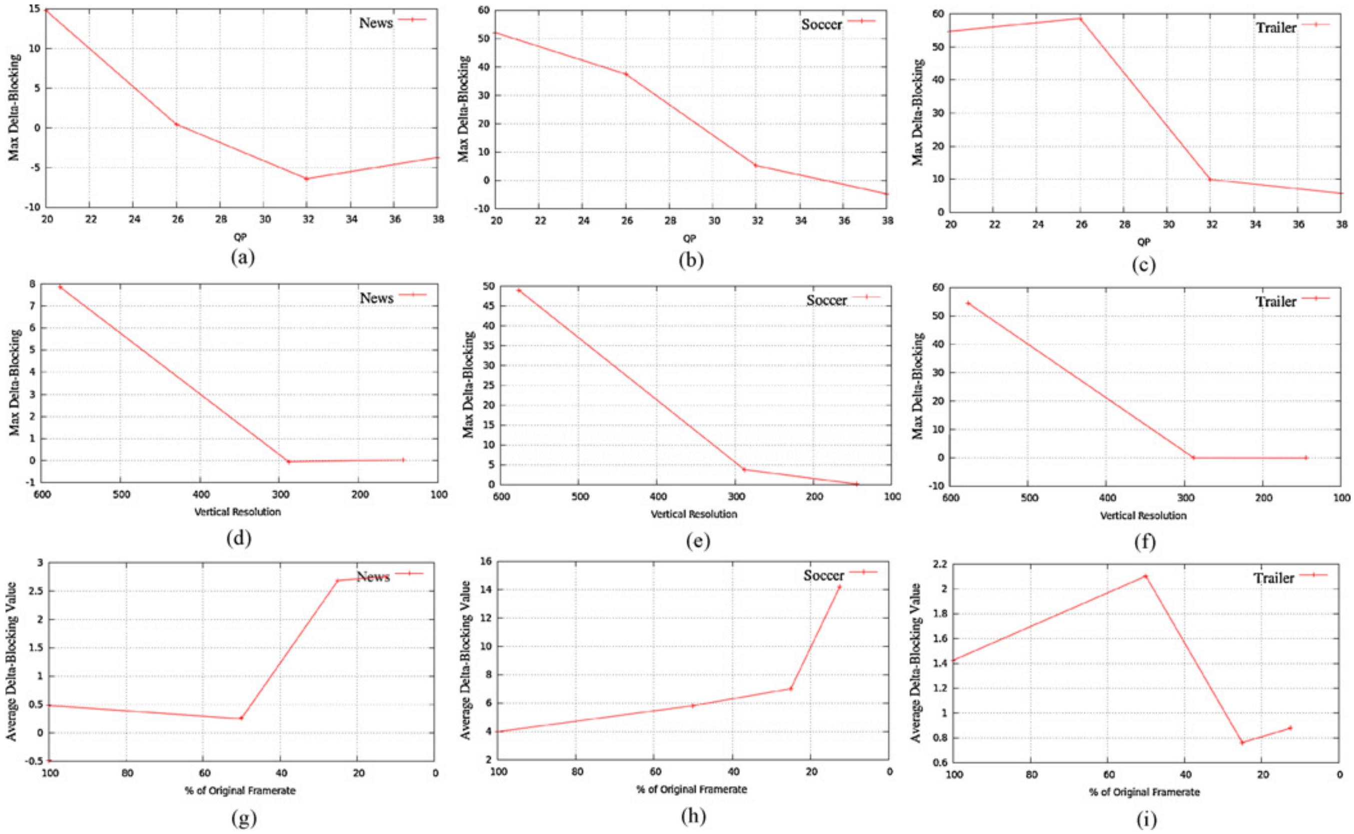

Fig. 16. Average and Max. Delta Blocking for each Video/Scalable Dimension Combination.

obtained when using a window size of 100. However, these loss rates are much higher than target loss rates for a deployed IPTV service and thus, detection of periodic events may not be of concern since the video content may be completely unwatchable at these loss rates. Without the use of *ADB_MF(i)* and simply using *ADB*, when a full frame is lost, quality monitoring may not continue in any meaningful manner. Hence, the applicability of *ADB* would be limited.

A. Characterizing the Effects of Loss using Reduced/No Reference Metrics

Figure 16 (a)-(i) presents the results of the impact of a fixed amount of loss on each video/scalable dimension combination. Due to the lack of a network based streaming server for H.264/SVC suitable for our purposes, these experiments were carried out using AVC encoded versions of the sequences. Previous works in the area have presented the use of an SVC streaming server [39], [40] however, these works were for simulation-based experiments only or were unable to be adopted for our work. In this work in order to enable a deployable solution, which takes actual decoder behavior in the face of loss into account, we require an experimental based approach. To expand on this point, we cannot guarantee that a simulation-based approach will capture the necessary behavior when faced with lost macroblock data, which is something that our solution requires.

While there may be small variations in the observed values between AVC and SVC, due to the fact that SVC is an extension of AVC and inherits its encoding techniques, we argue that the observations made here will hold true in the case of SVC. As stated above, there will be some variation in blocking values between the original sequence and the transmitted sequence (prior to macroblocking errors) due to the different encoders used. We can also observe that in some cases that the maximum blocking value in the received file is lower than that of the original video, again this can be partially attributed to the variations due to encoding, but the more likely case is that the original sequence frame with the highest blocking value was lost during transmission.

For the purpose of our experiment, as stated before we used the x264 software library to encode our content according to the same parameters used in the previous scalable video quality analysis. The outputted MPEG-4 file was the encapsulated into an MPEG Transport Stream file, in order to provide as close a representation to a deployed IPTV system. We then served the video using RTSP and RTP over the *localhost* interface with a specified loss rate of 0.5%. The server software used was the popular Live555MediaServer [41] and the client software used to request and capture the video was OpenRTSP [42]. The received video was then decoded to construct a raw YUV sequence which represented the “observed” video with errors in some frames due to loss macro-block data.

Across all cases, there were differing levels of damage to frame or, in some cases, total frame loss. This is where the entire frame (or multiple frames) could not be decoded and thus the frame length of the received sequence was lower than the original, requiring re-alignment of values.

As can be seen in Figures 16 (a)-(c), as we increase the levels of quantization of a particular sequence, the overall trend is that the difference between the frame with the highest blocking value in the original sequence and the highest blocking value (the frame with the highest levels of macroblocking) in the lossy sequence is decreasing.

We also observe that in some cases that the maximum delta blocking is less than zero. As detailed above, this can be explained by the loss of the frame with the highest blocking value during transmission. However, we can still observe that the overall trend indicates that the effect of blocking due to errors is lower as we decrease quality due to increased quantization.

A similar trend can be observed in Figures 16 (d)-(f), here we can see that, as we decrease the resolution of the sequence the difference between the frame with the highest levels of blocking due to errors and that of the original sequence, the observed impact of loss is lower. In our experiments, we scaled between 3 different resolutions; 1024x576, 512x288, 256x144, corresponding to "high", "medium" and "low" resolution respectively. We can observe that the impact of loss due to macroblocking is almost non-existent at the "medium" resolution compared to the "high" resolution and further degradation to the "low" resolution provides a much lower benefit in terms of robustness to loss. However, we may still degrade to the low "resolution" in order to increase bandwidth efficiency.

In the case of temporal scalability, we chose to use the difference between the average blocking value for the entire sequence. The reason for doing this is to demonstrate how the overall sequence is affected when a frame becomes damaged due to loss and using the difference between the averages is the most effective method. As stated above under temporal scalability, the frames that remain at the lower temporal layer must be repeated in order to match the same framerate found at the full temporal quality. However, as shown above this can cause a degradation of quality in terms of SSIM / PSNR.

Furthermore, as we have shown in Figures 16 (g)-(i) this also leads to damaged frames being repeated in the case of loss, this has the effect of increasing the overall blocking value and thus the perceived experience is lowered. We can observe in (g) and (h) that as we lower the framerate, the number of times a damaged frame is repeated increases and thus the average blocking value is raised. There is only a small degree of variance in the *news* sequence for the first temporal degradation step, however this is more pronounced as the framerate decreases further. In the case of (i) we can see that in 2 of the 3 degradation steps the average increases but in one case it is lowered. This could possibly be attributed to lower levels of losses in this case or perhaps damage to macroblocks which do not produce a large impact on the blocking value.

The effects of temporal scalability are likely to be minimized for low motion sequences. However, low motion sequences will not have a large number of motion vectors and therefore the actual bandwidth savings for temporal sequences with minimal quality degradation is not as great as other sequences with higher motion, which may suffer greater degradations in quality.

B. Use of the Solution in a Quality Monitoring System

The use of scalable video coding and the “delta-blocking” mechanism could be used in a number of target application scenarios. For example, an IPTV service provider could be delivering high quality video content to customers, satisfying their quality of experience (QoE) expectations of the service. However, should some factor limit the network’s ability to allocate sufficient bandwidth for the full-quality video, packets belonging to the video stream may be lost. This loss will manifest itself in terms of visible blocking errors at the customers’ playback devices.

If the IPTV service provider has already computed and transmitted the required $B_(S)$ information to the STBs (or perhaps a selection of customers in different regional areas), the delta blocking metric could be used to detect these visible blocking errors. This can then be provided as feedback to the service operator who may choose to switch to a lower scalable layer. The construction of the different SVC layers will depend on the particular type of content, allowing the operator to maintain the maximum quality for a given bandwidth. Upon resolution of the issue the delta-blocking metric can then be used to verify that the issue has been resolved.

VI. Conclusion and Future Work

In this paper we investigated the effects of scalability in all 3 dimensions for H.264 SVC using complex, broadcast content. We carried out this investigation using 2 full-reference metrics and 2 no/reduced–reference metrics in order to ascertain the observed video quality relative to a particular saving of bandwidth. The results give insight into degradation of video quality based on scalable dimension, which is important for setting the right parameters for encoding scalable video. We characterized the effect of different content types on performance with scalable video.

We also motivated the use of two no-reference metrics, namely Blocking and Blurring, to ascertain the effect on these as we progress through the degradation path for each scalable dimension. Furthermore, we investigated the effect on video quality for each scalable dimension in the presence of loss. Our findings indicate that as quantization increases or spatial resolution decreases the overall impact on video quality through loss is decreased. However, in the case of temporal degradation, due to its nature, the impact of loss leads to a greater impact on quality.

Finally, we presented the use of a no-reference metric in a reduced reference fashion, which can be used to detect visual damage and frame loss due to packet loss. This data can then be used to trigger a modification of the video stream using H.264 SVC. We also presented an algorithm which can be used to re-align the reduced reference data when frame loss does occur.

This works also highlights the use of no-reference metrics to assess the impact of loss on video quality, where full-reference metrics (due to their dependency on a frame-by-frame matching) cannot be employed.

It is hoped that this work could be used as a part of a framework to select the appropriate dimension(s) to scale (and by how much) when constructing the layers for a SVC sequence to maximize video quality, while saving the largest amount of bandwidth.

ACKNOWLEDGMENT

This work has received support from Science Foundation Ireland via the "Federated, Autonomic Management of End-to-End Communications Services" (grant no. 08/SRC/11403). This work is also partially supported by the National Science Foundation under Grant #1019343 to the Computing Research Association for the CIFellows Project.

REFERENCES

[1] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H. 264/AVC: Tools, performance, and complexity," *IEEE Circuits Syst. Mag.*, vol. 4, no. 1, pp. 7–28, Jan.–Mar. 2004.

[2] N. Cranley, L. Murphy, and P. Perry, "User-perceived qauality-aware adaptive delivery of MPEG-4 content," in *Proc. 13th Int. Workshop Netw. Operating Syst. Support Digital Audio Video*, 2003, pp. 42–49.

[3] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hoßfeld, "Impact of frame rate and resolution on objective QoE metrics," in *Proc. 2nd IEEE Int. Workshop Quality Multimedia Experience*, 2010, pp. 29–34.

[4] H. Sohn, H. Yoo, W. De Neve, C. Kim, and Y. Ro, "Full-reference video quality metric for fully scalable and mobile SVC content," *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 269–280, Sep. 2010.

[5] A. Khan, L. Sun, and E. Ifeachor, "Impact of video content on video quality for video over wireless networks," in *Proc. 5th IEEE Int. Conf. Autonomic Autonomous Syst.*, 2009, pp. 277–282.

[6] F. De Vito and J. De Martin, "Psnr control for gop-level constant quality in H.264 video coding," in *Proc. 5th IEEE Int. Symp. Signal Process. Inform. Technol.*, 2005, pp. 612–617.

[7] P. McDonagh, C. Vallati, A. Pande, P. Mohapatra, P. Perry, and E. Mingozzi, "Investigation of scalable video delivery using H.264 SVC on an LTE network," in *Proc. 14th Int. Symp. WPMC*, 2011, pp. 1–5.

[8] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[9] S. Hemami and A. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Process. Image Commun.*, vol. 25, no. 7, pp. 469–481, 2010.

[10] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[11] Y. Zhou, Y. Sun, X. Yin, and S. Sun, "Effective frame level rate control for H.264/AVC video coding," in *Proc. IEEE GLOBECOM*, 2008, pp. 1–5.

[12] G. Muntean, P. Perry, and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks," *IEEE Trans. Broadcast.*, vol. 50, no. 1, pp. 1–10, Mar. 2004.

[13] G.-M. Muntean, P. Perry, and L. Murphy, "Objective and subjective evaluation of QOAS video streaming over broadband networks," *IEEE Trasn. Network Service Manage.*, vol. 2, no. 1, pp. 19–28, Nov. 2005.

[14] A. Ksentini, M. Naimi, and A. Guéroui, "Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture," *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 107–114, Jan. 2006.

[15] G. Van der Auwera, P. David, and M. Reisslein, "Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG-4 advanced video coding standard and scalable video coding extension," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 698–718, Sep. 2008.

[16] J. Monteiro, C. Calafate, and M. Nunes, "Evaluation of the H.264 scalable video coding in error prone IP networks," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 652–659, Sep. 2008.

[17] X. Zhu, T. Schierl, T. Wiegand, and B. Girod, "Video multicast over wireless mesh networks with scalable video coding (SVC)," in *Proc. Visual Commun. Image Process.*, 2008, pp. 682 205–682 205.

[18] Y. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. Alnuweiri, "A link adaptation scheme for efficient transmission of H.264 scalable video over multirate WLANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 875–887, Jul. 2008.

[19] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004.

[20] J. Klaue, B. Rathke, and A. Wolisz, "Evalvid: A framework for video transmission and quality evaluation," in *Proc. Comput. Performance Eval. Modelling Tech. Tools*, 2003, pp. 255–272.

[21] T. Le, H. Nguyen, and H. Zhang, "EvalSVC: An evaluation platform for scalable video coding transmission," in *Proc. 14th IEEE ISCE*, 2010, pp. 1–6.

[22] J. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurutepe, T. Sikora, J. Ostermann, E. Izquierdo, and T. Ebrahimi, "Subjective evaluation of scalable video coding for content distribution," in *Proc. Int. Conf. Multimedia*, 2010, pp. 65–72.

[23] P. Seeling, F. Fitzek, G. Ertli, A. Pulipaka, and M. Reisslein, "Video network traffic and quality comparison of VP8 and H.264 SVC," in *Proc. 3rd Workshop Mobile Video Delivery*, 2010, pp. 33–38.

[24] K. Singh, A. Ksentini, and B. Marienval, "Quality of experience measurement tool for SVC video coding," in *Proc. IEEE ICC*, Jun. 2011, pp. 1–5.

[25] X. Li, P. Amon, A. Hutter, and A. Kaup, "Performance analysis of inter-layer prediction in scalable video coding extension of H.264/AVC," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 66–74, Mar. 2011.

[26] MSU. (2011). *MSU Video Quality Measurement Tool* [Online]. Available: http://compression.ru/index_en.htm

[27] J. Reichel, H. Schwarz, and M. Wien, *Joint Scalable Video Model 11 (JSVM 11)*, Joint Video Team, document JVT-X202, 2007.

[28] D. Vatolin, S. Grishin, A. Moskvin, and A. Parshin, "Objective measurements of artifacts produced by modern video coding standards," in *Proc. Spring Conf. Comput. Graph.*, 2006.

[29] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real-time system for adaptive video streaming based on SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1227–1237, Sep. 2007.

[30] H. Huang, W. Peng, T. Chiang, and H. Hang, "Advances in the scalable amendment of H.264/AVC," *IEEE Commun. Mag.*, vol. 45, no. 1, pp. 68–76, Jan. 2007.

[31] J. Maisonneuve, M. Deschanel, J. Heiles, W. Li, H. Liu, R. Sharpe, and Y. Wu, "An overview of IPTV standards development," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 315–328, Jun. 2009.

[32] Y. Xiao, X. Du, J. Zhang, F. Hu, and S. Guizani, "Internet protocol television (IPTV): The Killer application for the next-generation Internet," *IEEE Commun. Mag.*, vol. 45, no. 11, pp. 126–134, Nov. 2007.

[33] A. Leontaris and A. Reibman, "Comparison of blocking and blurring metrics for video compression," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 2. 2005, pp. 585–588.

[34] Z. Wang, A. Bovik, and B. Evan, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. 2000, pp. 981–984.

[35] X. Huang and K.-M. Chao, "A generalized global alignment algorithm," *Bioinformatics*, vol. 19, no. 2, pp. 228–233, 2003.

[36] A. Agrawal, V. P. Brendel, and X. Huang, "Pairwise statistical significance and empirical determination of effective gap opening penalties for protein local sequence alignment," *Int. J. Comput. Biol. Drug Design*, vol. 1, no. 4, pp. 347–367, 2008.

[37] A. Agrawal and X. Huang, "Pairwise statistical significance of local sequence alignment using multiple parameter sets and empirical justification of parameter set change penalty," *BMC Bioinformatics*, vol. 10, no. 3, p. S1, 2009.

[38] A. Agrawal and X. Huang, "Pairwise statistical significance of local sequence alignment using sequence-specific and position-specific substitution matrices," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 8, no. 1, pp. 194–205, Jan.–Feb. 2011.

[39] M. Ghareeb, A. Ksentini, and C. Viho, "Scalable video coding (SVC) for multipath video streaming over video distribution networks (VDN)," in *Proc. IEEE Int. Conf. Inform. Networking*, 2011, pp. 206–211.

[40] L. Zhang, C. Yuan, and Y. Zhong, "Reliable and efficient adaptive streaming mechanism for multi-user SVC VoD system over GPRS/EDGE network," in *Proc. IEEE Int. Conf. Comput. Sci. Softw. Eng.*, vol. 3. 2008, pp. 232–235.
[41] Live555. (2011). *Live555mediaserver* [Online]. Available: http://www.live555.com/mediaServer
[42] openRTSP. (2011). *openrtsp* [Online]. Available: http://www.live555.com/openRTSP

Patrick McDonagh received the B.Sc. degree in computer science from the University College Dublin, Dublin, Ireland, in 2009. He joined the Performance Engineering Laboratory (http://pel.ucd.ie) as a Ph.D. student in 2009. He was a Visiting Researcher at the University of California, Davis, CA, USA, in 2011. He also conducts research into video quality analysis with a focus on objective and no-reference methods. His current research interests include methods of monitoring and managing carrier-grade services in wired and wireless access networks.

Amit Pande (M'10) received the Ph.D. degree from Iowa State University, Ames, IA, USA, in 2010. He is currently a Project Scientist at the Department of Computer Science, University of California, Davis (UC Davis), CA, USA. He was awarded with Excellence in Post-Doctoral Research in 2012 by UC Davis, NSF CI Fellowship 2010–2012 by CRA, Zaffarano Award, hon. mention, in 2010 by ISU, Research Excellence Award in 2010 by ISU, Best Paper Award by WPMC in 2012, Design Contest Winner (VLSID 2012, 2010), Institute Silver Medal in 2007 by IIT Roorkee, Agilent Engineering and Technology Awards in 2007, India, and other distinctions. His current research interests include multimedia systems, wireless networks, embedded systems, security, and trust.

Liam Murphy (M'85) received the B.E. degree in electrical engineering from the University College Dublin, Dublin, Ireland, in 1985, and the M.Sc. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1988 and 1992, respectively. He is currently a Professor of computer science and informatics at the University College Dublin, where he is the Director of the Performance Engineering Laboratory. He has published almost 150 refereed journal and conference papers on various topics, including multimedia transmissions, dynamic and adaptive resource allocation algorithms in computer/communication networks, and software development. His current research interests include computer network convergence and software performance engineering. Prof. Murphy is a member of the IEEE (Communications, Broadcasting, and Computer societies) and a fellow of the Irish Computer Society.

Prasant Mohapatra (F'10) received the Doctoral degree from Pennsylvania State University, University Park, PA, USA, in 1993. He is currently the Tim Bucher Family Endowed Chair Professor and the Chairman of the Department of Computer Science, University of California, Davis, CA, USA. Previously, he was on the faculty at Iowa State University, Ames, IA, USA, and Michigan State University, East Lansing, MI, USA. He has also been a Visiting Scientist at Intel Corporation, Panasonic Technologies, Institute of Infocomm Research (I2R), Singapore, and National ICT Australia (NICTA), Sydney, Australia. He has been a Visiting Professor at the University of Padova, Padova, Italy, and Yonsei University, Seoul, Korea. His current research interests include wireless networks, sensor networks, Internet protocols, and QoS. Dr. Mohapatra was a recipient of the Outstanding Engineering Alumni Award in 2008. He was a recipient of the Outstanding Research Faculty Award from the College of Engineering, University of California, Davis, CA, USA.