Cross-dimensional Perceptual Quality Assessment for Low Bitrate Videos

Guangtao Zhai, Jianfei Cai, *Senior Member, IEEE,* Weisi Lin, *Senior Member, IEEE,*
Xiaokang Yang, *Senior Member, IEEE,* Wenjun Zhang, *Senior Member, IEEE,* Minoru Etoh, *Senior Member, IEEE*

***Abstract*—Most studies in the literature for video quality assessment have been focused on the evaluation of quantized video sequences at fixed and high spatial and temporal resolutions. Only limited work has been reported for assessing video quality under different spatial and temporal resolutions. In this paper, we consider a wider scope of video quality assessment in the sense of considering multiple dimensions. In particular, we address the problem of evaluating perceptual visual quality of low bitrate videos under different settings and requirements. Extensive subjective view tests for assessing the perceptual quality of low bitrate videos have been conducted, which cover 150 test scenarios and include five distinctive dimensions: encoder type, video content, bitrate, frame size and frame rate. Based on the obtained subjective testing results, we perform thorough statistical analysis to study the influence of different dimensions on the perceptual quality and some interesting observations are pointed out. We believe such a study brings new knowledge into the topic of cross-dimensional video quality assessment and it has immediate applications in perceptual video adaptation for scalable video over mobile networks.**

***Index Terms*—Video quality assessment, perceptual visual quality, subjective view test, video adaptation.**

I. Introduction

VIDEO quality assessment (VQA) has been an active research area for the last two decades. The most systematic attempts have been made by the Video Quality Expert Group (VQEG) [1], which was formed in 1997 aiming at achieving an international standardization/recommendation of objective video quality metric (VQM). Until now, only the Full Reference Television (FR-TV) project of VQEG has been completed. The FR-TV Phase I [2] drew the conclusions that subjective VQA cannot be replaced by the objective ones and no other full reference (FR) VQM in the FR-TV Phase I can statistically outperform the metric of PSNR. During the FR-TV Phase I, a test data set consisting of 16 hypothetical reference circuits (HRCs) for each of the 20 original sequences was released together with their mean opinion scores (MOS), which thereon greatly facilitates the research on VQA. As a result, the best VQM candidates in the FR-TV Phase II [3] have substantially outperformed PSNR.

Most methods in the literature for VQA, including the ones proposed by VQEG, have been focused on the evaluation of quantized video sequences but at fixed and high spatial and temporal resolutions such as 720×576 @ 50 fps and 720×486 @ 60 fps used in the VQEG video database for PAL and NTSC TV formats, respectively. It is well known that the quantization brings out intra-frame distortions such as blockiness, ringing, blurring and so on [4], [5]. These intra-frame artifacts have been thoroughly studied over the years, and various metrics have been proposed to evaluate the distortions and predict the perceptual quality.

For video transmission over resource-constrained networks such as wireless networks, it is hard to maintain high spatial and temporal resolutions. Usually, in addition to heavy quantization, temporal resolution reduction such as frame dropping and spatial down-sampling are often used to reduce the data size, which leads to inevitable quality degradation. In particular, frame dropping causes jitter/jerkiness and spatial down-sampling brings blurring (when the video is up-sampled and played back at the original spatial resolution). The inter-frame artifacts, though less investigated in the literature than the intra-frame artifacts, have also been considered and modelled [6], [7].

However, when the aforementioned intra and inter frame artifacts are combined, only limited work has been reported for assessing video quality under different spatial and temporal resolutions. Such an issue of cross-dimensional video quality assessment is frequently encountered in the field of perceptual video adaptation [8]–[10], where a major problem is to determine the best combination of frame rate, frame size and SNR levels that maximizes the visual quality given a certain bitrate budget. Typically, the cross-dimensional VQA is based on empirical spatiotemporal models constructed through off-line subjective-viewing tests. For instance, for MPEG-4 FGS video coding, Rajendran *et al.* [8] suggested a frame rate selection model, which prefers high temporal resolution under high SNR levels. Wang *et al.* [9] conducted subjective viewing tests on CIF videos coded at bitrates ranging from 50 kbps to 1 Mbps with different frame rates. They found that as available

Manuscript received Nov. 13, 2007; revised on Feb. 22, 2008 and Apr. 23, 2008. This research was partially supported by Singapore A*STAR SERC Grant (062 130 0059). This paper has been presented in part in IEEE International Symposium on Circuits and Systems (ISCAS) 2008. This paper was recommended by Guest Editor Dr. Ling Guan.

G. Zhai is with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, 200240, China. This work was done during his visit at the School of Computer Engineering, Nanyang Technological University, 639798, Singapore. e-mail: zhaiguangtao@sjtu.edu.cn

J. Cai and W. Lin are with the School of Computer Engineering Nanyang Technological University, 639798, Singapore. e-mail: {asjfcai,wslin}@ntu.edu.sg

X. Yang and W. Zhang are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, 200240, China. e-mail: {xkyang,zhangwenjun}@sjtu.edu.cn

M. Etoh is with the NTT DoCoMo Research Laboratories, Yokosuka, Japan. e-mail: etoh@ieee.org

Copyright (c) 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

bandwidth drops, for the minimal visual disturbance, 440 kbps and 175 kbps are the optimal points for the frame rate to be halved. However, this algorithm only considers the quality degradation caused by frame dropping. More recently, Cranley and Murphy's [10] subjective test proved that given certain bandwidth constraints, there exists an optimal combination of spatial and temporal resolutions that maximizes the visual quality.

In addition, the concept of multidimensional modelling of image quality has been proposed by Martens [11], who investigated the synthesized influence of noise and blurring on the perceptual quality of images. Shnayderman *et al.* [12] also constructed a multidimensional model based on singular value decomposition (SVD) to measure images with six types of distortions at various levels. Though using the term of 'multidimensional', these algorithms are actually dealing with only the intra-frame artifacts.

In this paper, we consider a much wider scope of "multidimensional' video quality assessment. In particular, we address the problem of evaluating perceptual visual quality of low bitrate videos under different settings and requirements. Extensive subjective view tests for assessing the perceptual quality of low bitrate videos have been conducted, which cover 150 test scenarios and include five distinctive dimensions: encoder type, video content, bitrate, frame size and frame rate. Based on the obtained subjective testing results, we perform thorough statistical analysis to study the influence of different dimensions on MOS and some interesting observations are pointed out. We believe such a study brings new knowledge into the topic of cross-dimensional VQA and it has immediate applications in perceptual video adaptation for scalable video over mobile networks.

The rest of this paper is organized as follows. The video quality assessment problem is formulated in section II. The details of the subjective viewing test are described in section III. We analyze the subjective test results in section IV. Finally, a conclusion is given in section V.

II. Problem Statement

For cross-dimensional VQA, in order to uniquely characterize a compressed video stream, we construct a video feature space, defined as a vector space $\mathbf{F}^n$ with n distinctive dimensions. In this way, any video bitstream is represented as a point in the vector space:

$$f = (f_1, f_2, \ldots, f_n) \in F^n. \tag{1}$$

Similarly, a quality space characterizing the perceptual quality of video sequences can be constructed as a vector space $\mathbf{Q}^m$, with

$$q = (q_1, q_2, \ldots, q_m) \in Q^m. \tag{2}$$

The problem of VQA can then be formulated as a mapping from $\mathbf{F}^n$ to $\mathbf{Q}^m$, denoted as

$$F^n \xrightarrow{vqa} Q^m. \tag{3}$$

When the mapping function vqa is linear, hypothetically of course, it can be expressed as a matrix multiplication $q = A\mathbf{f}$, with A being the $m \times n$ quality assessment matrix. This abstraction allows both the feature and quality spaces to be multidimensional.

TABLE I
Different combinations of bitrate, frame size (FS) and frame rate (FR)

FS \ FR	7.5 fps	15 fps	30 fps
CIF	64, 128 Kbps	64, 128 Kbps	128, 384 Kbps
QCIF	24, 48, 64 Kbps	24, 48, 64 Kbps	48, 64, 128 Kbps

In this research, we characterize a video bitstream by 5 distinctive dimensions: encoder type, video content, bitrate, frame rate and frame size, denoted as

$$F^5 = \{ET, VC, BR, FR, FS\}. \tag{4}$$

It should be noted that the encoder and content dimensions are highly conceptual. They can be further divided into sub-dimensions describing the specifications of the encoder (e.g. motion estimation/compensation, transformation and quantization) and the nature of the sequence (e.g. motion, color and texture). In this research, for simplicity we use the spatial and temporal activities of a video to represent the video content dimension. As for the quality space, commonly a single MOS is used to indicate the overall quality of a video sequence according to the ITU standard BT500-11 [13]. It could be extended to multiple dimensions. For example, Ghinea and Thomas [14] defined that the Quality of Perception (QoP) includes two components: the satisfaction (QoP_S) and understanding (QoP_U) aspects of viewers on the video. This can be regarded as a two dimensional modelling of video quality space, i.e. $Q^2 = \{QoP_S, QoP_U\}$. In this research, we only use a single MOS value to describe perceptual quality.

III. Subjective Viewing Test

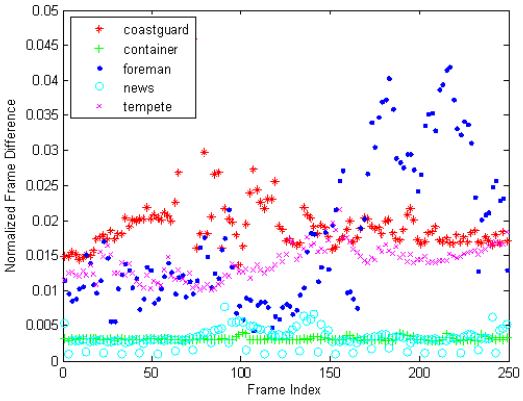
A. Testing Materials

Five 250-frame test sequences, namely 'Container', 'Coastguard', 'Foreman', 'News' and 'Tempete', are employed in the experiments. The snapshots of the sequences are shown in Fig. 1. To demonstrate the spatial and temporal activities of the five sequences, i.e. the video content dimension, their normalized absolute inter-frame difference and intra-frame variance are shown in Fig. 2(a) and 2(b), respectively. It can be seen that 'Container' has the least overall spatiotemporal activities, and 'News' has higher intra-frame activities. The sequence 'Tempete' has moderate spatiotemporal activity, and Coastguard' and 'Foreman' have the largest overall motion and intra-frame variance. All the sequences are compressed using the H.263 and H.264 encoders at bitrates ranging from 24k to 382k bps with frame sizes of QCIF and CIF and frame rates of 7.5 to 30 fps. For one video sequence coded by one encoder, the different combinations of bitrate, frame size and frame rate form totally 15 different test scenarios, as shown in Table I.

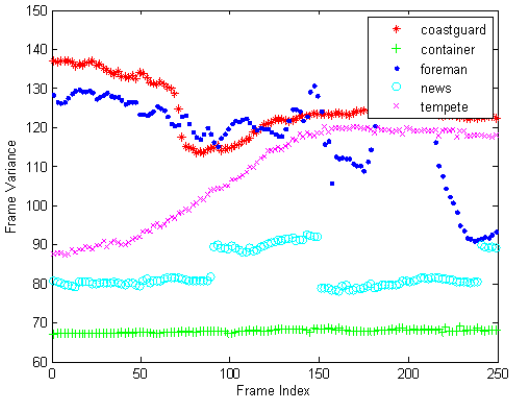
B. Testing Method

In the subjective test, different reconstructions of a video sequence are displayed at the same relatively high spatial and

Fig. 1. Snapshots of the test video sequences. (From left to right: 'Coastguard', Container', 'Foreman', 'News' and 'Tempete'.)



(a) Inter-frame difference



(b) Intra-frame variance

Fig. 2. The normalized absolute inter-frame difference and intra-frame variance for different sequences.

temporal resolutions, i.e. CIF and 30 fps. Lower resolution sequences are converted into the highest resolution through up-sampling (using the H.264/AVC 6-tap half-sample interpolation filter [15]) and frame repeat. Furthermore, we use the method of the double stimulus impairment scale variant II (DSIS II) for the subjective test experiments, which is suggested by ITU-R Recommendation BT 500-11 [13]. In particular, in the DSIS II method a reference sequence is first displayed followed by a distorted one, and then this process is repeated once more before the viewers are asked to score the perceptual quality of the distorted sequence within a four-second time interval. We use the five-level quality scale to describe the video quality in English, where the scores of $1, 2, 3, 4, 5$ represent 'bad', 'poor', 'fair', 'good' or 'excellent' quality, respectively. All the sequences were viewed by 20 participants (young university students with 10 males and 10 females), who have sufficient knowledge of English to make reasonable votes on the quality scale sheet.

TABLE II
THE MARKERS USED IN FIG. 3

VC \ BR	24 kbps	48 kbps	64 kbps	128 kbps	384 kbps
Coastguard	.	+	×	*	○
Container	.	+	×	*	○
Foreman	.	+	×	*	○
News	.	+	×	*	○
Tempete	.	+	×	*	○

C. Testing Environment

The laboratory has been set-up according to the ITU-R Recommendation 500-11 [13]. The monitors used are professional SONY 21' BVM 21F with 6000^{o} color temperature. The walls behind the monitors are covered with photographic papers to prevent distracting the viewers during the test. The lighting is provided by fluorescent lamps operating at 100 Hz with 6000^{o} color temperature so as to cast minimum inference on the monitors. The viewing distance is set to 3 to 4 times of the display screen height.

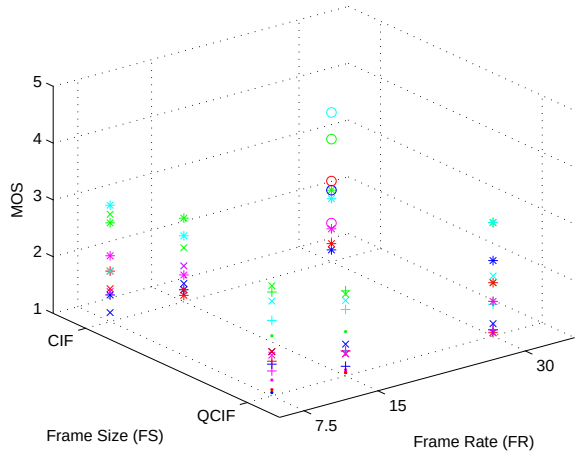
D. Testing results

Considering the five test sequences and the two types of encoders plus the 15 combinations listed in Table I, totally we have 150 test scenarios. Since the H.263 encoder does not use some advanced coding techniques such as quarter-pixel-accurate motion estimation / compensation, CAVLC / CABAC (context-adaptive variable length coding / binary arithmetic coding) [16], it is easy to imagine that its performance of mean opinion scores (MOS) is substantially lower than that of H.264. Thus, to better visualize the large number of MOS results, we show the performance of H.263 and H.264 separately. Fig. 3 shows the MOS results, where each MOS data point is drawn in a 3D space defined by frame size, frame rate and MOS and different sequences and bitrates are highlighted by various markers defined in Table II.

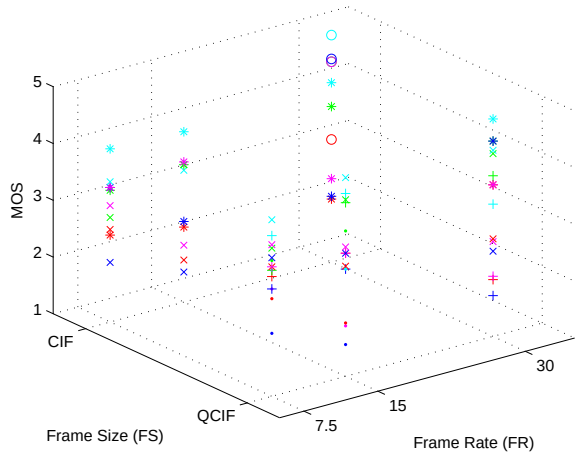
IV. Analysis of Subjective Testing Results

A. Influence of Different Dimensions

By roughly viewing the results shown in Fig. 3, we have the following simple observations. First, as expected, H.264 outperforms H.263. Second, in general, higher MOS are associated with higher spatial and temporal resolutions.



(a) using H.263



(b) using H.264

Fig. 3. The mean opinion scores (MOS) results.

This does not imply that sequences with low spatial and/or temporal resolutions cannot have good perceptual quality. In addition, between-sequence disparities can also be noticed in the results. For example, 'Foreman' has generally lower scores than 'News'.

In order to thoroughly study the influence of different dimensions on MOS, we perform ANOVA (analysis on variance) [17] on the MOS data set. In particular, we first perform a one-way ANOVA using the encoder types (ET) as the index. Table III shows the results, where the first column is the Sum of Squares between different treatments of ET. The second column is the Degrees of Freedom associated with the model of ET, which is defined as the number of treatments minus 1. The third column is the Mean Squares for the treatments, i.e. the ratio of Sum of Squares to degrees of freedom. The fourth column shows the F statistic, and the fifth column gives the p-value, which is derived from the cumulative distribution function (cdf) of F (refer to [17] for detailed explanation of ANOVA). As shown in Table III, the p-value is almost zero, which implies that the MOS results are severely affected by the encoder type. According to our experiment, we find the qualitative conclusions drawn for H.263 is almost the same as those for H.264. Therefore, without loss of generality, we will only consider the newer encoder type, i.e. H.264, for the rest parts of the paper.

Note that the bitrate shown in Fig. 3 is defined on the sequence level. To take into account different frame rates and frame sizes, for the following analysis we average the bitrate down to the pixel level, i.e. pixel bitrate (PB) defined as

$$PB = \frac{BR}{FR \cdot FS}. \tag{5}$$

If FR and FS are fixed, augmenting PB corresponds to increasing pixel coding quality, which is equivalent to improve the SNR level in SVC.

We then perform a four-way ANOVA on the MOS results with VC, FR, FS, PB as the variables. The analysis results are listed in Table IV. The small p-values ($p \leq 0.01$) indicate that the MOS is substantially affected by all the four dimensions. Furthermore, based on the magnitudes of the p-values, we can make a further claim that in general VC impacts the MOS results the most, followed by PB and then FR, while FS has the least influence. Our studies numerically substantiate the following observations reported in the literature of video quality assessment and video adaptation:

1) An accurate video quality assessment algorithm must be content-related. It cannot predict the visual quality distinctly by only using the bitrate and the spatial and temporal resolutions of the video stream;
2) The optimal combination of spatial and temporal resolutions that gives the best perceptual quality under a bitrate constraint varies from sequence to sequence. Hence, an effective video adaptation algorithm must take video content into account;
3) Given some extra bitrate budget, between the augmentations of FR and FS, increasing FR generally brings more perceptual quality improvement than increasing FS.

Because the MOS is found to be mostly inconsistent across the dimensions of VC and PB, we further categorize the sequences and pixel bitrates using the multiple comparison test, which is based on Tukey's honestly significant difference (HSD) criterion [18]. The results of the comparison test for VC and PB are shown in Fig. 4(a) and 4(b), where the center and the span of each horizontal bar indicate the mean and the 95% confidence interval, respectively. These results can be regarded as the projections of the data points residing within the feature space $\mathbf{F}^n$ onto the axes of VC and PB. From Fig. 4(a), we can group the test videos into two sets:

$$V_1 = \{Coastguard, Foreman, Tempete\}$$
$$V_2 = \{Container, News\}.$$

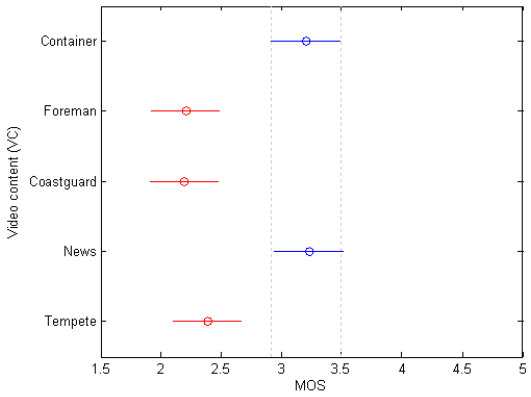
This classification matches the dissimilar levels of spatiotemporal activities of the five test video sequences. In particular, comparing the classification with the spatiotemporal activities indicated in Fig. 2, we can see that group V_1 has much higher frame difference and variance than group V_2. Thus, group V_1 requires more bits to encode their video content to reach the same quality. In other words, under the same bitrate

TABLE III
ONE-WAY ANOVA ON MOS WITH ENCODER TYPE (ET)

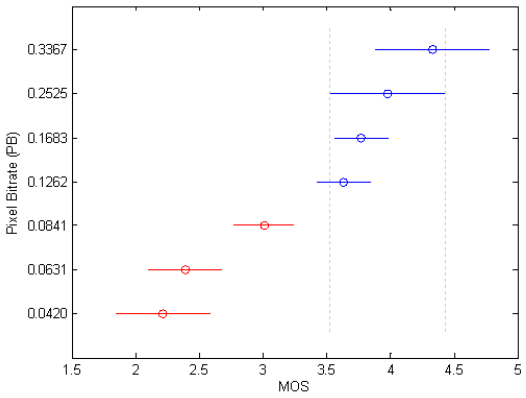
Sum of Squares	Degrees of Freedom	Mean Squares	F statistic	p-value
73.7943	1	73.7943	133.4745	$< 10^{-32}$

TABLE IV
FOUR-WAY ANOVA ON MOS WITH VIDEO CONTENT (VC), FRAME RATE (FR), FRAME SIZE (FS) AND PIXEL BITRATE (PB)

Dimensions	Sum of Squares	Degrees of Freedom	Mean Squares	F statistic	p-value
VC	20.3975	4	5.09937	35.08	3.4417e-015
FR	8.4808	2	4.24038	29.17	1.2894e-009
FS	1.1681	1	1.16806	8.04	0.0062
PB	20.8286	6	3.47143	23.88	2.5313e-014



(a) Video Content vs. MOS



(b) Pixel Bitrate vs. MOS

Fig. 4. The multiple comparison test for different video sequences and pixel bitrates.

constraint, the performance of group V_1 is inferior to that of group V_2. This well explains why group V_2 outperforms group V_1 in terms of MOS. This further justifies the claim that the perceptual quality of a video is highly related to its content.

Similarly, according to the multiple comparison results shown in Fig. 4(b), the pixel bitrates can also be categorized into two groups

$$B_1 = \{0.0420, 0.0631, 0.0841\}$$
$$B_2 = \{0.1262, 0.1682, 0.2525, 0.3367\}.$$

Group B_1 has MOS equal to or lower than 3, corresponding to the 'bad', 'poor' and 'fair' ranks of the five-level quality scale, whereas group B_2 has MOS larger than 3, corresponding to the 'good' and 'excellent' ranks. This classification implies that for the five test sequences, despite of their frame rate and frame size, the given pixel bitrate should be at least around 0.1 bpp in order to achieve a 'good' or 'excellent' perceptual quality. This finding also suggests that when other information is not available, the PB alone can serve as a rough quantitative gauge for video quality.

B. Optimal Combinations of Spatial and Temporal Resolutions

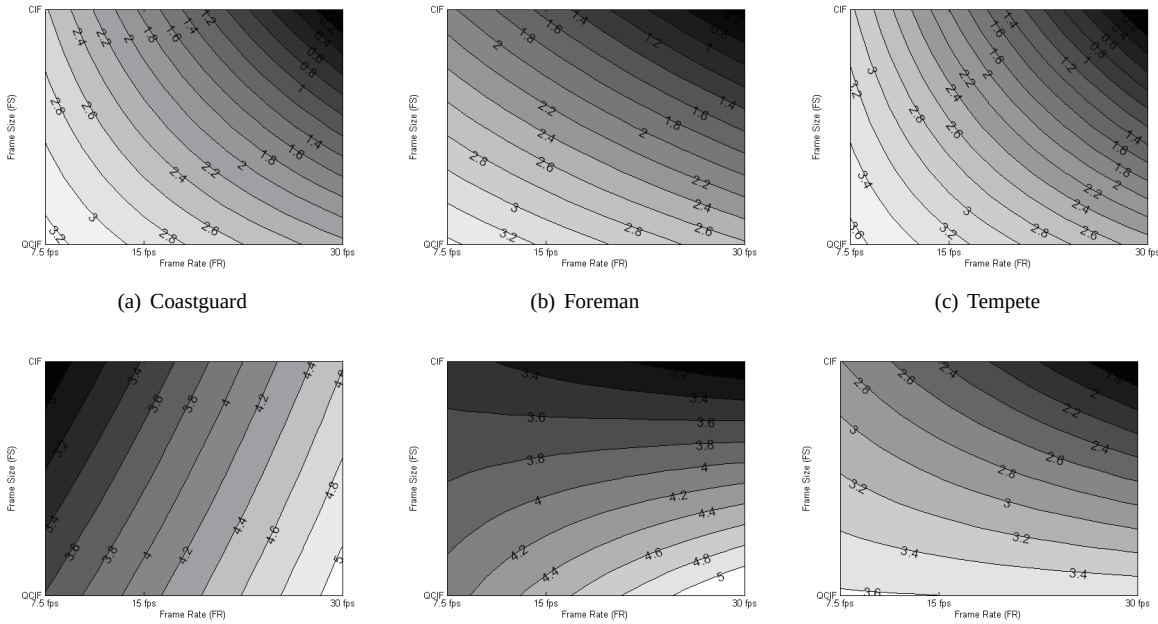
A direct application of the multidimensional VQA is to determine the best combination of multiple SVC scalabilities for perceptual video adaptation. In particular, given the video encoder ET, a particular video VC and the channel bandwidth BR, the problem of perceptual video adaptation is to select the optimal combinations of frame rate and frame size so as to maximize the perceptual quality $\mathbf{Q}$, i.e.

$$\begin{aligned}\{FR^*, FS^*\} &= \arg\max_{FR,FS} Q \\ &= \arg\max_{FR,FS}\{\mathbf{vqa}(F^5)|ET, VC, BR\}.\end{aligned} \quad (6)$$

Figs. 5(a)∼5(c) show the MOS vs. FR and FS plots for the five test sequences with $ET = \{H.264\}$, $BR = \{64\ kbps\}$, where the 'white' belt represents the best MOS. Note that the intermediate results are generated by using the spline based 2D interpolation [19], which has also been employed to generate Fig. 6 and Fig. 7.

For the results of group V_1 shown in Fig. 5(a), 5(b) and 5(c), it can be observed that in general MOS drops as FR and/or FS increases, and the best MOS results occur a the region of $FS = \{QCIF\}$ and $FR = \{7.5\ fps\}$. This is because higher spatiotemporal resolution at a certain low bitrate such as 64 kbps leads to a lower PB value, which is insufficient to describe the type of video sequences with large spatiotemporal activities and thus causes severe intra-frame degradation. On the other hand, with lower spatiotemporal resolution, more bits can be saved to achieve higher intra-frame quality and thereby effectively enhance the overall visual quality. This is in line with some recent studies in [14], [20].

For the results of group V_2 shown in Fig. 5(d) and 5(e), it is interesting to see that MOS increases with the increase



(a) Coastguard (b) Foreman (c) Tempete

(d) Container (e) News (f) Average behavior

Fig. 5. Equal-MOS contours for the five test sequences coded by H.264 at 64 kbps.

(decrease) of FR (FS), and the best MOS results is achieved at $FS = \{QCIF\}$ and $FR = \{30\ fps\}$. This is because the sequences in group V_2 have relatively less textural details and very low motion, for which high FR can be easily achieved without much cost on bitrate. Fig. 5(f) shows the average results over the five video sequences. Combining the analysis for groups V_1 and V_2, we can conclude that for perceptual adaptation of nature videos, under limited bandwidth, in general FS should be kept low while FR should be low (high) for the sequences with high (low) temporal activity.

C. Fixed Spatial or Temporal Resolution

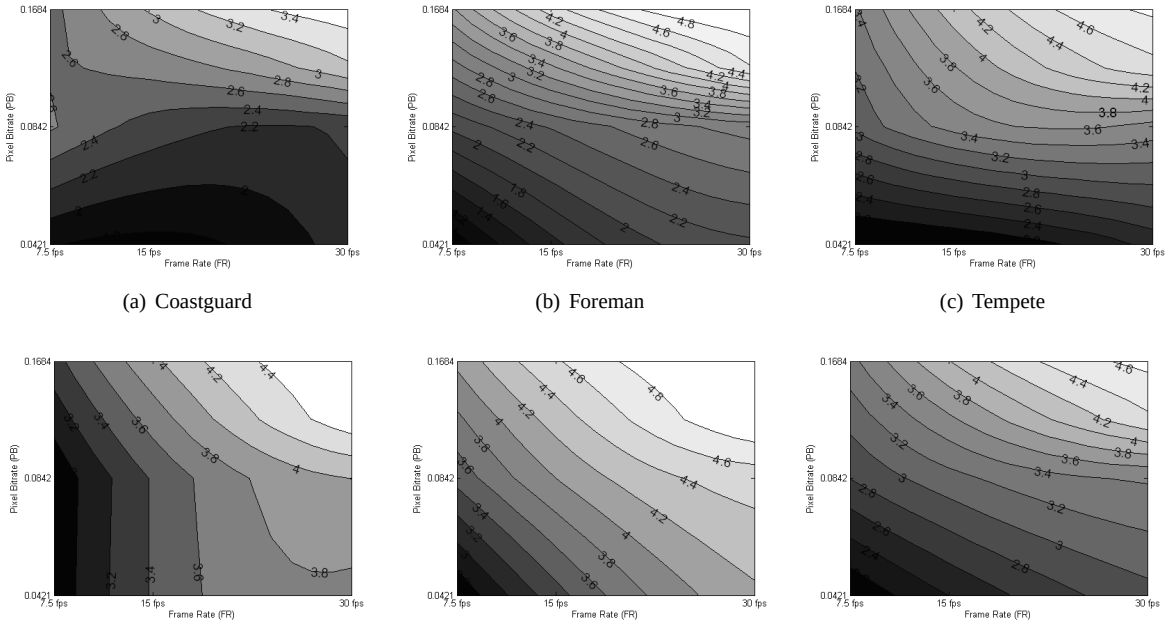
For some specific applications, the spatial or temporal resolution is usually fixed and tends to be maximized. For instance, in video surveillance applications FS is typically more important than FR, while in sports videos high FR is generally preferred. Thus, in this section, we further analyze the MOS results under different pixel bitrates and different spatial or temporal resolutions (but with one dimension being fixed). Note that as explained in section IV-A we use pixel bitrate (PB) instead of the common bitrate. Thus, when frame rate and frame size are fixed, the increment of pixel bitrate is equivalent to the enhancement of intra-frame SNR.

Fig. 6 shows the equal-MOS contours on the 2D plane of PB vs. FR. It can be see that in general MOS increases as PB or FR goes up. In particular, for the results of group V_1, the stripes in the figures roughly take a horizontal direction, which indicates that increasing PB brings more significant perceptual quality improvement than increasing the frame rate. This phenomenon suggests that for videos with high spatiotemporal activity coded at low bit rates with the CIF resolution, more bits should be allocated to improve the intra-frame quality. For the results of group V_2, their stripes are roughly vertical or oblique, which shows that the influence on MOS from FR becomes larger, at least as significant as PB. This is because of the characteristics of the video content in group V_2. For example, the 'container' sequence has extremely low spatial activity as shown in Fig. 2(b), and thus only increasing PB does not enhance the visual quality much.

Fig. 7 shows the equal-MOS contours on the 2D plane of PB vs. FS. Unlike the previous MOS results in Fig. 5 and Fig. 6, the performance disparity between groups V_1 and V_2 in Fig. 7 is not distinctive. The directions of all the stripes are approximately horizontal, which indicates increasing PB brings more significant perceptual quality improvement than increasing FS. This phenomenon can be explained as follows. Since the frame rate is fixed at 30 fps in this case, changing PB or FS only affects intra-frame distortions. As we mention in section III-B, in our experiments QCIF images are interpolated to CIF before displaying, and this up-sampling process inevitably causes image blurs. However, at low bitrate conditions, such blurs generally have less negative impact than other types of intra-frame distortions such as blockiness and ringing. As a consequence, more bits should be allocated to reduce the more significant distortions.

V. Conclusion

In this paper, the extensive subjective view tests for assessing the perceptual quality of low bitrate videos have been conducted, which cover 150 test scenarios and include five distinctive dimensions: encoder type, video content, bitrate, frame size and frame rate. Through statistical analysis, we have made the following interesting observations. First, we found that in general the perceptual quality of a decoded video is affected by the encoder type, video content, bitrate, frame rate and frame size in a descending order of significance.



(a) Coastguard (b) Foreman (c) Tempete

(d) Container (e) News (f) Average behavior

Fig. 6. Equal-MOS contours for the five sequences under different pixel bitrate (PB) and frame rate (FR) but with a fixed frame size (FS) of CIF.

Second, for nature videos coded by H.264, despite their frame rate and frame size (QCIF or CIF), generally the given pixel bitrate should be at least around 0.1 bpp in order to achieve 'good' or 'excellent' perceptual quality. Third, for the optimal combination of frame rate and frame size, we found that under a low bitrate constraint, small frame size is often preferred while frame rate should typically be kept low (high) for video sequences with high (low) temporal activity. Fourth, in the cases of using relatively high spatial resolution (CIF) or temporal resolution (30 fps) at low bitrates, we found that in general improving intra-frame SNR becomes the most efficient way to enhance the perceptual quality except for video sequences containing very low spatial activity. We believe our reported results can provide general guidelines for cross-dimensional video assessment and adaptation at low bitrates.

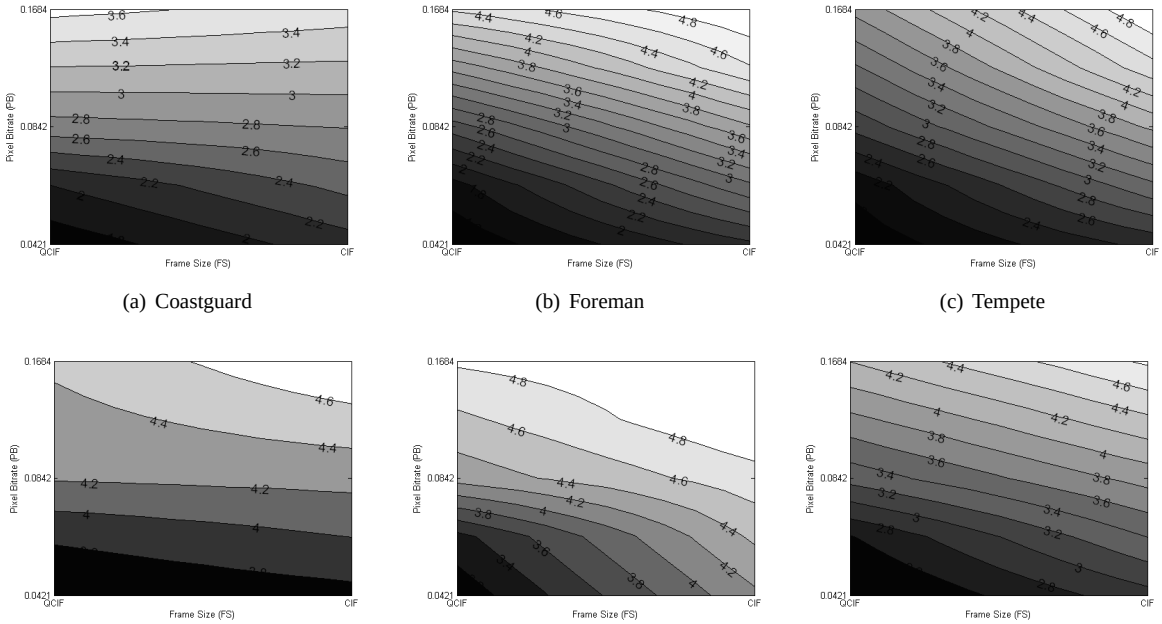
In the future, we want to test more video sequences, especially those containing large motions. We would also like to design new video quality assessment algorithms based on this study. In addition, it would be interesting to describe the video encoder type in terms of complexity so as to study the tradeoff between the complexity of video coding and the video quality.

APPENDIX

Table V and Table VI list out the MOS data shown in Fig. 3(a) and Fig. 3(b).

REFERENCES

[1] VQEG, "Video quality expert group," *www.vqeg.org*, 1997.
[2] ——, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *online available, www.vqeg.org*, 2000.
[3] ——, "VQEG final report of FR-TV phase II validation test," *online available, www.vqeg.org*, 2003.
[4] M. Yuen, *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL: CRC Press, 2006, ch. Coding Artifacts and Visual Distortions, pp. 87–122.
[5] M. Yuen and H. R. Wu, "Survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–278, 1998.
[6] Z. Lu, W. Lin, B. C. Seng, S. Kato, S. Yao, E. Ong, and X. Yang, "Measuring the negative impact of frame dropping on perceptual visual quality," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5666, pp. 554–562, 2005.
[7] R. Pastrana-Vidal, J. Gicquel, C. Colomes, and H. Cherifi, "Sporadic frame dropping impact on quality perception," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 5292, no. 1, pp. 182–193, 2004.
[8] R. Rajendran, M. van der Schaar, and S.-F. Chang, "FGS+: optimizing the joint SNR-temporal video quality in MPEG-4 fine grained scalable coding," *2002 IEEE International Symposium on Circuits and Systems. Proceedings*, vol. 1, pp. 445–448, 2002.
[9] Y. Wang, S.-F. Chang, and A. Loui, "Subjective preference of spatio-temporal rate in video adaptation using multi-dimensional scalable coding," *2004 IEEE International Conference on Multimedia and Expo (ICME)*, vol. 3, pp. 1119–1122, 2004.
[10] N. Cranley, P. Perry, and L. Murphy, "Optimum adaptation trajectories for streamed multimedia," *Multimedia Systems*, vol. 10, no. 5, pp. 392–401, 2005.
[11] J. Martens, "Multidimensional modeling of image quality," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 133–153, 2002.
[12] S. Aleksandr, G. Alexander, and A. M. Eskicioglu, "Multidimensional image quality measure using singular value decomposition," *Proceedings of SPIE*, vol. 5294, pp. 82–92, 2003.
[13] ITU, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT. 500-11*, 2002.
[14] G. Ghinea and J. Thomas, "Quality of perception: user quality of service in multimedia presentations," *IEEE Transactions on Multimedia*, vol. 7, no. 4, pp. 786–789, 2005.
[15] MPEG-4 AVC/H.264 Video Group, "Advanced video coding for generic audiovisual services," *ITU-T Rec. H.264 (03/2005)*, 2005.
[16] T. Wiegand, G. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
[17] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 8th ed. Iowa State University Press, 1989.
[18] R. V. Hogg and J. Ledolter, *Engineering Statistics*. New York: MacMillan, 1987.



(a) Coastguard (b) Foreman (c) Tempete

(d) Container (e) News (f) Average behavior

Fig. 7. Equal-MOS contours of the five sequences under different pixel bitrate (PB) and frame size (FS) with a fixed frame rate (FR) of 30 fps.

[19] P. Sankar and L. Ferrari, "Simple algorithms and architectures for b-spline interpolation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 2, pp. 271–276, 1988.

[20] A. Vatakis and C. Spence, "Evaluating the influence of frame rate on the temporal aspects of audiovisual speech perception," *Neuroscience Letters*, vol. 405, no. 1-2, pp. 132–136, 2006.

TABLE V
H.263 MOS RESUTLS

Sequence	Frame Size	Frame Rate	Bit Rate	MOS
container	CIF	30	128	2.26
container	CIF	15	128	2.47
container	CIF	30	384	3.16
foreman	CIF	15	128	1.21
foreman	CIF	30	128	1.21
foreman	CIF	30	384	2.26
coastguard	CIF	15	128	1.11
coastguard	CIF	30	128	1.32
coastguard	CIF	30	384	2.42
news	CIF	30	128	2.11
news	CIF	15	128	2.16
news	CIF	30	384	3.63
tempete	CIF	15	128	1.47
tempete	CIF	30	128	1.58
tempete	CIF	30	384	1.68
container	CIF	15	64	1.95
container	CIF	7.5	128	2.74
container	CIF	7.5	64	2.89
foreman	CIF	7.5	64	1.16
foreman	CIF	15	64	1.32
foreman	CIF	7.5	128	1.47
coastguard	CIF	15	64	1.21
coastguard	CIF	7.5	64	1.58
coastguard	CIF	7.5	128	1.89
news	CIF	15	64	1.63
news	CIF	7.5	64	1.89
news	CIF	7.5	128	3.05
tempete	CIF	7.5	64	1.53
tempete	CIF	15	64	1.63
tempete	CIF	7.5	128	2.16
container	QCIF	15	24	1.78
container	QCIF	7.5	24	2.06
container	QCIF	15	48	2.5
foreman	QCIF	15	24	1.06
foreman	QCIF	7.5	24	1.06
foreman	QCIF	15	48	1.17
coastguard	QCIF	15	24	1.06
coastguard	QCIF	7.5	24	1.11
coastguard	QCIF	15	48	1.44
news	QCIF	15	24	1.44
news	QCIF	7.5	24	1.78
news	QCIF	15	48	2.17
tempete	QCIF	15	24	1.11
tempete	QCIF	7.5	24	1.28
tempete	QCIF	15	48	1.44
container	QCIF	15	64	2.44
container	QCIF	7.5	48	2.83
container	QCIF	7.5	64	2.94
foreman	QCIF	7.5	48	1.56
foreman	QCIF	15	64	1.56
foreman	QCIF	7.5	64	1.78
coastguard	QCIF	15	64	1.39
coastguard	QCIF	7.5	48	1.61
coastguard	QCIF	7.5	64	1.78
news	QCIF	15	64	2.33
news	QCIF	7.5	48	2.33
news	QCIF	7.5	64	2.67
tempete	QCIF	15	64	1.39
tempete	QCIF	7.5	48	1.44
tempete	QCIF	7.5	64	1.72
container	QCIF	30	48	1.61
container	QCIF	30	64	1.94
container	QCIF	30	128	3
foreman	QCIF	30	48	1.11
foreman	QCIF	30	64	1.22
foreman	QCIF	30	128	2.33
coastguard	QCIF	30	48	1.06
coastguard	QCIF	30	64	1.06
coastguard	QCIF	30	128	1.94
news	QCIF	30	48	1.56
news	QCIF	30	64	2.06
news	QCIF	30	128	3
tempete	QCIF	30	64	1.06
tempete	QCIF	30	48	1.11
tempete	QCIF	30	128	1.61

TABLE VI
H.264 MOS RESULTS

Sequence	Frame Size	Frame Rate	Bit Rate	MOS
container	CIF	15	128	3.42
container	CIF	30	128	3.74
container	CIF	30	384	4.53
foreman	CIF	30	128	2.16
foreman	CIF	15	128	2.42
foreman	CIF	30	384	4.58
coastguard	CIF	30	128	2.11
coastguard	CIF	15	128	2.32
coastguard	CIF	30	384	3.16
news	CIF	15	128	4
news	CIF	30	128	4.16
news	CIF	30	384	5
tempete	CIF	30	128	2.47
tempete	CIF	15	128	3.47
tempete	CIF	30	384	4.53
container	CIF	7.5	64	2.84
container	CIF	7.5	128	3.32
container	CIF	15	64	3.42
foreman	CIF	15	64	1.53
foreman	CIF	7.5	64	2.05
foreman	CIF	7.5	128	3.37
coastguard	CIF	15	64	1.74
coastguard	CIF	7.5	128	2.53
coastguard	CIF	7.5	64	2.63
news	CIF	15	64	3.32
news	CIF	7.5	64	3.47
news	CIF	7.5	128	4.05
tempete	CIF	15	64	2
tempete	CIF	7.5	64	3.05
tempete	CIF	7.5	128	3.37
container	QCIF	7.5	24	3.39
container	QCIF	15	24	3.56
container	QCIF	15	48	4.06
foreman	QCIF	15	24	1.56
foreman	QCIF	7.5	24	2.11
foreman	QCIF	15	48	2.89
coastguard	QCIF	15	24	1.94
coastguard	QCIF	7.5	24	2.72
coastguard	QCIF	15	48	3.17
news	QCIF	15	24	2.89
news	QCIF	7.5	24	3.22
news	QCIF	15	48	4.22
tempete	QCIF	15	24	1.89
tempete	QCIF	7.5	24	2.89
tempete	QCIF	15	48	3.17
container	QCIF	7.5	48	3.22
container	QCIF	7.5	64	3.61
container	QCIF	15	64	4.11
foreman	QCIF	7.5	48	2.89
foreman	QCIF	15	64	3.17
foreman	QCIF	7.5	64	3.44
coastguard	QCIF	15	64	2.94
coastguard	QCIF	7.5	48	3.11
coastguard	QCIF	7.5	64	3.28
news	QCIF	7.5	48	3.83
news	QCIF	7.5	64	4.11
news	QCIF	15	64	4.5
tempete	QCIF	15	64	3.28
tempete	QCIF	7.5	48	3.28
tempete	QCIF	7.5	64	3.67
container	QCIF	30	48	3.83
container	QCIF	30	64	4.22
container	QCIF	30	128	4.44
foreman	QCIF	30	48	1.72
foreman	QCIF	30	64	2.5
foreman	QCIF	30	128	4.44
coastguard	QCIF	30	48	2
coastguard	QCIF	30	64	2.72
coastguard	QCIF	30	128	3.67
news	QCIF	30	48	3.33
news	QCIF	30	64	4.28
news	QCIF	30	128	4.83
tempete	QCIF	30	48	2.06
tempete	QCIF	30	64	2.67
tempete	QCIF	30	128	3.67