

SUBJECTIVE QUALITY ASSESSMENT OF SCALABLE VIDEO CODING: A SURVEY

Jong-Seok Lee, Francesca De Simone, Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{jong-seok.lee, francesca.desimone, touradj.ebrahimi}@epfl.ch

ABSTRACT

Scalable video coding is a promising solution for efficient video content distribution to users having heterogeneous network and terminal capabilities. Thanks to its inherent multi-dimensional adaptability, a scalable bit stream can be used to simultaneously transmit multiple video sequences having different bit rates to corresponding target users, without necessity of re-encoding or transcoding. In order to exploit such an advantage effectively, it is crucial to understand the effects of multidimensional scalability options on the perceived quality and the trade-off between the scalability dimensions through subjective quality assessment. This paper reviews existing studies aiming at achieving this goal in order to summarize their results and common findings.

Index Terms— Scalable video coding, subjective quality assessment, quality of experience

1. INTRODUCTION

Nowadays, applications of video content delivery are very popular thanks to the advances of hardware and software technologies for video production and processing. Usually, a content, e.g. a video clip on a video sharing web site, is consumed by many users. This brings an important research issue: how to efficiently deliver video content to consumers having diverse communication environments. Network resources (e.g. bandwidth) available for different users may be quite different and time-varying. In addition, the characteristics of users' terminals may vary significantly in terms of display resolution, processing power, etc. Therefore, the same content needs to be delivered at the same time in different formats according to these variables.

Scalable video coding is a useful concept that can deal with such difficulty in multimedia content delivery. A scal-

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2011) under grant agreement no. 216444 (PetaMedia), the Swiss National Foundation for Scientific Research in the framework of the NCCR Interactive Multimodal Information Management (IM2), and the COST Action IC1003 European Network on Quality of Experience in Multimedia Systems and Services (Qualinet).

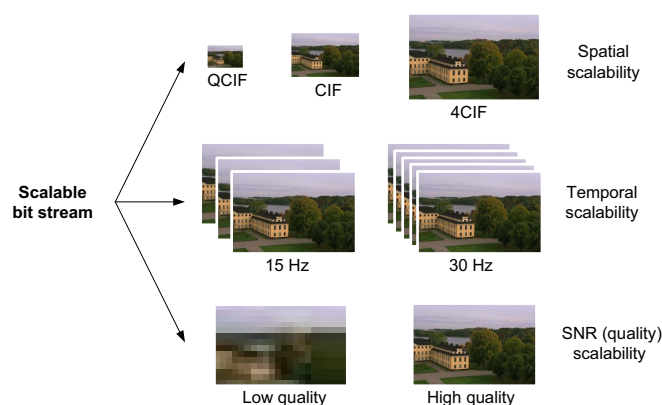


Fig. 1. Three-dimensional video scalability.

able video bit stream is composed of various layers, which can be adapted to a given bit rate constraint through truncation of parts of the bit stream. In general, the video scalability is realized in the following three dimensions (Fig. 1):

- Spatial scalability, i.e. possibility of reducing the frame size;
- Temporal scalability, i.e. possibility of reducing the frame rate;
- Signal-to-noise ratio (SNR) scalability (or quality scalability), i.e. possibility of reducing the frame quality.

Having the three different scalability dimensions, video transmission using scalable video coding schemes requires an adaptive strategy to determine which scalability options to be used for given resource constraints. When a bit rate limitation is given, the three dimensions have trade-off relationship, e.g. increasing the spatial resolution can be done only at the cost of the decreased temporal resolution and/or frame quality. The ultimate goal of such a strategy is to maximize end-users' quality of experience for the delivered content. Therefore, it is important to understand the impact of different scalability options and their combinations on human

observers' quality perception through subjective quality assessment. This is still a challenging open research topic because it is not straightforward for human subjects to judge the overall quality of video content across scalability dimensions and provide quality scores of the content in a single rating scale. In addition, subjective quality evaluation of video scalability inherently involves heterogeneous video consumption environments, which brings additional factors that may affect perceived quality of scalable video content. As a result, many research questions related to subjective perception of video scalability are still open. Also, it is not easy to compare different studies that sometimes show contradictory results and to figure out their common conclusions.

In this paper, we review existing studies on subjective quality assessment of video scalability, with slight emphasis on our own recent work. Results of the studies are thoroughly analyzed in various viewpoints such as considered scalability dimensions, used codecs, test environments and methodologies, and used stimuli, so that general conclusions can be identified. In addition, perspectives for future work are also given.

The rest of the paper is organized as follows. Section 2 briefly introduces the concept of scalable video coding. In Section 3, a review of the existing work on subjective evaluation of scalable video coding is provided. Finally, concluding remarks along with topics for future work are given in Section 4.

2. SCALABLE VIDEO CODING

In the last decades, a significant amount of research has been devoted to scalable coding for low complexity video adaptation.

The latest hybrid video coding standard, H.264/MPEG-4 Advanced Video Coding (H.264/AVC), provides a scalable extension called SVC [1]. An SVC bit stream is organized in a base layer, which is compatible to a non-scalable profile of H.264/AVC and corresponds to the lowest spatio-temporal resolution and quality, and several enhancement layers that add spatial, temporal, and/or SNR quality to the reconstructed base layer. Each layer is coded according to the design of single-layer H.264/AVC coding (intra-layer coding) by using intra-picture prediction and motion-compensated inter-picture prediction at the macroblock level. In addition, inter-layer prediction methods are used, which predict enhancement layer data from previously reconstructed data of a lower resolution layer. These methods exploit the statistical dependencies between different layers for improving coding efficiency.

The spatial scalability is achieved by applying inter-layer motion, residual and intra-prediction techniques. Coarse-to-fine image pyramid relationships across different resolutions and single-loop decoding are considered. This allows an effective propagation of the information found in the coded mo-

tion vectors and prediction residuals from each lower resolution layer to the higher resolution layer. Moreover, the lower resolution signal may consist of a cropped version of the higher resolution data, thus flexible cropping, scaling and alignment of the origins of the picture regions can be performed across layers.

The temporal scalability is achieved by using hierarchical bidirectional prediction. While the pictures of the base layer are only predicted from previous pictures of this layer, the enhancement layer pictures can be predicted by using the pictures of a lower temporal layer as references.

Finally, the SNR scalability is provided by coarse-grain scalability (CGS) and fine-grain scalability (FGS). CGS is achieved by encoding the texture information in the base layer and producing the enhancement layers by decreasing the quantization step size and encoding successive refinements of the transform coefficients. For FGS, a bit plane coding method of discrete cosine transform (DCT) coefficients in the enhancement layers is used, which allows the enhancement layer bit stream to be truncated at any point.

As an alternative to block-based coding, wavelet-based scalable coding solutions have been also developed [2]. A wavelet transform combined with motion-compensated temporal filtering (MCTF) is applied to perform a spatio-temporal decomposition of the input video sequence. The motion vectors and wavelet coefficients are compressed to remove redundancy and organized in a bit stream having a layered representation. Four different approaches can be distinguished: (1) schemes where MCTF is performed first and then a two-dimensional spatial transform on each temporal subband is applied; (2) schemes where the spatial transform is applied before temporal filtering; (3) schemes based on content-adaptive spatio-temporal decompositions; (4) schemes with pyramidal decomposition applied before or after MCTF.

In multimedia applications employing scalable video coding, only a part of the scalable video stream is sent or decoded at the receiver side, depending on the available network resources or terminal decoding and display capabilities. In addition, when the given communication channel is time-varying, the scalable stream enables fast bit rate adaptation.

The problem of quality assessment of scalable video coding aims at investigating the optimal combination of scalability options that maximize the perceived quality of the delivered content for given resource constraints. Scaling in each scalability dimension may degrade the overall video quality perceived at the end-user side. Frame rate reduction may result in motion jerkiness, frame size reduction causes blurring when upscaling is done for a fixed viewing window, and frame quality reduction typically results in blockiness for block-based encoding or blurring for wavelet-based encoding. Furthermore, fluctuations of video quality due to scalability option switching may negatively affect the quality of users' experience.

3. SUBJECTIVE QUALITY ASSESSMENT OF SCALABLE VIDEO CODING

Existing studies on subjective quality assessment of scalable video coding can be divided into groups with respect to the considered scalability options, which are presented separately below.

3.1. Temporal scalability

Human perception of different temporal resolutions has been investigated for a relatively long time. Especially, much work has been done to find the minimum acceptable frame rates of video stimuli for various tasks such as target tracking, target detection/recognition, lip reading, orientation judgment, etc. As for the scenario of video consumption, various factors affect the perceived quality of different frame rates, e.g. content type, viewing condition, display type, auditory cue, and observer characteristics. For example, a scene with fast motion would need a relatively high frame rate in order to prevent jerkiness artifacts in comparison to a scene with slow motion. A conclusion encompassing the results of the existing studies seems to be that the threshold of a subjective satisfaction level is around 15 Hz, although the exact value will vary significantly according to the aforementioned factors [3].

3.2. Temporal and SNR scalability

For a certain target bit rate, once the spatial resolution is fixed, the temporal resolution can be increased only at the cost of decreased frame quality. A significant amount of work has been done in order to examine the optimal trade-off between these two dimensions in terms of perceived quality of the final video sequence.

Traditionally, it is believed that a high frame rate is more important for a content with fast motion than a high frame quality, which is supported by [4, 5]. In [4], subjective experiments were conducted for video sequences encoded by using three different codecs (i.e. the Sorenson codec 2.1, H.263+ and a wavelet-based codec) for eight content types. For a fixed resolution of 352×240 pixels, three frame rates (10, 15 and 30 Hz) were considered. Overall, a frame rate of 15 Hz was most preferred across different coding conditions. However, content-dependence was observed, i.e. for content with slow (or fast) motion, preference of a frame rate of 10 Hz (or 30 Hz) was nearly as high as 15 Hz. Similarly, a double stimulus continuous quality scale (DSCQS) experiment in [5] compared H.263+ sequences coded at three different frame rates (7.5, 15 and 30 Hz) with five quantization parameter (QP) values, for a fixed spatial resolution of 320×192 pixels. The results showed that, for slow motion content, subjective quality degradation due to frame rate reduction was only minor.

Results contradicting this traditional belief have been reported in [6, 7]. In [6], experiments were conducted under

a desktop environment (CIF display resolution) and a palm-top environment (QCIF display resolution). Frame rates from 6 to 24 Hz were examined for three football scenes containing fast motion, encoded with H.263. A so-called “method of limits” was used, in which subjects were asked to indicate when the quality became acceptable or unacceptable, while the quality parameters were increased or decreased, respectively. Although slight differences were found with respect to the spatial resolution, the subjects were more sensitive to frame quality reduction than frame rate reduction. It was concluded that, in small screen devices, reducing frame quality removes important information about the football players and the ball. Finally, an extensive experiment in [7] using 128 contents at CIF resolution investigated the priority between frame rate and frame quality for a wide range of bit rate conditions (50 kbps to 1 Mbps). Motion compensated wavelet/subband video coding was used for encoding video sequences. The results based on the double stimulus impairment scale (DSIS) methodology showed bit rate-dependent preference, i.e. the three frame rates (7.5, 15 and 30 Hz) were most preferred for low, middle and high bit rate ranges, respectively. The boundaries between the three bit rate ranges were higher for complex scenes.

3.3. Spatial and temporal scalability

In [8], the trade-off between the spatial and temporal dimensions was studied based on the paired comparison (PC) methodology. For each bit rate condition, feasible combinations of different spatial and temporal resolutions were compared by 120 subjects. The two resolutions ranged 40-100% of QCIF and 5-25 Hz, respectively, and MPEG-4 encoding was used. Sequences having lower resolutions than QCIF were up-scaled up to the original frame size for presentation. It was found that, in the two-dimensional space of spatial and temporal scalability options, there exists a so-called “optimal adaptation trajectory” (OAT) that ensures the maximal perceived quality for each fixed bit rate when the trade-off of spatial and temporal alternatives is considered. Content-dependence of the OAT was also shown, i.e. the frame rate has higher priority over the frame size for content containing fast motion.

3.4. Spatial and SNR scalability

The trade-off between the spatial and SNR dimensions was studied in [5]. Combinations of three spatial resolutions (50, 75 and 100% of an original resolution of 320×192 pixels) and five QP values in H.263+ were compared for a fixed frame rate of 30 Hz. An important finding of the work was that for low bit rate conditions, a small frame size with smaller quantization errors is preferable to a large frame size with large quantization errors. It should be noted that, unlike other studies, the lower spatial resolution sequences were not up-scaled to their full original size. Thus, it is not possible to

directly compare these results with those reported in other studies where spatial upscaling is used.

3.5. Spatial, temporal and SNR scalability

Studies considering the three-dimensional scalability in subjective quality assessment has been presented only recently [9–12]. In all the studies presented below, sequences having lower frame sizes were always up-scaled to the largest original sizes by assuming fixed viewing windows.

In [9], subjective quality assessment of low bit rate video sequences (up to 382 kbps) was conducted. Two non-scalable encoding schemes, H.263 and H.264/AVC, were used to produce test stimuli for two spatial resolutions (QCIF and CIF) and three temporal resolutions (7.5, 15 and 30 Hz). From the results obtained using the double stimulus impairment scale variant II (DSIS II) methodology, it was concluded that the perceptual quality is affected by the encoder type, video content, bit rate, frame rate and frame size in a descending order of significance. As for the optimal combination of the spatial and temporal resolutions, a small frame size was preferred while the frame rate should be kept low (high) for a high (low) motion activity. When the spatial or temporal resolution was relatively high at low bit rates, the frame quality needed to be improved to efficiently enhance the perceived quality except for the content with a very low spatial complexity. The frame quality was especially important for H.264/AVC-encoded video sequences depicting nature scenes, so that the pixel bit rate should be at least 0.1 bit per pixel despite their frame rate and frame size.

Subjective evaluation of video sequences produced by SVC for mobile environments was performed in [10]. The bit rate conditions were chosen to be representative of wireless networks such as HSDPA and DVB-H. The audio channel was accompanied during the stimulus presentation. The results obtained by using the DSCQS methodology showed that scaling in the spatial dimension influences the perceived quality more significantly than scaling along the temporal or SNR dimensions. This is partially because the spatial scaling yielded larger bit rate changes in the chosen operation points of SVC.

In [11], a series of subjective experiments using stimuli obtained by SVC were presented. It was shown that, when the spatial resolution was fixed, subjects preferred a higher frame rate over a higher frame quality. When the temporal and spatial dimensions were compared, most of the subjects preferred the sequences having blurring artifacts due to up-scaling for small frame sizes against the jerky sequences having low frame rates. Content-dependence of the perceived quality was also examined, from which it was concluded that the correlation between the content and perceived quality decreases according to the increase of the frame size, frame rate or frame quality.

In [12], we presented an extensive subjective experiment

Bit rate	SVC	wavelet-based
Low	$T > R$	-
High	$T > R$	$T > R$

(a)

Bit rate	SVC	wavelet-based
Low	$T < S$	-
High	$T > S$	$T > S$

(b)

Bit rate	SVC	wavelet-based
Low	-	$S > R$
High	$S < R$	$S > R$

(c)

Fig. 2. Summary of the subjective evaluation results for SVC and a wavelet-based scalable codec as presented in [12]. The bit rate conditions are divided into ‘low’ and ‘high’ conditions with thresholds of about 700 kbps and 900 kbps for the two codecs, respectively. $A > B$ means an improvement in dimension A is preferable to an improvement in dimension B . S , T and R indicates the spatial, temporal and SNR dimensions, respectively. (a) When the frame size is fixed. (b) When the frame size and rate vary simultaneously. (c) When the frame rate is fixed.

where two scalable codecs (SVC and a wavelet-based codec) were used to produce stimuli of wide ranges of bit rates (up to 4 Mbps), frame rates (up to 50 Hz), frame size (up to 1280×720). The results from the PC-based tests can be summarized as shown in Fig. 2. For fixed frame sizes, the frame rate and quality are compared, where the frame quality is governed by coding artifacts (Fig. 2(a)). The results showed that a higher frame rate was always preferred against a better quality. The frame rate also appeared to be important when the frame rate and size varied simultaneously (Fig. 2(b)). An exception was observed when the bit rate was small, which indicates that for a low bit rate condition, enhancing the frame quality through increase of the frame size is more important than increasing the frame rate. For fixed frame rates (Fig. 2(c)), the comparison is between the frame size, related to the amount of blurring artifacts, and the frame quality, affected by coding artifacts. For SVC, a better quality was preferred over a larger frame size, while a larger size was more important in the wavelet-based codec. This is due to the fact that the coding artifacts of the two codecs are fundamentally different, i.e. blockiness in SVC versus blurring in the wavelet-based codec, and blurring artifacts are usually less annoying than blockiness artifacts.

Although it is not straightforward to directly compare the studies presented up to now, they can be roughly summarized as follows. Since a small frame size is usually up-scaled to the full original resolution by assuming a fixed viewing win-

low, the trade-off among the three scalability dimensions is in effect between the frame rate and the frame quality affected by blurring and coding artifacts. It seems that there exists a bit rate threshold where the subjective preference among the scalability options is switched. When the bit rate is smaller than the threshold, a better frame quality level achieved in the spatial or SNR dimensions is preferable against a higher frame rate. Above the threshold, the frame quality becomes acceptable and thus increasing the frame rate is effective for better perceived quality. The threshold highly depends on the content type, but other factors such as the codec and viewing condition are also involved in determining the threshold.

3.6. Dynamic scalability variations

In [13], the aspect of temporally varying operation points of scalability for adaptive video transmission was considered. It was shown that frequent operation point switching or a large variation of quality switching has an undesirable impact on the perceived quality.

3.7. Other factors

Not only the three scalability dimensions govern the perceived quality of scalable video coding, but also other elements significantly contribute to the perceived quality.

The encoder type is one of such elements. In [9], it was shown that the coder type is the most significant factor affecting perceived quality among five factors (i.e. encoder type, content, bit rate, frame rate and frame size). In [12], general conclusions remained consistent for the employed two encoders, while different types of artifacts generated by the encoders played different roles for some cases.

The viewing environment is also an important factor. The work in [6] compared the results for desktop and palmtop environments, and notable differences in the influence of the frame rate on perceived quality were reported.

Additionally, users' prior experience and expectation are factors affecting perceived quality of scalable video coding [10, 11], although measuring and analyzing their influence quantitatively are not straightforward.

4. DISCUSSION

In this paper, we surveyed existing studies on subjective quality assessment of scalable video coding. We compared and analyzed them with respect to the considered scalability dimensions. As a general conclusion, there exists subjective preference among the scalability options, which varies mainly according to bit rate conditions, but also depends on several factors such as content, encoder types, viewing environments, and even users' experience and expectation.

The findings of subjective studies can be exploited to design adaptive content distribution strategies where the opera-

tion point of scalability is automatically adjusted according to the static or dynamic bit rate constraint (e.g. [14–17]). Moreover, they can be used as a basis for developing objective video quality measures that automatically evaluate the quality of multimedia content as perceived by end-users (e.g. [18, 19]). The objective metrics can be used to perform in-service video quality monitoring so that the service provider can obtain information about the quality of the delivered video content and react accordingly in real-time.

In real-world applications of scalable video coding, factors causing quality degradation may exist, e.g. packet loss and transmission delay, which has been rarely considered in quality assessment of scalable video coding. Consideration of such errors together with multidimensional scalability would be beneficial for designing real-world applications in error-prone network environments.

Additionally, multimedia content accompanies audio channels along with visual channels in most cases. It is well-known that the two modalities directly and indirectly interact during the human perception process of audio-visual content and its quality, e.g. masking effects, synchronization, and focus of attention [20–22], which has been usually neglected in quality assessment of scalable video coding. Thus, it would be desirable to perform visual and audio-visual quality assessment under the presence of the audio channel in the framework of scalable video coding.

In the research field of quality assessment, it is important to have common databases that can be used for benchmarking and comparison of different studies. To the best knowledge of the authors, there exists only one publicly available dataset for quality assessment of scalable video coding, which contains scalable video sequences and subjective data described in [12]¹. Thus, we believe that it is necessary to encourage researchers to publish their datasets for diverse conditions and applications in order to facilitate further advances in this field.

5. REFERENCES

- [1] 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, 2007.
- [2] N. Adami, A. Signoroni, and R. Leonardi, "State-of-the-art and trends in scalable video compression with wavelet-based approaches," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1238–1255, Sept. 2007.
- [3] J. Y. C. Chen and J. E. Thropp, "Review of low frame rate effects on human performance," *IEEE Trans. Syst.*,

¹<http://mmspg.epfl.ch/svd>

- Man, Cybern. A*, vol. 37, no. 6, pp. 1063–1076, Nov. 2007.
- [4] G. Yadavalli, M. Masry, and S. S. Hemami, “Frame rate preferences in low bit rate video,” in *Proc. ICIP*, Barcelona, Spain, Sept. 2003, pp. 441–444.
- [5] D. Wang, F. Speranza, A. Vincent, T. Martin, and P. Blanchfield, “Towards optimal rate control: A study of the impact of spatial resolution, frame rate, and quantization on subjective video quality and bit rate,” in *Proc. SPIE*, 2003, vol. 5150, pp. 198–209.
- [6] J. D. McCarthy, M. A. Sasse, and D. Miras, “Sharp or smooth? comparing the effects of quantization vs. frame rate for streamed video,” in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, Vienna, Austria, Apr. 2004, pp. 535–542.
- [7] Y. Wang, S.-F. Chang, and A. C. Lou, “Subjective preference of spatio-temporal rate in video adaptation using multi-dimensional scalable coding,” in *Proc. ICME*, 2004, pp. 1119–1122.
- [8] N. Cranley, P. Perry, and L. Murphy, “User perception of adapting video quality,” *Int. J. Human-Computer Studies*, vol. 64, pp. 637–647, 2006.
- [9] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, “Cross-dimensional perceptual quality assessment for low bit-rate videos,” *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.
- [10] A. Eichhorn and P. Ni, “Pick your layers wisely- a quality assessment of H.264 scalable video coding for mobile devices,” in *Proc. Int. Conf. Communications*, Dresden, Germany, June 2009, pp. 1–6.
- [11] W. Song, D. Tjondronegoro, and S. Azad, “User-centered video quality assessment for scalable video coding of H.264/AVC standard,” in *Proc. Int. Conf. Multimedia Modeling*, Chong Qing, China, Jan. 2010, pp. 55–65.
- [12] J.-S. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurupepe, T. Sikora, J. Ostermann, E. Izquierdo, and T. Ebrahimi, “Subjective evaluation of scalable video coding for content distribution,” in *Proc. ACM Multimedia*, Firenze, Italy, Oct. 2010, pp. 65–72.
- [13] M. Zink, J. Schmitt, and R. Steinmetz, “Layer-encoded video in scalable adaptive streaming,” *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 75–84, Feb. 2005.
- [14] Y. Wang, M. van der Schaar, S.-F. Chang, and A. C. Loui, “Classification-based multidimensional adaptation prediction for scalable video coding using subjective quality evaluation,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1270–1279, 2005.
- [15] N. Cranley, P. Perry, and L. Murphy, “Dynamic content-based adaptation of streamed multimedia,” *J. Netw. Comput. Appl.*, vol. 30, pp. 983–1006, 2007.
- [16] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, “Three dimensional scalable video adaptation via user-end perceptual quality assessment,” *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 719–727, Sept. 2008.
- [17] C. S. Kim, H. Sohn, W. De Neve, and Y. M. Ro, “An objective perceptual quality-based ADTE for adapting mobile SVC video content,” *IEICE Trans. Info. Syst.*, vol. E92-D, no. 1, pp. 93–96, Jan. 2009.
- [18] H. Sohn, H. Yoo, W. De Neve, C. S. Kim, and Y. M. Ro, “Full-reference video quality metric for fully scalable and mobile SVC content,” *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 269–280, Sept. 2010.
- [19] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, “Perceptual quality assessment of video considering both frame rate and quantization artifacts,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 286–298, Mar. 2011.
- [20] S. Winkler and C. Faller, “Perceived audiovisual quality of low-bitrate multimedia content,” *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 973–980, Oct. 2006.
- [21] J.-S. Lee, F. De Simone, and T. Ebrahimi, “Influence of audio-visual attention on perceived quality of standard definition multimedia content,” in *Proc. Int. Workshop on Quality of Multimedia Experience*, San Diego, CA, July 2009, pp. 13–18.
- [22] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, “Perceptual-based quality assessment for audio-visual services: a survey,” *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482–501, Aug. 2010.