Perceptual Asymmetric Video Coding for 3D-HEVC

Yongfang Wang[1(✉)], Kanghua Zhu[1], Yawen Shi[1], and Pamela C. Cosman[2]

[1] School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China
yfw@shu.edu.cn

[2] Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA

Abstract. In the scope of stereoscopic 3D video, asymmetric video coding is an effective method in terms of maintaining the perceived quality while reducing the required transmission bandwidth by exploiting the perceptual phenomenon of binocular suppression. On the other hand, just noticeable distortion (JND) has been applied successfully in improving the video coding efficiency by removing human visual redundancies. However perceptual asymmetric coding combined with JND has not been studied. In this paper, the effectiveness of using JND aided asymmetric 3D video coding is explored for 3D-HEVC. We conducted extensive subjective tests, which indicate that if the base view is encoded at perceptual high quality and the dependent view is encoded at a lower perceptual quality, then the degradation in 3D video quality is unnoticeable and asymmetric just-noticeable distortion threshold is gained. Furthermore, we proposed a novel perceptually asymmetric 3D video coding framework by taking full advantage of these observations and subjective test results. Experimental results demonstrate that, compared with HTM, the proposed asymmetric 3D-HEVC video coding demonstrates comparable 3D perceived visual quality with about 13% bitrates savings in the whole view and about 32% bitrates savings for dependent view.

Keywords: Asymmetric 3D video coding · JND · 3D-HEVC

1 Introduction

Recent improvements in 3D video technology led to a growing interest in 3D video. While an efficient compression algorithm for 3D content is vital for the adoption of 3D technology. The state-of-the-art standard for multiview video coding (MVC) was mainly developed for efficient compression of scenes from different viewpoints. Since the bit rate required for MVC increases approximately linearly with the number of coded views, Multiview Video plus Depth (MVD) could be specified as an extension of MVC in order to support different stereoscopic as well as multi-view displays. In 2012, the standardization projects for MVD video were developed by MPEG and by ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) with the goal of developing a new standard for future applications.

E. Chen et al. (Eds.): PCM 2016, Part II, LNCS 9917, pp. 498–507, 2016.
DOI: 10.1007/978-3-319-48896-7_49

JCT-3V drafted two test models for a 3D video coding standard including the AVC-based (3D-AVC) and the HEVC-based (3D-HEVC) [1]. 3D-HEVC provides coding gains better than 3D-AVC, whose reference software implementation is called HTM [2]. Currently, 3D-HEVC is a popular choice for 3D entertainment media distribution. However, the delivery over bandwidth constrained networks exhibits challenges, which enforce the transmission system to perform perception-aware coding to save bitrates.

Binocular suppression show that the binocular perception of stereoscopic image pair is dominated by the high quality component [3], which is better supported with an asymmetric quality. Binocular suppression visual characteristics have been extensively investigated in asymmetric coding [4–10], which is a promising method for stereo compression to reduce bandwidth by the unequal quantization of views or unequal subsampling. In [4], a stereoscopic video coding method is proposed with asymmetric luminance and chrominance qualities based on the suppression theory of binocular vision. For asymmetry by spatial resolution reduction, Fehn et al. [5] have proposed an integration method of mixed-resolution with MVC by modifying the Disparity Compensated Prediction (DCP) loop across the views of different spatial resolutions and achieved significant average bit-rate gains. The study in [6] reports little quality and depth sensation degradations even when the spatial frequency was reduced to half of its original bandwidth. Authors in [7] had subjectively compared the performances of symmetric, asymmetric, and spatially mixed-resolution stereoscopic video coding. It was concluded that the performance of spatially mixed-resolution coding, where the auxiliary view is encoded at half resolution in both dimensions, is similar to that of the symmetric and asymmetric quality coding at full resolution. Low-resolution video coding has been found to be advantageous in terms of processing complexity. For temporal asymmetry, Anil et al. [8] proposed a low-weight frame skipping method for the secondary view to decrease the overall bitrate. Based on the subjective analysis done in [9], the performance of asymmetric quality stereoscopic video coding against that of the symmetric quality coding differs at different operating ranges. Asymmetric coding provides bit-rate gains without perceptually noticeable differences, if the auxiliary view is encoded at a higher quality than a certain threshold, while the other view is encoded at a lower quality but above a certain PSNR threshold.

Erhan et al. [10] proposed a visual attention aided ROI coding method in the context of asymmetric stereoscopic 3D video compression, in terms of achievable bit-rate gains and the dependency to the characteristics of the stereoscopic content. A mixed resolution coding method in [11] have been developed on the presumption of a certain visual fatigue response, which compared two methods of mixed resolution coding, single-eye and alternating-eye blur, in terms of overall quality for short exposures and visual fatigue level for long exposures. It is reported in [9] have concluded that PSNR reduction method is more suitable for asymmetric stereo video coding in the context of adaptive streaming at sufficiently high bitrates.

The aforementioned methods for asymmetric stereoscopic video coding have been proven to be effective as a mean for network adaptation. But only a limited number of works had incorporated just noticeable distortion (JND) model in the scope of asymmetric stereoscopic video coding. In the paper, we investigated the effectiveness of JND coding in the context of asymmetric stereoscopic 3D video compression by jointly considering the perceptual effect of binocular suppression and spatial-temporal JND model. Furthermore, the effect of 3D JND threshold levels based on binocular suppression within the asymmetric coding approach is tested subjectively. A large scale subjective test is conducted and the results are analyzed. The presented scheme can be used in a 3D-HEVC framework and achieve very high compression gains without reducing the overall perceived 3D video quality compared with state-of-the-art algorithm.

2 Perceptual Asymmetric Coding

2.1 Proposed Perceptual Asymmetric Coding Method

A flowchart of the proposed perceptual asymmetric coding method is shown in Fig. 1. Suppose that the left eye is considered as the dominant eye and it is encoded in HTM guided by JND model - J_{st}, which has not perceptual distortion in 2D display, and the right view is compressed with HTM guided by the proposed JND model - J_{3d}, which has perceptual distortion in 2D display but has not perceptual distortion combined with left view in 3D display. We define J_{3d} as following

$$J_{3d}(i,j) = J_{st}(i,j) \times Jb(i,j) \tag{1}$$

where J_{3d} denotes the 3D JND threshold of pixel (i,j),

$J_{st}(i,j)$ is the threshold of spatial and temporal JND model [12]. $Jb(i,j,n)$ is binocular JND threshold based on binocular suppression, which is obtained by the experiment in Sect. 2.2.

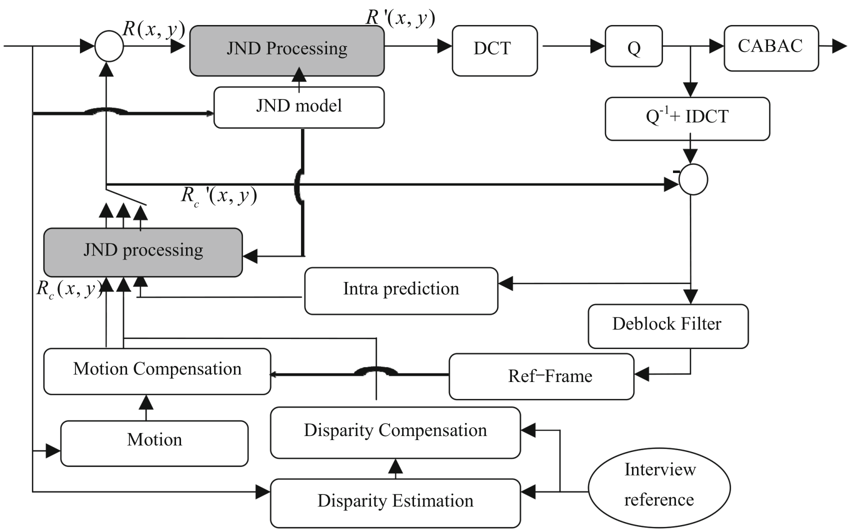


Fig. 1. A flowchart of the proposed perceptual asymmetric coding

The coding method guided by JND model is to pre-process residual coefficients and the distortion coefficient between reconstructed frame and original frame. If the residuals are less than the JND threshold, they can be discarded to save bit rate, or directly subtract the JND threshold from residual value if JND threshold is less than residual value, while maintains subjective quality.

$$R'(i,j) = \begin{cases} R(i,j) - J(i,j), & R(i,j) > J(i,j) \\ 0, & |R(i,j)| \leq J(i,j) \\ R(i,j) + J(i,j), & -R(i,j) > J(i,j) \end{cases} \tag{2}$$

where $R(i,j)$ and $R'(i,j)$ is the original residual signal and JND-processed residual signal in pixel domain, $J(i,j)$ is the JND threshold, which is $J_{st}(i,j)$ for left view and $J_{3d}(i,j)$ for right view. Besides, In order to further remove the perceptual redundancy, the distortion coefficient $R_c(i,j)$ between reconstructed frame and original frame will be pre-process. The JND processed distortion coefficient $R'_c(i,j)$ is denoted as follow:

$$R'_c(i,j) = \begin{cases} R_c(i,j) - J(i,j), & R_c(i,j) > J(i,j) \\ 0, & |R_c(i,j)| \leq J(i,j) \\ R_c(i,j) + J(i,j), & -R_c(i,j) > J(i,j) \end{cases} \tag{3}$$

2.2 JND Threshold Based on Binocular Suppression

JND refers to the maximum possible distortion in the signal, which is imperceptible to human eyes. According to the suppression theory of stereo human vision, the HVS can tolerate absence of high frequency information in one of the views; therefore, the two views can be represented at unequal resolutions or bitrates. This means that there exists the unequal maximum possible distortion for two views, which is JND threshold based on binocular suppression. One view is considered as the dominant eye and has not perceptual distortion in 2D display by JND processing. The other view which has perceptual distortion in 2D display but has not perceptual distortion combined with other view in 3D display by 3D JND processing based on binocular suppression.

In order to obtain JND threshold based on binocular suppression $Jb(i,j)$, our experiment is conducted in HTM platform, a typical prediction structure of 3D video coding is a hierarchical B picture (HBP) prediction structure. Left view is independently encoded and right view is encoded with disparity compensated prediction between views. For reference stereoscopic video, we encode in HTM for both left and right views. For asymmetric stereoscopic video, suppose that the left eye is encoded in HTM guided by JND model - J_{st}, and the right view is compressed with HTM guided by the proposed JND model - JND_{3d}. For the left video at each possible quality level (PSNR), Jb has been adjusted from 1 to increased by 1 until the distortion was becoming noticeable in 3D-display compared to reference stereoscopic video.

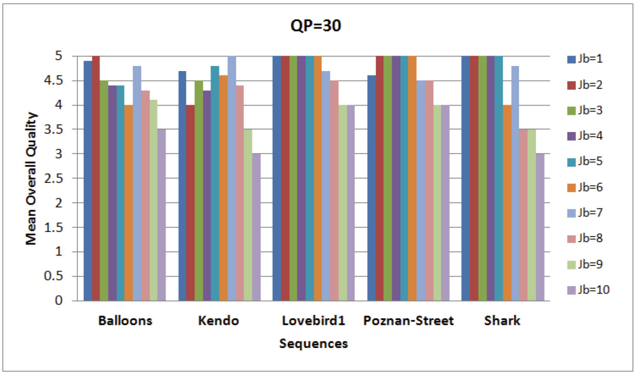
We have conducted a subjective experiment using 5 stereoscopic video sequences [13]. The test stimuli were shown on a 50-inch Panasonic TH-P50ST30C stereoscopic display with a 2D equivalent resolution of $1920 * 1080$ using polarized glasses. The aspect ratio of this display is 16:9, and the contrast ratio is 2000:1, where the picture height is 74.4 cm, the picture width is 120.4 cm. The distance of the subjects from the 3D display was 3.7 m that is approximately 5 times the picture height. The studio illumination was set to 200 lx to mimic home viewing conditions. This is an interactive subjective test that starts by displaying both views at reference stereoscopic video. A total of 13 subjects were tested, ranging in age from 23 to 29 years, 9 males and 4 female.

During the test, viewers reduce the perceptual quality of right view down to about 30 dB by incrementing JND value processing. The viewers always compare the current quality level against the reference stereoscopic video and continue to decrement the quality until coding artifacts become noticeable. This PSNR value and *Jb* is recorded. Figure 2 shows the average *Jb* threshold values for different quality stereoscopic video (over all subjects). Each of the 13 tests was 30 s long, consisting of the 8-second reference video followed by the processed test video shown twice. Preceding each video clip was a 2-second gray screen indicating whether the reference or test video was going to be shown. After each test, the subjects rated the overall quality of the test clip relative to the reference clip, indicating the level of difference or degradation on the following scale: (5) imperceptible, (4) perceptible, but not annoying, (3) slightly annoying, (2) annoying, (1) very annoying. This test design is adapted from the Double-Stimulus Impairment Scale (DSIS) method recommended in [14]. The scores are averaged across subjects and both trials of each video. If the variance between trials was large for any particular subject, that subject's data was discarded for video.

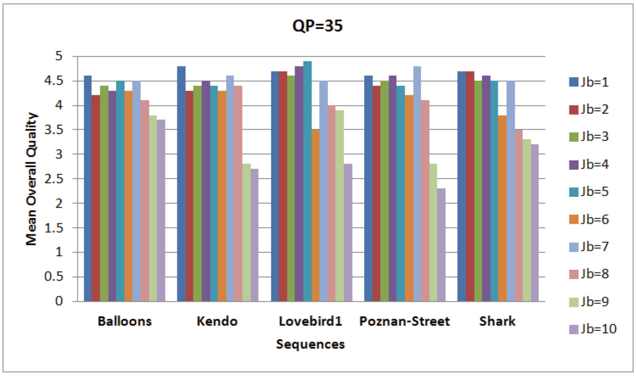
From Fig. 2, we find when $1 < Jb \leq 7$ stereoscopic video has perceptual distortion in 2D display but has not perceptual distortion in 3D-display compared to reference stereoscopic videos. Therefore, $Jb = 7$ is the maximum distortion, which is HVS can tolerate for two views.

3 Experimental Results

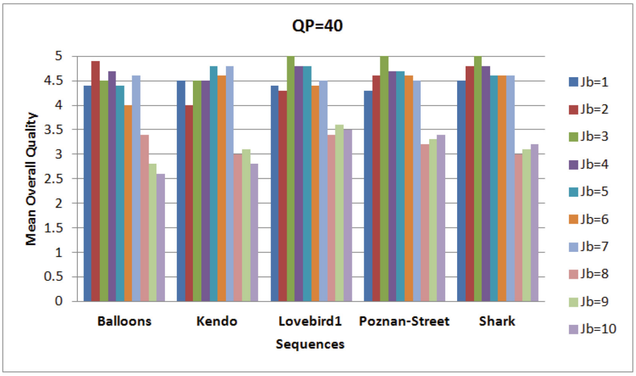
To evaluate the performance of the proposed algorithm, it is implemented on HTM12.1 of 3D-HEVC standard [2]. The software first codes center view0 (lift view), then side views1 (right view). The coding order is the following: T0, T1, T2 (in which Ti is the texture frames in the ith view). We mainly followed the common test condition (CTC) [13]. Most of the encoder configuration including QP setting was inherited from the CTC. A group of pictures (GOP) of 8 was considered with an Intra period of 24. The maximum coding unit depth was set to 4, and the maximum coding unit size was set to 64 × 64. For side view coding, disparity compensated prediction and multiview motion vector prediction were enabled. The following QP combinations for texture and depth respectively were considered: (30;39), (35;42), and (40;45). We have tested our algorithm on five sequences defined in the CTCs (1920 × 1088 and 1024 × 768) including Balloons, Kendo, Lovebird1, Poznan-Streets and Shark. The mean opinion score (MOS) scales for the DSIS protocol range from 1 to 5 for the quality from bad to excellent.



(a)



(b)



(c)

Fig. 2. 3D perceived visual quality evaluation for different *Jb* threshold values

In order to objectively assess how the proposed asymmetric stereoscopic video coding method affects bitrates saving performance, three schemes are designed as follows:

Scheme-I: Traditional 3D-HEVC video coding method in HTM

Scheme-II: The traditional perceptual symmetric 3D-HEVC video coding method in HTM, which is preprocessed by JND model - J_{st} for two views.

Scheme-III: The proposed perceptual asymmetric 3D-HEVC video coding method, which is preprocessed by JND model - J_{st} for left views and preprocessed by JND model - J_{3d} for right views.

Tables 1 and 2 show the PSNR, Bitrates and 3D DSIS scores of three methods. From Table 1 we can see that the proposed method present similar 3D perceived visual quality compared to other method. Suppose Scheme-I to be a benchmark, we can obtain the saving percentage with respect to Scheme-I. As shown in Table 3, Scheme-II contributes to average 11% and 19% bitrates savings for the whole views and dependent view, respectively. Scheme-III can achieve more bitrates savings ranging from 8% to 64% compared to Scheme-II for dependent view. It is because that Scheme-II improves coding efficiency by removing human visual spatial temporal redundancies. While Scheme-III further considers the vision binocular suppression masking effect. We also find that the proposed method can gain more bitrates savings when the left view have better video quality. Besides, we suggest that the modification of the spatial temporal JND model can slightly increase the efficiency of the asymmetric coding.

Table 1. Results comparison of Scheme-III, Scheme-II and Scheme-I in terms of PSNR and 3D DSIS

QP	Test seq.	Scheme-III			Scheme-II			Scheme-I		
		PSNR (dB)		3D MOS	PSNR (dB)		3D MOS	PSNR (dB)		3D MOS
		L	R		L	R		L	R	
30	Balloons	39.60	33.97	4.8	39.59	38.14	4.9	41.14	39.91	5
	Keno	39.57	33.81	5	39.57	38.42	4.7	41.77	40.66	5
	Lovebird1	38.32	32.55	4.7	38.32	35.85	5	38.85	36.84	5
	Poznan-Street	37.29	34.16	4.5	37.29	36.31	4.6	37.88	37.13	5
	Shark	37.05	34.73	4.8	37.05	36.57	5	38.40	37.89	5
35	Balloons	37.35	32.81	4.5	37.35	36.02	4.6	38.53	37.28	5
	Keno	37.55	32.86	4.6	37.55	36.49	4.8	39.25	38.17	4.5
	Lovebird1	35.76	31.46	4.5	35.76	33.50	4.7	36.11	34.27	4.7
	Poznan-Street	35.47	33.16	4.8	35.48	34.73	4.6	35.87	35.25	4.9
	Shark	34.57	33.08	4.5	34.57	34.23	4.7	35.51	35.15	5
40	Balloons	34.88	31.26	4.6	34.87	33.61	4.4	35.70	34.51	4.5
	Keno	35.31	31.59	4.8	35.31	34.27	4.5	36.55	35.54	4.6
	Lovebird1	33.15	30.03	4.5	33.15	31.34	4.4	33.78	31.81	4.8
	Poznan-Street	33.51	31.72	4.5	33.51	32.93	4.3	33.77	33.28	4.7
	Shark	32.22	31.27	4.6	32.22	31.96	4.5	32.86	32.60	4.8

Table 2. Results comparison of Scheme-III, Scheme-II and Scheme-I in terms of Bitrates

QP	Test seq.	Scheme-III		Scheme-II		Scheme-I	
		Bitrates (kbps)		Bitrates (kbps)		Bitrates (kbps)	
		L	R	L	R	L	R
30	Balloons	410.78	95.63	418.78	108.84	476.71	128.95
	Keno	360.70	74.87	360.70	87.08	420.79	110.88
	Lovebird1	567.80	116.63	567.80	152.04	602.07	199.54
	Poznan-Street	881.04	77.32	881.04	131.45	1013.2	202.64
	Shark	2034.7	150.97	2034.7	209.69	2335.9	269.96
35	Balloons	250.17	56.22	250.18	61.21	275.20	68.61
	Keno	217.43	44.76	217.44	49.86	245.70	59.38
	Lovebird1	303.44	57.90	303.44	68.99	313.24	84.85
	Poznan-Street	445.45	38.73	445.45	57.38	484.16	77.60
	Shark	1064.63	78.60	1064.6	102.6	1199.99	124.32
40	Balloons	155.60	34.10	155.60	33.81	165.91	36.97
	Keno	135.69	25.85	135.69	28.2	150.11	32.23
	Lovebird1	163.57	27.73	163.57	31.16	97.35	27.33
	Poznan-Street	238.63	20.15	238.63	27.05	252.19	34.85
	Shark	537.08	42.99	537.08	48.60	595.29	56.95

Table 3. Results comparison of proposed Scheme-III, Scheme-II to HTM encoder

QP	Test seq.	Scheme-II and Scheme-I		Scheme-III and Scheme-I	
		Save bitrates (%) for the whole views	Save bitrates (%) for the dependent views	Save bitrates (%) for the whole views	Save bitrates (%) for the dependent views
30	Balloons	13%	16%	15%	26%
	Keno	16%	22%	18%	32%
	Lovebird1	10%	24%	15%	42%
	Poznan-Street	17%	35%	21%	62%
	Shark	14%	22%	16%	43%
35	Balloons	9%	11%	11%	18%
	Keno	12%	16%	14%	25%
	Lovebird1	6%	19%	9%	32%
	Poznan-Street	10%	26%	14%	50%
	Shark	12%	17%	14%	37%
40	Balloons	7%	9%	7%	8%
	Keno	10%	13%	11%	20%
	Lovebird1	5%	16%	6%	25%
	Poznan-Street	7%	22%	10%	42%
	Shark	10%	15%	11%	25%
Ave.		11%	19%	13%	32%

4 Conclusion

A new perceptually asymmetric 3D video coding method based 3D-JND was presented in this paper. We first analyzed and presented the impact of utilizing JND coding within the framework of asymmetric stereoscopic video coding. Then, extensive subjective tests were evaluated to obtain asymmetric JND threshold considering different operating ranges and content types. Experimental results clearly demonstrated that the proposed algorithm achieves a significant bitrates savings than the conventional perceptually symmetric 3D video coding.

Acknowledgments. This work was supported by National Natural Science Foundation of China (61301113, 6167010863), Natural Science Foundation of Shanghai (13ZR1416500), and Open Project of Key Laboratory of Advanced Display and System Application.

References

1. Domanski, M., Grajek, T., Karwowski, D., Klimaszewski, K., Konieczny, J., Kurc, M., Luczak, A., Ratajczak, R., Siast, J., Stankiewicz, O., Stankowski, J., Wegner, K.: Multi-view HEVC—experimental results. Document JCT-VC (MPEG/VCEG), JCTVCG582, Geneva, Switzerland. Tech Rec, November 2011
2. Joint Collaborative Team on 3D Video (JCT-3V). https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-12.1/
3. Seuntiens, P., Meesters, L., Ijsselsteijn, W.: Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation. ACM Trans. Appl. Percept. **3**(2), 95–109 (2006)
4. Shao, F., Jiang, G., Wang, X., Yu, M., Chen, K.: Stereoscopic video coding with asymmetric luminance and chrominance qualities. IEEE Trans. Consum. Electron. **56**(4), 2460–2468 (2010)
5. Fehn, C., Kauff, P., Cho, S., Kwon, H., Hur, N., Kim, J.: Asymmetric coding of stereoscopic video for transmission over T-DMB. In: Proceedings of 3DTV-CON, Kos, Greece, May 2007
6. Stelmach, L., Tam, W.J., Meegan, D., Vincent, A.: Stereo image quality: effects of mixed spatio-temporal resolution. IEEE Trans. Circuits Syst. Video Technol. **10**(2), 188–193 (2000)
7. Aflaki, P., Hannuksela, M.M., Gabbouj, M.: Subjective quality assessment of asymmetric stereoscopic 3D video. Sig. Image Video Process. **9**, 331–345 (2013). ISSN 1863–1703
8. Aksay, A., Bilen, C., Kurutepe, E., Ozcelebi, T., Akar, G.B., Civanlar, M.R., Tekalp, A.M.: Temporal and spatial scaling for stereoscopic video compression. In: Proceedings of EUSIPCO, Florence, Italy, pp. 4–8 (2006)
9. Saygili, G., Gurler, C.G., Tekalp, A.M.: Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming. IEEE Trans. Broadcast. **57**(2), 593–601 (2011)
10. Ekmekcioglu, E., De Silva, V., Pesch, P.T., Kondoz, K.: Visual attention model aided non-uniform asymmetric coding of stereoscopic video. IEEE J. Sel. Top. Sig. Process. **8**(3), 402–414 (2014)
11. Jain, A.K., Robinson, A.E., Nguyen, T.Q.: Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding. IEEE Trans. Circ. Syst. Video Technol. **24**(3), 418–429 (2014)

12. Yang, X., Lin, W., Lu, Z., Lin, X., Rahardja, S., Ong, E.P., Yao, S.: Rate control for videophone using local perceptual cues. IEEE Trans. Circ. Syst. Video Technol. **15**(4), 496–507 (2005)
13. Muller, K., Vetro, A.: Common Test Conditions of 3DV Core Experiments. Document JCT3V-G1100, ITU-T SG 16 WP 3 and ISO/IEC JTC1/SC 29/WG 11, January 2014
14. ITU-R: Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union