SUBJECTIVE EVALUATION OF EFFECTS OF SPECTRAL AND SPATIAL REDUNDANCY REDUCTION ON STEREO IMAGES

Anil Aksay, Cagdas Bilen, Gozde Bozdagi Akar

Electrical and Electronics Department, Middle East Technical University
Ankara, 06531, Turkey
phone: +90 312 2104509, fax: +90 312 2101261, email: {anil, cbilen, bozdagi}@eee.metu.edu.tr

ABSTRACT

Human visual system is more sensitive to luminance than to chrominance. In order to reduce information that is not perceived by human visual system, color channels are downsampled while keeping luminance as original. Similarly in stereo case, human visual system uses high frequency information from the high resolution image of the mixed resolution image pair. By downsampling one of the pair, higher compression is achieved in stereo image coding. In this paper, we have examined downsampling color channels in higher ratios in color stereo image pairs. In our experiments, we have used "double-stimulus continuous-quality scale" (DSCQS) method. We have found out that the depth perception is not changed by compression or filtering. However, in order to keep perceived image quality similar to the original stereo pair, filtering should be applied to chrominance but not to luminance channels.

1. INTRODUCTION

Human Visual System (HVS) is more sensitive to luminance information than to chrominance (color) information. Thus, the full range of color information cannot be perceived by human visual system. Image compression systems use this to reduce spectral redundancy, by downsampling color channels. If image is in RGB color space, it has to be converted to YUV or YCrCb color space. Y channel holds the luminance information whereas UV (CrCb) channels hold color information. Downsampling color channels reduce information without any perceivable visual artifacts.

Another important point about HVS is the perception of the 3D world. HVS uses monocular and binocular cues in order to perceive 3D world. Binocular cues are the two slightly different images entering left and right eyes. Similar as in single image case, HVS does not process all the information supplied by the 3D world. If two images have different frequency characteristics, namely one image is high resolution whereas the other is low pass filtered, HVS uses the one with high resolution information. This is called binocular rivalry. By mixed resolution coding, [1] higher compression is achieved in stereo image coding.

In several experiments [2, 3] it is found out that coding distortions can be less perceivable in stereo image pairs than single images. If distortion results only from low-pass filtering one of the sequences, then perceived image quality is related with the high quality sequence whereas with MPEG-2 coding scheme, perceived image quality is related with the average of the qualities of right and left channel sequences. [4, 5]

In [6], the duration to observe 3D sequence and accuracy of depth perception is experimented for compressed and uncompressed sequences. The results show that compressed images are perceived similarly as uncompressed images.

ITU has recommendations for subjective assessment of stereoscopic television pictures (ITU-R BT.1438) [7] and for conventional monoscopic television (ITU-R BT.500-10) [8]. In order to evaluate image quality and depth perception, proposed assessment methods are used. These methods are described in Section II in more detail.

In most of the researches, low pass filtering is used for all color channels of image. In our research, we investigate how much we can filter chrominance channels without decreasing perceived image quality and depth perception. We also investigate asymmetric coding with different bitrates to minimize the distortion in perceived image quality.

In our tests, we have found out that the depth perception is not changed by compression or filtering. However, in order to keep perceived image quality similar to original stereo pair, filtering should be applied to chrominance but not to luminance channels. Similarly for asymmetric compression, differences between bitrates of right and left channels should be kept as low as possible.

2. ASSESSMENT OF STEREOSCOPIC IMAGE PAIRS

In [8], three methods are recommended for conventional monoscopic television: Double Stimulus Impairment Scale Method (DSIS or EBU), Double-Stimulus Continuous-Quality Scale Method (DSCQS), Single-Stimulus and

Stimulus-Comparison methods. In [7], DSCQS method is recommended for stereoscopic image pair assessment.

In DSIS method, both reference and distorted set are shown to the assessor. Assessor is expected to vote on the distorted set, while "keeping in mind the reference" [7].

In single-stimulus scaling, assessment is done on the image individually without any reference. In stimulus-comparison scaling, assessment is done by comparing different distortions against each other.

In DSIS method, single-stimulus and stimulus-comparison methods, assessor is expected to grade using a list of possible values described verbally (imperceptible, perceptible but not annoying, slightly annoying, annoying, very annoying). In DSCQS method grading is on a continuous scale.

3. DOUBLE-STIMULUS CONTINUOUS-QUALITY SCALE (DSCQS) METHOD

In this test method at least 15 assessors, which are chosen among non-experts and inexperienced assessors, should be used. The evaluation should be on a continuous scale such as in Figure 2.

The method can be applied in two variants:

Variant1: Each assessor is let to switch between two conditions, A and B (two stereoscopic images), one of which is always the source and the other is the tested condition applied on the source. The identity of the images, whether it is the source or the test condition, should be known by the experimenter but not by the assessors. After evaluating the conditions the assessor moves to the next pair of images.

Variant2: Multiple assessors are shown two conditions, A and B (two stereoscopic images), consecutively one of which is always the source and the other is the tested condition applied on the source. The identity of the images, whether it is the source or the test condition, should be known by the experimenter but not by the assessors. The next pair of conditions is shown after the assessors establish an opinion.

Analysis Method: For the analysis of the test results, each evaluation is graded between 0-100 and the difference between the scores of source image and the test condition is calculated to find the score of that test condition on that image by the assessor. After all these scores are calculated, the values are normalized to fit in 0-100. And as a final step, to find the scores of each algorithm (test condition) the average of all the scores over the assessors and images are taken. Scores of the algorithms can be compared with their closeness to the number to which zero score is mapped during the normalization process.

4. EXPERIMENTS

We have performed two sets of experiments for different distortions. Each experiment is explained in the following sections.

4.1. Experiment #1

In this experiment, we investigated the effects of downsampling in chrominance and luminance channels. We also investigated acceptance of JPEG and J2K coding at two bitrates. In order to meet requirements of assessment test, we use only 3 image sets with 9 algorithms.

Assessors: 16 assessors (4 female, 12 male with average age 22) with ages ranging from 20 to 25, volunteered to participate in the experiment. The participants were non-experts in the area of picture quality and were screened for color vision, stereo depth perception and visual acuity.

Equipment and Method: During the whole experiment eDimentional 3D glasses along with 19” Philips 109S Monitor are used to produce the necessary viewing conditions. The assessors were seated at a viewing distance of 2.5 H (H: height of tested images).

Each assessor is well informed on the test process and test materials (possible quality defects) before the test and they are assisted during the whole test procedure. DSQCS test method with the first variant mentioned above is used as the test methodology. At each step two image sequences, original left and right images and processed left and right images are used. We will call those 4 images an evaluation pair. In the experiments, original images are also repeated as a processed image in order to test the performance of the test.

At the beginning of the test, 5 random evaluation pairs are shown to the assessors and these 5 evaluation pairs are not evaluated since they provide stabilization of the perception of assessors. The test material is shown in a random order for each assessor. The randomization is done both among evaluation pairs and among the set of image sequences in the pair. In experiment 1, both image sets in the evaluation pair are marked with either a small red or green rectangle on the top left corner and the assessors are instructed to evaluate on the forms such as in Figure 2.

Test Material: As the test material, three different color source stereoscopic image pairs are used which are *tsukuba (384x288), sawtooth (434x380)* and *venus (434x383)*. 9 different distortions are applied on these images as shown in Table 1. As a result a total of 35 evaluation pairs, including first 5 stabilizing pairs, are shown to the assessors and it is assured that each test does not take more than 30 minutes.

The distortions in the Table 1 are as follows: x:y:z represents downsampling in Y, U and V channels. 1:4:4 means,

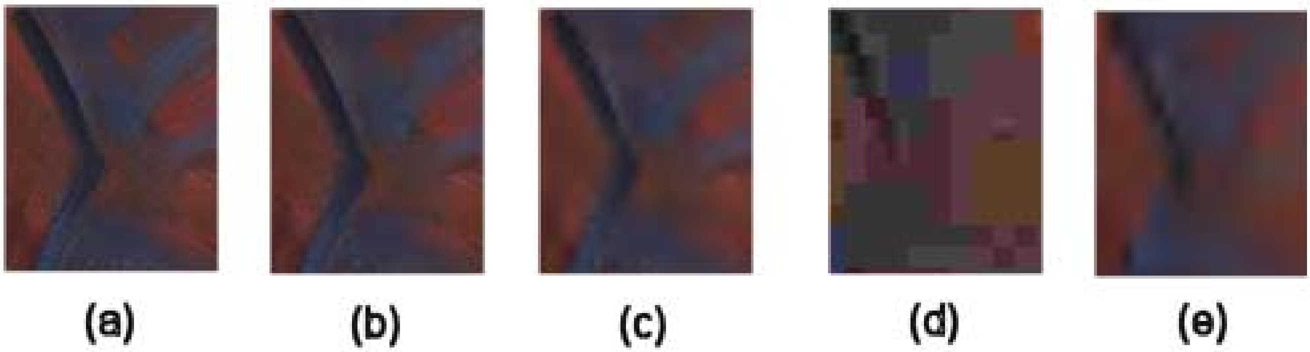

Figure 1 Portion of *sawtooth* image (a) original, (b) JPEG 0.05 bpp, (c) J2K 0.05 bpp, (d) JPEG 0.01 bpp, (e) J2K 0.01 bpp

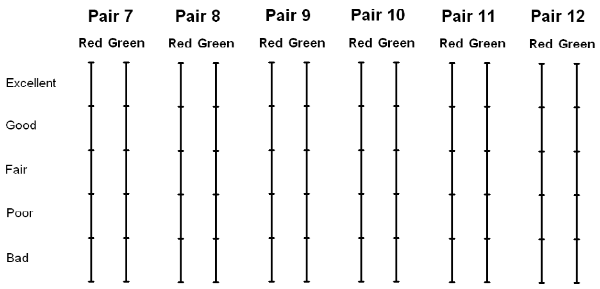

Figure 2 Assessment Form for Experiment#1

Y is kept original and U and V channels are downsampled by 4 in both directions, (1:2.667 compression). Downsampling is done by nearest-neighbour interpolation and upsampling is done by bilinear interpolation. JPEG stands for baseline JPEG compressor and J2K stands for JPEG2000 compressor with 6 resolution levels and integer mode using JasPer software [9].

PSNR values are all in dB and calculated according to the following formulas where D_l and D_r represent the distortions in right and left images [10].

$$PSNR_{all} = 10\log_{10}\frac{255^2}{(D_l + D_r)/2}$$

$$PSNR_l = 10\log_{10}\frac{255^2}{D_l/2}$$

$$PSNR_r = 10\log_{10}\frac{255^2}{D_r/2}$$

Table 1. PSNR (dB) and Mean Opinion Score (MOS) results for Test 1

Left Image	Right Image	MOS	PSNR (All)	PSNR (L)	PSNR (R)
original	original	15,42	INF	INF	INF
YUV 1:2:2	original	15,92	41,35	38,34	INF
JPEG 0.05 bpp	original	18,88	33,62	30,61	INF
J2K 0.05 bpp	original	19,63	36,67	33,66	INF
YUV 1:4:4	original	23,04	37,51	34,50	INF
YUV 1:5:5	original	24,42	36,56	33,55	INF
YUV 2:2:2	original	45,92	30,48	27,47	INF
J2K 0.01 bpp	original	48,33	29,09	26,08	INF
grayscale	original	48,38	26,30	23,29	INF
JPEG 0.01 bpp	original	57,79	27,21	24,20	INF

Due to the normalization, 0 (best quality) is mapped to 14, and the success of the algorithms can be measured by closeness of their mean to 14. As it can be seen, the mean of the original image is also not exactly 14, which is due to the misjudgement of the assessors and it is expected.

We concluded that JPEG coding performs better than J2K at higher bitrates because of binocular rivalry. Portion of sawtooth image with both JPEG and J2K coded at 0.05 and 0.01 bpp is shown in Figure 1.

4.2. Experiment #2

According to the results of Experiment #1, we decided not to downsample luminance channel. We have seen that we cannot downsample chrominance channel as much as we expected. We decided to increase the image sets with bigger image sizes for better assessment. To eliminate calculation errors, we implemented voting process in software as well. We also investigate the effect of adding distortion to different channel.

Assessors: 17 assessors (6 female, 11 male with average age 23) with ages ranging from 19 to 30, volunteered to participate in the experiment. The participants were non-experts in the area of picture quality and were screened for color vision, stereo depth perception and visual acuity.

Equipment and Method: The equipment and method used in this experiment are almost same with the experiment 1 except the evaluation process. In this experiment, the assessors are instructed to evaluate on software instead of forms as in the Figure 3.

Test Material: In this experiment, 6 color stereoscopic image pairs are used as source, which are *acrobats (642x480), mountain (512x488), teddy_bear (640x453), touring (1023x698), purple_car (1023x768)* and *truck (512x390)*. We applied 7 different distortions and results are shown in Table 2. A total of 53 evaluation pairs are shown to the assessors in their original resolution throughout the experiment and it is assured that each test does not take more than 30 minutes.

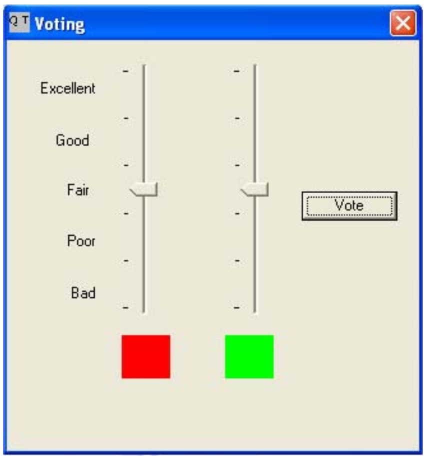

Figure 3 Assessment Dialog for Experiment#2

Table 2. PSNR (dB) and Mean Opinion Score results for Test 2

Left Image	Right Image	MOS	PSNR (All)	PSNR (L)	PSNR (R)
original	original	36,71	INF	INF	INF
JPEG 0.05 bpp	JPEG 0.1bpp	36,01	35,07	32,49	43,40
YUV 1:16:16	original	47,75	34,68	31,67	INF
YUV 1:16:16	YUV 1:8:8	48,53	32,95	31,67	34,83
YUV 1:8:8	YUV 1:16:16	49,24	33,06	34,61	31,95
JPEG 0.02 bpp	JPEG 0.02 bpp	53,44	27,01	27,07	27,00
JPEG 0.05 bpp	JPEG 0.01 bpp	63,88	26,02	32,49	23,59
JPEG 0.02 bpp	JPEG 0.01 bpp	66,81	24,97	27,07	23,59

In this experiment, 0 (best quality) is mapped to 36 and original sequence score is 36,71.

5. CONCLUSIONS AND FUTURE WORK

By comparing the results in Table 1 and Table 2, we can conclude that by downsampling Y channel, perceived image quality decreases significantly. In order to keep perceived image quality high, we need to keep Y channel original and downsample U and V channels. We can see that downsampling ratio depends on the image size and image content as well.

We have also tested the same distortion applied to either right or left image. Both PSNR values and Mean Opinion Scores (MOS) are similar to each other in both cases.

MOS and PSNR values mostly correlate with each other. On higher bitrate, JPEG performs better than J2K even though its PSNR value is lower. This is mostly due to binocular rivalry. By comparing Figure 1 (b) and (c), we can see that JPEG preserves edges but introduces more distortion to mostly uniform areas, whereas J2K removes high frequency component on the whole image. Binocular rivalry eliminates more distorted areas easier than the distorted whole image.

In future, we will try to incorporate our findings into stereo video coding. We will also try to combine chrominance downsampling with image compressors and disparity compensated coding techniques.

6. ACKNOWLEDGMENTS

This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

We would also like to thank Mehmet Oğuz Bici for his help in developing quality assessment software.

REFERENCES

[1] M. G. Perkins, "Data compression of stereo pairs," IEEE Trans. Comm., vol. 40, pp. 684-696, Apr. 1992.
[2] J. Y. Chen, Z. Liwei, and S. Q. Ding, "The effect of JPEG Coding Scheme on the perceived quality of 3D im-ages", SID symposium, Vol. 29, pp. 1211-1214, 1998.
[3] A. Schertz, “Source coding of stereoscopic television pictures,” in Proc. IEE Inter. Conf. Image Processing and its Applications, Maastricht, The Netherlands, Apr. 7–9, 1992, pp. 462–464.
[4] L. B. Stelmach, W. J. Tam, and D. V. Meegan, "Stereo Image Quality: Effects of Spatio-Temporal Resolution", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, pp. 188-193, 2000.
[5] L. B. Stelmach, W. J. Tam, D. V. Meegan, A. Vincent and P. Corriveau, “Human Perception of Mismatched Stereoscopic 3D Inputs", IEEE Signal Processing Society, 2000 International Conference on Image Processing, Van-couver, Canada, September 10-13th, 2000
[6] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov, and A. Henik, ‘‘On stereo image coding,’’ in Proc. Int. Conf. on Pattern Recognition, Vol. 1, pp.357–359 Nov. 1988.
[7] “Subjective Assessment of Stereoscopic Television Pictures,” ITU, Recommendation BT.1438, 2000.
[8] “Methodology for the Subjective Assessment of the Quality of Television Pictures,” ITU, Recommendation BT.500–10, 2000.
[9] M. D. Adams and F. Kossentini, “JasPer: A software-based JPEG-2000 codec implementation,” In Proc. of IEEE International Conference on Image Processing, Van-couver, BC, Canada, October 2000.
[10] N. V. Boulgouris and M. G. Strintzis, “A family of wavelet-based stereo image coders", IEEE Trans. on CSVT, Vol. 12, No. 10, pp.898-903, October 2002.