IMPACT OF ACQUISITION DISTORTIONS ON THE QUALITY OF STEREOSCOPIC IMAGES

Lutz Goldmann, Francesca De Simone and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

While objective and subjective quality assessment of 2D images and video has been an active research topic in the recent years, emerging 3D technologies require new quality metrics and methodologies taking into account the fundamental differences in the human visual perception and typical distortions of stereoscopic content. Therefore, this paper presents a comprehensive stereoscopic image database that contains a large variety of scenes captured using a stereoscopic camera setup consisting of two HD camcorder with different capture parameters. In addition to the images, the database also provides subjective quality scores obtained using an adapted single stimulus continuous quality scale (SSCQS) method. The resulting mean opinion scores can be used to evaluate the performance of visual quality metrics as well as for the comparison and for the design of new metrics.

1. INTRODUCTION

The introduction of three dimensional television (3DTV) on the public consumer market is believed to be just a matter of time and has been compared to the transition from black-and-white to color TV. To be a success, both visual quality and comfort must at least be comparable to conventional standards to guarantee a strain free viewing experience. Since 3DTV involves both 2D and 3D visual perception new distortions have to be considered beside the classical ones.

The quality of 3D content may be affected by artifacts arising from each of the stages within a typical 3D processing chain [1]: While the creation and restitution stages have a direct influence on the 3D perception, the other stages may also introduce artifacts leading to a decrease in perceived 3D quality.

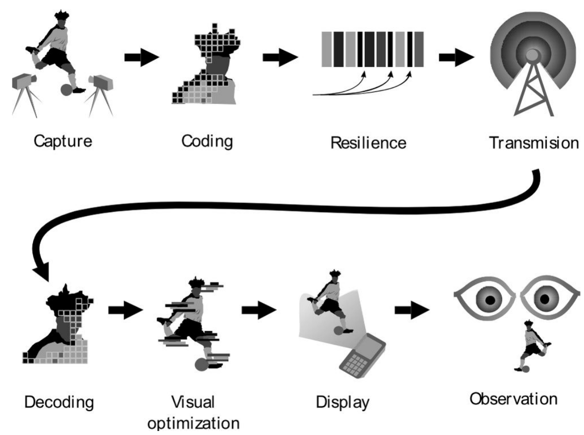

Fig. 1: Overview of a typical 3D processing chain from the acquisition to the restitution of a 3D scene [1].

Several parameters have an influence on the 3D effect of the captured stereoscopic video including the interocular distance between the camera lenses, the angle between the cameras which can be either parallel or converged and the distance of the real scene. In order to provide the observer with an optimal 3D effect it is important to understand the influence of these parameters on the quality of the perceived 3D content.

While the effects of the different acquisition parameters on the 3D perception have been studied and described before, their influence on the perceived quality for a human observer has not been assessed quantitatively. Furthermore, there is a lack of comprehensive stereoscopic image and video datasets that support the development and evaluation of 3D processing techniques. The objective of this work is to develop a comprehensive 3D image database with associated subjective quality scores comparable to the LIVE Image Quality Database[1], the Tampere Image Database (TID)

The research leading to these results has been performed in the frameworks of European Community's Seventh Framework Program (FP7/2007-2011) under grant agreement no. 216444 (PetaMedia) and Swiss National Centre of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM2).

[1] `http://live.ece.utexas.edu/research/quality/subjective.htm`

2008[2], and the EPFL Image Quality Database[3] for 2D images. Therefore, we have created a stereoscopic image dataset with various contents captured with different acquisition parameters. Given that dataset, we have conducted extensive subjective tests to study the influence of the acquisition parameters on the perceived 3D quality.

The paper is organized as follows. Section 2 discusses the novel stereoscopic image database including the initial requirements, the acquisition and the postprocessing of the images, and provides a description of the individual scene characteristics. Section 3 describes the goal and procedure of the subjective quality evaluation. Section 4 discusses the statistical analysis of the subjective quality scores and analyzes the results of the subjective quality test. Finally, section 5 summarizes the major findings and proposes directions for the use and extension of the proposed database.

2. DATABASE

2.1. Requirements

In order to make a 3D image database suitable for the development and evaluation of a complete 3D processing chain and a wide variety of applications, it has to fullfil several requirements. The following list of requirements is inspired by the recent calls of MPEG for FTV[2] test material:

- The cameras shall be in a 1D parallel arrangement.
- The type of cameras should be reported with data sheets and intrinsic information if possible.
- The temporal synchronization of the multiple cameras should be as accurate as possible.
- The minimum spatial resolution should be 640x480 pixels. However, higher spatial resolutions to simulate current and future media standards.
- Color consistency between the multiple views is very important and the cameras should be white balanced.
- The content should be representative and challenging with variations in texture, scene (depth) structures and lighting conditions.

As it will be shown below, most of the requirements are addressed by the proposed database.

2.2. Acquisition

For acquiring high quality stereoscopic images and video the following aspects have to be considered[3]:

[2] `http://www.ponomarenko.info/tid2008.htm`
[3] `http://mmspg.epfl.ch/iqa`

Fig. 2: HD stereo camera setup consisting of two identical camcorders and an adjustable stereo mount.

Matching cameras: Two identical cameras with the same firmware and settings should be used for complete interchangeability.

Matching optics: Both camera have to be optically matched in focal length and focus point. Since zoom lenses may differ only the extreme ends of the zoom range should be used.

Matching geometry: Both cameras have to be in perfectly controlled relative position to each other. From all possible translations and rotations, only translation along the horizontal axis and rotation along the vertical axis are allowed.

Matching photography: Full manual mode of the cameras should be used to ensure matching in white balance, sensitivity, shutter speed, aperture, gain, and framerate.

Synchronization: Accurate synchronization between the two cameras is essential and can be achieved by using a single remote control or electronically synchronized cameras.

Considering the different aspects mentioned above, we have built the stereo camera setup shown in figure 2, which consists of two identical HD camcorders (Canon HG-20) and an adjustable stereo mount. The mount ensures that optical axes of the cameras are parallel and supports the continuous adjustment of the camera distance in the range 7–50 cm. To ensure matching of the focal length the wide angle end of the zoom lens with a focal length of 43 mm has been used. In order to match the cameras with each other the focal length, white balance and shutter speed have been set manually. The synchronized operation of the two camcorder is ensured through the use of a single remote control. The camcorders support the capture of images with a resolution of 1920x1080 pixels and store them as high quality JPEG files.

2.3. Postprocessing

Because no 3D camera setup is perfect, a postprocessing step is needed to correct small calibration errors. This includes geometrical and color adjustment [3].

2.3.1. *Spatial alignment*

In a stereoscopic camera setup spatial distortions may be caused within the individual cameras (e.g. barrel/ pincushion distortion) or by the camera setup and calibration (e.g. relative positions). The spatial alignment focuses on the correction of the inter-camera distortions which are specific for the parallel camera setup.

The goal of the spatial alignment is to compensate small vertical disparities caused by the camera setup and adjust the depth position to avoid stereo window violations. This is achieved by applying a relative vertical and horizontal translation between the video pairs based on point correspondences. For a reliable adjustment of the depth position the control points for the nearest object are manually selected.

2.3.2. *Color adjustment*

Even with a manual control of white balance and exposure, luminance and chrominance components may vary globally between the different views. These discrepancies may originate from the use of heterogeneous cameras, calibration errors and appearance changes due to the different viewing angles. The goal of color adjustment step is to correct these color differences between the two stereo images.

Histogram matching as proposed by Hekstra et al. [4] is used to adapt the right camera view to the left camera view. This method assumes that a good fit of the distorted image to the reference image can be obtained by adapting the cumulative histogram. It makes no assumptions about the type of distortion such as brightness or contrast variations and considers nonlinear mappings. Since the correction is applied to the entire image, it is especially useful to correct global luminance and chrominance differences.

2.4. Description

The proposed database contains stereoscopic images with a resolution of 1920x1080 pixels. Various indoor and outdoor scenes with a large variety of colors, textures, and depth structures have been captured. Each of the scenes has been captured with different camera distances in the range 10–50 cm. Since the acquisition was done in a sequential way the content of a single scene may vary slightly across the different camera distance. However, the general 2D (color, texture, motion) and 3D (depth) characteristics are preserved.

Fig. 3: Visual samples of the different scenes considered within the database. From top left to bottom right: sofa, tables, sculpture, trees, moped, grass, bikes, monument, closeup, construction.

A subset of 10 scenes, shown in figure 3, with different characteristics has been selected for the subjective quality evaluation within this paper.

Table 1 provides an overview of the selected scenes together with the 3D characteristics such as near distance n and far distance l, and the maximum permissible camera distance b. The latter can be theoretically computed based on a simplified Bercovitz equation [5] as

$$b = \frac{p}{f} \cdot \frac{l \cdot n}{l - n} \tag{1}$$

with the focal length $f = 43\,\mathrm{mm}$ and a typical parallax limit $p = 1.2\,\mathrm{mm}$. The simplification of the Bercovitz equation is valid for $f \ll n$, which is true for all the scenes in the database.

3. SUBJECTIVE QUALITY ASSESSMENT

3.1. Equipment

The subjective test campaign was conducted at the Multimedia Signal Processing Group (MMSPG) quality test lab-

Table 1: Overview of the different scenes and their characteristics.

Id	Title	Near (m)	Far (m)	Distance (cm)
3	sofa	4	12	17
4	tables	4	15	15
5	sculpture	6	50	19
6	trees	6	25	22
7	moped	5	80	15
8	grass	4	80	12
9	bikes	4	12	17
10	monument	9	100	28
11	closeup	5	30	17
12	construction	7	100	21

Fig. 4: Quality test laboratory and 46" polarized 3D display used for the subjective quality evaluation.

oratory at EPFL (shown in figure 4), which is compliant with the recommendations for subjective evaluation of visual data issued by ITU-R [6]. The laboratory is setup in a way to assure the reproducibility of results by avoiding involuntary influence of external factors.

A 46" polarized stereoscopic display (Hyundai S465D) with a native resolution of 1920x1080 pixels has been used to display the test stimuli. The monitor was calibrated using an EyeOne Display2 color calibration device according to the following profile: sRGB Gamut, D65 white point, 120 cd/m² brightness and minimum black level. The room is further equipped with a controlled lighting system that consists of neon lamps with 6500 K color temperature. The illumination level measured on the screen is 30 lux and the ambient black level is 0.5 cd/m².

The experiments involved only one subject per session assessing the test material. The subject was seated in line with the center of the monitor, at a distance of approximatively 2 m which is equal to the height of the screen multiplied by factor 3 as suggested in the ITU-R BT.1438[7] for HDTV.

3.2. Observers

Twenty subjects (6 female, 14 male) participated in the test. All of them were non-expert viewers with a marginal experience of 3D image and video viewing. The age distribution ranged from 24 to 37 with an average of 27.

All the subjects were screened for visual acuity using the Snellen chart and color vision using the Ishihara test. Furthermore, the stereo vision was tested according to the ITU-R BT.1438 recommendation [7].

3.3. Stimuli

For the subjective evaluation, the stereoscopic image database described in section 2 has been split into a training set with 1 scene (grass) and a testing set with 9 scenes (sofa, tables, sculpture, trees, moped, grass, bikes, monument, closeup, construction). For each of the scenes 6 different stimuli have been considered corresponding to different camera distances (10, 20, 30, 40, 50, 60 cm).

3.4. Procedure

The subjective evaluation of 2D visual quality according to standardized methods has a long history. Several recommendations have been issued by the International Telecommunication Union (ITU) including the widely used ITU-R BT.500[6]. It describes methods for the subjective quality assessment of standard definition television (SDTV) pictures. The most prominent methods are the double stimulus continuous quality scale (DSCQS), the double stimulus impairment scale (DSIS) and the single stimulus (SS) method. While DSCQS and DSIS can be used for a direct comparison of impaired and unimpaired stimuli, SS is suitable to assess the quality without a reference. With respect to the subjective quality evaluation of 3-dimensional TV (3DTV) the same methods are recommended in ITU-R BT.1438[7].

Since the optimal acquisition settings for 3D content may vary depending on the scene, the display and the observer, it is difficult to select one of the stimuli as a reference. Therefore, a single stimulus (SS) method has been adopted for the subjective quality evaluation. In order to determine the influence of the camera distance on the 3D quality a continuous quality scale with 5 levels (excellent, good, fair, poor, bad), as described in ITU-R BT.500[6], has been used.

During the training session the subjective test methodology was introduced to the subjects and the range of quality levels was explained through a set of training stimuli. The stimuli have been selected by an expert viewer in such a way that each quality level is represented by an example and that the full range of quality levels within the set of test stimuli is covered. The training stimuli were presented in the same way as the test material to familiarize the subjects with the methodology.

During the testing session the subjects evaluated the quality of the 54 test stimuli in random order together with 5 dummy stimuli to stabilize the subjective evaluation at the beginning of the session. Each stimulus was shown once with a duration chosen by the subjects and a break between the stimuli, during which the subjects provided their scores.

4. ANALYSIS AND RESULTS

4.1. Outlier detection

The screening of subjects was performed according to the guidelines described in ITU-R BT.500 [6].

First, for each stimulus, it is tested whether the distribution of scores across subjects is normal or not. This is done by calculating the kurtosis coefficient of the distribution: if the coefficient is between 2 and 4, the distribution is assumed to be normal. Then, the score of each observer is compared with an upper and a lower threshold computed as the mean value plus and minus the standard deviation associated to that stimulus (times two, if normal, or times 20, if non-normal). For each subject, every time his/her score is found above the upper threshold a counter, P_i, is incremented. Similarly, every time his/her score is found below the lower threshold, a counter, Q_i, is incremented. Finally, the following two ratios are calculated: $P_i + Q_i$ divided by the total number of scores from each subject for the whole session, and $P_i - Q_i$ divided by $P_i + Q_i$ as an absolute value. If the first ratio is greater than 5% and the second ratio is less than 30%, then subject i is an outlier and all his/her scores are discarded.

Using the outlier detection described above, none of the 17 subjects have been discarded as an outlier. Thus the statistical analysis is based on the scores from 17 subjects.

4.2. Score computation

After the outlier removal, the mean opinion score is computed for each test condition j as:

$$MOS_j = \frac{\sum_{i=1}^{N} s_{ij}}{N} \tag{2}$$

where N is the number of valid subjects and s_{ij} is the score by subject i for the test condition j.

The relationship between the estimated mean values based on a sample of the population (i.e. the subjects who took part in our experiments) and the true mean values of the entire population is given by the confidence interval of the estimated mean. Due to the small number of subjects, the $100 \times (1 - \alpha)\%$ confidence intervals (CI) for mean opinion scores are computed using the Student's t-distribution, as follows:

$$CI_j = t(1 - \alpha/2, N) \cdot \frac{\sigma_j}{\sqrt{N}} \tag{3}$$

where $t(1 - \alpha/2, N)$ is the t-value corresponding to a two-tailed t-Student distribution with $N - 1$ degrees of freedom and a desired significance level α (equal to 1-degree of confidence). N corresponds to the number of subjects after outlier detection, and σ_j is the standard deviation of a single test condition across the subjects. The interpretation of a confidence interval is that if the same test is repeated for a large number of times, using each time a random sample of the population, and a confidence interval is constructed every time, then $100 \times (1-\alpha)\%$ of these intervals will contain the true value. We computed our confidence intervals for an α equal to 0.05, which corresponds to a significance level of 95%.

Figure 5 plots the mean opinion scores and confidence intervals vs. the camera distance for each of the individual scenes. Due to the page limitations only 6 representative scenes are shown. The maximum theoretical camera distances from table 1 are also shown as a dashed vertical line. In general the small size of the confidence intervals shows that the complexity of the subjective evaluation tasks was appropriate and that the ratings are quite consistent across the subjects. However it is interesting to analyze the size of the confidence intervals more closely. For most scenes the confidence intervals for the high and low quality levels are smaller than for the middle quality levels. Discussions with the subjects after the test lead to the conclusion that it is very hard to distinguish between intermediate quality level. As expected the influence of the camera distance on the quality of the stereoscopic quality is largely scene-dependent. While for some scenes (sofa, tables, moped, bikes) the 3D quality decreases rapidly with increasing camera distances, it decreases more slowly for other scenes (sculpture, construction). A closer look at the individual curves reveals three different groups: the quality of the first group (sculpture, construction) is between "fair" and "excellent" for most of the considered camera distances, the second group (sofa, bikes, moped) covers the quality range between "poor" and "good", the third group (tables) is mostly below "good". Although a large range of camera distances has been considered, not all the quality levels are equally covered. Especially the quality levels at both ends of the scale are only reached for a few scenes (excellent for sculpture, construction, and bad for sofa, tables). This shows that for non-expert viewers it is quite difficult to distinguish 5 quality levels (excellent, good, fair, poor, bad) and that 3 quality levels (good, fair, bad) may be more appropriate. The corresponding ranges are shown through the dashed horizontal lines.

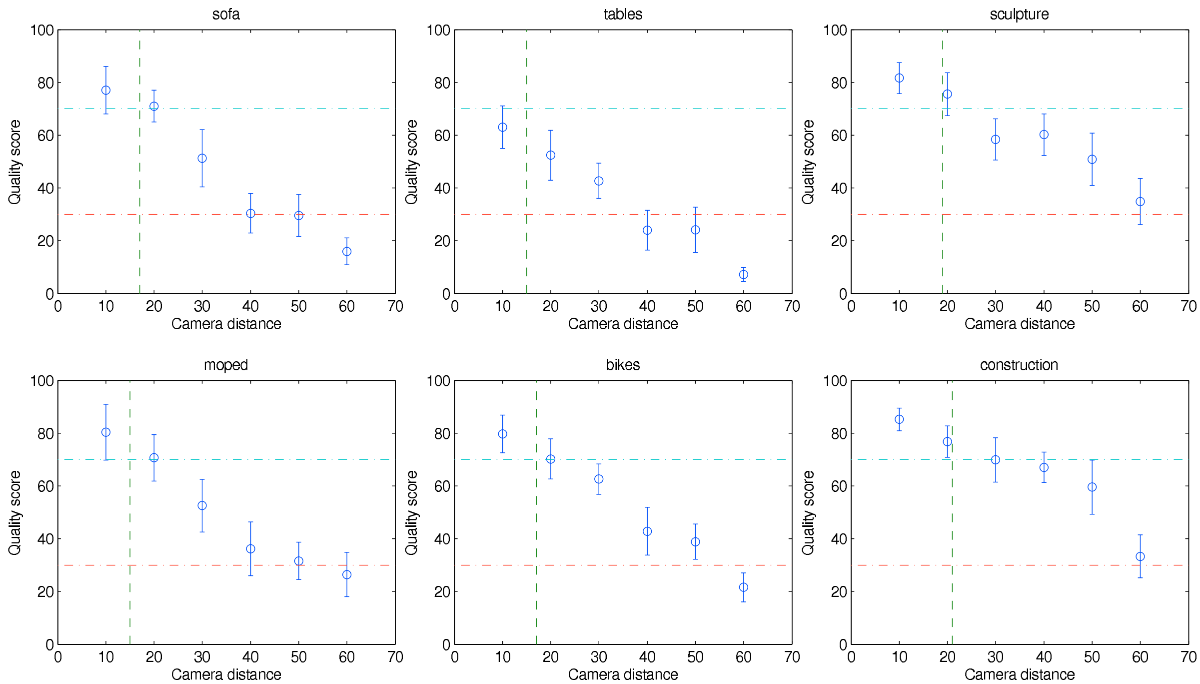

Fig. 5: Mean opinion scores and confidence intervals vs. camera distance for a subset of the scenes.

5. CONCLUSION

The success of 3DTV will largely depend on the improved quality of experience for the consumer. Since this type of media involves a combination of 2D and 3D perception it is important to understand the influence of whole 3D processing chain and the introduced artifacts on the subjective quality. As for 2D images a comprehensive database with associated subjective quality scores is crucial for the optimization of individual processing steps and the development of objective quality metrics. Since already the acquisition of stereoscopic images may have a large influence on the perceived 3D quality a novel database of high definition stereoscopic images captured with different camera distances has been created. A subjective test has been conducted to quantitatively assess the influence of the camera distance on the perceived quality. It is shown that depending on the scene structure and content, subjects are not very sensitive to camera distance changes and that only a few levels of 3D quality can be distinguished. The images and corresponding subjective quality scores are publicly available[4].

The database will be extended with more scenes to make it more comprehensible. Furthermore, new distortions related to the remaining processing steps such as coding, transmission, and restitution will be added. Based on that, it will be possible to compare the influence of different distortions and processing steps on the perceived quality. The resulting subjective scores will be used for the development and evaluation of objective quality metrics for 3DTV.

[4] `http://mmspg.epfl.ch/3diqa`

6. REFERENCES

[1] A. Gotchev, S. Jumisko-Pyykkö, A. Boev, and D. Strohmeier, "Mobile 3DTV System: Quality and User Perspective," *Proceedings of EUMOB*, 2008.

[2] MPEG, "Call for contributions on ftv test material," Tech. Rep. N9468, ISO/IEC JTC1/SC29/WG11, Shenzhen, China, Oct 2007.

[3] Bernard Mendiburu, *3D Movie Making*, Focal Press, 2009.

[4] AP Hekstra, JG Beerends, D. Ledermann, FE De Caluwe, S. Kohler, RH Koenen, S. Rihs, M. Ehrsam, and D. Schlauss, "PVQM – A perceptual video quality measure," *Signal processing: Image communication*, vol. 17, no. 10, pp. 781–798, 2002.

[5] J. Bercovitz, "Image-side perspective and stereoscopy," *Proceedings of the SPIE*, vol. 3295, 1998.

[6] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. BT.500-11, ITU-R, 2002.

[7] ITU-R, "Subjective assessment of stereoscopic television pictures," Tech. Rep. BT.1438, 2000.