
http://dx.doi.org/10.1117/12.838975

© Copyright 2010 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.
Multiview image coding scheme transformations: artifact characteristics and effects on perceived 3D quality

Roger Olsson and Mårten Sjöström
Dept. of Information Technology and Media, Mid Sweden Univ., SE-851 70 Sundsvall, Sweden

ABSTRACT

Autostereoscopic multiview 3D displays have been available for a number of years, capable of producing a perception of depth in a 3D image without requiring user-worn glasses. Different approaches to compress these 3D images exist. Two compression schemes, and how they affect the 3D image with respect to induced distortion, is investigated in this paper: JPEG 2000 and H.264/AVC. The investigation is conducted in three parts: objective measurement, qualitative subjective evaluation, and a quantitative user test. The objective measurement shows that the Rate-Distortion (RD) characteristic of the two compression schemes differ in character as well as in level of PSNR. The qualitative evaluation is performed at bitrates where the two schemes have the same RD fraction and a number of distortion characteristics are found to be significantly different. However, the quantitative evaluation, performed using 14 non-expert viewers, indicates that the different distortion types do not significantly contribute to the overall perceived 3D quality. The used bitrate, and the content of the original 3D image, is the two factors that most significantly affect the perceived 3D image quality. In addition, the evaluation results suggest that viewers prefer less apparent depth and motion parallax when being exposed to compressed 3D images on an autostereoscopic multiview display.

Keywords: 3D image, compression, quality, subjective test, autostereoscopic, multiview, JPEG 2000, H.264/AVC

1. INTRODUCTION

A fundamental property that must be fulfilled in order to provide the perception of depth in a 3D image is that the user perceives two different views of the presented 3D scene in the left and right eye simultaneously, so called binocular parallax. Autostereoscopic 3D display systems provide this 3D image property without requiring the user to wear special spectacles. Displays that can present a number of view pairs concurrently into different directions of the viewing space are, in addition, capable of providing another fundamental 3D property: motion parallax or "look-around capability". That is, the user can observe the scene from different perspectives merely by changing viewing position, just like when watching the scene in real life.

Many aspects contribute in providing the perceived quality of a 3D image beside the two types of parallax mentioned. Important factors are for example how the image is captured; with what compression scheme it is later compressed, and on what type of display system it is finally presented. The work presented here focus on investigating a set of compression scheme classes, the distortion characteristics they impose on a 3D image, and how the user perceives this distortion.

The paper is organized as follows. Section 2 sets the context in which a comprehensive picture is provided by the 3D image format used, the compression methods that may be applied to these 3D images and how the compression induced distortion may be quantified. Section 3 delimits the work by describing the user group that was assembled and the environmental assessment that was used. We also present the 3D images evaluated and how the selected compression schemes were parameterized for the evaluation. In Section 4 the results from the evaluation is presented, both in terms of objective quality as well as quantitative and qualitative subjective quality. Finally, Section 5 concludes the results.

Further author information: (Send correspondence to Roger Olsson)
Roger Olsson: E-mail: Roger.Olsson@miun.se, Telephone: + (46) 60 14 86 98
2. BACKGROUND

A number of autostereoscopic 3D display technologies exist that present a 3D image to one or several observers. The work presented in this paper focuses on the subset of displays that are built around common 2D pixel panels (e.g. LCD or Plasma) in combination with an adjacent lenticular lens array or barrier slit sheet.

The perceived 3D image is produced from a 2D pixel map presented on the pixel panel. Specific constraints are imposed on the pixel map facilitating the transformation of the 2D pixel map content into a perceived 3D image by the lens array. For the images used in this work a periodic structure is imposed stemming from the spatial multiplex of 25 different two-dimensional views. This high correlation between neighboring views results in the periodic structure exhibiting a high spatial redundancy corresponding.

Different compression approaches exist to compress this kind of multiview 3D images, ranging from pixel based methods to algorithms utilizing approximations of the scene geometry to increase the compression efficiency [1]. Regardless of approach an image compression scheme must be able to address this spatial redundancy in order to be efficient. In this work two different compression approaches that fulfill this requirement are used: wavelet transformation and DCT/DPCM-based hybrid coding.

Wavelet transformation is commonly used to reveal redundancy in 2D images, especially within the JPEG 2000 image compression standard. It is designed to efficiently handle 2D images by localizing the compression-induced distortion to high frequency components, which is less apparent to the human eye. However, the periodic structure of the 3D image produces high energy in high spatial frequencies. Hence, the potential quality reduction for JPEG 2000 compressed 3D images is primarily due to the compression affecting these high frequencies, breaking the interlaced property and causing interview cross talk and depth reduction.

The combination of discrete cosine transform (DCT) and differential pulse code modulation (DPCM) is used in most of today’s video compression standards. Considering that the 3D images consist of a sequence of 25 view images, the video compression approach is commonly adopted [1, 2]. The view sequence is then a function of space rather than time; hereafter denoted a pseudo video sequence (PVS) to indicate the difference. The current state-of-the-art video compression standard is H.264/AVC and was therefore selected as the second compression scheme to be evaluated. The potential quality reduction of H.264/AVC when applied to a PVS is primarily localized to large disparities, and sharp edges and high frequency content within each view.

The optical sub-system of the display is often disregarded in quality evaluation of 3D images corrupted by e.g. compression-induced distortion. Applying the Peak Signal to Noise Ratio (PSNR) to the 2D pixel map results in a global quality metric that produces a single scalar quality value enclosing the whole 3D image. The PSNR, measured in dB, for full color RGB-images is defined as

\[
PSNR(X, \hat{X}) = 20 \cdot \log_{10} \left( \frac{255}{\sqrt{MSE}} \right),
\]

where

\[
MSE = \frac{1}{3 \cdot M \cdot N} \left\| X - \hat{X} \right\|_F^2
\]

where \( \left\| X - \hat{X} \right\|_F \) denotes the Frobenius norm of the difference between matrix \( X \) and \( \hat{X} \), which are full color images of size \( M \times N \) pixels. The metric commonly assumes that \( \hat{X} \) has undergone some kind of distortion-inducing operation making it different from the original \( X \), which in the case of this work corresponds to lossy compression.

3. EVALUATION METHOD

The quality aspects of 3D images were in this work evaluated using objective quality metrics; by a single expert conducting a qualitative evaluation; and with an observer group quantitatively assessing the quality in a subjective test. In the following subsections we will describe the qualitative evaluation criteria and how the subjective test was conducted with respect to observers, test environment, compression schemes, 3D images, and test procedure.
3.1 Qualitative evaluation criteria
A pre-trial was conducted where strongly compressed images was examined for distortion. This test reveled a number of distortion characteristics that was grouped into the following types:

**Pixelation**  pixels that constitute each view became visible

**Ringing**  noise manifested around edges and sharp object boundaries, similar to the mosquito noise common in 2D image compression

**Blur**  loss of spatial detail within each views

**Depth**  reduction in the amount of perceived depth (binocular parallax)

**Perspective**  reduction in the amount of perspective change (motion parallax)

**Distribution**  the depth distribution of most of the distortion

These distortion types were used in the qualitative evaluation to describe the type of artifacts manifesting in the compressed images.

3.2 Observers
The majority of the 14 observers participating in the test were neither experienced in the field of image compression nor knowledgeable about the workings of autostereoscopic displays and 3D images. The average age of the observers was 40 years. Prior to the test each observer was presented with the background and aim of the test: evaluate the overall perceived 3D quality of 3D images compressed with different compression schemes. Out of the 14 observers, 3 was identified as having reduced visual acuity ($\leq 0.8$), reduced color vision, reduced stereo vision ($\geq 2$ arcmin) or a combination of the above. These observers where still allowed to conduct the test but marked for the statistical evaluation.

3.3 Environment
The test room was set-up according to the home environment defined in ITU-BT Rec. 500.11 [3]. A 42” autostereoscopic 3D display from LG Electronics (Flatron M4210D) was used as the display system on which the 3D images were presented. The display was chosen for its interface that allows full access to R, G, and B subpixels, which thereby gives complete control over the display’s 3D image presentation properties. In addition, the display has the potential to provide a significant amount of horizontal motion-parallax, or so called "look-around effect". Its lenticular array optically demultiplexes the full high definition 2D pixel map into a set of 25 independent views, each view with a resolution of 384x216 pixels. Other similar studies have used a similar display from Philips, which has the restriction of not allowing subpixel access and more views than 9 [4, 5].

The optimal viewing distance of the display is defined by the manufacturer to be four meters and the observers were consequently positioned at approximately this distance, see Figure 1 (a). At this distance they where also instructed on how to change their viewing position laterally in order to perceive any of the 24 different perspectives of the scene inherent in the 3D image, and the horizontal motion parallax this provides. Figure 1 (b) shows an observer conducting an evaluation.

3.4 3D images
Three 3D images were used in the subjective test: Fruit, Bottles, and Bike. The full 2D resolution of the underlying pixel maps is 1920x1080 pixels, corresponding to the 3D display’s pixel panel resolution. The 3D images differ relative to each other for example with regards to inherent motion parallax. Bike and Bottles exhibit a high degree of motion parallax whereas the perspective change of Fruit is relatively small when the viewpoint of the observer is altered.
3.5 Test session procedure

The test session was conducted according to the Double Stimulus Continuous Quality-Scale (DSCQS) method also defined in Rec. ITU-R BT.500 [3]. Each observer was presented with an image ‘A’ followed by a second image ‘B’, denoted a trial. For each trial ‘A’ or ‘B’ was selected as a reference image, i.e. it consisted of an uncompressed original image, and the other selected as the test image containing compression artifacts. The order of reference vs. test image was independently randomized for each trial and observer throughout the session. The observer was asked to assess the absolute quality of image ‘A’ and ‘B’. Prior to each test the meaning of the method’s terms ‘Bad’, ‘Poor’, ‘Fair’, ‘Good’, and ‘Excellent’ was explained as a mean to normalize the evaluation statements given by the observer group. A specific test sequence was also shown initially to all observers, which contained a set of images where the images containing compression artifacts showed the worst possible distortion for the two compression schemes. Thus, defining a distortion space within which all trials in the test session would be located thereby further facilitating comparability between observer evaluations.

3.6 Experimental design

The experiment was set up as a factorial design according to Table 1, i.e. with three main factors (Coding, Image, and Bitrate). The design was randomized with regards to the order in which the different trials were performed. A replication
of 14 was achieved by letting each observer (factor Observer in Table 1) assess the quality of all 18 main factor combina-
tions. Blocking was used and applied to the factor Observer in order address the heterogeneous set of observers, holding
uninteresting nuisance factors constant within each block or observer.

Two response variables were defined in the experiment design:

**Opinion_Original** the assessed quality for the original uncompressed image

**Opinion_Difference** the difference in assessed quality between the original and the compressed image

The Opinion_Original variable allows an estimate of the observation accuracy of each trial to be performed. Given that each
image was shown 6 times (3 · 2) for each observer, the assessed quality of the uncompressed original image within those 6
trials should ideally be the same. The Opinion_Difference was calculated as the difference between Opinion_Original and
Opinion_Compressed, where Opinion_Compressed was the assessed quality of the compressed image in each trial. Hence,
a low value of the variable Opinion_Difference would indicate that the quality of the compressed image is similar to the
uncompressed image.

4. RESULTS

4.1 Objective evaluation

Figure 2 shows the rate-distortion (RD) properties of the compressed images, which is constructed from applying PSNR
according to (1) to the 2D pixel map. Standing out among the three curves is the RD-curve for Fruit. Relative to the
other images Fruit’s two RD-curves are much higher with a minimum of approximately 24.6 dB for H.264/AVC; this being
similar to the maximum PSNR for the other images. This is explained by the lack of motion parallax, and hence large
redundancy between the views within Fruit. Large redundancy is more easily compressed resulting in a more favorable RD
relationship.

Note also the difference in RD-characteristics of the two compression schemes. The quality increase for JPEG 2000, as
a function of increased bitrate, is similar throughout the evaluated range of bitrates. This is in contrast with the characteristic
of H.264/AVC, which shows a steep change in quality at low bitrates but an almost constant quality for higher bitrates.
This is explained by the difference in how JPEG 2000 and H.264/AVC provide their compression ratios in their respective
quantization steps.

The transform coefficients of the 2D wavelet decomposition in JPEG 2000 is quantized using a scalar quantizer, which
can reduce the step size down to 1 and thereby achieve lossless compression. Hence, the PSNR can increase indefinitely
for increasing bitrate; a rate of 24 bpp would theoretically produce an infinite PSNR. The quantization used in H.264/AVC
is instead performed by quantizing the transform coefficients using different step sizes for different frequencies within
a 8x8 block; low frequencies are quantized finer whereas high frequencies are quantized coarser. Hence, in order for
H.264/AVC to produce a continuously increasing RD-characteristic similar to JPEG 2000 all quantizers would have to set
their step size to 1. However, this would counteract the use of applying block-wise DCT-transformation that is adapted to
the characteristics of the human visual system (HVS). In addition it would impair the overall RD-characteristics. Thus, the
approach of using a HVS weighted quantization matrix will always introduce distortion in high frequency content, even for
higher bitrates. This is evident in the H.264/AVC RD-characteristics of Figure 2 where the PSNR levels out at a specific
value over which this compression scheme can not perform.

4.2 Subjective evaluation

The result from the subjective evaluation is divided into two parts: qualitative and quantitative. The evaluation results
will describe the different distortion characteristics and how they relate to the used 3D images and selected compression
schemes. That is, a causal link will be described between the different compression methods and the distortion that may be
observed within the 3D images. The quantitative results stemming from the observer tests will elaborate on the importance
of each distortion type with regard to perceived 3D image quality.

Based on the objective evaluation the three bitrate classes Low, Medium and High were quantified in order to be re-
presentative for the three different images as well as showing significant differences in distortion between the compression
schemes in the pre-trial described in Section 3.1. The Low and High bitrates were selected as the rates corresponding to
the intersections between the RD-functions of Figure 2 (a) and (c) respectively, which according to the pre-trial exhibits markedly different distortion properties. The Medium bitrate was selected to be the same for all evaluated images. Table 2 summarizes the bitrate classes and their corresponding bitrates.

4.2.1 Qualitative

The compressed 3D images differ with respect to both inherent amount of motion parallax as well as the range of depth they contain. The difference is a result of how much these 3D properties are present in the original image as well as to what
Table 3. Summarizing distortion characteristics for evaluated 3D images

<table>
<thead>
<tr>
<th>Coding</th>
<th>Bitrate</th>
<th>Pixelation</th>
<th>Ringing</th>
<th>Blur</th>
<th>Depth</th>
<th>Perspective</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG 2000</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>uneven</td>
</tr>
<tr>
<td></td>
<td>medium</td>
<td>low</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>uneven</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>indeterminable</td>
</tr>
<tr>
<td>H.264/AVC</td>
<td>low</td>
<td>high</td>
<td>medium</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>even</td>
</tr>
<tr>
<td></td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>even</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>medium</td>
<td>low</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>even</td>
</tr>
</tbody>
</table>

degree the compression schemes reduce these properties. The evaluation results for the properties listed in Section 3.1 are summarized in Table 3 and to what degree the evaluated compression schemes affect them. "Low" means that that specific characteristic is now significantly present; "medium" corresponding to a perceived characteristic, and "high" indicates a strong presence of that distortion characteristic. An "uneven" distribution means that most of the distortion is present at larger distances within the 3D image whereas an "even" distribution indicates that most noise is distributed evenly over the image depth. In the case of "none" the distortion characteristic cannot be perceived and hence not quantified. An important difference between JPEG 2000 and H.264/AVC is that the distortion for the former tends to be localized to objects further into the scene resulting in an uneven depth distribution of the distortion. H.264/AVC on the other hand tend to distribute the distortion evenly throughout the 3D image depth range. Figure 3 illustrates the difference in distortion for the Low bitrate by presenting stereoscopic views extracted from the PVS of 3D image Bike. Additional graphs and images can be found at the paper’s accompanying web page.

The two views corresponding to each compression scheme, and the original uncompressed image, are presented with the right view to the left and vice versa, enabling the viewer to by herself perceive the depth of the 3D images using cross-eyed viewing. Note that in Figure 3 (b) a large portion of the background have been translated to the plane of zero disparity, especially noticeable at the rear wheel and the blue and white objects in the floor. This reduction in depth is further illustrated in Figure 4 where the absolute difference between two views are shown for the original, JPEG 2000 and H.264/AVC PVSs respectively. Figure 4 (a) shows significant differences in the background, which essentially is a feature of the H.264/AVC-compressed images as well (Figure 4 (b)). As shown in Figure 4 (c), JPEG 2000 reduces the change between views, which thereby reduces the depth. This applies to all view pairs, which thereby also reduces the motion parallax.

The distortion perceived in Figure 3 is further clarified in Figure 5 where the left view of JPEG 2000 and H.264/AVC is compared to the corresponding view of the original using absolute difference. In this form it is apparent that H.264/AVC (Figure 5 (b)) introduce compression distortion that is distributed within the complete view. Even though the amount of distortion is apparently higher for H.264/AVC, given the less white areas in Figure 5 (b), it is less visible as seen in Figure 3 (c). One explanation for this discrepancy is a lower correlation of distortion between the views for H.264/AVC than for JPEG 2000. Furthermore, the whiter areas in Figure 5 (a) (corresponding to less distortion) are mainly located at zero disparity for JPEG 2000; e.g. around the front wheel spokes, the text Stop!, and the front part of the frame.

4.2.2 Quantitative

The qualitative evaluation showed that the different compression methods manifest differently with regards to distortion characteristics. The following section will elaborate on how the different distortion types are perceived when being presented to a group of non-expert 3D-video observers.

Initially an analysis of the normality assumption was performed resulting in an estimated probability plot. Figure 6 shows that the deviation from a normal distribution was only moderate and the subsequent analysis can be performed assuming normality.

A summary of how the three different original uncompressed images were perceived is presented as a histogram of the assessed quality, shown in Figure 7. The standard deviation of the quality is to a large degree similar for all of the three images. Consequently, the reliability of the observers’ quality assessment is independent of image type. Worth noting is
Figure 3. Stereoscopic views (for cross-eyed viewing) extracted from the PVS of 3D image Bike compressed at Low bitrate.
Figure 4. Images showing the absolute difference between the left and right view of the stereoscopic views in Figure 3. Images have been inverted for presentation purposes.

Figure 5. Images that show the absolute difference between a view from the original and a corresponding view from the compressed version of the 3D image. Images have been inverted for presentation purposes.
that the image Fruit exhibits the highest mean quality despite having the lowest degree of motion parallax, as described in Section 3.4. However, further studies are required before this result can be used to conclude that 3D images with less motion parallax is perceived as having higher quality.

Table 4 summarizes the result from the analysis of variance, i.e. how the control factors defined in the experiment affect the response variable Opinion_Difference. The controlled factors that most significantly affect Opinion_Difference are Image and Bitrate, both with a P-value of < 0.001. An interesting result is that the choice of compression scheme (indicated by controlled factor Coding) shows no statistically significant effect on the assessed quality. This is in contrast to the qualitative analysis, which stated that the two compression schemes had quite different characteristics with respect to a number of distortion types. Taking into consideration the distortion characteristics in Table 3, a conclusion that can be made is that ringing is the distortion type most noticeable by the observers as it being the only distortion type with similar bitrate correlation regardless of selected compression scheme. Reducing the bitrate increases the ringing for both JPEG 2000 and H.264/AVC. The P-value of < 0.001 for the source Observer validates the hypothesis made in the experimental design that it is beneficial to block the observers.
Figure 8 illustrates the main effects of the controlled factors. As shown, there is a clear (and unsurprising) relationship between increased bitrate and reduced Opinion_Difference, i.e. the image quality is assessed to be higher for higher bitrates by the observers. A more surprising result is that the image with the highest assessed quality is Fruit, which according to Section 3.4 has the lowest degree of motion parallax of the three images. However, this result is similar to other studies of multiview 3D image quality where moderate depth range, and therefore a low degree of motion parallax, is preferred over 3D images with more distinctive depth [5]. When it comes to compression scheme, a small advantage is seen for JPEG 2000 when comparing the mean values for Opinion_Difference with respect to the factor Coding. However, given the high P-value for Coding in Table 4, this is a difference that is not statistically significant at the selected significance level.

5. CONCLUSIONS

This paper set out to investigate two classes of compression schemes: wavelet- and DCT/DPCM-based. The compression standards JPEG 2000 and H.264/AVC were used to exemplify the classes and identified as having different distortion characteristics when evaluated qualitatively, especially with respect to:

- distortion distribution as a function of depth
- induced blurring
These differences were however not validated when a quantitative evaluation was conducted using 14 non-expert viewers. The statistical analysis instead showed that image content and used bitrate significantly affected the perceived quality and not the identified distortion characteristics. Furthermore it was shown that the image with the least inherent 3D effect, in terms of motion parallax, was deemed having the best quality by the group of viewers.

The discrepancy between the two evaluation methods is supported by similar work conducted by other research groups. However, future studies will explore why the 3D effects that these images exhibit are not fully appreciated by non-expert viewers when being displayed on contemporary generations of autostereoscopic 3D displays. In this study a stereoscopic display will be included in order to isolate compression as the main parameter influencing perceived quality.

ACKNOWLEDGMENTS

This work is supported by the EU Objective 2 - project Digital 3D Signage and the Visualization project Medi3D which is partly financed by KK-stiftelsen, Vinnova, SSF, ISA and Vårdalstiftelsen.

References


