Simultaneous 2D and 3D perception for stereoscopic displays based on polarized or active shutter glasses

Payman Aflaki[a], Miska M. Hannuksela[b], Hamed Sarbolandi[a], Moncef Gabbouj[a]
[a]Department of Signal Processing, Tampere University of Technology, Tampere, Finland
[b]Nokia Research Center, Tampere, Finland

Abstract— **Viewing stereoscopic 3D content is typically enabled either by using polarizing or active shutter glasses. In certain cases, some viewers may not wear viewing glasses and hence, it would be desirable to tune the stereoscopic 3D content so that it could be simultaneously watched with and without viewing glasses. In this paper we propose a video post-processing technique which enables good quality 3D and 2D perception of the same content. This is done through manipulation of one view by making it more similar to the other view to reduce the ghosting artifact perceived without viewing glasses while 3D perception is maintained. The proposed technique includes three steps: disparity selection, contrast adjustment, and low-pass filtering. The proposed approach was evaluated through an extensive series of subjective tests, which also revealed good adjustment parameters to suit viewing with and without viewing glasses with an acceptable 3D and 2D quality, respectively.**

Index Terms—Stereoscopic; depth perception; subjective quality assessment; 3DV; 2DV.

1. INTRODUCTION

In the recent years, the number of 3D movie titles has increased considerably both at cinemas and as Blu-ray 3D discs. Moreover, broadcast of stereoscopic video content is provided commercially on a few television channels. Hence, many user side devices are already capable of processing stereoscopic 3D content whose volume is expected to rise sharply in the coming years. Preferences of customers drive the direction of improvements and novelties in different presentation methods of the 3D content and it is therefore important to understand the habits of viewing 3D content and mechanisms of the human vision. Psycho-visual aspects must therefore be considered when displaying 3D content.

The human vision system (HVS) perceives color images using receptors on the retina of the eye which respond to three broad color bands in the regions of red, green and blue in the color spectrum. HVS is more sensitive to overall luminance changes than to color changes. The major challenge in understanding and modeling visual perception is that what people see is not simply a translation of retinal stimuli (i.e., the image on the retina). Moreover, HVS has a limited sensitivity; it does not react to small stimuli, it is not able to discriminate between signals with an infinite precision, and it also presents saturation effects. In general one could say it achieves a compression process in order to keep visual stimuli for the brain in an interpretable range.

Stereoscopic vision is one of the principal methods by which humans extract 3D information from a scene. HVS is able to fuse the sensory information from the two eyes in such a way that a 3D perception of the scene is formed in a process called stereopsis. In stereoscopic presentation, the brain registers slight perspective differences between left and right views to create a 3D representation incorporating both views. In other words, the visual cortex receives information from each eye and combines this information to form a single stereoscopic image. Presenting different views for each eye (stereoscopic presentation) usually results into binocular rivalry where the two monocular patterns are perceived alternately [1]. In such a case, where dissimilar monocular stimuli are presented to corresponding retinal locations of the two eyes, rather than perceiving stable single stimuli, two stimuli compete for perceptual dominance. Rivalry can be triggered by very simple stimulus differences or by differences between complex images. These include differences in color, luminance, contrast polarity, form, size, and velocity. Stronger, high-contrast stimuli lead to stronger perceptual competition. In particular cases, one of the two stimuli dominates the field. This effect is known as binocular suppression [2], [3]. It is assumed according to the binocular suppression theory that the HVS fuses the two images with different levels of sharpness such that the perceived quality is close to that of the sharper view [4]. In contrast, if both views show different amounts of blocking artifacts, no considerable binocular suppression is observed and the binocular quality of a stereoscopic sequence is rated close to the mean quality of both views [5].

Binocular suppression has been exploited in asymmetric stereoscopic video coding, for example by providing one of the views with lower spatial resolution [6] or with lower frequency bandwidth [7], fewer color quantization steps [8], or coarser transform-domain quantization [9], [10]. In this paper we exploit binocular suppression and asymmetric quality between views in another domain, namely presentation of stereoscopic 3D content simultaneously on a single display for viewers with and without viewing glasses. Such a viewing situation may occur, for example, when television viewing is not active, but the television set is just being kept on as a habit. The television may be located in a central place at home, where many family members are spending their free time. Consequently, there might be viewers actively watching the television with glasses and while others are primarily doing something else (without glasses) and just momentarily peeking at the television. Furthermore, the price of the glasses, particularly the active ones, might constrain the number of glasses households are willing to buy. Hence, in some occasions, households might not have a sufficient number of glasses for family members and visitors watching the television. While glasses-based stereoscopic display systems provide a good stereoscopic viewing quality, the perceived quality of the stereo picture or picture sequence viewed without glasses is intolerable. Recently, authors in [11] presented a system for automatic 2D/3D display mode selection based on whether the users in front of the 3D display wear viewing glasses. In the research presented in [11] a combination of special viewing glasses and a camera on top of the display enables such display mode selection. However, this approach does not solve the problem of a mixed group of observers, some with and some without viewing glasses and only enables switching between 2D and 3D presentation based on the number of subjects with or without viewing glasses in front of the display.

We enable the same content to be simultaneously viewed both in 3D with viewing glasses and in 2D without viewing glasses by digital signal processing of the decoded stereoscopic video content, making the perceived quality in glasses-based stereoscopic viewing systems acceptable for viewers with and without 3D viewing glasses simultaneously. Viewers with glasses should be able to perceive stereoscopic pictures with acceptable quality and good depth perception, while viewers without glasses should be able to perceive single-view pictures i.e. one of the views of the stereoscopic video. The proposed processing is intended to take place at the display and can be adapted for example based on the ratio of users with and without viewing glasses. In the proposed algorithm, one of the views is processed so that its presence becomes harder to perceive when viewing the content without viewing glasses, while the quality and 3D perception is not compromised much thanks to binocular suppression. The proposed method includes three steps, namely disparity adaptation, low-pass filtering of the non-dominant view, and contrast adjustment. While known methods are used for each processing step, we are not aware of previous research works tackling the same problem, i.e. stereoscopic 3D content being simultaneously viewed with viewing glasses by some users and without viewing glasses by other users.

The rest of this paper is organized as follows. In section 2 we present a literature review of the research fields related to the algorithm proposed in the paper, while the proposed post-processing algorithm is described in section 3. Test setup and results are presented in sections 4 and 5, respectively. Finally the paper concludes in section 6.

2. LITERATURE REVIEW

In this section, we provide an extensive literature review focused on the operation of human visual system when observing an asymmetric quality stereoscopic video. Different types of asymmetry are classified and subjective assessment results are reported in sections 2.1 and 2.2 from perception and video compression viewpoints, respectively. Moreover, in section 2.3, we discuss the effect of camera separation on the depth perception. These techniques provide a basis for rendering algorithms utilized in this study. In section 2.4 we summarize some key aspects affecting the perceived 3D video quality, which are subsequently taken into consideration in the performed subjective viewing experiment. Finally, in section 2.5, the concept of depth-enhanced multiview video coding is described, as it can provide an unlimited number of rendered views at the 3D display. This coding approach can be exploited to display stereoscopic video with arbitrary camera separations, hence facilitating the disparity adaptation step of the method proposed in this paper.

2.1. Visual perception of asymmetric stereoscopic video

Binocular suppression provides an opportunity to use different types of asymmetry between views. Many research works have been carried out to study which types of asymmetry are subjectively most pleasing to human observers or closest to the symmetric stereoscopic video and to find optimal settings for various parameters related to the strength of asymmetry.

Typically the greater the amount of high frequency components (more detail), the better the 3D perception of the objects. This means that the stereo acuity decreases when the amount of blurring increases [12]. However, [13] studied this topic in more detail showing that within certain limits, it is possible to perceive stimuli well in 3D even when one eye sees a blurred image while the other eye sees a sharper one.

The capability of the HVS to fuse stereo pairs of different sharpness has been studied in many papers. Authors in [6] subjectively assessed the quality of uncompressed mixed-resolution asymmetric stereoscopic video by downsampling one view with ratios 1/2, 3/8, and 1/4. The results show that while downsampling ratio is equal to 1/2 the average subjective score has sufficient subjective quality which is comparable to that of full resolution stereo pair. A similar experiment was conducted by Stelmach in [14] where the response of HVS to mixed-resolution stereo video sequences where one view was low-pass filtered was explored by performing a series of subjective tests. Subjects rated the overall quality, sharpness, and depth perception of stereo video clips. The results show that the overall sensation of depth was unaffected by low-pass filtering, while ratings of quality and sharpness were strongly weighted towards the eye with the greater spatial resolution. Moreover, authors in [7] evaluated the perceptual impact of low-pass filtering applied to one view of a stereo image pairs and stereoscopic video sequences in order to achieve an asymmetric stereo scenario. The results showed that binocular perception was dominated by the high quality view when the other view was low-pass filtered.

2.2. Asymmetric stereoscopic video coding

The types of asymmetric video coding can be coarsely classified into mixed-resolution, asymmetric sample-domain quantization, asymmetric transform-domain quantization and asymmetric temporal resolution. Furthermore, a combination of different types of scalabilities can be used. The different types of asymmetric stereoscopic video coding are reviewed briefly in the sequel.

Mixed-resolution stereoscopic video coding [15], also referred to as resolution-asymmetric stereoscopic video coding, introduces asymmetry between views by low-pass filtering one view and hence providing smaller amount of spatial details or a lower spatial resolution. Furthermore, usually a coarser sampling grid is utilized for the low-pass-filtered image, i.e. the content is represented with fewer pixels. Mixed-resolution coding can also be applied for a subset of color components. For example, in [16], luma pictures of both views had equal resolution while chroma pictures of one view were represented by fewer samples than the respective chroma pictures of the other view.

In asymmetric transform-domain quantization the transform coefficients of the two views are quantized with a different step size. As a result, one of the views has a lower fidelity and may be subject to a greater amount of visible coding artifacts, such as blocking and ringing. In [9], the authors performed a series of subjective test experiments on coded stereoscopic video clips with asymmetric luminance qualities. Asymmetric luminance was achieved with coarser quantization of transform coefficient values in one luma view. Subjective results show that stereoscopic video coding with asymmetric luminance information achieved a bitrate reduction from 9% to 34% while maintaining the just noticeable distortion as introduced in [17]. Moreover, authors in [10] subjectively compared the quality of coded mixed-resolution stereoscopic video with that of compressed full-resolution video. The results revealed that under the same bitrate constraint, the same subjective quality can be expected while decreasing the spatial resolution of one view by a factor of 1/2 horizontally and vertically.

In asymmetric sample-domain quantization [8] the sample values of each view are quantized with a different step size. A higher compression ratio can be achieved for the quantized view compared to the other view, due to fewer quantization steps. Both luma and chroma samples can be processed with different quantization step sizes. If the number of quantization steps in each view matches a power of two, a special case of asymmetric sample-domain quantization, called bit-depth-asymmetric stereoscopic video, can be achieved. [8] presents a video coding scheme based on uneven quantization steps for luma sample values of left and right views along with spatial downsampling. Results of subjective quality assessment showed that the average ratings of proposed method outperformed full resolution symmetric and mixed resolution asymmetric stereoscopic video coding schemes with different downsampling ratios.

To our knowledge, asymmetric contrast has not been utilized in stereoscopic video compression. However, authors in [18] subjectively assessed the subjective quality of a wide range of binocular image imperfections by pointing out asymmetry threshold values which provide equal visual comfort. It was found that the contrast difference between views should not exceed 25% to prevent eye strain in subjects.

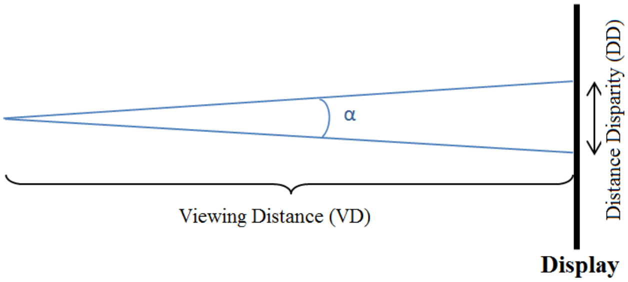
2.3. Impact of parallax on depth perception

Screen parallax is created by the difference between the left and right eye images on the 3D display. We need to converge and accommodate (focus) the eyes in order to project the object of interest to the fovea in both eyes. The distance between us and the object of interest defines the amount of convergence and accommodation in our eyes. Convergence can be defined as a process that is basically disparity driven and consists of the movement of the two eyes in opposite direction to locate correctly the area of interest on the fovea. Accommodation tries to remove blur and hence, alters the lens to focus the area of interest on the fovea [19].

Under natural conditions the accommodation and convergence systems are reflexively linked. The amount of accommodation needed to focus on an object changes proportionally to the amount of convergence required to project the same object on the fovea of the eyes. Under conditions of binocular fusion, for a certain amount of convergence, accommodation has a certain depth of focus, in which it can move freely and objects are perceived properly [20].

An area defining an absolute limit for disparities that can be fused in HVS is known as Panum’s fusional area [21], [22]. It describes an area, within which different points projected on to the left and right retinas produce binocular fusion and sensation of depth. Hence, horizontal disparity should be limited within Panum’s fusional area. Otherwise, excessive disparity could cause double vision or severe visual fatigue. The limits of Panum's fusional area are affected by many factors e.g. including stimulus size, spatial frequency, exposure duration, temporal effects, continuous features, and amount of luminance [21]. Disparities beyond 60 to 70 arcmin are assumed to cause visual discomfort and eye strain [23], [24].

Camera separation creates a disparity between the same object on the left- and right-view images on a display, which can be expressed in terms of number of pixels. Based on the display width and resolution, the disparity can be converted from a number of pixels to a distance disparity e.g. in centimeters as shown in (1) and (2).



$$\alpha = 2 \times \operatorname{atan}\left(\frac{DD}{2 \times VD}\right)$$

Figure 1. Disparity calculation in arcmin based on different disparities in number of pixels on display

$$w = W_{cm} / W_{pixels} \qquad (1)$$

where W_{cm} is the display width in cm and W_{pixels} is the display width in pixels. Hence, w presents one pixel width in cm.

$$DD = w \times PD \qquad (2)$$

where DD is the distance disparity and PD is the disparity in number of pixels

Considering the viewing distance (VD), the disparity in arcmin can be calculated for different objects in the scene using (3). This is depicted in Figure 1.

$$D_{Arcmin} = 2 \times \operatorname{atan}\left(\frac{DD}{2 \times VD}\right) \qquad (3)$$

where D_{Arcmin} is the disparity in arcmin and atan calculates the Arc Tangent in arcmin.

Pastoor in [17] assessed the viewing comfort when watching a series of stereoscopic images with disparities ranging from 0 to 140 arcmin. The results show that disparities up to 35 arcmin do not cause any discomfort while disparities above 70 arcmin should be avoided.

2.4. 3D video quality

Considering asymmetric stereoscopic video, artifacts causing contradictory depth cues are sent to each eye. Similarly to asymmetric video encoding which results in the masking of the artifacts of the worst view, the risk is to suppress the stereopsis because there might be no correspondences between the left and right views.

Even though it has been shown that image quality is important for visual comfort, it is not the only factor for great 3D visual experience. New concepts such as depth perception and presence i.e. the feeling of being there have to be considered too. These concepts are extensively studied in [25], [26], and [27].

One annoying artifact while observing 3D content with glasses is crosstalk [28]. It is perceived as shadow or double contours (ghosting artifact) due to imperfect optical separation between the left and the right images by filters of passive glasses or slight imperfection in synchronization between shutters in active glasses and the displayed left and right views [29]. This will cause perception of opposite view by each eye causing the ghosting artifact while it should have been blocked by the viewing glasses. Crosstalk has been mentioned as one of the main disturbing perceptual display related factors for 3D viewers [30]. The ghosting artifact is most visible when watching a stereoscopic video on a 3D display without glasses (2D presentation), since both left and right views are visible to both eyes. Hence, the subjective quality of stereoscopic video in 2D presentation is not acceptable due to this artifact as depicted in Figure 2.

2.5. Depth-enhanced multiview video coding

Multiview autostereoscopic displays (ASDs) require many high-quality views to be available at the decoder/display side prior to displaying. Due to the natural limitations of content production and content distribution technologies, there is no way that a large number of views can be delivered to users with existing video compression standards. Moreover, due to differing subjective preferences on the amount of depth in 3D displaying as well as different 3D displays and viewing environments, it is desirable to enable depth or disparity of the content in the decoder/display side. Therefore, the Moving Picture Experts Group (MPEG) issued a Call for Proposals for 3D video coding (hereafter referred to as the 3DV CfP) [31] for a new standard which enables rendering of a selectable number of views without increasing the required bitrate. The work initiated by MPEG has been continued in the 3D video coding standardization in the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) [32] and aims at enabling a variety of display types and

Figure 2. Subjective quality of stereoscopic video without glasses

preferences including varying camera separation to adjust the depth perception.

In ASD and other 3D display applications many views should be available at the decoder side. A multiview video plus depth (MVD) format [33], where each video data pixel is associated with a corresponding depth map value, allows reducing the input data for the 3DV systems significantly, since most of the views can be rendered from the available decoded views and depth maps using a depth-image-based rendering (DIBR) [34] algorithm. Such a scenario ensures the availability of a sufficient number of views for display where different disparities based on the targeted application can be achieved. Hence, as proposed by the 3DV CfP, a 3-view MVD coding scenario is suitable for creation of a wide range of required views for multiview ASD rendering while a suitable pair of synthesized views can also be used for rendering on a stereoscopic display.

3. PROPOSED RENDERING ALGORITHM

In this section, a set of adaptation methods, taking advantage of the binocular suppression theory and achieving a tradeoff between stereoscopic viewing with glasses and single-view viewing without glasses, are introduced. In these adaptation methods, one view is chosen as the dominant view while the other view will be the non-dominant view. The aim of the methods is to let the dominant view be perceived clearly and the ghosting effect caused by the non-dominant view to be close to imperceptible when viewing without glasses, while the perceived quality in viewing with glasses is only slightly degraded. The adaptation processes the non-dominant view and the disparity of the stereo pair with three methods. The selection of these methods was based on the previous conclusions in the literature showing that none of the methods is expected to affect the subjective quality of stereoscopic video considerably. In the first step, disparity is selected in agreement with [5], [17] and without sacrificing the depth perception in stereoscopic 3D presentation. Following this, the non-dominant view is low pass filtered, as it is shown in [6], [7], [13], and [14] that this does not affect the 3D perceived subjective quality. In the final step, a contrast adjustment algorithm is applied on the non-dominant view in favor of better quality in presentation without glasses. It has been confirmed in [18] that contrast adjustment of one view does not decrease the visual quality of stereoscopic video noticeably while watched with glasses. Figure 3 depicts the block diagram of the rendering process. As can be seen from Figure 3, the proposed processing takes place after decoding the stereoscopic video content and could be implemented in a television set or a display capable of stereoscopic rendering. All processing steps can be made adjustable, so that the viewers can be given the option of controlling the strength or the amount of processing. In the following sub-sections, each of the three processing steps is described in more details.

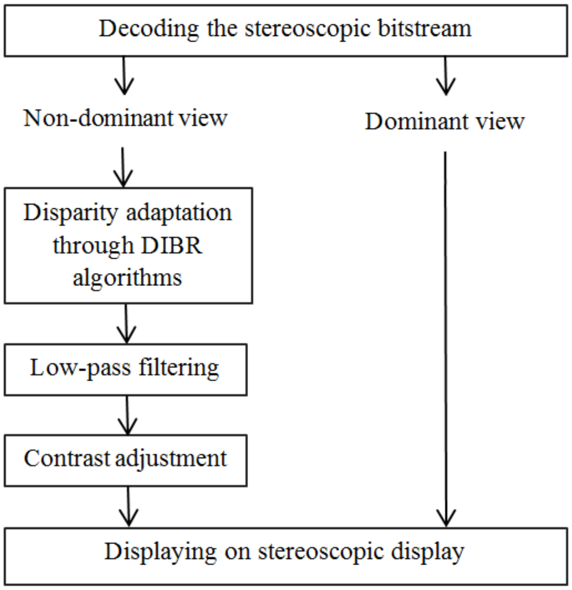


Figure 3. Block diagram of the rendering process

3.1. Disparity selection

It is important to control the disparity between the views in a stereoscopic presentation of 3D content in such a manner that the content is comfortable to view while a desired depth perception is also obtained. Clearly, while increasing the distance between left and right views, the ghosting artifact in 2D presentation of stereoscopic video increases and thus, more annoying subjective quality is expected when the content is viewed without viewing glasses. On the other hand, if the small disparity between views is chosen, the depth perception in 3D presentation decreases.

Disparity selection between the views is initially determined at the time of generating the content, for example through the camera baseline separation and the distance from the camera to the filmed objects. Disparity selection at the rendering device is enabled if a depth-

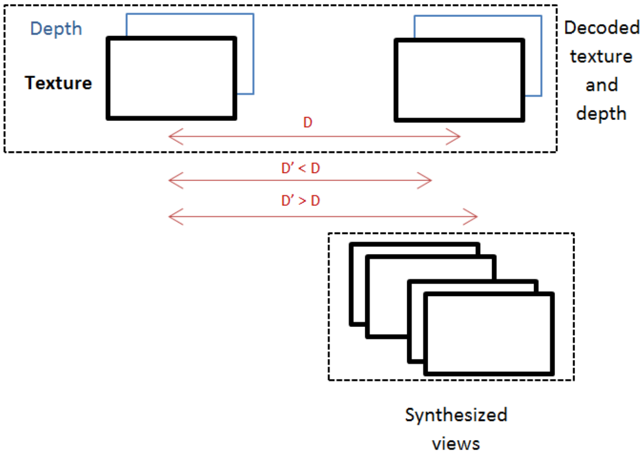


Figure 4. Enabling disparity selection through view synthesis process where D′ represents the view separation achieved by view synthesis process compared to view separation of decoded views (D)

enhanced multiview video coding is used as a distribution format or if the rendering device is capable of a disparity or depth estimation from decoded stereo pairs. Consequently, by means of DIBR algorithms, a view at a desired location can be synthesized. Considering the selected disparity and hence, the estimated view separation, a combination of one coded view and one synthesized view can be exploited to create the displayed stereoscopic video. This is illustrated in Figure 4.

The proposed adaptation methods presented in the next two sub-sections aims at rendering the non-dominant view as invisible as possible in the presentation of stereoscopic video without glasses. Nevertheless, having a smaller disparity still provides a smoother subjective quality for a 2D presentation of the content.

3.2. Low pass filtering

Low-pass filtering decreases the number of high frequency components (HFCs) in the non-dominant view by removing some details. Hence, in the created asymmetric stereoscopic video, the non-dominant view will be blurred compared to the dominant view. This will favor better 2D presentation of the stereo pair, as the dominant view will be sharper compared to the blurred non-dominant view and therefore it will be better perceived by HVS. Yet, as verified extensively in previous studies [4], [6], [7], and [14] asymmetric stereoscopic video where one view has been low pass filtered provides similar subjective quality and depth perception to those of stereoscopic video where both views have the same high quality.

In our experiments, the applied low-pass filter (LPF) was a 2D circular averaging filter (pillbox) within a square matrix having 2×radius+1 elements in each row and column, as it resulted in a better subjective performance compared to a few other tested LPFs. The equation used for this filter is MATLAB implementation of a simple pillbox filter presented in [35]. In general, any LPF can also be selected for example on the basis of memory access and complexity constraints. The level of HFC reduction depends on the radius defined for the filter such that increasing the radius results in more reduction of HFCs. The 2D matrix presenting the LPF coefficients of the used LPF for radius 6 is depicted in (4).

$$f = 10^{-4} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 13 & 36 & 44 & 36 & 13 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 61 & 88 & 88 & 88 & 88 & 88 & 61 & 8 & 0 & 0 \\ 0 & 8 & 76 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 76 & 8 & 0 \\ 0 & 61 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 61 & 0 \\ 13 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 13 \\ 36 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 36 \\ 44 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 44 \\ 36 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 36 \\ 13 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 13 \\ 0 & 61 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 61 & 0 \\ 0 & 8 & 76 & 88 & 88 & 88 & 88 & 88 & 88 & 88 & 76 & 8 & 0 \\ 0 & 0 & 8 & 61 & 88 & 88 & 88 & 88 & 88 & 61 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 13 & 36 & 44 & 36 & 13 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

3.3. Contrast adjustment

The response of HVS depends much more on the relation of luminance local variations compared to the surrounding values than absolute luminance. Contrast is a measure for this relative variation of luminance. In the visual perception of different scenes, contrast is determined by the difference in color and brightness of each object and other objects in the same viewing field. Hence, contrast adjustment is related to brightness and color settings i.e. how the luminance and chrominance differ and change.

The approach utilized in this experiment is to decrease the contrast of luma and chroma components of the non-dominant view while keeping the contrast of the dominant view unchanged. The contrast decrease of the non-dominant view will help the 2D presentation of the stereoscopic view that has more similarity to the dominant view while the stereoscopic presentation is not influenced considerably.

The contrast adjustment of an image can be done in various ways. We follow the same algorithm as used for the weighted prediction mode of the Advanced Video Coding (H.264/AVC) standard [36], that is:

$$O = round\left(\frac{i \times w}{2^d}\right) = (i \times w + 2^{d-1}) \gg d \tag{5}$$

where:

O is the adjusted luma or chroma contrast value
$round$ is a function returning the closest integer
i is the input sample value
w and d are the parameters utilized to create the adjustment weight
>> is a bit shift operation to the right

4. TEST SETUP

The performed tests targeted at verifying that the proposed method has potential to tackle the presented problem satisfactorily, i.e. that the same stereoscopic 3D content can be viewed with viewing glasses with acceptable 3D quality and depth perception and without viewing glasses with acceptable 2D quality and a tolerable level of ghosting artifacts. Furthermore, the performed tests aimed at discovering how to tune the processing steps of the proposed algorithm optimally, i.e. which are good trade-offs for the three processing components, disparity selection, low-pass filtering and contrast adjustment. As no objective video quality metrics are applicable to the presented problem as far as the authors are aware of, a large-scale subjective assessment was performed with four sequences: Poznan Hall2, Poznan Street [37], Ghost Town Fly (GT Fly), Undo Dancer, which were used in the 3DV CfP [31]. For GT Fly and Undo Dancer sequences, 500 frames were used while 250 and 200 frames were used for Street and Hall2, respectively. No encoding was applied to the

TABLE 1. INPUT VIEWS AND CAMERA DISTANCES FOR SMALL AND BIG CAMERA SEPARATIONS

	Left view-Right view , (Camera separation in cm)	
Sequence	Small disparity	Big disparity
Poznan Hall2	7-6.5 , (6.87)	7-6 , (13.75)
Poznan Street	5-4.5, (6.87)	5-4 , (13.75)
GT Fly	3-1 , (4)	5-1 , (8)
Undo Dancer	1-3 , (4)	1-5 , (8)

sequences. The frame rate was fixed to 25 Hz for all sequences. Each sequence was evaluated at two different disparities or camera separations, referred to as small and big disparity subsequently. The camera separation of the big disparity is the same as those introduced in MPEG 3DV CfP for the 3-view coding scenario and can be considered to represent a typical disparity for stereoscopic viewing, while in the small disparity scheme the camera separation distance is halved. The input views and the relative camera separation distances used in the experiments, for both small and large disparity stereoscopic sequences, are shown in Table 1.

4.1. Preparation of Test Stimuli

To prepare test material, the three adaptation methods presented in section 3 were used and various test cases based on different combinations of adaptation methods were created. In the experiments, we tested contrast reduction to 50% and 75% of the original values for different combinations by fixing the value of d to 4 and setting the value of w equal to 8 and 12, respectively, in equation (5). Moreover, all non-dominant views for different schemes except Original 2D were low pass filtered using the circular averaging filter with radius equal to 6 as presented in equation (4).

Two different disparities between the left and the right views were selected for different sequences. In the test sequences the disparity was always positive i.e. the objects are always behind the display level. Disparity selection was limited so that the results were in agreement with previous findings in the literature to prevent eye strain due to excessive disparities.

Disparity can be calculated from depth map by converting it to disparity. Table 2 presents the average and the maximum disparities for each sequence. Moreover, Table 1 presents the selected views and corresponding camera separations for different disparities of the sequences. For Poznan Hall2 and Poznan Street sequences, views 6.5 and 4.5, respectively, were synthesized from the original texture and depth views using the MPEG View Synthesis Reference Software (VSRS) version 3.5 [38]. The subjective quality of synthesized views was comparable to that of the original views. Moreover, since the synthesized artifacts were subjectively negligible, we assume that the synthesizing process did not affect the subjective ratings.

Combining the above-mentioned tested parameters, the following seven test cases were prepared and subjectively assessed. The combinations for each scheme are presented in the format of (disparity, contrast) where for disparity the values *0, Small, Big* refer to *0* disparity (identical left and right views), *Small* disparity, and *Big* disparity, respectively. For contrast the values *X%* present the contrast ratio of the non-dominant view relative to the dominant view. Seven different test schemes, as presented in Table 3, were used in the subjective tests.

TABLE 2. DISPARITIES FOR SMALL AND BIG CAMERA SEPARATION

	Average disparity (Maximum disparity) in arcmin	
Sequence	Small disparity	Big disparity
Poznan Hall2	18.6(22.2)	37.2(44.3)
Poznan Street	19.3(23.6)	38.6(47.2)
GT Fly	12.1(42.2)	24.3(84.3)
Undo Dancer	13.6(22.2)	27.2(47.2)

TABLE 3. DIFFERENT SCHEMES AND THEIR CHARACTERISTICS

Scheme	Disparity	Contrast adjustment
O → (Original 2D)	*0*	*100%*
S1	*Small*	*100%*
S2	*Small*	*75%*
S3 → (Best 2D quality)	*Small*	*50%*
B1 → (Best 3D quality)	*Big*	*100%*
B2	*Big*	*75%*
B3	*Big*	*50%*

4.2. Test Procedure and Subjects

Subjective viewing was conducted according to the conditions suggested in MPEG 3DV CfP. The polarized 46'' Vuon E465SV 3D display manufactured by Hyundai was used. The display has a total resolution of 1920×1200 pixels and a resolution of 1920×600 per view when used in the stereoscopic mode was used for displaying the test material. The viewing distance was equal to 4 times the displayed image height (2.29m).

Subjective quality assessment was done according to the Double Stimulus Impairment Scale (DSIS) method [39] with a discrete unlabeled quality scale from 0 to 10 for quality assessment. The test was divided into two sessions where in the first session, subjects assessed the subjective quality of video clips with glasses and in the second session, the test was performed without glasses. Two questions for each session of the test were considered and the subjects wrote their ratings after each clip was played. These questions are presented in Table 4. Each question is associated with its short term for simplicity in reporting the results. Prior to each test, subjects were familiarized with the test task, the test sequences and the variation in the quality to be expected in the actual tests. The subjects were instructed that 0 stands for the lowest quality and 10 for the highest.

Subjective viewing was conducted with 20 subjects, (16 males, 4 females), aged between 21-31 years (mean: 24.2). All subjects passed the test for stereovision prior to the actual test. Moreover, they were all considered naïve as they did not work or study in fields related to information technology, television or video processing. To prevent subjects from getting exhausted during the evaluation sessions, the duration of the test was limited to 45 minutes.

5. RESULTS AND DISCUSSION

In this section we present the results of the conducted subjective tests and an analysis of the statistics of the quantitative viewing experience ratings.

Figure 5 shows the subjective viewing experience ratings with 95% confidence interval (CI) for all sequences. The results are provided for four questions that subjects

Table 5. Flag table presenting significant differences for different test schemes presented in Table 3

Flags -1, 0, and 1 present significantly lower, similar, and significantly higher quality compared to other schemes, respectively

		Test scheme combinations					
	Flags	S1	S2	S3	B1	B2	B3
Dancer	-1	2	**1**	0	3	2	3
	0	17	**17**	16	16	18	14
	1	1	**2**	4	1	0	3
GT Fly	-1	4	**1**	1	5	2	1
	0	16	**17**	13	13	17	16
	1	0	**2**	6	2	1	3
Street	-1	4	**2**	3	7	4	4
	0	13	**13**	9	10	15	12
	1	3	**5**	8	3	1	4
Hall2	-1	4	**3**	6	4	0	4
	0	12	**14**	12	12	14	14
	1	4	**3**	2	4	6	2
Sum for all sequences	-1	14	**7**	10	19	8	12
	0	58	**61**	50	51	64	56
	1	8	**12**	20	10	8	12

were asked during the test sessions (see sub-section 4.2). The naming introduced for the different schemes in sub-section 4.1 is used in the figures for simplicity. Subjective ratings show that scheme O achieved the highest value in 2D evaluation (i.e. the session where viewing took place without glasses) and in the general quality of 3D presentation. However, because depth perception was rated the smallest in this scheme, it cannot be considered as a competitor for an acceptable trade-off for simultaneous 2D and 3D perception. Hence, it was excluded from the analysis presented next. For the other tested schemes, the following general trend was observed. In both small and big disparities, while decreasing the contrast ratio of the non-dominant view, the ratings of the 2D evaluation session increase and at the same time the 3D evaluation ratings decrease. This was expected as reducing the contrast of the non-dominant view targets ideal 2D subjective quality while compromising the 3D perception. Moreover, in all sequences, the ghosting effect in the 2D presentation of stereoscopic video clips without any contrast adjustment annoyed subjects more in the big disparity scheme when compared to the small disparity schemes. Considering the large amount of viewing experience ratings, it is hard to make many logical conclusions based on Figure 5. Hence, significant differences between the schemes were further analyzed using statistical analysis as presented in the paragraphs below.

The Wilcoxon's signed-rank test [40] was used as the data did not reach normal distribution (Kolmogorov-Smirnov: $p<0.05$). Wilcoxon's test is used to measure differences between two related and ordinal data sets [41]. A significance level of $p < 0.05$ was used in the analysis.

The following conclusions were obtained with this statistical significance analysis mentioned above. In the analysis, we compared pairwise the ratings of each two test case combinations resulting in fifteen flags per question and per sequence, indicating whether the subjective quality between different test cases have any statistically significant difference. Considering four sequences, four questions per sequence, and fifteen two-sided pairwise comparisons per question, we obtained $4\times4\times15\times2 = 480$ flags. Table 5 reports a summary of the distribution of these flags. Each cell presents the total number of flags from different questions where -1, 0, and 1 present significantly lower, similar, and significantly higher quality compared to other schemes, respectively. From this Table it is clear that only S2 provides similar or better subjective results for all sequences while other schemes have a lower performance at least in one sequence. Hence, the combination used in S2 seems to be a well-designed potential candidate for simultaneous 2D and 3D presentation. Moreover, Table 5 reports the cumulative flag counts over all sequences. It can be observed that the cumulative counts for S3 are comparable or better than those for S2. However, by studying the performance of S3 for individual sequences, it can be observed that the performance of S3 for Hall2 is inferior to the results obtained with S2. To analyze the subjective performance of each test scheme combination for 2D and 3D viewing separately, similar flag tables as the one presented in Table 5 are presented in Table 6, reporting results for 2D and 3D viewing experiments separately. Considering the two summaries provided in Table 6, S2 is the only test scheme for which the number of test cases where its performance was statistically superior to the another test scheme (flag value equal to 1) was greater than the number of test cases where its performance was statistically inferior to another test scheme (flag value equal to -1) in both 2D and 3D viewing experiments.

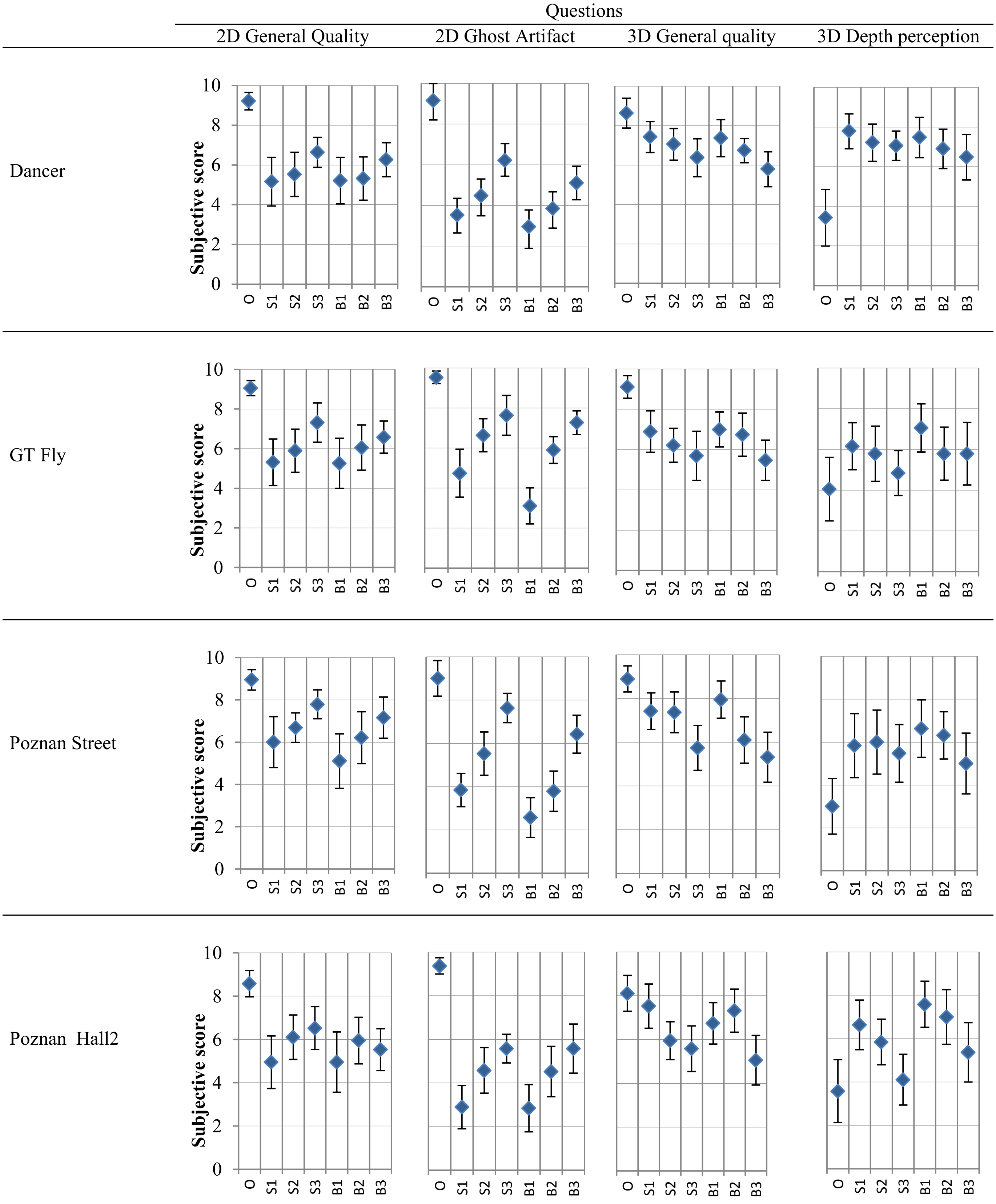


Figure 5. Viewing experience ratings with 95% confidence interval. The schemes are named according to Table 3.

TABLE 6. FLAG TABLE PRESENTING SIGNIFICANT DIFFERENCES FOR DIFFERENT TEST SCHEMES PRESENTED IN TABLE 3 FOR (A) 2D AND (B) 3D EXPERIMENTS

FLAGS -1, 0, AND 1 PRESENT SIGNIFICANTLY LOWER, SIMILAR, AND SIGNIFICANTLY HIGHER QUALITY COMPARED TO OTHER SCHEMES, RESPECTIVELY

(A)

		Test scheme combinations					
	Flags	S1	S2	S3	B1	B2	B3
Dancer	-1	2	**1**	0	3	2	0
	0	8	**8**	6	7	8	7
	1	0	**1**	4	0	0	3
GT Fly	-1	4	**1**	0	5	2	0
	0	6	**7**	4	5	7	7
	1	0	**2**	6	0	1	3
Street	-1	4	**2**	0	7	3	1
	0	5	**5**	2	3	6	5
	1	1	**3**	8	0	1	4
Hall2	-1	4	**0**	0	4	0	0
	0	6	**8**	8	6	8	8
	1	0	**2**	2	0	2	2
Sum for all sequences	-1	14	**4**	0	19	7	1
	0	25	**28**	20	21	29	27
	1	1	**8**	20	0	4	12

(B)

		Test scheme combinations					
	Flags	S1	S2	S3	B1	B2	B3
Dancer	-1	0	**0**	0	0	0	3
	0	9	**9**	10	9	10	7
	1	1	**1**	0	1	0	0
GT Fly	-1	0	**0**	1	0	0	1
	0	10	**10**	9	8	10	9
	1	0	**0**	0	2	0	0
Street	-1	0	**0**	3	0	1	3
	0	8	**8**	7	7	9	7
	1	2	**2**	0	3	0	0
Hall2	-1	0	**3**	6	0	0	4
	0	6	**6**	4	6	6	6
	1	4	**1**	0	4	4	0
Sum for all sequences	-1	0	**3**	10	0	1	11
	0	33	**33**	30	30	35	29
	1	7	**4**	0	10	4	0

The conclusion that S2 provides the most acceptable trade-off for simultaneous 2D and 3D viewing is in agreement with previous findings on contrast asymmetry in [18], where the contrast difference limit between the left and the right views was found to be equal to or less than 25% to provide equal viewing comfort. Moreover, considering camera separations presented in Table 2, the perceived disparity for all sequences was aligned with the results presented in [17], [23], and [24], where the limit for the maximum disparity between the left and right views was found to be 70 arcmin. Only the maximum disparity of the big camera separation for GT Fly is above this limit. This big disparity happens for 0.06 seconds in the 20 second sequence (3 frames in 500 frames). Figure 6 depicts a sample frame from a 2D presentation of a stereoscopic video from scheme S2 and the corresponding stereoscopic video frame with equal disparity and without any LPF or contrast adjustment applied.

After the test, the participants were asked whether they experienced any fatigue or eye strain during and/or after the test. Subjects seemed quite comfortable and there were no complaints regarding the 3D content and the asymmetric nature of the stereoscopic video clips. However, five subjects complained that sometimes it was difficult to distinguish the differences between the observed clips.

6. CONCLUSION

Stereoscopic video provides 3D perception by presenting slightly different views for each eye. Ghosting artifacts make it almost intolerable to watch the content without glasses for both active and passive glasses/displays. In this paper we tackled the problem of viewing 3D content simultaneously with and without viewing glasses by proposing a technique which makes it quite acceptable to watch stereoscopic content without glasses while the 3D perception is not sacrificed much. In the proposed approach, one dominant view is selected and then the non-dominant view is adjusted through disparity selection, contrast adjustment, and low-pass-filtering. These steps increase the similarity of the non-dominant view to the dominant view.

The performance of the proposed technique was assessed through extensive subjective tests. The statistical analysis of scores showed that combination of a disparity smaller than what is conventionally used for stereoscopic video along with low-pass-filtering the non-dominant view and decreasing its contrast to 75% provides the best trade-off between 3D and 2D perception of a stereoscopic 3D content. This is a new topic introduced in 3D research field and as a future plan we intend to do more research on other potential approaches to be used in the process.

Hall2

Street

Dancer

(a) (b)

Figure 6. 2D presentation of stereoscopic video combinations from (a) Original scheme and (b) Selected scheme i.e. S2

ACKNOWLEDGEMENT

The authors thank Prof. M. Domañski, et al. for providing Poznan sequences and their camera parameters [37].

REFERENCES

[1] C. Wheatstone, “On some remarkable, and hitherto unobserved, phenomena of binocular vision,” Philos. Trans. R. Soc. Lond., 1838.

[2] H. von Helmholtz, “Handbuch der physiologischen Optik,” Leopold Voss 1866, (English ed. 1962, Dover New York).

[3] H. Asher, “Suppression theory of binocular vision,” The British Journal of Ophthalmology, vol. 37, no. 1, pp. 37-49, 1953.

[4] B. Julesz, “Foundations of Cyclopean Perception,” University of Chicago Press, Chicago, IL, USA, 1971.

[5] P. Seuntiens, L. Meesters, and W. IJsselsteijn, “Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation,” ACM Trans. Appl. Perception (TAP), vol. 3, pp. 95–109, 2006.

[6] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, “Impact of downsampling ratio in mixed-resolution stereoscopic video,” Proc. of 3DTV Conference, June 2010.

[7] L. B. Stelmach, W. J. Tam, D. V. Meegan, A. Vincent, and P. Corriveau, “Human perception of mismatched stereoscopic 3D inputs,” Proc. of IEEE Int. Conf. on Image Processing (ICIP), Sep. 2000.

[8] P. Aflaki, M. M. Hannuksela, J. Hakala, J. Häkkinen, and M. Gabbouj, “Joint adaptation of spatial resolution and sample value quantization for asymmetric stereoscopic video compression: a subjective study,” Proc. of International Symposium on Image and Signal Processing and Analysis, Sep. 2011.

[9] F. Shao, G. Jiang, X. Wang, M. Yu, and K. Chen, “Stereoscopic video coding with asymmetric luminance and chrominance qualities,” IEEE Trans. on Consumer Electronics, vol.56, no.4, pp.2460-2468, Nov. 2010.

[10] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, “Subjective study on compressed asymmetric stereoscopic video,” Proc. of IEEE Int. Conf. on Image Processing (ICIP), Sep. 2010.

[11] Z. Zivkovic, E. Bellers, “Efficient and robust detection of users wearing stereoscopic glasses for automatic 2D/3D display mode selection”, International Conference on Consumer Electronics (ICCE), pp. 682-683, Jan, 2012

[12] L. M. Wilcox, J. H. Elder, and R. F. Hess, “The effects of blur and size on monocular and stereoscopic localization,” Vision Research, vol. 40, pp. 3575–3584, 2000.

[13] G. Papelba, I. Cipane, and M. Ozolinsh, “Stereovision studies by disbalanced images,” Advanced Optical Devices, Technologies, and Medical Applications, Proc. of SPIE vol. 5123, pp. 323–329, 2003.

[14] L. Stelmach, W. J. Tam; D. Meegan, and A. Vincent, “Stereo image quality: effects of mixed spatio-temporal resolution,” IEEE Trans. Circuits and Systems for Video Technology, Vol. 10, No. 2, pp. 188-193, March 2000.

[15] M. G. Perkins, “Data compression of stereopairs,” IEEE Trans. on Communications, vol. 40, no. 4, pp. 684-696, Apr. 1992.

[16] A. Aksay, C. Bilen, and G. Bozdagi Akar, “Subjective evaluation of effects of spectral and spatial redundancy reduction on stereo images,” European Signal Processing Conference, EUSIPCO, Sep. 2005.

[17] S. Pastoor, “Human factors of 3D imaging: results of recent research at Heinrich-Hertz-Institut Berlin,” Proceedings of the International Display Workshop (Asia Display) ,1995.

[18] F. L. Kooi and A. Toet, “Visual Comfort of binocular and 3D displays,” in Displays, Elsevier, vol. 25, pp. 99-108,2004.

[19] R. Suryakumar, “Study of the Dynamic Interactions between Vergence and Accommodation,” University of Waterloo, 2005.

[20] P. A. Howarth, “Empirical studies of accommodation, convergence, and HMD use,” Proceedings of the Hoso-Bunka Foundation Symposium: The Human Factors in 3-D Imaging, 1996.

[21] A. Coltekin, “Foveation for 3D visualisation and stereo imaging,” Ph.D. thesis, ISBN 951-22-8017-5, Helsinki University of Technology, 2006.

[22] D.Qin, M.Takamatsu, Y.Nakashima, “Measurement for the Panum's Fusional Area in Retinal Fovea Using a Three-Dimention Display Device,” Journal of Light & Visual Environment, pp. 126-131, Vol. 28, Issue 3, 2004.

[23] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, “Effect of disparity and motion on visual comfort of stereoscopic images,” Proc. Stereoscopic Displays and Virtual Reality Syst. XIII, vol. 6055, pp. 60550B-1–60550B-9, 2006.

[24] M. Wopking, “Viewing comfort with stereoscopic pictures: an experimental study on the subjective effects of disparity magnitude and depth of focus,” Journal of the Society for Information Display 3: 101-103, 1995.

[25] W. A. IJsselsteijn, H. de Ridder, J. Freeman, and S. E. Avons, “Presence: concept, determinants, and measurement,” Proc. of the SPIE, Vol. 3959, pp. 520-529, 2000.

[26] W. A. IJsselsteijn, H. de Ridder, J. Freeman, S. E. Avons., and D. Bouwhuis, “Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence,” Presence-Teleoperators and Virtual Environments, vol. 10, no. 3, pp. 298-311, 2001.

[27] P. Seuntiens, Visual Experience of 3D TV, Ph.D. thesis, 2006.

[28] J. Konrad and M. Halle, “3-D displays and signal processing: An answer to 3-D ills?,” IEEE Signal Process. Mag., vol. 24, pp. 97-111, Nov. 2007.

[29] D. Strohmeier, S. Jumisko-Pyykkö, U. Reiter, “Profiling experienced quality factors of audiovisual 3D perception”, Proc. of the Int. Workshop on Quality of Multimedia Experience, pp. 70-75, June 2010

[30] W. A. IJsselsteijn, P. H. J. Seuntiens, and L. M. J. Meesters, “Human Factors of 3D Displays,” 3D Video communication, algorithms, concepts, and real-time systems in human-centred communication, John Wiley & Sons, Ltd.:219-234, 2005.

[31] “Call for Proposals on 3D Video Coding Technology,” ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036, March 2011. http://mpeg.chiariglione.org/working_documents/explorations/3dav/3dv-cfp.zip

[32] http://phenix.int-evry.fr/jct3v/

[33] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, “Multi-view video plus depth representation and coding,” Proc. of IEEE Int. Conf. on Image Processing (ICIP), vol. 1, pp. 201-204, Oct. 2007.

[34] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV,” in Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems XI, vol. 5291, pp. 93–104, Jan. 2004.

[35] Xin Wang, Baofeng Tian, Chao Liang, “Blind image quality assessment for measuring image blur,” IEEE, 2008 Congress on Image and Signal Processing, pp. 467-470, May, 2008.

[36] ITU-T Recommendation H.264, “Advanced video coding for generic audiovisual services,” Jan. 2012.

[37] M. Domañski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner, “Poznan Multiview Video Test Sequences and Camera Parameters,” MPEG 2009/M17050, Oct. 2009.

[38] “View synthesis software manual,” MPEG ISO/IEC JTC1/SC29/WG11, Sept. 2009.

[39] ITU-R Rec. BT.500-11, Methodology for the subjective assessment of the quality of television pictures, 2002.

[40] F. Wilcoxon. Individual comparisons by ranking methods. Biometrics, 1:80–83, 1945.

[41] H. Cooligan “Research methods and statistics in psychology,” (4th ed.). London: Arrowsmith., 2004.