QUALITY ASSESSMENT OF 3D ASYMMETRIC VIEW CODING USING SPATIAL FREQUENCY DOMINANCE MODEL

Feng Lu[1]*, Haoqian Wang*[2]*, Xiangyang Ji*[1] *and Guihua Er*[1]

[1]TNList and Department of Automation, Tsinghua University, Beijing, P. R. China
[2]Graduate School at Shenzhen, Tsinghua University, Shenzhen, P. R. China
lu-f07@mails.tsinghua.edu.cn

ABSTRACT

To save bit-rate in stereo video application, asymmetric view coding is introduced, which encodes the stereo views with different qualities. However, quality assessment on asymmetric view coding is difficult, because the impact of the degraded view upon the 3D percept depends on Human Visual System (HVS) and cannot be indicated by conventional metrics. This paper introduces a quality assessment model based on the observed phenomenon that spatial frequency determines view domination under the action of HVS. A metric is proposed based on this model for assessing the quality of asymmetric view coding. Experimental results are presented to show that the proposed metric provides accordant assessment with the subjective evaluation.

Index Terms— quality assessment, spatial frequency, asymmetric view coding, HVS

1. INTRODUCTION

The concept of 3D video appeared a long time ago, and has received significant attention in recent years. The rapid growth of industrial requirements and the developments of key technologies make it practical within these years [1]. 3D video is distinguished from conventional 2D video by providing the visual cue of depth, which enables viewers to have the same visual experiences as they perceive in the real world. More and more applications use autostereoscopic displays instead of traditional 3D glasses, because it is more comfortable and natural for viewers [2]. Whatever technology is used in display, at least a pair of views is required to be presented to the left- and right-eye to form 3D percept. Therefore, compared to 2D video coding, 3D video coding needs to encode more than one view.

There are two key approaches for encoding multi-view 3D video: 1) each view is encoded using 2D video coding scheme usually with inter-view prediction and 2) a single view and its depth map are encoded and other views are reconstructed by depth image based rendering (DIBR) [3]. Both approaches encode at least one view in high quality as the key view and then encode the non-key views by inter-view prediction or DIBR. For symmetric video coding, the reconstructed quality of non-key views is as high as that of the key view, and the total bit-rate could be very high. Asymmetric view coding is introduced to save bit-rate without degrading the 3D experience. In this approach, the views for left- and right-eye are encoded with different qualities. Specifically, the non-key view for one eye is encoded as low-quality view or degraded view with lower bit-rate. Human Visual System (HVS) is exploited to ensure acceptable perceptual quality for asymmetric view coding. The reason is that HVS can combine two eyes' views with different qualities resulting in a relatively high quality 3D percept in brain. During the combination, also known as binocular fusion, the lost information in the degraded view can be well compensated and the visual noise can be concealed. It's therefore no longer appropriate to use conventional quality metrics for 2D images such as PSNR in 3D asymmetric view coding.

Some works [4, 5] have been done to study the impact of asymmetric view coding. Subjective evaluation is used to understand how the 3D video quality will be influenced while PSNR of the degraded view is varying. However, there seems no direct relationship between 3D perceptual quality and PSNR of degraded view. This is probably because it is not the PSNR but some certain characteristics of the degraded view to play direct roles in binocular fusion under the action of HVS. In this paper, we study the effect of image spatial frequency on view-dominance and propose a quality assessment model for the degraded image in stereoscopic image pair. This model can be used to find which areas in the degraded view dominate the other view and thus influence the 3D percept most. The goal of this work is to exploit the relationship between 3D perceptual quality and specific image characteristics, to make effective quality assessment and to guide 3D video coding.

The rest of the paper is organized as follows. Section 2 gives a brief discussion on how spatial frequency influences the view dominance, image quality assessment method is presented in Section 3, experimental results are shown and discussed in Section 4 and Section 5 concludes this paper.

2. EFFECT OF SPATIAL FREQUENCY ON VIEW-DOMINANCE

Visual effects of 2D video mostly depend on the characteristics of stimulus. With different colors, spatial frequencies and temporal frequencies, one region in the video may be very attractive while the other may be hardly noticeable. In the case of 3D percept, it is very similar to that of 2D percept. When the left- and right- eye views have some differences in the corresponding areas, HVS may choose one side view to dominate the other according to specific characteristics of stimulus of the two side views. Typically, for a pair of left- and right-eye luminance images captured in the same time, the most important factor to determine the dominant side is the spatial frequency.

Spatial frequency is used to measure how often the structure repeats. For digital image, it is a measure of how fast the values of adjacent pixels change. Generally speaking, high spatial frequencies represent abrupt spatial changes (e.g. edges), while low spatial frequencies represent global information of the image or its sub-areas.

As illustrated in Figure 1, the rectangle in the middle only contains low spatial frequencies, while the other two contain very high spatial frequencies (three lines for each). Now assume that the left rectangle represents the high quality image for left eye, and the other two are degraded images for right eye. Experimental results show that we can easily perceive the correct image (rectangle with three vertical lines) from the left and middle images but may be seriously disturbed by the horizontal lines from the right one.

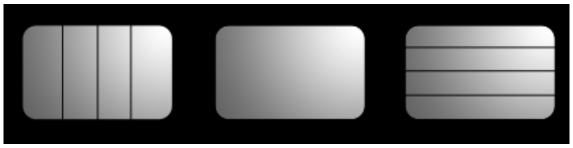


Figure 1: Illustration of eye dominance

Another case is shown in Figure 2, if the high quality image (left) only contains low spatial frequencies and the degraded image (right) contains high spatial frequencies, the artifacts will present in the final perceived image. In this case the degraded view dominates the high quality view.

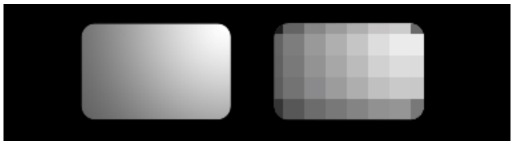


Figure 2: Another case of eye dominance

Several simple principles can be derived from the above phenomena: images with high spatial frequencies are more dominant than the images with very low spatial frequencies. Loss of high spatial frequency information in degraded images is tolerable since it can be well recovered by the high quality images. On the contrary, incorrect information containing high spatial frequencies such as coding artifacts in the degraded images could weigh heavily against the effectiveness of correct 3D percept.

3. QUALITY ASSESSMENT FOR DEGRADED IMAGE

Conventional metrics used to assess the image quality include sum of absolute (SAD), mean absolute difference (MAD) and peak signal-to-noise ratio (PSNR), etc. These methods measure the difference between two images by calculating the difference between every corresponding pixel values. However, these methods do not involve the content of images which may influence the quality greatly.

To include the ability of assessing the impact of eye domination based on spatial frequency, one term should be added to the original assessment function:

$$D(X_d, \lambda) = MAD(X_d) + \lambda\{\Delta E(r(X_d))\} \qquad (1)$$

where $X_d = \{x_{i,j}^d\}$ indicates pixel values of a certain area (e.g. a 4×4 block) in the degraded image. $MAD(X_d)$ is the mean absolute difference serving as conventional quality assessment while $\Delta E(r(X_d))$ is the new term added to measure the effect of spatial frequencies' imparity between degraded and original images. The final assessment of image degradation $D(X_d, \lambda)$ is the weighted sum of the two parts with the multiplier λ adjusting the weights.

3.1. Modeling for spatial frequency effect

To calculate the spatial frequency $r(X_d)$ for certain image area with size M×N, horizontal frequency r_h and vertical frequency r_v are first defined as follows:

$$\begin{aligned} r_h &= \sqrt{\frac{1}{M(N-1)}\sum_{i=0}^{M-1}\sum_{j=1}^{N-1}(x_{i,j}^d - x_{i,j-1}^d)^2} \\ r_v &= \sqrt{\frac{1}{(M-1)N}\sum_{i=1}^{M-1}\sum_{j=0}^{N-1}(x_{i,j}^d - x_{i-1,j}^d)^2} \end{aligned} \qquad (2)$$

Then the spatial frequency of the area can be given by the square root of a quadratic sum:

$$r(X_d) = \sqrt{r_h^2 + r_v^2} \qquad (3)$$

As discussed in Section 2, high spatial frequencies always lead to domination. However, the exact relationship between the disparity of spatial frequencies and the strength of the eye domination is very complex and still unknown. In order to find a simple approximation, one phenomenon should be noticed that image areas with very low spatial frequencies are much easier to be dominated than those with very high spatial frequencies. In other words, the effect of eye domination not only depends on the absolute difference but also the contrast difference of the spatial frequencies. Thus when real spatial frequency increases from r to $r+\Delta r$, it is supposed that the perceived difference $\Delta E(r)$ has the form of:

$$\Delta E(r) = \Delta r / r \tag{4}$$

Then the perceived effect of spatial frequency is:

$$E(r) = \ln(r) + K \tag{5}$$

where K is an undetermined constant.

Now let r_o and r_d be the spatial frequencies for a pair of corresponding areas in the original and the degraded images, respectively. It can be considered that the area from degraded image has negligible impact to the high quality image on condition that r_d is less than r_o. Otherwise, the negative influence from the degraded image is:

$$E(r_d) - E(r_o) = \ln(r_d) - \ln(r_o) = \ln(\frac{r_d}{r_o}) \tag{6}$$

Then the final expression is:

$$\Delta E(r(X_d)) = \begin{cases} \ln\left(\dfrac{r(X_d)}{r(X_o)}\right) & , r(X_d) \geq r(X_o) \\ 0 & , otherwise \end{cases} \tag{7}$$

where $X_o = \{x_{i,j}^o\}$ indicates the pixel values of the same area as X_d in the original high quality image. In the rest of the paper, the area size is set to 4×4.

3.2. Quality assessment method

To assess the degraded image quality, we can substitute Equation (7) into (1) and use it as a metric directly. However, it should be taken into consideration first that how to determine the value of multiplier λ. This work can be hard since the two terms in Equation (1) have different dimensions and value ranges. To avoid this difficulty, we propose a two-step method in quality assessing:

Step 1: For a new 4×4 block, MAD is first calculated. If the result d is greater than a fixed value D_0 (large enough), then regard d as the assessment result and go to step 1 to process the next block, otherwise go to step 2.

Step 2: Calculate $\Delta E(r(X_d))$ for current 4×4 block and regard it as the assessment result. Go to step 1 for next block.

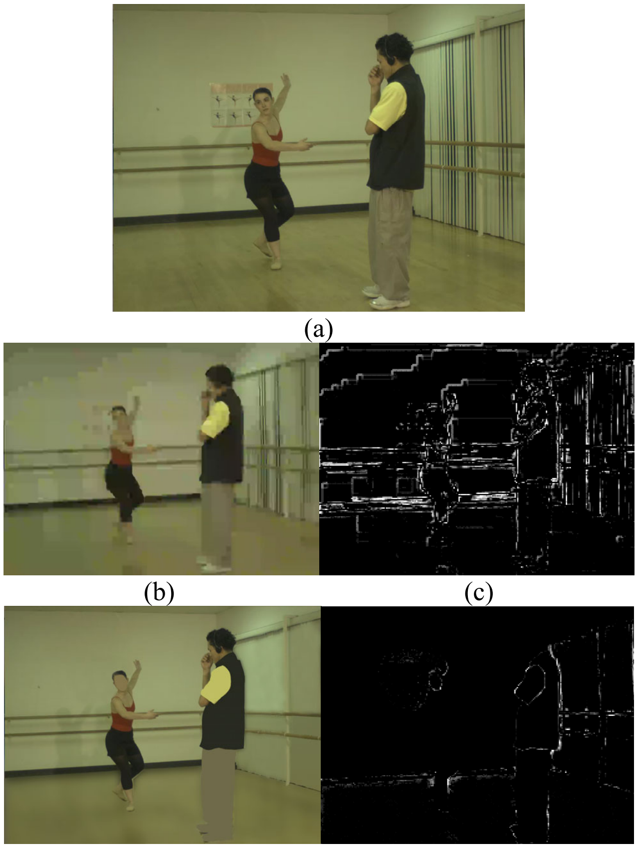
The proposed method can be used to assess the quality of the degraded image in the stereoscopic image pair effectively. Degraded area with high spatial frequencies are punished for its great influence on the 3D percept, while for smooth area, information loss is considered tolerable because it can be well recovered by the high quality view.

4. EXPERIMENTAL RESULTS

To show the superiority of our proposal to conventional metrics in assessing the image quality in asymmetric view coding, two experiments are conducted.

4.1. Quality assessment of degraded images

As discussed in Section 2, it is acceptable for image presented to one eye to lose some high spatial frequency information without degrading the 3D experience too much. On the contrary, introduction of error information with high spatial frequencies influences the 3D percept observably. Figure 3 gives examples of this effect and shows the capability of our proposed method to find out the areas which influence the 3D experience most.


(a)

(b) (c)

(d) (e)

Figure 3: Original image (a), degraded images (b and d) and degradation intensity maps (c and e)

In Figure 3, (a) is the first frame of View 1 from the ballet sequence, and (b) and (d) are two kinds of degraded images: (b) is reconstructed after H.264 encoding while (d) loses many details after image processing. The PSNR of image (b) is about 29 dB compared to image (d) with 26 dB, indicating the superiority of (b) in quality. However, in the case of 3D video, where a high quality image from adjacent view is also provided to form a 3D percept, the opposite result is reported. Because the coding artifacts and the mismatched edges containing high spatial frequencies in (b) influence the visual experience much more heavily.

In Figure 3, (c) and (e) show the areas where the degradation is more noticeable in (b) and (d). Notice that these highlighted areas point out the contours of the coding artifacts and the mismatched edges, and they are exactly the unpleasant but conspicuous areas. The judgment is based on the value of $\Delta E(r(X_d))$ for each 4×4 area introduced in Section 3 and the average values for (b) and (d) are 0.055 and 0.012, respectively, thus using our quality metric, the quality of image (b) is inferior to that of image (d) and this is in accordance with the subject evaluation result.

4.2. Comparison with subjective evaluation

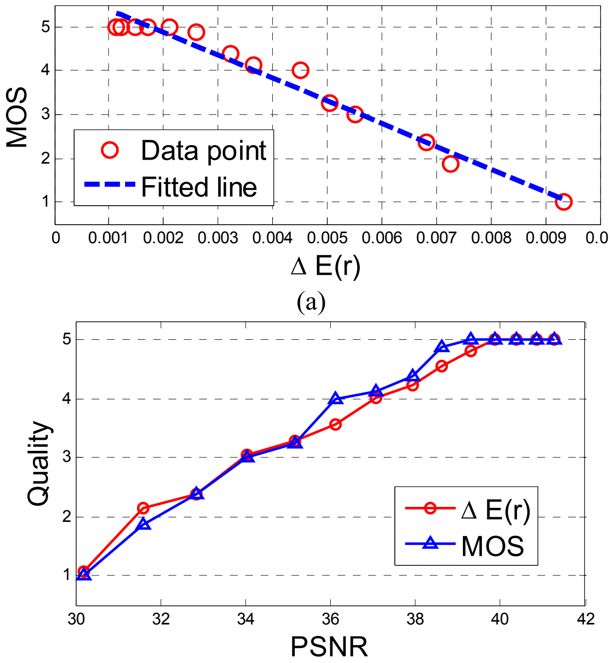
The sequence used for subjective evaluation is ballet. View 0 and View 1 were chosen to render stereoscopic images. View 0 was encoded with PSNR of 41.28 dB as the high quality view for right eye, while View 1 was encoded as the degraded view for the left eye with PSNR varied from 30.28 to 41.27 dB. The display resolution is 1024×768 and the video is displayed in a 20-inch stereo monitor.

The subjective evaluation was done by 12 participants who were asked to score the quality of 3D experience from 1 to 5. Figure 4(a) illustrates the correspondence between the Mean Opinion Score (MOS) and $\Delta E(r(X_d))$. Each data point corresponds to one PSNR value. It can be seen that the points fit well to the regression line, which indicates a high correlation between our proposed assessing method and the subjective quality evaluation.

Calculated regression equation (y=-518.2x+5.9) is used to convert $\Delta E(r(X_d))$ and the result is plotted together with MOS in Figure 4(b). Although it is very difficult to make rigorous comparison between theoretical result and subjective result, it can still be noticed that the curves fit well and have very similar variation trends at most segments.

5. CONCLUSION

This paper introduces a practical quality metric for stereo images in 3D video based on the proposed spatial frequency dominance model. This model is based on the ability of HVS that one view could dominate the other if it has obviously higher spatial frequencies in the corresponding areas. One term introduced in Section 3 is used to assess the effect of frequency dominance on 3D percept. Experimental results show that the proposed quality model corresponds well with the subjective evaluation. Further work includes integrating this model into video encoder to optimize the 3D video quality.



(a)

(b)

Figure 4: Results of subjective evaluation and proposed assessment method

6. REFERENCES

[1] H. Kalva, L. Christodoulou, L. Mayron, et al., “Challenges and opportunities in video coding for 3D TV”, *IEEE International Conference on Multimedia & Expo (ICME) 2006*, July 9-12, 2006, Toronto, Canada.

[2] P. Benzie, J. Watson, P. Surman, et al., “A survey of 3DTV displays: techniques and technologies,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1647–1658, 2007.

[3] A. Smolic, K.Mueller, N. Stefanoski, et al., “Coding algorithms for 3DTV—a survey,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1606–1620, 2007.

[4] H. Kalva, L. Christodoulou and B. Furht, “Evaluation of 3DTV Service Using Asymmetric View Coding Based on MPEG-2”, *3DTV Conference*, Kos Island, Greece, May 2007.

[5] Lew B. Stelmach, W. James Tam, “Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views”, *Signal Processing: Image Communication*, Vol. 14, pp. 111-117, 1998.