# Ego Motion Induced Visual Discomfort of Stereoscopic Video

Jongyoo Kim<sup>\*</sup> Kwanghyun Lee Taegeun Oh and Sanghoon Lee<sup>†</sup>

\*Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea.

E-mail: jongky@yonsei.ac.kr

<sup>†</sup>Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea.

E-mail: slee@yonsei.ac.kr

*Abstract*—When each video sequence is captured, an inappropriate camera motion should be one of crucial factors leading to visual discomfort and distortion. The well known symptom, visually induced motion sickness (VIMS) is caused by the illusion of self motion by perceiving the video with ego motion. In particular, for the stereoscopic 3D video, it can be easily observed that the viewers have dominantly feel much more severe symptoms of visual discomfort. In this paper, we analyze the ego motion of the stereoscopic video and predict the effects. We attempt a novel approach by exploiting the computer vision algorithm. We propose a novel method which can estimate the perceptual 3D ego motion from the stereoscopic video. Then we analyze the ego motion components to predict the visual discomfort of stereoscopic video.

## I. INTRODUCTION

Motion has been well known as one of the most challenging factors dealt with in the research of video quality assessment, because it is difficult to analyze the response of the HVS (human visual system) along the temporal activity of the video sequence. Apparently, it can be observed that fast motion due to a rapid camera movement induces motion blur in each frame and decrease visual quality. However, the human eyes have been well used to such quality degradation in common life so that sometime, viewers tend not to recognize the quality degradation due to the rapid camera motion at the subjective quality assessment. However, for 3-D stereoscopic video, the visual discomfort can be generally observed so that the viewers naturally avoid watching the 3D content by closing their eyes or by putting their fixation points out of the screen.

There have been a lot of researches on visual discomfort in the stereoscopic images. The confliction in accommodation and convergence control due to excessive disparity is the well known factor [1], [2]. Many researches have validated the effects of the disparity. However, there are only a few researches about psychophysical symptoms produced by motion factors in stereoscopic video. Some experiments have shown that the object motion in depth can induce the visual discomfort even in the zone of comfort area [3]. Still, little is known about relation of motion and visual discomfort. The usual motion factors in stereoscopic video can be categorized into two types: the object motion and the ego motion. This paper is focussing on the ego motion of the stereoscopic video.

In the virtual environment, there exist researches about motion sickness caused by the ego motion. When people are

exposed to motion, they may feel discomfort. The degree and the characteristics of the sick may differ for each human. However, it is common that the strong ego motion induce motion sickness. The phenomenon is called as visually induced motion sickness (VIMS) [4], [5], [6]. Though VIMS is known to be severer in the condition of large filed of view, it occur also in usual HDTVs. In addition, there is a research that the symptom is stronger in stereoscopic environment [7]. In general, VIMS is caused by the sensory conflict in the visual input and the vestibular input. While we are watching the video, only the visual input exists.

Based on the researches of VIMS, we attempt to quantify the effect of the camera movement on the visual discomfort by developing a visual metric for 3D stereoscopic video. The perceptual 3D ego motion estimation method is proposed. We obtain the movement of the stereo camera by employing the technique used for 3D reconstruction. The 3D reconstruction technique uses the image streams to estimate camera poses and the 3D structure [8]. The conventional 3D reconstruction requires the accurate calibration information of the camera. We propose an approximation method of the calibration information by analyzing the viewing geometry and the shooting geometry. The proposed methods reflect the depth perception while watching 3D TVs. This results in more flexible application to various stereoscopic videos, since we cannot know the calibration information only from the video. Utilizing the 3D ego motion via 3D reconstruction, we propose the metric that predict the visual discomfort.

In this paper, we propose a new application using the 3D reconstruction to predict the visual discomfort of 3D stereoscopic video when the camera movement occurs. We extend the conventional research to a stereoscopic environment. By comparing the shooting space and the viewing space, we use the approximated calibration information, which is more similar in human perception. Then we analyze the structure of each scene geometrically. Two factors are scored independently.

### II. 3D EGO MOTION ESTIMATION

The ego motion of the video is caused by the camera movement. In this section, we propose a novel perceptual 3D ego motion estimation method. The conventional 3D reconstruction technique is described first.



(b) shooting condition

Fig. 1. Geometry of the viewing and shooting condition

SfM (structure from motion) is a classical method to obtain the 3D scene. It starts from the geometry of two views. Using the epipolar geometry, it estimates the camera poses and 3D points of two images [8]. To expand to the multi-views, an iterative optimization is utilized. For each new image, we match the existing 3D points **X** and the new 2D feature points **x**. Then we find the camera matrix that minimizes the projection error. An accurate calibration information is needed to remove the projective ambiguity. The calibration matrix Kcontains the intrinsic parameters of the camera.

$$K = \begin{bmatrix} \alpha_x & p_x \\ & \alpha_y & p_y \\ & & 1 \end{bmatrix}$$
(1)

$$\alpha_x = \vec{f} \cdot m_x \tag{2}$$

$$m_x = \frac{Image \ width[pixel]}{CCD \ width[mm]} \tag{3}$$

 $\alpha_x$  and  $\alpha_y$  describe the focal length f in pixel unit along xand y axis.  $p_x$  and  $p_y$  describe the coordinate of the principal point which the image plane intersect the optical axis. In general case, we cannot know the calibration information from a stereoscopic video. To know parameters, we should know which camera has been used while shooting. Or we should operate the special calibration process using the same video camera.

However, in the respect of human perception, calibration information is not important. We want to know how human perceives the ego motion from the video not the ground-truth camera movement. While watching the video, the brain is stimulated from visual input without vestibular input. R. J. V. Bertin has shown that human can mispercept the trajectory from only with optical flow information [9]. We propose the alternation of the camera calibration information by using the geometry of eyes and the display. So that it uses the only information that human can obtain while watching the video.

Fig. 1 shows the viewing environment of stereoscopic video. When the two eyes see the object on the screen with disparity d, the convergence point is fixed on the point **X**. The convergence distance  $Z_d$  is the perceived distance by the human. We compare the condition with shooting environment. The stereo camera projects the 3D scene onto the left and right CCD sensors. The centers of two cameras are matched to the two eyes. The CCD sensor is corresponding to the display plane. As the triangulation reconstruct the 3D coordinate from the two image planes, the convergence reconstruct the 3D coordinate from the L-R images on the display. Based on this relationship, we can derive the components of the calibration matrices as following.

$$\alpha_x = Z_v \cdot m_x \tag{4}$$

(5)

$$\alpha_y = Z_v \cdot m_y \tag{5}$$
$$n^L = (Image width - t + m_z)/2 \tag{6}$$

$$p_x = (Image with - t_e \cdot m_x)/2$$

$$p_z^R = (Image width + t_e \cdot m_x)/2$$
(7)

$$p_{u} = (Image \ height)/2 \tag{8}$$

$$m_x = \frac{1}{Display \ width[mm]} \tag{9}$$

 $Z_v$  is the viewing distance and  $t_e$  is the inter-eye distance. Generally, inter-eye distance of adults is about 6.35cm. These parameters forms each left and right calibration matrix,  $K^L$ and  $K^R$ .

Now, we want to estimate the rotation and the translation of  $i^{th}$  camera. The  $j^{th}$  3D points  $\mathbf{X}_j$  are generated from the  $(i-1)^{th}$  frame by triangulation.  $\mathbf{x}_{ij}^L$  and  $\mathbf{x}_{ij}^R$  are corresponding 3D points for each left and right image. The cost function can be represented as following.

$$e_{ij}^{L} = D\left(P_{i}^{L}\left(\mathbf{r}_{i}, \mathbf{c}_{i}\right)\mathbf{X}_{j} - \mathbf{x}_{ij}^{L}\right)$$
(10)

$$e_{ij}^{R} = D\left(P_{i}^{R}\left(\mathbf{r}_{i}, \mathbf{c}_{i}\right)\mathbf{X}_{j} - \mathbf{x}_{ij}^{R}\right)$$
(11)

We should find  $\mathbf{r}_i$  and  $\mathbf{c}_i$  that minimizes both  $e_{ij}^L$  and  $e_{ij}^R$ .

$$\min_{\mathbf{r}_{i},\mathbf{c}_{i}}\sum_{j}D\left(e_{ij}^{L}+e_{ij}^{R}\right)$$
(12)

The problem can be solved by non-linear iterative optimization. The more detailed flow of 3D ego motion estimation is described in Algorithm 1. The estimated 3D ego motion is shown in Fig. 2. It shows the reconstructed camera path of the stereoscopic video. The result contains the error since the algorithm does not use the accurate calibration and only use the neighbor frame. However the result still shows the acceptable path and has sufficient information to predict the visual discomfort of the video.

## Algorithm 1 Perceptual ego motion estimation

**Require:**  $nFrames \ge 2$ Initialize  $K^L$  and  $K^R$ ; Initialize  $P_0^L$  and  $P_0^R$ ; Find and match features for  $I_0^L$  and  $I_0^R$ ;  $W_0 \leftarrow \{ \text{triangulated points from } I_0^L \text{ and } I_0^R \};$ for n = 1 to nFrames - 1 do Find and match features for  $I_n^L$  and  $I_n^R$ ; Find matches between  $I_n^L$  and  $W_{n-1}$ ; Find  $P_n^L$  that minimizes projection error  $e^L$  and  $e^R$ ; Generate  $P_n^R$  from  $P_n^L$ ;  $W_n \leftarrow \{ \text{triangulated points from } I_n^L \text{ and } I_n^R \};$ end for



Fig. 2. Reconstructed 3D camera motion

#### **III. ANALYSIS OF EGO MOTION**

Human can recognize the self motion only using the visual information. The visual motion field stimulate the neurons in the brain. Another important area in recognizing the self motion is the vestibular system. When our head moves, it provide the information of rotation and the acceleration. If we are walking, the visual input and the vestibular input are combined normally. However, in an artificial case like watching a stereoscopic video, we may feel VIMS by a sensory conflict. To quantize and analyze the effects of ego motion on visual discomfort, we define following three attributes.

- Point movement : Perceived motion along the path
- Rotation movement : Variation of optical axis
- Roll movement : Variation of gravity direction

We have obtained the camera movements in section II. Each camera movement contains 6 D.O.F motion: 3 translation and 3 rotation. The three attributes are characterized by these movements.

The point movement represent the camera's translation. This is related to the velocity of the camera. When the velocity increases, the vection will also increase. The strong vection is close relationship with VIMS. We model the point movement as the distance between cameras.

$$d_i = \left\| \mathbf{c}_i - \mathbf{c}_{i-1} \right\| \tag{13}$$

The rotation movement shows the change of the optical axis. It reflects the yaw and pitch rotation of the camera. We assume that the unit vector  $\mathbf{u}$  is parallel to the optical axis.  $\mathbf{u} = (0, 0, 1)^T$ . Since the  $\mathbf{u}$  is in its own coordinate system, we should transform it into the world coordinate system. If the *i*<sup>th</sup> camera matrix has a form of  $P_i = [R_i | \mathbf{t_i}]$ , the optical vector of the *i*<sup>th</sup> frame is converted to the world coordinate by following:  $\mathbf{v}_i = R_i^T \mathbf{u}$ .  $R_i$  is the rotation matrix and  $\mathbf{t_i}$  is the translation vector. Using optical vectors of *i*<sup>th</sup> and  $(i-1)^{th}$  frames, we can calculate the angle variation.

$$\theta_i = \cos^{-1} \frac{\mathbf{v}_i \cdot \mathbf{v}_{i-1}}{|\mathbf{v}_i| |\mathbf{v}_{i-1}|} \tag{14}$$

Finally, we consider the roll movement containing the direction of gravity. We choose the reference vectors in the world coordinate as  $\mathbf{r}_y = (0, 1, 0)^T$  and  $\mathbf{r}_z = (0, 0, 1)^T$ . The unit vectors with direction of *y*-axis and *z*-axis of the *i*<sup>th</sup> camera are converted into world coordinate.

$$\mathbf{r}'_{yi} = R_i^T \mathbf{r}_y \quad \mathbf{r}'_{zi} = R_i^T \mathbf{r}_z \tag{15}$$

The roll angle can be represented by the difference between  $\mathbf{r}_{yi}$  and the plane which is made with  $\mathbf{r}'_{zi}$  and  $\mathbf{r}_y$ . The plane is made of z-axis of  $i^{th}$  frame and y-axis of world coordinate. The normal vector of the plane is obtained by

$$\mathbf{n} = [\mathbf{r}_y]_{\times} \mathbf{r}'_{zi} \tag{16}$$

Then the roll angle can be calculated. And we obtain the roll movement by subtracting the roll angles of of  $i^{th}$  and  $(i-1)^{th}$  frames.

$$\omega_i = \frac{\pi}{2} - \cos^{-1} \frac{\mathbf{r'}_{yi} \cdot \mathbf{n}}{|\mathbf{r'}_{vi}||\mathbf{n}|} \tag{17}$$

$$\phi_i = \omega_i - \omega_{i-1} \tag{18}$$

The three attributes of the 3D ego motion are quantized. The point movement is represented by  $d_i$ , the rotation movement is represented by  $\theta_i$ , and the roll movement is represented by  $\phi_i$ .

## IV. EXPERIMENTAL RESULTS

The subjective test has been processed to analyze the effects of the ego motion. A subjective should decide immediately since the scene is changing continuously. Therefore we exploited the discrete 4 scale score to make decision easier. Three subjects were asked to record the score. The subjects are instructed to decide the score properly. The higher score means that the video induces more discomfort. Since the continuous subjective scores are dependent on the response times of the subjects, the recorded scores are delayed with 0.5 seconds. We have shot the two test sequences with a hand-held stereo camera. The test sequence includes the various camera movements. The stimuli 1 was shot with almost fixed center. Instead, the rotation and roll movements were major. In case of stimuli 2, translation, yaw and pitch were major. The lengths of the sequences were 30 seconds. Every  $5^{th}$  frame were used to estimate the 3D ego motion.

Fig. 3 (a) presents the calculated movements. The absolute values were used since the score should be positive for the rotation and the roll. In the stimuli 1, it has two significant roll movement around  $50^{th}$  to  $80^{th}$  frame and  $140^{th}$  to  $170^{th}$ frame. In the stimuli 2, it has lots of point and rotation movement. Fig. 3 (b) shows the averaged subjective scores with 90% confidence interval, and the predicted discomfort scores which were calculated by simple linear summation of point, rotation and roll movement scores. There are some areas that the prediction is not correct. In the continuous subjective test, it is hard to decide the score with quick response if the discomfort is sudden and short. Therefore some peaks are mismatched in the graph. However we can see that the important peaks are sufficiently predicted. For the stimuli 1, two roll movements affected subjective discomfort critically. These areas are well predicted in graph. Also, the overall shapes are similar between subjective discomfort score and the predicted score. In the stimuli 2, it has complex movement in the first half. Both subjective and predicted scores show higher score on these areas.

## V. CONCLUSION

The VIMS has been the significant issue in a video with ego motion. The phenomenon is even more important in a stereoscopic video. We analyzed the ego motion components of the stereoscopic video and showed how to quantize the ego motion. The novel method to estimate the perceptual ego motion was utilized. The proposed algorithm can derive the 6 D.O.F ego motion in the respect of human perception. Though the proposed ego motion estimation is not very precise with comparison to ground truth, its reconstructed results have sufficient information to predict the discomfort of the video. We derived the three simple attributes from the sequence of reconstructed camera matrices. By combining the attributes, critical regions of the video can be predicted well.

#### ACKNOWLEDGMENT

This work was supported by the IT R&D program of MSIP/KEIT [10044844, Development of ODM-interactive Software Technology supporting Live-Virtual Soldier Exercises]

#### REFERENCES

- M. Wöpking, "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*, vol. 3, no. 3, pp. 101-103, 1995.
- [2] M. Lambooij, W. Ijsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A Review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 0302011-03020114, 2009.





(b) Subjective test score and predicted score

Fig. 3. Score of test videos

- [3] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, no. 4, pp. 141-150, 2004.
- [4] J. E. Bos, W. Bles, and E. L. Groen, "A theory on visually induced motion sickness," *Displays*, vol. 29, no. 2, pp. 47-57, 2008.
- [5] R. So, and H. Ujike, "Visually induced motion sickness, visual stress and photosensitive epileptic seizures: what do they have in common?-preface to the special issue," *Applied ergonomics*, vol. 41, no. 4, pp. 491, 2010.
- [6] C. Diels, and P. A. Howarth, "Visually induced motion sickness: Singleversus dual-axis motion," *Displays*, vol. 32, no. 4, pp. 175-180, 2011.
- [7] J. Kuze, and K. Ukai, "Subjective evaluation of visual fatigue caused by motion images," *Displays*, vol. 29, no. 2, pp. 159-166, 2008.
- [8] R. Hartley, and A. Zisserman, "Multiple view geometry in computer vision", Cambridge Univ Press, 2000.
- [9] R. J. V. Bertin, I. Israël, and M. Lappe. "Perception of two-dimensional, simulated ego-motion trajectories from optic flow," *Vision Research*. 40.21, 2000.