INTEGRATING A COMPUTATIONAL MODEL OF OPTICAL FLOW INTO THE CYBERSICKNESS DOSE VALUE PREDICTION MODEL

Jennifer T.T. Ji*, Felix W.K. Lor, and Richard H.Y. So
Computational Ergonomics and Simulation Laboratory
Department of Industrial Engineering and Engineering Management
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong SAR
** ieemjtt@ust.hk*

This paper describes a computational model of optical flow and demonstrates how it can be used to estimate the navigation velocity of the user's viewpoint during a virtual reality (VR) simulation. The model takes as input a VR simulation video sequence and outputs the estimated average navigation velocity. When this estimated velocity is integrated into the published cybersickness dose value model (CSDV: So, 1999), the severity level of the cybersickness associated with the VR simulation can be predicted. Results of two simulation studies indicated that the average navigation velocities in lateral and vertical axes can be estimated to within 10% error. The simulations used a virtual environment (VE) with a known single depth. Discussion of an extension of the computational model to accommodate VEs with unknown multiple depths is included. The reported model is consistent with the biological neuro-pathway of the perception of visual motion velocities in humans. The benefits associated with this consistency are discussed.

INTRODUCTION

A cybersickness dose value (CSDV) model was proposed by So (1999) to predict the severity levels of cybersickness. The equation for calculating CSDV is:

$$CSDV = \int_0^T (SF \times V) dt \,. \quad (1)$$

As shown in Equation 1, CSDV is an integral of the product of two components (SF and V) over the duration (T) of the virtual reality (VR) simulation. The SF (spatial frequency) metric measures the average scene complexity of a virtual environment (VE), while the V (scene movement velocity relative to the viewer) metric measures the average navigation velocity of the user's viewpoint during a VR simulation (So *et al.,* 2001). Yuen *et al.* (2002) reported that the CSDV model can explain over 70% of the data variances in nausea ratings obtained from over 200 participants exposed to various VR simulations. In all previous studies concerning CSDV, the scene movement velocity relative to the viewer (V) which is the same as the average navigation velocity was assumed to be known or could easily be measured. This assumption is not valid in situations in which the VR simulation is generated by off-the-shelf software such as video games. In order to eliminate this assumption, a two-stage computational model is proposed here to estimate the navigation velocity from video sequences of a VR simulation when the velocity is not known. The integration of the proposed velocity estimation model into the CSDV model has two additional advantages. Firstly, after integration, CSDV can be directly estimated from the video sequences of any VR simulation because the other two components of CSDV (SF and T) have already been shown to be readily extracted from the video sequence (So *et al.*, 2001). Secondly, the two-stage velocity estimation model follows the biological mechanism of human perception of visual motion from a sequence of images. The former increases the ease of application of CSDV while the latter brings us a step closer to modeling the true mechanism of cybersickness.

OVERVIEW OF THE TWO-STAGE COMPUTATIONAL MODEL

The Human Visual-motion Processing Pathway and the Two-stage Computational Model

Figure 1 depicts the consistency between the two-stage computational model and the neuro-pathway of visual motion processing in humans.

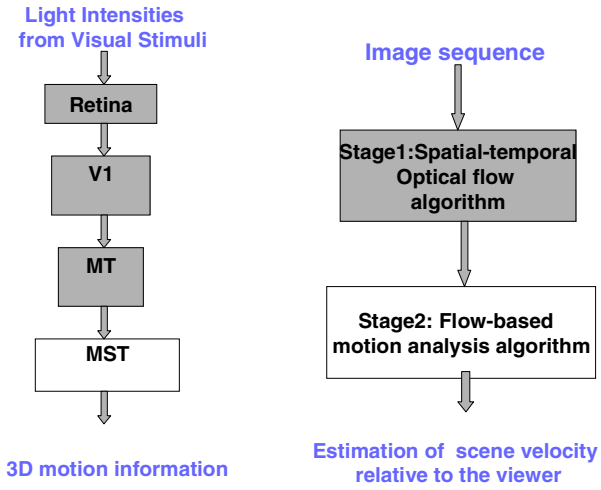

Figure 1. The two-stage computational model to estimate scene velocity relative to the viewer (i.e., navigation velocity) is shown on the right and the human visual motion processing pathway is shown on the left. The optical flow calculating module (stage 1) corresponds to the motion detection functions of retina, the primary visual cortex (V1), and the middle temporal area (MT) of the visual cortex while the optical flow analyzing module (stage 2) corresponds to the motion detection function of the medial superior temporal area (MST) of the visual cortex.

Figure 1 shows that light intensity information from a visual scene propagates from the retina (e.g., magno ganglion cells) to the V1 area of the brain (for spatial processing of visual information) before passing on to the MT area of the visual cortex for temporal processing. Outputs from the MT area are then passed on to the MST area for estimating the perceived motion velocity of the visual scene. This visual motion detection pathway has been thoroughly studied by neuroscientists and the understanding is that the visual-motion perception proceeds in two stages: firstly, the retina-to-V1-to-MT pathway results in some localized optical flow information. Secondly, in the MST area, the perceived motion velocities are estimated by extracting independent motion vector components that are representatives of the localized optical flow information. (Heeger, 1987; Jabri *et al.*, 2000, So and Lor, 2004). Thanks to researchers in the field of computational neuro-sciences, open-source biologically plausible computational algorithms are available to simulate the generation of localized optical flow information from video sequences. These algorithms use 3D spatial-temporal Gabor filters to simulate the processing of visual information from the retina to the MT area via the V1 area (Abramoff, 2000; Fleet and Jepson, 1990; Heeger, 1987). In this study, the spatial-temporal algorithm reported by Fleet and Jepson (1990) was used in the first stage. For the second stage, a set of computational rules was used to simulate the motion detection function of the MST area. In the rest of the paper, two simulations applying the two-stage computational model to estimate the average navigation velocity from a video clip of a VR simulation are presented. Since the computational model involves much processing of optical flow information, the fundamentals of how optical flow vectors can be used to represent visual motion information are reviewed in the next section.

Optical Flow Vectors

Optical flow is defined as "the apparent motion of brightness patterns in an image sequence" (Horn and Schunk, 1993). In a visual environment with stationary uniform light sources of constant brightness, the motion of objects with Lambertian (i.e., diffused) surfaces cause spatial and temporal changes in the intensities of the light projected on the human retina. These changes in light intensities are then transformed into localized optical flow vectors distributed throughout the visual field-of-view. The current biological understanding is that the visual field-of-view is divided into many localized areas called 'receptive fields' and, within each receptive field, optical flow vector information is constructed in the MT area of the brain based on the spatial and temporal changes of the perceived light intensities. Each optical flow vector can be described as a vector sum of two components: u and v. The u component represents the horizontal component of the optical flow vector and the v component represents the vertical component of the optical flow vector. The units of u and v are pixels / frame.

As explained above, these localized optical flow vectors are then processed in the MST area to form one or more independent motion flow vector(s), which represent the perceived visual motion information. However, in some situations, optical flow vectors lead to an underestimation of the visual motion information. This occurs when there is lack of spatial varying texture along the direction of the visual movement. In such a case, the visual movement results in inadequate spatial and temporal changes of light intensities that, in turn, lead to an underestimation of visual motion information. This situation has been referred to as the "aperture problem" (Horn and Schunk, 1993).

SIMULATION EXPERIMENTS

Objectives and Hypothesis

Two simulation experiments were conducted to estimate the average navigation velocity from video clips taken from two VR simulations. The actual navigation velocity in the two VR simulations was 3m/s in the lateral axis and 3m/s in the vertical axis. It was hypothesized that (i) the estimated navigation velocities extracted by the two-stage computational model would be along the same orientation as the pre-set navigation velocity (H1) and (ii) the magnitudes of the estimated velocities would be close to the pre-set navigation velocities (H2). The latter hypothesis (H2) was based on the assumption that the VE has enough textured

scene content to enable the estimation of navigation velocity.

Experimental Procedure, Simulation Settings and Software Implementation

Consistent with the two-stage computational model depicted in Figure 1, a video sequence of a VR simulation was used as the input to the optical flow calculating module in Stage 1. As explained in the overview section, the spatial temporal Gabor filtering algorithm by Fleet and Jepson (1990) was used in Stage 1. In this study, the open-source implementation of the Fleet and Jepson algorithm by Abramoff (2000) under the 'FlowJ' simulation environment (http://bij.isi.uu.nl/flowj.htm) was used. The VR simulation was rendered using the WorldToolKit (Sense8 Incorporated, San Rafael, California) software running on a Silicon Graphics ONXY Ⅱ (Silicon Graphics, Mountain View, CA) workstation. The architecture of the VE is illustrated in Figure 2. This VE has been used in previous studies of cybersickness (Chen and So, 2004). The VE is a large in-door acoustic chamber. The authors admit that this VE is not very exciting but it provides a highly controllable environment and, in this case, a single depth environment because the viewpoint only translates in front of the front wall along either the lateral and vertical axes.

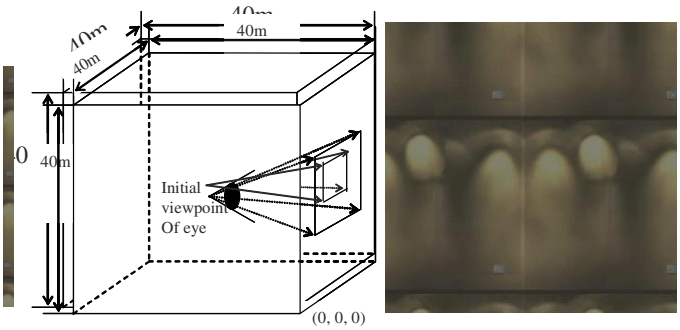

Figure 2. The architecture of the VE (a 40m×40m×40m empty room) is shown on the left and a snapshot of what the viewer saw from the initial viewpoint is shown on the right. The wall is made up of repeated panels. The oval shaped patterns on the panels were formed from illumination from wall-mounted lamps. These patterns were captured on a digital camera and pasted as a wall-paper mapping inside this VE. Inside the VE, uniform ambient lighting was used.

In order to generate the movie clips for the two simulations, the navigation paths in the VEs had been pre-determined to be moving left at a constant velocity of 3m/s and moving up at a constant velocity of 3m/s. The frame rate of the two videos was 30 frames/second and each video clip contained 21 frames (i.e., a duration of 0.7 seconds). The choice of 21 frames was recommended by Abramoff (2000) who tested the performance of Fleet and Jepson's Gabor filtering algorithm and found that the algorithm is more robust with an input video sequence of 21 image frames. The resolution of each frame was 290 pixels by 290 pixels.

RESULTS

Optical Flow Maps: Outputs after Stage 1 Processing

The outputs of the Stage 1 optical flow module are illustrated in Figure 3.

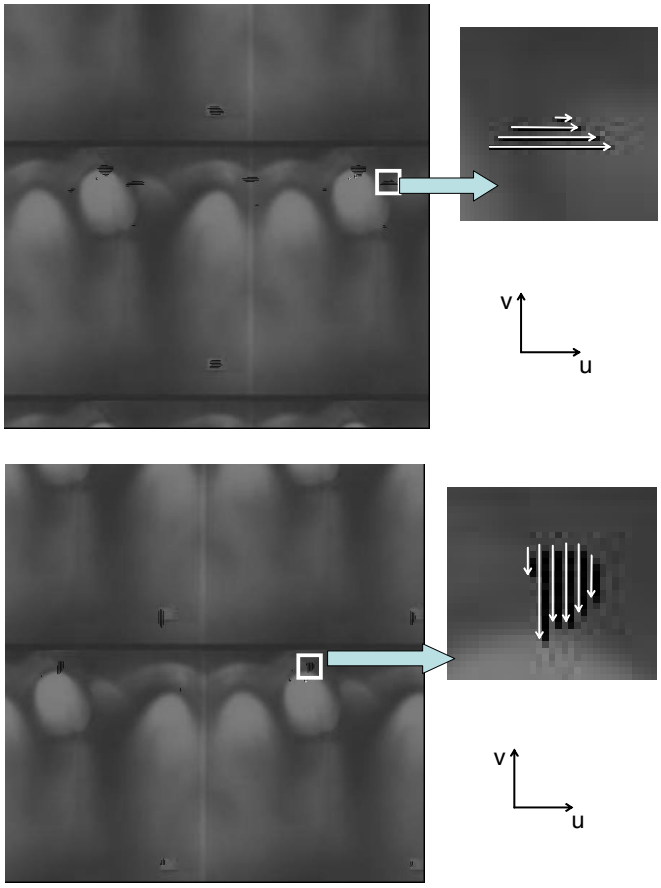

Figure 3. The optical flow maps generated after Stage 1 of the computational model are shown above. The maps on the top are for the simulation with lateral velocity and the maps on the bottom are for the simulation with vertical velocity. Both maps have been overlaid on top of a snapshot of the VE for illustration purposes. The directions of the components u and v of an optical flow vector are also shown as a reference. In the enlarged optical flow maps, each optical flow vector has been highlighted in white for ease of visualization (the original optical flow maps were directly generated by the FlowJ program).

The sparseness of the localized optical flow vectors can easily be observed in Figure 3. There are large areas in the visual field where no optical flow vector was generated. Further inspection of the figure indicates that the visual areas in which no optical flow vectors were generated were associated with lower contrast areas. This suggests that the localized spatial and temporal (over the 21 frames) changes of light intensity in the pixels in those areas were small. In

Fleet and Jepson's spatial temporal Gabor filtering algorithm, when the spatial and temporal changes in pixel intensity were small, the outputs were deemed to be unreliable and the algorithm removed any optical flow vectors that were below a certain reliability threshold (Fleet and Jepson, 1990). The threshold setting in this study was made according to settings recommended by Abramoff (2000).

Reduction and Processing of Optical Flow Vector Data: Part of Stage 2 of the Computational Model

From the two optical flow maps shown in Figure 3, there are 301 optical flow vectors in the simulation with lateral velocity, and 160 optical flow vectors in the simulation with vertical velocity. Two computational rules were applied in this stage. Firstly, in order to extract the unknown independent visual motion components from the optical flow vectors, independent component analyses should be used. However, in this study, because the visual motion is in a single direction, the dominant orientation was readily extracted by just checking the distribution of the orientation data. For the simulation with lateral velocity, 100% optical flow vectors have 0 ± 10.0 degree orientation (i.e., all pointing to the right); for vertical movement, 98.8% optical flow vectors have 270 ± 10.0 degree orientation (i.e., all pointing downward). These overwhelming percentages suggest that hypothesis H1 is supported.

Secondly, the optical flow vector with the maximum magnitude (i.e., sqrt $(u^2 + v^2)$) value among all the vectors in the dominant orientation cluster is chosen as the best approximation of the optical flow vector representing the navigation velocity. The reason for choosing the maximum value is that, with the aperture problem, the optical flow vectors tend to underestimate the visual flow information when there is insufficient texture in the picture (see 'Optical Flow Vectors'). Therefore, it is reasonable to choose the maximum value to estimate the visual motion velocity. As shown in Figure 3, the optical flow vectors with larger magnitudes are located in areas with high contrast. In the simulations with lateral and vertical velocities, the maximum magnitude of the optical flow vectors are 3.79 pixels/frame and 3.74 pixels/frame, respectively.

Recover the Navigation Velocity from the Maximum Optical Flow Vector: Part of Stage 2 of the Computational Model

The units of the optical flow vectors are in pixels/frame. In order to scale these values back to meters/second, information on the image depth, the focal length of the perspective projection, the image resolution, and the frame rate is needed. After scaling, the estimated navigation velocity becomes 3.25m/s in the lateral direction and 3.21m/s in the vertical direction. Compared with the preset navigation velocity of 3m/s in both the lateral and vertical directions, the estimation error is 8% for the lateral velocity and 7% for the vertical velocity. Although the results support the second hypothesis (H2), the authors were surprised by the overestimation of the navigation velocities. Possible reasons for the overestimation might have been caused by truncation and rounding errors in the optical flow calculations.

DISCUSSION

Depth Estimation and Motion Segmentation

As explained in the results section, the depth information for objects inside a VE is required to recover the visual motion velocity from the optical flow vector. In this study, the VE has a known single depth. For VEs with unknown multiple depths, the depth information needs to be estimated from the video sequences. Fortunately, there are already standard techniques for estimating depth information from video sequences. Examples include the correlation-based method (Huang and Netravali, 1994) and the parallax-based method (Vieville and Faugeras, 1995). As the depth information may vary in different areas of the visual field, a technique called motion segmentation can be used to classify the visual field into different segments for visual motion estimation. Again, algorithms to implement this technique are available.

Other Optical Flow Algorithms

Besides the spatial-temporal Gabor filtering method adopted in this work, there exist numerous computational algorithms for estimating optical flow, such as gradient-based methods and correlation-based methods. It has been reported that both the gradient-based methods (e.g., Luscas and Kanade, 1981) and the correlation-based methods (e.g., Singh, 1992) are much more computationally efficient than the spatial-temporal methods (e.g., Fleet and Jepson, 1990) (Abramoff, 2000). However, spatial-temporal algorithms are more consistent with the biological neuro-pathway of human visual motion perception. As a result, the spatial-temporal method has been used in this study.

Benefits of the Proposed Two-stage Computational Model

With the reported two-stage computational model, it is now possible to estimate the average navigation velocity from a video clip of a VR simulation. This has important implications. So *et al.* (2001) reported that the scene complexity component (SF) of the CSDV model can be calculated from image sequences of a VR simulation. In fact, the calculation algorithm has been made available on www.cybersickness.org and users can upload sequences of images to the Web server to calculate the corresponding values of SF. With the successful implementation of the two-stage computational model for velocity estimation, all three components of CSDV can now be estimated from a video sequence of a VR simulation. This means that if a video of any VR simulation is available, it is now possible to

estimate the CSDV value associated with the VR simulation and, mostly importantly, the predicted severity levels of cybersickness.

Moreover, as explained in this paper, the two-stage computational model is consistent with the biological process of human perception of navigation velocity during a VR simulation. This implies that in VR simulations in which the viewers will underestimate the navigation velocity because there is a lack of contrasting objects in the VE, the optical flow map as produced by the first stage of the computational model should truly reflect the perceived velocity, which is an underestimation of the actual navigation velocity. In the reported simulation studies, this advantage was eliminated because only the optical flow vector with the maximum magnitude was used in estimating the navigation velocity. Enhancements to the selection rules for processing optical flow vectors are desirable.

CONCLUSIONS, LIMITATIONS AND FUTURE WORKS

A two-stage computational model involving optical flow calculation and analysis modules, has been shown to estimate the navigation velocities of two VR simulations captured on video clips. This demonstrates the possibility of extracting the CSDV values for any VR simulation that has been captured as a video.

The authors acknowledge that the reported simulation studies have some limitations. The VE used is static with no independent moving object. Only navigation velocities in the lateral and vertical directions have been tested and the VEs had a known and single depth. As explained in the discussion, depth estimation algorithms and motion segmentation algorithms can be added to the two-stage computational model so that it can also work for VEs with unknown and multiple depths. Also, if independent component analyses were used in the second stage of the model, it could deal with VR simulations with navigation in multiple directions. The development of the two-stage computational model is continuing and the authors will present an enhanced model during the conference.

ACKNOWLEDGEMENT

The authors are grateful for the support of the Hong Kong Research Grants Council through the competitive earmarked grant HKUST6128/03E and DAG02/03.EG47. The authors are also grateful for the constructive comments from the two reviewers on the draft of this manuscript.

REFERENCES

Abramoff, M.D., Niessen W.J., and Viergever, M.A. (2000). Objective Quantification of the Motion of Soft Tissues in the Orbit. IEEE Transactions on medical imaging, 19, 986-995.

Chen, R.W. and So, R.H.Y. (2004) Effects of translational navigation on simulator sickness in a virtual environment. Submitted.

Heeger. D.J. (1987) Model for the Extraction of Image Flow. Journal of the Optical Society of America. A, 4, 1455-1471.

Horn, Berthold K.P. and Schunck, B.G.. (1993) Determining Optical Flow": a Retrospective, Artificial Intelligence, 59, 81-87.

Huang, T.S. and Netravali, A.N. (1994) Motion and Structure from Feature Correspondences: A Review. Proceedings of the IEEE, 82, 2, 252-268.

Jabri, M.A., Park, K.Y., Lee, S.Y., Sejnowski, T.J. (2000) Properties of Independent Components of Self-Motion Optical flow. 30th IEEE International Symposium on Multiple-Valued Logic.

Lucas, B. and Kanade, T. (1981) An iterative image registration technique with an application to stereo vision. Proceedings of 7th International Joint Conference on Artificial Intelligence. Vancouver, BC, Canada, 24-28 August, 674-679.

Singh, A. (1992) Image-flow computation: an estimation-theoretic frame-work. Comput. Vis. Graphics Image Understanding, 56, 152-177.

So, R. H. Y. (1999) The search for a cybersickness dose value. Proceeding of the 8th International. Conference on Human-Computer Interaction. Munich, Germany, 152-156.

So, R. H. Y., Ho, A., & Lo, W. T. (2001) A metric to quantify virtual scene movement for the study of cybersickness: definition, implementation, and verification, Presence, 10, 193-215.

So, R.H.Y. and Lor, F.W.K. (2004) Computational Ergonomics-a Possible Extension of Computational Neuroscience? Definitions, Potential Benefits, and a Case Study on Cybersickness. Contemporary Ergonomics, 405-409.

Vieville, T. and Faugeras, O.D. (1995) Motion Analysis with a Camera with Unknown, and Possibly Varying Intrinsic Parameters. Proceedings of the 5th Internationall Conference on Computer Vision, 750-756.

Yuen, S.L., Chen, R.W. and So, R.H.Y. (2002) A progress report on the quest for a cybersickness dose value. Proc. of 46th HFES Annual Meeting, Baltimore, MD, 2189-2192.