A Model for Visio-Haptic Attention for Efficient Resource Allocation in Multimodal Environments

Priyamvada Tripathi, Kanav Kahol, Anusha Sridaran, and Sethuraman Panchanathan

Center for Cognitive Ubiquitous Computing, Arizona State University, Tempe, AZ, USA
{priyamvada.tripathi, Kanav.Kahol, Anusha.Sridaran, panch}@asu.edu

Abstract. Sequences of visual and haptic exploration were obtained on surfaces of different curvature from human subjects. We then extracted regions of interest (ROI) from the data as a function of number of times a subject fixated on a certain location on object and amount of time spent on such each location. Simple models like a plane, cone, cylinder, paraboloid, hyperboloid, ellipsoid, simple-saddle and a monkey-saddle were generated. Gaussian curvature representation of each point on all the surfaces was pre-computed. The surfaces have been previously tested for haptic and visual realism and distinctness by human subjects in a separate experiment. Both visual and haptic rendering were subsequently used for exploration by human subjects to study whether there is a similarity between the visual ROI and haptic ROIs. Additionally, we wanted to see if there is a correlation between curvature values and the ROIs thus obtained. A multiple regression model was further developed to see if this data can be used to predict the visual exploration path using haptic curvature saliency measures.

Keywords: Vision, Haptics, Eye movements, Attention, Saliency, Regions of Interest.

1 Introduction

Humans interact with their environment through several modalities: vision, touch, audition, gustatory, olfactory. Among these modalities, vision and touch play an important role in gathering information from the surrounding environment and store representation of the environmental objects. This is so because most of the everyday objects such as cups, glasses, bowls, etc., have two essential modal components: a visual component and a tactile component. These two components overlap extensively and interact synergistically in order to produce coherent representations of these object and to enable humans to make consistent responses to them (Spence & Driver, 2004). For example, a purely visual component such as color may not interact with tactile information at all but a 'composite' feature such as texture will need

D.D. Schmorrow, L.M. Reeves (Eds.): Augmented Cognition, HCII 2007, LNAI 4565, pp. 329–336, 2007.
© Springer-Verlag Berlin Heidelberg 2007

contribution of both tactile and visual system to arrive at a coherent judgment. This distribution can be visualized as a Venn diagram as shown in Fig. 1.

Since it is computationally infeasible to enumerate all the objects with all their pertaining characteristics, attention serves as a prerequisite to any computations that may be performed on such multimodal (in this paper multimodality will only mean visuo-tactile) objects (Ballard, 1986; Tsotsos, 1990, 1991). An image or an object can thus be described in terms of its saliency or regions of interest (ROI). Some attributes in an object are assumed to be more salient than others and tend to attract more attention (tactile or visual) of the observer. Similarly, there are some regions of the object that are more focused upon than others in order to gain the representation of the object. Thus, selective attention processes serve an important role in exploration of novel everyday objects. Figure 2 shows how attention can act as 'moderator' in multimodal explorations by humans.

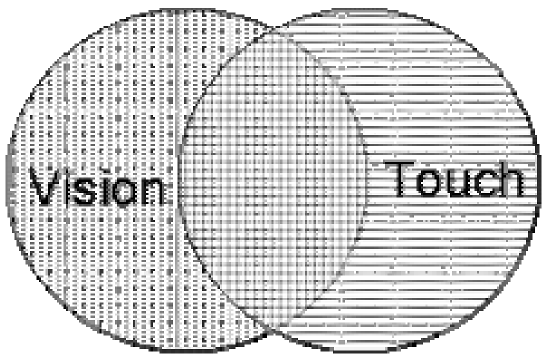

Fig. 1. Venn diagram showing distribution of features in a visuo-tactile object. Vision and Touch can be visualized as overlapping modalities wherein they convey information about purely visual (e.g. color) or tactual features (e.g. hardness) and also of those that combine vision and touch for coherent representation (e.g. texture).

In this paper, we asked the question whether purely spatio-temporal bottom-up exploration of a 3D geometrical feature such as Gaussian curvature will yield any correlation in vision and touch. If a correlation exists, then it paves way for optimal allocation of computational resources in visual and haptic interfaces due to the similarity in ROIs and overlap of perceptual load requirements. Furthermore, a predictive sequence can be developed that uses information from one modality to maximize the performance in other modality.

2 Background and Related Work

Human perceptual system comprises of several modality sub-systems that dictate our interactions with our environment. This distribution in order to extract information from the environment employs mainly two attentional mechanisms to handle the 'bottleneck' of the perceptual system: (a) bottom-up saliency of the attributes or features and (b) top-down or goal driven template search. Bottom-up saliency measures for images and objects have been studied extensively (Wolfe, 1998). In 2001, Itti et al. (Itti, Koch, & Neibur, 1998) constructed a biologically plausible model of visual selection attention that uses 'saliency maps'. Saliency maps, also called conspicuity maps, are explicit representation of saliency values for particular features such as color, luminous intensity, orientation. These features are traditionally assumed to constitute the set of 'basic features' for visual selective attention models. These models scan the locations in the images in order of decreasing saliency. Other features that have been explored for search mechanisms in vision are size, spatial frequency, scale, motion, shape, and depth cues (Pashler, 1998). Curvature has been explored as a possible candidate by several researchers in visual search task where curved object is to be identified from a from several distracters without any curvature (Brown, Weisstein, & May, 1992; Cheal & Lyon, 1992; Sun & Perona, 1993). These studies, unfortunately, are very limited in their exploration. In our point of view, curvature is a 3D concept and needs to be rendered in a 3D environment for results with reasonable perceptual accuracy. Curvature has occasionally been confused with

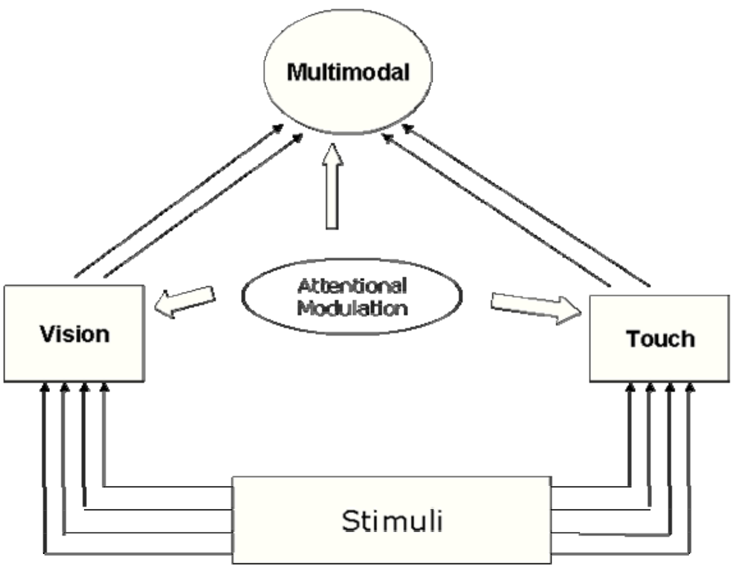

Fig. 2. Attention as a moderator in human perceptual system

change in orientation or illumination changes (Riggs, 1973). Additionally, no studies have actually explored the effect of curvature of haptic attention. Most contemporary studies point out that attention is most likely a post-sensory phenomenon (Mesulam, 1998; Posner & Dehaene, 1994),hence, it is needed that exploration be done of these 3D parameters on both haptics as well as vision.

Ouerhani and Hugli (Ouerhani & Hugli, 2000) explored the usefulness of depth (defined as distance from camera to the object) as a feature in saliency maps in a neurally plausible model of attention and pointed out that the depth contributed immensely in the focus of visual attention since the "depth enhanced model detects depth locations which stand out from their surrounding". Unfortunately, they did not include curvature in the model. Some studies (Jenmalm, Birznieks, Goodwin, & Johansson) have indicated that human subjects automatically adjusted the grasping force to correspond to the amount of curvature present on the object. This seems to signify that there is a very likely correlation between visual and haptic assessment of curvature and also, attentional exploration is sometimes dictated by curvature values of the object.

Thus, it is possible that curvature is an important signifier of the object exploration. In order to study its bottom-up saliency pattern in human subject, we conducted experiments and related the measures so conducted with a multiple regression model. Stimuli were carefully rendered using varying Gaussian curvature values and tested for distinctness and realism (Sridaran, Hansford, Kahol, & Panchanathan, 2007). In order to remove all other features from consideration, we opted for stimuli that will have only curvature as its parameter for exploration.

3 Methodology

A repeated measure design was used in the experiment to extract exploration strategy of humans in (1) touch and (2) vision in an undirected or free exploration condition. The goal was to extract the regions of interest (ROI) for vision and touch in separate explorations and correlate these two. In addition, we wanted to see if any specific curvature values correlated well with any ROIs.

3.1 Experiment 1: Haptic Exploration

Participants. 5 sighted right handed individuals (2 males, 3 females) participated in the experiment. The average age of participants was 27.8 years with a spread of 10.2 years. No subjects had any motor impairment. All subjected had normal or corrected to normal vision.

Material. Simple models like a plane, cone, cylinder, paraboloid, hyperboloid, ellipsoid, simple-saddle and a monkey-saddle were generated for the system. The Gaussian curvature representation of each point on all the surfaces was pre-computed. The surfaces rendered to the user were haptically parameterized according to their Gaussian curvatures in the haptic environment. The models and the corresponding Gaussian curvature signs are listed below:

- Planar Surface : plane, $K = 0$
- Parabolic Surface: Cylinder and Cone: $K = 0$.
- Elliptic Surface: Paraboloid, Ellipsoid , Sphere: $K > 0$
- Hyperbolic Surface: Elliptic Hyperboloid, Simple-saddle, Monkey–saddle: $K < 0$.

Each of these models has a homogenous Gaussian curvature sign. The haptic interface from Sensable® (Phantom Joystick) was used for the surface exploration during the haptic rendering process.

Methodology. The subjects were blindfolded and guided to the center of the surface before the start of exploration and a haptic snap effect was provided for the surfaces to guide the user to stay in the vicinity of the surface throughout. Surfaces were presented in a randomized fashion.

Data was obtained on the number of fixations (defined as a location where subjects stayed for more than 3 ms) and amount of time spent on each fixation point. The data was normalized and relative fixation data was thus obtained.

3.2 Experiment 2: Visual Exploration

Participants. Same participants from the haptic experiment were included.

Material. The images from the haptic objects as rendered were used as a slideshow in Clearview® software in Tobii® tracker. The screen resolution was the same as the haptic rendering screen resolution (1280x800).

Methodology. Subjects were seated in a hands free stool in front of the screen so that their center of fixation is the center of the image. They were required not to move during the experiment. When done, they could press any key to go to the next image. Image order was randomized to avoid any order effect. Relative fixation data was calculated from the gaze data thus obtained.

4 Results

4.1 Calculation of ROIs

Visual ROIs were calculated as a multiplicative function of normalized values of total time spent on a region and number of times the location is visited. Thus, locations were ordered according to the number of times it was visited and then multiplied with their corresponding dwell time i.e. amount of time spent on each location.

Haptic ROIs were calculated in similar to visual ROIs as a function of total time spent on a location and the number of times the location is visited

4.2 Correlation Analysis and Multiple Regression Model

ANOVA(Analysis of Variance) was performed on the visual ROIs and haptic ROIs. The F value obtained was equal to 0.82 and hence, null hypothesis that the means were equal could not be rejected. Typically, ANOVA is used to signify interaction effects between treatments but in this case our objective was to show that distribution of the two data groups (visual ROIs and hapticROIs) is overlapping. A high correlation of the points of ROIs was also observed (r= 0.849). Fig. 3. shows the correlation between Gaussian curvature and haptic ROIs.

Multiple regression model was employed to determine the visual ROIs for the given values of Gaussian curvature and haptic ROIs and used to predict the values of visual ROIs. The visual ROIs and the haptic ROIs values were divided into a training set and testing set. The values were normalized with respect to the total time spent.

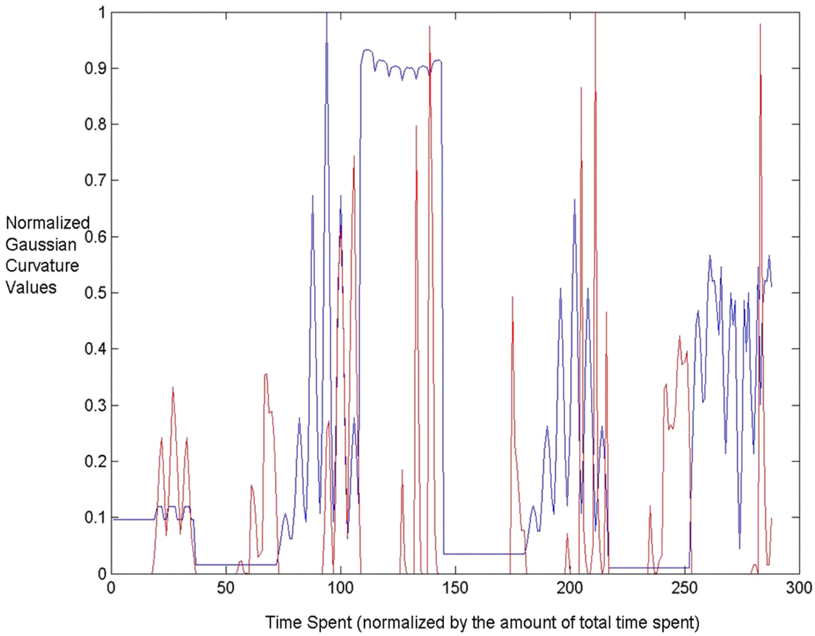

Fig. 3. Gaussian curvature plotted against time spent on each value. Red shows Time spent; blue shows the values for normalized Gaussian curvature.

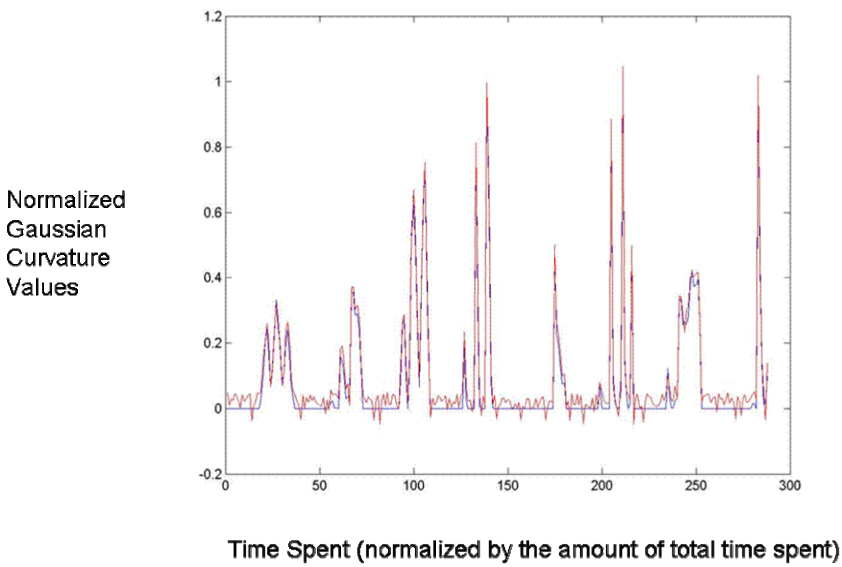

Fig. 4. The predicted values (shown in red) and actual values (shown in blue) for a visual ROI list

The training set employed 70% of all the data capture sessions and the remaining 30% of the data capture sessions were testing set. The test data values were fed to the linear model determined by the regression analysis to determine the predicted values of visual ROIs. Correlation coefficient between predicted values and actual values was determined. In all the ROIs, the correlation between predicted and actual measurement was above 0.89 for each subject and across subjects average correlation value was 0.85. Figure 4 shows an example curve of actual visual ROI (shown in blue) for a subject and the predicted visual ROI from the regression model. The correlation between predicted and actual values in this case was 0.97.

The experiment showed that the concept of saliency maps might be valid for curvature values. Furthermore, this allows allocation of attentional resources across vision and haptics based on the perceptual saliency as a function of curvature in each modality. This can prove very crucial to multimodal environments with high computation and response demands such as tele-surgery and tele-perception which require real-time realistic rendering.

5 Discussion

In this paper, we conduct experiments to compare the regions of interest (ROI) for visual and haptic exploration of surfaces with varying Gaussian curvature. Although, the analysis is somewhat preliminary, it is strongly indicative of the overlap between geometrical features in haptics to its visual exploration. This shows that there is a 'behind the scene' mechanism that attracts our eye gaze in 3D objects. More importantly, such a correlation also supports that interaction between haptics and visual modalities exist beyond the surface characteristics that are confined to visual or haptics modality alone.

Further exploration is needed that can quantitatively evaluate the perceptual load characteristics for such a system. For example, if a feature were to be presented that both modalities *add*, then it can be deduced that variance of such a feature is actually lesser (Ernst & Banks, 2002). Decision noise theory is a possible paradigm to explore the quantitative thresholds for such a distribution.

6 Conclusion

New augmented cognition applications such that flight simulations, remote perception and operation, and surgical simulations have created high demand for interfaces that tackle the information bottleneck for both computer and the user in most efficient manner. This brings in two concepts critical for human perception and human-computer interaction design: attention and multimodality. In spite of the plethora of information that overwhelms our surroundings, we perceive only a small amount that is relevant to us. A fascinating opportunity exists in development of attentive interfaces that employ computational models capable of predicting behavior using multimodal attention. Such systems will be highly beneficial for multimodal human computer interfaces.

References

1. Ballard, D.H.: Cortical connections and parallel processing: Structure and function. In: Arbib, M.A., Hanson, A.R. (eds.) Vision, brain, and cooperative computation, pp. 563–621. MIT Press, Cambridge (1986)
2. Brown, J.M., Weisstein, N., May, J.G.: Visual search for simple volumetric shapes. Perception and Psychophysics 51(4), 42–48 (1992)
3. Cheal, M., Lyon, D.: Attention in visual search: Multiple search classes. Perception and Psychophysics 52(2), 113–138 (1992)
4. Ernst, M.O., Banks, M.S.: Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870), 429–433 (2002)
5. Itti, L., Koch, C., Neibur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
6. Jenmalm, P., Birznieks, I.I., Goodwin, A., Johansson, R.: Differential responses in populations of fingertip tactile afferents to objects' surface curvatures. Acta. Physiologica Scandinavica A24, 181
7. Mesulam, M.-M.: From sensation to cognition. Brain 121, 1013–1052 (1998)
8. Ouerhani, N., Hugli, H.: Computing visual attention from scene depth. In: Paper presented at the 15th International Conference on Pattern Recognition (2000)
9. Pashler, H. (ed.): Attention. University College London Press, London (1998)
10. Posner, M.I., Dehaene, S.: Attentional networks. Trends in Neuroscience (17), 75–79 (1994)
11. Riggs, L.A.: Curvature as a feature of pattern vision. Science 181, 1070–1072 (1973)
12. Spence, C., Driver, J. (eds.): Crossmodal space and crossmodal attention. Oxford University Press Inc, USA (2004)
13. Sridaran, A., Hansford, D., Kahol, K., Panchanathan, S.: Surface interrogation methods for haptic rendering of virtual objects. Paper presented at the World Haptics 2007, Tsukuba, Japan (March 22-24, 2007)
14. Sun, J., Perona, P.: Preattentive perception of elementary three dimensional shapes. Invest. Ophtalmol. Vis. Sci. 41(Supplement), 1083 (1993)
15. Tsotsos, J.K.: Analyzing vision at the complexity level. Brain and Behavioral Sciences 13, 423 (1990)
16. Tsotsos, J.K.: Computaional resources do constrain behavior. Brain and Behavioral Sciences 13, 506 (1991)
17. Wolfe, J.: Visual search. In: Pashler, H. (ed.) Attention, University College London Press, London, UK (1998)