Combining Sensory Information to Improve Visualization

Organizer
Marc Ernst, Max-Planck für biol. Kybernetik

Panelists
Martin Banks, University of California, Berkeley
Felix Wichmann, Max-Planck für biol. Kybernetik
Laurence Maloney, New York University
Heinrich Bülthoff, Max-Planck für biol. Kybernetik

Abstract

Seemingly effortlessly the human brain reconstructs the three-dimensional environment surrounding us from the light pattern striking the eyes. This seems to be true across almost all viewing and lighting conditions. One important factor for this apparent easiness is the redundancy of information provided by the sensory organs. For example, perspective distortions, shading, motion parallax, or the disparity between the two eyes' images are all, at least partly, redundant signals which provide us with information about the three-dimensional layout of the visual scene. Our brain uses all these different sensory signals and combines the available information into a coherent percept. In displays visualizing data, however, the information is often highly reduced and abstracted, which may lead to an altered perception and therefore a misinterpretation of the visualized data. In this panel we will discuss mechanisms involved in the combination of sensory information and their implications for simulations using computer displays, as well as problems resulting from current display technology such as cathode-ray tubes.

STATEMENTS

Marc Ernst

Humans use all their senses, such as vision, audition, and touch, to interact with the environment. The information derived from these various senses has to converge in order to form a coherent percept. In general, this information can either be complementary in the different sensory modalities, such as information about an objects color or weight, or it can be redundant, such as the information about an objects shape or size derived from sight and touch. For the purpose of visualization one might want to utilize the complementary aspects of the different sensory modalities in order to increase the bandwidth for conveying information. It has to be considered, however, that in spite of the complementary nature of this information there are still often interactions that occur between the sensory modalities. The bases for these interactions are frequently statistical correlations in the occurrence of the different sources of information in the natural environment. The size-weight illusion, which I will demonstrate in this panel, is an example of such an interaction between vision and touch.

The main focus of my presentation however, will be on the combination of redundant information across the visual and haptic modality [1-3]. For example, when a person looks at an object while exploring it with the hand, vision and touch both provide useful information for estimating the object's size, shape or position. One observation is that frequently, vision dominates the integrated, visual-haptic percept. In some circumstances however, such as when judging an objects texture, the combined percept is clearly affected by haptics. I propose a general statistical principle, which is based on minimizing the variance in the final combined estimate, to determine the degree to which vision or haptics dominates. This principle is realized by using maximum-likelihood estimation (MLE) to combine the inputs. In general, each sensory estimate, $\hat{S}_i$, is corrupted by noise. If the noises are independent and Gaussian with variance $\boldsymbol{s}_i^2$, the MLE of the environmental property is given by a weighted linear sum of the individual estimates:

$$\hat{S} = \sum_i w_i \hat{S}_i \quad \text{with normalized weights:} \quad w_i = \frac{1/\boldsymbol{s}_i^2}{\sum_j 1/\boldsymbol{s}_j^2} \qquad (1)$$

When using the MLE rule to combine visual and haptic estimates, $\hat{S}_V$ and $\hat{S}_H$, the variance of the final (visual-haptic) estimate, $\hat{S}$, is given by:

$$\boldsymbol{s}_{VH}^2 = \frac{\boldsymbol{s}_V^2 \boldsymbol{s}_H^2}{\boldsymbol{s}_V^2 + \boldsymbol{s}_H^2} \qquad (2)$$

To investigate cue combination quantitatively, I first measured the variances associated with visual and haptic estimation of height. Those measurements were then used to construct predictions derived from the MLE integrator. The models predictions and the human's behavior were very similar in a visual-haptic task. Thus, the nervous system seems to combine visual and haptic information in a fashion quite similar to MLE integration. Visual dominance occurs when the variance associated with visual estimation is lower than that associated with haptic estimation and vise versa.

To conclude, I will be discussing the benefits of multimodal computer simulations for evoking univocal percepts, which is important not only for generating more convincing VR setups but also for the unmistakable visualization of scientific data.

Martin Banks

Current 3d display systems create a depth impression from a variety of cues including binocular disparity, perspective, motion parallax, and shading. Indeed, current systems allow one to reproduce the patterns of light in the retinal images that the portrayed scene would have created. Nonetheless, such displays

are often perceived as different from the portrayed scene. In principle, there are at least three depth cues created by digital displays that could contribute to such distortions: 1) inappropriate focus cues, 2) pixelization, and 3) inappropriate motion parallax during head movements. We measured the contribution of these inappropriate screen cues to perceived slant by varying independently the slant specified by the computer graphics algorithm ("computed slant") and the physical slant of the CRT on which the stimuli were presented ("screen slant"). Planes with different computed and screen slants were presented (tilt = 0 deg) and observers indicated the amount of perceived slant. Precise spatial calibration ensured that retinal-image shapes, texture gradients, and disparity gradients were determined by only the computed slant. Across different experiments, we examined the influence of display type (monocular vs. binocular), screen distance, amount of slant, conflict between computed and screen slant, and the observers' awareness that the stimuli were presented on a CRT. Screen slant had a significant effect on perceived slant in a wide variety of conditions. The effect was larger in monocular than in binocular viewing conditions, at short distances, at large screen slants, and when observers were aware of the use of a CRT. We used regression analyses to determine the effective weight given inappropriate screen cues across the various conditions. These results show that inappropriate screen cues can have a significant effect on 3d percepts and that the size of the effect depends strongly on viewing condition. These findings are highly relevant to the use of 3d displays in basic vision research and in applications such as scientific visualization, entertainment, teleoperation, and more.

Felix Wichmann

Most visualization panels today are still built around cathode-ray tubes (CRTs), certainly on personal desktops at work and at home. Whilst capable of producing pleasing images for common applications ranging from email writing to TV and DVD presentation, it is as well to note that there are a number of nonlinear transformations between input (voltage) and output (luminance) which distort the digital and/or analogue images send to a CRT. Some of them are input-independent and hence easy to fix, e.g. gamma correction, but others, such as pixel interactions, depend on the content of the input stimulus and are thus harder to compensate for. CRT-induced image distortions cause problems not only in basic vision research but also for applications where image fidelity is critical, most notably in medicine (digitization of X-ray images for diagnostic purposes) and in forms of online commerce, such as the online sale of images, where the image must be reproduced on some output device which will not have the same transfer function as the customer's CRT. I will present measurements from a number of CRTs and illustrate how some of their shortcomings may be problematic for the aforementioned applications.

Laurence Maloney

Typical rendering packages operate on 3-dimensional 'RGB' vectors that are intended to represent the spectral character of surfaces and lights in a scene [4]. The mathematical operations performed on these vectors include component-wise addition and multiplication, and multiplication of a vector by a scalar. With this mathematical vocabulary the package attempts to simulate the flow of light through the scene to the observer's eye. The superposition of lights $[R, G, B]$ and $[R', G', B']$ is $[R + R', G + G', B + B']$, the absorption and reemission of light $[R, G, B]$ by a matte surface $[r, g, b]$ is $\boldsymbol{g}[rR, gG, bB]$ where

$\boldsymbol{g}$ is a scalar determined by the direction of incidence, the surface normal, and the direction of emission (the *geometry* of the scene).

The $[R, G, B]$ code assigned to an illuminant is intended to capture its color appearance to a human observer. Two light sources that are identical in appearance to a human observer must, therefore, be assigned the same $[R, G, B]$ code. However, the objects in a scene illuminated by light sources that are identical in color appearance can take on very different color appearances under the different illuminants (Fig. 1). The $[R, G, B]$ code assigned to an illuminant accurately predicts the perceived color of a light viewed directly in isolation, but does not permit prediction of how the spectral power distribution of the illuminant is altered by interaction with surfaces and active media on its passage through the scene to the eye. A physical light assigned a code $[R, G, B]$ that is absorbed and re-emitted by a surface with code $[r, g, b]$ will not, in general, emerge as light with a code $\boldsymbol{g}[rR, gG, bB]$. The mathematical operations used in rendering do not mimic the physical operations of light-surface interaction with the usual interpretation of $[R, G, B]$.

Figure 1: *A Scene Illuminated under Two Lights that are Identical in Appearance when Directly Viewed.* Although the human observer cannot discriminate between the lights, the interactions of the lights with surfaces lead to different outcomes (Taken from [5], Color Plate XIII. With Permission of Wiley).

572

We can, however, assign N-dimensional vectors to lights and surfaces in such a way that the operations of component-wise addition and multiplication of vectors and scalar multiplication do serve to mimic physical light-surface interactions. Let I_i, $i = 1, 2, \cdots, n$ be non-overlapping intervals in the electromagnetic spectrum. Let j_i, $i = 1, 2, \cdots, n$ be the characteristic functions of the n intervals. Define the linear function space $\mathrm{F}(I_1, \cdots, I_n)$ of step functions,

$$f_{\boldsymbol{n}}(\boldsymbol{1}) \quad = \sum_{i=1}^{n} \boldsymbol{n} \boldsymbol{j}_i(\boldsymbol{1}) \tag{3}$$

for any choice of $\boldsymbol{n} = (\boldsymbol{n}_1, \cdots, \boldsymbol{n}_n)' \in R^n$. Note that $\mathrm{F}(I_1, \cdots, I_n)$ is closed under addition and multiplication as well as scalar multiplication; the real vector space R^n is closed under component-wise addition and multiplication as well as scalar multiplication. The correspondence $\boldsymbol{n} \rightarrow f_{\boldsymbol{n}}$ is an isomorphism of algebras [6]. The operations typically applied to vectors in rendering mirror the addition, multiplication and scalar multiplication of functions in $\mathrm{F}(I_1, \cdots, I_n)$. For example, if $f_{\boldsymbol{n}}$ is the spectral power distribution of a light and $f_{\boldsymbol{n}'}$ is the surface reflectance function of a surface, then the product of spectral power distribution and surface reflectance is $f_{\boldsymbol{n}}(\boldsymbol{1}) f_{\boldsymbol{n}'}(\boldsymbol{1}) = f_{\boldsymbol{n} \mathbf{e} \boldsymbol{n}'}(\boldsymbol{1})$ where $\mathbf{e}$ denotes component-wise multiplication of vectors.

If all of the spectral power distributions of lights and all of the spectral reflectance functions of scenes were confined to a *step-function algebra* $\mathrm{F}(I_1, \cdots, I_n)$, then we could view the operations of rendering as operations applied to the corresponding functions. By increasing n, we can create step function families that provide arbitrarily good approximations to the spectral power distributions of lights and the spectral reflectance functions of surfaces in scenes, and, consequently, we can reduce the error in rendering below any specified bound.

I will describe the 'fit' of n-dimensional step-function algebras to collections of spectral power distributions of daylight illuminants and reflectance functions of naturally-occurring surfaces as a function of n. I will also discuss the optimal choice of intervals. These methods also permit compression of such collections without loss of information essential to rendering applications.

Heinrich Bülthoff

The study of human perception and cognition has changed a lot since researchers started to use more sophisticated visualization tools and virtual environments (VE's) for their experimental studies. Since human behavior is intimately tied to the environment, in which it occurs, and the physical environment is a rich source of information an understanding of everyday behavior, then, can benefit from examining performance in similarly rich environments. In recent years it became possible to control and manipulate the environment in which perceptual experiments are conducted without having to resort to simplified and unrealistic laboratory environments. I will describe experiments conducted in high-fidelity virtual environments in which the primary focus is on the intimate relationship between perception and action. In particular, we have studied in our virtual environment laboratory driving, bicycling, as well as navigation in general, to determine what information is necessary for particular tasks. We also examine what the spatial and temporal relationships are necessary for feedback about our actions to be useful.

To enable a realistic experience, our participants are seated on a recumbent bicycle positioned in the centre of a large half-cylindrical projection screen (7m diameter) (Fig. 2). The bicycle requires them to actively steer and pedal in order to move through the environment. The design of the bicycle allows for a realistic simulation of the physical aspects of bicycle riding. A 3-pipe SGI Onyx2 InfiniteReality system is used to generate the 3500 by 1000 pixel image that is projected on the 180 by 50-degree projection screen.

Figure 2: *Virtual reality setup.* 180°, 7m diameter VR theatre with recumbent bicycle positioned in the centre displaying "Virtual Tübingen".

BIOGRAPHIES

Marc Ernst

Marc Ernst is a research scientist at the Max-Planck-Institute for Biological Cybernetics in Tübingen, Germany. He studied Physics in Heidelberg and Frankfurt/Main. In 2000 he got his doctoral degree from the University Tübingen for work he conducted at the Max-Planck-Institute for Biological Cybernetics Tübingen/ Germany. In his thesis, for which the Max-Planck Society awarded him the Otto-Hahn-Medaille, he investigated human visuomotor behavior. For his work on perceptual learning he was granted the year 2000 Attempto-Prize of the University of Tübingen. Marc Ernst spent almost 2 years as a postdoc at the University of California, Berkeley working with Prof. Martin Banks on psychophysical experiments and computational models investigating the integration of simultaneous visual and haptic information.

Contact:
Marc O. Ernst
Max-Planck-Institut für biologische Kybernetik
Spemannstr. 38
72076 Tübingen / Germany
marc.ernst@tuebingen.mpg.de

Martin Banks

Martin Banks is Professor of Optometry, Vision Science, Psychology, and Bioengineering at the University of California,

Berkeley. He is currently the chairman of the Vision Science Graduate Program at Berkeley. Dr. Banks received his BA degree in Psychology at Occidental College in Los Angeles, his MS degree in Experimental Psychology at the University of California, San Diego, and his PhD in Developmental Psychology at the University of Minnesota. His thesis work was on the development of spatial vision in human infants. He was Assistant and Associate Professor of Psychology at the University of Texas at Austin from 1976-1985 before moving to Berkeley where he has been since. His research interests include visual space perception, visual development, and cross-modal perception.

Contact:
Martin S. Banks
University of California, Berkeley
School of Optometry
360 Minor Hall
Berkeley, CA 94720-2020, USA
marty@john.berkeley.edu

Felix Wichmann

Felix Wichmann joined the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, as a research scientist in the Empirical Inference Department in 2001. He received his BA in 1994, and his D.Phil. in 1999 from the University of Oxford, England, working with Bruce Henning in Experimental Psychology. At Oxford he held a Jubilee Senior Scholarship at St. Hugh's College and a Junior Research Fellowship at Magdalen College, and won the George Humphreys Prize. He spent two years as a postdoctoral fellow jointly at Oxford and at the KUL in Leuven, Belgium. His research interests are (1) understanding the initial coding and processing of sensory data ("early vision"), (2) how coding and processing are tuned to the statistics of naturally occurring stimuli, and, (3) how the human visual system represents and categorizes visual patterns.

Contact:
Felix A. Wichmann
Max-Planck-Institut für biologische Kybernetik
Spemannstr. 38
72076 Tübingen / Germany
felix.wichmann@tuebingen.mpg.de

Laurence Maloney

Laurence Maloney is Associate Professor in the Psychology and Neural Science Department at the New York University. He received his BA degree in Mathematics at Yale University and his MS degree in Mathematical Statistics at Stanford University. From Stanford University he also got his Ph.D. in Psychology in 1985. In 1987 he was awarded the Troland Research Award of the National Academy of Sciences. He was Assistant Professor of Psychology and of Electrical Engineering and Computer Sciences at the University of Michigan at Ann Arbor and at the New York University. His research interests include (1) color vision, (2) depth vision and depth fusion, and (3) visual calibration. He is also interested in how perceptually derived information is represented, and in developing and applications of measurement theory to problems of representation.

Contact:
Laurence T. Maloney
New York University
Center for Neural Science
4 Washington Place, Room 809
New York, NY 10003, USA
ltm1@nyu.edu

Heinrich Bülthoff

Heinrich Bülthoff is scientific member of the Max Planck Society and director at the Max Planck Institute for Biological Cybernetics in Tübingen. He is head of the Psychophysics Department in which a group of about 40 biologists, computer scientists, mathematicians, physicists and psychologists work on psychophysical and computational aspects of higher level visual processes in the following areas: object and face recognition, sensory-motor integration, spatial cognition, computer graphics psychophysics, and perception and behavior in virtual environments. Prof.. Bülthoff is involved in many international collaborations on these topics and member of the European research network ECVision. He is partner in the following projects funded by the European Commission: CogVis, Comic, TOUCH-HapSys, POEMS and PRA. He holds a Ph.D. degree in the natural sciences from the Eberhard-Karls-Universität in Tübingen. From 1980 to 1988 he worked as a research scientist at the Max Planck Institute for Biological Cybernetics and the Massachusetts Institute of Technology. He was Assistant, Associate and Full Professor of Cognitive Sciences at Brown University in Providence from 1988-1993 before becoming director at the Max Planck Institute for Biological Cybernetics. Since 1996 he is also Honorary Professor at the Eberhard-Karls-Universität in Tübingen.

Contact:
Heinrich H. Bülthoff
Max-Planck-Institut für biologische Kybernetik
Spemannstr.38
72076 Tübingen / Germany
heinrich.buelthoff@tuebingen.mpg.de

References

[1]   Marc O. Ernst, Martin S. Banks & Heinrich H. Bülthoff. Touch can change visual slant perception. *Nature Neuroscience* **3** (1), 69-73, Jan. 2000.

[2]   Marc O. Ernst & Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429-433, Jan. 2002.

[3]   James M. Hillis, Marc O. Ernst, Martin S. Banks & Michael S. Landy. Combining sensory information: Mandatory fusion within, but not between, senses. (submitted)

[4]   M. F. Cohen & J. R. Wallace. *Radiosity and Realistic Image Synthesis*. (San Francisco: Morgan Kaufmann, 1993).

[5]   R. M. Evans. *An Introduction to Color*. (New York: Wiley 1948).

[6]   S. MacLane & G. Birkhoff. *Algebra, 2nd Ed.* p. 334 (New York: Macmillan, 1979).