CAMPFIRE

Perceptually Adaptive Graphics

Preconference Proceedings

Co-Chairs
Ann McNamara
Carol O'Sullivan

Program Advisory
Jim Ferwerda
Victoria Interrante
David Luebke
Holly Rushmeier
William Thompson
Tom Troscianko

Snowbird
26th - 29th May 2001

CAMPFIRES: Synergies for the Future

ACM SIGGRAPH and EUROGRAPHICS

INTRODUCTION

In recent years the realisation has been growing within the computer graphics community of the advantages to be gained by using knowledge of human perception. This Campfire will bring together researchers from the fields of computer graphics and visualisation, psychology, eye-movement analysis, and other related fields to discuss how such knowledge may be exploited to enhance the realism of computer-generated scenes, animations, and virtual environments.

There were many hard problems to solve in the field of Computer Graphics during the second half of the last century, and the challenges posed were sufficient to keep a large number of researchers very busy. For a long time, the emphasis was on producing physically valid representations of objects, lighting and motion, and on improving the efficiency of such algorithms. To enhance real-time performance, techniques that traded accuracy for speed were developed, rendering objects and animations at variable levels of detail in order to keep a high and constant frame-rate. In more recent years, the computer graphics community has started asking itself questions such as: How do I know how real an image is? What parts of the image can be rendered at a lower level of detail, or even eliminated, without the change being perceptible to the viewer? People are extremely sensitive to tiny anomalies in certain types of simulation (such as facial animation) and will barely notice significant inaccuracy in others (such as a tree blowing in the wind). Why is this, what factors influence it, and how can it be quantified and exploited? In the field of Virtual Reality, what will enhance the feeling of immersion in a virtual world?

Meanwhile, psychologists studying aspects of the Human Visual System (HVS), from both the neurological and psychophysical points of view, have made significant progress, but there is still much to learn. Many of the functions of the human brain, including the visual system, still remain a mystery. Therefore, there is no complete and rigorous theory that the computer graphics community can study and apply. Some computational models of different visual functions have been developed, but they rarely generalise to handle the complex visual tasks typical in graphical scenes and simulations. Most psychophysical experiments consider only one property of the visual system in isolation, (such as Vernier Acuity: the ability to detect whether two lines are displaced or not), and these experiments are often conducted in restrictive laboratory conditions e.g. with reduced lighting, monochrome display, fixed head position and restricted yes-no responses. Neurological investigations reduce even further, to

the reaction of a single brain cell to simple stimuli. Natural, higher level tasks need to be devised, that more truly reflect the situations being simulated.

We can see that there is a wonderful opportunity for the psychological community and the graphics community to co-operate and learn new truths about human perception. Semir Zeki, one of the world's foremost researchers into the visual functions of the brain, has written: "I hope that no one will be deterred from asking new questions and suggesting new experiments simply because they are not specialists in brain studies." The questions asked by CG practitioners when trying to reproduce reality often leave psychologists scratching their heads, thus potentially triggering new directions of research for them. Together, exciting experiments can be designed that provide deep insights into the higher-level processes involved in visual processing: Scenarios which would be almost impossible to set up in the real world are easily simulated and controlled using computer graphics, and psychological expertise is imperative to ensure the robustness of experiments and the validity of conclusions drawn.

We would like to thank the people who helped to make this Campfire possible: Firstly our program advisors, Jim, Victoria, David, Holly, Bill and Tom. Thanks also to Alan and Erica at ACM for all their help and to Shawn, Norma and Annie in Snowbird for looking after us so well. A special thanks goes to John in Dublin for the web design.

We hope you have a wonderful time and that this Campfire will stimulate interesting discussion and encourage future collaborations. Above all, we hope you all have a lot of fun!

Co-chairs

Ann McNamara
Carol O'Sullivan
Utah, May 2001

CONTENTS

IMAGE FIDELITY

DISTANCE & SCALE IN CG

INTERACTIVE GRAPHICS

APPLICATIONS

SCENE PERCEPTION

VISUALIZATION

Image Fidelity

Chair: Jim Ferwerda

Hi-Fi Rendering

Jim Ferwerda, Program of Computer Graphics, University of Cornell

The goal of realistic image synthesis is to produce images that capture the appearance of real objects and scenes. Over the past 30 years, great progress has been made toward this goal with the development of physically-based rendering algorithms that accurately simulate the propagation of light in complex environments. However, even with these advances, generating visually realistic images is still more of an art than a science. One of the major problems is that while failures of realism are immediately evident, we don't have good guidelines for achieving success. One of the tacit assumptions often made in graphics is that if physically accurate simulations can be generated then visual realism will follow. However physical accuracy is neither necessary nor sufficient to produce visually realistic images. The following paragraphs describe three standards of realism that that might help us define the criteria we need to meet to generate images that capture the appearance of the world around us

The first standard is physical realism. Here the criterion for realism is that the image has to provide the same visual stimulation as the scene. If we neglect optical filtering and scattering in the eyeball, this means that the image has to be an accurate point-by-point representation of the spectral radiance values at a particular viewpoint in the scene. Although physically-based rendering algorithms can create accurate simulations of light transport, and these may be of great use in engineering and design applications, adopting physical realism as the criterion for generating visually realistic images has a number of important drawbacks. First, physically-based rendering algorithms are extremely computationally expensive, which limits their usefulness for many graphics applications. Second, even if the efficiency of the algorithms could be improved, it's often impossible to accurately display their results, because the color gamuts and intensity ranges in the simulations frequently exceed the capabilities of existing display devices. Finally, in most cases, doing physically-accurate rendering is overkill if your goal is to make images for people to look at because of the limits of human visual processing.

These considerations lead to the second standard for realism and that's photorealism. In computer graphics, photorealistic rendering typically refers to generating an image that's indistinguishable from a photograph of a scene. This is a fine goal, but unfortunately it begs the question of realism because it doesn't explain why a photograph is realistic. This is largely an unanswered question, but at least one approach is to restrict the scope of photorealism to photometric realism. Here the criterion for realism is that the image has to produce the same visual response as the scene even if the physical energy coming off the image is different from the scene. Adopting this criterion allows us to include the observer in the rendering process, and take advantage of the well known spatial, temporal, chromatic, and dynamic range limitations of vision to relax the precision we need to create visually realistic images. One example is standard color imaging which takes advantage of the trichromatic nature of color vision to create the appearance of a wide range of hues by combining three primaries. Another is the use of threshold visibility metrics to determine sampling parameters in perceptually-driven rendering algorithms. However, using photorealism as a standard for visual realism also has a number of drawbacks. First, although photorealistic rendering algorithms can be faster than physically realistic algorithms, they are still far too computationally expensive to be useful in many applications. Second, it's not clear that photorealism is necessary or even desirable in a wide range of applications. Finally, adopting photorealism as the standard for visual realism in computer graphics categorizes most computer-generated images as failures, yet says nothing about their obvious utility.

These problems suggest a third standard for realism and that's functional realism. Here the criterion for realism is that the image has to provide the same visual information as the scene. Information here means knowledge about the properties and relations of objects (e.g. shape, size, position, motion, material, etc.) that allows a user to make reliable visual judgments and perform useful visual tasks. Here realism is measured in terms of the fidelity of the information

the image provides. If an image lets you do the task you need to do, and allows you to perform as well as you could in the real world, then for that task, the image is realistic. The beauty of a functional definition for realism is that it admits a wide range of rendering styles, from physically-accurate simulation, through photorealism, to more abstract approaches. One example of functionally-realistic images are the renderings used in flight simulators. These images aren't physically accurate nor are they photorealistic, but they are functionally realistic because they provide the user with much of the same visual information they would get if they were flying a real plane. The proof of the realism of these images is that users can learn skills that transfer into the real world at high levels of performance.

An important direction for future research in realistic image synthesis will be to develop metrics that will allow us to integrate these different standards of realism into a coherent framework. One promising approach could be to conduct psychophysical experiments that explore the relationship between physical accuracy and visual fidelity in image synthesis. On the basis of these studies we should be able to develop efficient realistic rendering algorithms that adapt gracefully depending on the resources available and the demands of the application, balancing accuracy and efficiency, but always maintaining visual fidelity.

Inverse Rendering Algorithms

Simon Gibson, University of Manchester, UK

My current research interests lie in the development of inverse rendering algorithms for virtual and mixed-reality applications. In particular, I have recently been examining the problem of reconstructing geometry and illumination information from images of real-world environments. One particular application of these inverse illumination algorithms is in capturing illumination data for use in augmented reality applications, where synthetic objects need to be illuminated by the same light as the real objects in the scene.

Figure: Digital photograph of a real scene (left), and synthetic renderings using reconstructed illumination and material data (middle and right).

Accurately estimating illumination in general environments is a difficult problem. In many cases, such as outdoor or large indoor scenes, obtaining a complete geometric model is not possible. Additionally, capturing radiometric information (in the form of photographs) for each surface is a time-consuming task. The environment could also be illuminated by many different types of light sources, or a mixture of natural and artificial illumination. A flexible inverse illumination algorithm should work with partial geometric and radiometric models, and be able to handle arbitrary types and mixtures of illumination.

Because of the complexity of the task, any algorithm that will work robustly under these conditions must make simplifying assumptions about the way which illumination is captured and used. It is here that knowledge of the human visual system will be of great benefit, both in simplifying the process of capturing illumination data and accelerating the rendering of synthetic objects. If we require our augmented reality renderings to be visually accurate, rather than physically accurate, can we use this to ease the process of capturing illumination data? Interesting questions arise, such as how accurately must indirect illumination be represented? Do we need to know the exact position and size of each light-source in order to render accurate shadow boundaries, or will approximations suffice? Also, given the fact that we have limited resources available for rendering, which aspects of the illumination should take precedence when interactive frame-rates are required?

It is problems like these that I feel need to be addressed if visually realistic augmented reality is to be achieved. The issue of perceptually accurate illumination in mixed-reality applications has received little attention compared to other fundamental problems like tracking and image display, and I hope that discussions at this campfire will start to raise to some of these issues.

Evaluating Realism

Ann McNamara, Image Synthesis Group, Trinity College Dublin

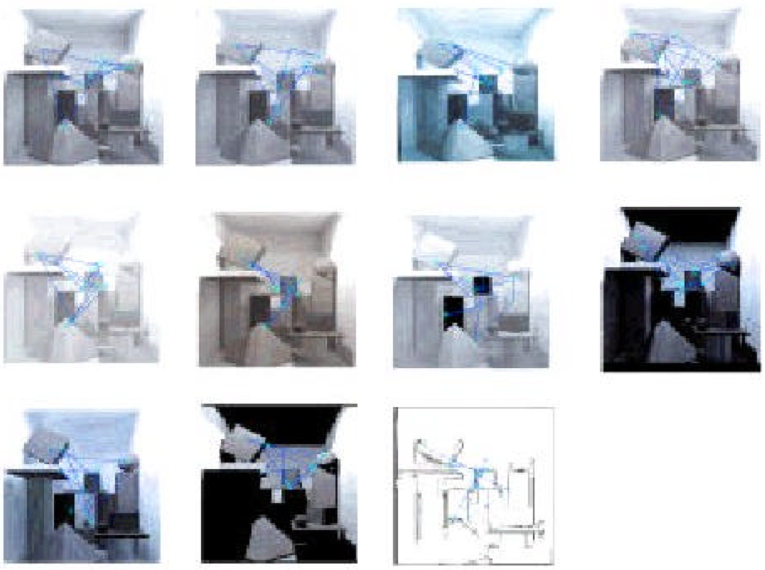
Increasing ubiquity of computer graphics demands high levels of realism. If computer images are to be useful beyond entertainment and simulation, then realism is the key. An important, yet largely unanswered question, is how do we measure the realism of computer generated imagery?

One approach, the approach taken in my research, is to use psychophysics to study the relationship between stimulus (image) and response (visual perception). By conducting a series of psychophysical experiments it is possible to assess the fidelity of graphical reconstruction of real scenes. Experimentation also shows that given a real scene and a faithful synthetic representation of that scene, the visual response function in both cases is similar. This gives a measure of the realism of the synthetic image. These experiments were conducted using a specific set of images – how can these results be generalised to other images, or if they cant how should I design future experiments which will yield results which can be applied to other image scenarios.

Setting up and running experiments is a (painfully) slow process, so it is important to get it right first time or you end up with a set of useless results. I am interested to learn how others go about conducting perceptual experiments to ensure robust results, including designing the experiment, choosing/setting the task, participants, analyses and generalisation of results.

Also, the role of eye movements in image quality assessments is largely unexplored. Can the pattern of eye-movements be used as an indication of the realism in rendered images? To investigate this, eye-movements were monitored while participants compared images of varying quality, including a photograph. This aims to establish if there is a relationship between the pattern of eye-movements and fixations with the quality of a rendered image.

This project is in its infancy, however, preliminary results show that there are indeed common fixation areas among images of similar quality. The figure below shows the patterns recorded for one participant. Images are shown in the rating order given by that participant (on a scale of 0-10). It can be seen that gaze is concentrated primarily in the centre of the image, with fixations mostly on the objects in direct lighting.



Perceptual Methods in Computer Graphics: Evaluation

Holly Rushmeier, IBM TJ Watson Research Center, USA

How do we evaluate perceptually-based methods in computer graphics? A difficulty in making progress in applying perceptual science to graphics is that the assessment of the value of a method is different from what graphics researchers are used to. We can't prove that we reduce the computational complexity of an algorithm, our goal is not to produce dramatically different visual effects, and we can't use a physical instrument to measure the fidelity of our simulations with respect to physical reality.

There is a history in computer graphics of applying perceptual principles ranging from general observations to the use of very specific data. In some cases the observations border on folklore, and other cases they are very sound. In some cases the specific data are applied appropriately, in other cases they are used to give the aura of scientific results to a questionable application. Simply stating that perceptual principles are used doesn't establish the utility of a method. We need user studies, but what are the requirements for these studies? Anyone who has tried to publish these results know the criticisms that are returned -- there should be more subjects, the subjects should be a cross-section of society (not just computer scientists), more variations should have been studied, and what about this, that and the other thing. On the other hand, there have also been publications where very poor studies are published, just because reviewers were so glad to see any effort at validation. Since the field of graphics has no standards or examples for what is adequate, publishing results is a hit or miss proposition.

In graphics we tend to want to solve everything now. There is a desire for a comprehensive method that can resolve substantial issues such as how should an object be represented or what is important in an image. In general, progress is much slower, and involves many experiments examining many subtle issues. How great a contribution is enough to warrant attention?

Clearly we need to look to psychology for guidance on conducting studies. That needs to be combined with our own history of what constitutes a useful contribution to computer graphics.

In the past there was a problem in publishing systems papers in computer graphics, because systems inherently do not use new algorithms, but creatively combine existing techniques. The community has worked on establishing what the criteria are for a systems paper, e.g. documentation of the decisions that were made in the design, report of experience in using the system. Similarly we need to establish some guidelines on how do we as a community evaluate new methods that apply perceptual principles to computer graphics.

Visual arts, perception and NPR

Fredo Durand MIT LCS

There are two classes of reasons that can lead to a divergence from realism: efficiency reasons, and aesthetic reasons. Taking into account perception has large potentials to produce not only more realistic pictures, but also clearer, more relevant or more beautiful pictures

A large class of pictures need to represent reality in a schematic way to improve clarity. Maps, manuals or scientific illustration are examples where a non-photorealistic picture is often clearer than a photograph, because it can emphasize the relevant information and provide straighter clues. Understanding how schematic visual information is perceived and how visual clues are processed are crucial issues to produce more efficient illustrations. Similarly, a major bottleneck of 3D authoring is the difficulty to deal with a 3-dimensional space and the resulting user-interface problems. Effective solution implies to produce interactive images that provide the human perceptual system with functionally optimal inputs.

Since the landmark writings of Von Helmholtz, Arnheim and Gombrich, studying the human visual system has proved fruitful to understand issues and techniques of the visual arts. Our perception of 3D scenes explains the limitations of linear perspective and why alternative projections can be desirable. The different levels of cognitive representation of a scene can explain dramatically different styles of depiction. Gestalt psychology provides fruitful insights to the art of image composition. Low-level and high-level cognitive mechanisms help us understand how an image can attract the gaze of the beholder to important regions. Other examples are numerous.

Most of these results are unfortunately mainly qualitative and will be challenging to translate into automatic image-generation techniques. Truly beautiful images will remain the unutterable domain of sensitive artists. Nonetheless, the insight into art techniques offered by perception can help improve the quality of images produced by non-artists. This includes offering more relevant high-level controls that directly relate to the aesthetic dimension of images. Another potential avenue is to use simple computational models of perception to provide the user with crude feedback and alarms about the quality of images, e.g. with respect to balance or gaze attraction. The development of such techniques has great cross-fertilization potentials between perception and computer graphics, and can help to devise and validate more quantitative models.

Perceptual Limits and CIE Lab-Based Error Metrics

Yinlong Sun, University of Purdue

Abstract
The ultimate goal of realistic image synthesis is to generate computer imagery that is perceptually indistinguishable from its real-world counterpart. Many rendering techniques have been proposed, making tremendous progress to reach this goal. Now we are faced with two basic questions. How far we are from the goal? How well the existing techniques perform? No doubt a quantitative definition or error metric of image realism is important for systematically evaluating synthesized images and comparing different techniques. However, such a definition or metric has not been available. This problem will remain open in the near future considering that the process of human evaluation of image realism, which involves complicated physical as well as physiological factors, has not been understood sufficiently so far. Nevertheless, we can adopt another approach that does not rely on the entire understanding of the human evaluation process. The solution is based upon perceptual limits, the maximal capabilities of the human visual system. It is known that the basic aspects of realism include space, color, and motion. As beginning one can focus on one aspect and establish a simple metric based upon the reference of the perceptual limit. This metric can be improved gradually within the aspect by including more factors and considering the application feedback, and be eventually incorporated with other aspects to form a comprehensive metric. Here we present an example of this approach: a CIELab-based error metric for the color aspect.

Perceptual Limits
The perceptual limit for space, which is also called the minimal resolvable spatial distance, is about 0.084 mm. This value can be calculated by applying Rayleigh's criterion for the eye (cf. page 463 in [Hechtt98]) to a viewing distance of 25 cm, for which the eye has the best acuity. For color, the minimal resolvable color distance is about 1 unit in CIELab color space [Hunt95]. Although the color appearance model [Fairchild98] measures colors more accurately, as beginning the CIELab space is good enough. Finally, for motion, the critical flicker frequency is about 60 Hz (page 18 in [Glassner95]). The critical flicker frequency has been widely adopted as the basis for motion realism or smoothness in computer animation, but it appears that the minimal resolvable distances for space and color have not been considered as references in realistic image synthesis.

Thus, one may regard perfect realism as the point where a synthesized image has just over-reached the perceptual limits. At this point, the difference between a synthesized image and its real-world counterpart is beyond the maximal recognizing powers of the eye. Putting it in another way, further improving the rendering accuracy beyond the perceptual limits will not be differentiated by the eye and is therefore unnecessary. For this reason, it is reasonable that we base on the perceptual limits to establish error metrics for image evaluation.

CIELab-Based Error Metric
We propose a simple error metric based on color differences measured in CIELab space. When comparing two images (of the same size), we compute the CIELab color distances between all corresponding pixel pairs. Thus, given a CIELab distance value, we know how many pixel pairs are differed by this color distance. This forms a histogram, where the horizontal and vertical axes are respectively CIELab distance and the corresponding percentage of pixel pairs. The reference image is best acquired from measurement without losing any color information, such as a spectral image. It is also possible to use a synthesized image with very high accuracy as the reference image. If a histogram distributes mainly on small color distances, say below 3 units, the two images have a good correspondence. In particular, when a histogram distributes completely below 1 unit, the two images have reached a perfect correspondence.

Note, however, that this error metric has not considered all relevant factors for image realism. For example, pixel-pixel correlations such as object boundaries and color contrast are important. This means that this metric can be improved by including contributions from these factors. The error metric can be applied in two ways in computer graphics. First, one can use

it to evaluate the impacts of different parameter values within the same technique. Second, it can be used to compare different techniques. In both cases, the spatial parameters of the camera and objects should be fixed. Thus, the corresponding pixels in the two compared images correspond to the same spatial point, and their color deviation is caused only by using different parameter values or methods.

Example: Comparing Composite Model and Sampling

We apply the CIELab-based error metric to comparing the composite model [Sun00] and the sampling method on representing spectral functions. The composite model decomposes any spectral function into a smooth background and a collection of spikes. A spike is thus represented with a delta function, and the smooth background with uniform sampling. The decomposition makes the composite model much more compact than the sampling method to achieve the same accuracy. This advantage is particularly obvious for fluorescent spectra that have narrow but strong peaks.

Our test scene contains six opaque spheres with reflectances "white", "red", "orange", "yellow", "green" and "blue" of the Macbeth ColorChecker [Wyczecki82]. The light source is a Mercury Arc lamp, which is fluorescent. Figure 1 displays the images rendered with raytracing using the sampling method of 1000 points (used as the reference image), the composite model with 12 sampling points for the smooth background, and the sampling method of 64 and 32 points, respectively. The CIELab distance histograms are displayed in Figure 2. When using the sampling method, even with sample points as many as 64, the error is still noticeable and the histogram has considerable distribution on large color distances. The reason is that the spikes are narrow and missing sampling a spike will result in a significant error. In contrast, the composite model using only 12 sample points has much better accuracy. All such differences are quantitatively shown by the CIELab-based histograms in Figure 2.

Future Work

Improve the CIELab-based error metric by including pixel-neighborhood factors such as object boundaries and color contrast. Measure color distances with color appearance model. If possible, map a color distance histogram to a single value. This can be done by integrate the histogram with a weighting function over color distance. The question is to find a reasonable eight function. Based on the color distances between the corresponding pixel pairs, we can also show the mage regions with large color disparities. This information is useful for analysis. Establish a spatial error metric based upon the minimal resolvable spatial distance. The perceptual limit for space can be used as a criterion for multi-resolution rendering. Combine the color aspect with space and motion to form a comprehensive error metric.

Summary Statements

A formal definition of image realism founded on the understanding of the human evaluation process will remain a open problem in the near future. This is because the human evaluation process is very complicated and has not been understood sufficiently.

As a practical approach, one can establish an error metric for image evaluation based upon the perceptual limits. This approach does not rely on the entire understanding of the human evaluation process.

As beginning, one can focus on one realism aspect and propose a simple error metric. The metric can be improved gradually within the aspect by including more factors and considering the application feedback, and eventually incorporated with other aspects to form a comprehensive metric.

As the perceptual limits, the minimal resolvable spatial distance is about 0.084 mm, the minimal resolvable color distance is about 1 unit in CIELab color space, and the critical flicker frequency for motion is about 60 Hz.

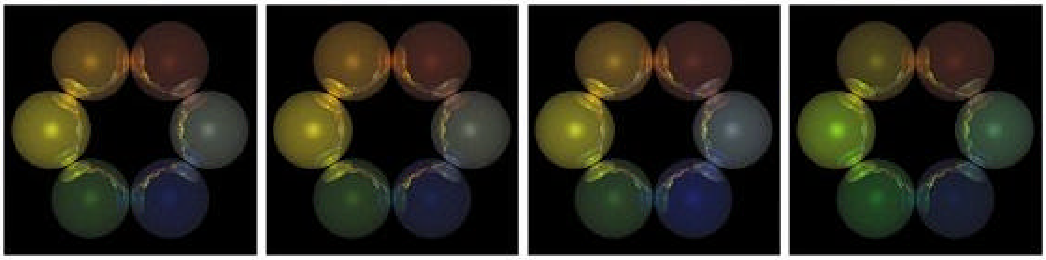


Figure 1: Rendered images using the composite model and the sampling method. The light source is a mercury arc lamp. From left to right, the images are rendered using the sampling method with 1000 points(used as the reference image), the composite model with 12 sample points for the smooth background, and the sampling method with 64 and 32 points, respectively. Compared to the first image, the gray and blue spheres in the fourth image have obvious errors and the red sphere in the third image is too dark. But the second image matches the first very well.

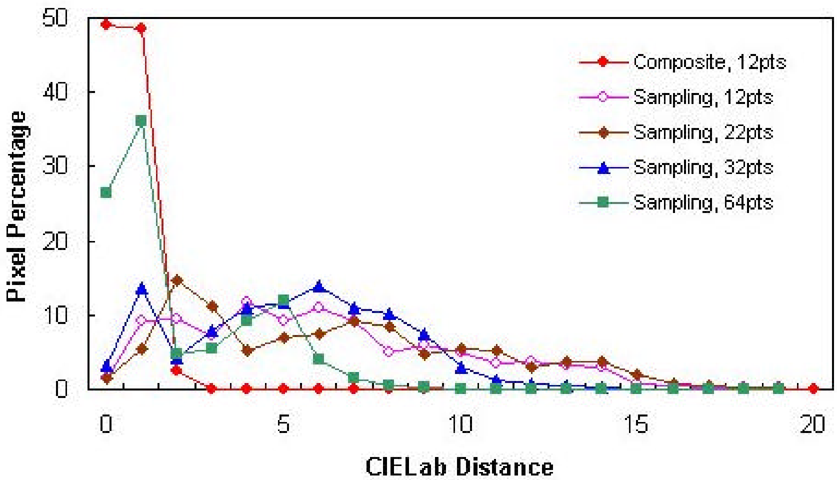


Figure 2: Comparing errors of test images rendered with the composite model and sampling method. The illumination source is Mercury Arc (spiky SPD).

One may regard perfect realism as the point where a synthesized image has just over-reached the perceptual limits.

A CIELab-based error metric is a histogram, where the horizontal and vertical axes are respectively CIELab distance and the corresponding percentage of pixel pairs of two images. If a histogram distributes mainly on small color distances, the images have a good correspondence.

The CIELab-based error metric can be used to compare different parameter values as well as different rendering techniques.

As an example, the CIELab-based error metric is applied to comparing the composite model and the sampling method on representing spectral functions.

References:
Hecht98 E. Hecht, Optics, Third Edition, Addison-Wesley, Reading, Massachusetts, 1998.
Hunt95 R.W.G. Hunt, The Reproduction of Color, Fifth Edition, Fountain Press, England, 1995.
Fairchild98 M. D. Fairchild, Color Appearance Models, Addison-Wesley, Reading, MA, 1998.
Glassner95 A. S. Glassner, Principles of Digital Image Synthesis, Morgan Kaufmann, San Francisco, 1995.
Sun00 Y. Sun, F. D. Fracchia, and M. S. Drew, "A Physically-Based Dual Representation of Spectral Functions," Optical Engineering, Vol. 39, No. 11, Nov. 2000, pp. 2931-2942.
Wyszecki67 G. Wyszecki and W. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulas, Second Edition, Wiley, New York, 1982.

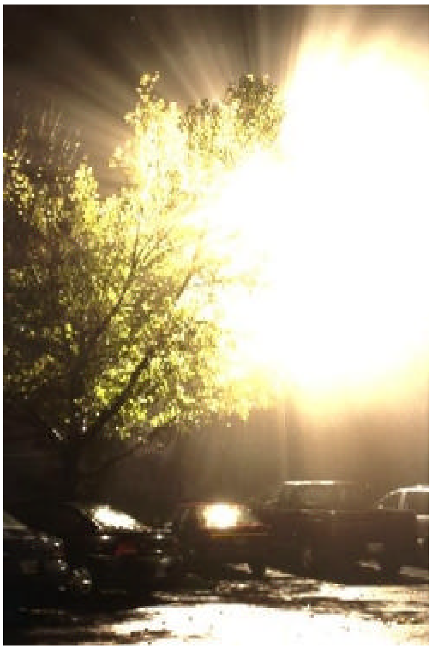
Make Pictures, Not Images

Jack Tumblin, Program of Computer Graphics, Cornell University

All great pictures, no matter how they are made, recreate some desirable set of visual sensations and perceptions. Picture-making processes are human attempts to capture and control what we see, or what we would like to see. Accordingly, we should regard pictures as containers for perceived quantities rather than dismissing them as a simple set of recorded of light measurements.

However, machine-assisted picture-making processes such as photography and computer graphics usually make just one narrow, restrictive class of pictures known as images. Our machines copy scaled 2D scene intensities to a display device--most other forms of picture-making such as sketching, painting, or carving are more directly governed by visual appearance instead, and are often much more expressive and successful.

I think we should move beyond digital image-making to develop new perceptually-guided digital picture-making processes instead.



Image

Picture

Often images are poor pictures. A simple 2-D map of scaled scene intensities often doesn't capture visual appearance very well for at least two important reasons:

- Intensity scaling affects visual appearance; there are many light-dependent changes in the way we see (a forest lit by the sun looks very different when lit by the moon at the same position in the sky), and
- Displays can't recreate most real image intensities and contrasts; common display methods (paint, charcoal, pencil, ink, photographic prints, LCDs, CRTs, DMDs, etc.) can't create displayed contrasts greater than about 100:1. Scene contrasts sometimes exceed 1,000,000:1.

Furthermore, a one-to-one mapping from scene to display intensities is neither necessary nor sufficient to recreate a scene's most important visual sensations; astonishingly 'realistic' watercolor paintings and charcoal drawings can express high contrast scenes with very limited display contrasts, and skilled artists and illustrators frequently use the same display intensities at different display locations to represent wildly different scene intensities.

Computer-aided picture-making methods currently ignore two-thirds of the picture-making process. Picture-making is not finished after we compute or measure the amounts of light in a scene; next we should compute estimates of the most visually important perceived quantities in the viewed scene (e.g. perceived boundaries, shading, textures, reflectances, shadows, illuminants, etc.). Finally, we should compute the best way to express these perceived quantities within the limited abilities of the available display device.

All of our current picture-making processes usually control visual appearance indirectly, by manipulating display controls such as paint, ink, pencil strokes, or pixel values. Digital pixels are usually just a quantized substitute for film density, and current usage for pixel values hopelessly tangle together scene intensities, display intensities, tri-stimulus color models, and vaguely defined perceived quantities (e.g. '0=black and 255=white'). Each should be expressed separately.

Digital picture-making can and should be profoundly different from all previous methods, including film. A digital picture is a purely numerical abstraction, a data structure of our own choosing that _could_ include a model of visual appearance, but usually does not. Our current 'grid-of-pixels' format is very nearly useless for any purpose except display. I think we need new digital picture descriptions that store the most important components of visual appearance in a machine-readable form, making pictures at least as meaningful and distinguishable to a search engine as it might be to a human observer.

Local Illumination Is All that Matters for Graphics Realism

Gregory J. Ward

Much has been made of the importance of global illumination for producing realistic computer graphics renderings. The presence of shadows, specular reflections, caustics, and diffuse interreflection provide important cues to the human visual system and help determine relationships between objects. But how correct do these details need to be in order to be convincing? What does the human visual system really require in terms of accuracy to make its judgements? How true to life does an image need to be to appear real? We observe that distant global lighting interactions may be randomly perturbed and in some cases eliminated with no change in the visual perception of a scene's realism. The same cannot be said for local interactions, which are vital for low-level and high-level visual processing. We propose a simple method for determining when an interaction is local and therefore visually relevant for presenting a plausible (though not necessarily accurate) rendition of a scene. Specifically, we conjecture that the following criteria must be met for accurate lighting interactions between two objects to be relevant to perceived realism:

- The interacting objects must be "nearby" each other and both visible.
- The interactions must be consistent with a plausible context (environment).

We define objects to be nearby one another if the distance between them is less than the larger of the two objects, and visible if at least some part of each is unoccluded and distinguishable from the background. The second rule is more difficult and requires some explanation; two interactions are said to be contextually linked if they relate to a third object, which may be nearby, distant, or fictional. A ready example is a shadow generated by a particular light source. The shadows cast by two local objects onto a third object, such as the floor, must be consistent with one another; i.e., they must correspond to roughly the same light source position. Likewise, specular highlights in the two objects must be consistent with a single light source configuration or environment. However, these two separate phenomena, shadows and highlights, need not be consistent with one another, or even plausible, to appear realistic. Furthermore, objects not nearby to the two objects in question may be consistent with a different environment with little or no loss in realism.

Based on the above conjecture, we propose the following redefinition of local illumination in graphics rendering, and suggest methods whereby rendering costs may be reduced while satisfying the need for realism. (Indeed, many CG lighters and shader writers regularly employ these tricks, but we seek to formalize them for the purposes of automation.) Comprehensive Local Illumination is the interaction of light from some plausible but unidentified illuminating environment with nearby (local) objects, including caustics and interreflections between these local objects.

We specifically include higher order interactions in our definition of local illumination because color bleeding and caustics do matter for nearby objects. This suggests that global illumination methods based on linking distant interacting groups (e.g., Rushmeier, Patterson and Veerasamy, GI '93) may have a good approach. In fact, we believe that a simple modification to this approach will provide equally realistic results at a substantially reduced cost, by dropping all links whose length significantly greater than the largest dimension of the two groups involved.

Another simplification to rendering complexity is to group light sources in a hierarchy, replacing individual light sources with group sources and environment maps past a certain distance. To maintain consistency on nearby objects, we suggest using the same grouping methods used for global illumination to assign environment maps and surrogate light sources to locally interacting subsets of the environment. When two groups are linked by the above proximity criterion, heuristics may be employed to determine tolerable differences between the associated environment maps and source groupings. In this way, each group will be linked to a set of nearby groups for the purposes of local interreflection, and to an environment map and surrogate source list for the purposes of generating shadows and reflections. Between distant, unlinked groups, there need be no restriction on how different

the source grouping and environment maps may be. Conventional ray tracing or scanline rendering techniques may then be employed with the following advantages:

1. Each group requires only local geometry for rendering -- memory requirements are greatly reduced by rendering groups one at a time and compositing them afterwards.

2. Global illumination is reduced to a small, sparse matrix for interacting groups, which may be solved quickly with little memory overhead using full-matrix radiosity.

3. Local interreflection may be computed using any combination of object-space and image-space algorithms, since complexity is held in check by grouping and image segmentation.

4. Expensive caustic rendering may be applied to obtain local specular-to-diffuse illumination, since distant interactions are ignored and combinatorial complexity is thus avoided.

Extending the methods described to animation requires a method to smoothly link and unlink groups as they move together and apart, and morph environment maps and source groupings in such a way that objectionable time-discontinuities are avoided. One very simple approach to this problem is to track positions of objects forward and backward in time, and at each frame interpolate between discontinuous links and maps from the past and future. Links may then be forged and broken at will, and discontinuous changes made to environment maps and source groupings, provided that each parameter maintains its value for a minimum span of time. Worst case, this doubles the amount of time spent rendering each frame, but this is a small price to pay for smooth animation.

None of this should be construed as an indictment of global illumination, a field that is very close to my heart. However, the goal of realism has never been very well-defined in graphics, and hence a good deal of effort has been spent to achieve physical accuracy, which is most assuredly overkill if all you want is to produce something that looks real to a human observer. Physical accuracy is indispensable if one is attempting to determine object visibility, glare, or evaluate aesthetics of a particular design or configuration. Fields such as architecture, lighting design, and flight and driving simulation therefore need to target all effects, local and global, in an illuminated environment. They may incorporate perceptual models to compute tone mapping or even to cut corners during the calculation, but they must in the end reach the same percept as a real environment to achieve their goal. However, for graphics rendering for special effects and other "make-believe" applications, the goal is very different. The distinction between these two applications is paramount, and different techniques are appropriate to each.

Distance and Scale in CG

Chair: William Thompson

Size Matters: Distance and Scale in Computer Graphics

Sarah Creem, Amy Gooch, Helen Hu, Peter Shirley, William Thompson, Peter Willemsen
University of Utah

Daniel Kersten, Cindee Madison, Herbert Pick, Jr. University of Minnesota

John Rieser, Vanderbilt University

Margarita Bratkova, Mount Holyoke College

Andrew Beall, UC Santa Barbara

Overview

Despite recent impressive gains in realism, computer graphics is currently unable to effectively generate images of objects and environments that convey an accurate sense of distance and size. This problem is most apparent in the failure of computer graphics imagery to simulate the appearance of something that is really large.

While some of the difficulty lies in limitations of the various display technologies available to us, computer graphics imagery itself appears to lack critical visual cues to distance and scale. Although we understand that a problem exists, almost nothing is yet known about the nature of this missing information. Our research therefore focuses on both determining visual cues that are perceptually relevant for computer graphics and developing methods for rendering images that include these cues.

Accomplishing these goals requires new insights into the 3D information extractable from 2D images, modifications to graphics algorithms in order to better render salient information, and sophisticated perceptual experimentation to validate that people can actually perceive the intended 3D space. Our approach is intrinsically interdisciplinary, involving perceptual psychology, computational vision, and robotics, as well as computer science and computer graphics.

Accurate perception of size and distance from computer graphics is particularly important in immersive interfaces that aim to give a person the sensation of actually being in a virtual world that in fact only exists as a computer model. In such situations, veridical visual perception is required for tasks ranging from fine manipulation to locomotion through large outdoor spaces. This will require new computationally tractable ways to simulate illumination effects, material properties, and complex geometry in both man-made and natural environments.

Immersive interfaces provide the opportunity to explore the coupling between visual perception and action that is critical to human interaction with complex spaces. We believe that when a user is interacting with a computer simulated spatial environment, it is important that the user be able to walk through the environment and manipulate objects of interest in a manner that is as natural and realistic as possible.

Current Research and Open Questions

Lots of questions, few answers...

Visual Realism and Depth Estimation

Surprisingly (at least to us), it is still an open question whether or not visual realism aids in accurate distance judgments based on computer graphics imagery. A number of different research groups have now reported that even with visual immersion, appropriate linear perspective, and binocular stereo, distance judgments while viewing computer graphics are significantly underestimated compared to those made while viewing the "real world," at least

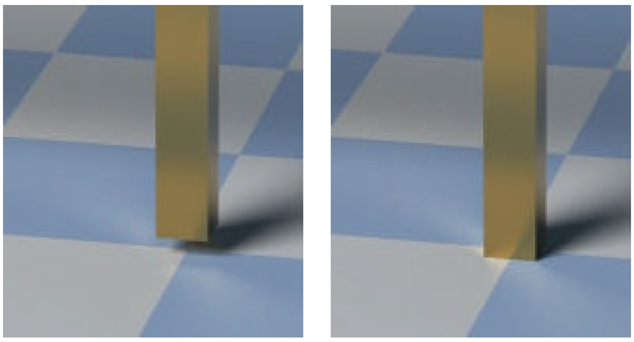
for distances ranging out to 10m-20m. The effect is likely even more pronounced for longer distances, though we are unaware of formal studies confirming this.

Estimating distance in a virtual world

A computational analysis shows that there are actually only a few visual cues that can provide information about absolute depth in the 10m-20m range. We are exploring the hypothesis that one such cue -- familiar size -- depends on levels of visual realism that have not been used in prior experiments measuring depth judgments based on computer graphics.

Global Illumination as a Contact Cue

Renderings based on only local illumination often have a "cookie cutter" appearance, with objects tending to visually float in space rather than being seen as in contact with other objects and surfaces. It is well known that cast shadows are a visual cue to both the 3-D position of objects in a scene and to the fact that an object is in proximity to another object or an extended surface. The evidence is conflicting, however, on the accuracy with which shadows need to be rendered in computer graphics to achieve this effect. Understanding which aspects of shadow realism are perceptually salient is critical in order to balance effectiveness and efficiency in interactive applications.



Illumination cues for object contact

In computer graphics, interreflection is understood to be important for achieving subjective realism. Until recently, however, interreflection has not been investigated as a spatial cue. We have now shown that interreflection can function as a contact cue in much the same way as shadows. In at least some situations, barely perceptible interreflections can generate a visual sense of contact that is comparable in effectiveness to prominent shadows.

Interaction of Visual Perception and Locomotion

People naturally walk through the complex real world while maintaining spatial orientation in an effective manner. Travelling through virtual worlds using current computer interfaces is far less easy. We know that physical locomotion through an environment greatly assists in keeping track of where you are in the environment and where other locations are with respect

to your own position. We are developing a locomotion interface with the goal of creating veridical visual and motor perception of distance, speed, and slope, and a realistic experience of turning and translating for accurate updating of locations in the environment. One key research question is whether we can build displays that synergistically combine visual information, generated by computer graphics, with biomechanical information on locomotion, such as generated by walking on a treadmill. In particular, can a user of such a system get a more accurate sense of a simulated space than would be possible using computer graphics alone?

Infrastructural Issues

In conducting experiments we have found ourselves making custom software to control the parameters of interest. For example, few if any packages allow one to create images with specific levels of direct lighting or interreflection turned on or off. We would like to develop a public-domain code infrastructure to make this process easier. Such an infrastructure should 1) avoid all discretization errors, and 2) have a clean and flexible code base that is easy for researchers to modify. We will discuss a plan to design such a framework, and seek guidance on what is needed in such a framework.

Interactive Graphics

Chair: David Luebke

Eye Tracking Techniques for Perceptually Adaptive Graphics

Andrew T. Duchowski, Clemson University

Introduction

A primary goal of Perceptually-Adaptive Computer Graphics (PACG) research is to develop CG methods which in some way adapt to the human observer. On the one hand, interactive approaches aim to dynamically alter the rendered image in real-time to match the perceptual limits of the Human Visual System (HVS) at the viewer's Point Of Regard (POR). On the other hand, non-interactive approaches typically involve generating static imagery designed to match the HVS. For example, adaptive image codecs and renderers have been developed to efficiently encode similar spatial or color regions which are below the human Just Noticeable Difference (JND) threshold. Both approaches can benefit from, and in fact may rely on, the use of eye trackers. The aim of this paper is to survey current PACG-related eye tracking techniques, and where appropriate, highlight potential future research opportunities.

Eye Tracking Techniques

Advancements in eye tracking technology, specifically the availability of cheaper, faster, more accurate and easier to use trackers, have inspired increased interdisciplinary eye movement and eye tracking research efforts. A good example of this type of diverse interest can be found in the proceedings of the recent Eye Tracking Research & Applications Symposium (organized and co-chaired by the author) [DKS00]. Although these proceedings contain exemplary work from several disciplines, only a few papers are directly related to CG research. Work most closely related to CG is currently conducted by the Human Computer Interaction (HCI) community, however, here CG is often used as a means to an end, and is not often the objective of the research.

The use of eye trackers in graphical systems falls into two general application types: diagnostic and interactive. Interactive eye tracking systems typically respond in some way to the location of the user's gaze. Such interactive systems may be classified by two application sub-types: selective and gaze-contingent. In selective applications, the user's gaze acts as an alternate mode of input, often compared to a pointing device. For example, Tanriverdi and Jacob have recently used an eye tracker as a selection device in VR [TJ00]. While these types of interactive applications are interesting in their own right, they are not necessarily relevant to the development of PACG techniques. Interactive techniques more concerned with developing perceptually-adaptive graphics fall under the gaze-contingent systems category, as shown in the hierarchy in Figure 1.

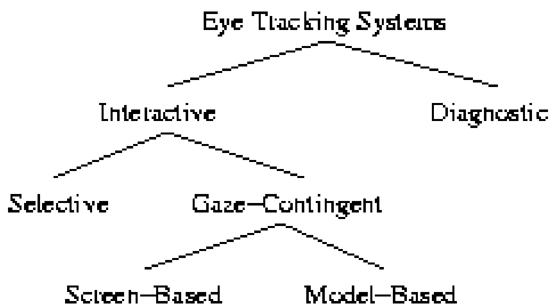


Figure 1: Hierarchy of eye tracking applications.

In contrast to interactive applications, diagnostic eye tracking systems typically involve the recording of eye movements over time, known as scanpaths, for analysis of the user's overt visual attention over a given stimulus. These systems generally do not require the display to react to the user's gaze. Numerous examples of general diagnostic applications can be found (see for example [DV00]), however, diagnostic eye tracking methods for PACG have not been widely adopted. This paper focuses on eye tracking techniques relevant to PACG, surveying interactive gaze-contingent and non-interactive diagnostic applications.

Interactive Systems

Interactive eye tracking methods for perceptually-adaptive graphics typically rely on the use of an eye tracker to manipulate the CG scene, contingent on the viewer's location of gaze. The motivation behind gaze-contingent systems is to minimize overall display bandwidth by reducing peripheral information in concordance with the perceptual limits of the HVS. Gaze-contingent displays are generally partitioned into two imperceptibly distinct spatial regions, a high-resolution foveal Region Of Interest (ROI) surrounded by a low-resolution peripheral region. There are two main approaches: screen-based and model-based. The screen-based approach deals with the manipulation of framebuffer contents just prior to display. The periphery is often masked or smoothed in some way, reducing the bandwidth by compressing the information (in bits-per-pixel) required to display or transmit the final image. Since early peripheral degradation attempts, development of screen-based approaches has continued to show promise for gaze-contingent display with minimal cost to either perception or performance. The model-based approach, while not as common, aims at reducing resolution by directly manipulating the geometric models prior to rendering. To increase display rates above those currently provided by view-dependent Level Of Detail (LOD) rendering methods, it has been suggested that an eye tracker is required to enable the presentation of high resolution portions of the scene or object only at the point of highest visual acuity, i.e., at the foveal ROI [LVCW+00].

Screen-Based Approaches

The idea of gaze-contingent displays is not new and dates back to early military applications, specifically eye-slaved flight simulators [Koc87,LTFW+89]. In the Super Cockpit Visual World Subsystem, Kocian considered visual factors including contrast, resolution, and color in the design of a head-tracked display. In their Simulator Complexity Testbed (SCTB), Longridge et al. included an eye-slaved ROI as a major component of the Helmet Mounted Fiber Optic Display (HMFOD). This ROI provided a high resolution inset in a low resolution (presumably homogeneous) field which followed the user's gaze. The precise method of peripheral degradation was not described apart from the criteria of low resolution. However, the authors did point out that a smooth transition between the ROI and background was necessary in order to circumvent the possibility of a perceptually disruptive edge artifact.

Since early flight simulators, various approaches have been developed for ROI-based image and video coding [LM00,PCN00,KG96,TEHM96,NLO94,ST94]. For screen-based Virtual Reality rendering, the work of Watson et al. is particularly relevant [WWH+97]. The authors studied the effects of LOD peripheral degradation on visual search performance. Both spatial and chrominance detail degradation effects were evaluated in Head Mounted Displays (HMDs). To sustain acceptable frame rates, two polygons were texture mapped in real-time to generate a high resolution inset within a low-resolution display field. The authors suggested that spatial and chrominance complexity can be reduced by almost half without degrading performance.

In an approach similar to Watson's, Reddy used a view-dependent screen-based LOD technique to evaluate both perceptual effects and system performance gains [Red98]. The author reported a perceptually modulated LOD system which afforded a factor 4.5 improvement in frame rate. It is not entirely clear how the LOD model was constructed, e.g., what was the method of degradation, nor is it clear what kind of apparatus was used. Reddy's empirical evaluation of the LOD model was performed on a 43.6 x 33.4 degree Field Of View (FOV) display, presumably a desktop monitor without the use of an eye tracker.

Although these reports are encouraging, it should be noted that head tracking alone appears to have been used as an estimate of gaze direction. The lack of an eye tracker may pose potentially serious implications on the interpretations of results since the true point of regard could not be verified.

Model-Based Approaches

The technique of simplifying the resolution of geometric objects as they recede from the viewer, as originally proposed by Clarke [Cla76], is now standard practice, particularly in real-time applications such as VR [Vin95]. Clarke's original criteria of using the projected area covered by the object for descending the object's LOD hierarchy is still widely used today.

However, as Clarke suggested, the LOD management scheme typically employed by these polygonal simplification schemes relies on pre-computed fine-to-coarse hierarchies of an object. This leads to uniform, or isotropic in terms of resolution degradation, simplification of objects.

A gaze-contingent adaptation of a model-based adaptive rendering scheme was proposed by Ohshima et al., where three visual characteristics were considered: central/peripheral vision, kinetic vision, and fusional vision [OYT96]. Their LOD algorithm generated isotropically degraded objects at different visual angles. Although the use of a binocular eye tracker was proposed, the system as discussed used only head tracking as a substitute for gaze tracking.

Isotropic object degradation is not always desirable, especially when viewing large objects at close distances. Numerous multiresolution mesh modeling techniques suitable for gaze-contingent viewing have recently been developed [ZS00]. Techniques range from multiresolution representation of arbitrary meshes to the management of LOD through peripheral degradation within an HMD where gaze position is assumed to coincide with head direction [LKRH+96,MJ96,Hop97,ZSS97,SS97]. Although some of these authors address view and gaze dependent object representation, few results concerning display speedup are as yet available showing successful adaptation of these techniques within a true gaze-contingent system, i.e., one where an eye tracker is employed. Due to the advancements of multiresolution modeling techniques and to the increased affordability of eye trackers, it is now becoming feasible to extend the LOD approach to gaze-contingent displays, where models are rendered nonisotropically.

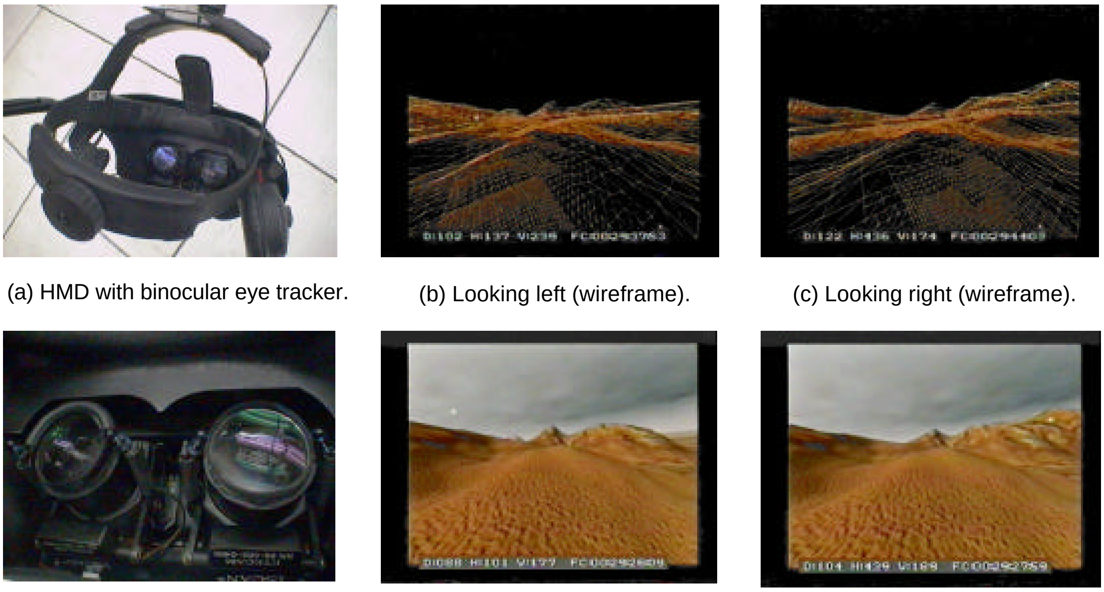
An early example of a nonisotropical model-based gaze-contingent system, where gaze direction is directly applied to the rendering algorithm, was presented by Levoy and Whitaker [LW90]. The authors' spatially adaptive near real-time ray tracer for volume data displayed an eye-slaved ROI by modulating both the number of rays cast per unit area on the image plane and the number of samples drawn per unit length along each ray as a function of local retinal acuity. The ray-traced image was sampled by a nonisotropic convolution filter to generate a 12deg. foveal ROI within a 20deg. mid-resolution transitional region. Based on preliminary estimates, the authors suggested a reduction in image generation time by a factor of up to 5. An NAC Eye Mark eye tracker was used to determine the user's Point Of Regard (POR) while viewing a conventional 19" TV monitor. A chin rest and immobilization strap were used to eliminate the need for head tracking.

Recently, Luebke and Erikson developed a gaze-directed LOD technique to facilitate the gaze-contingent display of geometric objects [LHNW00]. To test their rendering approach, the authors employed a table-mounted monocular eye tracker to measure the viewer's real-time location of gaze over a desktop display. While this work shows the feasibility of employing an eye tracker, the implementation framework used by the authors lacked a head tracker and required a chin rest to ensure tracker accuracy.

In our own work, we are developing gaze-contingent geometric modeling techniques for VR display. We use an off-the-shelf binocular eye tracker built into an HMD. In one scenario, we use the eye tracker as an indicator of gaze in a gaze-contingent multiresolution terrain navigation environment [DDGM00]. A surface, represented as a quadrilateral mesh, is divided into fixed-size (number of vertices) sub-blocks, which permits rendering for variable LOD on a per-sub-block basis. Resolution level is chosen per sub-block and is based on viewer distance. The resolution level is not discrete; it is interpolated between the pre-computed discrete levels to avoid ``popping" effects. The approach used is reasonably effective; however, it is not clear whether the technique is applicable to arbitrary meshes. More recently, we have developed an object-based LOD method similar to that of Ohshima et al. where objects are modeled for gaze-contingent viewing. However, unlike Ohshima's approach, our resolution degradation method is applied nonisotropically, i.e., objects are not necessarily degraded uniformly. Our modeling technique is based on that of Eck et al. [EDDH+95], where the original mesh is partitioned into Voronoi tiles and remeshed into multiresolution form. A 3D spatial degradation function is used for LOD selection. This function differs significantly from the area-based criteria originally proposed by Clarke. Instead

of evaluating the screen coverage of the projected object, our degradation function is based on the evaluation of visual angle in world coordinates.

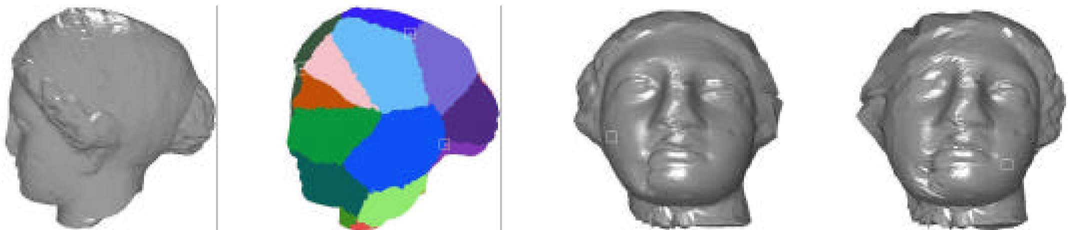
Besides the need to further develop nonisotropic model-based PACG techniques, a good deal of subsidiary problems remain. Due to the speed of saccadic eye movements, gaze-contingent displays often result in jerky, or noisy scene rendering. Real-time analysis of fixations would no doubt promote greater rendering stability. In addition, if the eye movement shows little variation, as is the case during a fixation, only the geometry subtended by the fovea need be reconstructed; cached geometry from previously rendered frames could be used in the periphery. Promising techniques are available which have yet to be exploited in gaze-contingent displays. Particularly exciting is real-time eye movement prediction, which can be used to alleviate the response lag inherent in all eye tracking systems.



(a) HMD with binocular eye tracker. (b) Looking left (wireframe). (c) Looking right (wireframe).

(d)Close up of binocular optics. (e)Looking left (rendered). (f)Looking right (rendered).

Figure 2: Martian terrain: gaze-contingent rendering.



(a)Original igea mesh. (b)Voronoi tiling (c)POR over left cheek (d)POR below right cheek.

Figure 3: Nonisotropic gaze-contingent geometric LOD rendering.

Diagnostic Systems

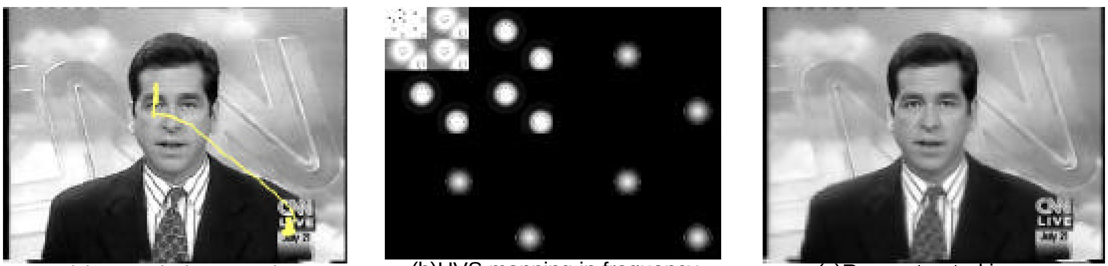
Human visual perception of imagery is an important contributing factor to the design of perceptually-based image and video display systems. Human observers have been used in various aspects of display design, including the development of corrective display functions (e.g., gamma function) dependent on models of human color and luminance perception, color spaces (e.g., CIE Lab color space), and image and video codecs. JPEG and MPEG both use quantization tables based on the notion of JNDs to quantize colors of perceptually similar hue [Wal91]. Perceptually adaptive image renderers (e.g., ray tracers and radiosity engines) have

been developed to efficiently encode similar spatial or color regions which are below the human JND threshold [BM98, FPSG96]. Often, however, these studies are based on automatically located image regions, which may or may not correspond to foveally viewed segments of the scene. That is, these studies do not necessarily employ an eye tracker to verify the ROI-based coding schemes. Instead, a figure-ground assumption is often used to argue for more or less obvious foveal candidates in the scene, or other signal-based approaches are used to identify candidate foveal regions. This research area is particularly suitable for diagnostic eye movement studies which can be used to corroborate the figure-ground assumption.

Diagnostic eye tracking systems for perceptually-adaptive graphics are not widespread, although these systems are generally easier to assemble than interactive ones. For non-immersive displays (e.g., a monitor vs. an HMD), a table-mounted remote monocular eye tracker is often sufficient to monitor gaze over the CG-generated stimuli. A scanpath recorded by such a device provides compelling evidence of the viewer's overt visual attention, which in turn facilitates fairly simple comparisons between ROIs automatically identified by the perceptually-adaptive algorithm against those actually foveated by the observer.

Instead of assuming a feature-based approach to foveal (high-resolution) image encoding, an eye tracker can be used to directly establish the foveal ROI. A suitable image degradation scheme may then be employed to render detail within this region (or regions). This motivated research by the author to find a suitable image degradation scheme which would match foveal acuity [Duc00]. Using a short video clip of a TV news anchor, foveal regions were selected over the TV anchor's right eye and another over the “timebox” located in the bottom right corner of the image. These ROIs were identified as commonly foveated regions by human subjects. The resolution of the reconstructed image drops off smoothly, matching visual acuity on the television display used during experimentation.

Spatial visual acuity is but one characteristic of vision that can be exploited in the design of non-interactive CG displays. Further degradation may be possible in terms of color, contrast, and motion, gaining greater perceptually-adaptive compression savings.



(a)Recorded scanpath (b)HVS mapping in frequency domain. (c)Reconstructed image

Figure 4: Image reconstruction and wavelet coefficient resolution mapping (assuming 50dpi screen resolution).

Conclusion

Examples of Perceptually-Adaptive Computer Graphics techniques where eye tracking can and does play a vital role have been discussed. Interactive gaze-contingent techniques are still relatively new, and many opportunities exist for advancing both screen-based and particularly model-based modeling approaches matching perception. Diagnostic eye tracking techniques for PACG are as yet underutilized. Clearly, objective evidence of where viewers actually foveate presents a benefit to any future PACG methodology where Regions Of Interest are selected automatically for perceptual compression.

This work was supported in part by a University Innovation grant (#1-20-1906-51-4087) and NSF CAREER award #9984278.

References

[BM98] Bolin, M. R., and Meyer, G. W., A Perceptually Based Adaptive Sampling Algorithm. In Computer Graphics SIGGRAPH'98 Proceedings (1998), ACM, pp.299--309.

[Cla76] Clarke, J. H., Hierarchical Geometric Models for Visible Surface Algorithms. Communications of the ACM 19, 10 (October 1976), 547--554.

[DDGM00] Danforth, R., Duchowski, A., Geist, R., and McAliley, E., A Platform for Gaze-Contingent Virtual Environments. In Smart Graphics (Papers from the 2000 AAAI Spring Symposium, Technical Report SS-00-04) (Menlo Park, CA, 2000), AAAI, pp.66--70.

[DKS00] Duchowski, A., Karn, K. S., and Senders, J. W., Eds., Eye Tracking Research & Applications (ETRA) (Palm Beach Gardens, FL, 2000), ACM. URL: http://www.vr.clemson.edu/eyetracking/et-conf/

[Duc00] Duchowski, A. T., Acuity-Matching Resolution Degradation Through Wavelet Coefficient Scaling, IEEE Transactions on Image Processing 9, 8 (August 2000), 1437--1440.

[DV00] Duchowski, A. T., and Vertegaal, R., Course 05: Eye-Based Interaction in Graphical Systems: Theory & Practice. ACM SIGGRAPH, New York, NY, July 2000. SIGGRAPH 2000 Course Notes, URL: http://www.vr.clemson.edu/eyetracking/sigcourse/

[EDDH+95] Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., and Stuetzle, W., Multiresolution Analysis of Arbitrary Meshes. In Computer Graphics (SIGGRAPH '95) (New York, NY, 1995), ACM, pp.173--182.

[FPSG96] Ferwerda, J. A., Pattanaik, S. N., Shirley, P., and Greenberg, D. P., A Model of Visual Adaptation for Realistic Image Synthesis. In Computer Graphics (SIGGRAPH '96) (New York, NY, 1996), ACM, pp.249--258.

[Hop97] Hoppe, H., View-Dependent Refinement of Progressive Meshes. In Computer Graphics (SIGGRAPH '97) (New York, NY, 1997), ACM.

[Koc87] Kocian, D., Visual World Subsystem. In Super Cockpit Industry Days: Super Cockpit/Virtual Crew Systems (Air Force Museum, Wright-Patterson AFB, OH, 31 March--1 April 1987), Air Force Systems Command/Human Systems Division/Armstrong Aerospace Medical Research Laboratory.

[KG96] Kortum, P., and Geisler, W. S., Implementation of a foveated image coding system for bandwidth reduction of video images. In Human Vision and Electronic Imaging (Bellingham, WA, January 1996), SPIE, pp.350--360.

[LW90] Levoy, M., and Whitaker, R. Gaze-Directed Volume Rendering. In Computer Graphics (SIGGRAPH '90) (New York, NY, 1990), ACM, pp.217--223.

[LKRH+96] Lindstrom, P., Koller, D., Ribarsky, W., Hodges, L. F., Faust, N., and Turner, G. A., Real-Time, Continuous Level of Detail Rendering of Height Fields. In Computer Graphics (SIGGRAPH '96) (New York, NY, 1996), ACM, pp.109--118.

[LTFW+89] Longridge, T., Thomas, M., Fernie, A., Williams, T., and Wetzel, P. Design of an Eye Slaved Area of Interest System for the Simulator Complexity Testbed. In Area of Interest/Field-Of-View Research Using ASPT (Interservice/Industry Training Systems Conference) (Brooks Air Force Base, TX, 1989), T. Longridge, Ed., National Security Industrial Association, Air Force Human Resources Laboratory, Air Force Systems Command, pp.275--283.

[LM00] Loschky, L. C., and McConkie, G. W., User Performance With Gaze Contingent Multiresolutional Displays. In Eye Tracking Research & Applications Symposium (Palm Beach Gardens, FL, 2000), ACM, pp.97--103.

[LHNW00] Luebke, D., Hallen, B., Newfield, D., and Watson, B. Perceptually Driven Simplification Using Gaze-Directed Rendering. Tech. Rep. CS-2000-04, University of Virginia, 2000.

[LVCW+00] Luebke, D., Varshney, A., Cohen, J., Watson, B., and Reddy, M. Course 41: Advanced Issues In Level Of Detail. ACM SIGGRAPH, New York, NY, 2000. SIGGRAPH 2000 Course Notes.

[MJ96] MacCracken, R., and Joy, K. Free-From Deformations With Lattices of Arbitrary Topology. In Computer Graphics (SIGGRAPH '96) (New York, NY, 1996), ACM, pp.181--188.

[NLO94] Nguyen, E., Labit, C., and Odobez, J.-M. A ROI Approach for Hybrid Image Sequence Coding. In International Conference on Image Processing (ICIP)'94 (November 1994), IEEE, pp.245--249.

[OYT96] Ohshima, T., Yamamoto, H., and Tamura, H. Gaze-Directed Adaptive Rendering for Interacting with Virtual Space. In Proceedings of VRAIS'96 (March 30--April 3 1996), IEEE, pp.103--110.

[PCN00] Parkhurst, D., Culurciello, E., and Niebur, E. Evaluating Variable Resolution Displays with Visual Search: Task Performance and Eye Movements. In Eye Tracking Research & Applications Symposium (Palm Beach Gardens, FL, 2000), ACM, pp.105--109.

[Red98] Reddy, M. Specification and Evaluation of Level of Detail Selection Criteria. Virtual Reality: Research, Development and Application 3, 2 (1998), 132--143.

[SS97] Schmalstieg, D., and Schaufler, G. Smooth Levels of Detail. In Proceedings of VRAIS'97 (March 1--5 1997), IEEE, pp.12--19.

[ST94] Stelmach, L. B., and Tam, W. J. Processing Image Sequences Based on Eye Movements. In Conference on Human Vision, Visual Processing, and Digital Display V (San Jose, CA, February 8-10 1994), SPIE, pp.90--98.

[TJ00] Tanriverdi, V., and Jacob, R. J. K. Interacting with Eye Movements in Virtual Environments. In Human Factors in Computing Systems: CHI 2000 Conference Proceedings (2000), ACM Press, pp.265--272.

[TEHM96] Tsumura, N., Endo, C., Haneishi, H., and Miyake, Y. Image compression and decompression based on gazing area. In Human Vision and Electronic Imaging (Bellingham, WA, January 1996), SPIE.

[Vin95] Vince, J. A. Virtual Reality Systems. Addison-Wesley, Reading, MA, 1995.

[Wal91] Wallace, G. K. The JPEG Still Picture Compression Standard. Communications of the ACM 34, 4 (April 1991), 30--45.

[WWH+97] Watson, B., Walker, N., Hodges, L. F., and Worden, A. Managing Level of Detail through Peripheral Degradation: Effects on Search Performance with a Head-Mounted Display. ACM Transactions on Computer-Human Interaction 4, 4 (December 1997), 323--346.

[ZS00] Zorin, D., and Schroder, P. Course 23: Subdivision for Modeling and Animation. ACM SIGGRAPH, New York, NY, 2000. URL: http://www.mrl.nyu.edu/dzorin/sig00course/

[ZSS97] Zorin, D., Schroder, P., and Sweldens, W. Interactive Multiresolution Mesh Editing. In Computer Graphics (SIGGRAPH '97) (New York, NY, 1997), ACM.

Perceptually driven criteria to accelerate interactive rendering

David Luebke, University of Virginia

We are interested in using perceptually driven criteria to accelerate interactive rendering. A great deal of excellent research has studied how to exploit the limitations of the human visual system to accelerate realistic rendering, such as global illumination algorithms. Metrics such as those published by Bolin and Meyer [Bolin 98 and Ramasubramanian et al [Ramasubramanian 00] bring sophisticated perceptual models to guide those algorithms. Their models account for coarse effects such as Weber's aw (threshold-vs-intensity) and contrast sensitivity over varying spatial frequency, as well as more subtle effects such as contrast masking. While many issues certainly remain to be studied in this realm, that research is at least well on its way.

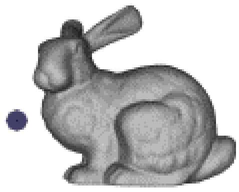


Figure 1: Perceptually driven polygonal simplification, illustrated on the Stanford Bunny with 3 different gaze points (point of user's gaze shown in blue). The simplifications shown here contain about 60,000 polygons, 34,000 polygons, and 12,000 polygons. At the appropriate field of view, these simplifications re imperceptible from the original. Both figures are courtesy the Stanford 3-D Scanning Repository.

On the other hand, very little work has investigated applying such metrics to the field of interactive rendering. Acceleration schemes in interactive rendering, such as the geometric simplification of polygonal models, is largely ad hoc. For example, the polygonal simplification community has long recognized that the human visual system deduces much about an object's shape from its silhouette, and so simplification algorithms have long included heuristics to preserve regions of high curvature, which are likely to form silhouettes [e.g., Rossignac 93]. However, the perception of shape also depends on the appearance and movement of specular highlights, so researchers have proposed heuristics to preserve regions where specular highlights appear and disappear [Xia 96].

Figure 2: We are investigating applying our perceptual metrics to other interactive rendering paradigms, such as the QSplat point-based renderer. Splats drawn in blue have been simplified. Perceptually driven rendering traverses nodes only where dictated by the local contrast and spatial frequency. Our model can also account for gaze direction: here, the user's gaze rests on Lucy's torch.

We are attempting to approach the integration of simplification and perceptual issues from a new and more systematic direction. One of our first principles is that taking advantage of perceptual effects generally requires the fine-grained control provided by a view-dependent

framework. These frameworks depart from the traditional approach to simplification, which creates several discrete levels of detail (LODs) in a preprocess. Instead, view-dependent simplification systems create a hierarchy of local simplification operations in which each operation eliminates only a few polygons [Luebke 97, Hoppe 97, Xia 96, DeFloriani 98]. In our work we consider the perceptual effect of each local simplification operation on the rendered image. We can achieve imperceptible simplification by only performing those operations for which the resulting change in the image is judged imperceptible. Our system is built as a set of callbacks on top of VDSlib, a flexible public-domain toolkit for performing view-dependent polygonal simplification.

In our preliminary technical report [Hallen 00], we describe a set of view-dependent simplification criteria that evaluate vertex merge operations according to the worst-case contrast and spatial frequency the merge could introduce in the final image. Using well-known measures from the psychophysical literature (e.g., [Campbell 68], [Kelly 75]), we derive a metric for the perceptibility of the merge operation, and only perform the merge when the resulting image will be indistinguishable from the original. User studies were used to verify the imperceptibility of the resulting view-dependent simplification. If "best-effort" rather than imperceptible simplification is the goal, our equations can order the merge operations according to the distance at which the user may be able to perceive the result. A greedy algorithm can then simplify the model to a set polygon budget while minimizing perceptual artifacts.

A particularly interesting aspect of this work is the optional incorporation of an eye-tracker to provide gaze-directed rendering. Gaze-directed rendering is not a new idea, but the fine granularity of view-dependent simplification seems to greatly improve on previous results by Oshima and Reddy [Oshima 96] [Reddy 97]. We have extended the perceptual model to account for eccentricity, or falloff in visual acuity in the peripheral vision [Figure 1], and performed user studies to verify the imperceptibility of the resulting simplifications. These preliminary experiments are encouraging indicators of the potential of gaze-directed rendering.

We have also applied our perceptually-driven framework to a completely different rendering paradigm, namely the point-based rendering system QSplat. Initial results are encouraging: while QSplat's standard high-quality mode only simplifies the model until each averaged sample is no larger than a pixel, we are often able to use much larger sample sizes (several pixels across) imperceptibly because of locally low contrast [Figure 2].

This research is in nascent form; while we believe that it breaks important intellectual ground, the overall rendering speedup is not nearly what we believe possible. Our immediate avenues of future work include studying more accurate techniques for evaluating the spatial frequency and contrast induced by a vertex merge, as well as implementing a more sophisticated perceptual model. Our current model is simple and somewhat bewhiskered, deriving from classic but decades-old psychophysical research. We would like to increase our efficiency and accuracy and incorporate additional perceptual phenomenon such as velocity and contrast masking.

References:

[Bolin 98] M. Bolin and G. Meyer. "A Perceptually Based Adaptive Sampling Algorithm", Computer Graphics, Vol. 32 (SIGGRAPH 98).

[Campbell 68] Campbell, F.W. and Robson, J.G. An Application of Fourier Analysis to the Visibility of Contrast Gratings, Journal of Physiology, 187 (1968)

[DeFloriani 98] L. De Floriani, P. Magillo, E. Puppo, Efficient Implementation of Multi-Triangulations, Proceedings IEEE Visualization '98, Research Triangle Park, NC, October 18-23, 1998.

[Hallen 00] Hallen, B., Luebke, D., Newfield, D., and Watson, B. "Perceptually Driven Simplification Using Gaze-Directed Rendering", University of Virginia Technical Report CS-2000-04, 2000.

[Hoppe 97] Hoppe, Hughes. "View-Dependent Refinement of Progressive Meshes", Computer Graphics, Vol. 31 (SIGGRAPH 97).

[Kelly 75] Kelly, D.H. Spatial Frequency Selectivity in the Retina, Vision Research, 15 (1975).

[Luebke 97] Luebke, David, and C. Erikson. "View-Dependent Simplification of Arbitrary Polygonal Environments", Computer Graphics, Vol. 31 (SIGGRAPH 97).

[Oshima 96] Oshima, Toshikazu, H. Yamammoto, and H. Tamura. "Gaze-Directed Adaptive Rendering for Interacting with Virtual Space", Proceedings of VRAIS 96 (1996).

[Ramasubramanian 99], Mahesh Ramasubramanian, S. Pattanaik, and D. Greenberg. "A perceptually based physical error metric for realistic image synthesis", Computer Graphics, Vol. 33 (SIGGRAPH 99).

[Reddy 97] Reddy, Martin. "Perceptually-Modulated Level of Detail for Virtual Environments", Ph.D. thesis, University of Edinburgh, 1997.

[Rossignac 93] Rossignac, Jarek, and P. Borrel. "Multi-Resolution 3D Approximations for Rendering Complex Scenes", pp. 455-465 in Geometric Modeling in Computer Graphics, Springer-Verlag (1993), Genova, Italy. Also published as IBM Research Report RC17697 (77951) 2/19/92.

[Xia 96] Xia, Julie, and A. Varshney. "Dynamic View-Dependent Simplification for Polygonal Models", proceedings of IEEE Visualization 96, 1996.

Perception-driven global illumination solutions

Karol Myszkowski, Max-Planck Institute for Computer Science

One of the basic goals of realistic rendering is to create images that are perceptually indistinguishable from real scenes. Since the fidelity and quality of the resulting images are judged by the human observer, the perceivable differences between the appearance of a virtual world (reconstructed on a computer) and its real world counterpart should be minimized. Thus, perception issues are clearly involved in realistic rendering and should be considered at various stages of computation such as global illumination computation, rendering, and image display. In this document we would like to focus on the problem of perception-guided global illumination computation and perception-based error metrics of resulting images.

Current global illumination algorithms usually rely on energy-based metrics of solution errors, which do not necessarily correspond to the visible improvements of the image quality. One of the research directions towards the perception-driven improvement of the global illumination computation performance relies on direct embedding of some simple error metrics to find the adequate level of light interactions between surfaces. Using such perceptual error metrics on the atomic level (e.g., every light interaction between patches, every mesh element subdivision) introduces a significant overhead on procedures that are repeated thousands of times in the course of the global illumination solution. This imposes severe limitations on the complexity of human spatial vision models, which in practice are restricted to models of brightness and contrast perception.

The scenario of embedding advanced HVS models into global illumination and rendering algorithms is very attractive, because more robust perception-driven guidance of the global illumination computation can be achieved. A practical problem arises that the computational costs incurred by the HVS models introduce an overhead to the actual lighting computation, which may become the more significant the more rapid the lighting computation becomes. The potential gains of such perception-driven computation can be easily cancelled out by this overhead, depending on many factors, such as scene complexity, performance of a given lighting simulation algorithm for a given type of scene, image resolution, and so on. The HVS models can be simplified to reduce the overhead, but then underestimation of visible image artifacts becomes more likely. To prevent such problems and to compensate for ignored perceptual mechanisms, more conservative (sensitive) settings of the HVS models should be applied, which may also reduce gains in the lighting computation driven by such models.

It seems that keeping the HVS models at some high level of sophistication and embedding them into rendering algorithms, which are supposed to provide a meaningful response rapidly, e.g., in tens of seconds or single minutes, may be a difficult task. The question arises what approach should be taken:

- Using advanced HVS models at the design stage of the global illumination algorithms and the tuning of their parameters. In such a case, the resulting algorithms can spend 100% of their computation time in lighting simulation, and the costs of HVS processing (which is performed off-line) are of secondary importance.
- Embedding advanced HVS models directly into the global illumination computation, which incurs significant overhead, but on the other hand better adaptation to the scene specific characteristics can be achieved.
- Using a hybrid of inexpensive energy-based metrics and more costly perception-based metrics as a trade-off solution.

Another question arises how to generalize the approaches chosen for static image generation toward the global illumination computation for dynamic environments, which becomes important for the high-quality animation rendering. I am looking forward discussing those issues during the campfire.

Collision Handling

Carol O'Sullivan and John Dingliana, Image Synthesis Group, Trinity College Dublin

People are very sensitive to physical events occurring around them... We know that one solid object cannot merge into another; We make decisions about the properties of objects based on the way in which they interact with each other; We judge whether objects are animate or inanimate depending on whether we perceive them as moving of their own volition, or being "caused" to move by another object (referred to as the perception of causality [Mic68]). Many studies have shown that these perceptual mechanisms are established very early on in infancy (e.g. see [BSW85])... but how accurate are they? Research in the realm of physics education has shown that most people have erroneous, yet very robust, pre-conceptions regarding the physical behaviour of objects [Cle82]. This is obviously not a good thing if you are trying to teach introductory mechanics, but it could be very useful if you are trying to get away with fast, yet plausible, physically-based simulations for real-time applications.

Despite significant advances in technology, real-time applications are still quite far from achieving photo-realistic rendering and mathematically precise simulation. Only through a great deal of simplification in the scene or the procedural models used to generate the animation, is it possible to even come close to achieving real-time frame rates. However, in most cases, the complexity of an animated scene, and the visibility of objects in the scene tend to change significantly over the course of the animation as objects move in and out of view and user focus. Many approaches try to interactively modulate levels of detail based on visibility information, culling the behaviours of objects outside the viewing frustum [CF97] or simplifying the procedures used to generate the motions of faraway entities [CH97]. Wherever tradeoffs are required, it is desirable to achieve them in the most efficient way possible. In the Physically Based Animation domain this entails:

1. That we guarantee target frame rates for real-time performance regardless of scene complexity
2. That we handle as much complexity as possible within the allowable time limit
3. That we get the best return for any simplifications that become necessary
4. That we minimize the overhead costs for the part of the system which manages the tradeoffs between complexity and speed

The first requirement for a system which will trade-off accuracy and speed is a mechanism that can handle processing at different resolutions and return consistent results. Refinable solutions process input data and return increasingly better results as more time is spent on the solution. A time critical or interruptible mechanism needs to be able to return a result as soon as it is instructed that it has exhausted all the processing time that has been allocated to it [Hub96]. For the purposes of Collision Detection, Hubbard's approach based on Hierarchical Sphere Trees is a prime example of a time-critical mechanism which returns refinable data. The approach is based on a Hierarchical Multiresolution sphere-tree representation of an object's volume (See Figure 1). Collision detection involves intersection tests between the nodes of the tree. When such collisions are found, and if there is time remaining, then a more accurate intersection test is performed between the children of the two colliding nodes.

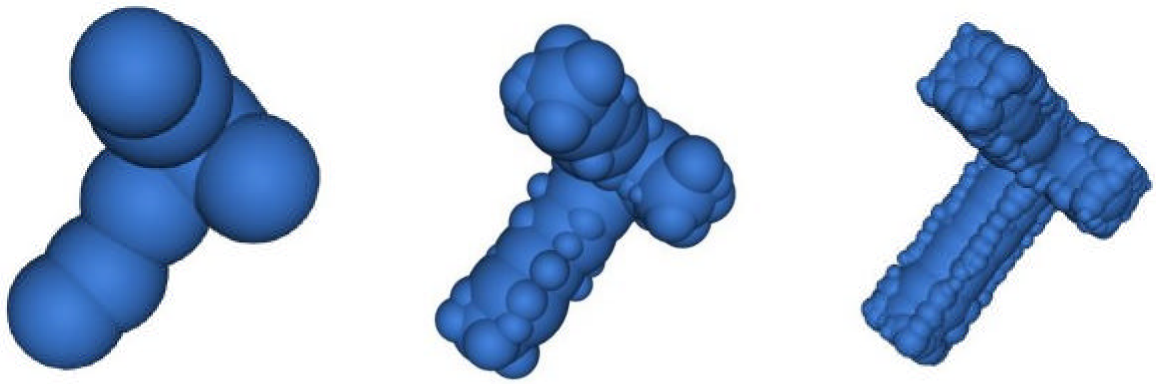


Figure 1: Representing an object with multiple levels of spheres.

We have extended this technique [DO00][DOB01] to return collision data, which can be used in computing physically based responses to object collisions. As with the collision detection process the accuracy of the data (such as details of contact points) is improved as the mechanism traverses deeper into the sphere tree. As we deal with higher resolutions of the volume model, the approximated contact points become increasingly accurate (see Figure 2). If the mechanism is interrupted, e.g. when it has used up its allocated time quota, then it immediately returns the most accurate approximation it has so far computed.

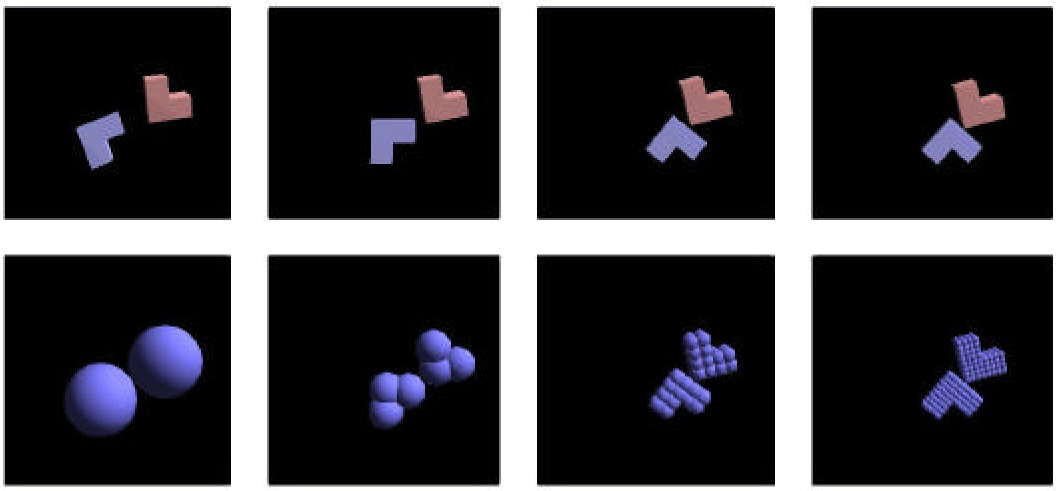


Figure 2: Multi-resolution collisions between objects The bottom row of images show the volumes used to perform collision detection at increasing levels of detail, while the top row shows what is actually seen by the viewer at the moment of impact.

We must remember that we are dealing with a viewer-centric model and that what we are trying to optimise is the perception of the animation rather than it's mathematical accuracy [BHW96][CF00]. In the viewer-centric model, the first thing we should note is that objects (and collisions between objects) further away from the user, due to occlusion or perspective fore-shortening become more difficult to judge and, as a result, errors and approximations in their states and behaviours become more difficult to notice. Furthermore, even a casual study will show that users' primary awareness of events in a scene focuses in a small radius around the point of fixation (see Figure 3). Other factors, which might influence the user's ability to judge an event on the scene, are properties of the objects e.g. size, shape, velocity; or properties of the scene and surroundings e.g. crowdedness, lighting, etc. Since the user's ability to notice error is non-uniformly spread across the scene, this suggests that processing time (and as a result simulation accuracy) for different events across the scene should also not be uniform. Instead, the solution we advocate is to prioritise events across the scene and distribute processing time based on this prioritisation.

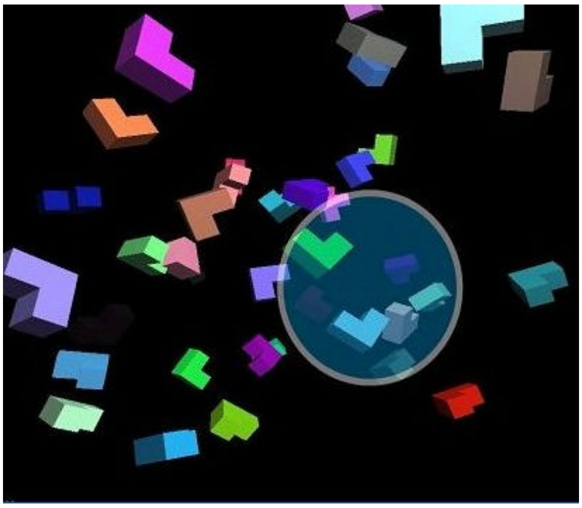


Figure 3: Important collisions, e.g. those close to the viewer's fixation position, should be processed first.

Early results indicate that the overhead from a full prioritisation and sorting of events in the scene on a per-frame basis becomes too high. A more fruitful approach is to use a small number of different priority groups into which events are interactively distributed. Each priority group is then allocated its share of processing time by the scheduler, with more processing being spent on higher priority groups. This method, whilst preserving a prioritisation scheme, bears considerably less overhead expense than a full continuous sort and in practice delivers good results even with very small numbers of priority groups [ORC99].

The level of optimisation depends a great deal on the quality of the metrics used to perform the prioritisation of objects in the scene. Although good results have been achieved by using "obvious" metrics such as object velocities, projected screen distances and distance from the user's fixation point (determined with the use of an interactive eye-tracker, see Figure 4) more extensive studies need to be performed to identify the most important factors which affect user perception of events and to determine how the influence of all such factors varies across a simulation scene.

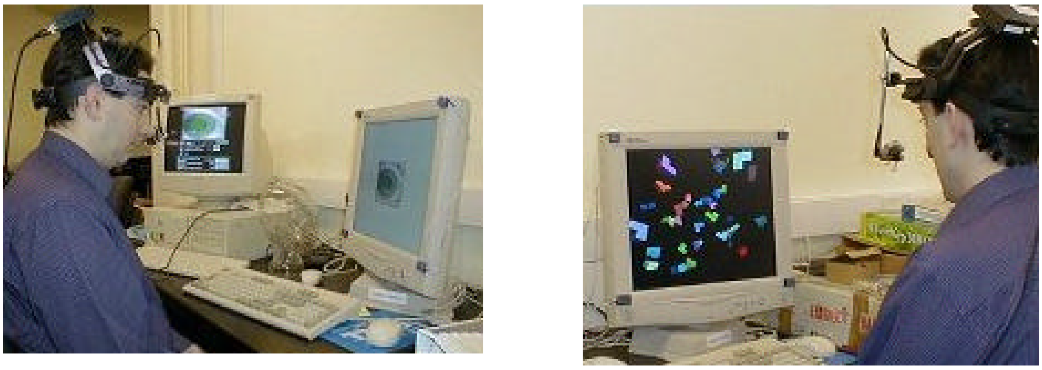


Figure 4: An eye-tracker is used to determine the viewer's point of fixation

We use psychophysical experiments to determine the factors that influence people's perception of dynamic events such as collisions and physical behaviours [OD01], with the purpose of developing dynamically-calculated metrics to drive the perceptual scheduling of our real-time adaptive physical simulations. Such experiments are very difficult to design, due to the high number of variables which need to be taken into account. Either the experiment needs to be reduced down to such a restrictive level of conditions, that the task is no longer representative of the real world (which is actually the case in most psychophysical

investigations in the vision literature), or more natural tasks are devised, thereby introducing an inevitable subjectivity or "fuzziness". How can we examine the effect of "bad" physical events (e.g. objects not touching before they bounce) without actually directing people's attention to them? We are also currently involved in designing "real vs. simulated" psychological experiments, where participants compare "real" dynamic scenes of multiple interacting objects with corresponding simulations at varying levels of detail.

References

[BHW96] Barzel R., Hughes J.F. and Wood D.N. - Plausible Simulation for Computer Graphics. Animation and Simulation '96. pp. 183-197. (1996)

[BSW85] Baillargeon, R. Spelke, E. and Wasserman, S. - Object Permanence in Five-month-old Infants. Cognition, 20:191-208. (1985)

[CH97] Carlson D.A. and Hodgins J.K. – Simulation Levels of Detail for Real-Time Animation. Proc. Of Graphics Interface '97. pp. 1-8. (1997)

[CF97] Chenny S. and Forsyth D. – View Dependant Culling of Dynamic Systems in Virtual Environments. In Proc. 1997 Symposium on Interactive 3D Graphics. (1997)

[CF00] Chenny S. and Forsyth D. – Sampling Plausible Solutions to Multi-Body Constraint Problems. In Proc. SIGGRAPH 2000 pp 219-228. (2000)

[Cle82] Clement, J. - Students' Preconceptions in Introductory Mechanics. American Journal of Physics, Vol 50. No. 1. 66-71. (1982).

[DO00] Dingliana, J. O'Sullivan, C - Graceful Degradation of Collision Handling in Physically Based Animation. - Computer Graphics Forum. Vol 19, Number 3 (Proc. Eurographics 2000), pp 239-247 (2000)

[DOB01] Dingliana, J. O'Sullivan, C. Bradshaw, G. - Collisions and Adaptive Levels of Detail - SIGGRAPH 2001 Sketches Program, LA (To appear) (2001)

[Hub96] Hubbard, P. - Approximating Polyhedra with Spheres for Time-Critical Collision Detection. - ACM Transactions on Graphics, Vol. 15, No. 3, July 1996, pp. 179-210 (1996)

[Mic63] Michotte, A. - The Perception of Causality - New York: Basic Books. (1963).

[OD01] O'Sullivan, C. Dingliana, J. - Real vs. Approximate Collisions: When can we tell the difference - SIGGRAPH 2001 Sketches Program, LA (To appear) (2001)

[ORC99] O'Sullivan C. Radach, R. Collins, S. - A Model of Collision Perception for Real-Time Animation - Animation and Simulation '99, pp. 67-76 (1999)

Measuring Visual and Temporal Fidelity

Benjamin Watson, Northwestern University, Computer Science

My research in perception and graphics has been motivated by the goal of using knowledge of perception -- and thus display fidelity -- to improve rendering for interactive computer graphics. In this effort, I found myself addressing many of the questions not answered (at least not directly) in existing perceptual results, and have performed many of my own experiments in collaboration with psychologists. Some of the results are of interest for non-interactive as well as interactive graphics applications. These collaborations have been fairly unique, involving engineering and cognitive psychologists rather than psychophysicists.

I believe that research using the goal-directed and top-down approaches of engineering and cognitive psychologists form a crucial complement to the psychophysically-based studies and systems described to date. As the organizers of this campfire have noted, the applied sorts of questions we graphics researchers tend to pose involve very many experimental variables, making it difficult to generalize from highly controlled psychophysically oriented results. In recognition of this, psychology has already made room within itself for “high-” and “low-level” psychology.

The first series of experiments we performed addressed the effectiveness of reducing detail in the periphery of interactive display, leveraging the drop in visual sensitivity with eccentricity. Participants performed a demanding search task that required heavy use of display peripheries. Results were encouraging [Wats97a]. Perhaps the most interesting result addressed the effectiveness of using only head tracking rather than eye tracking with this display technique. We found that modeling a head tracker as an eye tracker with 15 degrees of error worked quite effectively [Wats97b]. This confirmed previous results from the psychological literature in our applied graphics context.

An additional series of experiments examined the importance of temporal detail, in the form of frame rate and delay. In several studies [Wats97c, Wats98], we examined the relationship of temporal detail to the performance of several tasks. Among many other things, we found that temporal detail is more important in tasks that require constant feedback (e.g. placing something) than in tasks that do not (e.g. shooting). We also found that variation in temporal detail is relatively unimportant, though users are not completely insensitive to it.

These results provide guidelines for detail and fidelity management. However, they do not specifically address the tradeoff of visual and temporal detail. We began to tackle this issue by introducing the concept of I/O differencing. The basic insight is that delay of any sort introduces spatial inaccuracies, just as visual approximation does: objects are displayed where they *were*, rather than where they *are*. We implemented and experimented with a 3D rotation technique using I/O differencing in [Meru99]. We believe that although I/O differencing is a simple insight, it is a powerful one, and our experiments with it continue.

Finally, we have confronted the issue of visual fidelity measurement directly in a collaboration with cognitive psychologists [Wats00]. In our latest experiment, we display 3D models to participants and obtain three experimental measures of quality (naming time, ratings, and forced choice) in response to manipulations of three experimental variables (animal or object model, and type and amount of model simplification). We then used these results to evaluate three automatic measures of quality and fidelity (image MSE, the measure from Bolin and Meyer [Boli98], and 3D distance [Cign98]). The results are surprising and provocative, indicating among other things that at least for these stimuli, MSE is as good as the much more involved Bolin and Meyer measure. Perhaps more important, none of the automatic measures was very successful in predicting naming time – the measure which most closely indexes the process of recognition.

Our current research efforts are directed towards implementation of I/O differencing in a more complex 3D context, and following up our experimental results on fidelity measures with experiments using more varied stimuli, and ultimately improved automatic fidelity measures.

We believe the following questions deserve discussion at the campfire:

- How should visual fidelity be experimentally measured?
- For interactive applications, how can temporal fidelity be experimentally measured?
- How can automatic measures of fidelity be improved?
- How can fidelity be automatically measured during real time applications?
- What does fidelity mean in unnatural settings such as information visualization?

References

[Boli98] M. Bolin & G. Meyer. (1998). A perceptually based adaptive sampling algorithm. Proc. Of SIGGRAPH 98. In Computer Graphics Proceedings, Annual Conference Series, 1998, ACM SIGGRAPH, 299-309.

[Cign98] P. Cignoni, C. Rocchini & R. Scopigno. (1998). Metro: measuring error on simplified surfaces. Computer Graphics Forum, 17, 2, 167-174.

[Meru99] O. Meruvia (1999). Level of detail selection and interactivity. Masters thesis, University of Alberta, Dept. Computing Science. Available at: http://www.cs.nwu.edu/~watsonb/school/docs/ theses/iodifference.thesis.zip

[Wats97a] B.A. Watson, N. Walker, L.F. Hodges & A. Worden (1997). Managing level of detail through peripheral degradation: effects on search performance with a head-mounted display. ACM Trans. On Computer-Human Interaction, 4, 4, 323-346.

[Wats97b] B.A. Watson, N. Walker, L.F. Hodges (1997). Managing level of detail through head-tracked peripheral degradation: a model and resulting design principles. Proceedings ACM VRST '97, Virtual Reality Software Technology (Lausanne, Switzerland, September), 59-64.

[Wats97c] B.A. Watson, V. Spaulding, N. Walker & M.W. Ribarsky (1997). Evaluation of the effects of frame time variation on VR task performance. VRAIS '97, IEEE Virtual Reality Annual Symposium (Albuquerque, April, 1997), 38-44.

[Wats98] B.A. Watson, N. Walker, W.R. Ribarsky V. Spaulding (1998). The effects of variation in system responsiveness on user performance in virtual environments. Human Factors, Special Section on Virtual Environments, 40, 3 (Sept), 403-414.

[Wats00] B.A. Watson, A. Friedman, A. McGaffey (2000). Using naming time to evaluate quality predictors for model simplification. Proc. ACM Computer Human Interaction (CHI), 113-120.

Applications

Chair: Holly Rushmeier

High Fidelity Perception in Dynamic Scenes

Kirsten Cater, University of Bristol, UK

I am currently in the first year of my PhD research, entitled High Fidelity Perception of Dynamic Scenes, in the Department of Computer Science, University of Bristol. My research builds upon the techniques of computer graphics, visualisation and virtual reality which I learnt as an undergraduate in the department, graduating in July 2000 with a 2:1 BSc Honours degree.

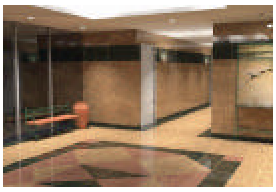
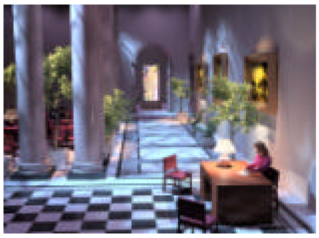
My main field of interest throughout my undergraduate training was that of computer graphics: for my undergraduate dissertation I chose to work on an ongoing project in the department, that of producing an aid for the visually impaired [1 & 2]. This project was of great interest to me as it also demonstrated the need for interdisciplinary research, i.e. it linked psychology with computer science, and my PhD continues with this interdisciplinary theme as it will utilise perceptual research.

My research will contribute to the significant interdisciplinary research being conducted in The Digital Media Research Laboratory [3] at the University of Bristol in realistic three-dimensional media content acquisition, processing and display. This research will be based on the use of the Explorer/1 simulator [4 & 5]. This simulator offers a new challenging delivery platform for investigating the high fidelity perception of scenes as the user is undergoing motion.

The major challenge in Virtual Reality is to achieve realism whilst keeping the computational time small and keeping up with real time control movements made by the user. With the Explorer/1 it is possible to obtain realistic motion and display 'blocky graphics' in real time. However, as soon as the rendering of these graphics is improved the rendering lags behind the control movements made by the user, thus destroying the effect of realism.



Some examples of the 'blocky graphics' achieved on the visual displays in the simulator with motion.

Here are some still Radiance images [6] which show some 'realistic' graphics, what we would like the display visuals to look like with motion in the simulator.

The main area therefore for my PhD research is to improve computer graphics in dynamic scenes for the motion simulator without simultaneously increasing the computational time taken to render the present graphics. One obvious way to solve this is by using a far more powerful system, but the economic constraints often preclude this. I think that perceptual research will provide the information to help me accomplish this task.

Currently I am looking into the use of the flaws in the human visual system as possible advantages in producing realistic computer graphics, for the human eye is good but it isn't perfect! It is interesting to consider how the eye's inability to perceive certain objects in images might be used to reduce the level of rendering detail, without the change being

perceptible to the viewer. The final goal, therefore, would be to use these flaws to save computational time rendering the 3D world we have created.

What I would like to learn from this campfire

Coming from a Computer Science background I understand how to create the graphics and program the simulator but my knowledge is lacking when it comes to the psychological and visualisation experiment techniques I will need to know and understand to perform the experiments. As I have just started my PhD most of my work has been researching into what people have already accomplished in this field rather than producing any results of my own. Thus I would be interested if anyone has done similar work in this field, how they accomplished it, what difficulties they came across, and any suggestions of how I should set up my own experiments.

Thus the main thing I wish to take advantage of at the campfire is the knowledge of other attendees.

The issues which I would like to discuss

- How important is it that the computer graphics in simulators are extremely realistic?
- Is it more important that in a simulator you have realistic motion and graphics that tie up with that motion than just realistic graphics?
- If we flood the senses, i.e. use motion, visuals and sound, can we get away with a lower level of rendering detail in the visuals, than if we just use the visuals?
- What sense of motion can you get just from the visuals? e.g. In a real life example, if a lorry starts rolling backwards when you are stationary in another vehicle next to it, you feel like you are moving forwards. Is it possible to simulate this effect?
- How can information about change blindness & blindsight be used to make sure that the user perceives important objects/icons in the case of user interfaces?

References

[1] http://www.cs.bris.ac.uk/Research/Vision/lowvision.html - ACRC Research Projects - Mobility aid for Low Vision.

[2] http://www.caterk.freeserve.co.uk/WebPage.html - Image Processing using Neural Networks - Producing an aid for the Visually impaired.

[3] http://www.cs.bris.ac.uk/Research/Digitalmedia/ - The Digital Media Research Group, University of Bristol.

[4] http://www.motionbase.com - The motionbase website.

[5] http://www.cs.bris.ac.uk/~cater/MotionSim.html - Some pictures showing the motion platform.

[6] http://radsite.lbl.gov/radiance/gallery/gallery.html - A gallery of images and animations produced using Radiance.

High Fidelity Image Synthesis of Archaeological Reconstructions

Kate Devlin, University of Bristol, UK

I am in the first year of PhD research in the Department of Computer Science at the University of Bristol, investigating high fidelity graphics with particular regard to archaeological reconstructions. My undergraduate degree was a BA Hons in archaeology, and following two years of working in field archaeology with an emphasis on draughting and illustrating, I returned to university to complete an MSc in Computer Science and Applications.

I am interested in the potential of realistic graphics as an aid to the archaeologist, both as a research tool and as a method of interpretive display. It stems from a desire to go beyond photo-realistic modelling to create physically realistic scenes, so that a numerical simulation rather than an artist's conception is produced, offering us a more valid view of past environments. This involves precision modelling of the distribution of light energy in such scenes, and my work will focus on flame lit environments. This interdisciplinary perspective offers me the unique opportunity to engage in research into realistic graphics and also to answer questions raised from an archaeological perspective.

The aim of realistic image synthesis is the creation of accurate, high quality imagery that faithfully represents a scene, be it real or imaginary. Computer generated imagery indistinguishable from the real physical environment will be of substantial benefit in a wide variety of fields, and in particular the reconstruction of archaeological sites. Advances in rendering techniques allow us to simulate the distribution of light energy in a scene with great precision, however a number of fundamental issues still have to be addressed if we are to ensure that the displayed image does in fact have a high fidelity visual appearance. These issues include the limited dynamic range of displays, any residual shortcomings of the rendering process, and the extent to which human vision encodes such departures from perfect physical realism. A good quantitative understanding of the spatial, temporal, chromatic and three dimensional properties of human vision will lead to more realistic and efficient rendering algorithms by concentrating computational effort on those areas of the image which are important to the human eye. The need to establish a metric for realism through comparison is essential if the images are to be of full use, and work in the future will attempt to quantify how "real" realistic graphics actually are.

What I wish to learn from this campfire

As this is the first year of my PhD my work so far has involved research and reading into the theory and the existing applications and techniques of realistic graphics. This campfire offers me the opportunity to pose questions to the experts who will be attending and discuss the relevant issues of my research subject, thus allowing me to progress with the practical aspects of my study.

Issues I wish to discuss

How "real" can an archaeological reconstruction be, given that all archaeology is in a sense "virtual" - we are trying to recreate scenes that no longer exist. How can a metric for realism be established if we have no physical scene with which to compare it? In flame-lit environments the range of light can vary greatly over short distances. How can we compensate for this? Is tone mapping the answer? Does increased realism lead to misinterpretation? Are we more likely to take a reconstruction at face value and ignore alternative interpretations if it appears "real" to us?

Integrative Displays for Anesthesiology

Frank A. Drews, Department of Psychology, University of Utah

Background

I began my training as a cognitive psychologist interested in how different ways of presenting information influence peoples' ability to make judgments and decisions. In my dissertation I showed that the more complex the presentation of information, the more judgment errors people made based on that information. However, when information was integrated in presentation, thus reducing the cognitive 'work' that people had to do, their judgments improved. Because our visual system is fast and powerful, I believe that complex information can best be integrated for people via visual presentation. My post-doctoral position (since March 2000) has allowed me to pursue this notion in the domain of anesthesiology.

Research

In anesthesiology information about the status of the patient is presented on computer displays as a combination of digital and chart-based information. The design of traditional displays has been driven by the single sensor single indicator approach: Information about the values of variables measured by a single sensor is displayed with a single indicator on the monitor. Given that a typical display shows over thirty variables in real-time, the information-processing capacity of the average human is clearly exceeded. One reason for the high number of anesthesiology related errors is that information is presented in sub-optimal ways. Existing displays do not support the anesthesiologist in detecting problems, making diagnoses and treating the patient.

An alternative approach is to develop an integrative display in which information from different sensors is integrated into a single graphical object (e.g. a sphere) (Figure). The basic assumption is that integrated information reduces the cognitive workload of the anesthesiologist, because she need not integrate the information herself. Moreover, using a graphical object permits exploitation of the human visual system's sensitivity to change in objects. Thus, deviations from perfect geometric shapes can be used to indicate changes. Our empirical evaluation supports the notion that visual integration is a good strategy. For example, in an interdisciplinary collaboration with Dwayne Westenskow (Bioengineering, Anesthesiology), Noah Syroid (Bioengineering), Julio Bermudez (Architecture), I ran a study where we compared anesthesiologists' performance in simulated surgeries using an integrated display and a standard display. We were especially interested in the time needed to detect changes in variables and to diagnose the cause for changes. Using integrated objects to provide information reduced detection time and improved diagnosis.

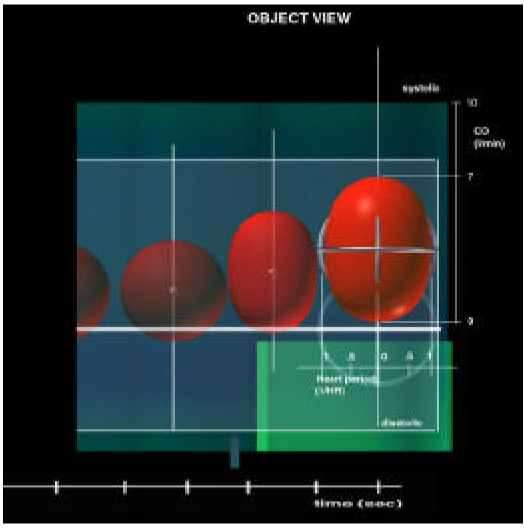


Figure : Integration of variables (heart object)

In another study we examined the effect of a new and innovative way of displaying drug information to anesthesiologists. In addition to other tasks anesthesiologists must constantly supervise patients' drug dosages to maintain appropriate levels of analgesia, anesthesia, and neuromuscular blockade. We wanted to see if integration f information assists them in doing so. We provided anesthesiologsts with integrated information about effects of drugs on simulated patients. Here we were interested in he effect of visually displaying the pharmacodynamic effects of drugs on nesthesiologists cognitive workload and drug dosage management. Providing nesthesiologists with the drug display reduced workload, and improved drug dosage anagement.

In both studies, providing physicians with information that is already integrated improved their performance. In the applied context of the operating room, the result would be faster detection and diagnosis of problems, more effective drug administration, and an overall reduction of human error.

In addition to the studies reported above we ran several experiments where we psychophysically scaled the objects, which will be part of a new anesthesia display. Additionally, we did some experiments where we tested the controllability of these objects. Currently, we are running subjects in using eye tracking (ASL 501) to determine precisely which information anesthesiologist's extract from the newly designed displays. The participants in these experimental studies are performing anesthesia in a high fidelity simulator in the Patient Simulation Center at the University of Utah. They have to treat two evolving critical scenarios. By recording fixations, pupil diameter and other performance data like detection times, diagnosis time and quality of diagnosis, we will be able to determine the improvement in performance when using the integrated display. Further it will be possible to determine precisely which parts of the display are used to extract relevant information, and which parts are not relevant for the anesthesiologist.

All reported studies are part of a larger interdisciplinary research program (Grant National Institute of Health Grant (IRO1 HL64590). The goal of this project is to develop and to evaluate a new, integrated display for anesthesiologists, which uses principles of graphically integrating information. The project is an interdisciplinary collaboration between Bioengineers, Anesthesiologists, Cognitive Psychologists, Computer Scientists and Architects.

Interests

The topics I am interested in, are focussed around experimental issues and issues of perception and visualization.

With regard to perception and visualization I am interested in questions about ways of presenting integrated information more efficiently. This includes questions about using space, texture, and color on a display to provide information about differences and similarities of elements. Further, I am interested in using variables like eye-movements and fixations to determine extraction of information for decision-making.

With regard to experimental issues I am interested in evaluation techniques of displays and methods of validating the effectiveness of new experimental techniques.

Shared Situational Awareness in Cockpits

Lucy Emery, Defence Evaluation and Research Agency (DERA), UK

I am attending the campfire in place of Dr. Helen Dudfield (AD, Cockpit System Engineering and Integration, Air Systems, Defence Evaluation and Research Agency). Helen and I have been working on a number of different projects together, mainly investigating the benefits and aiding the design of different display types (mainly in military environments), such as large screen (or panoramic) displays for command and control teams.

My background has mainly been focussed around cognitive psychology. I have a BSc in Applied Psychology and I am nearing the end of a taught MSc in Cognitive Neuroscience. I have two and a half years experience working for the Defence Evaluation and Research Agency; firstly in Air Systems, in the Human Factors group, now for the Centre for Human Sciences. The Centre for Human Sciences is a well-established group with a wide range of experience and expertise. The centre's capabilities include engineering and system psychology, and this is the main focus of my work. Particular interests include neuroscience, perception, vision, aircraft and other military displays, visualisation, attention, eye-movements and eye-tracking.

I am attending the campfire to the benefit of two current projects. The first project is funded by MoD Corporate Research Programme Technology Group 5: Human Sciences and Synthetic Environments: Campaign Combat Information Management For Future Command (CCIMFC). The second is an EC 5th Framework project - Visual INTeraction and Human Effectiveness in the Cockpit, Part II (VINTHEC II).

The aim of CCIMFC is to develop a novel visualisation tool for command teams. Command and control operations depend heavily on the interpretation, processing and conveying of large amounts of data. This often results in command personnel becoming overloaded with information. Currently, no adequate visualisation tool exists to support command teams in situation assessment and facilitate their selection of appropriate courses of action. This MoD funded research is being conducted in response to this identified gap between the existing technology and the human operator. The problem is, how to design the best display to achieve this aim. It seems logical that a key factor in designing optimal displays must be consideration of how the brain processes information. On this basis, the project started by conducting literature reviews on the visual system. The objective here was that if we can understand the human visual system we should be able to design displays to be most appropriate for it. However, the gulf between research into visual systems and display design appears to be large, and one that applied designers often struggle with. By attending the campfire, I am hoping that we can discuss how attempts may be made to integrate the two, how research as diverse as single cell recording can be applied directly to real-world problems and how we can successfully build collaborations between neuropsychology and applied design. I believe this to be a logical way to advance further with display design.

VINTHEC aims to provide an objective means of assessing shared situational awareness, in the cockpit or in flight simulators, and to assess eye-point-of-gaze (EPOG) measurements in these settings. Another aim is to find new and better ways to analyse and present the data recorded, especially with respect to EPOG data and data fusion. The immediate problems faced by this objective are:

??How do we approach the measurement of shared data (i.e. two pilots simultaneously, in real time)?

??How should the data be collected and analysed?

??What is the optimal way to display the data?

??What do eye movements really tell us about situational awareness?

It is hoped that campfire discussions will elucidate some answers to these questions, identify potential data visualisation techniques appropriate to the needs of the project and improved analysis techniques, specifically related to the concept of shared situation awareness.

New Display Technology for Anesthesiologists

Dwayne Westenskow, University of Utah

This project seeks to develop new displays for visually representing physiologic variables, to enhance a clinician's ability to see and rapidly respond to critical events. Unexpected incidents are common; anesthesiologists face them during 20 percent of all anesthetics. One quarter of these incidents pose significant danger to patients. Human factors research has shown that graphic displays improve an anesthesiologist's ability to detect and identify critical events. Groups that observed graphic displays saw changes 3.1 minutes sooner than those observing traditional displays. Erroneous decisions were reduced from 4.1% to 1.4% and human response times (the time used to correct the problem) were one-third the original time.

The research plan is to develop a physiologic display that increases situation awareness and shortens the time needed to detect a change, diagnose the cause of the change and treat the event. The display is to provide a comprehensive view of the surgical patient's physiologic state. If some of the variables are not being monitored, the display will use models to predict "population normal" values, thus completing the physiologic picture. The physiologic icons in the display will be organized to enhance the vision of the interrelationship between organ system function. The display will highlight input/output relationships, by showing drug concentrations and accompanying physiologic changes. Simulated sounds, which accompany the display, will enhance the serial interpretation of patient variables. The display design process will be iterative, with six cycles of design, computer implementation, evaluation, critique, and redesign.

The display will be evaluated in a full-scale patient simulator environment. Twenty-four anesthesiologists will be asked to treat simulated patients in cases where 12 critical events occur. They will be asked to think aloud as they treat the critical event, so that four stages of situation awareness can be detected. At the conclusion of six of the simulation scenarios, the anesthesiologist will be asked a set of 12 questions to assess their situation awareness. Each display will be evaluated with eye-tracking to identify the parts of the display that are most useful in decision-making.

The anesthesiologist "sees" the patient on the operating table through the monitoring display. The proposed research will identify the type of display, which best helps the anesthesiologists to rapidly detect physiologic changes, to make accurate diagnostic decisions, and to efficiently treat critical events. The final display should enhance patient safety during anesthesia.

Scene Perception

Chair: Tom Troscianko

Use of synthetic imagery for camouflage assessment

Marilyn Gilmore, DSTL

CAMEO-SIM has been developed as a physics-based, broadband, scene simulation toolset to enable the quantitative evaluation of both current and future camouflage in visible and infrared wavebands. This is necessary because field trials are expensive, subject to the vagaries of the weather and cannot be used to design new systems. The software development has been undertaken by Hunting Engineering Ltd under contract to DERA for the UK MoD.

The goal of the CAMEO-SIM system is to produce synthetic, high resolution, physically accurate radiance images of target vehicles in operational scenarios, at any wavelength between 0.4 and 14 microns. The software was developed with a scaleable rendering kernel in which imagery can be produced at different fidelities and frame rates depending on the image application and wavelength of operation.

CAMEO-SIM has two key elements that distinguish it from conventional ray-tracing packages. Firstly, there is a complete audit trail between the material properties used by the renderer and the final rendered image. This means that it is possible to conduct carefully controlled parametric manipulations of scene properties, which can be repeated. It allows a scientific investigation of important variables. Secondly because CAMEO-SIM is physics based it is predictive. That is, once a particular scenario has been created it can then be used to predict effects under different conditions such as diurnal variations and weather conditions. The same scene geometry can be used and different material properties assigned for different times of the year. Traditional ray-tracers rely on arbitrary parameter manipulation in order to simulate changes in the environment (e.g. change of atmosphere). CAMEO-SIM on the other hand incorporates these directly as a result of solving the underlying physical equations.

All geometric objects forming the synthetic environment are modelled using textured faceted structures. Texel values in these textures are mapped to real materials, which have measured physical properties, associated with them e.g. bi-directional reflectance, solar absorptivity, conductivity, and density. Each texel is then considered as a mixture of up to three different materials. This means that the physical properties of both soil and grass are modelled when using a ‘grass’ texture. Complex objects are modelled as a number of polygons. This means that the three dimensional effects of trees, including shadow effects can be simulated.

CAMEO-SIM computes the radiance in specified sub-bands for each pixel in the image. These sub band radiance images can then summed to produce an ‘in-band’ radiance image.

The software can display each sub-band as a grey scale image or any three sub-bands as a false colour image. CAMEO-SIM can also display visible band true-colour imagery. This is done by first evaluating the spectral radiance image cube for a user-defined number of sub-bands between 0.38 and 0.78 microns. The spectral image cube is then converted into device independent colour space represented by the tristimulus values X, Y, Z, using the CIE 1931 Colorimetric Standard Observer. The monitor that is used for the image display is calibrated both in terms of luminance and phosphor radiance allowing the X, Y, Z values to be converted to R, G, B values. Various luminance transforms can be employed to make best use of the limited CRT dynamic range.

The scaleable rendering kernel means that images can be rendered at different levels of fidelity. In the lowest fidelity mode all surfaces are diffuse and no shadow effects are modelled. The high fidelity renderer can model directional and bidirectional reflectance effects, geometric occlusion of point and extended sources, and ensures spectral integration of the optical properties with the atmospheric illumination terms to ensure correct radiative transport. As may be expected the higher fidelity images will take longer to render so the level of fidelity required can be selected before rendering the image, so that appropriate imagery can be generated for different applications.

Observer trials
CAMEO-SIM produces a cube of data with the spectral and spatial information. The images produced by CAMEO-SIM can also be displayed on a monitor for observer trials These observer trials are conducted under controlled conditions in the laboratory. The CAMEO-SIM imagery is played back onto a computer display that has been carefully calibrated both colorimetrically and for luminance output in order to provide true-colour imagery to observers (see input from Dr I R Moorhead).

Fidelity issues/validation
CAMEO-SIM has been successfully subjected to range of verification tests. Validation work has included tests where the imagery is compared to measurements made on a simple scene. The scene was created within CAMEO-SIM. However validating a model against the real world is notoriously difficult. All the input parameters must be carefully measured, so that the model has the best chance of producing a solution that is close to that which exists in reality.

One of the difficult problems is assessing what is an appropriate level of fidelity for different tasks, such as target detection. A research programme is investigating different methods of quantifying image fidelity. A range of techniques is being evaluated including models based on human vision performance (see input by Dr T Troscianko and Dr D Tolhurst) and statistical analysis of images using higher order statistics.

In the first year a large data set of imagery of a site on Salisbury Plain in the UK was obtained. Similar imagery was created using CAMEO-SIM, in different wavebands, and rendered at different levels of fidelity e.g. using different fidelity trees, with and without textures, and with different shadow options.

Statistical analysis of the images has looked at local effects by comparing a synthetic image with a similar real image, and global effects by comparing image statistics of real and synthetic scenes. The statistics can use different 'filters' such as blob and horizontal or vertical bar filters.

Analysis using the different metrics has shown that similar trends are observed using all metrics. However the impact of the 'global' fidelity on the ability to detect a target (a 'local' effect), has yet to be assessed.

Further work/areas of interest
The main issue is how to evaluate how 'real' synthetic imagery is. Methods to quantify the effect of different rendering techniques on image realism must be identified. The impact of human perception on the required fidelity must be assessed so that appropriate imagery can be rendered for different applications. For example; different effects might have to be captured when using imagery in a flight simulator compared with when the imagery is used to assess how easily different targets can be detected in a cluttered environment. Another issue is: how important is movement? Is the overall image fidelity different when the whole image is moving, compared with when an object is moving within the scene?

Target signature assessment methodologies for synthetic imagery

Ian Moorhead, DERA

Generation of synthetic imagery is not generally the end of the process. The imagery needs to be viewed, either by humans or by machine depending upon the application. Our work is primarily concerned with making use of synthetic imagery to assess target conspicuity.

Out imagery is generated by CAMEO-SIM which is a physics based image rendering system and therefore allows parametric studies to be undertaken (see the presentation by Marilyn Gilmore for more detail). That is, once a particular scenario has been created it can then be used to predict effects under different conditions such as diurnal variations and weather conditions. The same scene geometry can be used and different material properties assigned for different times of the year.

Three methodologies are being developed to assess target signatures.

Observer Trials

This is a traditional psychophysical approach. Images of static or moving target are displayed to subject and performance is measured, usually in terms of probability of detection and detection time. We are currently looking at methods to achieve more realistic light levels within the laboratory (typical monitor luminances are 10 – 50 cd m^{-2}, whereas light levels in the outside world can be 3-4 orders of magnitude higher). One method being considered is direct viewing of images produced by an LCD projector which ahs the potential to generate considerably higher light levels.
Another issue that arises in these experiments is how to map the very large dynamic range values produced by CAMEO-SIM into the limited range available on displays.

Contrast Metrics

This is a computational method of assessing target signatures that uses a multiscale contrast calculation (Peli), and in some sense approximates early stages of vision.

Vision Modelling

This third method uses a fully developed image based vision model – GTV developed at the Georgia Tech research Institute. The model simulates preattentive and attentive vision processes and includes processing of colour and simple motion. DERA is currently working with Georgia Tech on validating and developing the model with the aim of applying it to analyse synthetic imagery.

Each of these topics will be discussed within the context of assessing synthetic imagery and examples using images of moving aircraft and static land vehicle target detection will be presented.

Statistical approaches to image and scene manipulation

Erik Reinhard, University of Utah

The field of visual perception aims to study the human visual system. It is recognized that in order to understand a system, it is important to understand its input. This has led to the sub-field of natural image statistics, which tries to understand images in statistical terms. Images' statistical properties are usually studied by collecting a number of such images in an ensemble and computing first, second or higher order statistics on them.

One of the most well studied natural image statistics is that of the power spectrum, a second order statistic. By computing the Fourier transform of an image and multiplying each element of the transform by its complex conjugate, the power spectrum is obtained. Averaging over all directions gives power as function of frequency. For most natural image ensembles, plots of this quantity result in straight lines with a slope of around 1/f^2, where f is the spatial frequency (usually measured in cycles per image) [2,3,4]. This is a special property of natural images, which is generally not obtained for random images (such as for example random noise images, which would produce a flat power spectrum). The 1/f^2 spectral slope of natural images means that equal power is encoded in each frequency band. It also implies that natural images are statistically scale-invariant.

While the reason for this to occur is still being debated (e.g. [1]), these results are important for graphics applications. It has been shown that certain image interpretation tasks are negatively affected when the spectral slope deviates too much from -2 [6,7]. It is therefore argued that the human visual system expects to see images that conform to this statistic. As graphics applications produce input to the human visual system, it makes sense to optimize this input to be as easy to interpret as possible. Conforming to second order statistics is one step in that direction.

Although individual images may have spectral slopes that are somewhat different from 1/f^2, the variation between images is relatively small. Hence, the power spectrum provides a simple and elegant means to assess the quality of an image: if the power spectrum does not yield a straight line and if the slope of this line deviates much from -2, then the image is less realistic in a statistical sense.

Recent research has shown that the power spectrum is sensitive to modeling, but is quite insensitive to variations in rendering. Different levels of image compression schemes (gif, jpeg), anti-aliasing, gamma-correction, and choice of rendering algorithm all have at most a small impact on this second order statistic [9]. On the other hand, changes in geometry tend to have a much larger influence on the power spectrum. This establishes the power spectrum as a useful tool to assess quality of modeling. It may therefore be used to guide modelers to construct scenes with content that is perceived to be natural. Applications include parameter optimization for fractal terrains, displacement mapping and solid texture creation.

The quality of modeling is only one factor that influences the perceived naturalness of rendered images. A second issue with many synthetic images involves the selection of color which may affect image quality. Here, statistical methods may be used to impose the look and feel of a photograph upon a rendered image, thus improving perceived realism. Computing simple statistics on the colors of a photograph and applying those to the colors of a rendered image, may convert an obviously synthetic image into a more realistic looking image [9].

The method to transfer the look and feel of one image to another, involves conversion of both source and target images to a perceptually based color space where the three color axes are decorrelated. Decorrelation can be achieved by applying a Principal Components Analysis to each image. However, Ruderman et. al. [10] have shown that for natural image ensembles, the resulting axes have simple forms and interpretations, forming a new color space. The first principal component is an achromatic channel, while the second and third are yellow-blue and red-green color opponent pairs. Rather than computing PCA for each image separately, images given in LMS color space can be converted to a color space which on average decorrelates its axes using the following conversion matrix [5,10]:

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \log L \\ \log M \\ \log S \end{bmatrix}$$

Note that the conversion also takes the pixels to log space, which conditions the data by no longer requiring the data values to be positive. The data representation is more compact and in log space uniform changes in stimulus tend to be equally detectable. Thus, the l-alpha-beta color space promises to be useful for color manipulation.

The fact that the axes in l-alpha-beta color space are decorrelated is important, because it allows data along each axis to be modified separately, without affecting the other two axes. To transfer statistical properties between images, we compute mean and standard deviation of the pixels along each axis separately for both rendering and photograph. Then, all data points of the rendering are scaled and shifted to assume mean and standard deviation of the photograph. By properly matching a rendering with a photograph, convincing results may be obtained, as shown in Figure 1.

The work summarized in this abstract indicates that research into the statistical properties of natural scenes provides important clues as to how graphics algorithms may be optimized to generate imagery that is well-suited to be interpreted by the human visual system. Statistical approaches may provide a good alternative to assess image quality, generate scenes and modify existing images. Many more applications for this work can be anticipated.

Figure 1: The atmosphere of Vincent van Gogh's "Cafe Terrace on the Place du Forum, Arles, at Night" (Arles, September 1888, Oil on canvas) applied to a photograph of Lednice castle near Brno. In this example, the colors of the sky in both images, the yellows of the cafe and the castle and the browns of the tables at the café and the people at the castle were matched separately using small swatches [8].

References:

[1] Rosario M Balboa, Christopher W Tyler, and Norberto M Grzywacz. ``Occlusions contribute to scaling in natural images". Vision Research, 2001. In press.

[2] G J Burton and I R Moorhead. ``Color and spatial structure in natural scenes." Applied Optics, 26(1):157--170, January 1987.

[3] Dawei W Dong and Joseph J Atick. ``Statistics of natural time-varying images." Network: Computation in Neural Systems, 6(3):345--358, 1995.

[4] David J Field. ``Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems." In M Farge, J C R Hunt, and J C Vassilicos, editors, ``Wavelets, fractals and Fourier transforms", pages 151--193. Clarendon Press, Oxford, 1993.

[5] P Flanagan, P Cavanagh, and O E Favreau. ``Independent orientation-selective mechanism for the cardinal directions of colour space". Vision Research, 30:769--778, 1990.

[6] C A Parraga, T Troscianko, and D J Tolhurst. ``The human visual system is op-timised for processing the spatial information in natural visual images". Current Biology, 10(1):35--38, January 2000.

[7] Stephane J MRainville and Frederick A A Kingdom. ``Spatial-scale contribution to the detection of mirror symmetry in fractal noise". J. Opt. Soc. Am. A, 16(9):2112--2123, September 1999.

[8] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. ``Statistical color processing". Submitted.

[9] Erik Reinhard, Peter Shirley, and Tom Troscianko. ``Natural image statistics for computer graphics". Submitted.

[10] D L Ruderman, T W Cronin, and GC Chiao. ``Statistics of cone responses to natural images: Implications for visual coding. J. Opt. Soc. Am. A, 15(8):2036--2045, 1998.

Computational and psychophysical approaches to image comparison

Tom Troscianko, University of Sussex, UK

Abstract

How can we compare two images? This type of question may arise when a synthetic image is compared to the equivalent natural scene, or, more generally, in any situation where there is either a localised or a distributed difference between two images.

Two kinds of approach have been adopted by us (in collaboration with the various people mentioned, who should all be attending the Campfire meeting).

Computational model based on properties of cortical neurons

(with David Tolhurst, Cambridge. Application with Ian Moorhead and Marilyn Gilmore, DERA).

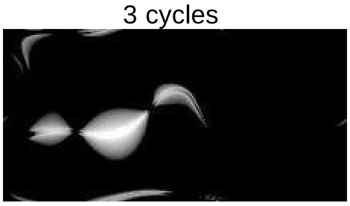
This approach is based on the notion that cortical neurons, which are tuned to orientation and spatial frequency (sf), respond to the contrast (at a given orientation and sf) of a local patch of the image. The human contrast discrimination function allows us to predict when a local patch of Image 1 will be discriminably different from the contrast of the same patch of Image 2. A comparison of the contrast discriminability across all patches of the image gives a numerical value which relates to the degree of perceived difference between the images. Such a number is often called a visual difference predictor. As well as a number, this method also allows computation of the distribution of discriminability in different parts of the image. This relates to the degree to which the differences are localised or distributed.

Synthetic image with vehicle

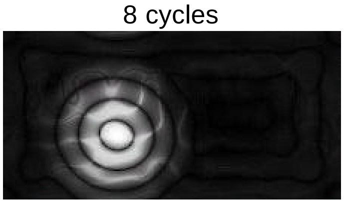
and without the vehicle

1. Find the differences in several frequency bands

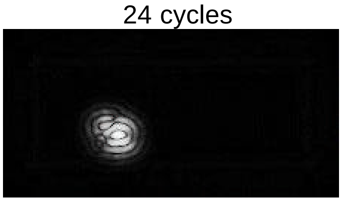
3 cycles



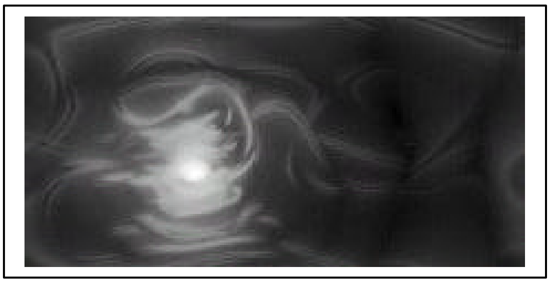
8 cycles



24 cycles



2. Combine the cues from all the bands to form

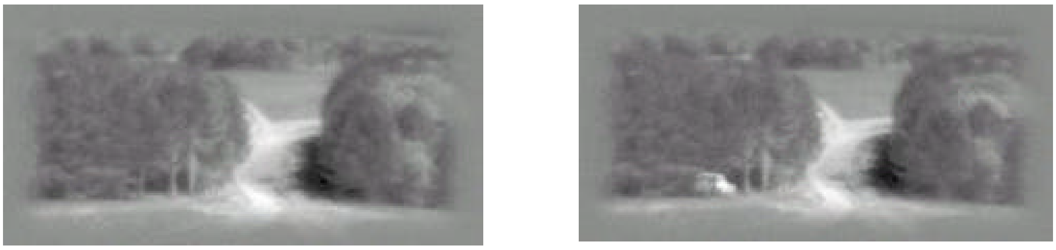


= 66.03

3. Add all pixel values to obtain a single "difference" number

Two questions arise about such an approach: first, is it valid to assume that such a low-level model will successfully predict the perceived difference between two images, or is there some class of task in which this model will fail? Second, is the model itself an accurate representation of what happens in the visual cortex?

We have carried out a preliminary experimental validation study of the model's predictions were compared to perceived differences using a magnitude-estimation method in which subjects were asked to assign numerical ratings of the difference between two images presented simultaneously on the screen. A subject was shown two images, one above the other, and asked to give a number (such as "50") which describes the overall perceived difference between the images. If a subsequently-viewed pair seemed twice as different, the difference rating would become "100" and so on. No time limit was set for the inspection of the images, and normal eye scanning was allowed within and between the images. The results showed a reasonable correlation of about 0.6 with the model's predictions. However, there were some "outliers" in which the model singularly failed to predict the degree of difference. Inspection of the outliers suggested that these images contained large regions of different lightness. Human subjects readily perceive such differences, whereas the model responds only weakly to them. With these outliers removed, the correlation increased to around 0.7.



Magnitude Estimation: "Let this difference be 50 units."

At present, an unanswered question is whether this correlation of around 0.7 is limited by (a) the extent to which the model is a plausible version of cortical processes (this question forms the basis of David Tolhurst's talk), or (b) whether the magnitude-estimation method itself lacks precision and therefore introduces noise. Studies are planned to address these issues, and feedback is welcome. Other important issues concern the viewing mode - a simultaneous comparison of two scenes is not what usually happens, so the role of memory for a previously-viewed reference scene is an unexplored issue. Inspection time is also crucial, and perhaps the information that the subject is trying to extract from the image.

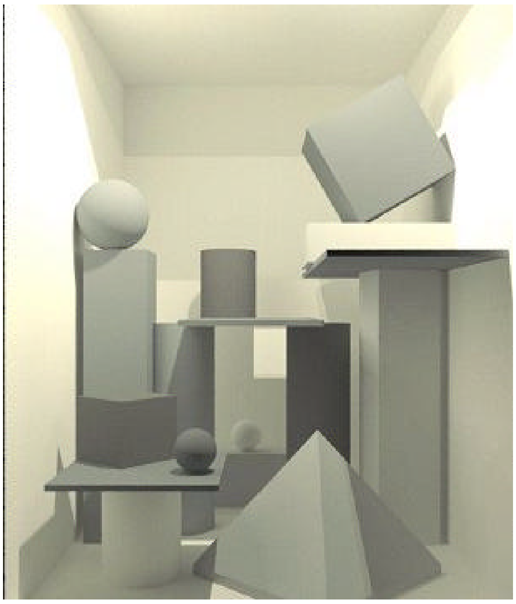
From the existence of the outliers in our pilot study, it seems clear that there are certain scenes that may respond poorly to such a low-level predictive algorithm. Why is this? A possible answer comes from a consideration of results using a very different methodology.

Psychophysical study of perceived lightness in original scenes and their graphic equivalents
(with Ann McNamara, TCD, Dublin).

Another method of comparing two scenes (real versus synthetic, or two synthetic scenes) is to as human subjects to rate the perceived lightness of parts of each scene. It is possible to train subjects quite quickly to give precise ratings of lightness. A similar pattern of lightness ratings in two scenes implies that the two appear similar.

Perceived lightness, in spite of its apparent simplicity as a percept, actually relies on the operation of high-level processes which know about 3-D shape and lighting - otherwise shadows would be confused with lightness differences. The combination of simplicity of use, and high-level scene analysis by the visual system, makes this approach interesting. Currently, we are using this method to compare tone-mapping algorithms. In the past, we have compared the quality of images made with different rendering algorithms. The method is sensitive to many of the parameters of rendering (such as indirect illumination) and therefore

is capable of predicting image quality. However - the method cannot be reduced to a computational algorithm, because of the need for a high-level representation.



High quality graphics reconstruction

Overview

The two methods above appear complementary since the lightness method works for exactly those stimuli which cause problems for the computational approach. Our argument, therefore, is that one needs a variety of methods for predicting the degree to which two images are different. We hope to stimulate debate as to

(a) what these methods should be, and
(b) in what cases a given method should be applied.

Modelling the visual system's ability to discriminate complex visual scenes

D.J. Tolhurst, University of Cambridge, UK

I have been involved with Tom Troscianko, Ian Moorhead and Marilyn Gilmore in developing and evaluating a model of the human visual cortex which is intended to allow us to predict whether a human observer would be able to distinguish between pairs of real or simulated photographs of complex natural scenes. Such models could be used to determine the visibility of camouflaged targets, for instance; or they could be used to assess the fidelity of reproduction of photographs by computer graphics hardware or printers.

The model is based on the known neurophysiological properties of simple cells in the visual cortex of experimental animals, and it is an aim of the model to provide a veridical record of real neurophysiological observations. It assumes that each point in a photograph is evaluated by a set of simple cells, whose optimal stimulus orientations and spatial frequencies cover a range of values. There are thus many thousands of model cells, differing in receptive-field location (compared to the picture borders) as well as in orientation and spatial frequency. The responses of the model cells are not just linear dot-products with picture, but include division by the local mean luminance in the picture to give the nonlinearity of contrast responsiveness. In order to determine whether two complex scenes are discriminable, we must calculate the response of each of the thousands of cells to the two pictures under comparison. It is then necessary to determine how much the cells' responses differ between the pictures, and whether the differences would be large enough to alert an observer to the fact that the pictures were different. We are also interested in whether we can provide a useful metric of how discriminable pictures are when the differences are clearly suprathreshold.

The idea of this model is, by no means, unique and our implementation is probably similar to that of others. But, the several groups will have had to re-invent the same wheels and to face the same assumptions, and it may be that we (as well as other groups) may not always have recognised the assumptions. I hope that, at the campfire, we could search out these assumptions and the solutions in order to determine which ones are critical and which ones are niceties. The development of our model and some of its kludges and "features" have been closely affected by the particular pairs of pictures that we were asked to compare: the same natural scene photographed at different times of day (with and without shadowing), or the presence or absence of a vehicle in the natural scene. However, we would want the model to be truly versatile and, at the campfire, I would be particularly interested to know whether other participants could suggest other kinds of task for the model to solve and to be tested against - other paired comparisons. Perhaps, our model is not versatile enough and its implementation is too closely focussed on a few tasks.

The most difficult and (probably) most contentious part of the model comes when we compare the responses of each simple cell to the two pictures. How do we model whether the paired responses of one simple cell are different enough in magnitude to be detectable? How do we combine the tiny cues provided by each of the many simple cells to provide an overall measure of discriminability? Usually, in such models a single cell's responses are thought to be discriminable if they differ by an amount predicted from a curve like the human contrast discrimination curve - the "dipper" function. The dipper function is usually thought of as a reflection of a non-linear relation between response magnitude and contrast. It is also usually thought that a single dipper function will describe a cell's coding, irrespective of whether a particular response comes from a low-contrast optimal stimulus or a high-contrast non-optimal stimulus. We have been studying how the "dipper" function might arise from the population response of cortical neurons whose limited dynamic ranges cover different parts of the contrast scale. When we also model contrast normalisation, which seems to be an important nonlinearity in cortical processing, it seems that the shape of the dipper may vary under different circumstances. At the campfire, I should like to discuss the implications for discriminations models of having to use more complicated discrimination rules than can be modelled by a simple "dipper". I would also like to discuss how we should pool small cues from different cells.

Visualization

Chair: Victoria Interrante

Cognitive and Computational Processes in Visualization

David Duke, University of Bath, UK

Fundamental to visualization is the idea of recruiting the human visual system and its mechanisms, and so it is unsurprising that there this a body of work in this area. However, much of the work to date has focused on comparatively low-level issues, such as the use of colour, the role of lighting, and the effects of transparency. The role of higher levels of cognitive information processing are still largely unexplored in this domain. So while, for example, Hibbard et al. [HLH+95] acknowledge the importance of feedback between "bottom-up" and "top-down" processing in perception, most of their concerns, such as the need for responsiveness during interaction, are actually generic and say little about how the information in the display is being extracted, used or understood. Similarly, Rheingans and Landreth [HL95] concern themselves with issues that apply mostly at the pre-attentive or early phases of visual information processing.

In visualization, as in other areas of interaction between humans and computers, research has concentrated largely on the development of practical tools and techniques, leaving a comparative dearth of theory about how particular visual representations operate to mediate understanding and exploration. Although analyses can be carried out on specific systems, we are a long way from having a detailed theory that would, for example, guide the development of techniques for visualising novel domains or addressing trade-offs in methods for working with large-scale data in abstract spaces (for example, graph visualization [HMM00]).

An important step in describing visualization, and how visualizations work, is a conceptual separation of the underlying data domain, the structure of the representations used to visualise that data, and the operations that can be performed on those representations. Mention should be made here of the work of Casner [C91], Zhang [Z97], and Chi and Riedl [CH98]. However, although research such as Casner's acknowledges the role of cognitive processing in using visualization systems, it essentially treats the cognitive system as a "black box". Fundamental constraints and capabilities of the human to process and respond to information presented by the system are left implicit, or at best treated as basic assumptions.

A key part of the problem is that unpacking the cognitive box, and relating its contents to the concerns of visualization, is a task that requires multidisciplinary research. And the disciplines involved, computing on the one hand, and cognitive psychology on the other, have traditionally operated with quite different models and theories. Within HCI, a number of "bridging representations" have been developed to try and integrate the products of multi-disciplinary analyses [BD97], but at the cost of separating results from modelling theory.

Over the last six years, the author of this position paper has been working with three other researchers

- David Duce from Oxford Brookes University, whose background is in computer graphics, distributed systems, and formal techniques,

and two psychologists,

- Jon May from the University of Sheffield, and
- Phil Barnard from the Medical Research Council's Cognition and Brain Sciences Unit in Cambridge,

on an approach to building theories of interaction between computating devices and an explicit account ofhuman cognitive processing. This work, which is called "syndetic modelling", represents one possible bridge between the understanding developed within computer graphics of techniques for building interfaces, and the body of knowledge within psychology about human cognitive resources and processing constraints.

Our approach has two main elements. First, as our cognitive model we utilise ICS (Interacting

Cognitive Subsystems), which provides a distributed architecture for human information processing and embodies constraints about the flow and nature of processing that can be carried out [BM95,M00]. In contrast to small-scale "local" theories, for example of visual adaptation or colour perception, ICS deals with larger-scale patterns of activity, while providing place-holders in which specific local theories can be situated. Figure 1 gives a flavour of the model in the context of interacting with a system via gestural input recognised by data glove. In ICS, cognitive processing is distributed over nine distinct subsystems, each specialised to deal with a particular level of representation. In the figure, top down goals directing gesture formation are originating at a "propositional" level (1) and are then mapped through a level of spatial orientation (2) to coordinate limb control (3) and hand position (4). In parallel, a bottom up stream derived from visual input (5) is being used to support this activity by feeding back a propositional awareness of system state and gesture recognition (6). A third stream of information carries proprioceptive feedback (7) from the hand back to coordinate limb control (8).

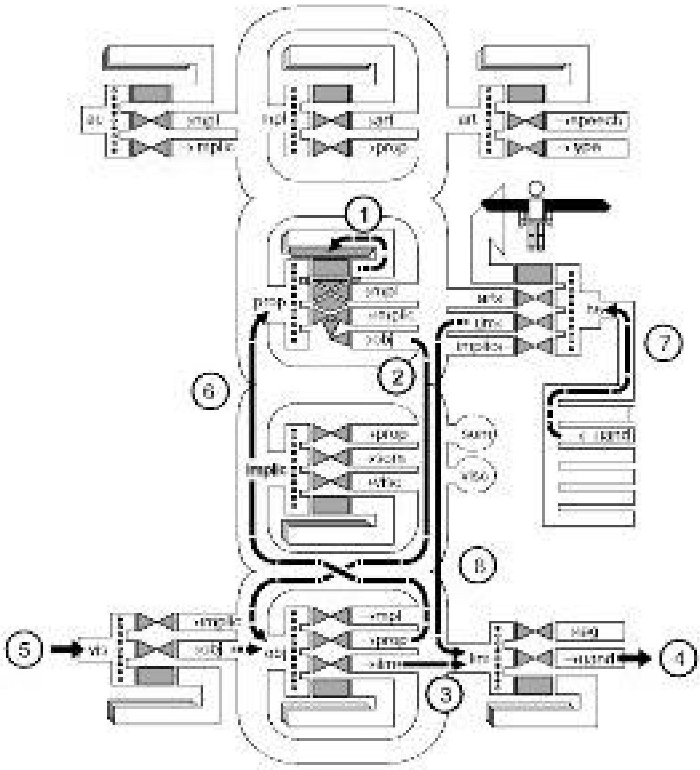


Figure 1. The ICS architecture (from [D95])

The second element of our work is the observation that to apply results from cognitive modelling to the design of computing systems, we need a common framework in which we can study how information is generated, exchanged and processed by both sides of the interaction. Figure 2 shows how the cognitive processes described above are situated in a larger system of processes including those responsible for gesture recognition and rendering. Rather than attempting to simulate this exchange, we have used mathematical modelling to specify the behaviour of both cognitive and computing systems, and then explore the possible behaviours allowed by the conjoint system.

Applications of this technique to date have been in the area of multimodal interfaces [DBDM99] and novel interaction techniques [D95]. However, we believe that the same approach could provide a useful tool for understanding visualization, and in particular the design of representations for abstract information. For example, Duce and Duke [DD95] argues that this kind of framework has the potential to support arguments about whether the reasoning afforded by particular visual representations is "truthful" with respect to the underlying domain

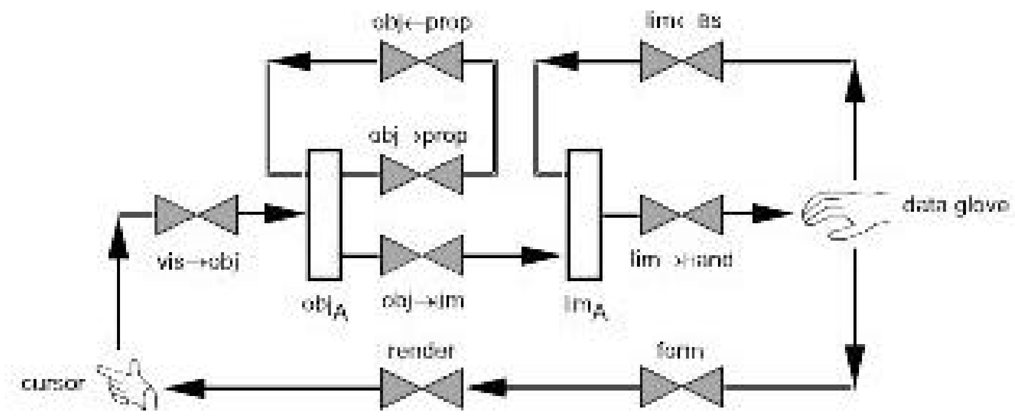


Figure 2. Information flow between computing and cognitive resources (from [D95])

This line of work presents a number of challenges. For example, how can the information carried by visual representation be characterised in a way that can be related to the cognitive processes that operate on these representations? Building on Casner's work [C91] linking "logical operations" on abstract data with "perceptual operations" on presented data, [DH95] developed an approach of using mathematical structures to characterise the presentations. There are relationships between these structures and the representation of information within ICS, however much remains to be done to develop the theory to a point that it can address interesting questions.

An example of the kind of design question that is of interest is represented by work on tree and graph visualization. Figure 3 shows a tree structure visualised using a tool called "Latour", developed by Ivan Herman and colleagues at CWI Amsterdam. One issue that is being explored within the visualization community is how to render such structures once the scale of the data set increases significantly. In the case of tree-like structures, one solution explored in [HMM+99] and shown in Figure 4 is to develop metrics for identifying the "importance" of sub-trees, and providing a simple means of systematically eliding sub-trees by introducing a "skeletal" representation of the elided detail. This approach though raises the question of how effectively users will be able to navigate and understand such overviews, and indeed how such overviews might be modified or tailored to support specific tasks in the application domain. More fundamentally still, can such techniques be generalised beyond trees to more general graph-like structures, and what are the cognitive implications of doing so?

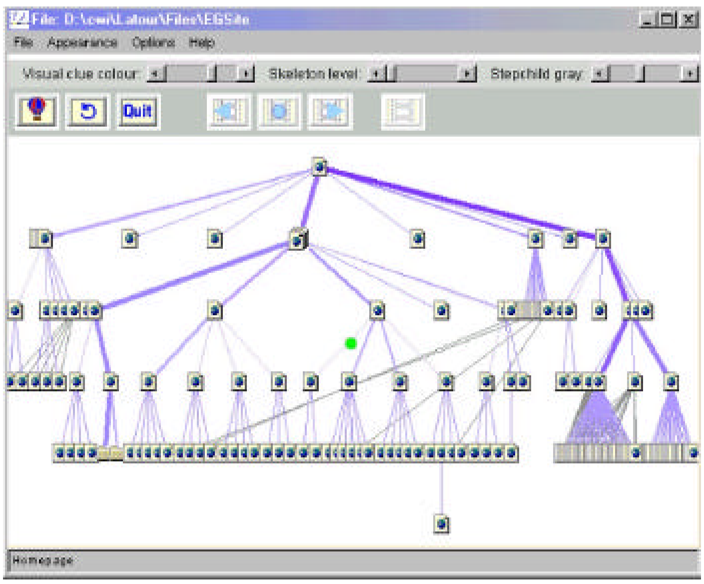


Figure 3. Viewing a tree in Latour

Syndetic modelling is still at an early stage of development, and much remains to be done, both in terms of the capturing more of the underlying cognitive theory, and in understanding how to represent interactions such as those between a visualization system and its user. It may yet be that such a theory does not provide the right insights to understand many of the questions about the design of visual representations and visualization interfaces. But the questions raised by visualization are interdisciplinary, and it's only by building bridges between the disciplines that deeper, long-term results will emerge.

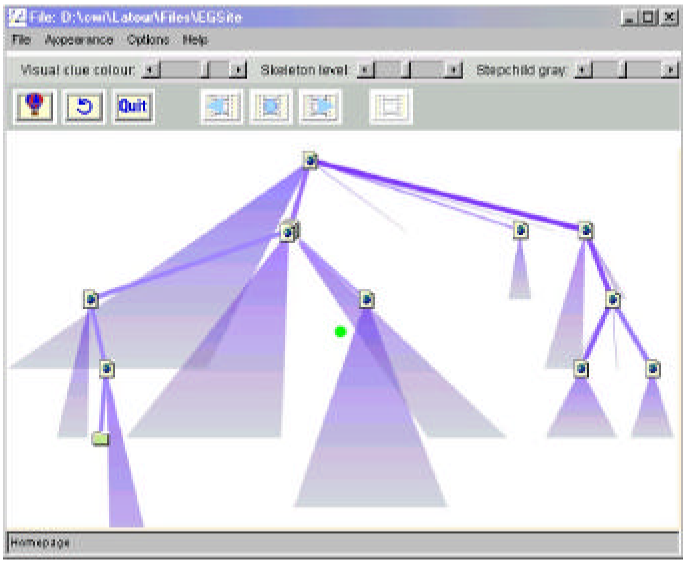


Figure 4. Skeletal view of a tree

Acknowledgement

Figures 1 and 2 are reproduced from reference [D95] by kind permission of the European Association for Computer Graphics.

References:

[BM95] P.J. Barnard and J. May, "Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge", in F. Paterno (ed), Interactive Systems: Design, Specification, and Verification, pp. 15-49, Springer, 1995.

[BD97] A.E. Blandford and D.J. Duke, "Integrating User and Computer System Concerns in the Design of Interactive Systems", International Journal of Human-Computer Studies, 46, pp. 653-679, Academic Press, 1997.

[C91] S.M. Casner, "A Task-Analytic Approach to the Automated Design of Graphics Presentations", ACM Transactions on Graphics, 10(2), pp. 111-151, ACM Press, 1991.

[CR98] E.H. Chi and J.T. Riedl, "An Operator Interaction Framework for Visualization Systems", Proceedings IEEE Symposium on Information Visualization, IEEE Computer Society, 1998.

[DD95] D.A. Duce and D.J. Duke, "Interaction, Cognition, and Visualization", in Second Eurographics Workshop on Design, Specification and Verification of Interactive Systems, P. Palanque and R. Bastide (Eds), pp. 1-20, Springer, 1995.

[D95] D.J. Duke, "Reasoning about Gestural Interaction", Computer Graphics Forum, 14(3), pp. 55-66, NCC/Blackwell, 1995.

[DH95] D.J. Duke and M.D. Harrison, "A Theory of Presentations", in Proc. Formal Methods Europe (FME)'94, M. Naftalin, T. Denvir and M. Bertran (Eds), Volume 873 of Lecture Notes in Computer Science, pp. 271-290, Springer-Verlag, 1994.

[DBDM99] D.J. Duke, P.J. Barnard, D.A. Duce and J. May, "Syndetic Modelling", Human-Computer Interaction, Vol 13(4), pp. 337-393, Lawrence Erlbaum Associates, 1999.

[HMM+99] I. Herman, M.S. Marshall, G. Melançon, D.J. Duke, M. Delest, J.-P. Domenger, "Skeletal Images as Visual Cues in Graphs Visualization", in Data Visualization '99, Proceedings of the joint Eurographics and IEEE TCVG Symposium on Visualization, E. Gröller, H. Löffelmann and W. Ribarsky (Eds), Springer-Verlag, pp. 13-22, 1999.

[HMM00] I. Herman, G. Melançon, and M.S. Marshall, "Graph Visualization and Navigation in Information Visualization: A Survey", IEEE Transactions on Visualization and Computer Graphics, 6(1), IEEE Computer Society, 2000.

[HLH+95] W. Hibbard, H. Levkowitz, J. Haswell, P. Rheingans amd F. Schroeder, "Interaction in perceptually-based Visualization", in G. Grinstein and H. Levkowitz (Eds.), Perceptual Issues in Visualization, pp. 23-32, Springer, 1995.

[M00] J. May, "Perceptual Principles and Computer Graphics", Computer Graphics Forum, 19(4), pp. 71-280, Blackwell, 2000.

[RL95] P. Rheingans and C. Landreth, "Perceptual Principles for Effective Visualizations", in G. rinstein and H. Levkowitz (Eds.), Perceptual Issues in Visualization, pp. 59-73, Springer, 1995.

[Z97] J. Zhang, "The Nature of External Representations in Problem Solving", Cognitive Science, 21(2), pp. 179-217, 1997.

Perception for Visualization

Victoria Interrante, University of Minnesota

My primary research focus is in visualization - designing and implementing algorithms for portraying large, complicated datasets in ways that facilitate the rapid, intuitive appreciation of the essential features of their contents. In the design component of my work I try to integrate inspiration from art and illustration with insights from research in human vision and perception to elucidate an objective understanding of how to define effective techniques for visually communicating the essential contents of a dataset. For example, I am currently working on a project whose goal is to investigate the effect of texture pattern anisotropy on surface shape perception; the purpose in this effort is to attain a understanding of how to select (or custom-define) and apply a texture pattern to a smoothly curving surface in order to enable a more accurate and intuitive understanding of its 3D shape than might be possible either in the absence of texture or if a standard alternative texturing method were used. Of course all of my work requires significant implementation efforts, and this component of the work is where I often must draw from a broad range of disciplines including applied mathematics, computer vision, image processing and computer graphics to define the algorithms that are necessary to achieve the results that I desire.

For example, one of the hypotheses in the texturing project that I mentioned above is that the orientation of an anisotropic pattern with respect to the lines of maximum and minimum normal curvature over the surface of an object may affect an observer's perception of the underlying surface shape. If this is true, then it means that there are better and worse ways to apply a given texture pattern to a surface in order to emphasize its shape. In order to test this hypothesis it is first necessary to be able to compute the principal directions everywhere over a surface and then to synthesize an anisotropic texture pattern over the surface such that the direction of the anisotropy either follows the principal directions or does not. If the hypothesis is true, in order for the results to be of practical benefit for applications involving acquired data it is necessary to be able to accurately estimate the principal direction and lines of curvature over arbitrary, polygonally-defined meshes, and to be able to synthesize a wide variety of useful and interesting textures (ideally from a provided 2D example). With the help of my students and collaborators, I am currently working on all of these different aspects of this shape-representation-with-texture project.

A second project that I am currently working on involves the design, synthesis and use of texture patterns for multivariate data visualization. The essence of this idea is that if we can define a set of independent perceptual dimensions for texture, we may be able to synthesize a multidimensional texture palette in which observers have the ability to accurately and intuitively identify the position of any individual pattern in the distribution. This collection of texture patterns could then potentially be used to effectively communicate a vector of values at all points in a 2D map. There are again multiple components to this project: we have to verify that human observers can reliably and independently perceive the individual texture features that are traditionally thought to comprise the perceptual dimensions of texture, then we must determine how to identify and how to select or synthesize a perceptually equidistant, multidimensional texture palette, and finally we must realize the system and test the results for effectiveness.

In addition to these two projects that have a significant component involving perception, I am working on several other projects in which the main emphasis is in feature extraction for visualization. These include a joint effort with aerospace and mechanical engineering faculty to automatically segment and track the motion, evolution and interactions of possibly densely packed individual vortices in a 3D turbulent boundary layer, and a joint effort with researchers in child development and ophthalmology that involves determining information about the gaze direction of infants from uncalibrated photographs, which I mention for completeness in communicating my interests and background.

There are several things that I would like to learn from this campfire. I am looking forward to the opportunity to meet other people who are working on problems that involve perception and graphics and to find out more about what everyone else is doing. I also hope to gain

insight into what the most important directions are for future work in perception-based visual representation, and how to more effectively develop productive collaborative relationships that can advance these efforts.

Other potential discussion questions that interest me are:

- How can we best mine the store of existing knowledge in visual perception for guidance in visualization design - how can we make the leap from what psychologists have learned about vision and attention through simple controlled experiments to the design of techniques for communicating complicated information in a more effective way?

- Also, to what extent do we need to and how might we more effectively assess the benefits of new visualization techniques? The current culture in our field is that experimental evaluations are not required for publication; what are the obstacles in conducting effective experiments, and how might we facilitate the dissemination of information or software that could assist more people in experimental design and the analysis of results?

In conclusion, my philosophy of research in visualization is that there is a science behind the art of effective visual representation, an explanation for why some representations are more effective than others. The more clearly we understand how our visual system works, the better equipped we are to be able to wisely choose appropriate directions for our efforts in designing effective approaches to the visual communication of information. However, design is much more than the straightforward application of principles, and I believe that insights for visualization from perception can be enriched by inspiration from the works of artists and illustrators, who have a long history of experience and their own unique insights into perception and visual representation.

Perceptually Inspired and Guided Visualization

Penny Rheingans, University of Maryland Baltimore County

Since the mechanisms of human perception have an enormous impact on the effectiveness of visual representation, an awareness of the characteristics of perception should provide the foundation for visualization design. Careful attention to the mechanisms and characteristics of human perception can yield more effective visualizations by exploiting the strengths of the visual system and avoiding its weaknesses. Additionally, visualization design should ensure that the most striking aspects of visualization are also the most important. Representations that draw the viewer's eye to unimportant features may cause more interesting features to be overlooked. Consideration of the characteristics of human perception can be a valuable guide in predicting which aspects of visualization will draw the attention of the average viewer. Feature attributes that influence attention include color, size, opacity/density, order, motion, and style.

Perceptual inspiration for effective visualization techniques need not come directly from the workings of the visual (and other sensory) system. Practitioners of a variety of fields of visual communication, such as technical illustration, graphic design, and art, have been developing techniques that exploit human perception and cognition for hundreds of years. These fields provide fertile sources of examples, techniques, heuristics, and tricks that represent the application of understanding of perceptual capabilities to visual communication. This understanding of perceptual capabilities may be explicit, lowing from direct study of the perceptual mechanisms, or implicit, based in intuition or the study of visual media.

Issues I would like to discuss include:

- How can we increase the accessibility of knowledge and experience in other fields to visualization researchers and practitioners?
- How can we apply knowledge about perception more directly and productively to visualization challenges?
- How can we get the spectrum color scale replaced as the defacto standard?