

Perceptually Adaptive Graphics

Carol O’Sullivan, Sarah Howlett, Rachel McDonnell, Yann Morvan, Keith O’Conor

Image Synthesis Group, Trinity College Dublin

Abstract

In recent years, the Graphics community has come to realise the importance of taking human perception into account when striving for realism in images, animations and Virtual Environments. In May 2001, a EUROGRAPHICS/SIGGRAPH Campfire brought together a group of researchers from various fields, including computer graphics and visualisation, psychology, neuroscience, eye-movements and medicine to discuss the future and current state of the field. Since then, many researchers have been very active in furthering the field of perceptually adaptive graphics. In this report, we outline the state of the art as discussed at that event and the progress that has been made since.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism *Keywords: Graphics and Perception, Perceptual Metrics, Psychophysics*

1. Introduction

At this stage in the development of the field of Computer Graphics, great progress has been made in generating highly realistic renderings and simulations of objects, characters, fluid, smoke, fire and many other effects. New hardware architectures enable users of commodity PCs to run high-quality graphical applications heretofore only available on costly high-performance graphics workstations. Meanwhile, the demand for visually appealing applications on low-end mobile devices such as PDAs and cell phones has increased dramatically, leading to the development of new architectures and libraries to achieve this purpose. This myriad of computational and display opportunities brings with it more than algorithmic and technical challenges. For photorealistic images and movies, how do we know that we are not simply producing pretty pictures and actually representing reality in a faithful manner? For real-time rendering and simulation, how do we make speed-accuracy trade-offs while minimising the perceptibility of any resulting anomalies? What types of anomalies are most noticeable and when can we fake reality and get away with it? How can we quantify these factors and use them in a methodical way to adapt our graphics to the perception of the viewer? Perceptually adaptive Graphics involves the investigation of these issues and will be the subject of this State of the Art Report.

Graphics researchers can learn much by studying the literature from experimental psychology, psychophysics, neurophysiology and related areas. Many computational models of various perceptual functions have been developed and used in recent years by the graphics community. However, such models are often difficult to apply or generalise to handle the complex visual tasks typical in graphical scenes and simulations. Experiments are often designed to study quite specific factors in isolation and in unnatural laboratory situations. Therefore, many researchers from the field of psychology have realised the potential of using Computer Graphics to devise more natural scenes and higher level tasks that more truly reflect the situations being simulated. Therefore, interdisciplinary studies involving Graphics and Perception work to the benefit of both communities.

The increasing interdisciplinary interest in graphics and perception in recent years has resulted in new insights into both areas. Those working on the borders of both disciplines needed a forum in which to meet like-minded people and discuss the state of the art and, in response to this need, a SIGGRAPH/EUROGRAPHICS Campfire on Perceptually Adaptive Graphics [MO01] was held in Snowbird Utah. Prior to this informal workshop, participants were asked to submit position statements, which were subsequently categorised under 6 headings: Image Fidelity, Scene Perception,

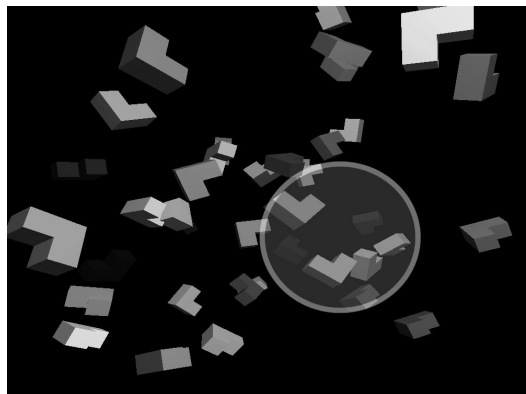
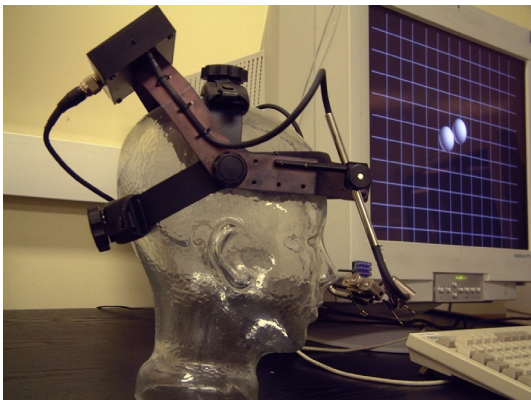


Figure 1: Eye-tracking can be used both to evaluate metrics and for perceptually adaptive collision handling

Applications, Visualization, Interactive Graphics and finally, Distance and Scale in Computer Graphics. The resulting position papers provide a good snap-shot of the state of the art in perceptually adaptive graphics at that time. In this report, we will start our review of recent work from that point onwards, and provide some indicative citations of progress in the meantime. Recent and ongoing research will be presented under the following main headings:

Interactive Graphics: We will discuss how perceptually driven criteria can be used to accelerate interactive rendering, including perceptually adaptive LOD techniques, and approaches that take attention and eye-movements into account.

Image Fidelity: Topics will include image fidelity metrics, perceptual display of high dynamic range images, and perception-driven global illumination techniques.

Animation: Recent work on perceptual metrics for evaluating physically based animation and character animation will be presented, including methods that use eye-tracking (see Figure 1).

Virtual Environments: We consider the problem of effectively generating images of objects and environments that convey an accurate sense of distance and size, along with fidelity metrics for Virtual Environments and methods for measuring and predicting Presence.

Visualisation and Non-Photorealistic Rendering (NPR) A major challenge in visualisation is the design of algorithms that represent large, complex datasets in such a way that people can quickly understand and grasp their main features. An understanding of human perception is integral to the success of such systems and some ongoing research in this area will be presented. Recent work in the application of perceptual principles to non-photorealistic rendering is also discussed.

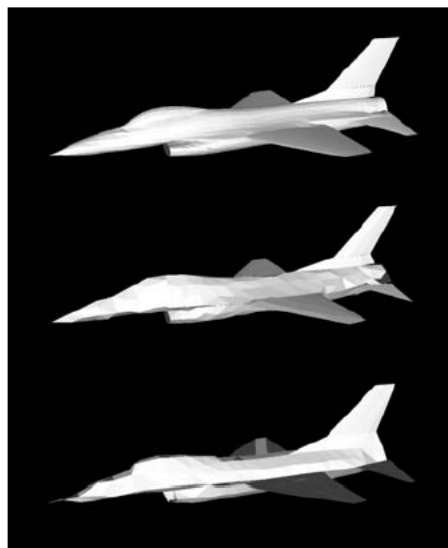


Figure 2: One set of stimuli from Watson's experiment: Original (top), QSlim at 80% (middle), Vclust 80% (bottom) (Image from [WFM01] courtesy of Ben Watson.)

2. Interactive Graphics

In interactive graphics applications, the ideal situation would be to render a fully detailed and photorealistic scene in real-time. However, despite recent advances in current computer graphics technology, this is not yet a feasible option in general. Therefore, the aim is to produce the best perceived image in the time available. An important measurement of visual quality is perceptual i.e., how much does the simulated scene look like the original. The properties of human vision are therefore very important in the design of perceptually based graphics. In this section, we review some of the most recent work in this field. Many of these approaches rely on



Figure 3: A view presented in the second experiment. Here the periphery uses the 20 x 15 LOD, while the lowest contrast background is used. The central area is (always) displayed at the highest HMD resolution. Four distractors are shown. (Image from [WWH04] courtesy of Ben Watson.)

models of visual attention or on an explicit measurement of gaze direction using eye-tracking.

2.1. Gaze-contingent approaches

There is an increasing demand for better display resolution with greater numbers of pixels, posing a constant challenge as large numbers of pixels consume a lot of computational resources. However, this is one scenario in which failures of the visual system can be exploited. Baudisch *et al.* [BDDG03] present several different approaches to take advantage of this using *attentive displays*, discussed in the following paragraphs.

When there is not sufficient computer power for a given task, there is a negative effect of large-screen technology on user productivity. As a step towards compensating for the demand for rendering power and display resolution, the user's attentional focus can be taken into account. Instead of having to add more hardware, display and computation resources can be instead directed to where they are needed most. A single user can only ever focus on a small portion of the display at a time. Approaches using an eye-tracker are examined, which aim to match the subjective quality of a non-degraded display. Alternatively, image content can be removed to achieve a different effect; by presenting viewers with only the most important information, their cognitive load can be reduced.

Gaze-contingent displays degrade the resolution in the peripheral image regions. The high resolution area moves with the user's focus, so the area under scrutiny is always rendered at a higher resolution. An imperceptible degradation is difficult to achieve but often, in visual search tasks, the re-

duction in quality has no effect on performance even if it is noticeable.

Focus Plus Context Screens are one result of new research into extending gaze-contingent displays. Foveal regions of arbitrary shape or size can be created, with peripheral regions degraded by arbitrary means such as colour or contrast and not simply resolution. Additionally, the simultaneous display of multiple foveal regions is possible, which can be used for prediction. Usually, when peripheral content is rendered at low resolution, the display hardware is still the same resolution as any other part of the screen surface. However, in the case of a focus plus context screen, there is a difference in resolution between the focus and the context area. It contains a wall sized low-resolution display with an embedded high-resolution screen. When the user moves the mouse, the display content pans and can be brought into high resolution focus as required. This is interesting for large maps or chip design where certain areas need to be focused upon.

Baudisch *et al.* also describe attentive user interface techniques for directing a system's resources towards the scene components in real-time 3D graphics. Specifically, attentive 3D-rendering engines are discussed, which uses a viewer's gaze position to vary the LOD at which an object is drawn (see Luebke *et al.*, for further details [LRC*02]). Although similar to a gaze-contingent display, such approaches have one main difference; objects in an attentive 3D-rendering engine are simplified at the object geometry level instead of the image level.

Finally, they describe displays in which the user's attention is captured as opposed to followed, referred to as *Easily Perceived Displays*. In such systems, the idea is to guide the user's attention and allow computer generation of aesthetically pleasing images. Such systems use a perceptual model that works from the gaze information of one user to decide which parts of a photograph should be removed. The result is stylized using smooth black lines and colored regions and not just blurred, thus guiding attention to what the original user found important.

In all cases described by Baudisch *et al.*, the intent is to match the characteristics of human vision to the characteristics of computer displays, namely its distinction between foveal and peripheral vision. Gaze-contingent displays and 3D approaches improve display frame rates and responsiveness given certain rendering hardware; focus plus context screens achieve better immersion and visual context given certain display hardware; and nonphotorealistic rendering saves the resources of the user's attention.

In some cases, a model of visual attention has been used to predict fixations instead of tracking the user's gaze. However, the success of the resulting system depends strongly on the similarity between the predicted and actual regions foveated by the observer. Marmitt and Duchowski [MD03]

have developed and evaluated a new method for the comparison of human and artificial scanpaths recorded in virtual reality. They use a string editing methodology for the evaluation of human-human or human-artificial scanpaths. They compare the sequence of regions of interest identified using Itti *et al.*'s attentional model [IKN98] with those recorded from a human observer. The experiment examined three different scenarios; a simple cube, a panorama, and a more complex graphical environment, which participants were allowed to free-view.

They showed that, for all three situations, the similarities between the human and the artificial scanpaths are less than expected. Although this attentional model works reasonably well for still images, it does not accurately predict human fixations in a virtual reality environment. They found that the attentional model assigns attention to a wider area of the image, whereas observers pay more attention to the central region of the display.

Recent work by Cater *et al.* [CCW03] supports the suggestion that visual attention is largely controlled by the task. One key advantage to this approach is that attention is therefore only dependent upon a specific task and not on the user. Therefore, no eye-tracker would be needed as different people performing the same task should, the authors claim, be using similar visual processes. They show how task semantics can be used to selectively render in high quality only the details of the scene that are attended to.

They carried out experiments involving a task on a still image. Participants were required to count the number of teapots in a computer generated office scene, which was rendered at three different levels of resolution; high (3072x3072), low (1024x1024) and selective level. At the selective level the scene was mostly rendered at a low level except for the visual angle of the fovea (2 degrees) centered on each teapot. All scenes were exactly the same except for the position of the teapots.

Results showed that, when carrying out a task, participants consistently failed to notice any difference between the high and the selective quality image. Twenty percent of observers even failed to notice the difference between the high and low quality images. Furthermore, when there was no task involved, the difference almost always went unnoticed. This demonstrates that people primarily attend to task-related objects and the authors postulate that such objects can often be identified in advance, depending on the task. They show experimentally that it is possible to render scene objects not related to the task at lower resolution without the viewer noticing any reduction in quality.

As a confirmation and a demonstration that this was indeed the effect of inattentive blindness and not peripheral vision, they used an eye-tracker to show that attention was fully consumed by the counting task. Even though participants fixated on non-teapot objects, these fixations were not enough to distinguish the difference in quality. They noted

that task driven focus seems to override low-level visual attention when it comes to noticing artifacts. Therefore, they claim, intentional blindness can be exploited to significantly reduce the rendered quality of a large portion of the scene, without having any significant effect on the overall perception.

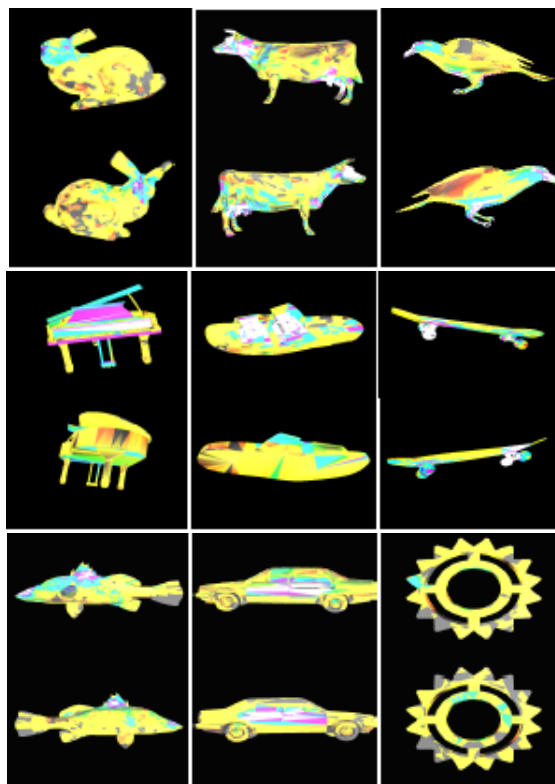


Figure 4: Results from the saliency experiment depicted as a colour-map (white representing the greatest number): the total length of fixations on the familiar natural objects (top), the duration of the first fixations on the man-made artifacts (middle), the total number of fixations on the unfamiliar objects (bottom).

2.2. Perceptually guided polygonal simplification

In the recent work of Luebke *et al.* [LH01], the primary objective was to improve interactive rendering rather than develop offline rendering approaches such as ray tracing, which are typically too computationally expensive to be evaluated interactively. They demonstrate a novel approach to reducing model complexity that is driven by perceptual criteria. They use a psychophysical model of visual perception to create a framework that improves interactive rendering and is used for multiresolution rendering techniques. The circumstances under which simplification will be perceptible are determined, and those that are deemed perceptible are not carried out. Their framework is applied to

view-dependent polygonal simplification and factors such as imperceptible simplification, silhouette preservation and gaze-directed rendering are taken into account. Their results demonstrate that imperceptible simplification was achieved with a limited reduction in polygon count when this method was used. In their evaluation it was found that the probability of seeing a difference was no better than chance. They claim that models could potentially be reduced even more i.e., up to three times further, without a degradation in perception due to the conservative estimate of the spatial frequency at present.

Closely related is work from Williams *et al.* [WLC*03], who describe a best-effort simplification of polygonal meshes based on rules of visual perception. Best-effort rendering is a form of time-critical computing where processing must occur within a certain time budget. This work applies to a wider range of models and accounts for textures and dynamic lighting. They use parameterized texture deviation to measure distortion more accurately, leading to better simplifications for a certain number of polygons. The simplification of lit models is improved by accounting for both specular and diffuse effects, under both Gouraud-shaded vertex lighting and per-pixel normal-map lighting. Here the focus is not so much on imperceptible simplification, but on the approach of perceptually-guided best-effort rendering to a budget. The most obvious advantage of this approach is on vertex-lit models, because the distortion and tessellation artifacts in specular highlights are highly perceptible. Normal maps are used to maintain smooth highlights even at low resolutions. The system has the ability to simplify low-contrast regions and to preserve high-contrast areas such as silhouettes.

How can we tell if one simplification is actually better than another? To answer this question, Watson *et al.* [WFM01] looked at techniques that experimentally and automatically measured and predicted the visual fidelity of simplified models. A set of 36 3D polygonal models were simplified using two different simplification methods (QSlim and Vclust) to two levels of detail (50% and 80% of the original detail), as shown in Figure 2. The stimuli were divided into two different object categories; natural objects and man-made artifacts. Three experimental measures were used to measure the fidelity of these images; naming time (i.e., time taken to verbalise the name of an object), ratings and forced choice preferences. All measures were affected by simplification level and type of simplification. Naming times were longer with increasing simplification and it took longer to name objects simplified using Vclust. When ratings were measured, participants were sensitive to simplification level and also rated objects simplified by QSlim as closer to the ideal. The preference results showed that there was a greater preference for QSlim-simplified stimuli, which was greater for the animal models and greater for the objects at 80% detail.

The effect of object type was particularly interesting. Firstly, it took longer to name the natural objects, which was consistent with earlier results. Furthermore, the ratings results showed that the animal models were more like the standard when simplified using QSlim, but that the artifacts were more like the standard when Vclust had been used during simplification. Regarding preferences, the preference for QSlim-simplified stimuli was greater for the animal models than for the artifact models and for the 80% simplified models than the 50% objects.

Watson *et al.* also examined three automatic techniques to see how well they predicted these experimental measures. They found that they were generally good predictors of quality as judged by ratings and preferences, but were not as successful at predicting naming times. In general the automatic measures correlate well to preferences, less well to ratings and poorly to differences in naming times. Correlations were worse for animals than for artifacts.

Pojar and Schmalstieg [PS03] present a tool for user-controlled creation of multiresolution meshes, allowing selective control of simplification. The user is allowed to identify mesh regions of high semantic or functional importance. These importance weightings are supplied by the user through a Maya plug-in interactively. In this approach, the original Quadric Error Metric of Garland and Heckbert [GH97] is weighted by the user input during simplification. The resulting framework allows the user to improve the quality of a multiresolution mesh by taking semantic and functional importance into account.

In a similar approach, Kho and Garland [KG03] provide another user-guided simplification system that allows the user to interactively control an automatic simplification method. As before, they use a weighted quadric error metric. However, as well as adaptive simplification they provide a tool that allows the user to apply geometric constraints. These constraints preserve features by guiding the placement of vertices on the approximation. The resulting system needs only a short interaction time to produce better semantic or functional models and there is no time difference between user-guided simplification and the original automatic simplification.

More recently, Howlett *et al.* [HHO04] attempted to determine salient features by using an eye-tracking device to capture human gaze data and then investigated if the visual fidelity of simplified polygonal models can be improved by emphasizing the detail of salient features identified in this way (see Figure 4). To try to evaluate the visual fidelity of models simplified using both metrics, a set of naming time, matching time and forced-choice preference experiments were carried out.

It was found that the perceptually weighted metric led to a significant increase in visual fidelity for the lower levels of detail of the natural objects, but that for the man-made artifacts the opposite was true. It was concluded that

visually prominent features may be predicted in this way for natural objects, but not for synthetic objects. A reason for this may be that man-made artifacts are more generally related to a task, thus defining prominent features differently. As described by Hayhoe [Hay00], when a participant's eye-movements were tracked while making a snack, results showed that almost all of the fixations focused on the task, rarely focusing elsewhere; This suggests that visual activity is largely controlled by the task.

Much of the previous work on LOD control is based upon perception at the threshold i.e. when exactly an artifact becomes perceivable. Watson *et al.* [WWH04] point out that, however, most LOD control happens above threshold. They highlight results from perception research that shows how supra-threshold perception differs from perception at threshold and carried out two experiments to investigate this apparent anomaly.

The first experiment they carried out studied the connection between supra-threshold LOD, contrast, eccentricity, and efficient location of complex potential target objects in the visual periphery during search. It had a three way design; 4 LODs, 3 target/background contrasts and 5 target eccentricities. The display environment consisted of a background image indicated by the current level of contrast. If present, the target was placed over the background at the peripheral location decided by the eccentricity. The target was a 12 degree square polygon textured with the monochromatic image of a human face. Participants were instructed to press one of 3 mouse buttons if the target was visible to the left, right or not at all.

As expected, results followed supra-threshold predictions: LOD became more important as contrast declined and, as eccentricity increased, all perceivable LODs were equally important. They speculated that below a certain level of reliable target discrimination, all LODs would have equal impact on accuracy, while above this level, all LODs would have no impact and results confirmed this.

Additionally, in a second experiment they measured performance by time and used variable LOD control. Participants had to find one target among four distractors in a random location (Figure 3). A head-tracked display was used as peripheral LOD and contrast were varied. Results demonstrate the effects on search time. Mostly equal improvements in LOD resulted in equal improvement in performance, following supra-threshold perceptual research. Surprisingly they found a level of minimum reliability. Even in a more dynamic and higher level display, threshold predictions were contradicted. Instead of requiring less detail with low contrast, participants required more. Again, participants found coarse and fine details equally important.

Overall, results were consistent with their hypothesis and showed that it is dangerous to base supra-threshold LOD control on perception at threshold. Results show that LOD

should often be increased in difficult situations, compensating for the challenging environment and maintaining a sufficient level of perceptibility. Threshold-based LOD control should only be used when supra-threshold contrast is low. When LOD control begins to affect task performance, detail should be preserved where sensitivity is lowest. Detail should be added to low contrast regions before high, and to eccentric regions before foveal.

2.3. Interruptible Rendering

Interruptible rendering, as described by Woolley *et al.* [WLWD03], is a trade-off between fidelity and performance. It combines spatial error caused by rendering and temporal error caused by delay to create a single image-space error measure called dynamic visual error. Basically, a progressive rendering framework is used, which draws a coarse image to the back buffer. This is continuously refined while the temporal error is simultaneously checked. When the error due to the time delay becomes greater than the error due to the coarseness of the image, reducing the quality of the image any further is pointless, so the image is rendered. They produce a system which results in a coarse, high frame-rate display when input is changing rapidly, and finely detailed, low frame rate display when input is static.

Different rendering algorithms are described, along with implementations based on polygonal rendering and ray-casting. For depth-buffered interruptible rendering, a method that has containment is required i.e. each refined version of a model should completely contain the simplified version. An ideal interactive renderer simulates off-line renderings as a gold standard for comparisons in evaluation. Their results from this evaluation show that interruptible rendering is more similar to an ideal, instantaneous render than other fidelity control schemes. They also show that interruptible ray-casting also works significantly better than unmanaged ray casting. Although not as significantly, depth-buffered interruptible rendering also improves dynamic visual accuracy compared to traditional fidelity-control schemes.

3. Image Fidelity

In computer graphics, particular effort has been expended to develop perceptual metrics and heuristics to measure or predict the fidelity of images. At the Campfire on perceptually adaptive graphics, Ferwerda discussed "Hi-Fi rendering" and noted that physical accuracy is neither necessary nor sufficient to produce visually realistic images. He described three standards of realism that might help to define the criteria needed for good Image Fidelity metrics and has since elaborated on these ideas [Fer03]. Three types of realism are defined:

1. Physical realism, where the image provides the same *visual stimulation* as the scene depicted

2. Photorealism, in which the image produces the same *visual response* as the scene, and
3. Functional realism, where the focus is on providing the same *visual information*.

Also at the campfire, Ward noted that, while a good deal of effort has been spent in achieving physical accuracy, this is overkill if the aim is simply to produce something that looks real to a human observer. Under this same topic, Gibson discussed the benefits of knowledge of the human visual system for simplifying the process of capturing illumination data in inverse rendering techniques and accelerating the rendering of synthetic objects. McNamara emphasized the importance of design when running perceptual experiments to evaluate realism, while Rushmeier emphasized the need to establish guidelines on how to evaluate new methods that apply perceptual principles to computer graphics. Myzkowski presented perception-driven global illumination solutions while Tumblin discussed the problem that computer-generated images, when represented as a simple 2-D map of scaled scene intensities, often make poor representations of reality because of display short-comings in reproducing intensities and contrasts.

Our discussions are organized as follows: firstly recent contributions and discussions on image fidelity metrics are presented. This is followed by a survey of new perceptually guided rendering techniques. Finally, progress in the reproduction of High Dynamic Range images is discussed. Throughout this section, the exposé continues on from where McNamara's EG2000 State of the Art Report on visual perception in realistic image synthesis [McN01] concludes.

3.1. Metrics

Taking insights from a study of user preferences in consumer photography [SEL00], Winkler [Win01] proposes to enhance visual fidelity metrics with a visual *appeal* component. Among the criteria that the study found, correlated to *appeal*, Winkler singles out two that do not rely on semantics and can be efficiently computed from image data, namely sharpness and colorfulness. He bases his evaluation of the sharpness of an image on analysis of isotropic local contrast using a previously presented technique combining analytic oriented filter responses [WV99]. For colorfulness, he uses the sum of the mean distance of pixel chroma to a neutral grey and the standard deviation of chroma in the image. He finds that chroma computed in the CIE $L^*u^*v^*$ color space give the best results. Both indicators are combined to derive a Perceptual Distortion Metric [Win99, Win00] and are tested on video sequences using data gathered from the Video Quality Experts Group [VQE00], as well as data obtained from experiments by the author. Results show a significant increase of correlation with opinion scores for both indicators.

Recalling the discouraging findings of the VQEG's assessment of visual fidelity metrics over a wide range of video

sequences, which show that their performance is not statistically distinguishable from PSNR [VQE00], Wang *et al.* [WBL02] investigate possible causes. They enumerate the assumptions made by traditional metrics based on error sensitivity and perceptual channel decomposition and discuss their validity. They particularly find fault with the assumption that interaction between channels is weak. They also illustrate the limitations of Minkowski error pooling, which is widely used to combine information over channels in those metrics. Stating that "*The main function of the human eyes is to extract structural information from the viewing field, and the human visual system is highly adapted for this purpose. Therefore, a measurement of structural distortion should be a good approximation of perceived image distortion.*", they propose a simple metric integrating three structural factors, namely loss of correlation, mean distortion and variance distortion. They obtain encouraging preliminary results, at a fraction of the cost of metrics simulating the response of the visual system.

Walter *et al.* [WGP02] present a technique to make error tolerance elevation maps resulting from visual masking available through the standard texture mapping pipeline. Computation of each texture's masking properties is inspired from the JPEG process, allowing for construction of the map as the image is decoded. The elevation maps are stored in each texture's mip-map. At rendering time, standard texture lookup and interpolation make error tolerance elevations available on a per pixel basis to any algorithm that makes use of them. Applications similar to those discussed in the next section are presented.

3.2. Rendering

3.2.1. Interactive rendering of global illumination solutions

Dumont *et al.* [DPF03] present a general framework, based on a decision theory approach, which uses perceptual criteria to handle resource constraints in interactive rendering of pre-computed global illumination solutions. The rendering of a frame is seen as the result of a set of rendering actions, each with an associated cost and utility. The cost represents the amount of resources needed to take the action. The utility measures the contribution of the action to the result. Resource constraints can then be met by running a resource allocation algorithm that will maximize utility.

They describe three applications of their framework, outlined in the following paragraphs. A common property is that a gold standard is available in the form of the global illumination solution. The utility is therefore defined as a measure of fidelity to it, provided by a Visual Difference Predictor (VDP). A key point is that, at equal cost, an ordering of the utilities of rendering actions is sufficient, as opposed to an absolute estimation.

The first application deals with diffuse texture manage-

ment. The constraint here is the amount of texture memory available. The rendering actions are then the choice of mip-map level at which each texture is stored. The utility function used is a customized version of the VDP proposed by Ramasubramanian *et al.* [RPG99] in their global illumination guiding framework: the spatial frequencies component, which accounts for visual masking, is pre-computed when the mip-map levels for each texture are generated.

The second application is an extension of the previous method to non-diffuse reflections. The authors choose to render them using prefiltered environment maps. The constraints are the memory space available for environment maps as well as the time needed to compute and filter them. Because of view dependence, the spatial frequency term of the VDP has to be approximated.

The last application aims at simplifying the radiosity mesh from the global illumination solution so as to meet rasterization limitations, which constitute the constraint here. Rendering actions are the display of radiosity elements, the more finely subdivided, the more costly. The utility function is again a specifically tailored VDP.

In each case, interactive frame rates are achieved with minimal perceptual impact out of configurations where no resource allocation led to catastrophic results.

3.2.2. Global illumination for static images

Farugia and Peroche [FP04] have recently modified a progressive radiance evaluation technique by Guo [Guo98] by substituting the refinement criterion based on discontinuity for a perceptual metric.

Guo's technique consists of computing radiance samples in an order that will maximize the speed of convergence to the final result. The process is driven through iterative construction of a Directional Coherence Map, which is an irregular subdivision of the image in basic blocks that either correspond to smooth regions (smooth block) or discontinuities (edge block, complex or simple). At each iteration step, the current configuration of the DCM guides where to take new samples in order to refine the block classification. Roughly, children of smooth blocks are considered smooth if none of their corner values vary sensibly from the value obtained by interpolating their parent's corners. Otherwise, they are classified as edge blocks. Edge blocks are sampled along their boundaries to determine if they are complex (i.e., they contain more than one image edge). For simple edge blocks, a discrepancy direction is computed, which is used for interpolating the values to be compared with the samples taken at the next iteration for classification of the children. Rendering is done by linear interpolation of the samples, following the discrepancy direction in the case of edge blocks. How blocks are flagged (smooth or edge) in the initial regular grid has a great impact on convergence speed. Guo uses a perceptual criterion based on contrast over the samples corresponding to the corners of each block.

Farugia and Peroche classify their subdivision cells by evaluating a VDP between an interpolation computed using the previously computed samples and an interpolation, taking into account the samples added at the next refinement step. They build their metric by computing a distance on the output of Pattanaik *et al.*'s Multiscale Model of Adaptation and Spatial Vision [PFFG98], which consists of contrast maps corresponding to spatial frequency bands. For efficiency, following Albin *et al.* [ARPT02], they evaluate their metric over each cell pair using a statistical approach.

Stokes *et al.* [SFWG04] present a new approach to efficient, high quality global illumination rendering using perceptual illumination components. Their work is based on the fact that the illumination of a surface can be split into components that are separately computable, namely: direct, indirect glossy, indirect diffuse and indirect specular illuminations (if one is to neglect the interaction between the latter three).

Their aim was to produce a perceptual metric functioning on those terms that they could use to drive rendering. They started by conducting a perceptual experiment to obtain data on the relative visual importance of each term. A test scene representative of typical global illumination scenarios was rendered from different viewpoints, each component separately, then the results blended in all possible combinations to serve as visual stimuli for the experiment (a full global illumination solution was also computed to serve as gold standard). Participants were asked to sort the images by perceived quality. The results confirmed the marginal perceptual contribution of light path interactions between the three indirect components (only present in the gold standard). The paper explains thoroughly how a mathematical model was fitted to the experimental data to formulate the metric. Rendering was then driven using the metric to predict the relative importance of each component as a function of the materials visible from the desired viewpoint.

3.2.3. Global illumination for animation

Myszkowski [Mys02] describes an Animation Quality Metric (AQM), which is based on the widely used original VDP by Daly [Dal93]. The time dimension is introduced in the metric by replacing the purely spatial Contrast Sensitivity Function (CSF) with a spatiovelocity CSF, which expresses the sensitivity of the eye to contrast in a visual stimulus as a function of both its spatial frequency and its speed over the retina. A spatiovelocity CSF, as opposed to spatiotemporal, is chosen principally because the AQM is expected to be applied for the rendering of animation sequences. This means that both camera pose and range data are available, therefore allowing for easy computation of pixel flow through image warping, which yields velocities for each pixel since the frame rate is known. (Another reason is that accounting for eye movement is made easier).

Myszkowski's first application of the AQM is to speed-

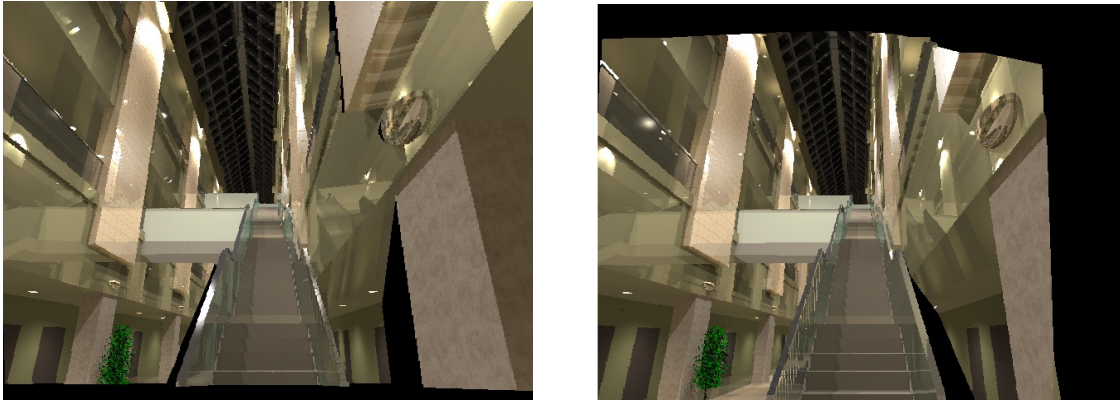


Figure 5: Views warped from keyframes at both ends of a frame interval. (Image courtesy of Karol Myszkowski)

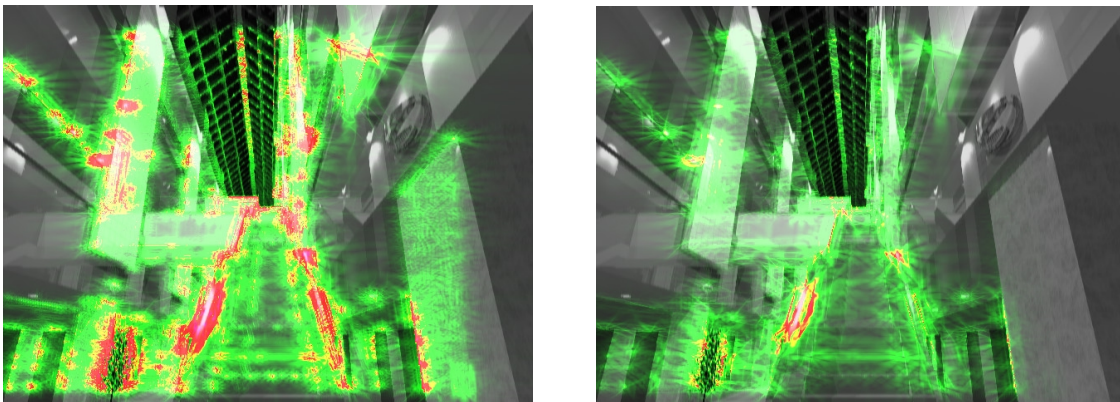


Figure 6: Probability of difference detection between the two images (transparent = 0, green = low, red = high): Original pixel flow (left) and accelerated pixel flow (right). (Image courtesy of Karol Myszkowski)

up the rendering of walk-throughs of static environments. The main idea is to use cheap Image Based Rendering techniques to compute as many pixels of the in-between frames as acceptable by using keyframes as reference views. Warping gaps and areas where the pixel flow is slow are accurately rendered. Keyframes have to be chosen intelligently so as to minimize the number of pixels that have to be rendered. From an initial coarse regular placement of keyframes along the sequence, new keyframes are added by recursively splitting those intervals where the AQM gives a high difference between the two instances of the middle in-between view warped respectively from each bound of the interval (See Figures 5 and 6). When run on test walk-through, the method provides a speed-up by a factor of two, with 50 to 30 percent of pixels needing accurate rendering for each frame. The second application, first described in [MTAS01], aims at keeping noise below noticeable thresholds in stochastic global illumination rendering of dynamic scenes.

Yee *et al.* [YPG01] present a perceptual method to guide

the RADIANCE [War94] irradiance caching mechanism. RADIANCE uses an *ambient accuracy* parameter as an error tolerance threshold when interpolating values from its irradiance cache, as opposed to computing them accurately. Yee *et al.* modify this parameter on a per pixel basis using a spatiotemporal error tolerance map computed on estimate renderings of each frame. The computation of the tolerance map is similar to Miszkowski's AQM, the major difference being that they incorporate an optional saliency estimation using Itti *et al.*'s model [IKN98].

3.3. High Dynamic Range image reproduction

Through multi-exposure photography and the use of new devices that allow for exposure adaptation on a per pixel basis [NM00], the capture of High Dynamic Range (HDR) images is fast becoming common. Current display devices are only able to reproduce luminance ranges that are several orders of magnitude below what can be captured. The human visual system however, through the mechanism of local adaptation,

is able to perceive local contrast in stimuli that exhibit wide luminance disparities. HDR display devices are just starting to appear. In media such as hard-copy, prospects are not as bright and contrast reduction methods, which have recently aimed at narrowing the perceptual response gap between the captured and reproduced signal by simulating local adaptation, will remain dominant.

3.3.1. High Dynamic Range display device

In the design of their display device, Seetzen *et al.* [SHS*04] exploit limitations of the human eye in handling high contrast in small regions. Because of its scattering properties, the eye is unable to clearly make out the edges or judge relative intensities of local luminance discontinuities of a ratio above 150 to 1. Some of the energy from the luminous side will be transmitted to cells that should be perceiving the dark side near the edge.

The main idea is to replace the uniform back-light of an LCD device with a modulated one capable of much higher brightness. The theoretical reproduced dynamic range is the product of the dynamic ranges of the two modulators. However, light path blending reduces that figure.

They present two designs, one using a Digital Mirror Device and one using an array of LEDs as modulated back-lights. In both cases, the HDR signal has to be decomposed into two signals to drive each modulator. Psychophysical results describing Just Noticeable luminance Difference at given luminance levels are used to determine the appropriate number of JND steps in each case, from which the optimal number of bits necessary to drive each modulated back-light is derived. In the case of the LED array, the minimal number of LEDs to maintain the same visual response is chosen based on perceptual experiments carried out by the authors [SWW03]. The blur introduced by the digital projector's point spread function is made acceptable by the same results.

On a related note, Mantiuk *et al.* [MKMS04] also exploit the limitations of the Human Visual System in high contrast perception to propose an efficient encoding method for HDR video.

3.3.2. Progress in local adaptation contrast reduction methods

The aim of local adaptation contrast reduction methods is to construct a map of appropriate adaptation levels to local luminance for each pixel. A psychophysically based tone-reproduction technique, such as [PFFG98] can then take advantage of it to produce an image resulting in a visual response similar to the response to the HDR stimulus.

Ideally, each pixel should be given an adaptation luminance in accordance with the perceived luminances of pixels of its group as segmented by the eye's local adaptation mechanism. This amounts to an edge-preserving smoothing

of luminance values: luminance values on one side of a discontinuity should not contribute to the smoothing of values on the other side.

Yee and Pattanaik [YP03] provide a four step method to achieve this. First, the HDR image is segmented by classifying each pixel in luminance interval bins. Then, contiguous pixels belonging to the same bin are grouped using a flood fill, and the mean of log luminance values computed for each group. Groups of insignificant size are then merged into bigger neighbour groups whose mean luminance is close. Those three first steps are repeated for increasing bin sizes, producing several layers. The last step of the method averages the values over layers corresponding to the closest bin sizes. The method is fast and gives nice results, getting rid of the halo artefacts that appear when sharp luminance discontinuities are not preserved. A drawback is that five user specified parameters are needed.

Edge-preserving smoothing can be seen as adapting the support of the smoothing function so as to avoid considering outlier values (i.e., values that belong to the other side of a discontinuity). Tomasi and Manduchi's [TM98] non-linear bilateral filter does just that, by treating the weights used for smoothing using a Gaussian distribution centered on the current pixel's luminance value. Durand and Dorsey [DD02] apply it to contrast reduction with good results. Pattanaik and Yee [PY02] propose a more direct approach for outlier rejection inspired by the field of photography. Simply put, they discard values from neighboring pixels whose luminance is outside a factor of 5 intervals centered on the current pixel's value. Building upon both Fattal *et al.*'s [FLW02] use of gradient information for contrast reduction and Tomasi and Manduchi's bilateral filter, Choudhury and Tumblin's [CT03] trilateral filter further adapts the support of the smoothing function to the signal by tilting it along the bilaterally-smoothed luminosity gradient. The neighbourhood used for smoothing is also adaptively grown.

Seetzen *et al.* [SHS*04] note that there has been little psychophysical validation of contrast reduction methods so far and suggest that the availability of HDR display devices will help design experiments to investigate that topic. Their display device has already been used by other researchers to compare tone mapping operators. Recent work by Drago *et al.* [DMMS02] addressed this issue by subjecting seven tone-mapping operators to perceptual evaluation based on similarity and naturalness ratings.

Other work that has followed on from research presented at the campfire include Reinhard and Devlin's approach to dynamic range reduction inspired by photoreceptor physiology [RD04], Reinhard *et al.*'s method of perceptually inspired colour transfer between images [RAGS01] and Paraga *et al.*'s [PTT02] investigation of spatio-chromatic properties of natural images and human vision.

4. Animation

In tandem with investigations into image fidelity, the importance of animation and simulation fidelity has also grown. We now review some work in the field of simulation and animation that aims to evaluate and/or improve animations by applying perceptual principles.

4.1. Physical simulation

Barzel and Hughes [BHW96] first introduced the concept of plausible simulation, and pondered the meaning of physically plausible vs. visually plausible motion. They suggested that there are situations in which inaccurate or probabilistic techniques implemented in a noisy or textured environment are likely to look more realistic than the physically correct solutions, which have the tendency to appear sterile and repetitive. Building on these ideas, Cheney and Forsyth [CF00] developed a scheme for sampling plausible solutions to constrained physical simulations. They allow a user to provide a function that describes physical plausibility, which is then used to generate a range of animations that satisfy both this user-defined definition of plausibility and any physical constraints to be imposed upon the system. For example, the product of unnormalised Gaussian bell-shaped functions is used to describe the physical plausibility of a bouncing ball animation - each collision normal is thus considered to be normally distributed around the perfect, vertical, direction.

Popovic *et al.* [PSE*00] also allow for the manipulation of physically-based animations by letting an animator play with the physical properties of a simulation until a desired end-result is achieved. In this case, the definition of plausibility is purely under the control of the animator. However, only recently has there been a concerted effort to examine issues of perceptibility and plausibility with respect to the simulation of motion for animations. In order to guarantee plausibility, an awareness of the perceptual impact of simplifications and distortions is imperative. In recent work, we investigated the role of various factors on human perception of anomalous collisions and the exploitation of these factors for developing adaptive simulation systems [OD01, ODGK03]. We also investigated the use of eye-tracking for evaluation and adaptive control of interactive graphical systems [ODH03], and to investigate the role of attention when viewing collision events [OL04] (see Figure 1).

Humans 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 [Mic63]). Many studies have shown that these perceptual mechanisms are established very early on in infancy [BSW85] but it is not clear how accurate they are. Research in the realm of physics education

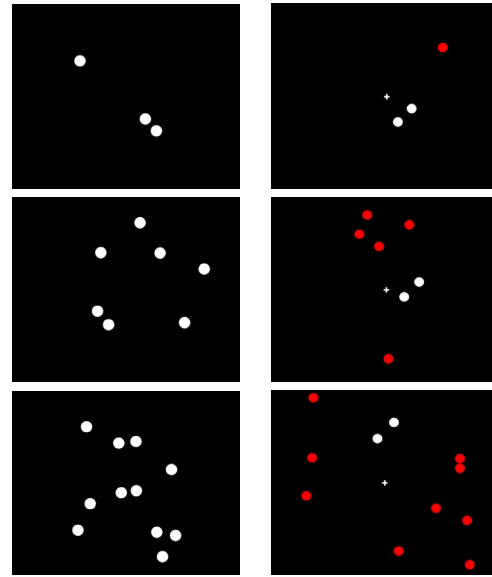


Figure 7: Experiments with visually similar (left) and dissimilar (right) distractors (from [OD01])

has shown that most people have erroneous, yet very robust, pre-conceptions regarding the physical behaviour of objects [Cle82].

In real-time animation, if fully-accurate processing of dynamic events such as collisions is performed, this can often lead to long delays if the objects are complex or there are many colliding groups. We found that the effect of such a delay at the moment of impact on the perception of the user was highly detrimental to the perceived realism [OD01]. To ameliorate this effect, simplified volumes are often used to compute collisions between objects in real-time simulation, but this can also lead to a variety of dynamic anomalies - separation between the objects when they collide, or less accurate physical response. We found that the negative effect of both of these factors was significant, but found that perception could be improved by adding more distractors (other objects moving in the scene - See Figure 7), occluding the view or adding a random spin after collision.

These studies provided some interesting insights into the factors that affect our perception of certain dynamic events. They were not, however, sufficient to provide the basis for an empirical metric, as they were more qualitative than quantitative in nature. Therefore, more recently we ran some psychophysical experiments that allowed thresholds to be found for human sensitivity to dynamic anomalies [ODGK03]. These studies were inspired by some earlier psychophysical experiments carried out by Kaiser and Proffitt [KP87].

We showed participants a range of simple dynamic events i.e., collisions between spheres, or between a sphere and a more complex object (see Figure 8), and applied distortions

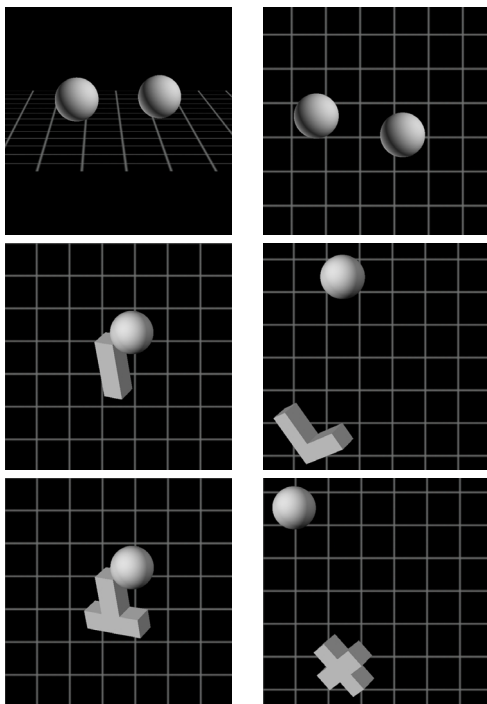


Figure 8: Screenshots from the experiments to examine perception of dynamic anomalies in collisions [ODGK03].

tions in a methodical manner in order to find the thresholds at which these distortions became perceptible. Such distortions included linear and angular velocity errors, delays or separation between objects at the moment of impact and erroneous changes to the post-collision trajectories of the objects. Some interesting biases were found and this information, along with the thresholds measured, was used to define a visual plausibility function. This was similar to that proposed by Cheney and Forsyth [CF00], but was now based on psychophysical data and hence took the perception of the viewer into account. To illustrate its potential, the metric was used in two case studies: simulation levels of detail using collision proxies, and constrained dynamics.

In the future, the perception of sound [vdDPA*02, KPK00] and haptics [OL03] should be included in any metrics that aim to evaluate the fidelity of simulations. Multisensory displays are becoming more prevalent and related perceptual issues are also recently receiving more attention - see the notes from Fels *et al.*'s recent course on design of interactive multimodal media systems [FFM*03] for a full discussion of these issues.

4.2. Animation of Human Motion

A particularly challenging new area of research is evaluating the perceptual impact of anomalies in human animation.

Incorporating perceptual principles into such simulations is a promising future area of research, with a view to improving quality and gaining insights into where to concentrate limited computational time in human and crowd animation. Human animations can look unbelievable for many reasons: perhaps the technique used to generate the motion is not accurate enough, or motion editing has introduced noticeable artefacts; important behavioural features may not be present or have not been modelled correctly; or maybe there simply was not enough time or computational power available to perform the computations necessary to produce the required simulation. Oesker *et al.* [OHJ00] demonstrated that there is psychological evidence that observers were affected by the level of detail of a character's animation in making judgments about the skill of virtual football players, while Hodgins *et al.* [HOT98] showed that the geometric model type used to represent the human affected people's ability to perceive the difference between two human motions. In this section, we provide an overview of some recent studies that further investigated several of these issues.

4.2.1. Motion Transitions

Lee *et al.* [LCR*02] developed a technique for keyframe transitioning. Their aim was to retain the realism of motion capture, while allowing new motions to be created. More specifically, this involves the creation of a motion capture database and a graph to incorporate the interconnections of the motion frames in this database, along with a search strategy for traversing the graph to form new paths.

Wang and Bodenheimer [WB03] conducted a set of experiments to evaluate the motion transition cost function proposed by Lee *et al.* The goal of the paper was to produce a set of optimised weights that select good transitions and validate these results with a cross validation study and a user study. Lee *et al.* set weights to one for shoulders; hips, knees, elbows, pelvis and spine, and all others had a value of zero. Wang and Bodenheimer used 16 different segments of captured motion and manually selected 16 good transitions and 26 bad ones and then solved for the optimal values of the weights using a constrained least square minimisation. The optimisation was constrained such that the weights were non-negative and symmetric.

A cross validation study was performed, where the weights were optimised with one set of training data deleted. The resulting weights were then used to compute the optimal value of a transition for the deleted data set. The algorithm proved to be both robust and general enough to pick reasonable transitions for a wide variety of motions.

A user study was then conducted to evaluate the weighting determined by the optimisation. Participants were shown a series of motion clips and asked to rate how noticeable and natural the discontinuities were. Half of the motion clips used Lee *et al.*'s weights and the other half used optimal weights. The comparison was made on clips without the use

of interpolation or smoothing. The participants were then asked to complete a post-sequence questionnaire where they were asked to compare and rate their impressions of the motions using a 5 point Likert Scale. It was found that, in general, participants considered the optimally weighted motions better and more natural than the original weighted motions.

A further study of motion transitions was conducted in Wang and Bodenheimer [WB04], with the focus on computing the optimal duration of blending two motion clips to form the motion transition. The goal of the paper is to give guidance to designers of animations systems as to which blend-lengths to use when transitioning between motions. Two methods were developed to calculate the blend length. The first used geodesic distance to compute the best blend length for blending between two arbitrary frames. The second method predicted the optimal blend length by calculating joint difference between the start and end frame for each major joint. The velocity of the joint with maximum difference between start and end frames was then used to find the optimal blend length. Four experiments were conducted to compare the two blend methods described to a typical transition scheme that employed a fixed blend length.

The first experiment used the geodesic distance method to compute transition lengths, and participants were shown pairs of these motion transitions. They were then asked to determine whether the pair of motions was "identical" or "different". The Just Noticeable Difference was determined from this experiment, and results showed that people could differentiate between transition lengths that differ by seven or eight frames.

The second experiment tested whether participants found the transitions generated by the geodesic distance method more natural than the transitions that used a fixed blend-length of 10 frames. Participants were shown pairs of motions and asked to judge which they found the most natural, and results showed that a statistically significant number of observers chose the geodesic method as being more natural than the fixed blend-length method. However, many participants commented that the motions seemed very similar.

A further study was conducted to compare the velocity method to the fixed blend-length method. This experiment used more active motions like boxing and dancing as it was felt they were more suitable for the velocity method. Again, participants were presented with pairs of motions and asked to choose the most natural looking motion transition. It was found that there were no individual motions for which participants favoured the 10-frame blend.

Finally, the geodesic method was altered to include *time-warping* and this method was compared to the geodesic method without timewarping. Again, there were no motions for which observers favoured the fixed blend method.

4.2.2. Perceptual Metrics for Character Animation

Reitsma and Pollard [RP03] investigated human sensitivity to errors in the motion of an object due to gravity (ballistic motion), to obtain perceptual metrics for character animation. Motion capture produces very realistic human motion, but the processing of this data can result in some unrealistic artefacts, such as foot sliding, interbody penetration, and also violations of the laws of physics. The metric outlined in this paper would be useful in the evaluation of post-processing techniques, motion graph construction and also in setting bounds for optimising animations.

The particular study presented focuses on errors in the ballistic phase of human jumping motion (Figure 9). Once the character leaves the ground at the start of the jump, the trajectory of the centre of mass is fully determined. Changes to that trajectory violate the laws of physics, and these types of changes often occur due to processing of motion capture data.

Two experiments were performed. The first tested the perception of anomalous accelerations and decelerations. Seven base motion capture animations were manipulated to incorporate error variables. Three levels (small, medium and large) of error variable, two varieties (horizontal or vertical acceleration), and two directions (acceleration or deceleration) were added. Participants were shown a sequence of animated human jumping motions and asked to categorize each motion as having either "no error" or "error" and mark their level of confidence in their answer on a scale of 0 to 9. Participants were told the number of animations that contained errors, but they were not told which ones contained errors. Results found that subjects were better able to detect added acceleration rather than added deceleration. Also, subjects found errors in horizontal velocities easier to detect than errors in vertical velocities.

The second experiment investigated perception of errors in effective gravity, and was conducted in the same manner as the first. The difference between this experiment and the first was that the added errors in motion were in the form of increasing or decreasing gravity. Results showed that subjects found low gravity easier to detect than high gravity. Motions with vertical, horizontal, and composite errors were also included in this experiment to test the validity of comparing data across studies, and the results were consistent with the observations in the first study. Finally, a description on how a metric might be designed based on the results is also included in the paper.

Harrison *et al.* [HRD04] feel that Reitsma and Pollard's metric may over or underestimate the sensitivity of observers, because the metric does not take expectation and attention into account. They investigated the role of attention in visual perception. In particular, they questioned the extent to which the lengths of the links in an animated figure could be altered, without the viewer being able to detect the

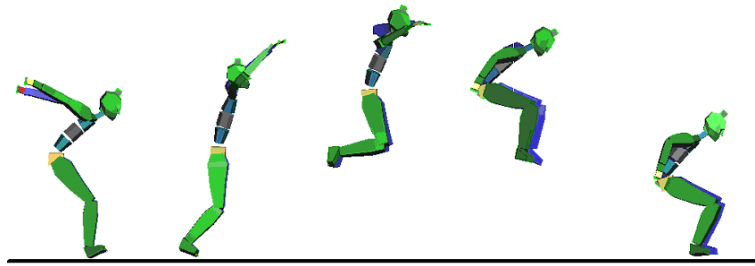


Figure 9: *Human jumping motion*

change. Five perceptual experiments were conducted to investigate a range of different factors that influence viewer's perception. A simple "two-link articulation" was used as a stimulus for the experiments. The shoulder joint was a fixed point, and the line between the shoulder and the elbow was described as the "upper segment", while the distance from the elbow to the wrist was described as the "lower segment". The wrist moved along a circular path, and the elbow moved such that the two links formed an articulated pair.

The first set of experiments tested Expectation. Animations of the arm moving around a circular path were created, with temporary increases and decreases in length of the entire arm. The first set of participants were shown pairs of animations and asked to judge which one of the two contained a change. These participants received feedback and the trials were grouped, thus they had an expectation as to how the change would affect the motion of the arm. Results showed that observers could detect very small changes in arm length; a threshold of 2.7% was achieved (detection thresholds are in terms of relative changes in length (Weber Fractions)). A second set of participants were presented with randomly interleaved changes in length of the entire arm. These naïve participants achieved a threshold of 5.6%, a worsening of sensitivity by a factor of 2.2, showing that expectation does affect the ability of the observer to tell the difference between motions of this kind.

The second experiment investigated the effect of Task Interference on the ability of the observers to detect a change in arm length. Participants were asked to perform a primary and a secondary task. The primary task was to count the number of times the wrist changed direction, while moving along a circular path, and the secondary task was to detect a length change. A threshold of 10.5% was achieved for this experiment, a worsening of sensitivity by 7.8, showing that task interference has a huge effect on the abilities of the observers.

Increases and Decreases in Length were examined in the third experiment. Results showed that for increases in length the threshold was 6.9%, whereas a much higher value of 19.7% was achieved for decreases in length. They Extended this last experiment by adding a distractor as a primary task.

Task interference caused a further decrease in sensitivity for slowly growing arms, from 6.9% to 12.7%.

Finally Duration of Length Change was examined. A strong linear relationship between log duration and threshold of detection was established. A list of guidelines for obscuring length changes during animation were formulated from the results of the five experiments, as follows: 1) Changes in arm length of up to 2.7% will likely go unnoticed, even when expected; 2) Length changes should never exceed 20%; 3) Decreases are less noticeable than increases; 4) Slower changes are more difficult to detect; 5) During fast motions, changes are less likely to be perceived; 6) A distractor effectively masks changes; 7) Expectation of change results in a greater likelihood of detection. The authors contend that any metric for evaluating the visual fidelity of human animation, or in fact any form of animation, must take attention and task into account.

4.2.3. Specifying and Modifying Motion

Harrison *et al.* [HBF02] focused on higher-level techniques for specifying and modifying motions. Three types of motion spaces were used to discuss the relationship between the parameters of movement and human perceptions (Figure 10). The first motion space is the "mechanical motion space", where the kinematics of movement are described by a vector space of motion signals. The second motion space is a hypothesized space, referred to as the "psychological motion space", where humans organise motions according to their features (e.g. humans would classify a motion as a ballet motion if they saw that the feet of the moving character were pointed and the motion appeared to be a graceful dance). The third motion space is referred to as the "linguistic motion space", where humans use words to describe motions (e.g., fast, slow, graceful, bouncy). Harrison *et al.* attempts a first step at finding a mapping between these spaces by running a set of perceptual experiments. The goal of finding the mapping would be to allow animators to specify motion directly using the linguistic motion space.

Experiments were conducted where participants were asked to rate motions (e.g., Figure 9) on a number of different continuous scales. The first experiment was a broad

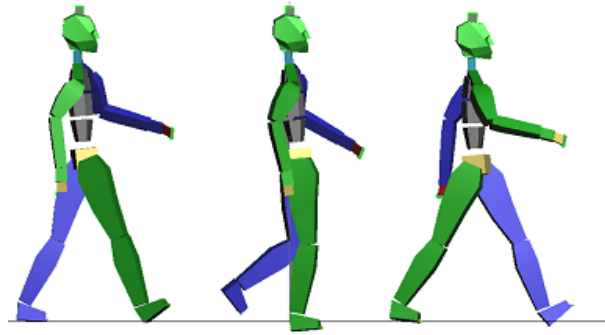


Figure 10: Human walking motion, an example of the type of motion used in [HBF02]

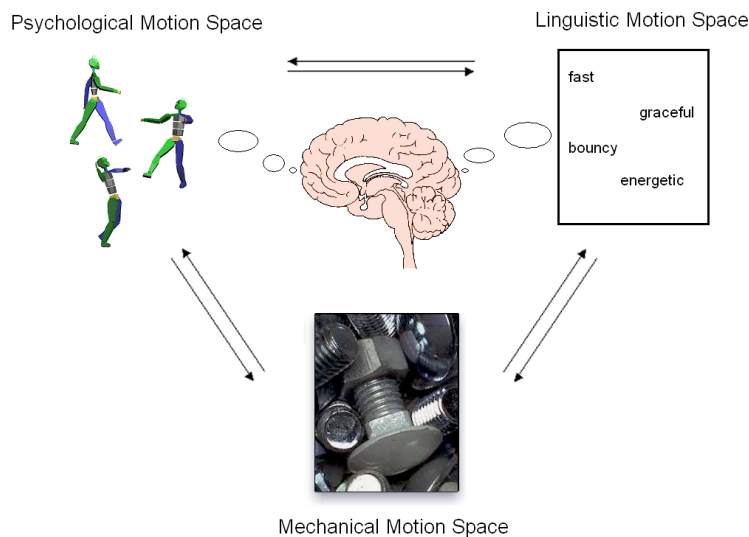


Figure 11: Mapping between the three motion spaces. Image reproduced from [HBF02]

experiment, which gathered information on the types of motions that humans find similar, and also subjective descriptions of the motions. A moderately strong to strong linear correlation was found between the three motion spaces, for each participant. They also found that there were agreements between participants as to the types of parameters that they used to judge similarity of the motions, and also to describe the motions (e.g., observers classified a motion as flexible-stiff from looking at the upper torso and pelvis rotations).

The second experiment was a more complex experiment where a more limited range of movements was tested. Interpolated gaits were created in order to make networks between primary and interpolated gaits. The purpose of these networks was to examine if the links between the motions in the mechanical motion space are similar to that of the psychological motion space, thus exploring the metric proper-

ties of motion similarity judgements. Participants were asked to rate pairs of motions that were linked in these networks over a number of rating scales. When similarity judgements collected from this experiment were tested against the metric properties, evidence showed that similarity judgements do not have all of the metric properties (non-degeneracy, symmetry, triangular inequality). However, they did find that evaluation was similar across participants.

Harrison *et al.* also analysed the relationships between motion spaces, in the hope of finding correlations. They observed inter-participant differences, which they conclude suggests that "animation systems should be customisable not only for the user's preferences, but for their perceptual abilities and movement categories as well".

5. Virtual Environments

At the Campfire, a session was dedicated to the problem of effectively generating images of objects and environments that convey an accurate sense of distance and size [MO01]. Thompson, Gooch, Hu and Willemsen discussed the issues of scale and distance in computer graphics while Beall and Madison asked: how do we know what people really see? Cream addressed the issue of interaction of visual perception and locomotion. She noted that, while people naturally walk through the complex real world while maintaining spatial orientation in an effective manner, when they try to move through virtual worlds using current computer interfaces this is more problematic. She noted that one key research question is whether displays can be built that combine computer-generated visual information with biomechanical information on locomotion, such as generated by walking on a treadmill, to create a more accurate sense of a simulated space than would be possible using computer graphics alone. Shirley discussed the infrastructure necessary to conduct experiments and presented plans to develop a public-domain code infrastructure to make this process easier.

In recent work, Thompson *et al.* [TWG*03] investigated if image quality affects the perception of distance in virtual environments. They conducted a user study where participants performed a triangulated walking task. A target was shown and then removed, and participants were first told to walk at an oblique angle to its position and then to walk towards its remembered position i.e., in a triangular formation. Three environments with drastically different image quality were displayed i.e., a high-quality panorama rendering, a low-quality textured environment and a wire-frame representation. Performance was compared to that in the real-world environment and it was found that image quality played no role, with under-estimation of distances in all three environments being equal, compared with the real-world scenario where estimation was far more accurate. However, Hu *et al.* [HGCRT02] showed that distance perception is affected by cues such as shadows and interreflections when carrying out a placement task in a VE.

Creem-Regehr *et al.* [CRWGT03] carried out a further set of experiments to examine the influence of factors such as field of view and binocular viewing restrictions on people's ability to perceive distance in the real world. From these results, some conclusions were drawn that could be useful in the design of immersive virtual environments. Their motivation was to find further insights into why distance estimations when using HMDs tend to be greatly compressed for directed walking tasks, compared to estimations in real world environments. They artificially created similar restrictions in terms of field of view and binocular viewing in the real world task and discovered that this under-estimation was not due to not being able to see one's own body, or to having a restricted field of view as long as head movement was allowed. They also found that monocular viewing did not

produce poorer performance than binocular viewing. They therefore concluded that these restrictions do not explain the poor performance in distance estimation for such tasks in immersive environments displayed using a HMD.

It seems that the need to see one's own body is not, therefore, paramount for achieving certain tasks in a VE. This finding is echoed in a study by Lok *et al.* [LNWB03], who investigated the role of avatar fidelity on task performance in virtual environments and found that the visual fidelity of the self-avatar had no impact on the reported sense of presence.

In comparison to the results for HMDs, in [CRGST03] the authors found that real-world performance was achieved for slope estimation tasks when using a large scale virtual environment using a locomotion interface, called a Treadport. Large projection screens and a self-propelled treadmill are used along with a tether to simulate forces such as gravity. They conclude that this environment is therefore suitable for use in studies of perception and action.

Latency is another important problem in Virtual Environments, especially when using HMDs. Ellis *et al.* [EMAH04] conducted a series of psychophysical experiments in which participants wore a HMD and were presented with a range of virtual environments. Participants were told to rotate their heads from side to side in a controlled fashion and results indicate that latencies of under 16 ms are required to achieve stability in such circumstances.

Insights are clearly needed into the relationship between delay and user-performance for deciding when there should be a trade-off between delay and quality. Watson *et al.* [WWW03] investigated the relationship of delay to user performance in 3D object placement. They also examine the relationship of previewing to delay and difficulty in a placement task.

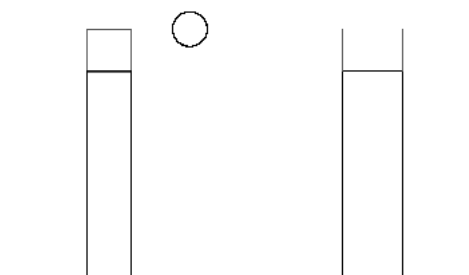


Figure 12: A schematic view of Watson's placement experiments - two pedestals with a translucent box on left and two translucent squares on the right. The spherical cursor is moving between them. (Image from [WWW03] courtesy of Ben Watson.)

Two experiments are described and the setup is shown in Figure 12. For the first experiment, subjects had to place the

sphere into box on the right pedestal by releasing the mouse. For the second experiment, previewing was implemented by a color change when the sphere was in the correct position over the box but had not yet been released by the mouse. The study had 3 levels of delay implemented by adding a delay during each frame and 6 levels of difficulty implemented by varying the width of the right pedestal. For the second experiment previewing was implemented by flashing the sphere when it was in position. Results for the first experiment showed that placement errors and time increased as delay and difficulty increased. Delay has a greater impact on performance when difficulty is high. Results for the second experiment showed that previewing reduces placement times and limits the effects of delay and difficulty. With previewing, more delay can be tolerated even when the task is difficult.

Simulation fidelity of Immersive Virtual Environments and, in particular, the measurement of *Presence* are active areas of research at present. Mania *et al.* [MTRC03] propose the use of memory tasks as a measurement of the fidelity of such environments. A memory task was given to participants, which they performed in a Virtual Environment displayed under a range of viewing conditions i.e., a HMD with or without head-tracking, mono or stereo, and a desktop PC. Performance was compared to that in the corresponding real world environment that the simulations depicted. The authors propose that the relative performance in memory recall, memory awareness states and presence in the real and virtual environments is a good metric for the fidelity of a simulation. Lathrop and Kaiser [LK02] also use task performance as an indication of the fidelity of a Virtual Environment in a similar set of experiments, where participants are required to point in the remembered direction of a previously seen target. Again, desktop, HMD and real-world conditions were presented and they found that performance in the real-world and HMD conditions were comparable, while participants' performance was much less consistent in the desktop condition.

Slater *et al.* propose an alternative approach to evaluating the fidelity of Virtual Environments, based on the concept of Presence, or "of being there" [SBS03, Sla03]. Up to now, questionnaires and interviews have been predominantly used to measure presence, but a new technique, based on physiological measurements, is proposed by the authors. Just as the shock of entering a room with a steep precipice induced a physiological response in users of Meehan *et al.*'s VE [MIWB02], Slater *et al.* postulate that physiological measures could also be used to record and predict *Breaks in Presence (BIP)*. Preliminary results from a VE, where the BIP induced was a white-out of the scene for a 2-second interval, appear to support this hypothesis but it remains to be seen if more subtle artefacts can also be caught using such methods.

6. Visualisation and Non-Photorealistic Rendering

At the Campfire, Interrante talked about designing and implementing algorithms for displaying large, complicated datasets in ways that facilitate the rapid, intuitive appreciation of the essential features of their contents [MO01]. She presented ways in which 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. Her current work on the effect of texture pattern anisotropy on surface shape perception was also presented. Rheingans also discussed perceptually inspired and guided visualization while Duke talked about cognitive and computational processes in visualization. Durand considered the potential of perceptually adaptive Non-Photorealistic Rendering (NPR) techniques in producing clearer, more relevant or more beautiful pictures.

6.1. Perceiving Shape from Texture

When visualising models or datasets in computer graphics, the chosen representation used for displaying the data has a great impact on how easy it is to comprehend its shape and features accurately. This is especially true when displaying unfamiliar structures such as isosurfaces or molecular structures, where the user might not know what general form the data could take.

Kim *et al.* [KHSI03a] presented a series of experiments designed to investigate the impact of various texture pattern characteristics on shape perception. A total of five participants took part in 200 trials, in which they were asked to manipulate probes on an irregular B-spline surface, orienting them to match the surface normal as closely as possible.

In order to examine the effect of texture patterns on this task, three different textures were synthesised and applied to the surface: a one-directional texture in which evenly spaced lines were oriented along the first principal direction of the surface, a two-directional texture where cross-hatched lines followed both principal directions, and a singly oriented Line Integral Convolution (LIC) like texture generated by a blurred noise pattern (Figure 13). A fourth condition where no texture was applied was used as a control. The textures were applied using a method developed recently by Gora *et al.* which allows the orientation of the pattern to follow a specified directional field (in this case the principal directions of the surface) at a per pixel level [GIS03].

The results of these experiments showed that the best performance came from the two-directional texture, followed by the LIC-like pattern, then the one-directional texture, and finally the no-texture condition as expected. From these results Kim *et al.* draw the conclusion that the various principal direction-oriented textures do seem to facilitate accurate shape perception. However, they suspect that several other uncontrolled characteristics of the texture patterns such as

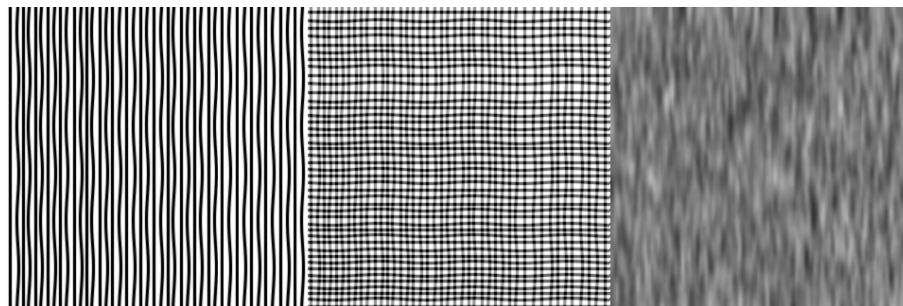


Figure 13: 1-directional, 2-directional, and the lic-like texture patterns, as in [KHSI03b]

spatial frequency and image contrast could also be influencing the shape perception, as well as the interaction between contrast and shape-from-shading information present due to the Phong shading model used in the experiment.

Kim *et al.* [KHSI03a] expand on these results and perform a further two studies involving a variety of surface patch shapes (elliptical, hyperbolic, cylindrical or flat), asking participants to identify both the shape and its surface orientation (convex, concave, both or neither). The aim of the first study is to determine whether shape identification is enabled under some texture conditions but not others, and which texture patterns are more likely to be helpful in facilitating shape perception. It compares the performances of 8 different texture types such as one- and two-directional such as in the above paper, as well as rotated variations of these textures, 3-directional textures, and other types such as swirls and noise. The study is performed under orthographic and perspective projection conditions, and with both head-on and oblique viewpoints, resulting in a total of 592 individual trials.

The results of this first experiment confirmed that texture type does indeed have an effect on shape perception, and that accuracy is generally highest under the principal direction grid condition. Also apparent is that while shape identification is possible in orthographic images with an oblique viewing direction, the best results are obtained using a perspective projection.

The second experiment seeks to further explore the multi-directional texture types, and asks whether the introduction of a third direction to the two principal directions would be an improvement, since it had been discovered in previous research that not all principal directions were equally effective for showing shape. The results indicate that a possible necessity for shape perception is 'shape indicators', contour lines that trace the essential features of a shape, whether they are explicit lines present in the texture, or lines implied from the texture pattern.

7. Non-Photorealistic Rendering

NPR techniques are particularly suited to achieving *functional realism*, as described by Ferwerda [Fer03], and psychological studies are needed to assess this. Ashurst Gooch and Willemsen [GW02] conducted a user study to evaluate how space was perceived in NPR immersive environments. Participants wore a Head Mounted Display and were asked to walk to a target while immersed in an NPR environment depicting a corridor with only edges rendered (e.g., see Figure 14). As a baseline, the participants also performed the task in the real scene. In the NPR environment, perceived distance was found to be 66% of the intended distance, versus 97.0% in the real corridor. This was found to compare well with previous results for traditionally rendered environments. While the authors are not recommending NPR techniques as being particularly effective for conveying spatial information in Virtual Environments, they do emphasise the importance of using sound psychological knowledge and techniques in the evaluation of NPR methods.

Non-photorealistic rendering techniques can also be applied to photographs of scenes and faces (e.g., see Figure 15). Gooch *et al.* [GRG04] also examined aspects of functional realism for NPR techniques, but this time in the context of facial illustrations. They presented a new technique to automatically generate NPR portraits and caricatures from photographs of people (Figure 16). It was noted that such representations could be useful for certain types of applications, such as interactive multi-media, education or psychology research. They evaluated the effectiveness of these representations in conveying information by measuring recognition time for each representation, along with learning speed and accuracy. They found that recognition time for the portraits and caricatures was not significantly degraded with respect to the photographs, while with the learning task, performance was significantly improved with illustrations and caricatures being learned twice or 1.5 times as fast as the photographs respectively.

Duke *et al.* [DBHM03] summarise a series of recent experiments that investigate the emotive content of rendered images. These experiments are concerned with the links be-

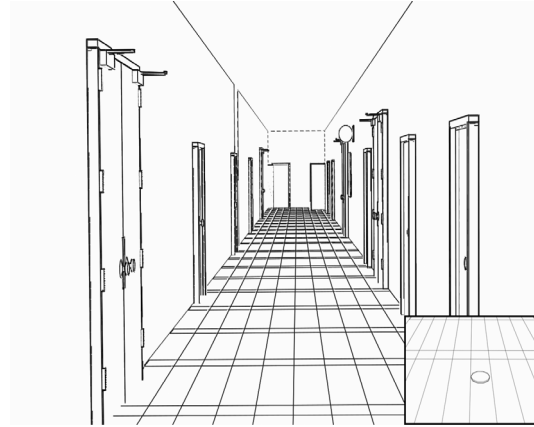


Figure 14: Left: a non-photorealistic representation of a street scene; Right: the corridor scene used in [GW02]. (right image courtesy of Amy Ashurst Gooch.)



Figure 15: A non-photorealistic rendering of a face in pencil style.



Figure 16: Non-photorealistic caricatures of a face automatically generated from a photograph (courtesy of Bruce Gooch).

tween non-photorealistic rendering (NPR), cognition and invariants [BGG96]. Invariants are properties that can be shared by entities and behaviours, which can apply across a range of abstraction levels, and can be restricted in some way by the context in which they are used. The example given in the paper is that of a radar display containing circles and triangles. When asked which group represented hostile units, nearly all subjects identified the triangles - the invariant of sharp pointed shapes invokes a connotation of threat.

They perform a total of nine experiments which explore similar themes, manipulating rendering methods in order to evoke an affective response in the participants. These experiments are detailed in Halper *et al.* [HMDS] and cover topics such as the perception of strength, weakness, danger, safety, and goal-related judgement. They explore participants' different responses to various different stimuli such as sharp jagged edges and other NPR effects.

Duke *et al.* surmise that the results of these simple tests can be extrapolated from and utilised to form a generalised understanding of the association between these rendering invariants and cognitive processes. They conclude that rendering style can convey meaning and influence judgement in a controllable fashion, as well as stressing that stronger links need to be forged between the fields of graphics and cognitive psychology.

8. Conclusions

In this report, we have aimed to give a flavour of recent and ongoing activity in the field of perceptually adaptive graphics. By its nature, such a report is selective and there is much excellent relevant work being carried out in these and related fields that have not been reviewed in this report. For example, Gilmore, Moorhead, Troscianko and Tolhurst presented research at the Campfire on the use of graphics for

camouflage assessment and target signature assessment for military applications, which they have since expanded on [GMW*02].

The issues that we have raised here are becoming increasingly important, as shown by the fact that a new symposium on Applied Perception in Graphics and Visualisation and a journal, the ACM Transactions on Applied Perception, have been started. Courses have been presented at SIGGRAPH that cover perception-related subjects [FRW03, FFM*03, DAG*02]. Such courses and their accompanying notes provide a rich source of information for those interested in learning more about the issues explored in this report. Clearly, the graphics and perception communities have come to realise the importance of working together to improve the quality of computer generated simulations, and also to further our knowledge of human perception.

References

- [ARPT02] ALBIN S., ROUGERON G., PEROCHE B., TREMEAU A.: Quality image metrics for synthetic images based on perceptual color differences. *IEEE Transactions on Image Processing 11* (2002), 961–971. 148
- [BDDG03] BAUDISCH P., DECARLO D., DUCHOWSKI A. T., GEISLER W. S.: Focusing on the essential: considering attention in display design. *Commun. ACM 46*, 3 (2003), 60–66. 143
- [BGG96] BRUCE V., GREEN P., GEORGESON M.: *Visual Perception: Physiology, Psychology, and Ecology*. Psychology Press, 1996. 159
- [BHW96] BARZEL R., HUGHES J. F., WOOD D. N.: Plausible motion simulation for computer graphics animation. In *Computer Animation and Simulation '96* (1996), pp. 184–197. 151
- [BSW85] BAILLARGEON R., SPELKE E., WASSERMAN S.: Object permanence in five-month-old infants. *Cognition 20* (1985), 191–208. 151
- [CCW03] CATER K., CHALMERS A., WARD G.: Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings of the 2003 EUROGRAPHICS Symposium on Rendering* (2003), Christensen P., Cohen-Or D., (Eds.), EUROGRAPHICS, pp. 270–280. 144
- [CF00] CHENNEY S., FORSYTH D.: Sampling plausible solutions to multi-body constraint problems. In *Proceedings of ACM SIGGRAPH 2000* (2000), ACM Press / ACM SIGGRAPH, pp. 219–228. 151, 152
- [Cle82] CLEMENT J.: Students' preconceptions in introductory mechanics. *American Journal of Physics 50*, 1 (1982), 66–71. 151
- [CRGST03] CREEM-REGEHR S. H., GOOCH A. A., SAHM C. S., THOMPSON W. B.: *Perceiving Virtual Geographical Slant: Action Influences Perception*. Technical report uucs-03-012, University of Utah, 2003. 156
- [CRWGT03] CREEM-REGEHR S. H., WILLEMSSEN P., GOOCH A. A., THOMPSON W. B.: *The Influence of Restricted Viewing Conditions on Egocentric Distance Perception: Implications for Real and Virtual Environments*. Technical report uucs-03-016, University of Utah, 2003. 156
- [CT03] CHOUDHURY P., TUMBLIN J.: The trilateral filter for high contrast images and meshes. In *Proceedings of the 14th Eurographics workshop on Rendering* (2003), Eurographics Association, pp. 186–196. 150
- [DAG*02] DURAND F., AGRAWALA M., GOOCH B., INTERRANTE V., OSTROMOUKHOV V., ZORIN D.: *Perceptual and Artistic Principles for Effective Computer Depiction*. Course 12, SIGGRAPH 2002, 2002. 160
- [Dal93] DALY S.: The visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital images and human vision* (1993), pp. 179–206. 148
- [DBHM03] DUKE D., BARNARD P., HALPER N., MELLIN M.: Rendering and affect. *Computer Graphics Forum (EUROGRAPHICS 2003)* 22, 3 (2003). 158
- [DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (2002), ACM Press, pp. 257–266. 150
- [DMMS02] DRAGO F., MARTENS W., MYSZKOWSKI K., SEIDEL H.-P.: *Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference*. Technical report mpi-i-2002-4-002, MPI Informatik, Saarbruecken, 2002. 150
- [DPF03] DUMONT R., PELLACINI F., FERWERDA J. A.: Perceptually-driven decision theory for interactive realistic rendering. *ACM Transactions on Graphics 22*, 2 (2003), 152–181. 147
- [EMAH04] ELLIS S., MANIA K., ADELSTEIN B., HILL M.: Generalizeability of latency detection in

- a variety of virtual environments. In *Human Factors and Ergonomics Society 48th Annual meeting (to appear)* (2004). 156
- [Fer03] FERWERDA J. A.: Three varieties of realism in computer graphics. In *Proceedings SPIE Human Vision and Electronic Imaging '03* (2003). 146, 158
- [FFM*03] FELS S., FISHER B., MACLEAN K., MUNZNER T., RENSINK R.: *Design of Interactive Multimodal Media Systems*. Course 5, SIGGRAPH 2003, 2003. 152, 160
- [FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (2002), ACM Press, pp. 249–256. 150
- [FP04] FARUGIA J.-P., PEROCHE B.: A progressive rendering algorithm using an adaptive perceptually based image metric. In *Eurographics conference proceedings* (2004). 148
- [FRW03] FERWERDA J., RUSHMEIER H., WATSON B.: *Frontiers in Perceptually Based Image Synthesis: Modeling, Rendering, Display, Validation*. Course 3, SIGGRAPH 2003, 2003. 160
- [GH97] GARLAND M., HECKBERT P.: Surface simplification using quadric error metrics. In *Computer Graphics Proceedings, Annual Conference Series* (1997), pp. 209–216. 145
- [GIS03] GORLA G., INTERRANTE V., SAPIRO G.: Texture synthesis for 3d shape representation. *IEEE Transactions on Visualization and Computer Graphics* 9, 4 (2003), 512–524. 157
- [GMW*02] GILMORE M., MOORHEARD I., WATSON G., THOMSON M., YATES T., TROSCIANKO T., TOLHURST D., FILBEE D.: Assessment of synthetic image fidelity. *SPIE Targets and backgrounds VIII: Characterisation and Representation 4718* (2002), 23–34. 160
- [GRG04] GOOCH B., REINHARD E., GOOCH A.: Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph.* 23, 1 (2004), 27–44. 158
- [Guo98] GUO B.: Progressive radiance evaluation using directional coherence maps. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM Press, pp. 255–266. 148
- [GW02] GOOCH A. A., WILLEMSSEN P.: Evaluating space perception in npr immersive environments. In *Proceedings of the second international symposium on Non-photorealistic animation and rendering* (2002), ACM Press, pp. 105–110. 158, 159
- [Hay00] HAYHOE M.: Vision using routines: A functional account of vision. In *Visual Cognition 2000* (2000), vol. 7(1/2/3), pp. 43–64. 146
- [HBF02] HARRISON J., BOOTH K. S., FISHER B. D.: Experimental investigation of linguistic and parametric descriptions of human motion for animation. In *Computer Graphics International* (2002). 154, 155
- [HGCR02] HU H. H., GOOCH A. A., CREEM-REGEHR S. H., THOMPSON W. B.: Visual cues for perceiving distances from objects to surfaces. *Presence: Teleoperators and Virtual Environments* (2002). 156
- [HHO04] HOWLETT S., HAMILL J., O'SULLIVAN C.: An experimental approach to predicting saliency for simplified polygonal models. In *To be—AGPV04* (2004). 145
- [HMDS] HALPER N., MELLIN M., DUKE D., STROTHOTTE T.: Implicational rendering: Drawing on latent human knowledge. Submitted, 2004. 159
- [HOT98] HODGINS J., O'BRIEN J., TUMBLIN J.: Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics*. 4, 4 (1998), 307–316. 152
- [HRD04] HARRISON J., RENSINK R. A., DEPANNE M. V.: Obscuring length changes during animated motion. *ACM Transactions on Graphics (Siggraph 2004)(To Appear)* (2004). 153
- [IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 11 (1998), 1254–1259. 144, 149
- [KG03] KHO Y., GARLAND M.: User-guided simplification. In *Proceedings of ACM Symposium on Interactive 3D Graphics* (2003). 145
- [KHSI03a] KIM S., HAGH-SHENAS H., INTERRANTE V.: Conveying shape with texture: an experimental investigation of the impact of texture type in shape categorization judgments. In *IEEE Information Visualization* (2003), pp. 163–170. 157, 158
- [KHSI03b] KIM S., HAGH-SHENAS H., INTERRANTE

- V.: Showing shape with texture: two directions seem better than one. In *Human Vision and Electronic Imaging VIII, SPIE 5007* (2003), pp. 332–339. 158
- [KP87] KAISER M. K., PROFFITT D. R.: Observers' sensitivity to dynamic anomalies in collisions. *Perception and Psychophysics* 42, 3 (1987), 275–280. 151
- [KPK00] KLATZKY R. L., PAI D. K., KROTKOV E. P.: Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments* 9, 4 (2000), 399–410. 152
- [LCR*02] LEE J., CHAI J., REITSMA P., HODGINS J. K., POLLARD N.: Interactive control of avatars animated with human motion data. In *Proceedings of SIGGRAPH 2002* (July 2002). 152
- [LH01] LUEBKE D., HALLEN B.: Perceptually driven simplification for interactive rendering. In *Rendering Techniques* (2001), Gortler S., Myszkowski K., (Eds.), Springer-Verlag, London, pp. 223–234. 144
- [LK02] LATHROP W. B., KAISER M. K.: Perceived orientation in physical and virtual environments: changes in perceived orientation as a function of idiothetic information available. *Presence: Teleoper. Virtual Environ.* 11, 1 (2002), 19–32. 157
- [LNWB03] LOK B., NAIK S., WHITTON M., BROOKS F.: Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments. In *Proceedings of IEEE Virtual Reality* (2003), pp. 125–132. 156
- [LRC*02] LUEBKE D., REDDY M., COHEN J., VARSHNEY A., WATSON B., HUEBNER R.: *Level of Detail for 3D Graphics*. Morgan-Kaufmann, 2002. 143
- [McN01] MCNAMARA A.: Visual perception in realistic image synthesis. *Computer Graphics Forum* 20, 4 (2001), 211–224. 147
- [MD03] MARMITT G., DUCHOWSKI A. T.: Modeling visual attention in vr: Measuring the accuracy of predicted scanpaths. In *Proceedings of the 14th Eurographics workshop on Rendering* (2003). 143
- [Mic63] MICHOTTE A.: *The Perception of Causality*. Basic Books, New York, 1963. 151
- [MIWB02] MEEHAN M., INSKO B., WHITTON M., BROOKS F.: Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics (SIGGRAPH 2002)* 21, 2 (2002), 645–653. 157
- [MKMS04] MANTIUK R., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics* 23, 3 (2004). 150
- [MO01] MCNAMARA A., O'SULLIVAN C. (Eds.): *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Campfire on Perceptually Adaptive Graphics*. 2001. Abstracts available online: <http://isg.cs.tcd.ie/Campfire/>. 141, 156, 157
- [MTAS01] MYSZKOWSKI K., TAWARA T., AKAMINE H., SEIDEL H.-P.: Perception-guided global illumination solution for animation rendering. In *Proceedings of ACM SIGGRAPH 2001* (2001), ACM Press / ACM SIGGRAPH, pp. 221–230. 149
- [MTRC03] MANIA K., TROSCIANKO T., R.HAWKES, CHALMERS A.: Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence, Teleoperators and Virtual Environments* 12, 3 (2003). 157
- [Mys02] MYSZKOWSKI K.: Perception-based global illumination, rendering, and animation techniques. In *Proceedings of the 18th spring conference on Computer graphics* (2002), ACM Press, pp. 13–24. 148
- [NM00] NAYAR S. K., MITSUNAGA T.: High dynamic range imaging: Spatially varying pixel exposures. In *Proc. Computer Vision and Pattern Recognition (CVPR)* (2000), pp. 473–479. 149
- [OD01] O'SULLIVAN C., DINGLIANA J.: Collisions and perception. *ACM Transactions on Graphics (TOG)* 20, 3 (2001), 151–168. 151
- [ODGK03] O'SULLIVAN C., DINGLIANA J., GIANG T., KAISER M. K.: Evaluating the visual fidelity of physically based animations. *ACM Trans. Graph.* 22, 3 (2003), 527–536. 151, 152
- [ODH03] O'SULLIVAN C., DINGLIANA J., HOWLETT S.: Eye-movements and interactive graphics. In *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research*, Hyönä J., Radach R., Deubel H., (Eds.). Elsevier Science, Oxford, 2003, pp. 555–571. 151
- [OHJ00] OESKER M., HECHT H., JUNG B.: Psychological evidence for unconscious processing of detail in real-time animation of multiple

- characters. *Journal of Visualization and Computer Animation 11* (2000), 105–112. 152
- [OL03] OTADUY M. A., LIN M. C.: Sensation preserving simplification for haptic rendering. *ACM Transactions on Graphics (SIGGRAPH 2003) 22*, 3 (2003), 543–553. 152
- [OL04] O'SULLIVAN C., LEE R.: Collisions and attention. In *Poster Abstract: Proceedings of the ACM SIGGRAPH Symposium on Applied Graphics in Graphics and Visualization (APGV'04)* (2004). 151
- [PFFG98] PATTANAIK S. N., FERWERDA J. A., FAIRCHILD M. D., GREENBERG D. P.: A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM Press, pp. 287–298. 148, 150
- [PS03] POJAR E., SCHMALSTIEG D.: User-controlled creation of multiresolution meshes. In *Proceedings of ACM Symposium on Interactive 3D Graphics* (2003). 145
- [PSE*00] POPOVIC J., SEITZ S. M., ERDMANN M., POPOVIC Z., WITKIN A.: Interactive manipulation of rigid body simulations. In *Proceedings of ACM SIGGRAPH 2000* (2000), ACM Press / ACM SIGGRAPH, pp. 209–217. 151
- [PTT02] PARRAGA C., TROSCIANKO T., TOLHURST D.: Spatio-chromatic properties of natural images and human vision. *Current Biology 12* (2002), 483–487. 150
- [PY02] PATTANIAK S., YEE H.: Adaptive gain control for high dynamic range image display. In *Spring Conference in Computer Graphics* (2002). 150
- [RAGS01] REINHARD E., ASHIKHMIN M., GOOCH B., SHIRLEY P.: Color transfer between images. *IEEE Computer Graphics and Applications: special issue on Applied Perception 21*, 5 (2001), 34–41. 150
- [RD04] REINHARD E., DEVLIN K.: Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics* (2004). 150
- [RP03] REITSMA P. S. A., POLLARD N. S.: Perceptual metrics for character animation: sensitivity to errors in ballistic motion. *ACM Trans. Graph.* 22, 3 (2003), 537–542. 153
- [RPG99] RAMASUBRAMANIAN M., PATTANAIK S. N., GREENBERG D. P.: A perceptually based physical error metric for realistic image synthesis. In *Proceedings of ACM SIGGRAPH 1999* (1999), ACM Press / ACM SIGGRAPH, pp. 73–82. 148
- [SBS03] SLATER M., BROGNI A., STEED A.: Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th Annual International Workshop on Presence* (2003). 157
- [SEL00] SAVAKIS A. E., ETZ S. P., LOUI A. C.: Evaluation of image appeal in consumer photography. In *Proc. SPIE* (2000), pp. 111–120. 147
- [SFWG04] STOKES W. A., FERWERDA J. A., WALTER B., GREENBERG D. P.: Perceptual illumination components: A new approach to efficient, high quality global illumination rendering. In *ACM SIGGRAPH conference proceedings* (2004), ACM Press. 148
- [SHS*04] SEETZEN H., HEIDRICH W., STUEZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A.: High dynamic range display systems. In *ACM SIGGRAPH conference proceedings* (2004). 150
- [Sla03] SLATER M.: How colourful was your day? why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments, in press* (2003). 157
- [SWW03] SEETZEN H., WHITEHEAD L., WARD G.: A high dynamic range display using low and high resolution modulators. *Society for Information Display International Symposium Digest of Technical Papers* (2003), 1450–1453. 150
- [TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision* (1998), IEEE Computer Society, p. 839. 150
- [TWG*03] THOMPSON W. B., WILLEMSEN P., GOOCH A. A., CREEM-REGEHR S. H., LOOMIS J. M., BEALL A. C.: Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments (to appear)* (2003). 156
- [vdDPA*02] VAN DEN DOEL K., PAI D. K., ADAM T., KORTCHMAR L., PICHORA-FULLER K.: Measurements of perceptual quality of con-

- tact sound models. In *Proceedings of the International Conference on Auditory Display* (2002). 152
- [VQE00] VQEG: *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. Tech. rep., VQEG, 2000. 147
- [War94] WARD G.: The radiance lighting simulation and rendering system. In *Proc. SIGGRAPH* (1994), ACM Press, pp. 459–472. 149
- [WB03] WANG J., BODENHEIMER B.: An evaluation of a cost metric for selecting transitions between motion segments. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2003), Eurographics Association, pp. 232–238. 152
- [WB04] WANG J., BODENHEIMER B.: Computing the duration of motion transitions: An empirical approach. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2004), Eurographics Association, p. To Appear. 153
- [WBL02] WANG Z., BOVIK A., LU L.: Why is image quality assessment so difficult? In *Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing* (2002), vol. 4, pp. 3313–3316. 147
- [WFM01] WATSON B., FRIEDMAN A., MCGAFFEY A.: Measuring and predicting visual fidelity. In *Proceedings of ACM SIGGRAPH 2001* (2001), ACM Press / ACM SIGGRAPH, pp. 213–220. 142, 145
- [WGP02] WALTER B., GREENBERG D. P., PATTANAIAK S. N.: Using perceptual texture masking for efficient image synthesis. In *Eurographics Computer Graphics Forum* (2002), vol. 21. 147
- [Win99] WINKLER S.: A perceptual distortion metric for digital color video. In *Proc. SPIE* (1999), pp. 175–184. 147
- [Win00] WINKLER S.: Quality metric design: A closer look. In *Proc. SPIE* (2000), pp. 37–44. 147
- [Win01] WINKLER S.: Visual fidelity and perceived quality: Towards comprehensive metrics. In *Proc. SPIE Human Vision and Electronic Imaging* (2001), pp. 114–125. 147
- [WLC*03] WILLIAMS N., LUEBKE D., COHEN J., KELLEY M., SCHUBERT B.: Perceptually guided simplification of lit, textured meshes. In *Proceedings of the 2003 ACM SIGGRAPH Symposium on Interactive 3D Graphics* (2003). 145
- [WLWD03] WOOLLEY C., LUEBKE D., WATSON B., DAYAL A.: Interruptible rendering. In *Proceedings of the 2003 ACM SIGGRAPH Symposium on Interactive 3D Graphics* (2003). 146
- [WV99] WINKLER S., VANDERGHEYNST P.: Computing isotropic local contrast from oriented pyramid decompositions. In *Proc. ICIP* (1999), pp. 420–424. 147
- [WWH04] WATSON B., WALKER N., HODGES L. F.: Supra-threshold control of peripheral lod. In *Proceedings of ACM Siggraph 2004* (2004). 143, 146
- [WWW03] WATSON B. A., WALKER N., WOYTIUK P., RIBARSKY W.: Maintaining usability during 3d placement despite delay. In *Proceedings of IEEE VR* (2003). 156
- [YP03] YEE Y. H., PATTANAIAK S.: Segmentation and adaptive assimilation for detail-preserving display of high dynamic range images. *The Visual Computer* 19, 7-8 (2003), 457–466. 150
- [YPG01] YEE H., PATTANAIAK S., GREENBERG D. P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph.* 20, 1 (2001), 39–65. 149