Research Goals Statement

Jack Tumblin

I am interested in developing models of human visual perception and appearance to improve the accuracy, quality, and usefulness of computer graphics and digital imaging systems.

Computer graphics allows us to compute the amounts of light we would see in a simulated world, a world that we can create and control by computer. Global illumination methods can accurately compute these amounts of light, and if they could be reproduced by existing display devices, we could look at them to see the exact visual appearance of our synthetic world. However, this approach is impractical because our eyes are far better than our displays: display outputs cover only a small portion of the usable range of our eyes. Instead, we must resort to models of visual perception, and compute a displayed image that best matches the *visual appearance* of the original scene within the abilities of the display device. Similar problems exist in digital imaging systems; limited camera and display capabilities make many scenes such as Figure 1 difficult or impossible to photograph by ordinary means. Good models of visual appearance can help us solve these problems.

I define "visual appearance" as the conscious, picture-like mental estimates of the scene contents we see depicted in a displayed image. Visual appearance is not just an estimate of the scene intensities, but also includes our estimates of reflectance, lighting, shadows, shapes, surfaces, and textures. I am searching for ways to compute the visual appearance of any scene and to make these appearance attributes available in machine-readable forms.

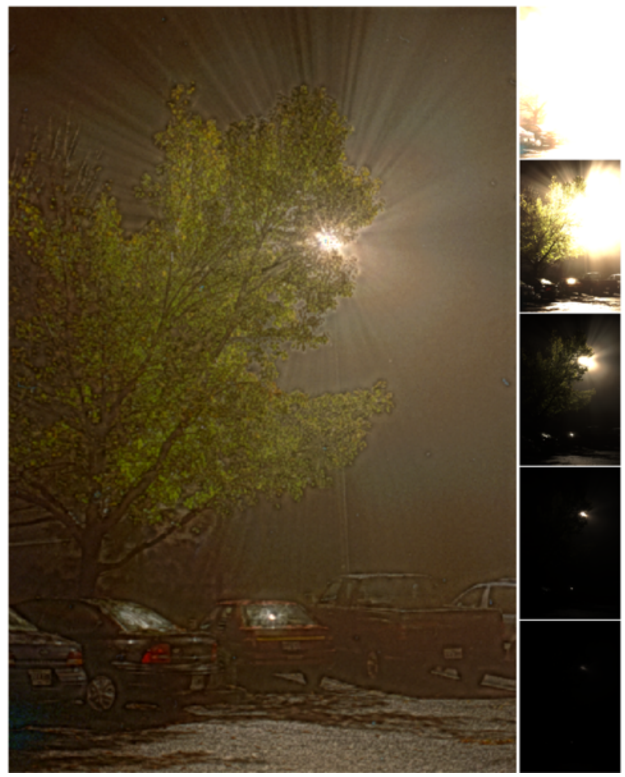

Figure 1: *This low contrast image of a streetlight on a foggy night was made by LCIS methods from an extremely high contrast radiance map. Small images show the original scene radiances scaled by progressive factors of 10. Despite scene contrasts greater than 100,000:1, LCIS methods preserve details impossible to capture in a single photograph, including long, dramatic fog streaks, asphalt texture, and tree details in highlight and shadow. (From SIGGRAPH'99 paper.)*

I am particularly interested in addressing these visual appearance-related problems:

Contrast Reduction: Much of my previous work addresses the display contrast problem; contrasts in real-world scenes can reach 1,000,000:1 or more, but most displays are limited to less than 50:1. How can we reduce contrasts for display without introducing artifacts or destroying low contrast details that are an important part of visual appearance? I found three workable methods. The first uses a sigmoid-shaped function similar to the response of film or retinal ganglia to compress scene illumination components in computer graphics renderings; reflectances and transparencies are preserved. The second method interactively adjusts the displayed image according to the user's direction of gaze to imitate foveally dominated visual adaptation; this method requires a computerized display. Third, the "LCIS" method demonstrated in Figure 1 separates fine details of a scene from its large features using an ordered hierarchy (a scale space) of scene boundaries, then compresses only the large features. I want to further investigate these LCIS hierarchies as a method for finding and encoding boundaries with subpixel precision. I also want to extend LCIS to higher dimensions to mark frameless boundaries of moving objects, for scientific visualization, and other uses.

High Contrast Image Capture and Display: I want to escape the contrast limitations of existing cameras and displays. Like current displays, most cameras also suffer from limited contrast range, and force the user either to adjust scene lighting or to lose important visible details in highlights or shadows. Recently Debevec and others offered ways to capture scene radiances by merging a sequence of photographs taken with different exposure times, but this procedure is often tedious and slow. Few interesting high contrast scenes are truly static; clouds move, trees sway gently, and other movements confound his method with image mismatches. How can we record scene radiances quickly enough to capture more dynamic and interesting scenes, such as sunset at the beach, a campfire, or oncoming traffic at night? I have several approaches I would like to investigate, including multi-pack film, electronic CMOS sensor adaptation, and active optical attenuators.

High contrast display hardware is difficult but not impossible to construct, and directly recreating original scene radiances is sometimes extremely important to verify models of visual appearance. Recently, Greg Ward-Larson used multiple layers of back-lit film to achieve high display contrasts. He distributed high scene contrasts among several different emulsions using spatial frequency distributions to reduce image alignment problems. I would like to extend LCIS methods to find a one-sided, boundary-based image decomposition that would improve such displays. I also plan to investigate other approaches.

Synthesizing Subjective Effects: Many familiar subjective effects of visual appearance are currently difficult to create with computer graphics rendering. For example, visual acuity falls with decreasing light, making reading and other detailed visual tasks more difficult. Though acuity is often modeled as a loss of high frequency response in the visual system, its visual appearance is quite different. Though we might not be able to read the largest letter on an eye chart lit only by starlight, the letter does not appear blurred; we can see that the chart and letters contains sharp boundaries, but we are unsure of their shape or position. How can we visually represent this uncertainty in a displayed image? Other interesting effects include murkiness, clarity, visual after-images, and the soft, slowly changing noise processes seen in night vision or on closing our eyes. Recent work on time-dependent intensity effects led to a paper written with Sumant Pattanaik.

Perceptual Validity: I have long been interested in perceptual validity in computer graphics. How can we ensure that an ordinary displayed image does not reveal any more or any less detail, color, or texture than would be visible to a viewer of the original scene? What components of a displayed image are genuinely important to the viewer? Though our eyes are exquisitely sensitive to image distortions and changes, not all changes are visually interesting. For example, despite available color-management software and hardware, few people bother to carefully calibrate their CRT monitors. However, very small changes to the walking or running gait of a computer animated character strongly affects assessments of its realism. Can we construct visual importance metrics to complement existing visual difference metrics?

Improved Image Representations: Currently, images are expressed as inputs to display devices instead of the human visual system. Displays use pixels, scanlines, and RGB triplets, but we see boundaries, shadings, textures, and movements: shouldn't images directly describe these more visually important components instead? Precomputed, bit-mapped fonts were a good example of display-oriented images, but outline fonts quickly replaced them, because outline fonts accurately describe the boundaries of the typeface and permit visually accurate renditions on any display at any size. Can we find a similarly precise description of other visually salient components to permit scale- and device-independent display for all images?

My current work addresses boundaries, because pixels and scene boundaries are almost antithetical. Pixel-based images are often assumed to be smooth between pixel sample points, and are interpolated with linear reconstruction filters. However, the visual appearance of an image is almost never smooth; we see sharp occlusions and object boundaries in the scene. Visually important boundaries include sharp discontinuities in intensity and its first derivatives, and form piecewise smooth contours that are not constrained by the pixel grid. Visual boundary information is especially useful for accurate compositing operations, and image enlargement or texture mapping without excessive blurring.