Research Statement of Jack Tumblin

Northwestern University, EECS Department, 2006

Why are static photographs of an object or scene so often inadequate as a substitute for direct visual examination? Our eyes sense light, and our cameras are imaging light meters, but this is not enough: light is only a carrier for the information we seek. Our eyes measure absolute amounts of light quite poorly, but are exquisitely sensitive to a very curious set of changes in light intensity, color, and direction, changes that we decode into a partial description of the scene before us. How can we make 'smarter' digital devices that capture those changes better? Can't we do better than simply replacing film with pixels? What, exactly, is the 'visual essence' of a viewed object or a viewed scene, the 'photographic signal' we wish to capture? What parts of visual appearance are missing from conventional photos? What simple, low-cost devices can we make to change 'visual appearance capture' from a vague goal into an objective, straightforward procedure?

Figure 1: *Oil paintings are not 2D images, but "very short sculptures" of many layers, and computational photography methods can reveal them. I developed this rig for low-cost 4-D lighting / capture (funded by NSF, NU) during my collaboration with the Art Institute of Chicago. We found pentimenti ('regrets' or hidden revisions) in several paintings, to augment X-ray, UV fluoroscopic, and IR studies. For example, in Pablo Picasso's "Untitled (Man with Moustache, Buttoned Vest, and Pipe, Seated in an Armchair)" (1915) shown here, the rough paint textures include sand, dried paint scrapings, and possibly even coffee grounds. The black-and-white jester's cap on the man's head (left side) completed a series of revisions that included a curled-brim, bowler-style hat!*

Research Goals

The goals of my research are 1) to devise practicable forms of `computational photography' with new devices, methods, and algorithms that capture visual appearance in machine-readable form; 2) to publish academic articles in the computer graphics and computer vision literature that push digital photography well beyond the intrinsic limitations of film, and 3) to share my scholarship through widely deploying these technologies in real world contexts such as museums and historical sites.

Earlier work: Tone Mapping

These research goals are a substantial expansion of my earlier work, (e.g. [IEEE93, SIGG99, SIGG00]) which is often cited as the roots of a computer graphics (CG) research area now widely known as *tone mapping*. It alerted the CG research community to confront severe dynamic range limitations in digital imaging, and CG rendering and lighting. Tone mapping methods compute displayable images from supposedly 'undisplayable' scenes, ones that include contrasts that exceed the display's ability to reproduce the highlights and shadow details simultaneously. Tone mapping methods perform a perceptual match instead, by computing displayable, non-saturated intensities that still accurately recreate the visual appearance of the original scene. It is this essential step of detail-preserving contrast reduction that makes displayable any computed 'high dynamic range' (HDR) photography or computer graphics renderings.

Before this work, the huge mismatch between display, camera and scene intensities was mostly ignored. Even textbooks assumed that display intensities made directly proportional to scene intensities would be perceptually equivalent, and followed the electrically- and chemically-constrained conventions of TV, film, and printing (e.g. exposure, gamma, saturation). My early work 1) showed that this assumption fails very badly at visual extremes, such as very dark, very bright, or very high dynamic range(HDR) scenes, and 2) defines a computable framework for tone mapping that can be built with psychophysical models, and this framework is still widely regarded as the ‘gold standard’ approach.

Tone mapping and HDR topics have drawn steady and growing research attention. While my own published tone mapping papers[SIGG99, SIGG00,EGSR03 and others] are now dwarfed by a much larger follow-on literature from a broad cross-section of researchers, they continue to be influential and widely cited. For example, our ‘trilateral filter’ [EGSR03] was considered in 2004 by a CIE committee for use as an international tone-mapping standard. Related research and applications in video cards and displays have blossomed, and tone-mapping remains an active research topic even today.

Current Work: Computational Photography

I have continued to push my research work beyond HDR and tone mapping questions to explore the nature of photography itself. Photography should record what humans can see, or would like to see, in machine-readable form. Despite its optical means, the purpose and intent of photography is perceptual. We are trying to capture a transient visual experience and make it tangible, to somehow ‘hold in our hands’ (our computers) the visual essence of what we see. This includes computed partial estimates of reflectance and illumination, sensed visual boundaries from silhouettes, shadows, and occlusion boundaries, joints and seams, as I demonstrated in [EGSR04]. So far, digital cameras don’t help much; they capture exactly what our film cameras did, with roughly the same abilities, limitations and conventions. We choose one moment to ‘click the shutter’, and this fixes a single ‘instantaneous’ film-like decision for focus, viewpoint, field-of-view, exposure, color, lighting, pose, motion blur, movement, timing and more. But something far more interesting and useful is now underway…

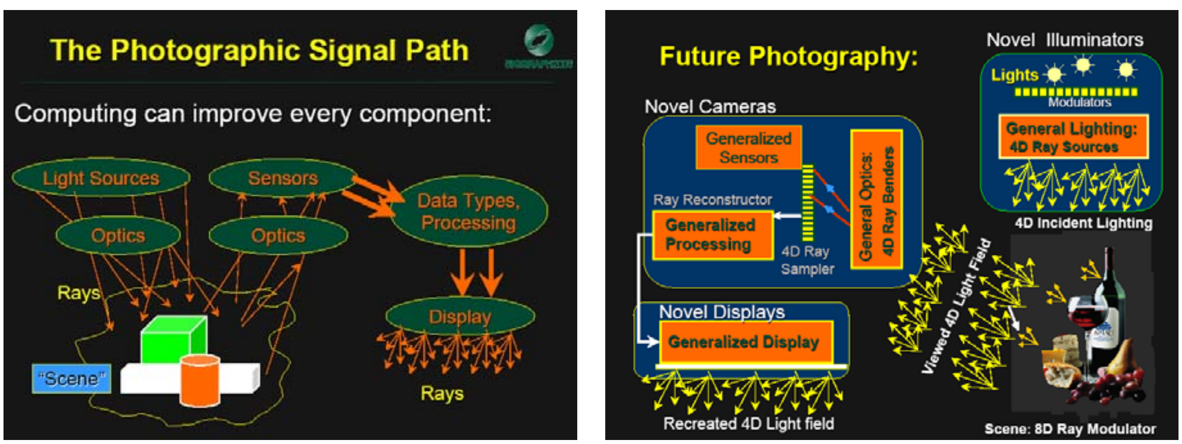

Figure 2: *We can classify the ‘photographic signal’ as measurable changes in 4D ray bundles in 3D space. `Computational Photography’ generalizes each step (light sources; optics; sensors; processing / data structures; displays) to find the most visually expressive changes among these ray bundles.*

My work is leading to a growing awareness that we can apply abundant low-cost digital storage and computing to shift away from these familiar film-like decisions towards higher-dimensional, ray-based methods to capture a visual experience, a research field becoming known as ‘computational photography’. By measuring changes in 4-D ray bundles rather than a 2-D map of pixel intensities, I am among those developing a growing variety of novel methods to compute film-like adjustments *after* we ‘take the picture’. These 4-D ray bundles are liberating because they need not be veridical: we are no longer restricted to use lighting that looks good, nor optics that form the most perfect, aberration-free images [SIGG06, others], nor sensors that measure point-wise grids of pixels[CVPR05, others]. Any mixture of measurements from which we can still compute a useful visual experience is sufficient.

My work is helping to establish and draw attention to this nascent field. I have already:

1. --built **awareness** by constructing a framework for computational photography, with a research 'roadmap'(greatly inspired by Shree Nayar's insights) which has already proved highly influential [SCPV05, SIGG05cp, digital humanities summit05, SIGG06, EG06,] (See Fig 2).
2. --contributed concrete, patentable **advances** to each area of the roadmap:
    - **Light Sources** e.g. [EG05],[EGSR05],[IEEE06]
    - **Sensor Methods** e.g. [CVPR05],[SIGG06],
    - **Optical Methods** e.g. [SIGG06],
    - **Data Representations** e.g. [EGSR04][SIGG05]
    - **Display Methods** [EGSR03],NSF 'Gigapixels' grant, and previous tone-mapping work).
3. --established and obtained funding for a research program to devise practical new tools for **historical preservation** tasks, building new computational photography applications that we can distribute freely (Fig 1) on the web to chronically under-funded museums anywhere in the world.

Awareness: I have organized and taught three tutorial courses at two of the top international graphics conferences ([SIGG05cp] [SIGG06cp] [EG06st]) in collaboration with three of the field's founders: Ramesh Raskar at MERL, Shree Nayar at Columbia, and Marc Levoy at Stanford. My invited contributions as a panel member at last year's ACM Symposium on Computational Photography and Video at MIT [SCPV05] generated wide discussion, and together have prompted 4 invitations for book proposals, another short-course (in preparation with Sylvain Paris (post-doc, MIT) and Pierre Kornprobst (INRIA-Sophia-Antpolis, FR). An invited talk at Adobe Systems (San Jose' CA) "What's Wrong with Pixels?" led to funding ($30,000 unrestricted gift, Dec. 2005) and on-going collaboration. Other recent invited talks included banquet speaker for this year's Human Vision and Electronic Imaging Conference (Jan. 2006) "Re-thinking Photography: Digital Devices to Capture Appearance" (San Jose, CA), related talks at both Cambridge University, Max-Planck Institute für Informatik, Saarbrucken GER (Mar. 2006), and four other upcoming talks this fall. My work has prompted interest and invitations from conservation officials from New York's Metropolitan Museum of Art (May,2006) and the National Gallery(UK) in London (Mar. 2006).

Advances:

The field of computational photography is advancing rapidly, and in a short time I have contributed new methods to each part of the 'roadmap'. My approach is to lower entry barriers as much as possible: devise simple, low-cost devices, anyone can build, extend off-the-shelf items and put all the sophistication in software, rather than high-precision customized lab-grade devices few can obtain.

Light Sources: In our paper "Desktop Photography…" published in ACM / Eurographics Symposium on Rendering in 2005 [EGSR05], [IEEE05] we showed that even the humblest museum can gather computationally relightable photographs and use on-screen sketching to place shadows and highlights at will. Optimization makes calibrated mechanical structures unnecessary(e.g. Debevec's light-stage designs, the Stanford robotic arm); an off-the-shelf steerable spotlight or 'disco light' and a large paper-lined box (~2 meter cube) will suffice for automatic data capture. In this way, any museum can afford to capture their collections in-house without risk from hot overhead lights or malfunctioning robotics.

This advance formed a core component of my proposal to NSF ("Thick Photography: Tools for Rich Digital Archives" ($328,874), which was funded Dec. 2005. Next steps in this research have consisted of a refined device (similar to Fig. 1) for building visual archives of Moche' pottery that are extensible, in the sense that data gathered now can be merged seamlessly with data gathered 20 years later with different equipment. I am conducting this research in collaboration with the Field Museum of Chicago and NU Anthropology Professor Mary Weismantel, and a related device (Fig 1) for paintings.

Curiosity about patterns in this system's raw data led to a broader discovery: in the Eurographics 2005 paper "Light Waving…"[EG05]. We showed how to recover light source angles from uncalibrated video sequences of an object illuminated with a single moving light source, even though the camera cannot see the light source itself. This suggests our simple, cheap 'reductionist' equipment may yet achieve the advantages of much more elaborate and expensive calibrated rigs previously used for higher-dimensional lighting.

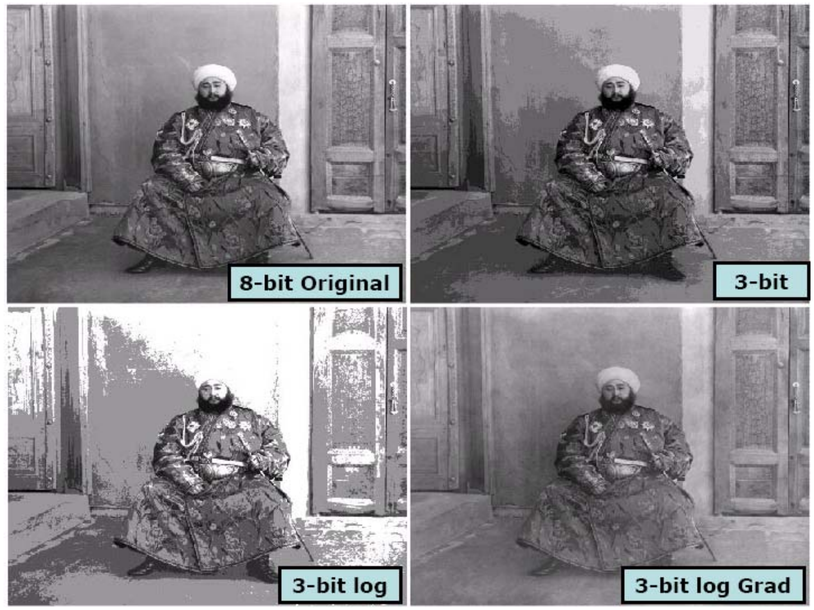

Figure 3: *Gradient sensors recover image boundaries well even under severe quantization, and hide errors as low-frequency noise with a smooth, 'cloudy' appearance. (Enlarge in your PDF viewer to see details)*

Sensor Methods: A key 'computational photography' goal is the notion of replacing pixel sensors with more direct measurements of change in ray bundles. In collaboration with my colleagues at MERL, my paper "Why I want a Gradient Camera" at CVPR 2005 presented a new sensor architecture that measures intensity gradients directly as forward differences, and uses a Poisson solver to reconstruct the displayable image. This new form of computational sensor enables a new kink of HDR sensor that never saturates (no under- or over-exposure), that detects and corrects it's own 'dead pixels', and tolerates noise and quantization errors by hiding them in low-frequency "cloudy" noise that is less disruptive to scene edges, and (to my eye) far less visually objectionable (See Fig 3). The paper led to immediate interest from MERL and Mitsubishi Electric engineering teams, who filed and was granted US and EU patents (I am lead inventor), interest from Cypress Semiconductor to pursue VLSI circuit designs, and new collaboration with Dr. Pradeep Sen, now at University of New Mexico, on sensor cell design and GPU-based reconstruction methods.

Optical Methods: While film-like cameras keep their optics fixed (or stabilized) during exposure for the sharpest result, my colleagues at MERL and I demonstrated that if we rapidly 'flutter' the shutter open and closed during the exposure time, we can computationally remove severe motion [SIGG06]. This research introduces 'coded aperture' techniques into the temporal domain, where the 'flutter' preserves spatial high-frequency components lost in ordinary shutters. We showed that image sequences or 'movies' are not strictly necessary to capture motion, and instead our method captures and decodes linear movement within a single photograph. This idea can also be extended into the sensor itself, applying temporal coding to individual sensors, groups of sensors (e.g. R,G,B) or interleaved sensors in spatially-coded patterns (e.g. Hadamard patterns). My sponsors at MERL and their engineering teams filed for US and EU patent applications immediately, and invited my PhD student Ankit Mohan to a 6-month internship at their Cambridge, MA labs for work on two offshoots of this idea that will be central to his PhD dissertation(2007). Taken together with the 'gradient camera' sensor, these two papers overcome long-standing stumbling blocks in development of low-frame-rate, low-light video cameras, allowing future research work to more directly capture visual appearance of motion.

Data Representations for real and synthetic photography (CG) may eventually describe what our eyes care about most; not pixels, not changes, but the scene features that cause those changes. Often those features aren't detectable in a single image (e.g. occlusion boundaries, transparency, metamerism), but emergent from multiple measurements. Returning to work begun during my 2000-

2001 post-doc, I have explored extensible alternatives to pixel-only image descriptors, finally published as “bixels” [EGSR04]. During that time, others recognized the value of boundaries embedded pixels to enable fast, sparse ray tracing, blur-free GPU-compatible shadow maps, and blur-free features in texture maps. My work on “bixels,” published in the 2004 Eurographics Symposium on Rendering, aimed more broadly at perceived visual boundaries, and offered a modestly better interpolant to reconstruct images. With Prasun Choudhury and Ankit Mohan’s help, I also developed a substantially improved, gradient-based form suitable for editing shadows and scene object boundaries [IEEE06(under review)]. This area’s importance is confirmed by continued strong publishing activity, by boundary representations implemented in GPUs, by proposals of texture/boundary hybrids for specialized image transmission (such as cell phones), prompting invited discussions with Nokia.

In another data representation project, my student Amy Gooch (PhD 2006) recognized that the luminance channel alone is a poor descriptor of the scene contents of images, but is routinely used for grayscale printing of color images. Working together, she and I developed a ‘color2grey’ conversion [SIGG05] that measures and preserves high-dimensional *perceptual* distances between pixels rather than preserving luminance. Optimization ensures that scene features with visibly different colors will retain their visual differences in black-and-white. This work drew immediate attention from Adobe’s Photoshop group, printer manufacturers, and color-blindness researchers; Northwestern Univ. filed for US patents, and several follow-on publications on the topic have already appeared in the literature.

Service to the Field and to the Community at Large

My work with the Art Institute and the Field Museum demonstrates my commitment to making these tools available to experts in the field, and increasingly to lay people through the open-source software community. Already funded by NSF, Adobe Systems, and Northwestern Corporate Partner grants and looking for more, our 4D pentimenti work (Fig. 1) contributes to catalogue preparations for an international exhibition to be announced in 2007-8. As we discover software solutions to curatorial problems, we have begun porting them to the open-source NIPS and VIPS software, used for archival image processing tasks for >15 years. These new tools shouldn’t require computer scientists or technical specialists to use, but instead will be easily available to anyone.

Conclusion

My research contributions began by setting the roots of a research area now widely known as tone mapping and HDR. I progressed to explore how abundant computing and storage can change the nature of photography itself. I have built awareness of computational photography with a research roadmap, and made patentable contributions to each of its areas: computationally enhanced light sources, sensors, optics, displays, and data representations. We are making a leap forwards, towards capturing much more direct descriptions of visual appearance in machine-readable form.