Prakash: Lighting Aware Motion Capture using Photosensing Markers and Multiplexed Illuminators

Ramesh Raskar[*]   Hideaki Nii[†]   Bert deDecker[0]   Yuki Hashimoto[‡]   Jay Summet[§]   Dylan Moore[♭]
Yong Zhao[¶]   Jonathan Westhues[*]   Paul Dietz[*]   John Barnwell[*]   Shree Nayar[♯]   Masahiko Inami[‡]
Philippe Bekaert[0]   Michael Noland[‖]   Vlad Branzoi[♯]   Erich Bruns[♯]

Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA[**]

Figure 1: *Motion and incident illumination capture in a natural setting with imperceptible marker tags. (Left) Traces of hidden tags superimposed on an actor running in sunlight. (Middle) A video frame from a recording session. (Right) A virtual sword inserted with matching location, orientation, lighting and motion blur.*

Abstract

In this paper, we present a high speed optical motion capture method that can measure three dimensional motion, orientation, and incident illumination at tagged points in a scene. We use tracking tags that work in natural lighting conditions and can be imperceptibly embedded in attire or other objects. Our system supports an unlimited number of tags in a scene, with each tag uniquely identified to eliminate marker reacquisition issues. Our tags also provide incident illumination data which can be used to match scene lighting when inserting synthetic elements. The technique is therefore ideal for on-set motion capture or real-time broadcasting of virtual sets.

---

[*]MERL, 201 Broadway, Cambridge, MA, USA

[†]University of Tokyo   [0]Universiteit Hasselt, Belgium

[‡]U of Electro-Communications, Tokyo

[§]Georgia Institute of Technology   [♭]Syracuse University

[¶]Brown University [♯]Columbia University

[‖]U of North Carolina at Chapel Hill [♯]Bauhaus University, Weimar

[**]e-mail:raskar(at)merl.com,web:http://www.merl.com/people/raskar/

Unlike previous methods that employ high speed cameras or scanning lasers, we capture the scene appearance using the simplest possible optical devices – a light-emitting diode (LED) with a passive binary mask used as the transmitter and a photosensor used as the receiver. We strategically place a set of optical transmitters to spatio-temporally encode the volume of interest. Photosensors attached to scene points demultiplex the coded optical signals from multiple transmitters, allowing us to compute not only receiver location and orientation but also their incident illumination and the reflectance of the surfaces to which the photosensors are attached. We use our untethered tag system, called Prakash, to demonstrate methods of adding special effects to captured videos that cannot be accomplished using pure vision techniques that rely on camera images.

1 Introduction

Motion capture is a key component in film and television special effects, user interfaces and analysis for injury rehabilitation. For high speed tracking, the majority of optical motion capture systems use high speed cameras [Robertson 2006]. These camera-based systems require special sensors and high bandwidth, are expensive and use a sanitized environment to maintain a high-contrast between the marker and its background. In this paper, we reverse the traditional approach. Instead of high speed cameras, we use high speed projectors to optically encode the space. Instead of retro-reflective or active light emitting diode (LED) markers, we use photosensitive tags to decode the optical signals. We show that by attaching tags with photosensors to scene points, we can capture not only the locations of each point, but also its orientation, incident illumination, and reflectance.

We describe an economical and scalable system, called *Prakash* (which means "illumination" in Sanskrit), where the light transmitters are **space-labeling** beamers and beacons. Each *beamer* is simply a LED with a passive binary film (mask) set in front. The light intensity sequencing provides a temporal modulation, and the mask provides a spatial modulation. We use a rigid array of such beamers, called *projectors*. The binary masks of individual beamers are carefully chosen to exploit the epipolar geometry of the complete beamer arrangement. Each beamer projects invisible (near infrared) binary patterns thousands of times per second. Photosensing tags determine their location by decoding the transmitted space-dependent labels. They also measure the strength of the signal received from four or more LED *beacons*, which are used to compute tag orientations. The location and orientation data for each tag are computed hundreds of times per second. Finally, the tags measure their incident ambient illumination. When the tagged scene points are imaged with an external camera, we can factor in real time the radiances measured at the corresponding camera pixels into the incident illuminations and the intrinsic reflectance of the corresponding scene points. Because the entire working volume of the system is optically labeled, the speed of the system remains constant, regardless of how many tags (scene points) are tracked.

The main features of our approach are:

- a novel spatio-temporal modulation that exploits the epipolar geometry of a cluster of transmitters to compute location in 3D space;

- an intensity-based technique for computing the orientation of sensor tags - a feature not available in most existing optical markers;

- a simple method to compute the intrinsic reflectance of each scene point by sensing its irradiance and factorizing radiance captured by a camera;

- the use of markers that can be imperceptibly integrated with a performer's costume and shot under natural or theatrical lighting conditions, and that can simultaneously measure photometric attributes; and

- new algorithms to support graphics/vision applications that exploit the sensed geometric and photometric quantities.

A key advantage of our approach is that it is based on components developed by the rapidly advancing fields of optical communications and solid-state lighting. It allows us to easily capture both geometric and photometric quantities. Marker-based techniques that use other physical media cannot capture photometric properties. A disadvantage of our approach is that the tags must be in the line of sight of the projectors. In addition, our system performance degrades with limited dynamic range when the ambient lighting is very strong and when confronted with multi-path distortions due to secondary scattering of light Hence, the approach is not suitable for all scenarios. However, our technique is effective in cases where traditional systems are incompatible with the capture environment.

Since our interest is in solving graphics problems, we will focus on the recovery of geometric and photometric scene properties; we do not discuss optical communication issues such as power, range, synchronization, noise and interference. We refer the reader to our supplementary material for detailed discussions on several aspects of hardware design and implementation that lie beyond the scope of our contributions in this paper but are necessary for reproducing our results.

1.1 Related Work

**Location Tracking**   Several options for motion tracking using magnetic, acoustic, optical, inertial, or radio frequency signals are available (as discussed in a survey by Welch and Foxlin [Welch and Foxlin 2002] and in Hightower and Borriello [Hightower and Borriello 2001]). Optical systems typically experience lower latencies and provide greater accuracy. We exploit the optical channel to investigate the photometric aspect of scene capture.

Let us consider optical location tracking with tags. Motion capture systems used in movie studios commonly employ high-speed cameras to observe passive visible markers or active light emitting diode (LEDs) markers [ViconPeak 2006; Motion Analysis Corporation 2006; Phase Space Inc 2007; PTI Inc 2006; Optotrak 2007; Codamotion 2007]. For example, the Vicon MX13 camera can record 1280x1024 full-frame Gray-scale pixels at speeds of up to 484 frames per second with onboard processing to detect the marker position. These devices provide highly reliable output data and have been developed over the last three decades. However, expensive high-speed cameras pose several scalability issues. Bandwidth limits resolution as well as frame-rate. Higher frame-rate (i.e., shorter exposure time) requires either brighter controlled scene lighting for passive markers or the use of power hungry active LED markers. To robustly segment the markers from the background, these systems also use methods for increasing marker contrast. This usually involves requiring the actor to wear dark clothing under controlled lighting. The use of photosensing allows capture in **natural settings**. Since the photosensors are barely discernible, they can be embedded in a wide range of natural clothing so long as the photosensing element is exposed. The power of emitters is comparable to the IR emission from TV remote controls. Instead of high-power emission, we exploit high-frequency modulation to robustly communicate with the photosensors. Similar to photosensors in TVs, our sensors will work in many lighting conditions. So, in studio settings, the actor may wear the final costume, and he/she can be shot under theatrical lighting.

In camera-based motion tracking, it is difficult to create tags that are invisible to the recording camera but discernible to the high speed motion-capturing camera, complicating the use of on-set motion capture or real-time virtual set broadcast. One approach uses high power infrared LEDs, contrasted with a dark background, and tracks the markers with a near infrared (IR) sensitive camera in an IR-free environment. IR light can be reduced by avoiding sunlight or by affixing IR-blockers to indoor lights. To produce sufficient power for the markers, bright lights must be powered on throughout the short exposure time of a high-speed camera's frame. Our approach avoids this problem by using 455 kHz modulation, which is detectable in full sunlight, similar to that of a TV remote control. However, a high speed camera cannot perform demodulation. More recent 1D linear array cameras can perform some low-frequency modulation [Phase Space Inc 2007; Optotrak 2007] because they are about 1000 times faster. One can view our approach as an extension of the trend from a 2D high speed camera to a 1D higher speed camera; our system comprises a very high speed zero-D (point photo-sampling) camera.

Let us now consider the update rate. Systems that use cameras are limited by their frame-rate. Active beacons must use time division multiplexing [PTI Inc 2006] so that only one LED can be turned on at a time. Each additional tag requires a new time slot. Hence, the total update rate is inversely proportional to number of tags. Such systems are ideal for applications requiring high speed tracking of a small number of tags, e.g. head or hand tracking in virtual reality systems. Passive markers need to resolve correspondence to avoid the "marker swapping" problem.

Although our current prototype lacks the engineering sophistication to compare with state-of-the art motion capture systems, our system should be capable of update rates approaching 1MHz using inexpensive, off-the-shelf optical communication components and provide orientation plus incident illumination measurements.
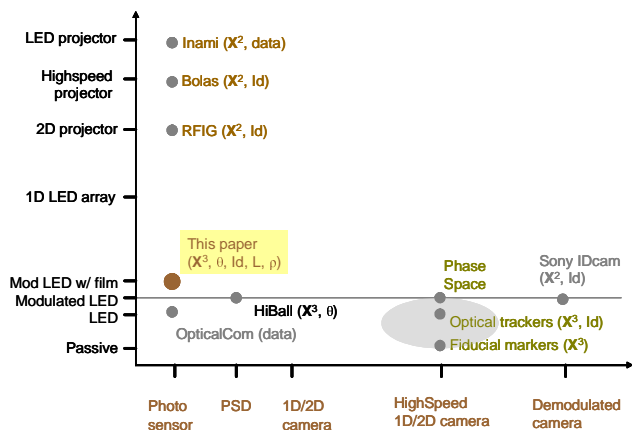
Figure 2: *Communication choices between light sources and receivers plotted as complexity of receivers (X-axis) versus complexity of transmitters (Y-axis).*

Photosensing Without cameras observing the scene, how can we find the location of a marker (tag)? The solution is to optically label the space using a collection of inexpensive high-speed beamers. Previous efforts to locate objects by employing attached photosensors include Ringwald's PDA-based interaction and the system by Patel et al. which use active RF tags that respond to laser pointers [Ringwald 2002; Patel and Abowd 2003]. Alternatively, the FindIT flashlight uses a one-way interaction and an indicator light on the tag to signal that the desired object has been found [Ma and Paradiso 2002].

Some have presented the idea of locating photosensing tags with a traditional video projector[Nii et al. 2005] [Raskar et al. 2004] [Lee et al. 2005]. Their goal was pose estimation and augmented reality. Rather than using an expensive video projector, we exploit the symmetry of the projected pattern and achieve the same with a set of light sources at distinct strategic locations. It is tempting to use a TI Digital Light Processing (DLP) system or Gated Light Valves (GLV) for high-speed binary coding [Bolas et al. 2004; Cotting et al. 2004; Silicon Light Machines 2006]. However, the DLP is designed for general purpose image projection as opposed to repeated pattern projection. By limiting ourselves to a specific goal, we are able to use a more common light source, an LED, making our system scalable, versatile, compact, and inexpensive. We showed our system in operation at Siggraph 2006 Emerging Technologies [Nii et al. 2006].

The UNC HiBall system [Welch and Bishop 1997] uses a group of 6 rigidly fixed position sensitive detectors (PSD) to find location and orientation with respect to actively blinking LEDs. Each LED provides a single under-constrained reading at a time. The system requires a large ceiling installation. We use a single photosensor in place of the 6 PSDs and do not require active control of the LEDs; they run open-loop. Thus, multiple receivers can simultaneously operate in the same working volume.

Systems such as Indoor GPS [Kang and Tesar 2004; Sorensen et al. 1989], Spatiotrack [Iltanen et al. 1998] or Shadow Track [Palovuori et al. 2001] use low-cost photosensors and two or more mechanically rotating light beam sources mounted in the environment. The spinning light sources sweep out distinct planes of light that periodically hit the optical sensors, and the system uses the temporal phase of the hits to derive the position at 60 Hz. However, fast-switching, solid-state emitters have the potential to achieve extremely high rate at a very low cost and form-factor compared to mechanically rotating devices. In addition, we believe our system is unique in also

exploiting the photodetector to measure photometric attributes.

2 Communication with Optical Tags

Our choice of spatio-temporal optical modulation is influenced by developments in bandwidth growth in optical communication, and opportunities in the projective geometry. A range of optical emitter and receiver options exist for employment in an optical capture system. Typically, there is a tradeoff in complexity of the transmitter and the receiver. The Figure 2 shows a few systems and indicates the type of measurements: dimensions for location sensing (X^d), orientation (θ), identity (Id), illumination (L) and reflectance (ρ). At one end, Sony ID Cam [Matsushita et al. 2003] which is a specially built 12 kHz demodulating camera, can track modulated LEDs; at the opposite end a per-pixel modulated LED projector can track a photosensor [Nii et al. 2005]. This asymmetric complexity makes it difficult to achieve high-speed capture with low-cost off-the-shelf or easily built components. Further, most systems report only tag ID and location of the tags. Using a simple photosensor receiver, our system is able to achieve several magnitudes of cost and bandwidth savings by switching to a unique multi-beamer configuration: distinct light emitters that geometrically behave as a single logical emitter. For simplicity of implementation, our method employs time-division multiplexing using readily available components. Unlike radio frequency communication, in the optical domain amplitude, frequency (FDM) and code division (CDM) multiplexing are still evolving [Azizoglu et al. 1992]. With these later multiplexing schemes, one can build 'always-on' unsynchronized multi-beamers in the future.
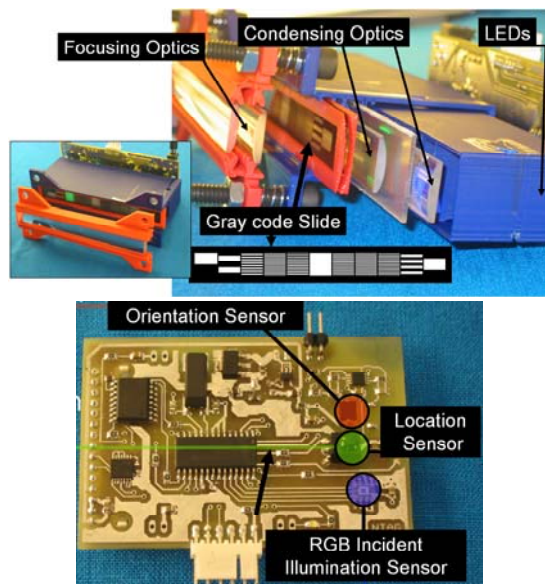
Figure 3: *Our prototype (Top) projector (a multi-beamer array), and (Bottom) tag receiver.*

2.1 Receiver Tag

Although various combinations are possible, we suggest the simplest possible optical receiver, a photosensor. Our photosensor excludes the customary lens so that when the sensor is tilted, the resulting photocurrent exhibits a cosine falloff. The receiver can report three quantities: the demodulated signal (which yields the position data), the received signal strength (which is used to determine orientation), and the base, DC level when no projectors are transmitting (which reveals the ambient light level). To simplify

the design and electronics, our current prototype uses three distinct photo-sensors - one for each task and senses these three values in three separate time slots (Figure 3). Each tag is equipped with flash memory for storing data and an RF antenna for wireless communication.

2.2 Transmitter

Our transmitter, a beamer, is similarly simple. It consists of a light emitting diode (LED) with a passive film set in front. The LED is temporally modulated at 455 kHz. The binary film achieves fixed spatial modulation–the optical signal is transmitted in parts where the film is transparent and blocked where the film is opaque. The binary film represents one bit position of the binary Gray code and represents one of the patterns of the conventional binary coded structured light illumination. An RF system is used to retrieve data from the tags. Since we are using time division multiplexing, each LED beamer is assigned a time slot. The receiver decodes presence or absence of modulated light in this time slot with synchronization. The transmitters run in an open loop. The tag stores the decoded data in onboard flash memory or transmits via the RF channel. A camera is optionally included in the system to compute reflectance.
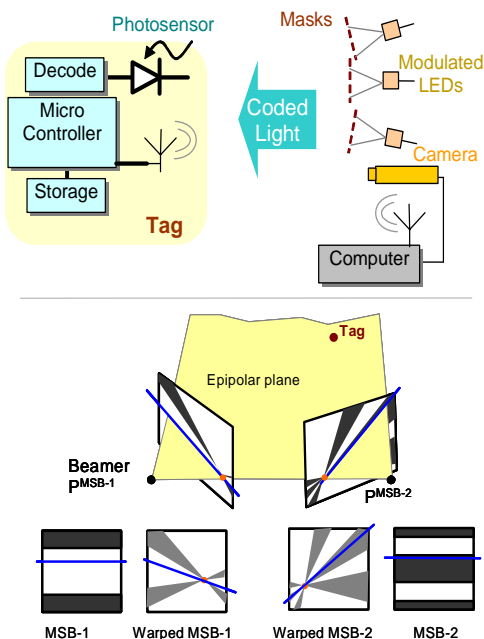
Figure 4: *A conceptual location tracking system using unstructured beamers. (Top) The interaction of a photosensing tag with multiple 1-bit beamers. (Bottom) Exploiting epipolar geometry for superimposed projection. Two patterns of traditional Gray-coded sequence can be aligned even if they are projected from non-colocated centers of projection. The Gray-coded patterns, shown here for bits MSB-1 and MSB-2, are warped and presented as passive masks in front of the LED. A shared epipolar line (blue) before warping, projects into a shared epipolar plane in yellow.*

3 Estimating Scene Parameters

The optical tags offer the ability to directly compute the location, surface orientation, and aggregate incident illumination from multiple ambient light sources. These three quantities can in turn be used to compute the surface reflectance.

3.1 Location

In location computation from multiple beamers that are optically emitting only a single spatial bit, the core idea is in exploiting the epipolar relationship between emitters and their patterns.

While we use the traditional Gray coded patterns, each bit of the pattern is projected by a different beamer. In traditional Gray coding projection, all patterns originate from a single projector. Our goal to is to project a different binary code from each of the multiple non-colocated beamers. (In a FDM or CDM technique, all the beamers will be on simultaneously). As seen in Figure 4, two LEDs set behind different Gray-code patterns can emit code patterns that create an appropriate pencil of planes if and only if the patterns are aligned along the corresponding epipolar lines. A new LED beamer can be added to this set by ensuring the corresponding epipolar planes are identical. This is possible only if the center of projection of the new LED beamer is collinear with the first 2 LED beamers. We can see that, to build a multi-bit projector, the center of projection of all LEDs beamers must be collinear as well. For collinear centers, the epipoles are at infinity and the corresponding epipolar lines are parallel to each other.

It would, however, be quite difficult to build N different 1-bit beamers with collinear centers of projection and align them such that the patterns share the parallel epipolar constraints. Each pattern must be aligned to within the angular resolution of a Least Significant Bit (LSB) pattern. For a 10 bit code on a 10 mm-wide mask, the LSB pattern resolution is 10 μm. To bypass the need for precise mechanical alignment we observe that each beamer pattern remains constant along the epipolar line direction and changes only along the direction perpendicular to the epipolar lines. Therefore, rather than using spherical lenses, which would focus in both directions, we can use cylindrical lenses, oriented to focus along the direction perpendicular to epipolar lines. We guarantee precision by using a single lens and a single multi-pattern mask rather than N lenses and N masks.

The collection of N light sources behind a common lens and mask create a compact array for coding one dimension of the geometry. Each beamer array is effectively a projector emitting 10,000 binary Gray-coded patterns per second. In addition, the binary patterns are modulated with a 455 kHz carrier. The tags decode the presence or absence of a 455 kHz carrier as 1's or 0's to directly compute the projector space coordinate. By using 3 or more such beamer arrays, we compute the 3D location of the tag.

Validation The location sensing is affected by focusing ability of the projectors, noise in sensing the modulated carrier due to ambient IR, square distance fall-off of LED intensity, occlusions and time-skewing. The depth of field for our system is 2.0 meters centered at 3 meters from the projector. The decoding photosensing modules are robust with respect to ambient IR and work indoors at the range of up to 5 meters and in bright outdoor conditions for up to 3 meters from the projector. The tags may receive spurious data due to partial occlusion or secondary scattering which can result in a signal that is plausible but incorrect. Camera-based systems can possibly reject such samples.

Our current system uses time division multiplexing which violates the simultaneity assumption in location estimation [Welch and Bishop 1997]. The various location bits are measured with a minimum delay of 100 microseconds and a maximum delay of 2 milliseconds between them in our current 10 kHz optical bit-rate, 500 Hz location tracking system. Due to this time skewing, the bits will not correspond to a single world location if the photosensor is displaced during this finite time. Some tags may also miss a part of the temporal code due to occlusions. A Kalman-filter based predictive
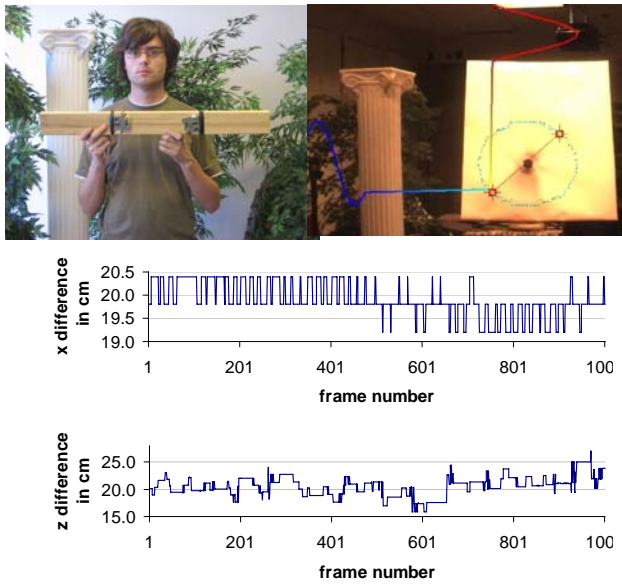
Figure 5: *Verification of 1D resolution. (Top Left) We computed distance between two tags on a rigid object for quantitative verification. (Top Right) Tags attached to fan blades provide a visual verification via sinusoids for x (red) and y (blue) coordinates. (Middle and Bottom) Distance between two tags when the movement is along x axis and along z axis, respectively.*

Figure 6: *Verification of orientation estimation. (Top Left) Cosine fall off as a function of orientation for the photosensor. Blue: Ideal cosine curve, Black: Observed fall off, Brown: Mean error from the ideal, the error is below 5% for +/- 60 degrees, Green: Standard deviation in error from the ideal. (Top Right) Verifying orientation of parallel tags. (Bottom) Verification of relative orientation between two tags in presence of large 3D motion.*
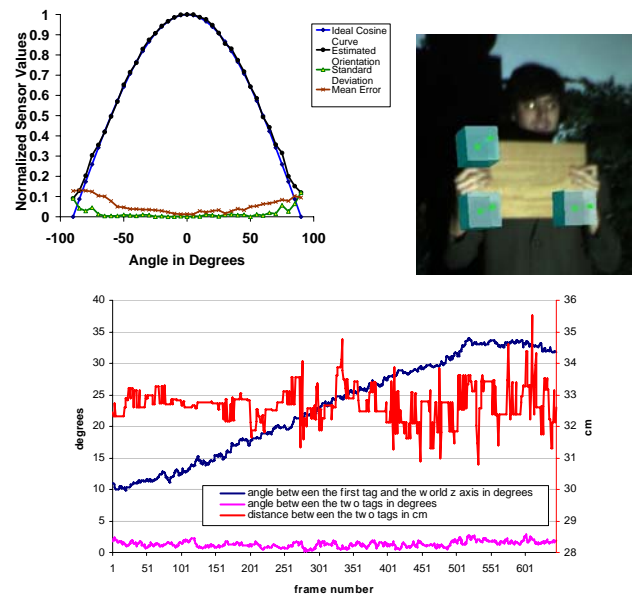
approach is a common solution to deal with dynamic systems.

The maximum marker frequency is 1000 Hz for 1D sensing. At 3.0 meters, the field of view is 1.5m x 1.5m, or 27 degrees. We verified the resolution of 1D location by translating two tags mounted on a rigid structure and computing the 1D distance between them, which should remain constant. The plot in Figure 5 shows that the average distance is 199.01 mm, the standard deviation is 3.70 mm and the maximum deviation is 7.01 mm. This agrees with the theoretical resolution of 5.8 mm at 3 m in *x-y* direction for a 8 bit location system.

Accuracy in the z direction is a function of the baseline. If the angular resolution of the projector is α and the baseline between projectors is d, then *z*-axis resolution at depth p is $\alpha \left(d^2 + p^2 \right) / (d - \alpha p)$. In our case the angular resolution is 0.0017 radians (0.1 degree), and the baseline is 1.3m. Hence at 3 meters the theoretical resolution is 22 mm. We moved the same two tags in *z*-direction away from the projectors. The plots below show that the average difference in z value is 205.58 mm, the standard deviation is 18.33 mm and the maximum deviation is 64.29 mm.

3.2 Orientation

Conventional techniques to compute orientation of tags are composed of magnetic trackers or based on integrating readings from inertial sensors. Optical solutions include position sensitive detectors (PSD) [Welch and Bishop 1997] but require a large form factor due to the lens. The Hamamatsu S6560 sensor [Hamamatsu Photonics 2005] detects incident light angle by differentially processing the output current of the built-in dual photodiode, but it can sense tilt in only one known dimension. One could estimate tag orientation by sensing the relative location of 3 or more sensors rigidly mounted on the tag. However, this becomes unreliable as the distance between the mounted sensors approaches resolvable location resolution. We describe a solution that fits within our framework of

a single photosensor tag to estimate instantaneous orientation. Our method estimates photosensor orientation by exploiting the natural cosine falloff due to foreshortening and employing the known instantaneous location estimation.

We assume the photosensor (without lens) is attached parallel to the surface. We then find the surface normal–i.e. orientation up to two degrees of freedom. We cannot estimate the rotation around the optical axis of the photosensor. The incident angles between the incident rays from distinct light sources and the surface normal attenuate the received strengths from those sources sensed at the photosensor by the cosine falloff factor. By observing the intensity from four or more light sources with known location and brightness, we determine the surface normal associated with the sensor.

However, the relative brightness of the transmitter beacons is a function of electro-optical characteristics and is often unknown. Can we use the multiple values measured at a single moving tag to achieve self-calibration, i.e. to simultaneously estimate the relative brightness of emitters and the tag orientation? We show that the intensity due to a minimum of 3 light sources measured at 3 or more distinct locations is required. Then we unambiguously establish the orientation of the tag with respect to each of the illumination sources.

Note that this problem differs from triangulation of location using a set of received signal strengths. Let's say we must estimate brightness of l light sources by moving a tag to m positions. At each tag position we have 2 unknowns for orientation, and hence we have $l + 2m$ unknowns. At each of the m positions, we have l readings, and hence $l * m$ equations. Since, $l * m > l + 2m$ for $l >= 3$ and $m >= l/(l-2)$, it follows that we need a minimum of 3 light sources and 3 tag positions. Consider the light source L_i with power P_i and the vector from tag to this light source $V_i = [V_{ix}, V_{iy}, V_{iz}]$ nor-

malized with the distance d_i between the tag and the light source. We can estimate the normal N using the intensities I_i recorded from the i-th light source.

$$I_i = k \cdot (V_i \bullet N) \frac{P_i}{d_i^2}$$

Where k is an unknown gain of the sensor. Let us substitute,

$Q_i = I_i d_i^2 / k P_i$, and

$$\mathbf{V} = \begin{bmatrix} V_1 \\ \vdots \\ V_l \end{bmatrix} \quad N = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \quad b = \begin{bmatrix} Q_1 \\ \vdots \\ Q_l \end{bmatrix}$$

We have, $V \cdot N = b$ and $N = V^+ b$, where V^+ is the pseudo-inverse of V.

Since, $|N|^2 = 1$, we have $N^T N = 1$, i.e., $b^T V^{+T} V^+ b = 1$

Let us substitute, $V^{+T} V^+ = C$, so that, for each position j of the tag we have the quadratic constraint on the unknown light source powers,

$$b^T C^j b = 1, j = 0..m-1$$

From three or more positions, we can estimate Q_i, using a nonlinear optimization. This problem has similarities to photometric stereo under unknown lighting [Basri and Jacobs 2001]. To find the initial guess, we restate the quadratic constraint in terms of $C_2^{l+1} = (l(l+1)/2)$ scalars c_i from the $l \times l$ symmetric matrix C. In practice, with 6 or more positions, we estimate the quadratic terms and employ the estimated brightness values as an initial estimate for nonlinear optimization. Once the brightness values are known, the tag orientation estimation is a linear process. More light source beacons improve the estimate at the cost of reducing the frame rate due to additional time slots required. In our simulations we found the solution with four beacons to be sufficiently stable. We typically perform the self-calibration only at the beginning and use the estimated brightness values for the remainder of the session. To improve signal strength, we used bright IR LEDs as beacons and spread them around the origin of the projectors.

Validation The intensity values are prone to distortions due to interreflections in realistic settings. The sensitivity is also low due to minimal changes in cosine values when the sensor is orthogonal to the illumination direction. In our current system, we use eight orientation beacons to minimize occlusions issues. We verified the cosine profile in a calibration setup by carefully reading the photocurrent 20 times and changing the orientation in 5 degree increments. Figure 6 plots the difference between ideal and measured cosine values. We found the photosensor to be consistent with cosine curve upto +/- 60 degrees (with mean error less than 5%). Hence, we ignore intensities values from a beacon that are below half the maximum value estimated at a given 3D location.

Note that the orientation estimation accuracy is affected by errors in estimated 3D location. We simultaneously verified 3D location difference and orientation difference by attaching two parallel tags to a rigid structure that was then moved and rotated. Figure 6 plots (in blue) the angle between the first tag and the z axis of the world coordinate system showing a variation from 10 to 35 degrees during our recording. The angle between the two tags (in magenta) has a mean 1.32 degrees with a standard deviation of 0.46 and a maximum deviation of 1.49 degrees. Thus, the orientation estimation at 3 m with 8 beacons spread over 2.5 m is reliable to within 1 degree. The 3D Euclidean distance between the two tags was an average of 327.02 mm with a standard deviation of 5.51 mm and a maximum deviation of 28.07 mm.

3.3 Illumination

We measure the flux arriving at the photosensor using the photocurrent, which is a measure of the ambient irradiance multiplied by the sensor area. The ambient flux is measured in the visible range, and hence, it is not affected by near-IR emission of beamers or beacons. The area of the detector is fixed and the photocurrent is proportional to the integration of the irradiance over the hemisphere. One can use a temporal low pass filtered version of the photocurrent to estimate the ambient illumination since the ambient illumination is temporally smooth compared to the 455 kHz modulating signal. However, to sense color, we used a separate triplet of sensors, for red, green and blue wavelength. The sensors have a broad wavelength response with overlapping lobes. If the sensor area is dA, the photocurrent for given irradiance E is given by the product of incident light's radiance multiplied by the cosine of the angle between the light vector, L and the tag normal, N, integrated over the hemisphere Ω.

$$Photocurrent \propto dA * E = dA \int_{\Omega} P_i (N \cdot L_i) d\omega_i$$

Note that irradiance integrated over the whole hemisphere includes direct as well as global illumination. We show several demonstrations to validate the accuracy of RGB sensing. Figure 7 shows ability to change the color of virtual labels using the sensed color of incident illumination.

3.4 Reflectance

Traditional scene capture methods use cameras to understand the interaction between the geometric and photometric attributes of a scene. However, robust scene understanding using camera images is known to be a difficult inverse light transport problem. The radiance measured at each camera pixel is a complex function of geometry, illumination, and reflectance. Since our system provides estimates of location, surface orientation, and incident irradiance, we can compute the reflectance of a scene point in a straightforward way if we can also sample the reflected radiance using an external camera. Factoring radiance measured at a camera into illumination and surface reflectance can be accomplished at very high speeds (much faster than with a typical camera), allowing the manipulation of individual video frames at an intra-image level.

The radiance, B, is related to the irradiance, E, and the Lambertian reflectance, ρ, and measured by the camera with an unknown scale.

Since, $B \propto \rho E$, for each wavelength, λ, we have

$$\rho_\lambda \propto \frac{B_\lambda}{E_\lambda} \propto CameraPixReading_\lambda / \Gamma(SensorPhotocurrent_\lambda / dA)$$

Where, $\Gamma(.)$ is the color transform matching the RGB sensor reading to the measured camera tri-color intensities. This ratio greatly simplifies the surface albedo computation.. We estimate the albedo up to an unknown scale by taking the ratio of the camera pixel intensities and RGB sensor values on the tag.

Although the incident illumination and albedo measurements are made at a sparse set of scene points, their richness allows extrapolation within a small neighborhood. We also solve the correspondence problem over successive frames via assigned tag IDs. We compute intrinsic reflectance for pixels around the tag sensor assuming a low frequency ambient illumination. The instantaneous photocurrent, however, is noisy. So we exploit the fact that the scene point is visible over several camera frames under varying location, orientation, and illumination values, and take a weighted average to get a more robust estimate of the reflectance.
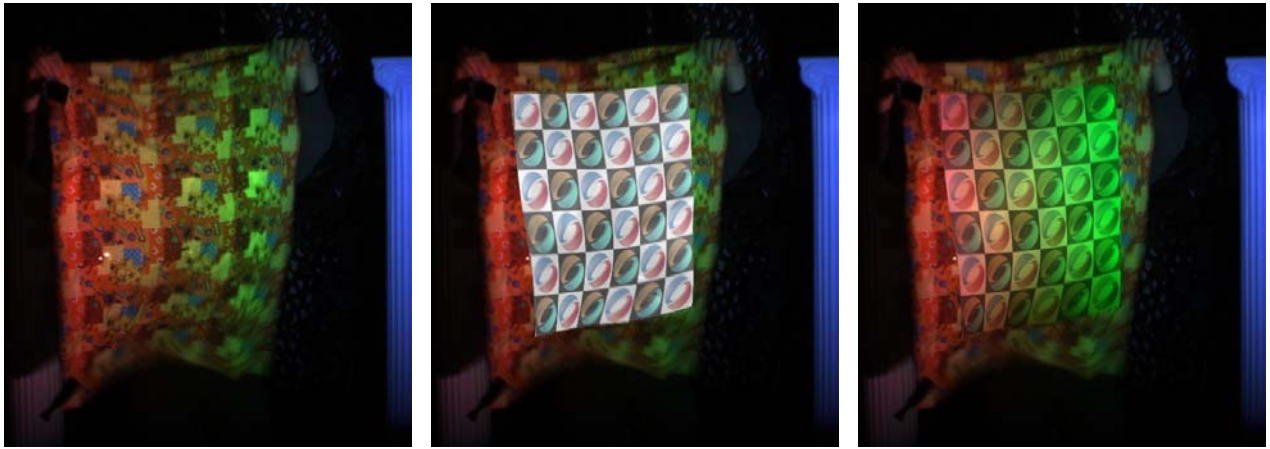
Figure 7: *Synthetic texture with matched illumination. (Left) A cloth with 12 hidden tags. (Middle) Overlaying texture using the location and orientation of the tags. (Right) Texture modulated with incident illumination.*

To perform photometric and geometric calibration, we compute the color transfer function $\Gamma(.)$ to match the color response of the camera and the RGB sensors on the tag via a color chart. We place the RGB sensor of the tag next to a color chart and capture images under varying colored illumination. Then we compute the color transfer function $\Gamma(.)$ as a 3×3 color transformation matrix.

Unlike conventional systems, the camera does not directly see the tag. Hence, we compute Euclidean calibration between the world coordinates, the projector array coordinates, and the camera coordinates. Given the projector coordinates, we triangulate to find the tag position in 3D, then use the camera perspective projection matrix to estimate the 2D pixel location.


Figure 8: *Computing intrinsic reflectance. (Left column) A color chart seen under purple and green illumination. The RGB sensor on the tag to the right senses the incident illumination spectrum. (Right column) We apply color correction to the center rectangle based on RGB sensor values.*

Validation The incident illumination as well as camera-based reflectance computation is highly sensitive to underexposure or saturation of the photosensors. The limited dynamic range of the camera is restricts us to estimating reflectance for Lambertian scene points and their smooth neighborhoods. We can not estimate reflectance for specular, transparent or surfaces patches with geometric complexity in the vicinity of the tag. Nevertheless, the RGB sensor proves effective for video manipulation, from adding virtual objects with matched lighting to changing albedo of real world objects.

jects.

To verify our method, we attempted to estimate the intrinsic reflectance of a color chart under varying illumination, orientation, and location values. We used a LCD projector to project varying illuminations and collected the illumination and appearance information by RGB sensor and camera, respectively. We divide the camera pixel values by the transformed RGB sensor values in the image neighborhood of the RGB sensor. The color-corrected rectangle shown in Figure 8, visually confirms that, after color correction, under purple as well as green illumination, the color chart appears similar. However, if the irradiance has no component in a given wavelength, reflectance recovery is erroneous.
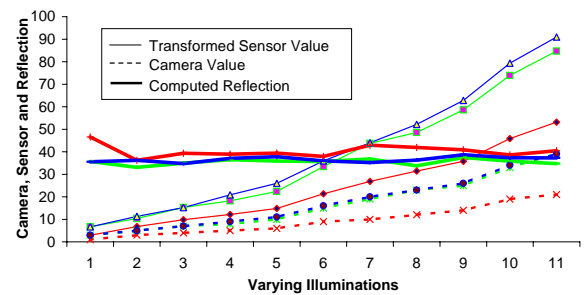
Figure 9: *Verification of reflectance estimation. The camera pixel values correspond to a gray square of the color chart illuminated under different illumination. The ratio of color pixel intensities and RGB sensor values transformed via $\Gamma(.)$ remains nearly constant. The standard deviation is under 7% of the mean value.*

In Figure 9, we show quantitative results of reflectance estimation. We plot the camera intensity values at a gray square in the color chart as well as the RGB sensor values under varying illumination. After matching via color transform $\Gamma(.)$, we plot the estimated R, G and B intensities at this gray square. The computed reflectance remains nearly constant and the standard deviation is less 6.8% of the mean value.

4 Applications

We show various video effects to demonstrate the effectiveness of our space labeling and sensing approach.

Figure 10: *Robustness under different conditions. (Left) Shooting from a moving car in daylight. (Right) Highly specular scene: Three tags are attached behind aluminum foil. We track tags behind a bottle of Coca-Cola.*

4.1 Tracking

We show high speed tracking of several objects. Figure 11 shows a tracking of a frisbee along with its rotation at 500 Hz. Note the dense trail of scene points we are able to capture.

We can robustly track the tags under various challenging conditions. Thanks to 455kHz modulation, we can track in **bright outdoors settings** even in the presence of DC IR light from the sun or in complete darkness. This is similar to the operation of IR photosensor on a TV which can receive and decode binary codes from a remote control in sunlight. Because our system is lightweight and portable, we can also track from a projector in a moving car (Figure 1 and 10). We track features on an actor running at high speeds. The actor is wearing a jacket with 8 wireless tags that are visually imperceptible. We used a recording camera rigidly attached to two projectors. We shot the video handheld from backseat of a moving car (Honda Accord). Note that without modulation and imperceptible tags, it will be difficult to collect motion-capture data outdoors where the actor appears to wear no special high contrast clothing or tags.

Camera-based systems can be challenged by **specularities**. By contrast, our system can track tags even if they are attached behind highly specular surfaces such as aluminum foil. We can also locate tags occluded by objects that are opaque in the visible spectrum but transparent in IR, such as a cola soft drink (Figure 10), fog or IR-transparent clothing.
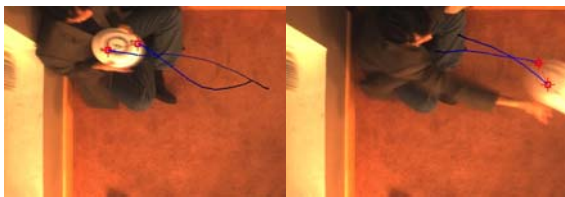

Figure 11: *Tracking two tags on a frisbee.*

4.2 Video Enhancements

By moving beyond location sensing, live or recorded videos can be significantly enhanced with captured ambient lighting. Adding virtual labels to real objects in video requires precise location data.

With additional orientation and illumination data, we can demonstrate several complex effects.

Figure 7 shows how a virtual label attached to flexing cloth can realistically match changes in position, shape and appearance of the cloth. We attached 12 tags to the cloth with only the photosensors exposed. We show a virtual checkered logo attached to this deformable cloth. Next, we project a rainbow pattern onto this cloth with a LCD projector. As the cloth travels across the projected pattern, the appearance of the virtual label changes automatically. The RGB sensors compute the irradiance and we use the estimated 3D positions and orientations to simulate different materials. The update rate for data capture is independent of the number of tags used in a scene. The attribute capture can be combined with cloth capture and simulation efforts [Bhat et al. 2003].

Our system is ideal for virtual sets where it can be used in a natural setting and the actors can wear the photosensing tags under their clothing. In addition, the projectors used are light weight and portable, and can be mounted on a video camera to acquire the camera pixel coordinates for each tag with minimal calibration.
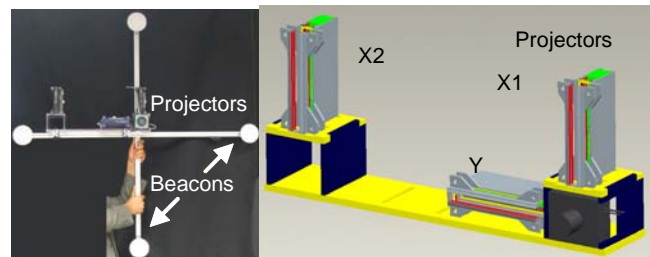
Figure 12: *Our setup with three projectors and four beacons. We use eight beacons to compute orientation in our experiments.*

Figure 13: *Rendering a synthetic sword (Left) without blur and (Right) with blur.*
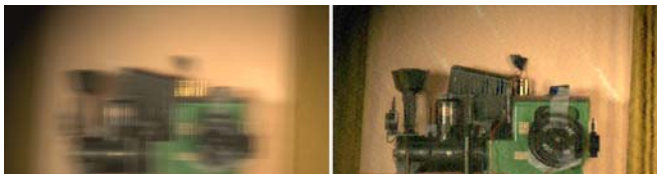
4.3 Blurring and Deblurring

The photosensors capture geometric and photometric data at a much higher speed than camera. We can use this higher temporal resolution information in various ways. We can render synthetic elements with more realistic motion and color blur. Figure 1 and 13 show a sword moving through blue light into red light. We attached three tags to the actor's sword prop and computed the point spread function (PSF) based on the geometric information as well as the incident illumination color. By combing this geometric and photometric data, we synthesized a motion and color smear for the sword so that it matches the motion of the character as well as the backdrop colors. High speed acquisition of incident illumination may allows one to recover temporally aliased scene properties.

It may be argued that this effect can be achieved using a pure video processing approach via a white-colored diffuse proxy to estimate

the incident illumination. However, this requires segmentation and possibly interactive marking per frame. In addition, authoring is required if there is more than one sword. Location and orientation needs to be estimated separately. Our system acquires ID and (image) location automatically providing the back-projected shape of the sword's 3D model in the image. We estimate the orientation allowing correct photometric effects. By contract, it is difficult to estimate the orientation of a sword from a narrow strip in a video.

We can also use the PSF for deblurring. Deblurring is an ill-posed problem but if the PSF is accurately known, some spatial frequencies of the original signal can be recovered. As shown in Figure 14, the sharp features of the moving train are clearly recovered, although there are some artifacts in the dark noisy regions of the scene.

Figure 14: *Deblurring using a precisely estimated point spread function.*
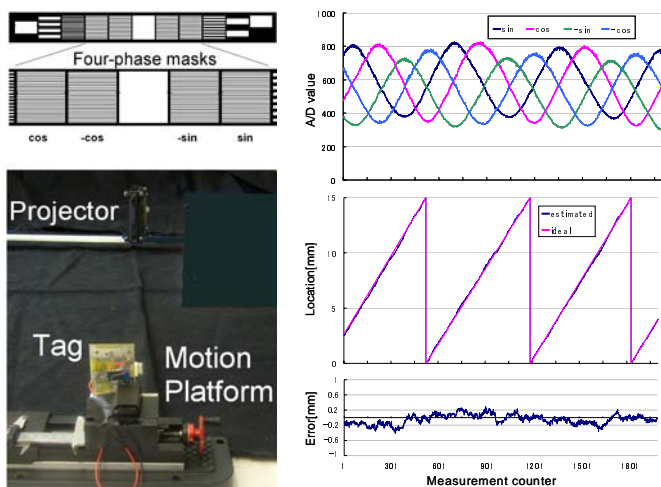
Figure 15: *High accuracy location tracking using analog masks. (Left top) Modified Gray code mask. (Left Bottom) Experimental setup to move and track the tag with a 100 micron resolution. (Right Column) Top: Intensity values recorded by the sensor due to the four phases. Middle: The computed phase between $[0, \pi]$ maps to location within a zone defined by 6 MSBs. Bottom: The error in position is computed based on linear motion assumption.*

5   Implementation

The tags use off-the-shelf IR decoding modules and PIC18 microcontrollers. Each beamer is on for just 33 microseconds with 33 microsecond delay at 455 kHz carrier. A set of 10 beamers can complete the frame in less than 800 microseconds including guard delays. Thus, temporal bandwidth for a single beamer is approximately 10,000 bits per second. One can treat this as a projector projecting 10,000 gray-coded patterns per second. A set of beamers on a single projector can project patterns at 1000 Hz. Thus the spatiotemporal labeling bandwidth is 10^3 Hz $\times 10^3$ locations $= 10^6$ bits per second. When orientation beacons are used, the additional

timeslots reduce the update rate to 124 Hz. Figure 12 shows our setup with four beacons.

The latency of our system is less than 1 millisecond. The optical packet length is 600 microseconds. The PIC microcontroller on the tag adds a delay of 100 microseconds. Our bottleneck is wireless communication; the Bluetooth wireless modules work at 78 kbits/sec but bandwidth is shared between the tags. A wired version can operate at a higher rate. The cost of our tags and projectors, including printed circuit boards, body of the projector (constructed using a 3D printer) and printed mask is in a few 10's of US dollars. The cylindrical lenses are off-the-shelf and cost about US$100 at Melles Griot. But molded lenses will be cheaper.

5.1   Performance Improvements

High Accuracy Location   Camera based systems use blob centroid analysis for achieving superresolution. Using part binary and part analog masks, we show that we can achieve 12 bit angular resolution (Figure 15). Our technique is similar to quadrature rotory encoders. We use 4 sinusoidal masks, instead of 2, which are successively out of phase by $\pi/2$. The masks in front of the first 6 most significant bits (MSB) are binary as in the previous design. They subdivide the field of view into 64 angular zones. Within each zone, we compute sub-zone precision location by using phase information from the sinusoidal masks for the remaining 4 beamers. The intensities received at a photosensor due to these four beamers are $I_i = k(sin(\phi + i * \pi/2) + q)$, $i = 0..3$, where k is the unknown gain and q is the unknown DC offset. After normalizing the raw values, we calculate the phase, ϕ, as,

$$\phi = tan^{-1}((I_0 - I_2)(I_1 + I_3)/((I_0 + I_2)(I_1 - I_3)))$$

The phase is linearly related to the location within the zone defined by the 6 MSB bits. The demonstration in the video shows that we can achieve 100 micron *x*-direction resolution (standard deviation 123 micron) at 1 meter distance from the projector, where the working volume is 50 cm wide.
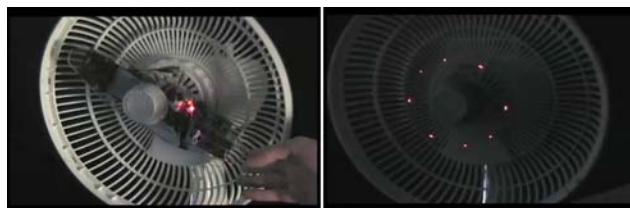

Figure 16: *IrDA based module for 20,000 locations per second tracking. The tag mounted on a fan blade blinks an LED at 8 predetermined positions.*

High Speed and Frequency Optical Communication   Our choice of 455kHz time division multiplexing is driven by availability of low cost off-the-shelf IR sensors and programmable 20 MHz PIC microcontrollers. However, our system is highly scalable. We can achieve 1D update rates of over 20,000Hz using tranceivers intended for use with IrDA communication. We used Linear technology LT1328 as IrDA receiver and Siemens BPW34F remote control photodiode as the transmitter. We tested the system at 250 kbits per second which means an effective frame rate of 20,000 frames per second for a single axis. We verify the speed by attaching the IRDA tag to a fan blade (Figure 16). An LED on the tag illuminates only when it is in front of eight predetermined locations. As the fan reaches its highest speed at 1200 revolutions per minute, these eight dots remain stable and visible at the same world location, indicating that the tag is still aware of its own location at these speeds.

For higher speeds, free space optics can achieve bandwidth of 1.25 Gbps over several meters [LightPointe Communications 2007].

With improvement in the dynamic range of photosensors, our method can be transitioned from time-division to optical amplitude, frequency or code-division multiplexing (CDMA) [Azizoglu et al. 1992] for 'always on' space labeling that does not require temporal synchronization. In an amplitude modulation optical communication, we can transmit three different modulated codes at 1MHz, 1.25 MHz and 1.66 MHz simultaneously without any synchronization. Figure 17 shows the three transmitters at the center of colored disks. The photosensor current is passed through a filter bank of 3 tuners to decode intensity from the three transmitters. As demonstrated in our video, we currently track these intensities simultaneously (shown on a scope) at 100 Hz update rate. With asynchronous multiplexed communication, a many(projectors)-to-many(receivers) distributed environment could be realized without temporal trade-off against the number of projectors or sensors.
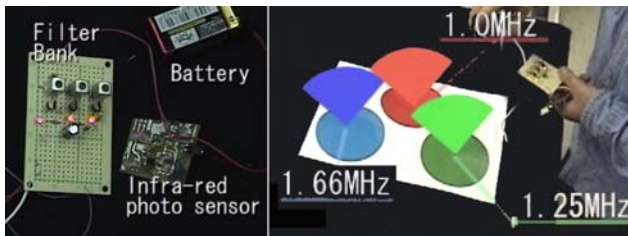
Figure 17: *Amplitude division multiplexing support three beamers to communicate with a tag simultaneously. The beamers are on the table. The plot shows signal strength received at the tag in hand.*

6 Future Directions

Photosensing Tags to Photosensing Surfaces Our setup allows sampling of incident illumination at sparse locations. But emerging technologies such as optically sensitive smart fabric can sense light over the entire surface [Bayindir et al. 2004]. Tiny radio frequency identification (RFID) transceivers tags can operate without batteries when an RF reader is within a few meters and some can sense incident light [Smith et al. 2006]. A sprinkling of such wireless transceivers embedded in light sensing fabric may create a highly versatile sampling surface. Dense photosensing could advance the art from discrete sampling to a more continuous sampling for precision research applications, realistic graphics integrations in the film industry, or eventual real time lighting applications.

Capturing Higher Dimensional Reflectance By moving a surface patch in our system, the estimation of its bidirectional reflectance distribution function (BRDF) is possible. In every camera frame, we obtain a direct estimate of scene irradiance and radiance. Earlier we assumed Lambertian reflectance, where this ratio is constant. A single fixed illumination source, a fixed camera, a fixed scene point position but varying surface orientation produces a 2D slice of the 4D reflectance function. By varying the position of the scene point, several 2D slices can be recovered. Since we know the irradiance, this probe-based procedure is more robust than a pure camera-based measurement. In addition, all the data is automatically annotated with location and ID allowing BRDF capture of a non-homogeneous reflectance surface.

Capturing Participating Media We can estimate properties of a participating media such as fog or liquids by positioning two or more sensors in the environment to measure the attenuation. Conventional systems measure only local attenuation–e.g. to compute practical visibility distance at the airports. In our case, the location computation system will work as is, and by measuring the attenuation of the same source at two or more distinct tag locations, one can find the factor of attenuation.

Illumination Manipulation Photographers and videographers typically take a lightmeter reading to estimate illumination and determine camera or flash parameters. Our method offers real time estimates of illumination at multiple locations, including correction terms such as orientation and reflectance. The intensity, color, or configuration of illuminators can be changed dynamically based on feedback from the tags.

Camera Adaptation and Factorization Optimal camera parameters can be estimated by taking readings from the tags. With several light meters wirelessly transmitting irradiance and location information, parameter settings can be made more accurately and with added functionality. For example, the focus settings, typically estimated from previous frames are incorrect in a dynamic scene. Since our tags are fixed in world coordinates rather than camera coordinates, they can provide appropriate data even before the camera views a point in any frame. Reflectance-illumination factorization of camera images holds promise for computer and machine vision.

Spatial Encoding When the projected patterns have one dimensional symmetry, our design creates a single logical projector with a shared center of projection from multiple non-colocated beamers. We built a low-cost high-frame rate projector capable of projecting 10,000 passive patterns per second. Such patterns are useful in structured light projection for 3D shape scanning, among other applications. We have chosen to exploit a geometric constraint on the projected gray-code patterns and simplify the position decoding process on the tag. This optimal coding uses the minimum number of bits to uniquely label the volumes in space. But it is possible to label the space using an **arbitrary arrangement** of individual LED beamers with masks corresponding to a random bit, trading off sub-optimality and flexibility in encoding. Given a perspective projection matrix of the N LED-beamers along with the pattern, one can find the space label by backprojecting 1's and 0's into the working volume and finding their intersection. An exciting area of future research is in calibration algorithms for perspective projection parameters (and non-linear distortions) of arbitrarily scattered LED-mask beamers. The current tag is a point-sampling photosensor, but one can imagine upgrading to a 1D or 2D optical sensor such as a PSD or a low-resolution camera to capture angular variation in incident light.

7 Conclusion

We have described a method for spatio-temporal coded projection of light to label 2D or 3D space. Our use of passive binary spatial masks in a strategic configuration that exploits epipolar constraints results in a receiver that is effective and yet simple. We note that in light based communication, the optical bandwidth, which is a product of temporal and spatial bandwidth, has been increasing annually by a factor of 10 [Silicon Light Machines 2006]. We have presented a scheme in which the modern solid-state light source can be employed to estimate useful parameters related to points in the environment – both geometric (location, orientation) as well as photometric (incident intensity, incident spectrum, surface reflectance).

Motion capture to date has been limited to exclusive and special purpose environments. Our low cost, easily portable, and visually imperceptible tracking system makes motion capture practical in

a much wider application area. Interwoven tags on can help analyze patient rehabilitation progress after injuries. Tracking may be supportable in home video games and other casual virtual and augmented reality interfaces. The dream of filmmakers and game developers is 'on-set motion capture'. One of the recent examples is the motion and appearance capture in the movie 'Pirates of the Caribbean' [Robertson 2006]. Our system can support this operation with unlimited number of imperceptible and interactive tags.

We believe our research can lead to several new areas of exploration. These include a deeper analysis of the projective geometry of randomly dispersed individual LED emitters, the measurement of BRDFs using scene probes, video enhancement via data streams faster than camera frame-rates, and more advanced algorithms for reflectance-illumination factorization. We hope that promising technologies such as light sensing fabrics for dense sampling, radio frequency identification transceivers for battery-less tags, and optical CDMA for uninterrupted space labeling will further stimulate techniques based on scene point photosensing.

References

AZIZOGLU, M. Y., SALEHI, J. A., AND LI, Y. 1992. Optical CDMA via Temporal Codes. *IEEE Transactions on Communications 40*, 7, 1162–1170.

BASRI, R., AND JACOBS, D. 2001. Photometric Stereo with General, Unknown Lighting. *IEEE CVPR 02*, 374–379.

BAYINDIR, M., SORIN, F., ABOURADDY, A. F., VIENS, J., HART, S. D., JOANNOPOULOS, J. D., AND FINK, Y. 2004. Metal-insulator-semiconductor Optoelectronic Fibres. *Nature 431, 826-829.*

BHAT, K. S., TWIGG, C. D., HODGINS, J. K., KHOSLA, P. K., POPOVIC, Z., AND SEITZ, S. M. 2003. Estimating Cloth Simulation Parameters from Video. In *Symposium on Computer animation (SCA)*, 37–51.

BOLAS, M. T., MCDOWALL, I. E., HOBERMAN, P., AND FISHER, S. S., 2004. Snared Illumination.

CODAMOTION, 2007. Charnwood Dynamics Ltd. http://www.charndyn.com/.

COTTING, D., NAEF, M., GROSS, M., AND FUCHS, H. 2004. Embedding imperceptible patterns into projected images for simultaneous acquisition and display. In *International Symposium on Mixed and Augmented Reality (ISMAR'04)*, 100–109.

HAMAMATSU PHOTONICS, 2005. Si PIN Photodiode S6560.

HIGHTOWER, J., AND BORRIELLO, G. 2001. Location Systems for Ubiquitous Computing. *Computer 34*, 8, 57–66.

ILTANEN, M., KOSOLA, H., PALOVUORI, K., AND VANHALA, J. 1998. Optical Positioning and Tracking System for a Head Mounted Display Based on Spread Spectrum Technology. In *2nd International Conference on Machine Automation (ICMA)*, 597–608.

KANG, S., AND TESAR, D. 2004. Indoor GPS Metrology System with 3D Probe for Precision Applications. In *Proceedings of ASME IMECE 2004 International Mechanical Engineering Congress and RD&D Expo*.

LEE, J. C., HUDSON, S. E., SUMMET, J. W., AND DIETZ, P. H. 2005. Moveable Interactive Projected Displays using Projector-based Tracking. In *ACM symposium on User interface software and technology (UIST)*, 63–72.

LIGHTPOINTE COMMUNICATIONS, 2007. Free Space Optics. http://www.freespaceoptics.org/.

MA, H., AND PARADISO, J. A. 2002. The FindIT Flashlight: Responsive Tagging Based on Optically Triggered Microprocessor Wakeup. In *Ubicomp*, 160–167.

MATSUSHITA, N., HIHARA, D., USHIRO, T., YOSHIMURA, S., REKIMOTO, J., AND YAMAMOTO, Y. 2003. ID CAM: A Smart Camera for Scene Capturing and ID Recognition. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, 227–234.

MOTION ANALYSIS CORPORATION, 2006. Hawk-I Digital System.

NII, H., SUGIMOTO, M., AND INAMI, M. 2005. Smart Light Ultra High Speed Projector for Spatial Multiplexing Optical Transmission. *Procams Workshop (held with IEEE CVPR)*.

NII, H., SUMMET, J., ZHAO, Y., WESTHUES, J., DIETZ, P., NAYAR, S., BARNWELL, J., NOLAND, M., BRANZOI, V., BRUNS, E., INAMI, M., AND RASKAR, R. 2006. Instant replay using high speed motion capture and projected overlay. In *ACM SIGGRAPH Emerging Technologies*, 111.

OPTOTRAK, 2007. NDI Optotrak Certus Spatial Measurement. http://www.ndigital.com/certus.php.

PALOVUORI, K., J., V., AND M., K. 2001. Shadowtrack: A Novel Tracking System Based on Spreadspectrum Spatio-Temporal Illumination. *Presence - Teleoperators and Virtual Environments. 9:6* (December), 581–592.

PATEL, S. N., AND ABOWD, G. D. 2003. A 2-Way Laser-Assisted Selection Scheme for Handhelds in a Physical Environment. In *Ubicomp*, 200–207.

PHASE SPACE INC, 2007. Impulse Camera. http://www.phasespace.com.

PTI INC, 2006. VisualEyez VZ 4000.

RASKAR, R., BEARDSLEY, P., VAN BAAR, J., WANG, Y., DIETZ, P., LEE, J., LEIGH, D., AND WILLWACHER, T. 2004. RFIG Lamps: Interacting with a Self-describing World via Photosensing Wireless Tags and Projectors. *ACM Transactions on Graphics (SIGGRAPH) 23*, 3 (Aug.), 406–415.

RINGWALD, M. 2002. Spontaneous Interaction with Everyday Devices Using a PDA Workshop on Supporting Spontaneous Interaction in Ubiquitous Computing Settings. In *UbiComp*.

ROBERTSON, B. 2006. Big moves. *Computer Graphics World 29*, 11 (Nov).

SILICON LIGHT MACHINES, 2006. Gated Light Valve. http://www.siliconlight.com.

SMITH, J., SAMPLE, A., POWLEDGE, P., ROY, S., AND MAMISHEV, A. 2006. A wirelessly-powered platform for sensing and computation. In *Ubicomp*.

SORENSEN, B. R., DONATH, M., YANG, G.-B., AND STARR, R. C. 1989. The Minnesota Scanner: A Prototype Sensor for Three-dimensional Tracking of Moving Body Segments. *IEEE Transactions on Robotics and Automation 45*, 4, 499–509.

VICONPEAK, 2006. Camera MX 40. http://www.vicon.com/products/mx40.html.

WELCH, G., AND BISHOP, G. 1997. SCAAT: Incremental Tracking with Incomplete Information. In *Proceedings of SIGGRAPH 97*, Computer Graphics Proceedings, Annual Conference Series, 333–344.

WELCH, G., AND FOXLIN, E. 2002. Motion Tracking: No Silver Bullet, but a Respectable Arsenal. *IEEE Comput. Graph. Appl. 22*, 6, 24–38.