Eurographics Symposium on Rendering (2006)
Tomas Akenine-Möller and Wolfgang Heidrich (Editors)

Relighting Human Locomotion with Flowed Reflectance Fields

Per Einarsson Charles-Felix Chabert Andrew Jones Wan-Chun Ma† Bruce Lamond
Tim Hawkins Mark Bolas‡ Sebastian Sylwan Paul Debevec

University of Southern California Centers for Creative Technologies
National Taiwan University†
University of Southern California School of Cinema-Television‡

Abstract

We present an image-based approach for capturing the appearance of a walking or running person so they can be rendered realistically under variable viewpoint and illumination. In our approach, a person walks on a treadmill at a regular rate as a turntable slowly rotates the person's direction. As this happens, the person is filmed with a vertical array of high-speed cameras under a time-multiplexed lighting basis, acquiring a seven-dimensional dataset of the person under variable time, illumination, and viewing direction in approximately forty seconds. We process this data into a flowed reflectance field using an optical flow algorithm to correspond pixels in neighboring camera views and time samples to each other, and we use image compression to reduce the size of this data. We then use image-based relighting and a hardware-accelerated combination of view morphing and light field rendering to render the subject under user-specified viewpoint and lighting conditions. To composite the person into a scene, we use an alpha channel derived from back lighting and a retroreflective treadmill surface and a visual hull process to render the shadows the person would cast onto the ground. We demonstrate realistic composites of several subjects into real and virtual environments using our technique.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism -- Color, shading, shadowing, and texture I.4.8 [Image Processing and Computer Vision]: Digitization and Image Capture -- Reflectance I.4.8 [Image Processing and Computer Vision]: Scene Analysis -- Tracking

1. Introduction

Many computer graphics applications involve combining photographic imagery of people with rendered 3D environments or background imagery. When photorealistic results are desired, the challenge is to make the foreground subject appear to be photographed from the same viewpoint, by the same camera, lit with illumination consistent with the desired environment. While photographic images are inherently realistic and efficient to acquire, they are fixed to a specific viewpoint and lighting. As a result, realistic composites of live-action performances into new environments are typically achieved only through careful planning, on-set documentation, and postproduction manipulation during compositing. While this process can be effective for producing realistic linear content, the possibilities for postproduction control of viewpoint and lighting are limited. Furthermore, the cost of filming on location is high, and the process is not easily applicable to interactive 3D content, where the viewpoint and lighting is determined in real time. As a result, photographic elements are rarely used when significant control of the viewpoint and illumination is required. Instead, laboriously modeled, textured, animated, and rendered 3D models are used in their place.

Considerable work has addressed aspects of postproduction control of viewpoint and illumination, though most proposed systems address only control of either the viewpoint or the illumination. Viewpoint control has been achieved by filming the subject from a relatively sparse array of cameras (e.g. [RNK97, MKKJ96, MBR*00, CTMS03, VBK05]) and projecting images from these viewpoints onto approx-

imate geometric models derived from or fit to the video. Such systems provide wide control of the virtual viewpoint but are prone to artifacts resulting from image misregistration and sparse sampling of the subject's directional reflectance. Other systems use relatively dense camera arrays (e.g. [YEBM02, WJV*05]) and light field rendering [LH96] to produce novel views of the subject. Such systems avoid the need for explicit scene geometry and achieve high-quality novel views, but for a restricted range of viewpoints corresponding to the size of the camera array. So far, both approaches only show the subject under original illumination conditions. Thus, it is not straightforward to realistically place the subject into an environment with novel illumination.

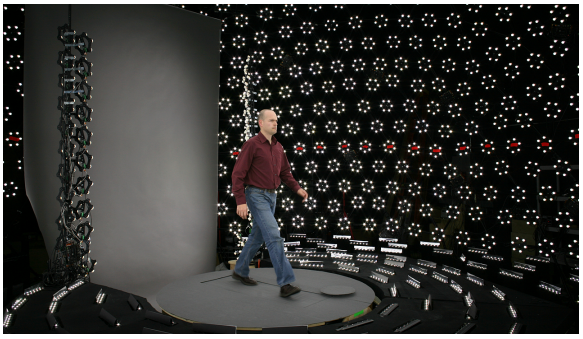


Figure 1: *The treadmill, turntable, and lighting apparatus used for capturing human locomotion from multiple viewpoints under time-multiplexed lighting. Behind the person is the background matte; the vertical array of high-speed cameras is out of frame to the right.*

Postproduction control over illumination has also been addressed in recent work. [WGT*05] uses a sphere of LED light sources to capture a person's performance in approximately 100 lighting conditions every 24th of a second using a high-speed camera, and image-based relighting to render the subject under new illumination environments. While this work produced realistic results, it did not address postproduction control over the viewpoint on the subject and limited the view to the subject's head and shoulders. Notably, [TAdA*05] addresses control over both viewpoint and illumination of a human performance using a sparse array of cameras, model fitting, and reflectometry. However, results shown from the technique are less realistic than real video of the subject owing to the difficulty of performing surface reflectance measurement with only one available lighting condition, imperfect scene geometry, and a sparse sampling of viewpoints.

1.1. Our Contributions

In this paper, we take a step toward an image-based approach to obtaining postproduction control over both viewpoint and illumination of *cyclic* full-body human motion by combining the performance relighting technique of [WGT*05] with a novel view generation technique based on a *flowed reflectance field*. We advance beyond [WGT*05] by capturing the whole human body using a novel lighting apparatus designed to simulate both a distant lighting environment and local illumination from a ground plane. By restricting our consideration to cyclic motion such as walking and running, we are able to acquire a two-dimensional array of views by slowly rotating the subject in front of a one-dimensional array of cameras. We thus obtain a set of views sparser than typical light field rendering approaches but denser than typical model-based approaches. We compute optical flow between neighboring viewpoints and use a combination of view interpolation [CW93, SD96, ZKU*04] and light field rendering [LH96, GGSC96] to generate views of the subject from novel 3D positions, generalizing the technique of [WJV*05] which used view interpolation to smoothly move the viewpoint within the plane of the camera array. We obtain a matte for the subject using both backlighting and retroreflection, and we use a visual hull process to simulate the shadows the subject would cast in a complex lighting environment. To demonstrate our technique, we show motions captured for different subjects realistically composited into both synthetic and real 3D environments.

2. Background and Related Work

Our technique builds on a variety of image-based modeling and rendering techniques in the following principal areas:

2.1. View Interpolation and Light Field Rendering

Significant work has described techniques to generate novel camera positions from previously captured or rendered images. [CW93] warps rendered images using depth maps to generate novel views while [LF94, MB95] use stereo correspondence to compute depth for altering the viewpoint of real scenes. [SD96] presents a view morphing technique for synthesizing correct perspective for novel viewpoints between corresponded original views; light field rendering techniques [LH96, GGSC96] directly synthesize views of a scene from a new 3D viewpoint by sampling rays from a 2D array of densely spaced viewpoints. Visual fidelity can be improved by projecting image samples onto scene geometry [GGSC96] or explicitly forming a *surface light field* [MRP98].

2.2. Dynamic Light Field Acquisition

Several works have constructed 2D arrays of cameras to capture light fields of dynamic events. Among them, [YEBM02] used an array of cameras and distributed rendering to allow multiple viewers to observe virtual views in real time; [YMG02] extended this work to *surface cameras* that allow the light field to focus on non-planar geometry. [ZC04] also

used depth information to focus a real-time light field from self-reconfigurable camera array. [WJV*05] used video from array of cameras to perform *spatiotemporal* view interpolation: generating novel views from intermediate camera positions and points in time by computing optical flow between views staggered in both time and space.

Building on and extending such techniques, our work acquires a moderately sampled 2D array of images surrounding the subject and combines light field rendering with view morphing from optical flow to generate focused views of the subject from new 3D viewpoints, both in front of and behind the original viewing surface. We make use only of the optical flow maps rather than explicitly reconstructed scene geometry, which allows our technique to work effectively with imperfectly repeating motion cycles for which no globally consistent geometric model exists. [BBM*01] captures a "motion lumigraph", wherein a time-varying light field is acquired across multiple cycles of repeating subject motion: in their case, a toy helicopter with a spinning rotor. In our work, we use this approach to acquire a time-varying reflectance field by acquiring multiple cycles of human locomotion with a rotating viewpoint. In this way, we construct a 2D array of viewpoints of approximately the same subject motion using just a 1D array of cameras.

2.3. Image-Based Relighting

Simulating novel illumination by forming linear combinations of images with different basis lighting conditions has been used in the context of rendered images [DAG95, NSD94] and human faces [DHT*00]. Such relightable data can be mapped onto traditional 3D models [SKS02, RH02] for real-time rendering. [DHT*00] describes a *non-local reflectance field* as a 6D function $R = R(\theta_i, \phi_i; u_r, v_r, \theta_r, \phi_r)$ as the space of radiant light fields $R_r(u_r, v_r, \theta_r, \phi_r)$ that result from illuminating a subject from the set of distant lighting directions (θ_i, ϕ_i). Sampled datasets of such 6D functions have been used to render arbitrary viewpoints and illumination of virtual objects such as trees [MNP01], real objects [MPN*02,MPZ*02], and faces [HWTD04]. However, unlike our work, such techniques have not addressed the problem of dynamic scenes or simulating the photometric interaction an object would have with its environment. [WGT*05] captures relightable datasets of human facial performances but does not change the viewpoint.

2.4. Free-Viewpoint Video

Several authors have presented techniques for novel view generation for a live-action performance by filming with a sparse array of cameras and mapping these images onto basic geometry of the subject using stereo correspondence, [RNK97], silhouette intersection [MKKJ96], image-based visual hulls [MBR*00], or fitting a human surface model to image silhouettes [CTMS03]. The central challenge is that differences between the recovered and modeled scene geometry can cause visible texture misalignments, the appearance of which can be reduced to some extent using view-dependent texture mapping [DTM96]. [ZKU*04] produces extremely high-quality virtual views of dynamic scenes using a layered representation for stereo correspondence with high-quality alpha channels, but limits its investigation to motion along a 1D array of cameras. [VBK02,VBK05] compute *scene flow* for a human performance by deriving scene geometry from a relatively sparse camera array and computing the movement of surface points in 3D with respect to the geometry; this allows rendering at novel time instants as well as from novel camera viewpoints. [Mag05] provides overviews of a wide variety of such video-based rendering techniques. To date, few such techniques have addressed changing the illumination, with the exception of [TAdA*05] (described earlier); relighting is important for realistically compositing a captured performance into a new environment. Our work takes a different approach to the problem in that we use time-multiplexed lighting to provide richer information to the relighting process and a flowed light field view interpolation approach that avoids the need for deriving an accurate, unified model of scene geometry. However, our technique's need for more viewpoints and higher frame rates limits our investigation to short sequences of repetitive motion such as walking and running. We use collections of viewpoints similar to those rendered for synthetic avatars in [TC00], but we capture this data for real subjects, use view morphing and light field rendering to increase the quality and generality of novel viewpoint generation, and we are able to relight the subject.

3. Apparatus

Our acquisition setup (Figs. 1 & 2) is designed to efficiently capture 2D images of a moving person sampled across time (1D), viewpoint (2D), and illumination (2D), yielding a 7D dataset. At the center of the setup, the subject walks on a treadmill placed on a rotating turntable. The treadmill belt and turntable top are covered with Reflecmedia Chromatte material used in the matting process. Shallow channels are cut into the board beneath the treadmill belt to give the subject a tactile reference for staying centered and performing repeatable motion. When recording, an operator monitors the speed and cutoff switches for both the turntable and treadmill, and a 125cm-wide area around the turntable is covered with 20cm of gymnastic foam for additional safety. We use a general-purpose lighting apparatus similar to that described in [WGT*05] to produce our illumination basis. The device is the top two thirds of an 8m 6th-frequency geodesic sphere with edge lengths optimized to obtain an even distribution of 901 controllable light sources. Each light consists of six LumiLEDs Luxeon V LEDs arranged in an 18cm-diameter hexagon to better span the incident illumination space. Each LED uses a Fraen "single wide" optic allowing each light to deliver 100 lux to the center of the

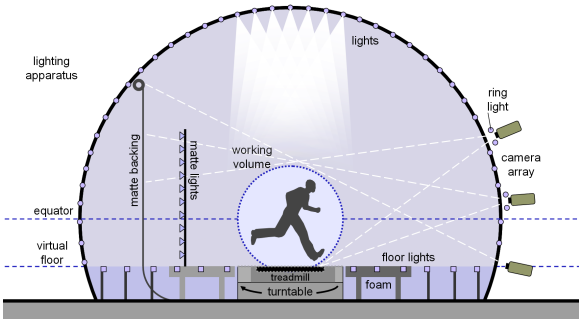


Figure 2: *A side view of the apparatus showing the treadmill, turntable, dome lights, floor lights, and camera array.*

device 4m away; the working illuminated area is designed to be a 2m sphere centered within the device. Each lighting basis condition used in this work comprises an average of 40 lights, allowing well-exposed images to be acquired at 990fps at an aperture of f/2.8. The lights are controlled by 75 microcontroller boards based on Microchip's PIC 18F8627 running at 40MHz. A master controller sends a global sync pulse to drive the lighting sequence and trigger the high-speed cameras; it also emits an audible metronome to indicate to the subject the desired walk-cycle rhythm.

Departing from [WGT*05], our apparatus includes 140 evenly spaced floor light units, also consisting of six LEDs each, at the height of the turntable 85cm below the equator to simulate illumination from a Lambertian ground plane beneath the subject. These lights omit the Fraen optics to maintain their original Lambertian light distribution; they thus mimic the illumination from a local ground plane in that they deliver negligible light to the subject's feet and increasingly greater light to the subject's midsection and head. Small vertical mirrors behind each LED increase the light cast toward the subject while maintaining the effective Lambertian distribution. The intensities of the dome and floor lights are calibrated by acquiring lighting basis images of a 30cm, 33% gray sphere at a sampling of positions above the turntable.

We photograph the subject using a vertical array of three Vision Research Phantom v7.1 high-speed cameras placed just outside the lighting structure: the lowest positioned at the level of the virtual floor, the other two at 17° and 34° above the floor relative to the subject (Fig. 2). The top two cameras are fitted with ring lights of six LEDs with Fraen single narrow optics to illuminate the retroreflective material on the turntable for matting the feet and legs of the subject (Sec. 5.1). We form a matte for the subject's body using a 3m × 4m sheet of 18% gray background paper behind the subject, illuminated by two stands of 22 additional light sources placed to the sides of the camera view.

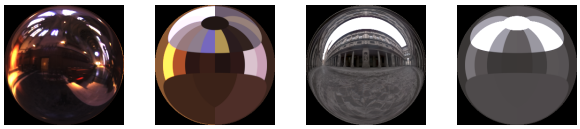
	walking	running
vertically spaced cameras	3	3
locomotion cycles per 360° rotation	36	36
frames in lighting sequence	33	33
basis conditions in sequence	26	26
sequences per second	30	30
sequences per locomotion cycle	32	25
seconds per locomotion cycle	1.07	0.78
pixels recorded per frame	320×448	320×448

Table 1: *Temporal, spatial, and angular sampling parameters.*

4. Acquisition

To capture a subject, we first measure his or her natural walking (or running) speed and cycle time and then set the treadmill and turntable speeds accordingly to acquire $n = 36$ motion cycles in 360° of rotation. We activate the time-multiplexed illumination conditions and, once the subject is walking comfortably at their measured rate (indicated by the audible beep from the controller), we enable the cameras to begin recording.

We record the subject with a sequence of illumination conditions repeating at 30 Hz. We use a 33-frame lighting sequence with 26 basis lighting conditions, three evenly-spaced tracking frames, three corresponding matte frames (Fig. 3, above), and a stripe pattern frame (not used at present). The lighting, tracking, and matting frames are similar in form and function to those in [WGT*05]. The frequency of the basis is chosen for straightforward rendering to 30fps video, and the relatively small number of lighting conditions was chosen as a tradeoff to be able to capture 36 walk cycles at 320×448 pixels within the 8GB memory capacity of the cameras. Higher-resolution lighting bases (or images) could be captured by trading off the number of cycles filmed, which would necessitate capturing different angular ranges across multiple sessions. The $33 \times 30 = 990$ frames per second we capture at 320×448 resolution is well below the cameras' maximum capability of recording 800×600 resolution at 4800 fps. Capture settings for our walking and running subjects are summarized in Table 1.



(a) (b) (c) (d)

Figure 4: *Two lighting environments (a,c) and their projections (b,d) onto the 26-element lighting basis.*

Our 26-element lighting basis (Fig. 4) is chosen to maximize its effectiveness for representing real-world lighting environments in a small number of lighting conditions and

Figure 3: *(a) Lighting sequence images taken in* $1/30^{th}$ *sec. with the middle camera. The 33-frame sequence contains 26 basis lighting conditions as well as three tracking frames T (center column) and matting frames M (to their right), and a stripe frame. (b) The* 36×3 *array of viewpoints for one instant in the recorded set of walk cycles.*

to be largely symmetrical with respect to the vertical axis (our rendering technique requires rotating the lighting basis according to the person's angle with respect to the virtual viewpoint). Finer angular resolution is devoted to the upper hemisphere of lighting directions versus the light from the ground plane, which is divided into just three regions (front, left, and right). As in [WGT*05], the lighting sequence is ordered to minimize the appearance of strobing, and subjects are screened for history of intolerance to strobing lights.

Capture times for our subjects were 38.4 seconds for walking or 30 seconds for running. During and after each capture, the subject's motion is monitored for consistency using an infrared mini-DV camera looking down at the subject from the top of the apparatus. If the subject's motion is notably inconsistent, the capture session is repeated. The 36 locomotion cycles recorded by the three cameras yield 108 relightable walk cycles from different viewpoints; a tracking frame from each such viewpoint is shown in Fig. 3(b).

At the end of a recording session, a *clean plate* image sequence of the treadmill and turntable under the lighting sequence without the actor present is acquired. Finally, geometric calibration data is acquired using a human-sized calibration checkerboard for use with the technique of [Zha00]; photometric calibration data is acquired by photographing a MacBeth ColorChecker chart with all three cameras.

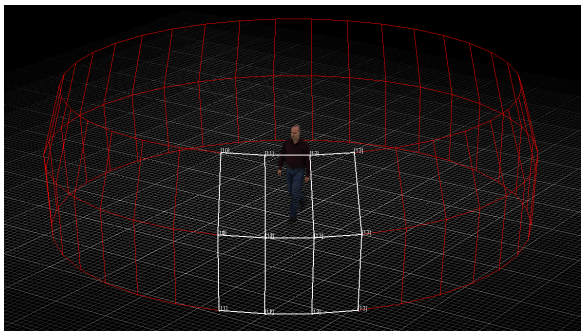


Figure 5: *The cylindrical array of viewpoints shown in relation to a virtual rendering of the subject in a real-time rendering system.*

5. Generating the Flowed Reflectance Field

Once data for a subject is captured, we derive alpha channels for the images and compute optical flow to spatially and temporally register the dataset. Finally, we compress the data into the flowed reflectance field.

5.1. Generating Mattes

We begin by generating an alpha channel [PD84], or matte, for each of the tracking frames T_0, T_1, T_2 in each lighting sequence (Fig. 3(a)). The alpha channels are derived from the

matte frames following each tracking frame. In the matte frames, the main lights used to generate the lighting basis images are turned off, and special matte lights are turned on to light the gray background behind the subject and the retroreflective material on the turntable (Fig. 6(b)). To form these mattes, we compare each tracking frame T (Fig. 6(a)) to its consecutive matting frame M and infer the foreground element pixels to be those whose monochrome brightnesses are greater in the diffusely lit tracking frame than in the back-lit matting frame. We then reduce stray matte elements by eliminating foreground regions not part of the central connected matte component and apply a 1-pixel Gaussian blur to model the filtering introduced by the image sensing and color interpolation processes to yield the final matte α (Fig. 6(c)). We then matte each tracking frame T onto a black background using the clean plate image C and the alpha channel α, forming the matted tracking frame $T' = T - (1-\alpha)C$. Finally, we zero pixel values near black to reduce the effect of camera noise.

T M α T'

Figure 6: *The matting process, showing tracking frame T, T's consecutive matte frame M, alpha matte image α, and T matted onto a black background and color-corrected, forming T'.*

5.2. Lighting Basis Registration

With the images matted, we temporally register the images in each lighting sequence using a process similar to that in [WGT*05] (the technique is also similar to how optical flow is used to register differently exposed images in an HDR video sequence in [KUWS03]). We wish to warp all 33 images so that they are temporally aligned with the tracking frame in the middle of the sequence (Fig. 3(a)); this will yield sharp rather than blurred images in the image-based relighting process. We compute optical flow vectors using the algorithm from [BA93] from the middle tracking frame T'_1 to the other tracking frames T'_0 and T'_2 using matted tracking frames to reduce image clutter and maximize the robustness of the flow process. We then interpolate each flow to warp each frame to T_1 using a reverse pixel lookup. Since some of the basis lighting images occur before the first or after the last tracking frame, we must mildly extrapolate the flow for these images. However, since no image is more than 1/60th sec from the central tracking frame, the flow maps are generally very accurate. At this point, the alpha channel for the middle tracking frame can be used as the alpha channel for every frame in the lighting basis. Using α, we matte each of the warped basis images onto black using their corresponding clean plate images.

5.3. Flow Between Viewpoints

For each frame of the locomotion cycle, we have a 36×3 grid (Fig. 3(b)) of 4D reflectance fields $R_{u,v}(\mathbf{s})$ where (u,v) is the horizontal and vertical viewpoint index and $\mathbf{s}$ is the 2D image coordinate in that view. To create the flowed reflectance field, we compute flow fields between each viewpoint and its 4-neighbors; images on the top and bottom row have only three flow fields. We denote these flow fields as $F^{\rightarrow}_{u,v}(\mathbf{s})$ where the arrow indicates the direction of the image toward which the flow has been computed. We store these flow fields relative to pixel coordinate $\mathbf{s}$ so that $\mathbf{s}$ in reflectance field $R_{u,v}$ corresponds to pixel coordinate $(\mathbf{s} + F^{\rightarrow}(u,v,\mathbf{s}))$ in reflectance field $R_{u^+,v}$. Thus, if there is no motion between two images, the flow field is zero.

We first compute bidirectional flow between each pair of neighboring vertical viewpoints $R_{u,v}$ and R_{u,v^+}. Since these views are acquired with differently aimed cameras, we first project the images R_{u,v^+} onto the frontoparallel plane H through the origin as viewed by the camera of $R_{u,v}$ to produce a warped field R'_{u,v^+}. This step minimizes the parallax between the images to aid optical flow as suggested in [Saw94]. In this work, we use just one correspondingly illuminated image from $R_{u,v}$ and R_{u,v^+} to compute flow; we choose a lighting direction from the front upper right to maximize the image texture due to shading while minimizing textureless shadowed areas. We then compute the corresponding pixel coordinate $\mathbf{t}'$ in R'_{u,v^+} for each pixel $R_{u,v}(\mathbf{s})$ and project this coordinate through the inverse of the homography to obtain the corresponding pixel $R_{u,v^+}\mathbf{t}$ to $R_{u,v^+}(\mathbf{s})$. We finally set $F^{\uparrow}_{u,v} = \mathbf{t} - \mathbf{s}$. We then reverse the roles of the images to compute $F^{\downarrow}_{u,v^+}$.

Next, we compute bidirectional flow between each pair of neighboring horizontal viewpoints. Since the horizontal viewpoints are captured from consecutive walk cycles and thus do not capture the subject in precisely the same position, we widen the search space in computing these flow maps. However, since the images are taken from nearby viewpoints with the same camera, we omitted using the homography to rectify these images to each other. The horizontal flow computations yield $F^{\leftarrow}_{u,v}$ and $F^{\rightarrow}_{u,v}$ for each $R_{u,v}$. Fig. 7 shows a visualization of such a flowed reflectance field.

5.4. Computing Shadows

To realistically composite the subject into an environment, we model the shadows the subject would cast on the surrounding environment. Our technique provides only a first-order approximation of the photometric interaction between the subject and the environment: indirect light from the subject is not cast onto the environment, and the shadows cast

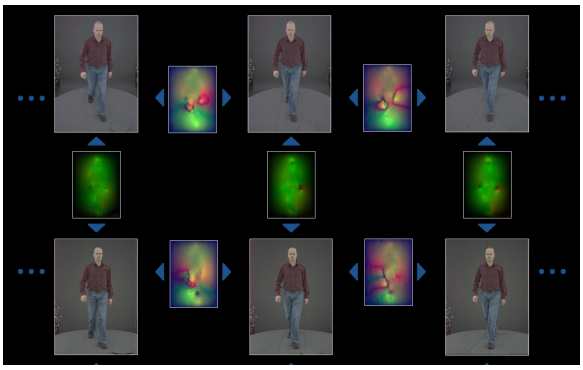


Figure 7: *Bidirectional optical flow maps are shown as computed between neighboring viewpoints in the dataset. Up/down displacement is visualized in green and left/right in red. Only one of the two flow maps is shown for each pair of views.*

on the environment do not reduce the illumination reaching the subject. Nonetheless, these approximate shadows significantly improve the realism of the composites.

To compute shadows, we model a 3m square below the subject with 128×128 pixels, and we form the visual hull of the subject from their alpha mattes using volumetric intersection [Sze93]. We trace rays from each vertex on the ground toward the light sources to determine the percentage of light from each basis lighting direction that remains unoccluded by the subject. The result is a set of 23 shadow maps for each frame of the walk cycle, one for each of the lighting bases above the floor (Fig. 8). These shadow maps are re-lit during the image-based relighting process described in Sec. 6.1.

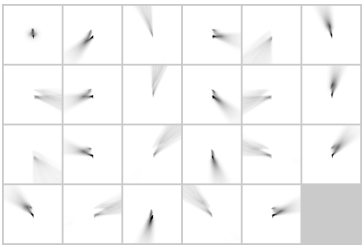


Figure 8: *Shadow maps computed for one frame in a subject's walk cycle for the 23 basis lighting conditions above the ground plane.*

5.5. Compression

For each data capture we record a total of 24 GB, 12-bit raw pixel data. Our generation of mattes, flow maps, shadow maps, and lighting basis registration takes approximately 6 hours using 5 computers with 3.2 GHz Xeon processors. In this process we compress the resulting data to approximately 1.5 GB of reflectance field data and 1 GB of flow maps.

For the flow maps, we quantize the pixel displacements to 8 bits using a signed logarithmic scale that is able to represent $1/4$ pixel flows up to 16 pixels, $1/2$ pixel flows up to 32 pixels, and integral pixel flows up to 64 pixels. These quantized maps are further compressed using Huffman coding, reducing the quantized maps to 34% of their size.

We implemented two techniques for compressing the reflectance field. In both, we compress the *images* of the reflectance field, rather than their reflectance functions, since our lighting resolution is relatively coarse. (A survey of reflectance function compression techniques is given in [MPDW04].) The first technique is a straightforward use of traditional JPEG compression to leverage fast integer DCT and IDCT algorithms. Although the DCT is orthogonal, one can only relight images before decompression if they are stored linearly, which cannot be done with high visual fidelity in 8 bits. We thus compress mosaics of the lighting basis images in gamma 2.2-corrected space and combine the images according to the relighting coefficients with floating point precision after decompressing and applying gamma. The process produced 16:1 compression ratio with only slight image blurring. Results shown in this paper use this JPEG-based technique.

In the second technique, each of the lighting basis images in the reflectance field are compressed with Daubechies D4 wavelet compression using a variant of JPEG 2000. We chose this orthogonal wavelet, rather than the standard JPEG 2000 biorthogonal wavelet, to allow relighting the reflectance field before inverting the wavelet transform. While this saves some computational expense, it necessitates encoding linear-response pixel values, which requires 16 bits of precision and thus somewhat lower compression ratios. Another disadvantage compared to traditional JPEG was the high computational overhead for JPEG 2000's arithmetic coding operation. Nonetheless, at a 16:1 ratio the image quality was somewhat higher than traditional JPEG, recommends continued investigation into an efficient wavelet-based compression scheme.

6. Rendering

Our rendering process consists of five steps: lighting, image warping, light field interpolation, shadow rendering, and compositing. In this section, we describe the process at a general level and then provide implementation details that make the process efficient.

6.1. Relighting

We first light the reflectance field according to the desired image-based lighting environment. Since our subject rotates during the capture, we first rotate the lighting environment

so that it is properly oriented toward the subject. We do this by projecting the environment onto our lighting basis to produce image-based relighting coefficients (Fig. 4). At this point, the Flowed Reflectance Field $R_{u,v}(\mathbf{s})$ becomes a Flowed Light Field $I_{u,v}(\mathbf{s})$ consisting of a set of 36×3 arrays of pre-lit images, one for each point in time in the locomotion cycle. The cylindrical-shaped polygon surface formed by these views is shown in Fig. 5.

6.2. Flowed Light Field Interpolation

Our rendering process morphs between images in the dataset according to interpolation coefficients computed from a light field rendering process. In our notation, linear interpolation according to scalar value β between a particular pre-lit view $I_{u,v}(\mathbf{s})$ and the view to its right is:

$$I'_{u,v,\beta}(\mathbf{s}) = \bar{\beta} I_{u,v}(\mathbf{s}) + \beta I_{u^+,v}(\mathbf{s})$$

Where $\bar{\beta} = 1 - \beta$ and $u^+ = u + 1$. Morphing [BN92,SD96] between the two images based on their flow maps is similarly:

$$I'_{u,v,\beta}(\mathbf{s}) = \bar{\beta} I_{u,v}(\mathbf{s} + \beta F^{\leftarrow}_{u^+,v}(\mathbf{s})) + \beta I_{u^+,v}(\mathbf{s} + \bar{\beta} F^{\rightarrow}_{u,v}(\mathbf{s}))$$

This process is visualized in Fig. 9. Note that the displacement of the pixel coordinate sampled from image $I_{u,v}$ is taken from the flow map $F^{\leftarrow}_{u^+,v}$ from the *other* image $I_{u^+,v}$ to $I_{u,v}$, and vice-versa. This is so that as β approaches 1, the pixel sampled from $I_{u,v}$ approaches the pixel that corresponds to $I_{u^+,v}$'s pixel at $\mathbf{s}$. We can generalize this process to morph between images at the four vertices of the cylindrical polygon according to bilinear interpolation coefficients β and γ as follows:

$$\begin{aligned} I'_{u,v,\beta,\gamma}(\mathbf{s}) = \; & \bar{\beta}\bar{\gamma} I_{u,v}(\mathbf{s} + \beta\bar{\gamma} F^{\leftarrow}_{u^+,v}(\mathbf{s}) + \bar{\beta}\gamma F^{\downarrow}_{u,v^+}(\mathbf{s})) \; + \\ & \beta\bar{\gamma} I_{u^+,v}(\mathbf{s} + \bar{\beta}\bar{\gamma} F^{\rightarrow}_{u,v}(\mathbf{s}) + \beta\gamma F^{\downarrow}_{u^+,v^+}(\mathbf{s})) \; + \\ & \bar{\beta}\gamma I_{u,v^+}(\mathbf{s} + \beta\gamma F^{\leftarrow}_{u^+,v^+}(\mathbf{s}) + \bar{\beta}\bar{\gamma} F^{\uparrow}_{u,v}(\mathbf{s})) \; + \\ & \beta\gamma I_{u^+,v^+}(\mathbf{s} + \bar{\beta}\gamma F^{\rightarrow}_{u,v^+}(\mathbf{s}) + \bar{\beta}\bar{\gamma} F^{\uparrow}_{u^+,v}(\mathbf{s})) \end{aligned}$$

Using this bilinear warping process, we can generate novel views of the person from anywhere on the cylindrical viewing surface. If we were to generate a dense sampling of such views on all of the polygons, we could use traditional light field rendering to re-bin rays from this viewing surface to generate a view from an arbitrary point in 3D space, including points inside and outside the viewing surface. This is precisely our rendering algorithm, except that we avoid generating this dense sampling of views by computing only the pixels $\mathbf{s}$ of the morphed views I' that comprise the final rendered pixels as follows: for each pixel $\mathbf{t}$ in novel view V:

1. Cast a ray R though $\mathbf{t}$ to intersect the cylinder at point p on polygon (u,v).
2. Determine the bilinear interpolation coefficients β,γ corresponding to p's position within the polygon.
3. Set $V(\mathbf{t}) = I'_{u,v,\beta,\gamma}(\mathbf{s})$ where $\mathbf{s}$ is the pixel in the image plane of I' intersected by ray R through p.

To apply this last step, one must be able to infer the camera parameters corresponding to a virtual view $I'_{u,v,\beta,\gamma}$ in order to project a ray R through the camera's center p to a point on its image plane. We do this by bilinearly interpolating the intrinsic parameters and the quaternion rotations of the real viewpoints at the polygon vertices. We found this to work well since our intrinsic parameters are consistent and since our relative camera rotations are small; were this not the case it would be advisable to use the more general camera interpolation technique of [SD96]. We apply this same process to the alpha channels of the original viewpoints to produce a rendered alpha channel α for the synthesized viewpoint.

We implement flowed light field interpolation on the GPU using OpenGL. Our flowed light fields (36×3 RGBA images and 180 RGBA flow maps) for one time instant can be kept in 256MB of GPU memory, which allows us to render a static pose at interactive frame rates as seen on the accompanying video. For rendering an animated walk cycle, we dynamically send the images and flow maps for each virtual camera to the GPU. To render the subject from a virtual camera, we use a two pass rendering algorithm. First, we render the polygonal geometry of the cylindrical array of cameras. For each pixel, we pass the ray intersection point and texture coordinates to the fragment shader, where we infer virtual camera parameters for $I'_{u,v}$ and determine interpolation coefficients (β,γ) to warp and blend the contribution from the four closest input images.

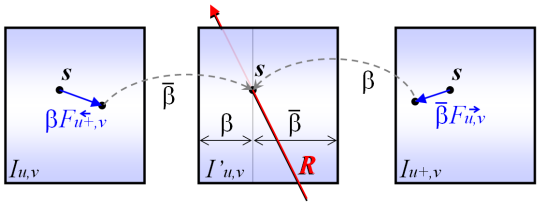


Figure 9: *Novel view generation in a flowed light field.*

6.3. Rendering Shadows

To render shadows, we apply a similar image-based relighting process to the corresponding shadow map basis (Fig. 8) for each frame of the animation. This produces a re-illuminated shadow map (for each color channel) indicating the relative ground irradiance below the subject. We map the shadow onto a horizontal polygon beneath the subject and multiply the background image by the shadow map to attenuate the observed radiance from the ground by the relative irradiance in the map.

6.4. Compositing

With the shadow applied to the background, we composite the rendering of the subject into the scene using its rendered alpha channel and the *over* operator [PD84].

7. Results

We captured and processed three datasets using our technique, a male walking, a female walking, and a male running. Fig. 10(a-c) shows three frames of the male walking through the image-based lighting environment in Fig. 4(d). Since the lighting environment is relatively distant, we light him with the same environment for the whole sequence. We composite the subject and his shadows onto a virtual camera move across a high-resolution image of the environment. The environmental lighting can be seen in the reflection on his skin and the diffused shadows below him. The perspective of the man remains consistent as his distance varies from 3m to 8m away from the virtual camera.

Fig. 10(d-f) shows renderings of the female subject in a virtual environment computed using global illumination. Her viewpoint is matched to an animated camera within the scene, and in each frame she is lit with varying illumination from omnidirectional HDR images rendered from the position of her torso (Fig. 10(d-f), inset). As she walks through the scene, she exhibits various dominant illumination directions and noticeably reflects indirect illumination from the colored walls. In motion, the animation shows occasional warping artifacts from errors in the optical flow, particularly when her arms cross in front of her pants, which are a similar color.

Fig. 10(g-i) shows three frames from an animation of the male's running dataset composited into another image-based lighting environment. The lighting was captured near sunset, yielding warm indirect light on the subject from a building behind the camera.

Fig. 11 shows several instances of the running subject composited into another image-based lighting environment. Each subject casts shadows on the ground from the environment, but our technique does not simulate the lighting interactions between the different subjects. Computing the shadows and indirect illumination that the subjects would cast on each other is a topic of interest for future work.

8. Discussion and Future Work

Our results demonstrate the effectiveness of the technique, but at the same time suggest several avenues for future work. The most noticeable artifacts in our renderings result from inaccurate image correspondences in the optical flow process, producing doubled or warped image elements in places. Currently, our optical flow comparisons are made using using only one image from the lighting basis. It seems likely that improved correspondences could be determined by comparing feature vectors derived from the multiple lighting conditions available in the dataset. Correspondences in textureless areas (such as when the female subject's arm passes in front of her pants) could be improved by including structured light patterns in the lighting basis (e.g. [ZSCS04]). In fact, we included one such condition before the middle tracking frame in our lighting basis (see Fig. 3(a)), but have not yet leveraged this image for computing flow.

Figure 11: *Multiple instances of the running subject are rendered into a photogrammetrically reconstructed lighting environment.*

Another challenge for our technique is the need to acquire a sequence of many consistent locomotion cycles. Even with the tactile treadmill, our subjects drifted slightly with respect to the turntable center and needed to make slight balance adjustments during the capture. The result is that when interpolating between certain viewpoints, the subject appears to widen or narrow due to their unmodeled shift in location. Future work could track the center of mass of the subject, perhaps using data from additional standard video cameras, to rectify the subject's image position accordingly.

We capture cyclic motions in this work to provide an early opportunity to investigate the use of 7D image-based datasets spanning time, view, and illumination; to our knowledge these are the first such datasets recorded. However, the restriction to cyclic motion currently limits the real-world applicability of the process: such characters could be used in crowd simulations or perhaps for a video game character (standing, punching, jumping, and kicking could also be captured), but the process is not able to capture a dramatic performance for a motion picture. One could explore synthesizing a wider range of motions by blending between different captured datasets in a manner analogous to Video Textures [SSSE00] or Motion Graphs [KGP02], though the range of new motions that can be simulated will still be limited. The most straightforward generalization of the technique to non-cyclic performances would employ a 2D array of cameras to capture all viewing directions simultaneously. While this would increase the system cost, cameras able

Figure 10: ***Relighting and View Transformation Results*** *(a-c) A walking subject is composited into an image-based lighting environment. (d-f) Another subject is composited into a virtual global illumination environment; her lighting changes as she walks down the hallway. (g-i) A final subject runs through another image-based lighting environment, receiving yellow-orange indirect light from sunlit buildings behind the camera.*

to capture 512×512 images at 1000fps are not extremely expensive. Fortunately, it is unlikely that capturing the full 360 degrees of the performance is necessary for most motion picture applications. Very likely, a practically-sized array of 3×2 or 5×3 cameras would capture a sufficient range of viewpoints to fine-tune camera angles chosen by the director and cinematographer. With such an array, one could conceivably reduce the amount of data captured by having each different cameras record different interleaved subsets of the lighting basis, analogously to how [WJV*05] staggers their camera exposures across time and space.

Currently, the low resolution of the lighting basis prevents us from being able to accurately simulate sharp sources such as sunlight. As cameras with more recording capacity will soon be available, it will become practical to capture substantially higher lighting resolution. Another approach would be to capture the subject with high-resolution lighting in various poses and use this information to augment the reflectance properties of a dynamic performance captured under a sparser lighting basis.

While our technique is highly data-intensive, we do not yet compress the data along all available dimensions. For example, the MPEG standard could be used to compress our animated reflectance field images spatiotemporally and perhaps across camera viewpoints. The flow vectors already computed for rendering could also be used as motion vectors for image compression.

9. Conclusion

We have presented a new technique for creating image-based renderings of moving people with freely controllable viewpoint and illumination. By leveraging the cyclic nature of locomotion, we have acquired a time-varying 6D reflectance field using a 1D array of cameras; furthermore, we have demonstrated the use of optical flow to effectively interpolate and extrapolate moderately spaced viewpoints in the dataset. Our technique has several limitations: its need for specialized equipment, its limitation to cyclic motion, and (with current recording capacities) its relatively low lighting resolution. Nonetheless, the technique provides notable realism compared to previous image-based approaches for controlling the lighting and viewpoint of a live-action subject. It is also substantially more automatic than traditional virtual human modeling processes, directly capturing motion, appearance, and reflectance from the real world.

Acknowledgements

Andreas Wenger, John Biondo, Maya Martinez, Laurie Swanson, Ian McDowell, Niko Bolas, Tomas Pereira, Krishna Mamidibathula, Larry Vladic, Vision Research, Inc., Marcos Fajardo, Peter Uys, Alexander Singer, Randal Kleiser, Rob Groome, Ramon Gonzales, Ramon Paz, Paul Vu, Drew Weiss, Bill Swartout, David Wertheimer, Neil Sullivan, Randolph Hall, and Max Nikias for their support and assistance with this work. This work was sponsored by the University of Southern California Office of the Provost and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

[BA93] BLACK M. J., ANANDAN P.: A framework for the robust estimation of optical flow. In *Fourth International Conf. on Computer Vision* (May 1993), pp. 231–236.

[BBM*01] BUEHLER C., BOSSE M., MCMILLAN L., GORTLER S. J., COHEN M. F.: Unstructured lumigraph rendering. In *Proceedings of ACM SIGGRAPH 2001* (Aug. 2001), Computer Graphics Proceedings, Annual Conference Series, pp. 425–432.

[BN92] BEIER T., NEELY S.: Feature-based image metamorphosis. *Computer Graphics (Proceedings of SIGGRAPH 92) 26*, 2 (July 1992), 35–42.

[CTMS03] CARRANZA J., THEOBALT C., MAGNOR M. A., SEIDEL H.-P.: Free-viewpoint video of human actors. *ACM Transactions on Graphics 22*, 3 (July 2003), 569–577.

[CW93] CHEN S. E., WILLIAMS L.: View interpolation for image synthesis. In *Proceedings of SIGGRAPH 93* (Aug. 1993), Computer Graphics Proceedings, Annual Conference Series, pp. 279–288.

[DAG95] DORSEY J., ARVO J., GREENBERG D.: Interactive design of complex time dependent lighting. *IEEE Computer Graphics & Applications 15*, 2 (Mar. 1995), 26–36.

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000* (July 2000), 145–156.

[DTM96] DEBEVEC P. E., TAYLOR C. J., MALIK J.: Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of SIGGRAPH 96* (Aug. 1996), Computer Graphics Proceedings, Annual Conference Series, pp. 11–20.

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. In *Proceedings of SIGGRAPH 96* (Aug. 1996), Computer Graphics Proceedings, Annual Conference Series, pp. 43–54.

[HWTD04] HAWKINS T., WENGER A., TCHOU C., DEBEVEC A. G. F. G. P.: Animatable facial reflectance fields. In *Eurographics Symposium on Rendering: 15th Eurographics Workshop on Rendering* (June 2004).

[KGP02] KOVAR L., GLEICHER M., PIGHIN F.: Motion graphs. *ACM Transactions on Graphics 21*, 3 (July 2002), 473–482.

[KUWS03] KANG S. B., UYTTENDAELE M., WINDER S., SZELISKI R.: High dynamic range video. *ACM Transactions on Graphics 22*, 3 (July 2003), 319–325.

[LF94] LAVEAU S., FAUGERAS O.: 3-D scene representation as a collection of images. In *Proceedings of 12th International Conference on Pattern Recognition* (1994), vol. 1, pp. 689–691.

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proc. of SIGGRAPH 96* (Aug. 1996), Computer Graphics Proceedings, Annual Conference Series, pp. 31–42.

[Mag05] MAGNOR M. A.: *Video-Based Rendering*. A K Peters, 2005.

[MB95] MCMILLAN L., BISHOP G.: Plenoptic modeling: An image-based rendering system. In *Proceedings of SIGGRAPH 95* (Aug. 1995), Computer Graphics Proceedings, Annual Conference Series, pp. 39–46.

[MBR*00] MATUSIK W., BUEHLER C., RASKAR R., GORTLER S. J., MCMILLAN L.: Image-based visual hulls. In *Proc. SIGGRAPH 2000* (July 2000), pp. 369–374.

[MKKJ96] MOEZZI S., KATKERE A., KURAMURA D. Y., JAIN R.: Reality modeling and visualization from multiple video sequences. *IEEE Computer Graphics & Applications 16*, 6 (Nov. 1996), 58–63.

[MNP01] MEYER A., NEYRET F., POULIN P.: Interactive rendering of trees with shading and shadows. In *Rendering Techniques 2001: 12th Eurographics Workshop on Rendering* (June 2001), pp. 183–196.

[MPDW04] MASSELUS V., PEERS P., DUTRE P., WILLEMS Y. D.: Smooth reconstruction and compact representation of reflectance functions for image-based relighting. In *15th Eurographics Symposium on Rendering* (Norrkoping, Sweden, June 2004).

[MPN*02] MATUSIK W., PFISTER H., NGAN A., BEARDSLEY P., ZIEGLER R., MCMILLAN L.: Image-based 3d photography using opacity hulls. *ACM Transactions on Graphics 21*, 3 (July 2002), 427–437.

[MPZ*02] MATUSIK W., PFISTER H., ZIEGLER R., NGAN A., MCMILLAN L.: Acquisition and rendering of transparent and refractive objects. In *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering* (June 2002), pp. 267–278.

[MRP98] MILLER G. S. P., RUBIN S., PONCELEON D.: Lazy decompression of surface light fields for precomputed global illumination. *Eurographics Rendering Workshop 1998* (June 1998), 281–292.

[NSD94] NIMEROFF J. S., SIMONCELLI E., DORSEY J.: Efficient re-rendering of naturally illuminated environments. In *5th Eurographics Workshop on Rendering* (June 1994), pp. 359–373.

[PD84] PORTER T., DUFF T.: Compositing digital images. In *Computer Graphics (Proceedings of SIGGRAPH 84)* (Minneapolis, Minnesota, July 1984), vol. 18, pp. 253–259.

[RH02] RAMAMOORTHI R., HANRAHAN P.: Frequency space environment map rendering. *ACM Transactions on Graphics 21*, 3 (July 2002), 517–526.

[RNK97] RANDER P., NARAYANAN P. J., KANADE T.: Virtualized reality: Constructing time-varying virtual worlds from real events. In *Proceedings of IEEE Visualization* (Phoenix, Arizona, 1997), pp. 277–283.

[Saw94] SAWHNEY H. S.: Simplifying motion and structure analysis using planar parallax and image warping. In *International Conference on Pattern Recognition* (Jerusalem, Israel, Oct. 1994), pp. A:403–408.

[SD96] SEITZ S. M., DYER C. R.: View morphing: Synthesizing 3d metamorphoses using image transforms. In *Proceedings of SIGGRAPH 96* (Aug. 1996), Computer Graphics Proceedings, Annual Conference Series, pp. 21–30.

[SKS02] SLOAN P.-P., KAUTZ J., SNYDER J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics 21*, 3 (July 2002), 527–536.

[SSSE00] SCHÖDL A., SZELISKI R., SALESIN D. H., ESSA I.: Video textures. In *Proceedings of ACM SIGGRAPH 2000* (July 2000), Computer Graphics Proceedings, Annual Conference Series, pp. 489–498.

[Sze93] SZELISKI R.: Rapid octree construction from image sequences. *CVGIP: Image Understanding 58*, 1 (July 1993), 23–32.

[TAdA*05] THEOBALT C., AHMED N., DE AGUIAR E., ZIEGLER G., LENSCH H., MAGNOR M., SEIDEL H.-P.: *Joint Motion and Reflectance Capture for Creating Relightable 3D Videos*. Technical Report MPI-I-2005-4-004, Max-Planck-Institut fuer Informatik, 2005.

[TC00] TECCHIA F., CHRYSANTHOU Y.: Real-time rendering of densely populated urban environments. In *Rendering Techniques 2000: 11th Eurographics Workshop on Rendering* (June 2000), pp. 83–88.

[VBK02] VEDULA S., BAKER S., KANADE T.: Spatio-temporal view interpolation. In *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering* (June 2002), pp. 65–76.

[VBK05] VEDULA S., BAKER S., KANADE T.: Image-based spatio-temporal modeling and view interpolation of dynamic events. *ACM Transactions on Graphics 24*, 2 (Apr. 2005), 240–261.

[WGT*05] WENGER A., GARDNER A., TCHOU C., UNGER J., HAWKINS T., DEBEVEC P.: Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics 24*, 3 (Aug. 2005), 756–764.

[WJV*05] WILBURN B., JOSHI N., VAISH V., TALVALA E.-V., ANTUNEZ E., BARTH A., ADAMS A., HOROWITZ M., LEVOY M.: High performance imaging using large camera arrays. *ACM Transactions on Graphics 24*, 3 (Aug. 2005), 765–776.

[YEBM02] YANG J. C., EVERETT M., BUEHLER C., MCMILLAN L.: A real-time distributed light field camera. In *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering* (June 2002), pp. 77–86.

[YMG02] YU J., MCMILLAN L., GORTLER S.: Scam light field rendering. In *Pacific Graphics* (Beijing, China, Oct. 2002).

[ZC04] ZHANG C., CHEN T.: A self-reconfigurable camera array. In *Rendering Techniques 2004: 15th Eurographics Workshop on Rendering* (June 2004), pp. 243–254.

[Zha00] ZHANG Z.: A flexible new technique for camera calibration. *PAMI 22*, 11 (2000), 1330–1334.

[ZKU*04] ZITNICK C. L., KANG S. B., UYTTENDAELE M., WINDER S., SZELISKI R.: High-quality video view interpolation using a layered representation. *ACM Transactions on Graphics 23*, 3 (Aug. 2004), 600–608.

[ZSCS04] ZHANG L., SNAVELY N., CURLESS B., SEITZ S. M.: Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics 23*, 3 (Aug. 2004), 548–558.