

# Fast High-Resolution Appearance Editing Using Superimposed Projections

DANIEL G. ALIAGA, YU HONG YEUNG, and ALVIN LAW

Purdue University

BEHZAD SAJADI and ADITI MAJUMDER

University of California Irvine

We present a system that superimposes multiple projections onto an object of arbitrary shape and color to produce high-resolution appearance changes. Our system produces appearances at an improved resolution compared to prior works and can change appearances at near interactive rates. Three main components are central to our system. First, the problem of computing compensation images is formulated as a constrained optimization which yields high-resolution appearances. Second, decomposition of the target appearance into base and scale images enables fast swapping of appearances on the object by requiring the constrained optimization to be computed only once per object. Finally, to make high-quality appearance edits practical, an elliptical Gaussian is used to model projector pixels and their interaction between projectors. To the best of our knowledge, we build the first system that achieves high-resolution and high-quality appearance edits using multiple superimposed projectors on complex nonplanar colored objects. We demonstrate several appearance edits including specular lighting, subsurface scattering, inter-reflections, and color, texture, and geometry changes on objects with different shapes and colors.

Categories and Subject Descriptors: I.3.3 [Picture/Image Generation]; I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling; I.3.6 [Computer Graphics]: Methodology and Techniques

General Terms: Algorithms, Design

Additional Key Words and Phrases: Appearance editing, multi-projector, light transport, optimization, compensation

## ACM Reference Format:

Aliaga, D. G., Yeung, Y. H., Law, A., Sajadi, B., and Majumder, A. 2012. Fast high-resolution appearance editing using superimposed projections.

This work was funded by NSF CCF-0434398, NSF CNS-0913875, and by a Purdue-Indiana University Applied Research Grant.

Authors' addresses: D. Aliaga, Y. H. Yeung, and A. Law, Department of Computer Science, Purdue University; email: {aliaga, yyeung, ajlaw}@cs.purdue.edu; B. Sajadi, A. Majumder, Department of Computer Science, University of California, Irvine; email: {bsajadi, majumder}@ics.uci.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2012 ACM 0730-0301/2012/04-ART13 \$10.00

DOI 10.1145/2159516.2159518

<http://doi.acm.org/10.1145/2159516.2159518>

ACM Trans. Graph. 31, 2, Article 13 (April 2012), 13 pages.

DOI = 10.1145/2159516.2159518.

<http://doi.acm.org/10.1145/2159516.2159518>.

## 1. INTRODUCTION

The ability to change the appearance of a physical object in order to create entirely different appearances is intriguing and has been a goal of computer graphics and augmented reality. Appearance editing impacts several applications including cultural heritage, product design, entertainment, and architectural relighting.

### 1.1 Background

Appearance edits can be achieved by using digital projectors to project carefully controlled light onto an object. Three issues are important when considering such appearance changes:

- (i) the geometric (planar or nonplanar) and photometric (white or colored) properties of the object,
- (ii) the number of projectors used, and
- (iii) the configuration of the projectors which in turn decides the quality or resolution of the appearance achieved.

Earlier works on appearance editing have addressed complexity in only a few of the preceding aspects. For example, Raskar et al. [2001] presented a system for augmenting custom-made white-colored objects with colored imagery. When multiple projectors were used, they were configured to have little overlap in their fields-of-projection (e.g., 10% overlap). The projected images were geometrically registered across the overlap regions between the projectors. To hide unwanted color inconsistencies from projector to projector, the overlap regions were feathered to create a smooth transition between the imagery illuminated by adjacent projectors. However, no explicit photometric calibration was performed. Bandhopadhyay et al. [2002] extended the approach to provide an interactive painting application and Okazaki et al. [2009] included a high-quality geometric acquisition method.

Many single projector editing systems exist (e.g., Nayar et al. [2003], Grossberg et al. [2004], Grundhöfer and Bimber [2006], and Wetzstein and Bimber [2007]), however, a single projector is fundamentally inadequate for handling objects of arbitrary shape. Even if the projector is optimally placed, some projector pixels will inevitably illuminate the object's surface at grazing angles. In fact, achieving the highest resolution provided by the single projector at all points on the surface is impossible; in practice only a fraction of the maximum resolution is achieved.

Multi-projector editing systems also exist (e.g., Raskar et al. [2001], Bimber et al. [2005a, 2005b]), and this requires the projectors to be registered with each other to create a seamless appearance

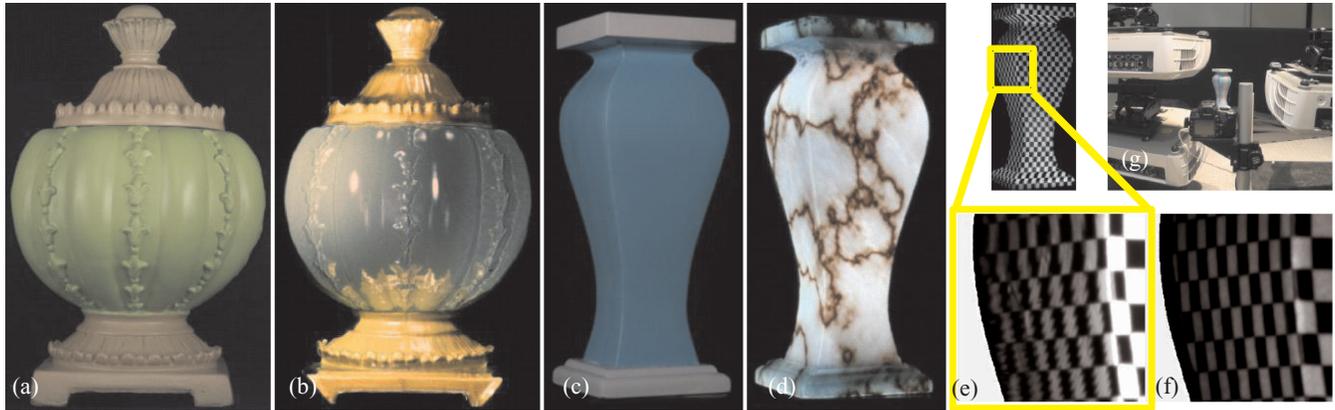


Fig. 1. Fast high-resolution appearance editing. We model the light interactions between multiple projectors with superimposed fields-of-projection over a surface of arbitrary color and geometry yielding fast, high-resolution appearance editing. All images shown are photos of objects visible by the naked eye. (a, c) The physical objects. (b) A glossy appearance of the object in (a) (note: glossiness is a static appearance and not view-dependent). (d) A subsurface-scattered marble appearance of (c). The resolution improvement achieved by our system: (e) one projector, traditional visual compensation and (f) our multiprojector method. (g) Our appearance editing setup.

(e.g., Raskar et al. [2003]). The additional projectors provide significant benefits. For example, the additional light radiance assists in achieving more difficult appearance edits (e.g., editing on darker colored surfaces), Bimber and Emmerling [2006] improved the focus and thus the quality of an appearance edit, and Aliaga et al. [2008] used additional projectors to control the maximum amount of light exposed to each surface point to minimize light exposure to fragile artifacts. While these works are capable of working with arbitrarily shaped objects, the issues presented by grazing angle illumination remain unaddressed.

We observe that using multiple projectors with superimposed fields-of-projection decreases the probability that a portion of the object's surface is illuminated only at grazing angles. In addition, the presence of overlapping pixels from multiple projectors can increase the resolution at head-on surface locations, allowing the entire target appearance to be achieved with improved resolution, not just at grazing angle locations.

Improving resolution or achieving super-resolution has been addressed in previous works. For example, Park et al. [2003] improved resolution using temporally adjacent video frames. In the context of improving resolution with superimposed fields-of-projection, previous works have only dealt with planar white surfaces (e.g., Majumder [2005] and Damera-Venkata et al. [2009]). Both these works addressed the issue of achieving a higher resolution than that of a single projector. However, by limiting their work to planar surfaces, the grazing angle issue is mitigated. Majumder [2005] showed theoretically that super-resolution on planar surfaces is not feasible without changing the size of the pixels. Damera-Venkata et al. [2009] showed that projector placements allowed limited control over the size of the pixels which can be exploited to achieve higher resolution via superimposition of projectors. The resolution of planar displays was doubled by superimposing 4–10 projectors.

Nonplanar surfaces pose an entirely different problem. As mentioned, superimposition from multiple projectors becomes necessary to achieve the resolution of a single projector at grazing angle areas. Further, nonplanar surfaces result in much greater variations in the focal distance of the projectors, resulting in pixel blurs due to the defocus during superimposition. Finally, while white surfaces have an almost constant albedo, colored surfaces generally consist

of significant variation in surface albedos. In designing our method, we consider the variations in pixel size, pixel blur, and surface albedos while striving to minimize degradation from a single projector resolution via superimposed multiple projectors. Thus, our content-dependent method allows us greater leverage to custom optimize the reduction of degradation in resolution in a spatially varying manner based on the desired spatially varying resolution of the target appearance.

Our appearance editing framework significantly advances the State-of-the-art by simultaneously considering complexity in all the aforementioned three aspects in order to achieve appearance edits of much higher quality than that possible by any earlier work. Our system achieves high-quality appearance edits on *complex nonplanar colored* objects by using multiple superimposed projectors (i.e., nearly 100% overlap of their fields-of-projection).

## 1.2 Approach Overview

We present an approach for using multiple superimposed projections to achieve fast high-resolution appearance editing (Figure 1). By high resolution, we imply a new appearance achieved at a resolution over the surface of the object that is higher than that accomplished using a single projector. By fast, we imply changing such appearances at nearly interactive update rates. We assume the surfaces to edit are diffuse or near-diffuse with no inter-reflections or other indirect illumination.

Our appearance modeling component uses a multi-projector light transport matrix that models the influence of each projector pixel on the camera image plane. The inverse of this multi-projector light transport matrix provides the compensation images to be projected to achieve a desired appearance. However, unlike a single projector system, when using multiple superimposed projectors to improve the resolution of appearance editing, there exist many possibilities of how light can be illuminated from the projectors to create the desired appearance. We model this problem as a large constrained optimization and show that an optimal solution to this optimization yields the best-quality appearance in the sense of producing a smooth, artifact-free appearance at the highest resolution possible while being within the illumination capability of the projectors.

The error metric we optimize provides a quantitative measure of the proximity of our achieved appearance to the desired appearance. Lastly, we include an accurate modeling process for individual projector pixels.

Our acceleration component tackles the issue that the constrained optimization system is huge, and solving such a large optimization once for each desired appearance is impractical. Our solution is to decompose the problem into two tasks: (a) a preprocessing task that is performed only once per object-projector setup; and (b) a fast runtime linear scaling task. The preprocessing task uses a parallelized solver to compute a solution for a base appearance. The computed solution is a base compensation image which, when projected, creates the base appearance. Then, for every new target appearance, the corresponding target compensation image is obtained by scaling the base compensation image at runtime by a scale image unique to the target appearance. This task only requires per-pixel multiplication and division operations thus enabling fast appearance edits at runtime.

### 1.3 Our Contributions

Succinctly, we present three main contributions.

- a formulation of high-resolution appearance editing as a constrained optimization whereby the constraints ensure the resulting solution is both feasible given the limited illumination ability of the projectors and smooth despite the limited resolution of the cameras and projectors and presence of acquisition noise and error,
- an algorithm to make computing and using the constrained optimization practical by decomposing the generation of new appearances into a one-time constrained optimization precomputation and a fast runtime linear scaling operation, and
- a projector pixel modeling process to accurately compute projector pixel properties (size, shape, center, and intensity distribution) and to model their interaction amongst several projectors, all of which are critical for our objective of high-resolution appearance editing and are also useful to other projector-camera applications.

To demonstrate our approach, we perform several high-resolution appearance edits and analyses on a variety of objects. Our appearance edits include adding subtle details to objects, performing illumination changes, modifying object colors, simulating material alterations (e.g., transforming an object to have subsurface scattering, converting a diffuse object to one with a static specular appearance), and quantitative and qualitative measurements of the improvement in resolution. Altogether, the a priori acquisition and modeling requires 2–3 hours. Generating a new target appearance at runtime takes less than a second.

### 1.4 Article Organization

In the remainder of this article, we describe the main components of our approach. Figure 2 shows a summary of our overall system pipeline. In the order of importance, we first describe our a priori computations for high-resolution appearance editing as a multi-projector constrained optimization (Section 2); we assume that the acquisition and pixel modeling steps are already completed. Then, we describe our runtime computations for quickly generating target appearances (Section 3). The acquisition and pixel modeling step is of great importance since it models the scene, geometrically and radiometrically calibrates the projectors, captures a multi-projector light transport matrix, and performs accurate per-projector-pixel modeling. The geometric and radiometric information is needed

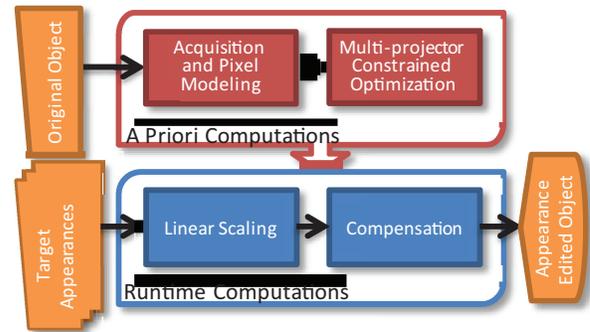


Fig. 2. System pipeline. A summary of our system pipeline.

in order to compute the necessary compensation images for a new appearance. Radiometric calibration serves to model the differences in color and intensity responses of the projectors and to linearize their behavior for more accurate appearances. In Section 4, we provide algorithm details about acquisition and pixel modeling. Finally, we present implementation details (Section 5), examples and analyses (Section 6), and conclusions (Section 7).

## 2. APPEARANCE MODELING

The goal of our appearance modeling component is to obtain a set of projector compensation images that, when simultaneously illuminated on the object, produce a high-resolution visual appearance as visually similar as possible to the target appearance. Our method builds upon prior work on light transport and/or projector-camera systems (e.g., Ng et al. [2003] and [Sen et al. 2005]) that use light transport to capture the light interactions with the object and then invert the light transport matrix  $T$  in order to calculate the desired projector compensation images. However, we face two new challenges.

- The light transport matrix  $T$  is very large. For example, for a multi-projector system, the size of  $T$  is on the order of  $10^7$  rows by  $10^6$  columns thus making computational efficiency to be of prime importance.
- Inverting a multi-projector light transport matrix  $T$  that explicitly models overlapping projector pixels requires solving a constrained optimization with many potential solutions, each producing of a set of compensation images yielding appearances of varying quality. For a single projector, a unique pseudo-inverse of  $T$  can be easily computed by bounding the intensity value of each pixel between  $[0,1]$ ; in practice, the intensity is then multiplied by the maximum pixel value, for example, 255. For the multi-projector case, we must carefully constrain the problem so as to arrive at an optimal solution, defined as a high-quality and high-resolution appearance. To illustrate this, Figure 3 shows a simple didactic example with several possible compensation image solutions when using multiple projectors. The figure intuitively shows that by using more projectors, a better solution can be achieved than with any single projector (Figures 3(a)–(g)). Further, the inversion of  $T$  must also take into account the illumination capabilities of the projectors or a solution might be computed that produces the target appearance but yields intensities that cannot be projected (Figures 3(h)–(i)). In addition, smoothness constraints are necessary to ensure the smooth areas of the target appearance appear as such; else the computed intensities will have unnecessary undulations (Figures 3(j)–(k)).

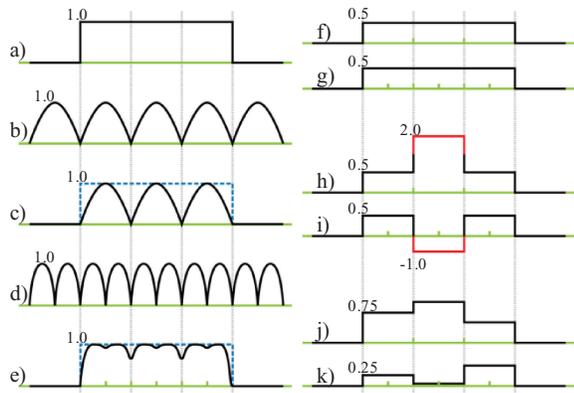


Fig. 3. Multiprojector constrained optimization. (a) Target intensities to achieve across an appearance. (b) Maximum surface illumination intensities from one projector. Each hump represents one projector pixel across the surface. (c) Reconstruction of (a) (dotted) using (b) and appropriate projector intensity values. (d) Maximum surface illumination intensities from a second projector which, in this example, is positioned at a different orientation relative to the object that yields smaller projected pixels. (e) Reconstruction of (a) using both projectors. The reconstruction is more accurate than (c) and also more accurate than using only the second projector. (f)–(g) Smooth projector intensities for the two projectors to achieve (e). (h)–(i) Without illumination constraints, intensity scales can overflow or underflow (in red) but theoretically still produce (e). (j)–(k) Without smoothness constraints, projector intensities may produce noise, shown as intensity undulations.

Our described method supports full RGB color, but for clarity we express our techniques for a single channel. Extending to RGB implies repeating the same calculations for three independent channels. Spectral overlap between the color channels of the projectors and the camera is handled by the radiometric calibration.

## 2.1 Light Transport

The light transport  $T$  between  $N$  projectors of resolution  $p \times q$  and a camera of resolution  $m \times n$  is modeled by a  $mn \times Npq$  matrix (e.g., Ng et al. [2003], Sen et al. [2005]). Mathematically, this is expressed as

$$C = TP, \quad (1)$$

where  $P$  is a  $Npq \times 1$  vector sequentially representing the pixels of all projectors and  $C$  is a  $mn \times 1$  vector representing the camera pixels. Given a target image  $C_t$  (i.e., the desired appearance as viewed from the camera's viewpoint), the target compensation image can be expressed as

$$P_t = T^+ C_t, \quad (2)$$

where  $T^+$  is the pseudo-inverse of  $T$  (the pseudo-inverse is needed since  $T$  is typically not square).  $P_t$  can then be partitioned into  $P$  segments, each containing the compensation image for a projector. A change in appearance can be obtained by replacing the target image  $C_t$  with a new one which in turn produces the new compensation image  $P_t$  needed to achieve the appearance.

## 2.2 Inverting the Light Transport Matrix

We formulate inverting the multi-projector light transport matrix  $T$  as a constrained optimization problem. Sen et al. [2005] also used light transport but did not require  $T$ 's inversion. Instead, they exploit Helmholtz reciprocity to change the viewpoint from the camera to

that of the projector by transposing the light transport matrix  $T$ . The work of Seitz et al. [2005] and of Wetzstein and Bimber [2007] is more closely related to our work. Seitz et al. define a theory of inverse light transport to create inter-reflection cancellation operators for removing the effects of inter-reflections and shadows from images. Wetzstein and Bimber invert the transport matrix of a single projector for radiometric compensation assuming no (significant) inter-pixel interaction. Further, since light transport is used, their method supports inter-reflection and global illumination effects. For these systems, only one way exists for assigning values to the projector compensation image pixels. To clarify this, let  $v_i \in [0, 1]$  be the intensity value assigned for projector pixel  $i$ . Each  $v_i$  is to be scaled just sufficiently to match the target image at the camera pixel which coincides with the projector pixel center. Hence, there is a unique inverse to  $T$  which when multiplied by  $C_t$  produces a unique compensation image  $P_t$ .

In contrast, when introducing pixels from one or more additional projectors, there are multiple combinations of values for the  $v_i$ 's that can create an appearance better than that produced by a single projector. To ensure the best quality, we need to find  $v_i$ 's which minimize the difference between the achieved appearance and the target high-resolution appearance. Hence, computing  $T^+$  is an optimization. Formally, let  $E_t = \|C_t - TP_t\|$ ; minimizing  $E_t$  yields a high-resolution appearance that is closest to the target image  $C_t$ .

## 2.3 Illumination Constraints

For multiple projectors, the solution values  $v_i$  must be constrained to be within the acceptable  $[0, 1]$  range. In the case of a single projector system,  $v_i$  can be clipped to  $[0, 1]$  as a postprocess. This clipping assures the best quality possible with a single projector since no additional projectors can compensate for the error caused by clipping. However, for multiple projectors, constraining is critical to avoid solutions that achieve optimality with  $v_i$ 's outside of a projector's capability (e.g., underflowing and overflowing intensities in Figures 3(h)–(i)). The ill effects that occur when ignoring illumination constraints are shown in Figure 8. Thus, computing  $T^+$  now requires solving a constrained linear optimization with a solver capable of constraining the solution vector (we use Matlab's lsqlin function which uses the reflective Newton method of Coleman et al. [1996]).

## 2.4 Smoothness Constraints

Our constrained optimization also enforces adjacent projector pixels from the same projector that illuminate a uniform color region of the target appearance to have similar compensation values. Consider a local uniform intensity area of the target image. In general, there are two solutions to achieve this appearance: (a) the color comes from all of the projectors which sum up to the uniform color of the target image, or (b) at some pixels one projector has a high contribution but the projector has a low contribution at an adjacent pixel. This second solution introduces unnecessary noise due to projector intensity changes from pixel to pixel in a small local neighborhood, exacerbating calibration errors. To remedy this situation, we introduce smoothing constraints which assure (i) neighboring projector pixels from any one projector have smoothly changing compensation values if the corresponding area in the target image is also smoothly varying, and (ii) if the target image has a sharp edge in a local region, the compensation values for the corresponding neighborhood in the projector compensation image need not be similar. In essence, the weights of the constraints should be inversely proportional to the difference of the intensities in the target image inspired by the idea of bilateral filtering (e.g., Durand et al. [2002]).

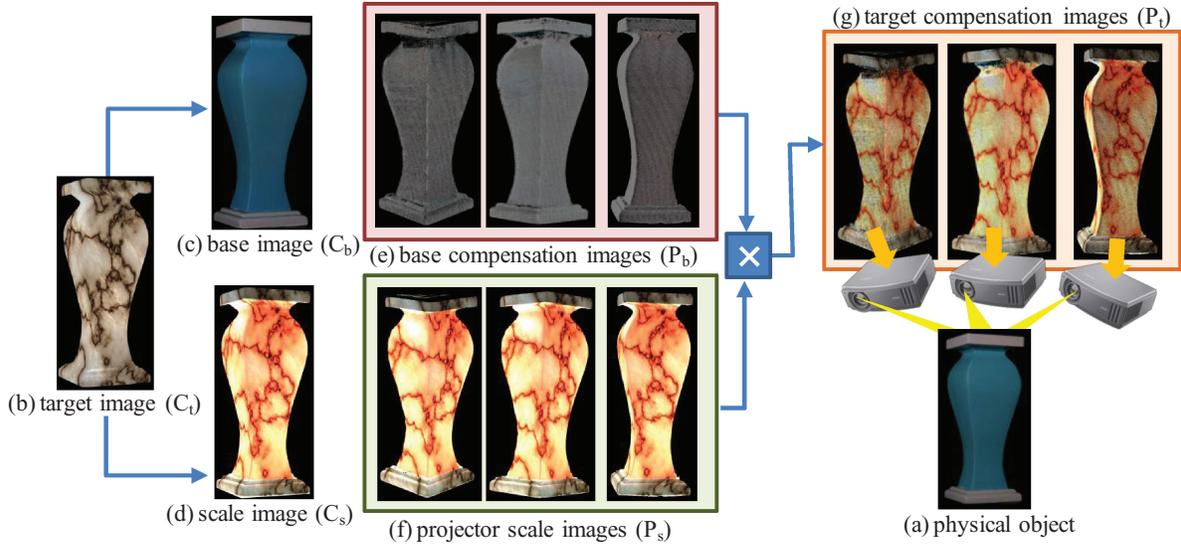


Fig. 4. Base-scale decomposition. Given a particular object (a) and a target image (b), superimposed projections are used to alter the appearance of the object to that of the target image. The target image is decomposed to a base image (c) and a scale image (d). The projector base compensation images (e) are computed once using the base image. The projector scale images (f) are computed using the scale image. (e) and (f) are multiplied together to create the final target compensation images (g) which are projected onto the physical object to create the target appearance. A photograph of the final appearance edited object is in Figure 1(d).

To achieve this smoothness, we augment Eq. (1) with additional linear equations (constraints). For any two adjacent projector pixels  $j_1$  and  $j_2$ , we add an equation  $\alpha_{j_2 j_2}(j_1 - j_2) = 0$  to the system in Eq. (1). The weight  $\alpha_{j_2 j_2}$  is defined as

$$\alpha_{j_2 j_2} = 1 - \frac{|\tau_{j_1} - \tau_{j_2}|}{\max(\tau_{j_1}, \tau_{j_2})} \quad (3)$$

and simulates an inverse bilateral filter.  $\tau$  is the ratio of the target image intensity to the observed intensity of the projector pixel. The effect of using this constraint is illustrated in Figure 8.

### 3. ACCELERATING APPEARANCE GENERATION

While previous works have focused on reducing the acquisition time and storage requirements of the light transport matrix  $T$  (e.g., Garg et al. [2006], Sen and Darabi [2009], Wang et al. [2009]), our acceleration component focuses on improving the process of inverting  $T$  so as to quickly generate the compensation image  $P_t$  for each new target image. Our observation is that computing  $P_t$  can be decomposed into a pixel-wise multiplication and a computation of two compensation images: (i) a base compensation image  $P_b$  computed once per object, and (ii) a scaled compensation image  $P_s$  calculated at runtime for each new appearance. Then, using only simple pixel-wise multiplication we obtain an approximation to  $P_t$  at close to interactive speeds on the CPU (and could be at real-time rates with a GPU implementation). For our diffuse surfaces, the decomposition separates geometric components of the compensation into the base image and reflectance components of the compensation into the scale image.

#### 3.1 Base-Scale Decomposition

Our method is based on decomposing any desired target image  $C_t$  into a base image  $C_b$ , unique to an object-projector setup, and a scale image  $C_s$ , unique to a target image. Thus, any target image  $C_t$

can be expressed by

$$C_t = C_b \otimes C_s, \quad (4)$$

where  $\otimes$  denotes a per-pixel multiplication operation. For the base image  $C_b$ , we compute its corresponding base compensation image  $P_b$  using Eq. (2) once a priori for the target object. Then, we seek to quickly compute the corresponding projector-space scale image  $P_s$  from the camera-space scale image  $C_s$  at runtime.

Each projector pixel of  $P_s$  is computed as a weighted sum of camera pixels from  $C_s$ . We assume each projector pixel  $j$  contributes light to a contiguous region of camera pixels  $A_j$  (defined by a 2D elliptical Gaussian in Section 4) and denote the relative contribution of each pixel in  $A_j$  (i.e.,  $A_{jk}$  for  $k \in [1, |A_j|]$ ) to  $P_s$  by the set of weights  $w_j = \{w_{j1}, w_{j2} - w_{j|A_j|}\}$ . Each scalar value  $P_s[j]$  (i.e., the scalar intensity value of the  $j$ 'th pixel of the projector-space scale image  $P_s$ ) is then computed as

$$P_s[j] = \sum_{k=1}^{|A_j|} C_s[A_{jk}] w_{jk}. \quad (5)$$

If we consider the weights  $w_j$  to approximately describe a normalized Gaussian weighted kernel, then  $P_s$  is equivalent to a low-pass-filtered version of  $C_s$ .

The target compensation image  $P_t$  is now approximated by  $P_a$ ,

$$P_t \approx P_a = P_b \otimes P_s, \quad (6)$$

where  $P_b = T^+ C_b$  is the base compensation image. The advantage of this decomposition is that we can precompute the base compensation image  $P_b$  for an object. Then, for every new target appearance  $C_t$ , we only need to multiply the base compensation image  $P_b$  with a corresponding scale image  $P_s$  to achieve the desired appearance. Figure 4 shows example images.

Different  $C_b$  base images provide different system capabilities. For example, using a base image which exactly matches a target

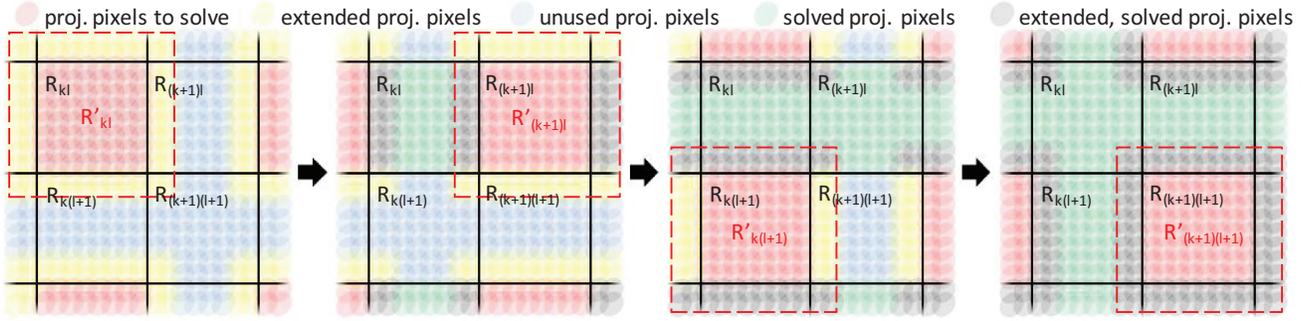


Fig. 5. Base compensation image computation. Our algorithm divides the projector pixels into four groups, each processed in parallel. In each step, one of the projector pixel groups (in red) is solved. Yellow pixels are part of  $R'_{kl}$ . Blue pixels are unused and yet to be solved, Green pixels have already been solved, and the grey pixels are pixels in  $R_{kl}$  which are already solved.

appearance will yield a high-quality appearance edit for the given target appearance, however, the solution may not be optimal for different appearances. To achieve a target appearance with high-resolution details (i.e., sharp changes), a base image containing sharp changes everywhere may be used with the hope that many of the sharp changes in the desired target experience are spatially coincident with those of the base image; however, smooth target appearances may suffer. Our informal experiments have shown that a good compromise base image  $C_b$  for a variety of alterations is an image that captures the surface albedo of the object without significant illumination effects (e.g., only diffusely reflected light is visible [Mallik et al. 2005], see Section 5 for details on acquiring such an image). This base image contains sharp changes precisely where the object albedo actually changes and thus ensures that a good compensation can be computed in those regions. This is beneficial to a target appearance (i.e., the smooth parts of the object’s physical albedo are easy to change, so we worry mostly about the boundary areas of the object’s albedo). Further, illumination changes (with the exception of sharp self-shadows) are known to be low frequency. Choosing an all-white appearance would be similar in effect since a smooth transition would have to be computed at albedo color edges.

Thus, our base-scale decomposition can yield visual accuracy comparable to that using a constrained optimization for each unique target appearance. Moreover, albedo changes (e.g., color/pattern changes) can be performed as well (see Figures 1 and 12 for a variety of strong albedo changes), and a base image which captures the color edges of the albedo allow the optimization to improve the appearance at these color edges. Nonetheless, the maximum resolution achievable by albedo changes when using  $P_a$  is subject to some resolution limitations (see Section 3.3).

### 3.2 Base Compensation Image Computation

To make computing the base compensation image  $P_b = T + C_b$  practical despite the large size of  $T$ , we use a parallelized solver to compute  $P_b$ . Our parallelized solver divides the camera space into a grid, converting a single huge constrained optimization into multiple smaller constrained optimizations, one for each grid cell. The smaller constrained optimizations are then solved (with the reflective Newton method of Coleman et al. [1996]) in a parallel four-phase fashion. Visual artifacts are avoided by introducing constraints from adjacent grid cells when solving a grid cell’s

constrained optimization. Since we assume there are no indirect illumination effects, no additional constraints are needed from non-adjacent grid cells. The main steps of our solver are as follows.

- (1) First, we divide the camera space image  $C_b$  into a  $K \times L$  grid of disjoint rectangular regions. Each region  $R_{kl}$ , for  $k \in [1, K]$  and  $l \in [1, L]$ , is of width  $g$  and height  $h$ , containing  $H_{kl} = g \times h$  camera pixels. We typically let  $R_{kl}$  be  $16 \times 16$  in size. Let  $G_{kl}$  be the number of projector pixels from all projectors in the setup whose centers lie inside  $R_{kl}$ . The rectangular regions  $R_{kl}$  are extended by a fraction of  $g$  and  $h$  (e.g., 25%) to create overlap between adjacent  $R_{kl}$ ’s and are called  $R'_{kl}$ . Let  $H'_{kl}$  be the number of camera pixels inside  $R'_{kl}$ , where  $H_{kl} < H'_{kl}$ . Let the number of projector pixels whose center lies in  $R'_{kl}$  be  $G'_{kl}$ . Since  $R_{kl} \subset R'_{kl}$  then  $G_{kl} < G'_{kl}$ . The extended  $R'_{kl}$ ’s ensure continuity of the values in the compensation image across the boundaries of  $R_{kl}$  during optimization (Figure 5).
- (2) Next, smaller sets of linear equations, one for each  $R'_{kl}$ , are created from the large set given by Eq. (1). All the equations involving the  $G'_{kl}$  projector pixels and  $H'_{kl}$  camera pixels are chosen to create a smaller system of equations,

$$C_{kl} = T_{kl} P_{kl}. \quad (7)$$

In the preceding equation, the sizes of  $T_{kl}$ ,  $P_{kl}$ , and  $C_{kl}$  are  $H'_{kl} \times G'_{kl}$ ,  $G_{kl} \times 1$ , and  $H_{kl} \times 1$  respectively. In the next step, we solve these smaller systems of equations in parallel while assuring continuity across boundaries.

- (3) Lastly, the parallel solver works in four phases. In each phase, the intensity values of projector pixels associated with  $\frac{K}{2} \times \frac{L}{2}$  regions are computed in parallel. These  $R'_{kl}$  of alternate rows and columns in the  $K \times L$  grid are solved to assure that no overlapping  $R'_{kl}$ ’s are processed together. Solving a  $R'_{kl}$  region finds the value of the compensation image at  $G'_{kl}$  projector pixels, but only  $G_{kl}$  of these pixels belonging to  $R_{kl} \subset R'_{kl}$  are taken as part of the final solution. When solving for the adjacent region  $R'_{(k+1)l}$  in the next phase, values of the compensation image at some of the  $G'_{(k+1)l}$  projector pixels in the boundary of  $R_{kl}$  and  $R_{(k+1)l}$  are already finalized. Hence, the equations being solved in this next phase are much more constrained in the overlapping boundaries than in the current phase. The same pattern repeats in the subsequent phases when  $R'_{(l+1)}$  and  $R'_{(k+1)(l+1)}$  are solved. In each phase the equations being solved get more constrained, assuring continuity at the boundaries.

### 3.3 Accuracy of the Decomposition

The base-scale decomposition opens up the question of accuracy: how close is the result of using  $P_a$  to using  $P_t$ ? To address this question, we analyze the potential frequency content of the involved images. In our discussion, the maximum frequency content of an image that can be illuminated by a single projector is denoted as  $F$  (e.g., for a projector of horizontal resolution 1024,  $F = 512$ ).

Consider the case when a target appearance  $C_t$  has content with a maximum frequency of  $F$ . When using a single projector, the appearance  $C_t$  can only be fully achieved on the subset of the object's surface that is head-on to the projection direction. In all other regions, the high-frequency details in  $C_t$  are aliased (e.g., Figure 1(e)). For most nonplanar objects, grazing angles are common, and the maximum frequency content over these portions of an object is a small fraction of  $F$ . In contrast, when using multiple projectors with superimposed projections, the limitation in the maximum possible frequency is removed because of the additional illumination directions (e.g., Figure 1(f)).

The compensation images  $P_a$  computed by our base-scale composition accurately yield a maximum frequency content of  $F$  and thus can be used to significantly improve appearance editing resolution as well as computation speed. Recall  $P_a = P_b \otimes P_s$ .  $P_b$  has the same maximum frequency content as  $P_t$  since it is obtained by computing  $T^+$  and multiplying with  $C_b$ ;  $C_b$  in turn is bandlimited by  $F$ , thus so is  $P_t$ . Also,  $P_s$  is bandlimited by  $F$  since it is created in projector space using Eq. (5). Therefore,  $P_a$  is bandlimited by  $F$  as would be  $P_t$ , and thus the base-scale decomposition can be used to significantly improve the computation speed of multiple appearances without sacrificing accuracy in the appearances.

However,  $C_t$  can theoretically have content with a frequency higher than  $F$  since the camera's resolution is usually higher than the projector's resolution (the frequency content is still limited by the camera's resolution, however). For surface locations that are head-on illuminated by two or more projectors,  $P_t$  can theoretically produce an appearance with frequencies higher than  $F$  but not more than that possible by the combined sampling ability of the  $N$  projectors. In this case, using the base-scale decomposition without sacrificing accuracy requires the high-frequency content of  $C_t$  to be shifted into  $C_b$  rather than leaving it in  $C_s$ . Then, the frequency content of the appearance that is higher than  $F$  is transferred to  $P_a$  via  $P_b$ , making  $P_t = P_a$ . If the high-frequency content is not placed in  $C_b$ , the higher frequencies will be lost during the low-pass filtering to create  $P_s$  and will result in aliasing artifacts. An experimental validation is illustrated in Figures 12(e)–(f) (a proof is also available in Appendix A). Since only a low resolution  $C_b$  allows us to create a large number of different appearances quickly, this shows that for target appearances that contain frequencies higher than  $F$ , the full constrained optimization must be performed (i.e.,  $C_t$  is used as the base image to compute  $P_t$ ).

## 4. ACQUISITION AND PIXEL MODELING

Our approach includes a comprehensive acquisition component to model and calibrate the appearance editing stage. The digitization and calibration methods are mostly based on previous methods. However, a key distinguishing factor of our acquisition component is an accurate projector pixel modeling method. Accurately detecting projector pixel properties, such as pixel shape, center, and intensity distribution as well as the overlap between projector pixels, is critical for our objective of high-resolution, high-quality appearance edits and for making high-resolution appearance editing feasible using off-the-shelf hardware. Since projector pixels

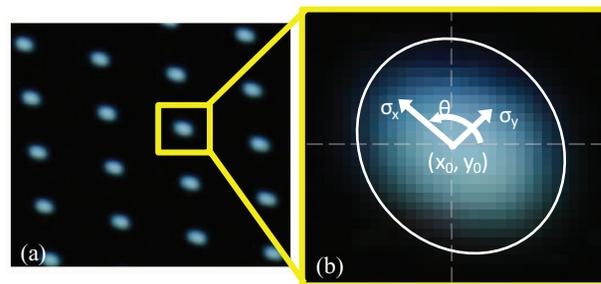


Fig. 6. Projector pixel modeling. (a) Projector pixels from an acquisition pattern image used to estimate metapixel parameters. (b) Close-up of a projector pixel on the camera plane and the five metapixel parameters to be discovered.

may overlap with other projector pixels from the same projector or from other projectors, a precise model that reflects the overlap between pixels is important. If the amount of overlap is over- or under-estimated, aliasing artifacts or excessive loss of contrast may appear in the resulting appearance.

Contemporary work on projector-camera systems seeks a good analytical model for representing projector pixels (e.g., Summet et al. [2006], Yang et al. [2005]). Similar to Chuang et al. [2000] and Ruzon and Tomasi [2000], our projector pixel modeling approach is based on using 2D elliptical Gaussians to define *metapixels*, but unlike those approaches we use 2D elliptical Gaussians to precisely model the interaction of projector pixels across multiple projectors. Our pixel modeling approach is useful for appearance editing and also for other projector-camera applications.

### 4.1 Calibration and Light Transport

Prior to appearance editing, we perform a self-calibrating reconstruction of the object (which also calibrates the poses of the projectors), compute a radiometric calibration, and sample the light transport matrix. Our self-calibrating object reconstruction method is based on the photogeometric structured-light method of Aliaga and Xu [2009] which supports diffuse and mildly specular objects. The projectors and camera are radiometrically calibrated using High Dynamic Range (HDR) imaging techniques that also output the camera transfer function [Debevec et al. 1997]. Further, to linearize the projectors, we use the work of Raji et al. [2003].

To acquire the light transport matrix  $T$ , a set of patterns is illuminated from each projector onto the scene, and a camera captures their projection (similar to Sen et al. [2005]). For these patterns, a projector image is divided into regions of  $r \times r$  pixels. In each pattern, only one pixel in each region is turned on (to white). The captured image for pattern  $i \in [1, r^2]$  is denoted by  $M_i$ .

### 4.2 Projector Pixel Modeling

Projectors attempt to emit discrete (square) pixels, but due to the relative orientation of the projector with respect to the local surface normal and the local dispersion of light on the surface, the appearance of a pixel projected on a diffuse surface on the camera plane more closely resembles a blob of pixels with a smooth fall-off (Figure 6(a)). We experimented with several pixel models (Figure 8) and found elliptical Gaussians to be good estimates of these metapixels. A metapixel is not a simple point but has associated properties such as axes lengths, amplitude, position, and orientation as defined by an elliptical Gaussian (Figure 6(b)). Further, a metapixel exists in camera space instead of projector space but is still indexed in the same way as projector pixels.

Another advantage of our analytical modeling of projector pixels is the improved ability to support the limited resolution of cameras. For an ideal noise-free camera at the same distance from the object as all projectors, the Nyquist sampling criterion states that a camera resolution of double the horizontal and vertical projector resolution is sufficient to estimate the projector properties to subpixel accuracy. However, when using consumer cameras, the presence of noise cannot be avoided and thus a denser sampling is required. Our experiments show that our model better estimates projector pixels' properties and their overlap amounts, despite relatively coarse camera pixel sampling, and obtains improved quality as compared to simpler approaches.

**4.2.1 Metapixel Acquisition.** To capture the properties of the metapixels, we exploit the patterns used to acquire the light transport and also capture an image  $W$  of the scene as illuminated by an all-white image. The intensity of the metapixels captured in image  $M_i$  is related to the poses of the projector, camera, and surface as well as the color and reflectance of the surface. Our initial modeling goal is to accurately detect the blob in  $M_i$  corresponding to each metapixel so that we can estimate its shape and location. However, if the nature of the object's reflectance makes the blob too dark, we cannot detect it well. To remove the dependency on the intensity of the surface reflectance, we divide each  $M_i$  by  $W$ . The resulting intensity-independent pattern image  $M'_i$ , has a normalized intensity for every metapixel. In the following, we describe how to process  $M'_i$  to estimate the elliptical Gaussian for each metapixel.

**4.2.2 Fitting Optimization.** An elliptical Gaussian  $G_e(x, y)$  is fitted to each metapixel blob in  $M'_i$ . We compute the elliptical Gaussian's amplitude  $A$ , camera image center  $(x_0, y_0)$ , orientation  $\theta$ , and standard deviations  $\sigma_x, \sigma_y$  along the ellipse's two semi-axes. Since  $M'_i$  is intensity independent, we cannot estimate the elliptical Gaussian's amplitude  $A$  from this image. However,  $M'_i$  is used to estimate all other properties of a metapixel using a nonlinear least squares optimization that minimizes

$$\| G_e(x, y) - M'_i(x, y) \|^2 \rightarrow 0, \quad (8)$$

where  $(x, y)$  is a camera pixel covered by the metapixel. To initialize the optimization, we use the projector-camera correspondences from the geometric calibration. These estimated properties are used to detect the blob corresponding to the metapixel in  $W$ , and then  $A$  is computed as a weighted average of the constituting pixel intensities in  $W$ . Since each metapixel blob is nonoverlapping in each  $M_i$ , the fitting optimizations for each metapixel are computed in parallel.

**4.2.3 Metapixel Interaction.** To accurately model the interaction between metapixels, we use the multi-projector light transport representation as described in Section 2. In this representation, the matrix element  $T[u, v]$  corresponds to the response of camera pixel  $u$  to the projected image of projector pixel  $v$  on the object surface. To create  $T$ , we assign  $G_s^v(u_x, u_y)$  to matrix element  $T[u, v]$  where  $G_s^v(\cdot, \cdot)$  is the elliptical Gaussian function for metapixel  $v$  and  $(u_x, u_y)$  are the camera coordinates of camera pixel  $u$ . Since the metapixels may overlap each other in camera space,  $T$  encodes the interaction between metapixels within and across different projectors and is generally a sparse matrix. Our parallel solver described in Section 3.2 breaks up  $T$  into multiple smaller matrices. While these smaller matrices are relatively denser than  $T$ , they are still sparse and benefit from our sparse constrained linear system solver.

**4.2.4 Metapixel Regularization.** To reduce noise in the system, we exploit the fact that metapixels from the same projector are uniformly spaced in the projector image. In general, errors are

present in the acquisition, calibration, digital cameras, and elliptical Gaussian optimizations. To reduce noise, we use a priority-based algorithm which ensures that the locations of the metapixels in the camera image exhibit the expected local uniformity in their spatial distribution. Consider metapixel  $j$  of a projector with estimated center  $(x_j, y_j)$ . Even for a nearly parallel camera image plane and projector image plane, the estimated metapixel locations computed in Section 4.2.1 deviate from perfect uniformity due to accumulated numerical errors. Our priority-based regularization algorithm regularizes the spacing between metapixels by perturbing the estimated  $(x_j, y_j)$ 's so as to improve their spatial uniformity with respect to immediately adjacent metapixels from the same projector. The local surface orientation and perspective foreshortening will make the distances between immediately adjacent metapixels not perfectly equal; nevertheless, we found our approximation sufficiently accurate to yield improved results.

This regularization method consists of the following steps, repeated for a number of iterations.

- (1) First, we identify the estimated centers of the eight connected neighbors of metapixel  $j$  in the camera image. Since we know the projector pixel that corresponds to each metapixel, we can obtain metapixel adjacency information directly from the pixel position on the projector image plane.
- (2) Next, we estimate the local uniformity in spatial distribution at  $(x_j, y_j)$  by computing the ratio  $\rho_j$  of the shortest to the longest distance of  $(x_j, y_j)$  from its neighboring metapixel centers in the camera image.
- (3) To prioritize the metapixels, we introduce a metric  $\beta_j$  defined as the change in  $\rho_j$  if metapixel  $j$  is moved from its current location to the average location of its neighbors in the camera image. Let  $\rho'_j$  be this new ratio and  $\beta_j = \rho_j - \rho'_j$ . A positive  $\beta_j$  indicates an improvement in the local uniformity while a negative  $\beta_j$  implies a reduction in the uniformity.
- (4) We place the metapixels in a max-heap. When the metapixel with the largest  $\beta_j$  is moved, the maximum benefit to achieve local uniformity in the spatial distribution is provided.
- (5) To perform the overall regularization, we follow a greedy optimization approach by removing a metapixel from the top of the heap, moving its estimated center  $(x_j, y_j)$  to the average of the centers of its connected neighbors, and updating  $\rho_j$  for its affected neighbors (e.g., all neighbors are also removed from the heap, updated, and reinserted). This process iterates until the largest  $\beta_j$  is too small to provide any benefit or the heap contains only negative  $\beta_j$ 's.

On rare occasions, some of the neighbors of metapixel  $j$  may not be detected in the camera image due to the object's reflectance. In this case, moving the estimated metapixel to the average of the centers of its neighbor might increase the size of the hole created by the undetected neighboring metapixel(s). To avoid this situation, metapixel  $j$  is connected to farther metapixels and the metapixel is moved to a weighted average of the estimated centers of its neighbors. The weights are inversely proportional to the distance between the neighbors in the projector space. If the metapixel lies on an edge or a corner (e.g., it is missing a row or a column of its neighbors), the metapixel's position is not adjusted. Figures 7 and 8 show the results of regularization.

## 5. IMPLEMENTATION DETAILS

**System Setup.** Our setup consists of three 1400 × 1050 Optoma EP910 projectors and a Canon Digital Rebel XTi 10MP camera.

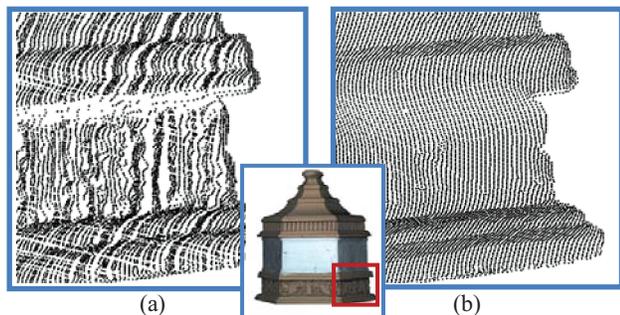


Fig. 7. Regularization. (a) Before regularization, the metapixels have nonuniform spacing. (b) After regularization, the metapixels have a more uniform pixel spacing.

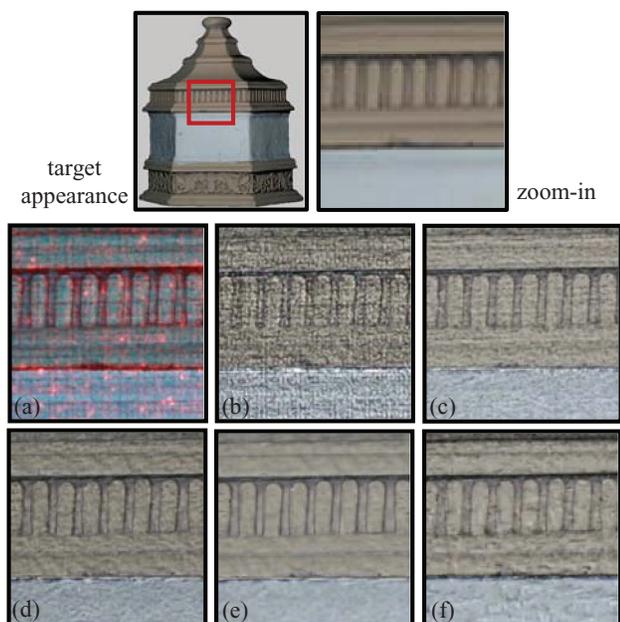


Fig. 8. Comparisons. (a) Computing a naïve inverse light transport optimization results in artifacts. (b) A constrained optimization restricts the pixel values to  $[0,1]$ , but severe noise and graininess still exists. (c) Projector pixel modeling using elliptical Gaussians is added for an improved image. (d) Regularization is added to reduce noise. (e) Smoothness constraints are added for a smooth, noise-free appearance. (f) Using quadrilaterals to model projector pixels results in more noise due to inaccurate projector pixel modeling.

The computer used contains four Dual core 3 GHz processors. The camera-object distance is approximately 1m. The three projectors are placed near the camera also approximately 1m away from the object, oriented directly at the object to achieve superimposed fields-of-projection (see Figure 1(g)). For one object, the total image acquisition time is about 1 hour, the projector-camera calibration process takes 10 minutes, the metapixel fitting optimizations complete in 3 minutes per projector, and the constrained optimization takes 1 hour. Once the base compensation image is computed, it takes less than one second on the CPU to compute the target compensation images for a new appearance.

**Metapixel Estimation.** For the  $M_i$  images, we use  $r = 16$  yielding 256 unique patterns. Metapixels typically span 10–40 camera pixels. We ignore results for metapixels that converged to variances or positions far from the initial guess. Such failures occur for projector pixels falling on a surface fragment at a large grazing angle with respect to the camera or for sharp discontinues on the object’s surface. In these cases we revert to the initial guess obtained from the 3D model. The projector pixel still participates in appearance editing, but its compensation quality is lower. In practice, very few metapixels fail the fitting optimization.

We use HDR imaging for capturing  $W$  in the metapixel acquisition stage during system setup. Since this image is used to compute the amplitude of the metapixels, HDR imaging assures higher accuracy by removing over- and undersaturation artifacts in the captured images. Hence, in Eq. (1) both  $T$  and  $C$  (formed by the amplitude and the base image respectively) are in HDR.

Also, concavities on the object’s surface negatively affect the fitting optimization of the metapixels only when the scale of the concavities is similar to the size of a projector pixel. With one-megapixel projectors, this is not too common for table-top size objects. Larger-scale concavity is illustrated in our examples.

**Base and Target Image Computation.** To obtain base images, we use a simple method for objects that do not deviate drastically from being Lambertian. Since our method does not require the base image to strictly contain only the albedo, we capture  $W$  at different exposures for each projector and form an HDR image for each projector. Then, we create a base image by taking the minimum pixel intensity at every pixel across the projector HDR images. This process removes most specularities which may arise from projector light illumination. To compute target images, we use Blender (<http://www.blender.org>), a 3D content creation suite.

## 6. RESULTS AND DISCUSSION

In this section, we provide some insight into the behavior of our algorithms as well as example appearance edits.

Figure 8 shows the per-algorithm step impact. We start by directly placing the pixel intensity values from the  $M_i$  images into a multi-projector light transport matrix  $T$ . Then, without any projector pixel modeling, denoising, or constraining of the optimization, we calculate  $T^+$  and use the resulting projector compensation images. The results show severe color and noise artifacts (Figure 8(a)). The color artifacts are due to overflow/underflow of projector intensities. Although we can clamp the intensities, the desired ratio of the color channel intensities would not be preserved, resulting in color artifacts. Next, we include the step of constraining the optimization to produce pixel intensity values within  $[0,1]$ . This removes the color clipping artifacts (Figure 8(b)). Subsequently, we incorporate the projector pixel modeling with elliptical Gaussians and obtain a better-quality appearance (Figure 8(c)). Adding priority-based regularization for the metapixels (Figure 8(d)) and smoothness constraints in the optimization of the compensation image results in a smooth and pleasing appearance (Figure 8(e)). As a comparison, we directly model the influence of a projector pixel by the corresponding patch of camera pixels (Figure 8(f)), typically forming a disc or quadrilateral, and find significantly more noise in this alternative solution due to inaccurate pixel modeling.

Figure 9 demonstrates the accuracy of using our parallel algorithm in computing the constrained optimization. We varied the size of  $R_{kl}$  between  $8 \times 8$ ,  $16 \times 16$ , and  $32 \times 32$  pixels for the object shown in Figure 9(a) when solving the constrained optimization for the base target appearance (described in Section 5). Solving the full constrained optimization would equate to a single region

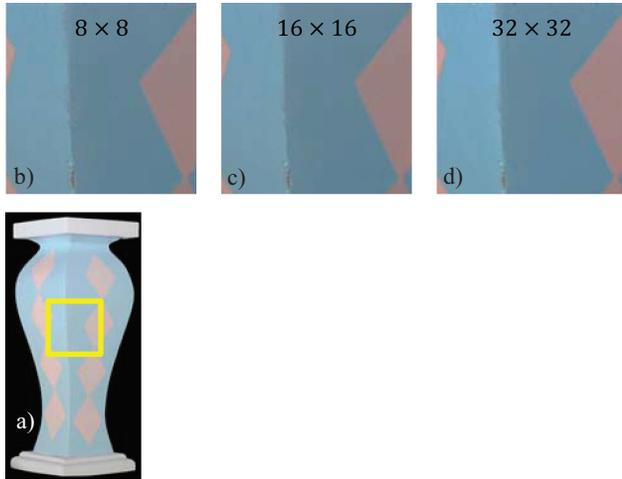


Fig. 9. Region size variation during optimization. We show our parallel algorithm’s accuracy by varying the region  $R_{kl}$ ’s size. No visual artifacts are produced in the resulting base appearance. (a) Base image of patterned pedestal used for the optimization. (b)–(d) Close-ups of a resulting synthetically created base appearance with  $R_{kl}$ ’s size varied amongst  $8 \times 8$ ,  $16 \times 16$ , and  $32 \times 32$  pixels.

encompassing the entire object in camera space. Figures 9(b)–(d) show synthetically created appearance edits for a portion of the base appearance. For all three  $R_{kl}$  region sizes, no visual artifacts arise due to our parallel solving algorithm.

We visualize the residual errors from the constrained optimization to demonstrate the improvement in the ability to achieve a high-resolution appearance as projectors are added to the system. Figure 10(a) shows a photograph of an object (Figures 1(e)–(f)) whose appearance was edited to contain a checkerboard pattern. For each of Figures 10(b)–(d), we show the per-region residual optimization error after computing compensation images via the constrained optimization (red is a large error, blue is a small error). From left to right, the figures use one to three projectors in the optimization. As the number of projectors increases, the overall optimization error decreases. In Figure 10(b), the single projector is at a grazing angle with the right-half of the object resulting in large errors. In Figure 10(c), the added projector is also at a grazing angle with the right half of the object; the errors are mildly reduced. For Figure 10(d), a third projector is positioned to illuminate the right half of the object resulting in an overall reduction of error.

The improved resolution of appearance edits due to adding additional projectors to our system is empirically shown in Figure 11. Here, another checkerboard pattern is applied to the same object used in Figure 9. In Figures 11(a)–(d), we show visualizations of the similarity between the captured image  $C_r$  of the appearance edited object and the target image  $C_t$  (red is high similarity, blue is low similarity). We use a local sharpness metric  $\lambda$  to measure the similarity of  $C_r$  to the high-resolution reference  $C_t$ ,

$$\lambda = \frac{\text{count}[(N)_r \otimes N_r]}{\max(\text{count}(N_t), \text{count}(N_r))}, \quad (9)$$

where  $N_t$  and  $N_r$  are small neighborhoods around a pixel in binary edge-detected  $C_t$  and  $C_r$  images containing at least one checker pattern respectively, and  $\text{count}(N)$  is a function which returns the number of white pixels in a neighborhood  $N$ . Naïvely-generated compensation images (i.e., generated without any projector pixel

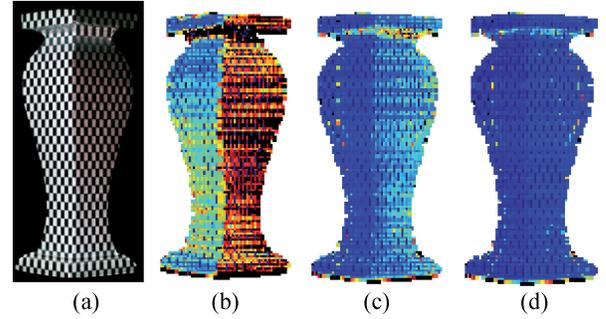


Fig. 10. Visualization of optimization error. (a) Photograph of an object with a checkerboard pattern appearance. (b)–(d) Show the per-region luminance residual error using 1, 2, and 3 projectors respectively (larger error in red). Error reduction with more projectors indicates the ability to provide a higher-resolution target appearance. The average normalized per-region residual error for 1, 2, and 3 projectors is 0.051, 0.014, and 0.007, respectively (e.g., for a 0-255 color channel, the average errors are 12, 5, and 2).

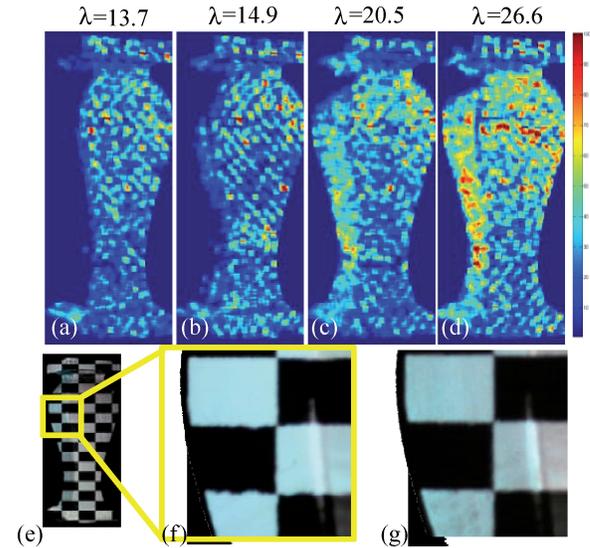


Fig. 11. Visualization of achieved resolution. Visualization of our quality metric (higher means better) for grazing angles using (a) one projector naïve method; (b) one projector our method; (c) three projectors naïve method; and (d) three projectors our method. (e)–(g) Head-on: object photo, one projector naïve method, and three projectors our method.

modeling, smoothness constraints, or metapixel regularization as in Figure 8(b)) and compensation images generated using our method are compared for one and three projector cases (Figures 11(a)–(d)). We show a spatially varying value for  $\lambda$  (normalized to 100), and the results show that using three projectors yields better appearance edits than a single projector. It is also shown that our method produces an improvement over the naïve approach when using either one or three projectors. Figures 11(f)–(g) compare  $C_r$  for a grazing angle case using compensation images generated by a naïve single projector approach and by our method.

An analysis of the effectiveness of our base-scale decomposition scheme for target appearances is shown in Figure 12. In Figure 12(a), we altered a strip of an object’s appearance to that of a horizontal

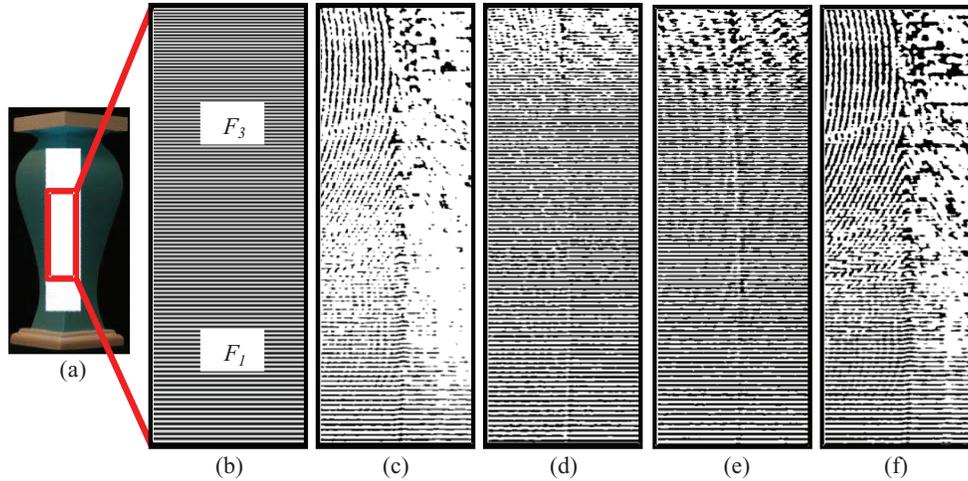


Fig. 12. Base-scale decomposition. (a) Photo of object with sinusoidal pattern projected on top. Figures (b)–(f) are thresholded photographs. (b) Target image with increasing sinusoidal frequency from bottom up. (c) Appearance using 1 projector, full optimization (aliasing starts at  $F_1$ ). (d) Appearance using 3 projectors, full optimization (aliasing starts at  $F_3$ ). (e) Using 3 projectors, base contains a high-pass-filtered version of (b), cutoff set at  $F_1$ . (f) Using 3 projectors, base contains a low-pass-filtered version of (b), cutoff set at  $F_1$ .

sinusoidal pattern (Figure 12(b)) of increasing frequency up the object. Figures 12(c)–(f) are thresholded photographs of the altered object using compensation images computed with different settings. In general, as the sinusoid’s frequency decreases, the ability of our system to recreate the sinusoids decreases, resulting in aliasing artifacts. Figures 12(c)–(d) show the resulting appearance using a full constrained optimization of one and three projectors respectively with no base-scale decomposition (the sinusoidal pattern is used as the target appearance). When only one projector is used, aliasing begins at the limiting frequency content of one projector, labeled as  $F_1$  in Figure 12(c) (which is very close to theoretical maximum frequency content  $F$ ). The projector used is fronto-parallel to the left side of the strip, so a higher resolution is attained on the left side of the strip. With three projectors, the maximum frequency content attained increases, and aliasing begins at  $F_2$  (Figure 12(d)). Based on the sinusoidal pattern, we estimate  $F_2 \approx 1.5F_1$  in this example. Figures 11(e)–(f) show the result of using our base-scale decomposition while varying the base image  $C_b$  to have different maximum levels of frequency content. In Figure 12(e), a high-pass filtered version of the target image (with cutoff set at  $F_1$ ) is used as  $C_b$ . The resulting quality is similar to that shown in Figure 12(d) for the full optimization. If instead  $C_b$  is low-pass filtered (with cutoff set at  $F_1$ ), the higher-frequency content of the appearance is aliased (Figure 12(f)) and the best frequency content achieved more closely resembles the best possible with a single projector. Altogether, this analysis shows that the base-scale decomposition always yields similar quality as the full constrained optimization for frequency content at or below  $F_1$ , thus empirically verifying Section 3. Further, frequency content in the target appearance greater than  $F_1$  is attainable using the base-scale decomposition so long as  $C_b$  has the frequency content above  $F_1$ .

Figure 13 demonstrates the improved quality of our appearance edits by comparing our method against previous multi-projector systems. These photos are subsections of the appearance in Figure 14. Figure 13(a) shows the resulting appearance edit using three superimposed projections but without constraining the pixel intensity values to  $[0, 1]$  when solving the light transport matrix and thus in a manner similar to Wetzstein and Bimber [2007];

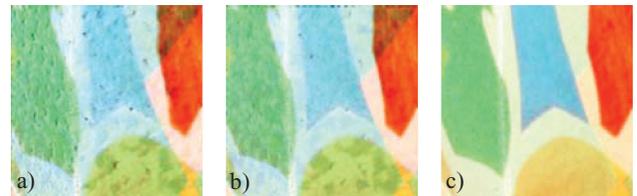


Fig. 13. Comparison against previous systems. (a) Appearance from solving the light transport matrix without constraining pixel intensity values to  $[0, 1]$  (i.e., Wetzstein and Bimber [2007]); visual artifacts are present. (b) Appearance from solving the light transport matrix constraining pixel values to  $[0, 1]$ ; visual artifacts are reduced but remain. (c) Appearance from our method; visual artifacts are mostly removed.

significant visual artifacts are present due to pixel value over/under flowing. By constraining the pixel intensity values to  $[0, 1]$ , the visual quality is improved, but visual errors are still observed (Figure 13(b)). Our system minimizes visual artifacts and achieves more consistent colors (Figure 13(c)).

In Figure 14, we show several altered objects demonstrating various appearances and capabilities of our system. These appearance edits include the effects of novel illumination and geometric and color edits. Some of these effects are dynamic which are better illustrated in our accompanying video. The dynamic appearances are achieved by quickly swapping between computed compensation images for each projector. Target appearances may also include view-dependent effects such as specular highlights and inter-reflections (e.g., Figure 1(b)). While these appearances are static (i.e., no viewer tracking), static view-dependent effects are still compelling as shown in our video. Figures 14(c)–(e) show an example of nonadditive appearance editing where the surface albedo’s color patterns are cancelled in addition to colors being edited. The inset next to Figure 14(c) shows the physical appearance of the patterned pedestal used in this example (i.e., not the same object as in Figure 1), on top of which a new and nonadditive pattern was placed for the target appearance.



Fig. 14. Example appearance editing. (a) Original photograph of a vase. (b) The vase with two leaves added. (c)–(e) Patterned pedestal (photo of original in inset) with three appearances imparted on it. (f) Beethoven model with changed appearance (photo of original in inset). (g) Beethoven model with the head appearing lowered as in a nodding motion, however, the actual object is unchanged.

## 7. CONCLUSIONS

We presented a framework for fast high-resolution appearance editing using multiple superimposed projections. Our solution builds upon multi-projector light transport and solves a constrained optimization which yields an appearance of the highest resolution possible. With our base-scale decomposition, the constrained optimization is done once, and later appearance modifications are done quickly by fast, runtime linear scaling operations. We also use an elliptical Gaussian model to obtain accurate, noise-free estimates of projector pixel centers, sizes, shapes, and interactions. Several compelling appearances are achieved.

### 7.1 Limitations and Future Work

Our approach is not without limitations and several interesting avenues of future work lay ahead. First, there is a loss of contrast as a result of the need to compensate for the weak presence of certain frequencies in the surface’s reflectance function. Hence, an item of future work is using optimal projector locations [Law et al. 2010] and minor target appearance modifications [Law et al. 2011] in order to better accommodate desired alterations. Second, our method is not view dependent, despite yielding some convincing but static view-dependent effects (e.g., Figure 1(b)). While using head-tracking and special viewing devices is an option, we look to other auto-stereoscopic methods to generate compelling view-dependent effects. Third, we are investigating how to support non-Lambertian objects (e.g., using Xu and Aliaga [2009]). Fourth, we are interested in incorporating depixelation and focus improvement algorithms (e.g., Zhang and Nayar [2006]). Fifth, we would like to explore techniques to reduce preprocessing time. In a real-world scenario, errors accumulate due to vibrations and drift; faster preprocessing would improve the practicality of our system.

## APPENDIX: ANALYSIS OF IMAGE DECOMPOSITION

When no base-scale decomposition is applied, the target compensation image  $P_t$  is given by

$$P_t = T^+ C_t = Q C_t,$$

where  $Q = T^+$ . Then, by considering  $Q[i]$  to be the  $i^{\text{th}}$  row of  $Q$ ,

$$P_t = \begin{bmatrix} Q[1] \\ Q[2] \\ \vdots \\ Q[p] \end{bmatrix} C_t \begin{bmatrix} Q[1] \cdot C_t \\ Q[2] \cdot C_t \\ \vdots \\ Q[p] \cdot C_t \end{bmatrix}$$

Assuming objects with no self-occlusion and no self-reflection, we can say that every pixel in the projector affects only one spatially contiguous area in the camera and vice versa. Hence, for both  $T$  and  $Q$ , every pixel  $i$  in  $P_t$  affects a single neighborhood denoted by  $A_i$ . Using  $C_t = C_b \otimes C_s$ , the value of the accurate compensation image at pixel  $i$ ,  $P_t[i]$  is given by

$$P_t[i] = Q[i] \cdot C_t = \sum_{j \in A_i} Q[i, j] C_t[j] = \sum_{j \in A_i} Q[i, j] C_b[j] \otimes C_s[j].$$

Next, when using our approximation,  $P_a$  is given by

$$P_a = \begin{bmatrix} Q[1] \cdot C_b \\ Q[2] \cdot C_b \\ \vdots \\ Q[p] \cdot C_b \end{bmatrix} \otimes P_s.$$

Hence,

$$P_a[i] \left( \sum_{j \in A_i} Q[i, j] C_b[j] \right) \otimes P_s[i].$$

Now, recall that

$$P_s[i] = \sum_{j \in A_i} C_s[j] w[j],$$

where  $w[j]$  are normalized weights (i.e., those of the elliptical Gaussian at that pixel). Thus,  $P_s$  is the image obtained by a low-pass filtering of  $C_s$ .  $P_a$  will be exactly equal to  $P_t$  if there is no loss of information when low-pass filtering  $C_s$ . This can only happen if  $C_s$  is bandlimited by the maximum frequency that can be sampled by one projector. Higher frequencies have to be encoded in  $C_b$ .

## REFERENCES

ALIAGA, D., LAW, A., AND YEUNG, Y. 2008. A virtual restoration stage for real-world objects. *ACM Trans. Graph.* 27, 5.

- ALIAGA, D. AND XU, Y. 2010. A self-calibrating method for photogeometric acquisition of 3D objects. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 4, 747–754.
- BANDYOPADHYAY, D., RASKAR, R., AND FUCHS, H. 2001. Dynamic shader lamps painting on moveable objects. In *Proceedings of the IEEE/ACM International Symposium on Augmented Reality (ISAR'01)*.
- BIMBER, O., CORIAND, F., KLEPPE, A., BRUNS, E., ZOLLMANN, S., AND LANGLOTZ, T. 2005a. Superimposing pictorial artwork with projected imagery. *IEEE Multimedia* 12, 1, 16–26.
- BIMBER, O., AND EMMERLING, A. 2006. Multifocal projection: A multiprojector technique for increasing focal depth. *IEEE Trans. Vis. Comput. Graph.* 12, 4, 658–667.
- BIMBER, O., EMMERLING, A., AND KLEMMER, T. 2005b. Embedded entertainment with smart projectors. *IEEE Comput.* 38, 1, 48–55.
- BIMBER, O. AND IWAI, D. 2008. Superimposing dynamic range. *ACM Trans. Graph.* 27, 5.
- COLEMAN, T. F. AND LI, Y. 1996. A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM J. Opt.* 6, 4, 1040–1058.
- CHUANG, Y., ZONGKER, D., HINDORFF, J., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2000. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proceedings of the ACM SIGGRAPH Conference*. 121–130.
- DAMERA-VENKATA, N. AND CHANG, N. L. 2009. Display supersampling. *ACM Trans. Graph.* 28, 3.
- DEBEVEC, P. AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the ACM SIGGRAPH Conference*. 369–378.
- DURAND, F. AND DORSEY, J. 2002. Fast bilateral filtering for the display of high dynamic range images. In *Proceedings of the ACM SIGGRAPH Conference*. 257–266.
- GARG, G., TALVALA, E., AND LEVOY, M. 2006. Symmetric photography: Exploiting data-sparseness in reflectance fields. In *Proceedings of the Eurographics Workshop on Rendering*. 251–262.
- GROSSBERG, M. D., PERI, H., NAYAR, S. K., AND BELHUMEUR, P. N. 2004. Making one object look like another: Controlling appearance using a projector-camera system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 1. 452–459.
- GRUNDHÖFER, A. AND BIMBER, O. 2006. Real-Time adaptive radiometric compensation. *IEEE Trans. Vis. Comput. Graph.* 14, 1, 97–108.
- LAW, A., ALIAGA, D., AND MAJUMDER, A. 2010. Projector placement planning for high quality visualizations on real-world colored objects. *IEEE Trans. Vis. Comput. Graph.* 16, 6.
- LAW, A., ALIAGA, D., SAJADI, B., MAJUMDER, A., AND PIZLO, Z. 2011. Perceptually-Based appearance modification for compliant appearance editing. *Comput. Graph. Forum* 0, 0.
- MAJUMDER, A. 2005. Is spatial super-resolution possible with multiple overlapping projectors? In *Proceedings of the IEEE International Conference on Audio, Speech and Signal Processing (ICASSP'05)*. Vol. 4, 209–212.
- MALLICK, S. P., ZICKLER, T., KRIEGMAN, D. J., AND BELHUMEUR, P. N. 2005. Beyond lambert: Reconstructing specular surfaces using color. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 619–626.
- NAYAR, S. K., PERI, H., GROSSBERG, M. D., AND BELHUMEUR, P. N. 2003. A projection system with radiometric compensation for screen imperfections. In *Proceedings of the ICCV Workshop on Projector-Camera Systems*.
- NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHY, R. 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM Trans. Graph.* 24, 3, 536–543.
- NG, R., RAMAMOORTHY, R., AND HANRAHAN, P. 2003. All-Frequency shadows using non-linear wavelet lighting approximation. *ACM Trans. Graph.* 22, 3, 376–381.
- OKAZAKI, T., OKATANI, T., AND DEGUCHI, K. 2009. Shape reconstruction by combination of structured-light projection and photometric stereo using projector-camera system. In *Proceedings of the Pacific Rim Symposium on Advances in Image and Video Technology*. 410–422.
- PARK, S. C., PARK, M. K., AND KANG, M. G. 2003. Super-Resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.* 20, 3, 21–36.
- RAU, A., GILL, G., MAJUMDER, A., TOWLES, H., AND FUCHS, H. 2003. PixelFlex2: A comprehensive automatic casually aligned multi-projector display. In *Proceedings of the IEEE International Workshop on Projector-Camera Systems*.
- RASKAR, R., VAN BAAR, J., BEARDSLEY, P., WILLWACHER, T., RAO, S., AND FORLINES, C. 2003. iLamps: Geometrically aware and self-configuring projectors. *ACM Trans. Graph.* 22, 3, 809–818.
- RASKAR, R., WELCH, G., LOW, K. L., AND BANDYOPADHYAY, D. 2001. Shader lamps: Animating real objects with image-based illumination. In *Proceedings of the 12<sup>th</sup> Eurographics Workshop on Rendering Techniques*. 89–102.
- RUZON, M. AND TOMASI, C. 2000. Alpha estimation in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 18–25.
- SEITZ, S. M., MATSUSHITA, Y., AND KUTULAKOS, K. N. 2005. A theory of inverse light transport. In *Proceedings of the IEEE International Conference on Computer Vision*. 1440–1447.
- SEN, P., CHEN, B., GARG, G., MARSCHNER, S. R., HOROWITZ, M., LEVOY, M., AND LENSCH, H. P. A. 2005. Dual photography. *ACM Trans. Graph.* 24, 3, 745–755.
- SEN, P. AND DARABI, S. 2009. Comprehensive dual photography. *Comput. Graph. Forum* 28, 2, 609–618.
- SUMMET, J., FLAGG, M., CHAM, T., REHG, J., AND SUKTHANKAR, R. 2006. Shadow elimination and blinding light suppression for interactive projected displays. *IEEE Trans. Vis. Comput. Graph.* 13, 3, 508–517.
- WANG, J., DONG, Y., TONG, X., LIN, Z., AND GUO, B. 2009. Kernel Nystrom method for light transport. *ACM Trans. Graph.* 28, 3.
- WETZSTEIN, G. AND BIMBER, O. 2007. Radiometric compensation through inverse light transport. In *Proceedings of the 15<sup>th</sup> Pacific Conference on Computer Graphics and Applications*. 391–399.
- XU, Y. AND ALIAGA, D. 2009. An adaptive correspondence algorithm for modeling scenes with strong inter-reflections. *IEEE Trans. Vis. Comput. Graph.* 15, 3, 465–480.
- YANG, R., MAJUMDER, A., AND BROWN, M. 2005. Camera based calibration techniques for seamless multi-projector displays. *IEEE Trans. Vis. Comput. Graph.* 11, 2, 193–206.
- ZHANG, L. AND NAYAR, S. 2006. Projection defocus analysis for scene capture and image display. *ACM Trans. Graph.* 25, 3, 907–915.

Received December 2009; revised December 2010; accepted October 2011