Multiscale Shape and Detail Enhancement from Multi-light Image Collections

Raanan Fattal
University of California, Berkeley

Maneesh Agrawala
University of California, Berkeley

Szymon Rusinkiewicz
Princeton University

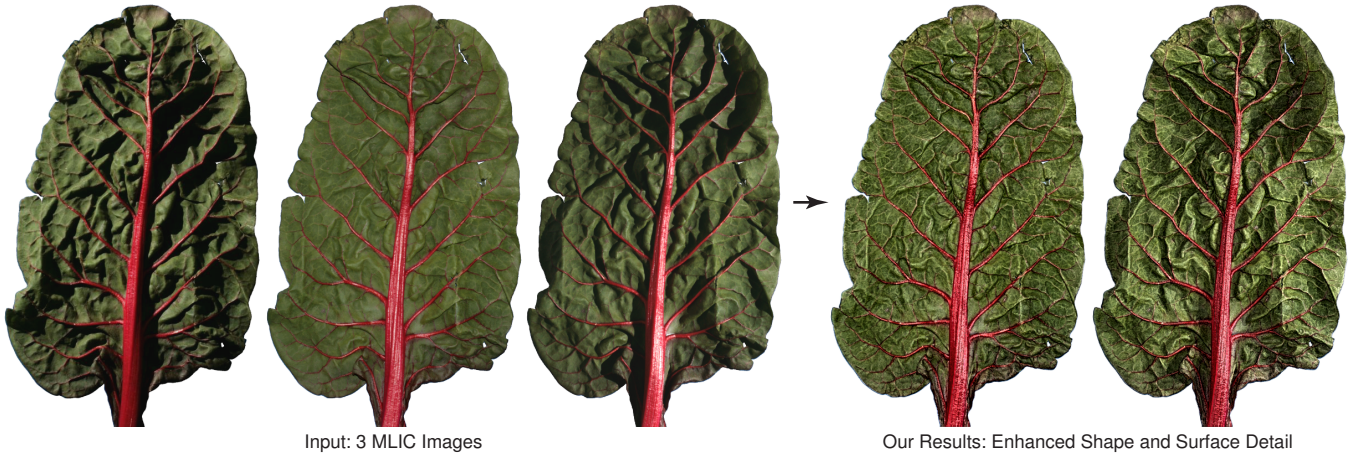


Figure 1: *The Multi-Light Image Collection for this chard leaf contains 3 images taken under varying lighting conditions. The shading in each input image reveals different aspects of its shape and surface details. We combine the shading at multiple scales across the input images to generate the enhanced results. The result on the left exaggerates surface details by eliminating shadows, but yields a flat look. The result on the right is less extreme and includes some shadows to increase the perception of depth, at the cost of reducing some visible detail in the shadow regions.*

Abstract

We present a new image-based technique for enhancing the shape and surface details of an object. The input to our system is a small set of photographs taken from a fixed viewpoint, but under varying lighting conditions. For each image we compute a multiscale decomposition based on the bilateral filter and then reconstruct an enhanced image that combines detail information at each scale across all the input images. Our approach does not require any information about light source positions, or camera calibration, and can produce good results with 3 to 5 input images. In addition our system provides a few high-level parameters for controlling the amount of enhancement and does not require pixel-level user input. We show that the bilateral filter is a good choice for our multiscale algorithm because it avoids the halo artifacts commonly associated with the traditional Laplacian image pyramid. We also develop a new scheme for computing our multiscale bilateral decomposition that is simple to implement, fast $O(N^2 \log N)$ and accurate.

Keywords: image enhancement, NPR, relighting, shape depiction, bilateral filter, multiscale image processing

1 Introduction

Shading is one of the strongest cues for depicting the shape of an object. As a result, artists, illustrators, and photographers commonly manipulate shading to emphasize both the shape and the fine-scale surface details of an object. For example, illustrators will emphasize surface bends by exaggerating the darkening as the surface turns away from the light source. Similarly, photographers will carefully position rim lights to highlight the object's silhouette and place grazing lights to increase self-shadowing due to surface mesostructure. Often the manipulations produce physically inconsistent or impossible shading effects, but nevertheless serve to increase the amount of shape information presented in a single view of the object.

Our key insight is that most of these shading manipulation techniques are based on the same underlying principle;

> ***Local shading adaptation:*** Artists increase local contrast and thereby bring out fine-scale shape details, while simultaneously reducing large-scale jumps in intensity to ensure that the resulting image maps to the full dynamic range of the display device.

While this principle of local adaptation is well known in the context of tonemapping high dynamic range images [Pattanaik et al. 1998; Durand and Dorsey 2002; Fattal et al. 2002; Reinhard et al. 2002; Li et al. 2005], we believe that it applies equally to the problem of manipulating shading to emphasize shape and surface details. The primary difference between tonemapping and non-photorealistic shape depiction is that the latter allows greater flexibility in manipulating image gradients due to shading.

We present a new, image-based technique for enhancing the shape and surface details that is directly based on this principle. The input to our system is a *Multi-Light Image Collection (MLIC)*, which consists of a set of photographs of an object taken from a fixed viewpoint, but under varying lighting conditions. For each input image we compute a multiscale decomposition based on the bilateral filter [Tomasi and Manduchi 1998; Durand and Dorsey 2002]. We construct the enhanced output image by combining detail information from all of the input images at each scale of the decomposition. Unlike earlier image-based relighting techniques, our approach does not require information about light positions or camera calibration. While our approach is similar in spirit to the multiscale exaggerated shading approach of Rusinkiewicz et al. [2006] we do not require any geometric information about the scene. Instead, our approach is purely image-based and acquiring the input MLIC is

relatively fast and easy. We show that we can produce good results with as few as 3 to 5 input images (see Figure 1).

Our technique provides a small set of parameters that allow users to interactively control the level of emphasis in the final image. We also provide default settings for these parameters so that an untrained user such as a botanist could take a small set of images in the field and produce an initial, enhanced result simply by pressing a button. We believe that this makes our approach suitable for quick creation of detailed yet understandable photographic illustrations in medical, technical, and archaeological domains.

The primary contribution of our work is a system for creating such photographic illustrations. In addition, we present two technical contributions. First, we use the bilateral filter to create a multiscale edge-preserving image decomposition and we demonstrate how this decomposition can be used to amplify shading detail while avoiding halo artifacts associated with the traditional Laplacian image pyramid [Burt and Adelson 1983]. Second, we develop a simple, fast $O(N^2 \log N)$, yet accurate approximation of the multiscale bilateral decomposition, using a technique based on Burt [1981] and the dyadic wavelet transform known as the *algorithme à trous* [Holschneider et al. 1989; Mallat 1998].

2 Previous Work

Inspired by artists and photographers, graphics researchers have developed a variety of image-based techniques for emphasizing shape and surface details.

Image-based relighting techniques take an MLIC as input and generate a realistically relit image that enhances the appearance of the scene as output. A common approach is to treat the input image collection as a basis and fit a local lighting model per image pixel. The scene can then be realistically relit under novel illumination conditions [Debevec et al. 2000; Malzbender et al. 2001]. However, these methods rely on knowing the position of the light source in each input image and therefore require specialized acquisition equipment that can be expensive and time-consuming to operate. Another approach is to estimate light positions by either instrumenting the scene with fiducials [Masselus et al. 2002] or using dimensionality reduction techniques [Winnemöller et al. 2005]. Our approach directly recombines the input images to form the output and does not need access to the light positions.

Raskar et al. [2004b] go a step further and use the known incident lighting in a set of multiflash images to extract silhouette edges and produce non-photorealistic images. Malzbender et al. [2001] similarly develop image-based non-photorealistic relighting methods to exaggerate surface shape and details. Given known light positions they reconstruct per-pixel surface normals using shape from shading techniques [Woodham 1980] and then apply *reflectance transformations* that enhance deviations in the normals. They demonstrate how such transformations can enhance images of archaeological artifacts and paintings [Mudge et al. 2005; Mudge et al. 2006]. Yet, such shape from shading techniques put strict requirements on the input images and objects (i.e. known light positions, Lambertian surfaces, etc.) and the resulting reconstructions are often noisy, contain holes, and are underconstrained at depth discontinuities.

Digital image painting interfaces such as Photoshop allow users to locally control contrast, brightness and exposure in the final image. Akers et al. [2003] extend the painting metaphor to the problem of relighting. In their system users manually paint spatially varying mattes to combine images taken under known lighting conditions. Photomontage [Agarwala et al. 2004] and Mohan et al. [2005] use more sophisticated paint-guided optimization procedures for such image-based relighting. An advantage of the two latter systems is that the input images do not need to be taken under known lighting conditions. However, all three of these systems force users to design the output image via painting. While such control may be desirable for trained designers, our goal is to develop a technique with a few simple parameters that is suitable for untrained users.

Our work falls into the general category of image fusion techniques [Ogden et al. 1985; Burt and Kolczynski 1993; Raskar et al. 2004a]. These methods combine multiple images taken under a variety of settings, such as multi-sensor, multi-exposure, multi-focus, etc. Many of these methods are based on the Laplacian image pyramid [Burt and Adelson 1983] or wavelets [Li et al. 1994]. For the purpose of shape and detail enhancement, where frequency bands are heavily modified, these linear band decompositions usually introduce haloing artifacts near edges (see Figure 11). This problem has been recognized in the tonemapping community [Tumblin and Turk 1999] and led to the use of edge-preserving filters to prevent such artifacts in tonemapping [Durand and Dorsey 2002; Choudhury and Tumblin 2003]. Li et al. [2005] work with a linear wavelet-based subband decomposition and avoid halos by limiting the manipulations that are performed at each scale to be spatially smooth (proportionally to the band). In contrast we use a nonlinear multiscale edge-preserving image decomposition which allows us to significantly amplify details at multiple scales without introducing halos.

3 Overview

Our shape and detail enhancement system is composed of two stages; analysis and synthesis (see Figure 2). In the analysis stage we compute a multiscale decomposition for each input image and in the synthesis stage we combine information within each scale of the decomposition, but across all of the input images to generate the enhanced output image. Here we consider the requirements of our input multi-light images and then describe the details of the analysis and synthesis procedures in sections 4 and 5 respectively.

Data Acquisition: The input to our system is a Multi-Light Image Collection in which the viewpoint remains fixed and the scene is static, but the lighting varies from image to image. Ideally we expect each surface in the scene to appear unshadowed in one or more of the input images so that we have samples of each surface under different shading conditions and can extract optimal local contrasts (i.e. surface details). Our approach does not use any information about the position of incident lighting. Thus, we acquire an MLIC by placing the camera on a tripod and taking several images while manually moving a hand-held flash to a new location between each shot. We minimize ambient light to preserve strong contrasts due to illumination from the flash. Although we leave the number of input images and the positioning of the light sources up to the photographer, we have achieved good results with 3 to 5 images; roughly positioning the flash near the camera viewpoint, and then at several points distributed along the upper semicircle around the camera. Our algorithms operate on single channel images $I = \log(Y)$, where Y is the luminance channel of the image in the YUV color space. The logarithmic conversion allows us to use spatial differencing to measure local contrast. Once we have applied our algorithms and constructed the enhanced single channel image, we invert these operations to produce a color image.

4 Analysis

The analysis stage of our system computes a multiscale edge-preserving decomposition for each image in the MLIC based on the bilateral filter [Tomasi and Manduchi 1998; Durand and Dorsey 2002]. The bilateral filter is often used to decompose an image into a base layer that smoothes small changes in intensity while preserving strong edges, and a detail layer that is the difference of the original image and the base layer. Applications of this two-layer decomposition include separating illumination from texture [Oh et al. 2001; Khan et al. 2006], tonemapping [Durand and Dorsey 2002; Choudhury and Tumblin 2003], flash-no-flash image

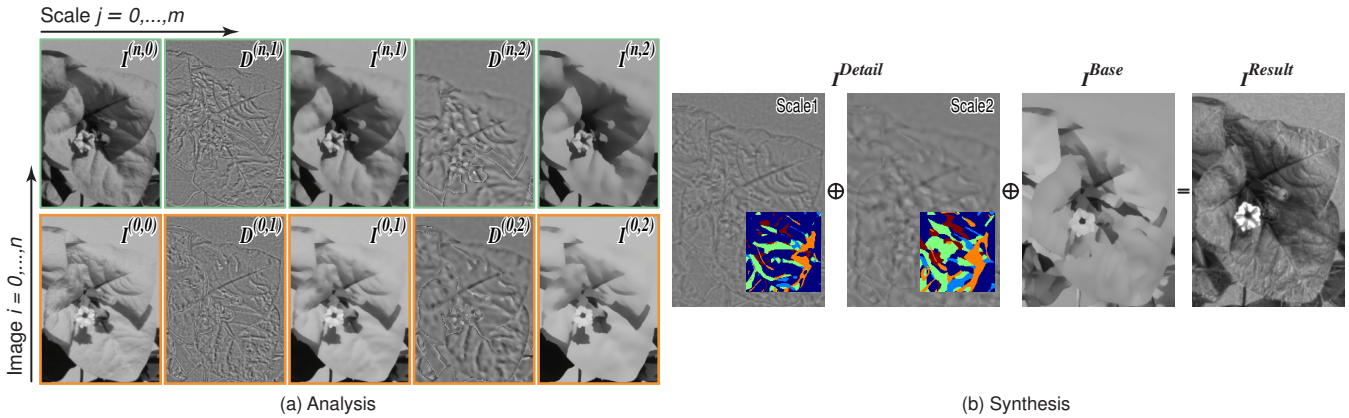


Figure 2: *Analysis and synthesis procedures. (a) We build a multiscale decomposition for each MLIC image. First we recursively apply the bilateral filter to form the filtered images $I^{(i,j)}$, where i indexes the image and j indexes the scale. The difference images $D^{(i,j)}$ are then computed as differences between $I^{(i,j)}$ and $I^{(i,j-1)}$. (b) We generate an enhanced image $I^{Result} = I^{Detail} + I^{Base}$. The detail image is constructed scale by scale as a weighted sum of the input difference images $D^{(i,j)}$. The color visualizations inset at each scale show the $D^{(i,j)}$ that was given the maximum weight for each pixel (though only 2 of 5 input images - orange and green - are shown in part (a)). The base image is constructed to provide coarse scale shading in the enhanced image.*

enhancement [Petschnigg et al. 2004; Eisemann and Durand 2004], video abstraction [Winnemöller et al. 2006] and photographic tone management [Bae et al. 2006]. We extend this two-scale decomposition into a multiscale decomposition in a manner similar to the construction of the Laplacian pyramid [Burt and Adelson 1983].

4.1 Basic Multiscale Bilateral Decomposition

For input image I the goal of the multiscale bilateral decomposition is to first build a series of filtered images I^j that preserve the strongest edges in I while smoothing small changes in intensity. At the finest scale $j = 0$ we set $I^0 = I$ and then iteratively apply the bilateral filter to compute

$$I_p^{j+1} = \frac{1}{k} \sum_{q \in \Omega} g_{\sigma_{s,j}}(\|q\|) \cdot g_{\sigma_{r,j}}(I_{p+q}^j - I_p^j) \cdot I_{p+q}^j, \quad (1)$$

$$\text{with} \quad k = \sum_{q \in \Omega} g_{\sigma_{s,j}}(\|q\|) \cdot g_{\sigma_{r,j}}(I_{p+q}^j - I_p^j), \quad (2)$$

where p is a pixel coordinate, $g_\sigma(x) = \exp(-x^2/\sigma^2)$, $\sigma_{s,j}$ and $\sigma_{r,j}$ are the widths of the spatial and range Gaussians respectively and q is an offset relative to p that runs across the support of the spatial Gaussian. The repeated convolution by $g_{\sigma_{s,j}}$ increases the spatial smoothing at each scale j and we choose the $\sigma_{s,j}$ so that the cumulative width of the spatial Gaussian doubles at each scale. Suppose that at finest scale we set the spatial kernel $\sigma_{s,0} = \sigma_s$. Then setting $\sigma_{s,1} = \sqrt{3}\sigma_{s,0}$ and $\sigma_{s,j} = 2^{j-1}\sigma_{s,j-1}$ for all $j > 1$ we obtain a width of $2^j\sigma_s$ for the cumulative spatial Gaussian[1]. Note however that because the bilateral filter is non-linear, the filtered image I^j is not identical to bilaterally filtering the original input image I with a spatial kernel of width $2^j\sigma_s$.

The range Gaussian $g_{\sigma_{r,j}}$ is an *edge-stopping* function. Ideally if an edge is strong enough to survive one iteration of the bilateral decomposition we would like it to survive all subsequent iterations. To ensure this property we set $\sigma_{r,j} = \sigma_r/2^j$. Reducing the width of the range Gaussian by a factor of 2 at every scale reduces the chance that an edge that barely survives one iteration will be smoothed away in later iterations. We typically set the initial width σ_r to $R/10$, where R is the intensity range of the image.

[1]Convolving two Gaussians of widths σ_a and σ_b respectively yields a new Gaussian of width $(\sigma_a^2 + \sigma_b^2)^{0.5}$. For $j = 1$ we want the cumulative Gaussian to have width $2\sigma_s$. Thus $(\sigma_{s,0}^2 + \sigma_{s,1}^2)^{0.5} = 2\sigma_s$ giving $\sigma_{s,1} = \sqrt{3}\sigma_s$. For $j > 1$ we solve $(\sigma_{s,0}^2 + \sum \sigma_{s,j}^2)^{0.5} = 2^j\sigma_s$ and find that $\sigma_{s,j} = 2^{j-1}\sigma_{s,j-1}$.

We compute a set of detail images as differences between successive levels of these bilateral filtered images $D^j = I^j - I^{j-1}$ for $j = 1...m$. Thus, the I^j retain the strongest edges in the image as preserved by the bilateral filter and the detail layers D^j contain the smaller changes in intensity. We can reconstruct the image I from this decomposition as

$$I = \sum_{j=1}^{m} D^j + I^m. \quad (3)$$

Unlike the more common Laplacian image pyramid [Burt and Adelson 1983], we do not subsample the I^j because such downsampling would blur the edges in I^j. In addition downsampling would prevent the decomposition from being translation invariant and could introduce grid artifacts when the coarser scales are manipulated. In our case, as shown in Figure 2(a), strong edges appear only in the I^j with minimal response in the D^j.

Computational Cost: Because the bilateral filter is non-linear, a brute-force implementation of the 2D convolution costs $O(k^2N^2)$ operations where k is the width of the spatial filter kernel and N is the width of the image. When computing the multiscale decomposition, the cost is dominated by the largest filter kernel, so $k \approx N$ and the total cost becomes $O(N^4)$. Fast approximations to the bilateral filter such as [Weiss 2006; Pham and van Vliet 2005] can significantly reduce this cost but introduce inaccuracies in the computation. An in-depth discussion of these variants can be found in [Paris and Durand 2006]. Recently, Barash [2004] has shown that the bilateral filter can be expressed exactly as a linear convolution in 3D. Based on this idea, Paris and Durand [2006] accelerate the filter by first downsampling the spatial dimension by σ_s^2 and the intensity dimension by σ_r and then convolving with a 3D Gaussian kernel of small fixed size. This method is fairly accurate and because it is independent of the kernel size k, it requires $O(N^2 \log N)$ operations to compute the full multiscale decomposition. While this approach is much faster than the brute-force algorithm, in the next section we introduce a new scheme for computing the multiscale bilateral decomposition that is simpler, requires less memory and is faster in practice than this approach.

4.2 Fast Multiscale Bilateral Decomposition

Burt [1981] has shown that wide Gaussian filters can be well approximated by repeated convolution with so-called *generating kernels* of much smaller support. In fact, these generating kernels have

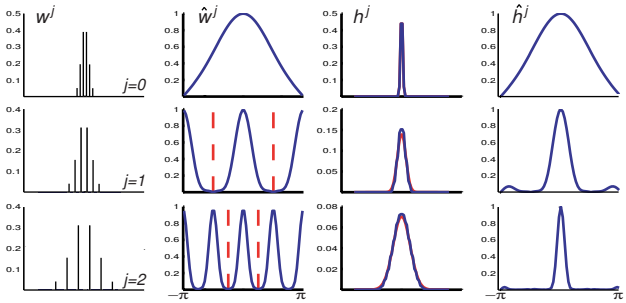


Figure 3: *Three scales of generating kernels w^j, cumulative filters h^j and their Fourier transforms $\hat{w}^j$ and $\hat{h}^j$. Red dotted lines in column 2 indicate smoothing condition boundaries $\pi/2^{j-1} > |\omega| > \pi/2^j$. In column 3 the blue curves representing h^j are within 2% of true Gaussians which are barely visible as red curves.*

a constant number of non-zero entries, and the number of convolutions involved depends logarithmically on the support size of the original Gaussian. More recently this construction has been analyzed in the context of the fast dyadic wavelet transform where it is known as the *algorithme à trous* [Holschneider et al. 1989; Mallat 1998]. We briefly derive and analyze this construction for Gaussian filters and then use it to build a fast multiscale bilateral decomposition. Consider the sequence of convolutions

$$h^{j+1}(p) = h^j(p) * w^j(q), \tag{4}$$

where w^j are the generating kernels, defined by

$$w^j(q) = \begin{cases} g_{\sigma_{s,j}}(\frac{q}{2^j}) & \text{if } \frac{q}{2^j} \in \mathbb{Z} \text{ and } |\frac{q}{2^j}| < k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

As shown in Figure 3, these generating kernels w^j have a fixed number $2k+1$, of non-zero coefficients, yet expand in space by the introduction of $2^j - 1$ zeros between these coefficients. By setting $h^0(p) = \delta(p)$ the h^j correspond to the cumulative filters that are obtained by applying the generating kernels repeatedly. Expanding equation 4 with the expression for w^j in equation 5 we obtain

$$h^{j+1}(p) = h^j(p) * g_{\sigma_{s,j}}(\frac{q}{2^j}) = h^j(p) * g_{2^j\sigma_{s,j}}(q). \tag{6}$$

As in the construction of the basic multiscale bilateral decomposition we choose the $\sigma_{s,j}$ so that the cumulative filter approximates a Gaussian that doubles in width at each scale *i.e.*, $h^{j+1} \approx g_{2^j\sigma_s}$. Solving for $\sigma_{s,j}$ we find that $\sigma_{s,0} = \sigma_s$ and $\sigma_{s,j} = \sigma_s\sqrt{3}/2$ for $j > 0$.

There is one complication in this construction. The generating kernels are essentially sampling Gaussians very sparsely and this introduces aliasing. In Fourier space, the repeated convolutions become

$$\hat{h}^{j+1}(\omega) = \hat{h}^j(\omega)\,\hat{w}^j(\omega) = \hat{h}^j(\omega)\hat{g}_{\sigma_{s,j}}(2^j\omega) = \prod_{k=0}^{j} \hat{g}_{\sigma_{s,k}}(2^k\omega), \tag{7}$$

where the $\hat{}$ notation represents the Fourier transform of the function. As shown in Figure 3, the expansion of the w^j by inserting zeros leads to the introduction of 2^j aliases in $\hat{w}^j(\omega)$. Thus, the generating kernels themselves do not operate as low-pass filters. However, as long as the following *smoothing condition*, $\hat{w}^j(\omega) < \varepsilon$ for $\pi/2^{j-1} > |\omega| > \pi/2^j$ holds, it follows from equation 7 that $\hat{h}^j(\omega) < \varepsilon$ for $|\omega| > \pi/2^j$ and h^j is essentially a low-pass filter. In our case, setting w^j as in equation 5 with $\sigma_s = 1$ we avoid the aliasing and obtain a relative error of less than 2% from the corresponding Gaussian of width $2^j\sigma_s$.

Another source of inaccuracy in this construction is the truncation of the Gaussian kernel to $2k+1$ non-zero coefficients. When the truncation keeps too few coefficients the cumulative filter h^j becomes box-like rather than Gaussian. Moreover, if k is very small, the smoothing condition may not be satisfied and the h^j would no longer act as low-pass filters. In practice we use 5×5 non-zero coefficients in order to approximate a Gaussian with $\sigma_s = 1$.

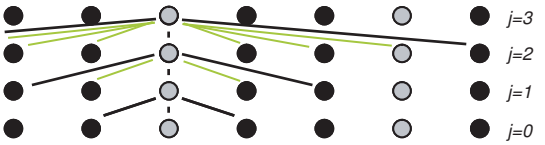


Figure 4: *Four scales of the multiscale bilateral decomposition on a 1D row of pixel intensities. Black lines indicate pixels at scale $j-1$ that are used by the fast algorithm to compute the value of the pixel in column 3 at scale j. In contrast, the basic algorithm also considers all the pixels marked with green lines. At $j = 3$ the basic algorithm averages in the gray pixel in column 6, but the fast algorithm never sees a contribution from that pixel.*

Finally, to compute the fast multiscale bilateral decomposition we replace the $g_{\sigma_{s,j}}$ in equations 1 and 2 with w^j so that

$$I_p^{j+1} = \frac{1}{k}\sum_{q\in\Omega} w^j(q) \cdot g_{\sigma_{r,j}}(I_{p+q}^j - I_p^j) \cdot I_{p+q}^j, \tag{8}$$

and k is modified similarly. Most importantly, we also modify Ω so that the offset q addresses *only* the points where w^j is non-zero, which is independent of j.

In comparison to the basic multiscale bilateral filter this scheme reduces the number of paths along which pixels from the finest level are averaged into the coarser levels of the decomposition. In most cases this reduction only changes the weights at which pixels are averaged and therefore it does not have a significant visual impact on the decomposition. However, as shown in Figure 4, there are scenarios in which this reduction excludes a pixel that should have been included in the coarse scale average.

Computational Cost: Our fast multiscale bilateral decomposition performs a brute-force convolution at each scale. However, because the size of the kernel remains fixed and very small the cost to compute the entire decomposition is $O(N^2 \log N)$. While our technique is asymptotically equivalent to that of Paris and Durand [2006], it provides two practical advantages. First, our method slightly modifies the brute force convolution algorithm and is therefore extremely easy to implement.

The second advantage is that our approach does not require transforming the images into and out of 3D. Therefore, for practical image resolutions our technique is faster and requires less memory. As shown in Table 1 for a 3MP image our approach maintains a constant time of 1.4s for each level. At fine scales $j = 1, 2$, Paris and Durand cannot substantially subsample the 3D space, and their running times are very long. We can pick the fastest method (brute-force or Paris and Durand) at each scale to build a combined strategy. However, the average running time for each scale of this combined approach is 8.2s which is 5.8 times longer than the 1.4s required by our scheme. We ran the same test for 1MP(and 6MP) images and found that the combined approach took an average of 3.5s(16.5s) while our approach took 0.5s(2.8s). All three methods are implemented as unoptimized C code (for Paris and Durand we use the code from their website) running on a 2.1GHz laptop.

In addition, for the 3MP image the peak memory requirement of the combined strategy occurs at $j = 4$ where the Paris and Durand approach requires more than 350 MB. In contrast our approach re-

Method	**Time (in sec) to compute scales** $j = 1...7$						
Brute-Force	1.4	3.6	11.5	39.6	*	*	*
Paris and Durand [2006]	*	*	27.5	15.7	10.5	8.0	6.7
Our Scheme	1.4	1.4	1.4	1.4	1.4	1.4	1.4

Table 1: *Timing comparison of three methods for computing the multiscale bilateral decomposition for a 3MP image with $\sigma_{s,j} = 2^{j-1}$ and $\sigma_{r,j} = 0.1/2^{j-1}$. The *'s represent long computation times > 60s.*



Figure 5: *The β parameter controls the relative contribution of the base and detail images. (a) At $\beta = 3.0$ the flower is very smooth. (b) Reducing β to 0.8 (the default value) and then (c) to 0.2 brings in more detail.*

quires about 50 MB. In terms of accuracy our approach is very similar to that of Paris and Durand. For a 1MP image Paris and Durand obtain peak-signal-to-noise ratios or PSNRs of (53.57, 47.87, 45.98, 45.31, 45.05, 44.91, 44.82) at the seven successive scales of the decomposition while we obtain PSNRs of (321.14, 56.72, 53.63, 50.38, 48.52, 47.67, 47.26). Finally, applying our method on color images (all 3 channels simultaneously) would triple our computation time and memory. Paris and Durand would have to deal with a 5D space and costs would increase substantially.

5 Synthesis

Our synthesis algorithm is designed to enhance shape and surface detail by combining shading information across all of the input MLIC images. We generate the enhanced output image as

$$I^{Result} = I^{Detail} + \beta \cdot I^{Base}, \qquad 0 < \beta \leq 1 \tag{9}$$

where I^{Detail} maximizes the detail at each scale j, across the difference images $D^{(i,j)}$ of all of the input images i. Similarly I^{Base} combines shading information at the coarsest scale $j = m$, across all of the input base images $I^{(i,m)}$. The parameter β is used to trade-off emphasis of the detail image versus the base images (see Figure 5). While we have found that $\beta = 0.8$ is a good default setting, our experience is that the tradeoff between detail and base is easy for novice users to appreciate and manipulate.

5.1 Constructing I^{Detail}

Our goal in building I^{Detail} is to maximize the smaller changes in intensity associated with smooth shading and surface details while minimizing large jumps in intensity. As shown in equation 3, the multiscale bilateral decomposition is designed so that the filtered images $I^{(i,j)}$ contain strong intensity edges while the difference layers $D^{(i,j)}$ primarily contain the smaller changes in intensity. Thus, we compute I^{Detail} as a weighted sum of the difference images,

$$I^{Detail} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} U^{(i,j)} D^{(i,j)}}{\sum_{i=1}^{n}\sum_{j=1}^{m} U^{(i,j)}}, \tag{10}$$

where the weight is computed as

$$U^{(i,j)} = g_{\sigma_d} * e^{(|D^{(i,j)}| - C^{(i,j)})}. \tag{11}$$

Since we wish to maximize small changes in intensity, the weight $U^{(i,j)}$ rewards pixels with large $|D^{(i,j)}|$. However, because the bilateral filter is not a perfect edge preserving filter, remnants of strong edges can also appear in the difference images. To reduce the effect of such remnants the second term in the exponent $C^{(i,j)}$, penalizes pixels containing such edges. This term is given by

$$C_p^{(i,j)} = \frac{\|\nabla I_p^{(i,j)}\|}{\min_q I_{p+q}^{(i,j)} + \varepsilon} \tag{12}$$

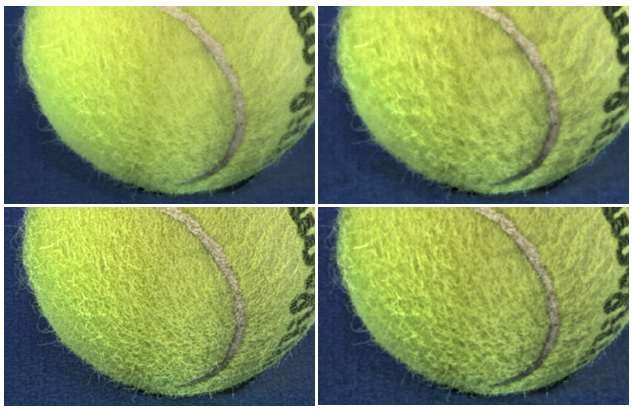


Figure 6: *The parameters λ_{lo},λ_{mid} and λ_{hi} control the detail amplification at different scales. In clockwise order starting at the top left, detail is amplified at all scales equally ($\lambda_{lo} = 1.0, \lambda_{mid} = 1.0, \lambda_{hi} = 1.0$). Next we amplify coarse detail only ($\lambda_{lo} = 0.7, \lambda_{mid} = 1.0, \lambda_{hi} = 1.0$), mid-level detail ($\lambda_{lo} = 1.0, \lambda_{mid} = 0.7, \lambda_{hi} = 1.0$) and fine detail ($\lambda_{lo} = 1.0, \lambda_{mid} = 1.0, \lambda_{hi} = 0.7$).*

where q is a pixel offset that runs over a 3×3 neighborhood and ε is small constant which ensures that the denominator is non-zero. The gradient magnitude in the numerator finds strong edges at each scale, while the denominator further strengthens the term wherever the image is dark. Our assumption is that images are taken under relatively low ambient light so that the darkest pixels are likely to correspond to shadow regions. Finally the spatial convolution with a Gaussian g_{σ_d} serves to locally smooth the weight and reduce noise. We typically set $\sigma_d = 8$.

To further reduce the ratio between the largest and smallest values in I^{Detail} we follow the approach of [Fattal et al. 2002; Li et al. 2005] and perform a scale-dependent dynamic range compression on the difference layers. We replace $D^{(i,j)}$ in equation 10 with

$$sign\left(D^{(i,j)}\right) \cdot |D^{(i,j)}|^{\lambda_j} \qquad 0 < \lambda_j \leq 1 \tag{13}$$

The exponents λ_j control amplification of weak details at each scale. In practice we evenly divide the scales into 3 ranges (lo, mid, hi) and set one exponent for each range $\lambda_{lo}, \lambda_{mid}$ and λ_{hi} (see Figure 6). The default values are $\lambda_{lo} = 0.95$, $\lambda_{mid} = 0.80$, and $\lambda_{hi} = 0.75$. This interface is similar to the visual graphic equalizer of Bae et al. [2006].

5.2 Constructing I^{Base}

The base image I^{Base} sets the coarsest level shading information and can have a strong effect on the overall look of the resulting image. Here we propose two alternative options for constructing the base image. The first allows users to directly combine shading from the input base images $I^{(i,m)}$. While this technique offers the most direct control, the resulting base image will often contain dark shadow regions that reveal very little shape or surface detail. Thus, we present a second technique that avoids this problem by removing shadows from the base image.

User Combined: The simplest approach is to set the base image as a weighted sum of the input base images $I^{(i,m)}$ so that

$$I^{Base} = \frac{\sum_{i=1}^{N} \alpha_i I^{(i,m)}}{\sum_{i=1}^{N} \alpha_i}, \tag{14}$$

where the user specifies the set of weights $\alpha_i, (i = 1...N)$ This option mimics the effect of superpositioning the light sources in the input images. Shadows often provide a strong sense of depth in an image, at the cost of introducing a large intensity difference between the shadowed and lit regions. This approach gives users some control over the strength and location of shadows in the final image.



(a) 1of 5 MLIC Inputs (b) Result: User Combined (c) Result: Robust Max. (d) Robust Max. versus Max.

Figure 7: *The user selected the input image (a) as the only base image to generate the result in (b). While the overall lighting of (a) is preserved, the image is enhanced by detail from all the other MLIC images. In (c) the robust maximum is used to construct the base image. Shadows are minimal making the image appear a bit flatter, but detail is revealed everywhere. In (d) the robust maximum attenuates the specular highlights on the upper leg of this metallic object, while the true maximum accentuates these highlights.*

Robust Maximum: Shadows regions are often much darker than the rest of the image and can make it difficult to see the smaller intensity gradients due to shading and surface detail. One simple way to reduce shadows is to construct I^{Base} by selecting the brightest pixel across the input base images. However, this approach will include oversaturated specular pixels in the resulting base image, which is undesirable because such pixels do not contain useful shading information. Instead, for each pixel p we identify the brightest pixel $b_1(p)$ and the second brightest pixel $b_2(p)$ across all of the the input base images. Rather than always taking the brightest pixel we favor the brightest pixel $b_1(p)$ as long as it is not too much brighter than $b_2(p)$. That is

$$I_p^{Base} = \frac{(b_1(p)+b_2(p)r(p))}{(1+r(p))}, \text{ and } r(p) = \eta\frac{b_1(p)}{b_2(p)}, \qquad (15)$$

where η is between 0 and 1 and is used to switch between using brightest pixel (i.e. the true maximum) and the robust maximum. In practice we set $\eta = 1.0$.

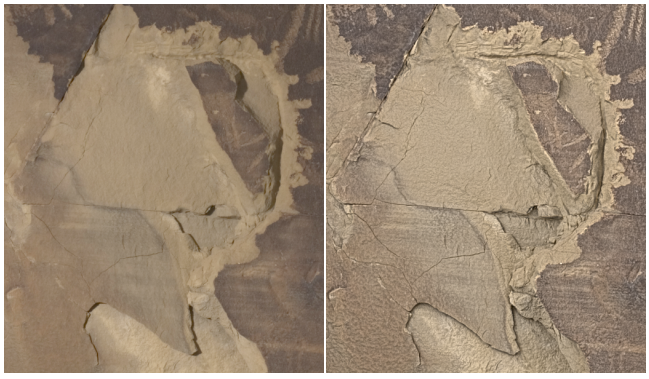
In Figure 7 we compare the user-combined approach for setting the base image to the robust maximum. The robust maximum is designed to reveal detail everywhere in the image by eliminating strong shadows. Thus we use it as the default method for constructing the base image. In Figure 7(d) we compare the robust maximum to using the true maximum. The robust maximum avoids the specular highlights on the upper part of the leg.

6 Results and Discussion

Figures 1, and 5-13 show examples of images generated using our system. In all of these cases we used a 5 level decomposition and applied the the default parameters unless otherwise stated. For all of the matted results we created the mattes by hand.

Our approach is designed to fuse information from all of the source MLIC images into a single enhanced result. For example, in Figure 8 the rough surface texture and crack patterns on the rock face become visible. Similarly in Figure 9 the fur is sharper in our result than in any of the input images. The parameters of our system allow users to control the photorealism of the results. As shown in Figures 1, 5 and 10 details can be exaggerated to give the images an illustrative appearance that resembles the look of botanical illustrations, with emphasis on the venation and surface texture in the petals and leaves.

Although our approach enhances shape and surface details, images that are rich in local contrast may appear somewhat flat compared to the input images. This flatness is due to the reduction of global contrast. However, our system does provide several parameters (the λ's and β) to control the balance between local and global contrast.



(a) 1 of 5 MLIC Inputs (b) Our Result

Figure 8: *Our shape and detail enhancement technique reveals the rough surface texture and crack patterns on this rock surface that are difficult to see in the input images.*

In the results shown in this paper we have opted to strengthen the local contrast to bring out the shape and surface details. We encourage readers to zoom into the images provided in the supplemental materials to evaluate the differences between our results and the input images.

In Figure 11 we show the importance of using the multiscale bilateral decomposition. We create an image using the Laplacian image pyramid instead of our decomposition and see that haloing artifacts are prevalent (white arrows point to halos). Because our decomposition excludes strong edges from the detail layers we can significantly enhance the shading and surface details without introducing such artifacts. In Figure 11(d) we apply Photoshop's unsharp mask filter to the input MLIC image in Figure 11(a). While this approach strengthens global contrast it also introduces halos at strong edges. In Figure 11(e) we apply our method to the same input image and although we cannot fuse detail from other images in this case, our approach is still able to increase detail while preventing halos.

In Figure 12 we compare our results to the gradient based approach suggested by Rasker et al. [2004a]. The max gradient image is computed by selecting the maximum gradient at each pixel across the input images and then reconstructing the image by solving the Poisson equation. An issue with this approach is that selecting the maximum gradient at each pixel individually cannot guarantee any global consistency. For example, the max gradient result contains a light to dark intensity ramp from the bottom to the top of the image (see also the green line in the graph which shows the average intensity across each row of the image). Our results reveal much



Input: 5 MLIC Images | Our Result

Figure 9: *None of the 5 images in the input MLIC individually capture the fine detail of the lion fur. By combining information from all of the input images our approach reveals this fine detail in the fur. Note that we show crops for all five inputs, with a full image for only input image.*



(a) 1 of 5 MLIC Inputs | (b) Our Results: Increasing Detail Enhancement | (c) Hand-drawn illustration

Figure 10: *We use a 5 image MLIC (a) to exaggerate the details of this pink rose. The default parameters with a user-combined base image generates the result image in (b, left). In (b, right) we amplify weak details by lowering λ_{mid} and λ_{hi}, and we reduce β to increase the influence of the detail image. Both results were created using the base image shading of the input image shown in (a). The exaggerated result emphasizes venation and surface texture on the petals much like a botanical illustration. The comparison illustration in (c) is by Catharina Lintheimer (1714-1741) – for more details see http://library.wur.nl/desktop/tulp/konstboeck/*



(a) 1 of 5 MLIC Inputs | (b) Laplacian Pyramid | (c) Our Result | (d) Unsharp Masking | (e) Result: One Input Image

Figure 11: *Comparison of our technique to linear filtering. In (b) we replace our multiscale bilateral decomposition with the Laplacian image pyramid. The resulting image suffers from haloing artifacts (see white arrows). In (c) we use our decomposition with the same synthesis parameters as in (b). (d) Photoshop's unsharp masking filter applied twice to the image in (a), first to enhance large scale detail (amount=50%, radius=50) and then to enhance fine scale detail (amount=200%, radius=4). (e) Our multiscale enhancement algorithm on a single input image.*

more detail than the max gradient approach and because our multiresolution approach is not a global method it does not introduce global inconsistencies such as the intensity ramp. In addition our approach provides control over the detail at each scale and also over the shadows introduced in the base image. However our results are also a bit noisier than the max gradient result.

In Figure 13 we compare our technique to the image-based relighting approach of Akers et al. [2003] and the exaggerated shading technique of Rusinkiewicz et al. [2006]. Akers et al. provide a direct painting interface for relighting and therefore rely on the artistic abilities of the user. While the overall shape of this baboon skull is enhanced, the resulting image lacks fine surface detail. To apply the technique of Rusinkiewicz et al. we first reconstructed surface normals for the skull using 576 MLIC images and their known lighting. The resulting image reveals the surface details but appears very flat and loses its photorealism. In contrast our result maintains a realistic appearance, but enhances the overall shape and surface details. Moreover our approach uses just 6 MLIC input images and does not require known lighting.

One of the limitations of our approach is that it can amplify noise

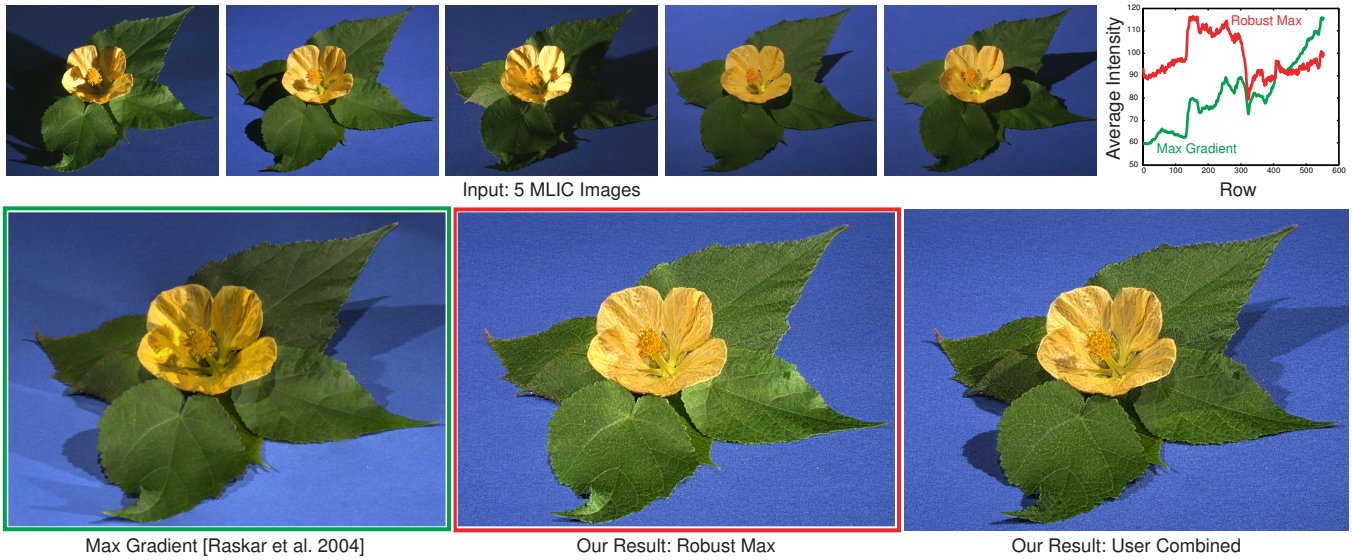


Figure 12: *Comparison of our technique to a gradient-based method [Raskar et al. 2004]. The max gradient result retains all of the shadows from the input images and it introduces a light to dark intensity ramp from bottom to top in the reconstructed image. While our results provide greater detail enhancement and do not introduce the intensity ramp they are also a bit noisier than the max gradient result.*

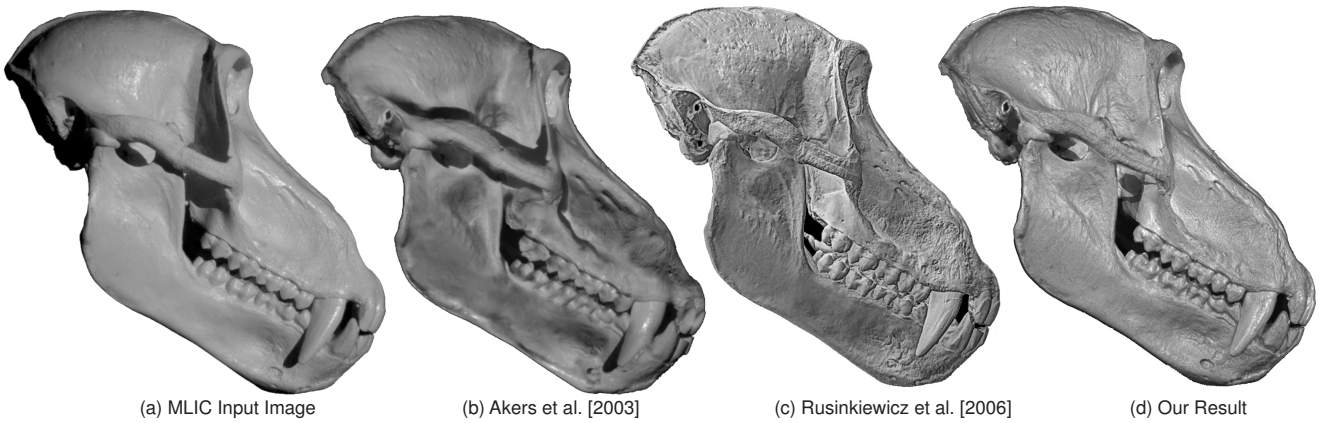


Figure 13: *Comparison to previous techniques. Both Akers et al. [2003] and Rusinkiewicz et al. [2006] use hundreds of source MLIC images with known lighting. Our result required just 6 MLIC images. It was created using the default parameters and the robust maximum base image.*

in addition to the true details in an image. The problem is that both noise and true detail can appear as small changes in image intensity. One way to reduce noise is to significantly reduce the weight of the finest detail layer in our synthesis algorithm. However, this approach would also reduce the true detail at those scales. In the long term, we believe the problem will be solved with more sensitive cameras that can capture the smallest variations in intensity without introducing noise.

A shortcoming of our robust maximum technique for constructing a shadow-free base image is that it assumes every pixel is lit in at least one of the input MLIC images. When a pixel lies in shadow in all the MLIC images it will remain in shadow in the resulting image. We believe that it may be possible to adapt more sophisticated gradient-based shadow removal techniques [Weiss 2001; Finalyson et al. 2006] to produce better shadow-free base images. We leave exploration of these techniques to future work.

7 Conclusions

We have demonstrated a new image-based approach for using a Multi-Light Image Collection to enhance shape and surface details. Our approach is designed so that untrained users can easily generate enhanced images after taking just a few input photographs. The resulting images are detail rich and range in appearance from photorealistic to illustrative. Our approach uses a new multiscale decomposition based on the bilateral filter. Our decomposition makes it possible to combine detail at multiple scales without introducing haloing artifacts. Finally, we have introduced an efficient scheme for incrementally computing this decomposition and shown that it produces very accurate results.

8 Acknowledgments

This work is partially supported by gifts from Adobe Systems Inc. and Apple Inc. Raanan Fattal is supported by the Miller Institute. Szymon Rusinkiewicz is supported by the Sloan Foundation and the National Science Foundation, grant CCF-0347427.

References

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Trans. Graph. (Proc. SIGGRAPH) 23*, 3, 294–302.

AKERS, D., LOSASSO, F., KLINGNER, J., AGRAWALA, M., RICK, J., AND HANRAHAN, P. 2003. Conveying shape and

features with image-based relighting. In *Proc. IEEE Visualization*, 349–254.

BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3, 637–645.

BARASH, D., AND COMANICIU, D. 2004. A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. *Journal of Image and Vision Computing 22*, 1, 73–81.

BURT, P. J., AND ADELSON, E. H. 1983. The Laplacian pyramid as a compact image code. *IEEE Trans. on Comm. 31*, 4, 532–540.

BURT, P., AND KOLCZYNSKI, R. 1993. Enhanced image capture through fusion. *Proc. ICCV*, 173–182.

BURT, P. J. 1981. Fast filter transforms for image processing. *Computer Graphics, Image Processing 6*, 20–51.

CHOUDHURY, P., AND TUMBLIN, J. 2003. The trilateral filter for high contrast images and meshes. In *Proc. Eurographics Symposium on Rendering*, 186–196.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proc. SIGGRAPH*, 145–156.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proc. SIGGRAPH*, 257–266.

EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Trans. Graph. (Proc. SIGGRAPH) 23*, 3, 673–678.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. In *Proc. SIGGRAPH*, 249–256.

FINALYSON, G. D., HORDLEY, S. D., LU, C., AND DREW, M. S. 2006. On the removal of shadows from images. *IEEE Trans. Pattern Anal. Mach. Intell. 28*, 1, 59–68.

HOLSCHNEIDER, M., KRONLAND-MARTINET, R., MORLET, J., AND TCHAMITCHIAN, P. 1989. A real-time algorithms for signal analysis with the help of the wavelet transform. *Wavelets, Time-Frequency Methods and Phase Space*, 286–297.

KHAN, E. A., REINHARD, E., FLEMING, R. W., AND BÜLTHOFF, H. H. 2006. Image-based material editing. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3, 654–663.

LI, H., MANJUNATH, B., AND MITRA, S. 1994. Multi-sensor image fusion using the wavelet transform. *Proc. ICIP 1*, 51–55.

LI, Y., SHARAN, L., AND ADELSON, E. H. 2005. Compressing and companding high dynamic range images with subband architectures. *ACM Trans. Graph. (Proc. SIGGRAPH) 24*, 3, 836–844.

MALLAT, S. 1998. *A Wavelet Tour of Signal Processing*. Academic Press.

MALZBENDER, T., GELB, D., AND WOLTERS, H. 2001. Polynomial texture maps. In *Proc. SIGGRAPH*, 519–528.

MASSELUS, V., DUTRÉ, P., AND ANRYS, F. 2002. The free-form light stage. In *Proc. Eurographics Symposium on Rendering*, 247–256.

MOHAN, A., TUMBLIN, J., BODENHEIMER, B., GRIMM, C., AND BAILEY, R. 2005. Table-top computed lighting for practical digital photography. In *Eurographics Symposium on Rendering*, 165–172.

MUDGE, M., VOUTAZ, J.-P., SCHROER, C., AND LUM, M. 2005. Reflection transformation imaging and virtual representations of coins from the Hospice of the Grand St. Bernard. In *Proc. VAST*.

MUDGE, M., MALZBENDER, T., SCHROER, C., AND LUM, M. 2006. New reflection transformation imaging methods for rock art and multiple-viewpoint display. In *Proc. VAST*.

OGDEN, J., ADELSON, E., BERGEN, J., AND BURT, P. 1985. Pyramid-based computer graphics. *RCA Engineer 30*, 5, 4–15.

OH, B. M., CHEN, M., DORSEY, J., AND DURAND, F. 2001. Image-based modeling and photo editing. In *Proc. SIGGRAPH*, 433–442.

PARIS, S., AND DURAND, F. 2006. A fast approximation of the bilateral filter using a signal processing approach. Tech. Rep. MIT-CSAIL-TR-2006-073.

PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proc. SIGGRAPH*, 287–298.

PETSCHNIGG, G., SZELISKI, R., AGRAWALA, M., COHEN, M., HOPPE, H., AND TOYAMA, K. 2004. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph. (Proc. SIGGRAPH) 23*, 3, 664–672.

PHAM, T., AND VAN VLIET, L. 2005. Separable Bilateral Filtering for Fast Video Preprocessing. *Proc. ICME*, 454–457.

RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. *Proc. NPAR*, 85–152.

RASKAR, R., TAN, K.-H., FERIS, R., YU, J., AND TURK, M. 2004. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Trans. Graph. (Proc. SIGGRAPH) 23*, 3, 679–688.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. In *Proc. SIGGRAPH*, 267–276.

RUSINKIEWICZ, S., BURNS, M., AND DECARLO, D. 2006. Exaggerated shading for depicting shape and detail. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3, 1199–1205.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *Proc. ICCV*, 839–846.

TUMBLIN, J., AND TURK, G. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *Proc. SIGGRAPH*, 83–90.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In *ICCV*, 68–75.

WEISS, B. 2006. Fast median and bilateral filtering. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3, 519–526.

WINNEMÖLLER, H., MOHAN, A., TUMBLIN, J., AND GOOCH, B. 2005. Light waving: Estimating light positions from photographs alone. *Computer Graphics Forum 24*, 3, 433–438.

WINNEMÖLLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-time video abstraction. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3, 1221–1226.

WOODHAM, R. 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering 19*, 1, 139–144.