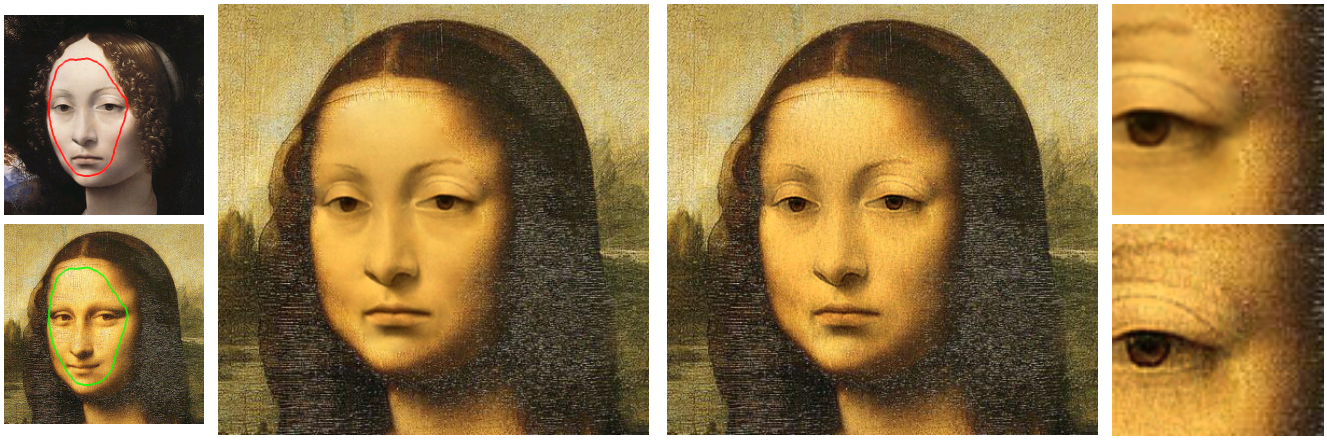
Multi-scale Image Harmonization

Kalyan Sunkavalli* Harvard University
Micah K. Johnson† MIT
Wojciech Matusik‡ Disney Research, Zurich
Hanspeter Pfister§ Harvard University



(a) Source / Target (b) Seamless Cloning (c) Harmonization (d) Close-ups

Figure 1: *In traditional image compositing (a) a user applies geometric transformations to a source image (top) and inserts it into a target image (bottom). Tools such as the Photoshop Healing Brush use gradient domain compositing to ensure that the composite is seamless (b) but the inconsistencies between the two images, make the result look unrealistic: the inserted face is much smoother than the rest of the image. Our method "harmonizes" the images before blending them, producing a composite that is seamless and realistic (c). The close-up images (d) compare traditional gradient-domain blending (top) to the harmonized result (bottom).*

Abstract

Traditional image compositing techniques, such as alpha matting and gradient domain compositing, are used to create composites that have plausible boundaries. But when applied to images taken from different sources or shot under different conditions, these techniques can produce unrealistic results. In this work, we present a framework that explicitly matches the visual appearance of images through a process we call *image harmonization*, before blending them. At the heart of this framework is a multi-scale technique that allows us to transfer the appearance of one image to another. We show that by carefully manipulating the scales of a pyramid decomposition of an image, we can match contrast, texture, noise, and blur, while avoiding image artifacts. The output composite can then be reconstructed from the modified pyramid coefficients while enforcing both alpha-based and seamless boundary constraints. We show how the proposed framework can be used to produce realistic composites with minimal user interaction in a number of different scenarios.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques I.4.3 [Image Processing and Computer Vision]: Enhancement—Filtering;

Keywords: Image compositing, alpha matting, gradient-domain compositing, Poisson blending, image pyramids, visual appearance transfer

*e-mail: kalyans@eecs.harvard.edu
†e-mail: kimo@csail.mit.edu
‡e-mail: matusik@disneyresearch.com
§e-mail: pfister@seas.harvard.edu

1 Introduction

Combining regions of multiple photographs or videos into a seamless composite is a fundamental problem in many vision and graphics applications, such as image compositing, mosaicing, scene completion, and texture synthesis. In order to produce realistic composites, it is important to ensure that the boundaries between the images being combined appear as seamless and natural as possible. This can be achieved through alpha matting, where pixel values are combined using a user-specified alpha matte, or through gradient-domain compositing techniques, which reconstruct pixel intensities from merged gradient vector fields.

While necessary, seamless boundaries are not always sufficient for creating realistic composites. Often the images being combined come from diverse sources and are shot by different cameras under different conditions. This is illustrated in Fig. 1a, where the user segments a novel face (top), and inserts it into another image (bottom). Gradient domain compositing (Fig. 1b) creates seamless boundaries in the composite. But because the two images are from different sources with different appearance, the two regions of the composite look inconsistent, detracting from the realism of the composite.

Currently, users fix these inconsistencies manually, and it takes even professional artists hours of work to produce highly realistic composites. In this paper, we address this problem by building tools to automatically *harmonize* images before compositing them (Fig. 1c). By building methods to automatically correct inconsistencies in images with minimal user interaction, this work takes the burden of compensating for inconsistencies away from the user and makes compositing effortless and user-friendly.

The main contribution of this work is a unified framework that harmonizes aspects of appearance, such as contrast, texture, noise, and blur. This is guided by the insight that a multi-resolution pyramid representation for images is useful for both transferring different aspects of visual appearance between images and compositing them. We show that we can transfer appearance by manipulating the different levels of the pyramid of the source and target images so that their histograms match. We also present a novel method to reconstruct the composite from the modified pyramids in conjunction with boundary constraints based on matting as well as gradient-domain compositing. To our knowledge, this is the first work that explicitly addresses the problem of harmonizing images during compositing.

This work does not deal with inconsistencies in viewpoint, lighting, or shadows. We assume that the images have been geometrically aligned and have compatible viewpoint and vanishing points.

2 Previous Work

Alpha Matting The simplest way to fuse images is to combine their absolute pixel values. This is often accomplished through alpha matting [Porter and Duff 1984], where the colors of the images are linearly interpolated using weights specified by the alpha matte. Recent work in this area has focused on making the matte creation as easy as possible [Wang et al. 2007; Sun et al. 2004], but has not corrected for appearance differences.

Gradient-Domain Compositing Often two images need to be merged *seamlessly*, i.e., the boundary between them should be imperceptible. Gradient-domain techniques accomplish this by combining image gradients (instead of absolute pixel values) and solving for the composite that would best produce the fused gradient field. These techniques were introduced to the imaging community by Pérez et al. [2003] and have since become the standard for seamless compositing [Agarwala et al. 2004; Levin et al. 2004] and a part of editing tools such as Photoshop [Georgiev 2004]. Perez et al. also propose variations of seamless cloning (such as mixing the source and target gradients) to handle differences in texture, but these solutions work only on very specific images. More recently, Farbman et al. [2009] showed that the solution to the Poisson linear system could be approximated using a novel interpolation scheme. This work did not consider issues related to harmonization of the source images, but did show that large image regions could be cloned at interactive rates. In general, our method extends gradient-domain techniques by reconstructing images from a much larger set of filter outputs and integrates harmonization into the compositing framework.

Transfer of Visual Appearance Most of the work on transferring visual appearance focuses on matching color distributions between images [Reinhard et al. 2001; Pitie et al. 2005; Lalonde and Efros 2007]. Cohen-Or et al. [2006] presented ways to transform images such that their color palettes are perceptually harmonic. Closely related to our work, is the work of Bae et al. [2006] on transferring tonal balance and level of detail from one image to another. They use a nonlinear bilateral filter to decompose the images into two scales and match the histograms of these scales to match the style of the images. We show that we can achieve similar effects with

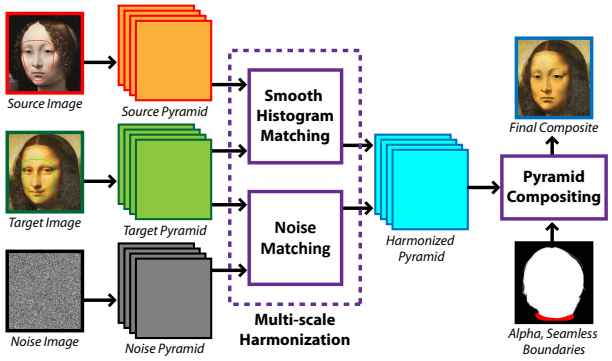


Figure 2: *An overview of the Multi-scale Image Harmonization framework. The input source and target images, and a uniform random noise image are decomposed into pyramids. Using a smooth histogram matching technique, the source and noise pyramids are iteratively shaped so that they match the target pyramid. This produces a harmonized pyramid from which the final composite is reconstructed by incorporating seamless and/or matte-based boundary conditions.*

linear filters and do this in the context of image compositing. Chen et al. [2009] present an interactive tool for separating the noise from an image; this noise can then be transferred to other images. In contrast, our approach automatically matches noise, contrast and blur using a single framework.

Multi-scale Methods Our paper is inspired by Burt and Adelson's seminal work [1983b] on using multi-scale representations such as Laplacian pyramids [1983a] to composite images. The statistics of each level of an image pyramid are known to be correlated with different aspects of visual appearance and pyramid based representations have been widely used for many problems in vision and graphics including texture analysis and synthesis, object recognition and image retrieval, and transferring visual appearance. In all these works, images are decomposed into multi-scale pyramids and the different levels of the pyramids are then analyzed or manipulated to achieve the desired objective. A classic example of this approach is the work of Heeger and Bergen [1995] who use pyramids for texture synthesis, and show that histogram matching the subband coefficients of a noise pyramid to those of a given texture can be used to generate synthetic stochastic textures.

A known problem with pyramids constructed using linear filters, is that applying nonlinear operations (such as tone-mapping and histogram matching) on the subband coefficients of images with structure often results in artifacts such as haloing along strong edges. As a result, recent work on multi-scale methods uses nonlinear edge-preserving filters like the bilateral filter [Tomasi and Manduchi 1998] to construct the pyramids [Bae et al. 2006; Fattal et al. 2007; Farbman et al. 2008] and avoid haloing. In contrast to this, Li et al. [2005] show that linear multi-scale decompositions used in conjunction with carefully controlled, smooth nonlinear operations (in their case, compressive transforms for high dynamic range tone mapping) do not lead to haloing artifacts.

Our work builds on previous uses of linear image pyramids in three ways. Firstly, we harmonize the appearance of the source and target images by histogram-matching the pyramid coefficients of the target to those of the source. Doing this naively could lead to artifacts but we show how regularizing the histogram transfer can minimize these artifacts. Secondly, we inject noise into the harmonization step and show how it can be shaped to handle differences in the noise and texture patterns between images. Finally, we introduce

(a) Input images (Source / Target) (b) Naive histogram matching (c) Smooth histogram matching (d) Smooth histogram and noise matching (e) Pyramid compositing with seamless boundaries

Figure 3: *In this example compositing scenario, the user clones a flat photograph ((a) top), onto a high-contrast and textured image ((a) bottom). Using naive histogram matching to modify the target subbands produces a result with blotches and haloing near strong edges (b). Using smooth histogram matching improves the result but the noise does not match the target image (c). Injecting noise into the harmonization resolves this (d). Finally, reconstructing the composite from the harmonized pyramid by enforcing seamless boundary conditions produces a highly realistic result (e). Photo credit: Flickr user Steve Wampler/*`Steve Wampler` *((a) top) and Starstock / Photoshot ((a) bottom).*

a novel way of computing the final composite from the histogram-matched pyramid coefficients by solving a linear system of equations while satisfying both seamless and matte based boundary conditions.

3 Overview

We assume that the user has a source image I^s with an object, or region, that they would like to insert into a target image I^t. The object in the source image may have different visual characteristics from objects in the target image, and our goal is to harmonize these characteristics to create a more compelling composite.

At a high level, we begin by building pyramids from the source and target images. We also synthesize a uniform random noise image and build a pyramid from the noise image. Next, we modify the source and noise pyramids to match the target pyramid – a process that harmonizes the images. Finally, we reconstruct the composite from the harmonized source and noise pyramids taking into account the appropriate boundary conditions (both alpha and seamless boundaries). An overview of this process is shown in Fig. 2. In this section, we provide an overview of our framework and in the sections that follow, we discuss each component in detail.

Our compositing framework uses a multi-resolution pyramid representation for all images. The pyramid is constructed by filtering each image with a set of n linear filters, f_1 to f_n; we use Haar filters. For a source image I^s and target image I^t, the subbands are:

$$\begin{aligned} B_i^s &= f_i \star I^s \\ B_i^t &= f_i \star I^t \ . \end{aligned} \tag{1}$$

A standard separable n-level pyramid has three subbands at every level in addition to a lowpass residue subband for a total of $3n+1$ subbands. Each level of the pyramid representation is created by filtering an image with three filters of the same scale. The statistics of pyramid subbands are known to be closely related to image appearance – a property that has been exploited in work on texture synthesis [Heeger and Bergen 1995; Portilla and Simoncelli 2000]. This makes the pyramid an ideal representation for us, and we harmonize the images by transforming the source subbands in a way that matches their statistics to those of the target subbands.

The main tool for modifying the source subbands in order that their statistics are similar to the target subbands is histogram matching [Heeger and Bergen 1995]. The harmonized subbands coefficients B_i^h can be computed as

$$B_i^h = \text{histmatch}(B_i^s, B_i^t), \tag{2}$$

where histmatch() denotes the transfer function that matches the histogram of B_i^s to that of B_i^t.

While the simple operation in Eqn. 2 is a powerful tool for matching the appearance of images, there are two fundamental problems with it. First, naive histogram matching is a nonlinear operation that distorts the shape of the subbands, and images reconstructed from these modified subbands often suffer from artifacts such as haloing along strong edges and the amplification of noise and blocking artifacts. For example, Fig. 3 shows different approaches to transferring the appearance of an older high-contrast and textured photograph to a newer flat and smooth photograph. Fig. 3b is the result of direct histogram matching – the gradients in the original source image have been over-sharpened and there are haloing artifacts near strong edges. Our smooth histogram matching technique – described in Sec. 4 – minimizes these artifacts by ensuring that the histogram matching process does not distort the shape of the subbands substantially (Fig. 3c).

The second problem with a direct application of Eqn. 2 relates to image noise. Natural images often have noise due to the camera, such as sensor and ISO noise, or due to compression, such as JPEG quantization noise. In addition, the target images might have textures that are missing in the source images. If the noise and texture patterns in the source and target images differ significantly, histogram matching the subbands alone will not harmonize them. To better model these differences, we introduce a noise term to our harmonization framework. In other words, we assume that the harmonized subbands we want to estimate are given by a sum of the

(a) Source image (b) Naive histogram matching (c) Regularized gain (d) Regularized gain and blending (e) [Bae et al. 2006]

Figure 4: *We would like to give the source image, the tulip photograph from Bae et al. (a), the appearance of Ansel Adams'* Clearing Winter Storm *(see Bae et al. [2006] Fig. 2a). Using naive histogram matching produces a result with haloing (b). Regularizing the gain removes these artifacts (c), but some of strong edges have been over-amplified. Blending in the source at these edges removes these problems producing a result (d) with the tones from the model image. The technique of Bae et al. [2006] (e) exaggerates these effects for a more stylized result.*

structure subbands B_i^h and *noise* subbands N_i^h, i.e.,

$$T_i^h = B_i^h + N_i^h. \tag{3}$$

Our intuition is that the structure components B_i^h can be estimated by shaping the source subbands to match the target subbands, while the noise components N^h can be estimated by shaping a noise image to match only the noise in the target subbands. Our harmonization step – covered in detail in Sec. 5 – does this iteratively to produce a set of harmonized subband coefficients that exhibit the properties we desire in the source image, including the appropriate contrast, texture, noise and blur (Fig. 3d).

The final harmonized image can be reconstructed from the modified pyramid coefficients T_i^h by collapsing the pyramid, i.e., applying synthesis filters (the inverse of the filters applied in Eqn. 1) and summing the results. There are fast and efficient algorithms to do this without explicitly solving the linear system of equations corresponding to Eqn. 1. However, to composite regions of the source image into the target image, we need to ensure that boundaries are appropriately handled and simply collapsing the pyramid will not satisfy the desired boundary constraints. Instead, for image compositing, we reconstruct the final composite I^h by solving a linear system of equations:

$$\mathbf{F}I^h = T^h - c\,, \tag{4}$$

where the matrix $\mathbf{F}$ contains the filters used to construct the pyramid, the vector T^h contains the harmonized subband coefficients, and the vector c specifies boundary constraints. In Sec. 6 we discuss how we set up this linear system and how c can be used to specify both seamless and alpha matting boundary constraints. While this linear system can be very large even for small images, we show how it can solved quickly and accurately using a quadtree subdivision.

4 Smooth Histogram Matching

As shown in Figs. 3 and 4, applying histogram matching naively on subband coefficients leads to haloing and the amplification of artifacts. Instead, we model histogram matching as a gain control that boosts or reduces subband coefficients depending on their magnitudes, and regularize it to avoid artifacts.

We first match the histograms of the source subbands B_i^s to the histograms of the target subbands B_i^t using Eqn. 2. To ensure that we modify the subband coefficient magnitudes without changing their sign, we apply the histogram matching on the absolute values of the coefficients and retain the sign. Matching the histograms produces the modified subbands B_i^{hist}.

The effect of the histogram matching can be modeled as a multiplicative gain that, in logarithmic units, is given as:

$$g_i(|B_i^s|) = \log(|B_i^{hist}|) - \log(|B_i^s|)\,. \tag{5}$$

A positive gain indicates an increase in the coefficient magnitude, i.e., the histogram matching enhanced detail in the source image, whereas a negative gain represents a decrease in the coefficient magnitude, i.e., the histogram matching dampened the detail. Up to this point, multiplying the source subband coefficients B_i^s by the gain function $\exp(g_i(|B_i^s|))$ recovers the histogram matched subbands B_i^{hist} perfectly.

In practice, three techniques help mitigate visible artifacts introduced by manipulating subband coefficients. The first is to use undecimated, or oversampled, pyramids; i.e., the subbands of the pyramid are not downsampled after filtering and are the same size as the original image [Li et al. 2005]. While pyramids based on any set of linear filters could be used to construct the pyramids, we use oversampled Haar pyramids [Gonzalez and Woods 2001] because of their ease of implementation.

The second method to minimize artifacts is to avoid large values in the gain function and we do this by controlling the maximum gain applied:

$$\hat{G}_i = \exp\left(\frac{\delta_k}{\|g_i\|_\infty} g_i\right). \tag{6}$$

Here δ_k indicates the maximum allowed gain for the subbands at level k and $\|g_i\|_\infty$ denotes the maximum value of g_i. δ_k controls the distortion that will be allowed in the subbands and is set to 1.5.

Finally, the third method to minimize artifacts is to ensure that the gain is spatially smooth and does not distort the shape of the subbands excessively. As in Li et al. [2005], we do not apply the computed gain map directly to the subband coefficients. Instead, at every level of the pyramid k, we compute an activity map that represents local coefficient magnitude by pooling all the rectified subbands (i.e., absolute values of the subband coefficients) at that level and blurring with a Gaussian:

$$\begin{aligned} A_k^s &= N(\sigma) \star \sum_{i \in \text{lev}(k)} |B_i^s|\,, \\ A_k^t &= N(\sigma) \star \sum_{i \in \text{lev}(k)} \left|B_i^t\right|. \end{aligned} \tag{7}$$

The parameter σ controls the width of the Gaussian N and it increases by a factor of two between levels with the value at the finest scale set to 4.

Since the activity maps are blurred, they are spatially smooth. Applying the gain function of Eqn. 6 to the activity maps thus produces a gain map $\hat{G}(A_k^s)$ that varies smoothly and does not distort the shape of the subbands excessively. The smooth histogram transfer for subband B_i^s is then given by:

$$B_i^h = m_i \hat{G}(A_k^s) \times B_i^s \,, \tag{8}$$

where m_i is a scaling factor related to the level of the pyramid and linearly reduces from 1.0 at the finest scale to 0.45 at the coarsest scale. Eqn. 8 describes the function that drives all the histogram matching operations we perform on subbands.

Regularizing the gain eliminates most of the artifacts from naive histogram matching. However, repeatedly manipulating pyramid coefficients in each iteration, might over-amplify strong edges in some cases. To avoid this, we compute an aggregate activity map:

$$A_{ag}^s = \sum_{k=1}^{m} A_k^s, \tag{9}$$

and convert it into an alpha map that is clamped to 0 at the 85^{th} percentile and 1 at the 95^{th} percentile, and varies linearly in between. We use this alpha map to blend the harmonized pyramid B^h with the original pyramid B^s. Since the activity maps are highest near strong edges, the blending removes over-amplified edges from the harmonized pyramid (Fig. 4d).

5 Structure and Noise Matching

As mentioned in Sec. 3, a composite will fail to look realistic if the noise pattern of the source image does not match the background in the target. We also found that histogram matching cannot successfully create noise to match a target image if the source image is too clean. To better match noise in the composited region, we inject noise into the harmonization process.

Let T_i^s represent the sum of the source subband and the corresponding noise subband, $T_i^s = B_i^s + N_i^s$. Similarly the harmonized subbands we wish to estimate T_i^h are also a sum of structure components and noise components. Following Eqn. 8, we construct a gain map $\hat{G}_b$ by matching the histogram of the summed source subbands to the target image.

For the noise subband, we construct a gain map, $\hat{G}_n$, designed specifically to shape the noise. We high-pass filter the target image to isolate the noise image I^n and construct a target noise pyramid N^t. This noise will also contain components of the image structure and cannot be used directly. Instead we assume that the noise components are more prominent in low-activity regions of the target image and we identify these by thresholding the target aggregate activity map as:

$$\Omega = A_{ag}^t < \text{percentile}(A_{ag}^t, \beta). \tag{10}$$

A_{ag}^t is computed by applying Eqn. 9 to the target image, and β is a user-specified parameter that enables us to differentiate between structure and the noise in the target image. We construct the gain map $\hat{G}_n$ using the process described in Sec. 4 by histogram matching the subbands N_i^s to the target noise pyramid subbands N_i^t, but restricted to the low-activity regions.

To summarize, the subband gain map $\hat{G}_b$ is computed by histogram matching the summed subband T_i^s to the target subband B_i^t using the entire compositing region. The noise gain map $\hat{G}_n$ is computed by histogram-matching the subbands N_i^s to the target noise pyramid subbands N_i^t while restricting the pixels to the low-activity region Ω. The structure and noise subbands are then updated as in Eqn. 8:

$$B_i^h = \hat{G}_b(A_i^s) B_i^s \tag{11}$$

$$N_i^h = \hat{G}_n(|N_i^s|) N_i^s \,. \tag{12}$$

After applying the gains, we collapse the source and noise pyramids to produce the corresponding images and repeat the entire harmonization loop for a fixed number of iterations (set to 5). We refer to this combination of smooth histogram and noise matching as harmonization.

After the final iteration, the harmonized pyramid T^h is given by:

$$T_i^h = B_i^h + N_i^h \,. \tag{13}$$

By collapsing this pyramid, we can reconstruct the final output image. If the goal is to composite the harmonized source and target images, we also need to impose the appropriate boundary conditions on the reconstruction. In the next section we describe how we achieve this.

6 Pyramid compositing

In the absence of any boundary conditions, the image corresponding to the harmonized subbands T^h is the solution to a linear system that comprises n separate linear systems, each corresponding to one subband in the harmonized pyramid:

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} I^h = \begin{bmatrix} T_1^h \\ T_2^h \\ \vdots \\ T_n^h \end{bmatrix}, \tag{14}$$

where f_i are the filters used to construct the pyramid, T_i^h are the harmonized subbands, and the vector I^h is the final composite.

Alpha matting and gradient-based compositing (also known as seamless cloning) are the two common ways of producing plausible boundaries in composites. While most compositing methods can handle one or the other – Drag and Drop Pasting [Jia et al. 2006] is a notable exception – in many cases, we would like to have both kinds of boundaries (see Fig. 8).

In alpha matting the composite is created by blending the foreground image with the background image (in our case the target image I^t) using the alpha matte α_m:

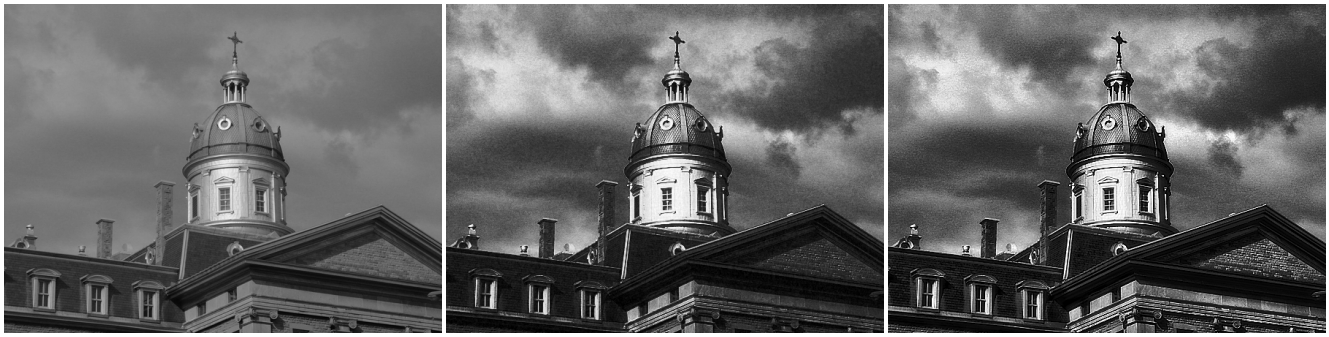
$$I^h = \alpha_m I^f + (1 - \alpha_m) I^t. \tag{15}$$

Combining the matting equation with Eqn. 14 gives us the relation:

$$\begin{bmatrix} \alpha_m f_1 \\ \alpha_m f_2 \\ \vdots \\ \alpha_m f_n \end{bmatrix} I^f = \begin{bmatrix} T_1^h - (1-\alpha_m) f_1 \star I^t \\ T_2^h - (1-\alpha_m) f_2 \star I^t \\ \vdots \\ T_n^h - (1-\alpha_m) f_n \star I^t \end{bmatrix}. \tag{16}$$

Since both the matte values and the target image are known, we can solve for I^f and compute the final composite I^h by substituting I^f in Eqn. 15.

We can incorporate seamless boundaries in Eqn. 16 by using the binary compositing mask as the alpha matte. Also, while imposing



(a) Source Image (b) Our result (c) [Bae et al. 2006]

Figure 5: *Using our harmonization framework to transfer the photographic look of Ansel Adams'* Clearing Winter Storm *to the source image (a) produces results (b) with similar effects to the system described by Bae et al. [2006] (c).*

seamless boundary conditions, we drop the equations corresponding to the coarsest lowpass subband, from Eqn. 16. This is similar to gradient domain techniques, where the composite is reconstructed solely from the (highpass) gradients.

To solve Eqn. 16 accurately, the subband coefficients T^h need to be consistent with the boundary conditions that we wish to impose. To ensure this, we combine the given alpha matte and seamless region into a single mask that is used to matte the source and target images to create a new image that is now used as the source image. The source subband coefficients T_i^s are computed by decomposing this image, and the harmonization as described in Sec. 5 is applied on them. Since the source pyramid is constructed on an image with the correct boundary conditions, the harmonized subband coefficients at the edges will encode these boundary conditions.

Quadtree Solver The size of the linear system we wish to solve in Eqn. 16 is quadratic in the number of pixels in the composited region, and as the size of the region increases, solving Eqn. 16 directly becomes prohibitively expensive. While this is true of most gradient-based techniques, this effect is amplified in our case because of the larger number of filters we employ.

Since we have chosen pyramid filters, we can reconstruct an image from the subband coefficients by collapsing the pyramid. This pyramid solution I^h_{pyr}, while fast to compute, does not satisfy the boundary constraints. On the other hand, the least-squares solution to Eqn. 16 I^h_{lsq}, satisfies the boundary constraints, but is slow to compute. The difference between these images is smoother than both I^h_{pyr} and I^h_{lsq} and can therefore be well approximated by an upsampled lower resolution image.

Agarwala [2007] made a similar observation in the context of panorama stitching, and proposed an algorithm that spatially subdivides the problem domain to create a reduced linear system. In our case, I^h_d still has some of the structure of the original image and we modify the Agarwala algorithm to allocate pixels to regions of high subband coefficient activity as described by the aggregate source activity map A^s_{ag}. Starting with the entire compositing region, we recursively subdivide every block of pixels into four quadrants as long as the aggregate activity in that block is greater than a threshold (set to 4). By basing the quadtree decomposition on the activity map, we are able to sample the difference image well. We solve for the difference image at the pixels at the corners of the quadtree decomposition and the pixels along seam boundaries. At all other pixels, we bilinearly interpolate these values and add it to the pyramid solution to produce an approximation that is visually identical but much faster to compute.

7 Results and Discussion

In this section, we describe how our framework can be used to easily create compelling composites in several common scenarios. The results in this paper have been scaled down and we request that the reader zoom into the images in the electronic version of the paper. All the results are also presented in the accompanying supplementary material where they can be viewed in higher resolution.

Except for Fig. 7, all the results shown in this paper were created using a 3-level pyramid. The one parameter in our system that is useful to control the final composite is the noise percentile β in Eqn. 10. The noise percentile enables us to distinguish between structure and noise and needs to be set according to how noisy the target image is. We used a value of 25% for all the results except for Figs. 6 and 9 where we used 50%.

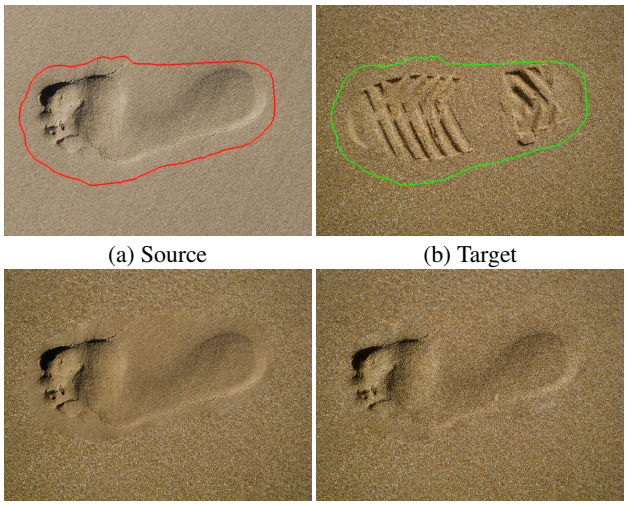
The run-times for our unoptimized Matlab implementation depend on the size of regions being composited and varied from 15 seconds for the result in Fig. 11a ($\approx$ 5500 pixels in the composited region) to 12 minutes for the example in Fig. 8 ($\approx$ 185500 pixels in the composited region). In most cases, almost 85% of the time is spent on solving the reduced version of the linear system in Eqn. 16. We used the *CSparse* library [Davis 2006] to solve the linear system. Recent work on fast sparse solvers [Szeliski 2006; McCann and Pollard 2008] and approximate solutions [Farbman et al. 2009] leads us to believe that an optimized implementation of our system can drastically reduce computation times.

Style transfer With smooth histogram matching on subbands, our harmonization framework is able to achieve effects similar to the style transfer technique described by Bae et al. [2006]. Their approach uses a two-level decomposition with nonlinear filters and has separate routines that allow it to exaggerate details. While our goal for harmonization is to improve realism rather than create a stylized result, our results in Figs. 4 and 5 suggest that some of these effects are possible within a linear pyramid framework.

Contrast Matching The source image in Fig. 11 has very different contrast from the target faces it has been composited into and the seamlessly cloned composite look unrealistic. By harmonizing the images, our method creates more natural composites.

Noise Matching In many cases, the noise characteristics of the source and target images are different. Injecting noise into our framework allows us to reproduce the noise characteristics of the target image and produces a more compelling result. This is illustrated in the examples in Figs. 8, 10, and 11.

While the harmonization framework can add noise to a image to match appearance, an interesting case is the problem of inserting a



(a) Source (b) Target

(c) Seamless cloning (d) Harmonized result

Figure 6: *The sand in the source image (a) has a different texture from that in the target image (b) leading to easily perceivable seams in the seamless cloning result (c). Harmonizing the two image matches the two textures so that the resulting composite (d) is more consistent. Photo credits: Flickr users Ivar Husevåg Døskeland/*`Scarto` *(a), and Christian Guthier/*`net_efekt` *(b).*

noisy source image into a smooth target region. This is similar to denoising, which is a long-standing problem in image processing. As seen in Fig. 10, matching the pyramid subbands decreases the noise and produces a better composite. Intuitively, harmonization suppresses the high frequencies of the noisy source image and automatically selects the bands to remove frequencies from based on the frequencies in the target image. However, harmonization will not be able to remove all the noise, and often, the final result will be slightly blurred compared to the original.

Texture matching In both Figs. 1 and 6, the target image has a textured appearance that the source does not have. This is especially pronounced in Fig. 6, where the images are of completely different kinds of sand. While gradient domain compositing produces seamless boundaries, the seam is still easily perceived. By shaping the noise we inject into our system to match the textures on the images, we are able to produce more compatible results.

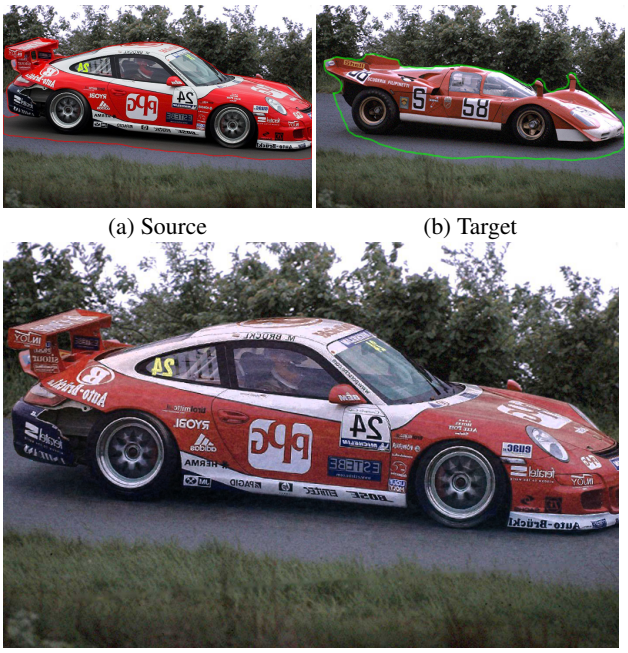
Color While our framework was described for grayscale images, it can be easily extended to color. It is important to manipulate color channels in a decorrelated color space so as not to create color shifts and we have found that CIELAB works well. We convert the images to CIELAB space and then harmonize and composite each channel separately. In some cases, the user might like to match the color palette of the source and target images and we use the N-dimensional PDF transfer method of Pitié et al. [2005] to match the a and b channels of the source image to those of the target before harmonizing them (Figs. 1, 8, and 10).

Blur Another scenario in compositing is when the user combines two regions with different blur. This is illustrated in Fig. 7 where the user segments a sharp object and clones it onto a blurred region expecting the inserted object to have the same defocus properties as the source. By harmonizing the inserted object with the defocused objects it is replacing, we are able to produce an image with realistic blur. We used a 4-level pyramid to generate this example because of the large amount of blur.

Mixed boundary constraints One of the advantages of pyramid compositing is the ability to incorporate boundary conditions for

(a) Source / Target (b) Harmonized result

Figure 7: *The region marked in red in the original image (a) is copied and pasted onto the regions marked in green. Cloning the pasted region seamlessly will not match the blur of the original image. Matching the blur produces a result (b) that preserves the shallow depth of field of the original photograph. Photo credit: Flickr user Brad T. Patterson/*`patterbt`.



(a) Source (b) Target

(c) Harmonized result

Figure 8: *In this example, the user clones a Porsche (a) into an old photograph of a Ferrari (b). Our result (c) matches the noise on the images, and alpha mattes the car while enforcing seamless boundaries on the road at the bottom. Photo credits: Flickr users Thomas Helbig/*`teliko82` *(a), and Jim Culp/*`prorallypix` *(b).*

both alpha matting and seamless cloning. This is illustrated by Figs. 8 and 9, where the final composite has seamless boundaries in some parts (the road and the sand) and alpha matte based boundaries elsewhere (the car and the hydrant).

Limitations Like Heeger-Bergen texture synthesis, our noise and texture matching technique makes the assumption that the target noise and texture can be matched by shaping the subbands of the noise image. Such techniques are known to work well on stochastic textures but do not reproduce every texture pattern accurately. In particular, it is known that histogram matching of pyramid subbands cannot be used to create textures that are correlated across

(a) Source / Target (b) Harmonized result

Figure 9: *Limitations. A hydrant in snow ((a) top) has been composited into sand ((a) bottom). Harmonization matches the snow to the sand, and compositing with mixed boundary conditions produces seamless boundaries along the sand and matting along the hydrant. However, the texture generated is not able to match the structure of the original sand. Also, because the target image does not have shadows or a hydrant, harmonization is not able to produce realistic shadows and has added excessive noise on the hydrant. Photo credits: Flickr users Robert Fornal/*`Bob.Fornal` *((a) top) and Luis Argerich/*`lrargerich` *((a) bottom).*

scales [Portilla and Simoncelli 2000]. Therefore, in some cases there might be differences in the noise between the target and harmonized images. For example, the harmonized image in Fig. 1 does not capture the small cracks in the painting and the result in Fig. 9 does not replicate the structure of the sand. In spite of this, harmonization leads to a substantial improvement in the realism of the composite, and in most cases, it is difficult to see the differences without looking at the original target image.

Also, a fundamental assumption of our approach is that matching the statistics of the source and target images will harmonize them. This may not always be the case, especially in situations where the objects being matched are completely different. This is illustrated in Fig. 9, where matching the images does not produce the right colors and leads to excessive noise on the foreground object.

8 Conclusions and Future Work

We have presented a framework that harmonizes the appearance of images before compositing them. By automatically matching different aspects of visual appearance, such as contrast, texture, noise, and blur, our technique takes the burden of correcting for them away from the user. We have also presented a novel compositing scheme that allows us to enforce both matte-based and seamless boundaries in the same framework.

There are other aspects of visual appearance that are important to the realism of a composite that our work does not address. The most important of these are shadows and shading. Automatically estimating and correcting the lighting in single images is a difficult vision problem and is an interesting avenue for future work.

The ability to realistically combine multiple images is important in many vision and graphics applications such as image mosaicing and digital photomontage, and we would like to apply our methods in their context too. In addition, we are interested in extending our work to the problem of video object insertion. Videos often have high levels of noise and compression and we believe our methods will be useful in creating realistic video composites.

Acknowledgements

Part of this work was done while Wojciech Matusik was a Senior Research Scientist, and Kalyan Sunkavalli was an intern at Adobe Systems, Inc. Micah K. Johnson would like to acknowledge support from the National Science Foundation under Grant No. 0739255 and a gift from Adobe Systems.

We would like to thank David Salesin and members of the Advanced Technology Labs for their support and feedback. We also thank the SIGGRAPH reviewers for their time and constructive comments. Finally, we would like to thank all the Flickr users whose photographs we have reproduced under Creative Commons.

References

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Transactions on Graphics 23*, 3, 294–302.

AGARWALA, A. 2007. Efficient gradient-domain compositing using quadtrees. *ACM Transactions on Graphics 26*, 3, 94.

BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. *ACM Transactions on Graphics 25*, 3, 637–645.

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

BURT, P. J., AND ADELSON, E. H. 1983. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics 2*, 4, 217–236.

CHEN, J., TANG, C.-K., AND WANG, J. 2009. Noise brush: interactive high quality image-noise separation. *ACM Transactions on Graphics 28*, 5, 1–10.

COHEN-OR, D., SORKINE, O., GAL, R., LEYVAND, T., AND XU, Y.-Q. 2006. Color harmonization. *ACM Transactions on Graphics 25*, 3, 624–630.

DAVIS, T. A. 2006. *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics 27*, 3, 1–10.

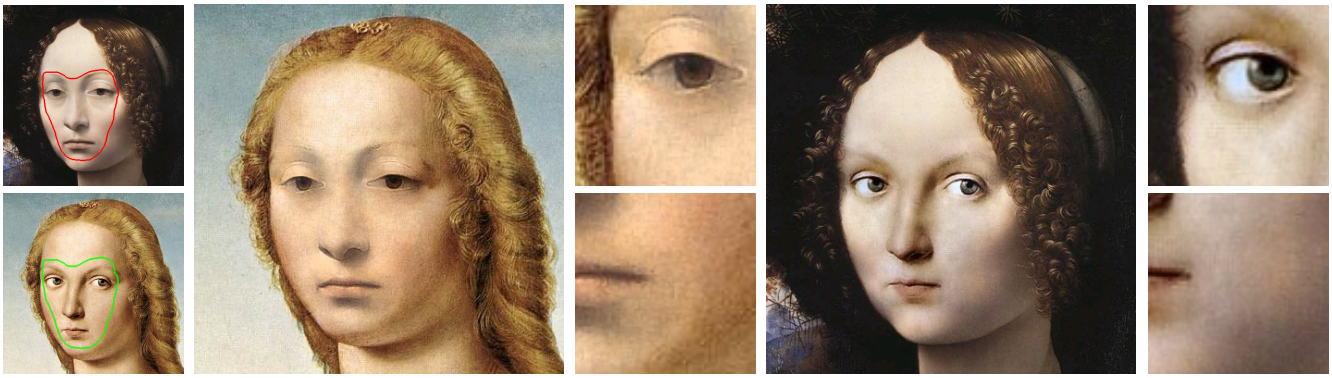
FARBMAN, Z., HOFFER, G., LIPMAN, Y., COHEN-OR, D., AND LISCHINSKI, D. 2009. Coordinates for instant image cloning. *ACM Transactions on Graphics 28*, 3, 1–9.

FATTAL, R., AGRAWALA, M., AND RUSINKIEWICZ, S. 2007. Multiscale shape and detail enhancement from multi-light image collections. *ACM Transactions on Graphics 26*, 3, 51.

GEORGIEV, T. 2004. Photoshop healing brush: a tool for seamless cloning. In *Workshop on Applications of Computer Vision (ECCV 2004)*, 1–8.

GONZALEZ, R. C., AND WOODS, R. E. 2001. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

HEEGER, D. J., AND BERGEN, J. R. 1995. Pyramid-based texture analysis/synthesis. *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, 229–238.



(a) Source / Target (b) Inserting source into target (c) Close-ups (d) Inserting target into source (e) Close-ups

Figure 10: *In this example, the source ((a) top) is smooth while the target ((a) bottom) is noisy. When inserting the source into the target, harmonization adds noise to produce a realistic composite (b). Conversely, when the target image is inserted into the source, harmonization removes most of the noise to match the images (d).*

JIA, J., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2006. Drag-and-drop pasting. *ACM Transactions on Graphics 25*, 3, 631–637.

LALONDE, J.-F., AND EFROS, A. A. 2007. Using color compatibility for assessing image realism. In *IEEE International Conference on Computer Vision*.

LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In *European Conference on Computer Vision*.

LI, Y., SHARAN, L., AND ADELSON, E. H. 2005. Compressing and companding high dynamic range images with subband architectures. *ACM Transactions on Graphics 24*, 3, 836–844.

MCCANN, J., AND POLLARD, N. S. 2008. Real-time gradient-domain painting. *ACM Transactions on Graphics 27*, 3, 1–7.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Transactions on Graphics 22*, 3, 313–318.

PITIE, F., KOKARAM, A. C., AND DAHYOT, R. 2005. N-dimensional probablility density function transfer and its application to colour transfer. In *IEEE International Conference on Computer Vision*.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 253–259.

PORTILLA, J., AND SIMONCELLI, E. P. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision 40*, 1, 49–70.

REINHARD, E., ASHIKHMIN, M., GOOCH, B., AND SHIRLEY, P. 2001. Color transfer between images. *IEEE Computer Graphics and Applications 21*, 5, 34–41.

SUN, J., JIA, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Poisson matting. *ACM Transactions on Graphics 23*, 3, 315–321.

SZELISKI, R. 2006. Locally adapted hierarchical basis preconditioning. *ACM Transactions on Graphics 25*, 3, 1135–1143.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *IEEE International Conference on Computer Vision*.

WANG, J., AGRAWALA, M., AND COHEN, M. F. 2007. Soft scissors: an interactive tool for realtime high quality matting. *ACM Transactions on Graphics 26*, 3, 9.

(a) Target images (b) Seamless cloning (c) Naive histogram matching (d) Smooth histogram matching (e) Smooth histogram and noise matching

Figure 11: *This figure illustrates how our method adapts the same source image to match different target images with markedly different contrast, noise, and texture. Gradient domain compositing (b) produces unrealistic results because of the discrepancies between the images being combined. Naive histogram matching (c) oversharpens the source image and creates ringing around strong gradients. Our smooth histogram matching method (d) automatically adapts the source image to each of the targets without these artifacts, but the noise is still inconsistent. Matching both the structure and the noise removes these inconsistencies and produces realistic results (e). Photo credits: Flickr users The Rob Oechsle Collection/*`Okinawa Soba` *(second row), Zsolt Botykai/*`zsoltika` *(third row), and David Flam/*`freeparking` *(fourth and fifth rows).*