Personal Photo Enhancement Using Example Images

NEEL JOSHI
Microsoft Research
WOJCIECH MATUSIK
Disney Research
EDWARD H. ADELSON
MIT CSAIL
and
DAVID J. KRIEGMAN
University of California, San Diego

We describe a framework for improving the quality of personal photos by using a person's favorite photographs as examples. We observe that the majority of a person's photographs include the faces of a photographer's family and friends and often the errors in these photographs are the most disconcerting. We focus on correcting these types of images and use common faces across images to automatically perform both global and face-specific corrections. Our system achieves this by using face detection to align faces between "good" and "bad" photos such that properties of the good examples can be used to correct a bad photo. These "personal" photos provide strong guidance for a number of operations and, as a result, enable a number of high-quality image processing operations. We illustrate the power and generality of our approach by presenting a novel deblurring algorithm, and we show corrections that perform sharpening, superresolution, in-painting of over- and underexposured regions, and white-balancing.

Categories and Subject Descriptors: I.4.3 [**Image Processing and Computer Vision**] Enhancement; I.4.4 [**Image Processing and Computer Vision**]: Restoration

General Terms: Algorithms

Additional Key Words and Phrases: Image enhancement, image processing, image-based priors, computational photography, image restoration

ACM Reference Format:

Joshi, N., Matusik, W., Adelson, E. H., and Kriegman, D. J. 2010. Personal photo enhancement using example images. ACM Trans. Graph. 29, 2, Article 12 (March 2010), 15 pages. DOI = 10.1145/ 1731047.1731050 http://doi.acm.org/10.1145/1731047.1731050

1. INTRODUCTION

Using cameras tucked away in pockets and handbags, proud parents, enthusiastic vacationers, and diligent amateur photographers are always at the ready to capture the precious, memorable events in their lives. However, these perfect photographic moments are often lost due to an inadvertent camera movement, an incorrect camera setting, or poor lighting. Such imperfections in the photographic process often cause a photograph to be a complete loss for all but the most experienced photo-retouchers. Recent advances in digital photography have made it easier to take photographs more often, providing more opportunities to capture the "perfect photograph", yet still too often an image is unceremoniously discarded with the photographer lamenting "this would have been a great photograph if only..."

There is still a large gap in quality of photographs between an advanced and casual photographer. Advances in digital camera technology have improved many aspects of photography, yet the ability to take good photographs has not increased proportionally with these technological advances. Cameras have increased resolution and sensitivity, and many consumer cameras have "scene modes" to help less experienced users take better photographs. While such improvements help during capture, they are of little help in correcting flaws after a photograph is taken.

Recent work has begun to address this issue. A common approach is to use image-based priors to guide the correction of common flaws

This work was completed while N. Joshi was a student at the University of California, San Diego and an intern at Adode Systems, and W. Matusik was an employee at Adobe Systems.
Authors' addresses: N. Joshi, Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399; email: neel@microsoft.com; W. Matusik, Disney Research; E. H. Adelson, MIT CSAIL, The Stata Center, Building 32, 32 Vassar Street, Cambridge, MA 02139; D. J. Kriegman, Department of Computer Science and Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0404.

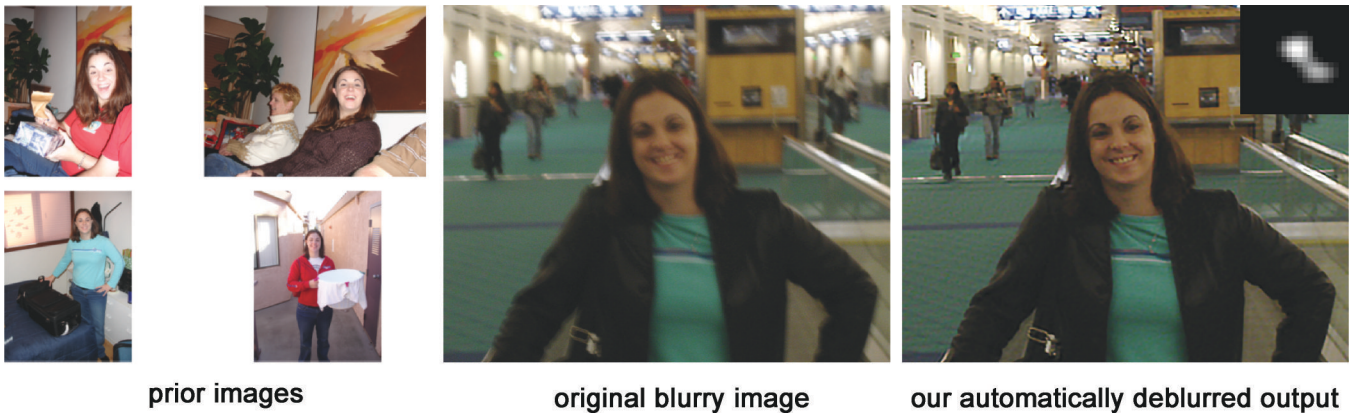

Fig. 1. Automatically correcting personal photos. We automatically enhance images using prior examples of "good" photos of a person. Here we deblur a blurry photo of a person where the blur is unknown. Using a set of other sharp images of the same person as priors (left), we automatically solve for the unknown blur kernel and deblur the original photo (middle) to produce a sharp image (right); the recovered blur kernel is shown in the top right enlarged 3×.

such as blurring [Fergus et al. 2006], lack of resolution [Freeman et al. 2002], and noise, due to lack of light [Roth and Black 2005]. Work in this area tends to rely on priors derived from a large number of images. These priors are specific to a particular domain, such as a face prior for superresolution of faces or a gradient distribution prior for natural images, but they tend to be general within the domain, that is, they capture properties of everyone's face or all natural images. Such methods are promising and have shown some impressive results; however, at times their generality limits their quality.

In this article, we take a different approach toward image correction. We note that many consumer photographs are of a personal nature, for example, holiday photographs and vacation snapshots are mostly populated with the faces of the camera owner's friends and family. Flaws in these types of photos are often the most noticeable and disconcerting. In this work, we seek to improve these types of photos and focus specifically on images containing faces. Our approach is to "personalize" the photographic process by using a person's past photos to improve future photos. By narrowing the domain to specific, known faces we can obtain high-quality results and perform a broad range of operations.

We implement this personalized correction paradigm as a postprocess using a small set of examples of good photos. The operations are designed to operate independently, so that a user can choose to transfer any number of image properties from the examples to a desired photograph, while still retaining certain desired qualities of the original photo. Our methods are automatic, and we believe this image correction paradigm is much more intuitive and easier to use than current image editing applications.

The primary challenges involved in developing our "personal image enhancement" framework are: (1) decomposing images such that a number of image enhancement operations can be performed independently from a small number of examples, (2) defining transfer functions so that only desired properties of the examples are transferred to an image, and (3) correcting nonface areas of images using the face and the example images as calibration objects. In order to accomplish this, we use an intrinsic image decomposition into shading, reflectance, and color layers and define transfer functions to operate on each layer. In this article, we show how to use our framework to perform the following operations.

—*Deblurring:* removing blur for images when the blur function is unknown by solving for the blur of a face,

—*Lighting transfer and enhancement:* transferring lighting color balance and correcting detail loss in faces due to underexposure or saturation,

—*Superresolution of faces:* creating high-resolution sharper faces from low-resolution images.

We integrate our system with face detection [Viola and Jones 2001] to obtain an automated system for performing the personalized enhancement tasks.

To summarize, the contributions of this article are: (1) the concept of the personal "prior": a small, identity-specific collection of good photos used for correcting flawed photographs, (2) a system that realizes this concept and corrects a number of the most common flaws in consumer photographs, and (3) a novel automatic multiimage deblurring method that can deblur photographs even when the blur function is unknown.

2. RELATED WORK

Digital image enhancement dates back to the late 60's with much of the original work in image restoration, such as denoising and deconvolution [Richardson 1972; Lucy 1974]. In contrast, the use of image-derived priors is a relatively recent development. Image-based priors have been exploited for superresolution [Baker and Kanade 2000; Liu et al. 2001; Freeman et al. 2002], deblurring [Fergus et al. 2006], denoising [Roth and Black 2005; Liu et al. 2006; Elad and Aharon 2006], view-interpolation [Fitzgibbon et al. 2005], in-painting [Levin et al. 2003], video matting [Apostoloff and Fitzgibbon 2004], and fitting 3D models [Blanz and Vetter 1999].

These priors range from statistical models to data-driven example sets, such as a face prior for face hallucination, a gradient distribution prior for natural images, or an example set of high- and low-resolution image patches; they are specific to a domain, but general within that domain. To the best of our knowledge, most work using image-based priors is derived from a large number of images that may be general or class/object specific, but there has been very little work in 2D image enhancement using *identity-specific* priors. Most work using identity-specific information is in the realm of detection, recognition, and tracking in computer vision and face animation and modeling in computer graphics. In the latter realm,

recent work by Gross et al. [2005] has shown that there are significant advantages to using person-specific models over generic ones.

A related area of work is photomontage and image compositing [Agarwala et al. 2004; Levin et al. 2004; Rother et al. 2006]. In the work of Agarwala et al. [2004], user interaction is combined with automatic vision methods to enable users to create composite photos that combine the best aspects of each photo in a set. Another related area of work is digital beautification. Leyvand et al. [2006] use training data for the location and size of facial features for "attractive" faces as a prior to improve the attractiveness of an input photo of an arbitrary person. We see our work as complementary the work of Agarwala et al. [2004] and Leyvand et al. [2006], as while we all share similar goals of improving the appearance of people in photographs, we focus more on overcoming photographic artifacts and do not seek to change the overall appearance of a subject.

Our individual corrections use gradient-domain operations pioneered by Perez et al. [2003]. Our work also shares similarities with image fusion methods and transfer methods [Reinhard et al. 2001; Eisemann and Durand 2004; Petschnigg et al. 2004; Agrawal et al. 2005; Bae et al. 2006] in that we use similar image decompositions and share similar goals of transferring photographic properties.

Our face-specific enhancements are inspired by the face-hallucination work of Liu et al. [2007]. Liu et al. [2007] use a set of generic faces as training data that are prealigned, evenly lit, and grayscale. Where our work differs is that we use identity-specific priors, automatic alignment, and a multilayered image decomposition that enables operating on a much wider range of images, where the images can vary in lighting in color, and we perform operations in the gradient domain. These extensions enable the use of a more realistic set of images (with varied lighting and color), improve matching, and give higher quality results. Furthermore, Liu et al. [2007] do not address in-painting and hallucinating entire missing regions, as we show in Figure 8.

Our deblurring algorithm is related to the work of Fergus et al. [2006] and multiimage deblurring methods [Bascle et al. 1996; Rav-Acha and Peleg 2005; Yuan et al. 2007]. Fergus et al. [2006] recover blur kernels assuming a prior on gradients on the unobserved sharp image and in essence only assume "correspondence" between the sharp image and prior information in the loosest sense, in that they assume the two have the same global edge content. Multi-image deblurring is on the other end of the spectrum. These methods use multiple images of a scene acquired in close sequence and generally assume strong correspondence between images. Our method resides between these two approaches with some similarities and several significant differences.

Relative to multi-image methods, we assume moderate correspondence, by using an aligned set of identity-specific images; however, we allow for variations in pose, lighting, and color. To the best of our knowledge, deblurring using any type of face-space as a prior, let alone our proposed identity-specific one, is novel. Both our method and Fergus et al.'s [2006] are in the general (and large) class of Expectation-Maximization (EM)-style deblurring methods. Where they differ is in the specific nature of the prior and that our method is completely automatic given a set of prior images. Fergus et al.'s [2006] work, on the other hand, requires user input to select a region of an image for computing a PSF. In our experience, this user input is not simple, as it often requires several tries to select a good region and must be done for every image. Furthermore, our work is computationally simpler using a *Maximum A Posteriori* (MAP) estimation instead of variational Bayes, which leads to a ten to twenty times speedup.

3. OVERVIEW

We present several image enhancement operations enabled by having a small number of prior examples of good photos of a person. The enhancements are grouped into two categories: global image corrections and face-specific enhancements. Global corrections are performed on the entire image by using the known faces as calibration objects. We perform global exposure and white-balancing and deblurring using a novel multi-image deconvolution algorithm. For faces in the image we can go beyond global correction and perform per-pixel operations that transfer desired aspects of the example images. We in-paint saturated and underexposed regions, correct lighting intensity variation, and perform face-hallucination to sharpen and super-resolve faces. Our system operates on base/detail image decomposition [Eisemann and Durand 2004] and therefore these operations can be performed independently. As illustrated in Figure 2, our system proceeds as follows.

(1) Automatically detect faces on target images and prior images.
(2) Align and segment faces in target and prior images.
(3) Decompose images into color, texture, and lighting layers.
(4) Perform global image corrections.
(5) Perform face-specific enhancements.

Step 1 outputs a set of nine feature points for each target and prior face and step 2 produces a set of prior images aligned to the target image with masks indicating the face on each image. Both steps are discussed in detail in Section 6. Step 3 is discussed in the next section, step 4 in Section 4, and step 5 in Section 5.

3.1 Prior Representation and Decomposition

In this work, we derive priors from a small collection of person-specific images. In contrast with previous work using large image collections [Hays and Efros 2007], our goal is to use data that is easily collected by even the most casual photographer, who may not have access to large databases of images.

Researchers have noted that the space spanned by the appearance of faces is relatively small [Turk and Pentland 1991]. This observation has been used for numerous tasks including face recognition and face hallucination [Liu et al. 2007]. We make the additional observation that the space spanned by images of a single person is significantly smaller; when examining a personal photo collection the range of photographed expressions and poses of faces is relatively limited. Thus we believe the use of a small set of person-specific photos to be a relatively powerful source for deriving priors for image corrections.

While expression and pose variations may be limited, lighting and color can vary significantly between photos. As a result, a central part of our framework is the use of a base/detail layer decomposition [Eisemann and Durand 2004] that we use as an approximate "intrinsic image" decompostion [Barrow and Tenenbaum 1978; Land and McCann 1971; Finlayson et al. 2004; Weiss 2001; Tappen et al. 2006]. In such as decomposition, an image is represented as a set of constituent images that capture intrinsic scene characteristics and extrinsic lighting characteristics. Intrinsic images are an ideal construct as they: (a) allow us to use a small set of prior images to correct a broad range of input images and (b) they enable modifying image characteristics independently.

We adopt the base/detail layer decomposition used by Eisemann and Durand [2004] that makes this separation based on the Retinex assumption and uses an edge-preserving filter to decompose lighting from texture. We decompose an RGB image I into a set of four images $[r, g, L, X]$, where $Y = R + G + B$ represents luminance

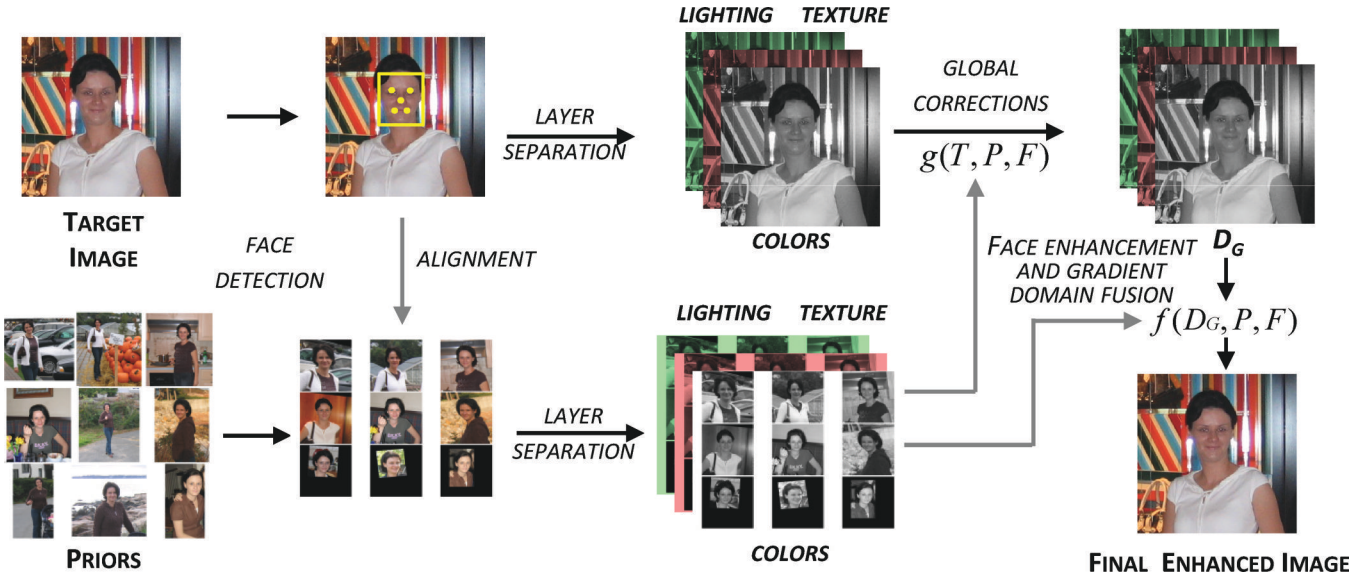

Fig. 2. Personal image enhancement pipeline. First we use face detection to find faces in each image and align the prior images to the person in the target photo. The images are then decomposed into intrinsic images: color, texture, and lighting. First global image corrections are performed and then face-specific enhancements. We combine the global and face-specific results using gradient domain fusion.

and $r = R/Y$ and $g = G/Y$ are red and green chromaticity. L, the lighting (or base) image, is a bilaterally filtered version of luminance Y. X, the shading image, is computed as $X = Y/L$. For the sake of simplicity of terminology, for the remainder of this article, we will refer to the (r, g) chromaticity reflectance images as "color layers," L, the base image as the "lighting layer," and X, the shading layer, as the "texture layer."

The layers from our example set are used for direct example-based techniques and to derive statistical priors. To achieve this, we follow the hybrid model of Liu et al. [2007] and perform corrections using both a linear eigenspace and a patch-based nonparametric approach.

3.2 Enhancement Framework

We create a desired processed image I with layers $I = [I_r, I_g, I_L, I_X]$ from a given observed image $O = [O_r, O_g, O_L, O_X]$ and an aligned set of prior images where one prior image is $P = [P_r, P_g, P_L, P_X]$. We automatically align the prior images P to O and compute a mask, F, for each face automatically. Alignment and mask computation is discussed in the next section.

The aligned, intrinsic prior layers are used directly for a patch-based method, and we also create eigenspaces for these layers. From each aligned and cropped intrinsic prior layer we create a set of orthogonal basis vectors using SVD. We denote $\mathscr{P}$ as a matrix of basis vectors and $\mathscr{P}_\mu$ as the mean vector that describes a feature space for the examples. An example of this is shown in Figure 4. Unlike previous work in this area, since our set of examples is small we do not use a subspace; our basis vectors capture all the variation in the data, and thus we are simply using Singular-Value Decomposition (SVD) to orthogonalize the data. Thus, our "personal prior" is the entire set of aligned layers and basis and mean vectors for each space.

As illustrated in Figure 2, we perform image enhancement by creating a desired image I, by first performing global corrections to obtain the image I^G, and then we perform face-specific corrections to obtain the final result I.

Face-specific enhancements are performed in the gradient domain using Poisson image editing techniques [Pérez et al. 2003], where an image is constructed from a specified 2D guidance gradient field, v, by solving a Poisson equation: $\partial I/\partial t = \nabla I - div(v)$. Specifically, this can be formed as a simple invertible linear system: $LI = div(v)$, where L is the Laplacian matrix. We refer the reader to the paper by Perez et al. [2003] for more details on gradient-domain editing.

3.3 Face Alignment and Mask Computation

To align faces in examples to a face in the input image, we use an implementation of the automatic face detection method of Viola and Jones [2001]. The detector outputs the locations of faces in an image along with nine facial features, the outside corner of each eye, the center of each eye, bridge and tip of the nose, and the left, right, and center of the mouth. From these features we align the faces using an affine transformation.

When performing face-specific enhancement, it is also necessary to have a mask for the face in the input and prior images. We automatically compute these by using the feature locations to initially compute a rough mask labeling face and nonface areas of the image. First our system creates a "trimap" by labeling the image as foreground, background, and unknown regions. The foreground area is marked as the pixel inside the convex hull of the nine detected features. The edge is labeled as background and the remaining pixels are in the unknown region. We compute an alpha-matte using the method of Levin et al. [2006] and threshold this soft-segmentation into a mask. The threshold is 50% and we eroded the mask pixels by ten pixels to get the final mask. An example of a mask is shown in Figure 3.

In the following sections, we describe our enhancement and correction functions and how each uses our prior.

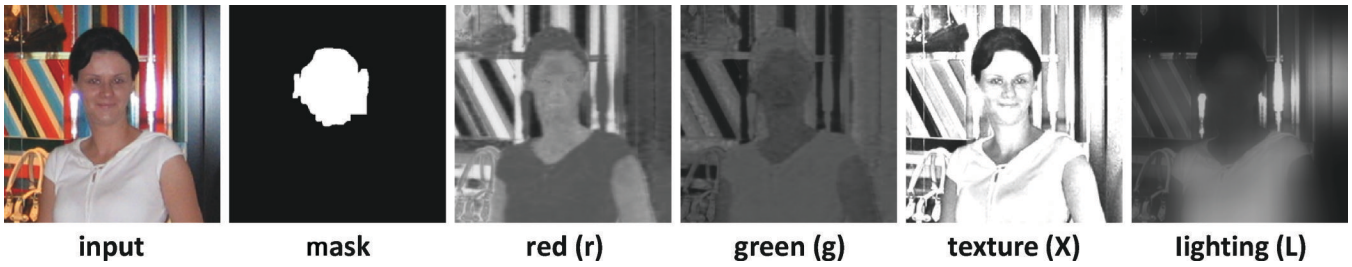

Fig. 3. Mask computation and layer decomposition. We perform our corrections on an "intrinsic image"-style decomposition of an image into color, lighting, and texture layers. This enables a small set of example images to be used to correct a broad range of input images. In addition, it allows us to modify image characteristics independently. We also automatically compute a mask for the face that we use as part of our face-specific corrections.

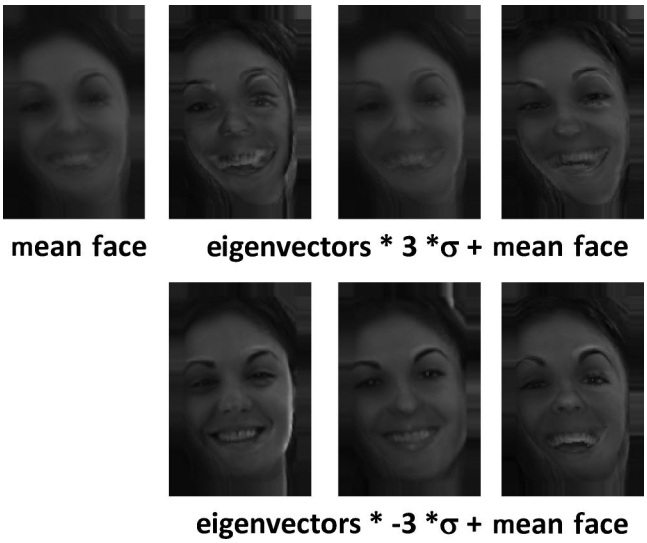

Fig. 4. Eigenfaces constraint. We use linear feature spaces built from an aligned set of good images of a person as a constraint in our image enhancement algorithms. Here we show the eigenfaces used as a prior for the delurring result shown in Figure 1.

4. GLOBAL CORRECTION OPERATIONS

Many aspects of a person's facial appearance, particularly skin color, albedo, and the location of features, such as the eyes and nose, remain largely unchanged over the course of time. By leveraging their relative constancy, one can globally correct a number of aspects of an image.

We consider global corrections to be those that are calculated using the face area of an image and are applied to the entire image. Our global corrections use basis and mean vectors constructed from the example images as a prior within a Bayesian estimation framework. Our goal is to find the most likely estimate of the uncorrupted image I given an observed image O. This is found by maximizing the probability distribution of the posterior using Bayes' rule, also known as *Maximum A Posteriori* (MAP) estimation. The posterior distribution is expressed as the probability of I given O.

$$P(I|O) = \frac{P(O|I)P(I)}{P(O)} \tag{1}$$

I can then be recovered by maximizing this posterior or minimizing a sum of negative log likelihoods.

$$I = \operatorname*{argmax}_I P(I|O) \tag{2}$$

$$= \operatorname*{argmax}_I P(O|I)P(I) \tag{3}$$

$$= \operatorname*{argmin}_I L(O|I) + L(I) \tag{4}$$

$L(O|I)$ is the "data" term and $L(I)$ is the image prior. The specific form of each value is different for each correction. In our system, we correct for overall lighting intensity and color balance and perform multiimage deconvolution to deblur an image.

4.1 Image Deblurring

We deblur an image of a person using our personal prior as a constraint during image deconvolution. While pixel-wise alignment of the blurred image and the prior images is difficult, a rough alignment is possible, as facial feature detection on down-sampled blurred images is reliable. The feature space for texture layers from our personal prior is then used to constrain the underlying sharp image during deconvolution. We rely on the variation across the prior images to span the range of facial expressions and poses.

We only consider blur parallel to the image plane and solve for a shift-invariant kernel. We model image blur as the convolution of an unknown sharp image with an unknown shift-invariant kernel plus additive white Gaussian noise:

$$O = I \otimes K + N, \tag{5}$$

where $N \sim \mathcal{N}(0, \sigma^2)$.

We formulate the image deconvolution problem using the Bayesian framework discussed before, except that we now have two unknowns I and K. We continue to minimize a sum of negative log likelihoods.

$$L(I, K|O) = L(O|I, K) + L(I) + L(K) \tag{6}$$

Given the blur formation model (Eq. (5)), we have

$$L(O|I, K) = ||O - I \otimes K||^2/\sigma^2. \tag{7}$$

We consider $O = O_X^F O_L^F$, which is the observed image's luminance, where the superscript F indicates that only the masked face region is considered (we will drop the ^{F} notation in later sections for the sake of readability). The sharp image I we recover is the deblurred luminance.

The negative log likelihood term for the image prior is

$$L(I) = \lambda_1 L(I|\mathcal{P}, \mathcal{P}_\mu) + ||\nabla I||^q, \tag{8}$$

which maintains that the image lies close to the examples' feature space by penalizing distance between the image and its projection onto the space, which is modeled by the eigenvectors and mean vector ($\mathscr{P}$, $\mathscr{P}_\mu$)). We have determined empirically that $\lambda_1 = 400$ works well.

The term also includes the sparse gradient prior of Levin et al. [2007]: $||\nabla I||^q$.

The feature-space used in the prior is built from the examples' texture layers times the observation's lighting layer, that is, $P^i = P^i_X O_L$, for each example i. This implicity assumes that the blurring process does not affect the lighting layer, that is, the prior's, examples and the observation have the same low-frequency lighting. While this assumption may not always be true, it holds in practice as lighting changes tend to be low frequency.

To use this feature space, we define a negative log likelihood term using a robust distance-to-feature space metric.

$$L(I\,|\,\mathscr{P},\,\mathscr{P}_\mu) = \rho\left([\,\mathscr{P}'\mathscr{P}(I - \mathscr{P}_\mu) + \mathscr{P}_\mu] - I\right) \tag{9}$$

$\mathscr{P}$ represents the matrix whose columns are the eigenvectors of the feature-space, and $\mathscr{P}'$ is the transpose of this matrix. This term enforces that the residual between the latent image I and the robust projection of I on to the feature-space [$\mathscr{P}$, $\mathscr{P}_\mu$] should be minimal. $\rho(.)$ is a robust error function described next.

We use a robust norm (rather than L_2 norm) to make this projection more robust to outliers (for example, specular reflections and deep shadows on the target face or feature variations not well-captured by the examples). For $\rho(.)$ we use the Huber norm.

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & |r| \leqq k \\ k|r| - \frac{1}{2}k^2 & |r| > k \end{cases} \tag{10}$$

k is estimated using the standard "median absolute deviation" heuristic. We use an Iterative Reweighed Least-Squares (IRLS) approach to minimize the error function.

For the sparse gradient prior, instead of using $q = 0.8$, as Levin et al. use, we recover the exponent by fitting a hyper-Laplacian to the histogram of gradients of the prior images' faces. To fit the exponent, consider that the p-norm distribution is $y = ce(x)^{-p}$ (c is a constant) , taking the log of both sides results in: $log(y) = log(c) + -plog(e(x))$. If $y = ||\nabla I^F||$ and x is the probability of different gradient values (as estimated using a histogram normalized to sum to one), p is the slope of the line fit to this data. By fitting the exponent in this way we constrain the gradients of the sharp image in a way that is consistent with the prior examples; in our experience, the recovered p is always between 0.5 and 0.6.

The prior on the kernel is modeled as a sparsity prior on the values and a smoothness prior on the kernel, which are common priors used during kernel estimation. The likelihood $L(K)$ is

$$L(K) = \lambda_2||K||^p + \lambda_3||\nabla K||^2, \tag{11}$$

where $p < 1$.[1]

Blind deconvolution is then performed using a multiscale, alternating minimization, where we first solve for I using an initial assumption for K (we use a 3x3 gaussian) by minimizing $L(I|B, K)$ and then use this I to solve for K by minimizing $L(K|B, I)$. Each subproblem is solved using iterative reweighted least-squares.

In performing debluring, we recover only the sharp image data for the face and the kernel describing the blur for the face. If the person in the target photograph did not move relative to the scene, this blur describes the camera-shake and we use the method of Levin et al. [2007] to deblur the whole image. A result from our method is shown in Figure 1.

[1]We have found our method is relatively insensitive to the value of p as long as it is < 1. $p = 0.8$ seems to work well.

4.2 Exposure and Color Correction

The goal of this part of our framework is to adjust the overall intensity and color-balance of the target photograph such that they are most similar to that of well-exposed, balanced prior images. We model this adjustment with scaling parameters for the lighting and color layers.

We robustly match the target face's lighting and color to mean lighting and color vector from the prior feature-spaces. We again formulate this using the Bayesian framework and minimize a sum of negative log likelihoods. For exposure correction, the data term is

$$L(O|I) = ||O_L - \omega_L I_L||^2, \tag{12}$$

where ω_L is a scalar value. The image prior is

$$L(I) = L(I_L\,|\,\mathscr{P}_{L\mu}) = \rho\left(\mathscr{P}_{L\mu} - I_L\right). \tag{13}$$

For white-balancing, we use an equation of the same form to compute scaling values ω_r and ω_g using the respective I_r, I_g, $\mathscr{P}_{r\mu}$, and $\mathscr{P}_{g\mu}$ values.

In practice, if $\omega_L > 1$ we set it equal to 1, so we do not scale down image exposure. For color-balancing, as skin tones do not span a large color gamut we perform a simple white-balance, that is, independent scaling of the color layers, as we have found it to be the only reliable transformation we can perform. In particular, we have found that a full linear transformation or a nonlinear transform, such as histogram matching, often has too many degrees of freedom and produces undesired results. Examples of global exposure and white-balancing are shown in Figure 5.

5. FACE-SPECIFIC ENHANCEMENT

In many cases, global corrections will not remove all the flaws in an image. While shortcomings of the corrections will generally be unnoticed for non-face regions, they are likely to be objectionable for faces, as people are much more sensitive to their appearance. Thus, we perform local corrections on faces.

5.1 Modifying Lighting and Texture

We address several image corrections under the umbrella of hallucinating high-frequency texture. Lack of detail due to defocus blur or oversmoothing during deconvolution, lack of resolution, or saturation of an image can be corrected by transferring high-frequency information from the aligned texture layers of the personal prior. In the case of over- and underexposure, saturated and clipped areas can be in-painted by hallucinating parts of all intrinsic image layers.

Restoring high-frequency texture. For this process, we decouple hallucinating a sharp-texture layer I into making global and local estimates, I^L and I^G respectively, where the $I = I^L + I^G$. The global component I^G captures the lower frequencies of the image and the local component I^L captures the highest-frequency data. This is the same decomposition used by Liu et al. [2007].

When the blur is unknown, such as for defocus and motion blur, I^G is the result of the blind-deconvolution method in Section 4.1. When the blur is known, such as with superresolution, we minimize

Fig. 5. Exposure and color correction. Using the same set of prior images our system automatically corrects exposure and white-balance for three different images containing the same person.

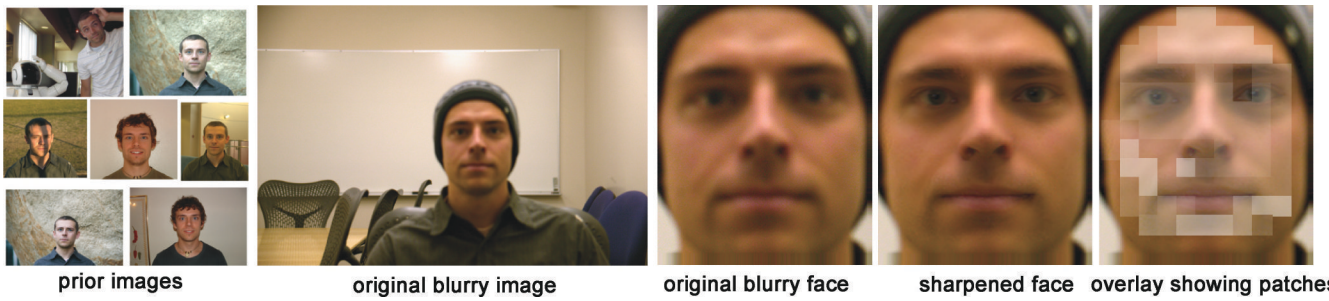

Fig. 6. Defocus blur. A set of sharp, in-focus priors (first image). An image suffering from blur due to misfocus (second image). A close up on original image (third image) and the corrected face after removing the defocus blur by hallucinating texture from priors (fourth image), and a visualization showing which parts of the face came from different patches (fifth image).

this equation.

$$L(I^G|O,K) = ||O - I^G \otimes K||^2/\sigma^2 + \lambda_1 L(I|\mathscr{P},\mathscr{P}_\mu) + ||\nabla I||^q \tag{14}$$

When performing superresolution, K is an antialiasing filter.

To recover I^L we use a patch-based nonparametric Markov network that is a combination of the method of Liu et al. and Freeman et al. [2002]. We model $I^L = I - I^G$, thus I^L is the highest-frequency component and depends on the low-frequency component I^G.

We compute training pairs from the prior texture layers of (P^{Li}, P_M^{Gi}) for all priors i, where $P_M^G = P^G - f \otimes P^G$, where f is a gaussian filter. We seek to find patches around each point that maximize the compatibility function.

$$\begin{aligned}\phi(I^L(m,n) &= P^{Li}(m,n), I_M^G(m,n)) \\ &= ||I_M^G(m,n) - P_M^{Gi}(m,n)||^2,\end{aligned} \tag{15}$$

where $I^L(m,n)$ denotes a patch centered at $(ms + s/2, ns + s/2)$. With patch $s + 2$ we use a patch size of $10x10$ pixels with a 2 pixel overlap. We have an additional affinity function that states that the overlapping region of patches must be similar.

$$\begin{aligned}\psi(I^L(m,n), I^L(m+i,n+j)) = ||\Omega(I^L(m,n)) \\ -I^L(m+i,n_+j))||^2,\end{aligned} \tag{16}$$

where Ω returns the overlapping regions on the two patches given the patch size. We refer the reader to the paper by Liu et al. [2007] for more details on this derivation.

Intuitively, maximizing the preceding function says that on a patch-by-patch basis, we predict the highest-frequency data based on how the mid-frequencies of the target and priors match each other. Just as in Liu et al.'s [2007] work, our priors are roughly aligned, so we only consider patches at the same location in the priors, and use a raster-scan technique to perform the energy minimization [Freeman et al. 2002; Hertzmann et al. 2001].

Where our method differs from the previous techniques is that we perform this patch-based prediction on separate color, lighting, and texture layers. An additional difference is that to assemble the final locally corrected image I^L, we composite the *gradients* of the $P^L(i)$ into O^L, which is the local component of the observed image O relative to the global correction: $O^L = O - I^G$. We have found the gradient-domain process to generate much cleaner composites.

Fig. 7. Superresolution. A set of sharp, in-focus priors (first image). An image with blur and JPEG artifacts at a low resolution (second image). A close up on original image up-sampled 2 times using bicubic interpolation (third image), the corrected face after 2x up-sampling and hallucinating texture from priors (fourth image), and a visualization showing which parts of the face came from different patches (fifth image).

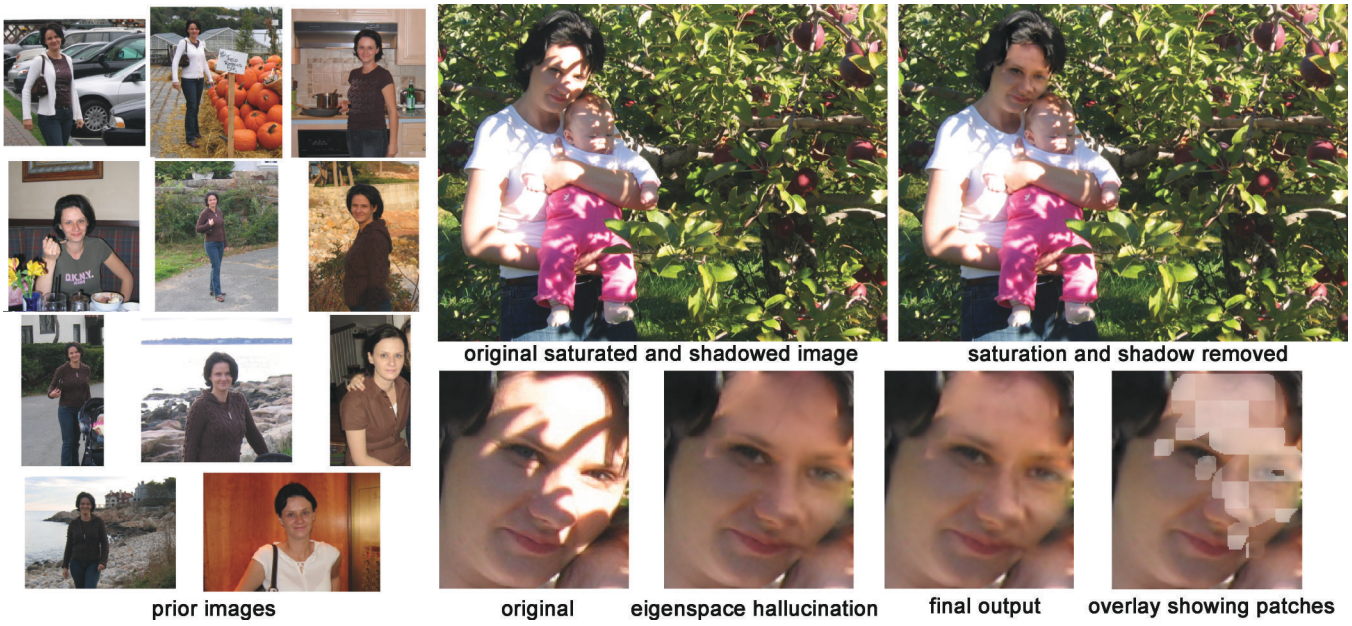

Fig. 8. Removing high-frequency shadows and uneven illumination. Prior images (on the left). A photo with uneven illumination, hard high-frequency shadows, and saturation (top row, second image). Our result (top row, third image). The shadow has been softened significantly and the saturated areas corrected. The bottom row shows close-up for the input image, our intermediate result after estimating the global texture layer, and the final result after using the patch-based method.

Examples of the methods presented here are shown in Figure 6 and Figure 7, where we add texture lost due to defocus blur and perform superresolution.

Restoring clipped data. When over- or underexposure causes pixel values to be clipped, there is a complete loss of texture, high and low frequency, in those regions. Thus, to in-paint these regions we use the algorithm described earlier with a few modifications. In the previous section, we discuss predicting a global estimate for the texture layer and then local estimates for texture and color layers. When restoring clipped data, the process must be altered slightly. This is because all of the data in the saturation region is unreliable. Thus, we must predict global estimates for all layers within the saturation region.

We construct a saturation/shadow mask for the clipped face region, where clipped pixels are those with original pixel values in any color channel above or below a threshold (we use $<= 10$ and $>= 240$ for images in the range $[0, 255]$). The saturation/shadow mask is incorporated into the global estimation process discussed previously, such that the algorithm fits the unclipped regions to the eigenspace and the masked out region is filled with data that is most consistent with this fit to the unmasked regions. This is achieved using a simple masking function when performing robust least-squares. Note that consistency is enforced with a sparse gradient penalty across the whole image.

We compute the global high-frequency texture layer X^g_{IH} and color layers r^g_{IH} and g^g_{IH} using a joint eigenspace of texture and color. Specifically, we perform the same operation as when restoring high-frequency texture, but instead of using an eigenspace for the texture layers alone we build the eigenspace by orthogonalizing the stacked vectors of $[X^g_{PH}, r^g_{PH}, g^g_{PH}]$ of the example high-resolution images (indicated by the subscript PH) and then solve Eq. (14) with these values for the vector $[X^g_{IH}, r^g_{IH}, g^g_{IH}]$. Note that for this application the kernel K in Eq. (14) is a delta function, since there

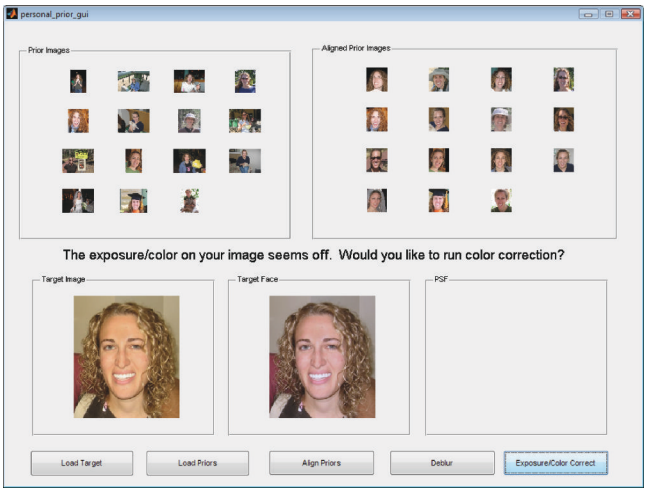

Fig. 9. Personal photo correction application. Here we show a screenshot of an initial prototype of our personal photo correction application. Currently the system performs automatic deblurring and automatic exposure and white-balancing. It can also suggest corrections by using the statistics of the personal prior.

is no blur. Before performing this correction, we first perform a global white balance (discussed in Section 4.2) such that the colors are similar between the examples and the target face. The global lighting correction is performed in a similar way using the lighting-space separately, that is, an eigenspace for the L^g_{PH} layers of the example images.

To predict high frequencies we then run our Restoring High-Frequency Texture algorithm (described before) on the global estimates for the texture and color layers, $[X^g_{IH}, r^g_{IH}, g^g_{IH}]$. For the lighting layer we only compute the global estimation and forgo the patch-based correction as the lighting tends to contain only low-frequency information, and it is generally undesirable to transfer high-frequency lighting (such as hard shadows and specularities).

An example of the method presented here is shown in Figure 8.

6. PERSONAL PHOTO CORRECTION APPLICATION

We have implemented a prototype application and user interface for performing the corrections discussed in the previous sections. In addition to the correction features, we have a "suggestion" system, where the application can suggest that the user performs a particular correction on a loaded photograph after computing some simple image statistics and comparing these to the statistics of the prior images.

For deblurring suggestions, our system computes the standard deviation of the magnitude of gradients for face region of all priors. Similarly, our system computes the magnitude and standard deviation of gradients in the face regions. If the latter value is less than 90% of the former, the system suggests performing deblurring. For the color and exposure suggestion, the application simply computes the correction discussed in Section 4.2, as this is a fast operation, and if the scalar adjustment for color or lighting is > 1.1 or < 0.9 (more than a 10% change) it suggests the correction. Currently, a photo must be actively loaded for a suggestion to be made; however, the process could easily be run offline as a way to automatically tag a collection of new photos with suggested corrections.

Figure 9 shows a screen-capture of the GUI.

7. RESULTS

We will now briefly recap some of our results that were presented in the body of the article and present several additional examples.

In Figures 1 and 10, we show two examples of our automatic blind deblurring method using our personal prior. In Figure 11 we compare our method to using Fergus et al.'s [2006] method. We recovered a PSF using the authors' code available online. The side-by-side comparison shows deconvolving the image using the method of Levin et al. [2007] with the PSF from our method and Fergus et al.'s. For the woman in Figure 1, deblurring with the recovered PSF using Fergus et al.'s method does not completely sharpen the image. For the man in Figure 10, the PSF from the Fergus et al. method seems to be overcompensating, and thus the result is overly sharpened with halo artifacts. For both images, the kernel recovered by our method appears more accurate and was recovered over ten times faster. Furthermore, our method does not require any manual input.

In Figures 6 and 7 we show two examples of face hallucination to remove image blur. Figure 6 shows a significant amount of defocus that is automatically removed using our method. Figure 7 shows hallucination for a 2x up-sampling compared to up-sampling with bicubic interpolation. For both results we show a visualization that indicates what regions of the final results came from different prior images when performing patch-based local hallucination. Both results are sharp with minimal artifacts. Furthermore, for the up-sampling result in Figure 7 our method has removed some of the JPEG artifacts in the original image.

In Figure 12, we compare results in Figures 6 and 7 to the result of performing hallucination using an implementation of Liu et al.'s [2007] method. Specifically, their work predicts high-frequency texture image from mid-frequency image data. They use a generic set of faces and use the raw image data directly (without an intrinsic layer decomposition or gradient-domain editing). We used images from the public Caltech and GeorgiaTech image databases as our set of generic faces examples. The results with the Liu et al. [2007] method have artifacts similar to those shown in their paper. We believe our results are more convincing. Also in Figure 12 we compare using our enhancement algorithm method with a set of generic faces instead of faces of the same person. Thus instead of using 5 to 10 hand-selected good images of a person, we used the 10 and 50 best matching generic faces as priors. We automatically selected these images by first aligning the generic faces to the input image and then compute a match score that is the L_2 norm of the difference of down-sampled/contrast-normalized versions of the image. For both results, compared to using generic faces, our results have fewer artifacts and appear to retain the original look and expression of the person in the input image. The woman's face in Figure 12 is particularly of note. When using generic faces and Liu et al.'s method, a mole is introduced into the up-sampled result, even when the input image shows no mole. When using our method with generic faces, the woman no longer looks completely like the same person, and the expression of the woman is altered, as she no longer appears to be smiling as much.

Figure 13 shows the results of two synthetic deblurring experiments, where we blurred a sharp image with a blur kernel, added 0.5% noise, and then solved for the PSF using our deblurring method and the method of Fergus et al. We then used the method of Levin et al. [2007] to deconvolve the blurred image with the ground-truth kernel, our recovered kernel, and the kernel resulting from running Fergus et al.'s code. The side-by-side comparisons show that the deconvolution results with our recovered kernels are fairly close to the quality of the results that use the known kernels, which shows that

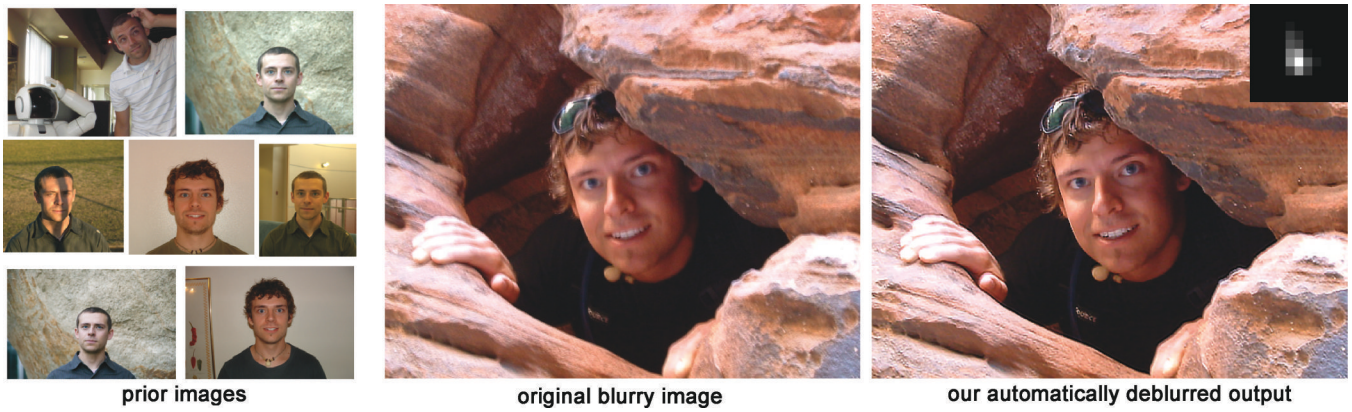

Fig. 10. Additional deblurring example. Our method automatically performs blind-deconvolution to recover the blur kernel. The entire image is then deblurred using the method of Levin et al.; the recovered blur kernel is shown enlarged 5× in the top right of the rightmost image.

Fig. 11. Comparison to Fergus et al.'s PSF estimation method. For the images in Figure 10, we hand-picked a good region for estimating the PSF using Fergus et al.'s method. We ran their code and then deblurred each image using the method of Levin et al. Fergus et al.'s system took over an hour and a half to recover each kernel. Our method recovers more accurate kernels and produce better deconvolution results.

Fig. 12. Face hallucination comparisons. We compare our result to performing hallucination using an implementation of Liu et al.'s method and to using our enhancement algorithm with a set of (the 10 and 50 best matching) generic faces instead of faces of the same person. Compared to Liu et al.'s approach and compared to using generic faces, our results have fewer artifacts and appear to retain the original look and expression of the person in the input image.

Fig. 13. Synthetic deblurring experiments. We blurred a sharp image with two different blur kernels and added 0.5% noise (second column), and then solved for the PSF using our deblurring method and the method of Fergus et al. We then deconvolved the blurred image with the ground-truth kernel (third column), our recovered kernel (fourth column), and the kernel resulting from running Fergus et al.'s code (fifth column). Each kernel is shown in the top right corner of each image.

our kernels are accurate. The results using Fergus et al.'s method are not very accurate for these images.

Figure 14 shows the results of two synthetic up-sampling/ hallucination experiments, where we down-sampled a sharp image by 2× and 4× and then up-sampled those images by 2× and 4× using our enhancement algorithm, an implementation of Liu et al.'s method, and our enhancement algorithm with a set of (the 10 and 50 best matching) generic faces. With the 2× up-sampling result, our method produces a convincing result that is sharper than traditional bicubic upsampling. Similar to the result in Figure 12, when using Liu et al.'s method, a mole is introduced into the up-sampled result, and when using our method with generic faces, there are a number of artifacts including that woman's identity and expression are significantly changed. When performing 4× up-sampling, our method produces an image that is a bit sharper than the others; however, this result shows the limits of our method; while we are able to hallucinate high frequencies, there is not enough mid-frequency information to predict the highest frequencies well, thus the result does not match the ground-truth image very closely.

In Figure 15, we show results of experiments of using our face hallucination algorithms without using an intrinsic image decomposition and gradient-domain editing operations, two of our improvements over the work of Liu et al. [2007]. The results without using an intrinsic image decomposition have more artifacts as the effects of the lighting and color differences between the input image and examples are no longer factored out of the matching process. The results without gradient-domain edits show very noticeable seams. Our modifications create enhanced images that are seamless and convincing.

In Figure 5, we show automatic exposure correction and white-balancing for three images of one woman using the same set of prior images. In Figure 16, we compare our results to using several current color constancy algorithms. Specifically, we compared to algorithms discussed by van de Weijer et al. [2007] and use the author's code and recommended parameters. Our results are more consistent across images, appear better white-balanced, and did not require any parameter tuning.

In Figure 8, we show an image of a woman in an apple orchard where her face has a hard high-frequency shadow edge across it and our algorithm reduces the shadow and recovered texture in the saturated region of the face.

8. DISCUSSION AND FUTURE WORK

We have presented a powerful framework for improving personal images and shown how to correct a number of the most common errors in photographs using what we believe is a simple, yet powerful concept of a "personal prior." While recent work in data-driven methods for photograph correction has tended towards using large generic databases and automated methods for picking "good"

Fig. 14. Synthetic upsampling experiments. We down-sampled a sharp image by 2× and 4× and then up-sampled those images by 2× and 4×, respectively, using bicubic upsampling (second column), our enhancement algorithm (third column), an implementation of Liu et al.'s method (fourth and fifth columns), and our enhancement algorithm with a set of (the 10 and 50 best matching) generic faces (six and seventh columns).

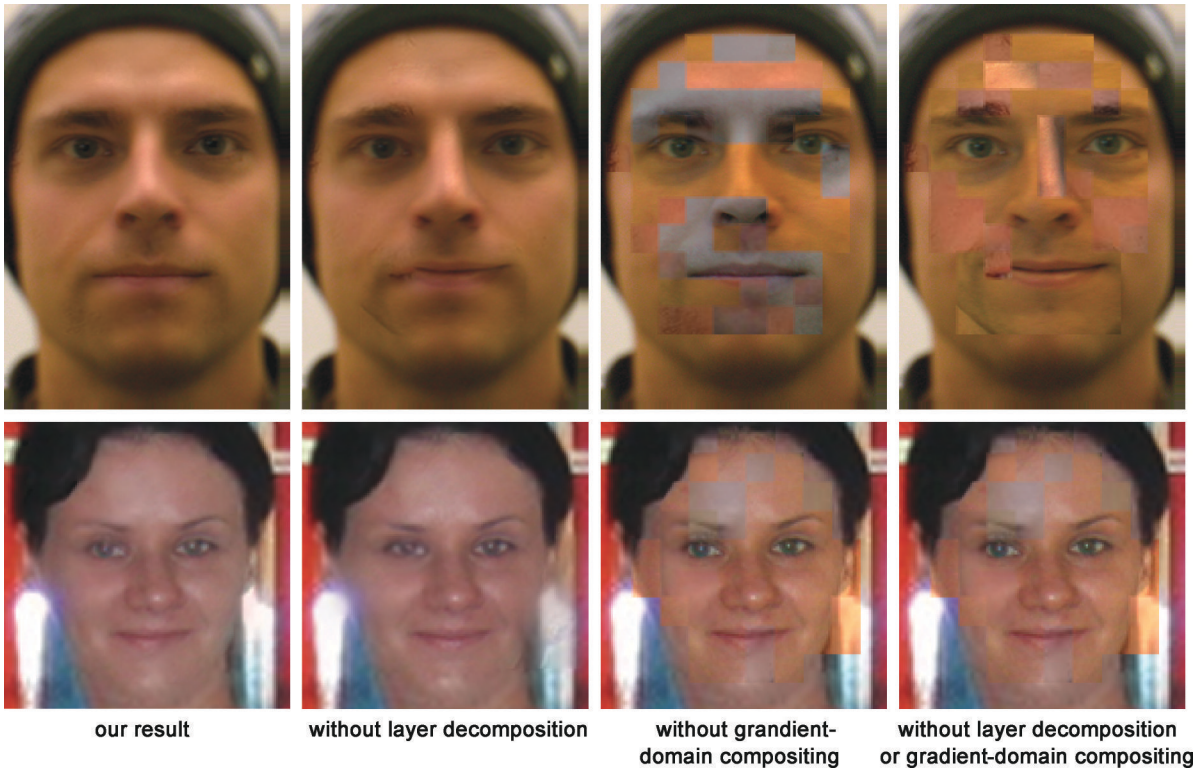

Fig. 15. Face hallucination algorithms without using an intrinsic image decomposition and gradient-domain editing. The results without using an intrinsic image decomposition have more artifacts as the effects of the lighting and color differences between the input image and examples are no longer factored out of the matching process. The results without gradient-domain edits show very noticeable seams.

Fig. 16. Comparisons to color constancy. We compare our results to the color constancy algorithms discussed by van de Weijer et al. Our results are more consistent across images, appear better white-balanced, and did not require any parameter tuning.

photographs [Hays and Efros 2007]. We have taken a very different approach. We have focused on correcting images with faces and have left the step of choosing good photographs to the user and automated the difficult part of editing the photo. We believe this is a very natural and intuitive way to think about correcting images of people.

A natural question for our work is: How many example images are needed? We have found that this depends on the type of the correction performed. Exposure and color correction are not very sensitive to expression and pose changes and thus very few, even one, photograph can be enough. Deblurring can require more images, but often not very many due to the robust estimation process we use; the algorithm will reject outlier regions of the face and favor matching to the more invariant parts of the face that are well-captured by the eigenspace. Hallucination is the most demanding, as it is not always an option to ignore parts of the face; however, due to our combined subspace and local patch approach, we have gotten good results with as few as seven images. While some analysis has been done for the dimensionality of the generic "face space" [Penev and Sirovich 2000], we are not aware of an analysis for specific individuals; as future work, we are very interested in such an analysis.

A general limitation of facial appearance modification, which our work is susceptible to, is the sensitivity of people to the appearance of faces. In our experience, with our system we have found that users are very sensitive to even subtle changes of photographs of people, especially when the person is known to the user. Generally the only pleasing and acceptable corrections are subtle and small changes, and it is difficult to make significant changes to an image without altering the fundamental mood or feel of the photograph. Often a large change sends a photo into the "uncanny valley". This concept states that aesthetic qualities related to modification of human appearance are subject to a curve where improvements in appearance are positive until a point when suddenly there is a negative reaction when the modified appearance becomes disturbingly "uncanny". This is a danger with our methods just as with any other work that modifies images of faces.

There are numerous possible avenues for future work. We think our system and methods could be easily incorporated into commercial photo editing products and could leverage photo-tagging and -rating systems that are already available, such as image rating in Windows Vista, ratings and labels in Adobe Bridge, or tags on Web sites like Flickr and Facebook. Our work could be paired with a simple labeling/rating system so that users could mark images with tags

such as "good color," "sharp," "good lighting." Then images with these labels could be automatically chosen as priors. We are very interested in integrating our system into such an application. We also note that as an extension to our current "correction suggestion" system, we could use machine learning techniques to try to automatically pick examples of good priors once the user has tagged a few.

We note that our system is currently limited to mostly frontal photographs. This is primarily due to our face-detector having been trained and tuned for frontal faces. Nonfrontal face detection is more difficult; however, there is much work in this area and we are interested in investigating improvements that would allow nonfrontal face detection. We could then have a richer set of pose-specific priors. A related and particularly useful extension that would build on our paradigm is to extend our personal prior concept to detection, whereby feature detectors are tuned to a specific person. Our system also does not currently use recognition, thus if multiple people are in an image, the user must perform the identification to pair the priors. We have experimented with face recognition in our framework and hope to add this shortly to our application.

Another interesting question is: What other forms of person-specific or class-specific prior information could be used for image manipulation? In this work, we have used a set of images, the main motivation being that everybody can easily acquire and select images that they like. We believe that our framework could be extended to perform even better with full 3D geometry, detailed reflectance properties (for example, spatially varying BRDF and subsurface scattering properties), or a linear morphable face model. One could include priors for the whole body rather than the faces only, and in principle, it would also be possible to store and use the priors about the environment (for example, the places where the photos are typically taken). An obvious disadvantage of using such information is that, at least currently, it can be difficult to acquire this type of data; however, using more sophisticated datasets presents several directions for future work.

Lastly, while we have focused on improving images of faces, we note that our framework is actually more general. Our fundamental framework could be applied to any object-specific appearance enhancement where one has detectors and example images for a specific object.

REFERENCES

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

AGRAWAL, A., RASKAR, R., NAYAR, S. K., AND LI, Y. 2005. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Trans. Graph. 24,* 3, 828–835.

APOSTOLOFF, N. AND FITZGIBBON, A. 2004. Bayesian video matting using learnt image priors. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'04)*. Vol. 01. 407–414.

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

BAKER, S. AND KANADE, T. 2000. Hallucinating faces. In *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition*.

BARROW, H. AND TENENBAUM, J. 1978. Recovering intrinsic scene characteristics from images. *Comput. Vision Syst.*, 3–26.

BASCLE, B., BLAKE, A., AND ZISSERMAN, A. 1996. Motion deblurring and super-resolution from an image sequence. In *Proceedings of the 4th European Conference on Computer Vision-Volume II (ECCV'96)*. Springer, 573–582.

BLANZ, V. AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Proceedings of SIGGRAPH 99*. 187–194.

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

ELAD, M. AND AHARON, M. 2006. Image denoising via learned dictionaries and sparse representation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. IEEE Computer Society, 895–900.

FERGUS, R., SINGH, B., HERTZMANN, A., ROWEIS, S. T., AND FREEMAN, W. T. 2006. Removing camera shake from a single photograph. *ACM Trans. Graph. 25,* 787–794.

FINLAYSON, G. D., DREW, M. S., AND LU, C. 2004. Intrinsic images by entropy minimization. In *Proceedings of the IEEE International Conference on Computer Vision (ECCV)*. 582–595.

FITZGIBBON, A., WEXLER, Y., AND ZISSERMAN, A. 2005. Image-based rendering using image-based priors. *Int. J. Comput. Vision 63,* 2, 141–151.

FREEMAN, W. T., JONES, T. R., AND PASZTOR, E. C. 2002. Example-based super-resolution. *IEEE Comput. Graph. Appl. 22,* 2, 56–65.

GROSS, R., MATTHEWS, I., AND BAKER, S. 2005. Generic vs. person specific active appearance models. *Image Vision Comput. 23,* 11, 1080–1093.

HAYS, J. AND EFROS, A. A. 2007. Scene completion using millions of photographs. In *ACM SIGGRAPH Papers (SIGGRAPH'07)*. ACM, New York, 4.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'01)*. ACM, New York, 327–340.

LAND, E. H. AND MCCANN, J. J. 1971. Lightness and retinex theory. *J. Optical Soc. Amer. 61*, 1–11.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. In *ACM SIGGRAPH Papers (SIGGRAPH'07)*. ACM Press, New York, 70.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2006. A closed form solution to natural image matting. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. IEEE Computer Society, 61–68.

LEVIN, A., ZOMAT, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In *IEEE International Conference on Computer Vision (ECCV'04)*.

LEVIN, A., ZOMET, A., AND WEISS, Y. 2003. Learning how to inpaint from global image statistics. In *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV'03)*. 305.

LEYVAND, T., COHEN-OR, D., DROR, G., AND LISCHINSKI, D. 2006. Digital face beautification. In *ACM SIGGRAPH Sketches (SIGGRAPH'06)*. 169.

LIU, C., FREEMAN, W. T., SZELISKI, R., AND KANG, S. B. 2006. Noise estimation from a single image. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. IEEE Computer Society, 901–908.

LIU, C., SHUM, H.-Y., AND FREEMAN, W. T. 2007. Face hallucination: Theory and practice. *Int. J. Comput. Vision 75,* 1, 115–134.

LIU, C., SHUM, H.-Y., AND ZHANG, C. 2001. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'01)*. 192–198.

LUCY, L. 1974. Bayesian-Based iterative method of image restoration. *J. Astro. 79*, 745–754.

PENEV, P. S. AND SIROVICH, L. 2000. The global dimensionality of face space. In *Proceedings of the 4th IEEE International Conference on*

Automatic Face and Gesture Recognition 2000 (FG'00). IEEE Computer Society, 264.

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

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

RAV-ACHA, A. AND PELEG, S. 2005. Two motion-blurred images are better than one. *Pattern Recogn. Lett. 26,* 3, 311–317.

REINHARD, E., ASHIKHMIN, M., GOOCH, B., AND SHIRLEY, P. 2001. Color transfer between images. *IEEE Comput. Graph. Appl. 21,* 5, 34–41.

RICHARDSON, W. 1972. Bayesian-Based iterative method of image restoration. *J. Optical Soc. Amer. A 62*, 55–59.

ROTH, S. AND BLACK, M. J. 2005. Fields of experts: A framework for learning image priors. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 860–867.

ROTHER, C., BORDEAUX, L., HAMADI, Y., AND BLAKE, A. 2006. Autocollage. In *ACM SIGGRAPH Papers (SIGGRAPH'06)*. ACM Press, New York, 847–852.

TAPPEN, M. F., ADELSON, E. H., AND FREEMAN, W. T. 2006. Estimating intrinsic component images using non-linear regression. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'06)*. 1992–1999.

TURK, M. A. AND PENTLAND, A. P. 1991. Face recognition using eigenfaces. In *Proceedings of the Computer Vision and Pattern Recognition (CVPR'91)*. 586–591.

VAN DE WEIJER, J., GEVERS, T., AND GIJSENIJ, A. 2007. Edge-Based color constancy. In *Proceedings of the IEEE Conference on IEEE Trans. Image Process. 16,* 9, 2207–2214.

VIOLA, P. AND JONES, M. 2001. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'61)*. 609–615.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. http://www.ai.mit.edu/courses/6.899/papers/13-02.PDF

YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2007. Image deblurring with blurred/noisy image pairs. In *ACM SIGGRAPH Papers (SIGGRAPH'07)*. ACM, New York, 1.

Received September 2008; accepted November 2009