Portrait Lighting Transfer using a Mass Transport Approach

ZHIXIN SHU, Stony Brook University
SUNIL HADAP, ELI SHECHTMAN, KALYAN SUNKAVALLI, and SYLVAIN PARIS, Adobe Research
DIMITRIS SAMARAS, Stony Brook University and CentraleSupélec, Université Paris-Saclay

Lighting is a critical element of portrait photography. However, good lighting design typically requires complex equipment and significant time and expertise. Our work simplifies this task using a relighting technique that transfers the desired illumination of one portrait onto another. The novelty in our approach to this challenging problem is our formulation of relighting as a mass transport problem. We start from standard color histogram matching that only captures the overall tone of the illumination, and show how to use the mass-transport formulation to make it dependent on facial geometry. We fit a 3D morphable face model to the portrait, and for each pixel, combine the color value with the corresponding 3D position and normal. We then solve a mass-transport problem in this augmented space to generate a color remapping that achieves localized, geometry-aware relighting. Our technique is robust to variations in facial appearance and small errors in face reconstruction. As we demonstrate, this allows our technique to handle a variety of portraits and illumination conditions, including scenarios that are challenging for previous methods.

CCS Concepts: •**Computing methodologies** →**Image manipulation;**

Additional Key Words and Phrases: face relighting, mass transport, histogram matching

ACM Reference format:
Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. 2010. Portrait Lighting Transfer using a Mass Transport Approach. *ACM Trans. Graph.* 9, 4, Article 39 (March 2010), 15 pages.
DOI: 0000001.0000001_2

1 INTRODUCTION

Good lighting is a key component of portrait photography. Professional photographers design complex configurations of strobe lights and reflectors to accentuate different aspects of a subject's appearance, and achieve a particular look. Designing these lighting setups requires significant time and expertise, making high-quality portrait photography challenging for casual photographers and time-consuming for professionals.

The goal of our work is to make portrait lighting easier by allowing users to transfer the illumination from a *reference* portrait to an *input* photograph to create high-quality relit images . We wish to do this without any calibration of the lighting or any additional meta-data, thereby enabling post-capture portrait relighting. In addition to allowing users to easily explore different lighting configurations, our work has applications in portrait retouching, and post-production editing and compositing.

This work was supported by a gift from Adobe, NSF IIS-1161876, the Stony Brook SensorCAT and the Partner University Fund 4DVision project.
Author's addresses: Z. Shu and D. Samaras, Department of Computer Science, Stony Brook University; S. Hadap and K. Sunkavalli, Adobe Systems Inc., 345 Park Avenue, San Jose, CA; E. Shechtman, Adobe Systems Inc., 801 N 34th St, Seattle, WA; S. Paris, Adobe Systems Inc., Cambridge Innovation Center, MA.

While face relighting has been studied extensively [Shashua and Riklin-Raviv 2001; Wang et al. 2009; Wen et al. 2003], it remains a challenging problem due to, for instance, variations in appearance, pose, identity, and expression. Furthermore, human observers are sensitive to the subtleties of facial appearance, and have a low tolerance to errors in processed face images. Standard face editing approaches fit low-dimensional parametric models to facial appearance data to achieve robustness, but these models often do not achieve high visual fidelity. We address these challenges with a novel approach to portrait relighting: we pose it as a multi-dimensional mass transport problem that computes a non-parametric mapping between the input and reference images.

We start with standard color histogram matching which captures the global color and tone of lighting by transferring the color distribution of the reference onto the input portrait. This approach ignores the fact that shading depends on face geometry, and unsurprisingly produces subpar results. We extend this technique to make it aware of the geometry of the face. First, we fit a generic 3D model to the portrait [Yang et al. 2011]. We use this model to augment the color at each pixel with position and normal information. We then exploit the known formulation of histogram matching as a mass-transport problem, and extend it from color space to the higher-dimensional $\{\text{colors}\} \times \{\text{positions}\} \times \{\text{normals}\}$ space. This generates a color mapping that is aware of face geometry, and is able to capture the localized, directional nature of lighting changes. We further make this process robust to variations in face appearance and geometry by smoothing the color distributions of the input and reference. Despite the high-dimensionality of the exact mapping, we explain how this can be achieved using stochastic sampling, which allows us to use an existing mass-transport solver [Rabin et al. 2012] to efficiently compute our results.

Our approach has several advantages. Being non-parametric, it makes few assumptions on face appearance and illumination, which enables it to handle a wide range of lighting conditions and subjects, including non-photorealistic images. The robustness resulting from the regularization allows us to rely on a generic 3D model and makes our results robust to possible minor misalignment. We demonstrate these properties on a variety of subjects and illuminations, and show via a user study that our algorithm produces plausible relighting results that are superior in many cases to other relighting techniques.

Contributions. In summary, our contributions are:

1. A novel approach to face relighting that uses a mass-transport formulation to transfer illumination between images.

2. A regularization scheme that makes the technique robust to variations in face appearance and geometry.

3. A complete pipeline that improves the state of the art in face relighting and compositing.

2 RELATED WORK

Single-Image Face Relighting. Face relighting using a single image has been studied extensively in the context of face recognition [Adini et al. 1997; Georghiades et al. 2001]. Shashua and Riklin-Raviv [2001] proposed the quotient (or ratio) image technique for face relighting, where an input face image is relit by multiplying it by the ratio of a known reference face captured under novel lighting and the original input lighting.

Subsequent work relaxed the requirement for reference images under calibrated illumination by reconstructing face geometry and using it to estimate low-frequency illumination [Wang et al. 2009; Wen et al. 2003]. Ratio images have also been used to transfer subtle shading variations caused by changes in expression [Liu et al. 2001], and match lighting for face compositing [Bitouk et al. 2008]. Chen et al. [2011] use edge-preserving filters to create base and detail illumination layers which are used to transfer low-frequency lighting between images. Blanz and Vetter [1999] proposed using low-dimensional shape and texture models to reconstruct face geometry, albedo, and scene lighting from a single image. They used a combination of ambient illumination and a single directional light source. This was later extended to handle complex illumination and harsh lighting [Wang et al. 2009].

These techniques assume specific appearance models, e.g., Lambertian shading under distant lighting, and require accurate 3D reconstruction. They perform well when these assumptions are satisfied but as we shall see in the result section, their accuracy decreases when these assumptions do not hold. For instance, some sophisticated lighting setups may not be well represented by these models and 3D reconstruction often suffers from minor inaccuracy and misalignment. In contrast, our formulation does not assume a specific illumination model and is robust to small imperfections, which allows it to perform well on cases where other techniques fail.

Lightstage Face Relighting. Debevec et al. [2000] estimated the reflectance field of a subject's face from images captured under a dense set of illumination directions. This data can be used to drive that subject's facial performances under arbitrary illumination [Alexander et al. 2009]. Peers et al. [2007] used the reflectance data of one subject to compute ratio images which are used to relight another subject's facial performance. While these techniques produce very impressive relit faces, they depend on complex calibrated acquisition setups. On the other hand, our work is a lightweight face relighting technique that does not require any additional data.

Color Transfer for Relighting. Color transfer techniques [Reinhard et al. 2001] match color and tone statistics between images and can capture the overall tone of the reference illumination. Pitié et al. [2005] proposed a multi-dimensional histogram matching scheme that transfers the full 3D color distributions of a reference photograph to the input. Our mass transport formulation extends these techniques by incorporating geometry, and by using a different optimization scheme to produce more robust transfer results. Recent work has used localized color transfer to produce results that are more representative of lighting variations [Laffont et al. 2014; Shih et al. 2013]. In particular, Shih et al. [2014] transfered a particular photographer's style, including lighting, to a given image. Their technique assumes the effect of lighting is low-frequency and as we show, it has limited ability to handle configurations like side illumination and high-contrast shading.

Mass Transport for Image Editing. Bonneel et al. [2011] used mass transport to interpolate displacements between high-dimensional distributions, and apply it to problems such as BRDF interpolation and histogram transfer. Solomon et al. [2015] proposed an algorithm for computing optimal mass transport on geometric domains using an approximate distance metric that can be evaluated efficiently using convolutions. Rabin et al. [2012] proposed an efficient approximate mass transport solver that uses a series of 1D histogram matching operations on the axes of the problem space to compute the Sliced Wasserstein distance. They applied this algorithm to the problem of texture mixing. Similarly to these works, our approach uses mass transport but for a different application, portrait relighting. Furthermore, while these existing techniques use mass transport in the domain on which their application data are defined, e.g., the 3D space of colors for histogram transfer, we cast the relighting problem in the higher-dimensional space $\{\text{colors}\} \times \{\text{positions}\} \times \{\text{normals}\}$ to make our algorithm aware of local face geometry.

3 MASS-TRANSPORT FORMULATION

In this section, we describe the core component of our approach, the mass transport formulation. We first express color histogram transfer as a mass-transport problem in the context of portrait relighting, and introduce an algorithm that incorporates positions and normals.

Given an input image I and a reference image R, we create a relit output image O with the lighting of the reference and the pose, identity, and expression of the input photograph. We fit a 3D face model to the 2D portrait using the Expression Flow algorithm [Yang et al. 2011]. We project the 3D positions and surface normals of the 3D model onto the image plane to get per-pixel positions and normals. This gives us a color–position–normal vector $(\mathbf{c}, \mathbf{p}, \mathbf{n})$ at every pixel of the input and reference images. We transfer lighting from the reference to the input image by matching the distributions of these high-dimensional vectors in the joint space $\{\text{colors}\} \times \{\text{positions}\} \times \{\text{normals}\}$. From this matching, we retain only the colors of the transformed input distribution and use them in conjunction with the original positions to create the final relit image. Figure 1 illustrates the pipeline of our technique.

3.1 Reformulating Color Histogram Transfer

Reinhard et al. [2001] matched the color of two images by transferring the color statistics from one image to the other. Pitié et al. [2005] extended this by transferring the full 3D color histogram, thereby encompassing all color-related statistics. As shown in Figure 2, for face relighting, this approach generates approximate results that only capture the overall color and tone of the lighting. We will discuss this issue in more detail later but first, we present the previously known interpretation of the color histogram matching process as a mass-transport problem. This interpretation will be the foundation of our solution to generating better face relighting results.

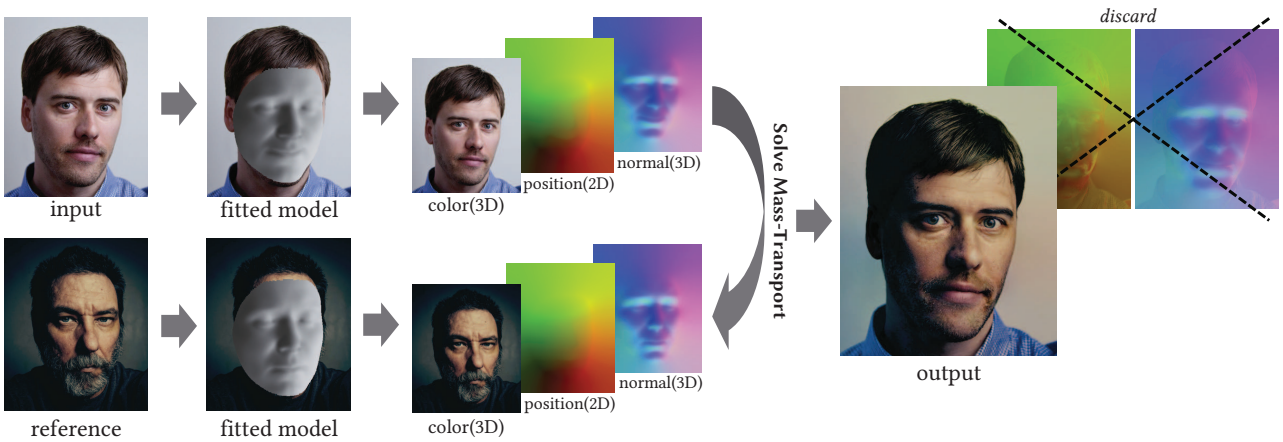


Fig. 1. Pipeline of portrait lighting transfer. We fit a generic 3D face model to the input and reference images. We use this model to extract per-pixel positions (2D) and normals (3D), extrapolate them from the face to the background, and concatenate them with the image RGB channels. We then compute the optimal mass transport from the input to the reference using the 8D data to obtain our output relit image. The output position and normal channels are discarded. Images courtesy: Flickr user *Geoff Stearns* (input), Flickr user *rpavich* (reference).

Histogram transfer is known to be related to mass transport, e.g., [Bonneel et al. 2011]. Intuitively, the input and reference histograms can be seen as sand heaps and one seeks to move the sand to transform the input heap into the reference heap while minimizing the amount of work (defined by the product of the transported mass by the distance over which it is transported). In the context of color histogram transfer, the mass-transport approach seeks to match the target histogram by modifying the input colors as little as possible, which is a desirable property for many applications. Formally, the mass-transport problem in the context of color histogram transfer is defined as follows: We use H_I and H_R to denote the normalized color histograms of the input and reference images, respectively, and i and j to index the input and reference colors. For the distance between input pixel with color i and reference pixel with color j, we use the L_2 norm in color space, $\|\mathbf{c}_i - \mathbf{c}_j\|$, where $\mathbf{c}_i$ and $\mathbf{c}_j$ are 3D vectors representing i and j colors. With this notation, the mass-transport problem seeks to minimize:

$$\arg\min_{T_{ij}} \sum_i \sum_j \|\mathbf{c}_i - \mathbf{c}_j\|^2 \, T_{ij} \tag{1a}$$

$$\text{such that:} \quad T_{ij} \geq 0 \tag{1b}$$

$$\textstyle\sum_j T_{ij} = H_I(\mathbf{c}_i) \tag{1c}$$

$$\textstyle\sum_i T_{ij} = H_R(\mathbf{c}_j) \tag{1d}$$

where the unknowns T_{ij} represent the proportion of the pixels with color i that are assigned to color j. The sum in Equation 1a represents the total amount of work needed to transform H_I into H_R. The constraint (1b) enforces the non-negativity of the masses, and (1c) and (1d) ensure that the entire input histogram is matched to the entire reference histogram. The minimal amount of work (Eq. 1a) is known as the Earth Mover's Distance [Rubner et al. 2000] or the Wasserstein Distance [Villani 2003, 2008] between the input and reference histograms. Equation 1 is often refered to as the Kantorovich formulation of the transport problem and amounts to a linear program [Villani 2003, 2008].

Deriving a Mapping. Transport T creates correspondences between the input and reference colors. The solution of the above mass-transport formulation (Eq. 1) is a *coupling*, i.e., an input color may be associated to more than one reference color. In our context, a coupling of this form is undesirable because it could introduce discontinuities in regions of uniform color. Instead, we seek a solution that is a *mapping*, i.e., all the pixels with the same color are associated to the same reference color. Formally, we are interested in the case where each input color $\mathbf{c}_i$ is assigned a single reference color. We name j_i the index of that reference color and f the function that maps $\mathbf{c}_i$ to $\mathbf{c}_{j_i}$, i.e., $f(\mathbf{c}_i) = \mathbf{c}_{j_i}$. In this context, the transport problem becomes:

$$\arg\min_f \sum_i \|\mathbf{c}_i - f(\mathbf{c}_i)\|^2 \, H_I(\mathbf{c}_i) \tag{2a}$$

$$\text{such that:} \quad H_{f(I)} = H_R \tag{2b}$$

The energy above (Eq. 2) is known as the Monge formulation of the transport problem and unlike the coupling case, it may not always have a solution, for instance when the input and reference images have different numbers of colors. That being said, there exist solvers that provide approximate solutions, e.g., [Bonneel et al. 2011; Rabin et al. 2012]. We use the Sliced Wasserstein Distance algorithm [Rabin et al. 2012] that estimates a mapping f such that $H_{f(I)} \approx H_R$. As we shall see in our validation section, this is sufficient to generate visually pleasing results. We further describe the Sliced Wasserstein Distance solver and its characteristics in Section 3.4.

3.2 Incorporating Positions and Normals

The problem with the color transfer technique discussed in the previous section is that it ignores the geometry of the face. For instance, a pixel on the forehead in the input image might have the same color as a pixel on the cheek. The color-only transfer will map both of these pixels to the same reference color leading to a relit result that does not capture the geometric dependence of lighting changes (Fig. 2c).

(a) Input (b) Reference (c) Color only (d) Color and position (e) Color, position, and normal (f) Our result

Fig. 2. Comparison of different transport strategies. Color-only transfer matches the tone of the lighting but not the angular distribution (c). Incorporating position leads to localized variations (d), which are further improved by adding normals (e). These results have some artifacts because of large jumps in the transport; adding regularization via sampling eliminates them to produce a visually pleasing result (f). Images courtesy: Flickr user *Sven Walter* (input), Flickr user *Brian Holland* (reference).

We address this problem by accounting for the position and normal at each pixel in addition to its color. The position $\mathbf{p}$ and normal $\mathbf{n}$ of a pixel are obtained by fitting a 3D face model to the image and projecting the 3D positions and normals onto the image. Our approach extends the formulation of Equation 2 by adding a position and a normal component to the distance function and to the histograms. For the sake of clarity, we introduce $\mathbf{s} = (\mathbf{c}, \mathbf{p}, \mathbf{n})$, the vector concatenating the color, position, and normal of a pixel. This leads to the following problem:

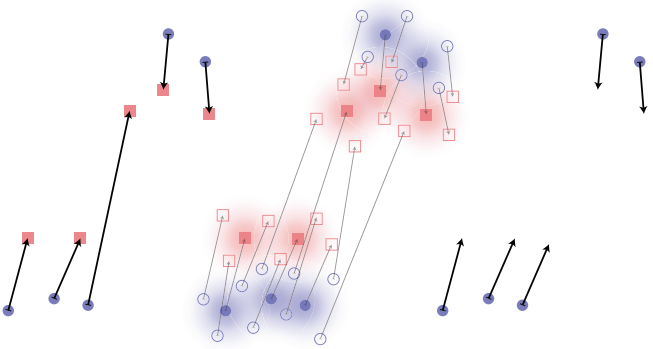
$$\arg\min_{\hat{f}} \sum_i \sum_j \left(w_c \, \|\mathbf{c}_i - \hat{f}_c(\mathbf{s}_i)\|^2 + w_p \, \|\mathbf{p}_i - \hat{f}_p(\mathbf{s}_i)\|^2 + w_n \, \|\mathbf{n}_i - \hat{f}_n(\mathbf{s}_i)\|^2 \right) \widehat{H}_I(\mathbf{s}_i) \tag{3a}$$

$$\text{such that:} \quad \widehat{H}_{\hat{f}(I)} = \widehat{H}_R \tag{3b}$$

where $\widehat{H}_I$ and $\widehat{H}_R$ denote the normalized input and reference histograms in the {colors} × {positions} × {normals} product space, $\hat{f}_c$, $\hat{f}_p$, and $\hat{f}_n$ are the color, position, and normal components of $\hat{f}$, and w_c, w_p, and w_n are weights that control the relative influence of the colors, positions, and normals in the transfer process.

The color-only approach (Eq. 2) minimizes only color variations, and can map spatially distant pixels to each other, as long as their colors are similar enough. Incorporating positions and normals into the transport formulation (Eq. 3) makes the mapping account for facial geometry. The position term $\|\mathbf{p}_i - \hat{f}_p(\mathbf{s}_i)\|$ penalizes long-range pairings, and favors local correspondences that preserve the spatial layout of the reference illumination. Similarly, the normal term $\|\mathbf{n}_i - \hat{f}_n(\mathbf{s}_i)\|$ discourages correspondences between points oriented differently. This is especially useful in regions such as the two sides of the nose that are spatially close but are often lit differently. Figure 2 illustrates the benefits of each term in Equation 3a.

Similar to Equation 2, the transport that minimizes Equation 3 creates a mapping $\hat{f}$ that generates an output with the histogram $\widehat{H}_R$ when applied to I. Because this transport is defined in the product space, it modifies the input pixel positions and normals in addition to their colors, i.e., it alters the colors and warps the face geometry. Since our goal is to transfer only the reference lighting while retaining the input face geometry, we restrict the effect of the transport to the color space c, that is, we only apply $\hat{f}_c$ to the input image I.



(a) Red and blue samples with mismatched distributions generate a discontinuous mapping (black arrows). (b) Stochastic samples (hollow) approximate a Gaussian convolution (shaded) and provide more matching options. (c) The resulting mapping is smoother.

Fig. 3. When two histograms differ, the mass-transport solution often generates an irregular mapping. For example, the input samples at the bottom are close to each other but are mapped to different reference values (a). Adding new samples to approximate Gaussian convolution (b) generates a smoother mapping, i.e., nearby samples are transformed similarly (c).

3.3 Regularization via Stochastic Sampling

So far, we have implicitly assumed that two faces under the same illumination generate images with similar histograms. But in practice, real-world variations in face geometry, pose, and lighting lead to differences in the histograms that can degrade the quality of the result if not handled properly. For example, if the shadow regions on the reference face are larger, a strict transfer introduces shadows in regions of the input face that should be lit. Variations in skin texture can also lead to local perturbations. In terms of transport, two different faces generate distributions with minor mismatches, e.g., with slightly different proportions of dark pixels. This generates

a mapping with discontinuities, which in turn creates the minor degradations seen in Figure 2.

We address this problem by regularizing the transport to prevent large local disparities in the correspondences. Appendix 1 details how we smooth the mapping function $\hat{f}_c$ using stochastic sampling. In short, we replace each sample $(\mathbf{c}, \mathbf{p}, \mathbf{n})$ by n_s stochastic samples with the same position $\mathbf{p}$ and $\mathbf{n}$ but color $\mathbf{c} + \boldsymbol{\nu}$ where the random vector $\boldsymbol{\nu}$ follows a Gaussian distribution G_σ. Intuitively, the additional samples offer new correspondence options that allow the solver to find a smoother mapping. Compared to entropy-based regularization [Solomon et al. 2015], our approach has the advantage that it still produces a mapping. Figure 3 shows the effect of this technique, and Figure 2(e) compares it to the no-regularization case.

3.4 Solving the Mass-Transport Problem

The mass transport formulation described in Equation 3 can be solved using linear program solvers. However these solvers are computationally intractable for a large number of points. Our particular problem is challenging for these methods; relighting a 640×480 image requires estimating the transport of a million points (with stochastic sampling) in an 8D space. Instead of solving for the exact transport, we use the solver proposed by Rabin et al. [2012], that approximates the transport energy (Eq. 1a), a.k.a. the Wasserstein distance, by the sum of 1D transport energies that correspond to projections onto arbitrary axes. This approximation is known as the Sliced Wasserstein Distance and leads to an iterative algorithm that solves a series of 1D mass-transport problems, which can be done efficiently using 1D histogram matching.

The algorithm works as follows: For simplicity, we assume that each pixel corresponds to a unique $(\mathbf{c}, \mathbf{p}, \mathbf{n})$ triplet and that i and j can also be used to index pixels. If two pixels share the sample triplet, we create a sample for each and our solver handles the situation seamlessly. We omit that degenerate case to simplify the presentation and pseudo-code (Alg. 1). We name $\mathbf{s}_i(t) = (\mathbf{c}_i(t), \mathbf{p}_i(t), \mathbf{n}_i(t))$ the input sample points after iteration t. At each iteration, the input and reference points are projected onto the axes of a randomly oriented coordinate system. A 1D histogram transfer along each axis of this coordinate system is performed between the projected input and reference points to create the intermediate input values $\tilde{\mathbf{s}}_i(t)$. The modified input points at the $(t+1)$ iteration are obtained by taking a partial step in this direction, i.e., $\mathbf{s}_i(t+1) = (1-\alpha)\mathbf{s}_i(t) + \alpha\tilde{\mathbf{s}}_i$. In practice, we use $\alpha = 0.2$ as in [Bonneel et al. 2015a] and $n_{\text{iter}} = 300$ iterations. Algorithm 1 summarizes this process.

This algorithm is equivalent to the N-dimensional PDF transfer technique of Pitié et al. [2005], the difference being that they move the input points by a full step (i.e., $\alpha = 1$) in each iteration. In practice, we found that the algorithm by Pitié et al. requires only about 20 iterations, which makes it faster, but its output varies between runs. In comparison, the Sliced Wasserstein solver is slower but produces consistent results. In our paper, we use the latter to demonstrate reproducible results but practitioners who can tolerate result variability and value speed may be better served by the former.

ALGORITHM 1: Face relighting pseudo-code

Input: Input image I and mask M_I
Reference image R and mask M_R

Output: Output image O

```
// Fit a 3D face model to the input and reference portraits
Estimate 3D positions and normals for I and R

// Create arrays of 8D points from the input and reference pixels
S_I ← empty array
for each pixel i of image I do
    if pixel i inside mask M_I then
        for k ← 0 to n_s − 1 do                    // Create n_s stochastic samples
            Draw random 3D vector ν according to G_σ
            S_I[n_s i + k] ← (c_i + ν, p_i, n_i)
        end
    end
end
Construct S_R similarly using the reference pixels

// Initialize the output points with the input pixels
S_O ← empty array
for each pixel i of I do
    S_O[i] ← (c_i, p_i, n_i)
end

// Repeatedly transform the points
for k ← 1 to n_iter do
    Select random 8D coordinate system F
    Express all the samples of S_I, S_R, and S_O in F
    for d ← 1 to 8 do
        Compute 1D histogram transfer function τ_d along d^th axis
    end
    for each sample s of S_I do
        s̃ ← (τ_1(s_1), τ_2(s_2), . . . )           // with s = (s_1, s_2, . . . )
        s ← α s̃ + (1 − α) s
    end
    Transform the samples in S_O similarly
end

// Keep only the color for the output
for each pixel o of image O do
    (c_o, p_o, n_o) ← S_O[o]
    O[o] ← c_o
end
```

3.5 Discussion

While our mass transport formulation globally matches the input and reference distributions, it also captures localized lighting variations. As previously noted, the geometric components of the distributions, positions and normals, encourage mappings between pixels with similar location and orientation, leading to color transfer between consistent geometric regions. Also, instead of committing to the original 3D geometry, our technique implicitly refines the position and normals to better align the images. This makes it robust to errors in the 3D face reconstruction that often compromise the quality of results from previous methods.

(a) Input (b) Reference (c) w_p=w_n=0, and regularization (d) **Our result**, w_p=w_n=1, and regularization (e) w_p=w_n=10, and regularization (f) w_p=w_n=50, and regularization (g) w_p=w_n=50, no regularization

Fig. 4. We show the effect of different position and normals weights on the relit result for the input (a) and reference image (b) shown above. Setting the weights to 0 results in color-only transfer where only the overall color and brightness of the lighting is captured (c). Our technique sets the weights to 1, to transfer the lighting from the left (d). Increasing the weight to 10 (e) and 50 (f) progressively approaches a pure warping technique and the result takes on the appearance of the reference subject. At high weights, our stochastic sampling regularization creates noisy results. Removing the sampling and setting the weights to 50 makes the warping of the reference clear (g). Images courtesy: Flickr user *John Ragai* (input), Flickr user *David Spinks* (reference).

Our technique can be thought of as a combination of two components: geometric warping (adjusting the positions and normals to better match colors) and color blending (manipulating the colors to better explain the geometric correspondences). The influence of each component on the final result is controlled by the weights (w_c, w_p, w_n) of the different terms in the mass transport formulation (Eq. 3). Setting the position and normal weights to zero, i.e, $w_p = w_n = 0$, results in color-only transfer. As seen in Figure 4, when these weights are set to high values, e.g., $w_p = w_n = 50$ and $w_c = 1$, the geometric warping takes over, and the output looks like a warped version of the reference subject. While our stochastic sampling-based regularization plays a crucial role in avoiding artifacts, at high geometric weights, it generates noisy results because the transfer starts approaching per-pixel warping. We found that $w_c = w_p = w_n = 1$ results is a good balance between the components and produces the best results.

4 IMPLEMENTATION

4.1 Implementation

In our prototype, each pixel is represented by its 3D RGB color, a 2D vector describing its position on the fitted face model, and a unit 3D vector for its normal [1]. For the position component, we use the x and y coordinates of the corresponding point on the 3D face model, where x stands for the left–right axis and y for the down–up axis. We also experimented with adding the z coordinate (for the back–front axis) but did not observe any difference. We did not include it in our final prototype to reduce its computational complexity. For the regularization, we use $n_s = 5$ stochastic samples at each pixel and a standard deviation $\sigma = 0.1$ for the Gaussian distribution G_σ (assuming that each RGB channel spans $[0; 1]$). When matching the input image to the reference, we do not explicitly build the 8D histograms or the 16D transport. The Sliced Wasserstein Distance algorithm allows us to maintain a list of all the samples, which can be modified in place. We set $w_c = w_p = w_n = 1$ throughout all experiments. Unless otherwise specified, we used 600×500 images, which took about 3 minutes to process using MATLAB code. We present a faster implementation using two-scale manipulation in section 4.2.

[1]Normals are inherently 2-dimensional quantities, i.e., $\mathbf{n}_z$ can be derived from $\mathbf{n}_x$ and $\mathbf{n}_y$, but for simplicity we use the full 3D normal representation.

Handling non-face regions. The 3D face model gives us positions and normals only on the face region of the portraits. We smoothly extrapolate the positions and normals to the rest of the image by solving a Poisson system with gradients set to zero. We compute the matching functions using only the face samples, but apply them to the entire image. In particular, in every iteration of the matching process, we compute the 1D histogram transfer functions only using the samples on the face mesh, and apply them to every sample. Algorithm 1 summarizes the entire pipeline of our portrait relighting technique.

Skin tone preserving relighting. Operating on the joint space of pixels' colors, positions and normals allows us to transfer the color of both light and skin reflectance from the reference to the input. If the input and chosen reference differ in skin reflectance, the corresponding result may exhibit a shift in the perceived light color (Fig. 5a, c). In our framework, the user can also transfer only the lighting direction, by simply replacing the pixel colors $\mathbf{c}$ with lightness, e.g. L channel in Lab color space, in Algorithm 1, and keeping the color channels (e.g. a and b channels) the same as in the input (Fig. 5d).

4.2 Two-Scale Manipulation

We now introduce our complete pipeline for portrait lighting transfer: a two-scale variant of Algorithm 1 (Fig. 6) that can accelerate lighting transfer for high resolution images, while also reducing the amount of noise in the result (Fig. 7).

With a fixed number of stochastic regularization samples and iterations, the complexity of Algorithm 1 is $O(N \log N)$, with N being the number of pixels in the input-reference image pair. Since this algorithm usually takes hundreds of iterations to converge to a satisfying result, it is prohibitively slow for large input images. However, face illumination effects are mostly large-scale or low-frequency [Haddon and Forsyth 1998; Jacobs et al. 1998]; for

Fig. 5. Skin tone preserving lighting transfer. Relighting the input (a) by the reference (b) in RGB space transfers skin color too (c), while doing so using lightness only (L in Lab) retains the original skin and light colors(d). Images courtesy: Flickr user *Dmitry Kolesnikov* (input), *Sabphoto*/Adobe Stock (reference).

Fig. 6. Two-scale manipulation pipeline. Input (a) and reference (b) are down-sampled to generate a low-resolution relit result 1 (c). Input and up-sampled result 1 are filtered by an edge-preserving filter [He et al. 2013]. The detail (represented by the additive residual (d)) of the input is transferred to the up-sampled-and-filtered result 1 to generate our lighting transfer result (e) at the original resolution. In a post-processing step, the user can choose to replace the background for better visual effect (f). Images courtesy: Flickr user *Geoff Stearns* (input), Flickr user *rpavich* (reference).

example, the residual detail in Figure 6 (d) contains very few illumination cues. As a result, it is possible to relight the image at a coarse scale, and retain the details at original scale.

Our proposed two-scale pipeline, illustrated in Fig. 6, is as follows:

(1) We first downsample the input and reference images to a coarser scale (lower resolution) and generate a relit result at this scale.
(2) Similar to the work of Bae et al. [2006], we decompose the input into two components using an edge-aware smoothing filter [He et al. 2013]: a smoothed *base* and an additive *residual*.
(3) We upsample the result from step (1), and apply step (2) to the upsampled image.
(4) We add up the relit *base* from (3) and the *residual* from (2) to synthesize the final result of the two-scale manipulation.

For input and reference images with resolution 1000×1320, the transport computation time is reduced from about 700 seconds to about 55 seconds with the two-scale manipulation and $\frac{1}{4} \times \frac{1}{4}$ down-sampling in our MATLAB implementation, using the same amount of regularization (4× stochastic sampling) and same number of iterations (200). This scheme also has the effect of smoothing out the contrast enhancement caused by the mass transport operation, leading to less noise and artifacts in the results (Fig. 7). Finally, users can choose to post-process the image by replacing the background of the output for better visual effect (Fig. 6(f)) with existing portrait segmentation tools (e.g. [Shen et al. 2016]).

5 EXPERIMENTAL RESULTS

We now demonstrate our portrait relighting algorithm and compare it to existing techniques. We then show additional results on related applications such as illumination painting, relighting from paintings, and realistic compositing. We provide more results in our online

(a) Result using a single scale (b) Result using two scales (0.25×0.25 down-sampling)

Fig. 7. Our two-scale method accelerates the relighting process, while also alleviating noise from over-sharpened contrast (see zoom-in on left).

supplementary material [2], where we also include examples in higher resolution.

5.1 Results and Qualitative Comparisons

In Figure 18, we transfer lighting from reference images that vary significantly in terms of identity, pose, expression, and gender. In all of these cases, we automatically produce artifact-free results that closely mimic the reference lighting.

Comparisons with Previous Relighting Methods. We compare our technique to three previous face relighting methods. The method of Wen et al. [2003] models facial appearance with spherical harmonics. They fit a morphable model to the portraits and estimate the input and reference radiance maps. With these radiance maps, they compute a ratio image used to relight the input image. For our comparison, we used the same morphable model used in our algorithm. As can be seen in Figure 8(c), this technique relies heavily on the accuracy of the 3D reconstruction and violations of the appearance model or errors in the reconstruction introduce artifacts in the results. In addition, this technique only relights the face region, where the geometry is available. Chen et al. [2011] (Fig. 8(d)) apply edge-preserving filters to separate shading into base and detail layers, and transfer the base illumination from the reference to relight the input. This technique is an improvement over that of Wen et al., but often transfers the reference appearance without capturing the lighting. The portrait style transfer of Shih et al. [2014] relights images by transferring low-frequency color variations. This leads to better results (Fig. 8(e)) than the other two techniques. However, this technique relies on accurate alignment between the two images, and errors in correspondence lead to artifacts when the reference lighting is strongly directional. Our results (Fig. 8(f)) are substantially better on these hard cases. They capture the reference lighting more accurately and do not have local artifacts. This is an advantage of our regularized mass-transport formulation that increases robustness to errors in alignment and challenging lighting conditions.

Comparison with Ground Truth. We also compared our relighting results to ground truth images. We used a multi-light dataset for two subjects, captured in a light stage [Weyrich et al. 2006], to render ground truth images under a set of reference illuminations. We then used images of Subject B, acquired under three different illuminations, to relight an input image of Subject A acquired under a fourth illumination. We then compared our results with the ground truth acquisition of Subject A under the three reference illuminations. As Figure 9 illustrates, our technique is able to capture the angular distribution of the reference illumination well, without any knowledge of the lighting in these images, and in spite of the differences between the two subjects.

Comparison with Regularized Discrete OT. We compare our approach to the regularized discrete optimal transport by Ferradans et al. [2014]. This technique introduces a relaxed mass-preserving constraint to the linear-programming formulation that alleviates visual artifacts caused by exact color matching (that is addressed using stochastic sampling regularization in our method). To make the linear programming solver tractable, they substantially reduce the number of particles by clustering the data (K-means) in color space. In the context of geometry-aware relighting, where position $\mathbf{p}$ and normal $\mathbf{n}$ dimensions are added to the data space, this method is less effective in capturing local color distributions and leads to piecewise color artifacts (Fig. 10). Also, while their results capture the global tone of lighting, they fail directional variations in lighting (Fig. 10(a)), even when geometric dimensions are added to the data (Fig. 10(b)). In contrast, our method faithfully captures the change of the color distribution and generates visually plausible results (Fig. 10(c)).

Recent work in fast regularized optimal transport [Solomon et al. 2015] makes use of Sinkhorn iterations [Cuturi 2013] and requires heat kernel convolutions in the space of the couplings. Because this space is 16-dimensional in the context of relighting (8D for each the input and output spaces), this approach is not practical for our application. In contrast, in our approximation using Sliced Wasserstein Distance [Rabin et al. 2012], the computation cost is only linear with the data dimensionality, and the 1D matching can be parallelized in each iteration.

Moreover, experimentally, we show that the regularization by stochastic sampling, not only reduces visual artifacts, but also increases numerical stability of the results. We generate relighting results multiple times with the proposed Algorithm 1, with different amount of stochastic regularization samples, and random coordinate systems (rotations) $\mathcal{F}$, as well as different update rates τ. The special case of $\tau = 1$ and without stochastic sampling, reproduces the N-dimensional probability distribution transfer algorithm of Pitié et al. [2005] (on color-position-normal space). Figure 11(a) shows the average of the pixel value variances across different trials. With larger τ and less amount of regularization, the results exhibit larger variance. We show the visual differences from errors of pixel values under different parameter settings on our online supplementary material [2] . Proper stochastic sampling regularization and a small update rate significantly reduce visual artifacts. From Figure 11(a) we observe that the variance reduces slowly with larger than 4× regularization. Therefore in our implementation, we fix the number of stochastic sampling to 4. We also show, in Figure 11(b), the convergence properties of the Sliced Wasserstein Distance algorithm

[2]http://www3.cs.stonybrook.edu/~cvl/content/portrait-relighting/prl.html

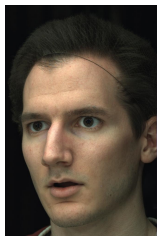
(a) Input (b) Reference (c) [Wen et al. 2003] (d) [Chen et al. 2011] (e) [Shih et al. 2014] (f) Our result

Fig. 8. Comparison with previous work. Input images (a), reference portraits (b), relighting using spherical harmonics-based radiance maps [Wen et al. 2003] (c), with edge-preserving filters [Chen et al. 2011] (d), large-scale illumination transfer [Shih et al. 2014] (e), and our technique (f). Our technique is able to produce robust relit results even when there are significant pose and/or identity variations and the lighting changes are very extreme (e.g., the red-blue split lighting at the bottom). Methods in (c) and (d) only allow manipulation inside the face region (i.e., a fitted morphable model mask region) with no straightforward extension that can apply to a background region. For (e) and (f) we use identical background and identical (manually defined) portrait mask within each row. Images courtesy: Flickr user *Geoff Stearns* ((a)-1), Flickr user *rpavich* ((b)-1), Flickr user *Eva Rinaldi* ((a)-2), *Sabphoto*/Adobe Stock ((a)-3, (b)-3), Flickr user *Rod Waddington* ((a)-4).

w.r.t the number of iterations and update rate. In our implementation, we fix the number of iterations to 300 and $\tau = 0.2$ to balance the running time and visual quality of the results.

5.2 User Study

We conducted a quantitative evaluation of our technique, as well as comparisons with previous state-of-the-art work [Shih et al. 2014], via a user study that evaluates the perceptual quality of the lighting transfer, and the visual realism of results. The input dataset we used is the Flickr dataset of headshot portraits, collected by Shih et al. [2014], which contains 98 high-resolution casual portraits. Our lighting reference dataset consists of 21 portrait photos spanning ethnicity, gender, age, pose, expression and lighting. The user study shows that our technique significantly outperforms prior art in both lighting transfer quality and visual realism of the results.



(a) Input

(b) Reference (c) Relit (Lab) (d) Relit (L) (e) Ground truth

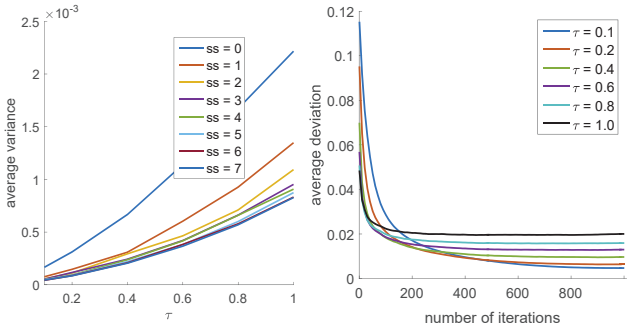
Fig. 9. Relighting the input image (a, rendered using a light stage dataset [Weyrich et al. 2006]) using the reference images (b, rendered under three different light setups) produces results (c, d) that reasonably approximate the angular distribution of the lighting in the ground truth images (e).

Quality of lighting transfer. In this questionnaire, we present users with three images: input, reference and at random, either the full-color transfer result generated by our algorithm with two-scale manipulation O_2^{full} or the result generated by the method of Shih et al. [2014] O^{style}. Both O_2^{full} and O^{style} transfer the skin tone and the light color. All results are generated completely automatically. We ask users to evaluate the quality of the lighting transfer (from the input to the example) using a single choice out of 4 options:

(1) Convincing lighting transfer
(2) Acceptable lighting transfer
(3) Limited lighting transfer
(4) Poor lighting transfer

(a) Regularized OT result on color space (b) Regularized OT result on color, position and normal space (c) Our result (color, position and normal, 1-scale)

Fig. 10. Comparison with regularized OT [Ferradans et al. 2014]. Input: Fig. 2(a); Reference: Fig. 2(b). (a) Regularized OT result on color space only; (b) Regularized OT result on the joint space of color, position and normal(the changes of position and normal are discarded after transport computation); (c) Our result. Our result is better in both lighting transfer and visual quality.



(a) (b)

Fig. 11. The variance/deviation of resulting pixel values, by Algorithm 1, w.r.t update rate (τ), amount of stochastic sampling (ss), and number of iterations. (a) Average of the resulting pixels' variances from 10 trials (with random $\mathcal{F}$ each time). Stochastic sampling and low update rate effectively reduce the variance of results. (b) Average of resulting pixels' deviations. The final choice of τ (0.2) and number of iterations (300) is a balance between result quality and running time. Please refer to the online supplementary material [2] for a visual comparison of results by varying τ and stochastic sampling.

Fig. 12. Statistics of user study responses for lighting transfer quality evaluation. Our technique significantly outperforms previous work [Shih et al. 2014] in the quality of the lighting transfer.

In the 5,629 responses that we collected (Fig. 12), 71% of O_2^{full} are rated 1 (Convincing) or 2 (Acceptable) (37% are rated 1), while 36% of O^{style} are rated 1 or 2 (13% are rated 1).

Visual realism of the results. We evaluate the visual realism of the relit results by comparing the results with unedited real images. In each comparison, we present users with two images simultaneously: (1) an unedited portrait (image from the input dataset); and (2) a relit result, randomly picked from the following

three options: O^{style}, O_2^{full}, and lightness transfer using our two-scale approximation $O_2^{\text{lightness}}$; image (1) and (2) are not necessarily from the same subject. We ask users to decide which image looks more realistic (allowing a tie). From the 2,400 comparisons that we collected, users indicated that 31.7% of O_2^{full}images look more or as realistic as the unedited images, while this number for $O_2^{\text{lightness}}$and $O^{\text{style}}$are 40% and 11.5% respectively.

5.3 Applications

Relighting with non-photorealistic examples. The robustness of our technique to differences in the input and reference images allows us to go beyond actual photographs, and to transfer illumination between real images and paintings. This is demonstrated in Figure 13.

Sketch-based relighting. We can even use a rough user sketch of shadows and highlights to drive the relighting process (Fig. 13-d,e). This could be used as a "lighting design" tool to allow easy exploration of different illuminations.

Compositing with relighting. Compositing images can lead to unnatural results when there are differences in their appearance. While previous work has proposed ways to fix inconsistencies along the boundary [Pérez et al. 2003] and differences in textures and noise [Sunkavalli et al. 2010], lighting is still an issue. By using our technique, users can eliminate lighting inconsistencies between images and create photorealistic composites (Fig. 14).

5.4 Discussion

In some cases, the face might be occluded by hair or accessories with widely different textures. This can lead to artifacts in the results. We can handle such cases by using a user-defined mask (in lieu of the face mask) to drive the transfer. Figure 15 illustrates how this can lead to significantly better results.

Matching a flat histogram to a histogram with strong peaks is challenging with our formulation; this means that we are not able to create sharp highlights, or sharp shadows when the input image is largely flat. We are also limited in our ability to remove very dark shadows; brightening such low SNR regions can render poor results (Fig. 16), which is outside the scope of this work. Moreover, our technique does not handle specularities well. Handling specularities will likely require an explicit separation of specular highlights on the face, and we leave this for future work.

In general, given an input face image, the reference portraits should ideally have similar pose in order to provide sufficient statistics of pixel positions and normals. With the mass transport formulation, our method is robust to a certain range of pose differences. In Figure 17 we show an example of relighting photos taken under different views, using a single reference. We observe that as the pose difference between the input and reference increases, the results become less convincing.

We show in the online supplementary material [2] that applying our technique to video data frame-by-frame generates reasonable relighting results, with the exception of subtle temporal inconsistency on the background. Such temporal fluctuations could be potentially addressed by another line of work [Bonneel et al. 2015b].

Finally, not every reference illumination is compatible with every input image. While our technique does a good transfer of illumination in many cases, some of the resulting outputs may not be aesthetically pleasing because of these incompatibilities (e.g., Fig. 8, bottom).

6 CONCLUSIONS

We have presented an effective algorithm to transfer the lighting from a reference portrait image onto another portrait image of a different subject possibly with a different pose and expression. We have introduced a new formulation of the relighting problem as a mass-transport problem and shown how to regularize it using stochastic sampling. Although our method is predominantly image-based and relies on a global solver, it is able to successfully account for local facial geometry to produce high-quality relit portraits as demonstrated in our results.

REFERENCES

Yael Adini, Yael Moses, and Shimon Ullman. 1997. Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *IEEE Trans. PAMI.* 19, 7 (1997), 721.

Oleg Alexander, Mike Rogers, William Lambeth, Matt Chiang, and Paul Debevec. 2009. The Digital Emily Project: Photoreal Facial Modeling and Animation. In *ACM SIGGRAPH 2009 Courses*. 12.

Soonmin Bae, Sylvain Paris, and Frédo Durand. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph.* 25, 3 (2006), 637.

Dmitri Bitouk, Neeraj Kumar, Samreen Dhillon, Peter Belhumeur, and Shree K Nayar. 2008. Face swapping: automatically replacing faces in photographs. *ACM Trans. Graph.* 27, 3 (2008), 39.

Volker Blanz and Thomas Vetter. 1999. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*. 187.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. 2015a. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision* 51, 1 (2015), 22.

Nicolas Bonneel, James Tompkin, Kalyan Sunkavalli, Deqing Sun, Sylvain Paris, and Hanspeter Pfister. 2015b. Blind Video Temporal Consistency. *ACM Trans. Graph.* 34, 6 (2015).

Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. 2011. Displacement interpolation using lagrangian mass transport. *ACM Trans. Graph.* 30, 6 (2011), 158.

Xiaowu Chen, Mengmeng Chen, Xin Jin, and Qinping Zhao. 2011. Face illumination transfer through edge-preserving filters. In *Proc. CVPR*. IEEE, 281–287.

Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In *Proc. NIPS*. 2292–2300.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the Reflectance Field of a Human Face. In *Proc. SIGGRAPH*. 145.

Sira Ferradans, Nicolas Papadakis, Gabriel Peyré, and Jean-François Aujol. 2014. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences* 7, 3 (2014), 1853.

Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. 2001. From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose. *IEEE Trans. PAMI.* 23, 6 (2001), 643.

John Haddon and David Forsyth. 1998. Shading primitives: Finding folds and shallow grooves. In *Proc. ICCV*. IEEE, 236–241.

Kaiming He, Jian Sun, and Xiaoou Tang. 2013. Guided image filtering. *IEEE Trans. PAMI.* 35, 6 (2013), 1397.

David W Jacobs, Peter N Belhumeur, and Ronen Basri. 1998. Comparing images under variable illumination. In *Proc. CVPR*. IEEE, 610–617.

Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient Attributes for High-level Understanding and Editing of Outdoor Scenes. *ACM Trans. Graph.* 33, 4 (2014), 1.

Zicheng Liu, Ying Shan, and Zhengyou Zhang. 2001. Expressive Expression Mapping with Ratio Images. In *Proc. SIGGRAPH*. 271–276.

Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. 2007. Post-production facial performance relighting using reflectance transfer. *ACM Trans. Graph.* 26, 3 (2007), 52.

(a) Input (b) Reference (c) Our result (d) Reference (e) Our result

Fig. 13. Our technique can also be used with non-photorealistic images. We can transfer lighting from a real image to a painting (top, left), between paintings (top, right), and from a painting to a real portrait (bottom, left). In addition, our technique allows for an intuitive way to "design" lighting; users can scribble shadows and bright regions on a photograph and use it as a reference to relight the image. Image courtesy: Flickr user *Eva Rinaldi* (portrait of *Julie Andrews)*.

(a) Reference (b) Input (c) Composite (d) Relit output (e) Relit composite

Fig. 14. Face Swapping: when the eyes and nose in this input image (b) are Poisson-blended with the reference (a), the difference in the illumination gradient causes unrealistic color shifts (e.g., the unnatural cheek color). Our technique eliminates the lighting inconsistency (d), leading to a photorealistic result (e). Images courtesy: Flickr user *Geoff Stearns* (input), Flickr user *rpavich* (reference).

Patrick Pérez, Michel Gangnet, and Andrew Blake. 2003. Poisson image editing. *ACM Trans. Graph.* 22, 3 (2003), 313.

Francois Pitie, C. Kokaram Anil, and Rozenn Dahyot. 2005. N-dimensional probability density function transfer and its application to color transfer. In *Proc. ICCV*. IEEE, 1434–1439.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. 2012. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*. Springer, 435.

E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. 2001. Color transfer between images. *Computer Graphics and Applications, IEEE* 21, 5 (2001), 34–41.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40, 2 (2000), 99.

Amnon Shashua and Tammy Riklin-Raviv. 2001. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Trans. PAMI.* 23, 2 (2001), 129.

Xiaoyong Shen, Aaron Hertzmann, Jiaya Jia, Sylvain Paris, Brian Price, Eli Shechtman, and Ian Sachs. 2016. Automatic portrait segmentation for image stylization. *Computer Graphics Forum* 35, 2 (2016), 93.

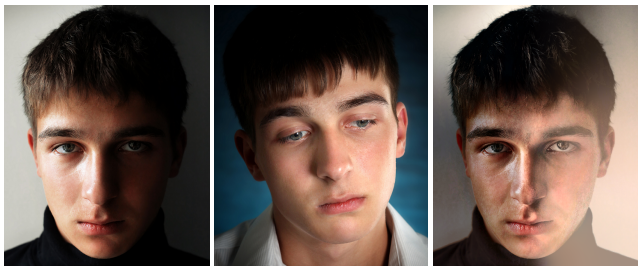
YiChang Shih, Sylvain Paris, Connelly Barnes, William T Freeman, and Frédo Durand. 2014. Style transfer for headshot portraits. *ACM Trans. Graph.* 33, 4 (2014), 148.

Yichang Shih, Sylvain Paris, Frédo Durand, and William T. Freeman. 2013. Data-driven Hallucination of Different Times of Day from a Single Outdoor Photo. *ACM Trans. Graph.* 32, 6 (2013), 200.

(a) Input and user mask (b) Reference and user mask

(c) Relit with face mask (d) Relit with user mask

Fig. 15. Input and reference portraits with overlaid user-specified masks (a,b). Matching using the face mask (including forehead) causes artifacts in the forehead region because of occlusions by the hat (c), but using the user specified mask avoids these issues (d). *Image courtesy: Flickr user Eva Rinaldi (Input), Flickr user Loren Kerns (Reference).*



(a) Input (b) Reference (c) Our result

Fig. 16. Relighting an input image with harsh lighting brightens the shadowed regions that are typically noisy. The low image quality in these regions leads to poor results. Images courtesy: *Sabphoto*/Adobe Stock (input and reference).

Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. 2015. Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains. *ACM Trans. Graph.* 34, 4 (2015), 66.

Kalyan Sunkavalli, Micah K Johnson, Wojciech Matusik, and Hanspeter Pfister. 2010. Multi-scale image harmonization. *ACM Trans. Graph.* 29, 4 (2010), 125.

Cédric Villani. 2003. *Topics in Optimal Transportation (Graduate Studies in Mathematics, Vol. 58).* American Mathematical Society.

Cédric Villani. 2008. *Optimal Transport: Old and New* (1 ed.).

Yang Wang, Lei Zhang, Zicheng Liu, Gang Hua, Zhen Wen, Zhengyou Zhang, and Dimitris Samaras. 2009. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Trans. PAMI.* 31, 11 (2009), 1968.

Zhen Wen, Zicheng Liu, and Thomas S Huang. 2003. Face relighting with radiance environment maps. In *Proc. CVPR*, Vol. 2. IEEE, II–158.

Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross. 2006. Analysis of Human Faces Using a Measurement-based Skin Reflectance Model. *ACM Trans. Graph.* 25, 3 (2006), 1013.

Fei Yang, Jue Wang, Eli Shechtman, Lubomir Bourdev, and Dimitri Metaxas. 2011. Expression Flow for 3D-aware Face Component Transfer. *ACM Trans. Graph.* 30, 4 (2011), 60.

APPENDIX: ADDING COLOR NOISE TO REGULARIZE THE MAPPING

In this section, we explain how we regularize the mapping $\hat{f}_c$ by adding color noise to the data. Intuitively, we seek to smooth $\hat{f}_c$ which we could naively do by convolving it with a Gaussian for instance. However, this would imply that we explicitly compute, store, and process an 8D function, which is not practical. Instead, in this section, we explain how to modify the input and reference data so that solving the mass transport problem with the Sliced Wasserstein algorithm directly generates a smooth function. While a result independent of the solver being used would be more desirable, the smoothness of the transport function is known to be a thorny problem even in seemingly simple cases [Villani 2008, Chapter 12].

Our approach builds upon the observation that continuous functions are smooth and that Gaussian convolution is an effective way to smooth a function. We first consider the case of two 1D datasets $\{U_i\}$ and $\{V_i\}$, e.g., gray-scale images. It is known that histogram transfer is achieved by computing the normalized histograms H_U and H_V, and their corresponding cumulative distribution functions $C_U(z) = \int_{-\infty}^{z} H_U$ and $C_V(z) = \int_{-\infty}^{z} H_V$, and composing them to get the transfer function $\tau = C_V^{-1} \circ C_U$. For τ to be continuous, it is sufficient that C_U and C_V^{-1} are continuous. For C_U, a convolution by a Gaussian kernel G_σ is sufficient, i.e., we use $G_\sigma \otimes C_U$ instead of C_U. And since C_V is monotonically increasing from 0 to 1, it is also sufficient to convolve C_V by a Gaussian kernel to ensure that C_V is invertible and C_V^{-1} is continuous. Last, we observe that the integral commutes with the convolution, i.e., $G_\sigma \otimes C_U(z) = G_\sigma \otimes \int_{-\infty}^{z} H_U = \int_{-\infty}^{z} (G_\sigma \otimes H_U)$, which means that it is sufficient to apply the convolution on the histograms. Further, since the Sliced Wasserstein solver handles high-dimensional datasets by repeatedly applying one-dimensional transfers, the above result on an 1D dataset is sufficient to guarantee that convolving the input data with a Gaussian kernel generates a continuous mapping.

In our context, the data are higher-dimensional because each data point is represented by a $(\mathbf{c}, \mathbf{p}, \mathbf{n})$ vector (Eq. 3). However, only the colors may have a discontinuous distribution — the positions form a uniform distribution over the image domain, and the normals are smoothly distributed because faces are smooth shapes. Because of this, convolving the positions and normals with a Gaussian has a negligible effect since it amounts to applying a low-pass filter to band-limited data. In practice, we only smooth the color components of the data. Because of the high dimensionality of the data, explicitly representing the histograms to perform the convolutions

Fig. 17. Effect of pose difference between input and reference. We relight photos under different views (top) using a single frontal-facing reference image (bottom, first). The results (bottom) look best when the input and reference image have similar pose, and become less plausible at large angles.

is impractical. Instead, we use stochastic sampling, that is, we replace each color value by n_s samples randomly generated from a G_σ distribution centered on the original color; we keep the position and normal unchanged. This achieves our goal, that is, by replacing each input and reference point $(\mathbf{c}, \mathbf{p}, \mathbf{n})$ by n_s stochastic samples $(\mathbf{c} + \boldsymbol{\nu}, \mathbf{p}, \mathbf{n})$ where $\boldsymbol{\nu}$ is a random 3D vector drawn from Gaussian distribution G_σ, we ensure that the solution of the mass-transport problem (Eq. 3) directly generates a smooth mapping $\hat{f}_c$ and we do this only by manipulating discrete samples and never explicitly represent an 8D function nor perform a 16D convolution.

In our prototype, we speed up the computation by using only $(n_s - 1)$ stochastic samples and adding the original samples to the set of samples used to compute the transport ($\mathcal{S}_I$ in Algorithm 1). This simplifies the computation because we do not need to maintain a separate sample set for the original samples and the stochastic samples. Formally, this amounts to sampling the distribution $\frac{1}{n_s}\delta + \frac{n_s-1}{n_s}G_\sigma$ with δ the Dirac distribution. This is a close approximation of G_σ and has little influence on the results.

Received February 2007; revised March 2009; final version June 2009; accepted July 2009

Fig. 18. Relighting results on a set of portrait images (shown on the diagonal in red boxes). The same original photograph is used as an input image (rows) and as a reference image (columns), i.e., the image on (row a, col b) uses (a,a) as the input and (b,b) as the reference. As this figure shows, our technique is able to transfer illumination between images, in spite of differences in pose, expression, gender, and ethnicity. Image courtesy: Flickr user *Eva Rinaldi* (1,1), Flickr user *Abhi* (2,2), *Ari Levinson* (3,3), Flickr user *DoD News* (4,4).