Face Swapping: Automatically Replacing Faces in Photographs

Dmitri Bitouk Neeraj Kumar Samreen Dhillon* Peter Belhumeur Shree K. Nayar
Columbia University†

(a) Original photographs (b) After automatic face replacement

Figure 1: *We have developed a system that automatically replaces faces in an input image with ones selected from a large collection of face images, obtained by applying face detection to publicly available photographs on the internet. In this example, the faces of (a) two people are shown after (b) automatic replacement with the top three ranked candidates. Our system for face replacement can be used for face de-identification, personalized face replacement, and creating an appealing group photograph from a set of "burst" mode images. Original images in (a) used with permission from Retna Ltd. (top) and Getty Images Inc. (bottom).*

Abstract

In this paper, we present a complete system for automatic face replacement in images. Our system uses a large library of face images created automatically by downloading images from the internet, extracting faces using face detection software, and aligning each extracted face to a common coordinate system. This library is constructed off-line, once, and can be efficiently accessed during face replacement. Our replacement algorithm has three main stages. First, given an input image, we detect all faces that are present, align them to the coordinate system used by our face library, and select candidate face images from our face library that are similar to the input face in appearance and pose. Second, we adjust the pose, lighting, and color of the candidate face images to match the appearance of those in the input image, and seamlessly blend in the results. Third, we rank the blended candidate replacements by computing a match distance over the overlap region. Our approach requires no 3D model, is fully automatic, and generates highly plausible results across a wide range of skin tones, lighting conditions, and viewpoints. We show how our approach can be used for a variety of applications including face de-identification and the creation of appealing group photographs from a set of images. We conclude with a user study that validates the high quality of our replacement results, and a discussion on the current limitations of our system.

CR Categories: I.4.9 [Artificial Intelligence]: Image Processing and Computer Vision—Applications H.3.3 [Information Systems]: Information Storage and Retrieval—Search and selection process

Keywords: Face Replacement, Image Databases, Image-Based Rendering, Computational Photography

1 Introduction

Advances in digital photography have made it possible to capture large collections of high-resolution images and share them on the internet. While the size and availability of these collections is leading to many exciting new applications, it is also creating new problems. One of the most important of these problems is privacy. Online systems such as Google Street View (http://maps.google.com/help/maps/streetview) and EveryScape (http://everyscape.com) allow users to interactively navigate through panoramic images of public places created using thousands of photographs. Many of the images contain people who have not consented to be photographed, much less to have these photographs publicly viewable. Identity protection by obfuscating the face regions in the acquired photographs using blurring, pixelation, or simply covering them with black pixels is often undesirable as it diminishes the visual appeal of the image. Furthermore, many

*samreendhillon@gmail.com
†{bitouk,neeraj,belhumeur,nayar}@cs.columbia.edu

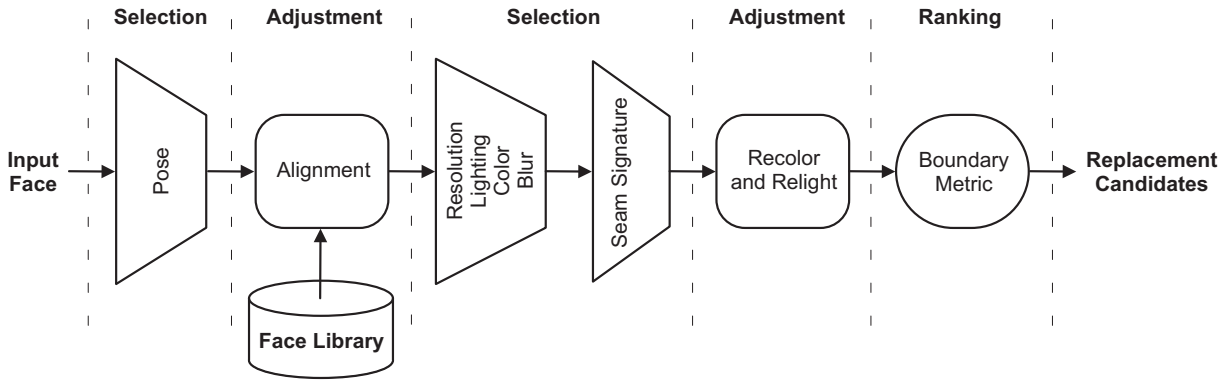

Figure 2: *The main steps of our automatic face replacement algorithm. Given an input face that needs to be replaced, we first search the face library for faces that have similar pose, resolution, image blur, lighting, and seam signature to the input face. Next, we adjust the color and lighting of the selected faces to match those of the input face. Finally, we rank the replacement candidates using a boundary metric. Given the large and diverse nature of our face library, the top ranked candidates almost always correspond to highly realistic face replacements.*

of these methods are currently applied manually, on an image-by-image basis. Since the number of images being captured is growing rapidly, any manual approach will soon be intractable. We believe that an attractive solution to the privacy problem is to remove the identities of people in photographs by automatically replacing their faces with ones from a collection of stock images.

Automatic face replacement has other compelling applications as well. For example, people commonly have large personal collections of photos on their computers. These collections often contain many photos of the same person(s) taken with different expressions, and under various poses and lighting conditions. One can use such collections to create novel images by replacing faces in one image with more appealing faces of the same person from other images. For group shots, the "burst" mode available in most cameras can be used to take several images at a time. With an automatic face replacement approach, one could create a single composite image with, for example, everyone smiling and with both eyes open.

In this paper, we present a complete system for fully-automatic face replacement in photographs. (Figure 1 shows example results.) We take advantage of the large number of publicly available images on the web, as well as a high-quality commercial face detector, to build a large library of face images for de-identification. Each face within the library is cropped from its original image, labeled with yaw and pitch pose angles estimated by the face detector, binned into one of several pose bins, and then aligned to a coordinate system common to all images in the chosen pose bin. For the personalized replacement application mentioned previously, we can build smaller, non-generic face libraries from the users' personal photo collections.

The basic steps of our replacement approach are shown in Figure 2. When our system is supplied with an input image containing a face to be replaced, it performs face detection to extract the face, estimates the pose, and aligns the face to the appropriate pose bin-specific coordinate system. The system then looks into the face library to select possible candidate faces to use for replacement. Note that only candidate faces within the same pose bin of the library are considered; this ensures that replacement faces will be relatively similar in pose, thus allowing the system to use simple 2D image-compositing instead of 3D model-based approaches which require precise alignment. In addition, the system requires that the selected candidate faces are similar to the input face in image quality, color, lighting, and the boundary of the replacement region. Once possible candidate faces have been selected, the system transforms the color and lighting of these candidate faces to match those of the input face and blends the results into the input photograph. As a final step, to weed out inferior replacements, the system ranks the resulting images according to how well the adjusted candidate replacement face fits the surrounding region in the original photograph and chooses the highest ranked replacement.

A key contribution of our work is that it enables automatic replacement of faces across different pose, lighting, facial expression, image resolution, image blur, and skin tone – all without using 3D reconstruction techniques or manual assistance. We demonstrate how our approach can be applied to large-scale face de-identification, as well as a number of image manipulation tasks such as face swapping and creating composite group photographs. Without automation, it would be very difficult (or even impossible) to tackle these applications, setting our work in a different class from previous non-automatic approaches such as [Blanz et al. 2004]. We present results of a user study which shows that people are almost equally likely to classify real face images and our replaced face images as being real. We conclude with a detailed discussion on the limitations of our current system and areas for future work.

2 Related Work

Our approach is related to previous work in several fields, including computer vision, computer graphics, and privacy protection.

Face Replacement: While there exists a rich body of work on replacing parts of images with new image data, the replacement of faces in images has received relatively little attention. To the best of our knowledge, the work of [Blanz et al. 2004] is the only published approach which allows one to replace faces in photographs. They fit a morphable 3D model to both the input and target face images by estimating shape, pose and the direction of the illumination. The 3D face reconstructed from the input image is rendered using pose and illumination parameters obtained from the target image. The major drawback of this approach is that it requires manual initialization in order to obtain accurate alignment between the morphable model and the faces in both the input and target images. Although this is acceptable for their goal (virtual hairstyle try-on), our de-identification application absolutely requires that there be no user intervention. A commercial system that also uses 3D models is currently in development at XiD Technologies (http://xidtech.com), but details regarding their technical approach, the degree of automation, and the quality of their results are not known. (The estimation of 3D face shape from a single image is an inherently under-constrained problem and by nature difficult to fully automate.) In contrast, our approach allows us to automatically replace faces across different pose and lighting conditions without resorting to 3D methods.

An unpublished work [Malik 2003] describes another 3D model-based approach for face replacement. This work focuses on im-

proved relighting and recoloring using histogram matching and color blending. However, this system requires manual face alignment to the 3D model, and some post-processing is needed to improve the visual quality of the results. Finally, [Liu et al. 2001] addresses the related, but slightly different problem of transferring expressions of faces between two images. This work introduces a relighting technique which is similar to the one we use for our system, described in Section 5.1.

Face De-Identification: The easiest and most well-known method for face de-identification is to distort the image either through pixelation or blur [Boyle et al. 2000], although the results in this case are not as visually appealing. In [Newton et al. 2005], a face de-identification algorithm is introduced which minimizes the probability of automatic face recognition while preserving details of the face. However, since this technique uses Principal Component Analysis to compute averages of images, the replaced faces contain blurring and ghosting artifacts. The work of [Gross et al. 2006] improves the quality of de-identified faces using Active Appearance Models [Cootes et al. 2001] but still suffers from blurring artifacts. Moreover, all of these face de-identification methods work only on faces in frontal pose, and produce images inside a pre-defined face region without any guarantee that the de-identified face will blend well with the rest of the original photograph. Our work differs in that we automatically select faces which yield realistic final results for input faces in different poses, and we perform critical appearance adjustments to create a seamless composite image.

Image Compositing: Some of the applications we describe can be addressed using image compositing approaches. For example, the Photomontage framework [Agarwala et al. 2004] allows a user to interactively create a composite image by combining faces or face parts taken from several source photographs. In [Wang et al. 2007a], faces are replaced using gradient domain image blending. Another approach which uses a large image library as we do is [Hays and Efros 2007], in which images can be "completed" using elements taken from similar scenes. These image compositing methods are not specifically targeted to face images, however, and they require user interaction to create plausible replacement results. In contrast, our algorithm focuses on faces and can automatically generate a ranked set of replacement results.

3 Creating a Face Library for Replacement

In order to replace a face in a photograph without using 3D reconstruction techniques, our approach is to find candidate replacement faces whose appearance attributes are similar to those of the input face. This requires us to create a rich library of face images. We construct our face replacement library by searching websites such as Flickr and Yahoo Images, using keyword queries likely to yield images containing faces. As the images are downloaded, we automatically process each one to extract faces using the OKAO Vision face detection software [Omron 2007]. Duplicates and tiny faces are discarded from the face replacement library. Saturated and under-exposed facial images are automatically filtered out as well. At present, our face replacement library contains about 33,000 faces, and this number is growing with time.

The OKAO Vision face detector provides us with an estimate of the face orientation in the photograph, given by the yaw, pitch, and roll angles. To replace faces across different orientations, we assign each face into one of 15 pose bins using its out-of-plane rotation angles. Since the current version of the OKAO detector is less reliable for extreme poses, we restricted ourselves to poses within $\pm 25^\circ$ in yaw and $\pm 15^\circ$ in pitch. The pose bins, shown in Figure 3, span intervals of 10° from -25° to 25° for the yaw and from -15° to 15° for the pitch. We also add the mirror image of each face, so as to increase the number of candidates in our library.

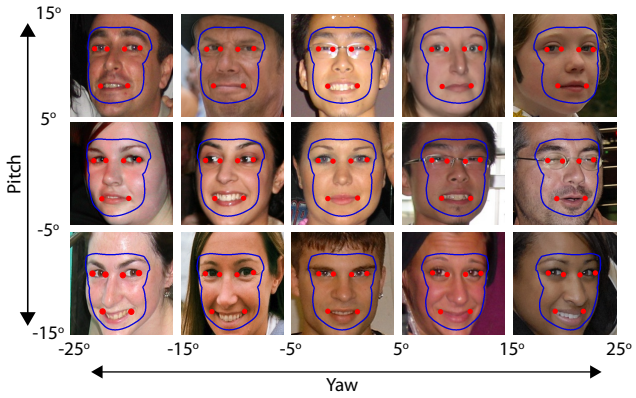

Figure 3: *The six fiducial points (red) and the outline of the replacement regions (blue) for each of the pose bins. Note that these are all defined in 2D, and require no 3D geometry.*

To replace an original face in a photograph with a candidate face from the library, we need to align both images to a common coordinate system. Our alignment is computed from the location of six fiducial points: the inner and outer corners of both eyes and the corners of the lips, as shown by the red dots in Figure 3. (These fiducial points are automatically found using the OKAO detector [Omron 2007].) For each of the pose bins, we define a generic face – with predetermined fiducial points – oriented at the center of the bin. To align a face image to the common coordinate system, we estimate an affine transformation between the fiducial points of our face image and the corresponding fiducial points of the generic face. We then apply this transformation to the face image to bring it into alignment.

Given an input face image, we first identify its pose bin and align it to the corresponding coordinate system. Our main technical challenges are to first select from the pose bin of the input image a number of good replacement candidates, then to apply adjustments to the candidates such that their appearances are consistent with the input image, and finally replace the input image with the candidates. The first two steps (selection and adjustment) will be discussed in the following sections. The final replacement step is simple. Since the input and candidate faces are aligned, we copy over the pixels in the corresponding replacement regions, outlined in blue in Figure 3, from the candidate face to the input face. To ensure that the result is seamless, we apply feathering over a small number of pixels along the region boundary. Since this replacement result is in the coordinate system used for alignment, it is then transformed back to the coordinate system of the original face image. Examples of replacements created after just this alignment step – without selection and adjustment – are shown in the first column of Figure 4c.

4 Appearance-Based Selection

When given an aligned input face to replace, the first step in our approach is to select candidate faces from our library which yield plausible replacements. We define a number of attributes that characterize similarity of face appearance in images. In this section, we describe these attributes and the corresponding match criteria used to select candidate replacement faces from the library.

4.1 Pose, Resolution, and Image Blur

In order to produce a perceptually realistic replacement image, the poses of the input and replacement faces must be quite similar – more similar even than would be guaranteed by belonging to the same pose bin. This is because, while an in-plane rotation between the two faces can be compensated using the alignment procedure

described in Section 3, large out-of-plane rotations, which are given by the yaw and pitch angles, are hard to adjust using image-based approaches. Therefore, we select faces from the library whose yaw and pitch angles differ by no more than 3° from the yaw and pitch of the original face.

It is also important to ensure that replacement faces have similar resolutions and blur properties. Significant differences in either attribute would cause a noticeable mismatch between the inside and outside regions of replaced faces. We define the resolution of facial images using the distance between the centers of the eyes. Since higher resolution images can always be downsampled, we only have to define a lower bound on the resolution of candidate faces. Therefore, we select faces from the library whose eye distance is at least 80% of the eye distance of the face to be replaced.

While there exists extensive work on estimating blur in images [Kundur and Hatzinakos 1996; Fergus et al. 2006], we use a simple heuristic metric to measure the similarity of the degree of blur in two images. This blur distance compares the histograms of the image gradient magnitude in the eye region. First, we normalize the grayscale intensity in the eye region for each of the aligned facial images to zero mean and unit variance. Second, we compute histograms $h^{(1)}$ and $h^{(2)}$ of the gradient magnitude in the normalized eye regions. Since high values of the gradient magnitude are usually associated with sharp edges, the higher-index bins of the histograms are more indicative of the blur amount. Therefore, we multiply the histograms by a weighting function which uses the square of the histogram bin index, n: $\tilde{h}^{(i)}(n) = n^2 h^{(i)}(n),\ i = 1, 2$. Finally, we compute the blur distance as the Histogram Intersection Distance (HID) [Rubner et al. 2000] between the two weighted histograms, $\tilde{h}^{(1)}$ and $\tilde{h}^{(2)}$, as follows: $d_B = HID(\tilde{h}^{(1)}, \tilde{h}^{(2)})$. Only images with a weighted distance in the top 50% are kept as candidates.

4.2 Color and Lighting

The appearance of a face in a photograph is greatly affected by the incident illumination in the scene and the skin color of the face. If we attempt to replace a face with another face which was captured under significantly different illumination or with a large difference in skin color, the replacement result would appear perceptually incorrect. Although our recoloring and relighting algorithm, presented in Section 5.1, allows us to adjust for small differences in color and lighting between the two faces, drastic variations of illumination in terms of shadows and dynamic range are much harder to handle. Instead, we take advantage of the fact that our library should already contain faces captured under illuminations similar to that of the face to be replaced, and with similar skin color. For example, frontal flash images are especially common in our library, and thus our relighting technique can easily handle such cases, given that we find suitable frontal flash candidate faces. Our approach is to estimate the lighting and average color within the replacement region for each of the aligned faces in the library and, given an input face, to select faces whose lighting and color are fairly similar to the input face.

Since we only have a single image of a face, illumination in the scene cannot be accurately recovered using traditional techniques that measure or control the lighting [Debevec 1998; Debevec et al. 2000]. Instead, we use a face relighting method similar to the ones used in [Wen et al. 2003] and [Wang et al. 2007b]. We represent the face shape as a cylinder-like "average face shape," aligned to the coordinate system of the corresponding pose bin. We use a simple orthographic projection to define the mapping from the surface to the face. Furthermore, we assume that faces are Lambertian, and the image intensity $I_c(x, y)$ of the face replacement region in each of the RGB color channels can be approximated as $\tilde{I}_c(x, y)$ using

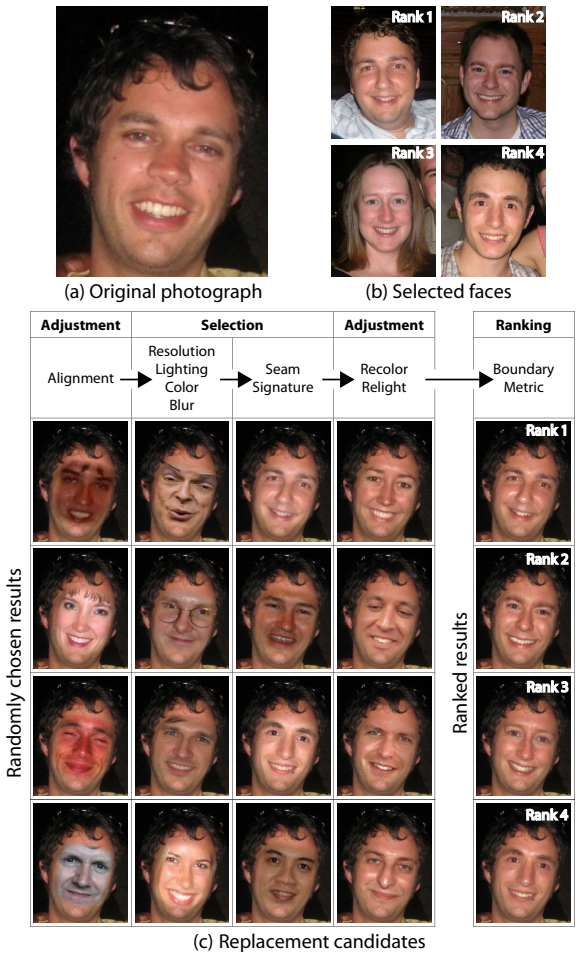

Figure 4: *(a) An original photograph, (b) the faces for the top ranked replacements and (c) face replacement results after each step in our algorithm (as illustrated in Figure 2). Thus, the first column shows replacement results matching only the pose of the two faces, without any kind of selection or appearance adjustment. The subsequent columns show the results after adding basic selection, seam-signature filtering, appearance adjustments, and ranking, respectively. The results get better in each column. Note that since there is no notion of order prior to the last column, we show randomly selected replacements.*

a linear combination of 9 spherical harmonics [Ramamoorthi and Hanrahan 2001; Basri and Jacobs 2003]:

$$\tilde{I}_c(x, y) = \rho_c \sum_{k=1}^{9} a_{c,k} H_k(\mathbf{n}(x, y)),\ c \in \{R, G, B\}, \qquad (1)$$

where $\mathbf{n}(x, y)$ is the surface normal at the image location (x, y), ρ_c are the constant albedos for each of the three color channels (which represent the average color within the replacement region), the coefficients $a_{c,k}$ describe the illumination conditions, and $H_k(\mathbf{n})$ are the spherical harmonic images.

Since the spherical harmonics $H_k(\mathbf{n})$ do not form an orthonormal basis in the replacement region, we cannot use the l_2 distance between the coefficients $a_{c,k}$ as a similarity measure of the lighting between two faces. Instead, we create an orthonormal basis $\psi_k(x, y)$ by applying the Gram-Schmidt orthonormalization to the harmonic basis $H_k(\mathbf{n})$. The approximate image intensity $\tilde{I}_c(x, y)$ can thus be expanded using this orthonormal basis as

$$\tilde{I}_c(x, y) = \rho_c \sum_{k=1}^{9} \beta_{c,k} \psi_k(x, y),\ c \in \{R, G, B\}. \qquad (2)$$

We estimate the 3 albedos ρ_c and the 27 illumination coefficients $\beta_{c,k}$ by minimizing the sum of squared differences (SSD) between the right-hand side of Equation 2 and the aligned face image $I_c(x, y)$ within the replacement region.

We convert the RGB albedos to the HSV color space and use the l_∞ metric to compare the average color within the replacement regions of the input face image $I^{(1)}$ and the replacement candidate $I^{(2)}$. Only those candidates whose hue and saturation are within 5% and brightness within 10% of the input image are kept. To compare the illuminations, we define the lighting distance d_L as the l_2 distance between corresponding lighting coefficients (and keep only the top 50%):

$$d_L(I^{(1)}, I^{(2)}) = \left(\sum_{c \in \{R,G,B\}} \sum_{k=1}^{9} \left(\beta_{c,k}^{(1)} - \beta_{c,k}^{(2)} \right)^2 \right)^{1/2} . \quad (3)$$

The second column of Figure 4c shows replacement results after selection based on resolution, blur, color, and lighting. Notice that the results are, in general, much better than those in the previous column (without pruning based on attributes).

4.3 Seam Signature

Although our selection process so far has already removed many unsuitable candidates for replacement, another important criteria to match is the appearance of the face along the boundary of the replacement region. Differences across this boundary (e.g., caused by facial hair, eyebrows, and hair covering the forehead) can produce visible artifacts in the final output, even after image blending. To avoid these problems, we introduce a simple filter which uses a "signature" of the seam along the replacement boundary. We first resize each aligned image in the replacement library to 256x256 pixels and then define the seam to be a strip containing all pixels inside the replacement region within a 6 pixel radius of its boundary. We create the seam signature by unfolding the seam into a rectangular image, and normalize it so that the average intensity is the same for all faces in the library. This seam signature provides an efficient representation of the texture along the replacement mask boundary. To reduce the dependence of the seam signatures on lighting, we compare the seam signatures using the L_2 distance of the absolute value of the gradient in the direction along the seam. To avoid penalizing gradual changes in appearance, we use a distance of 0 for all pixels within 8% of each other, only using the L_2 distance for pixels which differ by more than this amount. The better quality of replacement results in the third column of Figure 4c shows that this criteria is important for filtering faces with significant differences along the boundary of the replacement region.

4.4 Searching the Library

Selecting candidate faces from the library using the various appearance attributes introduced in this section is a nearest neighbor search problem. This can be computationally intensive due to the high dimensionality of the blur, illumination and seam signature features. To speed things up, we use a sequential selection approach. Given a query face, we first execute a fast SQL query to select faces whose pose, resolution and average colors (given by the albedo $\rho_c, c \in \{H, S, V\}$) are close to those of the input face. This step allows us to reduce the number of potential candidate replacements from 33,000 to just a few thousand faces. Next, we further prune the list of candidates using the blur distance d_B and, subsequently, the lighting distance d_L. Finally, we select the top 50 candidate faces which match the seam signature of the input face. By running these steps in increasing order of complexity, our C++ implementation of the appearance-based selection algorithm requires less than a second to generate a list of candidate replacements for an input face image.

Figure 5: *Color and lighting adjustment. We replace (a) the face in the input photograph with (b) the face selected from the library. Replacement results (c) without and (d) with recoloring and relighting. Notice the significantly improved realism in the final result.*

Figure 6: *Face replacement results. Each row contains (from left to right) the original photograph, a candidate face selected from the library, and the replacement result produced automatically using our algorithm. The age and gender mismatches in (c) and (d) could be avoided by enforcing consistency across those attributes (which our system does not currently do).*

5 Appearance Adjustment and Ranking

5.1 Color and Lighting Adjustment

While our selection algorithm described thus far is essential for finding candidate images to replace an input face with, it is not sufficient for creating realistic results – we must adjust the lighting and color properties of the candidates to match those of the input image. This is true even with very large face libraries because the chance of finding another image with *exactly* the same lighting (with matching pose and other attributes) is extremely small.

The first step in the adjustment is to use the quotient image formulation [Liu et al. 2001; Wen et al. 2003] to apply the lighting of the input image $I^{(1)}$ to the replacement candidate image $I^{(2)}$, within the replacement region. Using Equation 2, we can write the approximate image intensities for each of these images as

$$\tilde{I}_c^{(1,2)}(x,y) = \rho_c^{(1,2)} \sum_{k=1}^{9} \beta_{c,k}^{(1,2)} \psi_k(x,y), \;\; c \in \{R,G,B\}. \quad (4)$$

To obtain our relit replacement $\hat{I}^{(2)}$, we simply multiply the replacement candidate image by the ratio of the approximate images:

$$\hat{I}_c^{(2)} = I_c^{(2)} \left(\frac{\tilde{I}_c^{(1)}}{\tilde{I}_c^{(2)}} \right), \;\; c \in \{R,G,B\}. \quad (5)$$

Note that since $\tilde{I}_c^{(1)}$ and $\tilde{I}_c^{(2)}$ each capture only low-frequency lighting and color information, their ratio varies smoothly. Thus, the high-frequency content (e.g., highlights) of the replacement image is preserved during this relighting process. Finally, to match appearances even more closely, we transform the relit candidate image so that its RGB histogram matches that of the original input image within the replacement region.

Figure 5 shows the importance of our adjustment algorithm. Using the input image in Figure 5a and the replacement candidate in Figure 5b, we would obtain the result shown in Figure 5c if we simply blended in the candidate face without performing appearance transformation. Note that even though the faces in this example have somewhat similar skin colors, the replacement result looks noticeably incorrect without adjustment. In contrast, the final replacement with adjustment, shown in Figure 5d, looks highly realistic.

5.2 Boundary Ranking

To pick the best replacement results from our list of candidates, we rank the candidate replacements using a metric measuring their perceptual similarity to the input image along the boundary of the replacement region. The width of this strip is equal to 11% of the distance between the eyes (located inward from the region boundary), and this same strip is used for the final feathering operation. The ranking is computed using L_2 distance in CIE LAB space between the candidate replacement strip and the input strip. The last column of Figure 4c shows the highest ranked results for the input face shown in Figure 4a. The original candidate faces are shown in Figure 4b. One can see that these results look better than the ones shown in the previous column (without ranking).

6 Results

Figure 6 shows several examples of the results obtained using our system. Each example shows, in order from left to right, the input face image, a candidate face, and the replacement result. Note the realism of the results, despite differences in pose, lighting and facial appearance. Figures 6c and d show examples of face replacement across different ages and genders. While our system currently does not enforce consistency for these attributes, one could do so by incorporating classifiers such as [Moghaddam and Yang 2002; Lanitis et al. 2004].

(a)

(b)

Figure 7: *Face De-Identification and Switching. (a) Result of automatically replacing the input faces (top) with the top-ranked candidate from the face library to obtain the de-identified results (bottom). No user intervention was used to produce this result. (b) Result of switching the two input faces (left) with each other to obtain the de-identified output (right).*

6.1 Applications

Face De-Identification: To preserve privacy in online collections of photos, one can use our system to automatically replace each face in an input image with the top-ranked candidate taken from a collection of stock photographs. Figure 7a shows an example of such a replacement. We stress the fact that no user interaction was required to produce this result.

Switching Faces: As a special case of face de-identification (or for use as a special effect), we can limit the system to use candidates only within the same image, resulting in the switching of faces. Figure 7b shows the result of switching Elvis Presley and Richard Nixon's faces. (Here, we first reversed each face before replacement, so that the poses matched better.)

Composite Group Photographs: When taking group photographs, it is often difficult to get a "perfect" picture – where, for example, everyone is smiling, with eyes open, and looking at the camera. By taking several photos using the "burst" mode of a camera, we can construct a composite image in which everyone has the desired appearance. Since the relative position of faces does not change significantly during a burst mode shot, we limit the candidates to those with similar position in all images (thus avoiding the

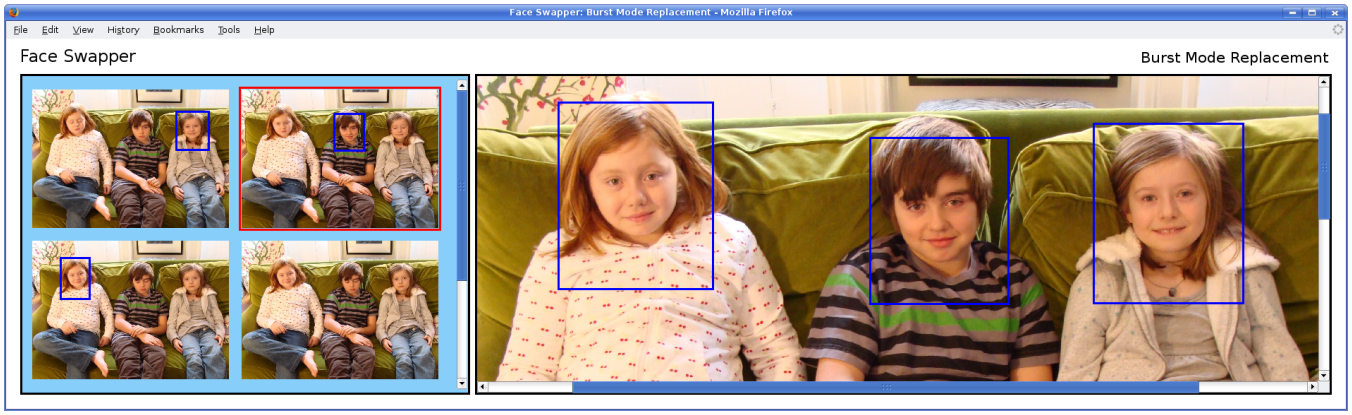

Figure 8: *Burst Mode Replacement. From a set of images taken using the "burst" mode of a camera (left panel), a composite image is created in which everyone is smiling and has their eyes open (right panel). The candidate faces for each child are constrained by the relative positions of the faces in all images, and thus no face recognition is required. While in this case the best replacement face for each child was selected manually (outlined in blue), blink and smile detection could be applied to select them automatically.*

need for face recognition algorithms). Furthermore, if we have access to smile and eye-blink detection (either by using classifiers or as given by the face detector), we can automatically select the best face to use for each person, resulting in single-click creation of the final image. Figure 8 shows our implementation of this application for creating a pleasing composite image of three children from a set of six input images. We see that the result, unlike each of the input images, has all of the children smiling and with eyes open.

6.2 User Study

The evaluation of face replacement results has so far been qualitative. To obtain quantitative results, we performed a formal user study, testing people's ability to distinguish between real images of faces and those generated by our system. For this evaluation, we showed users 50 images containing faces and asked them to classify them as "real," "fake," or "not sure," within a time limit of 5 seconds (roughly the time one would normally spend looking at a face). Exactly half of the images were real.

Across a total of 12 people tested, we found that 58% of our replaced face images were misidentified as real. In comparison, only 75% of real images were correctly marked real (full results are presented in Figure 9). These percentages were computed as $1 - \frac{\#\ marked\ fake}{total\ number\ of\ images/2}$. Note that a vote of "not sure" or no vote within the time limit was counted as real because it suggests that a user could not definitively mark an image as fake. (Forcing users to make a decision regarding the authenticity of the image raises their sensitivity to minor appearance effects they would not normally notice.) These numbers show the high quality of our results – users could not easily differentiate between real face images and those created by our system.

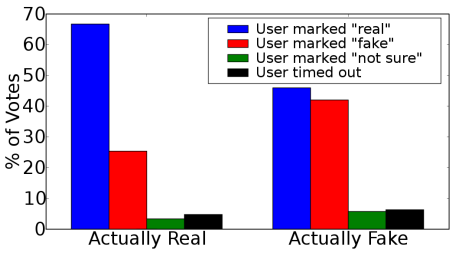

Figure 9: *Results of the User Study. Users were presented with 50 images, half real and half fake. Users were asked to mark each image as "real," "fake," or "not sure." The first set of bars shows how the users labeled images that were actually real. The second set shows how users labeled images created by our system. Notice that users marked fake images as real almost as frequently as they marked real ones as real (58% vs. 75%). See text for details.*

7 Discussion

We have created a comprehensive system for automatically replacing faces in photographs. Given an input image, we extract and align faces to a common coordinate system, search for similar faces from a library of candidates, match the appearance of the candidate faces using a photometric adjustment, blend the results into the input image, and finally rank the results. This entire process takes about 1 second using our C++ implementation. We note that our system is better suited for applications where the replacement face used is not specified by the user, but is allowed to be chosen (or at least narrowed) from a set of candidates. This is indeed the case with our target applications.

While we achieve very realistic results for a large variety of images, there are several limitations to our current system. These can be divided into two types: Those due to the face detector that our system depends upon, and those due to the replacement algorithm itself. Missed face detections, incorrect pose information, or misaligned fiducial point locations are all examples of problems caused by the face detector. Since the rest of the system depends on the accuracy of the detector outputs, we obtain worse results if there are any errors in them. As shown earlier in Figures 6c and d, our system does not currently enforce consistency for gender and age mismatches. However, this could also be fixed at the face detection stage; see again [Moghaddam and Yang 2002; Lanitis et al. 2004].

Figure 10 show several examples of limitations of our algorithm itself. In each case, we show the input and candidate faces on top, the replacement result in the middle, and a detailed inset highlighting the problem area on the bottom. Figure 10a shows that differences in face appearance, such as due to eyeglasses, can cause visual artifacts in replacement. Figure 10b shows our sensitivity to occlusions, where no faces could be found in the library with a similar occlusion. Figure 10c shows problems as we try to replace faces in extreme poses – the face starts getting blended into the background. Finally, Figure 10d shows a failure case for our relighting algorithm, where we forced a replacement between two faces with very different lighting. Our lighting selection step (bypassed here) would normally prevent such replacements.

In future work, we hope to remedy these issues in various ways. First, face detectors are expected to improve with time and our system will naturally benefit from this. Second, incorporating basic classifiers for gender, age, etc., will help in reducing mismatched replacements. On the algorithm side of things, our biggest limitation currently is the use of statically defined masks for each pose bin. Using dynamically-chosen optimal masks (e.g., as in [Avidan

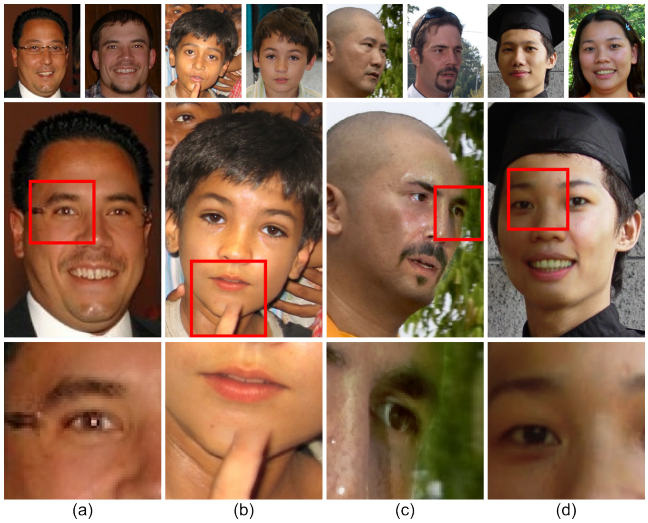

Figure 10: *Limitations. Input and replacement candidate faces are shown on the top, replacement results in the middle, and a detailed inset of the problem area on the bottom. The lack of eyeglasses in (a) and the occluding finger in (b) cause visual artifacts in the results. In (c), the extreme pose of the face results in it being blended into the background. These problems could be solved by dynamically selecting optimal replacement regions. (d) shows a relighting failure case, caused by forcing a replacement between images with very different lighting (skipping our lighting selection step).*

and Shamir 2007; Efros and Freeman 2001]) would help tremendously. These masks could also be defined hierarchically, to perform replacement on only parts of the face, thus avoiding problems with occlusions and extreme pose.

Photo Credits

Images in Figure 1a used with permission from Retna Ltd. (top) and Getty Images Inc. (bottom). Figure 7b is in the public domain. Figure 8 uses our own images. All other photographs taken from Creative Commons licensed (http://creativecommons.org/licenses/) images on flickr.com, shared by the following users: thedoorman, matt-meyer, sixsteps, astrolondon, torres21, bert52, rollerstan, captainsmurf, 676, bryanesque, katcha, sarahbaker, soldiersmediacenter, nesster, dtmagee, thepinkpantherz, derusha, 101206, oneiroi, josephbjames, srrf, nelsva, ehavir, shadowalker, silvers, kjirstinb, bike, g_w_y_n, adrianhon, sapnachiara, tipuri, piven, tracylee, killermonkeys, ciordia, chungwei, leeicep.

Acknowledgements

We are grateful to Omron Technologies for providing us the OKAO face detection system. We thank Tarandeep Singh for implementing the interface to the OKAO face detector. Neeraj Kumar was supported by the Department of Defense through the National Defense Science & Engineering Graduate Fellowship (NDSEG) program.

References

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

AVIDAN, S., AND SHAMIR, A. 2007. Seam carving for content-aware image resizing. *ACM Transactions on Graphics 26*.

BASRI, R., AND JACOBS, D. 2003. Lambertian reflectance and linear subspaces. *IEEE TPAMI 25*, 218–233.

BLANZ, V., SCHERBAUM, K., VETTER, T., AND SEIDEL, H.-P. 2004. Exchanging Faces in Images. *Computer Graphics Forum 23*, 669–676.

BOYLE, M., EDWARDS, C., AND GREENBERG, S. 2000. The Effects of Filtered Video on Awareness and Privacy. In *ACM Conference on Computer Supported Cooperative Work*.

COOTES, T., EDWARDS, G., AND TAYLOR, C. 2001. Active Appearance Models. *IEEE TPAMI 26*, 681–685.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the Reflectance Field of a Human Face. In *SIGGRAPH 00*, 145–156.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH 98*, 189–198.

EFROS, A. A., AND FREEMAN, W. T. 2001. Image quilting for texture synthesis and transfer. In *SIGGRAPH 01*, 341–346.

FERGUS, R., SINGH, B., HERTZMANN, A., ROWEIS, S., AND FREEMAN, W. 2006. Removing camera shake from a single photograph. *SIGGRAPH 06*, 787–794.

GROSS, R., SWEENEY, L., DE LA TORRE, F., AND BAKER, S. 2006. Model-Based Face De-Identification. 161–168.

HAYS, J., AND EFROS, A. A. 2007. Scene completion using millions of photographs. *ACM Transactions on Graphics 26*, 3.

KUNDUR, D., AND HATZINAKOS, D. 1996. Blind image deconvolution. *IEEE Signal Processing Magazine*, 3, 43–64.

LANITIS, I., DRAGANOVA, C., AND CHRISTODOULOU, C. 2004. Comparing different classifiers for automatic age estimation. *IEEE Trans. on Systems, Man, and Cybernetics, B 34*, 621–628.

LIU, Z., SHAN, Y., AND ZHANG, Z. 2001. Expressive expression mapping with ratio images. In *SIGGRAPH 01: Proc. of the 28th CGIT*, 271–276.

MALIK, S., 2003. Digital face replacement in photographs. http://www.cs.toronto.edu/∼smalik/2530/project/results.html.

MOGHADDAM, B., AND YANG, M.-H. 2002. Learning gender with support faces. *IEEE TPAMI 24*, 707–711.

NEWTON, E., SWEENEY, L., AND MALIN, B. 2005. Preserving Privacy by De-Identifying Face Images. *IEEE Trans. on Knowledge and Data Eng.*, 232–243.

OMRON, 2007. OKAO vision. http://omron.com/rd/vision/01.html.

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. An Efficient Representation for Irradiance Environment Maps. In *SIGGRAPH 01*, 497–500.

RUBNER, Y., TOMASI, C., AND GUIBAS, L. J. 2000. The earth mover's distance as a metric for image retrieval. *IJCV*, 99–121.

WANG, H., RASKAR, R., XU, N., AND AHUJA, N. 2007. Videoshop: A New Framework for Video Editing in Gradient Domain. *Graphical Models 69*, 57–70.

WANG, Y., LIU, Z., HUA, G., WEN, Z., ZHANG, Z., AND SAMARAS, D. 2007. Face re-lighting from a single image under harsh lighting conditions,. *CVPR '07*.

WEN, Z., LIU, Z., AND HUANG., T. S. 2003. Face Relighting with Radiance Environment Maps. In *CVPR '03*, 158–165.