Digital Face Makeup by Example

Dong Guo and Terence Sim
School of Computing
National University of Singapore
Singapore, 117417
{guodong,tsim}@comp.nus.edu.sg

Abstract

This paper introduces an approach of creating face makeup upon a face image with another image as the style example. Our approach is analogous to physical makeup, as we modify the color and skin detail while preserving the face structure. More precisely, we first decompose the two images into three layers: face structure layer, skin detail layer, and color layer. Thereafter, we transfer information from each layer of one image to corresponding layer of the other image. One major advantage of the proposed method lies in that only one example image is required. This renders face makeup by example very convenient and practical. Equally, this enables some additional interesting applications, such as applying makeup by a portraiture. The experiment results demonstrate the effectiveness of the proposed approach in faithfully transferring makeup.

1. Introduction

Face makeup is a technique to change one’s appearance with special cosmetics such as foundation, powder, cream etc. In most cases, especially for females, makeup is used to enhance one’s appearance. With physical face makeup, the foundation and loose powder are usually used to change the texture of face’s skin. Foundation is mainly used to conceal flaws and cover the original skin texture, while the loose powder is for introducing new, usually pleasant, texture to skin. Afterwards, applications of other color makeup, such as rouge, eye liner and shadow, etc., follow on top of the powder.

Consider this scenario: when a customer enters a beauty salon, she selects an example image from a catalog and tells the makeup artist to apply the same makeup on her. Before actual task, it would be extremely helpful if she can preview the makeup effects on her own face. However, this is difficult. Traditionally, people have two choices for trying out makeup. One is to physically apply the makeup, which is time-consuming and requires the patience of the participants. Alternatively, one may try on makeup digitally by way of digital photography and with the help of photo editing software, such as Adobe Photoshop™. But using such photo editing software is tedious and relies heavily on the users’ expertise and effort.

In this paper, we present an approach of creating makeup upon a face image (Fig. 1(a)) with the prototype of another image (Fig. 1(b)) as the style example. This is very practical in the scenario of the beauty salon.

Our approach is inspired by the process of physical makeup. First, we decompose the subject and example images into three layers separately: face structure layer, skin detail layer, and color layer. Ideally, the face structure layer contains only the structure of every face component, such
as the eyes, nose, mouth, etc. The skin detail layer contains the skin texture, including flaws, moles, as well as any wrinkles. The color layer represents color alone. After the three layers are decomposed, face makeup by example can be considered as transferring the skin detail layer and color layer from the makeup example to the subject image while preserving the face structure layer of the subject image.

**Contribution.** Our proposed approach is effective in transferring face makeup from an example image. This significantly reduces the effort for previewing makeup effects using traditional methods. Moreover, only one single example image is required. This renders face makeup by example much more convenient and practical, compared to previous work which usually requires a pair of “before”-“after” makeup images as examples.

**2. Related Work**

There is not much previous work addressing digital face makeup. The most closely related work is that of Tong et al. [15]. In their work, the way makeup changes the appearance is learned from a pair of example images of the same face “before” and “after” makeup. The quotient of the “after” divided by the “before” is used to represent the change. Then the quotient is multiplied by another image in order to achieve the makeup result. In contrast, our approach requires only one “after” example. This is more convenient and practical, as providing the “before” image is rather difficult in most cases. In addition, it is quite common in the makeup to change the texture of face skin (conceal the original and introduce new one). Because the original texture varies from person to person, the change from “before” to “after” is different between different faces. Thus, it is inappropriate to apply the change across two faces. In contrast, our approach directly transfer the skin texture of the example to the subject image, concealing the original texture. Note that our approach can also keep the texture of the subject image if needed.

Another approach by Ojima et al. [10] also uses a pair of “before”-“after” makeup examples. However, only the foundation effect is addressed in their work. In contrast, Tsumura et al. [16] employed a physical model to extract hemoglobin and melanin components. Changes in facial appearance are simulated by adjusting the amount of hemoglobin and melanin. The effects they demonstrated include tanning, reddening due to alcohol consumption, aging and cosmetic effects. However, the cosmetic effects are quite limited, and are much simpler than real makeup. Besides, an online commercial software, Taaz [13], provides users with virtual makeup on face photos by simulating the effects of specified cosmetics.

Besides makeup, some existing works also focus on the beautification of face photos. For example, Brand and Pletscher [2] proposed an automatic face photo retouching method aiming to detect and remove flaws, moles, and acne from faces. Another interesting work by Leyvand et al. [6] introduced a technique of modifying face structure to enhance the attractiveness. However, this may also change the identity of the subject, as face structure is usually considered to be a key representation of identity. Conversely, we achieve the goal of beautification by modifying only skin detail and the color, while faithfully preserving the face structure.

The idea of image processing by example can be found in image analogies [5]. Image analogies provide a general framework of rendering an image in different styles. This method learns how pixel values change from a pair of “before”-“after” images as example. This idea was used in Tong et al.’s work [15]. As mentioned earlier, the difference is that our approach learns the effects after alteration, while their approach learns the way of altering image pixels.

Our method can also be considered as texture and color transfer. Some previous work on texture transfer is also related. Shan et al. [12] proposed a method to transfer the texture across two images. Their method of transferring fine texture is similar to ours in spirit. They used the quotient of the original image and a Gaussian smoothed version of the image to represent the texture. The quotient is multiplied to the smoothed version of another image. This is similar to our layer decomposition. However, there are many differences. First, the Gaussian blur they used may produce halo effect at the strong edges. In contrast, we use an edge-preserving smooth [4] to separate the layers, which successfully suppresses the halo effects. Moreover, they only focus on texture transfer; but it is also important to keep color consistent. We separately transfer the skin detail in lightness channel and color in color channel. In addition, they focused on transferring texture, while our goal is to transfer makeup effect, which is more complex.

**3. Digital Makeup**

The input of our approach are, a subject image \(I\), a face image to be applied with makeup, and an example image \(E\), providing makeup example. The output is result image \(R\), in which the face structure of \(I\) is retained while the makeup style from \(E\) is applied.

The workflow of our approach is illustrated in Fig. 2. There are four main steps. First, face alignment should be done between the subject and example images. Because the information is transferred pixel by pixel, a fully alignment is necessary before transferring. Followed is layer decomposition (Section 3.2). Both \(I\) and \(E\) are decomposed into three layers: face structure layer, skin detail layer, and the color layer. Third, information from each layer of \(E\) is transferred to corresponding layer of \(I\) in different fashions: skin detail is transferred in an additive way (Section 3.3); color is transferred by alpha blending (Section 3.4); highlight and
3.1. Face alignment

For face alignment, we adopt the Thin Plate Spline (TPS) [1] to warp the example image $\mathcal{E}$ to subject image $\mathcal{I}$. The control points required by TPS are obtained using an extended Active Shape Model (ASM) [8]. Due to the diversity of face appearance under various possible makeup, ASM may not get the accurate position of points. Our system may still require user to refine the position of control points. Since the control points have been roughly located already, the refinement does not require much effort. It usually takes less than 1 minute to refine the control points for a face. An example of refined control points is shown in Fig. 3(a). There are 83 points in total on a face.

As shown in Fig. 3(b), these control points have defined different face components, viz eyebrows, eyes, nose, nostrils, lips, mouth cavity (the space between lips), and other facial skin (the rest of the face). These components are further divided into three classes to be treated in different ways during makeup. These three classes ($C_1 \sim C_3$) are illustrated in different colors in Fig. 3(b). $C_1$ (the skin region, the entire face excluding $C_2$ and $C_3$) follows the workflow illustrated in Fig. 2. Since the texture of $C_2$ (lips) varies greatly from person to person and the region of $C_2$ is easy to deform, we use a special method to transfer the makeup style in this region (discussed in Section 3.6). $C_3$ (eyes and mouth cavity) is kept untouched all the time during entire makeup process.

3.2. Layer decomposition

The subject image $\mathcal{I}$ and example image $\mathcal{E}$ (after warping) are first decomposed into color and lightness layers. We then further decompose the lightness layer into face structure and skin detail layers.

In the first step, $\mathcal{I}$ and $\mathcal{E}$ are decomposed into color and lightness layers by converting them to CIELAB colorspace. The $L^*$ channel is considered as lightness layer and $a^*$, $b^*$ channel the color layer. We choose the CIELAB colorspace because it performs better than other color spaces in separating lightness from color [17], and it is approximately perceptually uniform [7].

Second, the lightness layer is decomposed to large-scale and detail layers. The large-scale layer is considered as the face structure layer and the detail layer as skin detail layer. Large-scale/detail layer decomposition has been addressed in many works, such as [3] and [19]. The main idea is first to perform an edge-preserving smoothing on the lightness layer to obtain the large scale layer, and then to subtract (or divide) the large scale layer from the lightness layer to obtain the detail layer. In this approach, we adapt a weighted-least-squares (WLS) operator recently proposed by Farbman et al. [4]. An alternative method is bilateral filtering [14], which was used in many previous works. We choose this WLS operator because of its better performance compared to the bilateral filter, especially when the blur level
Suppose that the lightness layer and the face structure texture layer are denoted by \( l \) and \( s \), respectively. The problem of solving \( s \) can be formulated as minimization of the energy function:

\[
E = |s - l|^2 + \lambda H(\nabla s, \nabla l).
\]

The first term \(|s - l|^2\) is to keep \( s \) similar to \( l \), while the regularization term \( H(\nabla s, \nabla l) \) is trying to make \( s \) as smooth as possible.

The WLS operator described in [4] performs the same level of smoothing all over the image; but we expect different levels of smoothness in different regions. Thus, a spatial-variant coefficient \( \beta \) is added to \( H \). Then \( H \) is defined as

\[
H(\nabla s, \nabla l) = \sum_p \beta(p) \left( \frac{|s_x(p)|^2}{|l_x(p)|^\alpha + \epsilon} + \frac{|s_y(p)|^2}{|l_y(p)|^\alpha + \epsilon} \right),
\]

where \( p \) indexes the image pixel, \( \epsilon \) is a small constant preventing division by zero, \( \{ \cdot \}_x \) and \( \{ \cdot \}_y \) denote the partial derivative of \( \{ \cdot \} \) along \( x \) and \( y \) coordinates respectively, while \( \alpha \) is a coefficient for adjusting the effect of \( \nabla l \) on \( \nabla s \). We use \( \alpha = 1.2 \) and \( \lambda = 0.2 \) in all our experiments.

We expect that \( \beta(p) \) is low inside the facial components, and equal to \( 1 \) over the skin area. As shown in Fig. 3(c), \( \beta(p) \) is 0.3 in the eyebrow region, 0 in other face component region, and 1 in facial skin. In addition, we also expect \( \beta(p) \) changes smoothly over the whole image. Thus, we further define \( \beta(p) \) as

\[
\beta(p) = \min \left( 1 - k(q), e^{-\frac{(q-p)^2}{2\sigma^2}} \right),
\]

where \( q \) indexes the pixel over the image, \( k(q) \) is 0.7 for eyebrows, 0 for skin area, 1 for other facial components. The value of \( \sigma^2 \) is set to \( \min(\text{height}, \text{width})/25 \). Then, we have \( \beta(p) \) as shown in Fig. 3(d).

Since the \( L^* \) channel is approximately perceptually uniform, we use subtraction to obtain the skin detail layer \( d \) from lightness layer \( l \), i.e.

\[
d(p) = l(p) - s(p).
\]

Two examples of our decomposition results are shown in Fig. 2. We can see that detail is controlled well by \( \beta(p) \). Because \( \beta \) is zero in the eyes and mouth regions and outside the facial region, the skin detail layer is zero.

In the rest of the paper, we will use \( \{ \cdot \}_x, \{ \cdot \}_d, \{ \cdot \}_e \) to denote \( \{ \cdot \} \)’s the face structure layer, skin detail layer, and color layer, respectively.

### 3.3. Skin detail transfer

Skin detail transfer is straightforward. The resultant skin detail layer \( R_d \) is a weighted sum of \( I_d \) and \( E_d \), i.e.

\[
R_d = \delta_I I_d + \delta_E E_d,
\]

where \( 0 \leq \delta_I, \delta_E \leq 1 \). The values of \( \delta_I \) and \( \delta_E \) control the contribution of each component.

For different applications, different \( \delta_I \) and \( \delta_E \) values can be used. As we mentioned in Section 1, the purpose of foundation and loose powder in physical makeup is to conceal the original skin detail and to introduce new skin detail. Thus, we set \( \delta_I = 0 \) to conceal \( I_d \), and \( \delta_E = 1 \) to transfer \( E_d \) to \( R_d \). This is a typical setting for beauty makeup transfer. It is used in all our experiment results except the ones showing different manipulation of makeup effects (Fig. 4).

In some cases, we can also set \( \delta_I > 0 \) to keep some original skin detail. Note that the sum of two weights is not required to be 1 because \( R_d \) can come from any amount of \( I_d \) or \( E_d \). In addition, the sum should not be very small, otherwise the face in result image \( R \) would be not realistic due to lack of skin detail.

### 3.4. Color transfer

The resultant color layer \( R_c \) is an alpha-blending of color layers of \( I \) and \( E \), i.e.

\[
R_c(p) = \begin{cases} 
(1 - \gamma) I_c(p) + \gamma E_c(p) & p \in C_3 \\
I_c(p) & \text{otherwise}
\end{cases}
\]

The value of \( \gamma \) is to control blending effect of two color layers. The result in Fig. 1 is obtained with \( \gamma = 0.8 \).

### 3.5. Highlight and shading transfer

The highlight and shading effects of makeup lie in the \( L^* \) channel. Because the face structure layer is actually the large scale layer of \( L^* \) channel, the smooth change of the highlight and shading effects remain in the face structure layer. Since these effects are important for makeup, we should transfer them across this layer.

Because face structure layer contains identity information, we can neither directly copy \( E_s \) over \( I_s \) nor blend
them. Instead, we adapt a gradient-based editing method. The idea is to add only large changes of $E_s$ to $I_s$. Doing this, we assume that these changes are due to makeup. This assumption holds if the illumination of $E$ is approximately uniform.

Gradient-based editing can preserve the illumination of $I$, transfer the highlight and shading effects, and meanwhile yield smooth result. Editing an image in the gradient domain was introduced by Pérez et al. [11] and was employed in many later works. The gradient-based method used here is similar. The gradient of $R_s$ is defined as

$$
\nabla R_s(p) = \begin{cases} 
\nabla E_s(p) & \text{if } \beta(p)||\nabla E_s(p)|| > ||\nabla I_s(p)|| \\
\nabla I_s(p) & \text{otherwise}
\end{cases}
$$

Since only the gradient of the face region ($C_3$) is changed (but not its boundary nor regions outside $C_3$), the process of solving the resultant face structure layer $R_s$ from its gradient is equivalent to solving a Poisson equation with Dirichlet boundary condition. We use the Gauss-Seidel method with successive over-relaxation to solve the Poisson equation.

### 3.6. Lip makeup

The makeup effect of lip region ($C_2$) is quite different from that of face skin ($C_1$). In physical makeup, cosmetics on lips (e.g., lipstick) usually preserve or highlight the texture of lips, rather than conceal it as in face skin. Thus, lips region makeup needs to be treated in a different way.

The main idea is to fill each pixel of $R$ with pixel value from $E$ guided by $I$. Then the makeup effect is similar to $E$ and the texture is similar to $I$. Specifically, for each pixel $p$ in $I$, we search for a pixel $q$ in $E$ so that $E(q)$ and $I(p)$ are as similar as possible, while $q$ and $p$ are as close as possible.

Suppose that lip region after makeup is denoted by $M$. For each $p \in C_2$, we have

$$
M(p) = E(\hat{q}),
$$

where

$$
\hat{q} = \arg \max_{q \in C_2} \{ G(||q - p||)G(||E(q) - I(p)||) \}
$$

where $G(.)$ denotes Gaussian function. For $|E(q) - I(p)|$, we use the difference of pixel values in only $L^*$ channel after histogram equalization of $E$ and $I$ separately.

The makeup result of lips region $M$ is then merged to result image $R$. The $L^*$ channel of $M$ is added to $L^*$ channel of $R$ with a gradient replacement. The $a^*$, $b^*$ channel of $M$ replace corresponding region in $R$.

### 4. Experiments and Results

**Beauty makeup:** Our approach can manipulate makeup effect in various ways. In case of heavy foundation where the foundation covers the original skin, we use $\delta_I = 0$, $\delta_E = 1$ to transfer all skin detail from $E$ to $I$. An example of result is shown in Fig. 1(c). To simulate light foundation effect, we use $\delta_I = 1$, $\delta_E = 0$ to keep original skin detail. An example is Fig. 4(a). Adjusting $\delta_I$, $\delta_E$ can simulate different level of foundation effect. For a good result of concealing the skin detail of $I$, it is required that the skin detail of $E$ should be only due to makeup, i.e. $E$ should have heavy foundation (usually also with loose powder); otherwise $R$ may contain imperfection of skin detail from $E$. Another manipulation is on the level of makeup color by adjusting $\gamma$. As shown in Fig. 4(b), we used $\gamma = 0.5$ for a light makeup color.

A comparison of our result with that of Tong et al. is shown in Fig. 5. Note that they used an additional image as example of “before” makeup, which is not shown here. In contrast, we only use the “after” image. Our result shows vivid color and more similar makeup style to the example image. From the close-up view of the eye, we can see that our approach preserves the structure much better than theirs. Moreover, the highlight of lips in our result is successfully transferred while the texture is preserved. It is much more similar to real makeup. For skin tone, we have different representations. Our result takes the skin tone from example image, while their result keeps the original tone. As we presented before, our aim was to apply the makeup style faithfully from the example image, including color, skin detail. In the physical makeup, tone color is usually also considered to be a part of makeup. Another consideration is that color used in makeup should match the skin tone [18]: makeup transferring without skin tone may destroy the original harmony.

**Photo retouching:** Our approach can be also used to retouch a face photo. This is almost the same as face makeup by example. We address it separately because this application itself is interesting and useful. Instead of photographs
from others, users may also provide their own previously well-taken photograph, with or without makeup, as the example for retouching. We show an example in Fig. 6. A photo (Fig. 6(a)) with makeup taken in a professional studio is provided as the example. Two other photos of the same person (Fig. 6(b) and (d)) taken by an amateur user are provided as the subject images. In the results (Fig. 6(c) and (e)), the imperfection of the face skin was concealed and pleasant face skin was introduced successfully.

**Makeup by portraiture:** An interesting application is transferring makeup from a portraiture to a real face photo. This is similar to makeup transfer between real face photos. However, portraitures usually have non-realistic artifacts in the skin detail layer due to the drawing material, or aging and discoloration. These artifacts are not suitable for a face photo. For example, in Fig. 7, the portraiture has some noise. Transferring it makes result unnatural (Fig. 7(c)). If we restrain it by setting $\delta_E = 0$ and $\delta_I = 1$, the result becomes much better (see Fig. 7(d)).

5. **Conclusion and Future Work**

In this paper, we have presented an approach of creating makeup upon an image with another example image as the makeup style. The main idea is simple yet powerful. Promising results have demonstrated the effectiveness of our approach. One major advantage of our approach is that only one example image is required.

Limitations: The active shape model we adopt assumes that the face is frontal and upright. Our system is currently tested only with both the subject and example images being nearly frontal. But we envision that it can work well on any pose as long as the pose difference is not large between the subject and example images. One future work is to extend the approach to any pose.

Currently, only one image is used as example. It would be interesting and practical to extend our current approach to multiple example images. In this case, the consistency between different makeup styles may be an issue.

Since only skin detail and color are required from the example image, if we warped these information to a canonical face, the original face structure of the example image is not required any more. Doing so indeed helps to preserve the privacy of the makeup actors. Another future work is to collect more makeup examples and build up a makeup engine enabling users to browse different makeup styles with their own faces, as the previously introduced beauty salon case.

6. **Acknowledgments**

We thank the reviewers for their valuable comments, Michael S. Brown, Xiaopeng Zhang, Ning Ye, Shaojie Zhuo, and Hao Li for their insightful discussions, as well as Wenzhe Su and Jing Sun for providing their photographs. We acknowledge the generous support of NUS.
Figure 6. Photo retouching. (a) The example image, taken in a professional studio, provides the desired retouching result. (b)(d) Photos of the same person in (a), taken by an amateur. (c)(e) The retouching results of (b) and (d), respectively.

Figure 7. Makeup by portraiture. (a) An old portraiture scan, as the example image. (b) A photo, as the subject image. (c) The makeup result with skin detail form (a). (d) The makeup result with skin detail from (b).

References


