Digital Makeup Face Generation

Wut Yee Oo

Mechanical Engineering

Stanford University

`wutyee@stanford.edu`

Abstract—Make up applications offer photoshop tools to get users inputs in generating a make up face. While this has been a well established area of development, automatic transfer of make-up from a reference to a target image presents more challenges. This project applies the make up from a reference image to a target on a pixel by pixel basis. The reference and target were aligned by warping using 77 detected face landmarks. The images were decomposed into large scale layer, detail layer and color layer through which makeup highlights and color information were transferred by Poisson editing, weighted means and alpha blending. The test results showed that the techniques work well with reference images of a different age, a different skin color and even a hand-sketched reference image.

I. INTRODUCTION

There are many photoshop applications which rely on the user's skills to use makeup tools and apply them on their digital faces manually. Although this allows customization, this is a more involved process which requires time and skills from the user. In some scenario, the user with limited knowledge of makeup will just want to adopt the makeup look from a reference makeup model to her face. This project attempts to apply make up from a reference image to a target image without make up as shown in Figure 1.

There has been work done on digital face makeup transfer from a reference image by Guo, etl [4]. Some of their work are further developed by Chen, etl for face illumination transfer from a reference to a target [3]. This project consolidates these existing work and leverages the existing research on face detection. The algorithm for digital makeup face generation, as seen in Figure 2, includes four major components: (1) face features detection and extraction, (2) face alignment, (3) layer decomposition (4) transferring of makeup through each layer.

II. METHODOLOGY

Facial detection by Active Shape Model was performed in both reference and target images to recognize eyes, nose, lips and face contours. Using the positions of detected facial features, the two images are aligned by warping so that properties from the reference to the target can be transferred on a pixel by pixel basis. Eyes and lips were cut away from the main face of the target and the warped reference to be edited separately. The subsequent shapes were decomposed into lightness and color layers in CIE L*a*b space and the lightness layer was further broken down into a large scale layer and a detail layer by a Weighted Least Square filter. The large scale layer ideally contains the smoothed face image with only highlights and contours whereas the detail layer contains moles, skin texture, wrinkles and etc. The makeup highlights,

(a) Reference Image with Makeup   (b) Target Image without Makeup

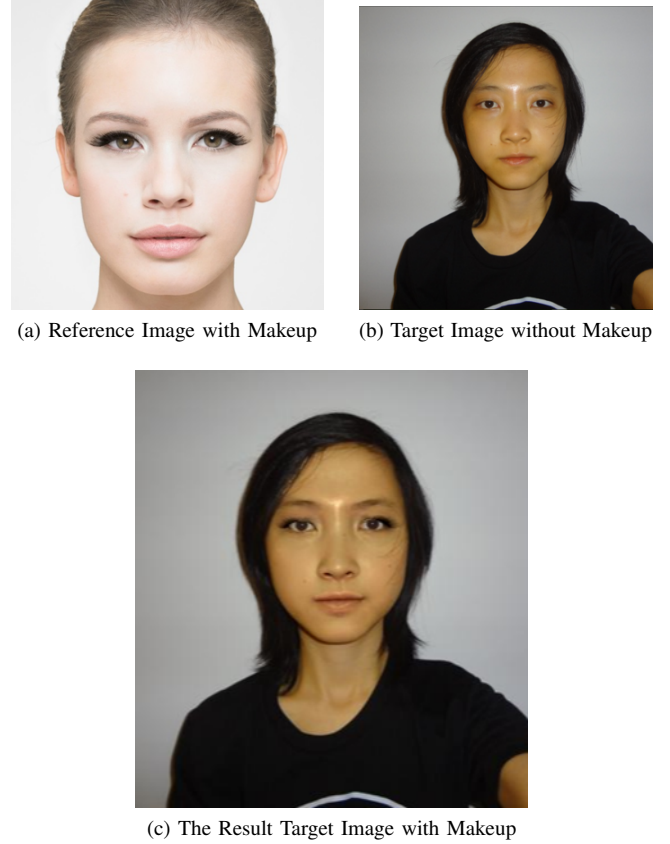(c) The Result Target Image with Makeup

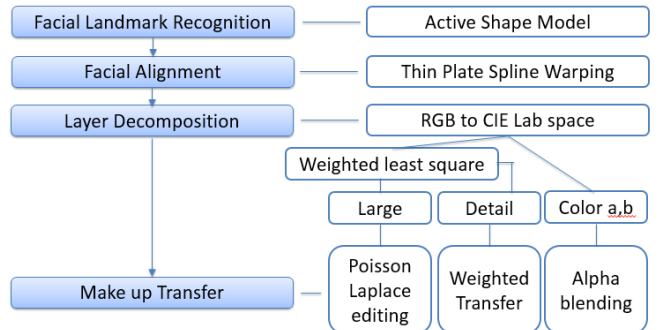Fig. 1: Makeup Transfer from Reference to Target

Fig. 2: Algorithm for Digital Makeup Face Generation

the skin details, and the colors are transferred through the large scale layer by gradient editing, detail layer by weighted means and color layer by alpha blending respectively. The layers were all merged together in the end and stitched with the original target background by Circular Averaging Filter to give the final result with makeup.

III. Face Detection

The landmarks, distinct points at the corner or with high curvature in an image, constitute a shape. Shapes are represented by vectors and the mean shape is calculated by aligning shapes so that they will have the least minimum Euclidean distance between them. Active shape model has trained data of the mean shapes of the face, eyes and lips. When active shape model is applied to the sample face profile, it will first create an estimated face shape based on its mean shape and global face detector. It will then try to deform the estimated shape within the model's limits by scaling, rotation and transformation so that there is the least minimum mean square distance between the two shapes.

The active shape model, implemented in Stasm library in C++ by Stephen Milborrow [1], was used to generate 77 landmarks on face, eyes and lips of the target and the reference as shown in Figure 3. The landmarks were obtained in the order as annotated by the numbers in Figure 3. The advantage of the Active Shape Model over other face detectors is that it can generate face contours effectively and this therefore eliminates the need to look at face segmentation.

IV. Face Alignment: Warping

The reference image needs to be aligned to the target image in 77 landmarks so that the two images will of the same size and shapes, and pixels can be directly transferred from one to another. Thin plate spline warping as described in [5] and implemented by Archibald in MATLAB was used. The bending energy of the plate,i.e the final warped image of the reference, were obtained from the position constraints set by the target landmarks.

## V. Pre-processing Methods before Layer Decomposition

A. Cutting and Masking

After the face detection, more data points were created by interpolation methods so as to have evenly spaced contour points between the sets of landmarks for face, lips and eyes as seen in 4. These data points were inputs to create a polygonal area of binary interest in MATLAB which were used for masking and cutting. Since the reference image was already aligned with the target, the same set of data points could be used to cut away the two areas of the interest in both target and reference images for make up transfer. One cut area of interest is the face without eyes and lips, and the other with just lips. It is important to separate out eyes and lips from the face layer so that the algorithm for make up transfer process, as discussed in Section VI and VII can hold.


(a) Reference Image
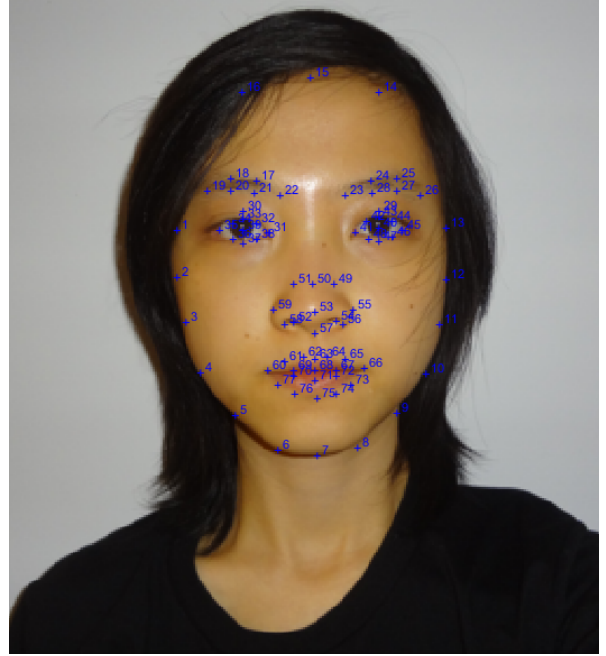
(b) Target Image

Fig. 3: Active Shape Model generates 77 landmarks on face contours, eyes, nose and lips.

B. Adaptive Thresholding

It was found that there were some components in the target's face area, particularly hair, remained after cutting and masking. This will give undesirable results in makeup transfer. Therefore, global adaptive thresholding was performed with Otsu's method to remove hair from the cut face area in the target image as seen in Figure 4(b). Narrowing down the threshold operation to just face cut area makes it easier to distinguish between foreground pixels (i.e. face) and the
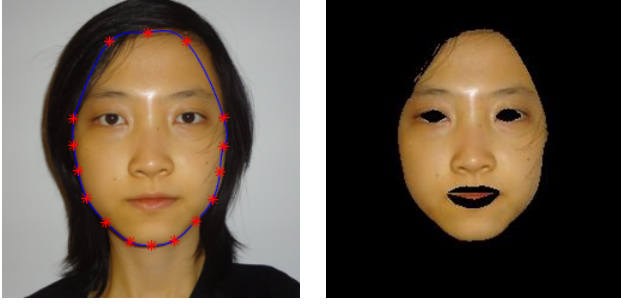
(a) Interpolated Contour on Target Face with * Representing Landmarks

(b) Cut Target Face after Hair Removal through Otsu's Method

(c) Cut Reference Face after Warping and Linear Color Scaling

(d) Cut Reference Face after Warping and Gamma Correction

Fig. 4: Pre-processing Methods: (a) Cutting and Masking, (b) Thresholding, Color Adjustment by (c) Linear Scaling, (d) Gamma Correction

background pixels (i.e. hair). The eyebrows which get removed as a result of thresholding were refilled by hole filling function in MATLAB.

C. Color Adjustment

The R,G,B color channels in the cut areas of warped reference image were individually scaled linearly by a factor so as to have the least minimum mean squared error with the corresponding color channel in the respective target areas (see Figure 4(c)). It was found that linear scaling of color channels is a better method as opposed to Gamma color correction of channels as shown in Figure 4(d) because the former gives a more calming color tone and emphasis.

VI. LAYER DECOMPOSITION

Figure 5 shows the work flow of layer decomposition, highlights and color transfer which were discussed further in Section VII-A and VII-B.

The project closely follows the techniques as introduced by Guo, etl for the process of makeup transfer. Make up transfer was achieved through highlight, detail and color layers [4]. As the first step, the RGB color space was transformed into CIE $L * a * b$ space where L refers to lightness and a and b color channels. CIE $L * a * b$ space allows for easy separation between highlight and color layers.

The highlight layer was further decomposed into large scale and detail layers by a Weighted Least Square Filter. Weighted Least Square filter is a type of edge-preserving filters and it was chosen over other filters such as bilateral filter because the former can maintain the shading information but may not preserve the structure of the image well [3]. This works fine for our case because the edges can be separated out in a detail layer which is not affected by makeup transfer. The large scale layer s was subtracted from the lightness layer L to give the detail layer d as shown in equation 1

$$s + d = L \qquad (1)$$

The standard weighted least square filter performs the same level of smoothing all over the image by minimizing Energy term E as shown in 2 [3] [4]. The first term $|s - l|^2$ keeps the large scale layer s similar to lightness layer L while regularizing H makes s as smooth as possible. The regularizing term was further represented in 3 [3] [4]. The nonlinear term λ was added for this particular WLS filter because in our case, the smoothing should be performed differently depending on the gradient changes in the selected areas [3] [4]. For instance, the eyebrows and the nose edges should be completely deposited into the detail layer so that the original face details of the target will not be manipulated by highlight transfer in the face structure layer. The non-linear WLS was implemented both in [3] and [4] but the author chose to adopt the method by [3] for easier implementation. Small λ is needed for flat regions and higher λ is for regions with more details to get smoother large scale layer. Increasing α preserves sharper edges. The subscript p denotes a spatial location of a pixel in 3.

$$E = |L - s|^2 + H(\nabla s, \nabla l) \qquad (2)$$

$$H(\nabla s, \nabla l) = \sum_p \lambda(p)\left(\frac{(\frac{\delta s}{\delta x})_p^2}{(\frac{\delta l}{\delta x})_p^\alpha + \epsilon} + \frac{(\frac{\delta s}{\delta y})_p^2}{(\frac{\delta l}{\delta y})_p^\alpha + \epsilon}\right) \qquad (3)$$

In equation 4, γ calculates the number of pixels with gradients greater than t_1 within a local window frame of w_p. The resultant number of pixels were then used to calculate λ as shown in equation 5. The parameters were chosen as belows for 3, 4 and 4.

$\lambda_s = 1$, $\lambda_l = 4$, $\alpha = 1.2$, $t_1 = 0.02$, $w_p = [5\ 5]$

$$\gamma(p) = \sum_{i \in w_p} \sqrt[2]{(\delta l/\delta x)_i^2 + (\delta l/\delta y)_i^2} \geq t_1 \qquad (4)$$

$$\lambda(p) = \lambda_s + (\lambda_l - \lambda_s) * \gamma(p) \qquad (5)$$

VII. MAKEUP TRANSFER

A. Highlight Transfer

Unlike other image blending process, the makeup from the reference cannot be directly copied and pasted onto the target. The target still needs to maintain its facial information while getting highlights and colors from the reference due to target. Therefore, only the gradients of each image need to be edited. Poisson gradient editing as developed by Perez et al.[2] was
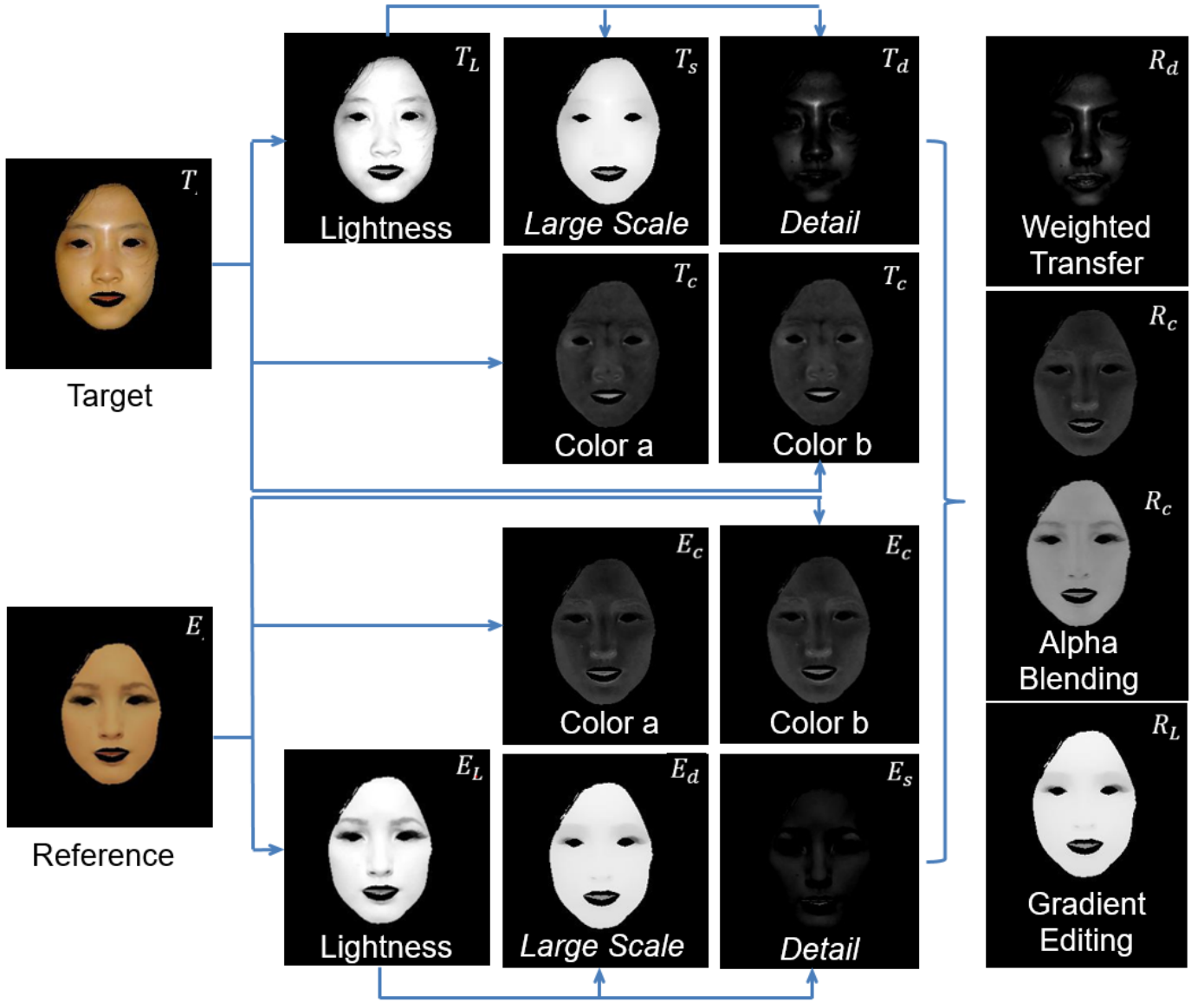
Fig. 5: Layer Decomposition and Makeup Transfer Workflow Model. Detail layers and color layers are multiplied by a factor of four to enhance the image

used for gradient editing as shown in 6. The changes in the target will use those in the reference if the reference's gradient is above our threshold range, assuming that large gradients in the reference are due only to makeup. There will be no gradient change in the target T_s if the reference's gradient $\nabla E_s(p)$ is below the thresholded range. This highlight transfer process was carried out only for face area with eyes and lips.

$$\nabla R_s(p) = \begin{cases} \nabla E_s(p) \text{ if } \beta(p)||\nabla E_s(p)|| > ||\nabla T_s(p)|| \\ \nabla T_s(p) \text{ otherwise} \end{cases} \quad (6)$$

$$\lambda_s + (\lambda_l - \lambda_s) * \gamma(p) \quad (7)$$

For lips, the highlights of the target will remain the same. It was assumed that lip coloring will only vary the lips' colors of the target while keeping the texture of the target's lips the same.

B. Details Transfer

Detail layer is not affected by makeup and describes the inherent facial characteristics of the target and the reference. Details layer can be obtained by the weighted sum of the target's detail and the reference's detail layer. The weight of the target w_T was set as 1 and that of the reference w_E was 0.5. The weight values ensure that the target still retains its original face features while having added makeup attributes such as eye lashes. The sum of the weights of the details also do not need to be equal to zero.

This same operation was carried out in lip detail transfer.

(a) Binary Mask Used to Blend Result's edges (b) Binary Mask Used to Blend target's background

Fig. 6: Circular Averaging Filtered Binary Mask

$$R_d = w_T T_d + w_E E_d \qquad (8)$$

C. Color Transfer

Color transfer was achieved by alpha blending between the reference and the target image for each color channel a and b. Unlike detail layer transfer, the weights of the target and the reference should add up to one as seen in 9. ζ was chosen to be 0.8 for both color channels a and b. This same operation was carried out for lips color transfer.

$$R_c(p) = \begin{cases} (1-\zeta)T_c(p) + \zeta E_c(p) \text{ if } p \in C_3 \\ T_c(p) \text{otherwise} \end{cases} \qquad (9)$$

D. Blending

The color layer a and b, the large scale layer R_s and the small scale layer R_d are all combined together to give R_sum and added with the remaining target background T_B. The edges of the edited layers R_{sum} are blurred by by using a circular averaging filtered binary mask with a 'disk' radius of 15 as shown in Figure 6 .

$$R = (1-f)T_B + f * R_{sum} \qquad (10)$$

VIII. TESTING

Testing was performed on the sample target face without makeup as shown in Figure 1 (b) by transferring makeup from models with a different age, different skin color and a hand-sketched model. The results closely aligned with makeup profile of the model.

In Figure 7 (a), it can be concluded that the model should be of the same age as the sample image because facial bone structures and aging effects were wrongly transferred as makeup, causing the target's appearance to look older. There were also some distortion in the reference image during warping and some artifacts got transferred to the target as a result.

The makeup transfer in figure 7 (b) proves that linear color scaling works well to change the face color profile because the target still retained her skin color tone. However due to

(a)
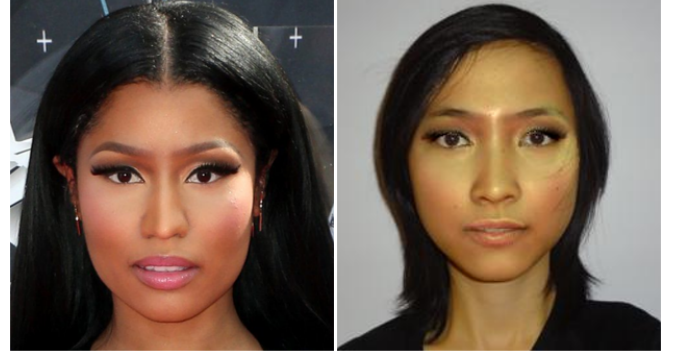
(b)

(c)

Fig. 7: Test Results - Reference Images on the Left and Test Images on the Right

assumption of uniform lighting, the strong lights from the reference also got transferred as makeup to the target.

Lastly, it can be seen that the algorithm also works on artwork of models with makeup. This is in fact an ideal makeup transfer scenario since it will be much easier to decompose the lightness layer into face structure layer and detail layer. The target can fully retain its shape while obtaining only makeup information from the reference.

In all of these images, it was found that lip transfer process can be improved to better represent the lip's style of the reference while preserving its own lip texture.

While testing with a target image of another person, the

testing couldn't be completed since there was shortage of MATLAB memory error while performing highlight gradient transfer. The test images successfully generated before up to Gradient Transfer were stored in MATLAB output files folder. There will be future investigation on whether the current algorithm is subject to the target image's size, image quality and attributes such as lighting and gradients.

IX. CONCLUSIONS

The makeup of the reference image was successfully transferred to a sample target image. Non-linear WLS filter as developed by [3] was also useful particularly for this application in separation of the highlight layer from the detail layer. However, there are limitations with face detection when the front area of the face is not completely visible or when its face contours were disturbed by uneven lighting and shadow. This will result in incorrect detection of face landmarks and thus misalignment between the reference and the target. The makeup transfer will not give desirable results consequently. The highlight transfer also requires the reference image to have uniform lighting. In addition, computation time for lightness layer decomposition into large scale and detail layers as well as Poisson Gradient editing is too high to be implemented in the actual user application. For future work,computation efficiency of gradient editing and weighted least square filter should be improved.

ACKNOWLEDGMENT

REFERENCES

[1] S. Milborrow and F. Nicolls, Locating Facial features with an Extended Active Shape Model, in EECV, 2008.

[2] Z. e. a. Farbman, Edge-preserving decompositions for multi-scale tone and detail manipulation, ACM Transactions on Graphics (TOG), vol. 27, no. 3, p. 67, 2008.

[3] X. Chen and e. al., Face illumination transfer through edge-preserving filters, in Computer Vision and Pattern Recognition (CVPR), 2011.

[4] D. Guo and T. Sim, Digital face makeup by example, in Computer Vision and Pattern Recognition, 2009.

[5] F. L. Bookstein, Principal Warps: Thin-Plate Splines and the Decomposition of Deformations, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pp. 567-584, 1989.