Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

Wavelet Flow: Optical Flow Guided Wavelet Facial Image Fusion

Hong Ding[1,2], Qingan Yan[3], Gang Fu [2] and Chunxia Xiao[2]

[1]School of Information and Statistics, Guangxi University of Finance and Economics, China, 530003
[2]School of Computer Science, Wuhan University, 430072
[3]JD.com American Technologies Corporation, CA, 94043

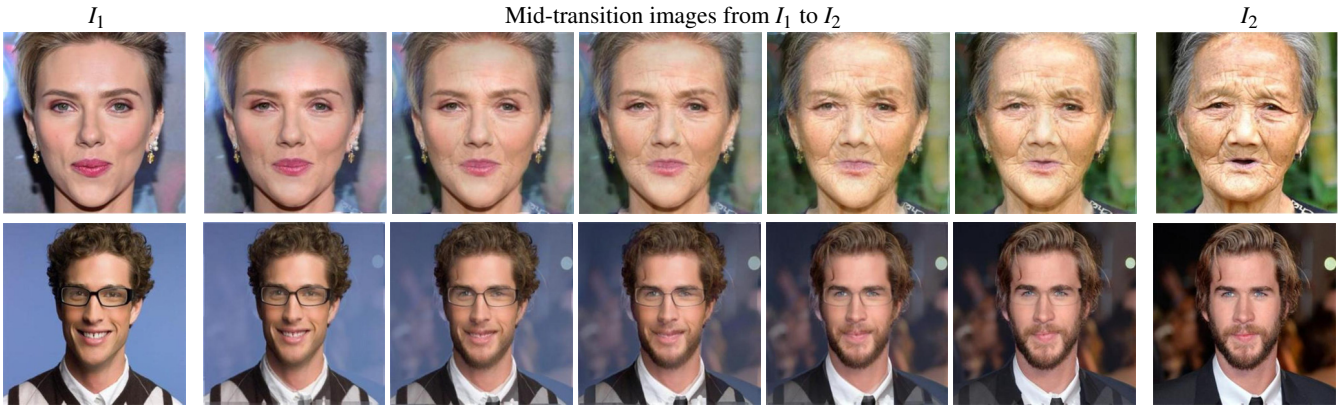


Figure 1: *Given a pair of images I_1 and I_2, our algorithm synthesizes natural in-between images. Note the significant variation in lighting, color, facial expression and background between the two input images.*

Abstract

Estimating the correspondence between the images using optical flow is the key component for image fusion, however, computing optical flow between a pair of facial images including backgrounds is challenging due to large differences in illumination, texture, color and background in the images. To improve optical flow results for image fusion, we propose a novel flow estimation method, wavelet flow, which can handle both the face and background in the input images. The key idea is that instead of computing flow directly between the input image pair, we estimate the image flow by incorporating multi-scale image transfer and optical flow guided wavelet fusion. Multi-scale image transfer helps to preserve the background and lighting detail of input, while optical flow guided wavelet fusion produces a series of intermediate images for further fusion quality optimizing. Our approach can significantly improve the performance of the optical flow algorithm and provide more natural fusion results for both faces and backgrounds in the images. We evaluate our method on a variety of datasets to show its high outperformance.

CCS Concepts

• ***Computing methodologies*** → ***Computational photography;***

1 Introduction

Facial images fusion and editing is an active research area in computer graphics and computer vision [MAL16]. Two input facial images, for example, the photos selected from Internet, may have different backgrounds, facial expressions and lighting conditions. Most methods of facial image editing are based on face mask, which fail to combine the edited face with background smoothly [KSS12]. Furthermore, the state of the art GAN methods [LPSB17, BCW*18] are unable to generate facial image with natural background as there are no training datasets for the background of particular facial images (as shown in Figure 6). Our method offers an intuitive way to directly transfer these valuable information. To the best of our knowledge, there are few works on establishing optical flow for fusing facial images including background regions.

There are several factors determining the fusion quality for images with both the faces and the backgrounds. First, if the faces are extracted and edited, and when the fused faces are blended to the background, the edited faces usually not fit at the boundaries with the background. Second, as the backgrounds are usually complex, producing satisfactory background fusion results are difficult. Third, if the faces and backgrounds are fused and edited independently, the overall illumination of the fused image will not be con-

sistent. Finally, if the edited facial image is dependent on the facial image dataset, such as the methods [KSS12, UGP*17], the quality of the dataset will have great impact on the results of optical flow. Thus, to produce satisfactory facial image fusion results, it is necessary to build a globally consistent fusion method for the facial images, which is essentially important for avoiding above problems.

In this paper, we propose a novel wavelet flow method to perform facial image fusion. Our method is based on the following observations: Since optical flow is defined as the two dimensional motion of brightness patterns between two images, if the background and the lighting of the input facial image pair can be processed similarly, the difficulty of estimating optical flow for facial image pair will be alleviated. By this means, we can obtain consistent optical flow for both the face and background of the input facial image pair.

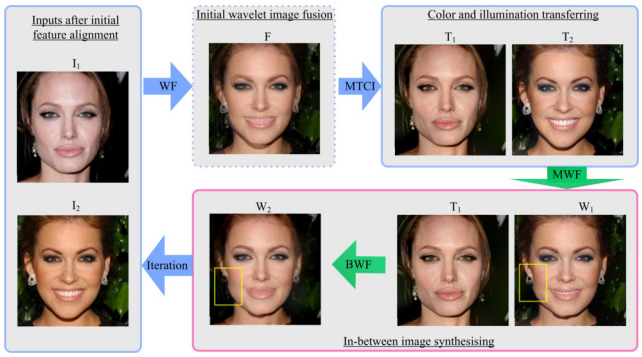


Figure 2: *Overview of wavelet flow system. WF is wavelet fusion. MTCI is multi-scale transfer of color and illumination from F to I_1 and I_2. MWF is optical flow guided monolateral wavelet fusion. BWF is optical flow guided bilateral wavelet fusion. T_1 and T_2 are the results of transferring the lighting and color from F to I_1 and I_2, respectively. W_1 is the monolateral wavelet fusion result from T_1 and T_2, and W_2 is the bilateral wavelet fusion result from T_1 and W_2.*

Our method can handle image pair with large difference in brightness, color and background, and can effectively avoid the blurring and ghosting artifacts often occurring in image fusion. We have evaluated the proposed wavelet flow approach on a large number of facial images downloaded from the Internet, or from the public data set, such as IMDB-WIKI dataset [RTG16]. Experimental results demonstrate that our approach substantially improves the results of existing optical flow estimation methods in both faces and backgrounds.

The main contributions of this paper are threefold: (1) Propose a multi-scale wavelet transform algorithm for multi-scale transfer of low frequency map. Thus there is no need to extract faces from facial images and blend them back redundantly. (2) Introduce optical flow guided wavelet fusion algorithms to produce high quality image fusion results, and avoid the blurring and ghosting artifacts occurred in the edited facial image. (3) Develop a wavelet flow algorithm to produce in-between facial images with gradual expression and texture changing.

2 Related work

Background editing and facial Relighting: Global statistics from one image to another can successfully mimic a visual look. For example, the global color transfer methods [XSXM13, ZYZ*19, ZYL*17, FG19] applied global transformation on the source regions to match color statistics of the target regions. whereas these methods are mainly for scene lighting editing, not for facial images, and their transfer effects are still similar to original images. Shih et al. [SPBF14] introduced a local and multiscale technique to transfer the local statistics of an example portrait onto a new one, but the transfer technique is local and mainly tailored for headshot portraits. Altering the illumination on a face is also a common operation for face recognition and face relighting [PTMD00], as well as the facial image style editing [FJS*17], however, they can not process the background of the images. As photographic images usually involve large illumination and background changes, different from above methods, we focus on changing the image background and relighting of the images by using optical flow guided wavelet transform, and aim at processing image fusion.

Wavelet transform: Wavelet transform has been widely used in image editing, such as image denoising [ZY11] and image fusion [WHZB16]. Image fusion techniques based on multi-scale wavelet transform also have been proposed, which improve the fusion performance, for example, Li et al. [LLYZ16] proposed two multifocus image fusion techniques based on multi-scale and multi-direction neighbor distance. In above works, wavelet transform is used to extract various features of the images, then various methods are used to deal with the high frequency wavelet coefficient of the images. The information represented by the low frequency coefficient, such as the color and illumination of the image, is ignored. Different from the existing image fusion methods using wavelet transform, our approach deals with not only the high frequency coefficient but also the low part, and edits both faces and backgrounds of the facial images. Furthermore, our method can address the problem of image blurring and ghosting produced in the fusion processing.

Flow estimation: Optical flow is defined as the two dimensional motion of brightness patterns between two images. The recent success of convolutional neural networks has led to the attempt of using high-level information for the optical flow problem. Dosovitskiy et al. [DFI*15] presented FlowNet to learn optical flow using CNN. An improved version of FlowNet, the FlowNet2, is proposed by Ilg et al. [IMS*17]. One problem in learning optical flow is the limited amount of training data, to receive higher accuracy, larger training data is required. For the specific case of facial images, there are many alignment literatures available. Zhu et al. [ZVGH09] estimated non-rigid facial motion under moderate illumination changes, by introducing an outlier model and allowing for local bias-gain variations. Kemelmacher et al. [KSS12] developed collection flow to compute optical flow between a pair of faces, where the corresponding database was needed to establish for each face. However, both methods do not provide optical flow for both face and background in an image simultaneously, and need to perform the tasks such as face extraction, image recombination, benchmark dataset establishing. It makes the application scope of these methods be limited.

In this paper, guided by optical flow based image warping, we

focus on fusing the facial images including both foreground (faces) and background.

3 Overview

Given a pair of images, to synthesize the in-between images (as illustrated in Figure 1), our system consists of the following main steps:

Initial feature alignment: Our inputs are a pair of facial images which may have different postures, lighting styles, background and color. In initial feature alignment, different from face alignment of 3D [ZX16], we detect the facial landmarks of the images using the method [SLC10], then roughly align the eyes and mouth of the inputs using an affine transform similar to [JMAK10].

Initial wavelet image fusion: We leverage wavelet multi-scale algorithm to decompose the input facial images into low frequency images and high frequency images. The low frequency images present the general characteristics of the input image, while the high frequency images reflect the detail of the image. We perform wavelet image fusion for the image pair to produce the initial fused image.

Color and illumination transferring: We apply multi-scale wavelet coefficients for color and illumination transfer, and apply wavelet coefficients transfer to make the color and illumination of the input images similar to the initial fused image.

In-between image synthesis: With above preprocessed images, we apply optical flow guided wavelet fusion to produce in-between synthesized images of the input facial image pair. Flow guided wavelet fusion includes monolateral wavelet fusion and bilateral wavelet fusion, which play different roles.

To produce more compelling results, Step 2 to Step 4 can be performed in the iterative way.

Figure 2 gives the system overview of the proposed method. Figure 3 is the overview of initial wavelet image fusion. Figure 4 and Figure 5 show the multi-scale background transfer process. Figure 7 and Figure 8 give the algorithm overview for optical flow guided monolateral wavelet fusion and bilateral wavelet fusion, respectively. More sophisticated system overview is presented in section 5.

4 Multi-scale transform of images

With the image pair I_1 and I_2, as shown in Figure 3, we intend to transfer the color and illumination of the I_2 into I_1. The goal in this section is to match the visual style of the I_2 without changing the identity of the I_1. This means we let the output to keep the same person as the I_1 while with the background (including lighting and color) matching that of the I_2.

Wavelet transform can extract various features of images, sub-images of low frequency presents the general characteristics of the original image, while the high frequency sub-images reflects the details of the original image.

With these advantages, we process the background and illumination using wavelet transformation. We perform multiple scales of wavelet transform for the images to deal with the wide range of

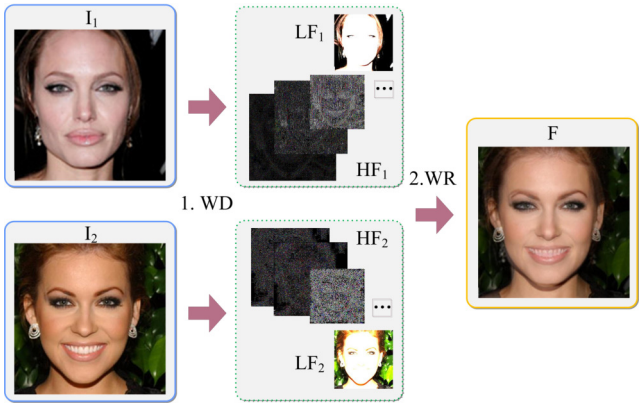


Figure 3: *The overview of wavelet image fusion. WD is wavelet decomposition, WR is wavelet reconstruction, LF_1 is the low frequency of I_1, LF_2 is the low frequency of I_2, HF_1 is the high frequency of I_1 and HF_2 is the high frequency of I_2.*

appearances that faces and backgrounds exhibit. Multiple scales of wavelet transform allow us to better capture the general appearance and details of these images for image editing.

Given an input image, multi-scale wavelet decomposition of the image is defined as follows:

$$\begin{cases} c_{k;n,m} = \sum_{l,j} h_{l-2n} h_{j-2m} c_{k+1;l,j}, \\ d^1_{k;n,m} = \sum_{l,j} h_{l-2n} g_{j-2m} c_{k+1;l,j}, \\ d^2_{k;n,m} = \sum_{l,j} g_{l-2n} h_{j-2m} c_{k+1;l,j}, \\ d^3_{k;n,m} = \sum_{l,j} g_{l-2n} g_{j-2m} c_{k+1;l,j}. \end{cases} \tag{1}$$

where n is the row subscript, m is the column subscript, $\{h_k\}_{k\in Z}$ satisfies the wavelet scale equation, $g_k = (-1)^k \bar{h}_{-k+1}$, h, g are called standard filter, $\bar{h}$ is the conjugate h, c is low frequency coefficient, d is high frequency coefficient and k is the layer number of wavelet transform.

The sub-images produced by wavelet decomposition with one level have four parts:

$$\begin{pmatrix} c_{k;n,m} & d^1_{k;n,m} \\ d^2_{k;n,m} & d^3_{k;n,m} \end{pmatrix}$$

where each sub-image is a quarter of the size of original image. The sub-image of low frequency for each level of transformation is recursively decomposed. The reconstruction process is similar. By using this means, the tower structure of the two-dimensional wavelet transform is constructed. Hence, the number of sub-images of the high frequency parts is $3 \times N$ times of that of the low frequency part, where N is the layer number of wavelet decomposition.

5 Wavelet Flow algorithm

In this section, we give the technical details for optical flow guided wavelet fusion.

5.1 Wavelet image fusion

Both Gaussian pyramid algorithm and wavelet analysis algorithm are widely used for images analysis. Gaussian pyramid produces multiple sets of signals with different scales through Gaussian blurring and down sampling. Two-dimensional Gaussian blurring function is

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{2}$$

where σ is the blurring radius, and x and y are the relative coordinates of the peripheral pixels to the center pixels. Gaussian pyramid is a series of images based on sampling from the Gaussian blurring.

However, the Gaussian pyramid algorithm is only for a single frequency. The most familiar analogy to the wavelet analysis is the digital microscope, as it combines multi-scale and multi-resolution techniques. In contrast, using wavelet analysis, we can obtain more sophisticated internal structure of images under different frequencies, which is preferred for image fusion. The overview of wavelet fusion images is shown in Figure 3.

In Figure 3, I_1 and I_2 are decomposed into high and low frequency parts by using wavelet transformation, respectively. Their coefficients can be expressed respectively as:

$$\begin{aligned} w_1 &= c_1 + d_1, \\ w_2 &= c_2 + d_2. \end{aligned} \tag{3}$$

Different from the multi-scale color and illumination transfer, where the low frequency part of one image is composed with the high frequency part of the other image, in wavelet image fusion, the two images are composed with different weights. For example, both the low and high frequency coefficients of one image are strengthened linearly and those of the other images are weakened linear by μ before the fusion. The process can be expressed as:

$$W_F = \mu(c_1 + d_1) + (1-\mu)(c_2 + d_2). \tag{4}$$

where μ is the strength weight, $0 \leq \mu < 1$. Then, the fused low and high frequency parts are reconstructed by using inverse wavelet transform to produce the output image F. The parameter μ is controlled by the iteration. The higher the μ is, the greater fusion strength of I_1 is, and the smaller that of I_2 is. In our experiment, if m in-between images are generated between I_1 and I_2, μ is set as $1/(m+1), 2/(m+1), ..., i/(m+1), ...m/(m+1)$. Here m is the total number of iterations. In the ith iteration, μ is equal to $i/(m+1)$.

5.2 Multi-scale transfer of color and illumination

In this section, we illustrate the principle of the color and illumination transfer between two facial images. The low frequency map contains the appearance of illumination and color of the image, and the high frequency map includes the detail of the image. We aim at matching the appearance of the low frequency map of I_1 to I_2. The wavelet coefficients of an image can be represented as:

$$w = c + d. \tag{5}$$

where w is the wavelet coefficient of the image, c is the low frequency coefficient of w, d is the high frequency coefficient of w. The operation + means the combination of coefficients. The coefficients of I_1 and I_2 can be presented as w_1 and w_2, respectively:

$$\begin{aligned} w_1 &= c_1 + d_1, \\ w_2 &= c_2 + d_2. \end{aligned} \tag{6}$$

The result of multi-scale transfer of background information of I_1 to I_2 is T_1, which can be expressed as:

$$w_{12} = c_1 + d_2. \tag{7}$$

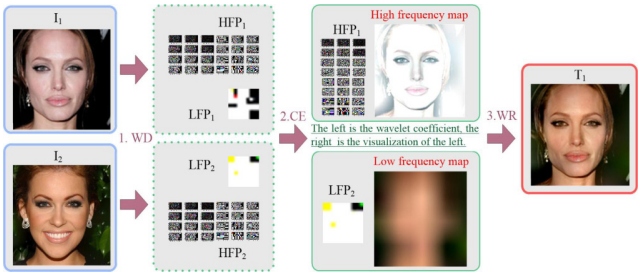


Figure 4: *The overview of multi-scale wavelet color and illumination transfer. WD is wavelet decomposition, CE is coefficient exchanging of wavelet, WR is wavelet reconstruction. LFP_1 is the low frequency part1. LFP_2 is the low frequency part2. HFP_1 is the high frequency part1 and HFP_2 is the high frequency part2.*

In Figure 4 we present the multi-scale color and illumination transfer results. Two facial images I_1 and I_2 are inputs. First, these two images are decomposed into high and low frequency parts by N-scale wavelet transformation, respectively. Then, the high frequency part of I_1 and the low frequency part of I_2 are reconstructed to produce output T_1. The T_1 keeps the identity of I_1 with the background and overall lighting of I_2. N controls the degree of multi-scale transform.

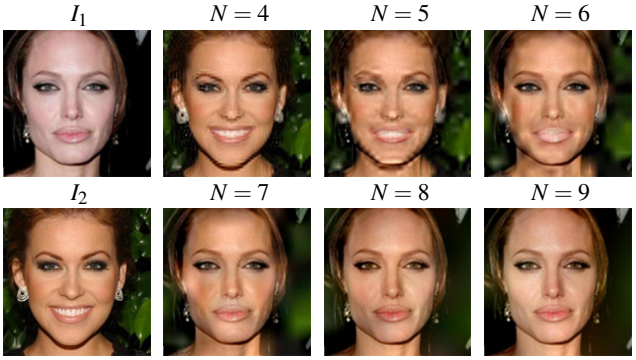


Figure 5: *Color and illumination transfer. For images with 1000x1000 pixels, N is set from 4 to 9. When N = 8, the color and lighting of I_2 are transferred to I_1 well.*

Figure 5 shows the results of multi-scale transfer of lighting and color from I_2 to I_1, respectively, where scale N is set from 4 to 9. N is usually set by experience. In our experiments, for facial images with 1000x1000 pixels, when $N = 8$, it is good enough to extract the color and illumination through the low frequency parts. If images have larger size, we set a larger value for N.

To illustrate the effectiveness of the multi-scale transfer of color and illumination process, we compared our method with the deep learning algorithm [LPSB17] and Reinhard algorithm [RAGS02] in experiments in Figure 6.

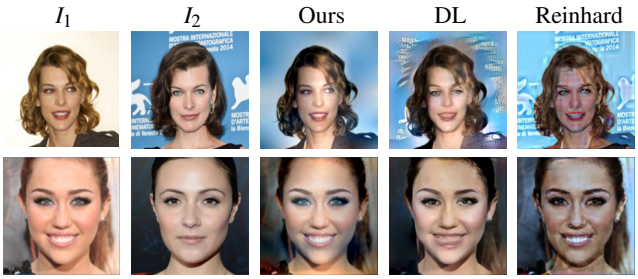


Figure 6: *Comparison with deep learning [LPSB17] and Reinhard algorithm [RAGS02] on color and illumintaion transfer for facial images. Column 1 and 2 are inputs, column 3, 4 and 5 are the results of transferring the color and illumintaion from I_2 to I_1 using our algorithm, deep learning and Reinhard algorithm, respectively.*

5.3 Optical flow guided wavelet fusion

Without dense correspondence, wavelet fusion images may be blurred, the following wavelet flow algorithm try to solve the above problems to some extent. We develop the wavelet flow by combining wavelet fusion with optical flow based image warping. According to the different degrees of wavelet fusion, we define optical flow guided monolateral wavelet fusion and optical flow guided bilateral wavelet fusion. In this paper, we use optical flow proposed by Liu [Liu09].

5.3.1 Flow guided monolateral wavelet fusion

Flow guided monolateral wavelet fusion generates warped image which changes the expression and shape of the face while keeping the texture and color unchanged. The procedure of optical flow guided monolateral wavelet fusion is shown in Figure 7, it contains the following five main steps:

Step 1: Initial feature alignment. We detect the facial landmarks of the images and roughly align the eyes and mouth of the inputs using an affine transform.

Step 2: Multi-scale color and illumination transfer. We decompose I_1 and I_2 into multi-scale sub-images. Then the low frequency of I_2 and the high frequency of I_1 are combined to produce T_1, with color and illumination matching those of I_2. The multi-scale background transfer to produce T_1 can be expressed as:

$$w_{T1} = c_2 + d_1. \tag{8}$$

where w_{T1} is the result of multi-scale transfer of background from I_2 to I_1, c_2 is the low frequency coefficient of I_2, d_1 is the high frequency coefficient of I_1.

Step 3: Monolateral fusion. This step is to blend T_1 and I_2 to generate F_1. Monolateral fusion for T_1 and I_2 can be expressed as:

$$w_{F1} = \mu_1 w_{T1} + (1-\mu_1) w_2 = \mu_1 (c_2 + d_1) + (1-\mu_1)(c_2 + d_2). \tag{9}$$

where w_{F1} is the coefficients of F_1, μ_1 is the weight of T_1, $1-\mu_1$ is the weight of I_2, $0 \le \mu_1 < 0.5$.

Step 4: Optical flow guided image warping. We compute the optical flow between the image F_1 to image T_1 using [Liu09], then based on the optical flow vector, we warp image F_1 to image T_1 and obtain image O_1.

Step 5: Iteration operation. I_2 is set as O_1, step 2 to 4 are repeated, until O_1 will not change or meet user's requirements. (The user is satisfied with the produced fused images on the background, facial expressions and facial features, etc.)

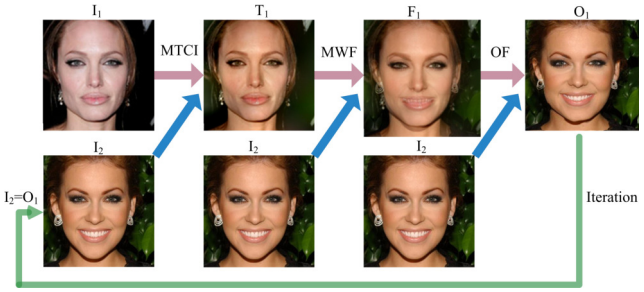


Figure 7: *The procedure of optical flow guided monolateral wavelet fusion. MTCI is multi-scale transfer of color and illumination. MWF is monolateral wavelet fusion. OF is optical flow guided warping.*

In the process of the monolateral fusion for I_2 and T_1, the weight μ_1 of T_1 is much larger than that of I_2. The purpose of the fusion is to generate F_1 which is closer to the I_2 than T_1 on both background and face. μ_1 is used to control the degree that the foreground of I_2 is close to that of I_1. The higher μ_1 the greater fusion strength of I_1 and the smaller that of I_2. We usually set μ_1=0.3 to produce satisfying results.

Using above optical flow guided monolateral wavelet fusion method, the image I_2 is warped to image I_1 keeping its identity and illumination unaltered, while its expression and face contour changes, as shown in Figure 7.

5.3.2 Flow guided bilateral wavelet fusion

Flow guided bilateral wavelet fusion generates warped image which changes not only the expression and shape of the face but also its texture and color. The procedure of optical flow guided bilateral wavelet fusion is shown in Figure 8, it contains the following five main steps:

Step 1: Initial feature alignment. Similar to step 1 in the flow guided monolateral wavelet fusion.

Step 2: Multi-scale color and illumination transfer. Similar to the Step 2 in Flow guided monolateral wavelet fusion, we decompose I_1 and I_2 into multi-scale sub-images. Then the sub-image of the low frequency of I_2 and those of the high frequency of I_1 are reconstructed to produce T_1, with color and illumination matching those of I_2.

Step 3: Bilateral fusion. One fusion is that the weight of T_1 is much less than that of I_2, in this way F_1 is produced. The other

fusion is the weight of I_2 is much less than that of T_1, in this way F_2 is produced. Bilateral fusion for T_1 and I_2 can be expressed as:

$$w_{F1} = \mu_1 w_{T1} + (1-\mu_1)w_2 = \mu_1(c_2+d_1) + (1-\mu_1)(c_2+d_2). \quad (10)$$

where w_{F1} is the coefficients of F_1 (the result of monolateral fusion), μ_1 is the weight of T_1, $1-\mu_1$ is the weight of I_2, and $0 \le \mu_1 < 0.5$.

$$w_{F2} = (1-\mu_2)w_{T1} + \mu_2 w_2 = (1-\mu_2)(c_2+d_1) + \mu_2(c_2+d_2). \quad (11)$$

where w_{F2} is the coefficients of F_2 (the result of monolateral fusion), and $0 \le \mu_2 < 0.5$. When $\mu_2 = 0$, $w_{F2} = w_2$, F_2 is I_2, wavelet flow with bilateral fusion is equal to wavelet flow with monolateral fusion.

In the process of the bilateral fusion between I_2 and T_1, the purpose is to generate F_1 and F_2 which are more similar to each other in both background and face.

Step 4: Optical flow guided image warping. The optical flow between the facial images F_1 and F_2 is computed using [Liu09]. With the optical flow vector, image F_1 is warped to F_2 to produce image O_2.

Step 5: Iteration operation. I_2 is set as O_2, step 2 to 4 are repeated, until O_2 will not change or meets user's requirements.

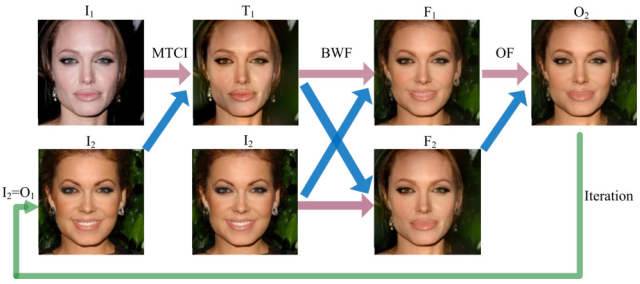


Figure 8: *The procedure of optical flow guided bilateral wavelet fusion. MTCI is multi-scale transfer of color and illumination, BWF is bilateral wavelet fusion, OF is optical flow guided warping.*

Note that comparing with flow guided monolateral wavelet fusion, the main difference is the step 3. In this process, μ_1 and μ_2 are used to control the similarity degree to I_1 and I_2 respectively. Then, the optical flow of F_1 to F_2 is calculated to generate O_2. We usually set $\mu_1 = 0.3, \mu_2 = 0.7$.

Using above optical flow guided bilateral wavelet fusion method, the image I_2 is warped to image I_1, and the expression and face contour of I_2 is also warped to those of I_1, while keeping its identity and illumination unaltered. Furthermore, the warped I_2 will have the texture detail information of I_1, as illustrated in Figure 8.

5.3.3 Similarities and differences

The common point of optical flow guided monolateral wavelet fusion and bilateral wavelet fusion is that both methods produce intermediate images between the input facial image pair. The difference between them is that monolateral wavelet fusion depends more on the effect of optical flow based warping, while bilateral wavelet fusion can generate images incorporating the texture detail information of both the input pair.

In Figure 9, the background of I_1 is first transferred to I_2 to generate T_1, and then the optical flow of T_1 and I_2 is computed to generate B_1. We can observe in Figure 9 the difference between these two types of wavelet flow. The results show that, compared with I_2, (B_1) I_2->I_1 has changed the facial eyes and mouth, while (B_2) I_2->I_1 has changed not only the face expressions, but also the texture detail of the face and background, such as the hair and moustache.

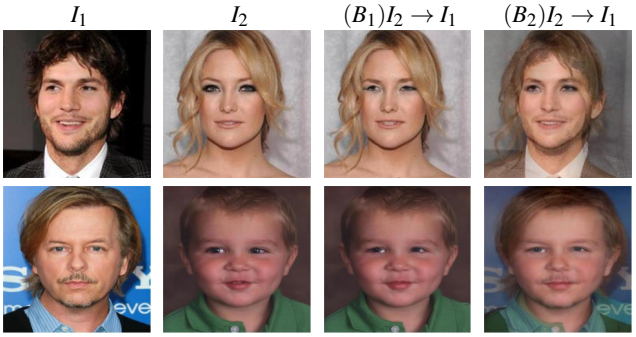


Figure 9: *The comparison of the effect between flow guided monolateral wavelet fusion and bilateral wavelet fusion. In each row, I_1 and I_2 are the input image pair. (B_1) I_2->I_1 is the wavelet flow results from I_2 to I_1 using monolateral wavelet fusion, and transfer the background and color from I_1 to I_2. (B_2) I_2->I_1 is the wavelet flow results from I_2 to I_1 using bilateral wavelet fusion, and transfer the background and color from I_1 to I_2. Two results have different texture and expression transfer effects.*

Wavelet flow can effectively alleviate some problems left by wavelet fusion. For example, when wavelet inverse transformation is used to generate the wavelet fusion image, the wavelet fusion image may be slightly blurred. In addition, when the facial features of the input images differ slightly, wavelet fusion may blur the facial features. Compared with wavelet fusion method, wavelet flow algorithm can improve the results of wavelet fusion, such as fuzzy quality, ghosting and other issues.

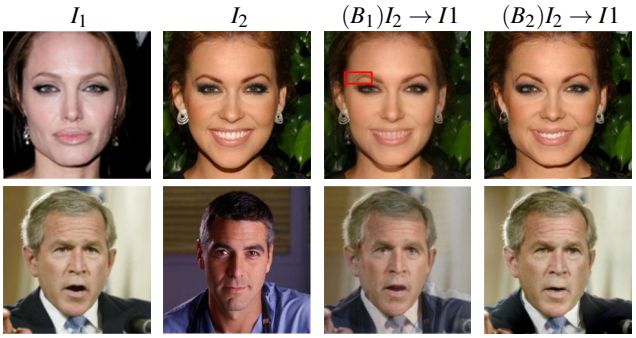


Figure 10: *Comparison results between wavelet fusion and wavelet flow. In each row, the first and the second are input images, the third is the wavelet fusion result of the input images, and the last is wavelet flow result of the input images.*

In Figure 10, we present some comparison results between wavelet fusion and wavelet flow. We can observe that the image of wavelet fusion has some ghosting artifacts because the wavelet fusion is the superposition of the coefficients of the input image pair. While the wavelet flow fusion works not as simple superposition of the wavelet coefficients, but incorporating both background illumination transfer and optical flow guided warping. The results of the wavelet flow significantly alleviate the ghosting artifacts.

5.3.4 Intermediate image synthesizing

By using optical flow guided monolateral wavelet fusion and bilateral wavelet fusion, we can synthesize intermediate images of I_1 and I_2 with significant variation in lighting, color, facial expression and texture, as illustrated in Figure 1 and 12. Figure 2 gives the system overview of the intermediate image synthesizing procedure. The synthesizing procedure has the following main steps:

Step 1: Initial feature alignment. We align the corresponding sense organs for I_1 and I_2.

Step 2: Wavelet fusion. We blend I_1 and I_2 to produce F_1 using different fusion weights for I_1 and I_2, and F_1 has blended lighting and color of I_1 and I_2.

Step 3: Multi-scale color and lighting transfer. Transfer the lighting and color from F to I_1 and I_2 to produce T_1 and T_2.

Step 4: Flow guided monolateral wavelet fusion. With the optical flow estimated from T_2 to T_1, we warp T_2 to T_1 and obtain W_1, which is close to T_1 in expression, etc. To obtain more expressive results, by setting T_2 as W_1, this step can performed in an iterative way.

Step 5: Flow guided bilateral wavelet fusion. With the optical flow estimated from T_1 to W_1, we warp T_1 to W_1 to produce W_2. W_2 has different texture detail level by using different fusion weights (see the red box marked in Figure 2).

The synthesized in-between results can be improved in progressive way: I_1 is set as T_1, I_2 is set as W_2, and step 2 to step 5 can be performed in iterative way. In each iteration, an in-between image is produced, which blends the lighting, color, facial expression and texture of I_1 and I_2 in a weighted way.

Note that by using different fusion weights in Step 4 and Step 5, we can obtain different synthesized images. The wavelet fusion can fuse two different kinds of background to realize the gradual transition of background. More synthesized in-between images are shown in Figure and Figure 12.

5.3.5 Discussions

For input images I_1 and I_2, if we first warp I_1 to I_2 using optical flow method [Liu09], then we fuse the warped image with I_2 using wavelet fusion, ghosting artifacts will exhibit in the fused result, as show in the third column in Figure 11. In fact, for facial images with different texture and background, warping face directly can not obtain natural synthetic results. Consequently after wavelet fusion, the result would become even worse. In contrast, we in advance moderately fuse the facial images after color and lighting transferring, and produce initial fused images, which have similar texture features and make image warping produce better results, as illustrated in Figure 11.

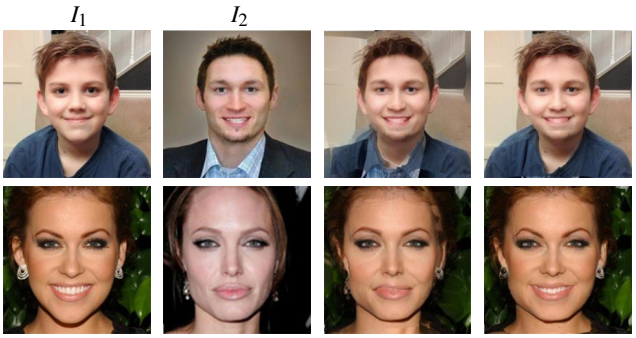


Figure 11: *In each row, the first and the second are input images, the third is the wavelet fusion result after direct image warping, and the last column is the results of optical flow guided bilateral wavelet fusion.*

6 Results

In this section, we present the results of our algorithm on a variety of photo collections. We also compare our method with the state of the art methods. The computer configuration used in this paper is as follows: Processor: Intel® Core^{TM} i7-5500 2.40 GHz CPU, Memory: 16.0 GB.

Our wavelet flow algorithm is divided into three main parts: the multi-scale color and illumination transfer, wavelet fusion and optical flow guided warping. For an input facial image pair with $1000 * 1000$ pixels, in the process of multi-scale decomposition, the scale of wavelet transform is set to 8. We use the optical flow algorithm [Liu09], which is the 'Coarse2FineTwoFrames' function of MATLAB, to compute dense optical flow field in the image warping step. The arguments for optical flow includes: $alpha$ is the regularization weight, $ratio$ is the downsample ratio, $minWidth$ is the width of the coarsest level, n_1 is the number of outer fixed point iterations, n_2 is the number of inner fixed point iterations, and n_3 is the number of SOR iterations. We set $alpha = 0.12$, $ratio = 0.6$, $n_1 = 7$, $n_2 = 1$, $n_3 = 30$ and $minWidth = 80$ for most images.

Wavelet flow results: In Figure 12, we present two examples to illustrate the performance of our method on in-between images production. In each example, there are notable variations in facial expression, lighting, color, background, and texture exhibits between the two input images. For example, in first row, the input facial image pair includes two women with different expression, background, color and lighting. The in-between images show the smiling woman changes to another woman remaining neutral expression, and the color of the background changes from green to black. Second row shows a smiling boy changes to man, remaining neutral expression with beard, and the content of background changes accordingly. Notable that, in all these examples, we generate consistent transition for the in-between images on both foreground and background.

The differences between wavelet flow, traditional image warping and image morphing are: the wavelet flow incorporates the advantages of the traditional image warping and morphing. It can not only realize image warping, but also image fusion. It skillfully combines warping and morphing to achieve better facial image editing effects.

Figure 12: *Wavelet flow for image morping. Column 1 and 7 are the input pair, column 2 to 6 are the fused in-between images.*

Table 1: *Discussion on the parameters of wavelet flow algorithm*

Parameter	Representation	Setting
wavelet scale of low frequency transfer	N	For images of 1000x1000 pixels, it is set to 8. For larger images, N is larger.
weight of wavelet fusion	μ	$1/(m+1)$, ..., $m/(m+1)$, m is the total of the in-between images.
weights of unilateral fusion	μ_1	0.25-0.3
weights of bilateral fusion	μ_2	0.25-0.3
parameters of optical flow	$alpha$, $ratio$, n_1, n_2, n_3, $minWidth$	the default values are 0.12, 0.6, 7, 1, 30, 80

Parameters setting: The main parameters used in this paper are summarized in Table 1. In the Table, the wavelet scale of low frequency transfer, N, controls the transfer of low frequency map, and further affects the result. If N is too large or small, it can not produce complete transfer map. The weight of wavelet fusion, μ, influences the changing process of the low-frequency map of the final result. The weights of unilateral fusion, μ_1, decides the final warped images. Increasing μ_1 will remain more features of T_1 in results. The weights of bilateral fusion, μ_2, controls the final results of warping and texture fusion directly. Large μ_2 will reserve more features of I_2 accordingly. Other parameters about optical flow are presented in [LYT11, Liu09].

Comparison with optical flow methods: We compare our method with Liu's optical flow [Liu09], SIFT flow [LYT11] and Sun's optical flow [SRB10]. Liu's optical flow estimates a flow vector for every pixel by matching intensities. SIFT Flow computes dense correspondence across scenes, and aligns an image to its nearest neighbors in a large image corpus containing a variety of scenes. The SIFT flow consists of matching densely sampled and pixel-wise SIFT features between two images, while preserving spatial discontinuities. Sun's optical flow develops a method that ranks at the top of the Middlebury benchmark by modifying information on flow and image boundaries.

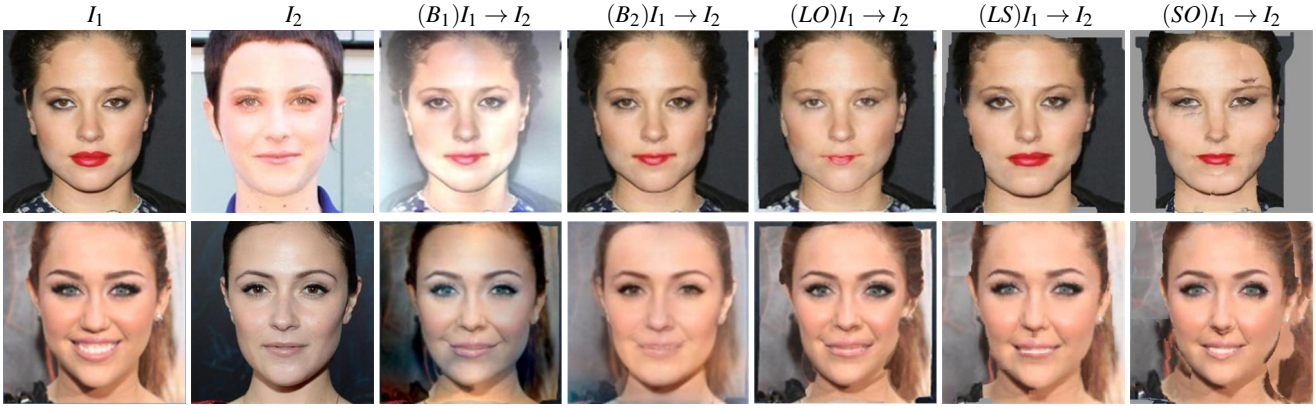
Figure 13(a) shows our comparison results with optical flow method [Liu09], SIFT flow [LYT11] and optical flow method [SRB10]. The effectiveness of our wavelet flow algorithm is reflected in the natural background, the eyes, mouth, face shape, etc. The mouth and eyes in the results, have changed compared with I_1 or I_2. Figure 13(b) shows the flow visualization results corresponding to Figure 13(a). In Figure 13(b), the flow visualization is produced by flow to color method [Mee04]. Different colors represent different motion directions and the darkness represents the speed of motion. The darker the flow color is, the greater the motion amplitude is. The flow visualization results of our method exhibit obvious facial feature change of the two input images. Color inconsistency is significantly weakened, as the background transformation process can effectively narrow the gap between two input images to improve the optical flow effect.

Table 2: *Qabf evaluation of Figure 13(a)*

Row	$(B_1)I_1 \rightarrow I_2$	$(B_2)I_1 \rightarrow I_2$	$(LO)I_1 \rightarrow I_2$	$(LS)I_1 \rightarrow I_2$	$(SO)I_1 \rightarrow I_2$
1	0.32	0.26	0.23	0.22	0.22
2	0.28	0.27	0.27	0.23	0.24

We also use Qabf [PH03] to evaluate above fused images. Qabf is an objective metric for evaluating image fusion quality, and it concerns more about the feature quality. The higher the Qabf value is, the better fusion quality the image exhibits. In Table 2, we present the Qabf values for the fused images in Figure 13(a). It can be observed that the Qabf values of the wavelet flow fusion images are larger than those of the optical flow and sift flow.

In Figure 14 and Table 3, we compare our wavelet flow with Liu's optical flow, SIFT flow and Sun's optical flow on another two examples. Our method effectively warps the source image to the target image, while maintaining the illumination and tone of the source image. For the mid-transition image, our results combine illumination and tone of the source and target images, and the warped features are also natural. As shown in Figure 14, for results of optical flow and sift flow, there are obvious artifacts in face or background (label with red box in the figure) regions. Qabf is also used to evaluate the above fused images. In Table 3, the Qabf values of the wavelet flow images are a little larger than those of op-



(a) Comparison results between wavelet flow, optical flow and sift flow.

(b) Flow visualization results corresponding to (a).

Figure 13: *Results comparing with the flow methods [Liu09], [LYT11] and [SRB10]. (a) is comparison results between wavelet flow, Liu's optical flow, sift flow and Sun's optical flow, (b) is the flow visualization results corresponding to (a). In each row of (a), I_1 and I_2 are input image pair. (B1) I_1->I_2 is the wavelet flow results from I_1 to I_2 with monolateral fusion, and transfer the background and color from I_1 to I_2. (B2) I_1->I_2 is the wavelet flow results from I_1 to I_2 with monolateral fusion, and transfer the background and color from I_2 to I_1. $(LO)I_1$->I_2 is the results from I_1 to I_2 using [Liu09]. $(LS)I_1$->I_2 is the results from I_1 to I_2 using [LYT11]. $(SO)I_1$->I_2 is the results from I_1 to I_2 using [SRB10].*

tical flow and sift flow. The qualitative and quantitative evaluation results show the advantage of our method.

Table 3: *Qabf evaluation of Figure 14*

Method	$(L)I_1 \rightarrow I_2$	$(L)M$	$(L)I_2 \rightarrow I_1$	$(R)I_1 \rightarrow I_2$	$(R)M$	$(R)I_2 \rightarrow I_1$
Ours	0.43	0.30	0.30	0.21	0.22	0.22
OF [Liu09]	0.34	0.29	0.30	0.18	0.20	0.20
SF [LYT11]	0.25	0.22	0.19	0.19	0.20	0.19
OF [SRB10]	0.25	0.22	0.19	0.17	0.18	0.19

Note: L and R mean the left and right example of Figure 14, respectively. OF and SF denote optical flow and sift flow, respectively.

Comparison with landmark-based image warping techniques: We compare our method with landmark-based image warping method [MAL16], which uses dlib to detect 80 characteristic point of the two images respectively, and intelligently blend them later. The comparison result is shown in Figure 15. We can see that the method of [MAL16] sometimes induce visual ghost artifacts while our method can produce better results with clearer details. Furthermore, landmark-based methods rely highly on the quality of input images and the accuracy of characteristic detection. These problems make the method [MAL16] work poorly for some facial examples, such as the images used in Figure 16.

Figure 15: *Comparison with landmark-based image warping method [MAL16]. The yellow boxes show some details of the comparison.*

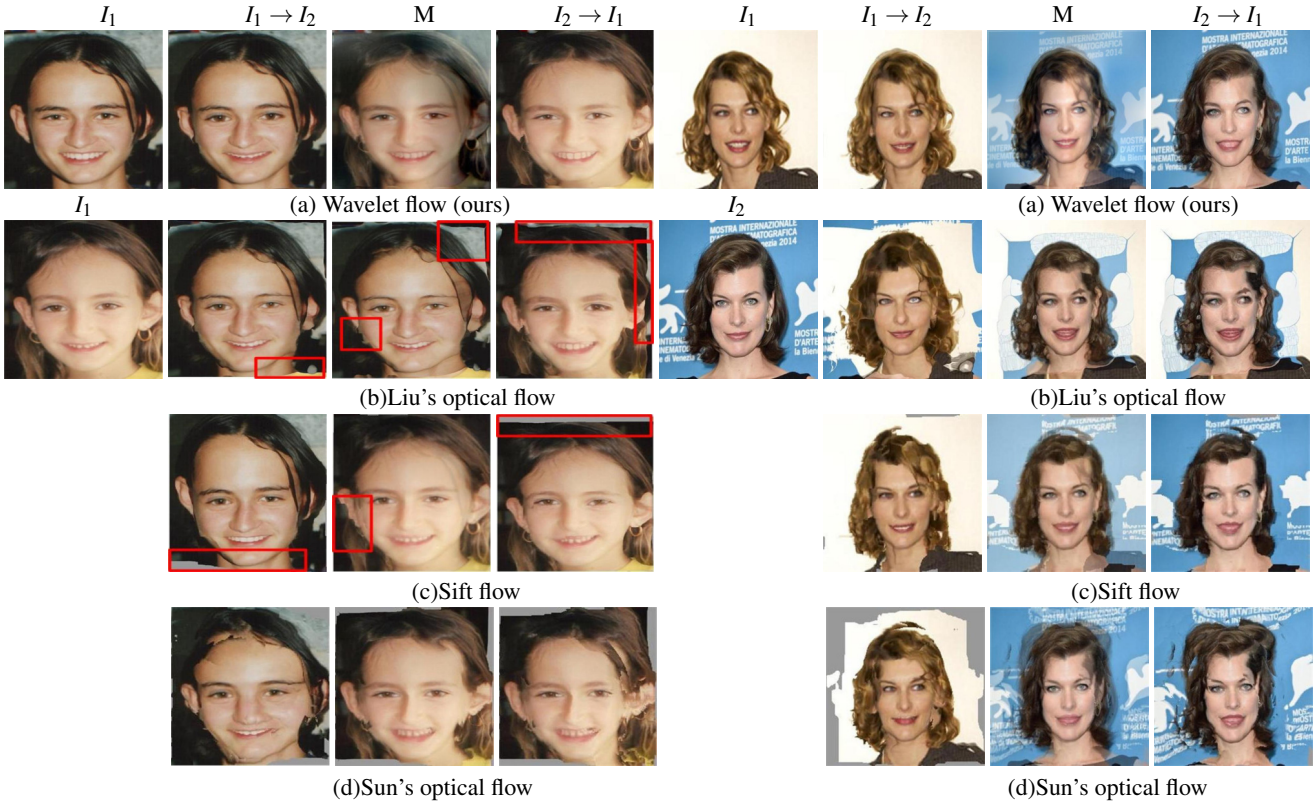


Figure 14: *Results comparing with the flow methods [Liu09], [LYT11] and [SRB10] in image warping and mid-transition image generation. I_1->I_2 is I_1 warped to I_2. I_2->I_1 is I_2 warped to I_1. M is the mid-transition image.*

Comparison with previous state-of-the-art image meld methods: We compare the results of wavelet flow against those of previous state-of-the-art methods, those include comparison results with deep feature interpolation for image content changes (DFI) [UGP*17], Poisson meld [BCCZ08], Darabi's image meld [DSB*12], Barnes' PatchMatch image editing, [BSFD09], Liu's optical flow [Liu09], Liu's sift flow [LYT11] and Sun's optical flow [SRB10]. DFI is for image content changes relies on simple linear interpolation of deep convolutional features from pretrained convnets. Darabi's image meld presents a new method for synthesizing a transition region between two source images. Bhat etal using a DCT-based screened Poisson solver to meld images. Darabi present a image melding method for synthesizing that builds upon a patch-based optimization foundation. Barnes presents interactive image editing tools using a randomized algorithm for quickly finding approximate nearestneighbor matches between image patches. Figure 16 and Table 4 shows the comparison results between our method and the three methods. In Figure 16, we present additional comparison results by adding glasses and opening mouth. In general, the results of DFI are more fuzzy than others. The results of Poisson meld and Darabi's image meld contain distorted pixels and the illumination is not consistent. The faces from Barnes' PatchMatch method have rare changes. Liu's optical flow, Liu's sift flow and Sun's optical flow produce many bad points, while our results are more nature. Table 4 shows the Qabf comparison between the wavelet flow and DFI.

Table 4: *Qabf evaluation of Figure 16*

Method	glasses 1	glasses 2	mouth 1	mouth 2
DFI [UGP*17]	0.56	0.59	0.52	0.57
Poisson [BCCZ08]	0.58	0.13	0.18	0.68
Darabi [DSB*12]	0.47	0.14	0.18	0.45
PatchM [BSFD09]	0.43	0.15	0.18	0.27
Liu's [Liu09]	0.21	0.15	0.19	0.31
Liu's [LYT11]	0.21	0.15	0.18	0.26
Sun's [SRB10]	0.14	0.11	0.14	0.15
Ours	0.61	0.66	0.50	0.41

User Study We also design a user study to measure how users choose between the results under different methods.

1. Methods: We compare our method to the seven previous methods DFI [UGP*17], Poisson meld [BCCZ08], Darabi's image meld [DSB*12], Barnes' PatchMatch image editing [BSFD09], Li-

Figure 16: *Comparison results with DFI [UGP*17], Poisson meld [BCCZ08], Darabi's image meld [DSB*12], Barnes' PatchMatch image editing, [BSFD09], Liu's optical flow [Liu09], Liu's sift flow [LYT11] and Sun's optical flow [SRB10]. The small image embedded in the bottom right corner of the result is the other input image.*

u's optical flow [Liu09], Liu's sift flow [LYT11] and Sun's optical flow [SRB10]. Results of [UGP*17], [BCCZ08], [DSB*12], [BSFD09], [Liu09], [LYT11] and [SRB10] are labeled using R_1, R_2, R_3, R_4, R_5, R_6 and R_7, respectively. Our results are labeled using R_8.

2. Study details: There are 4 facial images editing results in the user study, as shown in Figure 16: wearing glasses and opening mouth for two persons. We perform the user study (similar to [ZNZX16]) with 50 random volunteers to validate the effectiveness of the proposed method in Figure 16. Each volunteer browses the labeled images in Figure 16, a survey is conducted to collect the feedbacks on following questions: (1) Which one do you think exhibits the best overall visual appearance? (2) Which one do you think looks best natural? (3) Which one do you think preserves the clearest textures? (4) Which one do you think has the best harmonious background? For each question, the volunteer is asked to vote for only one result.

Let V_{ij} denotes the total votes of R_i on jth question. To evaluate each method on the individual question, we compute the percentage of votes (PoV) obtained by R_i on the jth question as follows:

$$PoV = (V_{ij}/200) * 100\%, \tag{12}$$

To provide an overall evaluation of different methods, we further calculate the percentage of votes obtained by R_i in all by

$$\overline{PoV} = (\sum_{j=1}^{4} V_{ij})/800 * 100\%. \tag{13}$$

3. Results: In Table 5, we give the percentage of votes obtained by different methods in the survey, where Qu. x denotes the x-th question. From Table 5, we can observe that most volunteers think that our results are visually better than that of [UGP*17], [BCCZ08], [DSB*12], [BSFD09], [Liu09], [LYT11] and [SRB10].

Table 5: *Vote results obtained by different methods.*

Labels	Qu.1	Qu.2	Qu.3	Qu.4	Overall
R_1	15%	12%	6%	10%	10.75%
R_2	3%	16%	18%	21%	14.50%
R_3	2%	4%	20%	19%	11.25%
R_4	0%	2%	1%	1%	1.00%
R_5	1%	0%	4%	2%	1.75%
R_6	0%	0%	3%	4%	1.75%
R_7	1%	0%	1%	5%	1.75%
R_8	78%	66%	47%	31%	55.50%

Limitations: If the background difference between input image pair is too significant, the fused background using our method may suffer from artifacts, as illustrated in Figure 17. In this example, the clothes and collars of the input images have differences in color, shape and content, and it is notable that there are fusion artifacts in these parts of the third image.

Figure 17: *Color inconsistency may happen in the presence of significant background variations.*

7 Conclusion and future work

In this paper, we have presented a wavelet flow method for facial image fusion. Our key idea is to estimate the flow between images by producing in-between images using multi-scale image transform and optical flow guided wavelet fusion. We have shown that lighting and background variations can be easily managed by multi-scale image transfer, and optical flow guided wavelet fusion can effectively warp and fuse the corresponding features of the image pair. In the future, we will extend our method to video situation, and work on facial video fusion.

Acknowledgment

This work was partly supported by the National Key Research and Development Program of China (2017YFB1002600), the NSFC (No. 61672390), Wuhan Science and Technology Plan Project (No. 2017010201010109), and Key Technological Innovation Projects of Hubei Province (2018AAA062). Chunxia Xiao is the corresponding author of this paper.

References

[BCCZ08] BHAT P., CURLESS B., COHEN M., ZITNICK C. L.: Fourier analysis of the 2d screened poisson equation for gradient domain problems. In *ECCV* (2008), pp. 114–128. 10, 11

[BCW*18] BAO J., CHEN D., WEN F., LI H., HUA G.: Towards open-set identity preserving face synthesis. In *CVPR* (2018). 1

[BSFD09] BARNES C., SHECHTMAN E., FINKELSTEIN A., DAN B. G.: Patchmatch: A randomized correspondence algorithm for structural image editing. *Acm Transactions on Graphics 28*, 3 (2009), 1–11. 10, 11

[DFI*15] DOSOVITSKIY A., FISCHER P., ILG E., HAUSSER P., HAZIRBAS C., GOLKOV V., VAN DER SMAGT P., CREMERS D., BROX T.: Flownet: Learning optical flow with convolutional networks. In *ICCV* (2015), pp. 2758–2766. 2

[DSB*12] DARABI S., SHECHTMAN E., BARNES C., DAN B. G., SEN P.: Image melding: Combining inconsistent images using patch-based synthesis. *Acm Transactions on Graphics 31*, 4 (2012), 1–10. 10, 11

[FG19] FU GANG ZHANG QING X. C.: Towards high-quality intrinsic images in the wild. In *IEEE International Conference on Multimedia and Expo (ICME)* (2019). 2

[FJS*17] FISER J., JAMRISKA O., SIMONS D., SHECHTMAN E., LU J., ASENTE P., LUKAC M., SYKORA D.: Example-based synthesis of stylized facial animations. *Acm Transactions on Graphics 36*, 4 (2017), 155. 2

[IMS*17] ILG E., MAYER N., SAIKIA T., KEUPER M., DOSOVITSKIY A., BROX T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In *CVPR* (2017), pp. 2462–2470. 2

[JMAK10] JOSHI N., MATUSIK W., ADELSON E. H., KRIEGMAN D. J.: Personal photo enhancement using example images. *Acm Transactions on Graphics 29*, 2 (2010), 1–15. 3

[KSS12] KEMELMACHER-SHLIZERMAN I., SEITZ S. M.: Collection flow. In *CVPR* (2012), pp. 1792–1799. 1, 2

[Liu09] LIU C.: *Beyond pixels: exploring new representations and applications for motion analysis*. Massachusetts Institute of Technology, 2009. 5, 6, 7, 8, 9, 10, 11

[LLYZ16] LI H., LIU X., YU Z., ZHANG Y.: Performance improvement scheme of multifocus image fusion derived by difference images. *Signal Processing 128*, 2 (2016), 474–493. 2

[LPSB17] LUAN F., PARIS S., SHECHTMAN E., BALA K.: Deep photo style transfer. In *CVPR* (2017), pp. 4990–4998. 1, 5

[LYT11] LIU C., YUEN J., TORRALBA A.: Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 5 (2011), 978–988. 8, 9, 10, 11

[MAL16] MALLICK S.: Face morph using opencv ął c++ / python. 1, 9

[Mee04] MEER P.: Robust techniques for computer vision. *Book Emerging Topics in Computer Vision* (2004). 8

[PH03] PIELLA G., HEIJMANS H.: A new quality metric for image fusion. In *ICIP 2003* (2003), vol. 3, IEEE, pp. III–173. 8

[PTMD00] PEERS P., TAMURA N., MATUSIK W., DEBEVEC P.: Post-production facial performance relighting using reflectance transfer. *Acm Transactions on Graphics 26*, 3 (2000), 52–63. 2

[RAGS02] REINHARD E., ASHIKHMIN M., GOOCH B., SHIRLEY P.: Color transfer between images. *IEEE Computer Graphics and Applications 21*, 5 (2002), 34–41. 5

[RTG16] ROTHE R., TIMOFTE R., GOOL L. V.: Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshop* (2016), pp. 252–257. 2

[SLC10] SARAGIH J. M., LUCEY S., COHN J. F.: Face alignment through subspace constrained mean-shifts. In *ICCV* (2010), pp. 1034–1041. 3

[SPBF14] SHIH Y. C., PARIS S., BARNES C., FREEMAN W. T.: Style transfer for headshot portraits. *Acm Transactions on Graphics 33*, 4 (2014), 1–14. 2

[SRB10] SUN D., ROTH S., BLACK M. J.: Secrets of optical flow estimation and their principles. In *CVPR* (2010), pp. 2432–2439. 8, 9, 10, 11

[UGP*17] UPCHURCH P., GARDNER J. R., PLEISS G., PLESS R., SNAVELY N., BALA K., WEINBERGER K. Q.: Deep feature interpolation for image content changes. In *CVPR* (2017), pp. 6090–6099. 2, 10, 11

[WHZB16] WANG L. J., HAN J., ZHANG Y., BAI L. F.: Image fusion via feature residual and statistical matching. *Iet Computer Vision 10*, 6 (2016), 551–558. 2

[XSXM13] XIAO C., SHE R., XIAO D., MA K. L.: Fast shadow removal using adaptive multi-scale illumination transfer. *Computer Graphics Forum 32*, 8 (2013), 207–218. 2

[ZNZX16] ZHANG Q., NIE Y., ZHANG L., XIAO C.: Underexposed video enhancement via perception-driven progressive fusion. *IEEE Transactions on Visualization and Computer Graphics 22*, 6 (2016), 1773–1785. 11

[ZVGH09] ZHU J., VAN GOOL L., HOI S. C. H.: Unsupervised face alignment by robust nonrigid mapping. 1265–1272. 2

[ZX16] ZHU X LEI Z E. A. F. A. A. L. P. A. D. S.: Face alignment across large poses: A 3d solution. In *CVPR* (2016), pp. 2432–2439. 3

[ZY11] ZHU Y., YANG X.: A dyadic wavelet filtering method for 2-d image denoising. *Journal of Signal and Information Processing 2*, 4 (2011), 308–315. 2

[ZYL*17] ZHANG L., YAN Q., LIU Z., ZOU H., XIAO C.: Illumination decomposition for photograph with multiple light sources. *IEEE Trans Image Process PP*, 99 (2017), 1–1. 2

[ZYZ*19] ZHANGČŇLING, YAN Q., ZHU Y., ZHANG X., XIAO C.: Effective shadow removal via multi-scale image decomposition. In *Visual Computer* (2019), vol. 35, pp. 1091–1104. 2