Facial Image Deformation Based on Landmark Detection

Chaoyue Song, Yugang Chen, Shulai Zhang and Bingbing Ni
Shanghai Jiao Tong University, China
{beyondsong, cygashjd, zslzsl1998, nibingbing}@sjtu.edu.cn

Abstract

In this work, we use facial landmarks to make the deformation for facial images more authentic. The deformation includes the expansion of eyes and the shrinking of noses, mouths, and cheeks. An advanced 106-point facial landmark detector is utilized to provide control points for deformation. Bilinear interpolation is used in the expansion and Moving Least Squares methods (MLS) including Affine Deformation, Similarity Deformation and Rigid Deformation are used in the shrinking. We compare the running time as well as the quality of deformed images using different MLS methods. The experimental results show that the Rigid Deformation which can keep other parts of the images unchanged performs better even if it takes the longest time.

1. Introduction

Image deformation, as one of the most popular topics in the area of computer vision and image processing, has been discussed for many years. Recently, the emergence of artificial intelligence has enabled new techniques in image deformation and has achieved impressive achievements, especially in some specific scenarios. The deformation aimed for faces is one of the most popular areas in academe as well as industry. With more and more people pursuing beauty, verisimilar deformed facial images and an automatic process to generate these images are required to meet these people's needs. Accurate deformation methods are continuously in great demand. Before the emergence of artificial intelligence, common manipulations (e.g. expansion, shrinking, and blurring) on facial features have no differences from that on other objects. This is because those manipulations omit special characteristics of facial features. Deep learning methods can extract facial landmarks from facial images, which provides possibilities of warping based on these landmarks. With these landmarks, we can produce more accurate results and make the warped image more authentic. There are amounts of techniques in operating deformations onto facial images and the fundamental operation is image warping, especially warping with control points.

Figure 1. Examples of our deformation result. From top left to bottom right: original image, image with eye expansion, image with nose, mouth, and cheek shrinking, image with both expansion and shrinking.

Expansion and shrinking are two preferred operations on facial images. Expansion is often used in deformations onto eyes and shrinking is often used on noses and mouths. For an automatic facial deformation process, it is intuitive that more landmarks are detected accurately and more authentic

deformed images will be obtained. Dlib[1] can provide a 68-point facial landmark detection. Nevertheless, that is not enough. For example, there are too few points located near the nose which makes it difficult to perform deformations. Thus, to perform deformations with more accuracy, we need a landmark detector that provides more control points.

The key in the deformation stage is to find an accurate mapping function to map one reference point or several adjacent points if required to the wanted point, based on several control points. In addition, we may need to choose different mapping functions according to the demands of different deformations.

In this work, we trained an advanced 106-point facial landmarks detector based on the method proposed by [16], which can provide enough control points for the deformation. Then we implemented the expansion based on bilinear interpolation whose degree of expansion is adjustable. Shrinking is achieved by the Moving Least Squares algorithm (MLS) [7] which includes Affine Deformation, Similarity Deformation and Rigid Deformation. The experimental result shows that our method which combines facial landmark detection and image deformation can provide authentic deformed facial images.

2. Related Work

Facial Landmark detection The research on facial landmark detection can trace back to 1995 when an Active Shape Model (ASM) [3] was proposed. ASM is based on Point Distribution Model and it was improved into Active Appearance Models [2] which consists of Shape Model and Texture Model. With the development of deep learning, convolutional neural network is used in facial detection for the first time [10]. In [10], a deep convolutional neural network (DCNN) is proposed. Later, Face++ [16] improved the accuracy in DCNN and can detect and localize more landmarks with higher accuracy. After that, TCDCN (Tasks-Constrained Deep Convolutional Network) [15], MTCNN (Multi-task Cascaded Convolutional Networks) [14], TCNN(Tweaked Convolutional Neural Networks) [13], DAN(Deep Alignment Networks) [5] came up and performed better and better. Recently, the work [6], [9], and [12] have proposed more methods with higher accuracy and robustness.

Image Warping Image warping is a transformation which maps all positions in one image plane to positions in a second plane [4]. There are three forms of warping which are translation warping, scaling warping and rotation warping in [11]. The common ground for image warping is that a set of handles (also named as control points) is required. However, these methods do not consider the features of the image. In [1], a feature-based image deformation method is proposed to solve these problems. And an image deformation method based on linear Moving Least Squares was proposed in [7] to meet the smoothness, interpolation and identity demands for image deformation. Such deformation has the property that the amount of local scaling and shearing is minimized. Later, an image warping method based on artificial intelligence was proposed in [8] and techniques in artificial intelligence are used more widely in image warping.

[1] http://dlib.net/

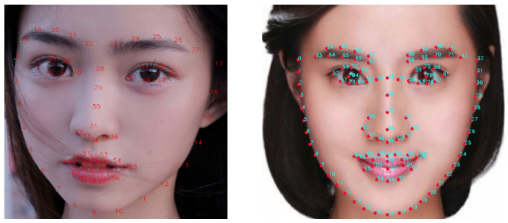

Figure 2. Facial landmark detection. Left: 68-point facial landmark detection. Right: 106-point facial landmark detection.

3. Method

3.1. Facial Landmark Extraction

Accurate facial landmark extraction is the prerequisite of successful facial image deformation. The model we use in this work is proposed in [16]. We improve the 68-point facial landmark detector provided by Dlib to a 106-point one. As shown in Figure 2, there are 33 landmarks for cheeks, 18 landmarks for eyes, 18 landmarks for eyebrows, 15 landmarks for the nose, and 20 landmarks for the mouth.

3.2. Expansion

In this section, we take the manipulation on the left eye E as an example for explanation. The manipulation on the right eye is exactly symmetric to the manipulation on the left eye. Before doing expansion, we need to first determine the control points, thereby making sure the area that needs to be adjusted, which is named as the deformation area in this paper. The center point $E_c(x_c, y_c)$ and the landmark $E_d(x_d, y_d)$ at the corner of eye E are used to determine the boundary of the deformation area. There are two schemes to determine E_c. One is regarding the center landmark as E_c, and the other is regarding the midpoint between the landmark for the outer canthus E_o and the landmark for the inner canthus E_d as E_c. These two schemes are illustrated in Figure 3 and 4.

The deformation area is a circle, whose center is E_c and radius $R_E = ((x_d - x_c)^2 + (y_d - y_c)^2)^{1/2}$. Thus, the pixel $P(x, y)$ within the deformation area satisfies $((x - x_c)^2 + (y - y_c)^2)^{1/2} < R_E$. For each pixel in the deformation area, there is a corresponding reference pixel $P_r(x_r, y_r)$. To find the corresponding reference pixel, we define a parameter

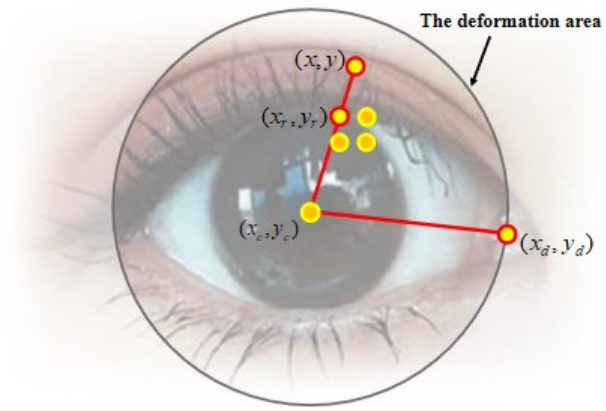

Figure 3. Scheme one for eye expansion. Using the center landmark as E_c. The four orange points whose outlines are yellow are the four pixels used in bilinear interpolation. The intervals between pixels are exaggerated.

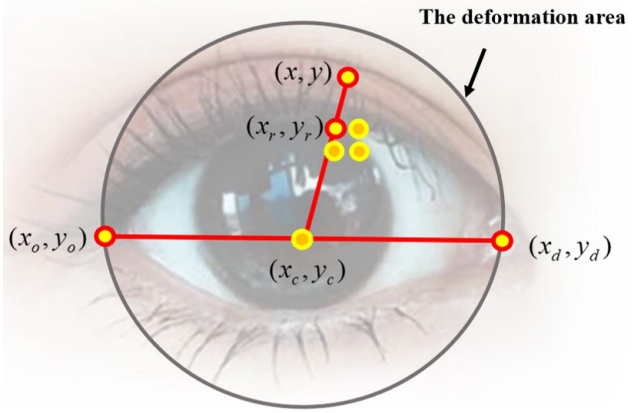

Figure 4. Scheme two for eye expansion. Using the middle point between E_o and E_d as E_c. The four orange points whose outlines are yellow are the four pixels used in bilinear interpolation.

a which is used to determine the expansion scale for each pixel $S(x, y)$. Then the expansion scale is expressed as

$$S(x,y) = 1 - \frac{a}{100} \times (1 - \frac{(x-x_c)^2 + (y-y_c)^2}{R_E^2}) \quad (1)$$

The reference pixel's position is

$$x_r = (x - x_c) \cdot S(x,y) + x_c \quad (2)$$

$$y_r = (y - y_c) \cdot S(x,y) + y_c \quad (3)$$

We use bilinear interpolation to compute the value of $P(x, y)$ after deformation, which is expressed as $f(x, y)$. Then

$$f(x,y) = \frac{\sum_i \sum_j f(x_r+i, y_r+j)(x - x_r - i)(y - y_r - j)}{\sum_i \sum_j (x - x_r - i)(y - y_r - j)} \quad (4)$$

where i and j's values are 0 and 1.

3.3. Shrinking

The methods we use in the shrinking are Moving Least Squares (MLS) [7].

3.3.1 Background

MLS views the deformation as a function f that maps all pixels in the undeformed image to pixels in the deformed image. It needs to apply the function f to each point v in the undeformed image. In [7], the authors consider building image deformations based on collections of points with which the user controls the deformation and this is natural in facial images. Let p be a set of control points and q be the deformed positions of the control points.

For a point v in the image, the best affine transformation $l_v(x)$ can be solved by minimizing

$$\sum_i w_i |l_v(p_i) - q_i|^2 \quad (5)$$

where p_i and q_i are row vectors and the weights w_i have the form

$$w_i = \frac{1}{|p_i - v|^{2\alpha}} \quad (6)$$

Therefore, a different transformation $l_v(x)$ for each v can be obtained.

Next, the deformation function f can be defined to be $f(v) = l_v(v)$. There are three kinds of functions that will induct different deformation effects which are Affine Deformation, Similarity Deformation and Rigid Deformation. The mapping function for Affine Deformation is

$$f_a(v) = (v - p_*)(\sum_i \hat{p}_i^T w_i \hat{p}_i)^{-1} \sum_j \hat{p}_j^T \hat{q}_j + q_* \quad (7)$$

The mapping function for the Similarity Deformation is

$$f_s(v) = \sum_i \hat{q}_i (\frac{1}{\mu_s} A_i) + q_* \quad (8)$$

where $\mu_s = \sum_i w_i \hat{p}_i \hat{p}_i^T$ and A_i depends only on the p_i, v and w_i which can be precomputed. A_i is

$$A_i = w_i \begin{pmatrix} \hat{p}_i \\ -\hat{p}_i^{\perp} \end{pmatrix} \begin{pmatrix} v - p_* \\ -(v - p_*)^{\perp} \end{pmatrix}^T \quad (9)$$

The mapping function for the Rigid Deformation is given by

$$f_r(v) = |v - p_*| \frac{\overrightarrow{f_r}(v)}{|\overrightarrow{f_r}(v)|} + q_* \tag{10}$$

where $\overrightarrow{f_r}(v) = \sum_i \hat{q}_i A_i$. Because the rigid deformation was proposed to make the deformation be as rigid as possible, it performs best in our task.

3.3.2 Implementation Details of Shrinking

In detail, the control points $C_i (i = 1 \cdots 51)$ are exactly the landmarks for noses (15 points), mouths (13 points) and cheeks (21 points). In this paper, we achieve the global adjustment by controlling the moving vector V_m^i for each control point.

Before applying the mapping function, we need to make sure if the facial image is in the right direction. We decide this by checking if the vector $\overrightarrow{V_e}$ from the center landmark of the left eye E_{left} to the center landmark of right eye E_{right} is horizontal. To compare, we define the horizontal vector is V_0. The angle $\beta =< V_e, V_0 >$ decides the direction of V_m^i. V_m^i can be computed after knowing β and the moving distance l_i. As shown in Figure 5, for control points on the left side of the face, $V_m^i = (l_i \cos\beta, l_i \sin\beta)$. For control points on the right side of the face, the situation is opposite and $V_m^i = (-l_i \cos\beta, -l_i \sin\beta)$. Keep the control points on the axes still which means $V_m^i = 0$.

4. Experimental Results

The experiments mainly focus on showing the performance of different methods in the expansion and the shrinking. Two center point determination methods in the expansion and three MLS methods in the shrinking will be discussed in this section.

4.1. Expansion Results

In general, the expansion operation is to make eyes bigger. Thus we set $a = 50$ in this experiment to testify the expansion effect. However, we can also set a to a negative value to make eyes smaller. We also regard this as a part of the expansion process and in this experiment, we set $a = -50$ for another set of results. We use the proposed two methods for center point determination. Three characteristic facial images are tested and the results are shown in Figure 6. The three images from left to right depict people looking straightforward (normal), looking sideways (the pupil is not in the middle), and showing only the side face, in sequence. In the first row, the center point of the eye is the center eye landmark. In the second row, the center point is the midpoint between E_o and E_d. For each set of images in each row, the first is the original image, the second is the deformed image when $a = 50$, and the third is the deformed image when $a = -50$.

As shown in Figure 6, the first set and the second set of images reveal nearly similar effects using different center point determination methods. However, two methods obtain different results when the input contains only the side face. The method using the midpoint of inner canthus and outer canthus as the center point is proposed to fight against the latent distortion that may be caused by the difference between the pupil and the real center point. But in some cases, the first method could also have a good performance as shown by the middle images of 6, when the pupil is different from the real center point, the expansion effect using the first method is better than the effect of the second effect, which is beyond our expectation.

The reason for this result is that the line between the inner canthus and the outer canthus is often lower than the pupil. Thus, the center point is not accurate which makes the deformation also inaccurate. However, in practice, the second method can obtain better effects for some specific images. Thus, the choice of methods depends on the images when doing the expansion.

4.2. Shrinking Results

We compare results and the run-times of three methods: Affine Deformation, Similarity Deformation and Rigid Deformation in this subsection. We then evaluate how the weight α influences deformation results in Rigid MLS deformation.

Deformation results comparison. As shown in Figure 7, for images in each row, the first one is the original image, the second is the deformed image using Affine MLS, the third is the deformed image using Similarity MLS and the fourth is the deformation image using Rigid MLS. Affine MLS and Similarity MLS can both obtain quite satisfying results on the face, but inevitably induce other unstabilizing factors. For example, some distortion appears in the area close to the boundary of the deformed image. Rigid MLS can achieve a satisfying result which keeps the part we don't want to deform unchanged. However, because of the fact that these methods are all based on facial landmarks, a new problem is raised. Will results be affected if the face is not symmetric, especially when only the side face is revealed? One may think there will be distortion on the figures. However, when the figures are clear and the face does not turn a lot, the nose, the mouth and the cheeks can be shrunken successfully and little distortion can be perceived, and it can be proved by the second row in Figure 7.

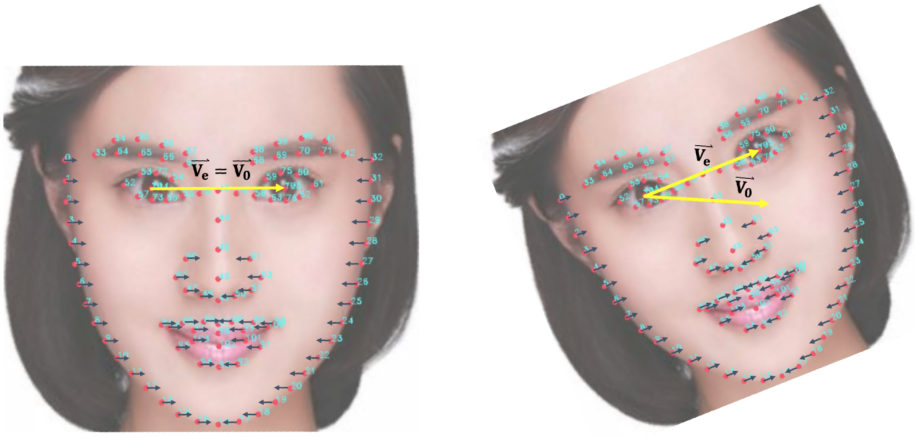

Figure 5. The shift of control points. Left: The face is horizontal. Right: The face is tilted and the moving vectors of the control points on the left side of the face have the same direction with $\overrightarrow{V_e}$

Figure 6. The results of the eye expansion process. Images in the first row are deformed images using the center eye landmark as the center point. Images in the second row are deformed images using the midpoint between E_o and E_d as the center point.

Method	Figure 7 (top)	Figure 7 (middle)	Figure 7 (bottom)
Affine MLS	0.49s	0.53s	0.64s
Similarity MLS	0.88s	0.89s	1.06s
Rigid MLS	0.90s	0.89s	1.06s

Table 1. Deformation times for the various methods.

Run-time Comparison. Three methods have different run-times. As shown in Table 1, Similarity MLS and Rigid MLS's run-time is longer than Affine MLS's, but they are still edurable and welcome because they have better preformance on deformation results.

Results with Different Weight. In this paper, we use the reciprocal of the distance from v to the control point q as the weight. The factor that affects the weight is α which be can known in Equation 6. To compare the effects of different α, we try $\alpha = 0.1, 1, 5$ for the same image. The different results are shown in Figure 8.

As we can see, the deformed image is almost the same as the original image when α is small. However, the extent of deformation is too much when α is large. Therefore, we choose the most ideal situation when $\alpha = 1$. Obviously, the deformation is exactly appropriate under such condition which can be shown by the third image in 8.

5. Conclusion

In this paper, we describe a complete pipeline for facial image deformation based on facial landmark detection. The image deformation consists of two parts which are the expansion, based on bilinear interpolation and the shrinking, based on Moving Least Squares methods. During the experiments, we notice that the center points of eyes influence the expansion effect a lot. In addition, we compare the quality as well as the running time of deformed images using different MLS methods and then explore how the weight in Rigid Deformation influences the results of the shrinking process. Based on the experimental results, we find that the methods in both parts have a good performance. However, we think there is still substantial room for improvement. In future works we plan to explore more advanced methods and re-

Figure 7. Comparison of three deformation methods. From left to right: origin, Affine MLS, Similarity MLS, Rigid MLS

Figure 8. Comparison of rigid MLS with different weight. From left to right: origin, $\alpha = 0.1$, $\alpha = 1$, $\alpha = 5$

duce the impact of inaccurate detection of facial landmarks.

References

[1] T. Beier and S. Neely. Feature-based image metamorphosis. *Computer graphics*, 26(2):35–42, 1992. 2

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *European conference on computer vision*, pages 484–498. Springer, 1998. 2

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 2

[4] C. A. Glasbey and K. V. Mardia. A review of image-warping methods. *Journal of applied statistics*, 25(2):155–171, 1998. 2

[5] M. Kowalski, J. Naruniec, and T. Trzcinski. Deep alignment network: A convolutional neural network for robust face alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 88–97, 2017. 2

[6] D. Merget, M. Rock, and G. Rigoll. Robust facial landmark detection via a fully-convolutional local-global context network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 781–790, 2018. 2

[7] S. Schaefer, T. McPhail, and J. Warren. Image deformation using moving least squares. In *ACM transactions on graphics (TOG)*, volume 25, pages 533–540. ACM, 2006. 2, 3

[8] J. Shiraishi, Q. Li, D. Appelbaum, and K. Doi. Computer-aided diagnosis and artificial intelligence in clinical imaging. In *Seminars in nuclear medicine*, volume 41, pages 449–462. Elsevier, 2011. 2

[9] G. Song, Y. Liu, M. Jiang, Y. Wang, J. Yan, and B. Leng. Beyond trade-off: Accelerate fcn-based face detector with higher accuracy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7756–7764, 2018. 2

[10] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3476–3483, 2013. 2

[11] G. Wolberg. *Digital image warping*, volume 10662. IEEE computer society press Los Alamitos, CA, 1990. 2

[12] W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, and Q. Zhou. Look at boundary: A boundary-aware face alignment algorithm. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2129–2138, 2018. 2

[13] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan. Facial landmark detection with tweaked convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):3067–3074, 2018. 2

[14] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 2

[15] Z. Zhang, P. Luo, C. C. Loy, and X. Tang. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*, pages 94–108. Springer, 2014. 2

[16] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 386–391, 2013. 2