Face Aging Synthesis Application Based on Feature Fusion

Danbei Wang[1], Zhengzheng Cui[1], Huiming Ding[2], Shuqi Yan[2], Zhifeng Xie[1, 2]
[1] *Shanghai Engineering Research Center of Motion Picture Special Effects*
[2] *Department of Film and Television Engineering, Shanghai University*
Shanghai, China
zhifeng_xie@shu.edu.cn

***Abstract*—Face aging is one of research hotspots in the field of computer vision, but there is no mature system yet that can be applied to digital entertainment. This paper proposes a face aging system based on feature fusion, which can synthesize aging facial images of young people efficiently and quickly by using facial texture features of elderly people's facial images, achieve the goal of simulating the appearance after aging. The system published online by the "eleme" app for the Mother's Day, showing great potential for digital entertainment.**

Keywords: face aging; digital entertainment; feature fusion

I. INTRODUCTION

With the continuous development of mobile Internet technologies and image processing technologies, people's entertainment methods have also undergone tremendous changes. In 2015, the TV series "WuMeiNiang Legend" was broadcast. TiantianPtu topped the mainland, Taiwan, Hong Kong and other Asia-Pacific App Stores by launching the "National COS WuMeiNiang". In 2017, to celebrate the 90th anniversary of the founding of the army, TiantianPtu and the People's Daily client jointly launched the "My Military Photo", the cumulative number of PV (Page View) reached billion. PV is an important indicator to measure the quality and dissemination of Internet products. A few hundred million PV in a short time demonstrate the great attraction for people when technology and entertainment combined. It can be said that the image processing technology, especially the technology development in face processing, not only helps people to liberate from the tedious PS, but also greatly enriches the social interest. However, the current work mainly focuses on face swap and face make-up. Face aging is a blank space in digital entertainment. Based on this foundation, this paper has designed an application system for face aging that can make use of the elderly facial features. The system published by the "eleme" app on Mother's Day. The accumulated amount of PV reached 40,828 times a day after launch, demonstrating the huge potential of the system.

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61303093 and 61402278, the Innovation Program of the Science and Technology Commission of Shanghai Municipality of China under Grant No. 16511101300, and the Gaofeng Film Discipline Grant of Shanghai Municipal Education Commission of China.

II. RELATED WORK

Face aging is not only a research hotspot but also a research difficulty in the field of computer vision. Because the cause of the face aging is very complicated, it is not only determined by the gene, but also has a great relationship with the living habits and living standards. In this respect, the predecessors have already had some effective research work. The traditional research results can be mainly divided into two major categories. The first type is based on the physical model. In the early years, A.C.Berg et al. [1] considered using a 3D model to simulate different faces of a person at different ages. In addition, L.Boissieux et al. [2] proposed a wrinkle model of human face to simulate aging skin and simulate physical and visual aging characteristics. In general, this type of method is computationally complicated and time costly. The second type is based on prototype. The goal of this method is to construct an average face for both young and elderly archetypes, migrating the archetype to test texture differences between images. The most representative is a simulation method of face aging based on illumination invariance proposed by I.Kemelmacher-Shlizerman et al. [3]. This method uses optical flow and singular value decomposition to deal with face alignment and aging face reconstruction. The results are good, but not obvious.

With the development of deep learning and its strong capabilities in image processing, some work has begun to apply it to face aging, for example, Wei Wang et al. [4] proposed a recurrent face aging framework (RFA) based on RNN, through the input of pictures age from 0 to 5 years old, to obtain the prediction results of face aging of 61-80 age group step by step. In addition, in 2017, Zhifei Zhang et al. [5] proposed a network structure called CAAE (conditional adversarial auto-encoder), which can learn the diversity of human faces and predict facial image including the fullness of any age by an input facial image.

In general, traditional methods are complex and time-consuming, or the results of aging are not obvious enough and can not apply to recreational activities. Although the method of using deep learning has achieved better results in the paper, we found that most of the results obtained by the training are quite different from the input images in the actual experimental

reproduction. The effect is not satisfactory, and the processing speed is very slow, unable to apply to commercial projects.

III. Face Aging Synthesis Application

The aging of the human face has certain commonalities, such as eye sockets depression, wrinkles in the corners of the eyes, deepening of the nasolabial fold, and facial muscles drooping. Based on this conclusion, this paper proposes a face aging image synthesis application system suitable for digital entertainment. Using the aging features of the elderly face template image, it is integrated into the face of young people to quickly and efficiently synthesize the image of the face after aging.

Referring to the good compatibility and interactivity displayed on the mobile terminal before, the H5 page was displayed on the mobile terminal in the "National Cos WuMeiNiang" and "My Military Photo" activities, the visualization of the entire system was implemented in the form of Web, using the Django framework, as shown in Fig. 1.

The input of the system is collected through the front end. It is a single frontal image with a single face uploaded by the user. If no face or multiple faces are detected, the front end will return a prompt to the user. This part uses the mainstream Html, CSS, and The JavaScript language is implemented in combination and uses jQuery to simplify JavaScript. The front end compresses the input image and transmits it to the back-end via Ajax (Asynchronous JavaScript and XML) technology. The back-end performs a series of processing and then transmits it to the front end for display to the user. After that, the user can perform operations such as saving and sharing.

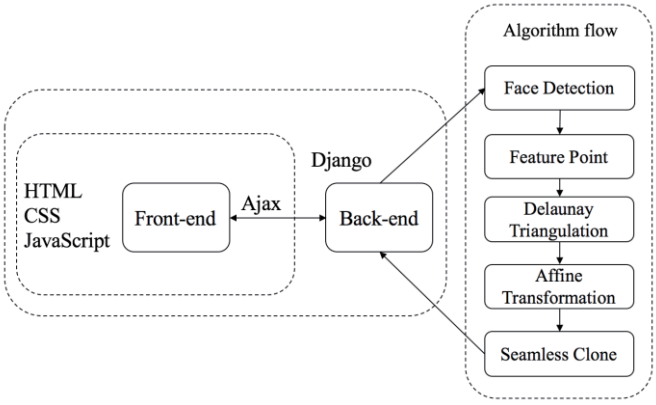

Fig. 1. Overall system architecture.

A. Page Design and Development

In order to respond the activities of Mother's Day, let young people cherish the time spent with their parents. The design theme is set to "If I am the same age as my mother", young people can use this system to feel the traces of sculpture on their faces after several decades. The page design adopts a nostalgic and concise style. As shown in Fig. 2, the yellowish main colors give people a sense of sensation. In order to evoke the user's interest, the operation design is simple, easy to promote, the process is as follows:

Step1: The user enters page 2 from the "eleme" app red envelope page, and clicks "点击生成照片" (click to synthesize a photo) on page 2 to jump to page 3;

Step2: Click plus sign on Page 3 to upload the image, select gender and click "点击生成" (click to synthesize) to jump to Page 4;

Step3: On page 4 the user can see the aged image, long press to save it.

Fig. 2. System page design.

H5 is an abbreviation of the 5th generation HTML and is applicable to mobile terminals. The two main concerns in development are adaptability and responsiveness.

- Adaptability. The user's main use scenario is the mobile terminal. The screen resolutions of mobile devices such as mobile phones and tablets are relatively diverse. The adaptive webpages are implemented using the percentage layout and the H5 "meta" tag.
- Responsiveness. The user's uploaded image is encoded using base64 encoding to reduce transmission time. The encoded image is transmitted using JSON (JavaScript Object Notation) data packets, and the efficiency of network transmission is effectively improved by using a lightweight text data exchange format. The transmission method between the front end and the back end adopts Ajax (Asynchronous JavaScript and XML) technology, and it is possible to update some pages without reloading the entire web page.

B. Algorithm Design

The algorithm flow chart of this system is shown in Fig. 3 (Note: the experimental photos are from the Internet). It is assumed that the input young people's facial image is I_i, the old

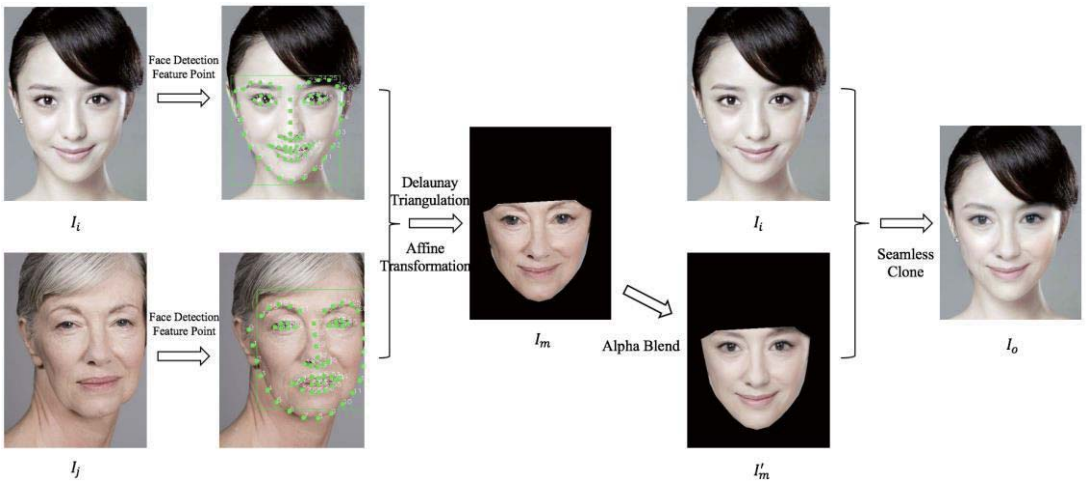

Fig. 3. Algorithm flow chart.

people's facial template image is I_j. First, the face detection and feature points are point positioning for I_i and I_j at the same time, and then triangulation is performed based on the position of the feature points. After the affine transformation by I_i, a deformed template image I_m is obtained, and I_m is used as a pixel value operation process to obtain I'_m as a foreground and I_i as a background, and a corresponding foreground mask is used for seamless fusion to obtain an output image I_o, which simulates the effect of aging.

1) Face detection and feature point positioning

First, use the open source Dlib library to implement face detection and feature point location. Dlib is an open source, cross-platform, general-purpose library written in C++ that contains many modules, such as linear algebra, machine learning, and image processing. In the face detection module, a simplified deep residual network ResNet-34[6] for image recognition is used, several convolutional layers are removed, and the number of filters per layer is reduced to half. Firstly, face detection is implemented using pre-trained face_recognition_models to form an approximate face detection frame, and then 68 feature points with a fixed order are generated. Among them, 1-16 positioning face facial features, 17-21 and 22-26 positioning eyebrows, 27-35 positioning nose, 36-41 and 42-47 positioning eyes, 48-68 positioning lips, as shown in Fig. 4.

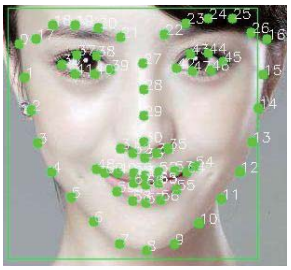

Fig. 4. Face feature points.

2) Trianulation and affine transformation

After obtaining the feature points of the two images, in order to fuse features such as wrinkles more naturally, we need to establish the correspondence between the two images through affine transformation. The Delaunay triangulation algorithm is first used to divide all the key points into several triangles, as shown in the leftmost column of Fig. 5. Suppose the triangle formed by a certain three feature points $P_k(x_k, y_k), k = 0,1,2$ in the I_j grid is S_j, and the triangle formed by three feature points $P'_k(x'_k, y'_k)$, $k = 0,1,2$ of the corresponding serial number in the I_i grid is S_i, and the Eq. 1 is obtained according to the principle of affine transformation, where M is a matrix of 2×3.

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = M \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \ i = 0,1,2 \tag{1}$$

From this, the transformation matrix M can be obtained, and the affine transformed position (a', b') of any pixel (a, b) in S_j can be seen, as shown in Eq. 2.

$$(a', b') = (M_{11}a + M_{12}b + M_{13}, M_{21}a + M_{22}b + M_{23}) \tag{2}$$

The process of the algorithm is illustrated in Fig. 5. This deformation process is repeated for all the triangles in the I_j grid, and each deformed triangle mask is put together to obtain a variant I_m of I_j relative to I_i.

3) Alpha blending and seamless clone

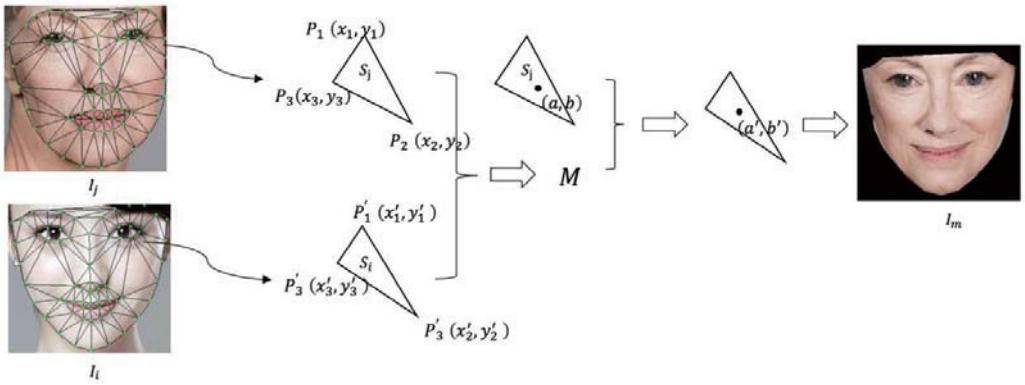

Fig. 5. Triangle affine transformation.

After obtaining I_m, The similarity of the fusion of I_m and I_i is controlled by the parameter α, which is obtained by Eq. 3 by calculating the RGB value of the pixel.

$$I_m^{'} = (1-\alpha)I_m + \alpha I_i,\ 0 \le \alpha \le 1 \tag{3}$$

For the problem of skin color fusion, a classical fusion algorithm is used, Poisson fusion, and the Poisson equation is used to solve the optimal pixel values. The gradient of the source image is preserved and the skin color transition is achieved at the same time. In Eq. 4, where Ω is the covered area of I_i after fusion, $\partial\Omega$ is the boundary of the area, ∇ is the gradient operator, and Δ is the Laplace operator. The fusion effect is shown in Fig. 6, which is also the final generation effect of the system.

$$\Delta I_o = div(\nabla I_m^{'}), \qquad s.t.\ \ I_o|_{\partial\Omega} = I_i|_{\partial\Omega} \tag{4}$$

Fig. 6. Before and after the fusion.

C. *Experiments and Application Effect*

According to Tencent Big Data, "Mobile Page User Behavior Report" in 2016, 74% of users will leave the page if they load more than 5 seconds, so the response time of the web page is very important for mobile users. Therefore, in this system, we regard time as an important performance indicator. In the process of getting the aged images from opening to the end of the webpage, the parts that are not affected by user operations are webpage loading time and image processing time. Therefore, when the backend is deployed, we take the shortening of these two parts as the main goal. The following efforts have been made for this purpose.

- Shorten the website opening time. Because the school network is a local area network, we deploy the domain name on Alibaba Cloud Server to avoid intranet penetration when the website is opened.
- Reduce the image transfer time. First of all, the H5 page needs background images and button maps, etc., to be controlled under 200KB. Second, the images uploaded by users are generally around several MB, and transmission takes too much time. Therefore, the images uploaded by users are compressed at the front end.
- Reduce the image processing time. Alibaba Cloud's image processing capability is poor. The processing time takes more than 10 seconds. We deployed servers on three Titan X and implemented diverse streaming on the first page of the front end using JavaScript. Users will randomly select a server to process images when they upload images. The maximum number of visits that can be tolerated in a minute is 360.

Because the response time of the web page is related to device performance, network speed, browser performance, image size uploaded by the user, etc. Before the system went online, we simulated the user's usage status, and selected mobile terminals of seven different models and browsers to open web pages. The recording time, as shown in Fig. 7, shows that both the page loading time and the image processing time are around 3 seconds, which is very fast and within the acceptable range of the user. In addition, for various screen sizes on the mobile side, it is very easy for the problem of typographical errors on the web page. We consider the adaptability of the H5 page as another important indicator. In this system, the web page self-adaptive effect is better, and the results are showed as Fig. 8.

Model	Huawei mate10	Huawei Honor v9	iPhone7	iPhone6s	Xiaomi note3	Vivo x6	iPad mini2
Web browser	System Browser	Quark	Quark	Ghrome	QQ Browser	System Browser	Safari
Network	4G	4G	4G	4G	4G	4G	Wi-Fi
The initial response time	2.66s	2.01s	3.33s	3.55s	2.93s	3.24s	2.20s
Image generation time	3.44s	3.28s	3.04s	3.05s	2.58s	2.81s	2.62s

Fig. 7. System response time for different devices.

Fig. 8. Adaptability of web pages with different screen sizes.

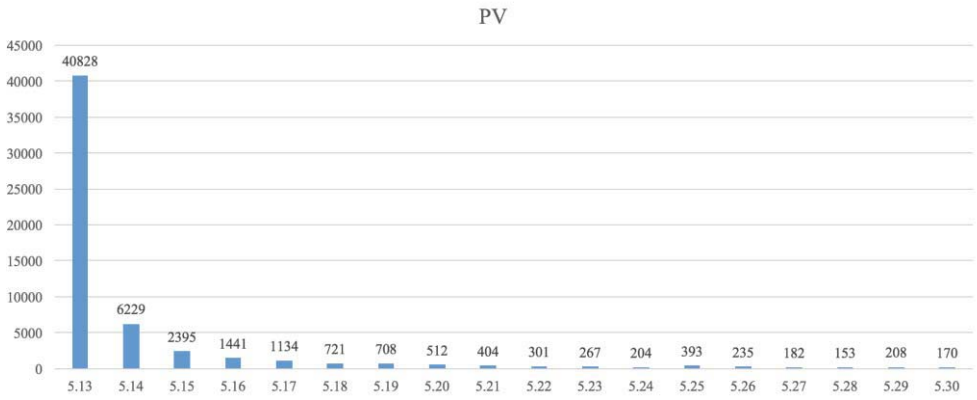

Fig. 9. PV statistics.

After the system went online, according to background statistics, the cumulative number of PV (Page View) to the mother's day on May 13 was as high as 40,828. On the day after Mother's Day, hundreds of people still visited every day, from May 13 to May 30 for a total of 18 days. The PV data statistics is shown in Fig. 9.

IV. Conclusion and Future Work

In this paper, we proposed a system for digital entertainment that can quickly and efficiently synthesize young people's aged face using the commonality of the elderly face. The system is launched on the Mother's Day as an "eleme" app's red envelope event, and is not promoted by any social platform. As of June 7th, the cumulative number of PV (Page View) reached 63,322 times. This data proves that the system has strong practicality and feasibility, and shows good application ability and commercial value. In the future, we will strive to improve the effect of aging and explore other interesting face processing application for digital entertainment.

References

[1] Berg A C, Justo S C. Aging of Orbicularis Muscle in Virtual Human Faces[C]// International Conference on Information Visualization, 2003. IV 2003. Proceedings. IEEE, 2003:164-168.

[2] Boissieux L, Kiss G, Thalmann N M, et al. Simulation of Skin Aging and Wrinkles with Cosmetics Insight[J]. Eurographics, 2000:15-27.

[3] Kemelmacher-Shlizerman I, Suwajanakorn S, Seitz S M. Illumination-Aware Age Progression[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:3334-3341.

[4] Wang W, Cui Z, Yan Y, et al. Recurrent Face Aging[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:2378-2386.

[5] Zhang Z, Song Y, Qi H. Age Progression/Regression by Conditional Adversarial Autoencoder[J]. 2017:4352-4360.

[6] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.

[7] Murray J E, Halberstadt J, Ruffman T. The face of aging: sensitivity to facial feature relations changes with age[J]. Psychology & Aging, 2010, 25(4):846.

[8] Xiaobin Y I, Chen Y. Frontal face synthesizing based on Poisson image fusion under piecewise affine warp[J]. Computer Engineering & Applications, 2016.

[9] Zhou H, Wong K W, Lam K M. Feature-aging for age-invariant face recognition[C]// Asia-Pacific Signal and Information Processing Association Summit and Conference. IEEE, 2015:1161-1165.

[10] Ren L, Liu S, Sun Y, et al. Time Traveler: A Real-time Face Aging System[C]// ACM, 2017:1245-1246.

[11] Esme B, Sankur B. Effects of Aging over Facial Feature Analysis and Face Recognition[J]. Bogazici Un.electronics Eng.dept, 2010.

[12] Li X, Xu Y, Lv Q, et al. Affine-Transformation Parameters Regression for Face Alignment[J]. IEEE Signal Processing Letters, 2015, 23(1):55-59.

[13] rez, Patrick, Gangnet, Michel, Blake, Andrew. Poisson image editing[J]. Acm Transactions on Graphics, 2003, 22(3):313-318.

[14] Song Q, Robinson J. A feature space for face image processing[C]// International Conference on Pattern Recognition, 2000. Proceedings. IEEE, 2000:97-100 vol.2.