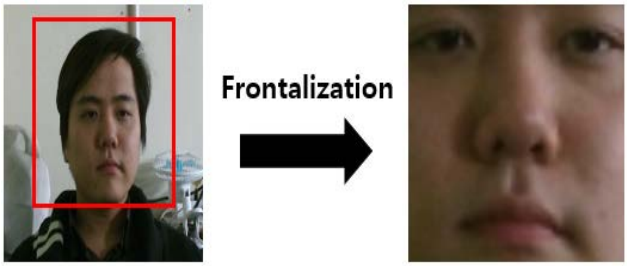
Face Frontatlization Method using Depth Information

Xingjie Zhang[1], Joongseok Song[2] and Jong-Il Park[3]
[1, 2, 3] Department of Computer Software, Hanyang University
1, 2, 3 222, Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea
E-mail: [1]sgzhang@mr.hanyang.ac.kr, [2]jssong@mr.hanyang.ac.kr, [3]jipark@hanyang.ac.kr

Abstract: Recently, the face recognition system have achieved excellent performance. However, the performance of face recognition is being affected by various factors such as illumination, expressions, occlusions and pose variations. And the pose variation is the most important of these. In this paper, we propose a face frontalization technique which used face landmark detection and depth information. This method could be applied to the preprocessing stage of the face recognition system or expression recognition system. Proposed method could transform the non-front face to the front face as well. In experiments, we verified that a set of faces with various pose could be transformed to the front face by using this method which is proposed in this paper.

Keywords—**Face frontalization, Face alignment, Face recognition, Face identification.**



(a) **(b)**

Figure 1. Face frontalization. (a) is the non-front face obtained from camera. (b) is the front face obtained from target face by face frontalization process.

1. Introduction

In order to improve the performance of face recognition or expression recognition system, we need to get the front faces from dataset, which are used to train face or expression classifier and test this recognition system.

The process of obtaining front face is called as face frontalization. We could gain a set of similar-front faces with the fixed size box-region by applying face frontalization. Figure 1. Shown the concept of face frontalization. This kinds of preprocessing will be helpful in face recognition system of expression system.

Some previous face frontalization techniques deal with this problem by just using similarity transformation with two eyes location. In this case, the two eyes location could obtain through an eye detection or face landmark localization technique.

This kinds of methods have high computational speed. However, the face, which obtain from similarity transformation, is not enough similar with front face due to the limitation of similarity transforms. Especially, if the pose of face occur the 3d rotation.

In order to solve this problem, the 3D face frontalization method is proposed [1]. This method frontalize human face by an exquisite generic 3D shape model.

This kinds of method have excellent performance. However, it need to consume much computational time due to the exquisite 3D shape model rendering process.

In this paper, we propose a face frontalization method using depth information and face land mark detection. Proposed method is relatively more exquisite and could not consume much computational time.

For purpose of this, the first thing we need to do is face landmark localization. The active shape model is adopted for this.

In this method which proposed in this paper, we only use four landmark points which include two eye-centers and two mouth-corners.

Next we compute the 3D location of such features and find four corners which are used as define the face region.

Finally, the new face region transformed to a regular face image using homography transformation.

The structure of this paper as follows: In Sect. 2, we introduce the whole process of the proposed algorithm. Sect. 3, we show the experimental results. Finally, in Sect. 4, we express conclusions of this paper and further research.

2. Face Frontalization Method

Firstly, in order to transform the non-front face to the front face, we get color and depth images by using a depth camera. This kinds of camera must have the ability to simultaneously capture color and depth image like Inter realsense F200 or Microsoft kinect. These device could capture color and depth information at real time.

Next, we detection the target face from input image and extract face landmark points from this region. Then we could obtain a set of face landmark points. Among these landmark

Corresponding author: Jong-Il Park

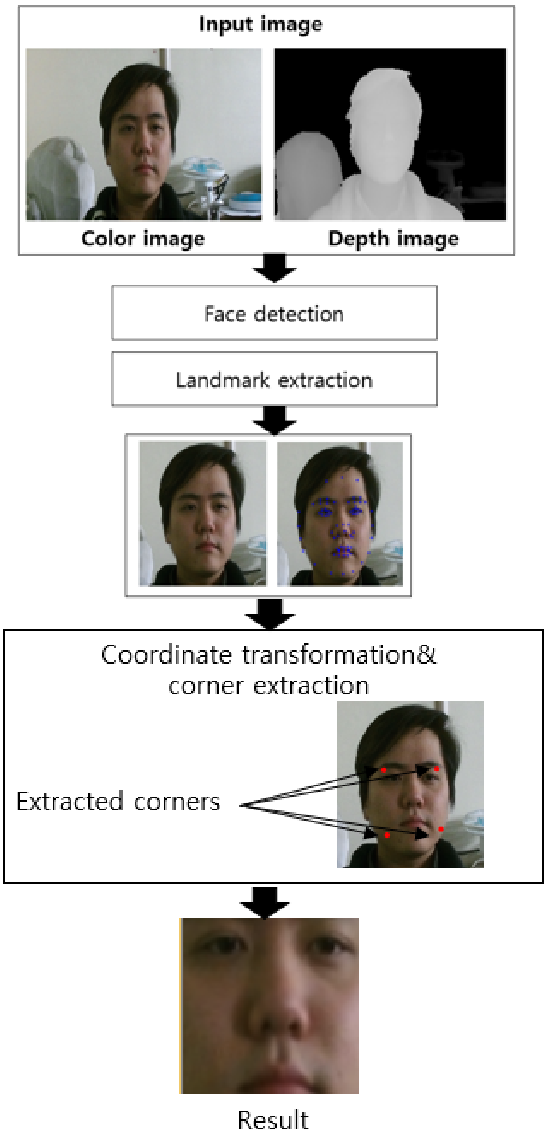


Figure 2. The process of face frontalization

points, we just select four points which include two eye-centers and two mouth-corners.

After that, we transform the coordinate of these four points from 2D to 3D. We can obtain a 3D plane using this 3D points. The plane will located in the front of the target face in the 3D space.

Then we projection these points to the 3D plane and find four corners using these projected points and its center.

Finally, these four corners re-project to the image plane and warping the face region to a regularized region by using homography transformation.

The whole process of our face frontalization method which proposed in this paper is shown in Figure 2.

2. 1 Face Detection

The first step of this algorithm which proposed in this paper is face detection, We adopt the Viola and Jones face detection algorithm proposed by Paul Viola and Michael J. Jones and extended by Rainer Lienhart and Jochen Maydt [2] [3].

This face detection system employs the Adaboost algorithm [4] used in training the face detector and using the Harr features to detect human face. This face detection system adopt the Integral image to compute the response of Haar filter, so the execution time of finding face is approximate to real time.

The result of face detection is shown in Figure 3. The red box indicate the detected face region.

2. 2 Landmark Localization

The research of Facial landmark extraction using active shape model have been receiving attention gradually in the past several decades. Many of improve version, like AAM (active appearance model), CLM (constrained local model), SDM (supervised descent method), ESR (Explicit shape regression), have got a remarkable achievement [5] [6] [7] [8].

After the face detection, we need to extract the face landmarks from the detected face region. In this paper we lacalize face landmark using STASM, which is the one of the proposed version of ASM [9], algorithm proposed by Milborrow, Stephen, and Fred Nicolls [10]. This method is based on ASM (active shape model) of proposed by Tim Cootes and his colleagues.

The results of face landmark detection is shown in Figure 4. This algorithm detect the face landmarks by using 68 points model.

2. 3 Coordinate Transformation

In this Section, we briefly introduce a simple coordinate transformation method which could convert 2D coordinate to 3D coordinate or convert 3D coordinate to 2D coordinate. As shown in Eq. 1, if we know the inner parameters of camera and corresponding depth value, then we can transform 3D coordinate to 2D and vice versa.

$$u = \frac{f_x X + C_x Z}{Z} \qquad v = \frac{f_y Y + C_y Z}{Z} \qquad \textbf{(1)}$$

Here, f_x, f_y are two focal-length of x axis and y axis and C_x, C_y is the principle point of camera obtained from camera calibration.

X, Y , Z are the coordinate of 3D points which have derived. u and v is the location of a point in 2D image coordinate.

2. 4 Face Frontalization

In this paper, we assume the four features - the centers of two eyes and two corners of mouth – are placed on the same plane. We select these four landmarks to estimate collect face region.

After obtain a set of facial landmark points using STASM, then we compute the 3D coordinate of two eyes center and two corners of mouth by using Eq. 1.

Next, we figure out a virtual plane using these points. We knew that any three non-collinear points can define a plane.

After this, we project these four points to the virtual plane be virtual points and compute the center of these virtual points.

After then, we using this virtual points and center to find four vertex of quadrilateral face region using Eq. 2. These vertexes indicate the four corners of face cropping region in 3D space.

Finally, we can get a homography matrix using Eq. 4

$$P_V^{(i)} = P_c + s\, h(i)\, norm(V_{(i)})\, k_{(i)} \qquad \textbf{(2)}$$

$$P_c = \frac{1}{4}\sum_{i}^{n} k_{(i)} \qquad \textbf{(3)}$$

Here, $P_V^{(i)}$ represent the vertex coordinate of cropping region.

P_c is the central point of two eyes center and two corners of mouth, that is received by using Eq. 3.

$h(i)$ is coefficient signs and equal to +1 when the index i is 1 or 3, and equal to -1 when i is 2 or 4.

$V_{(i)}$ is a vector from the central point P_c to the point of eye center. If i is 1 or 3, $V_{(i)}$ represent the vector from P_c to left eye, else i is 2 or 4, then $V_{(i)}$ represent the vector from P_c to right eye.

$k_{(i)}$ is unit vector of $V_{(i)}$ and s is scaling coefficient.

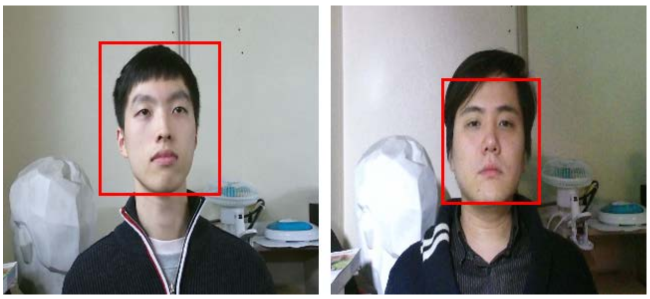
n is the number of points which have been choosen.

$$P_N^{(i)} = HP_V^{(i)} \qquad \textbf{(4)}$$

Here, $P_N^{(i)}$ is regular corner points which defined as a fixed size rectangle.

$P_V^{(i)}$ is virtual corner points which derived from Eq. 2.

H is the homography matrix. We convert non-front face to frontface using this matris.



(a) (b)

Figure 3. The result of face detection

(a) (b)

Figure 4. The result of face landmark detection

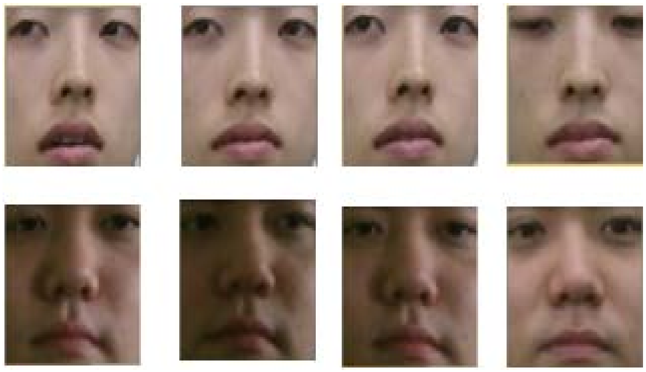
3. Experimental result

Our system was developed using Visual C++ and tested in 2.13 GHz processor. Moreover, the OpenCV library was also used.

The images which used in test captured by Microsoft Kinect v2. The coefficient s was defined as 1.5 in the test. In order to matching two coordinate of color and depth images, a rectification method have been applied to coincide depth image to color image.

The experiment result is shown in Figure. 5. The images (a) are the frontalization results which obtained by using simple similarity transform, and the images (b) represent the result through our method which proposed in this paper.

(a) Face frontlization using affine transformation



(b) Face frontlization using proposed method

Figure 3. Experiment result

4. Conclusions

In this paper, we proposed an efficient face frontalization method using depth information and facial landmarks. In experiment, the face images are more naturally frontalized by proposed method than previous work which using affine transformation.However, some problems such as the performance of propose method affected by the distance between camera and human face and lead to some image distort is not solved yet. Therefore, we plan solve these problems in the future studies.

Acknowledgment

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2016-H8601-16-1005)supervised by the IITP(Institute for Information & communications Technology Promotion).

References

[1] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[2] Viola, Paul, and Michael J. Jones. "Robust real-time face detection. "International journal of computer vision 57.2 (2004): 137-154.

[3] Lienhart, Rainer, and Jochen Maydt. "An extended set of haar-like features for rapid object detection." Image Processing. 2002. Proceedings. 2002 International Conference on. Vol. 1. IEEE, 2002.

[4] Freund, Yoav, Robert Schapire, and N. Abe. "A short introduction to boosting."Journal-Japanese Society For Artificial Intelligence 14.771-780 (1999): 1612.

[5] Cootes, Timothy F., Gareth J. Edwards, and Christopher J. Taylor. "Active appearance models." IEEE Transactions on Pattern Analysis & Machine Intelligence 6 (2001): 681-685.

[6] Cristinacce, David, and Tim Cootes. "Automatic feature localisation with constrained local models." Pattern Recognition 41.10 (2008): 3054-3067.

[7] Xiong, Xuehan, and Fernando Torre. "Supervised descent method and its applications to face alignment." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.

[8] Cao, Xudong, et al. "Face alignment by explicit shape regression." International Journal of Computer Vision 107.2 (2014): 177-190.

[9] Cootes, Timothy F., et al. "Active shape models-their training and application."Computer vision and image understanding 61.1 (1995): 38-59.

[10] Milborrow, Stephen, and Fred Nicolls. "Locating facial features with an extended active shape model." Computer Vision- ECCV 2008. Springer Berlin Heidelberg, 2008. 504-513.