Face Recognition based on CNN 2D-3D Reconstruction using Shape and Texture Vectors Combined

Edy Winarno[1], Imam Husni Al Amin[1], Sri Hartati[2], Prajanto Wahyu Adi[3]
[1] Faculty of Information Technology, Universitas Stikubank, Indonesia
[2] Department of Computer Sciences and Electronics, Universitas Gadjah Mada, Indonesia
[3] Department of Informatics Engineering, Universitas Dian Nuswantoro, Indonesia

Article Info

ABSTRACT

This study proposes a face recognition model using a combination of shape and texture vectors that are used to produce new face images on 2D-3D reconstruction images. The reconstruction process to produce 3D face images is carried out using the convolutional neural network (CNN) method on 2D face images. Merging shapes and textures vector is used to produce correlation points on new face images that have similarities to the initial image used. Principal Component Analysis (PCA) is used as a feature extraction method, and for the classification method, we use the Mahalanobis method. The results of the tests can produce a better recognition rate compared to face recognition testing using 2D images.

Corresponding Author:

Edy Winarno,
Faculty of Information Technology, Universitas Stikubank,
Jl. Tri Lomba Juang No.1 Semarang, Indonesia
Email: edywin@edu.unisbank.ac.id

1. INTRODUCTION

The human face recognition system is research in the field of computer vision that continues to develop today. Several studies on the development of facial recognition have been carried out by several previous researchers such as [1,2,3,4,5]. Some of these face recognition studies include many using facial recognition processing using 2-dimensional (2D) imagery in the process of image data acquisition, preprocessing, feature extraction and classification. Some facial recognition studies using 2D imagery have advantages in the face recognition computation process but they still have weaknesses, especially the problem of illumination variations and pose variations when taking facial images, resulting in less optimal facial recognition.

There are several studies of facial recognition systems that have used 3-dimensional (3D) imagery in the process of human face recognition as done by [6] and [7]. Some face recognition studies using 3D imagery require image acquisition using 3D scanners which are quite expensive. The process of facial image recognition using 3D imagery also has a weakness in the process of computing face recognition which is more complicated than the face recognition model using 2D imagery generated from ordinary cameras.

To improve the accuracy of maximum facial recognition and faster facial recognition computing time, several studies combining 2D and 3D techniques have been carried out. [8] conducted face recognition research to improve face recognition accuracy based on features-based 2D images on taking a single vision camera. The technique used is to reconstruct 2D to 3D images to produce a frontal single face image with neutral facial expressions and normal illuminance. This face recognition technology can produce facial recognition accuracy in severalconditions such as variations in poses, light illumination and facial expressions that are better than the general 2D face recognition system.

Face recognition research using 2D front-facing facial images to reconstruct the next 3D face image was carried out by [9]. This study uses a 3D generic elastic model (3D GEM) to reconstruct the 3D model of each human face obtained from a database of 2D face images taken using a single vision technique. The main parts of the face are extracted using texture, and the results of the reconstruction are based on the value of the depth on the 2D land-mark face. The results of this study canproduce a face recognition technologythat canhandle changes in facial expressions.

From several previous face recognition studies, there are still many methods and algorithms that have not been studied, specifically the use of reconstruction algorithms from 2D images to 3D forms that are used as databases in face recognition. This 2D to 3D image reconstruction method is expected to contribute robustly in face detection and recognition so that it has high accuracy and fast facial recognition computing. To solve some of these problems, this research conducted a study of facial recognition using an approach to develop 2D to 3D image reconstruction models using the Convolutional Neural Network (CNN). The CNN method is used to produce a 3D face image from 2D face images. The next step is to combine the shape and texture vector to produce a correlation of points on the new face image that has similarities to the initial image used. The results of the process of combining vector shapes and textures from 3D face images are then processed using a database for the face recognition process.

2. RESEARCH METHOD

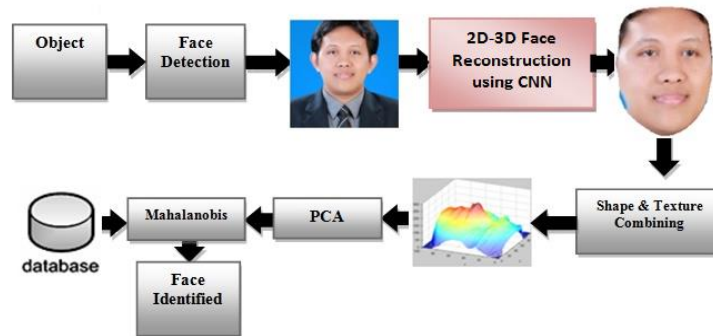The proposed method that used in this research as shown in Figure 1.

Figure 1. Proposed Method

This face recognition system uses Viola-Jones face detection to determine which faces are detected. The next process is to reconstruct 2D face images into 3D using CNN [10]. The CNN method can produce 3D face images which are then processed using a shape and texture combination to produce a correlation of points on a new face image that has similarities to the initial image used.

In the feature extraction process, the Principal Component Analysis method is used to reduce the dimensions of a facial image [5]. The classification process for comparing test face images with face images on a database used the Mahalanobis Distance method [5]. This method can produce a high recognition rate in the proposed facial recognition system

2.1. 2D-3D Reconstruction using CNN

Similar to [11, 12, 13, 14], a 3D face can be reconstructed from a 2D face using a formula represented in Eq.(1):

$$S = \acute{S} + U_d . \alpha_d + U_e . \alpha_e \qquad (1)$$

S is the 3D face output, $\acute{S}$ is the mean of facial's shape, U_d is the training of principal components on a 3D face, α_d shows the vector of parameter identification, U_e is the training of principal components on the offset 3D facial scans and α_e is the vector of expression parameter. The method is used to reconstruct a 2D facial image becoming a reconstructed 3D face. The model of 2D-3D face reconstruction using CNN is seen in Figure 2.
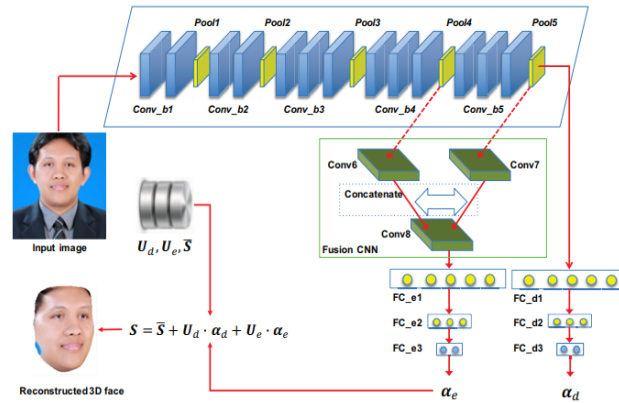
Figure 2. The architecture of the model for 2D-3D face reconstruction using CNN

2.2. Shape and Texture Combining

The result of 3D face reconstruction using CNN was then processed by shape and texture combined. It is used to recognize faces across a wide range of pose and illumination conditions. To solve the poses and illuminations problems, we can use the parameters separately between intrinsic and extrinsic models. The information of all poses of faces can be stored by estimating a 3D shape of faces.

We estimate the intrinsic and extrinsic parameters to recognize a face by comparing the nearest neighbouring face in the database. To determine a morphable model of 3D faces, we represent a face using 2 vectors shown in Eq.(2) and Eq.(3).

$$S_0 = (x_1, y_1, z_1, \ldots \ldots x_n, y_n, z_n)^T \qquad (2)$$

$$T_0 = (R_1, G_1, B_1, \ldots \ldots R_n, G_n, B_n)^T \qquad (3)$$

Eq.(2) and Eq.(3) explained that the pixel at (x_k, y_k, z_k) have colors (R_k, G_k, B_k). S_0 is the vector of shape and T_0 is the vector of texture. The two vectors are combined using a formula seen in Eq.(4):

$$\mathbf{I}(h, \phi) = (r(h, \phi), R(h, \phi), G(h, \phi), B(h, \phi))^T \qquad (4)$$

The combination of the shape vector and texture vector that produces a new face is shown as a formula in Eq.(5) and Eq.(6).

$$S = \sum_i a_i S_i \qquad (5)$$

$$T = \sum_i b_i T_i \qquad (6)$$

A 3D face resulted from a vector combination needs to correlate every point with the equivalent point on a reference face. A modified algorithm is used to determine the point to point correspondence. For parameters with (h, ϕ), a diagram that maps each point between the reference face and the new face is used to result in S and T vectors. The correlation of every point between a new 3D face and a reference face is shown in Figure 3.
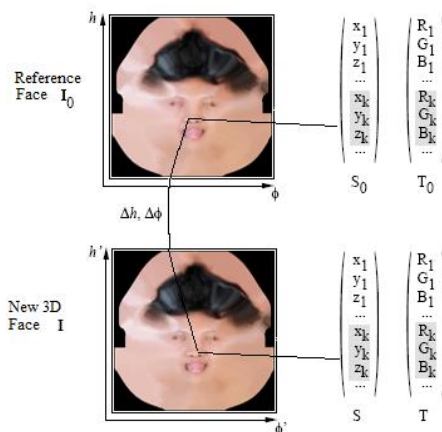

Figure 3. Correlation of every point between new a 3D face and a reference face

2.3. Principal Component Analysis (PCA)

In this research, we use PCA [5] to reduce the dimensionality on the set of shape and texture of vectors (S_i and T_i). Different types of faces can be produced if linear combinations of vector shapes and textures are formed separately for the eyes, nose, mouth, etc. PCA is a feature extraction model that serves to reduce image features before the classification process is carried out.

Using PCA, we can find c_x as the matrix characteristics and then calculate the covariance matrix $\sum_x$. The eigenvalues are provided by solving the characteristic equation of ($\lambda_i I - \sum_x$) = 0. The transformation characteristics are calculated using $c_y = c_x W^T$. The new characteristics will be processed in classification using Mahalanobis Distance.

2.4. Mahalanobis Distance

In the matching step, we use the Mahalanobis distance to improve recognition rate and computation time. Mahalanobis distance is used to determine the similarity among the features in face recognition.

Mahalanobis Distance (D^2) is determined using x as vector data, m as a vector of mean values of independent variables, C^{-1} as the inverse of the covariance matrix and T as transpose vector [5]. The formula of Mahalanobis distance is shown in Eq. (7) [5].

$$D^2 = (x - m)^T C^{-1} (x - m) \qquad (7)$$

The matching process to identify a face by comparing a morphable face model with a probe face is seen in Figure 4.
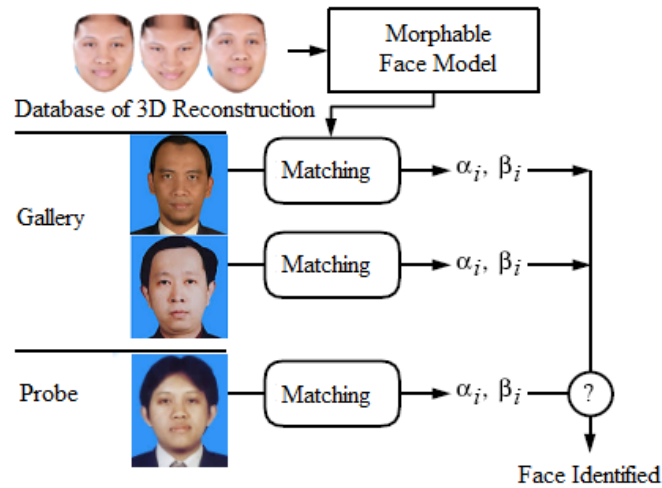

Figure 4. Matching process for face identification

3.   RESULTS AND DISCUSSION

This face recognition system uses the CNN 2D-3D Reconstruction method using Shape and Texture Vectors Combined. This system produces a face recognition system by producing a new face database. Examples of face images used and the representations of the results of face image correlation are shown in Figure 5.

Figure 5. Face image sample and correlation of every point

The results of the reconstruction of facial images that have been processed using CNN are seen in Figure 6.

Figure 6. Reconstructed 3D face image

The results of 3D image reconstruction using CNN are then processed using feature extraction by the PCA method to reduce image dimensions. The PCA reduction image is then processed using the Mahalanobis Distance method as its classification method. Several tests were carried out to produce an accurate facial recognition. Testing is done by comparing the use of the PCA method with 3Wavelet-PCA. A comparison between the Euclidean Distance method andthe Mahalanobis Distance method was used in the classification process. This test shows that the accuracy of facial recognition using PCA-Mahalanobis is better than other methods. The test results of the proposed model are seen in Table 1.

Table 1. Recognition Rate

| Object | Recognition Rate (%) | | | |
| | Euclidean | | Mahalanobis | |
	PCA	3Wavelet-PCA	PCA	3Wavelet-PCA
10	90.00	80.00	100.00	90.00
20	95.00	90.00	100.00	95.00
30	96.67	93.33	96.67	93.33

Subsequent tests were carried out by comparing the accuracy of the proposed face recognition from different angle/face rotation. Tests using face position comparisons were performed on left rotation, right rotation, top rotation, bottom rotation, and each rotation angle was 0 °, 22.5 °, 45 °. Tests carried out using 40 face sample data with each face 12 different rotations. The results of tests with various rotations of facial positions can be seen in Table 2.

Table 2. Recognition Rate based on face rotation

Angle (°)	Recognition rate (%)			
	Left rotation	Right Rotation	Up Rotation	Up Rotation
0	100	100	100	100
22.5	100	100	100	100
45	100	100	100	100

Test results show that at an angle of 0° (frontal face) to 45° can produce 100% accuracy of face recognition. These results indicate that the use of 2D-3D reconstruction using CNN does not affect the accuracy of facial recognition even though it is performed on various types of face.

4.   CONCLUSION

From several studies on face recognition systems, there are still many methods and algorithms that have not been examined specifically, especially the use of reconstruction of algorithms from 2D images to 3D shapes that are used as databases in face recognition. The face recognition system proposed uses an approach to develop 2D to 3D image reconstruction models using the Convolutional Neural Network (CNN). The CNN method is used to produce a 3D face image from 2D face images. The merging vector shapes and textures are used to produce correlation points on new face images that have similarities to the initial image used. The result of the process of combining vector shapes and textures from 3D face images is then used as a database for the face recognition process. The feature extraction method uses the PCA method, the classification method uses the Mahalanobis distance method. Both methods can work well on the proposed facial recognition system. The proposed method can produce a face recognition that has a high accuracy of up to 100%

REFERENCES

[1] Gupta, R., and Saxena, A. K., "Survey of Advanced Face Detection Techniques in Image Processing", *International Journal of Computer science and management Research*, *1*(2), 156-164, 2012.

[2] Choi, J., Dumortier, Y., Choi, S. I., Ahmad, M. B., and Medioni, G., "Real-time 3-D face tracking and modeling from a webcam", In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on* (pp. 33-40), IEEE, 2012.

[3] Xu, Y., Li, X., Yang, J., and Zhang, D., "Integrate the original face image and its mirror image for face recognition", *Neurocomputing*, 131, 191-199, 2014.

[4] Qiu, M., Liu, W., and Cao, J., "Integrating the original face images and "symmetrical faces" to perform face recognition", *Optik-International Journal for Light and Electron Optics*, 125(11), 2665-2670, 2014.

[5] Winarno, E., Harjoko, A., Arymurthy, A., M., and Winarko, E., "Improved Real-time Face Recognition Based On Three Level Wavelet Decomposition – Principal Component Analysis", *Journal of Computer Science* 10 (5): 844-851, 2014, doi:10.3844/jcssp.2014.844.851, 2014.

[6] Passalis, G., Perakis, P., Theoharis, T., and Kakadiaris, I. A., "Using facial symmetry to handle pose variations in real-world 3D face recognition", *Pattern Analysis and Machine Intelligence, IEEE Transactions* on, *33*(10), 1938-1951, 2011.

[7] Zafeiriou, S., Atkinson, G. A., Hansen, M. F., Smith, W. A., Argyriou, V., Petrou, M., and Smith, L. N., "Face recognition and verification using photometric stereo: The photoface database and a comprehensive evaluation", *Information Forensics and Security, IEEE Transactions* on, 8(1), 121-134, 2013.

[8] Fooprateepsiri, R., and Kurutach, W., "A general framework for face reconstruction using single still image based on 2D-to-3D transformation kernel", *Forensic science international*, *236*, 117-126, 2014.

[9] Moeini, A., Moeini, H., and Faez, K., "Expression-Invariant Face Recognition via 3D Face Reconstruction Using Gabor Filter Bank from a 2D Single Image", In *Pattern Recognition (ICPR), 2014 22nd International Conference* on (pp. 4708-4713). IEEE, 2014.

[10] Eigen, D., Puhrsch, C., and Fergus, R., "Depth map prediction from a single image using a multi-scale deep network", In *Advances in neural information processing systems* (pp. 2366-2374), 2014.

[11] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3D solution", In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 146 – 155, Las Vegas, NV, June 26-July 1 2016.

[12] A. Jourabloo and X. Liu, "Large-pose face alignment via CNN-based dense 3D model fitting", In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 4188 – 4196, Las Vegas, NV, June 26-July 1 2016.

[13] E. Richardson, M. Sela, and R. Kimmel, "3D face reconstruction by learning from synthetic data", In *Proc. International Conference on 3D Vision*, pages 460–469, California, USA, October 25-28 2016.