Qatawneh S, Ipson S, Qahwaji R and Ugail H (2008): "3D FACE RECOGNITION BASED ON MACHINE LEARNING", *Eighth IASTED International Conference on Visualization, Imaging and Image Processing (VIIP 2008), pp. 362-366*, Palma de Mallorca, Spain.

3D FACE RECOGNITION BASED ON MACHINE LEARNING

S. Qatawneh, S. Ipson, R. Qahwaji, H. Ugail
Department of Electronic Imaging and Media Communications
University of Bradford, Richmond Road, Bradford BD7 1DP, England, U.K.
{Smsalqat, S.S.Ipson, R.S.R.Qahwaji, H.Ugail}@bradford.ac.uk

ABSTRACT

3D facial data has a great potential for overcoming the problems of illumination and pose variation in face recognition. In this paper, we present a 3D facial system based on the machine learning. We used landmarks for feature extraction and Cascade Correlation neural network to make the final decision. Experiments are presented using 3D face images from the Face Recognition Grand Challenge database version 2.0. For CCNN using Jack-knife evaluation, an accuracy of 100% has been achieved for 7 faces with different expression, with 100% for both of specificity and sensitivity.

KEY WORDS

3D Face Recognition, CCNN, .ABS images, Feature extraction.

1. Introduction

The human face is one of the most important objects found in images. Detecting the location of faces and then extracting the facial features in an unconstrained image is a challenging process. It is very difficult to locate the positions of faces in an image accurately. There are several variables that affect the recognition performance, such as gender, facial hair, pose, illumination and facial expressions. Furthermore, the human face is a 3D object, which might be affected by perspective distortion and uneven illumination. Current 2D face recognition systems encounter difficulties in handling facial variations due to head poses and lighting condition [1], which introduce large amount of intra-class variations. Range image based 3D face recognition has been demonstrated to be effective in enhancing the face recognition accuracy [2][3][4]. Since each range image provides only a single view point of the face, instead of the full 3D view [5].

Many methods have been proposed for 3D face recognition over the last few years [6]. A prior study on curvature has been proposed for face recognition, which described slight features [7]; Chua et al. [8] treat face recognition as a 3D non-rigid surface matching problem and split the human face into rigid and non-rigid regions. Beumier et al. [9] develop a 3D acquisition prototype based on structure light and build a 3D face database. Bronstein et al. [10] propose a method capable of extracting the intrinsic geometric features of facial surfaces using geometric invariants, and the use of bending invariant canonical representation makes it robust to facial expressions and transformations typical of non-rigid objects. Lu et al. [11] construct many 3D models as registered templates, and then they match 2.5D images to these models using iterative closest point (ICP). Chang et al. [12] use principal component analysis (PCA) on both 2D intensity images and 3D depth images, and fuse 2D and 3D results to obtain the final performance. In their survey on state of the art in 3D and multi-modal face recognition, Mian et al. [13] used the Hotelling transform to correct the pose for 3D images and their corresponding 2D images in order to build up a rejection classifier for the recognition process. Bowyer et al. [14] conclude that 3D face recognition has the potential to overcome the limitations of its 2D counterpart.

Generally, it's not easy to compare or reproduce results of other research as many results are not reported using the same data. If there is a common database, such as the Face Recognition Grand Challenge (FRGC) database, different pre-processing operations may be used for different methods, which make direct comparisons difficult. However, the FRGC data set is the most challenging data set available supporting research on 3D face recognition [15]. Several recognition approaches are commonly investigated in 3D face recognition, for instance correlation, closest vector, PCA, SVM, EHMM and ICP approaches. In this work we carry out 3D face recognition with range data from FRGC Ver.2.0 data set using the CCNN, which has not been done before. We aim to introduce a robust 3D face recognition system by using the machine learning technique which is widely recognised as an appropriate and efficient validation method, as explained later.

In the following section, the dataset and machine learning algorithm (CCNN) used in this work are presented. In the third section we described our experimental work. Concluding remarks and suggestions for possible future work are given in the last section.

2. Methodology and System Design

This section presents a brief description of the data used, the feature extraction techniques employed and how these features are implemented for use with the Machine learning algorithm

2.1 Dataset

The work described in this paper is conducted using FRGC version 2.0 3D dataset [15]. FRGC data was collected at the University of Notre Dame and is part of an ongoing multi modal biometric data collection. The 3D images were taken under controlled illumination conditions appropriate for the Vivid 900/910 sensor. In the FRGC, the 3D image set includes 9,500 face images for both range and texture channels [15]. As a starting point we used a sample of 56 3-D near frontal images of faces for 7 people with 8 images each.

2.2 Cascade Correlation Neural Network

An important aspect of automatic Face Recognition is the pattern recognition problem, which is very hard to solve due to the nonlinearity of human faces. It can be treated as a matching problem, to be performed in a high-dimensional space. The neural network (NN) can be considered as a good solution to the face recognition problem. NNs are commonly used in many other pattern recognition problems and can be adapted to handle the people authentication task. NNs are becoming powerful tools for solving a variety of multi-class classification problem and are more appropriate than the SVM learning method. Advantages of NNs over linear methods are that they can reduce misclassifications among the neighbourhood classes and they can carry out nonlinear mappings from the input to the output nodes. Also, NNs can be used to correct for instrument drift, and they are robust in relation to noisy data [16].

The training of back-propagation neural networks is considered to be a slow process because of the step-size and moving target problems [17]. To overcome these problems cascade neural networks were developed. These are “self organizing” networks [17] with topologies which are not fixed. The supervised training begins with a minimal network topology and new hidden nodes are incrementally added to create a multi-layer construction. The new hidden nodes are added to make the most of the correlation between the new node’s output and the remaining error signal that the system is being adjusted to eliminate. The weights of a new hidden node is fixed and not changed later, hence making it a permanent feature detector in the network. This feature detector can then be used to generate outputs or to create other more complex feature detectors [16].

The Jack-knife technique [18] was employed to evaluate the performances of the CCNN. The Jack-knife technique is usually implemented to provide a correct statistical evaluation of the performance of a classifier when applied to a limited number of samples divided into two sets: a training set and testing set. In practice, a random number generator is used to select the samples used for training and the samples kept for testing. The classification error varies with the training and testing sample sets and, for a finite number of samples; an error-counting procedure is used to estimate the performance of the classifier [18].

In this work, 80% of the available samples were randomly selected and used for training while the remaining 20% were used for testing. Therefore 45 samples were used for training and 11 for testing. The results were then analyzed to assess the performance. The performance criteria used in our work are accuracy (the fraction of all correct predictions), sensitivity (the fraction of positive cases correctly classified) and specificity (the fraction of negative cases correctly classified

3. Experimental work

To investigate the impact of different registration techniques for correspondence estimation on the quality of the 3D model for face recognition, we have constructed a 3D statistical face model using 56 datasets. A typical face consists of about 20,000 points.

3.1 Interpret the 3D Data and Extract the Facial Area

The FRGC database range files have the extension .ABS, and contain data in ASCII representation. The .ABS file includes a header, which indicates the number of rows and columns of the range image. The third line shows how the data is stored in the file (flag X Y Z). This indicates four groups of data; the first block in the file is a flag image, where valid pixels have value of 1. The second block, of size rows × columns, entries contains X coordinate of each pixel, the third block contains all the Y values and the fourth block contains all the Z values.

The FRGC database also provided a texture (.PPM) file for each 3D image. We can use the texture by mapping each pixel from the 2D image with each pixel from the range image. For instance, the pixel (200, 200) from the texture file may be applied to the (x, y, z) point that is stored at pixel (200,200) in the range image. For the most part we can simply discard the (x, y) data as it falls fairly close to the image plane, this just gives a 2.5D range image of the subject remaining in the Z-data. At this stage of our experiments we just use the .ABS images without their associated texture. To extract the facial area from the background of a given a facial scan, the invalid X, Y, Z points can be distinguished using the mask data.

3.2 Standardized the Face Area and Holes filling

The nose tip is a distinctive point of the human face; we manually detect the nose tip in the first step in order to crop out the required facial area from the 3D face for further processing. We need this process to be sure we are dealing with the region of interest (i.e. human face) in the image. For a frontal facial scan, the nose tip usually has the largest Z value. Then we crop a face region from the raw 3D data to construct a 3D face image which centred at nose section, the size of each image is 480*640. Once the face is cropped, outlier points causing spikes in the 3D face are removed, leaving holes, which are filled using cubic interpolation [19]. Figure 1 shows an example 3D face after removing the spikes and filling the holes.

(a) (b) (c)

Figure 2 (a) extracted the facial area (b) the face after inserting the Z data and removing the spikes, (c) the face after cropping and hole filling.

3.3 Feature Extraction

Landmarks are one of the most important methods used to define features that are manually placed on the 3D face. In order to ensure correct correspondence the landmarks should be sited on anatomically distinct points of the face. However, parts of the face such as the cheeks are difficult to landmark because there are no uniquely distinguishable anatomical points common across all faces. It is important to choose landmarks that contain the local feature information such as the size of nose, as well as the overall sizes of the face for example the eyebrow locations. Previous work on 3D face modelling for classification has shown that there is not much difference resulting between the use of 11 and 59 landmarks [20].

In our experiments 10 landmarks are sufficient to capture the shape and size variations of the face appropriately. Table 1 shows the features that are used, and Figure 2 shows an example of a face that was manually landmarked [21].

Anatomical points landmarked	
Points	**Landmark Description**
Eyes	**Both the inner and outer corners of the eyelids (4 landmarks).**
Nasion	**The intersection of the frontal and two nasal bones of the human skull where there is a clearly depressed area directly between the eyes above the bridge of the nose (1landmark).**
Nose tip	**The most protruding part of the nose (1landmark).**
Subnasal	**The middle point at the base of the nose (1landmark).**
Nose extremes	**The outer corners of nose (2 landmarks).**
Gnathion	**The lowest and most protruding point on the chin (1 landmark).**

Table 1. The 10 manually selected landmarks chosen because of their anatomical distinctiveness.

Two kinds of features are computed over faces and compared in order to determine which of the two approaches gives best recognition; the first represents the distances between the chosen landmarks and the second uses ratios of distances. We varied the numbers of features, starting with 4 features of those shown in Figure 2. These features are the distance between outer corners of eyes (AB), the distance between the inner corners of the eyes (CD), the Nasion distance between nose tip and a point between the eyes (FE), and the distance between nose extremes (GH). We increases the numbers of features to 5, by adding the Gnathiona distance between the lowest point in the chin and the Subnasal middle point at the base of the nose (IJ).

The second approach represents the data extracted from the same dataset but the features are the ratios between symmetry line of face (FE) with the outer corners of eyes (AB), the ratios between symmetry line of face (FE) with inner corners of eyes (CD) and the ratios between symmetry line of face (FE) with the nose extremes (GH). Also we added a new feature corresponding to the ratio between symmetry line and the line connects between the lowest point in the chin (Gnathion) and the middle point at the base of the nose (Subnasal) (I J).

3.4 Results

We carried out experiments on 56 images, using a CCNN with a single hidden layer and with numbers of hidden nodes ranging from 1 to 10. For each hidden node case, the accuracy, specificity and sensitivity generated, were the average of 10 iterations carried out using the Jack-knife technique (80% randomly used for training and the rest for testing). Hence 100 learning and testing experiments were carried out for each case.

Our results are presented in the form of ROC (Receiver Operating Characteristic) curves for computing the rate of recognition. Figure 3 shows ROC curves for all the biometric features we compute. According to the experiments, the CCNN with 5 input nodes (distance features) and 2 hidden nodes gives the best results for face recognition, where TPR =1 and FPR=0.

4. Conclusion and future works

In this paper we presented a 3D face recognition method based to the CCNN technique. Firstly, we extracted the distinctive features by manually land marking; two sets of features were used, the distances between the landmark points and the ratios of these distances. Then the CCNN made the final decision. With the 5 distance approach we achieved an accuracy of 100 %.

Our future goal is to automated the system, expand the data set to use the whole 3D data provided in FRGC Ver.2.0, and combine more features such as spherical harmonic and Fourier series representations.

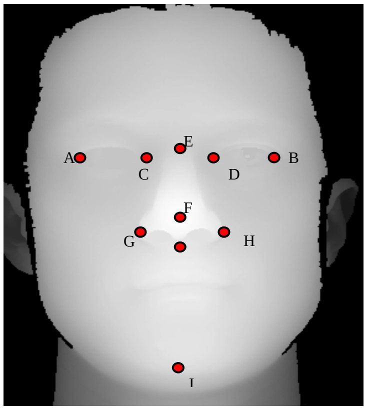

Figure 2. The 10 manually selected features chosen because of their anatomical distinctiveness

(a)
(b)
(c)
(d)

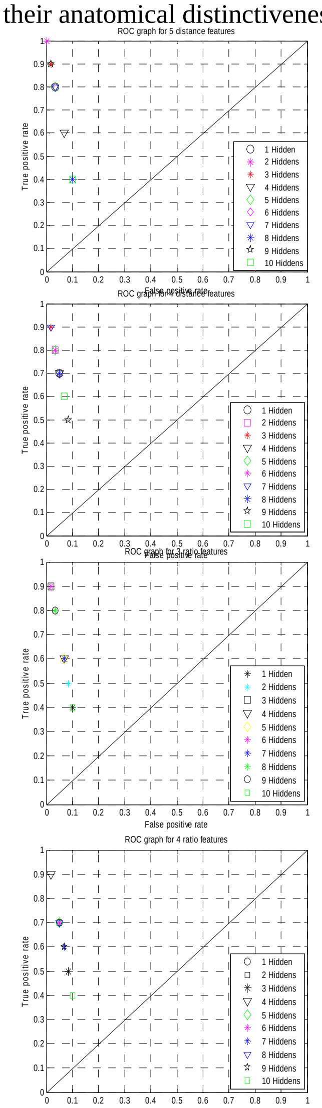

Figure 3. ROC curves for different biometric features, as explained at subsection 3.3 (a) 5 distance features, (b) 4 distance features, (c) 4 Ratio features, and (d) 3 Ratio features.

References

1. *Face Recognition Vendor Tes*t <http://www.frvt.org/>.

2. C. BenAbdelkader and P. Griffin. Comparing and combining depth and texture cues for face recognition. *Image and Vision Computing*, 23:339–352, 2005.

3. K. I. Chang, K. W. Bowyer, and P. J. Flynn. Multi-modal 2D and 3D biometrics for face recognition. In *Proc. IEEE Workshop on Analysis and Modeling of Faces and Gestures*, pages 187–194, France, Oct. 2003.

4. F. Tsalakanidou, S. Malassiotis, and M. Strintzis. Face localization and authentication using color and depth images. *IEEE Transactions on Image Processing*, 14(2):152–168, 2005.

5. X. Lu, A. K. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE Trans.PAMI*, 28(1):31–43, 2006.

6. K. W. Bowyer, K. I. Chang, and P. J. Flynn. A survey of approaches to three dimensional face recognition. *International Conference on Pattern Recognition*,**1**:358-361, 2004.

7. G. Gordon. Face recognition based on depth maps and surface curvature. *SPIE*, pages 234–274, 1991.

8. C. Chua, F. Han, and Y. Ho. 3d human face recognition using point signature. *IEEE Inetrnational Conference on Automatic Face and Gesture Recognition*, 233–238, 2000.

9. C. Beumier and M. Acheroy. Face verification from 3d and grey level cues. *Pattern Recognition Letters*, 22:1321–1329, 2001.

10. A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three dimensional face recognition. *IJCV*, 2005.

11. X. Lu, A. K. Jain, and D. Colbry. Matching 2.5d face scans to 3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):31–36, 2006.

12. K. I. Chang, K.W. Bowyer, and P. J. Flynn. An evaluation of multi-modal 2d+3d face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619 624, 2005.

13. Ajmal S. Mian, M. Bennamoun and R. Owens, Automatic 3D Face Detection, Normalization and Recognition, *Third International Symposium on 3D Data Processing, Visualization and Transmission*, 2006.

14. K. Bowyer, K. Chang, and P. Flynn. A Survey of Approaches and Challenges in 3D and Multi-Modal 3D+2D Face Recognition. In *CVIU*, 101:1–15, 2006.

15. P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, 2005, I:947–954

16. Y. S. Abu-Mostafa and D. Psaltis, "Optical Neural Computers", Scientific American, 256:88-95, 1987.

17. Fahlmann, S.E., Lebiere, C.:1989, In Touretzky, D.S.(ed.) Advances in Neural Information Processing System 2(NIPS-2), Morgan Kaufmann, Denver, p.524.

18. Fukunaga, K., Introduction to Statistical Pattern Recognition, (Academic, New York, 1990).

19.O'Neill C.,Cubic Spline Interpolation,MAE5093, 2002

20. Hutton, T. (2004). Dense Surface Models of the Human Face. PhD thesis, University College London.

21. Rueckert, D., Frangi, A. F., and Schnabel, J. A. (2003). Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025.