2D and 3D Face Recognition Using Convolutional Neural Network

Huiying Hu*, Syed Afaq Ali Shah, Mohammed Bennamoun, Michael Molton

School of Computer Science and Software Engineering, The University of Western Australia

35 Stirling Highway, Crawley 6009, Western Australia, Australia

*Corresponding author: Huiying Hu. E-mail: 21742778@student.uwa.edu.au

Abstract—Face recognition remains a challenge today as recognition performance is strongly affected by variability such as illumination, expressions and poses. In this work we apply Convolutional Neural Networks (CNNs) on the challenging task of both 2D and 3D face recognition. We constructed two CNN models, namely CNN-1 (two convolutional layers) and CNN-2 (one convolutional layer) for testing on 2D and 3D dataset. A comprehensive parametric study of two CNN models on face recognition is represented in which different combinations of activation function, learning rate and filter size are investigated. We find that CNN-2 has a better accuracy performance on both 2D and 3D face recognition. Our experimental results show that an accuracy of 85.15% was accomplished using CNN-2 on depth images with FRGCv2.0 dataset (4950 images with 557 objectives). An accuracy of 95% was achieved using CNN-2 on 2D raw image with the AT&T dataset (400 images with 40 objectives). The results indicate that the proposed CNN model is capable to handle complex information from facial images in different dimensions. These results provide valuable insights into further application of CNN on 3D face recognition.

Key words: Face Recognition; Convolutional Neural Networks; Depth Image

1. INTRODUCTION

Biometric systems are automated methods which are used to identify and verify humans through physiological or behavioral characteristics. Face recognition is one of the most important components of biometric system which has tremendous applications in security, control and entertainment applications [1, 2].

Most face recognition systems are focused on 2D face recognition and numerous recognition approaches and techniques have been proposed. These techniques could be mainly classified into appearance-based methods [3, 4], feature-based matching methods [5, 6] and artificial intelligence methods [7]. Specifically, appearance-based methods have two categorizations including holistic and hybrid approaches. In the holistic approach, the input is the information of whole face. Eigenfaces method[3] based on principal component analysis (PCA) and Fisherfaces method [4] based on linear discriminant analysis (LDA) were the two most successful methods. The hybrid approaches employ both holistic and local features. A few methods fusing various features such as fusion of Gabor and LBP [8], have attracted much attention. Compared to holistic approach, feature-based methods are more compatible with inaccuracy in face localization due to pose, expression and lighting. Feature-based methods use a number of pre-processed local facial features to identify faces. Local facial features include the facial distances and angles among eye corners, month extrema, nostrils and chip top. One of the most widely used examples in feature-based methods is the Elastic Bunch Graph Matching (EBGM) system [6]. However, feature-based methods are still not reliable enough when considering the impact of pose, expressions and illumination [9]. Artificial-intelligence based methods such as support vector machine, neural network/deep network, fuzzy logic and genetic algorithm have become extremely important in computer vision. Phillips [10] proposed a support vector machine (SVM) for 2D face recognition. SVM performs a non-linear classification, mapping inputs into high-dimensional feature space. SVM has shown to achieve outstanding performance compared to other methods including PCA.

However, the practical application of 2D face recognition techniques was greatly hindered by the disadvantages that derive from pose, expression and illumination variations [11, 12]. To overcome these limitations, 3D approach are proposed. The main advantage of the 3D based approaches is that all the information about the face geometry is processed. Two main representations for modelling 3D faces are 3D and 2.5D images. 3D images are a global representations of the whole facial geometry. Point cloud based approach is one widely used approach for 3D images. A 2.5D depth image consists of a two-dimensional representation of a 3D point set, but different viewpoints are scanned to ensure all possible information are included in 2.5D images. In 3D objective identification, Iterative Closest Point [13] (ICP) is proposed. Parameters are iteratively searched to align one object model to the other one.

978-1-5090-1134-6/17/$31.00 ©2017 IEEE

The limitation of ICP is that it cannot process non-rigid alignment. The differential geometry approach was also proposed for 3D face recognition which is not affected by translation and rotation. Gordon et al. [14] extracted 3D geometry face features by using Gaussian and mean curvatures. Tanaka et al. [15] calculated free-form curved surfaces using spherical correlation as a 3D shape recognition method for 3D face recognition.

For 2.5D depth images, classical 2D face recognition techniques such as Eigenfaces and Fisherfaces can be employed to identify depth images. Pan et al. [16] combined Eigenfaces and PCA for classification. An identification rate of 95% was achieved on FRGCv1.0 (943 images of 276 subjects). Xu et al. [17] employed LDA and AdaBoost for classification on Gabor image features. A verification accuracy of 95.3% on FRGCv2.0 was reached. Lv et al. [18] chose LBP as feature extractor and achieved an recognition rate of 97.8% using sparse representation classifier (SRC) on FRGCv2.0. In this work, 4007 images of 466 subjects were investigated. Every subject has around 22 images with pose variation. These works has provided useful insights into 3D face recognition based on depth images. However, further research efforts on 3D face recognition based on more complicated 2.5D dataset using more advanced methods are still needed to improve recognition performance in practical applications.

Deep learning based methods [19] are a class of machines that can learn levels of representation and abstraction by constructing high-level features. Convolutional neural network [20] are a type of deep learning models where local pooling and filters are applied to the raw images, leading to complex features. This method has demonstrated to be successful in many fields such as character recognition [21] and EEG signal recognition [22]. Lawrence et al. [7] constructed a hybrid neural network system which is consisted of sampling, self-organizing neural network (SOM) and convolutional neural network (CNN). An accuracy of 96.2% was achieved on ORL dataset (400 2D images of 40 subjects). Yi Sun et al. [23] recently constructed two CNNs rebuilt from stacked convolution and inception layers. Joint face identification-verification supervisory signals were added during training. An accuracy of 96.0% face recognition for 2D LFW dataset was achieved. Jun-Cheng Chen et al. [24] built a deep CNN to train real-world unconstrained 2D LFW faces. An accuracy of 97.45% was achieved. Current research efforts have been mainly focused on 2D face recognition using deep learning methods. There is a lack of systematic study in the applications of deep learning methods on 3D face recognition.

In this paper, we try to fill this research gap. We constructed two CNN models and trained the models on 2D image database (AT&T, 400 images of 40 subjects) and 2.5D depth image database (FRGCv2.0, 4950 images of 557 subjects) to extract high-level features to improve accuracy performance. The performance of two CNN models with raw image input and LBP processed features both in 2D and 3D face recognition was then investigated.

2. Methodology

To measure the improvement brought by the different structures of CNNs, all our CNNs layer configurations are designed using the same principles, inspired by LeCun et.al [25]. We first describe the generic layout of our CNN-1 and CNN-2 configurations and then detail the specific configuration used in the experiments. The details of CNNs training and evaluation are given next.

2.1 Layout of CNN models

During training, the input to our CNN-1 and CNN-2 is a fixed-size 90*90 2D gray-scale image. Then the 2D gray scale image is passed through a stack of convolutional layers, where we use filters, which are small receptive fields to capture the notion of left, right, up, down and center. In the configurations of CNN-1, we utilize six 11*11 convolution filters for the first convolutional layer (C1) and sixteen 5*5 filters for the second convolutional layer (C3), while six 11*11 filters are applied for the only convolutional layer (C1) in CNN-2. Spatial pooling is carried out by the one mean-pooling layer, which follows all the convolutional layers. Mean-pooling is performed over a 2*2 pixel window with stride 2. A stack of convolutional layers is followed by two fully-connected layers. In both CNN-1 and CNN-2 for 2D face recognition, the first fully-connected layer has 480 channels, while the second performs 40-way AT&T classification and thus contains 40 channels.

For 3D face classification using CNN-1 and CNN-2, the input depth image is fixed as 66*66 in CNN-1 while is 30*30 in CNN-2. In CNN-1, there are six 7*7 filters for first convolutional layer (C1) while sixteen 5*5 filters for second convolutional layer (C3). In CNN-2, there are three different sizes of six filters we test: 3*3, 5*5, and 7*7 convolved with the convolutional layer (C1). One mean-pooling is set up following all the convolutional layers. The last two layers both in CNN-1 and CNN-2 are two fully-

connected layers, one including 800 channels and the other including 557 channels (557 objects need to be identified in FRGCv2.0 dataset).

2.2 Experimental configuration

The CNN-1 and CNN-2 configurations for 2D grayscale image and depth image are shown in Fig. 1 and Fig. 2, respectively. All configurations follow the generic design presented in Section 2.1 and differ in the depth: CNN-1 has two convolutional layers and two pooing layers, while CNN-2 only has one convolutional layer and one pooling layer.

2.3 Training

The training procedure generally follows LeCun et.al [25]. In 2D gray-scale face recognition, there are two kinds of input (shown in Fig. 3): raw image and LBP image. Particularly, LBP image we use is a texture feature which is the result of combining 2D raw image with a Local Binary Patter filter using (8, 1) neighborhood [26]. There are 400 images in AT&T dataset, and we randomly choose 80% for training and the remaining 20% for testing. All input images are resampled into 90*90 and then the training is carried out by optimizing the multinomial logistic regression objective using mini-batch gradient descent based on back-propagation.

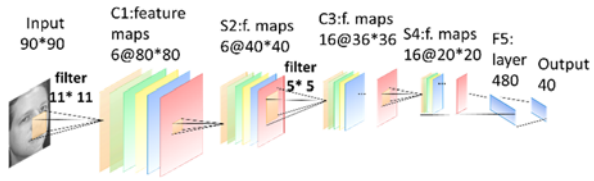

Fig. 1 Schematic configuration of CNN-1 model

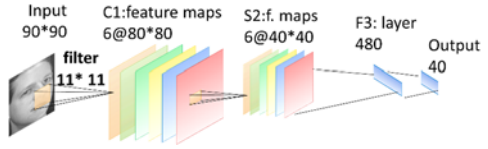

Fig. 2 Schematic configuration of CNN-2 model

Fig.3 2D gray scale image (top row) and LBP image (bottom row) in AT&T dataset

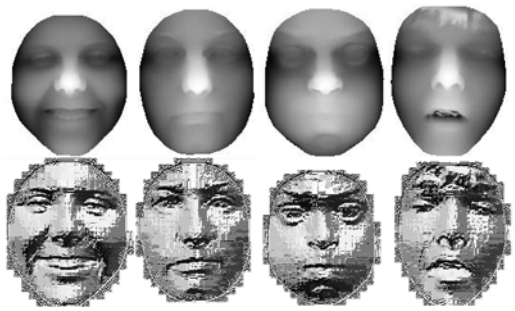

Fig.4 2.5D depth images (top row) and LBP depth image (bottom row) in FRGCv2.0 dataset

For 3D face recognition, there are two kinds of input (shown in Fig. 4): raw depth image and LBP depth image (processed by (8, 1) neighborhood LBP filter). In CNN-1, a rescaled image of size 66*66 was used whereas a rescaled image of size 30*30 was utilized in CNN-2. In FRGCv2.0 dataset, there are 4950 depth images in total, and we randomly used 80% for training and the remaining for testing.

3. Results and Discussion

In this section, we present the 2D and 3D face recognition results using the CNN-1 and CNN-2 models. For 2D face classification, we choose AT&T dataset, which includes faces of 40 classes and each class has 10 images. This dataset is randomly split into two sets: 80% for training (320 images), 20% for testing (80 images). For 3D face classification, we choose FRGCv2.0 depth image dataset, which includes 4950 faces of 557 classes. It also randomly split into 80% for training (3960 images) and 20% for testing (990 images). The classification performance is evaluated based on recognition accuracy.

3.1 CNN on 2D face recognition

3.1.1 CNN-1 Experiments

Table 1. CNN-1 Performance with different activation functions (learning rate = 0.01)

C1 activation function	relu	relu	tanh	tanh
C3 activation function	relu	tanh	tanh	relu
Raw image	2.5%	82.5%	**88.75%**	0%
LBP image	0%	**85%**	72.5%	0%

We begin evaluating the performance of CNN-1 model on 2D face recognition. As activation

function in convolution layer is an important parameter in neural network [27], we first investigated the effect of different combinations of activation functions in CNN-1 layers on the classification accuracy. As shown in Table 1, we noted that a best classification accuracy of 88.75% on raw image was achieved with C1 and C3 activation functions both being tanh. It was found that a best classification accuracy of 85% based on LBP image was reached with C1 and C3 activation functions being relu and tanh, respectively.

We further examined the effect of learning rate on the classification accuracy using CNN-1 model. Based on the previous results, in the following experiments tanh was chosen as C1 and C3 activation functions for raw image whereas relu and tanh was chosen as C1 and C3 activation functions for LBP image, respectively. As shown in Fig.5, we observed that the classification accuracy using raw image increased from 63.75% to 88.75% as the learning rate was increased from 0.006 to 0.010 but decreased to zero when learning rate was over 0.010. The best classification accuracy of 88.75% using raw image was reached when learning rate was 0.010. As shown in Fig.6, a similar trend using LBP image was also observed. The classification accuracy increased from 85.00% to 91.25% as the learning rate was increased from 0.01 to 0.03 but decreased to zero when learning rate was over 0.04. The best classification accuracy in CNN-1 model of 91.25% using LBP image was achieved when learning rate was 0.03. It can be seen that CNN-1 has a better classification performance using LBP processed image compared to raw image.

3.1.2 CNN-2 Experiments

We further evaluated the performance of CNN-2 model on 2D face recognition. The effect of activation functions on the classification accuracy using raw image and LBP image was also first examined. As shown in Table 2, a better classification accuracy of 86.25% using raw image was reached with tanh being C1 activation function. It also showed that a better classification of 92.5% using LBP image was achieved with relu being C1 activation function.

Next we assessed the effect of learning rate on the classification accuracy using CNN-2 model. Tanh was chosen as activation function when raw image was used whereas relu was chosen as activation function when LBP filter was employed. As shown in Fig.7, the classification accuracy (using raw image) was increased from 86.25% to 95.00% as learning rate increased from 0.01 to 0.05 and decreased to 92.50% when learning rate was 0.06. In Fig.8, the classification accuracy (using LBP image) had the best performance (92.50%) when learning rate was 0.01. The classification accuracy dropped to 83.75% when learning rate was 0.02 but slowly increased to 90% with learning rate increasing to 0.03. The highest classification of 95% was reached in CNN-2 model when raw image was used and learning rate was 0.05. It can also be seen that CNN-2 had a better and more stable performance using raw image than using LBP image.

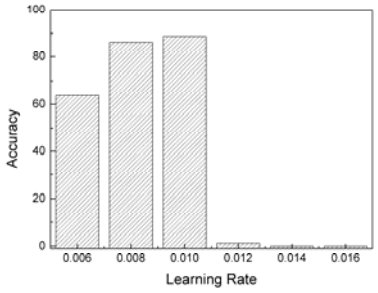

Fig. 5 CNN-1 performance with different learning rate (LR) using raw image

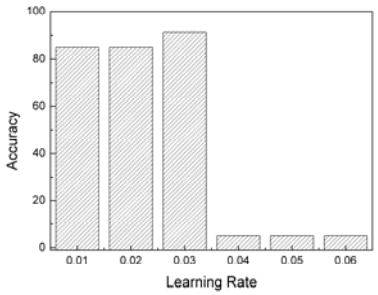

Fig. 6 CNN-1 performance with different learning rate using LBP image

Table 2. CNN-2 performance with different activation function (learning rate = 0.01)

C1 activation function	relu	tanh
Raw image	0.05%	**86.25%**
LBP image	**92.5%**	72.5%

3.1.3 Comparison and evaluation of the results

A detailed comparison of the results for the two CNN models proposed in this paper and other relevant work including Fisherfaces [28], Linear Regression Classification (LRC) [29] and Kernel Eigenfaces [28] is summarized in Table 3. These experiments were all performed on AT&T dataset. For CNN models, the algorithm achieves a comparable recognition accuracy of 95%. This

result outperformed the Fisherfaces approach, Linear Regression Classification approach and Kernel Eigenfaces approach under evaluation protocol 1(EP1[29]: where first five images are designated as the training set and the last five as probes). This shows that CNN approach is fairly comparable to the Fisherfaces, Kernel Eigenfaces and LRC on 2D face recognition.

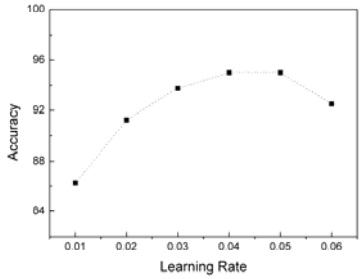

Fig. 7 CNN-2 performance with different learning rates (LR) using raw image

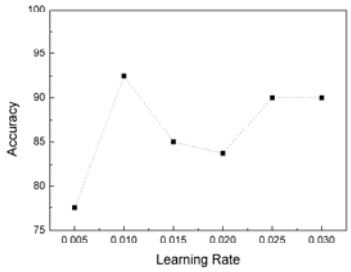

Fig. 8 CNN-2 performance with different learning rates (LR) using LBP image

Table 3 Comparison of results on 2D face recognition

Approach	Accuracy
CNN-1 + Raw image	88.75%
CNN-1 + LBP image	91.25%
CNN-2 + Raw image	**95.00%**
CNN-2 + LBP image	92.50%
Fisherfaces (EP1) [28]	94.50%
LRC (EP1) [29]	93.50%
Kernel Eigenfaces (EP1) [28]	94.00%

Table 4 CNN-2 performance with different learning rates (LR) and filter sizes using LBP image and raw image on FRGCv2.0 dataset

Accuracy		**3*3 filter**	**5*5 filter**	**7*7 filter**
image	***LR***			
LBP image	0.034	76.67%	74.14%	80.91%
	0.036	74.75%	**80.81%**	60.61%
	0.038	80.20%	79.80%	**81.92%**
	0.040	**80.40%**	79.09%	79.39%
	0.042	75.96	80.40%	78.89%
Raw image	0.034	80.71%	84.04%	**84.75%**
	0.036	82.63%	84.55%	82.02%
	0.038	81.72%	**85.15%**	83.03%
	0.040	81.62%	82.42%	83.13%
	0.042	**82.73%**	81.31%	83.45%

3.2 CNN on 3D face recognition

Finally we tested on CNN-1 and CNN-2 models for 3D face recognition. These experiments were performed on FRGCv2.0 dataset. In CNN-1 model, we found that under various combinations of parameters, recognition accuracy is no more than 60%. For CNN-2 model, we examined the effect of feature input, learning rate and filter size on the 3D recognition accuracy as shown in Table 4. With raw image input and LBP image input, a range of learning rate from 0.034 to 0.042 was studied while three filter sizes (3*3 filter, 5*5 filter and 7*7 filter) were investigated for CNN-2 model. For LBP image, we found that for tested combinations of learning rate and filter size, accuracy rate fluctuated around 80%. The best accuracy obtained in LBP image input is 81.92% when learning rate is 0.038 and filter size is 7*7. With raw image as input, when filter size was 3*3, classification accuracy fluctuated around 81% and 82% and reached the highest performance at 82.73% when learning rate was 0.042. With a 5*5 size filter, classification accuracy first increased from 84.04% to 85.15% with learning rate increasing from 0.034 to 0.038, but then dropped to 81.31% when learning rate increased to 0.042. Using a 7*7 filter size, the best classification accuracy was 84.75% with a learning rate of 0.034. While learning rate was over 0.034, the accuracy dropped to around 83%. It can be seen that CNN-2 model has a stable performance on 3D face dataset. A state-of-art accuracy of 85.15% was achieved with a 5*5 filter and a learning rate of 0.038 using raw image input. This is consistent with the results on 2D face

recognition that CNN-2 model has a better performance using raw image input.

4. Conclusions

In this work, we investigated the performance of convolutional neural networks on 2D and 3D face recognition. Two kinds of CNN models (namely CNN-1 and CNN-2) were constructed. Two kinds of image inputs namely raw image and LBP image were used. Effects of important parameters including activation function, learning rate and filter size on the classification accuracy were studied. Activation functions including tanh and RELU, learning rate ranging from 0.005 to 0.06 and three filter sizes (3*3, 5*5 and 7*7) were tested for CNN models. Our experimental results show that CNN-1 has a best classification accuracy of 91.25% on 2D face dataset while CNN-2 has a best classification accuracy of 95%. On 3D face dataset (FGRCv2.0), an accuracy of 85.15% was achieved using CNN-2 model. The results indicate that the proposed CNN model is capable to handle complex information from facial images in different dimensions. Our results provide valuable insights into the application of convolutional networks on 2D and 3D face recognition.

Acknowledgements

The authors acknowledge the support from The University of Western Australia, the Australian Research Council (ARC) linkage grant LP130100138 and also the AshuCNN toolbox provided by Ashutosh Kumar Upadhyay.

Reference

[1] W. Hariri, H. Tabia, N. Farah, A. Benouareth, D. Declercq, 3D face recognition using covariance based descriptors, *Pattern Recognition Letters*, 78 (**2016**) 1-7.
[2] S.A.A. Shah, M. Bennamoun, F. Boussaid, Iterative deep learning for image set based face and object recognition, *Neurocomputing*, 174 (**2016**) 866-874.
[3] M. Turk, A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, 3 (**2014**) 71-86.
[4] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 19 (**1997**) 711-720.
[5] I.J. Cox, J. Ghosn, P.N. Yianilos, Feature-based face recognition using mixture-distance, in: Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on, 1996, pp. 209-216.
[6] L. Wiskott, J.M. Fellous, N. Kuiger, V.D.M. Christoph, Face Recognition by Elastic Bunch Graph Matching, in: International Conference on Computer Analysis of Images and Patterns, 1997, pp. 456-463.
[7] S. Lawrence, C.L. Giles, T. Ah Chung, A.D. Back, Face recognition: a convolutional neural-network approach, *IEEE Transactions on Neural Networks*, 8 (**1997**) 98-113.
[8] S. Xie, S. Shan, X. Chen, J. Chen, Fusing local patterns of Gabor magnitude and phase for face recognition, *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 19 (**2010**) 1349-1361.
[9] I.A. Kakadiaris, G. Passalis, G. Toderici, E. Efraty, P. Perakis, D. Chu, S. Shah, T. Theoharis, Face Recognition Using 3D Images, in: S.Z. Li, A.K. Jain (Eds.) Handbook of Face Recognition, Springer London, London, 2011, pp. 429-459.
[10] P.J. Phillips, Support Vector Machines Applied to Face Recognition, *Advances in Neural Information Processing Systems*, 11 (**1998**) 803--809.
[11] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: A survey, *Pattern Recognition Letters*, 28 (**2007**) 1885-1906.
[12] T. David, H.B. Hons, Face Recognition: Two-Dimensional and Three-Dimensional Techniques, *University of York*, (**2005**).
[13] D. Chetverikov, D. Stepanov, P. Krsek, Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm, *Image & Vision Computing*, 23 (**2005**) 299-309.
[14] G.G. Gordon, Face recognition based on depth maps and surface curvature, *Proceedings of SPIE - The International Society for Optical Engineering*, 1570 (**1991**) 234-247.
[15] H.T. Tanaka, M. Ikeda, Curvature-Based Face Surface Recognition Using Spherical Correlation - Principal Directions for Curved Object Recognition, in: . International Conference on Face & Gesture Recognition, 1998, pp. 372.
[16] P. Gang, H. Shi, W. Zhaohui, W. Yueming, 3D Face Recognition using Mapped Depth Images, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, 2005, pp. 175-175.
[17] C. Xu, S. Li, T. Tan, L. Quan, Automatic 3D face recognition from depth and intensity Gabor features, *Pattern Recognition*, 42 (**2009**) 1895-1905.
[18] S. Lv, F. Da, X. Deng, A 3D face recognition method using region-based extended local binary pattern, in: 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 3635-3639.
[19] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, DeepFace: Closing the Gap to Human-Level Performance in Face Verification, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708.
[20] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional Neural Networks for No-Reference Image Quality Assessment, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733-1740.
[21] B. Graham, Sparse arrays of signatures for online character recognition, *Computer Science*, 135 (**2013**) 89-90.
[22] S. Ji, W. Xu, M. Yang, K. Yu, 3D Convolutional Neural Networks for Human Action Recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 35 (**2013**) 221-231.
[23] Y. Sun, D. Liang, X. Wang, X. Tang, DeepID3: Face Recognition with Very Deep Neural Networks, *Computer Science*, (**2015**).
[24] J.C. Chen, V.M. Patel, R. Chellappa, Unconstrained face verification using deep CNN features, in: IEEE Winter Conference on Applications of Computer Vision, 2015, pp. 1-9.
[25] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86 (**1998**) 2278-2324.
[26] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: application to face recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 28 (**2006**) 2037-2041.
[27] L. Ma, K. Khorasani, Constructive feedforward neural networks using hermite polynomial activation functions, *IEEE Transactions on Neural Networks*, 16 (**2005**) 821-833.
[28] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 26 (**2004**) 131-137.
[29] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 32 (**2010**) 2106-2112.