The 2nd International Conference on Artificial Intelligence, September 16-20, 2002, Crimea, Ukraine, pp. 428-436

ACCESS CONTROL BY FACE RECOGNITION USING NEURAL NETWORKS AND NEGATIVE EXAMPLES *

Dmitry Bryliuk, PhD student, bdv78@mail.ru

Valery Starovoitov, Doctor of Sciences, valerys@newman.bas-net.by

Institute of Engineering Cybernetics, Laboratory of Image Processing and Recognition, Surganov str., 6, 220012 Minsk, BELARUS

ABSTRACT

A Multilayer Perceptron Neural Network (NN) is considered for access control based on face image recognition. We studied robustness of NN classifiers with respect to the False Acceptance and False Rejection errors. A new thresholding approach for rejection of unauthorized persons is proposed. Ensembles of NN with different architectures were studied too. Advantages of the ensembles are shown, and the best architecture parameters are given. The usage of negative examples was explored. We have shown that by using negative examples we can improve performance for access control task. The explored NN architectures may be used in real-time applications.

INTRODUCTION

One of the most developing areas in modern world is a biometrics, intended to replace traditional identification systems. Biometrics systems are based on such unique human characteristics as fingertips, voice, retina pattern and others. Face recognition differs from other biometrics systems because there is no need for active interaction with such system and face recognition system can be built using cheap and widespread equipment.

This paper considers neural networks and its application for access control using face images. The main attention is given to access control task. It means we have group of people, authorized to access some resource. These are authorized persons. All other people are unauthorized, 'aliens'. System must prevent their access to resource.

We have used the ORL face database (www.cam-orl.co.uk/facedatabase.html). It has 40 persons and 10 different images for each person, 92x112 pixels with slightly varying lighting conditions, pose, scale, face expression and presence or absence of glasses. No attempts to normalize these images were made.

There are many works, considering face recognition [1-4]. But most of them studied only recognition rate without exploring how system performs during access control. We have studied not only recognition performance, but False Rejections Errors (FRR), False Acceptance Errors (FAR) and its Equality Errors. Several attempts to improve performance both of access control and recognition performance were made, and results are described in this paper.

* This work has been partially supported by project INTAS 00-626.

The 2nd International Conference on Artificial Intelligence, September 16-20, 2002, Crimea, Ukraine, pp. 428-436

1. THEORETICAL BACKGROUND

Multilayer Perceptron (MLP) Neural Network is a good tool for classification purposes [5,6]. It can approximate almost any regularity between its input and output. The NN weights are adjusted by supervised training procedure called backpropagation. Backpropagation is a kind of the gradient descent method, which search an acceptable local minimum in the NN weight space in order to achieve minimal error. Error is defined as a root mean square of differences between real and desired outputs of NN.

During the training procedure MLP builds separation hypersurfaces in the input space. After training MLP can successfully apply acquired skills to the previously unseen samples. It has good extrapolative and intrerpolative abilities.

Typical architecture has a number of layers following one by one [5,6]. MLP with one layer can build linear hypersurfaces, MLP with two layers can build convex hypersurfaces, and MLP with three layers – hypersurfaces of any shape.

In experiments we have used one hidden layer with 20 and 30 units in it. The number of units in the input layer is equal to the number of image pixels, 2576 in our case (i.e. 46x56). We have used images scaled down by factor 2 in order to speed up learning. Our previous experiments [7] has showed that such scaling did not change recognition rate on the ORL database. Gray level of every pixel was linearly scaled from range [0; 255] to [-0.05; +0.05] in order to avoid paralysis or surfeit of the network. The number of output units is equal to the number of classes, i.e. 40, the number of persons in the ORL database. Each output unit has corresponding "own" class.

We have used hyperbolic tangent as an activation function. It has output range [-1; +1] and outperforms the standard sigmoid function [5,6]. Each unit in the output layer is trained to give respond "+1" for the own class and "-1" for others. Thus MLP remaps the input space into the output space $\{O_i\}$, $O_i = \begin{cases} +1, i = class \\ -1, i \neq class \end{cases}$. In practice, real outputs are not exactly "+1" or "-1". They vary in the range [-1; +1] and the vicinity to the ideal values depends on the NN confidence. The closer output values to ideal, the more confidence to the NN decision. Recognition is performed by finding output neuron with the maximal value. The input image is considered as belonging to the class corresponding to this neuron. Then a thresholding algorithm is applied. It can reject or confirm decision of the NN.

For the training we have used a modification of the standard backpropagation procedure, an adaptive step suggested by Golovko [6]. When learning rate is small, training takes a long time. When learning rate is big, learning may never converge. Main advantage of this approach is that no need to manually select learning rate. Learning process is converged stable and fast (Fig. 1). In our experiments it takes from 50 to 100 training cycles to obtain the best performance and we used 100 cycles almost in all experiments, if otherwise is not mentioned. For the standard division of ORL (first five images of every person are used for training, last five – for testing) our system has recognition rate from 90% to 94%. Also we performed experiment to choose the number of hidden units for best performance (Fig. 2). As can be seen, adding more than 30 hidden units has no effect.

After training neural network produces practically ideal output for training samples (Fig. 3). Typical output for test sample is given on Fig. 4.

The typical output of NN for the case of unreliable classification, misclassification person looks like in Fig. 5. However, there can be cases such as in Fig. 6, where one person is more similar to another person than to itself (possibly due to similar pose or lighting conditions). For unauthorized persons in access control task output of the NN looks more diverse if there are no special algorithms were used for training.

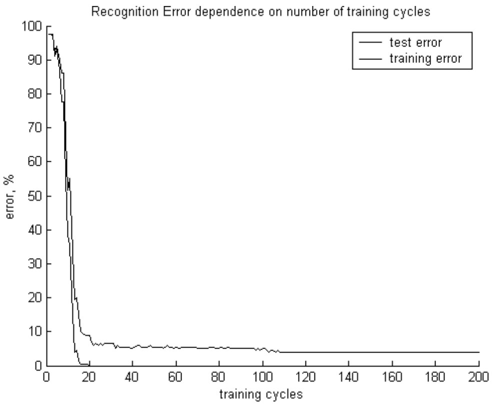

Fig. 1. Changing of recognition rate with training cycles

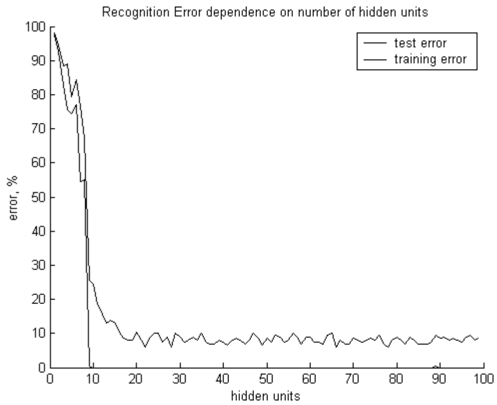

Fig. 2. Changing of recognition rate with number of hidden units

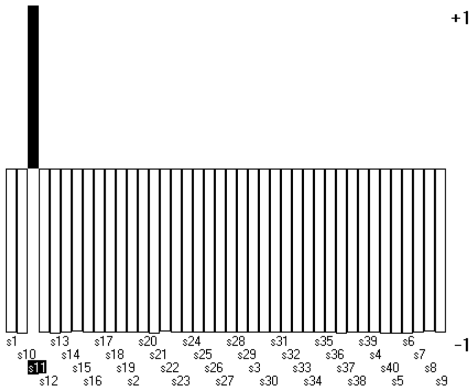

Fig. 3. Recognition of training sample, class "s11". (output of NN)

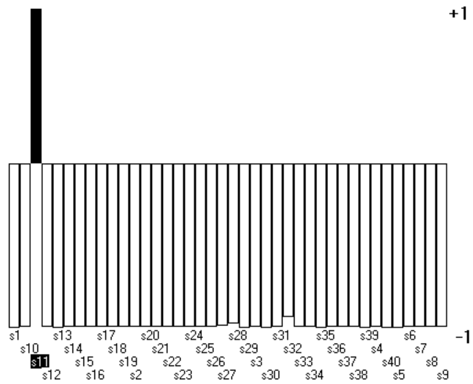

Fig. 4. Recognition of test sample, class "s11". (output of NN)

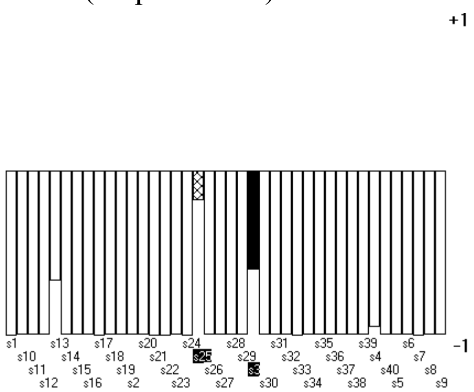

Fig. 5. Misclassification, class "s3" (black bar) recognized as "s25" (crossed bar)

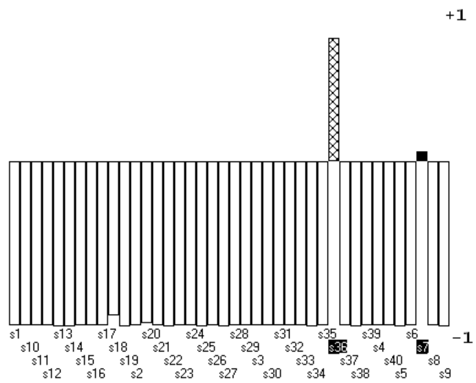

Fig. 6. Strong misclassification, class "s7" (black bar) recognized as "s36" (crossed bar)

2. EXPERIMENTAL RESULTS

For this set of experiments ORL database were divided on two parts. First part represented authorized persons and has 20 persons, from which random 5 images were used for training (total 100 images), other 5 for testing (total 100 images). Second part represented unauthorized persons. It has 20 persons and 10 images per person (total 200 images) only for testing purposes. Thus system has 100 images for training, and 300 for testing (100 authorized and 200 unauthorized).

In order to measure performance of algorithms we have used following factors. False Acceptance Rate (***FAR***), the number of unauthorized persons, considered as authorized, divided by total number of unauthorized attempts. False Rejection Rate (***FRR***), the number of authorized persons, considered as unauthorized, divided by total number of authorized attempts. FAR/FRR midpoints, the ratio when FAR and FRR is approximately equal. Recognition Rate, sum of misclassified authorized persons, false acceptance cases, false rejection cases divided by total number of access attempts. Also we considered stability of this factors. The system can achieve different recognition results on different training runs due to the probabilistic character of NN training. Stability is calculated as root mean square deviation (***RMSD***) of these factors on different runs from an average value. The less is ***RMSD*** the more stability the training algorithm has.

2.1. Experiment 1 – exploring thresholds

Besides the cases of correct and reliable classification the cases system must handle are misclassification, unreliable classification or attempt to access by unauthorized person for access control task. In this experiment we have explored two thresholding algorithms for rejection of such cases. The recognition system must reject such cases as much as possible, but perform well for authorized persons.

First thresholding algorithm (labeled ***'min'***) compares value of the maximal output neuron O_{max} with threshold t. When this output is lower than threshold, the decision of NN is rejected and person considered as unauthorized. Otherwise person considered as authorized. If we consider the output of NN as n-dimensional space (n – number of classes), then this algorithm will represent so-called ***"chess"*** metric. The value of threshold can be in range from "-1" (the lower value of NN output) to "+1" (the highest value of NN output).

The drawback of ***'min'*** algorithm is that it can't deal with situation such as on Fig. 6, when some class is similar to more than one class. Second thresholding algorithm (labeled ***'sqr'***) uses values of all output units and can cope with such situations. It calculates root mean square deviation from the real NN output to the ideal NN output: $d = \sqrt{\sum_{i=1}^{n} (O_i - \begin{Bmatrix} +1, i = \max \\ -1, i \neq \max \end{Bmatrix})^2}$. If d is less than threshold t, then NN decision is rejected, and person considered as unauthorized. Otherwise the person is authorized for access. This algorithm can be considered as Euclidean distance in NN output space, where each class has area, bordered by quarter of circle with radius t and center in its ideal position $\{O_i\}$ (Fig. 7). The minimal value of threshold is 0 and maximum is infinity, but practice has showed that for $t > 2$ there are can be no false rejections and we have used range [0; +2].

The comparative performance measure is shown on Fig. 8, 10, 11. Because the threshold ranges is different, we labeled ranges from 0 to 20 in order to place values in one graph. The graphs are showing averaged values for three different database divisions. For each division were performed three NN trainings with different initial weight seedings.

As can be seen, ***'sqr'*** thresholding algorithm gives better recognition rate for all threshold ranges and better FAR/FRR midpoints. Also ***'sqr'*** has significantly lower FAR and slightly higher FRR. In other words it much stricter to unauthorized persons and slightly stricter for authorized persons.

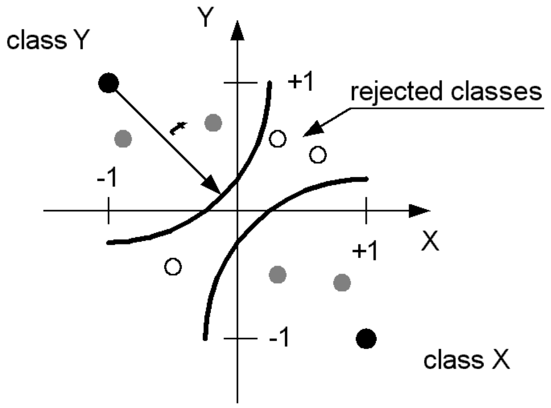

Fig. 7. NN output space and ***'sqr'*** thresholding algorithm

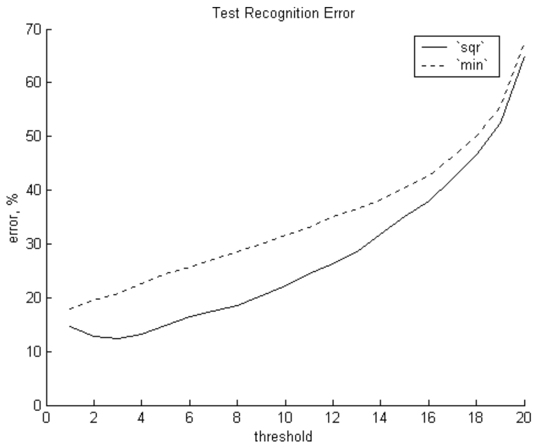

Fig. 8. Recognition Rate for both thresholding rules

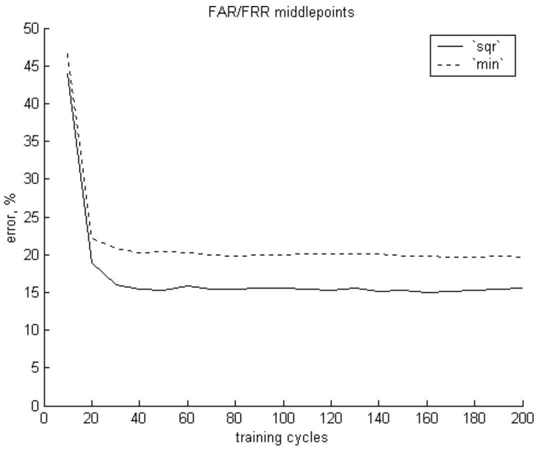

Fig. 9. FAR/FRR midpoints for both thresholding rules

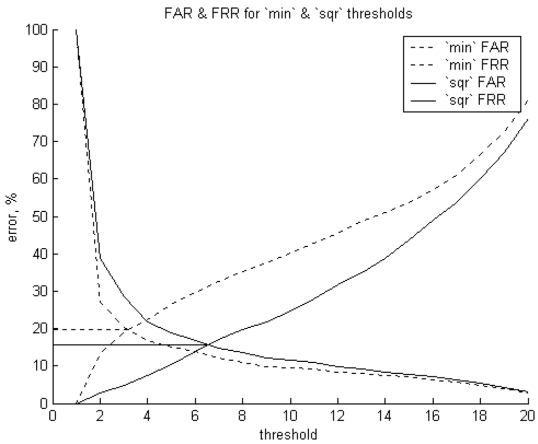

Fig. 10. FAR and FRR for both thresholding rules

2.2. Experiment 2 – exploring ensembles of neural networks

In this set of experiments we have measured performance of four different architectures of NN. First architecture (labeled ***'mlp'***) was usual multilayer perceptron with ***'sqr'*** thresholding rule.

Second architecture (***'one-one'***) is ensemble from 40 MLPs. Every MLP have one "own" class assigned with it. Each MLP have 2 layers, 20 hidden unit and one output unit. An output unit was trained to give "+1" for own person and "-1" for other persons. Outputs of ensemble were forming an aggregate network output, for example output of first MLP in ensemble were representing first output unit of aggregate NN. Then output of aggregate NN was considered as output of uniform MLP and ***'sqr'*** thresholding rule was applied as usual.

Third architecture (***'one-two'***) is like ***'one-one'***, but it has additional second output unit which was trained to give "-1" for own classes and "+1" for other classes. However in aggregate NN were considered only first output unit. By this architecture we have checked a statement that more different goals NN has, the eases the learning process and better the performance.

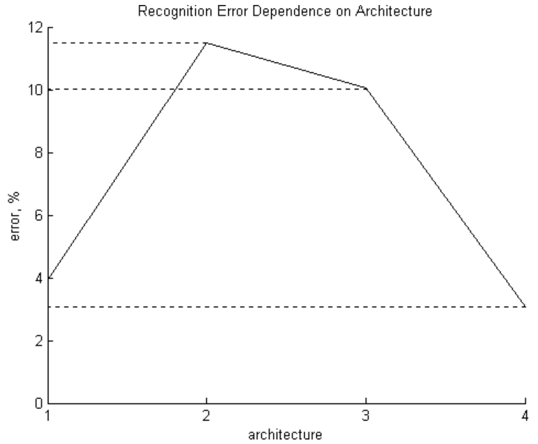

Fig. 11. Recognition performance for classification, all architectures

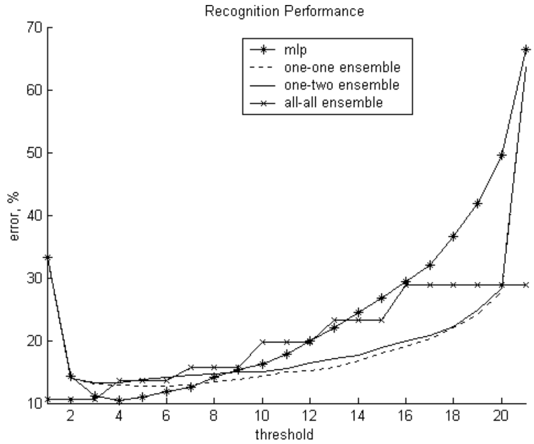

Fig. 12. Recognition performance for access control, all architectures, full thresholds range

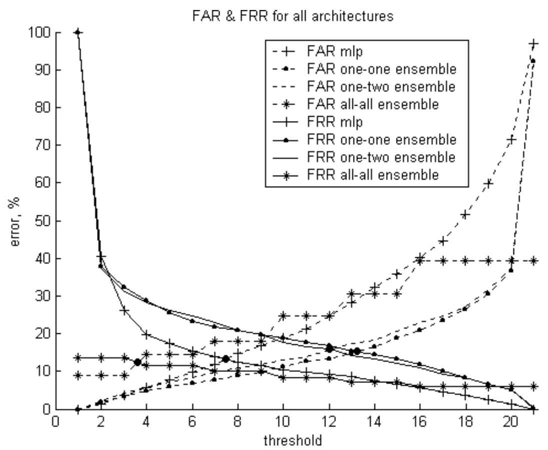

Fig. 13. FAR and FRR for all architectures, full thresholds range

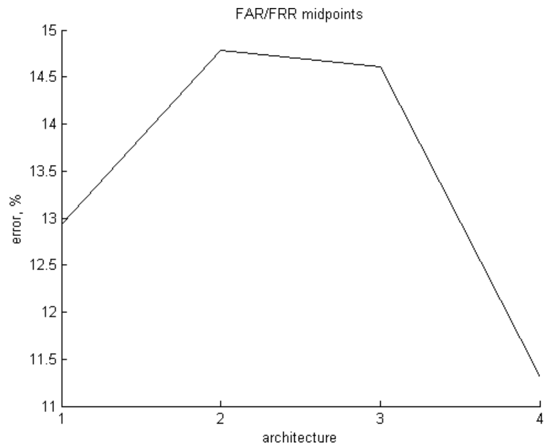

Fig. 14. FAR/FRR midpoints for all architectures

Fourth architecture (***'all-all'***) is an ensemble of NN of first (***'mlp'***) architecture. Voting makes the decision of such ensemble. Every MLP in ensemble gives one voice for person recognized by this ensemble. When the confidence for certain MLP is low, such MLP can abstain and give no voice at all. Confidence is checked by ***'sqr'*** thresholding rule. The threshold for abstaining was chosen experimentally and is equal to ***1.2***. Then decision rule is counting the number of voices for all persons. The person with maximum voices (it must be at least two voices) considered as recognized. Then we have compared the percent of voices for recognized person with thresholding percent (range [0; +1]). If the person has fewer voices than threshold, it considered as unauthorized and rejected. We have experimentally checked the performance of this architecture depending on number of ensembles. The performance was increasing with the number of MLPs and reaches optimum value for seven MLPs.

First, we have explored the ability of all architectures to classify face images (Fig. 11). All classes were used both for training and testing. Graph shows an averaged data for different database divisions and NN seedings. As can be seen from Fig. 11, the fourth architecture has best recognition performance. Second and third architectures are worse than usual MLP. As we expected the third architecture performs better than second. The stability of recognition results (RMSD) is the same. The behavior of FAR/FRR midpoints is practically the same, but the third architecture slightly better than second.

Second, we have explored the performance of these architectures for access control task. For the training were used 20 classes from 40 with five samples per class. Results can be seen on Fig. 12-15. First and fourth architecture has best recognition performance (Fig. 12). Fourth architecture has the best FAR/FRR relation (Fig. 14). Then follows first architecture. The second and third architectures is very strict to unauthorized persons (low FAR) but also strict for authorized persons.

Fourth architecture has significantly best performance as for recognition task either for access control. However it has less stability (RMSD) in recognition results than other architectures. The second and the third architectures are very strict to attempts for unauthorized access, but they have low performance for authorized persons (Fig. 13, 14).

2.3. Experiment 3 – use of negative examples

In this set of experiments we have explored the usage of negative examples. Negative example is an image of a person, which is always considered as alien. Such 'negative' persons were used only for training.

The database was divided into three parts. The first part consists of ten persons, which were the 'negative' persons. All ten images of each person were used only for training. The second part consists of ten 'authorized' persons. Five images of each person were used for training and five for testing. The third part consists of twenty 'alien' persons. Such persons were considered as unknown alien persons, and all ten images of each person were used only for testing.

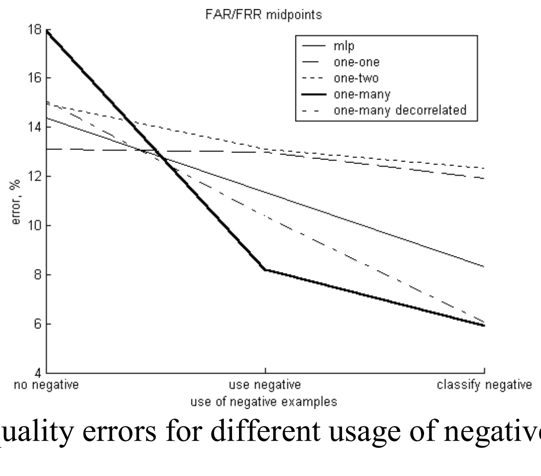

Fig. 15. Equality errors for different usage of negative examples

We have studied three cases of using negative examples. In the first case, the first, 'negative' part, was not used for training. We need such step to find out how performance is improved by using negative examples. In the second case, the neural network was trained to reject all negative examples. It means that each output of neural network trained to respond '-

1' for any negative example. And in the third case neural network was trained to classify all negative examples in the same manner as positive examples (authorized persons). It means that for each negative person neural network has separate output, which is trained to respond '+1' for its person and '-1' for others.

Fig. 15 shows performance of all architectures for each mentioned case. As can be seen, by using negative examples we can improve performance. When the unknown image is classified as one of the negative persons, we consider such person as alien and reject him. So, by using and classifying the negative examples, we can achieve better performance.

Also an attempt to decorrelate errors of neural networks in ensemble was made. As we have found, the best performance is achieved when the networks in ensemble have different initial weights and different example order.

CONCLUSION

As can be seen from the experimental results, the more different goals NN have to learn, the better performance is. A collective decision is better than a decision of one network. Also the introduced ***'sqr'*** thresholding rule has better performance for rejection an unauthorized persons than ***'min'*** thresholding rule. By using negative examples, we can significantly improve the performance for access control task.

Improvements presented in the paper are insufficient for creation a real access control system. First, an image must be normalized in brightness and contrast, face orientation and scale to bring it to uniform conditions. This is required to exclude systems reaction on similar shooting conditions that may be greater than difference between two different persons. Second, we need to develop technique, which will allow expanding training set in order to compensate lack of examples with different poses, lightning conditions, expressions and generation of 'synthetic' negative examples.

REFERENCES

1. Pan Z., Rust A. G., Bolouri H. Image Redundancy Reduction for Neural Network Classification using Discrete Cosine Transforms // Proceedings of the IJCNN. - 2000. - Vol. 3. - P. 149-154.
2. Lawrence S., Giles C. L., Tsoi A. C., Back A. D. Face Recognition: A Convolutional Neural Network Approach // IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition. - 1997. - P. 1-24. (http://www.neci.nec.com/~lawrence).
3. Eickeler S., Muller S., Rigoll G. High performance face recognition using Pseudo 2-D Hidden Markov Models. Gerhard-Mercator-University Duisburg, Germany, 1998. - 6 p.
4. Eickeler S., Jabs M., Rigoll G. Comparison of Confidence Measures for Face Recognition. Gerhard-Mercator-University Duisburg, Germany, 1999. - 6 p.
5. A.I. Wasserman Neural Computing: Theory and Practice – New York: Van Nostrand Reinhold, 1989.
6. Golovko V., Gladyschuk V. Recirculation Neural Network Training for Image Processing // Advanced Computer Systems. - 1999. - P. 73-78.
7. Bryliuk D., Starovoitov V. Application of Recirculation Neural Network and Principal Component Analysis for Face Recognition // The 2nd International Conference on Neural Networks and Artificial Intelligence. - Minsk: BSUIR, 2001. - P.136-142.