INFORMATION
TECHNOLOGY JOURNAL

A Study of Neural Network and its Properties of Training and Adaptability in Enhancing Accuracy in a Multimodal Biometrics Scenario

Fawaz Alsaade

College of Computer Sciences and Information Technology,
King Faisal University, Al-Ahsa, Saudi Arabia

Abstract: The main aim of this study was to present investigations for enhancing the accuracy of multimodal biometrics by introducing the neural network into the score-level fusion process. Presently, various fusion techniques are being widely used in combining separate information from different modalities to provide complementary data. The resilient backpropagation training algorithm was used for this purpose. The effectiveness of the proposed method is to benefit from the properties of training and adaptability of the neural network technique. The experimental investigations involved the recognition mode of verification in mixed-quality data conditions. It was found during the study that by deploying such technique at the score level, the system error rate can be reduced considerably. The study presented the motivation and the potential advantages of the proposed approach and the details of the experimental study.

Key words: Multimodal biometrics, score-level fusion, neural network, biometric, verification

INTRODUCTION

Presently, a rapid growth has been witnessed in the demand for effective identity verification methods for an increasing range of applications. This process, however, has been found to be considerably problematic due to the inherent limitations of traditional authentication methods (Bubeck, 2003). Because, the keys are easily misplaced, copied, stolen, or mechanically by-passed through various forms of stealth and artifice. Also, the Pin numbers slip too readily from memory or if recorded for later reference and they are potentially accessible to impostors. The biometric systems essentially supercede these knowledge-based verification controls since they are found in one who is the legitimate user by virtue. For example, of one's face, fingerprints, voice, hand-geometry or retina scans. However, the verification of identity using biometric data is replete with operational problems such as imperfect imaging conditions or changing conditions from one sample to another, changes in user's physiological or behavioral characteristics, user's interaction with the sensor and the ambient conditions such as temperature and humidity fluctuation. In recent years, an area of considerable interest in biometric recognition has been the use of multiple modalities. The main attraction of multimodal biometrics is that it provides the opportunity for enhancing the recognition accuracy beyond that achievable with unimodal biometrics.

A multimodal biometric system requires an integration scheme to fuse the information obtained from the individual modalities. Such information is proved to be complementary to each other (Jain *et al.*, 2004; Indovina *et al.*, 2003). There are various levels of fusion from sensor to decision level (Jain *et al.*, 2004). The fusion at the matching score level is the most popular and frequently used method due to its better performance, intuitiveness and simplicity (Indovina *et al.*, 2003).

In this study, a neural network approach is proposed for combining data obtained from face and voice modalities. The resilient backpropagation training algorithm was used for this purpose. The main aim of this study was to explore the potential usefulness of neural network technique and to investigate the possibility of benefiting from its properties of training and adaptability in enhancing accuracy in a multimodal biometrics scenario. To investigate the effectiveness of the suggested approach, it was compared with two well known fusion schemes namely Brute Force Search (BFS) and Genetic Algorithms (GA). The Mean Squared Error (MSE) was used as a performance indicator for these experiments.

Fusion techniques
Brute Force Search (BFS): This fusion technique can be used in the case of having two matcher types only. The approach is based on the following equation (Ma *et al.*, 2005).

$$u = x_1 w + x_2 (1-w) \qquad (1)$$

where u is the fused score, m^{th} is the normalized score of the mth matcher, m = 1 or 2 and w is a weighting (combination) factor in the range 0 to 1. The weight (w) is calculated heuristically, by exhaustive search in order to minimize the Mean Squared Error (MSE) on the given development data.

Genetic algorithms (GA): The genetic algorithms have proved their capability of elite preservation strategy to search for optimum solutions in multi-dimensional space without worrying about local minima. They were intensively investigated in the last decades in optimization problems and several variants have been proposed in the literature (Kamepalli, 2001; Jia *et al.*, 2003; Castillo *et al.*, 2007; Sun and Wan, 1995). They rely mainly on performing biological type operations such as reproduction, cross-over, mutation and selection according to some predefined fitness or cost function. The reproduction scheme generates a population of candidates in some region of the space; i.e., exploration, cross-over will give birth to offspring of the next generation, mutation simulate small random variation of the genotype and the selection will preserve only elite candidates or 'best candidate' solutions; i.e., exploitation.

The algorithm starts with generating an initial population of candidates $W_i^0 = (w_1, w_2)$ with a uniform distribution. Then, for each candidate W_i, compute MSE(u)
where:

$$u = w_1, x_1 + w_2, x_2 \qquad (2)$$

where, u is the fused score, x_m is the normalized score of the m-th matcher, m = 1 or 2 and w_m is the corresponding weight (obtained on some development data) in the interval of 0 to 1, with no constrains. Then, select best candidates W_i^0 for which fitness MSE(u)'s are minimal, others are discarded, these will represent the elites for which mutation and cross-over operations are performed. At this stage a new population i is created; W_i^j; for the next generation j, where, i = 1 to N, N is the population size and j = 1 to M, M is the maximum number of generations. This process will be iterative going back again to computing $MSE(u_i^j)$ where:

$$u_i^j = x_1 w_{i1}^j + x_2 w_{i2}^j \qquad (3)$$

where, u_i^j is the fused score for the ith population in the jth generation.

In this technique, it is ensured that the population size during all generations remain unchanged. Up to a number of predefined generations M, the best candidate

W for which $MSE(u_i^j)$ is minimal is then selected. However, so far, no formal method exists in the current GA literature as to predefine best GA parameters. They need performing several runs to find best-fit GA parameters, i.e., population size, number of generations, reproduction, cross-over, mutation schemes and most importantly the selection criteria, i.e., MSE for such fusion scores is minimal.

In this study, the cross-over probability is 0.90, the mutation rate is 0.10, the population size is 100, the number of generations is 10 and the fitness function is such that the vector weight for which the error rate of the fused scores be minimal.

Neural network: The most popular neural network in pattern recognition and decision is multilayer feed-forward type. The back-propagation is the most useful training algorithm for feed-forward networks and is used to calculate the gradient of the error of the network with respect to the network's modifiable weights. This gradient is then used in a simple stochastic gradient descent algorithm to find weights that minimize the system error rate. There exist faster algorithms that use heuristic techniques. One heuristic modification is the variable learning rate backpropagation and resilient backpropagation. In this study, the resilient backpropagation training algorithm was used (Freeman *et al.*, 1998; Zainuddin and Abu-Hassan, 2005). The purpose of using such algorithm is to eliminate the harmful effects of the magnitudes of the partial derivatives.

The proposed structure for fusing biometrics is shown in Fig. 1.

In this technique, a feed-forward multi-layer neural network (Freeman *et al.*, 1998; Zainuddin and Abu-Hassan, 2005) is used with the following parameters:
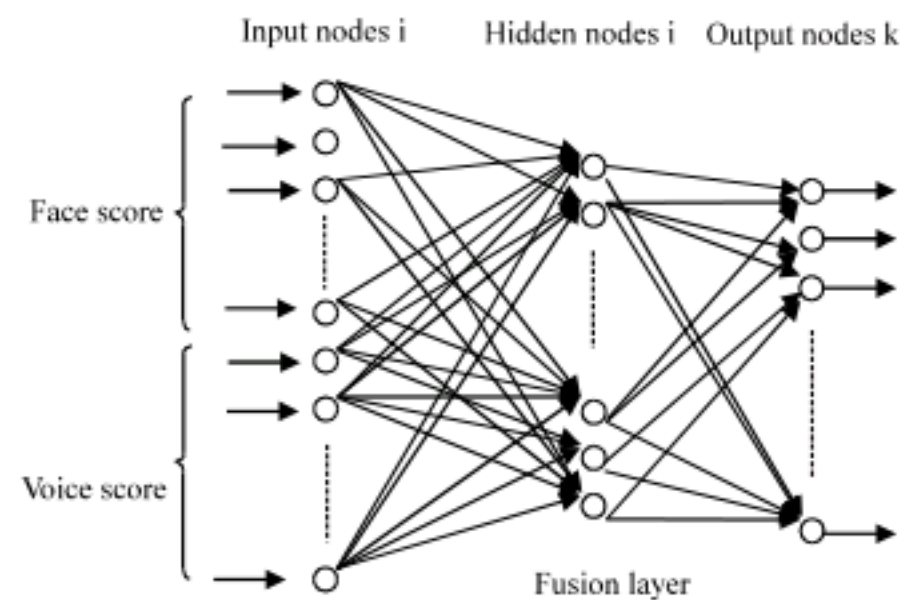

Fig. 1: Structure of the fusion neural network of scores from the individual biometric modalities

Input layer with 2N nodes, N nodes for face modality and N nodes for voice modality.

One hidden layer of M nodes is used for fusing biometrics; M being less than 2N. For each client represented by a concatenation vector of both modalities; one output node is assigned value 1 for sake of recognition. For impostors, the remaining output nodes are forced to zero. The output layer will be of N nodes, each of these for a particular client to be recognized. In this technique, the scores for face and voice modalities are of size N×N. Thus, the combined scores for fusion will be of size N×(2N) each row is score vector of a particular client.

The fusion process is achieved by the proposed scheme in this work as follows:

$$\theta_j = F\left(\sum_{i=1}^{2N}\left(w_{ij}S_i^{face} + w_{ij}S_i^{voice}\right)\right) \qquad (4)$$

$$\theta_k = F\left(\sum_{i=1}^{M}(w_{jk}\theta_j)\right) \qquad (5)$$

where, θ_j is the output of node j in the hidden layer, F is a simple transfer function (sigmoid) for all nodes, S_i^{face} is a score vector for a particular client from the face modality, S_i^{voice} is a score vector for a particular client from the voice modality, w_{ij} is the synaptic weight connection node i to node j. The part θ_k in Eq. 2 represents the output of node k in the output layer and w_{jk} being synaptic weight between node j in the hidden layer and node k in the output layer.

In the development stage, the weigh matrices w_{ij} and w_{jk} are rearranged in such a way that the network will perform a recognition of a particular client over impostors. As said earlier that the accelerated resilient back-propagation algorithm, in this study, aims to minimize the mean squared errors using the Delta rule [9,10].

$$w_i(t) - w_i(t-1) = \Delta(w_i) = -\eta\frac{\partial E}{\partial w} \qquad (6)$$

where, η is the momentum of the gradient and E is the sum squared error between the actual output and the desired one as follows:

$$E = \sum_{k=1}^{N}(\theta_k^{Desired} - \theta_k^{Actual})^2 \qquad (7)$$

Finally, the fusion is made explicitly within the network structure at the individual score level in the hidden layer. Therefore, the computational-based search of best-fit neural synaptic weights for fusion is simply generated by adapting the neural network for best-decision scores.

RESULTS AND DISCUSSION

The experimental studies are concerned with the score-level fusion of face and voice biometrics in the recognition mode of verification. The investigations were initially performed by using scores for clean face images together with scores for degraded utterances.

In each experiment, the individual biometric score types involved were subjected to the range equalization process using the Min-Max normalization (Indovina *et al.*, 2003). In this study, the process of score-level fusion is based on the use of brute force search, genetic algorithms as well as the neural network. The procedures followed for speech feature extraction and speaker classification were according to Fortuna *et al.* (2004). The face recognition scores were based on the approach described by Zafeiriou *et al.* (2006).

Fusion under varied data quality conditions: The datasets considered for the face and voice modalities were extracted from the XM2VTS (clean images) (Zafeiriou *et al.*, 2006) and from the 1-speaker detection task of the NIST Speaker Recognition Evaluation 2003 (degraded speech) databases, respectively (Fortuna *et al.*, 2004). Using these data sets, a total number of 140 client tests and 19460 (i.e., 140×[140-1]) non-client tests were used from the development data for investigating the performance of the proposed schemes.

The results of verification experiments were presented as Mean Squared Errors (MSEs) (Table 1). The proposed neural network approach was compared to two well known fusion schemes namely the Brute Force Search (BFS) and the Genetic Algorithm (GA). These three fusion approaches are mainly concerned with adjusting the balance of weighting in fusion in favour of the modalities of better quality. Therefore, it can be stated that the method that could introduce an appropriate weighting scheme should lead to better verification results.

In the first fusion technique (BFS), although weights are calculated heuristically, by exhaustive search in order to minimize the MSE on the given development data. It was noticed that the use of BFS has increased the MSE obtained with the best single modality. This increase in MSE could be due to highly degraded speech database involved in this study. However, it should be emphasized

Table 1: Effectiveness of BFS and GA in biometric verification based on mixed-quality data

Modality	MSE (%)
Voice (NIST)	35.56
Face (XM2VTS)	4.75
Fusing by BFS	4.86
Fusing by GA	0.68
Fusing by neural network	0.05

that the use of GA, successfully reduced the MSE for fused biometrics up to 86% which resulted from the characteristics of GA that aims to assign best fit weights to the biometric scores. It was also clear from the results that neural network approach leads to the best performance as compared to BFS and GA (Table 1). In this case, the reduction achieved in MSE with such a fusion scheme was in excess of 99% as compared to the better modality. These results confirmed the usefulness of neural network in enhancing accuracy in multimodal biometric systems. This could be due to the fact that during the training stage, neural network successfully tunes the connection weights to minimize the system error rates.

CONCLUSION

The study findings introduced the resilient backpropagation training algorithm for combining data obtained from face and voice modalities. Amongst the three fusion methods considered, neural network approach appears to offer considerable improvements to the accuracy of multimodal biometrics in varied data conditions which seems to be related to the individual characteristics of the proposed neural network approach. The fusion is made explicitly within the network structure at the individual score level. The network connection weights can successfully be tuned during the training stage for best decision scores. The preliminary results of the proposed approaches provides motivation for further research in order to exploit the properties of training and adaptability of neural network along with fuzzy systems to multi-modal fusion.

REFERENCES

Bubeck, U.M., 2003. Multibiometric Authentication an Overview of Recent Developments. Spring, San Diego State University, Diego State.

Castillo, V.C., M.K.E. El-Debs and M.C. Nicoletti, 2007. Using a modified genetic algorithm to minimize the production costs for slabs of precast prestressed concrete joists. Eng. Appl. Artificial Intel., 20: 519-530.

Fortuna, J., P. Sivakumaran, A. Ariyaeeinia and A. Malegaonkar, 2004. Relative effectiveness of score normalisation methods in open-set speaker identification. Proceedings of the ODYSSEY 2004 the Speaker and Language Recognition Workshop, May. 31-Jun. 3, Canon Research Centre Europe Ltd., Bracknell, Berkshire, UK, pp: 369-376.

Freeman, J., S. Ramakrishnan, K. Varnik, M. Neuhaus, P. Burk and D. Birchfield, 1998. Algorithms and Architecture: Neural Network Systems Techniques and Applications. 1st Edn., Academic Press, London, pp: 293-317.

Indovina, M., U. Uludag, R. Snelick, A. Mink and A. Jain, 2003. Multimodal biometric authentication methods a COTS approach. Proceedings of the MMUA 2003, Workshop on Multimodal User Authentication, Dec. 11-12, Santa Barbara CA., pp: 99-106.

Jain, A.K., A. Ross and S. Prabhakar, 2004. An introduction to biometric recognition. IEEE Trans. Circ. Syst. Video Technol. Special Issue Image VideoBased Biometr., 14: 4-20.

Jia, H.Z., A.Y.C. Nee, J.Y.H. Fuh and Y.F. Zhang, 2003. A modified genetic algorithm for distributed scheduling problems. J. Intel. Manuf., 14: 351-362.

Kamepalli, H.B., 2001. The optimal basics for GAs. Potentials IEEE, 20: 25-27.

Ma, Y., B. Cukic and H. Singh, 2005. A classification approach to multi-biometric score fusion. Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication, Jul. 20-22, Hilton Rye Town, New York, USA., pp: 484-493.

Sun, Z. and Q. Wan, 1995. A modified genetic algorithm: Meta-level control of migration in a distributed GA. Proceedings of the IEEE International Conference on Evolutionary Computation, Nov. 29-Dec. 1, Perth, WA, Australia, pp: 312-328.

Zafeiriou, S., A. Tefas, I. Buciu and I. Pitas, 2006. exploiting discriminant information in non-negative matrix factorization with application to frontal face verification. IEEE Trans. Neural Networks, 17: 683-695.

Zainuddin, N.M. and Y. Abu-Hassan, 2005. Improving the convergence of the backpropagation algorithm using local adaptive techniques. Proc. World Acad. Sci. Eng. Technol., 1: 3-6.