Face Recognition for Homeland Security: A Computational Intelligence Approach

Grant J. Scott, James M. Keller, Marjorie Skubic, Robert H Luke III

Department of Computer Engineering and Computer Science
University of Missouri-Columbia
{ScottGS, KellerJ, SkubicM, rhl3db}@missouri.edu

***Abstract*-By utilizing Morphological Shared-Weight Neural Networks (MSNN) that have been trained for face recognition, common access restriction points can be enhanced to identify particular individuals of interest. A trained MSNN is a computational intelligence structure that learns representation of a specific face that encodes in its connection weights the feature extraction and classification abilities needed to identify an instance of that face. It has been shown effective in analyzing images that contain the target in a group of faces, even with the target face at varying orientations and lighting, as well as occluded target faces. The experiments presented here show the possible application of the MSNN to perform watch-list scanning of faces as individuals pass through access screening areas.**

I. Introduction

A fundamental need in security is the ability to be alerted when a person of interest enters a secure or restricted area. Common in airports and many government buildings is the presence of a choke point, through which access to sensitive areas is monitored (e.g. a metal-detector). Using a face recognition system developed with Morphological Shared-Weight Neural Networks (***MSNN***)[2,3], these points of access can be automatically monitored for persons of interest. This paper presents the abilities of the MSNN to perform face recognition and its suitability as a security system component. MSNNs were first developed by Won and Gader [3,4] and have been used mostly for vehicle recognition under a variety of range modalities [1,3,4,5]. Some advantages of the MSNN for face recognition are that no segmentation and face component identification is required and input features and classification are learned simultaneously from a small set of face samples. As an example application, when an individual passes through the screening area of an airport, an automatic system can capture and search for matches on a terrorist watch-list. This can be done in a pervasive manner that requires no cooperation on the part of the individuals being scanned.

Other approaches to face recognition include shape decomposition with morphological methods, template matching, and feature vector extraction using PCA [5,6,7,8,9,10,11,13]. Many of these approaches require normalization of the face image prior to processing. EigenFaces are the basis for much of the current research in face recognition; however, this approach becomes increasingly error prone when the scanned face varies in orientation and lighting level from the training data [10]. In contrast the MSNN has been shown to be un-affected by variations in lighting level and robustly handles variation in target orientation [3,4]. Additionally, the MSNN has a proven ability to perform recognition under image occlusion [1,2,3,4]. Figures 1 and 2 provide examples of recognition of the target face under occluded conditions.

We have shown previously that the MSNN is robust enough to perform face recognition in natural environments and under various conditions [2]. In this paper, we examine the applicability of the MSNN to surveillance and security systems. Our experiments show a scenario where the MSNN scans a sequence of images as individuals pass through a doorway, simulating the metal-detector scenario. Fig. 1 shows the "hit point" identified on the target face in a test image after it was scanned by a trained MSNN. The target face is correctly identified with sunglasses even though no such images were included in the training set. Fig. 2 demonstrates the resiliency of the MSNN to occlusion and Fig. 3 shows the ability to detect the target face with a downward and to the side orientation. The MSNN's ability to perform the recognition under target occlusion is a key feature [4]. The MSNN handles occlusion and variances in target scale despite the fact that no such training data is supplied.

Figure 1 Testing of face recognition ability: input image is marked in white pixels by the MSNN where it detects the target face.

Figure 2 Example of target identified under occlusion from a second face in the image scene.

Figure 3 Recognition of the target face at a downward, turned orientation, wearing glasses. Note the single random false alarm.

II. USING THE MSNN FOR FACE RECOGNITION

The MSNN is a heterogeneous network composed of two cascaded sub-networks, the feature extraction network and the classification network (see Figure 4). The feature extraction layer takes a two dimensional array as input, which is the input sub-image. This input is passed through structuring elements, in the form of connection weights, into each feature map layer. All feature map nodes in a given feature map share the structuring element connection weights. The learning of these feature map input weights results in structuring elements that can perform linear or non-linear mappings (see [2,3] for details of MSNN training). In the context of the feature maps, these are the morphological ***hit*** and ***miss*** kernels. The propagation of the input signal through these structuring elements and the feature maps performs the gray-scale ***Hit-Miss Transform***, which is the output result passed to the classification phase of the MSNN. The Hit-Miss Transform provides output from two net inputs to a feature map node y as,

$$a_y = net_y^h - net_y^m \cdot$$

Where net_y^h *and* net_y^m inputs represent the *hit* and *miss* operations performed through the structuring elements, respectively. The *hit* and *miss* are defined as [3]

$$net_y^h = \left\{ \sum_{x \in D_{ty}} w_y^h(x) \left[r\left\{ a(x) - t_y^h(x) \right\}^{p_h} \right] \right\}^{\frac{1}{p_h}}$$

and

$$net_y^m = \left\{ \sum_{x \in D_{ty}} w_y^{\hat{m}}(x) \left[r\left\{ a(x) - t_y^{\hat{m}}(x) \right\}^{p_m} \right] \right\}^{\frac{1}{p_m}}.$$

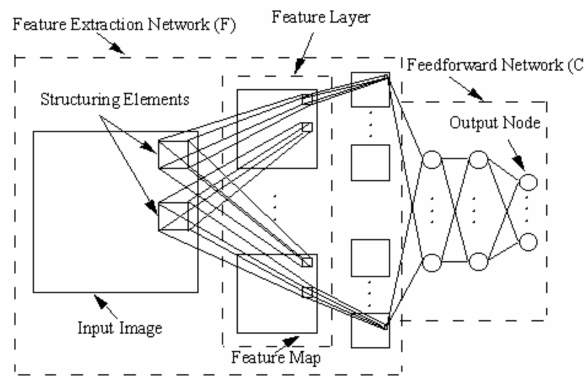

Figure 4 The Morphological Shared-Weight Neural Network (MSNN) architecture.

The output of the top-level feature maps becomes the direct input of the classification phase. This sub-network is a multi-layer feed forward network using back-propagation learning. The two layers of MSNN are trained together, resulting in the network simultaneously learning both the necessary feature extraction and classification. This is accomplished by propagating the error back from the classification into the feature extraction stage. The output of the MSNN is a two-node layer, one node representing confidence that the input sub-image is the intended target, and one node representing confidence that input is a non-target.

The output of the network is used to create a Detection Image Plane (***DIP***), which is a [0,1] set of confidence values the same dimension as the input image, corresponding to the MSNN confidence of a face centered at each pixel coordinate. Through thresholding and various other image processing techniques the location of the target face can be determined [2]. Figures 5,6,7 show a test image and the resulting DIP and output Image generated from the MSNN scan of the test image.

Figure 5. Standard testing input image of the MSNN

Figure 6. The Detection Image Plane (DIP) produced by the MSNN for the input image of Figure 5, the output in the confidence interval [0,1] is mapped to [0,255] gray-scale.

Figure 7. An output image produced by the MSNN for the input image in Figure 5 with face recognition points marked by white pixels, where MSNN confidence was over 0.94, or 240 gray-scale.

The MSNN for particular face recognition was trained using 34 PGM images of a 320x240 resolution, with the target face in 31 images. In training, sub-image size of the target face was generally about 30x40 pixels. The feature extraction phase of the MSNN was designed with a single feature extraction layer of four feature maps. The classification stage was 1200x8x8x2 feed forward neural network. Additional information on the training and structure of the network can be found in [2].

III. Security Scenario Experiment

The scenario we established for this experiment was a simulation of people passing through an access choke point similar to airport metal-detectors. We set a camera to capture a sequence of images as a group of people passed through a doorway. The participants were not instructed to look at the camera or behave in any special manner. The images were acquired using a Logitech 3000pro QuickCam, and all images were 320x240 pixels. Figure 7 shows an example of the MSNN recognizing the target individual as he passes through the choke point.

Figure 8. Video frame sample (1) from test scenario as target individual approaches the choke point.

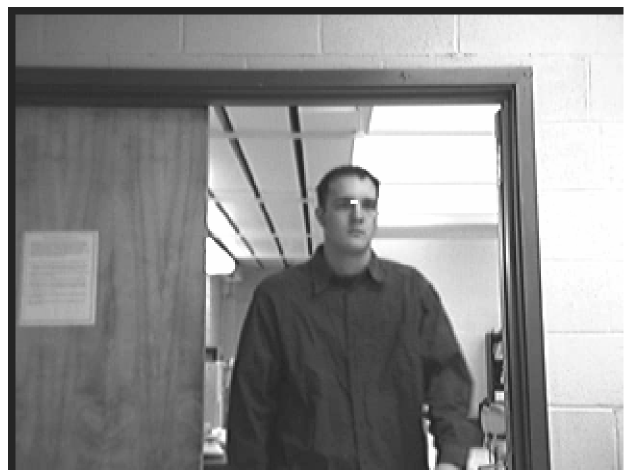

Figure 9. Video frame sample (2) from test scenario as target individual passes through the choke point.

Figures 8-11 show a sequence of images depicting the approach and transition of the individual through the choke point. In each frame sampled from the video feed the MSNN marks the individual's face, demonstrating its recognition. Currently, the MSNN is only trained on one particular face at a time. One possibility, for security screening, is to train multiple nets each to look for one person. Further experimentation will determine if the MSNN can be modified to use a single set of feature maps to classify multiple faces. This modification would result in the classification phase using multiple output nodes to encode the classification of the face, as opposed to the current target/non-target two-node output architecture. A face recognition system using the MSNN could be implemented to continuously sample frames from a video source and run them against a watch-list. Another application is the deployment of a watch-list face recognition system in a less structured environment, Figures 12-14 show the recognition in a scene with multiple faces. Here again, only the target face is identified in the imagery. In [2] we discuss the sensitivity and specificity of face recognition using this network-based approach. The MSNN has the ability to perform face recognition in natural settings. This allows it to sample video frames from various sources and still perform recognition. Possible sources include closed-circuit security cameras that are common in many places. Since the MSNN uses low-resolution gray-scale images, even the oldest security cameras can be used to feed a signal to the system. The use of these low resolution images, combined with varying frame sampling rates, would allow the application to operate in real time, without the aid of specialized hardware.

Figure 10. Video frame sample (3) from test scenario as target individual has passed through the choke point.

IV. FURTHER UTILIZATION OF SOFT COMPUTING

We are continually developing ideas to improve and extend the use of the Morphological Shared-Weight Neural Network. For further improvements in face recognition, the focus will be at the application level, where additional techniques of soft computing will be employed. An initial research direction is the extension of using a fuzzy integral to analyze and detect the face across a temporal spectrum. The application in the scenario presented in Section III would allow the face recognition system to develop confidence over time about a particular face as it is tracked across the image scene. One issue that such a temporal fusion will address is a reduction of false alarms. Our experience shows that false alarms tend to be more random and smaller than true positives. Hence, using temporal information should significantly reduce the potential for false alarms. Figure 15 provides a few images from an extended sequence that includes the target approaching, conversing, then departing in the image scene. Figures 15(b) and (e) show typical false alarms on the arm of a secondary individual; these could be removed using a fuzzy rule-base.

Another possible extension is the partitioning of faces into sub-components and having multiple MSNN trained on the various sub-components. This results in each face being encoded in a combination of MSNN. For example, a three-MSNN-segmentation might include training one MSNN for the eye region, another for the mouth-nose region, and a third for entire facial region. Then using some fuzzy fusion operator, all these confidences can be combined to produce a single confidence value. An extension of using MSNNs trained for partitions of a face could be the implementation of a fuzzy logic rule based mechanism for determining matches. Once a full system is implemented, a soft computing method will also be developed for tracking faces as they move across the image scene. This will allow the system to intelligently focus on possible matches in successive scans, thereby scanning a smaller area of the input signal.

V. CLOSING REMARKS

The MSNN provides the ability to learn a target face with a relatively small training set, and performs recognition of the target face under various scene conditions. The degree to which it handles occluded images and still performs recognition makes it ideal for real world situations where the subjects are unaware or uncooperative. The performance despite the variation in lighting levels, orientations, and scale makes a face recognition system using the MSNN a superb choice for integration into existing video security systems.

Figure 11. Video frame sample (4) from test scenario as target individual continues past the choke point.

REFERENCES

[1] M. Khabou, P. Gader, and J. Keller, "Morphological Shared-Weight Neural Networks: a Tool for Automatic Target Recognition Beyond the Visible Spectrum," in *Proc. of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, 1998.

[2] G. Scott, R.H. Luke III, M. Skubic, and J.M. Keller, "Face Recognition with Morphological Shared-Weight Neural Networks," Dept of Computer Engineering and Computer Science Technical Report, University of Missouri-Columbia 2002.

[3] Y. Won, *Nonlinear Correlation Filter and Morphology Neural Networks for Image Pattern and Automatic Target Recognition*, Ph. D. Dissertation, University of Missouri – Columbia, 1995.

[4] Y. Won, P. Gader, and P. Coffield, "Morphological Shared-Weight Neural Networks with Applications to Automatic Target Recognition," *IEEE Trans. Neural Networks*, Vol. 8, pp. 1195-1203, 1997.

[5] K.C. Chung, S.C. Kee, and S.R. Kim, "Face Recognition using Principal Component Analysis of Gabor Filter Responses," *Proc. of Int'l Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999, pp. 53-57.

[6] G.D. Guo and H.J. Zhang, "Boosting for Fast Face Recognition," *Proc. of IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 2001, pp. 96-100.

[7] B. Moghaddam, C. Nastar, A. Pentland, "A Bayesian Similarity Measure for Direct Image Matching," *Proc. of Int'l Conference on Pattern Recognition*, 1996.

[8] B. Moghaddam and A. Pentland, "An Automatic System for Model-Based Coding of Faces," *Proc. IEEE Data Compression Conference*, 1995.

[9] B. Moghaddam and A. Pentland, "Beyond EigenFaces: Probabilistic Matching for Face Recognition," *Proc. 3rd IEEE Int'l Conference Automatic Face & Gesture Recognition*, 1998.

[10] M. Turk and A. Pentland, "Face Recognition Using EigenFaces," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Analysis*, 1991, pp. 586-591.

[11] A. Tefas, C. Kotropoulos, and I. Pitas, "Face Verification based Morphological Shape Decomposition," *Proc. 3rd IEEE Int'l Conference Automatic Face & Gesture Recognition*, 1998, pp. 36-41.

[12] J. Keller and X. Wang, "A Fuzzy Rule-based Approach for Scene Description Involving Spatial Relationships," *Computer Vision and Image Understanding*, Vol. 80, 2000, pp. 21-41.

[13] M. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Anal. Machine Intell*, Vol 24, pp. 34-58, 2002.

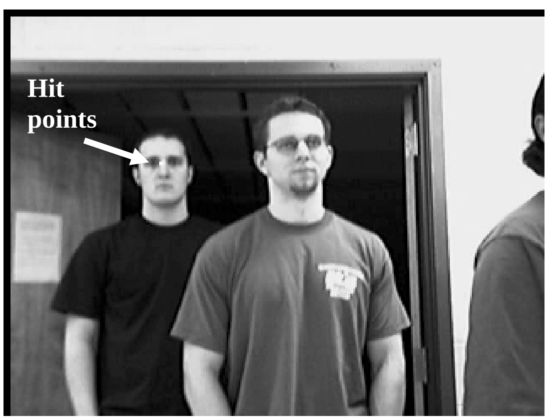

Figure 12. Test scenario image of target face recognition with other faces in the image scene. The target is marked on the edge of the nose from the MSNN scan.

Figure 13. Another multiple face test scene, the target is marked on the edge of the nose and corner of the eye.

Figure 14. A continuation of the scene from Figure 13 with the men moving across the scene.

Error!

(a)

(b)

(c)

(d)

(e)

Figure 15. A test scenario sequence with the target approaching and conversing with other individuals. Images (b) and (e) have typical false alarms on the secondary individuals arm.