An Artificial Neural Network Based Matching Metric for Iris Identification

Randy P. Broussard*[a], Lauren R. Kennell[b], Robert W. Ives[b], Ryan N. Rakvic[b]
[a]Systems Engineering Department, U. S. Naval Academy, Annapolis, MD 21402;
[b]Electrical and Computer Engineering Department, U. S. Naval Academy, Annapolis, MD 21402

ABSTRACT

The iris is currently believed to be the most accurate biometric for human identification. The majority of fielded iris identification systems are based on the highly accurate wavelet-based Daugman algorithm. Another promising recognition algorithm by Ives et al uses Directional Energy features to create the iris template. Both algorithms use Hamming distance to compare a new template to a stored database. Hamming distance is an extremely fast computation, but weights all regions of the iris equally. Work from multiple authors has shown that different regions of the iris contain varying levels of discriminatory information. This research evaluates four post-processing similarity metrics for accuracy impacts on the Directional Energy and wavelets based algorithms. Each metric builds on the Hamming distance method in an attempt to use the template information in a more salient manner. A similarity metric extracted from the output stage of a feed-forward multi-layer perceptron artificial neural network demonstrated the most promise. Accuracy tables and ROC curves of tests performed on the publicly available Chinese Academy of Sciences Institute of Automation database show that the neural network based distance achieves greater accuracy than Hamming distance at every operating point, while adding less than one percent computational overhead.

Keywords: Iris identification, neural networks, hamming distance, biometrics

1. INTRODUCTION

The majority of todays fielded iris identification systems are based on the highly accurate Daugman algorithm [1, 2, 3]. This algorithm first transforms the iris into a small rectangular image using a Cartesian to polar coordinate conversion (called unwrapping) of selected points. Wavelet based features are extracted from this image to form an iris signature (template). One attractive feature of this algorithm is that Hamming distance (bitwise agreement) is used to compare a new iris template to templates stored in a database. Database comparison is typically the largest time expenditure in any iris identification algorithm. Hamming distance computation is extremely fast which makes real-time searches of large databases practical, but Hamming distance may not be the most accurate way to utilize the information within the iris template. Other distance metrics have been proposed [4, 5, 6, 7, 8], and are in use, but many are too computationally intensive to be used in real-time systems containing large databases.

Optimizing the use of the information within the iris is desirable, but a mix of distance metrics may be required to achieve the desired computational speed. Using a fast, simple matching method at the pixel level on individual template regions can reduce the dimensionality of the problem immensely. This dimensionality reduction would produce a quickly computed score on each region within the iris. A more advanced metric can then be applied to the dimensionally reduced data to make improved use of the discriminatory content within the iris template.

A promising identification algorithm has been produced by Ives et al [9]. This algorithm uses an unwrapping method similar to the Daugman algorithm. From the rectangular iris image, directional energy features are extracted to create an iris template. This algorithm also uses Hamming distance as a match metric. No code is publicly available for this algorithm at the time of this writing, but is available to the authors. Since the Daugman algorithm is not publicly available (patented), we used Libor Masek's version of the Daugman algorithm [10, 11] for accuracy comparisons. This code is publicly available and the algorithm is documented in detail.

We used the Chinese Academy of Sciences Institute of Automation (CASIA) iris database [12]. This database, available

*broussar@usna.edu; phone 1 410 293-6131; fax 1 410 293-2215

from the Internet, is composed of images from 108 different eyes, with 7 images of each eye (totaling 756 iris images). These images are all 320x280 pixels in dimension. They are 8-bit bitmapped (.bmp) images, each occupying 92,160 bytes on a hard drive. The CASIA database contains some blurry images, as well as some images where the eye is only barely open such that a great deal of the iris is occluded. Only 662 of the 756 images were used in this research, after 94 poorly segmented images were identified and removed from the database.

It was shown in [7, 13-16] that individual regions within the iris contain varying degrees of discriminatory information. When selecting regions within the iris, annuli are convenient to work with because they allow the use of signature shifting to achieve rotation invariance. Fig. 1 shows the accuracy achieved by the Ives et al directional energy based system [9] using individual regions of the iris as compared to the entire iris. Fig. 2 shows the accuracy achieved by the Masek wavelet based system [10]. Several accuracy measurements are shown for comparison. Best accuracy is defined as the minimum number of incorrect decisions regardless of class. Decidability is a statistical distance between class distributions proposed by Daugman [3] where a higher number denotes better accuracy. As can be seen in each figure, at the majority of operating points, accuracy achieved by individual annulus varies. The goal of this research is to weight and combine this information in a computationally efficient manner to achieve greater accuracy than straight Hamming distance.

Region measured	Accuracy at EER (%)	Best Accuracy (%)	Decidability (Std Devs)
Full template	99.36	99.60	4.79
Annulus 1	83.74	98.73	2.04
Annulus 2	93.38	99.59	3.16
Annulus 3	96.54	99.59	3.92
Annulus 4	97.05	99.58	3.99
Annulus 5	96.93	99.61	3.90
Annulus 6	95.13	99.60	3.51
Annulus 7	94.84	99.60	3.24
Annulus 8	91.80	99.54	2.83
Annulus 9	82.88	97.48	1.95
Annulus 10	75.91	90.59	1.32

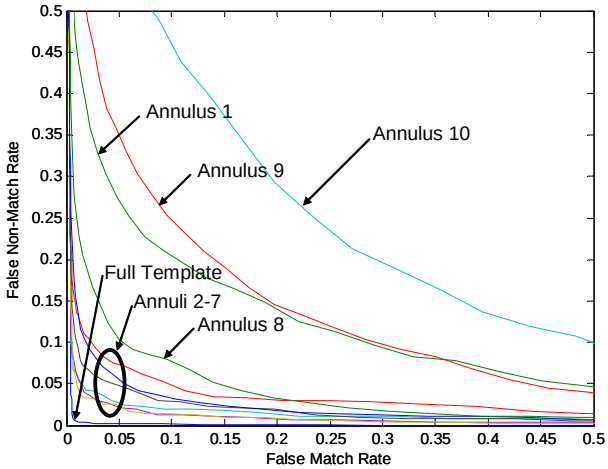

Fig. 1. Accuracy achieved by the Ives et al directional energy based system using individual regions of the iris as compared to the entire iris.

Region measured	Accuracy at EER (%)	Best Accuracy (%)	Decidability (Std Devs)
Full template	99.40	99.63	4.96
Annulus 1	94.82	99.61	3.32
Annulus 2	96.42	99.61	3.76
Annulus 3	96.84	99.59	3.93
Annulus 4	96.91	99.61	3.86
Annulus 5	95.13	99.62	3.45
Annulus 6	92.32	99.58	2.88
Annulus 7	89.50	99.42	2.51
Annulus 8	85.46	98.64	2.10
Annulus 9	80.54	97.01	1.71
Annulus 10	77.13	94.86	1.50

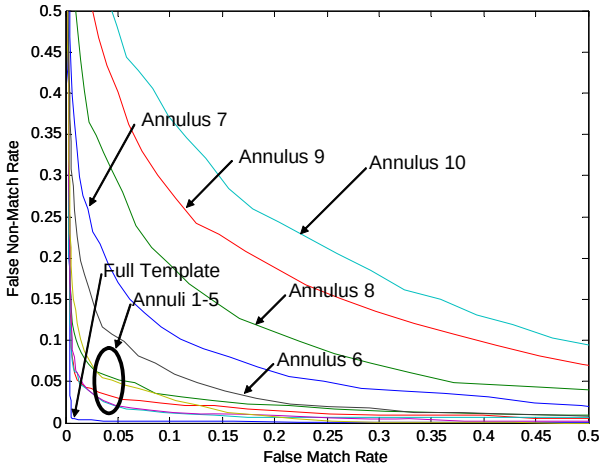

Fig. 2. Accuracy achieved by the Masek wavelet based system using individual regions of the iris as compared to the entire iris.

2. APPROACH

Although fast, Hamming distance equally weights each iris template pixel which may not be the most accurate way to utilize the information within the iris template. Many alternate metrics are too computationally time consuming to be used in real-world applications. Applying a complex distance calculation to each of the thousands of pixels within a template can prove computationally intensive when repeated over millions of templates. The dimensionality of the computation must be reduced if a complex calculation is to be used on a large database. In this research, Hamming distance is used to reduce the computational dimensionality of iris comparisons, followed by a more complex metric to form a match similarity score. Several authors have proposed using measurements taken from individual regions within the iris template as classification features [7, 16]. In this research, Hamming distance is computed on N regions of the iris template, thus reducing the computational dimensionality of the problem to N. Next, a more complex and computationally expensive distance metric is applied to this N dimensional data to form a match decision. This paper investigates the distance metrics used by four common statistical/non-statistical classifiers and compares the resulting accuracy to that achieved using Hamming distance. The highest accuracy method is discussed in detail. Accuracy gains are shown for the Masek, and the Ives et al algorithms. Accuracy results are presented for both clean and noisy images. The effects of the equation parameters on accuracy are shown. Last, computational cost is compared to the cost of using Hamming distance.

2.1 Distance metrics

Hamming distance is currently the standard metric used in much research and in the majority of fielded iris identification systems [1, 2, 3, 9, 10, 16, 17]. Hamming distance is simply an efficient Boolean method for determining the percentage of pixels that mismatch between two iris templates. Each bit in one template is exclusively ORed with the corresponding bit in a second template. The bits that mismatch are counted and divided by the total number of bits in the template. This gives a similarity score ranging from zero to one, with zero denoting all bits match and one denoting no bits match. Often bad or suspect regions within each template are removed from the computations which is a process known as masking.

The use of regional Hamming distance computations as a template preprocessing step reduces the dimensionality of any subsequent distance calculation. If Hamming distance values are computed from N regions of an iris template, they form an N dimensional signature that can be used for classification. Such a signature is ideal for use as input into a statistical or non-statistical pattern classifier for class determination. To avoid a lengthy discussion on classifier basics, the mathematical computation used within the classifier will be presented and the computed value will be used as a similarity score between two irises. Many classifiers (Bayesian, K-Nearest Neighbor, K-Means, etc) determine class membership by combining multidimensional data to compute a distance. A simple Bayesian classifier computes the distance from a class mean, in standard deviations, using the equation

$$D_j = \sqrt{\sum_n \frac{(x_i - \mu_i)^2}{\sigma_i^2}} \tag{1}$$

where D_j is the distance from the jth class, x_i is the ith dimensional measurement of the object to be classified, μ_i is the mean of class j in the ith dimension and σ_i is the standard deviation of class j in the ith dimension. Other classifiers, such as the K-Nearest Neighbor, use the computationally simpler Euclidean distance shown in equation 2

$$D_j = \sqrt{\sum_n (x1_i - x2_i)^2} \tag{2}$$

where $x1_i$ is the ith dimensional measurement of the object to be classified and $x2_i$ is i^{th} dimensional measurement of the is a known object to which it is being compared. The K-Means classifier combines the two previous approaches and computes Euclidean distance from the mean of each class to determine class membership. Equation 3 describes this distance computation.

$$D_j = \sqrt{\sum_n (x1_i - \mu_i)^2} \tag{3}$$

2.2 Neural network distance

Artificial neural networks have been explored for iris identification. Many approaches, such as [18], formulate the identification problem such that the neural network would identify the individual irises (i.e., the neural network contains an output class for each individual iris). This means a neural network that could identify 100 people would have 200 output nodes, each representing an individual iris. This architecture is typically useful for small databases, but requires enlarging and retraining of the neural network as the database grows.

In this research, a neural network is used to identify the statistical pattern present when one iris template matches or does not match another. The neural network can be small (thus fast), and will contain only two output nodes representing a match or a non-match. The neural network will not need to be retrained as individuals are added to the database. A feed forward multilayer perceptron (MLP) artificial neural network, shown in Fig. 3, was used to form the match decision. The error back-propagation training algorithm was used to adjust the internal neural network weights.

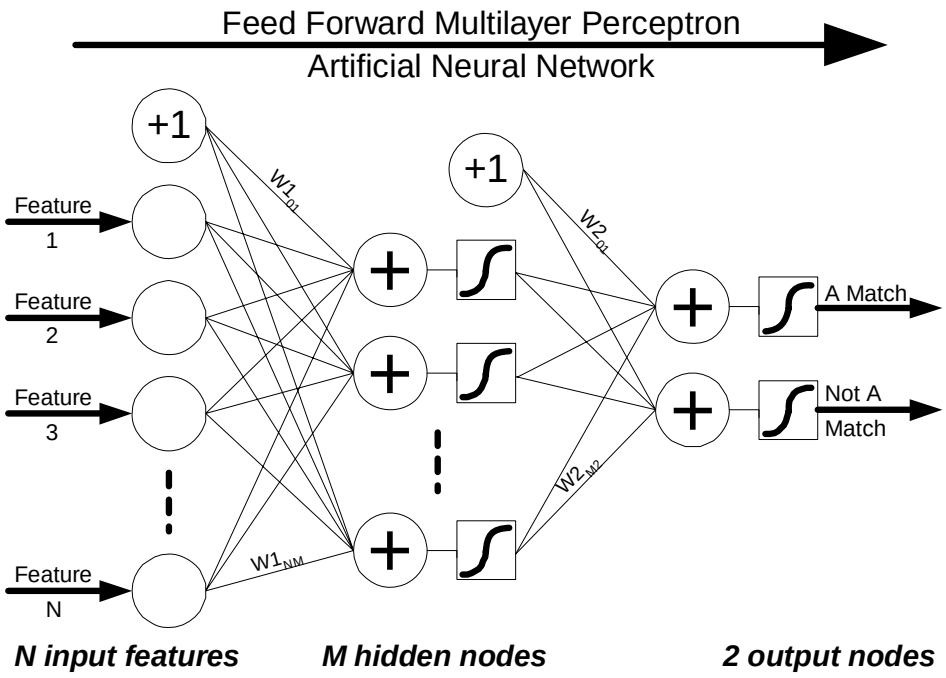

Fig. 3. The feed forward multilayer perceptron artificial neural network architecture used to train weights for use in computing a neural network metric.

Since a typical multilayer perceptron neural network produces a class membership decision as an output, the maximally responding output node represents the class membership of the input pattern. To achieve a more descriptive comparison to the previously described distance metrics, it is desirable that the neural network produce a distance value as an output. For iris template matching, this distance value would represent a similarity score between two iris templates. To achieve this goal, the execution version of the neural network was modified as shown in Fig. 4. One of the output nodes was removed, leaving one remaining output node which represents a similarity measurement. The sigmoid function on the remaining output node was also removed. The purpose of the output sigmoid function was to force and limit the output values to one or negative one. Removing the sigmoid function allowed the neural network to output numerical values that represent the degree to which it determined two iris templates match. This numerical value can be considered a similarity score between two iris templates and will be called the neural network distance. As with the other distance metrics, a smaller number from the neural network metric denotes a greater similarity between two templates. Accuracy results, identical to those presented in this paper, could be achieved by simply training and executing the unmodified neural network. These modifications simply allow the neural network to output a distance measurement that can be directly compared to the other distance metrics using Receiver Operator Characteristics (ROC) curves to demonstrate the added value at multiple operating points.

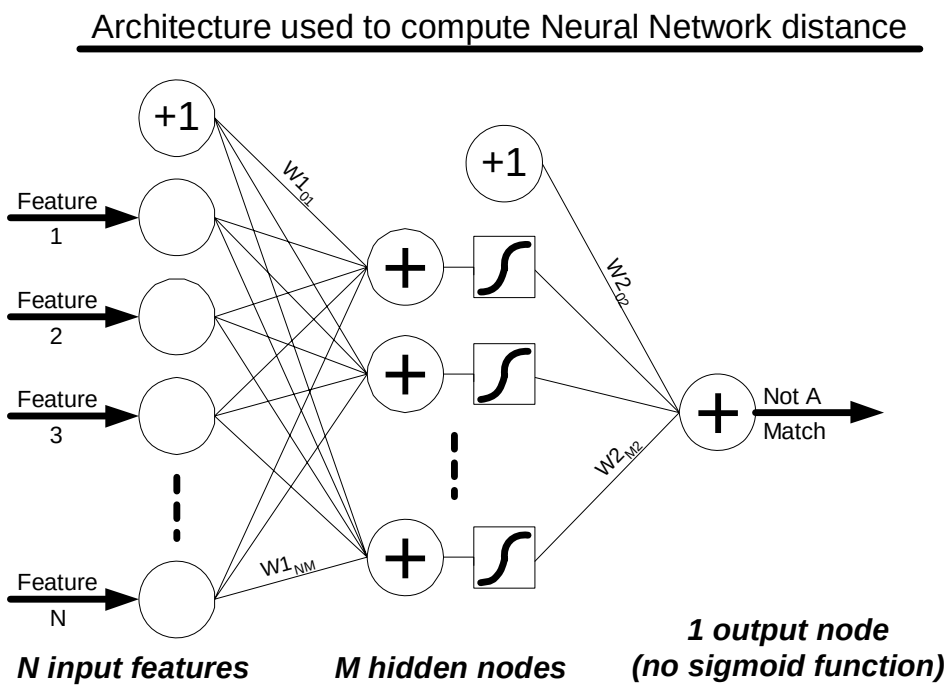

Fig. 4. Artificial neural network architecture used to compute a neural network metric for template matching.

Equation 4 shows the computation for the neural network distance where W_{1j} is the jth weight in the first layer of weights (connected to the input nodes) and W_{2i} is i^{th} weight in the second layer of weights (connected to the output nodes). By definition, the input to the $zero^{th}$ weight in any layer is one. The input features are the Hamming distance scores of individual regions within the iris templates.

$$D_{NNET} = \sum_{i=0-M} W_{2i} \tanh(\sum_{j=0-N} W_{1j} input_j) \tag{4}$$

The unmodified neural network was used when training the internal weights and the modified neural network (using the trained weights) was used to compute a similarity score. During the training phase of a neural network, a training algorithm adjusts the internal connection weights to minimize mean output error over the training set [19]. An artificial neural network can offer a performance advantage over other statistical classifiers if the feature distributions do not match the PDF assumptions used in these classifiers [20]. The error back-propagation training algorithm used to train the MLP traverses an error surface to minimize classification error [19]. The neural network adjusts mathematical hyper-plane boundaries to form a near-optimal discriminate between statistical class distributions in a multidimensional space. In this process, no PDF assumption is made. The neural network learns the statistical distributions of each class during training. A neural network's classification accuracy approaches the accuracy of a Bayes optimal solution [20] and therefore can near-optimally weight template regions, based on the discriminatory content of the individual regions, to form a near-optimal similarity score. Because of this weighting, regions that statistically have greater discriminatory information content will contribute more to the final match decision than regions that statistically have lower discriminatory information content. The training set consisted of $1/100^{th}$ of the template comparisons performed, and the test set consisted of the remaining $99/100^{th}$ of the template comparisons.

3. RESULTS

Fig. 5 and Fig. 6 show a comparison of the various match metrics using the Ives et al Directional Energy algorithm and the CASIA iris database. Each bar in Fig. 5 shows the total accuracy achieved by the Directional Energy templates when the corresponding match metric was applied during template matching. Each template was compared to all other templates for a total of 218,460 comparisons. The chart on the left side of Fig. 5 shows the accuracies achieved at the Equal Error Rate operating point. The chart on the right side of Fig. 5 shows the decidability value of each match metric

based on the statistics of the match and non-match distributions. Fig. 6 shows the ROC curve for each match metric. The ROC curve graph gives best perspective of accuracy gains because it demonstrates achieved accuracy over a range of operating points.

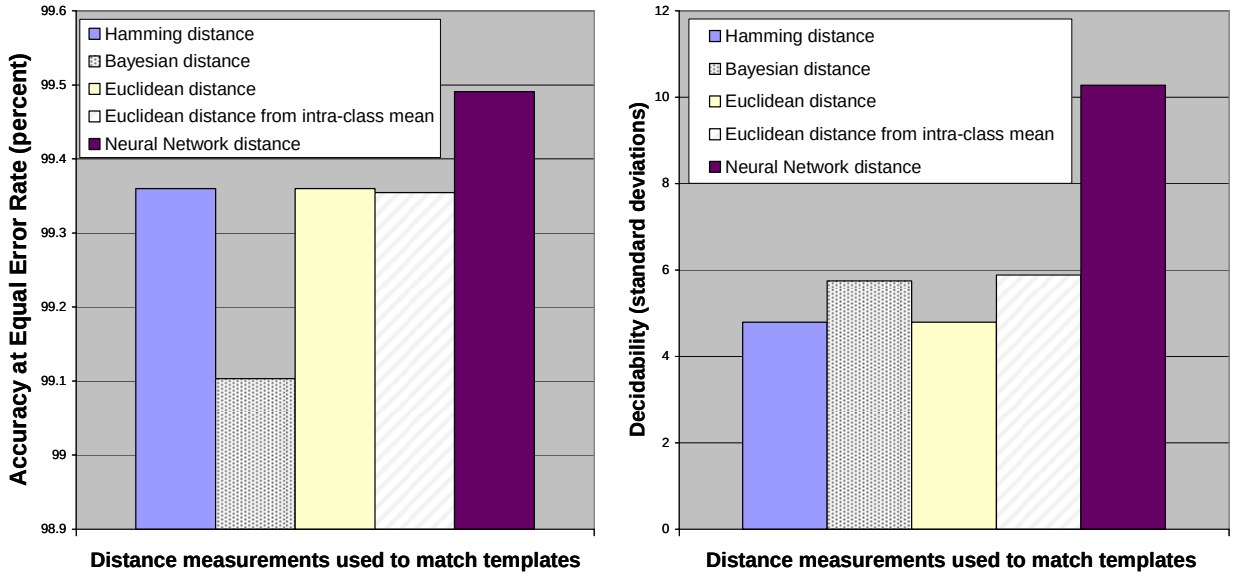

Fig. 5. Accuracy achieved by various similarity metrics as compared to Hamming distance.

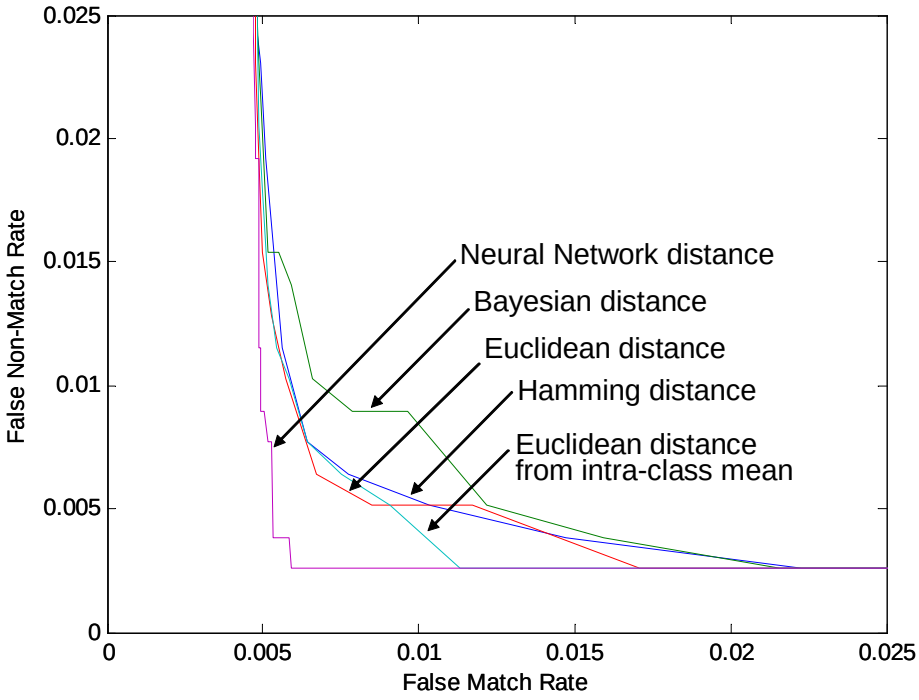

Fig. 6. Receiver Operating Curves of various similarity metrics as compared to Hamming distance.

As can be seen in Fig. 5 and Fig. 6, Bayesian distance degraded accuracy at most operating points. The two Euclidean distance based metrics offered only slight improvement or degradation in accuracy depending on the operating point.

The Neural network distance improved accuracy at most operating points. The operating region near the knee of the ROC curve graph showed significant accuracy improvements. EER, Decidability and ROC results for the Libor Masek algorithm were similar to the results of the Ives et al algorithm in both shape and accuracy. These results are omitted from this paper due to space considerations. The remainder of this paper will focus on the neural network distance and how it compares to straight Hamming distance. ROC curves will be used to present accuracy results. The minimum False Match Rate of 0.005 was due to the relative low resolution and quality of the CASIA database. Similar curves of test performed on higher resolution databases (Bath, etc.) approach a FMR of 0.0001. These results will be presented in future work.

3.1 Impact of number of hidden nodes on accuracy

One issue that arises when using neural networks is the question of data memorization. A neural network with a large number of hidden nodes has the ability to memorize input data points that are statistical outliers. Another issue is computational efficiency. A neural network with a large number of hidden nodes can be computationally expensive to execute over a large database. To address each of these issues, only eight hidden nodes were used to produce the results presented in this research, except for this section. To demonstrate the impact on accuracy, the number of hidden nodes contained within the neural network was systematically varied from two to twenty. First, each iris template was horizontally dissected into ten equal sized regions. Each horizontal region within a template represents an annular region within the iris. Next, the Hamming distance scores of the individual template regions were computed and used as the input to the neural network. Fig. 7 shows the impact that the number of hidden nodes has on identification accuracy. All hidden node configurations outperformed the straight Hamming distance metric. As expected, the trend was an accuracy improvement as the number hidden nodes was increase. Each neural network configuration was trained multiple times to remove some of the random aspects of training a neural network. As can be seen in Fig. 7, six hidden nodes produced virtually the same accuracy as the higher node configurations. This number of nodes is small enough to reduce the likelihood of memorization and offer a reasonable computational efficiency.

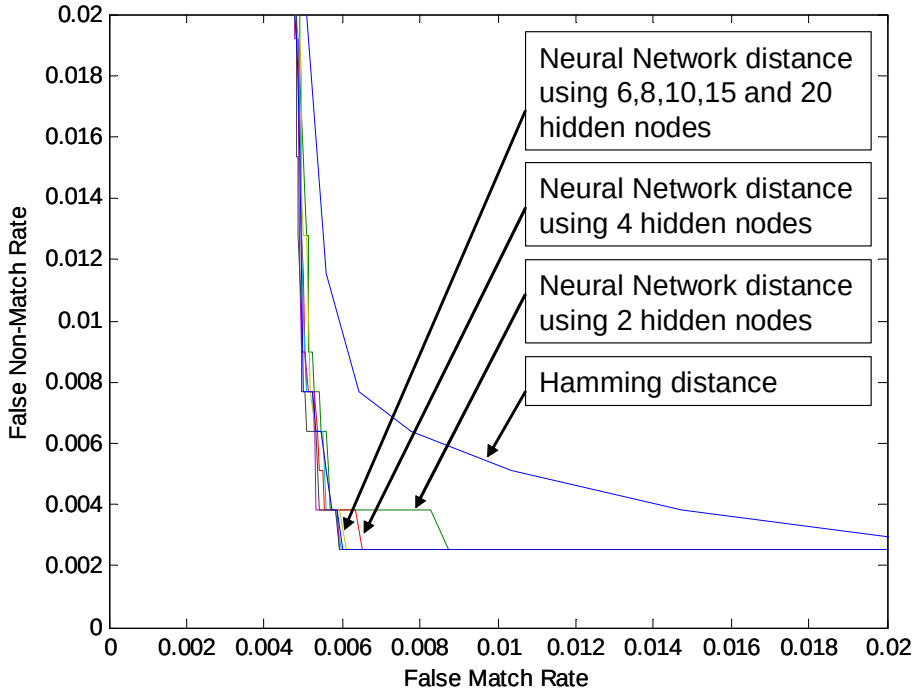

Fig. 7. Receiver Operating Curves showing accuracy achieved, using neural network distance, while varying the number of nodes used in the hidden layer of the neural network. Hamming distance accuracy is shown for comparison.

3.2 Impact of number of template regions on accuracy

Except for this section, all accuracy graphs presented in this paper arbitrarily used the Hamming distance from ten equally sized template regions to compute a neural network distance. The neural network contained one hidden layer with eight hidden nodes. Fig. 8 shows the impact that the number of template regions, used as input to the neural network, has on identification accuracy. The number of template regions was varied from two to twenty in steps of two. Additionally, tests with five and twenty five regions were added since they divided the fifty row template exactly. The entire iris template was used to form each curve; it was simply dissected into a differing number of equally sized horizontal regions. When the number of regions did not exactly divide the template size (i.e., 50 pixels divided into four regions), the region size was rounded up and allowed to overlap a bordering region. As the number of regions was varied, the neural network was trained with a multiple number of hidden nodes and the ROC curve from the best performing configuration was used. Only the ROC curves for the first six tests are presented in Fig. 8 to keep the plot uncluttered. As can be seen in the figure, dissecting a template into multiple regions offers an accuracy gain over computing Hamming distance over the entire template. All test runs outperformed the straight Hamming distance metric, except for the two region input. The accuracy of the neural network metric monotonically increased as the number of regions increased. All test configurations using greater than 10 regions exhibited diminishing returns.

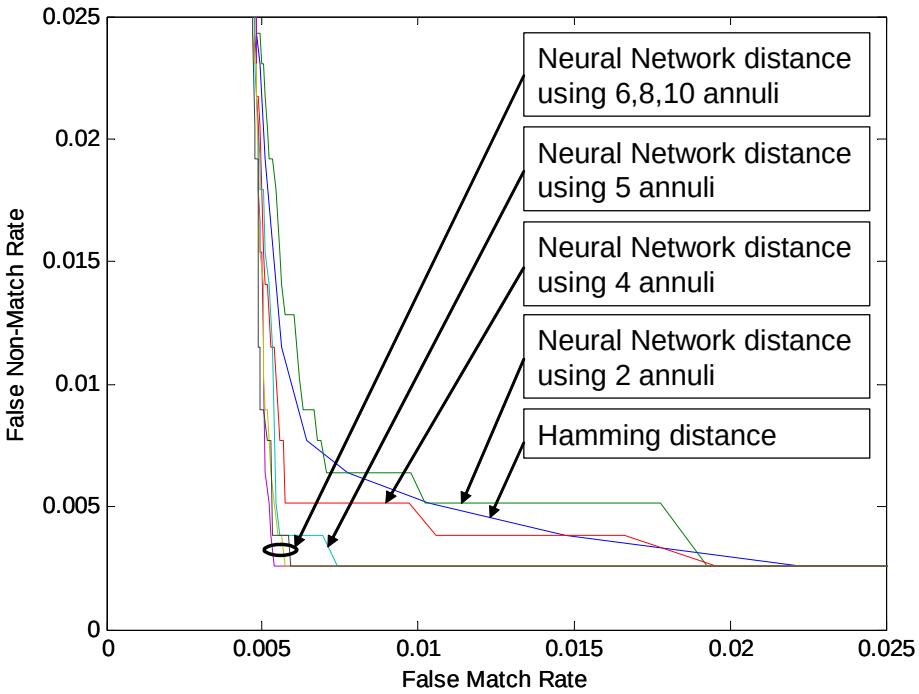

Fig. 8. Receiver Operating Curves showing accuracy achieved, using neural network distance, while varying the number of iris template regions (annuli within the iris) used as input to the neural network. Hamming distance accuracy is shown for comparison.

3.3 Neural network accuracy on templates without eyelid/noise removal

One of the promising aspects of individually weighting regions within an iris template is that areas that often exhibit noise, such as the eyelid and eyelash region, can be de-weighted. De-weighting noisy regions can possibly compensate for poor eyelid/eyelash localization and removal. The neural network distance applies statistically formed weights to each input region to form a weighted similarity metric. The error back-propagation training algorithm de-weights any input regions that do not contribute to a correct result. For the tests performed in this section, the Masek noise mask [6], which performed rudimentary eyelid, eyelash and reflection removal, was used. Both the directional energy and the wavelet based methods were used to test accuracy implications of adding the neural network distance as a similarity

metric. Fig. 9 shows the directional energy algorithm accuracy with and without the addition of the neural network metric. The algorithm was tested with and without the use of the noise mask. All tests showed an increase in performance when the neural network metric was added. The most dramatic marginal accuracy increase occurred when the neural network metric was added to the configurations where the noise mask was not used. The neural network metric accuracy increase approached the accuracy increase achieved by using rudimentary eyelid, eyelash and reflection removal. The neural network metric also increased performance when the noise mask was used.

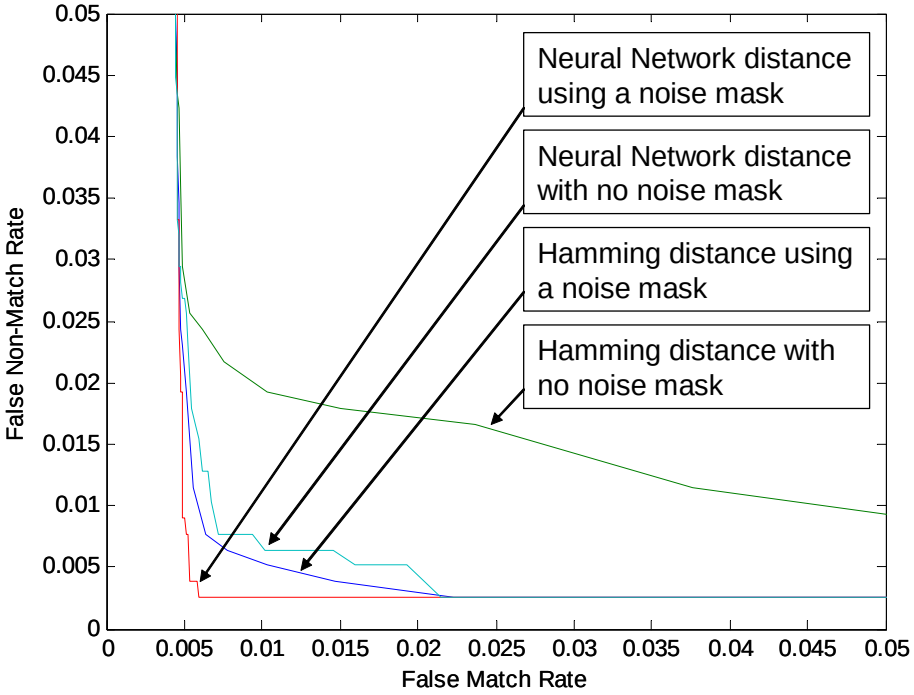

Fig. 9. Receiver Operating Curves generated using the Ives et al directional energy based system, comparing identification accuracy of the neural network distance and Hamming distance. The curves demonstrate accuracy with and without the application of a noise mask. The noise mask performs rudimentary eyelid, eyelash and reflection removal when applied in the template matching process.

Fig. 10 shows the wavelet based algorithm with and without the addition of the neural network metric. This algorithm was also tested without the use of the noise mask. For the results in Fig. 9 and Fig. 10, both algorithms used identical noise masks. All tests showed an increase in performance when the neural network metric was added. Again, the neural network metric accuracy increase approached the accuracy increase achieved by using rudimentary eyelid, eyelash and reflection removal. In both identification systems, the neural network metric also increased performance when the noise mask was used.

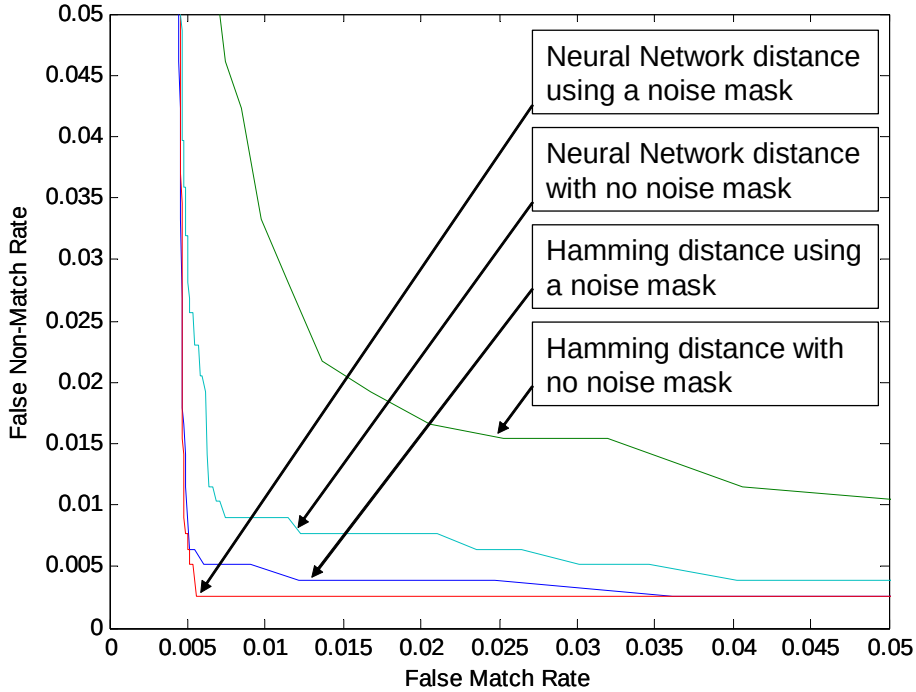

Fig. 10. Receiver Operating Curves generated using the Masek wavelet based system, comparing identification accuracy of the neural network distance and Hamming distance. The curves demonstrate accuracy with and without the application of a noise mask. The noise mask performs rudimentary eyelid, eyelash and reflection removal when applied in the template matching process.

3.4 Computational cost of the neural network metric

As with any post-processing step, the computational cost of applying the neural network distance should be quantified. From equation 4, it can be seen that a neural network distance requires (M+1) additions, multiplications and hyperbolic tangents in the outer summation and (N+1) multiplications and additions within the inner summation. All of these calculations are performed on floating point variables. For an approximation of computational cost, the hyperbolic tangent will be treated as a simple operation equivalent to three multiplications. Though the tangent operation is not computationally equivalent to three multiplications on a Pentium processor, the tangent function could be implemented in a lookup table which would require one multiplication for scaling, casting to an integer and indexing into the lookup table. Equation 5 gives an approximate computational cost for implementing a neural network distance where M is the number of hidden nodes and N is the number of regions used as input to the network.

$$Computations = (M+1) \times 5 \times (N+1) \times 2 \tag{5}$$

The neural network configuration used in the majority of this research contained eight nodes in the hidden layer and the hamming distance from ten regions as input data. Applying a neural network metric with this configuration (M=8 and N=10) imposes an additional computational cost of 990 floating point operations to a Hamming distance based algorithm. The Hamming distance computation alone takes approximately 268800 integer calculations (one bitwise OR for the data and two bitwise ANDs for the masks) for a CASIA size (280 x 320) image. Adding a neural network distance computation adds approximately 0.37 percent overhead to the computational cost of performing a Hamming distance computation. Using the values M=6 and N=6, shown in sections 3.1 and 3.2 to work well, produces a computational overhead of 0.18 percent. The computational cost of the neural network distance is not based on image size, thus the percent overhead will decrease as image size increases. The additional accuracy introduced by the neural network distance has a computational cost of less than one percent additional processing.

4. CONCLUSIONS

Of the four metrics investigated, the neural network metric provided the largest accuracy increase. Using Hamming distance scores of regions within the iris template as input to the neural network distance increased identification accuracy with less than one percent additional processing. All configurations of the neural network distance matched or exceeded the accuracy achieved using Hamming distance alone. The near-optimal weighting, of template regions, performed by the neural network can largely compensate for poor eyelid/eyelash localization and removal.

REFERENCES

1. J. Daugman, "How iris recognition works", *IEEE Trans. on Circuits and Systems for Video Technology.*, Vol. 14, No. 1, pp. 21-30.
2. J. Daugman, "Probing the Uniqueness and Randomness of IrisCodes: Results From 200 Billion Iris Pair Comparisons", *Proceedings of the IEEE*, Vol. 94, No. 11, pages 1927-1935, November 2006.
3. J. Daugman "Demodulation by Complex-Valued Wavelets for Stochastic Pattern Recognition", International Journal of Wavelets, Multiresolution and Information Processing, Vol. 1, No. 1, pages 17-35, January 2003.
4. C.H. Park, J.J. Lee, M.J.T. Smith, K.H. Park, "Iris-Based Personal Authentication Using a Normalized Directional Energy Feature", Audio and video based biometric person authentication conference proceedings, pp. 224-232, 2003.
5. Y.Chen, S.C. Dass, A.K. Jain, "Localized Iris Image Quality using 2-D Wavelets", *Proceedings of the International Conference on Biometrics (ICB'2006)*, Hong Kong, 2006.
6. Shuai Shao, Mei Xie, "Iris Recognition Based on Feature Extraction in Kernel Space", *IEEE Biometrics Symposium 2006*, Baltimore MD, September 2006.
7. R.P. Broussard, L.R. Kennell, R.W. Ives, "Identifying Discriminatory Information Content within the Iris", *SPIE Defense and Security Conference*, Orlando, Fl, March 2008, to be published.
8. G. Gupta, M. Agarwal, "Iris Recognition Using Non Filter-based Technique", Proceedings of Biometrics Symposium, pages 45-47, Arlington, VA, September 2005.
9. R.W. Ives, L. Kennell, R. Broussard, D. Soldan, "Iris Recognition using Directional Energy", *IEEE International Conference on Image Processing*, San Diego, CA, Oct. 2008, submitted.
10. L. Masek, "Recognition of Human Iris Patterns for Biometric Identification," M. Thesis, The University of Western Australia, 2003, www.csse.uwa.edu.au/~pk/studentprojects/libor/LiborMasekThesis.pdf, May 12, 2006.
11. Libor Masek, Peter Kovesi. *MATLAB Source Code for a Biometric Identification System Based on Iris Patterns*. The School of Computer Science and Software Engineering, The University of Western Australia, 2003.
12. CASIA Iris Image Database, http://www.cbsr.ia.ac.cn/Databases.htm, May 12, 2006.
13. Yingzi Du, Robert Ives, Bradford Bonney, and Delores Etter, "Analysis of partial iris recognition", *SPIE 5779: Biometric Technology for Human Identification II*, volume 5779, pages 31-40, 2005.
14. Y. Du et al., "Analysis of partial iris recognition using a 1-d approach", *International Conference on Acoustics, Speech, and Signal Processing*, 961-964, 2005.
15. Raul Sanchez-Reillo and Carmen Sanchez-Avila. "Iris recognition with low template size", *International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 324-329, 2001.
16. H. Proenca, L.A. Alexandre, "Toward Noncooperative Iris Recognition: A Classification Approach Using Multiple Signatures", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 29, pages 607-612, April 2007.
17. E.M. Newton, P. J. Phillips, "Meta-Analysis of Third-Party Evaluations of Iris Recognition", *Systems, Man and Cybernetics, Part A*, SMCA07-08-0235, Submitted.
18. P.E. Merloti, Experiments on Human Iris Recognition using Error Backpropagation Artificial Neural Network", CS553-Neural Networks Final Project, San Diego State University Computer Science Dept., Spring 2004.
19. Cybenko, G. "Approximation by Superpositions of Sigmodial Functions", *Mathematics of Control, Signals, and Systems*, 1982.
20. Ruck, D. W., et al., "The multilayer perceptron as an approximation to a Bayes optimal discriminant function", *IEEE Transactions on Neural Networks*, Vol 4, pp. 296-298, 1990.