Iris Biometrics for Secure Authentication

Abdullah Ali Alshehri, Tyler Daws, and Soundararajan Ezekiel

***Abstract*—Identification of individuals based on behavior or biology is known as biometrics. A recent and widely used development in the field of biometrics is iris recognition for identification. The average iris recognition algorithm requires taking an image of the iris, then testing and segmenting the image. This is commonly achieved through both statistical and classical techniques. Every method has strengths and weaknesses. In this proposal, we present a novel iris recognition approach. The objective of our study is to create a technique for accessing computers and computer networks. Due to widespread use of computers in many forms (desktop, handheld, etc.) our technique offers valuable security for information. This wavelet-based algorithm will be capable of enhancing the scanned image, reducing noise, and extracting important elements of the picture to check against data within a database. Additionally, our method may be generalized to a variety of applications including surveillance, e-commerce, ATM transactions, and others.**

***Index Terms*—DCNN, biometrics, multi-resolution transform, wavelet, image fusion.**

I. Introduction

In recent years, with the continued proliferation of biometric signatures used for security and authentication, the necessity for robust and accurate techniques has risen as well. Contemporary applications of biometric security such as fingerprints, veins, and iris scanning allow for a unique signature for everyone that cannot be stolen or phished for like traditional passwords. Traditional methods of iris recognition involve capturing an image of the iris upon an authentication attempt and segmenting the iris into regions, which are then compared to a database reference, usually through statistical techniques [1]-[4]. The cornea, ridges, and veins and other textures on the iris itself are specific to the individual and as such valid to be used as a unique biometric authenticator [5]. However, one of the most significant issues with biometric security, and particularly iris recognition, is false results, especially due to variations in the actual capturing of the iris image. Due to variations in the scanning itself, the probability that an iris captured during authentication will exactly match its database reference is near zero, simply due to the discrepancies such as the presence of noise, different angles, etc. In this paper, we investigate a method of biometric security which utilizes multi-resolution transformation image fusion techniques such as wavelets [6]-[8], which are especially adept at extracting key features of images and combining them into a single image, along with machine learning techniques, namely the heterogeneous fusion of deep convolutional neural networks [9], to create a single, unique feature vector signature based on multiple images, rather than a single reference image alone being used as a database reference.

Multiple iris images are first decomposed into their multi-resolution coefficients using various techniques such as wavelets, contourlets, and curvelets, the latter of which are particularly proficient in capturing the smooth curves and contours that are found in an iris. The decomposed coefficients are then denoised to enhance the actual features of the iris and then fused using image fusion techniques [10]-[12] to create a single set of coefficients. The denoised coefficients are then reconstructed, with one reconstructed image for each type of transformation. The reconstructed images are then used as input for pre-trained deep convolutional neural networks, however rather than being used for classification purposes, the unique features of the iris are extracted from the penultimate fully connected layer. This feature vector can then be stored in a database and used as a unique identification signature, rather than an image of an iris itself. When an iris is scanned for identification, the process is repeated without the image fusion techniques, resulting in a feature vector which can then be directly compared to the database reference within a tolerance of error.

The remainder of the paper is organized as follows: Section II outlines the conceptual and technical background for the multi-resolution transformations, statistical, and machine leaning techniques utilized in this paper. Section III explains the methodology and processes used in this study. Section IV shows the experimental results of our methodology, along with the identification accuracy of our proposed method. Lastly, Section V discusses the significance of the results and describes future work in the area.

II. Technical Background

A. Deep Convolutional Neural Networks

Alexnet is a convolutional neural network model originally designed by Alex Krizhevsky for the 2010 ImageNet Large Scale Visual Recognition Challenge [13]. The network was able to classify over 1.2 million images with top-5 and top-1 error rates that significantly outperformed previous models, at a rate of 17% and 37.5% [14]. The novelty of Alexnet came from the integration of GPUs for the computationally intensive training process, reducing the training time. Alexnet is comprised of eight layers total, five convolutional

Manuscript received November 25, 2019; revised March 11, 2020.

A. Alshehri is with the Electrical Engineering Department, King Abdulaziz University, Jeddah, KSA (e-mail: aaashehri@gmail.com).

S. Ezekiel and T. Daws are with the Computer Science Department, Indiana University of Pennsylvania, Indiana, PA USA.

doi: 10.18178/ijmlc.2020.10.3.961

and three fully connected layers, designed for CUDA architecture as a way of leveraging GPUs for the training process.

Alexnet opened a new avenue of investigation for neural networks in which the depth of the layers was studied. The University of Oxford's Visual Geometry Group began developing two architectures, VGG16 and VGG19 in 2015. Like Alexnet, both models leverage ImageNet's dataset for the actual training, and the main difference between the two networks is the number of hidden layers, 16 and 19 respectively. By adding more layers to the network, it allows it to extract more features from input images, however adding more layers to an architecture also involves longer computational times, especially regarding training. Despite the increasing computational time that comes with high complexity, both VGG16 and VGG19 are generally able to provide higher classification accuracy than AlexNet.

DCNNs are generally comprised of a feedforward, stacked layer architecture in which data is input into the first layer and is fed forward through each of the layers, with the output of the final layer being the actual classification. After the convolutional and pooling layers which are responsible for the actual feature extraction, the Fully Connected (FC) Layers are the final layers of the network and are responsible for high-level reasoning and classification. The input of the FC layers are the outputs and weights of the previous layers that representing the extracted high-level features' activation maps. The output of this is the probabilities of each of the categories by correlating the features to each of the classes, stored as a vector. The FC7 layer is the penultimate layer of the network, containing the extracted features and weights of all the classes. The weights contain the actual high-level correlations that are used for classification, making these layers useful for feature extraction.

B. Wavelet

A wavelet is defined as a finite oscillation, like a wave, that has an average value of zero. Moreover, a wavelet starts with an amplitude of zero then the amplitude will oscillate a finite number of times before ending with a final amplitude of zero. For a function such as, $\Psi(x)$, to be considered a wavelet, the following two conditions equations 1 and 2 must hold:

$$\int_{-\infty}^{\infty} \psi(x)\,dx = 0 \tag{1}$$

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\,d\omega = C_{\psi} < \infty \tag{2}$$

where $\Psi(\omega)$ is the Fourier transform of the selected wavelet function and C_{Ψ} is the *admissible constant*. Numerous wavelets have been constructed, most derived by Daubechies in ten lectures on wavelets [15]. Each can be categorized according to whether the wavelets are defined on a discrete grid vs. over continuous time or space, and whether they are real vs. complex valued. Two fundamental characteristics of wavelets are their rescale and translation. Given a mother wavelet $\Psi(x)$, an entire family of wavelets $\psi_{j,k}(x)$ is defined in equation 3:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{|j|}}\psi\left(\frac{x-k}{j}\right) \tag{3}$$

where j is the scaling variable and k is the translation variable. The rescale and translational characteristics of the wavelet allows for detection of small, abrupt changes in signals. This makes it an ideal transform for point-wise edge detection. The continuous wavelet transforms (CWT), defined as the inner product of a function $f(x) \in L_2(\mathbb{R})$ and a wavelet $\Psi(x)$, is expressed in equation 4:

$$f,\psi_{j,k} = \int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{|j|}}\psi\left(\frac{x-k}{j}\right)dx \tag{4}$$

for image processing, the function f would represent an image that has the wavelet transform applied to it. However, images are not generally processed as continuous-space functions, but rather as sampled discrete space functions. As a result, the discrete wavelet transforms (DWT) is generally used to process sampled images. Like the CWT, the DWT of the function f, which we will denote G_{ψ}, is be expressed in equation 5:

$$G_{\psi}(f,\psi) = \frac{1}{\sqrt{M}}\sum_{i,j=1}^{n} f(x)\psi_{j,k}(x) \tag{5}$$

where M is a scaling weight it is noted that there are some drawbacks as a result of moving from a continuous to discrete transformation. In the discrete domain, the wavelet transform loses directionality and shift-invariance. Hence, the wavelet can detect image edges, subsequently it does not see an entire contoured edge as one connected piece. The lack of shift-invariance refers to the discrete wavelet's inability to transform shifted versions of the function f in the time domain as shifted versions of G_{ψ} in the wavelet domain. These two shortcomings set in motion the creation of other multi-resolution transforms which attempt to outperform the DWT.

C. Fusion Methods

Pixel-level image fusion is a process whereby a single composite image is produced based on two reference images [16]. Specifically, by fusing multi resolution coefficients obtained from two different images, a newly fused image can be reconstructed. Many methods for multi-resolution image fusion have been advanced. Each are distinguished by different methods for merging approximation and detail coefficients. For example, in "max-min" fusion, the max criterion is used for the approximate coefficients and the min criterion for the detail coefficients. This is the primary notation used for all the fusion methods, except for those of the bandelet. Since the bandelet only has detail coefficients, only one fusion criterion can be used in that framework. In the following sections we will let F be the fused image, A be the first input image and B be the second input image, where $f_{i,j}$ are elements of form matrices.

The max criterion takes the maximum absolute value of each entry between the two matrices seen in equation 6:

$$f_{ij} = \begin{cases} a_{ij} & if\ |a_{ij}| \geq |b_{ij}| \\ b_{ij} & otherwise \end{cases} \tag{6}$$

Absolute values are essential in these decision statements

considering pixel values cannot be negative, however, multi-resolution coefficients can.

The min criterion takes the minimum absolute value of each entry between the two matrices in equation 7:

$$f_{ij} = \begin{cases} a_{ij} & if \left|a_{ij}\right| \le \left|b_{ij}\right| \\ b_{ij} & otherwise \end{cases} \quad (7)$$

The linear method requires a linear combination of the two matrices with each coefficient summing to one. If the constant parameter c is selected on the interval (0,1) then the resulting coefficient matrix will have values as represented in equation 8:

$$f_{ij} = c_{ij}a_{ij} + \left(1 - c_{ij}\right)b_{ij} \quad (8)$$

The mean method is a special case of the linear method where $c = 0.5$. It will take the mean between the two entries at each position in the matrix as shown in equation 9:

$$f_{ij} = \frac{a_{ij} + b_{ij}}{2} \quad (9)$$

Principal Component Analysis (PCA) is method that can be used to reduce the dimensions of matrices while keeping most of the information. The correlated values in the larger matrix are reduced into uncorrelated variables known as principal components. For image fusion, PCA can be used to pick out the most important features of an image and then fuse them. PCA was deemed as the most effective fusion method, being that it provides the most important features of an image. While the other methods are still viable, PCA fits the study at hand best. By implementing this method of fusion, the images produced consist of most important features, providing a more accurate database for the DCNNs to utilize. Consequently, this will decrease overall error produced.

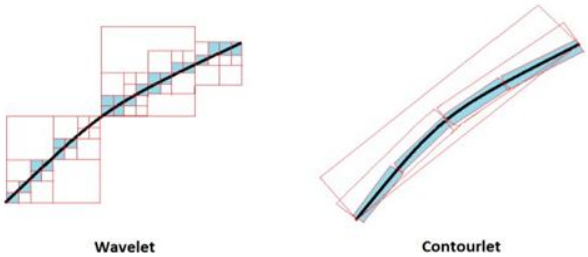

Fig. 1. Comparison of Wavelet and Contourlet.

D. Contourlet Transform

The contourlet transformation was originally proposed by Minh N. Do and Martin Vetterli in 2002 in response to the directional and anisotropic constraints of the wavelet and curvelet [17]. Contourlets are an improvement on the wavelet transformation as wavelets are applicable for following one-dimensional smooth signals [18]. The contourlet includes basis elements that cover more directions than the standard wavelet of just horizontal, vertical, and diagonal. The contourlet can effectively represent images by combining five desirable properties: anisotropy, directionality, multi-resolution, locality, and the ability to process sampled data. A major advantage of the contourlet transform compared to the curvelet transform is that the contourlet transform was specifically developed for the discrete domain. The contourlet transform provides a sparse representation for two-dimensional piecewise smooth signals that represent images. Fig. 1 shows the comparison between wavelet and a contourlet.

The basic two-dimensional wavelet transform provides multi-resolution and localization features but lacks the ability to efficiently model local edge direction and curvature. The multi-resolution transforms such as contourlet and curvelet provide a solution for this disadvantage. The contourlet is composed of two processing stages: a Laplacian pyramid (LP) and a Directional Filter Bank (DFB). In order to extract contourlet decomposition, the image has two transformations applied to it. The Laplacian Pyramid and the Directional Filter Bank produce the orientation edge components of a region within an image.

E. Laplacian Pyramid

The LP transform resembles a Fourier transformation, however while the Fourier transform is a function of real variables the LP is a function of complex variables. At each level the LP transform decomposes the image into a lowpass down sampled version of the original with the difference between the original and the prediction resulting in a bandpass image. In each level of a LP decomposition, only one bandpass image is generated [19]. Fig. 2 shows a standard LP operation on an image.


Fig. 2. Example of LP applied to Lena Image 4 Layers.

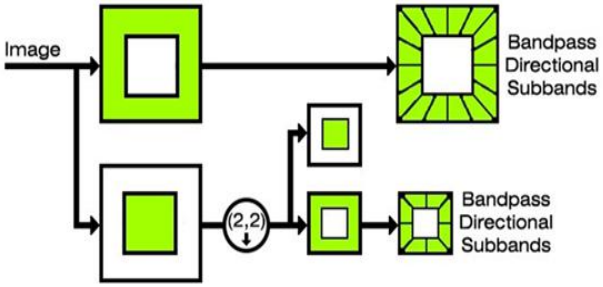

Fig. 3. Contourlet Filter Bank.

F. Directional Filter Bank

In order to generate the smooth contours of an image, a filter bank must be applied. A multiscale decomposition into octave bands using the LP is computed, followed by a DFB being applied to each bandpass channel as shown in Fig. 3. The contourlet transformation is a multiscale decomposition

that removes low-frequency due to the non-separable combination of the LP and DFB. The combination can also be a pyramidal directional filter bank (PDFB). The DFB identifies the directional information of the contours of an image, the results can be reconstructed into the image. The contourlet transform makes use of a discrete domain and constructs a multiresolution, local, and directional expansion of the image by utilizing contour segments.

The PDFB decomposition has the conservation function for an image *x* of:

$$\| x \|^2 = \sum_{j=1}^{J} \| d_j \|^2 + \| a_j \|^2 \tag{10}$$

where d_j are the directional coefficients and a_j is the low-pass image for bandpass images $j = 1, 2, ..., J$.

G. Performance Metrics

A multitude of metrics are available in order to test the performance of the fused image when compared to the originals. These methods include image structural similarity based, information theory based, image feature based, and human perception based [20]-[22]. Due to the uncertainty of human perception-based metrics, only the first three types will be used. As an example, the normalized mutual information metric:

$$QMI = \left[\frac{MI(A,F)}{H(A)+H(F)} + \frac{MI(B,F)}{H(B)+H(F)} \right] \tag{11}$$

would fall under the information theory-based category. For each metric, various methods of image fusion are used such as the mean, max, min, principal component analysis (PCA), and weighted average and data sets are collected from each. The performance metrics being used are Cvejie, Piella, Normalized Mutual Information (MI), Tsallis entropy, Multiscale, and Spatial Frequency (SF). In order to test the integrity of the fused image to the original images, Spatial Frequency Ratio (SFr) and Structural Similarity Index Metric (SSIM) will also be used. The methods can be used for image quality assessment.

H. Curvelet

As an extension of the wavelet transformation, the curvelet is like methods such as the double density wavelet and the dual tree wavelet. The difference between these transformation methods and the curvelet however is that the curvelet takes advantage of the ridgelet transformation [24] rather than a wavelet transformation. Furthermore, the curvelet can detect smooth curved edges within an image. Ridgelets exploit functions like wavelet to better represent smooth edges but lack the significant local information. They do however provide scale and translation characteristics, as well as orientation characteristics which assists in smooth edge data.

As stated, before the ridgelet transformation is unable to be applied globally. This led to the plan of partitioning the given domain, followed by adding the ridgelet transformation to each section. This plan led to the curvelet transformation. This transformation works in the frequency domain and uses polar coordinates to split it into sections. Then, using a series of concentric circles, these sections are further split into wedges. The relationship between the number of wedges, N, associated with one of the concentric circles is:

$$N_j = 4 \times 2^{\frac{j}{2}} \tag{12}$$

where j corresponds to the number of concentric circles. See the Fig. 4 below for an example of how the frequency domain is tiled.

Through a bandpass filtering of multiscale ridgelets is a curvelet, which can be defined as:

$$\psi_{a,b,\theta}(x) = a^{\frac{-3}{4}} \psi\left(M_a R_\theta (x-b)\right) \tag{13}$$

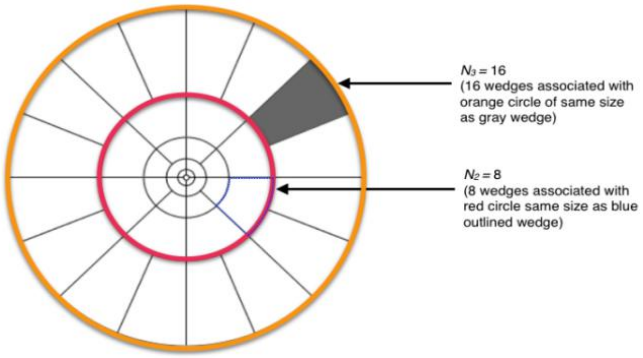

Fig. 4. Continuous curvelet tiling.

where ψ is a wavelet function, a is a scaling variable between zero and one, b is a translation variable, and Θ represents an orientation. Furthermore, R_Θ represents a rotation and M_a is the following scaling matrix:

$$M_a = \begin{bmatrix} 1/a & 0 \\ 0 & 1/\sqrt{a} \end{bmatrix} \tag{14}$$

The aspect ratio of each wedge is approximately the width equal to the length squared. The curvelet transform itself is a map from $L^2(\mathbb{R}^2) \to F$ where F is the frequency domain and L is the Laplacian operator. The transform itself has a similar structure to the wavelet transform and is expressed as

$$\psi_{a,b,\theta}, f = \int_{\mathbb{R}^2} \psi_{a,b,\theta}(x) f(x) dx \tag{15}$$

where f is the function to decompose the data (generally an image) into components. Unfortunately, for image processing a discrete version of the transform is needed, while it is currently defined in a continuous domain. For the curvelet transform being applied to a sampled image, *f*, the frequency domain cannot be tiled using concentric circles, and concentric squares are used instead. By using concentric squares, the sizes of the corresponding wedges are slightly different allowing the orientations and rotations to no longer reflect the physical scene or natural representations. Hence, in place of rotation, shearing operators are used to remedy the discrete curvelet transform's performance capabilities. The discrete curvelet transform is represented as: [25]-[27]

$$F_\psi(f, \psi) = \sum_n f(n) \psi_{a,b,\theta}(n) \tag{16}$$

The curvelet transform above is representative of a

directional multi-resolution transform; however, it was not widely used the field of image processing, as it is more naturally suited to continuous-space signals, rather than sampled signals. To remedy this shortcoming the contourlet transform was developed.

I. Performance Metric

The removal of "irrelevant visuals" causes a loss of real and or quantitative image information. Due to the loss of information, a process of quantifying what information is loss is required. There are two criteria for such assessment: objective fidelity criteria and subjective fidelity criteria. If information loss can be expressed by a mathematical function between the input and output of a compression process, it is based on the objective fidelity criteria. An example is root-mean-square error (RMSE) between two images. When representing a given input image each pixel is representing by $f(x,y)$ and a given output image is represented by $f(x,y)$. For any given value of x and y, the error $e(x,y)$ is between $f(x,y)$ and $f(x,y)$.

$$e(x,y) = f(x,y) - f(x,y) \tag{17}$$

The total error between two images is:

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y) - f(x,y)] \tag{18}$$

where $M \times N$ represents the size of the given images. The RMSE, e_{rms}, between $f(x,y)$ and $f(x,y)$ is then the square root of the squared error averaged over the $M \times N$ array.

$$e_{rms} = [\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y) - f(x,y)]^2]^{1/2} \tag{19}$$

The root-mean-square error (RMSE) is used as ground truth to objectively evaluate each metric. A low value means that the error in the image is low while a high value means there is more error in the image when compared to the original. Values are compared to the metric values being evaluated for each fusion method in order to evaluate the metric's effectiveness for each method. The goal is to determine a correlation between the ground truth measure and the image fusion metrics in order to judge RMSE's effectiveness [28].

Human visual perception is designed to assign high importance to structural information Image, structural similarity-based metric (SSIM) makes use of such findings. This metric quantifies the loss of this information as an approximation of the image blurring. A structure of this method given two images A and B is shown in equation 20 which can be expanded to equation 21.

$$SSIM = [l(A,B)]^{\alpha}[c(A,B)]^{\beta}[s(A,B)]^{\gamma} \tag{20}$$

$$SSIM = (\frac{2_{\mu A \mu B} + C_1}{\mu_A^2 + \mu_B^2 + C_1})^{\alpha}(\frac{2_{\sigma A \sigma B} + C_2}{\sigma_A^2 + \sigma_B^2 + C_2})^{\beta}(\frac{2_{\sigma AB} + C_3}{\sigma_A^2 + \sigma_B^2 + C_3})^{\gamma} \tag{21}$$

where μ_A and μ_B are the average values of images A and B respectively, σ_A, σ_B and σ_{AB}, the variance, and covariance respectively [29]. $l(A,B)$, $c(A,B)$ and $s(A,B)$ in formula 20 are the luminance, contrast and correlation components. The parameters are used to adjust the weight given to any of the components and the constants are included to avoid the divide by zero problem. Using the SSIM definition, Piella and Heijmans defined three fusion metrics. Only two of the three are employed in this study, Piella's metric and Cvejie's metric.

III. Methodology

The preliminary step in our study involved the necessary step of collecting and developing a suitable database of irises as a means to verify the real-world application of our methodology. Iris recognition applications have clearly highlighted two challenges, i.e. usability and scalability. As such, we developed the CASIA Iris Image Database (CASIA-Iris), containing sets of iris images captured to account for any variations in angles and other factors that would be present in real world applications.

The CASIA Iris Image Database Version 1.0 (CASIA-IrisV1) includes 756 iris images from 108 eyes. For each eye, 7 images are captured in two sessions with our self-developed device CASIA close-up iris camera. In order to protect the IPR in the design of our iris camera (especially the NIR illumination scheme), the pupil regions of all iris images in CASIA-IrisV1 were automatically detected and replaced with a circular region of constant intensity to mask out the specular reflections from the NIR illuminators in order to see the effect of the synthetic data on untouched labels. Iris images of CASIA-Iris-Interval were captured with a self-developed close-up iris camera. The most compelling feature of our iris camera is that we have designed a circular NIR LED array, with suitable luminous flux for iris imaging. Because of this novel design, the iris camera can capture very clear iris images (Fig. 5).

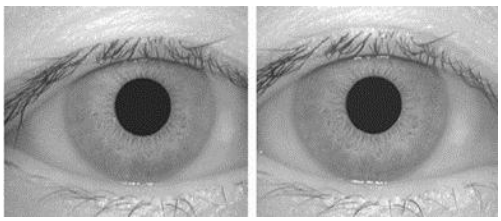

Fig. 5. Example iris images in CASIA-Iris-Interval.

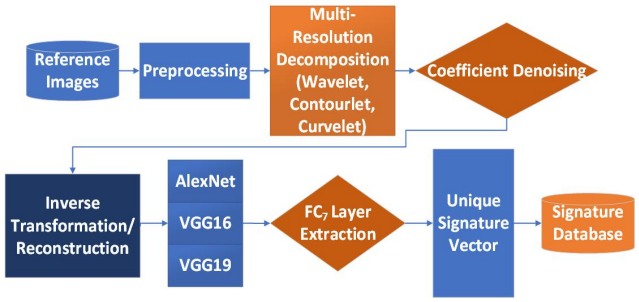

Fig. 6. General encoding methodology.

The general methodology for the study is divided into two overarching procedures, outlined in Fig. 6 and Fig. 7, an

encoding process and an authentication process. Encoding involves creating a unique individual signature feature vector using multiple iris images and registering it to a database, while the authentication process is essentially the same, without the fusion and with an added step for verification. Multiple iris samples are first captured from the same eye, for our purposes using an iris dataset containing three samples each from the same eye. Each image is then decomposed into its respective multi-resolution detail coefficients, which capture the directional coefficients and smooth contours and curves that are prevalent in irises. The coefficients are then thresholded and denoised in order to extract the key features of the iris and remove any outlier data that might affect the quality of the image, to ensure that only the significant features of the iris are used for verification.

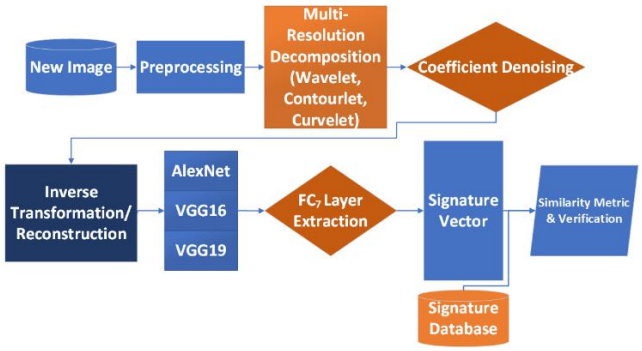

Fig. 7. Authentication methodology.

Algorithm 1

```
procedure ENCODEIRIS
    procedure IRISFUSION(irisSet)
        detailCoefficients ← MULTIRESDECOMP(irisSet, wavelet, contourlet, curvelet)
        denoisedCoefficients ← COEFFICIENTTHRESHOLDING(detailCoefficients)
        fusedCoefficients ← PCA(denoisedCoefficients)
        reconstructedIrisSet ← INVERSETRANSFORM(fusedCoefficients)
    procedure DCNNFUSION(reconstructedIrisSet)
        featureVectors ← FC7EXTRACTION(reconstructedIrisSet, AlexNet, VGG16, VGG19)
        signatureVector ← FEATUREFUSION(featureVectors)
```

The decomposed directional coefficients are then fused using various image fusion techniques, such as maximum, average, and PCA, creating a single set of coefficients from the three images. The fused coefficients are then reconstructed using their respective inverse transformations, creating a fused wavelet, contourlet, and curvelet image each capturing different key directional features from the iris. The images are then used as input for pre-trained AlexNet, VGG16, and VGG19 networks, but rather than being used for classification, the feature vectors containing the most significant details are extracted from the penultimate FC_7 Layers. The FC_7 layers are then fused using heterogeneous DCNN fusion to create a single feature vector, which is then stored and used as reference for authentication. The authentication process is similar, forgoing the image fusion step and instead including a verification process within a tolerance of error to compare it to the stored database reference. The multi-resolution decompositions are still utilized to denoise the iris image and capture significant directional information, still resulting in one reconstructed image for each transformation. The reconstructed images are then used as input in the neural networks and heterogeneously fused, creating an individual feature vector which can then be compared to the database reference.

Algorithm 2

```
procedure SIGNATUREVERIFICATION
    procedure MULTIRESDENOISING(irisSet)
        detailCoefficients ← MULTIRESDECOMP(irisSet, wavelet, contourlet, curvelet)
        denoisedCoefficients ← COEFFICIENTTHRESHOLDING(detailCoefficients)
        reconstructedIrisSet ← INVERSETRANSFORM(fusedCoefficients)
    procedure DCNNFUSION(reconstructedIrisSet)
        featureVectors ← FC7EXTRACTION(reconstructedIrisSet, AlexNet, VGG16, VGG19)
        signatureVector ← FEATUREFUSION(featureVectors)
    procedure IRISAUTHENTICATION(signatureVector, referenceVector)
        similarityScore ← RMSE(signatureVector, featureVector)
        if (similarityScore + ϵ) ≥ authenticationThreshold then
            RETURN positiveAuthentication
```

IV. Results

Two different sets of irises were tested for our study as a means to study not just the effects multi-resolution transformations had on the integrity of an image but also to identify any factors that could cause the similarities between two different irises to appear to rise. Upon performing the multi-resolution transforms and reconstructing the fused images, the SSIM scores, exampled in Table 1, indicated that although the fused images do generally have high scores between each other, they can occasionally have low similarities between the same eyes, even more so than between different eyes.

TABLE I: Sample SSIM Scores between Two Irises

	Contour1	Wavelet1
Contour1	1	0.92
Curvelet1	0.85	.94
Wavelet1	0.92	1
Contour2	0.73	0.74
Curvelet2	0.63	0.63
Wavelet2	0.79	0.80

The SSIM scores indicated that while the images fused using multi-resolution transforms generally were structurally similar to matching eyes, there were also cases where differing eyes actually had higher SSIM scores after being transformed and fused. The same process was applied to both images sets, creating a train set of fused images and a test set that consists of a single training image which has been denoised using the three transforms but not fused, resulting in three images for both sets.

The images were used as input for the three neural networks, AlexNet, VGG16, and VGG19. Before classification, the feature vectors of the sets were extracted, resulting in three feature vectors for each set. The feature vectors extracted from the fused training images represent the reference signature, which would then be stored into a database for later authentication. The test set in this case, represents a single iris presented for authentication, still having the multi-resolution transforms done for the purpose of denoising and feature extraction. Interestingly, the SSIMs between the matching iris sets were generally lower than the SSIM of simply comparing the reconstructed, fused images, however the risk of false positives also appeared to be much lower, with the differing iris sets having much lower SSIM scores than the fused reconstructions alone.

TABLE II: SSIM Scores between FC_7 Layers

	Fused1	Denoise1
Denoise1	0.74	1
Fused2	0.52	0.37

The results appear to show promise in using extracted feature vectors for the biometric authentication. While it is apparent that overall the SSIM scores between the matching pairs were lower, the difference between a denoised iris used for authentication and the fused, reference signature was only .37, indicating a much higher difference between the feature vectors than simply between the multi-resolution denoised irises.

V. Conclusion and Future Work

Using extracted FC7 layers as a mean for signature verification shows promise, however it appears more work must be done in terms of fine-tuning the multi-resolution transformations and image fusion steps. The similarity scores between the extracted FC7 layers were found to be much lower than the fused irises, indicating that the neural networks were able to extract significant, latent features from the irises and weight them in such a way that eyes with different features possibly have much different scores. Although differing pairs of eyes had consistently lower similarities when comparing the feature vectors to the irises, more work needs to be done to improve the similarity scores between the signature vector and a single, matching eye like in cases of verification. Although the signature vectors for matching irises had consistently higher similarity scores, they were still low overall possibly indicating errors occurring as a result of the image processing step which would need to be further refined. Future work would involve implementing a full database system for signature vectors to be stored and then attempting to measure the execution time of our method in real-time authentication, processing a single iris and comparing it to a stored signature. Additionally, improving the similarity scores between the fused images and a single iris or determining a proper tolerance of error for authentication to limit the possible of false negatives would also be necessary.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

T. Daws conducted the research with guidance of Professors S. Ezekiel and A. Alshehri

Acknowledgment

This work was funded by Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (829-104-D1435). The authors, therefore, acknowledge with thanks DSR technical and financial support. The authors would like to extend their gratitude to Indiana University of Pennsylvania's students Adam Lutz, Brennan Dumm for the integration of the MATLAB code and work on the project.

References

[1] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450-455, 2005.

[2] J. L. Wayman, "Fundamentals of biometric authentication technologies," *International Journal of Image and Graphics*, vol. 1, no. 1, pp. 93-113, 2001.

[3] J. Wayman, A. Jain, D. Maltoni, and D. Maio, "An introduction to biometric authentication systems," *Biometric Systems*, pp. 1-20, Springer, London, 2005.

[4] J. Daugman, "New methods in iris recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 5, pp. 1167-1175, 2007.

[5] R. P. Wildes, "Iris recognition: An emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348-1363, 1997.

[6] W. W. Boles and B. Boashash, "A human identification technique using images of the iris and wavelet transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp.1185-1188, 1998.

[7] C. Sanchez-Avila, R. Sanchez-Reillo, and D. de Martin-Roche, "Iris-based biometric recognition using dyadic wavelet transform," *IEEE Aerospace and Electronic Systems Magazine*, vol. 17, no. 10, pp. 3-6, 2002.

[8] Y. Chen, S. C. Dass, and A. K. Jain, "Localized iris image quality using 2-D wavelets," in *Proc. International Conference on Biometrics*, 2006, pp. 373-381, Springer, Berlin, Heidelberg.

[9] D. Kornish, S. Ezekiel, and M. Cornacchia, "Fusion based heterogeneous convolutional neural networks architecture," in *Proc. Applied Imagery Pattern Recognition Workshop*, pp. 1-6.

[10] A. Lutz, M. Giansiracusa, N. Messer, S. Ezekiel, E. Blasch, and M. Alford, "Optimal multi-focus contourlet-based image fusion algorithm selection," *Geospatial Informatics, Fusion, and Motion Video Analytics VI*, vol. 9841, International Society for Optics and Photonics.

[11] M. Giansiracusa, A. Lutz, N. Messer, S. Ezekiel, M. Alford, E. Blasch, and M. Manno, "A comparative study of multi-focus image fusion validation metrics," *Geospatial Informatics, Fusion, and Motion Video Analytics VI*, vol. 9841, International Society for Optics and Photonics.

[12] M. Giansiracusa, A. Lutz, S. Ezekiel, M. Alford, E. Blasch, A. Bubalo, and M. Thomas, "Multi-focus and multi-modal fusion: A study of multi-resolution transforms," *Geospatial Informatics, Fusion, and Motion Video Analytics VI*, vol. 9841, International Society for Optics and Photonics.

[13] K. Alex, *ImageNet Classification with Deep Convolutional Neural Networks*, November 2013.

[14] S. Ilya *et al*., "ImageNet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 60, no. 6, pp. 84–90, 2017.

[15] D. Ingrid, *Ten Lectures on Wavelets*, Philadelphia: Society for industrial and applied mathematics, 1992, vol. 61.

[16] Soundararajan Ezekiel, et.al., "Bandelet-based image fusion: A comparative study for multi-focus images," *2016 Geospatial Informatics, Fusion, and Video Analytics VI, SPIE Defense + Security Conference*, IEEE, 2016

[17] E. J. Candes and D. L. Donoho, "Ridgelets: A key to higher-dimensional intermittency?" *Philosophical Transactions of the Royal Society A*, pp. 2495–2509, 1999.

[18] R. C. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed., Prentice Hall, 2008.

[19] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091-2106, 2005.

[20] D. K. Sahu and P. Parsai, "Different image fusion techniques - A critial review," *Internationl Journal of Modern Engineering Research*, vol. 2, no. 5, 2012.

[21] E. Blasch, P. Valin, and E. Bossé, "Measures of effectiveness for high-level fusion," in *Proc. Int'l Conf. on Info Fusion*, 2010.

[22] E. Blasch, L. Z. Petkie *et al*., "Image fusion of the teraherts-visual NAECON grand challenge data," in *Proc. IEEE National Aerospace and Electronics Conf.*, 2012.

[23] E. Blasch, X. Li, G. Chen, and W. Li, "Image quality assessment for performance evaluation of image fusion," in *Proc. Int'l Conf. on Info Fusion*, 2008.

[24] X. M. Li, G. P. Yan, and L. Chen, "A new method of image denoise using contourlet transform," in *Proc. International Conference on Environmental Science and Information Application Technology*, 2009, vol. 2, pp. 390-393.

[25] P. Constantine, and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15-33, 2000.

[26] M. Stéphane and G. Peyré, "Orthogonal bandlet bases for geometric images approximation," *Communications on Pure and Applied Mathematics*, vol. 61, no. 9, pp. 1173-1212, 2008.

[27] L. P. Erwan and S. Mallat, "Bandelet image approximation and compression," *Multiscale Modeling & Simulation*, vol. 4, no. 3, pp. 992-1039, 2005.

[28] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1993.
[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 1, pp. 1-14, 2004.

Abdullah A. Alshehri received his B.S. in 1993 in electrical engineering from University of Detroit, Detroit, MI. USA. He received his M.S. and Ph.D. in electrical engineering from University of Pittsburgh, PA in 1999 and 2004 respectively. From 2005 to 2010, he worked as assistant professor in the College of Telecom and Electronics CTE and Jeddah College of Technology JCT, Saudi Arabia. In December 2010, he joined the Electrical Engineering Department at King Abdulaziz University-Rabigh KAU, Saudi Arabia and is now an associate professor. His areas of interests are in advanced signal and image processing such as time-frequency, wavelet transform, neural networks, and statistical signal processing. Dr. Alshehri is a member of IEEE since 1992 and a member of the Saudi Engineers Council SEC since 2005.

Soundararajan Ezekiel received his M.A and PhD degree from the Department of Mathematics, University of Pittsburgh, Pittsburgh, PA USA. He also received his MSc degree in mathematics at Loyola college, Post Graduate Diploma in Operations Research at Anna University, and his M. Phil at Madras Christian College, India.

He is currently a professor in computer science, Indiana University of Pennsylvania, PA. Professor Ezekiel is the recipient of three-time SFFP fellow and seven-time VFRP fellow. His research includes image processing, signal processing, wavelet analysis, artificial intelligence, machine vision, deep learning, and cyber security.