KSÜ Mühendislik Bilimleri Dergisi, 26(4), 2023
Araştırma Makalesi

KSU J Eng Sci, 26(4), 2023
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 17.08.2023
Kabul Tarihi : 21.09.2023

Received Date : 17.08.2023
Accepted Date : 21.09.2023

IMAGE FUSION AND DEEP LEARNING BASED EAR RECOGNITION USING THERMAL AND VISIBLE IMAGES

TERMAL VE GÖRÜNÜR GÖRÜNTÜLER KULLANILARAK GÖRÜNTÜ BİRLEŞTİRME VE DERİN ÖĞRENME TABANLI KULAK TANIMA

Mücahit CİHAN[1*] (ORCID: 0000-0002-1426-319X)
Murat CEYLAN[1] (ORCID: 0000-0001-6503-9668)

[1] Konya Technical University, Electrical and Electronics Engineering, Konya, Türkiye

* Corresponding Author / Sorumlu Yazar: Mücahit CİHAN, mcihan@ktun.edu.tr

ABSTRACT

Advances in imaging and deep learning have fueled interest in ear biometrics, as the structure of the ear offers unique identification features. Thermal and visible ear images capture different aspects of these features. Thermal images are light-independent, and visible images excel at capturing texture details. Combining these images creates more feature-rich composite images. This study examines the fusion of thermal and visible ear images taken under varying lighting conditions to enhance automatic ear recognition. The image fusion process involved three distinct multiresolution analysis methods: discrete wavelet transform, ridgelet transform, and curvelet transform. Subsequently, a specially designed deep learning model was used for ear recognition. The results of this study reveal that employing the complex-valued curvelet transform and thermal images achieved an impressive recognition rate of 96.82%, surpassing all other methods. Conversely, visible images exhibited the lowest recognition rate of 75.00%, especially in low-light conditions. In conclusion, the fusion of multiple data sources significantly enhances ear recognition effectiveness, and the proposed model consistently achieves remarkable recognition rates even when working with a limited number of fused ear images.

Keywords: Ear recognition, image fusion, deep learning, multi-resolution analysis methods, thermal and visible images.

ÖZET

Görüntüleme ve derin öğrenme alanındaki gelişmeler, kulağın yapısı benzersiz tanımlama özellikleri sunduğundan kulak biyometrisine olan ilgiyi artırmıştır. Termal ve görünür kulak görüntüleri bu özelliklerin farklı yönlerini yakalar. Termal görüntüler ışıktan bağımsızdır ve görünür görüntüler doku ayrıntılarını yakalamada mükemmeldir. Bu görüntülerin birleştirilmesi daha zengin özelliklere sahip kompozit görüntüler oluşturur. Bu çalışma, otomatik kulak tanımayı geliştirmek amacıyla farklı aydınlatma koşulları altında elde edilen termal ve görünür kulak görüntülerinin birleştirilmesini incelemektedir. Görüntü birleştirme işlemi üç farklı çok çözünürlüklü analiz yöntemini içermektedir: ayrık dalgacık dönüşümü, ridgelet dönüşümü ve curvelet dönüşümü. Ardından, kulak tanıma için özel olarak tasarlanmış derin öğrenme modeli kullanılmıştır. Bu çalışmanın sonuçları, karmaşık değerli curvelet dönüşümü ve termal görüntülerin kullanılmasının, diğer tüm yöntemleri geride bırakarak %96.82 gibi etkileyici bir tanıma oranı elde ettiğini ortaya koymaktadır. Buna karşılık, görünür görüntüler özellikle düşük ışık koşullarında %75.00 ile en düşük tanıma oranını sergilemiştir. Sonuç olarak, birden fazla veri kaynağının birleştirilmesi kulak tanıma etkinliğini önemli ölçüde artırmaktadır ve önerilen model, sınırlı sayıda birleştirilmiş kulak görüntüsüyle çalışırken bile tutarlı bir şekilde dikkate değer tanıma oranlarına ulaşmaktadır.

Anahtar Kelimeler: Kulak tanıma, görüntü birleştirme, derin öğrenme, çoklu çözünürlüklü analiz yöntemleri, termal ve görünür görüntüler

ToCite: CİHAN, M., & CEYLAN, M., (2023). IMAGE FUSION AND DEEP LEARNING BASED EAR RECOGNITION USING THERMAL AND VISIBLE IMAGES. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 26(4), 997-1009.

INTRODUCTION

Biometrics are biological measurements or physical characteristics that can be used to identify individuals (Jain et al., 2007). These traits can be chemical-based, such as DNA, physical-based, like fingerprinting, or behavioral-based, such as gait. These features are unique to each individual, akin to an individual's signature. Biometric systems offer the advantage of stability and durability. Biometric systems can identify a person despite minor differences and provide quick results. However, the primary disadvantage of biometric systems is that they are often considered invasive, and people may feel uncomfortable sharing their personal information. Additionally, the expense associated with biometric systems, including installation, operation, and maintenance costs, can be perceived as a drawback. Nevertheless, biometric systems possess the unique ability to recognize personal characteristics in various applications.

In addition to commonly used biometric systems such as fingerprint, iris, retina, and face recognition, extensive studies have been carried out on ear recognition. This is because the ear structure has some distinctive features for recognizing people. Initially, Iannarelli (1989) conducted experiments to prove that the ear structure is unique for all individuals. In his research, Iannarelli manually performed twelve measurements for different regions and thus detected differences. However, applying this approach to a real-world scenario is difficult. For this reason, many recent studies have focused on automating the ear recognition process by extracting new features for an effective system and developing new methods.

Ear images can be easily captured from profile or video images, making them more versatile compared to fingerprint and iris recognition, which require direct interaction with a sensor (Emeršič et al., 2017). These images can be obtained remotely, without requiring the subject's cooperation, making ear recognition technology similar to image-based biometric methods such as face and palm recognition. Some studies have even confirmed that even identical twins exhibit differences in certain features of their ear structure (Nejati et al., 2012). Furthermore, ear recognition systems can complement other biometric modalities, providing identification clues when information from other sources is unreliable or unavailable. Recent studies have demonstrated the importance of ear recognition technology in multimodal biometric systems (Sarangi et al., 2018; Sarangi et al., 2022; Ma et al., 2020; Maity et al., 2020).
While ear recognition offers many advantages, it also faces challenges that can impact the recognition process. One significant challenge is the illumination level in the environment where ear images are captured, which is a common drawback for biometric applications like ear and face recognition (Toygar et al., 2018; Jamil et al., 2014). To address this issue, an ear recognition method that can adapt to varying illumination conditions is crucial for effective ear biometry. An approach to tackle this problem is the active approach, which uses imaging modalities capable of acquiring images independently of illumination. For example, Abaza & Bourlai (2012) demonstrated that their proposed ear detection methods worked with promising results in both mid-wave infrared (MWIR) and visible bands, unlike previous approaches that only functioned in the visible spectrum. These two different spectrums each have their advantages and disadvantages. Thermal imaging is immune to illumination variations but may not capture texture features effectively, whereas visible images illuminated with sufficient light offer better texture feature representation. To harness the advantages of both methods and gather comprehensive information, this study combined thermal and visible images.

Ear recognition can be performed manually or automatically. Prior to automatic recognition, experimental studies highlighted the individual recognition potential of ears (Bertillon, 1896; Fields et al., 1960). Iannarelli made a significant contribution to this field with a long-term study on ear recognition (Iannarelli, 1989), using over ten thousand ear images. Subsequent to this pioneering work, the 1990s marked the advent of automatic ear recognition, with different methods being developed. Moreno et al. (1999) used geometric ear features and applied automatic classification with compression mesh. Victor et al. (2002) employed Principal Components Analysis (PCA) on ear images and achieved successful results. In another study, Gutiérrez et al. (2010) aimed to achieve recognition by segmenting ear image data into 12 different modules. Initially, they achieved a success rate of 91.85% without applying preprocessing. Subsequently, they applied preprocessing steps, including region of interest determination and wavelet transform. These preprocessing steps led to a notable increase in the recognition rate, reaching 97.5%. Pflug et al. (2014) conducted a comprehensive study using various feature extraction methods and achieved success with Linear Discriminant Analysis. More recently, deep learning methods, particularly Convolutional Neural Networks (CNNs), have gained prominence in automatic ear recognition, offering robustness and tolerance to shape and visual variations (Galdámez et al., 2017). These networks automatically extract features from images, eliminating the need for separate feature extraction algorithms.

Deep learning methods have gained increasing prominence in recent times for automatic ear recognition. Among these methods, CNNs stand out as one of the most favored choices. When compared to traditional feature-based methods like Local Binary Patterns (LBP) (Benzaoui et al., 2015) or Histogram of Oriented Gradients (HOG) (Ciresan et al., 2011), CNNs exhibit significantly greater robustness and tolerance to variations in shape and visual aspects within images intended for recognition (Galdámez et al., 2017). Additionally, CNNs' convolution layers automatically extract features from images, eliminating the need for separate feature extraction algorithms. Alshazly et al. (2019) presented and compared manually crafted and CNN-based ear recognition models. They initially extracted features using seven different feature extraction methods and classified them using Support Vector Machines (SVM). Subsequently, they inputted ear images into AlexNet, a CNN architecture, and trained the model. The results indicated that the AlexNet architecture achieved a 22% higher success rate. Similar research employing the AlexNet architecture is available in the literature (Abd Almisreb et al., 2018). In another study, Emersic et al. (2017) achieved high accuracy rates, even with limited training data, using different CNN architectures such as VGG-16 and SqueezeNet. Furthermore, El Naggar & Bourlai (2022) attained impressive success rates of 98.76% for visible images and 96.93% for thermal images through pre-trained CNN architectures and transfer learning on various ear datasets obtained under the same lighting conditions. These studies collectively demonstrate that deep learning methods consistently outperform feature extraction-based machine learning methods in ear recognition applications. In some face recognition studies, researchers have successfully addressed illumination challenges by combining thermal and visible images (Kong et al., 2007; Choi et al., 2012; Seal et al., 2017). For instance, Kong et al. (2007) proposed combining visible and thermal images using a multi-scale data fusion method based on the Discrete Wavelet Transform (DWT) to enhance face recognition performance. However, there is limited research exploring the fusion of thermal and visible images for ear recognition. Ariffin et al. (2017) conducted one such study, combining images using simple fusion rules like simple average and mean average along with DWT. They extracted features from these images using the Histogram of Oriented Gradients (HOG) method and classified them using SVM, reporting improved results with fused images. In our study, we took a different approach, combining thermal and visible ear images using Multiresolution Analysis (MRA) methods. To our knowledge, no previous research has combined thermal and visible ear images using MRA and classified them using deep learning methods. In this context, the findings of our study represent a significant advancement in the field of ear recognition.

In this study, the DIAST dataset, which comprises both thermal and visible ear images, was used. In this phase, it was employed various MRA techniques to combine thermal and visible ear images. These fused images were subsequently employed to train and assess the performance of our custom-designed CNN model. It was conducted experimental studies under diverse illumination conditions to evaluate the effectiveness of our proposed method, and recognition rates were calculated to gauge its performance across different modalities. The key contributions of this study can be summarized as follows:

- **MRA-Based Fusion:** It was employed three distinct MRA methods for image fusion, allowing us to capture and amalgamate distinctive features present in thermal and visible ear images acquired under varying illumination conditions. These fusion techniques excel in preserving details, demonstrating improved generalization capabilities, and delivering superior visual results.

- **Unique CNN Model:** It was designed a specialized CNN model tailored for extracting ear features from the fused images. Impressively, this model achieved outstanding results while using a minimal number of parameters, even with the constraints of limited available data.

- **Enhanced Recognition:** The findings underscore the significant improvement in the success of ear recognition applications achieved through the fusion process. This enhancement in performance is particularly noteworthy when dealing with variations in illumination conditions.

In essence, our study showcases the potential of MRA-based fusion techniques in conjunction with a carefully crafted CNN model to significantly enhance the efficacy of ear recognition, even in challenging scenarios characterized by varying illumination levels.

METHODOLOGY

This study aims to create a system capable of automatically determining the correct class or identity of the fusion image by combining thermal and visible ear images obtained under different lighting conditions using MRA-based methods. In this context, a unique CNN model was designed. This section comprises three separate sub-steps: (1)

Multiresolution analysis methods, (2) *Pixel-level fusion of thermal and visible ear images*, and (3) *the designed convolutional neural network Model and parameters*.

Multiresolution analysis methods

The development of wavelets has made Multiresolution Analysis (MRA) methods very popular. MRA methods, operating at various scales, are frequently employed in image processing applications to capture different features of images (Cihan & Ceylan, 2021). By displaying images at different scales, one can easily detect inconspicuous features at various levels (Morlet et al., 1982). In this study, three different MRA methods were used for the fusion of thermal and visible images: discrete wavelet transform, ridgelet transform, and curvelet transform. Due to their diverse capabilities in identifying points, edges, and curves, these methods are often preferred for tasks such as image denoising, object recognition, and image fusion. Moreover, each of these methods offers different levels and versions of transformations.

Discrete wavelet transform (DWT): The Wavelet Transform is among the MRA methods used for the analysis of both stationary and non-stationary signals. It serves as an effective tool in image analysis methods by enabling local analysis through the separation of data into various frequency components. This segmentation allows the examination of large signals in small areas (Cihan & Ceylan, 2021). Figure 1 illustrates the 2D-DWT process involving low-pass and high-pass filter banks. This process yields the low-resolution counterpart of the original image (LL subband), which contains the approximation coefficients, as well as images with detail coefficients (HL, LH, and HH subbands) that convey additional information. The LL subband can be further transformed to achieve higher levels of transformation.

Equation 1 is used to apply WT in discrete form, where i, w_i, s_i, p_i and $\Psi(t)$ are the numbers of the samples, weight coefficients, scales, positions and mother wavelet, respectively.

$$\hat{h}(x) \approx \sum_{i=1}^{K} w_i \Psi\left(\frac{x-p_i}{s_i}\right) \quad (1)$$

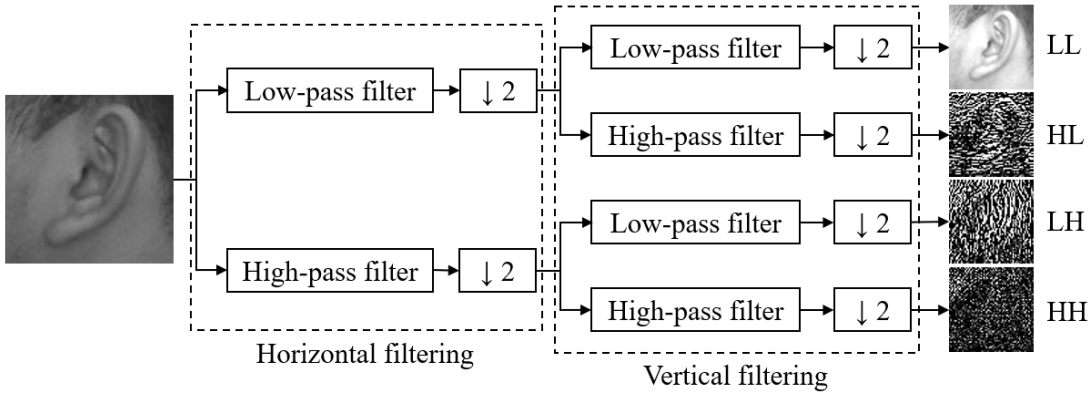

Figure 1. The Decomposition Of A Visible Ear Image Using 2D-DWT Involves A Two-Step Process. Initially, Filter Banks (Comprising Low-Pass And High-Pass Filters) Are Applied Horizontally To The Image. Following This Horizontal Filtering, Subbands Are Generated By Applying Filter Banks Vertically To The Results Obtained In The Previous Step.

Ridgelet transform (RT): Ridgelet transform (RT) offers several advantages over the wavelet transform. While DWT can capture horizontal, vertical, and diagonal components in an image pointwise, it cannot capture linear components and curvilinear structures, such as edges and corners, at various angles. Wavelets struggle to effectively represent objects with highly anisotropic elements due to their limited edge curve regularity and non-geometric nature. In contrast, RT is preferred in many computer vision applications because of its ability to capture linear components at various angles (Rane & Bhadade, 2020; Chen et al., 2021).

When applying RT (the method of choice in image analysis studies), the Ridgelet function (Equation 2) is used. Thanks to the included angle parameter, the Ridgelet function provides a versatile analysis along a straight line. RT is implemented similarly to WT. The RT coefficients of a 2D signal $f\ (x_1, x_2\ )$ are obtained from the product of the original signal and the Ridgelet function (Equation 3). In Equation 3, $\Psi(.)$ $r$epresents 1B Wavelet function so that

$x = (x_1 x_2) \in R^2$ condition is satisfied. In the same equation, θ ($\theta \in [0,2\pi)$) is the direction parameter (Do & Vetterli, 2003).

$$\Psi_{a,b,\theta}(x) = a^{-1/2}\Psi((x_1 cos\theta + x_2 sin\theta - b)/a) \quad (2)$$

$$R(a,b,\theta) = \int_{R^2} \Psi_{a,b,\theta}(x) f(x_1, x_2) dx_1 dx_2 \quad (3)$$

RT is fundamentally rooted in the Radon transform, which itself relies on the Fourier transform. To derive the Radon transform coefficients of an image, a two-dimensional Fourier transform is initially applied to the image. These coefficients are then interpolated along a straight line. Subsequently, applying a one-dimensional inverse Fourier transform to the interpolated result yields the Radon coefficient. Notably, the Radon transform serves to convert the curves present in an image into point discontinuities. The Radon transform of an image ($f(x,y)$) can also be expressed as follows to show a δ Dirac distribution:

$$P(t,\theta) = \int_{R^2} f(x,y)\delta(xcos\theta + ysin\theta - t)dxdy \quad (4)$$

If 1D WT is applied to the Radon coefficients, RT coefficients are reached (Equation 4). The application of RT depending on the Fourier transform is given in Figure 2.

$$R(a,b,\theta) = \int_{R^2} \Psi_{a,b}(t) P(\theta,t) dt \quad (5)$$

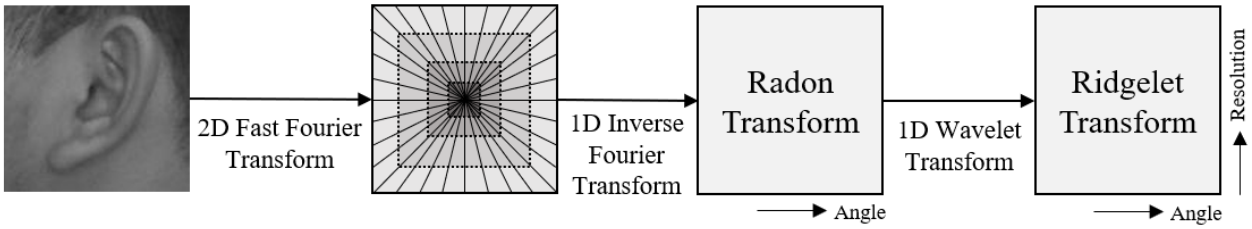

Figure 2. The Ridgelet Transform Flowchart Involves Processing Each Radial Line In The Fourier Domain Individually. First, A 1D Inverse Fast Fourier Transform Is Calculated Along Each Radial Line. Subsequently, The Ridgelet Coefficients Are Derived From The 1D Non-Orthogonal Wavelet Transform.

Curvelet transform (CT): While RT is effective in capturing regions containing edges and linear components in images, it falls short in capturing regions with curvilinear components. To address this limitation, the Curvelet Transform (CT) was developed by Fadili and Starck in 2009.

CT, first introduced by Candes & Donoho (1999) and later revised by Starck et al. (2003), represents a high-dimensional generalization of the WT. This MRA method is specifically designed to represent images at various angles and scales. CT can be visualized as a multi-scale pyramid, with its frame elements indexed based on parameters such as scale, direction, and position. Notably, the curvelet pyramid offers exceptional directional sensitivity and a degree of anisotropy, as demonstrated by AlZubi et al. (2011).

There are two types of CT methods: first-generation CT and second-generation CT. First-generation CT is designed to reduce noise in images but requires more processing time compared to second-generation CT. It has a more complex structure. Conversely, the numerical implementation of second-generation CT is simpler and can be executed in less time with fewer operations (Candes et al., 2006). In this study, second-generation CT was employed using both real and complex values. While real-valued CT provides only amplitude components, complex-valued CT additionally extracts phase components, enhancing its directional selectivity. However, working with complex numbers can introduce greater computational complexity, necessitating more computational resources and processing power. In contrast, real-valued CT, relying solely on real numbers, is known for its simplicity and speed in computations. Complex-valued CT is typically chosen when capturing fine details is critical, especially in tasks like high-resolution image processing and signal analysis. Conversely, real-valued CT is better suited for applications prioritizing efficient and rapid computations. Figure 3 illustrates the schematic of the second-generation CT.

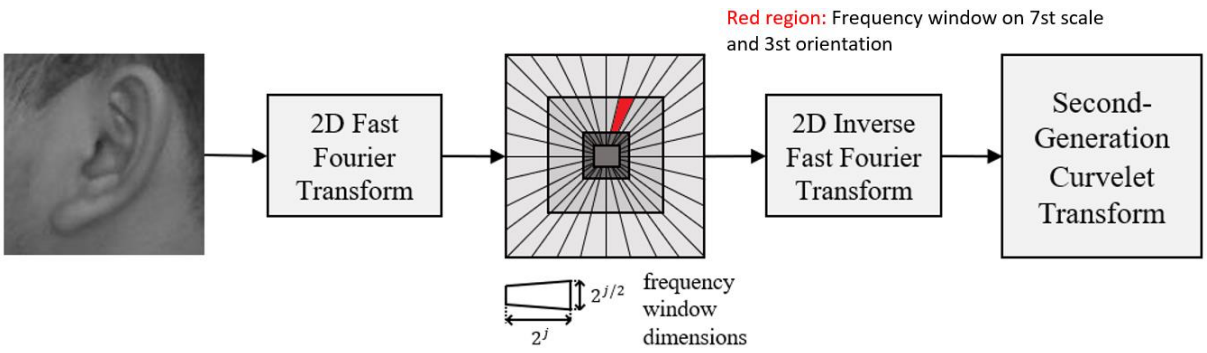

Figure 3. The Flowchart For The Second-Generation Curvelet Transform Involves The Separate Processing Of Frequency Windows Between Radial Lines In The Fourier Domain. Initially, The 2D Inverse Fast Fourier Transform Is Computed For Each Frequency Window, Resulting In The Derivation Of Second-Generation Curvelet Coefficients. These Frequency Windows Can Be Assigned Individual Names, And The Name Of The Frequency Window Within The Red Region Is Highlighted.

Pixel-level fusion of thermal and visible ear images

Image fusion is the process of combining images obtained from different sensors in a way that complements each other's missing aspects. The fusion is based on using datasets containing diverse information for the same data, which can be collected with different sensors or under varying conditions. In computer vision, image fusion is defined as the collection of critical information from multiple images, often resulting in a single, enriched image. The fused image contains more information than any individual input image (Haghighat et al., 2011). Previous studies have explored various techniques, including pixel-level fusion and MRA methods (Pajares & De La Cruz, 2004; Singh et al., 2004).

In this study, it was performed the fusion of thermal and visible ear images using pixel-level fusion in the multi-scale transform (MST) domain. MST-based fusion necessitates the conversion of source images into the MST domain before applying fusion rules. These fusion rules are then applied to the coefficients in the MST domain, resulting in a fused image obtained through inverse transformation. The specific methods employed for MST-based fusion are detailed in the subsection *Multiresolution Analysis Methods*.

In this study, it was used the mean selection rule to combine images after obtaining multi-scale coefficients of the thermal and visible ear images using MRA methods. Given the brightness of the thermal images, it was reasoned that the mean selection rule, as opposed to maximum or minimum rules, would yield more accurate results. In Equation 6, *T* and *V* represent the transform coefficients obtained from the thermal and visible images, respectively. The mean rule was applied to each value within the coefficient matrices, resulting in a fused image for each ear instead of two separate images. An example of a DWT-based fusion method is illustrated in Figure 4.

$$F = (T + V)/2 \tag{6}$$

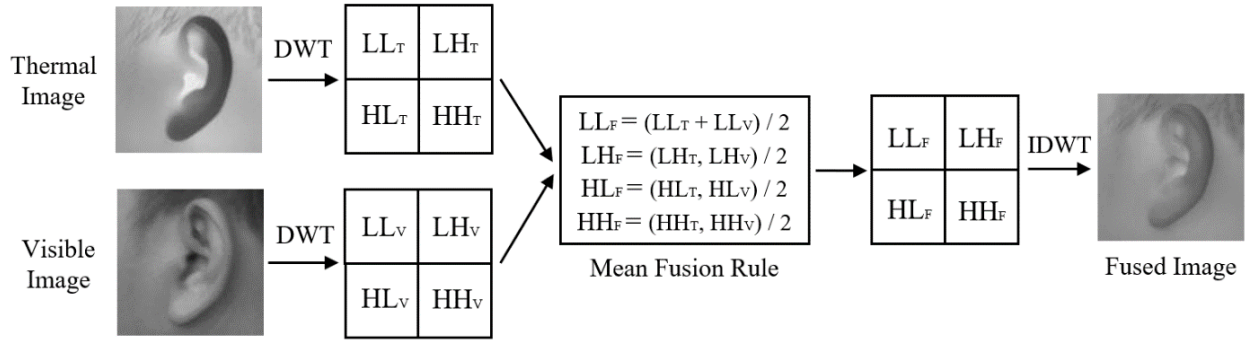

Figure 4. An Example Of A DWT-Based Fusion Method: First, The Coefficients Of The Thermal And Visible Images Is Obtained At Various Scales Using DWT. Next, The Mean Selection Rule To These Coefficients Is Applied. The Mean Selection Rule Is Used To Combine Information From Both Thermal And Visible Images Effectively. Finally, A Fusion Image Is Created By Performing An Inverse DWT On The Processed Coefficients.

The designed convolutional neural network model and parameters

Convolutional Neural Networks (CNNs) are a subset of deep neural networks and represent a specialized version of multilayer perceptrons. They find widespread use in various applications, including image classification (Cihan et al., 2022a), medical image analysis (Yu et al., 2021), image clustering (Guérin et al., 2021), and object recognition (Ashiq et al., 2022). A typical CNN architecture comprises several essential components, including a convolution layer, pooling layer, activation function, and fully connected (FC) layer (Cihan et al., 2022b).

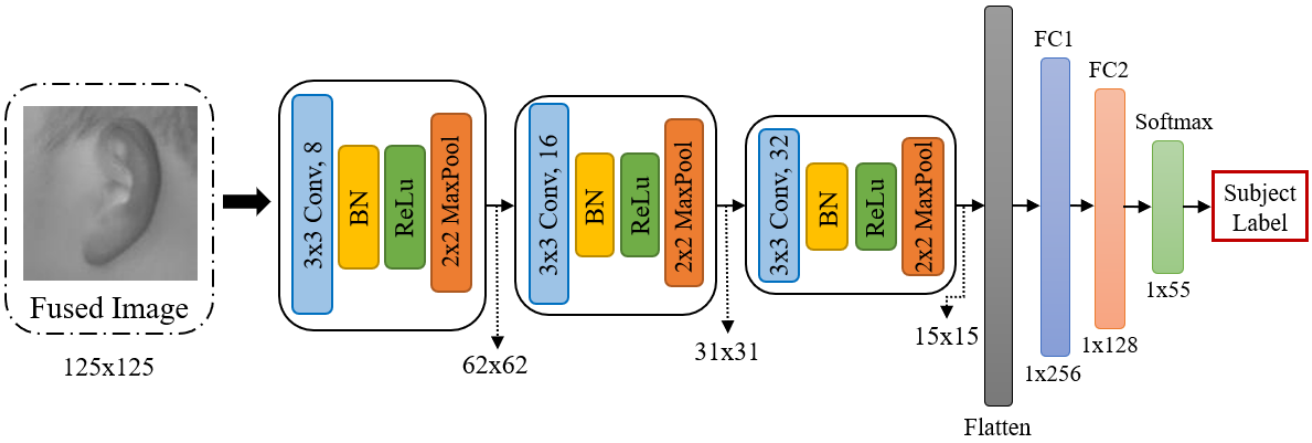

Figure 5. The Designed CNN Model Includes Several Key Components: Conv (Convolution), BN (Batch Normalization), Maxpool (Maximum Pooling), And FC (Fully Connected).

The designed CNN model comprises convolution and pooling layers that change during feature extraction, followed by deep fully connected layers for the final classification stage. Figure 5 illustrates the layer structure of the network, while Table 1 provides details about the parameters of the model. The network begins with three convolution layers. The fused images are fed into the first convolution layer, followed by subsequent layers with 8, 16, and 32 filters, respectively, each with a filter size of 3×3. Batch Normalization (BN) is applied after each convolution layer to enhance speed and stability. After each convolution layer, a 2×2 maximum pooling layer is used to reduce the size of the feature maps. Dropout layers are introduced after the second and third max-pooling layers, with a dropout rate of 0.5. The *ReLU* activation function is applied to each convolution layer. The data are then flattened and transformed into one-dimensional tensors. Three Fully Connected (FC) layers follow this step, with dropout layers introduced after each FC layer to prevent overfitting, each with a dropout rate of 0.4. The first two FC layers contain 256 and 128 neurons, respectively, while the third FC layer, serving as the output layer, consists of 55 neurons (corresponding to the number of classes). Finally, as there are multiple classes in total, the classification is performed using the *softmax* function.

Table 1. Parameters Of The Designed CNN Model.

Layer (type)	Output Shape	Parameter
input_1 (InputLayer)	(125, 125, 1)	0
conv2d_1 (Conv2D)	(123, 123, 8)	80
BN_1	(123, 123, 8)	32
max_pooling2d_1	(61, 61, 8)	0
conv2d_2 (Conv2D)	(59, 59, 16)	1168
BN_2	(59, 59, 16)	64
max_pooling2d_2	(29, 29, 16)	0
dropout_1 (Dropout)	(29, 29, 16)	0
conv2d_3 (Conv2D)	(27, 27, 32)	4640
BN_3	(27, 27, 32)	128
max_pooling2d_3	(13, 13, 32)	0
dropout_2 (Dropout)	(13, 13, 32)	0
flatten_1 (Flatten)	5408	0
fc_1 (FC)	256	1384704
dropout_3 (Dropout)	256	0
fc_2 (FC)	128	32896
dropout_4 (Dropout)	128	0
fc_3 (FC)	55	7095
Total parameters: 1,430,807		

EXPERIMENTAL SETUP AND RESULTS

In this section, the dataset used is first introduced. It is then merged thermal and visible ear images using the MRA-based fusion methods described in the previous section. Subsequently, it is employed the resulting fusion images to train the designed CNN model. The experimental outcomes and findings are then presented and discussed.

Dataset

In this study, it was used the DIAST dataset, which includes both thermal and visible ear images, as detailed by Ariffin et al. (2016). This dataset comprises a total of 2200 ear images, representing 55 different individuals. Specifically, the dataset consists of 1100 thermal images and 1100 visible images. Each individual's ear images were acquired separately for the left and right ears. The raw images in the dataset are grayscale and stored in jpg format, with a spatial size of 125x125 pixels. What sets this dataset apart is that images for each individual were captured under five distinct illumination levels, spanning a wide range of light intensities from 2 lux to 10700 lux. These images are categorized into three lighting conditions based on their lux values: 'dark' for lux values ranging from 0 to 20, 'average' for lux values between 21 and 100, and 'bright' for lux values exceeding 100. To provide a visual representation, Figure 6 showcases example images obtained under different illumination levels.

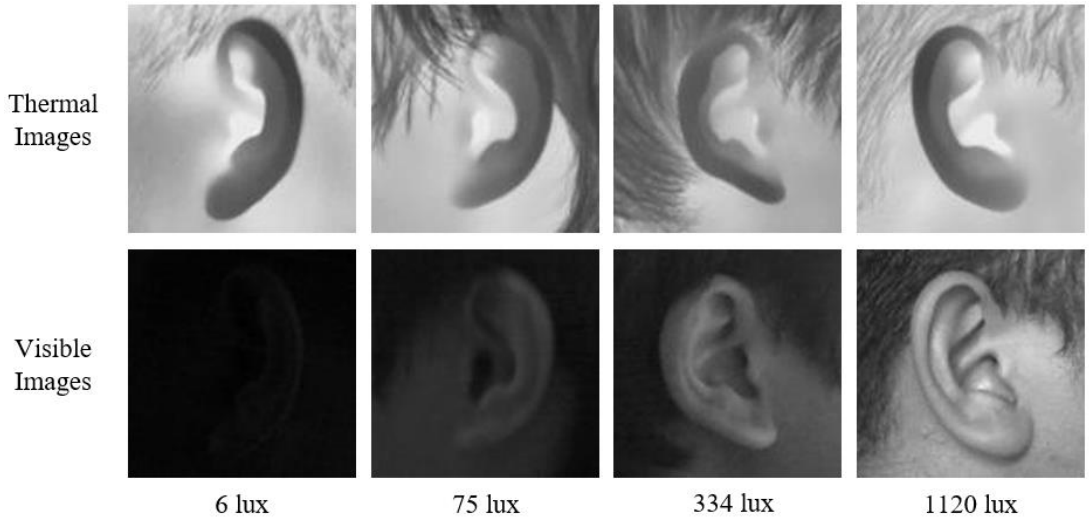

Figure 6. The Examples Of Thermal And Visible Ear Images At Different Illumination Levels In The DIAST Dataset.

Performance metrics and protocols

In this experiment, fusion images obtained through MRA methods were used for an identification task, and the results were evaluated using the designed CNN model. The dataset, comprising images captured at different illumination levels, was used to train the network, with the goal of assigning test images to one of the 55 available classes. The primary performance metric used in this study was the recognition rate, which calculates the ratio of correctly predicted data to the total number of data.

For the experiment, an 80%-20% split was employed, with 80% of the total data used for training and the remaining 20% for testing. To enhance model reliability and detect overfitting, 5-fold cross-validation was performed during training. The CNN model was trained with a batch size of 6 over 200 epochs.

Results and discussions

In this experiment, thermal and visible images captured under the same illumination conditions were combined using MRA methods. Each MRA method created 1100 fusion images, evenly distributed between right and left ears. Separate and joint recognitions were performed for the right and left ears. Figure 7 provides visual examples of fusion images obtained using different MRA methods, with a total of three MRA methods used in this study. These included 2-level and 3-level DWT fusion and RT fusion as well as real-valued CT (RCT) and complex-valued CT (CCT) fusion for second-generation CT.

Image features were automatically extracted using the convolution and maximum pooling layers within the CNN model. Ten images from each subject (either right or left ear) were available under varying illumination levels. Eight of these images per subject were used for training, resulting in a total of 440 training images. The remaining two images from each subject were reserved for testing, totaling 110 testing images. To ensure robustness and reliability, 5-fold cross-validation was employed, using images from all illumination levels for both training and testing. The same number of ear images was sampled from both the right and left ears.

The experimental results are summarized in Table 2. Notably, visible images achieved the lowest recognition rate at 75.00%, which can be attributed to the challenges posed by poorly illuminated images that hinder accurate feature extraction. Upon reviewing all results, the highest recognition rate of 96.82% was achieved using CCT with thermal images. Thermal images demonstrated superior recognition rates due to their consistent representations across different lighting conditions.

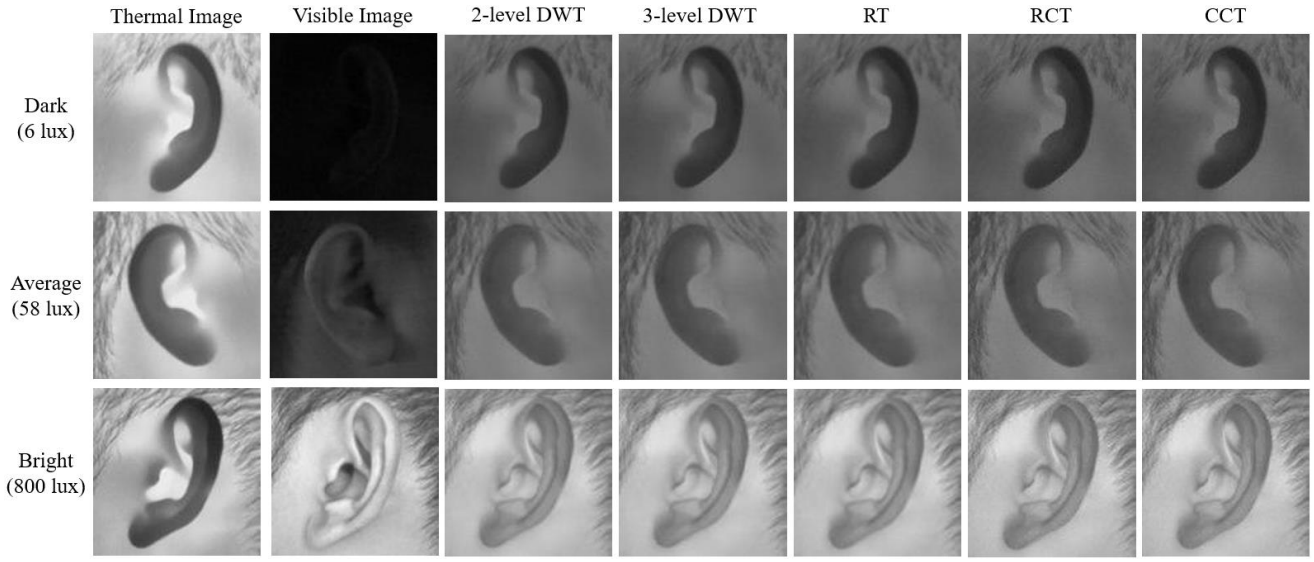

Figure 7. Original Thermal And Visible Images For Three Different Illumination Levels And Fusion Images Obtained Using Different MRA Methods.

Table 2. Experimental Results Including Recognition Of Ear Images Obtained Using Different Imaging Modalities And Pixel-Level Fusion With The Designed CNN Model. The Best Performances Are Marked In Bold.

Image Modalities	Recognition Rate (%)		
	Right	Left	Both
Visible	74.55	75.46	75.00
Thermal	**97.27**	96.36	**96.82**
2-level DWT	93.64	95.46	94.55
3-level DWT	95.46	96.36	95.91
RT	91.82	96.36	94.09
RCT	95.46	96.36	95.91
CCT	96.36	**97.27**	**96.82**

Figure 8 illustrates the model's loss and accuracy for the two most successful methods. When considering the use of thermal images, the model was trained in a shorter time compared to the CCT method. However, a closer look at test accuracy reveals that the CCT model achieves impressive results within a shorter training duration. This can be attributed to the incorporation of valuable textural information from visible images into the composite image after fusion. These graphs clearly show a simultaneous increase in model accuracy alongside a decrease in model losses for both methods. Importantly, there is no evidence of underfitting or overfitting, except in cases related to the recognition rate observed during cross-validation.

Extracting low-level features, particularly texture, from visible ear images captured under dark lighting conditions presents a significant challenge. The lack of distinctive features in these dark images hinders the fusion process, as they fail to complement the information from thermal images effectively. The experimental results in Table 2 confirm

that the deficiencies in features obtained from dark images negatively impacted the quality of fusion images, resulting in a decrease in the recognition rate.

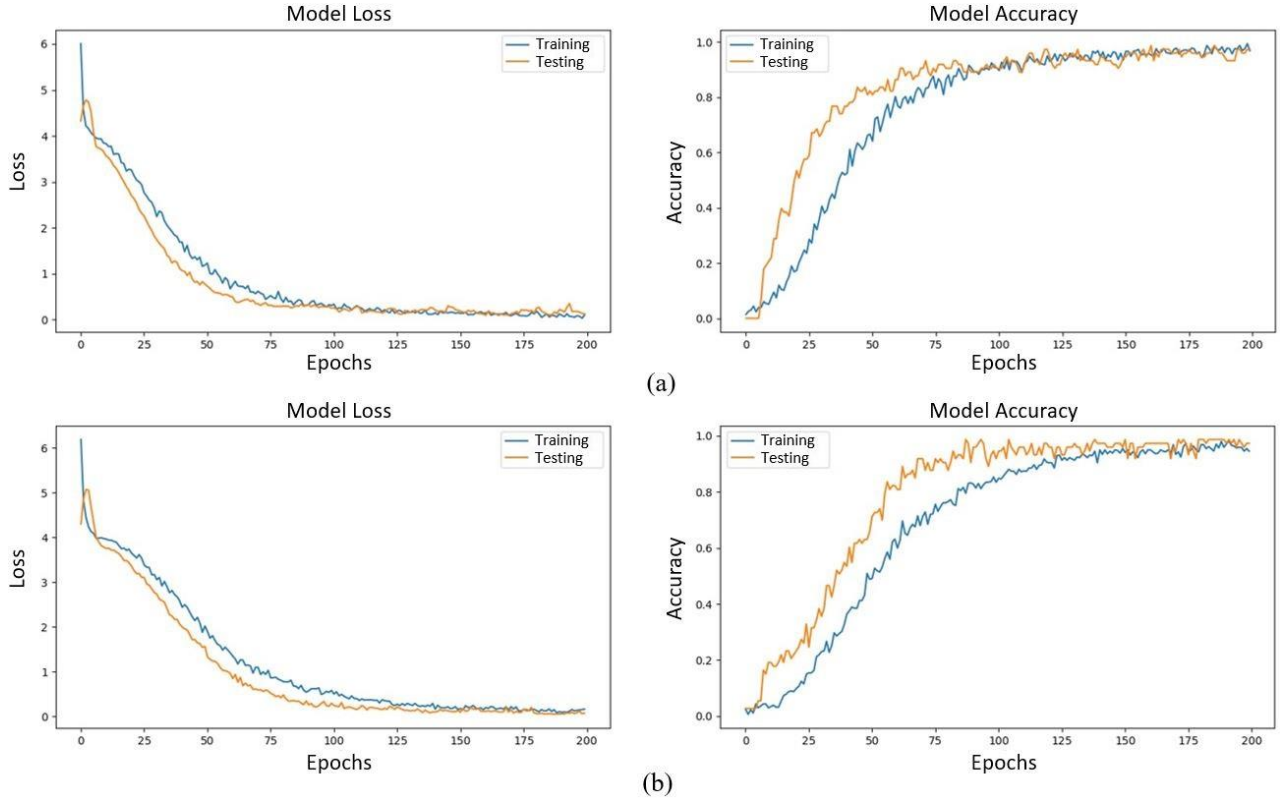

Figure 8. Model Losses And Accuracies **a.** Thermal Images **b.** CCT.

The proposed model has delivered remarkably successful results. With a total of six layers, including three convolution layers and three fully connected layers, this model showcases high-performance capabilities while maintaining a significantly reduced parameter count in comparison to established deep learning architectures such as VGG, AlexNet, ResNet, and others. The exceptional performance achieved by the proposed method eliminates the need for off-the-shelf architectures, emphasizing the effectiveness of this approach.

CONCLUSION AND FURTHER STUDY

The performances obtained through the mean fusion rule, employing three MRA methods (DWT, RT, CT), were compared in the proposed ear recognition system. Additionally, the effects of thermal and visible images on performance under different illumination conditions were evaluated. Subsequently, the obtained results are presented as follows:

- Thermal imaging remains unaffected by illumination variations but tends to struggle in capturing texture properties. Conversely, well-illuminated visible images excel at capturing textures. To leverage the strengths of both imaging modalities without sacrificing information, appropriate fusion techniques are employed. This fusion combines the distinctive features of thermal and visible images, resulting in successful outcomes.

- When dark images are used during the CNN model training, visible images exhibit lower recognition performance compared to thermal and fusion images.

- CCT stands out as the most efficient MRA method for ear recognition, as evident from the recognition rates. CCT's ability to effectively capture directional selectivity in ear images using both phase and amplitude information contributes to its superior performance. Thermal images yield the best results for the right ear, while CCT achieves the highest recognition rate for the left ear. Additionally, RCT and 3-level DWT demonstrate high recognition rates for both the right and left ears. These results underscore the enhancement of ear recognition through the fusion process. However, it's worth noting that there is no single optimal fusion technique for ear recognition.

This study worked with a limited dataset of ear data, yet our results showcase the potent application of deep learning and image fusion techniques even within data-limited domains. These methodologies empower us to achieve high-quality results, overcoming the constraints posed by limited data. Furthermore, they establish a crucial foundation for future research in this domain. The findings from this study offer valuable inspiration for all fields confronted with data limitations.

In further studies, expanding the dataset and training the CNN model with more data would be beneficial. Additionally, it would be valuable to compare results using CNN architectures like AlexNet, VGG, and SqueezeNet. Given the limited dataset size, the application of pre-trained models for transfer learning holds the potential to yield faster and more effective results in further research. Furthermore, exploring different fusion methods and fusion rules for evaluations could provide valuable insights.

REFERENCES

Abaza, A., & Bourlai, T. (2012, May). Human ear detection in the thermal infrared spectrum. *In Thermosense: Thermal Infrared Applications XXXIV*, 8354, 286-295. https://doi.org/10.1117/12.919285

Abd Almisreb, A., Jamil, N., & Din, N. M. (2018, March). Utilizing AlexNet deep transfer learning for ear recognition. *In 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, 1-5. DOI: 10.1109/INFRKM.2018.8464769

Alshazly, H., Linse, C., Barth, E., & Martinetz, T. (2019). Handcrafted versus CNN features for ear recognition. *Symmetry*, 11(12), 1493. https://doi.org/10.3390/sym11121493

AlZubi, S., Sharif, M. S., Islam, N., & Abbod, M. (2011, May). Multi-resolution analysis using curvelet and wavelet transforms for medical imaging. *In 2011 IEEE international symposium on medical measurements and applications*, 188-191. DOI: 10.1109/MeMeA.2011.5966687

Ariffin, S. M. Z. S. Z., Jamil, N., & Rahman, P. N. M. A. (2016, September). DIAST variability illuminated thermal and visible ear images datasets. *In 2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 191-195. DOI: 10.1109/SPA.2016.7763611

Ariffin, S. M. Z. S. Z., Jamil, N., & Rahman, P. N. M. A. (2017, May). Can thermal and visible image fusion improves ear recognition?. *In 2017 8th International Conference on Information Technology (ICIT)*, 780-784. DOI: 10.1109/ICITECH.2017.8079945

Ashiq, F., Asif, M., Ahmad, M. B., Zafar, S., Masood, K., Mahmood, T., Mahmood, M. T., & Lee, I. H. (2022). CNN-based object recognition and tracking system to assist visually impaired people. *IEEE Access*, 10, 14819-14834. DOI: 10.1109/ACCESS.2022.3148036

Benzaoui, A., Kheider, A., & Boukrouche, A. (2015, October). Ear description and recognition using ELBP and wavelets. *In 2015 International Conference on Applied Research In Computer Science And Engineering (Icar)*, 1-6. DOI: 10.1109/ARCSE.2015.7338146

Bertillon, A., & McClaughry, R. W. (1896). Signaletic instructions including the theory and practice of anthropometrical identification. *Werner Company*.

Candès, E. J., & Donoho, D. L. (1999). Ridgelets: A key to higher-dimensional intermittency?. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1760), 2495-2509. https://doi.org/10.1098/rsta.1999.0444

Candes, E., Demanet, L., Donoho, D., & Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3), 861-899. https://doi.org/10.1137/05064182

Chen, D., Tang, J., Xi, H., & Zhao, X. (2021). Image Recognition of Modern Agricultural Fruit Maturity Based on Internet of Things. *Traitement du Signal*, 38(4). DOI: 10.18280/ts.380435

Choi, J., Hu, S., Young, S. S., & Davis, L. S. (2012, May). Thermal to visible face recognition. *In Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX*, 8371, 252-261. https://doi.org/10.1117/12.920330

Cihan, M., & Ceylan, M. (2021). Fusion of CT and MR Liver Images Using Multiresolution Analysis Methods. *Avrupa Bilim ve Teknoloji Dergisi*, (30), 56-61. https://doi.org/10.31590/ejosat.1005858

Cihan, M., Ceylan, M., & Ornek, A. H. (2022a). Spectral-spatial classification for non-invasive health status detection of neonates using hyperspectral imaging and deep convolutional neural networks. *Spectroscopy Letters*, 1-14. https://doi.org/10.1080/00387010.2022.2076698

Cihan, M., Ceylan, M., Soylu, H., & Konak, M. (2022b). Fast Evaluation of Unhealthy and Healthy Neonates Using Hyperspectral Features on 700-850 Nm Wavelengths, ROI Extraction, and 3D-CNN. *IRBM*, 43(5), 362-371. https://doi.org/10.1016/j.irbm.2021.06.009

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011, June). Flexible, high performance convolutional neural networks for image classification. *In Twenty-second international joint conference on artificial intelligence*. DOI: 10.5591/978-1-57735-516-8/IJCAI11-210

Do, M. N., & Vetterli, M. (2003). The finite ridgelet transform for image representation. *IEEE Transactions on image Processing*, 12(1), 16-28. DOI: 10.1109/TIP.2002.806252

El-Naggar, S., & Bourlai, T. (2022). Exploring Deep Learning Ear Recognition in Thermal Images. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 5(1), 64-75. DOI: 10.1109/TBIOM.2022.3218151

Emeršič, Ž., Štepec, D., Štruc, V., & Peer, P. (2017). Training convolutional neural networks with limited training data for ear recognition in the wild. arXiv preprint arXiv:1711.09952. https://doi.org/10.48550/arXiv.1711.09952

Emeršič, Ž., Štruc, V., & Peer, P. (2017). Ear recognition: More than a survey. *Neurocomputing*, 255, 26-39. https://doi.org/10.1016/j.neucom.2016.08.139

Fadili, J. M., & Starck, J. L. (2009). Curvelets and ridgelets. https://doi.org/10.1007/978-0-387-30440-3_111

Fields, C., Falls, H. C., Warren, C. P., & Zimberoff, M. (1960). The ear of the newborn as an identification constant. *Obstetrics & Gynecology*, 16(1), 98-102.

Galdámez, P. L., Raveane, W., & Arrieta, A. G. (2017). A brief review of the ear recognition process using deep neural networks. *Journal of Applied Logic*, 24, 62-70. https://doi.org/10.1016/j.jal.2016.11.014

Guérin, J., Thiery, S., Nyiri, E., Gibaru, O., & Boots, B. (2021). Combining pretrained CNN feature extractors to enhance clustering of complex natural images. *Neurocomputing*, 423, 551-571. https://doi.org/10.1016/j.neucom.2020.10.068

Gutiérrez, L., Melin, P., & Lopez, M. (2010, July). Modular neural network integrator for human recognition from ear images. *In The 2010 International Joint Conference on Neural Networks (IJCNN)*, 1-5. DOI: 10.1109/IJCNN.2010.5596633

Haghighat, M. B. A., Aghagolzadeh, A., & Seyedarabi, H. (2011). Multi-focus image fusion for visual sensor networks in DCT domain. *Computers & Electrical Engineering*, 37(5), 789-797. https://doi.org/10.1016/j.compeleceng.2011.04.016

Jain, A. K., Flynn, P., & Ross, A. A. (Eds.). (2007). Handbook of biometrics. *Springer Science & Business Media*.

Jamil, N., AlMisreb, A., & Halin, A. A. (2014). Illumination-invariant ear authentication. *Procedia Computer Science*, 42, 271-278. https://doi.org/10.1016/j.procs.2014.11.062

Kong, S. G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B. R., Koschan, A., Yi, M., & Abidi, M. A. (2007). Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. *International Journal of Computer Vision*, 71(2), 215-233. https://doi.org/10.1007/s11263-006-6655-0

Lannarelli, A. (1989). Ear identification. *Forensic identification series*.

Ma, Y., Huang, Z., Wang, X., & Huang, K. (2020). An overview of multimodal biometrics using the face and ear. *Mathematical Problems in Engineering*, 2020. https://doi.org/10.1155/2020/6802905

Maity, S., Abdel-Mottaleb, M., & Asfour, S. S. (2020). Multimodal biometrics recognition from facial video with missing modalities using deep learning. *Journal of Information Processing Systems*, 16(1), 6-29. DOI: 10.3745/JIPS.02.0129

Moreno, B., Sanchez, A., & Vélez, J. F. (1999, October). On the use of outer ear images for personal identification in security applications. *In Proceedings IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology (Cat. No. 99CH36303)*, 469-476. DOI: 10.1109/CCST.1999.797956

Morlet, J., Arens, G., Fourgeau, E., & Glard, D. (1982). Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media. *Geophysics*, 47(2), 203-221. https://doi.org/10.1190/1.1441328

Nejati, H., Zhang, L., Sim, T., Martinez-Marroquin, E., & Dong, G. (2012, November). Wonder ears: Identification of identical twins from ear images. *In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 1201-1204.

Pajares, G., & De La Cruz, J. M. (2004). A wavelet-based image fusion tutorial. *Pattern recognition*, 37(9), 1855-1872. https://doi.org/10.1016/j.patcog.2004.03.010

Pflug, A., Paul, P. N., & Busch, C. (2014, October). A comparative study on texture and surface descriptors for ear biometrics. *In 2014 International carnahan conference on security technology (ICCST)*, 1-6. DOI: 10.1109/CCST.2014.6986993

Rane, M. E., & Bhadade, U. (2020, December). Face and palmprint Biometric recognition by using weighted score fusion technique. *In 2020 IEEE Pune Section International Conference (PuneCon)*, 11-16. DOI: 10.1109/PuneCon50868.2020.9362433

Sarangi, P. P., Mishra, B. P., & Dehuri, S. (2018, March). Multimodal biometric recognition using human ear and profile face. *In 2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, 1-6. DOI: 10.1109/RAIT.2018.8389035

Sarangi, P. P., Nayak, D. R., Panda, M., & Majhi, B. (2022). A feature-level fusion based improved multimodal biometric recognition system using ear and profile face. *Journal of Ambient Intelligence and Humanized Computing*, 13(4), 1867-1898. https://doi.org/10.1007/s12652-021-02952-0

Seal, A., Bhattacharjee, D., Nasipuri, M., Gonzalo-Martin, C., & Menasalvas, E. (2017). Fusion of visible and thermal images using a directed search method for face recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 31(04), 1756005. https://doi.org/10.1142/S0218001417560055

Singh, S., Gyaourova, A., Bebis, G., & Pavlidis, I. (2004, August). Infrared and visible image fusion for face recognition. *In Biometric technology for human identification*, 5404, 585-596. https://doi.org/10.1117/12.543549

Starck, J. L., Donoho, D. L., & Candès, E. J. (2003). Astronomical image representation by the curvelet transform. *Astronomy & Astrophysics*, 398(2), 785-800. DOI: 10.1051/0004-6361:20021571

Toygar, Ö., Alqaralleh, E., & Afaneh, A. (2018). Person identification using multimodal biometrics under different challenges. *Human-Robot Interaction-Theory and Application*, 81-96. DOI: 10.5772/intechopen.71667

Victor, B., Bowyer, K., & Sarkar, S. (2002, August). An evaluation of face and ear biometrics. *In 2002 International Conference on Pattern Recognition*, 1, 429-432. DOI: 10.1109/ICPR.2002.1044746

Yu, H., Yang, L. T., Zhang, Q., Armstrong, D., & Deen, M. J. (2021). Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing*, 444, 92-110. https://doi.org/10.1016/j.neucom.2020.04.157