Received: 21 September 2023 | Revised: 3 June 2024 | Accepted: 12 July 2024

DOI: 10.4218/etrij.2023-0395

ORIGINAL ARTICLE

ETRI Journal WILEY

Enhancement of eye socket recognition performance using inverse histogram fusion images and the Gabor transform

Harisu Abdullahi Shehu[1] | **Ibrahim Furkan Ince**[2] | **Faruk Bulut**[3,4]

[1]School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

[2]Department of Software Engineering, Istinye University, Istanbul, Turkey

[3]Department of Computer Engineering, Istanbul Esenyurt University, Istanbul, Turkey

[4]School of Computer Science and Electronic Engineering, University of Essex, Colchester, England

Correspondence

Harisu Abdullahi Shehu, School of Engineering and Computer Science, Victoria University of Wellington.

Email: harisushehu@ecs.vuw.ac.nz

Funding information

None reported.

Abstract

The eye socket is a cavity in the skull that encloses the eyeball and its surrounding muscles. It has unique shapes in individuals. This study proposes a new recognition method that relies on the eye socket shape and region. This method involves the utilization of an inverse histogram fusion image to generate Gabor features from the identified eye socket regions. These Gabor features are subsequently transformed into Gabor images and employed for recognition by utilizing both traditional methods and deep-learning models. Four distinct benchmark datasets (Flickr30, BioID, Masked AT & T, and CK+) were used to evaluate the method's performance. These datasets encompass a range of perspectives, including variations in eye shape, covering, and angles. Experimental results and comparative studies indicate that the proposed method achieved a significantly ($p < 0.001$) higher accuracy (average value greater than 92.18%) than that of the relevant identity recognition method and state-of-the-art deep networks (average value less than 78%). We conclude that this improved generalization has significant implications for advancing the methodologies employed for identity recognition.

KEYWORDS

classification, deep learning, eye socket, Gabor features, identity recognition, image matching, vector quantization

1 | INTRODUCTION

Biometric identification methods are becoming increasingly popular in various applications and allow the recognition of individuals based on their distinct physical characteristics. These biometrics encompass a broad range of features, including voice, ear, palmprint, fingerprint, face, iris, retina, hand geometry, posture, and walking style. Iris detection has gained significant attention because of its unique and unchanging nature throughout a person's lifetime. However, capturing high-resolution iris images in practical settings is challenging. Obtaining close-up images is imperative for precise detection because the intricate patterns within the iris contain crucial details. The overall performance of an iris detection system relies on accurate localization and segmentation of the iris from an eye image as well as the resolution of the image itself.

Eye socket recognition, also referred to as eye region detection or eye localization, is a computer vision technique used to identify and determine the position and shape of an eye region within an image or video frame [1].

This technique has been applied in various domains, including gaze tracking [2], face recognition [3], emotion detection [4], eye state recognition [5], gaze tracking [6], and exigency detection [7].

The eye region is an intricate structure that encompasses the eye itself, the eyebrows, and the surrounding area, known as the eye socket. The eye socket region consists of elements such as the eyelashes, eyelids, and sclera, as illustrated in Figure 1. Specifically, eye socket recognition entails the detection of the position and size of the eye socket in an image. This information was subsequently used to identify and analyze accurately the eye regions.

To detect accurately the eye socket, the initial step involves the segmentation of the eye image into distinct parts to extract the annular region located between the sclera and pupil. Failure to identify these regions precisely can lead to inaccurate outcomes in the identification process. Therefore, before proceeding with eye socket matching, it is crucial to localize effectively and segment the eye socket regions, including the inner boundary (pupil) and outer boundary (sclera along with the eyelids and eyelashes).

Eye socket recognition can be accomplished using various techniques, including feature-based methods and deep-learning algorithms [8]. Feature-based methods involve the extraction of specific attributes or patterns, such as color, texture, and shape, from images that are known to be associated with the eye region. These extracted features were then utilized to train a machine learning model to recognize the eye region in the new images. Conversely, deep-learning algorithms employ convolutional neural networks (CNNs) for eye socket recognition. CNNs are specialized artificial neural networks designed to process images by extracting relevant features and patterns. To train a CNN for eye socket recognition, a sizable dataset of labeled images was utilized, wherein each image was annotated according to the location and size of the eye region. Once the CNN is trained, it can accurately predict the location and size of the eye region in new images.

Various types of biometric studies have been conducted to explore gaze, iris, pupil, and emotional detection, among other areas of interest [9]. Several studies have been conducted on this topic. For example, Min-Allah and others [1] conducted a comprehensive investigation of recent research on pupil and iris recognition. Although these methods have achieved state-of-the-art performance, empirical findings indicate that obstacles, such as contact lenses, closed eyes, eye diseases, light reflections, and moments of pupil constriction or dilation, can hamper their effectiveness [10]. Similarly, Mahanama and others [11] suggested that many studies rely on trend analysis or first-order statistical features, such as the minimum, maximum, and skewness, rather than employing advanced measures, such as the eye region. These features, which are based on color, are expected to exhibit decreased performance and potential racial bias when identifying individuals of different races from those used to train the model. Moreover, the color of one's eyes does not serve as the sole basis for recognizing one's identity.

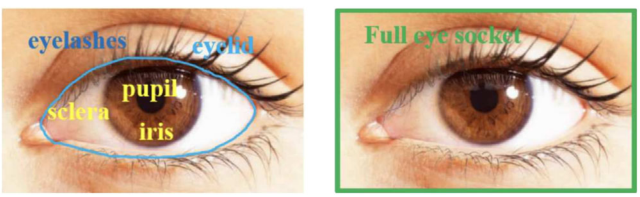

FIGURE 1 Basic parts of an eye and the full eye socket.

Conversely, a real-time vision-based system for eye state recognition was proposed using a dual CNN ensemble over eye patches [12]. Quality-based multimodal eye recognition systems that utilize the entire eye, scleral areas, and iris for quality measurements within the recognition system have been presented [13]. These multimodal systems enhance the accuracy of biometric recognition systems.

Many studies, including those mentioned above, have primarily focused on the full facial region, where the face is detected before extracting the eye region. However, the use of partial facial coverings, such as face masks, has added to the existing challenges of facial identification, especially given the recent impact of the coronavirus [14]. For example, the impacts of face masks on facial identity, gender, age, and emotion recognition were assessed using a single dataset [15]. Carragher and others examined facial recognition and perceptual face-matching systems and compared the performances of human observations and deep-neural networks (DNN) on the faces with surgical face masks [16]. Shehu and others [17] and Noyes and others [18] compared the familiarity, unfamiliarity with face matching, and classification of emotions of individuals wearing face masks and sunglasses. These studies predominantly explored the effects of partial face coverage on human recognition abilities and automated methods.

By contrast, a multimodal methodology for facial expression recognition using face masks was proposed utilizing deep-learning techniques [19]. This methodology addresses the critical step in face reading when the lower portion of the face is covered. Freud and others examined how the use of face masks alters face perception, providing qualitative and quantitative evidence of changes in masked face processing that can have meaningful effects on daily life and social interactions [20]. Furthermore, several studies employing deep-learning-based systems in the Internet-of-Things [21], ensemble deep-transfer learning [22], and YOLO models in public places for real-time detection systems [23] have been conducted. These studies highlighted the decrease in accuracy observed in automated recognition systems when the face is partially covered, such as in facemasks. However, no attempts have been expended to improve the accuracy of automated recognition systems.

Inspired by an iris recognition system that employs uniform histogram fusion images and deformable circular hollow kernels [24], this paper presents a novel recognition system that utilizes the eye socket region. This method involves the generation of Gabor features from the identified eye socket regions using an inverse histogram fusion image (IHFI) and the conversion of these features into Gabor images for recognition, with or without the integration of machine learning modules. The rationale behind using this specific image type stems from the challenge of differentiating numerous unique patterns distributed throughout the iris, which can be difficult to distinguish effectively.

Conventional recognition systems often rely on detailed and high-resolution images, particularly for iris recognition, which requires capturing intricate patterns within the iris. However, this approach has practical limitations in terms of consistently capturing high-resolution images, particularly from a distance or in less-controlled environments. Inspired by an iris recognition system that employs uniform histogram fusion images and deformable circular hollow kernels [24], this paper presents a novel recognition system that utilizes the eye socket region. This method involves the generation of Gabor features from the identified eye socket regions using an IHFI and the conversion of these features into Gabor images for recognition, with or without the integration of machine learning modules.

The rationale behind using this specific image type stems from the challenge of differentiating numerous unique patterns distributed throughout the iris, which can be difficult to distinguish effectively. It modifies the image histogram to emphasize the most frequent features within the eye socket region, countering the issues of skewed histograms owing to predominant skin colors. This enhanced the visibility and distinctiveness of the features necessary for accurate recognition.

The primary objective of this study was to propose a robust and innovative eye socket recognition method that focuses on the shape and surrounding regions of the eye socket. Unlike traditional iris-based systems that rely solely on the iris region, the proposed method provides a more dependable means of identification by considering the entire eye socket area. By leveraging the distinctive shape and surroundings of the eye socket, this method has the potential to offer a more universal approach to allow recognition applicable to individuals with diverse physical characteristics.

The contributions of this study are the following:

1. Propose a novel eye socket recognition method that employs inverse histogram fusion within the Gabor transform. This method, which considers the unique shape of the eye socket and its surrounding region, is expected to provide advantages over existing iris-based recognition systems. Specifically, it aims to improve accuracy in identifying individuals with obscured or damaged irises.
2. Analyze the generalizability of the proposed method using four different in-the-wild facial recognition datasets: (i) BioID [25] dataset with uncovered faces, (ii) MaskedAT&T [26] dataset with partially covered faces using face masks, (iii) Flickr30 [27] dataset with variations such as makeup, glasses, or eye infections, and (iv) CK+ [28] dataset with individuals exhibiting different facial expressions.
3. Highlight the advantages of the proposed method over the Gabor method and several state-of-the-art deep-learning approaches, including ResNet50 [29], InceptionV3 [30], VGG19 [31], face mesh deep neural network (FaceMesh_DNN) [32], and multitask network (MTN) [33], across datasets of varying difficulty. These findings underscore the efficacy and robustness of the proposed approach in diverse and challenging scenarios.

The remainder of this paper is organized as follows. Section 2 describes the proposed method in detail, and Section 3 provides comprehensive information on the experimental work, including hardware specifications, utilized datasets, and the obtained results. Section 4 presents a thorough discussion of the findings and potential avenues for future research. Finally, Section 5 concludes the paper.

2 | PROPOSED METHOD

Before introducing the proposed method, it is essential to address the potential limitations of eye-and-eye socket recognition systems. These considerations are crucial for a comprehensive understanding of the proposed method, as illustrated in Figure 2. The system operates through three distinct phases: region-of-interest (ROI) detection, IHFI, and recognition.

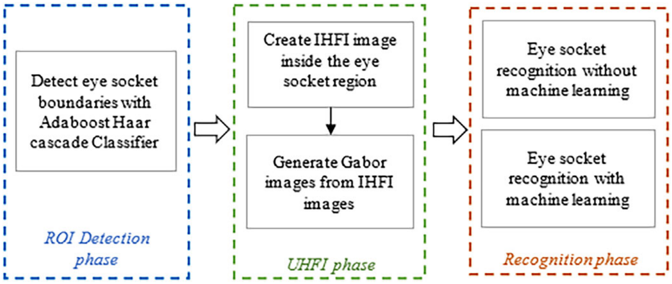

FIGURE 2 Flowchart of the proposed method.

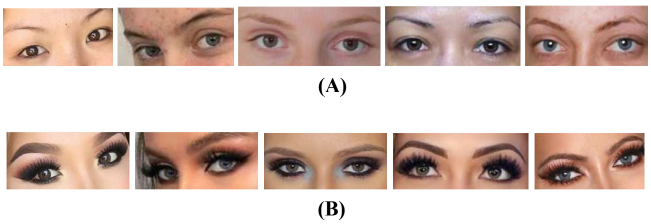

FIGURE 3 Comparison of images before and after eye makeup: (A) before and (B) after [34].

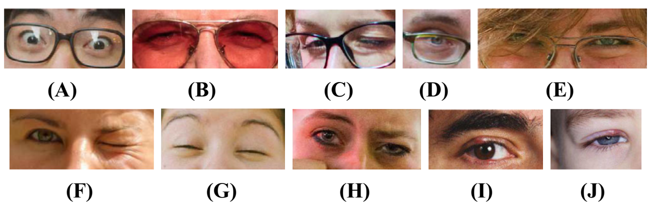

FIGURE 4 (A) Glasses reflecting specular light, (B) colorimetric glasses, (C, D) eyeglass frame preventing the eye socket from being seen, (E) hair partially obstructing the eye, (F) a partially open eye and eye blinking, (G) closed eyes, (H) nearly closed eye, (I, J) eye illnesses.

2.1 | Limitations

Several barriers affect the effectiveness of the detection systems. Significant limitations that can affect the performance of an eye socket recognition system include (i) makeup, (ii) glasses and varying light conditions, (iii) illnesses, (iv) emotions, and (v) different angles. Figure 3 presents a sample showing how the eye socket can be influenced by these factors in different individuals before and after applying eye makeup.

The use of eyeglasses, the presence of light, and eye illnesses can also serve as restrictions that can affect the eye socket (see Figure 4).

Different views of the emotions of the same person are shown in Figure 5. Each prominent emotional state directly affected the shape of the eye socket.

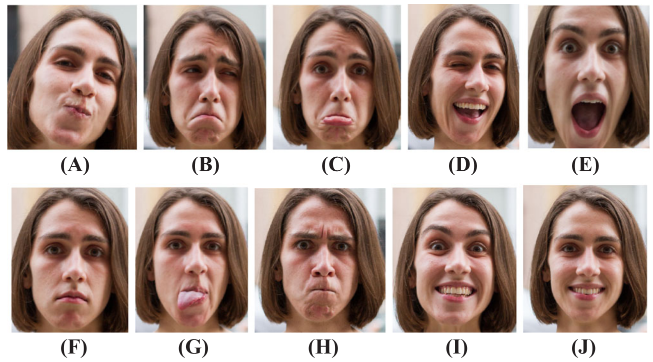

FIGURE 5 Photographs of a person's face expressing different emotions: (A) kiss, (B) crying, (C) sad, (D) winking, (E) astonished, (F) disappointed, (G) tongue, (H) angry, (I) grinning, and (J) smiling [35].

FIGURE 6 Images acquired from different angles of the same person:(A) down, (B) up, (C) 45° R, (D) 45° L, and (E) front.

Views of a person from different angles are shown in Figure 6. In these cases, the recognition system fails.

In experimental studies, a woman could cover her face with her hands in some photographs; similarly, individuals could bow their heads. Individuals wearing sunglasses were excluded. However, this issue must not be addressed. If eyes were not visible, they were excluded from the dataset. Contact lenses do not affect the proposed system because they are placed on the iris.

2.2 | Eye socket extraction

The key aspect of a system that recognizes human identity from the eye region is the accurate identification of the eye sockets. This can be achieved by using Haar-like object detectors, which were originally introduced by Viola [36] and subsequently improved by Lienhart [37]. These detectors enable a classifier trained with sample views of an object to detect the object in an entire image. This method offers speed, efficiency, and accuracy when implemented appropriately. It is particularly effective for detecting objects that are partially obscured or when video frames are noisy. The OpenCV library provides fully trained eye region Haar-cascade descriptors that were utilized in our system.

2.3 | IHFI

This study proposes a new conversion method known as IHFI. This approach is based on the fact that the iris contains numerous unique patterns, but the surrounding eye socket region is dominated by skin color, leading to a skewed histogram. To address this issue, an eye socket image was first inverted in terms of color, and a probability density function (PDF) was used to measure the frequency distribution of the image. An inverse PDF (which maps the intensity values back to the intensity levels) is then created from the intensity histogram of each color channel (red, green, blue, and alpha) of the eye socket image, as shown in (1). The resulting IHFI image captures the most probable and frequent features within the eye socket region, which are the iris patterns. This IHFI image was then used to generate Gabor features, which were compared with a dataset using normalized cross-correlation for identity recognition.

$$p_x(i) = \frac{n_{L-1-i}}{n}, \;\; 0 \le i \le L-1, \tag{1}$$

where $p_x(i)$ represents the probability of the pixel intensity i, n denotes the total number of pixel intensities, L is the maximum pixel intensity value, and i is the pixel intensity value, where $0 \le i \le L-1$.

Equation (2) shows the calculation of the PDF value $p_x(i)$, which is related to the occurrence probability of an intensity level for each color channel of the eye socket image. The maximum color intensity value is represented by L. To create a probability image for each channel, the PDF value of each pixel is multiplied by its intensity value. This process was performed separately for each pixel in the image for each color channel (red, green, blue, and alpha).

$$j = p_x(i) \times i, \tag{2}$$

where $j = p_x(i) \times i$, j represents the resulting value, $p_x(i)$ represents the probability of the pixel intensity i, and i is the pixel intensity value.

Considering that there was only one white-colored pixel in the eye socket image, the resulting PDF value for the white color would be nearly zero; this would create an intensity value near zero in the probability image created from it. This scenario represents the maximum information loss, where the white color almost turns black. To prevent this, the geometric means of the original pixel intensity value and the pixel intensity value of the conventional probability image were calculated. This method aims to avoid losing information in worst-case scenarios. The corresponding (3) is given as follows:

$$k = \sqrt{j \times i}, \tag{3}$$

where k represents the geometric mean, j denotes the resulting value, and i denotes the original pixel intensity value.

The geometric mean operation was performed to adjust the pixel intensity values in relation to the likelihood of occurrence in the eye socket image. This procedure is effective in reducing information loss that may occur during the conversion process. The geometric mean ensures that the pixel intensity values are properly calibrated according to their probabilistic distribution in the image. Equations (4) and (5) are the extended versions of (3).

$$k = \sqrt{p_x(i) \times i \times i}, \tag{4}$$

$$k = i \times \sqrt{p_x(i)}, \tag{5}$$

where k is the final histogram image created from the square root of the PDF of the eye socket. The square root operation is expected to produce better results in classifying natural images, such as eye socket images, compared with conventional probability methods because it allows for natural growth. However, the resulting histogram may not be uniform and may be difficult for the human eye to perceive. This makes them unsuitable for classification tasks in which uniformity is important. To address this issue, a standard histogram equalization operation was applied separately to each channel of the histogram image to achieve uniformity. Figure 7A, B presents the frequencies of the histograms extracted from the same image (presented in Figure 8) before and after histogram equalization. The histogram equalization operation employs the cumulative density function (CDF), which is calculated from the PDF value of each pixel in the histogram image k, as expressed by (6),

$$\mathrm{cdf}_x(k) = \sum_{l=0}^{k} p_x(l), \tag{6}$$

where $\mathrm{cdf}_x(k)$ represents the cumulative distribution function (CDF) of the pixel intensity up to k and $p_x(l)$ represents the probability of pixel intensity l. Summation was performed from $l = 0$ to k.

It is important to note that standard histogram equalization methods may lead to over- or under-enhancements and may produce visual artifacts. By contrast, adaptive histogram equalization methods can achieve better results. In this study, a newly introduced

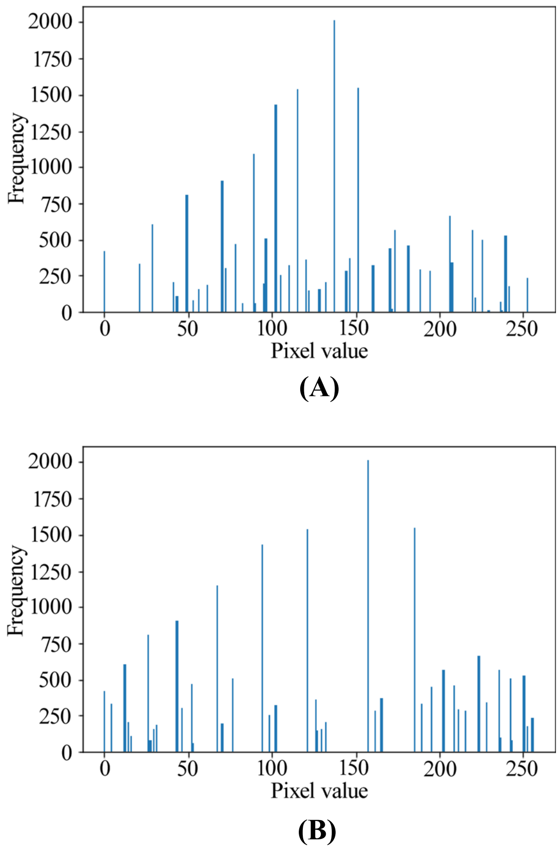

FIGURE 7 Histograms of the image presented in Figure 8 (A) Before and (B) after histogram equalization.

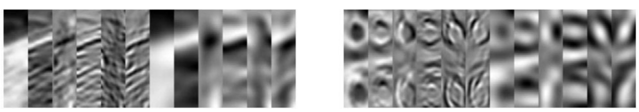

FIGURE 8 Extraction of the Gabor feature images for the same people shown in Figure 6 (from left to right).

method called low-dynamic range histogram equalization was used for contrast enhancement because it is expected to provide better results. However, because the main goal of this study was to present a robust approach for iris recognition, additional studies can be performed to explore the possibility of implementing fine-tuning procedures, such as using adaptive HE methods. After obtaining the CDF vector, which consists of 256 bins in the range [0, 255], an inverse transformation was employed using a particular transformation function $y = T(x)$, where T is the number of normalized intensity levels and y is the generated uniform histogram image (see Figure 9). Subsequently, CDF was linearized across the value range, that is, $\mathrm{cdf}_y(l) = l \times K$ for some constant K. Finally, a linear mapping operation was performed to map the values back into their original range using (7),

$$y' = y \times (\max(x) - \min(x)) + \min(x), \tag{7}$$

where y' represents the mapped value, y denotes the original value, $\max(x)$ is the maximum value in the original range of x, and $\min(x)$ is the minimum value in the original range of x.

These steps between (1) and (7) were performed for each image channel separately, and four different single-channel uniform histogram images (for red, green, blue, and alpha channels) were created. The generated images were merged into a single final image to create a four-channel RGBA output.

2.4 | Gabor filter

Gabor filters are a type of linear filter commonly used in image processing and computer vision applications. They are named after Dennis Gabor, a Hungarian physicist who won the Nobel Prize in Physics in 1971 for his work on holography. Gabor filters are banks of filters, each of which is tuned to a specific frequency and orientation. They are designed to mimic the receptive areas of simple cells in the visual cortex. By convolving an image with a bank of Gabor filters, features such as edges, textures, and corners can be extracted. Gabor filters are often used in applications such as face, fingerprint, and object recognition. They have the advantage of capturing both local and global features of an image, making them a powerful tool for feature extraction. However, they can be computationally expensive, especially when using a large number of filters. In (8) and (9), the two-dimensional (2D) Gabor filters are expressed as follows:

$$G_c[j,k] = B \cdot e - \frac{(j^2 + k^2)}{2\sigma^2} \cos(2\pi f (j\cos\theta + k \sin \sin\theta)), \tag{8}$$

where $G_c[j, k]$ represents the real part of the Gabor filter response at position (j, k), B is a constant, e is the base of the natural logarithm, σ is the standard deviation of the Gaussian envelope, f is the spatial frequency of the sinusoidal plane wave, and θ is the orientation of the sinusoidal plane.

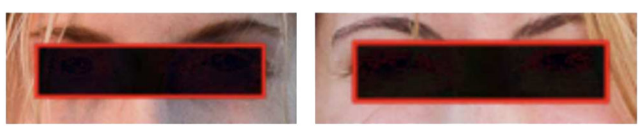

FIGURE 9 Extraction of the inverse histogram fusion images (IHFI) inside the eye socket boundaries of two people from left to right. The eye socket images are in circled in the black region.

$$G_s[j,k] = C \cdot e - \frac{(j^2+k^2)}{2\sigma^2} \cos(2\pi f(j\cos\theta + k \sin \sin\theta)), \quad (9)$$

where $G_s[j, k]$ represents the imaginary part of the Gabor filter response at position (j, k).

The use of Gabor filters in eye socket recognition provides a unique description of texture compared with other filtering algorithms, such as bilateral filtering, guided filtering, and wavelet-based filtering. Other feature extraction methods may extract features, but they do not provide a unique description, unlike Gabor features in the Gabor space, which can be utilized for eye socket pattern recognition. Once the eye socket region is segmented and Gabor features are extracted, the subsequent step is to normalize these features to generate iris codes for comparison. Normalization is necessary because there are variations in the eye socket location, size, and orientation among individuals, which require a common representation with similar dimensions. The normalized Gabor features of each eye socket produce a sample Gabor image, as shown in Figure 8.

Finally, the obtained Gabor images of the eye socket were employed for eye socket recognition by comparing the test images with the training images using normalized cross-correlation. In addition, these images were used in deep learning.

3 | EXPERIMENTAL WORK

3.1 | Hardware specification

A 24 GB graphical processing unit device (Nvidia RTX A5000) with the CUDA Toolkit (v11.3) was used throughout the experiments.

3.2 | Datasets and experimental design

In this study, four popular benchmarking datasets were used, as listed in Table 1. All datasets were obtained from the wild (Figure 10). Unknown/unprovided information is represented by the symbol "-." Although the resolution was low, the recognition system performed well. Each image in the dataset had a different resolution. When the rectangular eye pair part is extracted by ROI analysis, the area becomes almost too small and prohibits the recognition of the eye sockets efficiently.

Datasets were divided into test and training sets (80–20 split). Subsequently, the features produced using Gabor and histogram equalization were fed to a DNN learner for classification. The model was set up to run for 21000 epochs, the images were resized to 28 × 28 pixels, and normalization was applied to enable fast computation.

3.3 | Experimental results

In Table 2, the column "dataset" represents the datasets used in the experiment, the column "proposed" represents the accuracy achieved by the proposed method (i.e., Gabor features and histogram equalization), and the column "Gabor" represents the accuracy achieved by the method using only Gabor features without histogram equalization.

Owing to the nondeterministic mechanism of deep networks and the stochastic nature of the processes, the results obtained using the proposed method were obtained using 30 independent runs. However, the results presented are those obtained by the best model indicating the upper and lower bounds of the 95 % confidence interval. Conversely, because the Gabor method is deterministic, only a single result is presented.

FIGURE 10 Sample images of an exemplar participant from the CK+ dataset. As shown, different emotions affect the shape of the eye region. Note that images from Flickr30, BioID, and Mask AT & T are not shared owing to copyright reasons.

TABLE 1 Datasets used for eye socket recognition.

Name	No. of individuals	No. of images/individual	No. of images	Resolution	Specifications
BioID	23	[2, 150]	1526	384 × 286	Frontal face images
MaskedAT&T	-	-	-	-	Frontal face images with masks
Flickr30	30	-	-	500 × 500	Frontal face images
CK+	123	[10, 60]	5876	640 × 480	Frontal face images
				640 × 490	with different emotional expressions

TABLE 2 Results received by the presented method and other state-of-the-art methods on each dataset.

Dataset	Proposed	Gabor	Bilateral	Guided	Wavelength
BioID	100.0 ± 0.0	90.16	93.77	97.05	96.72
MaskedAT&T	75.0 ± 3.9	59.32	31.25	13.75	41.25
Flickr30	93.75 ± 3.3	69.23	90.0	96.67	93.33
CK+	100.0 ± 0.0	90.46	100.0	100.0	100.0

TABLE 3 Comparison of the proposed method with state-of-the-art deep networks on each tested dataset.

Dataset	Proposed	ResNet50	InceptionV3	VGG19	FaceMesh_DNN [32]	MTN [33]
BioID	100.0 ± 0.0	83.07 ± 1.6 ↓	66.54 ± 3.3 ↓	81.89 ± 1.1 ↓	97.92 ± 0.4 ↓	97.74 ± 0.2 ↓
MaskedAT&T	75.0 ± 3.9	28.81 ± 9.0 ↓	37.29 ± 3.6 ↓	35.59 ± 4.6 ↓	71.88 ± 5.6 →	56.25 ± 1.0 ↓
Flickr30	93.75 ± 3.3	48.08 ± 3.9 ↓	28.85 ± 4.0 ↓	50.0 ± 4.9 ↓	99.89 ± 0.6 ↑	61.62 ± 19.03 ↓
CK+	100.0 ± 0.0	76.53 ± 1.0 ↓	78.59 ± 0.9 ↓	76.67 ± 0.8 ↓	100.0 ± 0.0 →	100.0 ± 0.0 →

As shown in Table 2, results obtained (from all the datasets used) by the proposed method outperformed the results obtained by the Gabor method. Moreover, we performed a statistical test (two-sample unpaired t-test) to determine the significance of the differences. We found that the accuracy achieved by the proposed method outperformed that of the Gabor method ($p < 0.001$) on all four datasets and the bilateral, guided, and wavelength filtering-based methods in most cases by a significant amount.

Initially, analysis of variance was performed to test the significance of the interaction between the results. We found a significant main effect for the method [F (5, 434) $= 81.24$, $p < 0.001$, $\eta^2 = 0.65$] and dataset [F (3, 434) $= 236.92$, $p < 0.001$, $\eta^2 = 0.82$], indicating that the achieved accuracies vary across methods and datasets. We also found a significant interaction between the method and dataset [F (12, 434) $= 117.81$, $p < 0.001$, $\eta^2 = 0.87$], which indicates differences in the mean accuracies among the methods and datasets.

A post-hoc, two-sample, t-test using Bonferroni correction (corrected $\alpha = 0.0125$) was performed to test the significance of the differences in accuracy achieved by the proposed method compared with state-of-the-art deep networks.

In Table 3, the symbols "↑," "→," and "↓" are used in conjunction with each compared method to show how the proposed method performs in comparison with the compared method. The term "↓" implies that the compared method achieved a significantly lower accuracy compared with the proposed method, "→" implies either the lack of significant differences or indicates that the compared method achieved a similar result. Additionally, the term "↑" implies that the compared method achieved significantly higher accuracy compared with the proposed method.

As shown in Table 3, the performance of the proposed method is better than that of the state-of-the-art deep networks. The post-hoc, two-sample t-test showed that the accuracy achieved by the proposed method was significantly higher (all p's < 0.001) than the accuracy achieved by the state-of-the-art deep networks on all four datasets.

4 | DISCUSSIONS AND FUTURE STUDY

The proposed eye socket recognition system based on the IHFI (Figure 9) and Gabor features has yielded promising results in the evaluation using some benchmark datasets. Accordingly, it is expected that some additional studies will be conducted in this domain in the future. The performance evaluation of any recognition system relies heavily on the datasets used for testing. Although the proposed method utilizes four benchmark datasets with varying perspectives and difficulties, further expansion of the dataset could provide a more comprehensive assessment of its effectiveness. The inclusion of datasets from diverse populations, ethnicities, and age groups will help assess the generalizability and robustness of the method across different demographics.

Although the benchmark datasets used in this study covered a range of eye shapes, coverings, and angles, challenging conditions may still exist that have not been fully addressed. For instance, the presence of occlusions, such as eyeglasses or partial obstruction of the eye socket, can affect recognition accuracy. It would be valuable to investigate the performance of the proposed method under such challenging conditions and explore techniques to mitigate their effects.

This study primarily focused on identity recognition based on eye socket shapes. However, eye socket characteristics may change over time owing to factors such as aging, injury, or surgical intervention. Conducting longitudinal analyses to examine the stability and consistency of eye socket features within the same individual over an extended period would provide insights into the viability of eye socket recognition as a long-term biometric modality.

The proposed method focuses solely on eye socket shape and region without incorporating other biometric modalities. Investigating the potential benefits of fusing eye socket recognition with other biometric traits, such as iris texture or facial features, could enhance the accuracy and reliability of identification systems. Exploring multimodal fusion techniques and assessing their performance in comparison with unimodal approaches are valuable directions for future research.

It is also crucial to evaluate the real-time performance of the proposed method to facilitate its practical application. Considering the potential use of eye socket recognition in security systems, access control, and healthcare applications, future studies should aim to optimize the algorithm for efficient computation and explore hardware acceleration techniques to enable real-time processing. Therefore, this method is suitable for real-time applications.

Apart from technical aspects, the study of the user experience, user acceptance, and human factors associated with eye socket recognition is vital. Conducting user studies, feedback from end-users, and investigating usability aspects can provide valuable insights into system improvement and identify potential challenges in user adoption.

In conclusion, although the proposed eye socket recognition system achieved promising results, further research is required to address these issues. Future studies will contribute to a more comprehensive understanding of the capabilities, limitations, and potential applications of the proposed method, ultimately advancing the field of biometric identification based on eye socket shapes.

5 | CONCLUSION

This study introduces a fast and innovative method for eye socket detection that eliminates the need for iris boundaries. The proposed approach utilizes an IHFI to extract Gabor features from previously detected eye socket regions using Adaboost Haar cascade classifiers. These IHFI images captured the most probable and frequent features. The generated Gabor features were then compared with the dataset from the training set using a normalized cross-correlation. The proposed system was evaluated using different datasets including Flickr30, BioID, Masked At & T, and CK+. These datasets comprised benchmarking images of various emotions and perspectives. Additionally, some images featured individuals wearing glasses, makeup, or face masks, among other factors. The experimental results and comparative studies demonstrated that the proposed method outperformed existing state-of-the-art approaches. These findings highlight the effectiveness of eye socket recognition using IHFIs within the Gabor transform, as it exhibits high-recognition accuracy and robustness to variations in lighting and noise. This system has the potential for diverse applications in biometric identification and medical diagnosis.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Information on downloading the data used in this study can be found online (Github).

CODE AVAILABILITY

The code implemented in this study is as follows (Github).

ORCID

Harisu Abdullahi Shehu https://orcid.org/0000-0002-9689-3290

REFERENCES

1. N. Min-Allah, F. Jan, and S. Alrashed, *Pupil detection schemes in human eye: a review*, Multimed. Syst. **27** (2021), no. 4, 753–777.
2. M. Mukhiddinov, O. Djuraev, F. Akhmedov, A. Mukhamadiyev, and J. Cho, *Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people*, Sens. **23** (2023), no. 3, 1080.
3. F. Izhar, S. Ali, M. Ponum, M. T. Mahmood, H. Ilyas, and A. Iqbal, *Detection & recognition of veiled and unveiled human face on the basis of eyes using transfer learning*, Multimed. Tools Applicat. **82** (2023), no. 3, 4257–4287.
4. C.-L. Lee, W. Pei, Y.-C. Lin, A. Granmo, and K.-H. Liu, *Emotion detection based on pupil variation*, Healthcare, **11** (2023). https://doi.org/10.3390/healthcare11030322
5. I. Kayadibi, G. E. Güraksın, U. Ergün, and N. Özmen Süzme, *An eye state recognition system using transfer learning: Alexnet-based deep convolutional neural network*, Int.

J. Computat. Intell. Syst. **15** (2022), no. 1. https://doi.org/10.1007/s44196-022-00108-2

6. X. Wang, A. Haji Fathaliyan, and V. J. Santos, *Toward shared autonomy control schemes for human-robot systems: action primitive recognition using eye gaze features*, Frontiers in Neurorobot. **14** (2020). https://doi.org/10.3389/fnbot.2020.567571
7. P. S. Lamba, D. Virmani, and O. Castillo, *Multimodal human eye blink recognition method using feature level fusion for exigency detection*, Soft Comput. **24** (2020), no. 22, 16829–16845.
8. L. Zhao, Z. Wang, G. Zhang, Y. Qi, and X. Wang, *Eye state recognition based on deep integrated neural network and transfer learning*, Multimed. Tools Applicat. **77** (2018), 19415–19438.
9. V. K. Hahn and S. Marcel, *Biometric template protection for neural-network-based face recognition systems: a survey of methods and evaluation techniques*, IEEE Trans. Inform. Forensics Secur. **18** (2022), 639–666.
10. J. Z. Lim, J. Mountstephens, and J. Teo, *Emotion recognition using eye-tracking: taxonomy, review and current challenges*, Sens. **20** (2020), no. 8, 2384. https://doi.org/10.3390/s20082384
11. B. Mahanama, Y. Jayawardana, S. Rengarajan, G. Jayawardena, L. Chukoskie, J. Snider, and S. Jayarathna, *Eye movement and pupil measures: a review*, Frontiers Comput. Sci. **3** (2022), 733531. https://doi.org/10.3389/fcomp.2021.733531
12. S. Saurav, P. Gidde, R. Saini, and S. Singh, *Real-time eye state recognition using dual convolutional neural network ensemble*, J. Real-Time Image Process. **19** (2022), no. 3, 607–622.
13. S. Lee, C. Y. Low, J. Kim, and A. B. J. Teoh, *Robust sclera recognition based on a local spherical structure*, Expert Syst. Applicat. **189** (2022), 116081. https://doi.org/10.1016/j.eswa.2021.116081
14. C. K. Sharma, services across different countries, and the transitions experienced in their education sectors as well, in response to Covid-19. A holistic read on the pandemic, this book will be of interest to scholars and researchers of sociology, medical anthropology, sociology of health, pandemic and health studies, political studies, social anthropology, public.
15. D. Fitousi, N. Rotschild, C. Pnini, and O. Azizi, *Understanding the impact of face masks on the processing of facial identity, emotion, age, and gender*, Frontiers Psychol. **12** (2021), 4668. https://doi.org/10.3389/fpsyg.2021.743793
16. D. J. Carragher and P. J. B. Hancock, *Surgical face masks impair human face matching performance for familiar and unfamiliar faces*, Cognit. Research: Principles Implications **5** (2020), no. 1, 1–15.
17. H. A. Shehu, W. N. Browne, and H. Eisenbarth, *A comparison of humans and machine learning classifiers categorizing emotion from faces with different coverings*, Appl. Soft Comput. **130** (2022), 109701.
18. E. Noyes, J. P. Davis, N. Petrov, K. L. H. Gray, and K. L. Ritchie, *The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers*, Royal Soc. Open Sci. **8** (2021), no. 3. https://doi.org/10.1098/rsos.201169
19. H. M. Shahzad, S. M. Bhatti, A. Jaffar, and M. Rashid, *A multi-modal deep learning approach for emotion recognition*, Intell. Autom. & Soft Comput. **36** (2023), no. 2, 1561–1570.
20. E. Freud, A. Stajduhar, R. S. Rosenbaum, G. Avidan, and T. Ganel, *The COVID-19 pandemic masks the way people perceive faces*, Scientific reports **10** (2020), no. 1. https://doi.org/10.1038/s41598-020-78986-9
21. R. A. S. Naseri, A. Kurnaz, and H. M. Farhan, *Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach*, Appl. Soft Comput. **134** (2023), 109933. https://doi.org/10.1016/j.asoc.2022.109933
22. R. K. Bania, *Ensemble of deep transfer learning models for real-time automatic detection of face mask*, Multimed. Tools Applicat. **82** (2023), 1–23.
23. V. K. Kaliappan, R. Thangaraj, P. Pandiyan, K. Mohanasundaram, S. Anandamurugan, and D. Min, *Real-time face mask position recognition system using YOLO models for preventing COVID-19 disease spread in public places*, Int. J. Ad Hoc Ubiquitous Comput. **42** (2023), no. 2, 73–82.
24. F. Bulut, H. A. Shehu, and I. F. Ince, *Performance boosting of image matching-based iris recognition systems using deformable circular hollow kernels and uniform histogram fusion images*, J. Electron. Imag. **31** (2022), no. 5, 053036–053036.
25. https://www.bioid.com/About/BioID-Face-Database
26. F. S. Samaria and A. C. Harter, Parameterisation of a stochastic model for human face identification, (Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA), 1994, pp. 138–142.
27. F. I. Dataset. www.flickr.com/photos/thefacewemake/albums
28. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, *The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression*, (IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San- Francisco, CA, USA), 2010, pp. 94–101.
29. K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, (IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NLV, USA), 2016, pp. 770–778.
30. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, (IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA), 2016, pp. 2818–2826.
31. K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, arxiv preprint, 2014. https://doi.org/10.48550/arXiv.1409.1556
32. S. Hangaragi, T. Singh, and N. Neelima, *Face detection and recognition using face mesh and deep neural network*, Procedia Comput. Scie. **218** (2023), 741–749.
33. P. Foggia, A. Greco, A. Roberto, A. Saggese, and M. Vento, *Identity, gender, age, and emotion recognition from speaker voice with multi-task deep networks for cognitive robotics*, Cogn. Comput. **16** (2024), 2713–2723.
34. P. before and after makeup. https://104fashion.com/pictures-of-before-and-after-makeup/
35. T. face we make. https://pictures.dxtr.com/The-Face-We-Make/n-D8Ps72/TFWM-Cropped/
36. P. Viola and M. J. Jones, *Robust real-time face detection*, Int. J. Comput. Vis. **57** (2004), 137–154.
37. R. Lienhart and J. Maydt, *An extended set of Haar-like features for rapid object detection*, (Proceedings. International Conference on Image Processing, Rochestrer, NY, USA). https://doi.org/10.1109/ICIP.2002.1038171

AUTHOR BIOGRAPHIES

Harisu Abdullahi Shehu received his B.Sc. degree in Computer Engineering from Gediz University, Turkey, and was the top-ranked student in this department. He also received his M.Sc. degree in Computer Engineering from Pamukkale University, Turkey, and his Ph.D. degree from Victoria University of Wellington, New Zealand, in the same field. He serves as a reviewer for international journals, such as IEEE Transactions on Affective Computing and IEEE Transactions on Cybernetics, and for conferences, such as IEEE Congress on Evolutionary Computation and the Asian Joint Conference on Artificial Intelligence.
He is currently a Researcher with Victoria University of Wellington, New Zealand. His research interests include computer vision, machine learning, and emotion detection from facial movement patterns and following physiological changes.

Ibrahim Furkan Ince received his Ph.D. in Information Technology Convergence Design from the Graduate School of Digital Design, Kyungsung University, Busan, Republic of Korea, in 2010. He was enrolled in postdoctoral research at the University of Tokyo, Japan, between 2010 and 2012. He worked as a Chief Research Engineer at the Hanwul Multimedia Communication Co., Ltd., Busan, Republic of Korea, from May 2012 to May 2014. Additionally, he worked as an Assistant Professor of Computer Engineering at Gediz University, Izmir, Turkey, from November 2014 to July 2016. He was an Assistant Professor at Nisantasi University, Department of Computer Engineering, Istanbul, Turkey, between October 2019 and June 2024. He is currently an Assistant Professor in the Department of Software Engineering, Istinye University, Istanbul, Turkey. His research interests include image processing, computer vision, pattern recognition, and human-computer interaction.

Faruk Bulut is an Associate Professor in the Computer Engineering Department at Istanbul Esenyurt University. His research interests include machine learning, image processing, optimization methods, decision-support systems, and graph theory. Throughout his career, Bulut has assumed roles as an educator, instructor, engineer, project manager, and consultant and has contributed to numerous projects in both the private and government sectors.

How to cite this article: H. A. Shehu, I. F. Ince, and F. Bulut, *Enhancement of eye socket recognition performance using inverse histogram fusion images and the Gabor transform*, ETRI Journal (2024), 1–11, DOI 10.4218/etrij.2023-0395