Mathematical Biosciences and Engineering

Research article

Face emotion recognition based on infrared thermal imagery by applying machine learning and parallelism

Basem Assiri and Mohammad Alamgir Hossain*

Department of Computer Science, College of CS & IT, Jazan University, Kingdom of Saudi Arabia

* Correspondence: Email: alamgirhossain.nit@gmail.com.

Abstract: Over time for the past few years, facial expression identification has been a promising area. However, darkness, lighting conditions, and other factors make facial emotion identification challenging to detect. As a result, thermal images are suggested as a solution to such problems and for a variety of other benefits. Furthermore, focusing on significant regions of a face rather than the entire face is sufficient for reducing processing and improving accuracy at the same time. This research introduces novel infrared thermal image-based approaches for facial emotion recognition. First, the entire image of the face is separated into four pieces. Then, we accepted only four active regions (ARs) to prepare training and testing datasets. These four ARs are the left eye, right eye, and lips areas. In addition, ten-folded cross-validation is proposed to improve recognition accuracy using Convolutional Neural Network (CNN), a machine learning technique. Furthermore, we incorporated a parallelism technique to reduce processing-time in testing and training datasets. As a result, we have seen that the processing time reduces to 50%. Finally, a decision-level fusion is applied to improve the recognition accuracy. As a result, the proposed technique achieves a recognition accuracy of 96.87 %. The achieved accuracy ascertains the robustness of our proposed scheme.

Keywords: facial emotion recognition; infrared image; active region, machine learning; parallelism

1. Introduction

Using an infrared thermal image to recognize and analyze facial expressions has become a popular study topic in recent years [1–3]. People monitoring through visual surveillance[4], driver safety and roadside accident [5], homeland security [6] work for the military [7], firefighters [8],

medical applications diseases and diagnosis [9,10], human-computer interaction (HCI), and so on are just some of the emerging applications of infrared thermal and infrared images. The visible camera catches the normal scene, which aids in the detection of human identification [11]. The infrared thermal camera, on the other hand, uses infrared light sources, which is extremely useful in cases when there are lighting concerns, such as shades of darkness. Wearing masks [12] or classes, for example, can also assist you in discovering the hidden elements of photographs. Many studies [13,14] look at facial recognition based on the eye-tracking scheme. In addition, digital color, chemical, thermal infrared, and infrared images are used in some works [15]. In all instances, Eid et al. [16] fined that the accuracy of recognition using an infrared thermal image is higher than a digital image. Furthermore, the physical look and different facial expressions like panic [17], smiley [18], and normal face are significant since they communicate emotion. Different elements of the face and physical features such as lips appearance or blinking rate of eyes are taken into account in research for facial and expression recognition [19].

Novel infrared thermal image-based face expression detection techniques are presented in this research. It makes use of parallelism to increase precision and speed up execution. The image of the face is divided into four sections. Instead of the four components, just active regions (ARs) are employed to prepare the training and testing data set. The ARs for expression identification are the left-eye-retina, right-eye-retina, and lips. This study considers normal, joyful, terrified, sad, and astonished expressions. Figure 1(a),(b) shows color and infrared thermal images, respectively. Figure 1(c) shows the division of the 4 classes and 3 ARs.

The following are the major contributions:

- To separate frames from a movie, the recommended method for image registration is described.
- To avoid data redundancy, a centralized database called fused-image-data-mask is maintained.
- ARs were identified by considering the nose-tip as the key point of interest during the recognition process (POI). The Optimized Probability Density Function (OPDF) is used to do this assignment, which combines both PDF and time-series to achieve more precise recognition.
- When estimating pose, a weighted PDF is taken into account so that the estimation error may be quantified. This will aid us in lowering the measurement error rate.
- To speed up the procedure, the AR classification process works in tandem.
- To improve expression categorization accuracy, a decision level fusion is proposed.

The manuscript is organized as follows: Related work is described in Section 2. Sections 3 describes the proposed recognition method. Section 4 shows the experimental results. Lastly, a conclusion is addressed in Section 5.

2. Related works

Facial recognition has been both a useful and problematic subject for decades. Face tracking and identification methods for various scenarios are given. One method of recognition is to use the gaze recognition method in conjunction with a video, which has proven to be effective [20]. This approach can make use of infrared and/or digital RGB. In reality, of the two picture acquisition technologies for recognition and authentication, the infrared thermal technique is currently the more extensively utilized [21]. A small number of researchers have investigated several cutting-edge methods for identifying facial expressions from infrared thermal recordings using a thermal sensor. Infrared thermal imaging can be used to forecast the physiological impact of ordinary activities on the nervous system.

In work [22,23] to recognize facial expressions based on the facial patches, PCA, LBP, and CNN are applied in sensor networks and mobile computing. Khan [24] present a series of approaches for detecting faces and estimating based on landmark point and CNN. The work [25] applied an emergent technique for real-time face identification and emotion categorization using statistical and time-series methods. Zhang et al. [26] propose an infrared thermal video sequence for eye recognition that splits the movie into infrared thermal picture frames. Moreover, the complete face is divided into numerous sections to extract features for facial expression recognition. In work [27], a complete face picture is segmented into equal patches, from which the attributes are retrieved. Based on this data, Random Forest (RF), SVM is used to classify facial expressions. Islam et al. [28] propose scaling sub-window in face photos to fix patch sizes. Others look for active areas around the lips, nose and eyes. They make use of the idea of ROI (Region of Interest).

a. Color image   b. Infrared image

c. Deviden in 4 classes and 3 regions
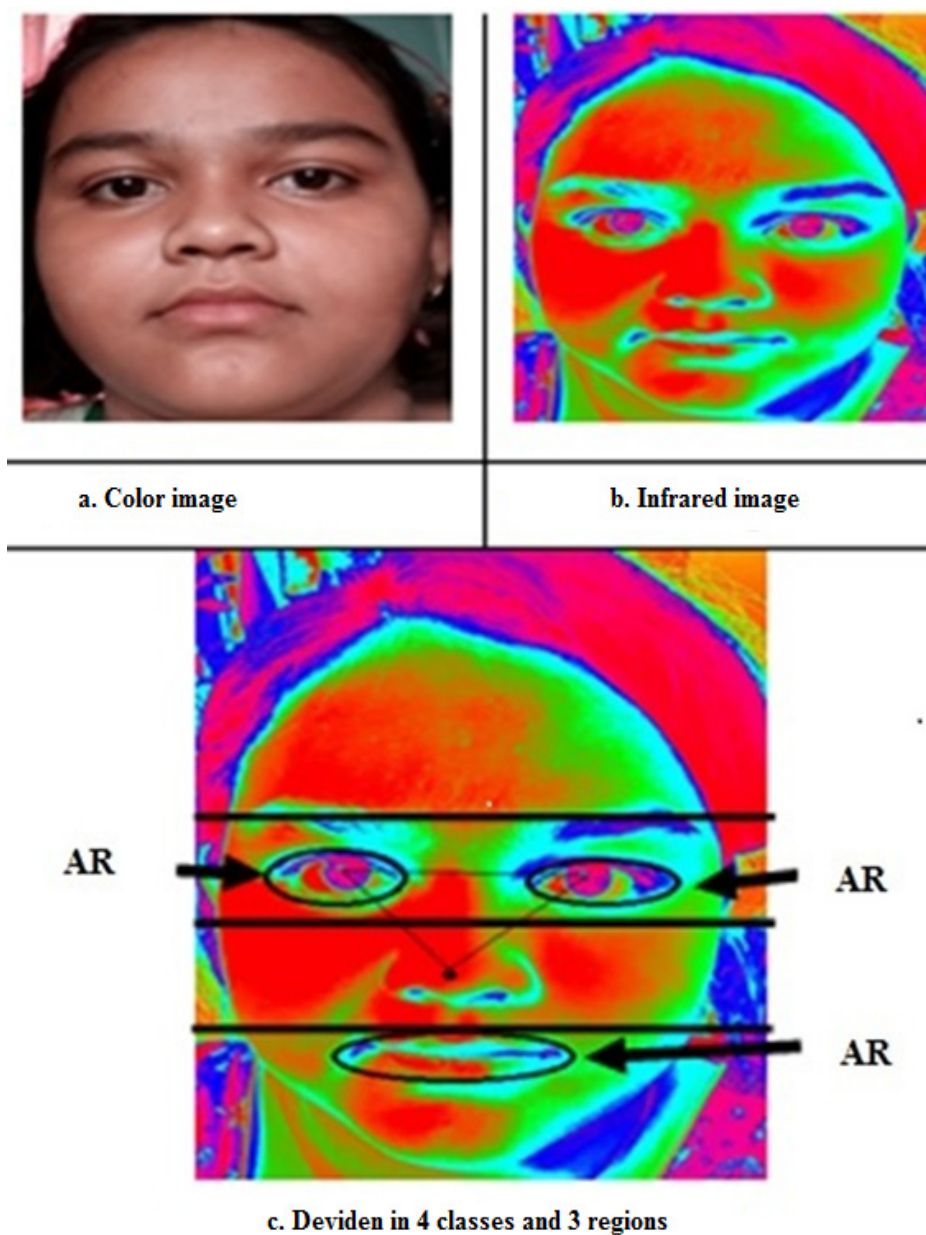
Figure 1. Shows color image, infrared thermal image division in 4 parts and the 3 ARs.

Moreover many works use machine learning-based recognition approaches. Researchers employ machine learning models such as SVM (Support Vector Machine), CNN and ANN (Artificial Neural Network) for facial identification [29]. A CNN is a deep learning approach that has a high level of performance and can extract features from training data. It achieves features from side to side using several convolutional layers and is often validated by a series of totally related layers [30]. On the CK+ database, researchers [31] raised CNN accuracy to 96.76%. Improved pre-processing processes such as sample formation and intensity normalization have improved the accuracy of the results. A Boosted Deep Belief Network (BDBN) [32] with numerous facial expression classifications was employed in another investigation. The accuracy on the CK+ database was 96.7%, whereas it was 91.8% on the JAFFE database. On the other hand, the execution period lasts for eight days.

In this study, the proposed technique saves processing time while improving recognition performance. The proposed method uses numerous pre-processing techniques in conjunction with CNN to achieve excellent accuracy. Apart from that, it is recommended that ARs focus on critical data in categorization to forecast projected expressions. This also helps to cut down on processing time. Parallelism is frequently employed to improve speed and precision.

3. Proposed methodology

The suggested method separates face photo frames from an infrared thermal video. New image frames are being added to the database. A recorded frame is separated into four sections to find ARs. The central point is then determined by detecting the tip of the nose. The tip of the nose is used to identify and distinguish other parts. A sophisticated posture estimate technique is used because poses have a direct impact on facial expression. The flow diagram is shown in Figure 2. It demonstrates how the procedure is divided into two sections. Phase-I includes image registration, pre-processing, central point (nose-tip) identification, face ARs detection, head position estimation and correction, feature weighting and extractions, keep sequence, and vector preparation and separation. A deep learning strategy is utilized in phase two to use a machine learning technique. It entails the gathering of training and testing datasets, as well as the classification procedure.

3.1. Image registration

To guarantee that an acceptable infrared image dataset series is collected from a facial image, image registration is required. To begin, the image must be registered in the database if it has not already been done so. As a result, the proposed technique can help detect picture duplication during image registration. During the image registration process, video images are converted into frames.

3.2. Pre-processing

Due to noises in the obtained images, traditional recognition methods have proved that obtaining correct portions of any part of a human face in unusual settings (such as low lighting, dark, raining period, or any sort of natural calamity) is exceedingly tough. We applied the notion of using histograms to avoid difficulties like this, which assures that the effect of any lighting does not cause too many complications. Normal and sensitive histograms can be used to implant 3D data.
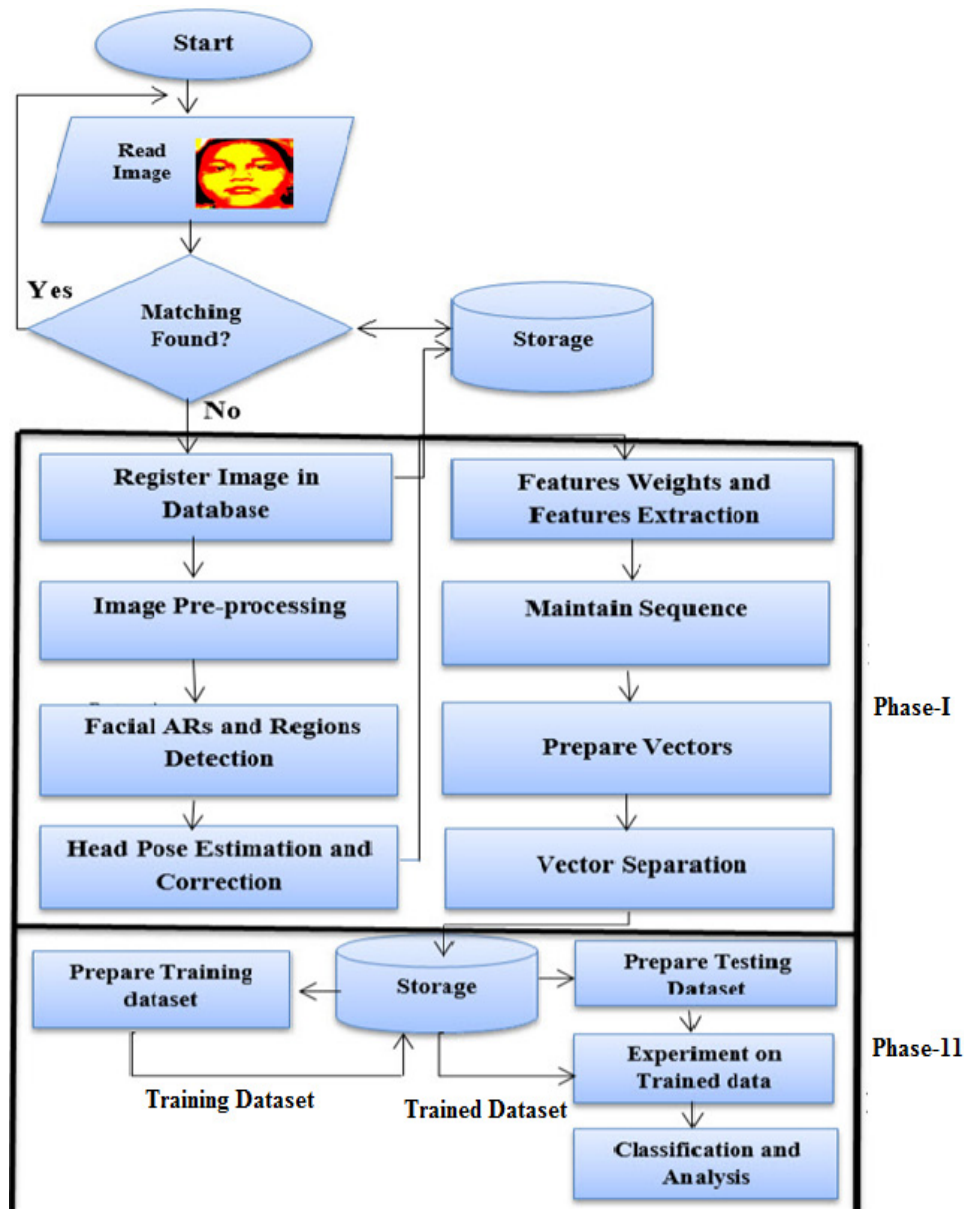
Figure 2. Proposed flow diagram.

3.3. Facial expression recognition as an application of ARs recognition

In recent years, facial expression recognition (FER) exploded as the new idea is to use a person's face for various purposes. Scholars are studying and proposing new face recognition techniques due to their growing importance. A pipeline method has been proposed by the work [13] to reduce tracking and verification costs. SHM and APDF have been used to approximate the subject's pose. Facial landmarks are key to FER methods. Lips, mouths, and brows have been used to keep tabs on people. As per researchers, facial expressions can be detected in several ways like genetic algorithm, mutation technique [33,34]. The author suggested FACS in their manuscript. From a facial landmark, AUs represented face motions. It is crucial to correctly identify all kinds of AUs while using FACS to evaluate them. Researchers [35,36] classified FER using FACS, geometric algorithms, region-based

systems, and appearance-based approaches. As with an appearance-based method, a geometry-based one should work. The work [37] has refined facial shape retrieval by dividing facial regions into blocks. These expressions correspond to face parts. Postures, positions, and shapes classify facial expressions. AROIs on a face image was discovered. Some scientists can spot smiles and panic [37]. Facial alignment improves FER. Some companies, like Microsoft and Face++, use cutting-edge technology to find facial landmarks. When identifying face landmarks, API was used. The face and body is used as input in FER. If the face is split up, it may be easier to treat. The 6×7 identical patches are shown in the image. Using these patches, Local Binary Patterns (LBP) functionality is extracted [38]. Faces are classified by SVM. With this method, it's difficult to fix the size and location of regions. Researchers [30] divided this same face into 8×8 cubes to find active areas of interest. High activity was found in the nose, mouth, eyes, and lips. We've included AROIs and created AROI as a new search algorithm. For object classification, Hossain et al. suggest Gabor-wavelet and GA. They showed that these method is more efficient than geometry. Mutation methods and GA help to increase recognition rate in FER. To recognize faces, Liu et al. used BDBN. BDBN had few poor classifiers. Each scholar's expression is recognized [30]. In their research, they tested their hypothesis on the CK+ and JAFFE databases, achieving 96.7 and 91.8% recognition accuracy, correspondingly. The BDBN's accuracy is high, but training its data sets takes longer time. Scholars combined LBP, grey value, and HOG classifier [39].

Recently Manda, Bhaskare and Muthuganapathy [40] used deep classifier to learn design models using CNN. Their experiment used GPUs to speed up calculations. New research shows that when ANNs and CNNs are used together to improve accuracy, the results are indeed stronger. Various monitoring systems can be used in image recognition. In the traditional model human movements were being identified and analyzed by using statistical method with time series [41]. According to their experiment, ANN is superior. Some strategies, such as CNN, use face recognition with poses, viewpoints, and illumination to track people. Deep learning methods like CNN extract features from training dataset. Convolution layer and sub-layers are used for features. We used an aligned layer. CNN's impressive performance has enabled researchers to use it. The results of integrating this method are introduced in a mixture. Because ML is an elevated method, CNN deep learning has been studied. It's crucial in AI and ML to (MI). Deep learning and ML methods [42] were used by many academics. Integrating CNN with pre-processing alike test generation, rotation correction, intensity normalization, etc. Using this method, the CK+ database's accuracy was improved to 96.76%. The proposed method is faster to direct connections and procedure and superior at identifying. Minimal training time leads to high correctness for the suggested CNN. Instead of the whole face, AROI classifies and identifies active regions for regular expressions. In work [43] CNN-based deep learning can recognize human faces. A CNN and LSTM-based method proposed by Rajan et al. [32] attained 81.60, 56.68 and 95.21% accurateness by MMI, SFEW, and ground truth data set.

3.4. Facial ARs and regions detection

Focusing primarily on the ARs, as shown in Figure 3, lowers costs in terms of execution time and processing overhead, as well as space and hardware management. The procedure starts with the location of the nose-center tip's point being determined. Depending on the nose-tip, other ARs, such as eyes and lips, are detected. After that, the two-phase experimental procedure is merged with a three-phase pipeline algorithm, which is expected to outperform the sequential method.
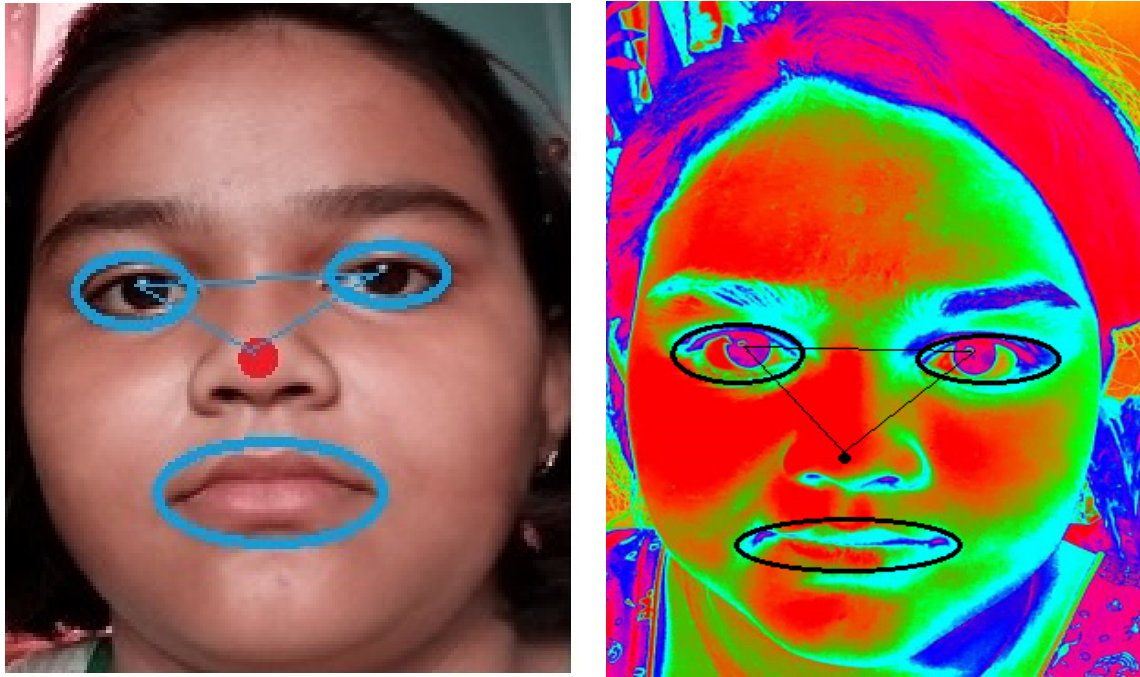
Figure 3. Recognition of ARs.

Figure 4. Dissimilar Illuminations with sensitive image histograms.

The sensitive picture histogram is exhibited as shown in Figure 4. There are four images in this set, each of which has been converted utilizing sensitive histograms and a variety of illuminations.

As previously indicated, the ARs are divided into three areas. The nose-tip is one zone that is made up to be the central point, whereas the left and right eyes along with lip regions are used for recognition, as seen in Figure 3. Figure 5 also depicts how various ARs are assessed. It recognizes landmarks before using nose-tips to identify the remaining portions. The four regions and landmark points are labelled A, B, C and D. The Gradient Minimization Method (GMM) is used to find the radius for levelling the eye and lip regions during the eye approximation. The cranny's edge detector is applied to locate eye regions. Filters are applied to landmark points to remove any superfluous edges. The last points found in the AR of the eyes, lips, and their associated features are shared. The indicated edges E1 and E2 will be estimated by applying equations one and two.

$$E_1 = \sum(P_r * Q_r) \tag{1}$$

$$E_2 = \sqrt{W_a^2 + W_b^2} \tag{2}$$

where P denotes the eye area and Q is the circular region from which the radius is calculated. Y (a, b) denotes pixel gradients measured based on (horizontal, vertical) values, where (a, b) denotes the POIs. As demonstrated in Figure 5, detection is based on landmark points and POIs.

3.5. Head pose estimation and correction

After detecting the central point and face ARs, the head pose estimation technique is used, as shown in the flow diagram (Figure 2). Inadvertently, head movement or different poses caused unwanted concerns, which have a stronger impact on the later analysis. This method helps reduce findings differences caused by head movement, as well as ensuring the validity of facial image analysis. As a result, head position is calculated using the Sinusoidal Head Model (SHM). After identifying the accurate measurements, we utilized the Optimized Probability Density Function (OPDF) to correct the head posture and then scaled them using Algorithm-2. This algorithm's name is algorithm for Pose Estimation.

3.6. Features weights extraction

As shown in Figure 5 the features' weights (FW) must now be determined, which establishes the image's critical pixels and weights. The importance of the features represented by the pixels is reflected in the weights. A set of features is used to represent an AR. The method of locating POI and important ARs such as the nose-tip, LE, RE and lip region is depicted in Figure 5. Important qualities are given larger weights to improve the dependability of information, which improves recognition.

The next step is to extract the features. The Stochastic Face Shape Model (SFSM) and Optimized Principal Component Analysis (OPCA) were used to compare patterns on original and altered pictures (after the corrections of angles). From there, the features are retrieved and expressed in vector form. When the vectors are formed, phase-I is done. It's worth mentioning that an image-based sequence should be kept. The procedure of takeout facial features and preparing vectors is described in detail below:

SFSM: In the training image sets, the number of landmark points (Figure 5) is displayed, and a new shape is created based on them. In the SFSM, there is an OPCA pattern. This is when the equation below comes in use.

$$X = X + P_S \tag{3}$$

where X represents as mean value on the shape, P contains the times (T) of Eigen vectors with biggest eigenvalues, where Si is shape parameter by limited value within the range ±3√λi and with different shapes.

$$S = (S_1, \ldots\ldots, S_n)^T \tag{4}$$
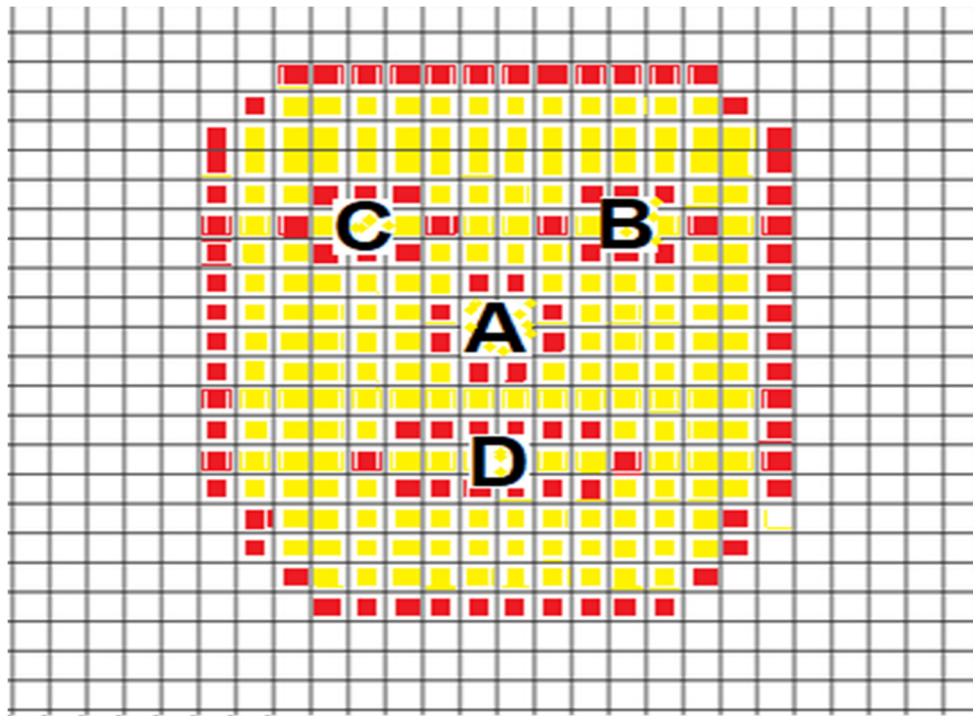
Figure 5. Finding Nose-tip (A), LE (B), RE (C) and Lip (D) regions.

Facial Feature Extraction Technique (FFET): Each point, represented as (xi, yi), will be in vector form in the procedure of the face feature extraction procedure, as demonstrated by the following equation.

$$P = (X_1, \ldots\ldots, X_n; Y_1, \ldots\ldots, Y_n)^T \tag{5}$$

It is difficult to represent 3D data obtained from the head measuring area (S). As a result, we must use the 3D to 2D conversion technique. A 3D shape equation is used in this method, and it is combined with the sinusoidal surface. The transformation of an image from a 3D video into a 2D frame is required in the conversion procedure in order to generate a vector that will be utilized for feature extraction.

Algorithm 1: Algorithm for pose estimation

1. OM = OHPEA (P$_{2D}$,FW, F)

   Input: P$_{2D}$, FW and F

   Accept Matrix (IM) for input

   Accept Matrix (OM) for output

2. N = A[P$_{2D}$,1] ; C = B[N,1]

   //Initializing by creating vector in N, C

3. IM= P$_{2Dx}$ /F; n = P$_{2Dy}$ /F;

   //Fixing start value of POI

4. Y = [P$_{3D}$, C]; O = pinv[Y]

   //Build vector of AP with values

5. while (not empty)

6. {J = O*IM; K = O*n;)

7. LOC$_z$=1/sqrt(1/||J||*1/||K||);

8. OM$_1$ = J *LOCz; OM$_2$ = K * LOCz;

9. AR$_1$ = OM $_1$(1 : 4);

10. AR$_2$ = OM $_2$(1 : 4);

11. AR$_3$ = AR$_1$/||AR$_1$|| *AR$_2$/||AR$_2$||;

12. AR$_4$ = AR$_2$/||AR$_2$|| * AR$_3$/||AR$_3$||;

13. OM $_4$ = [AR$_4$; LOCz];

14. C = Y*M$_3$/LOCz;

15. U = IM; V = n;

16. IM = C*FW*P$_{2Dx}$;

17. n = C*FW*P$_{2Dy}$;

18. E$_x$ = IM - U;

19. Ey = n –V;

20. if (||E||< E)

    { OM $_4$ = [OM $_1$(4), M$_2$(4), LOCz, 1]^{TM} ;

    break;   // Exit Loop   }

21. end of while loop

Output: OM

3.7. Classification, parallelism and by decision level fused method

A high rate of recognition accuracy is required by the new facial expression recognition technique. In this paper, CNN is used to classify ARs. Three ARs are evaluated, as shown in Figure 6, the left and right eye regions, as well as the lip region. These three ARs were classified using CNN. CNN trains and tests images using ten-folded cross-validation, in which the images are divided into ten groups. One group is chosen as testing set each time, while the rest are marked as training sets. This improves accuracy by removing the negative impacts of data set splitting.

Figure 6. Original images are in column-1. Images after regions detection are in column-2. Column-3 show left eyes detection. Fourth column shows right eyes detection and fifth column shows lip region detection.

Algorithm 2: ARs recognition

      Initialize:
a. 2D facial features
b. 3D shape & head pose OM

Start tracking from all frames

i. for (I= 1; I<= Total Frame(TF)) do

ii. Use OPDF to identify eye-region, eye-corner (Ec) & eyelid (El)

iii. Attained vector G = Ec – El

    Where we use the eyelid points(x,y) termed as Elx, Ely respectively. In doing this a mapping function of the landmark points is maintained.

iv. Identify the feature weight (FW) from P_{2D}.

v. Now get the FW and estimate points as OM = OHPEA (P_{2D}, FW, F).
    //Call algorithm 1

vi. Get the displacement points by derivation as follows: (∂Elx, ∂Ely) based on eyelid point(x,y).

vii. Now the final ARs is obtained as(S_x,S_y)

        // end of for-loop

In addition, the three ARs are classified simultaneously for each image to speed up the process, resulting in three outputs. As a result, as shown in Figure 7, a Decision Level Fusion technique is employed to obtain the outcome among the three outputs. The expressions are both labelled and converted to binary form using the One-hot code [35]. As shown in Figure 7, each AR is processed separately using CNN. At the same moment, the three ARs are processed. The findings are displayed using a five-bit one-hot coding. Each pixel represents a distinct emotion (normal, happiness, fear, sadness, or surprise). False and True are represented by the bits zero (0) and one (1), respectively. Finally, based on the three outcomes of the one-hot codes, the decision level fusion approach is used to discover the proper expression. The decision level fusion strategy (for the three parallel classification operations) determines the final findings based on a single majority of accuracy. The targeted expression is any expression that receives a bit (1) twice or more. In other words, if more than two classifiers are active at the same moment, we consider the result of the majority. When the outputs of the three classifiers differ, CNN is utilized to choose the most accurate result.

4. Experimental evaluation

During the experimentation, both qualitative and quantitative performance is assessed. The research is completed using an Intel laptop with a Core (TM) i7 CPU running at 3.60 GHz. To estimate the suggested OPDF's recognition accuracy, we use two categories of images. To begin, we take ground truth photos with an infrared thermal camera. Second, the CK+ dataset, which includes over 1000 grayscale images of ten people, is used.

In this experiment, two types of pictures are employed. With 93.98% accuracy out of thirty-five experimental images, the performance accurateness in recognizing the nose-tip is typically high enough for average color shots. The left and right eyes have 90.89 and 92.57% identification rates, respectively. Lip recognition is 93.02%, which is an excellent result. Recognition on infrared thermal

images of the nose tip is 94.35% among the training shots. The accurateness of left-eye recognition is around 91.47%, right-eye recognition is around 93.31%, and lip recognition is around 94.16%. On average, the proposed method detects 92.62% of color photographs and 93.85% of infrared images. Hossain and Assiri [2] previously estimated that the average recognition accuracy for color and infrared thermal images was 92.62 and 93.32%, respectively. These performances are presented in Table 1.
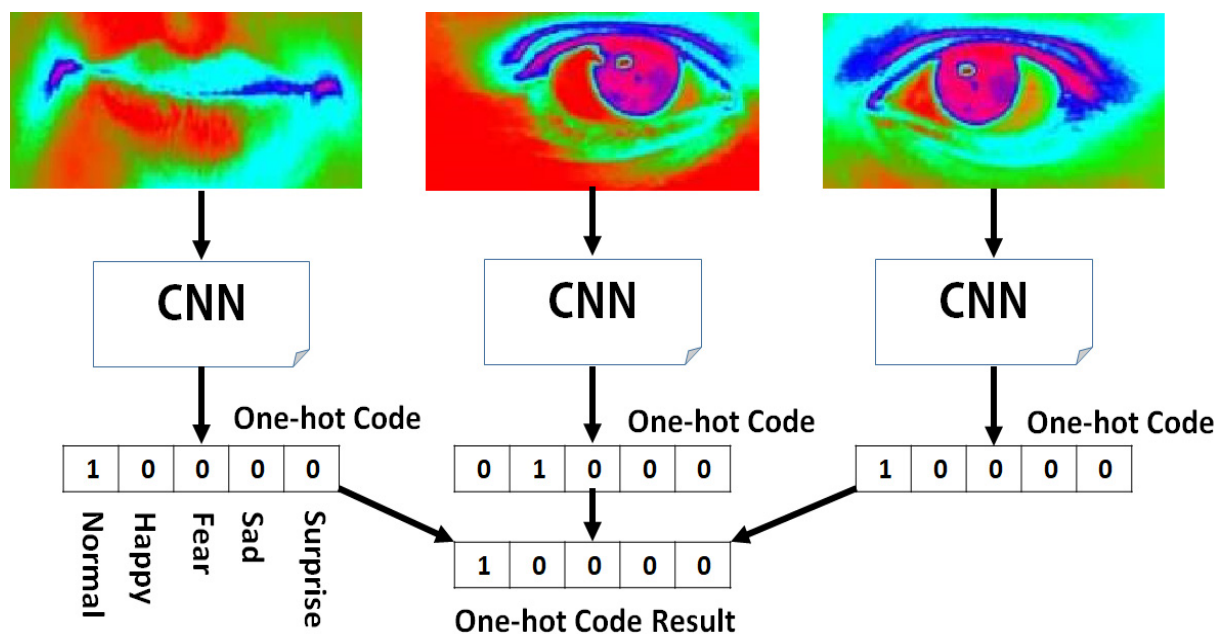
Figure 7. Decision level fusion method using 1-hot code.

Table 1. Performance accuracy of tracking.

Different Images	Nose-tip Accuracy	Left eye accuracy	Right eye accuracy	Lip accuracy	Average accuracy
Average colour image	93.98%	90.89%	92.57%	93.02%	92.62%
Average infrared image	94.35%	91.47%	93.31%	94.16%	93.32%
Proposed optimized method	94.51%	92.19%	93.71%	94.97%	93.85%

5. Conclusions

It has been discovered that using our proposed optimized method boosts the average recognition rate by about 1%. The use of infrared thermal pictures not only improves recognition rates, but also allows us to deal with unwelcome events such as natural disasters, harsh weather, low lighting conditions, and even no illumination conditions. Furthermore, it automatically tracks the ARs (left-eye, right-eye, nose-tip, and lips region). Additionally, the proposed OPDF and OHPEAs are applied to improve pose prediction. As a result, AR detection and vector information collection become

easier, with improved recognition and authentication accuracy. The experimental study employs both color and infrared photographs. On average, the proposed optimized method recognizes 92.62% of color photographs and 93.85% of infrared thermal images. Previously, Hossain and Assiri [2] proposed that the average recognition accuracy for color and infrared thermal pictures was 92.62 and 93.32%, respectively.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. A. Hossain, B. Assiri, Facial emotion verification by infrared image, in *International Conference on Emerging Smart Computing and Informatics (ESCI)*, (2020). https://doi.org/10.1109/ESCI48226.2020.9167616

2. M. A. Hossain, B. Assiri, Emotion specific human face authentication based on infrared thermal image, in *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, (2020). https://doi.org/10.1109/ICCIS49240.2020.9257683

3. M. Vollmer, K. P. Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications*, 2nd edition, WILEY-VCH Verlag, 2018. https://doi.org/10.1002/9783527693306

4. M. A. Hossain, G. Sanyal, Tracking humans based on interest point over span-space in multifarious situations, *Int. J. Software Eng. Appl.*, **10** (2016), 175–192. https://doi.org/10.14257/ijseia.2016.10.9.15

5. C. Myeon-gyun, A study on the obstacle recognition for autonomous driving RC car using lidar and infrared thermal camera, in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, (2019). http://doi.org/10.1109/ICUFN.2019.8806152

6. Q. Wan, S. P. Rao, A. Kaszowska, V. Voronin, K. Panetta, H. A, Taylor, et al., Face description using anisotropic gradient: infrared thermal to visible face recognition, in *Mobile Multimedia Image Processing, Security, and Applications*, (2018). https://doi.org/10.1117/12.2304898

7. T. Bae, K. Youngchoon, A. Sangho, IR-band conversion of target and background using surface temperature estimation and error compensation for military IR sensor simulation, *Sensors*, **19** (2019), 2455. https://doi.org/10.3390/s19112455

8. Y. Abdelrahman, P. Knierim, P. W. Wozniak, N. Henze, A. Schmidt, See through the fire: evaluating the augmentation of visual perception of firefighters using depth and thermal cameras, in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (2017), 693–696. https://doi.org/10.1145/3123024.3129269

9. M. A. Hossain, D. Samanta, G. Sanyal, Eye diseases detection based on covariance, *Int. J. Comput. Sci. Inf. Technol. Sec.*, **2** (2012), 376–379.

10. E. Sousa, R. Vardasca, S. Teixeira, A. Seixas, J. Mendes, A. Costa-Ferreira, A review on the application of medical infrared thermal imaging in hands, *Infrared Phy. Technol.*, **85** (2017), 315–323. https://doi.org/10.1016/j.infrared.2017.07.020

11. M. A. Hossain, B. Assiri, Facial expression recognition based on active region of interest using deep learning and parallelism, *Peer. J. Comput. Sci.*, **8** (2022), e894. https://doi.org/10.7717/peerj-cs.894

12. N. M. Moacdieh, N. Sarter, The effects of data density, display organization, and stress on search prformance: An eye tracking study of clutter, *IEEE Trans. Hmman Mach. Syst.*, **47** (2017), 886–895. https://doi.org/10.1109/THMS.2017.2717899

13. M. A. Hossain, B. Assiri, An enhanced eye tracking approach using pipeline computation, *Arab. J. Sci. Eng.*, **45** (2020), 1–14. https://doi.org/10.1007/s13369-019-04322-7

14. S. U. Mahmood, F. Crimbly, S. Khan, E. Choudry, S. Mehwish, Strategies for rational use of personal protective equipment (PPE) among healthcare providers during the COVID-19 crisis, *Cureus*, **12** (2020), e8248. http://doi.org/10.7759/cureus.8248

15. C. Filippini, D. Perpetuini, D. Cardone, A. M. Chiarelli, A. Merla, Thermal infrared imaging-based affective computing and its application to facilitate human robot interaction: A review, *Appl. Sci.*, **10** (2020), 2924. https://doi.org/10.3390/app10082924

16. M. A. Eid, N. Giakoumidis, A. El Saddik, A novel eye-gaze-controlled wheelchair system for navigating unknown environments: Case study with a person with ALS, *IEEE Access*, **4** (2016), 558–573. https://doi.org/10.1109/ACCESS.2016.2520093

17. M. A. Hossain, D. Samanta, G. Sanyal, Extraction of panic expression depending on lip detection, in *2012 International Conference on Computing Sciences*, (2012), 137–141, https://doi.org/10.1109/ICCS.2012.35

18. M. A. Hossain, D. Samanta, Automated smiley face extraction based on genetic algorithm, *Comput. Sci. Inf. Technol.*, **2012** (2012), 31–37. https://doi.org/10.5121/csit.2012.2304

19. S. S. Alam, R. Jianu, Analyzing eye-tracking information in visualization and data space: From where on the screen to what on the screen, *IEEE Trans. Visualization Comput. Graphics*, **23** (2017), 1492–1505. https://doi.org/10.1109/TVCG.2016.2535340

20. W. Zhang, H. Liu, Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications, *IEEE Trans. Image Process.*, **26** (2017), 2424–2437. https://doi.org/10.1109/TIP.2017.2681424

21. A. Torabi, G. Massé, G. A. Bilodeau, An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications, *Comput. Vision Image Understanding*, **116** (2012), 210–221. https://doi.org/10.1016/j.cviu.2011.10.006

22. Y. Liu, Y. Cao, Y. Li, M. Liu, R. Song, Y. Wang, et al., Facial expression recognition with PCA and LBP features extracting from active facial patches, in *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, (2016). https://doi.org/10.1109/RCAR.2016.7784056

23. W. R. Almeida, F. A. Andaló, R. Padilha, G. Bertocco, W. Dias, R. da S. Torres, et al., Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor aware loss function, *Plos One*, **4** (2020), 1–24. https://doi.org/doi.org/10.1155/2020/6385281

24. F. Khan, Facial expression recognition using facial landmark detection and feature extraction via neural networks, preprint, arXiv: 1812.04510.

25. M. A. Hossain, H. Zogan, Emotion tracking and grading based on sophisticated statistical approach, *Int. J. Adv. Electron. Comput. Sci.*, **5** (2018), 9–13. https://doi.org/12-451-152482928314-18

26. W. Zhang, X. Sui, G. Gu, Q. Chen, H. Cao, Infrared thermal imaging super-resolution via multiscale Spatio-Temporal feature fusion network, *IEEE Sensors J.*, **21** (2021), 19176–19185. https://doi.org/10.1109/JSEN.2021.3090021

27. H. Mady, S. M. S. Hilles, Face recognition and detection using Random forest and combination of LBP and HOG features, in *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, (2018). https://doi.org/10.1109/ICSCEE.2018.8538377

28. K. T. Islam, R. G. Raj, A. Al-Murad, Performance of SVM, CNN, and ANN with BoW, HOG, and image pixels in face recognition, in *2017 2nd International Conference on Electrical & Electronic Engineering (ICEEE)*, (2017). https://doi.org/10.1109/CEEE.2017.8412925

29. M. Sajjad, S. Zahir, A. Ullah, Z. Akhtar, K. Muhammad, Human behavior understanding in big multimedia data using CNN based facial expression recognition, *Mobile Networks Appl.*, **25** (2020), 1611–1621. https://doi.org/10.1007/s11036-019-01366-9

30. P. Liu, S. Han, Z. Meng, Y. Tong, Facial expression recognition via a boosted deep belief network, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2014). https://doi.org/10.1109/CVPR.2014.233

31. A. T. Lopes, E. Aguiar, A. F. D. Souza, T. O. Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, *Pattern Recognit.*, **61** (2017), 610–628. https://doi.org/10.1016/j.patcog.2016.07.026

32. S. Rajan, P. Chenniappan, S. Devaraj, N. Madian, Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM, *IET Image Process.*, **14** (2020), 1373–1381. https://doi.org/10.1049/iet-ipr.2019.1188

33. M. A. Hossain, G. Sanyal, A new improved tactic to extract facial expression based on genetic algorithm and WVDF, *Int. J. Adv. Inf. Technol.*, **2** (2012), 37. https://doi.org/10.5121/ijait.2012.2504

34. M. A. Hossain, D. Samanta, G. Sanyal, A novel approach for panic-face extraction based on mutation, in *2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, (2012), 473–477. https://doi.org/10.1109/ICACCCT.2012.6320825

35. J. Lee, S. Kim, S. Kim, K. Sohn, Multi-Modal recurrent attention networks for facial expression recognition, *IEEE Trans. Image Process.*, **29** (2020), 6977–6991. https://doi.org/10.1109/TIP.2020.2996086

36. Z. Kang, S. J. Landry, An eye movement analysis algorithm for a multielement target tracking task: Maximum transition-based agglomerative hierarchical clustering, *IEEE Trans. Hmman Mach. Syst.*, **45** (2015), 13–24. https://doi.org/10.1109/THMS.2014.2363121

37. M. A. Hossain, D. Samanta, G. Sanyal, Statistical approach for extraction of panic expression, in *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, (2012), 420–424. https://doi.org/10.1109/CICN.2012.189

38. R. Janarthanan, E. A. Refaee, K. Selvakumar, M. A. Hossain, S. Rajkumar, K. Marimuthu, Biomedical image retrieval using adaptive neuro-fuzzy optimized classifier system, *Math. Biosci. Eng.*, **19** (2022), 8132–8151. https://doi.org/10.3934/mbe.2022380

39. F. Bu, T. Pu, W. Huang, L. Zhu, Performance and evaluation of five-phase dual random SVPWM strategy with optimized probability density function, *IEEE Trans. Ind. Electron.*, **66** (2019), 3323–3332. https://doi.org/10.1109/TIE.2018.2854570

40. B. Manda, P. Bhaskare, R. Muthuganapathy, A convolutional neural network approach to the classification of engineering models, *IEEE Access*, **9** (2021), 22711–22723. https://doi.org/10.1109/ACCESS.2021.3055826

41. M. A. Hossain, G. Sanyal, A stochastic statistical approach for tracking human activity, *IJITMC*, **1** (2013), 33–42. https://doi.org/10.5121/ijitmc.2013.1304

42. A. J. A. AlBdairi, Z. Xiao, M. Alghaili, Identifying ethnics of people through face recognition: A deep CNN approach, *Sci. Prog.*, **2020** (2020), 6385281. https://doi.org/10.1155/2020/6385281

43. N. Alay, H. H. Al-Baity, Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits, *Sensor*, **20** (2020), 5523–5539. https://doi.org/10.3390/s20195523