Applied Intelligence (2023) 53:2471–2487
https://doi.org/10.1007/s10489-022-03517-0

MmLwThV framework: A masked face periocular recognition system using thermo-visible fusion

Nayaneesh Kumar Mishra[1] · **Sumit Kumar**[1] · **Satish Kumar Singh**[1]

Accepted: 15 March 2022 / Published online: 9 May 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In wake of COVID-19, the world has adapted to a new order. People have started wearing mask on their faces to prevent getting infected. The present face recognition models are no longer proving to be efficient in the current circumstances. This is because, most of the informative part of the face is covered by mask. The periocular recognition therefore holds the key to future of face recognition. However, the periocular region proves to be insufficiently enough to generate highly discriminative features. Also, most of the pre-COVID-19 algorithms fail to work in cases, where the number of training images available is very less. We propose a lightweight periocular recognition framework that uses thermo-visible features and ensemble subspace network classifier to improve upon the existing periocular recognition systems named as Masked Mobile Lightweight Thermo-visible Face Recognition (MmLwThV). The framework successfully improves the accuracy over a single visible modality by mitigating the effect of noise present in the thermo-visible features. The experiments on WHU-IIP dataset and an in-house collected dataset named, CVBL masked dataset, successfully validate the efficacy of our proposed framework. The MmLwFR framework is lightweight and can be easily deployed on mobile phones with a visible and an infrared camera.

Keywords Masked face recognition · COVID-19 · Periocular recognition · Random subspace sampling · Ensemble of networks · Thermo visible fusion

1 Introduction

Ever since the outburst of pandemic COVID-19, the whole world is struggling to overcome it. The world order has changed and people are developing new protocols to cope with the highly infectious disease. Recently the World Health Organization suggested that the world will have to learn to live with the COVID-19 disease even after the vaccine is launched. This will also involve modifying the existing systems to work with the new world protocol.

To prevent the spread of the highly infectious disease, people have to wear mask at all public places like airports. Mei Ngan et. al [32] from National Institute of Standards and Technology (NIST), recently published an exhaustive report on Face recognition accuracy with masks using pre-COVID-19 algorithms. This report contains a detailed analysis of the results obtained from 89 different face recognition algorithms from different companies while applied on masked faces. Some of the face recognition algorithms which have been tested on masked faces are [20, 30, 36, 44]. The report states that due to mask, the face recognition algorithms resulted in comparatively higher False non-match rate (FNMR) and False match rate (FMR), in comparison to when the pre-COVID-19 face recognition algorithms are applied on unmasked faces. The reason for the failure of present day face recognition algorithms on masked faces is that they require whole face image as input for recognition. However, when the mask on the face hides most of the area on the face below the eyes, the face recognition algorithms are not able to generate features that are rich enough to discriminate between the different faces. The second reason for the higher FNMR and FMR in the results of the pre-COVID-19 algorithms on masked faces is that the mask on the faces adds to the noise in the generated face features. The noise in the face features dominates the relevant face features and contributes to misclassification.

Though there are improvements in the primitive machine learning classifiers and deep learning networks for face

✉ Nayaneesh Kumar Mishra
nayaneesh@gmail.com

[1] Indian Institute of Information Technology, Allahabad, India

recognition [1, 12, 20, 31], the improvements have not been done keeping in mind the masked faces. Hence, there is an urgent need to develop a face recognition system that is able to recognise faces even with mask or in other words, we can say that there is need for a face recognition systems which can recognize faces using only the eye region. The system that recognizes a face using the eye region is called Periocular Recognition System. A Periocular recognition system has the advantage that the periocular region does not change because of various poses, aging, expression, facial changes and other artifacts. Because of this advantage, lot of research work is going on in the domain of periocular biometric recognition [11, 41]. Some have used the primitive features like local binary patterns while others have used modern deep learning models for extraction of features and its classification [3, 21, 24]. However, in all the cases, periocular recognition systems are not as robust as the full face recognition systems. The reason is very clear that due to the presence of the mask on the faces, the informative portion of the face is hidden. Robust face recognition therefore, cannot be done on the basis of features present in the periocular region alone.

Face recognition using the thermal modality is an emerging field of research [28]. This is because unlike visible images, the thermal modality is unaffected by the illumination variation and most of the occlusions (except occlusion caused by glass). Also the thermal images can be captured in dark which is not possible using the visible camera. This means that that the thermal and visible modality complement each other by providing for the missing information in the other modality. But the thermal images cannot be alone used for face recognition [6, 35] because the face detection algorithms work well only for visible images and not on thermal images. To use the thermal modality for face recognition, reconstruction of visible images from the thermal images has to be done [9, 19, 47]. This process is heavy weight and complex, and prone to error. The reconstruction of visible image from the thermal image may not always be correct because of the missing information in the thermal image required for the reconstruction of the visible image. This may lead to inconsistent face recognition results. Lot of research therefore has been done in the field of fusion of thermal and visible features [27, 37, 49]. Fusion of thermal and visible image has given good results for face recognition. This is because the both the thermal and visible modalities complement each other. Also, the process of fusion of thermal and visible features is a comparatively lightweight process when compared with the process of reconstruction of visible face from the thermal. In our survey, we did not find any research work that applied the concept of fusion of the visible and thermal modality for periocular recognition. In this paper, we therefore propose a Masked Mobile Lightweight Thermo-visible Periocular Recognition system (MmLwThV) that performs better and comparable to the existing face recognition systems by the use of fusion of thermal and visible features.

At most places such as airports, arranging large number of training images is not possible. Hence, we propose to develop a periocular recognition system that is lightweight and needs only few templates of the individual faces for recognition. Another challenge that limits the performance of the MmLwThV framework is the introduction of the noise due to mask. The noise due to mask, may affect the discriminative ability of the thermo-visible features and hence the periocular recognition accuracy. To negate the affect of noise, we propose to use ensemble of classifiers [34] along with random subspace sampling [7, 8] of the training samples in the proposed MmLwThV framework. Our proposed lightweight framework MmLwThV, can be deployed in any mobile or handheld device which has a visible and thermal camera.

Hence, to summarize, we make the following contributions in this paper:

- We proposed an efficient and robust periocular recognition framework for masked faces that uses the fusion of thermal and visible features and an ensemble of subspace networks to mitigate the effect of noise due to mask in the features. The proposed framework is called Masked Mobile Lightweight Thermo-visible Face Recognition (MmLwThV) framework.
- We collected an unconstrained, in-the-wild thermo-visible masked dataset for validation of our proposed MmLwThV framework.

2 Related work

Lot of work has been done in face recognition using visible images [20, 30, 36, 44]. However, all these works fail to work with masked faces [32]. Occlusion is one of the real world challenges that degrades the performance of face recognition. Various approaches have been proposed to handle the problem of occlusion [10, 23, 25, 26, 38, 42, 43, 46]. But all these works are based on random occlusion. These works are therefore not suited for conditions where most of the face is occluded with mask leaving only periocular region uncovered.

To overcome the challenges of face recognition in visible images, [18] and [40] reported face recognition using thermal images. Singh et.al. [40] used the wavelet domain and eigenspace domain to combine and fuse the features from visible and thermal images. He used Genetic algorithm to find an optimum fusion strategy. Madheswari et. al. [27] used feature fusion of thermal and visible images. To perform feature fusion from images of two different modalities,

the authors considered features such as discrete wavelet transform and Curvelet transform (CT). Then they employed particle swarm optimization, self-tuning particle swarm optimization and brain storm optimization algorithm to find optimal fusion coefficients. In the above such works where the features from the thermal and visible modalities were fused, finding an appropriate feature and subsequently finding an optimal fusion strategy was always a challenge. Also, we needed a feature that can be extracted from both thermal and visible image without any compatibility issues.

To reduce sensitivity to noise, illumination conditions, and facial expressions, texture analysis has time and again proven to be highly efficient. Ojala et. al. [33] introduced the local binary patterns (LBP) for capturing the texture in an image. LBP is computationally simple and yet is capable of capturing the fine details in the image. LBP has not only been used in visible images but also in thermal images for feature extraction. Several variants of LBP have been proposed like Local Derivative Patterns [5] and Local Variant Patterns (LVP) [17]. Most recently in the year 2015, in the paper[11, 41] used the Local Binary Pattern to extract feature from the eye region. The author performed bit shifting for feature matching, because the accuracy may degrade because of head movement. Dubey el. al. [15] in 2015 introduced a novel low dimensional and time efficient variant of LBP. This was called Local Bit-plane Decoded Pattern (LBDP). The LBDP encodes the relationship of each pixel in the image with its neighbours in each bit plane separately. In LBDP, encoding is done at the lowest level of an image, i.e. the bit level, LBDP can be efficiently used to capture the texture variation of an image from two different modalities. These works motivated us to use LBP and its different variants to extract features from both visible and thermal images and fuse them.

3 Datasets

To validate our proposed framework MmLwThV, we needed a dataset that contained masked and unmasked images of each subject. Each image in the dataset was required to have both the thermal and the visible modality. In our experiments for the MmLwThV framework, we needed the periocular region from both the thermal and visible images. However, cropping the periocular region from the thermal image is difficult as eye detection algorithms do not work on thermal images. The periocular region can be cropped from the thermal image using the same coordinates as that of the periocular region in the visible image, only when the thermal image is registered with the visible image. The dataset for our experiments therefore, was required to have pixel-to-pixel registration between the thermal and the corresponding visible images. WHU-IIP dataset contains registered thermal and visible face images of different subjects. However, it does not contain the masked images of subjects. So we collected and prepared a masked thermo-visible dataset and named it as CVBL Masked Face Recognition dataset. The details of the datasets are given below:

3.1 WHU-IIP dataset

WHU-IIP dataset is a thermo-visible dataset that contains thermal and visible images of 33 different unmasked subjects. For each of the 33 subjects, the dataset contains 24 thermal and 24 visible images. The thermal and the corresponding visible images are pixel-to-pixel registered. Some of the thermal and visible images from the WHU-IIT dataset are shown in Fig. 1.

3.2 CVBL masked face recognition dataset

To verify our proposed framework MmLwThV, we collected a masked dataset in unconstrained environment for face recognition. We named this dataset as CVBL Masked Face Recognition dataset. The dataset has both visible and corresponding thermal images for masked and unmasked images of all the subjects. The images from the dataset have been shown in Fig. 2. The images have been captured using Sonel KT150 Thermal Imager camera. The camera is capable of capturing both the thermal and visible images of a subject simultaneously. Both the thermal and visible images are pixel-to-pixel registered. The images have been captured in the real world environment with varying pose, illumination, resolution and distance of the subject from the camera. Because of the lighting variations, the quality of the thermal image is also varying and is not consistent throughout. This is because, the quality of thermal images is affected by the day light intensity which kept on varying during the entire session of the data collection. The visible images also exhibit wide variation in their quality. The periocular region is sometimes not very clearly captured in the visible images because of inconsistent lighting and varying pose as shown

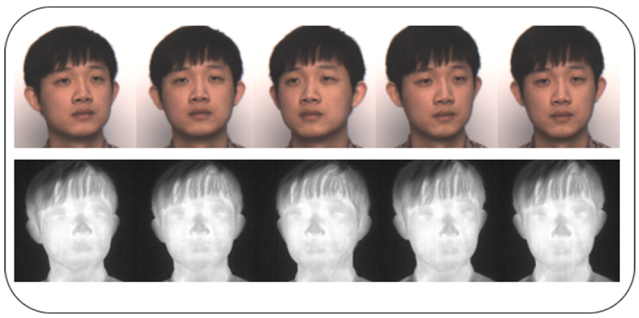

Fig. 1 Registered thermal and visible images of WHU-IIP dataset. The WHU-IIP dataset has no masked faces

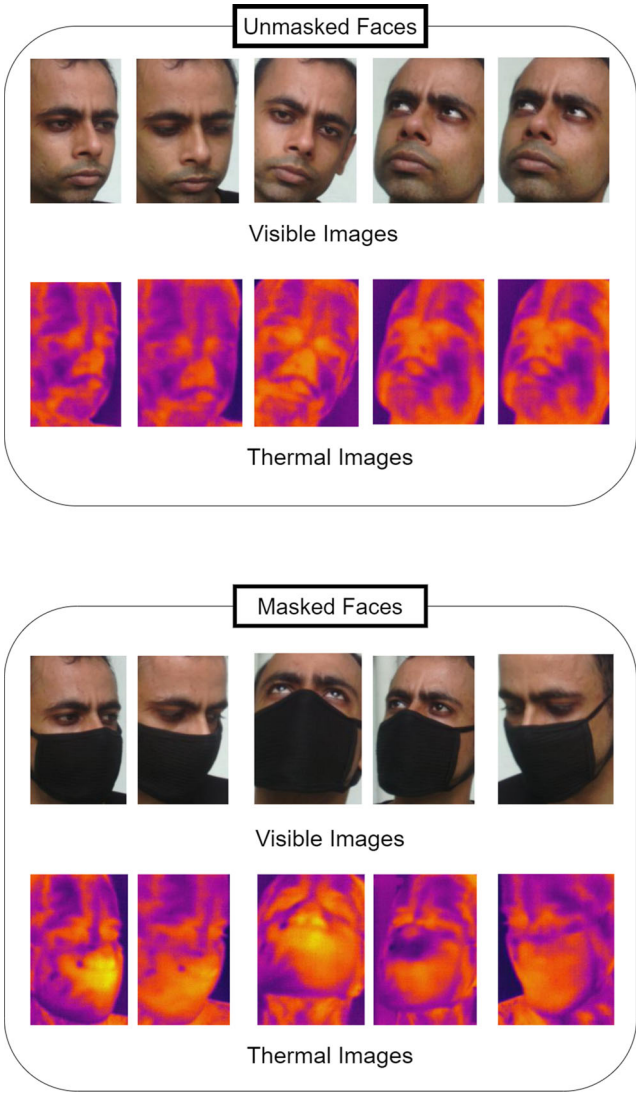

Fig. 2 Registered thermal and visible images of CVBL Masked dataset. The dataset is unconstrained and captured in real world environment. There is high variation in illumination, pose, distance of subject of the camera. The subjects have real masks on their faces. The quality of thermal images is also affected by the intensity of the day light when the image was captured

in Fig. 2. Most importantly, the masked dataset is realistic and not synthetic. The quality of masks and position of the masks on the face also varies from person to person. The number of images per subject also varies in the dataset. Thus the CVBL masked dataset represents the actual situation where the number of images available for training the classifier per subject may vary depending on the availability. Thus, CVBL masked dataset presents all such practical and real world challenges before the periocular recognition system. Thus, CVBL masked dataset will allow us to check the validity and robustness of our proposed method.

There are in total 235 visible images and 235 corresponding thermal images for 21 subjects. Thus there are on an average 11 images per subject each for visible and thermal modalities. The lowest number of images for any subject is 6 and the highest number of images for any subject is 16. In case of unmasked images, there are 236 images from the same 21 subjects. The lowest and highest number of unmasked images for any subject is 6 and 16 respectively.

4 Ablation study

4.1 Face recognition on masked faces

In this section, we have shown the experimental results when a masked face is used for face recognition using the pre-COVID-19 algorithms inspired by Local Binary Patterns (LBP) [2] and its variants. We then compared the results of face recognition on masked and unmasked full faces. The experiments have been performed on CVBL masked dataset using the thermal, visible and thermo-visible features. The results are summarized in Table 1. We used the LBP[2], LBDP [15], LDP [5] and LVP [17] features and used lightweight primitive classifiers such as Minimum Distance Classifier, Support Vector Machines (SVM) [13] and its variants, K-Nearest Neighbour(KNN) [16] and its variants, Ensemble Subspace Discriminant [4] and Ensemble Subspace KNN [48] for classification.

In all experiment, we trained the classifiers on the features from the unmasked face images. We used 5-fold cross validation to test the classifiers on unmasked images. Using the trained classifiers on unmasked faces, we also tested the classifiers on masked dataset. The results in Table 1 can be summarized in two headings: In the Table 1, we compare the corresponding values from the two rows 'M' and 'Un', for a classifier, we find that the face recognition accuracy on masked face is very less than the corresponding accuracy on unmasked face for a particular modality. The highest accuracy obtained on masked face is 27% with Ensemble Subspace Discriminant classifier and LBDP features of the masked thermal face image. The corresponding accuracy on unmasked face is 63.6%. The highest accuracy obtained on unmasked face is 92.80% when Ensemble Subspace KNN classified the thermo-visual features of LBDP for the unmasked images.

From the Table 1, it can be easily observed that, in many cases the fusion of thermal and visible features did not generate better results than the features for visible or the thermal modality when used individually. This pattern in the results is more prevalent in case of masked images than in case of unmasked images. This clearly means that when the features of thermal and visible modalities are fused together, the noise due to facial mask become more dominant than the facial features themselves. Hence,

Table 1 Comparative Results of Face Recognition for MASKED (M) and UNMASKED (Un) faces on CVBL Masked Dataset. The experiment has been performed on features from Visible (V) and Thermal (T) images

Classifiers	M/Un	LBP			LBDP			LDP			LVP		
		V	T	V+T	V	T	V+T	V	T	V+T	V	T	V+T
Minimum													
Distance	Un	81.78	81.78	81.78	87.71	72.46	92.80	71.19	58.90	84.32	76.69	55.51	86.44
Classifier	M	19.41	5.91	10.97	5.06	21.94	9.70	13.08	13.08	15.61	10.55	14.35	16.46
Linear SVM	Un	64.4	65.7	64.4	81.8	73.7	89.4	58.5	51.3	72.5	57.2	50.8	70.8
	M	21.1	4.64	7.59	10.97	21.52	15.19	13.92	14.77	12.66	12.24	16.03	12.24
Quadratic SVM	Un	67.8	67.8	68.2	84.7	77.5	90.3	64.4	56.4	77.1	67.8	55.1	79.2
	M	21.10	4.22	7.59	10.13	20.25	12.66	15.19	16.46	13.92	13.08	16.88	12.24
Cubic SVM	Un	68.2	66.9	67.4	84.3	75.8	89.4	62.3	50.0	76.3	64.8	54.2	76.7
	M	20.68	4.22	8.02	8.86	20.68	14.35	13.62	15.61	13.50	13.50	17.72	12.24
Fine	Un	7.6	8.1	7.2	9.7	11.4	7.6	6.4	7.2	6.4	6.8	6.4	6.8
Gaussian SVM	M	4.22	4.22	4.22	4.22	7.59	4.22	4.22	4.22	4.22	4.22	4.22	4.22
Medium	Un	55.5	56.8	57.2	72.5	69.1	75.8	53.0	41.5	64.8	49.2	48.3	62.3
Gaussian SVM	M	9.28	6.33	6.75	5.06	21.52	4.64	10.13	15.19	14.77	10.97	15.61	13.50
Coarse Gaussian	Un	29.2	29.2	30.1	46.2	32.6	46.6	24.2	24.6	30.9	22.9	28.0	29.2
SVM	M	15.61	3.80	8.86	6.33	20.68	13.50	9.28	12.24	13.50	6.75	11.81	17.72
Fine KNN	Un	63.1	62.3	60.6	85.6	62.3	90.3	58.5	55.1	78.0	63.6	57.2	76.3
	M	19.41	4.64	10.97	9.28	18.14	13.92	7.59	16.88	13.92	13.50	15.61	17.30
Medium KNN	Un	61.9	57.6	60.2	71.2	59.3	79.2	58.5	45.8	63.6	55.5	53.0	70.3
	M	17.30	5.91	11.39	4.64	14.77	11.81	12.66	18.57	13.50	12.24	20.25	19.41
Coarse KNN	Un	23.7	25.8	25.0	20.3	24.6	23.3	27.5	16.1	22.5	28.8	20.0	23.7
	M	18.99	6.75	14.35	5.91	16.03	9.28	6.75	12.24	10.13	8.44	17.72	14.35
Cosine KNN	Un	58.5	57.2	58.1	72.5	58.5	83.9	45.8	31.4	67.4	59.7	42.4	69.1
	M	15.19	7.17	11.39	4.22	19.83	13.92	9.28	14.77	16.03	11.39	16.88	21.70
Cubic KNN	Un	51.7	46.2	51.7	65.3	48.3	69.1	51.7	49.2	61.9	53.0	51.7	68.2
	M	15.19	5.49	11.39	3.38	13.50	7.59	13.92	18.57	14.35	12.24	16.88	17.72
Weighted KNN	Un	64.8	60.6	63.6	76.7	61.9	84.7	58.5	52.1	69.1	59.3	53.0	73.7
	M	14.77	7.17	11.81	5.49	18.14	12.24	10.97	18.99	12.24	11.39	20.25	20.25
Ensemble													
Subspace	Un	79.2	78.4	71.6	78.8	63.6	89.8	73.3	65.3	86.4	77.1	66.1	84.3
Discriminant	M	24.47	9.70	12.24	8.02	27.00	13.92	15.61	22.36	18.99	19.83	23.21	21.10
Ensemble	Un	82.6	82.6	83.1	87.7	75.4	92.8	71.2	62.7	83.1	79.7	58.1	86.4
Subspace KNN	M	22.36	5.91	10.13	3.80	22.36	12.66	13.50	15.61	15.19	11.39	14.77	16.88

face recognition accuracy on masked faces dropped down due to fusion of the features from thermal and visible modalities.

4.2 Strength of periocular recognition

In this section, we experimentally try to understand the strength of the periocular recognition. For this, we extracted the lightweight LBDP features of the full face, left eye and right eye from the unmasked images in the WHU-IIP dataset and compared the results. The results are summarized in Table 2. The predictions have been done using the minimum distance classifier which uses the Euclidean distance.

From the Table 2, we compare the results for left and right eye from that of the face. The periocular recognition for left and right eye is quite less than the face recognition

Table 2 Recognition Accuracy for different face regions for Thermal and Visible Images using Euclidean distance based prediction

Face Region	Thermal Image	Visible Image	Thermo-visible
Face	98.99	98.23	100
Left Eye	81.75	56.05	92.05
Right Eye	75.51	58.59	91.29

for the thermal or the visible modality. However, for thermo-visible modality, the periocular recognition accuracy for left and right eye become comparable to the face recognition accuracy. The accuracy for the left and right eyes are respectively 92.05% and 91.29%.

So we can conclude that periocular recognition can make a very strong contribution towards identification of a person by using thermo-visible features.

5 The proposed methodology

We describe our proposed Masked Mobile Lightweight Thermo-visible Face Recognition (MmLwThV) framework in this section. The entire framework is shown in Fig. 3. The MmLwThV framework follows the following steps for recognition of masked faces:

1. **Periocular Region Extraction**
   As diagrammatically described in the Fig. 3 the training set X contains registered masked thermal and visible images of n subjects. For each subject, the periocular region is extracted from the visible image using cropping based on coordinates provided by the eye detection algorithm. The corresponding periocular region from thermal image is extracted using the bounding box obtained from the visible image. The bounding box was obtained using the functions available in opencv.
2. **Thermo-visible Feature Extraction** After obtaining the periocular region from the thermal and visible images of a subject in the training set, the handcrafted features are then generated from periocular region in the thermal and visible modality and concatenated to generate features X_k such that k varies from 1 to n. Thus $X_1, X_2, X_3, X_4, ..., X_k$ are the fused features obtained from each subject in the training set X.
3. **Random Subspace Sampling** This is shown in block named **Random Feature Selection** in Fig. 3. The training set X is randomly sampled to produce D different samples subspaces $X^d{}_1, X^d{}_2, ..., X^d{}_k, ..., X^d{}_D$. D such random subspaces are generated for training D different classifier models.
4. **Classifier Training** Classifier Training is shown in block named **Random Feature Selection** in Fig. 3. D different models $C_1, C_2, C_3, C_4, ..., C_D$ are trained such that classifier model C_k is trained using the sample subspace X^d_k, *k lies between 1 and n*. Thus, the ensemble network consisting of D weak classifiers is trained using D random subspaces. The trained network can now be used for periocular recognition of masked faces.
5. **Periocular Recognition by the Ensemble Network** After the training, the ensemble of classifiers $C_1, C_2, C_3, C_4, ..., C_D$ is used for classification. Registered thermal and visible images of masked faces are captured and used to extract periocular region from both visible and thermal images. Features are then extracted from both thermal and visible images of the periocular region. The features obtained from thermal and visible periocular regions are then fused and is given as input to the ensemble network. The ensemble network classifies the input using the majority voting rule.

In Fig. 4, three different visible facial images I_A^V, $I_P^{V_i}$ and $I_N^{V_j}$ from a facial image dataset of size N such that $1 \leq i, j \leq N-1$ and $i \neq j$ for all i and j. I_A^V is the anchor image, $I_P^{V_i}$ is a positive image and it is having face of the person same as in I_A^V. $I_N^{V_j}$ is a negative image and contains a face image of a different person than that in I_A^V.

Similarly, I_A^T, $I_P^{T_i}$ and $I_N^{T_j}$ are three thermal images corresponding to the visible images I_A^V, $I_P^{V_i}$ and $I_N^{V_j}$ respectively such that I_A^V and I_A^T are pixel-to-pixel registered. Similarly, $I_P^{V_i}$ and $I_P^{T_i}$ as well as $I_N^{V_j}$ and $I_N^{T_j}$ are also having pixel-to-pixel correspondence.

Features of the visible images I_A^V, $I_P^{V_i}$ and $I_N^{V_j}$ are F_A^V, $F_P^{V_i}$ and $F_N^{V_j}$ respectively such that:

$$F_A^V = [F_{A_1}^V, F_{A_2}^V, F_{A_3}^V, F_{A_4}^V, F_{A_5}^V, ...F_{A_l}^V] \tag{1}$$

$$F_P^{V_i} = [F_{P_1}^{V_i}, F_{P_2}^{V_i}, F_{P_3}^{V_i}, F_{P_4}^{V_i}, F_{P_5}^{V_i}, ...F_{P_l}^{V_i}] \tag{2}$$

$$F_N^{V_j} = [F_{N_1}^{V_j}, F_{N_2}^{V_j}, F_{N_3}^{V_j}, F_{N_4}^{V_j}, F_{N_5}^{V_j}, ...F_{N_l}^{V_j}] \tag{3}$$

The distance between the anchor image I_A^V and positive image $I_P^{V_i}$ is denoted by $D_{AP}^{V_i}$ and the distance between the I_A^V and negative image $I_N^{V_j}$ is denoted by $D_{AN}^{V_j}$ where,

$$D_{AP}^{V_i} = F_A^V - F_P^{V_i} \tag{4}$$

$$D_{AN}^{V_j} = F_A^V - F_N^{V_j} \tag{5}$$

Calculating $D_{AP}^{V_i}$ and $D_{AN}^{V_j}$ using Euclidean distance, we get:

$$D_{AP}^{V_i} = \left\{ \sum_{k=1}^{l} (F_{A_k}^V - F_{P_k}^{V_i})^2 \right\}^{1/2} \tag{6}$$

$$D_{AN}^{V_j} = \left\{ \sum_{k=1}^{l} (F_{A_k}^V - F_{N_k}^{V_j})^2 \right\}^{1/2} \tag{7}$$

From the unmasked faces, we obtain discriminative features from the whole of the face. Hence, intra-class distance $D_{AP}^{V_i}$ is assumed less than inter-class distance $D_{AN}^{V_j}$. This can be written $\forall k : 1 \leq k \leq l$ as:

$$(F_{A_k}^V - F_{P_k}^{V_i}) <<< (F_{A_k}^V - F_{N_k}^{V_j}) \tag{8}$$

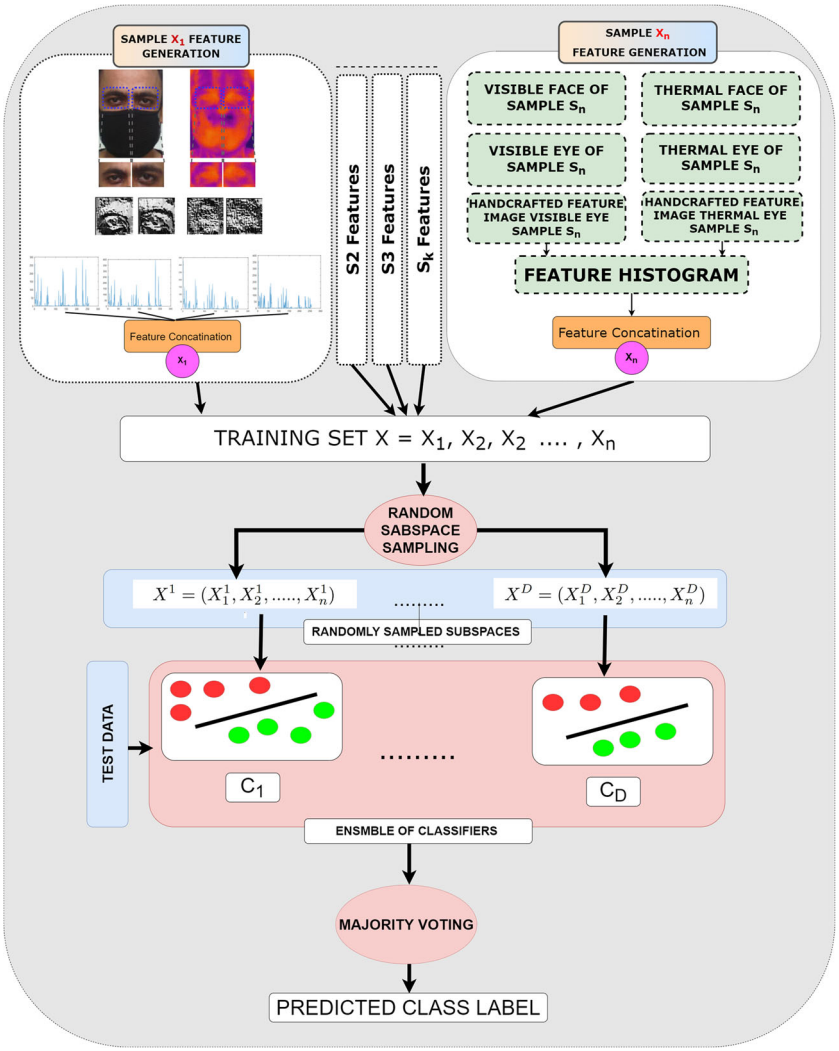

Fig. 3 Complete block diagram of MmLwThV framework

So, the classification can be done correctly as the demarcation line between the classes is clear.

5.1 Problem 1

For the masked faces, if a k^{th} feature $NF_{P_k}^{V_i}$ is a noise due to mask. Then it may affect the relation in equation (8). The equation may become:

$$(F_{A_k}^{V} - NF_{P_k}^{V_i}) >> (F_{A_k}^{V} - F_{N_k}^{V_j}) \tag{9}$$

If (9) is true for most of the k, then the result is:

$$D_{AP}^{V_i} >> D_{AN}^{V_j} \tag{10}$$

i.e. the discriminative ability of the features gets affected. The problem is to devise a method for masked face recognition so that the noise due to mask do not affect the discriminative ability of the features for correct classification.

5.2 Proposed solution stage 1: Thermo-visible fusion

We propose to use only the periocular region from the masked faces such that the noise due to mask is removed. For simplicity of nomenclature, we assume that F_A^V, $F_P^{V_i}$, $F_N^{V_j}$, F_A^T, $F_P^{T_i}$, $F_N^{T_j}$ are the thermal and visible features from the periocular region, then we get:

$$(F_{A_k}^{V} - F_{P_k}^{V_i}) \approx< (F_{A_k}^{V} - F_{N_k}^{V_j}) \tag{11}$$

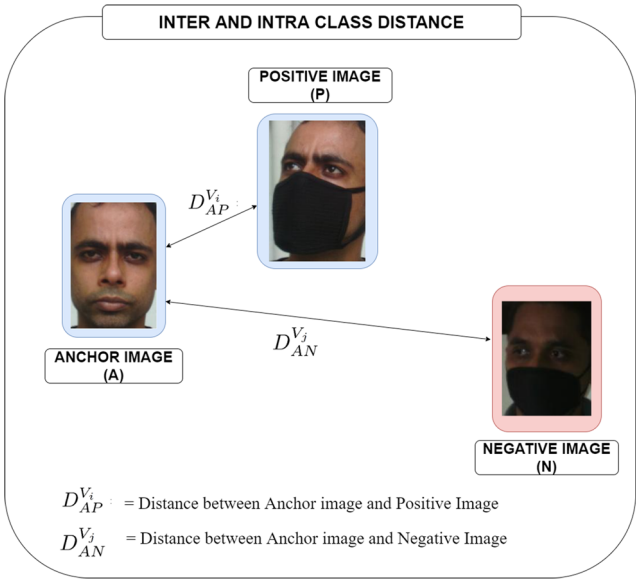

Fig. 4 The Anchor, Positive and Negative Visible Images. For correct classification, the Intra-class distance should be less than the Inter-class distance

However, since the features from periocular region are not highly discriminative, we get

$$D_{AP}^{V_i} \approx D_{AN}^{V_j} \tag{12}$$

This may lead to erratic classification results. Hence, the fusion of thermal features with the visible features is done to increase the discriminative power of the thermo-visible features of the periocular region.

Given the thermal and visible features, we fuse the features of the visible image from the features of the corresponding thermal image, then as shown in Fig. 5, the thermo-visible features can be written as:

$$\begin{aligned} F_A^{VT} = [&F_{A_1}^V, F_{A_2}^V, F_{A_3}^V, F_{A_4}^V, F_{A_5}^V, ...F_{A_l}^V, \\ &F_{A_1}^T, F_{A_2}^T, F_{A_3}^T, F_{A_4}^T, F_{A_5}^T, ...F_{A_l}^T] \end{aligned} \tag{13}$$

$$\begin{aligned} F_P^{VT_i} = [&F_{P_1}^{V_i}, F_{P_2}^{V_i}, F_{P_3}^{V_i}, F_{P_4}^{V_i}, F_{P_5}^{V_i}, ...F_{P_l}^{V_i}, \\ &F_{P_1}^{T_i}, F_{P_2}^{T_i}, F_{P_3}^{T_i}, F_{P_4}^{T_i}, F_{P_5}^{T_i}, ...F_{P_l}^{T_i}] \end{aligned} \tag{14}$$

$$\begin{aligned} F_N^{VT_j} = [&F_{N_1}^{V_j}, F_{N_2}^{V_j}, F_{N_3}^{V_j}, F_{N_4}^{V_j}, F_{N_5}^{V_j}, ...F_{N_l}^{V_j}, \\ &F_{N_1}^{T_j}, F_{N_2}^{T_j}, F_{N_3}^{T_j}, F_{N_4}^{T_j}, F_{N_5}^{T_j}, ...F_{N_l}^{T_j}] \end{aligned} \tag{15}$$

If we assume $\forall k : 1 \leq k \leq l$:

$$D_{AP_k}^{V_i} = (F_{A_k}^V - F_{P_k}^{V_i}) \tag{16}$$

we can write (6) and (7) as:

$$D_{AP}^{V_i} = \sqrt{\sum_{k=1}^{l} (D_{AP_k}^{V_i})^2} \tag{17}$$

$$D_{AN}^{V_j} = \sqrt{\sum_{k=1}^{l} (D_{AN_k}^{V_j})^2} \tag{18}$$

Similar to (4) and 5, the Euclidean distance between the fused features F_A^{VT} and $F_P^{VT_i}$, denoted by $D_{AP}^{VT_i}$ and the distance between F_A^{VT} and $F_N^{VT_j}$, denoted by $D_{AN}^{VT_j}$ is given by equations:

$$D_{AP}^{VT_i} = \sqrt{\sum_{k=1}^{l} (F_{A_k}^V - F_{P_k}^{V_i})^2 + \sum_{k=1}^{l} (F_{A_k}^T - F_{P_k}^{T_i})^2} \tag{19}$$

$$D_{AN}^{VT_j} = \sqrt{\sum_{k=1}^{l} (F_{A_k}^V - F_{N_k}^{V_j})^2 + \sum_{k=1}^{l} (F_{A_k}^T - F_{N_k}^{T_j})^2} \tag{20}$$

Squaring both sides of (6) and also of (7) and substituting in (19) and 20 respectively, we get:

$$D_{AP}^{VT_i} = \sqrt{(D_{AP}^{V_i})^2 + \sum_{k=1}^{l} (F_{A_k}^T - F_{P_k}^{T_i})^2} \tag{21}$$

$$D_{AN}^{VT_j} = \sqrt{(D_{AN}^{V_i})^2 + \sum_{k=1}^{l} (F_{A_k}^T - F_{N_k}^{T_j})^2} \tag{22}$$

When thermal features are added and concatenated with the visible features, they complete the missing information in the features. The thermal features complement the visible features. The term $(F_{A_k}^T - F_{P_k}^{T_i})$ in (21) is less than the term $(F_{A_k}^T - F_{N_k}^{T_j})$ in (22) i.e.:

$$F_{A_k}^T - F_{P_k}^{T_i} < F_{A_k}^T - F_{N_k}^{T_j} \tag{23}$$

$\forall k : 1 \leq k \leq l$. Hence, there is a small increment added to the intra-class distance $D_{AP}^{V_i}$ while, there is a larger increment in the inter-class distance $D_{AN}^{V_i}$. So, we get:

$$D_{AP}^{VT_i} <<< D_{AN}^{VT_j} \tag{24}$$

The equation clearly tells that, the fusion of the thermal and visible features, makes the fused feature more discriminative. The intra-class distance is much less than the inter-class distance.

5.3 Problem 2

In real world circumstances, the mask is randomly placed on the face. When the periocular region is cropped from the face in the thermal or visible images, a portion of the mask is left in the cropped periocular region. This is shown in Fig. 5. This small masked region in the cropped periocular part of the thermal and visible periocular image, adds to the noise and may affect the discriminative ability of features. For the recognition of the periocular region from the masked face, it is compared with periocular region of the unmasked face template. Here, the unmasked periocular template is

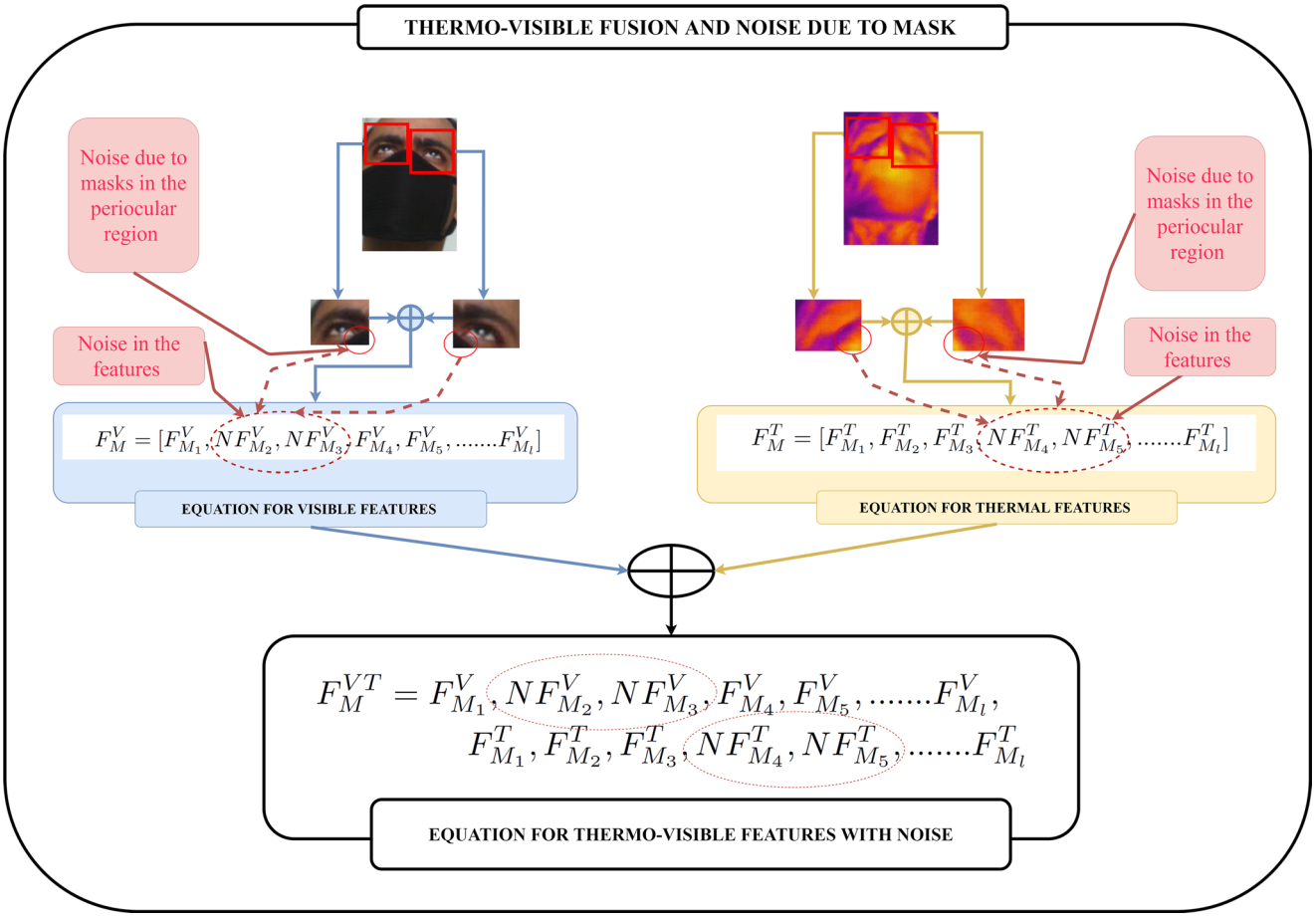

Fig. 5 The figure explains how the mask causes noise in the thermo-visible features. The noise affects the discriminative ability of the features

free from the noise due to no mask. While, the test image may be a periocular region from the masked Positive image or a Negative image.

We can therefore rewrite the (16) for noisy features as:

$$D_{AP_k}^{V_i} = (F_{A_k}^V - NF_{P_k}^{V_i}) \tag{25}$$

$$D_{AN_k}^{V_i} = (F_{A_k}^V - NF_{N_k}^{V_i}) \tag{26}$$

such that $NF_{P_k}^{V_i}$ and $NF_{N_k}^{V_i}$ are the noisy feature values from the periocular region of a positive and a negative visible image $I_P^{V_i}$ and $I_N^{V_j}$ respectively.

As stated before, if $D_{AP_k}^{V_i} < D_{AN_k}^{V_i}$ for most of values of k then the condition $D_{AP}^{V_i} < D_{AN}^{V_i}$ is satisfied and classification is correct.

However, the noisy feature values $NF_{P_k}^{V_i}$ and $NF_{N_k}^{V_i}$ can be any random value. This can affect the relation between $D_{AP_k}^{V_i}$ and $D_{AN_k}^{V_i}$ and therefore $D_{AP}^{V_i}$ and $D_{AN}^{V_i}$. The relation may transform to $D_{AP}^{V_i} > D_{AN}^{V_i}$, and subsequently result in misclassification.

The problem is to propose a method so that the noise in the features can be ignored for better classification.

5.4 Proposed solution stage 2: Random subspace sampling method and ensemble of networks

We propose to use Random Subspace Sampling Method to overcome the effect of noise in the thermo-visible features. Random Subspace Sampling Method is described below:

Let us suppose that we have a training set $X = (X_1, X_2, ..., X_n)$. Each of the training object X_i in the training set X is a p dimensional feature vector such that $X_i = (x_{i1}, x_{i2}, ..., x_{ip})$.

To construct random subspace X^d from the available training set X, we select r random features from each of the p dimensional feature vector X_i such that $r < p$. Thus the modified random subspace training set $X^d = (X_1^d, X_2^d, ..., X_n^d)$. Each of training object in random subspace training set, X_i^d for $i = 1, 2, ...n$ is a r-dimensional vector such that $X_i^d = (x_{i1}^d, x_{i2}^d, ..., x_{ir}^d)$ where each of x_{i1}^d for $i = 1, 2...r$ is selected from the p-dimensional vector $X_i = (x_{i1}, x_{i2}, ..., x_{ip})$.

We now construct an ensemble of classifiers $C^d(x)$ containing D classifiers, such that $d = 1, 2, ..., D$. Each classifier $C^d(x)$, is trained using a separate random subspace

X^d, $d = 1, 2, ..., D$. The results of each classifier $C^d(x)$ is combined using the majority voting rule. The algorithm for the training of the ensemble of classifiers using the random subspace samples from the original training set is shown in Figure 6.

The thermo-visible features of the anchor image is as shown in (13). The features of the masked face image contains noise due to mask, hence we can rewrite (13) to include the noise as shown in (27) and (28):

$$F_P^{VT_i} = [F_{P_1}^{V_i}, NF_{P_2}^{V_i}, NF_{P_3}^{V_i}, F_{P_4}^{V_i}, F_{P_5}^{V_i}, ...F_{P_l}^{V_i}, F_{P_1}^{T_i}, NF_{P_2}^{T_i}, NF_{P_3}^{T_i}, F_{P_4}^{T_i}, F_{P_5}^{T_i}, ...F_{P_l}^{T_i}] \quad (27)$$

$$F_N^{VT_j} = [F_{N_1}^{V_j}, F_{N_2}^{V_j}, NF_{N_3}^{V_j}, NF_{N_4}^{V_j}, F_{N_5}^{V_j}, ...F_{N_l}^{V_j}, F_{N_1}^{T_j}, F_{N_2}^{T_j}, NF_{N_3}^{T_j}, NF_{N_4}^{T_j}, F_{N_5}^{T_j}, ...F_{N_l}^{T_j}] \quad (28)$$

such that NF in the feature term represents the noisy feature. For example, $NF_{P_k}^{V_i}$ or $NF_{N_3}^{T_j}$, $k = 1, 2, 3, ...(2l))$, denote the noisy features. Lets say we generate a random subspace X^d, $d = 1, 2, 3, ..., D$ of dimension r, $r < 2l$, using the random subspace sampling method, then one of the sample objects, $F_P^{VT_i}$ can be expressed as:

$$F_P^{VT_i} = [F_{P_1}^{V_i}, F_{P_4}^{V_i}, F_{P_5}^{V_i}, ..., F_{P_r}^{T_i}] \quad (29)$$

$$F_N^{VT_j} = [F_{N_1}^{V_j}, F_{N_2}^{V_j}, NF_{N_4}^{T_j}, ..., F_{N_r}^{T_j}] \quad (30)$$

Because $r < S$, such that r is the dimension of the randomly selected subspace, and S is dimension of the features in the original training set, the probability that a features gets selected in the sampled subspace X^d, from the visible feature set is only r/l. Whereas, the probability that a feature gets selected in the sampled subspace X^d, from the thermo-visible feature set is only $r/(2l)$. Since $r/(l) >> r/(2l)$, we can say that fusion of thermal and visible features and random subspace sampling both together lessen the probability of appearance of noise in the final sampled subspace and hence the misclassification.

RANDOM SUBSPACE SAMPLING METHOD

1) Given a training set $X = (X_1, X_2, ..., X_n)$ such that X_i, $i = 1, 2, ..n$ and $dim(X_i) = p$
2) Iterate over d = 1,2,3... D:
   a) Construct randomly sampled subspace X^d where $X^d = (X_1^d, X_2^d, ..., X_n^d)$ such that $X_i^d \subset X_i$ for each $i = 1, 2.., n$ and $dim(X_i^d) = r$ and $r < p$
   b) Construct a classifier $C^d(x)$
   c) Train $C^d(x)$ on randomly sampled subspace X^d
3) Combine all the classifiers $C^d(x)$ $d = 1, 2, ..., D$ using the majority voting rule for the final decision.

Fig. 6 Random Subspace Sampling Method

The ensemble of networks has the following advantages in our proposed approach:

1. **Generalization property of the Ensemble**
   The ensemble of networks has a good generalization property. It takes weak classifiers and generalizes the result obtained from the weak classifiers to generate results better than the individual classifiers.
2. **Ensemble reduces the impact of noise**
   The probability of a noisy feature appearing in the randomly selected subspace is $r/(2l)$. The probability is low and hence the noisy feature may not appear in most of the sampled subspace. This causes most of classifiers networks in the ensemble to make correct prediction. The final decision about the classification, is done by majority voting. This allows to nullify the effect of noise on the final classification results of the ensemble network.

6 Implementation

To verify the efficacy of our proposed MmLwThV framework, we carried out the experiments on WHU-IIP dataset and CVBL masked face recognition dataset. As discussed before, WHU-IIT dataset has pixel-to-pixel registered thermal and visible images. But it does not have the masked images of the subjects. So using the WHU-IIP dataset, we performed the periocular recognition with visible, thermal and thermo-visible features extracted from the unmasked faces. The CVBL masked face recognition dataset contains the masked and unmasked faces in thermal and visible modalities. Hence we performed exhaustive experiments on the dataset for periocular recognition using the thermal, visible and thermo-visible features on masked and unmasked faces. We have used ensemble networks for classification. We have also used other basic classifiers such as minimum distance classifier(using Euclidean distance), Support Vector Machines(SVM), K-Nearest Neighbour(KNN) and their different versions for the purpose of comparison. The result on WHU-IIT dataset and CVBL dataset are summarized in Tables 3 and 4 respectively.

6.1 Periocular recognition using MmLwThV framework on WHU-IIP dataset

As discussed above, the WHU-IIT dataset does not contain masked faces. Hence, there is no introduction of noise due to mask in the thermo-visible features. We therefore performed the experiment on WHU-IIT dataset to validate the efficacy of fusion of thermal and visible features. We used 5-fold cross-validation for the test results. The results are summarized in Table 3.

Table 3 Results of MmLwThV framework on WHU-IIP dataset on Thermal(T) and Visible (V) Images for different Features

Classifiers	LBP			LBDP			LDP			LVP		
	V	T	V+T	V	T	V+T	V	T	V+T	V	T	V+T
Minimum Distance Classifier	97.1	97.73	98.86	83.46	95.33	97.60	49.87	26.26	54.8	91.54	54.17	91.41
Linear SVM	96.6	90.8	99.1	71.7	87.4	93.6	83.7	73.5	92.6	91.7	81.3	96.1
Quadratic SVM	97.0	90.9	99.2	73.6	86.9	93.3	85.2	75.4	94.1	92.6	83.3	96.7
Cubic SVM	96.7	89.9	98.9	70.8	84.8	91.9	84.5	73.7	93.8	92.6	82.6	96.8
Fine Gaussian SVM	5.2	5.1	4.8	6.3	6.1	6.4	3.4	2.8	3.7	3.0	2.8	3.3
Medium Gaussian SVM	95.7	90.7	98.4	74.5	83.0	90.7	82.8	68.3	92.7	92.7	81.4	96.2
Coarse Gaussian SVM	16.2	15.7	16.7	14.0	13.5	16.2	15.0	13.6	14.8	15.4	14.0	16.0
Fine KNN	74.2	46.1	73.7	59.6	76.4	85.9	39.5	19.4	44.2	71.5	34.8	74.7
Medium KNN	69.6	45.6	63.5	51.6	74.0	78.7	37.1	17.6	37.8	72.0	30.3	71.8
Coarse KNN	39.1	23.1	32.4	21.7	31.9	38.5	21.7	9.2	17.7	51.4	18.7	53.9
Cosine KNN	92.0	82.3	97.5	58.2	77.1	84.0	70.6	51.3	83.2	84.0	69.4	91.4
Cubic KNN	51.9	34.5	44.9	40.8	62.8	65.2	21.0	12.1	20.3	54.8	20.1	46.5
Weighted KNN	74.5	50.1	68.7	56.8	76.4	81.7	40.2	17.0	40.8	74.6	32.4	77.1
Ensemble Subspace Discriminant	99.1	97.2	**99.9**	68.6	85.1	88.8	91.4	76.4	96.7	98.1	89.5	98.5
Ensemble Subspace KNN	97.0	98.1	**99.1**	83.5	97.1	98.2	87.6	75.4	93.4	96.6	79.2	96.6

In order to evaluate different classifiers, it is necessary to understand if there is any statistical difference in the results of any two classifiers. If the results of any two classifiers are statistically different, then the two classifiers can be compared for better performance. If the results are statistically same, then we can say that the behavior of the

Table 4 Results of MmLwThV framework on CVBL Masked dataset on Thermal(T) and Visible (V) Images for different Features

Classifiers	LBP			LBDP			LDP			LVP		
	V	T	V+T	V	T	V+T	V	T	V+T	V	T	V+T
Minimum Distance Classifier	65.11	26.81	65.95	56.17	34.47	53.19	36.60	20.43	34.04	36.6	31.49	42.98
Linear SVM	45.11	29.36	42.98	31.91	28.51	37.02	34.04	21.70	34.89	34.04	25.53	34.47
Quadratic SVM	50.64	29.79	45.53	35.32	27.23	38.72	39.15	24.26	40.85	39.15	29.36	42.55
Cubic SVM	50.64	30.21	44.68	37.45	29.79	39.15	38.30	22.98	40.85	37.02	26.81	41.28
Fine Gaussian SVM	4.26	4.26	4.26	4.26	8.09	4.26	4.26	4.26	4.26	4.26	4.26	4.26
Medium Gaussian SVM	41.28	23.83	32.77	20.43	23.83	22.98	28.94	23.40	28.94	32.77	29.79	37.02
Coarse Gaussian SVM	17.02	13.19	17.02	17.02	20.00	19.57	10.64	16.17	17.45	11.49	16.17	17.87
Fine KNN	36.17	22.55	33.62	48.09	30.64	39.15	29.79	24.68	32.77	29.79	31.91	36.17
Medium KNN	34.47	22.55	34.89	37.87	30.21	35.32	31.91	25.11	33.62	32.77	28.51	36.60
Coarse KNN	18.30	14.89	17.02	12.77	15.74	16.60	12.34	14.89	25.96	14.47	14.47	19.57
Cosine KNN	42.13	33.62	45.53	40.00	25.96	40.43	34.04	22.13	34.04	31.91	27.23	34.89
Cubic KNN	23.40	23.40	28.09	34.04	21.28	29.36	30.64	24.26	31.06	31.91	27.66	36.17
Weighted KNN	33.62	23.83	37.02	40.00	32.34	40.00	34.04	25.11	31.06	32.77	29.36	38.30
Ensemble Subspace Discriminant	48.94	36.60	**63.40**	38.72	28.94	51.06	48.09	37.87	51.49	48.94	40.43	54.47
Ensemble Subspace KNN	68.94	28.94	**70.64**	54.47	35.74	54.47	36.17	22.13	36.60	37.87	30.64	42.98

two classifiers are same. Hence, we analysed the results of MmLwThV framework on WHU-IIT dataset using the Wilcoxon Signed rank test [14, 39].

6.1.1 Wilcoxon signed-ranks test

The Wilcoxon signed-rank test [14, 39] is a non-parametric statistical test. This means that the data is derived from a population that is non-parametric in nature. A non-parametric population can be ranked but does not have numerical values. The Wilcoxon signed-rank test is used to determine if two or more sets of pairs are different from one another in a statistically significant manner. The Wilcoxon sign test is performed with the assumption that the two samples are dependent observations of a case. The second assumption is that the paired observations are randomly and independently drawn from the population.

6.1.2 Evaluation of Wilcoxon signed-ranks test on WHU-IIT dataset

From the Table 3, we can see that Ensemble subspace discriminant is having the highest accuracy of 99.9%. Hence we compare Ensemble subspace discriminant with every other classifier as specified in Table 3. The comparison is done using Wilcoxon signed rank test to find if the results of Ensemble subspace discriminant is statistically different from the results of other classifiers. The results of the Wilcoxon Signed-rank Test are summarized in Table 5 where each column presents the results obtained from the comparison of Ensemble subspace discriminant with a different classifier. In the first column of Table 5, for example, Ensemble subspace discriminant is compared with Ensemble subspace KNN.

From the results in Table 5, it is very much clear that the results of the classifiers Ensemble subspace KNN, Minimum Distance classifier, Linear SVM and Quadratic SVM are similar to that of the Ensemble subspace discriminant classifier. The results of all other classifiers are different from that of the Ensemble subspace discriminant. Hence we can say that the Ensemble subspace discriminant shows a better performance in comparison to most of the classifiers on WHU-IIT dataset.

6.1.3 Results analysis and discussion

From the Table 3, it can be observed that, the results of MmLwThV framework on WHU-IIP dataset is highly encouraging. The highest accuracy for periocular recognition obtained is 99.9% for the LBP thermo-visible features using the Ensemble subspace discriminant classifier. The accuracy of 99.9% is comparable to any state-of-the-art result. For LBP feature, the periocular recognition accuracy of the Ensemble Subspace Discriminant for the visible modality is 99.1% and 98.1% in thermal mode. The fusion of thermal and visible LBP features increased the periocular recognition accuracy to 99.9% by the Ensemble Subspace Discriminant. Thus the result validates our proposed framework. Also, the ensemble subspace classifiers performs the best among all the classifiers. Ensemble subspace classifier gives the highest accuracy in each column of the Table 3.

6.2 Periocular recognition using MmLwThV framework on CVBL masked face recognition dataset

We performed experiments on CVBL masked dataset. Since both the masked and unmasked images are available in the dataset, we cropped the periocular region from the registered thermal and visible images for masked and unmasked faces. For the masked faces, features were extracted from the periocular region of both thermal and visible images and subsequently fused with each other. The same was done with features of the periocular regions of the unmasked images. We then trained the classifier on the thermo-visible features from the periocular region of the unmasked faces and tested the classifier on the thermo-visible features from the periocular regions of the masked faces. To understand the effect of fusion of thermal and visible features, we conducted separate experiments for face recognition using only thermal images for masked and unmasked faces and also face recognition using only visible images for masked and unmasked faces. The results are summarized in Table 4

6.2.1 Evaluation of Wilcoxon signed-ranks test on CVBL masked dataset

To see if the results of Ensemble Subspace KNN are same as or different from other classifiers as specified in Table 4, we conducted Wilcoxon Signed-ranks Test. The results of the Wilcoxon Signed-ranks Test are summarized in Table 6. In the first column for example, Ensemble subspace KNN is compared with Ensemble subspace discriminant.

From the results in Table 6, it is very much clear that the results of the classifiers Ensemble subspace discriminant, Quadratic SVM and Cubic SVM are similar to that of the Ensemble subspace KNN. The results of all other classifiers are different from that of the Ensemble subspace KNN. Hence we can say that the Ensemble subspace KNN shows a better performance in comparison to most of the classifiers on CVBL dataset.

Table 5 Results of Wilcoxon Signed-ranks Test for the MmLwThV framework on WHU-IIT masked dataset. Ensemble Subspace Discriminant is compared with all other classifiers. The other classifiers are represented by the letter in brackets. Ensemble subspace KNN (B), Minimum distance classifier (C), Linear SVM (D), Quadratic SVM(E), Cubic SVM (F), Fine Gaussian SVM (G) , Medium Gaussian SVM (H), Core Gaussian SVM (J), Fine KNN (K), Medium KNN (L), Course KNN (M), Cosine KNN (N), Cubic KNN (O), Weighted KNN (P). Also, S stands for similar and D stands for Different

Parameters	B	C	D	E	F	G	H	J	K	L	M	N	O	P
R+	44	56	62	59.5	66	78	68	78	78	78	78	78	78	78
R-	34	22	16	18.5	12	0	10	0	0	0	0	0	0	0
Stats	34	22	16	18.5	12	0	10	0	0	0	0	0	0	0
T	S	S	S	S	D	D	D	D	D	D	D	D	D	D

6.2.2 Results analysis and discussion

The results on CVBL masked dataset are a more realistic evaluation of the MmLwThV framework. Firstly because, as already discussed, the CVBL dataset has been captured in real world circumstances and therefore presents real world challenges for periocular recognition. Also, the masks in the dataset are real and no synthetic masks have been used. The second reason why the results on CVBL dataset are realistic is because the unmasked images from the dataset are used for training the classifiers and the masked images are used for testing the trained classifiers. This is unlike in case of WHU-IIT dataset, where we used the unmasked images for both training and testing of the classifiers using 5-fold validation.

As shown in the Table 4, the highest periocular accuracy on masked faces is 70.64%. The accuracy has been obtained using the Ensemble subspace KNN on the thermo-visible LBP features. The second in position of periocular recognition accuracy is Minimum Distance classifier on LBP thermo-visible features with 65.95%. Subsequently, the third rank in the order of accuracy is held by ensemble subspace classifiers on thermo-visible features indicates the efficacy of the classifier on thermo-visible fused features.

The efficacy of the ensemble subspace classifiers are further validated when we analyse the Table 4 column wise. When we say ensemble subspace classifiers, we mean either by Ensemble subspace Discriminant or Ensemble subspace KNN. For all the columns in the Table 4, the highest periocular accuracy is given by the ensemble subspace classifiers. There is one exception to this however, when the Minimum Distance classifier gives the highest accuracy of 56.17% on LBDP features and Ensemble subspace KNN follows next with an accuracy of 54.47% accuracy. But this happens for the visible modality. But for all the thermo-visible features, the highest periocular accuracy is given by the ensemble subspace classifiers.

In summary, the experiments on CVBL masked dataset re-validate all our conclusions drawn on results of MmLwThV framework on WHU-IIT dataset.

6.2.3 Effect of noise on the results of MmLwThV framework

In our paper, we have performed our experiments on two different datasets: WHU-IIT dataset and CVBL masked dataset. WHU-IIT dataset is a dataset that contains images without mask. On the other hand, CVBL masked dataset contains images that contain images with real masks. The images in CVBL dataset are captured in real world environment.

Hence, the periocular region from the images in WHU-IIT dataset contain no noise due to mask. However, periocular images from the images in CVBL dataset contain noise due to mask. This is because while the periocular region is

Table 6 Results of Wilcoxon Signed-ranks Test for the MmLwThV framework on CVBL masked dataset. Ensemble Subspace KNN is compared with all other classifiers. The other classifiers are represented by the letter in brackets. Ensemble subspace KNN(A), Ensemble subspace discriminant (B), Minimum distance classifier (C), Linear SVM (D), Quadratic SVM(E) , Cubic SVM (F), Fine Gaussian SVM (G) , Medium Gaussian SVM (H), Core Gaussian SVM (J), Fine KNN (K), Medium KNN (L), Course KNN (M), Cosine KNN (N), Cubic KNN (O), Weighted KNN (P). Also, S stands for similar and D stands for Different

Parameters	B	C	D	E	F	G	H	J	K	L	M	N	O	P
R+	29	56.5	77	54.5	61.5	77	76	78	75	75.5	78	72	77	75
R-	49	9.5	1	23.56	16.5	0	2	0	3	2.5	0	6	1	3
Stats	29	9.5	1	23.5	16.5	0	2	0	3	2.5	0	6	1	3
T	S	D	D	S	S	D	D	D	D	D	D	D	D	D

extracted from the faces in CVBL dataset, a portion of the mask is present in the periocular images.

Comparison of the results in Tables 3 and 4 clearly demonstrates the effect of noise due to mask. The highest accuracy of 99.9% in case of WHU-IIT dataset in comparison to 70.64% in case of CVBL dataset clearly demonstrates the effect of noise in the images due to mask. The effect of noise can also be understood by comparing cells in the Tables 3 and 4. It can be understood that the accuracy of a classifier on a feature in case of CVBL dataset is less than the corresponding accuracy of the same classifier on the same feature in case of WHU-IIT dataset. This means that the noise due to masks negatively impacts the accuracy of face recognition.

On analysing the results column wise in Table 4, we can see that the accuracy of either Ensemble subspace discriminant or Ensemble subspace KNN is highest. This clearly means that drop in the accuracy due to noise, is more in case of all other classifiers except Ensemble subspace discriminant or Ensemble subspace KNN. Hence we can say that, our proposed MmLwThV framework, is less affected due to noise in comparison to other classifiers. We can therefore conclude that MmLwThV framework is robust against noise to mask.

6.2.4 Complexity of the MmLwThV framework

For SVM based classifiers in the Table 1,the number of parameters in a Support Vector Machine (SVM) is equal to the number of pixels in the input image. Since we input handcrafted features such LBP, LDP etc in the classifier, the largest input size is 4096. Hence, we can say that largest number of trainable parameters in an SVM can be equal to 4096.

As we know that, the discriminant analysis classifier learns coefficients for projecting a sample into the correct class. If there are k classes, $k \times k$ structure of coefficient matrices are learnt by the classifier. In an Ensemble Subspace Discriminant classifier, there are 30 such Discriminant classifiers. Also, there are 33 classes in WHU-IIP dataset and 21 classes in CVBL Masked Face recognition dataset. Therefore, total number of parameters for WHU-IIP dataset can be calculated as 32,670. Similarly, for CVBL Masked Face recognition dataset total number of learnable parameters are 13,230. For an Ensemble Subspace KNN, there are no trainable parameters. This is because in KNN there is no requirement for training.

Now when we compare the Ensemble Subspace KNN or Ensemble Subspace Discriminant classifier any SVM based classifier, then we can say that in general all the classifiers mentioned in Table 4, are actually machine learning based classifiers. All the machine learning based classifiers are lightweight. This can be better understood if any of the machine learning based classifier is compared with any deep learning model. This is because a deep learning model contains at least a million of parameters.

If we compare Ensemble based classifiers with SVM based classifiers based on the number of learnable parameters, it may occur that Ensemble based classifiers are more complex. But it must be mentioned that the increase in the number of parameters is not very high. Also, a little increase in complexity of ensemble based classifiers is because of more number of learners in the model. The increase in the parameters of ensemble based classifiers comes with increase in robustness against noise and generalization capability. Hence we can finally conclude that our proposed MmLwThV framework is a lightweight solution for the masked face recognition.

6.2.5 Summary of discussion on MmLwThV framework

Putting together the results on WHU-IIT dataset, we can say that periocular recognition can be effectively used instead of full face identification if the visible features are used with the fusion of thermal features as well. However, upon the fusion of the thermal and visible features, it is likely that the noise due to mask can dominate the recognition accuracy and increase the false reject and false accept rates. The Ensemble subspace classifiers have been effectively able to combat the effect of noise and generalize well over the features from the periocular region.

Thus experimentally it is validated that our proposed MmLwThV framework is highly effective in improving the robustness and accuracy of the masked periocular recognition over the existing visible periocular recognition systems. MmLwThV framework accomplishes this by using the ensemble subspace networks over thermo-visible features.

7 Comparison with the state-of-the-art methods

Because of the advent of COVID-19 in recent past, little work has been published in the domain of masked face recognition. Diaz et al. [22] used deep learning networks for feature extraction from the periocular region of the masked faces. The features were then used for classification using Euclidean distance measure. Li et al. [29] used attention based network to recognize faces with masks. Li et al. [29] used spatial and channel attention modules within the deep learning networks. The attention modules within the deep learning network forced the network to give attention to those areas of the input face that could help in generation

Table 7 Comparison with state-of-the-art methods

Method	Accuracy in %
Alexnet [22]	42.55
GoogleNet [22]	43.83
VGG-16 [22]	39.15
ResNet-50 [22]	61.70
ResNet-18 [29]	55.31
ResNet-34 [29]	61.27
ResNet-18 [45]	15.74
ResNet-34 [45]	9.78
MmLwThV framework	70.64

of discriminative features. Another work by Wu et al. [45] used pyramidal attention module, apart from the spatial and channel attention modules for masked face recognition.

We implemented the above discussed works on our CVBL masked face dataset. Since the CVBL masked dataset is too small with only 21 classes, it cannot be used to train a deep learning network from scratch. Hence, we used pretrained deep learning models for our purpose and retrained them using transfer learning on CVBL masked face dataset. The fine-tuning of the deep network was carried for 100 epochs. The results are summarized in Table 7.

From the Table 7, it can be easily concluded that state-of-the-art methods perform poorly on the challenging CVBL masked face dataset. The reason is that the CVBL masked face dataset is a challenging real world dataset collected in real world environment. Such a dataset must be large enough to train a deep learning network properly. However, since the CVBL masked dataset is small and not enough to train a deep learning network, the results for masked face recognition are poor. Hence, we can say that our proposed method is an efficient, robust and a lightweight method for masked face recognition in case of a small dataset.

8 Conclusion

For COVID-19 like scenarios, we proposed a novel framework for periocular recognition which is robust and does not require much data for training. The proposed MmLwThV framework fuses the thermal and visible features from the periocular region and classifies it using ensemble subspace network. We used the ensemble subspace network for classification because of its ability to generalize and ignore the presence of noise due to mask. We tested our proposed MmLwThV framework on two thermo-visible datasets, WHU-IIT and CVBL masked dataset. On both the datasets, MmLwThV framework successfully improved the results of periocular recognition over the visible or thermal modality by using thermo-visible fused features and ensemble subspace network. We obtained the highest accuracy of 99.96% on the WHU-IIP dataset using the LBP feature. The result is comparable to any state-of-the-art face recognition systems. Again, on unconstrained and challenging CVBL masked face dataset, the MmLwThV framework successfully increases the accuracy of visible periocular recognition from 68.94% to 70.64%. Moreover, the MmLwThV framework makes the periocular recognition system robust to noise due to mask. The MmLwThV framework can be customized flexibly to work with any other feature other than LBP for better performance and suitability. The MmLwThV framework, being lightweight, can be easily deployed on any mobile phone which has an installed visible and an infrared camera on it.

Acknowledgments I would like to thank Indian Institute of Information Technology, Allahabad, India for supporting the work on recognition of masked faces.

Declarations

Conflict of Interests We declare that we have no conflict of interest.

References

1. Abhinav G (2018) Deep learning reading group. Squeezenet
2. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12):2037–2041
3. Alahmadi A, Hussain M, Aboalsamh H, Azmi A (2020) Convsrc: Smartphone-based periocular recognition using deep convolutional neural network and sparsity augmented collaborative representation. Journal of Intelligent & Fuzzy Systems (Preprint), 1–17
4. Ashour AS, Guo Y, Hawas AR, Xu G (2018) Ensemble of subspace discriminant classifiers for schistosomal liver fibrosis staging in mice microscopic images. Health Information Science and Systems 6(1):21
5. Baochang Z, Yongsheng G, Sanqiang Z, Jianzhuang L (2010) Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. IEEE Trans Image Process 19(2):533–544
6. Bhowmik MK, Saha K, Majumder S, Majumder G, Saha A, Sarma AN, Bhattacharjee D, Basu DK, Nasipuri M (2011) Thermal infrared face recognition—a biometric identification technique for robust security system. Reviews, Refinements and New Ideas in Face Recognition, 7
7. Bishop C, Tipping M (1998) Pattern analysis and machine intelligence. IEEE Transactions on 20(3):281–293
8. Bryll R, Gutierrez-Osuna R, Quek F (2003) Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. Pattern recognition 36(6):1291–1302
9. Chen C, Ross A (2019) Matching thermal to visible face images using a semantic-guided generative adversarial network. In: 2019 14Th IEEE international conference on automatic face gesture recognition (FG 2019), pp. 1–8. https://doi.org/10.1109/FG.2019.8756527

10. Chen W, Gao Y (2013) Face recognition using ensemble string matching. IEEE Transactions on Image Processing 22(12):4798–4808
11. Cho SR, Nam GP, Shin KY, Nguyen DT, Park KR (2015) Periocular recognition based on lbp method and matching by bit-shifting. In: Advanced multimedia and ubiquitous engineering. Springer, pp 99–104
12. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3):273–297
13. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3):273–297
14. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7:1–30
15. Dubey SR, Singh SK, Singh RK (2015) Local bit-plane decoded pattern: a novel feature descriptor for biomedical image retrieval. IEEE Journal of Biomedical and Health Informatics 20(4):1139–1147
16. Fix E (1951) Discriminatory analysis: nonparametric discrimination, consistency properties. USAF School of Aviation Medicine
17. Freitas PG, Akamine WYL, de Farias MCQ (2017) Blind image quality assessment using local variant patterns. In: 2017 Brazilian conference on intelligent systems (BRACIS), pp 252–257
18. Guzman AM, Goryawala M, Wang J, Barreto A, Andrian J, Rishe N, Adjouadi M (2012) Thermal imaging as a biometrics approach to facial signature authentication. IEEE Journal of Biomedical and Health Informatics 17(1):214–222
19. Haitao Z, Shaoyuan S, Zhongliang J (2007) Visible-information-aided eyeglasses removing for thermal image reconstruction. In: 2007 10Th international conference on information fusion, pp 1–7. https://doi.org/10.1109/ICIF.2007.4408092
20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
21. Hernandez-Diaz K, Alonso-Fernandez F, Bigun J (2018) Periocular recognition using cnn features off-the-shelf. In: 2018 International conference of the biometrics special interest group (BIOSIG), pp 1–5. https://doi.org/10.23919/BIOSIG.2018.8553348
22. Hernandez-Diaz K, Alonso-Fernandez F, Bigun J (2018) Periocular recognition using cnn features off-the-shelf. In: 2018 International conference of the biometrics special interest group (BIOSIG). IEEE, pp 1–5
23. Huang X, Zhao G, Zheng W, Pietikäinen M (2012) Towards a dynamic expression recognition system under facial occlusion. Pattern Recogn Lett 33(16):2181–2191
24. Ipe VM, Thomas T (2019) Cnn based periocular recognition using multispectral images. In: International symposium on signal processing and intelligent recognition systems. Springer, pp 94–105
25. Kanan HR, Faez K (2010) Recognizing faces using adaptively weighted sub-gabor array from a single sample image per enrolled subject. Image Vis Comput 28(3):438–448
26. Kanan HR, Faez K, Gao Y (2008) Face recognition using adaptively weighted patch pzm array from a single exemplar image per person. Pattern Recogn 41(12):3799–3812
27. Kanmani M, Narasimhan V (2020) Optimal fusion aided face recognition from visible and thermal face images. Multimed Tools Appl, 1–25
28. Krišto M, Ivasic-Kos M (2018) An overview of thermal face recognition methods. In: 2018 41St international convention on information and communication technology, electronics and microelectronics (MIPRO), pp 1098–1103
29. Li Y, Guo K, Lu Y, Liu L (2021) Cropping and attention based approach for masked face recognition. Appl Intell 51(5):3012–3025
30. Liu W, Wen Y, Yu Z, Li M, Raj B, Song L (2017) Sphereface: Deep hypersphere embedding for face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 212–220
31. Moghaddam VH, Hamidzadeh J (2016) New hermite orthogonal polynomial kernel and combined kernels in support vector machine classifier. Pattern Recogn 60:921–935
32. Ngan ML, Grother PJ, Hanaoka KK (2020) Ongoing face recognition vendor test (frvt) part 6a: Face recognition accuracy with masks using pre-covid-19 algorithms
33. Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29(1):51–59
34. Ryu JW, Kantardzic M, Walgampaya C (2010) Ensemble classifier based on misclassified streaming data. In: Proc. of the 10th IASTED int. Conf. on artificial intelligence and applications, austria, pp 347–354
35. Sancen-Plaza A, Contreras-Medina LM, Barranco-Gutiérrez AI, Villaseñor-Mora C, Martínez-nolasco JJ, Padilla-Medina JA (2020) Facial recognition for drunk people using thermal imaging. Mathematical Problems in Engineering, 2020
36. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 815–823
37. 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
38. Sharma M, Prakash S, Gupta P (2013) An efficient partial occluded face recognition system. Neurocomputing 116:231–241
39. Sheskin DJ (2003) Handbook of parametric and nonparametric statistical procedures. Chapman and hall/CRC
40. Singh S, Gyaourova A, Bebis G, Pavlidis I (2004) Infrared and visible image fusion for face recognition. In: Biometric technology for human identification, vol 5404. International Society for Optics and Photonics, pp 585–596
41. Tiong LCO, Lee Y, Teoh ABJ (2019) Periocular recognition in the wild: Implementation of rgb-oclbcp dual-stream cnn. Appl Sci 9(13):2709
42. Vijayalakshmi A, Raj P (2015) An efficient method to recognize human faces from video sequences with occlusion. World of Computer Science & Information Technology Journal 5(2)
43. Wen Y, Liu W, Yang M, Fu Y, Xiang Y, Hu R (2016) Structured occlusion coding for robust face recognition. Neurocomputing 178:11–24
44. Wen Y, Zhang K, Li Z, Qiao Y (2016) A discriminative feature learning approach for deep face recognition. In: European conference on computer vision. Springer, pp 499–515
45. Wu G (2021) Masked face recognition algorithm for a contactless distribution cabinet. Math Probl Eng, 2021
46. Yang G, Feng Y, Lu H (2015) Sparse error via reweighted low rank representation for face recognition with various illumination and occlusion. Optik 126(24):5376–5380
47. Yuan C, Sun C, Tang X, Liu R (2020) Flgc-fusion gan: an enhanced fusion gan model by importing fully learnable group convolution. Math Probl Eng, 2020
48. Zhang Y, Cao G, Wang B, Li X (2019) A novel ensemble method for k-nearest neighbor. Pattern Recogn 85:13–25
49. Zhao Y, Fu G, Wang H, Zhang S (2020) The fusion of unmatched infrared and visible images based on generative adversarial networks. Math Probl Eng, 2020

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Nayaneesh Kumar Mishra is a Ph.D. from the Indian Institute of Information Technology, Allahabad India. His areas of interest include Image Processing, Computer Vision, Biometrics, Deep Learning, and Pattern Recognition.

Sumit Kumar is a research scholar at the Indian Institute of Information Technology, Allahabad India. His areas of interest include Image Processing, Computer Vision, Biometrics, Deep Learning, and Pattern Recognition.

Satish Kumar Singh is an Associate Professor at the Department of Information Technology, Indian Institute of Information Technology Allahabad India. He has about 16+ years of professional experience in various capacities. Presently, he is heading the Computer Vision and Biometrics Lab (CVBL) at IIIT Allahabad from 2015 onwards. His group is involved in the R&D of Signal & Image Processing, Vision, and Biometrics Algorithms and Systems. His areas of interest include Image Processing, Computer Vision, Biometrics, Deep Learning, and Pattern Recognition.