Microprocessors and Microsystems 76 (2020) 103096

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Combining pixel selection with covariance similarity approach in hyperspectral face recognition based on convolution neural network

Ashok Kumar Rai*, Radha Senthilkumar, Aswin Kumar R

Department of Information Technology, Faculty of Information and Communication Engineering, Anna University, Chennai, India

ARTICLE INFO

Article history:
Received 24 January 2020
Revised 5 March 2020
Accepted 18 March 2020
Available online 19 March 2020

Keywords:
Face Recognition
Convolution Neural Network
Hyperspectral Imaging
Firefly Algorithm
Band Selection

ABSTRACT

A Hyperspectral camera provides discriminating features for capturing human faces that cannot be obtained by any other imaging technique. Nevertheless, it has new issues comprising curse of dimensionality, physical parameter retrieval, fast computing and inter band misalignment. As a result of, the literature of Hyperspectral Face Recognition is more scanty and confined to improvised dimensionality reduction and minimization of wide-ranging bands, and thus the objective can be obtained by the use of the Convolution Neural Network (ConvNet). Since ConvNet is of great attention in recent times, it can offer outstanding performance in face recognition systems, where the quantity of training data is amply large. We propose a Hyperspectral Face Recognition system using Firefly algorithm for band fusion and the Convolution Neural Network for classification. In addition to this, the present work is extended 11 exiting face recognition methods to perform Hyperspectral Face Recognition task. Thus the work has been framed as Hyperspectral Face Recognition problem to an image-set classification problem and assessment of the performance has been done on six state-of-the-art image-set problem techniques, and similarly it was examined on five state-of-the-art RGB and gray scale face recognition system, subsequently applied improved Firefly band selection algorithm on Hyperspectral Images to get appropriate band. Assessment with the eleven extended and five existing HSI Face Recognition system on two benchmark datasets (CMU-HSFD & UWA-HSFD) demonstrates that the proposed system overtakes all by a noteworthy margin. Lastly, we execute the band selection demonstration to get the novelty for most informative bands in Visible Near Infrared (VNIR).

1. Introduction

In spite of massive study, face recognition problem is still an interesting task since the fact is intra-person differences often beat the inter-person differences in the presence of posture and brightness deviations. Maximum of the existing study is focused on the human facial appearance taken by the colour camera, which produces spatial grayscale or RGB images typically picked up in the visible range of electromagnetic spectrum [1,2]. The essential volume of the colour or grayscale images are generally below than average, which is restricting the discriminative ability of face recognition techniques based on available data. Additionally, the spatial parameters extracted from grayscale/RGB images, as Principle Component Analysis, Support Vector Machine or Gabor filter, possibly will perform poor in the existence of posture and brightness deviations and resultant is very poor inter-person accuracy.

In order to gain high volume of the face recognition datasets, several researchers have studied to use of hyperspectral imaging facial datasets [3–7]. Hyperspectral imaging technique possibly will be gain better face recognition results by acquiring additional biometric data like spectral features of human faces. As the amount of objects to be verified increases, the accuracy of spatial image based techniques decreases cause of the reduction in the inter-object distance in the facial space. In this situation, hyperspectral imaging technique possibly will be beneficial for improving the performance because of the huge volume of features.

A HSI image is a data cube comprises two spatial along with one spectral aspect, which is acquired by an HSI camera which obtains the intensity of light or radiance for a huge amount (normally tens to some hundred) of continues wavelength of the electromagnetic spectrum. Thus, every pixel in the hyperspectral images comprises a contiguous band (in radiance or reflectance) and then it might be passed-down to describe the features in the particular section by excessive accuracy and elements., Hyperspectral imaging technique evidently delivers much more precise info about the

* Corresponding author.
E-mail address: akraiceg@gmail.com (A.K. Rai).

https://doi.org/10.1016/j.micpro.2020.103096
0141-9331/

images in comparison with a RGB image, which only obtains three dissimilar spectral bands equivalent to the visual prime colors red, green and blue. Henceforth, hyperspectral imaging technique leads to a massively enhanced capability to categorize the features in the images depending upon their spectral properties. Hyperspectral imaging techniques furthermore offer liveliness examination by differentiating the band signature of an actual human face from a fake concealment or an ordinary picture.

Even though the hyperspectral dataset is very huge, offers expected parameters for the growing uniqueness of human faces, but it includes new issues like inter band misalignments, curse of dimensionality and low signal to noise ratios in some particular bands. Inter-band misalignment is cause of object displacement, while acquiring the hyperspectral dataset. Low signal to noise ratio is normally perceived in spectra surrounding the blue band (at 410nm) causing short photon liveliness in these wavelengths. Hence, these bands possibly will contain quite higher noise. Further, huge volume of data creates discriminative feature extraction as an interesting challenge.

The foremost reason behind aforementioned issue is that the dataset used for face verification/recognition are scanty. One most basic works have completed by [8] who passed-down the 31 dimensional (31 wavelengths) electromagnetic spectral signatures at preferred positions of the face in the Visible Near Infrared (VNIR) spectrum (400–1000 nm) for face recognition. They address great verification and recognition ratio under posture dissimilarities on an exclusive datasets including 1400 hyperspectral images of nearly 200 objects. Though, their consequences are not reproducible on public hyperspectral facial databases. In the paper [9], researchers equated spectral signatures of diverse face sections by spectral angle measurements. Their tests were narrow on actual small datasets comprising of eight objects the researchers in a paper [10] estimated the hyperspectral cube to a small volume two dimensional Principle Component Analysis space and calculated recognition datasets [11]. The identical datasets were passed-down by [12] to integrate together the spatial as well as the spectral dimension. Nevertheless, they passed-down an improvised technique for spectral mixing and transformed the hyperspectral data cube to a colour image by picking up only one value from a particular wavelength at every pixel. By stating it clearly, they abandoned 96 percentages of pixels from all hyperspectral data cube without any justification, Euclidean distance for face verification and recognition in the paper [13] estimated Gabor Wavelets to hyperspectral data cube which made 52 cubes from each hyperspectral datasets. To decrease the vast quantity of data, an improvised sub-sampling method was passed-down.

Current hyperspectral face verification/recognition algorithm functions dimensionality reduction technique nonetheless in a quite improvised way. Furthermore, existing study shows that shortage of a wide-ranging opinion as they each use a very small quantity of evaluating objects or benchmark datasets that are not openly accessible. More extensively, present literature does not offer wide-ranging assessments by state of the art face recognition algorithms which are precisely constructed for hyperspectral face verification and recognition however may be tested by miniature variations. For instance, every portion of the HSI cube may be considered as a distinct image, and available grayscale/RGB face recognition algorithms may be applied for the classification. Besides, by considering a hyperspectral cube images as a set of images, image-set classification techniques can be tested on this type of challenges without any help of dimensionality reduction techniques. We have observed the result of six newly coming to the house image-set classification techniques [14–19] on two standard face recognition datasets [14,15]. Note that this is the worthwhile to mention that we have tested all the dataset on GPU (2 GB on board memory).

The main goal of this paper is to develop a novel hyperspectral image face verification or recognition technique based on improved firefly algorithm for band fusion. This anticipated technique extracts targeted and useful hyperspectral face band from the several bands of the hyperspectral face cube. Band selection has been completed by improved firefly band selection algorithm. Mainly, there are two essential features in Firefly Algorithm (FA), and that are illumination and magnetism. The fireflies' illumination is articulated by present location. Whenever the fireflies are shiny, the location will be near to accurate, the quantity of the objective function will be higher, and the movement will be from a darker area to shiny area. And if the fireflies' illumination is the similar as early, the movement will be random.

While the random points are getting done the movements of illumination and magnetism are kept on updating frequently and randomly on distributed locations and it moves bit by bit towards the desired location. After a certain time of experiments the desired positional point has been achieved.

In a Hyperspectral face recognition database the surrounding spectral wavelengths are deeply correlated, which leads to the low generalization ability for the classification techniques [20]. As per our best information, pixel wise band fusion and classification algorithm [21–23] have been reported so far for hyperspectral face recognition. Where PCA and fractal analysis are frequently used technique for dimensionality reduction in hyperspectral face recognition datasets. Separating the informative bands from hyperspectral facial images is challenging task using statistical techniques [24,25] due to the huge computational liability at the pixel-wise classification. Nevertheless, in our proposed technique we are selecting the most informative band from hyperspectral data cube and classifying the image as image set problem, where each and every wavelength in a hyperspectral facial image cube will be treated as a distinct image. We focused band selection issue at image based as we consider that this permits to accomplish high elevated info from contours and abstract theories from hyperspectral facial images in contrast with pixel-level chosen bands. On the other hand, we use highly successful one image classification technique, i.e. convolutional neural networks [26–29]. We propose a methodology for learning spectral information using ConvNet to obtain a parameter within the range of VNIR. In [44–46], discuss the intelligent computer based vision based security recognition system.

2. Method

The overall architecture for classification and band fusion is illustrated in Fig. 1, which contains four key parts: Formulation of Hyperspectral Imaging (HSI) problem as Image-Set, alignment of different posture image within the same person (Inter-person), constructing an HSI Face ConvNet to learn and classify the images and finally applying improved firefly algorithm to get the fused images for intra person classification of hyperspectral images.

2.1. Formulating HSI face recognition problem as an image-set

Image-sets schema is a combination of images and offers greater accuracy than conventional image based regression and classification [14,16–19,31]. We extend HSI face recognition problem as an image level classification since most informative information basically occurs among the entire bands of HSI cube. Image-level classification outline is vigorous to misalignments and is comfortable to permit to vary the bands throughout the library or probe sets.

Six state-of-the-art image-set classification problems have been evaluated in our study. Two of them are Discriminant Learning [14], Covariance Discriminant Learning similar in the structure and

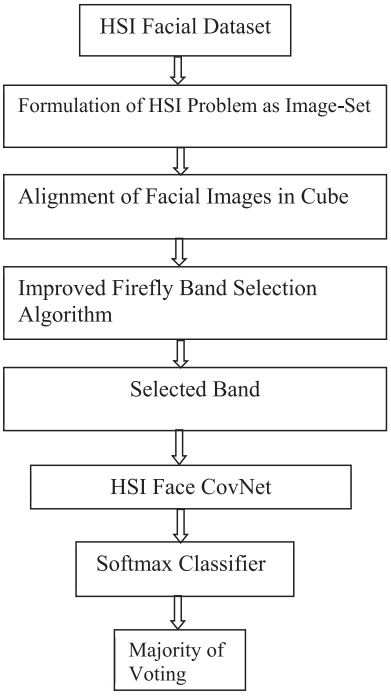

Fig. 1. Flowchart of proposed architecture.

the remaining algorithms are replica based and get the closet points throughout the image-set. The aforementioned algorithms also use Affine Hull, Convex Hull [18] and Sparse approximated closet points [19] techniques.

2.1.1. Transformation of HSI image to RGB

For the sake to utilize the existing classification technique on the visible band of the hyperspectral image, we convert the whole hyperspectral image into three band visible image by CIE 2006 tristimulus colour like functions [8]. We take advantage of the unique CIE method to map the visible band from the HSI image to the colour space.

Let I (λ) be the spectral response of illuminator of CIE Standard (D65) day vision brightness, r(λ), g(λ) along with b(λ) be the discrete basis functions for CIE 2006 corresponding to (Fig. 2), which represent the vision system of human and R(λ) be the reflectance of the object in colour space. Let R (p, q,$\lambda\iota$) be the brightness regularized reflectance on some 2D space (p, q) and wavelength $\lambda\iota$. The tristimulus value of CIE X, Y, and Z can be acquired on 2d space (p, q) by following method.

$$X(p,q) = \sum_{i=1}^{N} R(p,q,\lambda i) I(\lambda i) r(\lambda i) \tag{1}$$

$$Y(p,q) = \sum_{i=1}^{N} R(p,q,\lambda i) I(\lambda i) g(\lambda i) \tag{2}$$

$$Z(p,q) = \sum_{i=1}^{N} R(p,q,\lambda i) I(\lambda i) b(\lambda i) \tag{3}$$

Where, N is the band count in HSI data cube. In this paper (λ_i) is acquired near to the CIE standard brightness which contains a fixed spectral response within the human vision range. To acquire exact calorimetrically RGB Image, the tri-stimulus X, Y and Z has been converted in the direction of the sRGB 2D colour plane by taking advantage of

$$\begin{bmatrix} Rs \\ Gs \\ Bs \end{bmatrix} = \begin{bmatrix} 3.1336 & -1.6168 & -0.4907 \\ -0.9787 & 1.9161 & 0.0335 \\ 0.0721 & -0.2291 & 1.4057 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

It's worthwhile to mention that we are taking advantage of D55 white reference in present work while taking use of the D65 the accuracy of the classification result is similar. In some specific spectral band tri-stimulus shows the packets of sR, sG and sB negative value cause of narrow colour gamut. Though, in the facial spectral data it never appears in the archived RGB images. It was proved analytically for both the datasets used in this work.

2.1.2. Transformation of HSI image to grayscale

In spatial Imaging, in front of the sensor a colour filter is placed. Here the colour filter checks the sensitivity of each and every photocell within a particular area of the electromagnetic spectrum. Such that, at each and every pixel of the sensor archived image the object features in only one band is taken. Likewise technique is called as grayscale imaging technique [1] (Fig. 2). We resemble a grayscale imaging technique which is penetrating above the visible range of electromagnetic spectrum and accounts the whole intensity of radiance dropping down in all pixels.

Let I(λ) be the spectral response of the CIE standard sunshine brightness (D55), R(λ) be the reflectance of the object plane, and the electromagnetic spectral diffusion of the VNIR filter power distribution is T(λ)along with S(λ) which is electromagnetic spectral response of silicon sensor (Fig. 2). We acquire a single coverage by VNIR filters for fine-tune bands in the range of 450-1090nm. The grayscale camera sensor response for VNIR acquisition to all spatial positions (p, q) (HSI to RGB and grayscale image) of illuminating entity is acquired by

$$\rho(p,q) = \sum_{i=1}^{n} I(\lambda\iota) R(p,q,\lambda\iota) S(\lambda\iota) T(\lambda\iota) \tag{4}$$

Where, n is band count in hyperspectral data cube, I($\lambda\iota$), S($\lambda\iota$) and T($\lambda\iota$) are the CIE 2006 colour matching function, CIE standard sunshine brightness (D55) and Visible Near Infrared (VNIR)filters subsequently.

2.2. Improved Firefly band fusion algorithm

Current band fusion techniques perform using projection of hyperspectral datasets on three supportive vectors and that is parallel to weighted averaging with the spectral aspect. Like this type of estimation does not utilize the spatial info into justification and does not execute balancing along with the spatial measurements. Spatial data are highly capable to make fine distinctions and have the benefit of efficiently eliminating the noise of sensor and inter-band misalignment. In variation with this, we propose an improved firefly band technique that combines native spatial info along with well removed noise by balancing together with the spectral and spatial measurements. Additionally, it also solves little posture movements like eye blinking and little expression abnormality and take advantage of enhanced discrimination.

2.2.1. Searching for best position based on firefly

The Firefly Algorithm (FA) in a growing optimization algorithm suggested by Yang [9]; which is motivated by class searching depend on natural characteristic, illumination of fireflies. Fireflies have a unique twinkling nature that is used for interaction and magnetise to hidden victim. Illumination (I) and Magnetism (M) are two necessary components are there in fireflies. The Illumination I at a fixed distance d follows the law of Inverse Square that means if there is a decrement in illumination I then increment

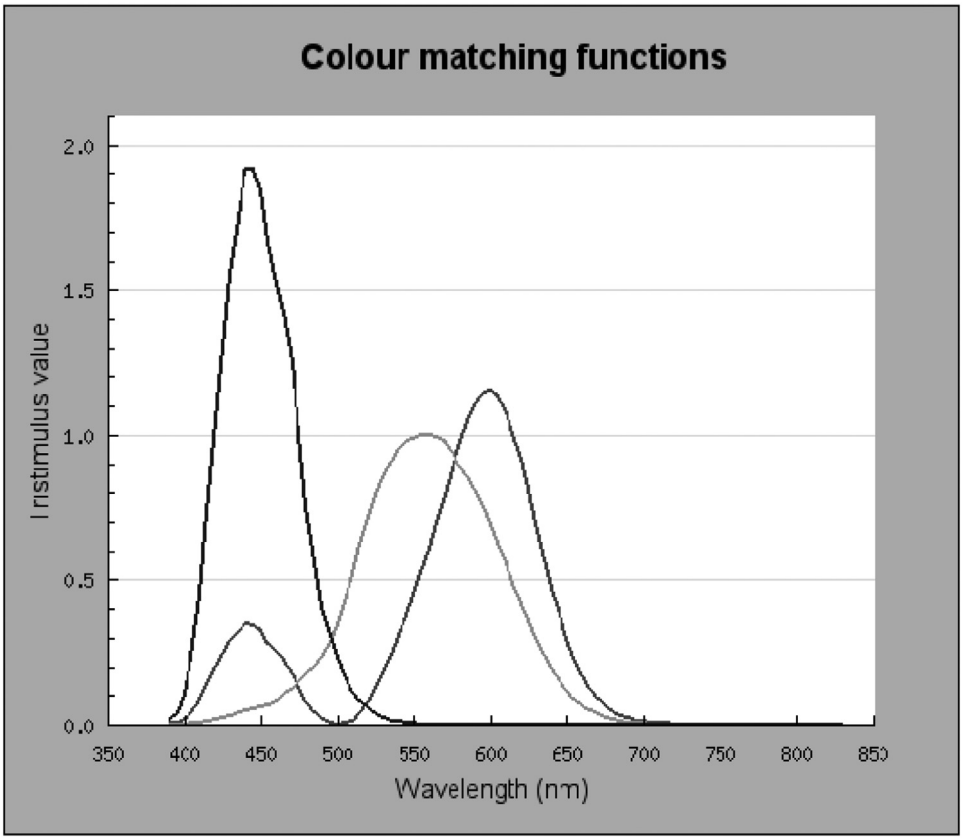

Fig. 2. The CIE 2006 RGB original function.

should be there in distance d. In the firefly algorithm, the illumination of a firefly is said in terms of the present location of fireflies: if the illumination of firefly is higher at the particular location, the location will be preferred; it also states that the quantity of the objective functions is higher. The less illuminating fireflies will step forward to the higher illuminating fireflies. If the illumination of fireflies has the identical value, the movement will be random. For explanation the following assumption is described for FA.

(1) Entire fireflies stay unisex so that the fireflies will be magnetise each other irrespective of their sexual characteristics;
(2) The magnetism of a fireflies is directly proportional to its illumination; and
(3) The illumination is proportional to its objective function [31,32]. Throughout the iteration of flies' movement, illumination I and magnetism M are updated repetitively, and on random generated locations are slowly stepping forward to the desired location, which is accounted as best positional location and not as much of desired locations are eliminated.

The illumination of fireflies at the distance d of its position differs with respect to its objective function that can be well-defined as

$$I(\gamma) = I_0^{-\gamma d^2} \tag{5}$$

Where I_o is the max illumination when $d = 0$; which is associated to the cost of the objective function, and the higher value shows higher illumination and γ indicates the illumination coefficient.

The magnetism of a firefly on distance d of its position is directly proportional to its illumination acquired by neighbouring fireflies; and it expressed as

$$M(\gamma) = M_0 e^{-\gamma d^2} \tag{6}$$

Where M_0 is magnetism, and the space between two fireflies is zero. The equation which shows that firefly i moves to jth position can be given below

$$p_i = p_j + M * (p_j - p_i) + \alpha\left(random - \frac{1}{2}\right) \tag{7}$$

Where p_i and p_j are the starting location of fireflies i, j and α is a constant within 0 and 1.

All the features and locations of fireflies are initialized, the illumination and magnetism of fireflies is designed. Achieved the higher illuminating area which is best for magnetize the other firefly. And finally the location is updated and calculated the final illumination. The iteration is repeated until the best result gotten. Lastly the best result and global extrema are accounted

2.3. Classification by ConvNet

Since the last decade, several researches have been performed on the feature learning and labeling of classes which is handcrafted in computer vision community, but maximum of them are not trainable for recent problems and unlike job. Recently, feature extraction from learning oriented modules have been observed to overtake in the labeling related issues, since they ensure the ability of finding and optimizing graphical description for the particular task to be resolved. In regards to this, Convolution Neural Networks are having a significant advantage in the domain of parameter learning based feature extraction techniques [27,28,30]. With the diverse architecture of the convolution neural network, scholars can gain new techniques that are suitable to their issues, in these papers a technique has been presented for hyperspectral facial data to train through the convolution neural network and get the labeled classes. The network can obtain significant graphical

information for most desired band in our study. The architecture of the proposed convolution neural network for classification of hyperspectral facial data is described below.

For the convenience, we mention the proposed convolution neural network as HSI-ConvNet (see Fig. 1). Further, we are extending the hyperspectral face recognition problem as image-set. Where each band is treated as separate image and they are having same class which is similar like the original hyperspectral facial image cube. Taking the ConvNet on a solo channel produces higher accuracy. Treating the dataset as an image-set, handles the requirement of an amply large dataset, that is essentially required in the training process of ConvNet. In this paper, we are utilizing the ConvNet such as eight convolution layers, followed by fully connected layer and finally softmax classifier is added to find the distribution over the output classes. All the Conv layers are processed by Batch-Norm function; a down-sampling layer, a non-linearity function Rectified Linear Unit (ReLu) do remove the non-linearity from it and finally added dropout to prevent overfitting problem in the network.

Notation: In this paper, the dimension of the hyperspectral facial image cube is $l \times b \times \lambda$, where the total number of channels (or wavelength) is N, l and b are the length and breadth of the hyperspectral facial cube, correspondingly. Here in this paper the dimension of each image are $(l \times b)_{\lambda\iota}$, i∈[1…..N], of the hyperspectral facial data cube, and each image considered as distinct image $p(l \times b)_{\lambda\iota}$, for colour image and $p(\lambda\iota)$ for panchromatic image for classification and recognition purpose.

2.3.1. Network architecture and training

The following portion defines the ConvNets used in the present paper and training process. Which is inspired by [33], where the networks are ''very deep'', that means they includes a long sequence convolutional layers and accomplished state-of-the-art result in some of event of the ImageNet 2014 I issues [34], likewise in some more events [35–37].

2.3.1.1. Learning the classifier. Originally, the deep neural network architectures ω are loaded as the face recognition problem where N=664 and 153 identities for both data (CMU-HSFD and UWA-HSFD) respectively, structured as N-ways classifier. The ConvNet connects to each and every training data l_t, $t = 1$……T, a resultant matrix $X_t = p\omega(l_1) + b$; b $\in\Re^{N,}$ Which shows that the last fully connected layer holds N linear predictors [p, b]; $p \in \Re^{N \times D}$ and b $\in\Re^{N}$. These results are compared with ground truth labelled identity c_t = {1…….N} by calculating the empirical softmax function

$$S(\omega) = -\frac{\sum \log(e^{e_{c_t} X_t})}{\sum_{q \to 1 \ldots N} e(e_q, X_t)} \tag{8}$$

Where, $X_t = \omega(l_1) \in \Re^{D,}$ and e_c represents one active matrix of label c.

Once learning got over the classification layer (p, b) may be eliminated and the resultant matrix $\omega(l_t)$ may be used for facial identity verification by the uses of the Euclidean Distance method to relate them. In the other hand the results may be meaningfully increased by tuning them for verification in Euclidian space by a "triplet loss" training method. While the goal is to obtain decent overall performance, loading the network as a classification technique is elaborated in the next section.

2.3.1.2. Triplet loss. This training objective is to learn resultant matrix that achieves sound in the concluding application i.e. isolates face verification using comparison of labels in Euclidian space. This is quite similar like metric learning approaches, which is used for learning projection. At the same time the system is able to accomplish the curse of dimensionality problem.

In this present work the triplet loss training model is Identical to [38]. The outcome $\omega(l_1) \in \Re^{D}$ of the ConvNet, pre-trained model as described in Section 2.3.1.1 is l_2 – normalized with the projection of L ≤ D dimensional distance using affine transform $X_t = p'\omega(l_t)/\|\omega(l_t)\|_2$, $W' \in \Re^{L \times D}$ though here the formula is identical to linear predictor lettered above, but some key differences are there, like L is not equal to the number of class identity D (L ≠ D), and the next one is that the transformation P' is trained for minimization of observed triplet loss.

$$S(p') = \sum_{a,w,n \in T} \max\{0, \alpha - \|X_a - X_n\|_2^2 + \|X_a - X_w\|_2^2, \} \tag{9}$$

$$X_i = W'\{\omega(l_i)/\|\omega(l_i)\|_2\}$$

It is noteworthy to mention, unlike the previous section, there is no bias has been learned due to dissimilarity in (Eq. (9)) would terminate it. Here learning rate $0 \le \alpha \le 1$ and T is a group of training triplets. A triplet (a, w, n) is the collection of anchor 'a', positive w ≠ 'a' along with negative image n.

2.3.1.3. Architecture. We are taking in account three architectures based on A, B along with D [28]. The architecture includes 11 layers, and all contain a linear function followed by non-linearities (ReLu and down sampling). The starting 8 layers are convolutional layers. The linear function is a collection of linear kernels. The ending three layers are Fully Connected (FC) layers; they are equal as convolutional layers, however the dimension of the kernel is equal to the dimension of the input dataset, Such that each kernel "convolves" information from the whole image. Each and every convolutional block are followed by non-linearity function (ReLu) as in [28], but dissimilar to [19]. The two fully connected block produces 1037 and 230 dimensional respectively to our dataset (CMU-HSFD, UWA-HSFD) and last fully connected block produces 664 and 153 dimensional based on the total loss function that is utilized for optimization. And finally subsequent vector is passed to softmax classifier. Network B and D are relatively same to A, but comprises 2 and 5 additional blocks respectively. The input image to the network is the size of 160 × 160 with a regular facial data, which is achieved after the training of the dataset.

2.3.1.4. Training. Learning the multi-way face classifier (in Section 2.3.1.1) succeeds the steps of [28] with the changes recommended by [33]. The overall aim is to hunt the features of the ConvNet that reduce the typical estimation of logrithmatic loss later to the softmax layer. We elaborate the outline for the ConvNet A and finally the differences for the B and D configurations. The optimization has been done using SGD (Stochastic Gradient Decent) with small batches of 20 images and the momentum coefficient is *0.9*. The technique is later on regularized by dropout along with weight loss; the latter the momentum coefficient was agreed to be *0.00005*, and dropout was adopted after two subsequent FC blocks with a learning rate of 0.5. The initial learning rate was 0.1 and after that reduced by a factor of 10, when the accuracy of validation is stopped increasing. The whole model was trained by three reducing factor learning rates. The kernel weight in the networks was initialized using a random sampling method by a Gaussian distribution with an average of zero, standard deviation of 0.01, and the bias was set to be zero in the initial phase. In the training procedure the ConvNet was initialized by 160 × 160 size of images cropped from the original images, where crops vary every time. Moreover, the images were flipped from left to right with a 50% probability, which is known as augmentation process. While in our experiment, any augmentation on color images was not performed as in [28] and [33]. The ConvNet A is trained from scratch, whereas the training of ConvNet B and D are started from trained A. This is achieved by the adding extra

Table 1
Detailed illustration of both datasets cmu-hsfd and uwa-hsfd.

Dataset	Identities	Images	No. of Channels	Wavelength Range	No. of Cubes
CMU	664	40,680	65	450–1090 nm	147
UWA	153	4997	33	400–720 nm	145

FC block in A, initialised randomly, and finally it was fine-tuned (i.e training with smaller learning rate). During the embedding process through triplet loss the ConvNet was frozen apart from the last FC block applying the discriminative transformation. After this the network was trained with a learning rate of 0.1 for 10 epochs using stochastic gradient descent. The positive pairs (a;w), where a is considered as anchor and w is its positive pair. Selecting a good triplet a challenging task and should be balanced. This is acquired by each pair to a triplet (a;w;n) by random sampling of n images, however it's done only between those who interrupt the marginal lass of triplets. Further hard-negative excavation is done, but it is not either beneficial or cheaper than selecting the *maximally violating* instance, as usually tested in structured learning of results. At the time of testing, the embedded labels $p'\omega(l_t)$ are evaluated in Euclidian space for the resolution of face verification. In this phase the aim is that to find weather two facial images are having the same character or not; this is achieved by the comparing of distance $\|p'\omega(l_1)-p'\omega(l_2)\|_2$ in between the embedded label should be less than a threshold value. This value of the threshold is learned independently to exploit the verification accuracy on the appropriate validation information.

3. Experiments and results

The present method is distinguished with 13 in the house and extended HSI face verification/recognition techniques which is categorized in three ways; four hyperspectral, six image-set and three colors and panchromatic face recognition technique. Where, the proposed system outperforms all with a significant margin.

3.1. Hyperspectral Facial data

We experimented on two benchmark datasets. Table 1 shows the summary of datasets. The primary dataset is Carnegie Mellon University Hyperspectral Facial Database (CMU-HSFD) [39] (Fig. 3) developed by prototype spectro -polarimetry camera. The entire hyperspectral facial cube comprises 65 channels within the range of 450–1090 nm with a spectral gap of 10 nm. Three similar lighting source of 600 W halogen bulbs was utilized for the brightness purpose. The data comprise 48 subjects; all the subjects have minimum 4 and maximum 20 cubes obtained in unlike sessions and different illuminating arrangements. Some particular channels have

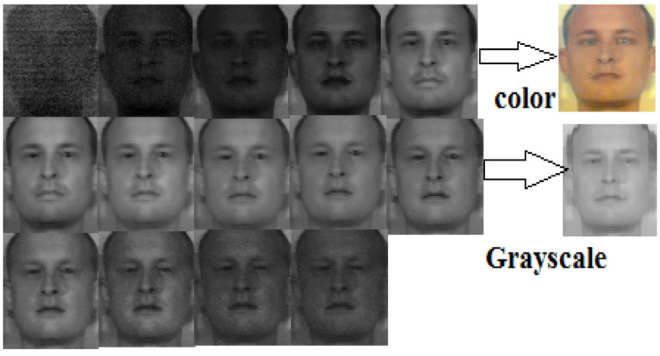

Fig. 3. CMU-HSFD (450–1090 nm) in VNIR range with a spectral difference of 10 nm.

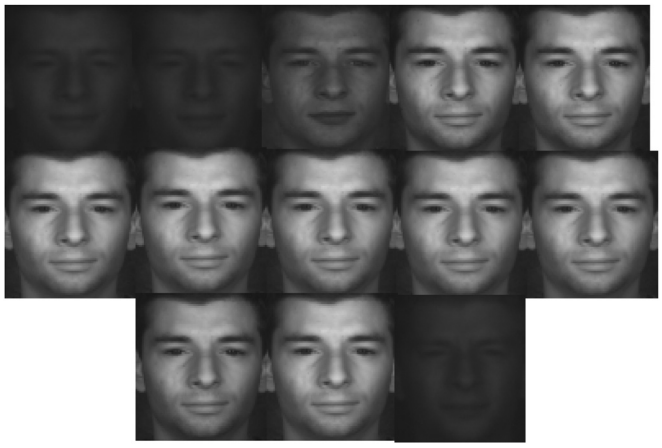

Fig. 4. UWA-HSFD (450–720 nm) in VNIR range with a spectral difference of 10 nm

a low Signal to Noise Ratio (SNR) and maximum subjects are not proper due to the eye blinking and movements while capturing the image. We utilized the data cubes which are obtained when all the lights were in working condition. Hence, our tentative data is a collection of 147 hyperspectral facial data cube and 48 subjects, where all the subjects have nearly 1to 5 cubes. The image gallery is the collection of one cube per identity (subject) randomly.

The second database is University of Western Australia- Hyperspectral facial data (UWA-HSFD) see (Fig. 4) obtained using CRI's VariSpecc LCTF incorporated photon focused camera. All HSI cube is within the range of 400–720 nm with a gap10nm with 33 channels. The noise is comparatively less because of the algorithm which mechanically accepts camera exposure time to the filter of transmittance, radiance intensity and due to the sensitivity in each channel. The respected location of the identity of the camera is not strictly static. Maximum identifies perform eye blinking and minor head movements during image acquisition. Hence alignment inaccuracies are exiting in respective channel. The UWA-HSFD contains 145 cubes of 78 identities and some of them are used as probes. Table 1 shows the summary of both the datasets. For both datasets, we used a cross validation (ten-fold) test by random selection of image from gallery in each fold.

3.2. Implementation details

All the experiments were conducted in a Windows environment with Intel Core i5 -7200 CPU running at clock speed of 2.5 GHz supported by 8GB DDR4 memory. To accelerate the training process the system also consists of the GPU (NVIDIA GeForce 940MX with 2GB on board memory) and CUDA 9.0 with CuDNN 7.0. We are also taking advantage of Tensorflow 1.6.0 based on Python 3.6 and Tensorboard 1.6.0 for the visualization purpose.

The HSI-ConvNet $\omega(l_t)$ comprises linear label predictor and layers of softmax, which provides D = 1037, and 339 respectively for both the datasets. Fed as input image is a dimension of 160 × 160 chunks are cropped from all the 4 corners and focal point by horizontal flip and taken the mean of the output vector. To allow N-Scale test, the facial images are resized into three dimensions 256, 384 and 512 pixels and the chopping process are constant for

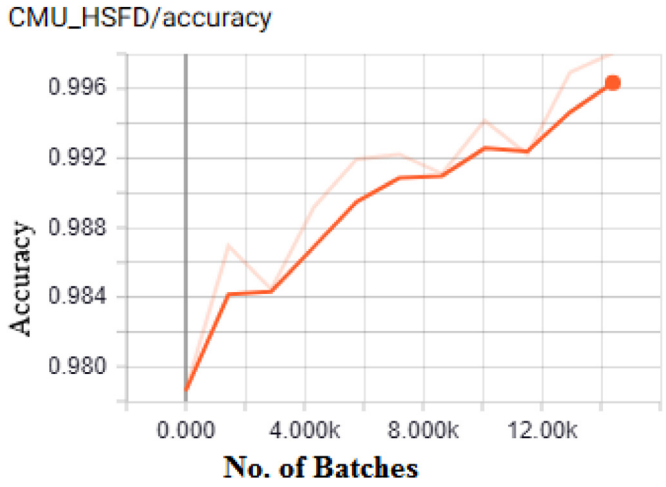

Fig. 5. The accuracy for CMU-HSFD dataset, where x-axis shows the number of batches (each batch contains 20 images), and y – axis is the accuracy.

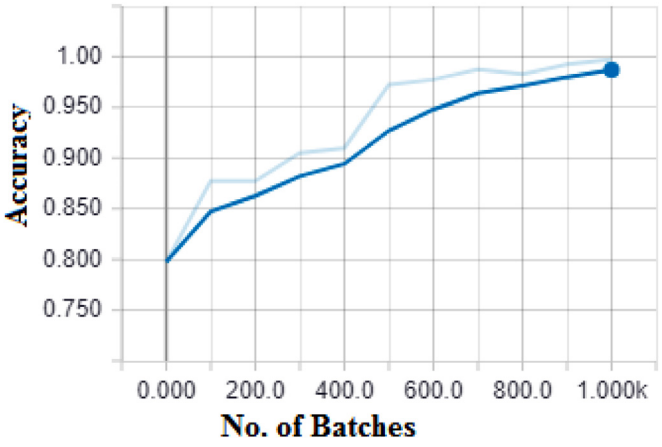

Fig. 6. The accuracy for CMU-HSFD dataset, where x-axis shows the number of batches (each batch contains 20 images), and y – axis is the accuracy.

each of these as in [39]. The resultant label for the facial images is mean of the entire feature vectors. Faces are detected with the help of the technique elaborated in [43] if the alignment of faces is utilized, and then facial features are calculated the technique described in [44] and spatial similarity projection is applied to record the face at canonical location. The prepared images are then fed as input to the HSI-ConvNet to detect the key features for identification of facial images and based on these key features the model encapsulates the face in a bounding box. The bounded region is cropped from the whole image which is rescaled in the dimension of 160 × 160 and stored separately as aligned image. This aligned image is preceded further with a softmax classifier on the top of the model. The training has been done for a constant 10 epochs for both the dataset. Each epoch contains 1440 batch with 20 random images per batch and 100 batches with 20 random images per batch respectively, for both the dataset CMU-HSFD and UWA-HSFD. The images used for training are also flipped horizontally in a random order to render the model to recognize the under flipped condition. The cross entropy loss has been utilized to fine tune the features and regularization loss is used to prevent the so called problem overfitting by dropping out the neurons. After each epoch of training 10 fold cross- validation is performed on a random generated small subset of total images to calculate the accuracy and validation rate. The present model has achieved maximum accuracy 99.80% and 99.75% for both the dataset CMU-HSFD and UWA-HSFD respectively. All the results are visualized as a graph with the help of Tensorboard 1.6.0. By looking to the (Figs. 5, and 6), we are clearly able to see that when the number of batches (i.e. images, each batch 20 images) are low, the accuracy was low and when the number of batch size got increased, the accuracy got increased which is nearly 99.80 and 99.75 respectively for both the dataset (CMU-HSFD & UWA-HSFD) (Table 2).

Table 2
Comparison of mean recognition ratio and standard deviations with 13 existing algorithm.

Techniques	CMU-HSFD	UWA-HSFD
Hyperspectral		
Spectral Signature for Diverse Faces[8]	38.2 ± 1.79	40.6 ± 1.00
2-Dimensional PCA [10]	72.2 ± 5.31	83.95 ± 2.27
Eigen face by Spectral [12]	84.8 ± 3.48	91.7 ± 3.02
3Dimensinal Gabor wavelets [13]	91.8 ± 2.96	91.5 ± 2.07
Image-set		
Discriminative learning [14]	87.70 ± 3.00	91.50 ± 3.18
Manifold -Manifold Distance [15]	90.00 ± 2.68	82.9 ± 4.16
Manifold Discriminant Analysis [16]	90.60 ± 2.77	91.00 ± 3.16
Covariance Discriminative Learning [17]	92.7 0± 2.21	93.2 0± 1.40
Affine Hull [18]	90.80 ±2.20	92.50 ± 2.10
Convex Hull [18]	91.10 ± 2.10	92.60 ± 2.26
Color and Panchromatic		
Face recognition by Eigenfaces [40]	82.7 ± 4.12	80.5 ± 2.20
Face recognition by Fisherfaces [41]	93.7 ± 1.87	96.20 ± 1.22
Face recognition by Local Binary Patterns [42]	96.8 ± 1.47	96.3 ± 1.00
Proposed Technique		
Improved Firefly + HSI-ConvNet	99.80 ± 0.05	99.75 ± 0.25

4. Conclusion

We offered a hyperspectral face verification/recognition (HSI-ConvNet) technique based on improved firefly and convolution neural network. The present technique was tested on two benchmarks datasets and compared with 11available state-of-the-art technique. We also offered a band selection technique to search the most informative band in the VNIR range of the electromagnetic spectrum. Overall the proposed technique marginally outperforms state-of-the-art hyperspectral face recognition techniques. Moreover, with the increasing utilization of the hyperspectral technology like HSI camera mounted cars, HSI camera for the use of medical practitioner and for precision farming can be possible enhancement for our society. The eye blinking error and posture detection can be more efficiently improved by the Capsule Network.

Declaration of Competing Interest

Thus the work has been framed as Hyperspectral Face Recognition problem to an image-set classification problem and assessment of the performance has been done on six state-of-the-art image-set problem techniques, and similarly it was examined on five state-of-the-art RGB and gray scale face recognition system, subsequently applied improved Firefly band selection algorithm on Hyperspectral Images to get appropriate band.

References

[1] Stan Z. Li, Anil K. Jain, Handbook of Face Recognition, second Ed., Springer, 2011 ISBN 978-0-85729-931-4.
[2] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (December 2003) 399–458.

[3] Wei Di, Lei Zhang, David Zhang, Quan Pan, Studies on hyperspectral face recognition in visible spectrum with feature band selection, IEEE Trans. SMC-A (2010), doi:10.1109/TSMCA.2010.2052603.
[4] Zhihong Pan, Glenn Healey, Manish Prasad, Bruce Tromberg, Face recog. inhyperspectral images, IEEE TPAMI (2003), doi:10.1109/TPAMI.2003.1251148.
[5] Zhihong Pan, Glenn Healey, Bruce Tromberg, Comparison of spectral-only and spectral/spatial face recog. for personal identity verification, EURASIP JASP (2009), doi:10.1155/2009/943602.
[6] A.Robila. Stefan, Toward hyperspectral face recognition., In Proceddings of SPIE ISTE Imaging, 2008, doi:10.1117/12.765268.
[7] Linlin Shen, Songhao Zheng, Hyperspectral face recognition using 3D Gabor wavelets, in: Proceedings - International Conference on Pattern Recognition, 2012, pp. 1574–1577.
[8] Z. Pan, G. Healey, M. Prasad, B.J. Tromberg, Face recognition in hyperspectral images, IEEE Trans. Pattern Anal. Mach. Intel. 25 (12) (2003) 1552–1560.
[9] D. Ryer, Quest hierarchy for hyperspectral face recognition, Air Force Institute of Tech, 2012, doi:10.1155/2012/203670.
[10] Z. Pan, G. Healey, M. Prasad, B.J. Tromberg, Comparison of spectral-only and spectral/spatial face recognition for personal identity verification, EURASIP J. Adv. Sig. Proc. (1) (2009) 2009.
[11] S. Robila, Toward hyperspectral face recognition, SPIE, Image Process. Algorithms Syst. VI 6812 (2008) 1–9.
[12] W. Di, L. Zhang, D. Zhang, Q. Pan, Studies on hyperspectral face recognition in visible spectrum with feature band selection, IEEE Trans. Syst. Man Cybern. Part A 40 (6) (2010) 1354–1361.
[13] L. Shen, S. Zheng, Hyperspectral face recognition using3d Gabor wavelets, in: International Conferenceon Pattern Recognition, 2012, pp. 1574–1577.
[14] T.-K. Kim, J. Kittler, R. Cipolla, Discriminative learning and recognition of image set classes using canonical correlations, IEEE Trans. Pattern Anal. Mach. Learn. Intel. 29 (6) (2007) 1005–1018.
[15] R. Wang, S. Shan, X. Chen, W. Gao, Manifold-manifold distance with application to face recognition based on image set, in: IEEE International Conference on Computer Vision and Pattern Reclognition., 2008, pp. 1–8.
[16] R. Wang, X. Chen, Manifold discriminant analysis, in: IEEE International Conference on Computer Vision and Pattern Recognition., 2009, pp. 429–436.
[17] R. Wang, H. Guo, L. Davis, Q. Dai, Covariance discriminative learning: a natural and efficient approach to image set classification, in: IEEE International Conference on Computer Vision and Pattern Recognition., 2012, pp. 2496–2503.
[18] H. Cevikalp, B. Triggs, Face recognition based on image sets, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 2567–2573.
[19] Y. Hu, A. Mian, R. Owens, Face recognition using sparse approximated nearest points between image sets, IEEE Trans. Pattern Anal. Mach. Learn. Intel. 34 (10) (2012) 1992–2004.
[20] CIE 2006 color matching functions. http://cvrl.ioo.ucl.ac.uk/cmfs.htm. 2016.
[21] G.P. Hughes., On the mean accuracy of statistical pattern recognizers, IEEE Trans. Inform. Theory (1968), doi:10.1109/TIT.1968.1054102.
[22] C. Cariou, K. Chehdi, S.L. Moan, Band clustering: an unsupervised band reduction method for hyperspectral remote sensing, GRSL (2011), doi:10.1109/LGRS.2010.2091673.
[23] C. Chang, An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis, IEEE Tran. Inform. Theory (2000), doi:10.1109/18.857802.
[24] S.B. Serpico, G. Moser, Extraction of spectral channels from hyperspectral images for classification purposes, TGRS (2007).
[25] D.A. Landgrebe, Signal Theory Methods in Multispectral Remote Sensing, John Wiley and Sons, 2005, doi:10.1002/0471723800.
[26] A. Richards J, J. Richards, Remote Sensing Digital Image Analysis, Spring, 1999 978-3-642-30061-5.
[27] Chatfield K., Simonyan K., Vedaldi A., and Zisserman A. "Return of the devil in the details: delving deep into convolutional neural networks." arXiv preprint arXiv:1405.3531, 2014
[28] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, NIPS (2012).
[29] B.B. Le Cun, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Handwritten digit recognition with a back-propagation network, NIPS (1990).
[30] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deep face: closing the gap to human-level performance in face verification, HANHAND In CVPR (2014).
[31] X. Yang, X. He, Firefly algorithm: Recent advances and applications, Int. J. Swarm Intell. 1 (1) (Aug. 2013) 36–50.
[32] X. Yang, Firefly algorithm for multimodal optimization, on Stochastic Algorithms, Found. Appl. 5792 (2009) 169–178.
[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Int. Conf. Learning Representations, 2015.
[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Imagenet large scale visual recognition challenge, IJCV (2015).
[35] Goodfellow I. J., Bulatov Y., Ibarz J., Arnoud S., and Shet V., Multi-digit number recognition from street view imagery using deep convolutional neural networks. 2014.
[36] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. P'erez, C. Schmid, Aggregating local image descriptors into compact codes, In IEEE PAMI, 2011.
[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, Going deeper with convolutions., 2014 CoRR, 1409.4842.
[38] F. Schroff, D. Kalenichenko, Philbin.J. Facenet, A unified embedding for face recog. and clust, In Proceeding CVPR, 2015.
[39] L. Denes, P. Metes, Y. Liu, Hyperspectral Face Database, Robotics Inst., Pittsburgh, PA, October 2002 Tech. Rep. CMU-RI-TR-02-25.
[40] V. Turk, Pentland M M., Face recognition using eigenfaces, in: IEEE International Conference on Computer Vision and Pattern Recognition., 1991, pp. 586–591.
[41] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intel. 19 (7) (1997) 711–720.
[42] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, in: Europ. Conf. on Comp. Vision, 2004, pp. 469–481.
[43] M. Mathias, R. Benenson, M. Pedersoli, Van L. Gool, Face detection without bells and whistles, Procedia ECCV (2014).
[44] M. Everingham, J. Sivic, A. Zisserman, Taking the bite out of automatic naming of characters in TV video, Image Vis. Comp. 27 (5) (2009).
[45] ChenQu et.al, Virtual reconstruction of random moving image capturing points based on chaos embedded particle swarm optimization algorithm, Microprocess. Microsyst. 75 (2020) 103069 June 2020.
[46] Sh Yufeng, Interactive design of intelligent machine vision based on human—computer interaction mode, Microprocess. Microsyst. 75 (2020) 103059 June 2020.

Ashok Kumar rai has received the B.E. degree in Computer Science and Engineering from Annamalai University, Tamilnadu, India, in 2012, and the M.E. degree in Computer Science and Engineering from Annamalai University, Tamilnadu, India in 2014. He has been doing his research since January 2015 in Anna University, MIT Campus Chennai. He is having expertise in Firmographic Technographic and Intent data, His research interests include Hyperspectral Imaging, Computer Vision, NLP, Deep Learning, Reinforcement Learning and Data Science.

Dr Radha SenthilKumar has received the B.E. degree in Computer Science and Engineering from Bharathidasan University, Tiruchirapalli, India, 1993, the M.E. degree in Computer Science and Engineering from Anna University, CEG Campus Tamilnadu, India in 2000 and Ph.D. degree in Faculty of I&C from Anna University, CEG Campus in 2012. Her research interests include Hyperspectral Imaging, Machine learning, Deep Learning, Reinforcement Learning XML Technology and Fuzzy logic.