Robust Facial Recognition System using One Shot Multispectral Filter Array Acquisition System

M. Eléonore Elvire HOUSSOU[1], A. Tidjani SANDA MAHAMA[2], Pierre GOUTON[3], Guy DEGLA[4]
ImVIA, University of Bourgogne Franche-Comté, Dijon France[1, 2, 3]
IMSP, University Abomey-Calavi, Dangbo Benin[1, 2, 4]

***Abstract*—Face recognition in the visible and Near Infrared range has received a lot of attention in recent years. The current Multispectral (MS) imaging systems used for facial recognition are based on multiple cameras having multiple sensors. These acquisition systems are normally slow because they take one MS image in several shots, which makes them unable to acquire images in real time and to capture moving scenes. On the other hand, currently there are snapshot multispectral imaging systems which integrate a single sensor with Multispectral Filter Arrays (MSFA) allow having at each acquisition an image on several spectra. These systems drastically reduce image acquisition time and are able to capture moving scenes in real time. This paper proposes a study of robust facial recognition using Multispectral Filter Array acquisition system. For this goal, a MSFA one-shot camera was used to collect the images and a robust facial recognition method based on Fast Discrete Curvelet Transform and Convolutional Neural Network is proposed. This camera covers the spectral range from 650 nm to 950 nm. A comparison of the facial recognition system using Multispectral Filter Arrays camera is made with those that using multiple cameras. Experimental results proved that face recognition systems whose acquisition systems are designed using MSFA perform more efficiently with an accuracy of 100%.**

***Keywords*—Multispectral image database; multispectral imaging; multispectral filter array (MSFA); one-shot camera; facial recognition system**

I. Introduction

The Biometric system is defined as an automatic measurement system based on the recognition of physiological and / or behavioral characteristics specific to a person. It is characterized by its uniqueness, its public nature, and its performance. There are several biometric modalities namely fingerprint, palm sign, face, iris, retina, DNA, voice etc.

Facial recognition is one of the widely used biometric identification methods because face is easy to capture in a controlled or not controlled environment, and in a cooperative or non-cooperative manner [1] [2]. Facial recognition systems performance depends on the electromagnetic spectra in which face images have been acquired. Facial recognition based on the visible spectrum generally use texture characteristics. Its performance is affected by light, occlusions, and pose variations. Infrared spectrum has several advantages over the visible spectrum; it is not perceptible to the human eye and at the same time, less sensitive to variations in light[3] [4]. The infrared spectrum is subdivided into near infrared spectrum (770-1400 nm), Short Wavelength Infrared (1,4–3 µm), mid-wave infrared spectrum (3 – 8 µm) and thermal infrared spectrum (8 - 15 µm). There have been reported some research showing that in environments with uncontrolled illumination the NIR approach remarkably has higher performance in comparison to VIS approach in the extraction of information in different aspects such as appearance and structure [5]. The use of visible and near infrared spectra in facial recognition combines the benefits of both spectra and improves the performance of facial identification systems. Face recognition in the visible and infrared range has received a lot of attention in recent years. A multispectral recognition system is a system using images acquired on 3 to 10 spectral bands. Each of the images acquired in a given band contains specific information that is very important and useful in facial recognition.

MS imaging [6] systems can be broadly grouped into three categories: multi-cameras systems, single-camera and multi-shot systems, and single-camera one-shot. The multispectral systems based on the first category consist of several cameras, at least one per inference filter (or spectral band filter), multispectral systems using a single camera and several shots are made up of several image sensors each equipped with a narrow bandwidth wavelength filter, which makes them heavy, bulky, energy-intensive and very expensive. The last category that of single camera one shot, uses a single sensor with Spectral Filter Array (SFA) or Multispectral Filter Array (MSFA) to acquire a single image on multiple spectral band simultaneously. These imaging systems are very fast and operate in real time.

Most current MS imaging facial recognition systems use acquisition systems consisting of multiple cameras or a single camera with multiple sensors. Considering the real time operations and benefits of these one-shot multispectral cameras, the research has been oriented towards facial recognition using multispectral images acquired with Multispectral Filter Array camera. This paper proposes a robust facial recognition system using MS image dataset collected with MSFA one shot camera that covers the visible and near infrared spectrum. Fast Discrete Curvelet Transform(FDCT)[7] ,VGG19 [8] and ResNet 101[9] convolutional neural networks have been used to develop recognition end.

This paper is organized as follows: Section 2 focuses on different MS facial recognition systems used in the literature. Section 3 describes our Multispectral Filter Array one shot acquisition system and the method. The experimental results are reported in Section 4. Section 5 is dedicated to the discussion and the conclusion is presented in the last section.

II. Related Work

The use of multiple spectral bands improves the performance of facial recognition system, this explains the interest of researchers on MS facial recognition in recent years [5] [10] [11]. A face recognition system can be represented by four main modules: capture, feature extraction, matching and decision. The capture module is mainly based on image acquisition system that enables the acquisition of images. The acquisition system consists of one or several cameras. The feature extraction module takes the acquired images and extracts only the relevant information in order to build a new data representation. The matching module compares the set of extracted features with the features of those images stored by the system in the database during enrollment. The decision module verifies the identity asserted by a user based on the degree of similarity between the extracted features and the ones from the database. The following figure (Fig. 1) illustrates the architecture of a face recognition system.

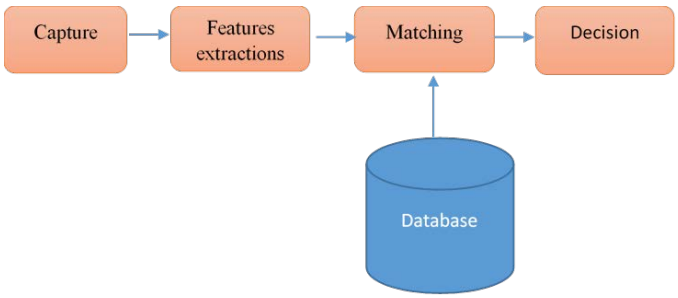

Fig. 1. Architecture of a Facial Recognition System.

There are several facial recognition systems operating in the visible and Near InfraRed range, but in the most of the cases images were not acquired in real time.

Abood et al. [12] proposed a face recognition system using Fusion of Multispectral Imaging to overcome the limitations of visible facial recognition. This recognition system merges features from visible, near infrared and thermal infrared images. This system did not acquire images in the capture module but used images from the Carl database for the other modules. Carl Face images Database contains visible, NIR and thermal images of 41 persons. This database had been collected with two cameras: customized Logitech Quickcam messenger E2500 with a Silicon based CMOS image sensor for near infrared image and thermal camera TESTO 880-3 (incorporating an uncooled detector with a spectral sensitivity range from 8 to 14 µm and provided with a germanium optical lens) for visible and thermal images. Auteurs used Gabor wavelet transform for feature extraction and Support Vector Machine (SVM) for classification. Experimental results achieve a recognition rate of 96, 4%.

Y.Jin et al. [13] have developed a Coupled Discriminative Feature Learning for Heterogeneous Face Recognition. They implement a method that represents the discriminative features of the face by building an optimal filter eigenvector with the raw pixels of the image. The developed approach uses Local Ternary Patterns for encoding local patterns and cosine metrics to estimate the similarities between images. The performance of this method was tested on CASIA 2.0 NIR-VIS database. This database is widely used in publications. CASIA NIR-VIS 2.0 was collected from 2007 to 2010 by Stan Z. Li et al.[5], it contains visible and near infrared frontal images of 735 subjects. Two cameras were used to acquire the visible and near infrared images. The visual color face images are captured using Canon A640 Camera and home-brew device were used for near infrared image acquisition. The NIR imaging device used to capture NIR image is a standard version for indoor use and under fluorescent lighting. A long pass optical filter is integrated in this camera and allows capturing images in wavelengths 720, 800, 850 and 880 nm. The spatial resolution of acquired images is 640 × 480 images. Experimental results indicated that the accurate recognition rate is less than 90%.

In 2021 R. He et al. [14] have proposed a Coupled Adversarial Learning (CAL) system for semi-supervised heterogeneous face recognition. This approach has used the VIS-NIR face matching by performing adversarial learning on both image and feature levels. VIS images had been generated from the unmatched VIS-NIR images. This system did not acquire images in the capture module but used images from CASIA NIR-VIS 2.0 database for the other modules. A series of end-to-end neural network composed with 29 convolution layers with residual blocks (LightCNN-29) have been used to extract and learn features. The experimental results indicated a rank 1 accuracy from 98.6% to 99.6%.

A. Yu et al [15] have implemented face recognition system that used Generative Adversarial Networks(GANs) in the VIS and NIR. The VIS images have been generated from the NIR images. In order to reduce the domain gap between the NIR and VIS data, an attention module has been developed by the authors. This system has acquired images for visible and NIR range with two camera. A Large-Scale Multi-pose High-Quality Database of NIR-VIS images called LAMP-HQ was collected. A LightCNN-29 has been used as classifier. The performance achieved with this system has showed a rank 1 accuracy from 94.94% to 97.91%.

Song et al. [16] have implemented an adversarial discriminative feature learning framework for VIS and NIR face recognition. In order to compensate the detection gap the authors have applied the methods based on adversarial learning on both the raw pixel space and the compact feature space. This approach combines ResNet and Light CNN. The experiment has been performed on three databases: CASIA NIR-VIS 2.0 database, BUAA-VisNir [5] face database and Oulu-CASIA NIR-VIS facial expression database. Experimental results achieved give respectively for the tree databases a rank 1 accuracy of 98,15%, 95.2% and 95.5%. The BUAA-VisNir face database has been created by D. Huang et al. It contains 162 images/person of 150 persons. The images were acquired in visible and NIR range with 9 different facial expressions.

Oulu-CASIA NIR&VIS facial expressions database was set up by Chinese Academy of Sciences. It contains videos with six typical expressions i.e. happiness, sadness, surprise, anger, fear and disgust from 80 subjects captured with two imaging systems (SN9C201 & 202) which combines a NIR camera and a visible camera. This imaging system captures NIR and visible images under three different illumination conditions

normal indoor illumination, weak illumination and dark illumination. The imaging hardware works at the rate of 25 frames per second and the image resolution is 320 × 240 pixels.

In order to correct misalignment problems between visible and NIR matched images P. Zao et al. [17] have developed a Self-Aligned Dual NIR-VIS Generation for Heterogeneous Face Recognition. The architecture proposed by the authors is based on GANS and allows generating semantically aligned dual NIR-VIS images with the same identity. This system has not acquired images in the capture module but has used images from CASIA NIR-VIS 2.0, Oulu-CASIA NIR-VIS and BUAA VIS-NIR. The features have been extracted with ligthCNN and two encoder networks for generation tasks. A rank 1 accuracy close to 99.9% has been performed for each datasets.

M. Diarra et al. [18] have proposed the MS-FRHF (Multispectral Face Recognition using Hybrid Feature) approach for visible and thermal. They used Robotics Intelligent System (IRIS) database for feature extraction. The points of interest and the texture have been extracted respectively with the Maximally Stable Extremal Region (MSER) keys points extractor and Gray Level Co-Occurrence Matrix (GLCM). Principal Component analysis (PCA) was used to fuse the feature. Authors have concluded this approach gives the best recognition rates than those obtained in the visible and thermal infrared.

In 2020, Zhihua Xie et al. [19] have developed the fusion methods based on the local binary model and the discrete cosine transform for face recognition in the infrared and visible range. For this purpose, the low frequency information is first extracted from the near-infrared images with the discrete cosine transform and the LBP is applied to represent the discriminative features. Then the features of the visible images have been extracted with the LBP and finally a fusion has been done. Experimental results have shown that the recognition rate has been improved with this approach.

Guo et al. [20] have proposed Face recognition system using both visible light image and near-infrared image and a deep network. This system uses two different cameras, one in the visible and the other in the near infrared. The visible and near infrared features have been first extracted with the neural network. Then the authors have used the cosine distance to determine the classification score. Finally, a fusion of the classification scores has been performed. They have used HIT LAB2 and SunWin Face database to test the performance of their model. Accuracy of 99.89% and 99.56% has been achieved on the two databases respectively in weak light change.

Hu et al. [21] have presented a Discriminant Deep Feature Learning based on joint supervision Loss and Multi-layer Feature Fusion for heterogeneous face recognition. This approach implements Convolutional Neural Networks by integrating a loss function called scatter loss in order to improve the discriminating power of the learned features in depth. The features extracted by the CNNs in the different visible and near infrared bands have then been merged. The performance of the system has been tested on CASIA NIR-VIS 2.0 and Oulu-CASIA NIR-VI databases. The experimental results have given a rank 1 accuracy of 98.5% to 98.8% on the CASIA NIR-VIS 2.0 dataset, and of 98.5% to 99.3% on Oulu-CASIA NIR-VIS database.

F. Wei et al. [22] have developed an intraspectrum discrimination and interspectrum correlation analysis deep network (IDICN) approach for facial recognition. This system has improved the performance of multispectral face recognition by including inter- and intra-spectral information. Authors didn't acquire images but have used three databases: Hong Kong Polytechnic University (HK PolyU) dataset, Carnegie Mellon University (CMU) dataset and the University of Australia (UWA) dataset. This approach consists of a set of spectrum-set-specific deep convolutional neural networks with a spectrum pooling layer. The convolutional neural networks extract features related to a set of spectra, and the spectrum pooling layer selects a group of spectra with discriminative capabilities.

The HK PolyU dataset consists of 48-subject hyper-spectral image cubes, which are acquired using CRIs VariSpec liquid crystal tunable filter (LCTF) under halogen light. The spectral range extends from 400 to 720 nm with a step size of 10 nm.

CMU database is collected with a prototype spectro-polarimetric camera developed by CMU. It contains the images of 54 subjects. The hyper-spectral range is between 450 and 1090 nm with a step length of 10 nm.

The UWA dataset consists of 120 hyper-spectral image cubes of 70 subjects acquired with the VariSpec LCTF CRIs integrated with a photonic focusing camera. Each hyperspectral image cube contains 33 bands covering the spectral range from 400 to 720 nm with a 10 nm step size.

Experimental results have achieved average recognition rates of 99.76%, 100% and 99.85 respectively on the three bases.

III. Materials and Methods

A. Our One-shot Multispectral Filter Array Acquisition System

This section describes the Multispectral Filter Array one shot camera used for image database collection.

In recent years, sustained research efforts have been carried in the field of multispectral imaging systems incorporating MSFA. Multi-spectral imaging using a single camera with MSFA is an efficient way to acquire spectral data. It has the potential to promote a fast and real time multispectral imaging system. The concept of Spectral (or Multispectral) Filter Arrays has been developed recently and enables one shot multispectral acquisition with a compact camera design. Generally, Multi-Spectral filter array (MSFA) is made up recurrent patterns of filtering elements. Multi-Spectral filter array (MSFA) cameras are a new single-shot spectral imaging technology that is defined by a basic repetitive pattern composed of filter elements. Each filtering element is sensitive to a specific spectral band. Multi-spectral Filters Array are filter matrix in which each filter corresponds to a spectral band. During the technical design of the cameras the filters are carefully selected. MSFA one-shot camera architecture shows

that the MSFA overlap the camera sensor so as to cover it. The light entering the camera is filtered with a band pass spectral filters on each pixel. A MSFA aims at object property estimation and/or objective color measurement. A MSFA might be defined by its moxel, mosaic element, which corresponds to the occurrence of a pre-defined pattern that consist of a set of filters arranged geometrically in a relative manner [23]. The moxels or multispectral pixels are the smallest patterns in the MSFA. An overview of the global approach is shown in following figure (Fig. 2).

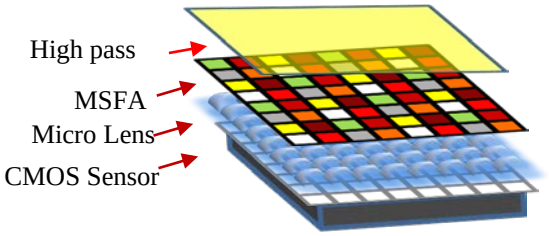

Fig. 2. Global Scheme of MS Imaging System with MSFA.

Multispectral Filter Array Camera was used to create the multispectral images database. The oneshot acquisition system has been designed in the Electronics, Computer and Image (LE2I) Laboratory which is now ImViA (Imaging and Artificial Vision) during UE H2020 project called EXIST (EXtended Image Sensing Technologies). It is a compact and lightweight acquisition system that integrates a single Viimagic 9220H sensor, MSFA for one-shot imaging system, Optic lenses, Electronic board for driving the sensor and Camera board for image acquisition. In order to correct the linearity of the sensor and also to measure the spectral sensitivity, a characterization of the CMOS sensor has been made before mounting the Multi-Spectral filter Arrays. The MSFA has been selected carefully considering a regular distribution of pixels in the moxel. Our personalized filter matrix was built using SILIOS technologies. SILIOS Technologies has developed the COLOR SHADES® technology, which use Fabry-Perot interferometer to manufacture multispectral transmittance filters. This technology is based on the combination of thin film deposition and micro- / nano-etching processes on a fused silica substrate. Standard micro-photolithography steps are used to define the cell geometry of the multispectral filter. COLOR SHADES® provides band pass filters originally in the visible range from 400 nm to 700 nm. SILIOS has developed filters in the NIR range in collaboration with LE2I (ImViA) laboratory, combining their technology with a classical thin film interference technology to realize our filters. The MSFA system, integrated into a camera with dedicated hardware and software computations, allows operating in real-time application with 30 fps. The filters used overcome the problems caused by lighting variation, motion blur noise and SNR noise which severely affect the performance of facial recognition systems using CMOS. These are 8 optimal filters selected in the wavelengths {685, 720,770, 810, 835, 870, 895, 930} (in nm) thanks to a technical study carried out at LE2I laboratory. Fig. 3 illustrates the spatial distribution of moxel.

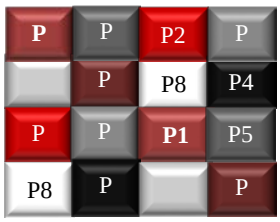

Fig. 3. Final Moxel of MSFA.

This acquisition system is a light robust, broadband multispectral system which tends to measure the physical properties of the object. It is also a real time system which covers visible and near infrared spectra (650 to 950 nm). The MSFA integrated in our acquisition system has a small moxel 4x4 with 2 pixels per band and the size of filter pitch is 5x5 µm2. At each acquisition, the MSFA one shot camera provides a best resolution of raw or mosaic image of size 2072 x1104 pixels in which the values of every channel are accessible at each pixel according to the MSFA. The missing channel values are there after estimated by demosaicking process. In order to provide an optimal solution for the loss of spatial resolution inherent to MSFA, specific algorithms have been developed for multispectral demosaicking. Indeed, the demosaicking process must be associated with the design of the MSFAs otherwise the loss of image resolution could be critical. As the acquisition system is an MSFA one shot camera, it privileges spectral resolution. The Fig. 4 presents the MSFA camera.

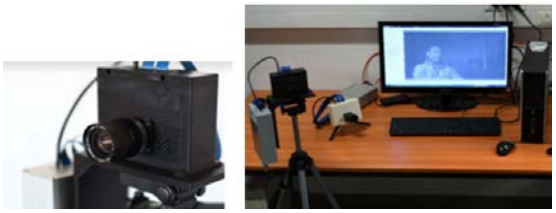

Fig. 4. MSFA One-Shot Camera.

B. Data Collection

Multispectrals face images were collected over two years in Imaging and Artificial Vision (ImViA) Laboratory in Faculty of Science and Technology of Burgundy University in France. The acquisition room is a black room, dedicated to MS imaging. The photos were taken with an illuminant light with different orientations. This light illuminates the subject's face to be photographed. The light is oriented from left to right and vice versa during the acquisition. The distance between the camera and the subject to be acquired is 1 meter. The relative position of camera is shown in Fig. 5.

The MS images database have been acquired in winter 2020, winter 2021 and spring 2021. Participants were made up of residents and international students, and 75% of subjects agreed to have their photo posted. They are men and women of all ages, black and white. Participants are from different continents namely Europe, Asia, Africa, Arabic and African. The multi-spectral image database named EXIST MS database contains faces images of 103 subjects. Face images MS database is structured as follows:

- All images are tiff format and size 2072 x 1104;

- Our database contains 20 different mosaic faces images per subject;
- 20x8 demosaic faces images per subject.

In total our MS images database contains 103x20x8 (16 480) MS images.

The following figure (Fig. 6) shows some mosaic face images of a subject in the database.

Fig. 5. Acquisition Set-up in the Black Room.

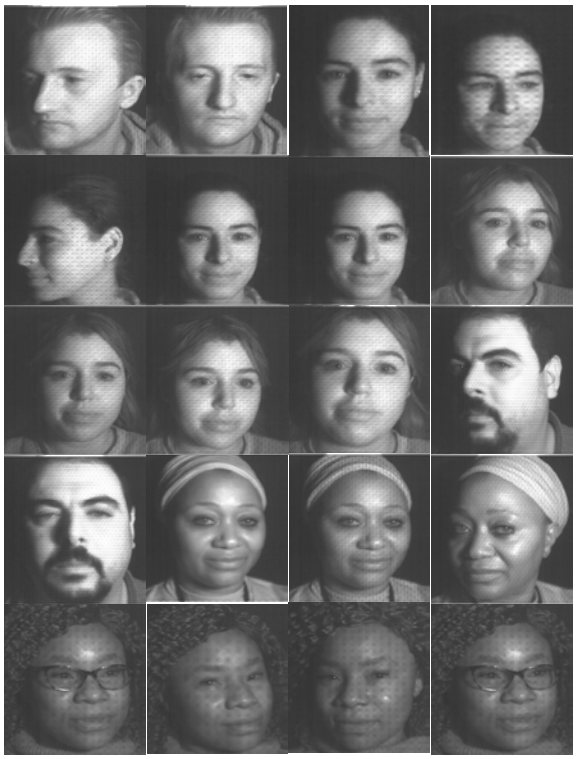

Fig. 6. Sample Images from our MS Database.

C. Image Demosaicking Algorithm

Having used a multispectral acquisition system which provides a mosaic image, a demosaicking is necessary to generate multi-band images. In fact, a mosaic or raw image is of size (X x Y) pixels, in which a single band $k \in \{1. . . K\}$ is refers to the value of each pixel p depending on the MSFA structure. Image demosaicking is a process that separates mosaic or raw images into multispectral images according to filter number in the MSFA. Before the demosaicking, a strip extraction was done. It consists of multiplying the mosaic image by different binary masks $M^k(x,y)$, $k \in \{1,2, \ldots K\}$ [24](pp. 31-34).

These masks have the value 1 at the positions where the component is available, and 0 at the other positions. Each plane component is obtained after multiplying the mosaic image by the corresponding mask M^k.

$$M^k_{(x,y)} = \begin{cases} 1 \; si \; \left(x \; mod \; \sqrt{K}\right) + \left(y \; mod \; \sqrt{K}\right) \times \sqrt{K} = k \\ 0 \; sinon \end{cases} \quad (1)$$

In this case, the multiplication of the mosaic image by each mask allows us to obtain 8 planes of shifted images in which only one component is available at each pixel. Each mask corresponds to an image plane I'^k.

$$I'^k = I \odot M^k \quad (2)$$

Where $\odot$ denotes the element-wise product and M^k is a binary mask defined at each pixel p.

Bilinear interpolation method is used for multispectral demosaicking. This method enables to estimate the missing in each pixel. Bilinear interpolation can be expressed as a resampling technique based on distance weighted average of the four nearest pixel values to evaluate each missing pixel value. Bilinear interpolation is a succession of two linear interpolations, each in one direction. The linear interpolations can be performed in multiple directions. For a missing pixel *P(i,j)* at position *(i,j)* , the linear interpolation is defined as follows:

- Diagonally

$$P(i,j) = \frac{1}{4} \sum_{(m,n)=(-1,-1),(-1,1),(1,-1),(1,1)} p(i+m, j+n) \quad (3)$$

- Vertically

$$P(i,j) = \frac{1}{2} \sum_{(m,n)=(-1,0),(1,0)} p(i+m, j+n), \quad (4)$$

- Horizontally

$$P(i,j) = \frac{1}{2} \sum_{(m,n)=(0,-1),(0,1)} p(i+m, j+n) \quad (5)$$

The multispectral image demosaicking using bilinear interpolation also consists in convolving each component plane obtained by an H filter. This filter is determined as a function of the spatial distance between the neighbors from the central pixel.

Two filters H1 and H2 (Fig. 7) were used to do the convolution. The interpolated image band is defined by

$$I^k = I'^k \odot H \quad (6)$$

With H=H1 or H=H2

1/9	2/9	1/3	2/9	1/9
2/9	4/9	2/3	4/9	2/9
1/3	2/3	1	2/3	1/3
2/9	4/9	2/3	4/9	2/9
1/9	2/9	1/3	2/9	1/9

1/5	1/4	1/5	1/3	0
0	1/3	1/5	1/4	1/5
1/5	1/3	1	1/3	1/5
1/5	1/4	1/5	1/3	0
0	1/3	1/5	1/4	1/5

Fig. 7. Filter H1 and H2.

To estimate the missing pixel value *P(i,j)*, the image I'^k is convolved with each of the H1 and H2 filters. If V1 and V2 are convolution results then the missing pixel value is defined by average of V1 and V2. This process is used to estimate each missing pixel in I'^k. The figure (Fig. 8) shows the missing pixel value estimation in demosaicing process.

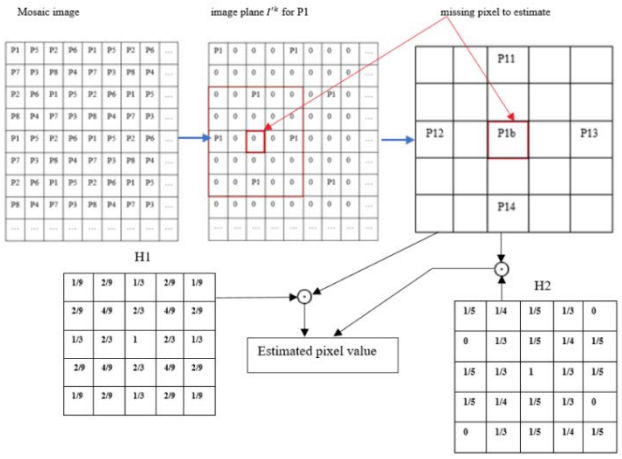

Fig. 8. Missing Pixel Estimation for Demosaicing Process.

The image demosaicing generates 8 images belonging to the wavelengths {685, 720,770, 810, 835, 870, 895, and 930} (in nm). The figure (Fig. 9) illustrates the demosaicking images process.

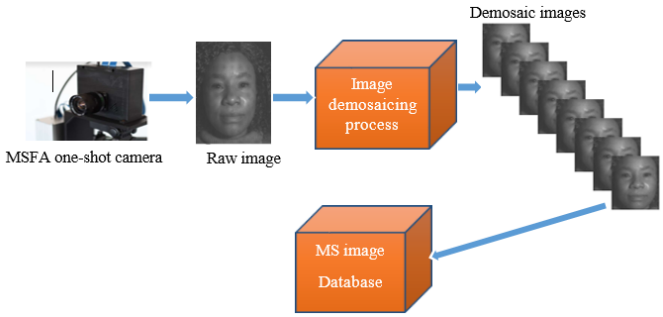

Fig. 9. The Process of Multispectrals Images Demosaicking.

D. Methodology

This section describes the algorithms used for facial recognition with the Multi-Spectral Filters Array camera described above. Fast Discrete Curvelet Transform (FDCT) and Convolutional Neural Networks are used to implement the facial recognition system. In order to exploit the important information contained in each spectral band we first use a fusion at the image level with the FDCT. First, the image-level fusion method is implemented with FDCT. Then, VGG19 and ResNet 101 neural networks are used to perform the recognition. The following figure (Fig. 10) illustrates the proposed method:

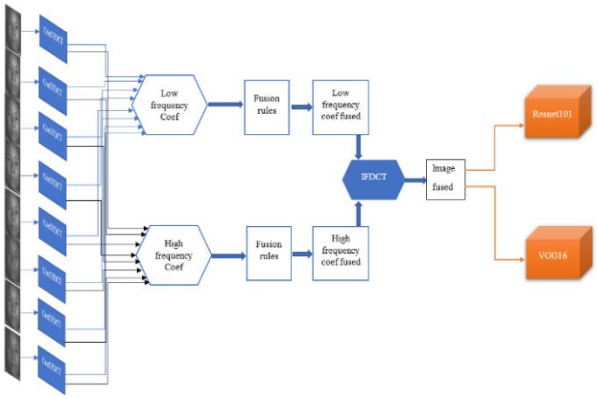

Fig. 10. Bloc Diagram of the Proposed Face Recognition Method using MSFA One-shot Camera.

The Fast Discrete Curvelet Transform (FDCT) is a wavelet transform that decomposes an image in low and high frequency. FDCT adopts local Fourier transform for the frequency domain decomposition. First, each of the eight demosaic image was decomposed with the FDCT. Then the low frequency coefficients are merged together and so for the high frequency coefficients. Inverse Fast Discrete Curvelet Transform was applied to obtain a merged image containing the information of all bands. Finally transfer learning for VGG19 and ResNet 101 are used to classify the images fused with FDCT method.

IV. Experiments and Results

EXIST is the image database acquired with the described acquisition system and the one used evaluation purpose.

All the images are acquired from 20 different positions per person; in total 2000 images in the multi-spectral images database are taken.

Experimentations are carried out on Microsoft System windows, version 2010, with two computers. The first one was equipped with an Intel (R) Core (TM) i7-8565U CPU, 8 GB of RAM memory. The second has a graphical processing unit (GPU) NVIDIA Quadro P400 with 32GB of Random Access Memory (RAM). All the code are developed in the programming language of Matlab 2020 and Python 3.7.

Table I describes the training parameters for VGG19 and ResNet101.

To analyse the results, the following performance evaluation metrics have been calculated: accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC) and Means Square Error (MSE).

An accuracy indicates the percentage of correct predictions.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

TABLE I. PARAMETERS USED IN THE TRAINING PROCEDURE

Parameters	CNN	
	VGG19	***ResNet101***
Batch size	32	16
Optimization algorithm	SGD[a]	ADAM
Learning rate	0.0001	0.0001
Epoch number	20	20

[a.] Stochastic Gradient Descent

Where *TP* (True Positive), *TN* (True Negative), *FP* (False Positive), *FN* (False Negative).

The precision is the proportion of true positives out of all detected positives.

$$precision = \frac{TP}{TP+FP} \quad (8)$$

The recall is the number of true positives that are correctly classified.

$$recall = \frac{TP}{TP+FN} \quad (9)$$

Component F1 score includes recall and precision and is calculated as

$$F1_{score} = \frac{2*precison*recall}{precision+recall}) \quad (10)$$

The Matthews Correlation Coefficient (MCC) is the method of calculating correlation coefficient between real and predicted values. MCC is more informative score and give best result in binary classification assessment [25]. The range of values of MCC is between -1 and 1. A model with a score of 1 is a perfect model and -1 is a poor model.

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \quad (11)$$

The Mean Squared Error (MSE) allows to calculate error between predict values $\hat{y}$ and reals value y.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (12)$$

Where N is the number of samples we are testing against.

The different metrics calculated allow for a better interpretation of the results. The accuracy of a model predicts the percentage of all persons that can be recognized by the model. The recall indicates the number of correct predictions that were actually recognized.

The fusion rules such as min, max and average are used to obtain the images. Three types of experiments were done: a first one with the images obtained with FDCT and the min fusion rule, a second one with the images obtained with FDCT and the max fusion rule and a third one with the average fusion rule.

In Table II and Table III, the performance of the VGG19 and ResNet101 model for min, max and average fusion are listed respectively.

TABLE II. VGG19 RESULTS

Fusion	Metrics				
	Precision	***MSE***	***F-score***	***Recall***	***MCC***
average	1.0	2.80e-07	1.0	1.0	1.0
min	0.9849	7.96e-07	0.98	0.98	0.98
max	0.9870	6.76e-07	0.98	0.98	0.98

TABLE III. RESNET101 RESULTS

Database	Metrics				
	Precision	***MSE***	***F-score***	***Recall***	***MCC***
average	0.99	4.91e-10	0.99	0.99	0.99
min	0.97	7.98e-10	0.97	0.97	0.97
max	0,96	91e-10	0,96	0,96	0,96

Note that the two models VGG19 and ResNet101 have been trained with the images of the databases, taking into account different batches size 8, 16 and 32.

By comparing the different metrics calculated on the two neural networks, the results show that the performance of average fusion rule method are superior to those of min and max fusion rule.

Table IV describes the recognition rate of stat of art recognition and the proposed system.

TABLE IV. THE RECOGNITION RATES

Acquisition system	Recognition methods	Recognition Rate
Visible camera + NIR camera + thermal camera	Gabor wavelet transform and Support Vector Machine (SVM) [12]	96.4%
Visible camera and NIR camera	Local Ternary Patterns and cosine metrics [13]	90%
Visible camera and NIR camera	LightCNN-29 [14]	98.6% to 99.6% [a]
Visible camera and NIR camera	LightCNN-29 [15]	94.94% to 97.91%.
Visible camera and NIR camera	ResNet and LightCNN-29[16]	98.15%, 95.2% and 95.5% [b]
Visible camera and NIR camera	ligthCNN and two encoder networks[17]	99.9%
Visible camera and NIR camera	Neural Network and cosine distance[20]	99.89% and 99.56%
Visible camera and NIR camera	CNN[21]	98.5% to 98.8% and 98.5% to 99.3%
Hyper-spectral camera	deep convolutional neural networks[22]	99.76%, 100% and 99.85
EXIST camera	FDCT and VGG19	100%
EXIST camera	FDCT and ResNet 101	99%

[a.] Results achieved between two values

[b.] Results achieved on different databases

Table IV indicates that most facial recognition systems in the visible and NIR range use multiple cameras. Depending to the image database used, the recognition rate range 95.3% to 100%. The rate of 100% was reached with the deep convolutional neural network algorithm. In general systems using neural networks algorithm have a rate close to 100%. In this case, depending to the recognition algorithm, the rate is range 90% to 100%. The face recognition systems that use an acquisition system integrating Multispectral Filters Arrays (MSFA) achieve perform as well as those using several cameras.

V. Discussion

In this article, face recognition based on MSFA oneshot camera were demonstrated. Most previous studies in the literature used multiple cameras and Convolutional neural to extract features for face identification. Y.Jin et al. in [13] used multiple cameras with Local Ternary patterns, cosine metrics and achieve recognition rate of 90%. The recognition systems presented in the literature that are based on Gabor wavelet Transform and Support Machine Vector (SVM) achieve a recognition rate of 96.4% [12]. The recognition systems based on Convolutional Neural Network [14],[15],[16],[18],[22] achieve respectively accuracy in [98.6% - 99.6%], [94.94% - 97.91%] and [98.15%, 95.2%,95.5%], [99.89%, 99.56%],[98.5%-98.8%, 98.5%-99.3%] depending on database. Also [23] used hyperspectrals camera with neural networks and get [99.76%, 100%, 99.6%] accuracies on three different databases. Results and experiments of the facial recognition using MSFA oneshot camera achieve accuracies of 99% and 100% with respectively Resnet101 and VGG19.

The comparison of the results shows that the different facial recognition algorithms implemented give good performance depending on the images and methods used. The proposed system reaches one of the best performances. But also because of its camera and its algorithms, it is the only system which proposes a real time acquisition of images.

VI. Conclusion

This paper presents a new multispectral facial recognition system using one shot multispectral imaging systems integrated Multispectral Filters Array for acquisition. It is a one-shot acquisition system that operates in real time on the visible and NIR spectra. A multispectral database containing images with spectral information has been created for this end. This recognition system is based on the Fast Discret Curvelet Transform, ResNet 1O1 and VGG19 Convolutional Neural Network. Experimental results show that face recognition systems using MSFA cameras perform as well as those using multiple cameras. This system is however more interesting as it is more economical, technically reliable and especially equipped with a real time acquisition system.

In future work, the image database will be extended; other multispectral demosaicing and recognition algorithms will be implemented.

References

[1] L. Kambi Beli and C. Guo, “Enhancing Face Identification Using Local Binary Patterns and K-Nearest Neighbors,” Journal of Imaging, vol. 3, no. 3, Art. no. 3, Sep. 2017, doi: 10.3390/jimaging3030037.

[2] M. Lal, K. Kumar, R. H. Arain, A. Maitlo, S. A. Ruk, and H. Shaikh, “Study of Face Recognition Techniques: A Survey,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, no. 6, Art. no. 6, 29 2018, doi: 10.14569/IJACSA.2018.090606.

[3] R. He, X. Wu, Z. Sun, and T. Tan, “Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp. 1761–1773, Jul. 2019, doi: 10.1109/TPAMI.2018.2842770.

[4] Z. Xie, P. Jiang, and S. Zhang, “Fusion of LBP and HOG using multiple kernel learning for infrared face recognition,” in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), May 2017, pp. 81–84. doi: 10.1109/ICIS.2017.7959973.

[5] L. L. Chambino, J. S. Silva, and A. Bernardino, “Multispectral Facial Recognition: A Review,” IEEE Access, vol. 8, pp. 207871–207883, 2020, doi: 10.1109/ACCESS.2020.3037451.

[6] M. Mateen, J. Wen, Nasrullah, and M. A. Akbar, “The Role of Hyperspectral Imaging: A Literature Review,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, no. 8, Art. no. 8, 49/01 2018, doi: 10.14569/IJACSA.2018.090808.

[7] Y. Yang, S. Tong, S. Huang, P. Lin, and Y. Fang, “A Hybrid Method for Multi-Focus Image Fusion Based on Fast Discrete Curvelet Transform,” IEEE Access, vol. 5, pp. 14898–14913, 2017, doi: 10.1109/ACCESS.2017.2698217.

[8] T. Gwyn, K. Roy, and M. Atay, “Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study,” Future Internet, vol. 13, no. 7, Art. no. 7, Jul. 2021, doi: 10.3390/fi13070164.

[9] H. Ling, J. Wu, L. Wu, J. Huang, J. Chen, and P. Li, “Self Residual Attention Network for Deep Face Recognition,” IEEE Access, vol. 7, pp. 55159–55168, 2019, doi: 10.1109/ACCESS.2019.2913205.

[10] Y. Park and B. Jeon, “An Acquisition Method for Visible and Near Infrared Images from Single CMYG Color Filter Array-Based Sensor,” Sensors, vol. 20, no. 19, p. 5578, Sep. 2020, doi: 10.3390/s20195578.

[11] X. Chen, H. Wang, Y. Liang, Y. Meng, and S. Wang, “A Novel Infrared and Visible Image Fusion Approach Based on Adversarial Neural Network,” Sensors, vol. 22, no. 1, Art. no. 1, Jan. 2022, doi: 10.3390/s22010304.

[12] Z. Abood, G. Karam, and R. Haleot, Face Recognition Using Fusion of Multispectral Imaging. 2017, p. 112. doi: 10.1109/AIC-MITCSA.2017.8722957.

[13] Y. Jin, J. Lu, and Q. Ruan, “Coupled Discriminative Feature Learning for Heterogeneous Face Recognition,” IEEE Transactions on Information Forensics and Security, vol. 10, no. 3, pp. 640–652, Mar. 2015, doi: 10.1109/TIFS.2015.2390414.

[14] R. He, Y. Li, X. Wu, L. Song, Z. Chai, and X. Wei, “Coupled adversarial learning for semi-supervised heterogeneous face recognition,” Pattern Recognition, vol. 110, p. 107618, Feb. 2021, doi: 10.1016/j.patcog.2020.107618.

[15] A. Yu, H. Wu, H. Huang, Z. Lei, and R. He, “LAMP-HQ: A Large-Scale Multi-pose High-Quality Database and Benchmark for NIR-VIS Face Recognition,” Int J Comput Vis, vol. 129, no. 5, pp. 1467–1483, May 2021, doi: 10.1007/s11263-021-01432-4.

[16] L. Song, M. Zhang, X. Wu, and R. He, “Adversarial Discriminative Heterogeneous Face Recognition,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, Art. no. 1, Apr. 2018.

[17] P. Zhao, F. Zhang, J. Wei, Y. Zhou, and X. Wei, “SADG: Self-Aligned Dual NIR-VIS Generation for Heterogeneous Face Recognition,” Applied Sciences, vol. 11, no. 3, Art. no. 3, Jan. 2021, doi: 10.3390/app11030987.

[18] M. Diarra, P. Gouton, and A. K. Jérôme, “Multispectral face recognition using hybrid feature,” Electronic Imaging, vol. 2017, no. 18, pp. 200–203, Jan. 2017, doi: 10.2352/ISSN.2470-1173.2017.18.COLOR-061.

[19] Z. Xie, L. Shi, and Y. Li, “Two-Stage Fusion of Local Binary Pattern and Discrete Cosine Transform for Infrared and Visible Face Recognition,” in Emerging Trends in Intelligent and Interactive Systems and Applications, Cham, 2021, pp. 967–975. doi: 10.1007/978-3-030-63784-2_117.

[20] K. Guo, S. Wu, and Y. Xu, “Face recognition using both visible light image and near-infrared image and a deep network,” CAAI Transactions

on Intelligence Technology, vol. 2, no. 1, pp. 39–47, Mar. 2017, doi: 10.1016/j.trit.2017.03.001.

[21] W. Hu and H. Hu, "Discriminant Deep Feature Learning based on joint supervision Loss and Multi-layer Feature Fusion for heterogeneous face recognition," Computer Vision and Image Understanding, vol. 184, pp. 9–21, Jul. 2019, doi: 10.1016/j.cviu.2019.04.003.

[22] F. Wu et al., "Intraspectrum Discrimination and Interspectrum Correlation Analysis Deep Network for Multispectral Face Recognition," IEEE Transactions on Cybernetics, vol. 50, no. 3, pp. 1009–1022, Mar. 2020, doi: 10.1109/TCYB.2018.2876591.

[23] P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, "Multispectral Filter Arrays: Recent Advances and Practical Implementation," Sensors, vol. 14, no. 11, pp. 21626–21659, Nov. 2014, doi: 10.3390/s141121626.

[24] S. Mihoubi, "Snapshot multispectral image demosaicing and classification," Theses, Université de Lille, 2018.

[25] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment," IEEE Access, vol. 9, pp. 78368–78381, 2021, doi: 10.1109/ACCESS.2021.3084050.