Article

Hyperspectral Face Recognition with adaptive and parallel SVMs in partially hidden face scenarios

Julián Caba [1], Jesús Barba [1], Fernando Rincón [1], José Antonio de la Torre [1], Soledad Escolar [1] and Juan Carlos López [1]

[1] University of Castilla-La Mancha, Ciudad Real 13071, Spain
* Correspondence: julian.caba@uclm.es

Version October 8, 2022 submitted to Sensors

Abstract: Hyperspectral imaging opens up new opportunities for masked face recognition via discrimination of the spectral information that are obtained by hyperspectral sensors. In this work, we present a novel algorithm to extract facial spectral-features from different regions of interests by performing computer vision techniques over the hyperspectral images, particularly Histogram of Oriented Gradients. We have applied this algorithm over the UWA-HSFD dataset to extract the facial spectral-features and then a set of parallel Support Vector Machines with custom kernels, based on the cosine similarity and euclidean distance, have been trained on fly to classify unknown subjects/faces according to the distance of the visible facial spectral-features, i.e. the region that are not behind a face mask or a scarf. The results draw up an optimal trade-off between recognition accuracy and compression ratio in accordance with the facial regions that are not occluded.

Keywords: facial recognition; hyperspectral compression; hyperspectral imaging; biometrics; SVM; computer vision

1. Introduction

Face recognition is a special branch of biometrics to identify faces and it is considered an easy task for humans but a challenge when a machine is employed to the automatic face recognition. Traditionally, this process has been performed through an analysis of the face features in which computer algorithms pick out specific, distinctive details about a person's face. These details, such as distance between the eyes or mouth, are then converted into a mathematical representation (face encoding vector) and compared to other faces previously collected in a database [1]. In recent times, computer vision applications have been highly engaged via deep learning techniques. On this basis, most recent works advocate for the use of neural networks for face recognition, whose results are very promising [2].

Although face recognition is no longer considered a challenge due to the good results obtained by the variety of techniques and algorithms published in the scientific community, this topic is back in the limelight when the scenario is not the usual one, e.g. in scenes where some details of the face are hidden. In this sense, the outbreak of COVID-19 pandemic has introduced a new way of life into our lives, e.g. the use of face masks in public and private places, such as public transport, is mandatory in some countries or workplaces according to restrictions imposed by health authorities, mostly based on the status of virus transmission. However, the use of face masks compromises security due to criminals can hide their face under them as well as we are not able to distinguish the person behind the mask. This fact opens up a new challenge in face recognition topic, where traditional state-of-the-art approaches lack essential information, hidden behind masks, that would allow them to achieve successful results. It is worth mentioning that this challenge does not born out of COVID-19, before it appeared, people usually wear clothing accessories, such as scarves or sunglasses, that results in the same effect; the face is partially hidden. Thus, new facial features must be extracted form faces to supply such lack of information, e.g. encoding micro-expressions extracted from the regions of interest of a face, as presented by Y.J. Liu et al. in [3], or extending the RGB information provided by conventional cameras through the use of wider spatial information obtained from hyperspectral sensors [4].

Hyperspectral imaging initially found its applications for remote sensing due to the richness of the spectral information that allows to apply techniques with greater visibility in the thorough analysis of land surfaces by means of the identification of visually similar materials and the estimation of physical parameters of many complex surfaces [4]. However, apart from the spectral information captured by hyperspectral sensors, it complements the data information collected by traditional sensors, such as RGB cameras. This kind of sensors have been improved in the last decade by reducing their cost and increasing in the imaging speed, which in turn has opened up the hyperspectral imaging to other applications, making it more popular than ever in recent decades [5] [6]. Hyperspectral imaging is widely used for a large variety of applications such as precision agriculture, forestry, city planning, urban surveillance and homeland security, chemistry, forensic examination and face recognition.

In recent years, masked face recognition has gained great importance due to COVID-19, which has been reflected in the number of articles published on this topic [7]. In this sense, there are many works on face mask detector to trigger an alarm when detecting a person who does not wear a mask or to analyse the degree to which health restrictions are enforced. On this basis, deep learning models has been used to automate the process of face mask detection. G.J. Chowdary et al. [8] have employed transfer learning of InceptionNet through augmented techniques to increase the diversity of the training data, as well as increase the performance of the proposed model. M. Loey et al. [9] have developed a hybrid deep transfer learning model that consists of two components for the detection of face mask; a component for feature extraction using ResNet50 [10] and a second component to classify face mask using decision trees and ensemble algorithm.

YOLO-based algorithms have also been used for face mask detection purposes, in which YOLOv3 is considered a major breakthrough in terms of the trade-off between detection precision and speed. S. Singh et al. [11] propose an efficient real-time deep learning-based technique to detect masked faces by using YOLOv3 architecture that has been trained by a small custom dataset, in which authors have provided the necessary labels and annotations. T. Q. Vinh and N. T. N. Anh [12] present an algorithm composed by a Haar cascade classifier that detects the faces in a picture and whose output feeds the YOLOv3 algorithm that determines whether a person wears a mask. To do so, the YOLOv3 has been previously trained with the MAFA dataset. Other works go one small step further and not only detect whether a person wears a face mask. P. Wu et al. [13] propose a YOLO-based framework to monitor whether people wear mask in a right mode, where the feature extractor, feature fusion operation and post-processing techniques are all specifically designed. Whilst, X. Su et al. [14] propose an efficient YOLOv3 algorithm, using EfficientNet as the backbone feature extraction network and reducing the number of network parameters, for mask detection and classify them into qualified masks (N95 and disposable medical masks) and unqualified masks (cotton, sponge, scarves, . . .).

The research efforts in masked face recognition have been increased since the COVID-19 pandemic by extending previous works related to face recognition or occluded face recognition methods. One of the approach adopted to face this challenge consists in restoring the part hidden by the mask and then use a face recognition alternative. In this sense, N. U. Din et al. [15] break the problem into two stages; firstly a binary segmentation of the mask region is performed and then the mask region is replaced with face textures retaining the global coherency of face structure. To do so, authors use a GAN-based network with a discriminators that learns the global structure of the face and other discriminator that comes in to focus learning on the deep missing region. Unfortunately, this kind of solutions result in failures cases when the map module is unable to produce a reasonable segmentation map of the mask object, i.e. the mask object are very different than those in the dataset. This kind of approaches follow the same strategy than older works in which restoration process from a gallery takes place [16] [17].

Other approaches only employ the visible part of the masked faces, i.e. these works extract the facial features from the upper part of the face or apply a filter to remove the mask area. W. Hariri [18] extracts deep features from the unmasked face regions through the last convolutional layer of three pre-trained deep CNN (VGG-16, AlexNet and ResNet50). Then a bag-of-features paradigm is applied to quantize the obtained features and, thus, a slight representation is obtained to, finally, feed a Multilayer Perceptron, that performs the classification process. F. Boutros et al. [19] propose an Embedding Unmasking Model (EUM) operated on top of existing face recognition models, such as ResNet50 [10] or MobileFaceNet [20]. These models do not require any modification or extra training. To do so, authors propose a loss function to guide the EUM during the training phase, minimizing and maximizing the distance between genuine and impostor pairs, respectively.

The lack of masked faces in well-known datasets has been supplied by extending them with fake versions that contain masks, i.e. synthetic masked faces are generated from existing faces. Moreover, some proposals also enrich the datasets through data augmentation to make variations in the images, such as cropping, flipping or rotation. Thus, A. Anwar and A. Raychowdhury [21] combine the VGG2 dataset [22] with augmented masked faces and train the model following the original pipeline described in FaceNet [23], this approach is also able to determine a masked face on the basis of the extracted features.

Despite the fact that hyperspectral imaging has not played a major role in face recognition because of other techniques have been very successful, there are several works with a variety of techniques that address this problem. M. Uzair et al. [24] use an algorithm based on spatiospectral covariance for band fusion to merge hyperspectral images into one, and propose the Partial Least Squares (PLS) regression algorithm to achieve face recognition and classification. In addition, authors perform band selection experiments to find the most discriminative bands in the visible and near infrared response spectrum. This band selection is followed by S. Bhattacharya et al. [25], they propose a face-specific band selection framework to identify the optimal band set that results in satisfactory face recognition performance. In the same line, Q. Chen et al. [26] emphasize on designing an efficient band selection method to reduce the spectral information without loss of recognition accuracy. V. Sharma et al. [27] propose hyperspectral CNN for image classification and band selection, where each band of the hyperspectral image is treated as a separate image. The architecture of the CNN is composed by 6 layers: 3 convolutional layers followed by 2 fully connected layers which are then connected to C-way softmax layer.

Pan et al. [28] study the reflectance of skin tissues for face recognition by analyzing near infrared spectral bands ($0.7\mu m$ - $1.0\mu m$), which vary from persons, thus these bands are employed for human recognition. In this sense, the problem of luminance affecting face recognition is decreased by manually selecting five facial regions of interest: hair, forehead, right and left cheeks and lips. However, the strong aspect of this work is that it can be used to recognize faces in the presence of changes in facial pose and expression. They also fuse the spatial information of the hyperspectral image, where each pixel in the fusion image is selected from a specific band in the same position, thus this method transforms a 3D hyperspectral image cube into a 2D image. In contrast, W. Di et al. [29] apply three techniques to analyze the efficiency over a different set of bands, from the whole bands to a single band, or, using a subset of bands. Thus, the three techniques comprise whole band $(2D)^2$ PCA, single band $(2D)^2$ PCA with decision level fusion, and band subset fusion-based $(2D)^2$ PCA, in which the latter two methods follow a simple efficient decision level fusion strategy. Authors conclude that the set of bands from $0.53\mu m$ to $0.59\mu m$ provides the most significant feature information since such bands correspond to the activity of human skin and absorption and reflection characteristics of carotene, hemoglobin and melanin.

In this work, we present a novel algorithm for facial features extraction (HyperFEA) from hyperspectral images, using a combination of computer vision techniques through histogram of oriented gradients (HOG) and hyperspectral transformations, to face recognition. In addition, a set of adaptive and parallel Support Vector Machines (AP-SVM) has been designed to classify unknown individuals. Thus, the extracted spatial information supply the lack of information that face masks or clothing accessories occlude. To the best of authors' knowledge, this is the first time computer vision techniques have been applied to hyperspectral images for that purpose. The main contributions of this work are listed as follows.

- An algorithm that uses computer vision techniques to extract facial regions of interests for face recognition of hyperspectral images.
- A significantly compression of the spatial information obtained from the facial regions of interest that maintains the uniqueness of the face hyperspectral signature.
- An adaptive and parallel Support Vector Machine tree to distinguish unknown individuals using only the regions of interests that are visible.
- An evaluation of the proposed model to analyze the recognition accuracy and an analysis of the similarity results.

The rest of the paper is organized as follows. In Section 2, the characteristics of the HyperFEA algorithm used in this work are introduced in detail to understand the extraction process of face features from hypersperctral images and the set of SVMs is also depicted. Section 3 describes the hyperspectral data sets, the performance assessment metrics used to evaluate the accuracy of the results provided by the proposed algorithm and shows the experimental results. Section 4 compares the results with the ones obtained for other state-of-the-art proposals. Finally, Section 5 draws up the main conclusions of this work.

2. Materials and Methods

2.1. Extracting Spectral Information

Feature extraction is a crucial stage in face recognition, whose main objective is to get a set of features that clearly represent a person. Typically, the set of features is composed by key facial attributes, such as eyes, mouth or nose, and/or the distance between them. From these features, a face encoding vector is generated and used to determine a similarity measure with other individuals in the recognition process. Unfortunately, the feature extraction process becomes more complicated when people wear face masks or clothing accessories, i.e. part of key facial features are hidden by them. This fact causes the existing face recognition methods to be adapted in order to extract representative facial features. On this basis, we propose the use of spectral information to complete the information lost. Thus, the proposed algorithm, HyperFEA extracts the relevant spatial information from hyperspectral faces. This algorithm has been developed for providing a good recognition accuracy by extracting facial regions of interests as well as providing a good compression performance of the spatial information of such regions. Additionally, the algorithm follows an unmixing-like strategy that selects the image pixels that are potentially more useful.

The process performed by the HyperFEA algorithm to hyperspectral images consists of four main stages, which are: (1) facial landmarks stage, which extracts the points where a face is located and optionally the face is rotated to horizontally align it; (2) extracting facial regions of interests (ROI) where the unused spatial information is removed; (3) a spectral transform; (4) a coding stage. The HyperFEA spectral transform selects the most different pixel and the average pixel of each facial ROI. Figure 1 graphically shows these four stages, as well as the data shared between them.

2.1.1. Algorithm notations

In the following, $\mathbf{HF} = \{\mathbf{HI}_i, i = 1, ..., ns\}$ is a sequence of ns hyperspectral frames, $\mathbf{HI}_i$, comprised by nb spectral bands that represents a hyperspectral image. Whereas $\mathbf{HI'}$ is the aligned hyperspectral image that is obtained from $\mathbf{HI}$ whose maximum deviation of the angle formed by the eyes is set by β (depicted in degrees). $L = [\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_\alpha]$ represents the facial landmark points, where α is the number of landmarks. Whilst $V = [\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_p]$ depicts the position of the points that delimits the facial regions

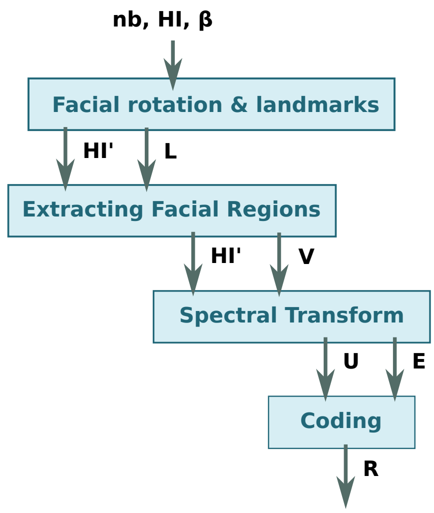


Figure 1. Diagram of the HyperFEA algorithm stages.

of interests extracted from L, where p is the number of regions; i.e. V_i corresponds to the points that set the limits of the i facial ROI. $\mathbf{HR}_i$ represents the i hyperspectral region whose location is stored in V_i. The average pixel, also called centroid, is represented by the symbol $\hat{\boldsymbol{\mu}}$, while $\hat{\boldsymbol{e}}$ represents the most different hyperspectral pixel extracted from such region. Each facial ROI can be represented as $\mathbf{R}_i = (\hat{e}, \hat{\mu})$.

Therefore, in addition to the hyperspectral image containing the face, the HyperFEA algorithm uses two main input parameters to extract the facial ROI and the spectral information.

- *Number of bands (nb).* This parameter denotes the number of bands that contains the hyperspectral images. It is provided to the algorithm in order to consider the whole spectral information.
- *Degrees threshold (β).* It determines a threshold of degrees up to the hyperspectral image must be rotated, i.e. all bands are rotated until all of them fulfil this requirement.

Algorithm 1 HyperFEA algorithm.

Inputs:
$\mathbf{HI} = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{nb}]$, nb, β
Outputs:
$\mathbf{R} = [\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_\alpha]; R_i = (\hat{\mu}_i, \hat{e}_i)$
Algorithm:
Face Alignment: $\mathbf{HI'} = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{nb}]$;
Facial Landmarks: $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_\alpha]$;
Location of Facial ROI: $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_p]$;
for i **in** V **do**
  Get hyperspectral region: $\mathbf{HR}_i \leftarrow getRegion(\mathbf{HI'}, \mathbf{V}_i)$;
  Centroid or average pixel: $\hat{\boldsymbol{\mu}}_i$;
  Centralization: $\boldsymbol{C} = \boldsymbol{HR}_i - \hat{\boldsymbol{\mu}}_i$;
  for j **in** HR_i **do**
    Brightness Calculation: $\boldsymbol{b}_j = \boldsymbol{c}'_j \cdot \boldsymbol{c}_j$;
  end for
  Maximum Brightness: $\hat{\boldsymbol{e}}_i = \boldsymbol{argmax}(\boldsymbol{b}_j)$;
  Save Spatial information: $\mathbf{R}_i \leftarrow (\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{e}}_i)$;
end for

2.1.2. HyperFEA algorithm

The *HyperFEA algorithm* is described in detail in Algorithm 1 for a hyperspectral image, **HI**. Firstly, the hyperspectral face is rotated ($\boldsymbol{HI'}$) and, then, the facial landmarks are extracted from it, $\boldsymbol{L}$, in lines 1 and 2, respectively. From the location of the face landmarks, the algorithm infers the facial ROI by obtaining the matrix $\mathbf{V}$ that contains the area of each region ($\mathbf{V}_i$), i.e. $\mathbf{V}_i$ represents the cloud of points that delimits the i^{th} facial ROI. The hyperspectral facial ROI is obtained from $\boldsymbol{HI'}$ by cropping the image according to the set of points stored in $\mathbf{V}_i$ (line 5 of Algorithm 1). Thus, the algorithm calculates the centroid and the brightest pixel of each hyperspectral facial ROI, $\mathbf{HF}_i$. The average pixel or centroid ($\hat{\boldsymbol{\mu}}_i$) is computed in line 6. Afterwards, the facial ROI is centralized by subtracting the average pixel to the original spectral information, i.e. each hyperspectral pixel that contains the facial ROI is subtracted by the average pixel (see line 7 of Algorithm 1). In addition, the most different pixel is extracted in line 11. In the remainder of this document, it is referred as brightness of a pixel. In this process, the dot product of each frame pixel within the centralized facial ROI with itself is first computed (lines 8 till 10 of Algorithm 1), whose the maximum value corresponds to the highest brightness ($\hat{\boldsymbol{e}}_i$). Finally, both spatial features, $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{e}}_i$, are stored in the matrix $\mathbf{R}$ in line 12, which contains the whole spatial information extracted from a hyperspectral face and it will be used to compare it with other matrix and determine its similarity.

2.1.3. Face alignment and extracting facial landmarks

Face alignment is an early stage of the modern face recognition pipeline that increases the recognition accuracy, which is optional in our proposal. Figure 2 shows the steps performed in the face alignment process. The first step is to detect the location of the eyes to extract the center of them and imaginatively draw a line between the two centres. Thus, the angle formed by the horizontal line with the previously one (ρ) gives the degree of inclination of the face. From this angle we can determine the rotation degrees by applying inverse trigonometry functions (*arc cosine* function), the result in degrees determines the angle to rotate the image, whenever its value is greater than β, i.e. the threshold degree parameter. Once the image is rotated, the algorithm checks that the face is horizontally aligned by a new iteration, it means the rotated image is the new input (orange arrow of Figure 2). In general, the constraint is fulfilled in the first iteration, which is an important issue when working with hyperspectral images, due to the computational cost required; the operations of the rotation stage are applied to all bands, i.e. the face alignment stage is repeated at least *nb* times. Thus, the hyperspectral face is horizontally aligned, $\boldsymbol{HI'}$ (line 1 of Algorithm 1).

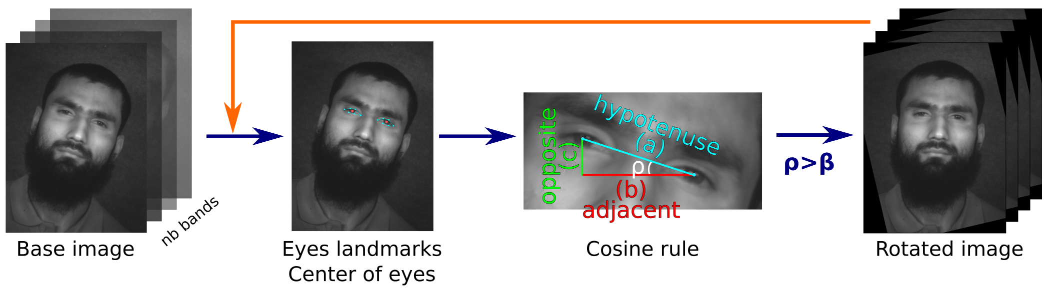


Figure 2. Face alignment process.

From a trigonometry point of view, the algorithm draws a rectangular triangle to calculate the angle between eyes (see multi-color triangle of Figure 2), whose sides corresponds as follows: line between the centers of the detected eyes (hypotenuse, blue line), horizontal line between the center of the detected eyes (adjacent, red line) and the line that close the triangle (opposite, green line). The length of the three lines are calculated with euclidean distance algorithm from the 2D points of the three edges of the triangle. Then, the cosine rule (see Equation 1) is performed to obtain the ρ angle.

$$cos(\rho) = (b^2 + c^2 - a^2)/2bc \tag{1}$$

The face alignment process can be carried out in two different modes depending on the speedup of the hyperspectral sensor to capture the images. The first mode considers each spectral band independent of the others, even if the captured face moves during the exposure of the picture taking, the alignment process corrects such small deviations. Meanwhile, the second mode takes as reference the first spectral band, which is aligned, and from the degree to which it is aligned, the rest of the bands are rotated. This mode is only suitable for those hyperspectral sensors whose time of exposure is small, i.e. it can be omitted. Thus, the facial alignment process is independent of the hyperspectral sensor speedup feature, but it is mandatory for the hyperspectral sensors with long time of exposure.

After the horizontal facial alignment phase, the next step is to extract the facial landmarks by providing a set of cloud that contains 2D points. These points represent and localize salient regions of the face, such as eyes, eyebrows or nose (see Figure 3(**b**)). The process is divided in two steps; firstly, the face must be localized in the image and then the key facial structures are detected. This method is widely used in RGB or gray scale images, so it is suitable for hyperspectral images, where both operations, face localization and facial landmark detection, are performed using the first spectral band. The algorithm considers that the rest of bands are aligned, either because the facial alignment process has been applied or because the hyperspectral sensor is able to capture all the spatial information in a shot. The result of this stage is a dictionary of lists, $\boldsymbol{L}$, where the location of α salient regions are stored (line 2 of Algorithm 1).

(**a**) (**b**) (**c**)

Figure 3. Extracting facial regions of interest from facial landmarks. (a) Base image. (b) Facial Landmarks. (c) Facial regions of interest.

2.1.4. Extracting facial regions of interests

The next stage is to obtain the location of facial ROI (line 3 of Algorithm 1). This process is similar that the one applied by F. Becattini et al. in [30], where 36 facial ROI are estimated. The solution proposed extracts 38 facial ROI from the location of the facial landmarks ($\boldsymbol{L}$), which were obtained in the previous stage (see Figure 3(**c**)). The facial regions are represented by a set of 2D points, which corresponds to the x-axis and y-axis position within the hyperspectral image ($\boldsymbol{HI'}$). The 2D points that delimit the facial ROI are stored in a list ($\boldsymbol{V}$) that will be used for the spectral transform process.

In turn, hyperspectral images have a problem caused by the position of the hyperspectral sensor, the light that falls on the face and the shape of the face itself, i.e. the quality spatial information depends on the reflection of the light over the surface of an object, a face in our case, and it also depends on the position of the hyperspectral sensor, it means the regions located in front of the sensor will have good quality of spatial information. Thus, the shape of the face is not flat, it seems a balloon in which the luminosity does not produce a good reflection in all parts. This fact is the reason that some facial ROI are divided in order to get useful spatial information instead of including them as

(**a**) (**b**) (**c**)

Figure 4. Quality of spectral information related to each facial ROI according to the visible parts. (a) Facial ROI no masked. (b) Facial ROI with mask/scarf. (c) Facial ROI with sunglasses and mask/scarf.

a single piece, e.g. the forehead is broken into ten subregions, where the lateral subregions do not provide good spatial information. Figure 4 shows the biometric areas of interest of a face that has been extracted in accordance with the visible face parts. In addition, the quality of spatial information of the collected facial ROI is also highlighted and classified as good (green), mid (yellow) and poor (red).

It is worth mentioning that clothing accessories and/or facial mask hide part of the face, so this work only takes the upper facial ROI for the experiments; from ten till nineteen regions. Figure 4(**b**) and 4(**c**) highlight the facial ROI considered, using the same colors that were used to mark the quality of the spatial information, whilst regions that are omitted are darkened.

2.1.5. Spectral transform

The spectral transform has a twofold objective: it contains enough information to discriminate against people and the spatial information is reduced. The second feature directly depends of the number of regions of interests and the number of spatial information, i.e. the number of bands.

Therefore, the HyperFEA transform sequentially selects the most different pixels ($\hat{\boldsymbol{e}}$) and the average pixel or centroid ($\hat{\boldsymbol{\mu}}$). For this purpose, a rectangular area of the hyperspectral image is extracted to apply a mask, which has been previously generated from the set of 2D points of the facial ROI that is being computed ($\mathbf{V}_i$). Therefore, the result is a hyperspectral image where the pixels out of the ROI have a value of 0, so spectral operations do not consider such pixels. Then, from each facial ROI a centroid is extracted ($\hat{\boldsymbol{\mu}}_i$) (lines 6 of Algorithm 1) by computing Equation 2, where $p^k_{x,y}$ represents the pixel located in the k^{th} band (spatial axis) at x,y position (x-axis and y-axis, respectively). The result is a hyperspectral pixel composed by nb bands whose values are the average value of each band.

$$\hat{\mu}_k = \frac{\sum_{x=0,y=0}^{x=W,y=H} \mathrm{p}^k_{x,y}}{NPixels_{valid}} \quad (2)$$

Afterwards, the hyperspectral facial ROI is centralized by subtracting the centroid, i.e. the subtraction operation is applied between each pixel in the ROI and the centroid (line 7 of Algorithm 1). In turn, the most different pixel of such ROI ($\hat{\boldsymbol{e}}_i$) is obtained by calculating the brightness of each pixel (lines 8 to 10 of Algorithm 1). The brightness is obtained by applying the $l^2 - norm$ vector normalization algorithm; it squares root of the sum of the squared element of the hyperspectral pixel. Then, the highest brightness pixel is selected (line 11 of Algorithm 1).

Figure 5 graphically shows the flow to extract the spatial information of the *mid_lower_forehead* facial ROI from a hyperspectral face. The algorithm extracts a rectangle that contains the hyperspectral pixels of the facial ROI to then apply a mask to crop the region, considering the pixels that are inside the region. Thus, the algorithm obtains the number of valid pixels and extracts the average value of each spectral band to build the centroid ($\hat{\mu}_{mid\_lower\_forehead}$) and extracts the brightest hyperspectral pixel ($\hat{e}_{mid\_lower\_forehead}$). It is worth mentioning that the spectral information extracted of each facial ROI can be performed in parallel.

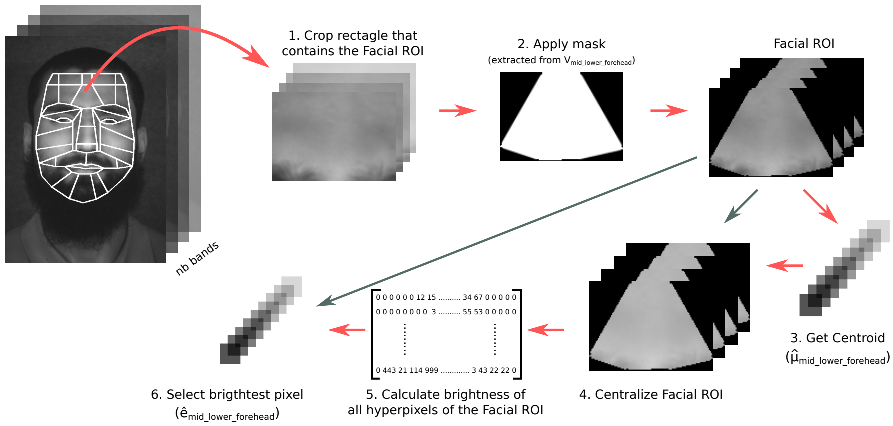


Figure 5. Flow of spatial information extraction process.

2.2. *A tree based on adaptive and parallel SVMs for face recognition*

Cascade of Support Vector Machines (C-SVM) has been introduced as an extension to classic SVM devoted to accelerate inference time by using an horizontal scaling strategy. The concept relies on the division of the problem into smaller problems, where each layer of SVMs is considered as a filter. This way, it is straightforward to get partial solutions leading towards the global optimum [31]. On this basis, the proposed solution leverages the advantages of C-SVM by proposing an adaptive and parallel layered-solution (AP-SVM). AP-SVM includes layers that may contain two or more SVMs, which are trained at run-time with the output of the previous layer. The output of one layer denotes the elements of the dataset the next layer of SVMs must be trained with. Since the size of the training dataset is getting smaller as the pipeline advances, overall latency of the process does not soar.

Figure 6 shows the AP-SVM tree structure for face recognition purposes using hyperspectral images. Classification time is sped-up due to the parallelization, following the C-SVM approach. For example, in layer one there is one SVM per ROI and per SVM kernel used on this work. Besides the independent and concurrent processing of each ROI, the size of the problem is smaller which leads to reduced latency.

Two kernels have been customized to obtain the closeness degree between the individual to be identified and the well-known persons in the dataset. AP-SVM uses the cosine similarity (i.e. computes the angle between two vectors) and euclidean distance (i.e. calculates the distance between two points). The cosine similarity is used to model the affinity in the reflectance realm whilst euclidean distance helps to model the spatial differences between concerning the morphology of the face.

2.2.1. Layer 1: Centroid Classification

The main goal of the first layer is to obtain a set of subjects that are close to the unknown individual. To do this, the problem is divided into as many SVMs as the number of facial ROIs used (horizontal division). In addition, the split is duplicated, in accordance with the number of kernels used (euclidean and cosine). The SVMs of this layer are previously trained and do not change during the classification process. The output of this layer is a list of potential candidates ($Pred_E$ and $Pred_C$ lists) composed by

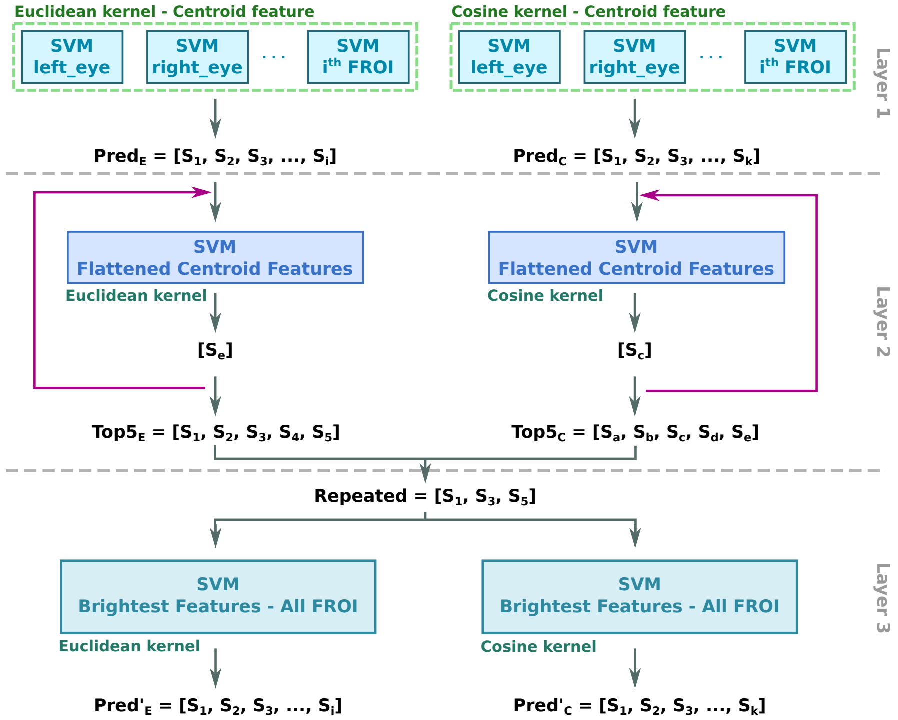


Figure 6. AP-SVM tree for face recognition.

the subjects whom spatial signature for a particular region is the closest to the individual that is being classified.

2.2.2. Layer 2: Flattened Centroid Classification

Once the first list of candidates is extracted by layer 1, it is necessary to measure the distance of the complete spatial signature between all the candidates and the unknown individual. For each ROI, the centroid and brightest pixels are selected and flattened in a single spatial features vector. Then, two SVM (one for each aforementioned kernel) are trained during run-time with the flattened spatial information vector in parallel. Afterwards, classification of the individual takes place. This process depends on the output of the previous layer, so the SVMs are trained every time an outsider is classified (adaptive feature). As a result, the two SVMs predict two possible candidates set; the ones whose euclidean distance and cosine similarity are the shortest, respectively ($[S_e]$ and $[S_c]$). Whenever the size of the $[S_e]$ and $[S_c]$ lists are not equal, they must be extended. Finally, the second layer of the AP-SVM extracts the top five candidates of both SVMs ($Top5_E$ and $Top5_C$ lists).

This layer is also considered as decision layer. The unknown individual can be considered identified anytime the candidate obtained by both SVMs is the same after the first iteration (i.e. $S\_1 = S\_a$). This means that the candidate has the shortest cosine and euclidean distance. The second rule for direct candidate selection establishes that if the first two candidates output by the euclidean SVM are equal (i.e. $S_1 = S_2$) that must be the identification of the unknown individual.

2.2.3. Layer 3: Brightest Classification

If the unknown individual could not be identified so far, the last layer proposes a unique candidate based on the analysis of the brightest features. Therefore, two SVMs are trained with the $Top5_E$ and $Top5_C$ candidate lists using the brightest features solely. In this case, the brightest features are considered in the same pool for the euclidean and cosine SVMs because of the brightest can be repeated or located in different facial ROI.

3. Experimental Results

In this section, the hyperspectral data used for evaluating the recognition accuracy of the proposed model is introduced. The hyperspectral dataset used in experiments has been provided by The University of Western Australia through a database that consists of 164 facial hyperspectral images of 79 different subjects (UWA-HSFD), where 75 individuals are males and 4 females; of those 75 males, 13 wear glasses, while in the case of females, only one wears glasses. The face database was sensed by CRI's Varispec LCTF, equipped with a photon focus camera helping in adjusting exposure time, luminance adaption and CCD sensitivity. Image cubes were captured over 33 bands from $0.4\mu m$ to $0.72\mu m$ with a difference of $0.1\mu m$; each band is stored in separate files. Figure 7 shows an example of a subject's face cube with the 33 bands [24]. The hyperspectral images have been organized in four sessions, i.e. the repeated faces have been taken on different days. Unfortunately, this dataset presents an important challenge; some subjects did not keep their head still during the process of capturing the image cube, so there are variations in the dataset.

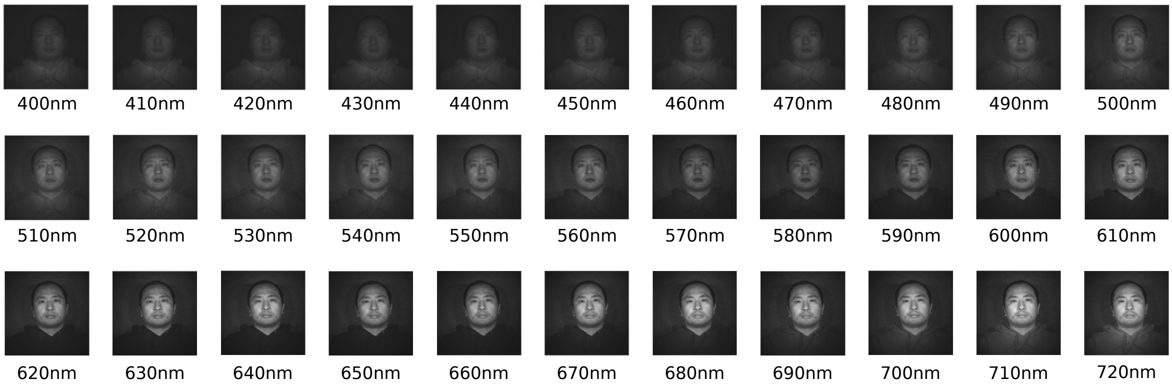


Figure 7. UWA-HSFD: Example of a subject's face cube.

To extract the spectral signatures of the images using the proposal presented in this work, we have set the maximum error of the face horizontal alignment to one degree, which has been applied to all spectral bands. This configuration requires high computational costs; the extraction of a spectral signature has taken an average of 248 seconds on a i7-10710U CPU with 32GB of RAM and SDD, i.e. the process to obtain all spectral signatures has taken 9 hours and 47 minutes. It is worth mentioning that this stage can be optimized in performance terms; the horizontal alignment of each band can be carried out in parallel as well as each facial ROI could be independently processed.

3.1. Partial results of the AP-SVM tree

For the shake of clarity, this section describes through a case study the intermediate results of the AP-SVM, using the 19 facial ROI highlighted in Figure 4(**b**). The SVMs of the first layer are trained with the two first sessions, in which certain individuals appear in both sessions, thus the spectral information is doubled. The individual to recognize is the labeled as 1 by using the spectral signature of the third session. Figure 8 displays the confusion matrix obtained after the classification performed by the SVMs of the first layer over the unknown individual. In this example, there are fourteen candidates; the subject 1 is the one that the distance of the spectral information is smallest. The SVMs of this stage are trained in parallel and, hence, it requires high computational resources to optimize the time performance of the AP-SVM. Thus, this process takes up to 2 seconds.

Although, the subject 1 contains more spatial features, we cannot discard other candidates that contains similarities. Thus, the second layer measures the entire spectral information providing two *Top5* lists. In this case, the top of these lists are not equal neither the first and second item of the *Top5* related to the Euclidean distance. Figure 9(**a**) and 9(**b**) display the confusion matrices using the euclidean and cosine custom kernels, respectively, which have been applied to the whole spectral signature. To obtain the list of candidates whose spectral signature is close to that of the unknown

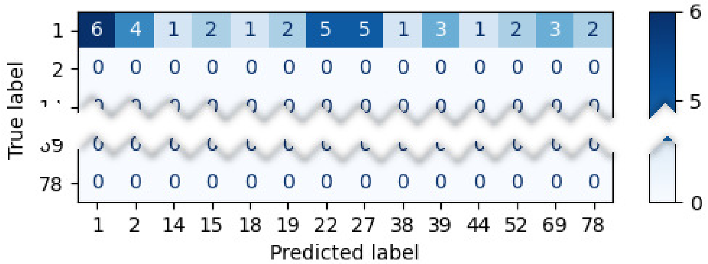
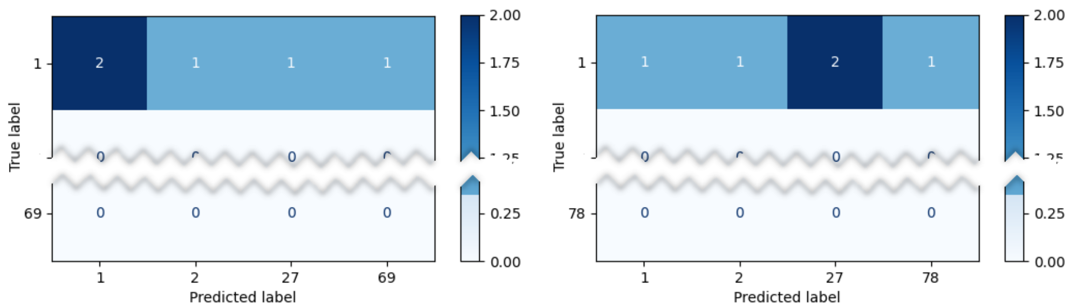


Figure 8. Example of confusion matrix obtained after the classification of layer 1.

individual, the closest one is not considered in the next iteration. Thus, the list of candidates was obtained in the following order $[2, 1, 1, 69, 27]$ and $[1, 2, 78, 27, 27]$ for the euclidean and cosine kernel, respectively.

The results draw differences between the two custom kernels, i.e. the candidates of the first step differs, the euclidean kernel stats that the unknown individual is the subject labeled as 2, whilst cosine kernel asserts that it is the subject 1, so the first rule to find out who is the unknown individual is not fulfilled. The second rule compares that the first and second candidates of the euclidean distance are the same, but this is not the case either.



(**a**) (**b**)

Figure 9. Example of separated confusion matrix obtained after the classification of layer 2. (a) Euclidean kernel. (b) Cosine kernel.

Afterwards, the repeated candidates are the input of the third layer, which calculates the brightest cosine and euclidean distance. Thus, the subjects $[1, 2, 27]$ are considered to identify the individual. Figure 10 shows the confusion matrix after measure the cosine and euclidean distances of the brightest pixels. Therefore, the AP-SVM determines that the unknown individual is the subject 1, because of it is the one that contains more similarities than the others, i.e. it has more coincidences in the brightest distance feature.

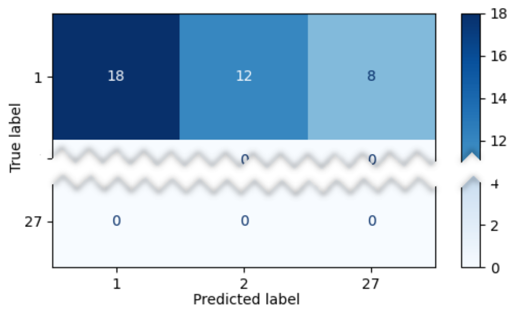


Figure 10. Example of confusion matrix obtained after the classification of layer 3.

3.2. Facial recognition accuracy

The achieved facial recognition accuracy directly depends on the visible facial ROI and the classification done by the first layer of the AP-SVM and the repeated items obtained from the second layer, i.e. the cosine and euclidean *Top5*. Table 1 lists the *Top5* and *Top3* in percentage for the three following scenarios: when there are no objects that hide the face (100%), when an object, such a scarf or a mask, occludes the lower part of the face (50%) and when the forehead is the only visible part (25%). On top of that, the aforementioned scenarios reduce the spectral information that can be extracted (see Figure 4) and, hence, the accuracy recognition achieved. In contrast, the compression ratio is greater and the computational costs are reduced. It is worth mentioning that the maximum recognition accuracy is as high as the ones obtained by the two first layers.

Table 1. $Top5$ and $Top3$ results obtained from the second layer of the C-SVM (depicted in percentage).

Top	All face (100% ROI visible)	Upper part (50% ROI visible)	Forehead (25% ROI visible)
Top5 Euclidean ($Top5_E$)	93%	93%	66%
Top5 Cosine ($Top5_C$)	80%	80%	60%
Top3 Euclidean ($Top3_E$)	80%	73%	60%
Top3 Cosine ($Top3_C$)	80%	73%	60%

The first scenario does not introduce objects that occlude any facial ROI, so there are 36 visible regions to extract the spatial information (see Figure 4(**a**)). In total 72 features are obtained, corresponding to the centroids and the brightest pixels of each of the visible areas. This fact reduces the spectral information roughly 99.9933% that is used to classify the unknown individuals. Figure 11 shows the confusion matrix in this scenario, in which the recognition accuracy achieved is 93%, i.e. 14 out of 15 individuals are recognised by the AP-SVM.

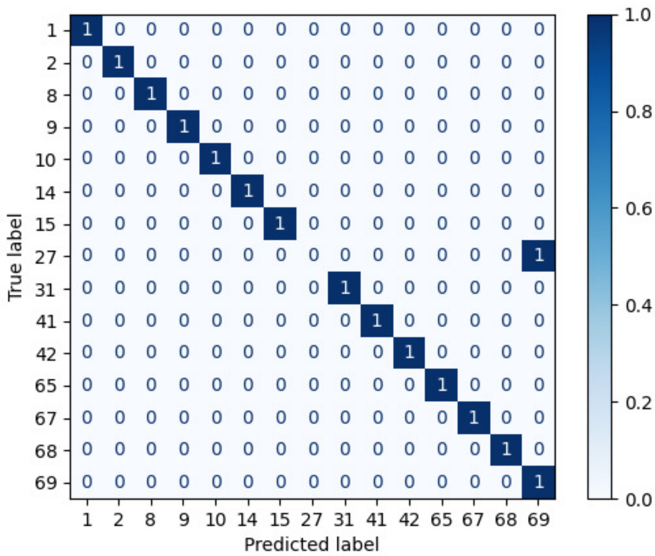


Figure 11. Confusion matrix with the 100% visible of the facial ROI.

The second scenario only makes visible the facial ROI located in the upper part of the face (see Figure 4(**b**)). Thus, 19 facial ROI are used to extract the hyperspectral signature composed by 38 features, that reduces the information roughly 99.9963%. The confusion matrix of the aforementioned scenario is shown in Figure 12; 13 of 15 individuals are recognized (86.67%). Due to the lack of a dataset that contains hyperspectral information with clothing accessories, such as scarves or sunglasses, we have only selected the facial ROI that are visible in the corresponding scenarios.

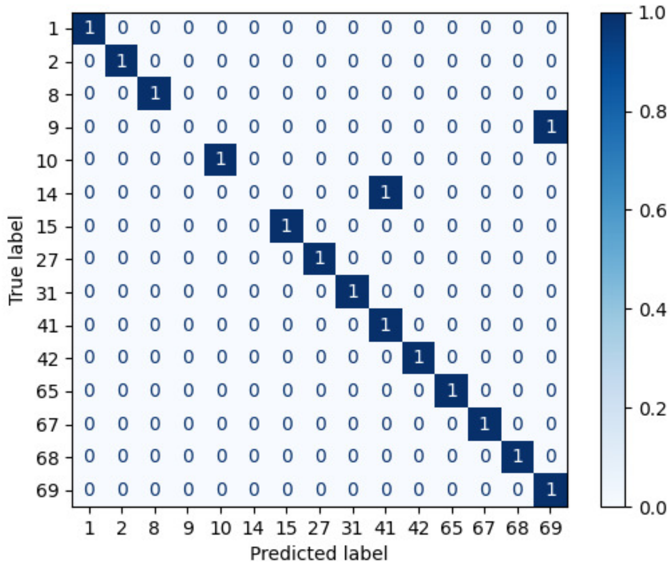


Figure 12. Confusion matrix with the 50% visible of the facial ROI.

The last scenario imposes most visibility restrictions in which 10 facial ROI are visible, which match with the regions of the forehead (see Figure 4(**c**)). This lack of spectral information results in lower facial recognition accuracy; only 6 of 15 individuals are recognized (40%) as is shown in Figure 13. In contrast, the spectral information stored is reduced up to 99.9981%.

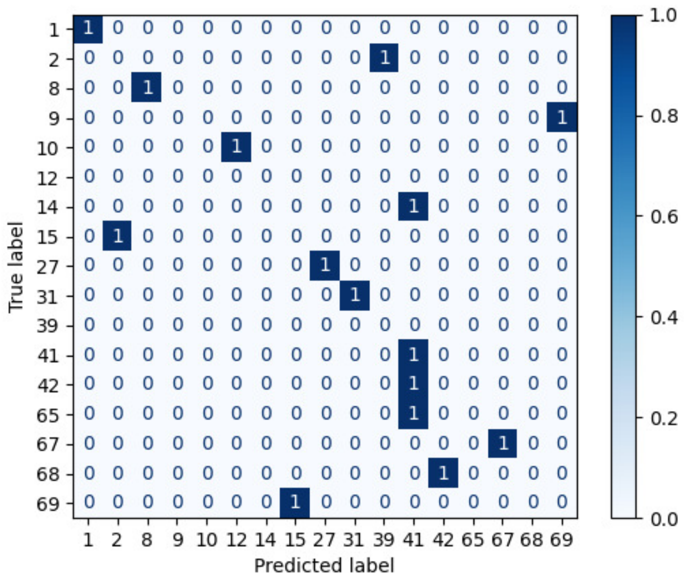


Figure 13. Confusion matrix with the 25% visible of the facial ROI.

3.3. Performance evaluation metrics for the AP-SVM tree

The performance of classification results has been exhibited through four key metrics such as precision, recall, f1-score and accuracy, whose calculation formulas are expressed in Equation 3, 4, 5 and 6, where *TP, TN, FP* and *FN* denoted the true positive, true negative, false positive and false negative, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

This evaluation criteria on face classification has been applied in the three aforementioned scenarios: all face (100% ROI), upper face (50% ROI) and forehead (25% ROI). Figure 14 graphically shows the results of the evaluation metrics used in the different layers of the AP-SVM tree and on the overall AP-SVM. It is worth mentioning that *recall* and *accuracy* has the same values, it means the model is somehow balanced, i.e. the AP-SVM is able to correctly classify positive unknown individuals as well as to correctly classify negative unknown individuals.

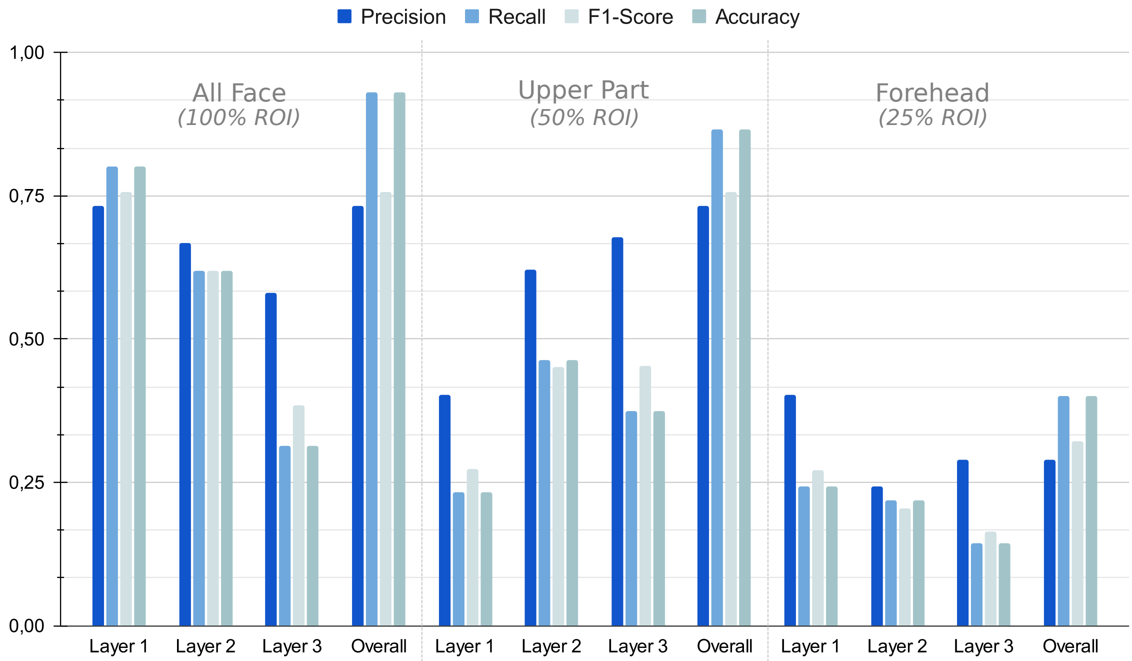


Figure 14. Performance metrics for the AP-SVM tree with different visible parts of the face.

4. Discussion

The trade-off between compression ratio and recognition accuracy has been compared with five state-of-the-art proposals that use hyperspectral images. Table 2 shows the recognition accuracy obtained by the different studies on hyperspectral face recognition as well as the compression ratio achieved by them. Z. Pan et al. [28] get a compression ratio up to 99.9% because they manually select five key regions for the frontal faces corresponding to the forehead, left cheek, right cheek, hair and lips by achieving roughly 75% correct coincidences. Unfortunately, only two of these regions are visible when a person wears a mask or a scarf, so the accuracy recognition results will be worse. In the same sense, W. Di et al. [29] manually locate the eyes position, from them the face is extracted, whose size is 162×150. Then, the extracted hyperspectral cube is normalized and scaled to 54×50 with the aim to save the computational costs. The rest of the works [24,27,32] also crop and resize the face area and perform an image fusion to transform the hyperspectral cube into a flatten image, which is obtained by band fusion. V. Sharma et al. [27] keep the whole spectrum range but the size of the face is resized to 263×263 pixels, so the compression ratio is worse than the other studies but the recognition accuracy is high. Meanwhile, authors of works [24,29,32] perform a band selection as well as resize the face area to reduce the spatial information obtaining high percentages of hits in facial recognition.

Table 2. Comparison of hyperspectral face recognition accuracy and compression ratio.

	Dataset			Extracted		Accuracy	Compression Ratio
	Dataset/size	Bands	Spectrum	Features	Bands		
[28]	200	31	$0.7\mu m - 1.0\mu m$	5	31	75%	99.9995%
[29]	25 (PolyU)	33	$0.4\mu m - 0.72\mu m$	2700 (54 × 50)	24	78%	99.2509%
[27]	CMU	65	$0.4\mu m - 0.72\mu m$	69169 (263 × 263)	65	86.1%	93.2264%
[24]	UWA	33	$0.4\mu m - 0.72\mu m$	900 (30 × 30)	4	98%	99.9895%
	PolyU	24	$0.45\mu m - 0.68\mu m$	1748 (46 × 38)	5	95.2%	99.8610%
[32]	PolyU	33	$0.4\mu m - 0.72\mu m$	4096 (64 × 64)	4	95%	99.8106%
	CMU	65	$0.4\mu m - 0.72\mu m$	4096 (64 × 64)	37	98%	99.7776%
Ours	UWA	33	$0.4\mu m - 0.72\mu m$	70	33	93%	99.9933%

Therefore, the state-of-the-art proposals applies one of the two following methods with/without band selection: band fusion through calculating the average of each band or select a key pixel of a facial ROI. Nevertheless, HyperFEA algorithm automatically delimits the face area and its facial ROI without a band selection to extract the average pixel (centroid) and a key pixel (brightest pixel) of each region. On this basis, the face area is not resized to save computational costs.

Figure 15 shows the compression factor normalized with respect to our proposal. For the shake of clarity the proposal presented by V. Sharma et al. [27] has not be considered in this study due to the compression ratio is the worst. The results draw up a good balance between recognition accuracy and compression ratio considering that our proposal is one of those that reduces the hyperspectral information the most.

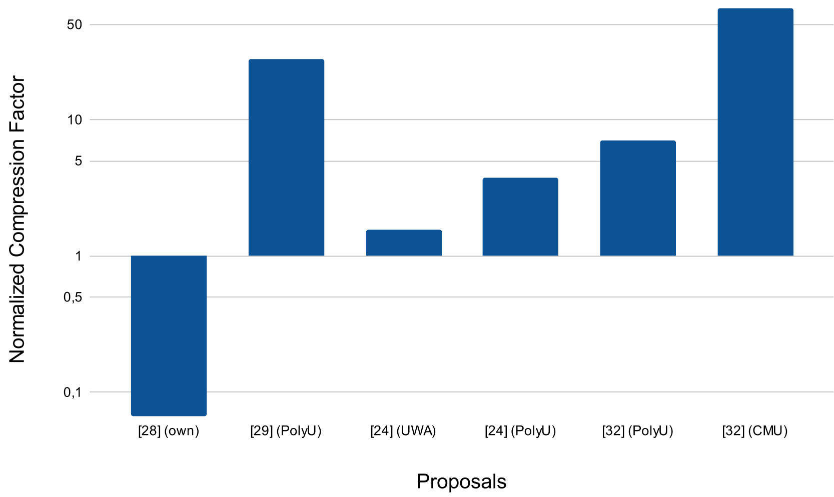


Figure 15. Normalized compression factor with respect to our proposal.

The recognition rate obtained by our proposal has also been compared with other state-of-the-art proposals, whose objective is to recognize unknown individuals that wears mask or other clothing accessories that occlude part of the face. Table 3 lists the recognition rates achieved by some state-of-the-art proposals in the three aforementioned scenarios. Moreover, Table 3 also depicts the method used by each proposal. The results reveal that our proposal is close to the state of the art when the face is hidden by a mask. Meanwhile, when the forehead is only the face region that is visible, the recognition rate of our solution doubles the one proposed by A.C. Tsai et al. [33].

Table 3. Comparison of recognition rates with state-of-the-art proposals.

Proposal	Method	All face (100% ROI visible)	Upper part (50% ROI visible)	Forehead (25% ROI visible)
[18]	CNN+BoF	NA	91.3%	NA
[34]	CNN+SVM	NA	87%	NA
[33]	CNN	97.36%	95.38%	23.07%
[35]	PCA+SVM	90.14%	67.82%	NA
[36]	SRC	91.92%	72.37%	NA
[37]	CNN	97.21%	68.69%	NA
[37]	DCGAN+CNN	97.36%	75.21%	NA
Ours	AP-SVM (SVM)	93%	86.67%	40%

5. Conclusion

This work has focused on extracting features from hyperspectral images by using computer vision techniques and classify unknown individuals through an AP-SVM tree. The process starts with the detection of the facial ROI. Then, the *centroid* or average pixel and the brightest pixel are extracted from each facial ROI. In contrast to the works in the literature, which use hyperspectral images to face recognition, HyperFEA algorithm automates the extraction of spectral characteristics of a face. In addition, it facilitates the extraction of such features in parallel by using parallel-programming architectures, such as GPUs or FPGAs, where each facial subregion will be processed by a different kernel.

Experimental results draw up an interesting trade off achieved by the HyperFEA algorithm in which the compression ratio is up to 99.99% and the recognition accuracy is 93%, when all facial ROI are visible, but the results are more interesting when several regions of the face are hidden by objects, such as masks, sunglasses or scarves, where the recognition accuracy achieve is up to 86.67%. This result could be improved by using hybrid hyperspectral and non-hyperspectral techniques, where they are self-complementary.

Author Contributions: Conceptualization, Julian Caba; Funding acquisition, Juan López López; Investigation, Julian Caba, Jesús Barba, Fernando Rincón and Soledad Escolar; Methodology, Julian Caba, Jesús Barba and Fernando Rincón; Software, Julian Caba; Validation, José Antonio de la Torre; Writing – original draft, Julian Caba and José Antºonio de la Torre; Writing – review & editing, Julian Caba, Jesús Barba, Fernando Rincón, Soledad Escolar and Juan López López.

Acknowledgments: This research was funded by H2020 European Union program under grant agreement No. 857159 (SHAPES project) and by MCIN/AEI/10.13039/501100011033 under grant TALENT-BELIEF (PID2020-116417RB-C44).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Klare, B.F.; Burge, M.J.; Klontz, J.C.; Vorder Bruegge, R.W.; Jain, A.K. Face Recognition Performance: Role of Demographic Information. *IEEE Transactions on Information Forensics and Security* **2012**, *7*, 1789–1801.
2. Zhou, Y.; Ni, H.; Ren, F.; Kang, X. Face and Gender Recognition System Based on Convolutional Neural networks. 2019 IEEE International Conference on Mechatronics and Automation (ICMA), 2019, pp. 1091–1095.
3. Liu, Y.J.; Zhang, J.K.; Yan, W.J.; Wang, S.J.; Zhao, G.; Fu, X. A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition. *IEEE Transactions on Affective Computing* **2016**, *7*, 299–310. doi:10.1109/TAFFC.2015.2485205.
4. Arellano, P.; Tansey, K.; Balzter, H.; Boyd, D.S. Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images. *Environmental Pollution* **2015**, *205*, 225 – 239. doi:10.1016/j.envpol.2015.05.041.

5. Calin, M.A.; Calin, A.C.; Nicolae, D.N. Application of airborne and spaceborne hyperspectral imaging techniques for atmospheric research: past, present, and future. *Applied Spectroscopy Reviews* **2020**, pp. 1–35.
6. Qureshi, R.; Uzair, M.; Zahra, A. Current Advances in Hyperspectral Face Recognition, 2020. doi:10.36227/techrxiv.12136425.v1.
7. Alzu'bi, A.; Albalas, F.; AL-Hadhrami, T.; Younis, L.B.; Bashayreh, A. Masked Face Recognition Using Deep Learning: A Review. *Electronics* **2021**, *10*. doi:10.3390/electronics10212666.
8. Jignesh Chowdary, G.; Punn, N.S.; Sonbhadra, S.K.; Agarwal, S. Face Mask Detection Using Transfer Learning of InceptionV3. *Lecture Notes in Computer Science* **2020**, p. 81–90. doi:10.1007/978-3-030-66665-1_6.
9. Loey, M.; Manogaran, G.; Taha, M.H.N.; Khalifa, N.E.M. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement* **2021**, *167*, 108288. doi:10.1016/j.measurement.2020.108288.
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.
11. Singh, S.; Ahuja, U.; Kumar, M.; Kumar, K.; Sachdeva, M. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimedia Tools and Applications* **2021**, *80*, 19753–19768. doi:10.1007/s11042-021-10711-8.
12. Vinh, T.Q.; Anh, N.T.N. Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier. 2020 International Conference on Advanced Computing and Applications (ACOMP), 2020, pp. 146–149. doi:10.1109/ACOMP50827.2020.00029.
13. Wu, P.; Li, H.; Zeng, N.; Li, F. FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image and Vision Computing* **2022**, *117*, 104341. doi:10.1016/j.imavis.2021.104341.
14. Su, X.; Gao, M.; Ren, J.; Li, Y.; Dong, M.; Liu, X. Face mask detection and classification via deep transfer learning. *Multimedia Tools and Applications* **2021**. doi:10.1007/s11042-021-11772-5.
15. Ud Din, N.; Javed, K.; Bae, S.; Yi, J. A Novel GAN-Based Network for Unmasking of Masked Face. *IEEE Access* **2020**, *8*, 44276–44287. doi:10.1109/ACCESS.2020.2977386.
16. Drira, H.; Ben Amor, B.; Srivastava, A.; Daoudi, M.; Slama, R. 3D Face Recognition under Expressions, Occlusions, and Pose Variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2013**, *35*, 2270–2283. doi:10.1109/TPAMI.2013.48.
17. Gawali, S.; Deshmukh, R. 3D Face Recognition Using Geodesic Facial Curves to Handle Expression, Occlusion and Pose Variations. *International Journal of Computer Science and IT* **2014**, *5*, 4284–4287.
18. Hariri, W. Efficient masked face recognition method during the COVID-19 pandemic. *Signal, Image and Video Processing* **2021**. doi:10.1007/s11760-021-02050-w.
19. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Unmasking Face Embeddings by Self-restrained Triplet Loss for Accurate Masked Face Recognition. *CoRR* **2021**, *abs/2103.01716*, [2103.01716].
20. Chen, S.; Liu, Y.; Gao, X.; Han, Z. MobileFaceNets: Efficient CNNs for Accurate Real-time Face Verification on Mobile Devices. 2018, pp. 1–10.
21. Anwar, A.; Raychowdhury, A. Masked Face Recognition for Secure Authentication. *CoRR* **2020**, *abs/2008.11104*, [2008.11104].
22. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A Dataset for Recognising Faces across Pose and Age. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); IEEE Computer Society: Los Alamitos, CA, USA, 2018; pp. 67–74. doi:10.1109/FG.2018.00020.
23. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823. doi:10.1109/CVPR.2015.7298682.
24. Uzair, M.; Mahmood, A.; Mian, A. Hyperspectral Face Recognition With Spatiospectral Information Fusion and PLS Regression. *IEEE Transactions on Image Processing* **2015**, *24*, 1127–1137.
25. Bhattacharya, S.; Das, S.; Routray, A. Graph Manifold Clustering based Band Selection for Hyperspectral Face Recognition. 2018 26th European Signal Processing Conference (EUSIPCO), 2018, pp. 1990–1994.
26. Chen, Q.; Sun, J.; Palade, V.; Shi, X.; Liu, L. Hierarchical Clustering Based Band Selection Algorithm for Hyperspectral Face Recognition. *IEEE Access* **2019**, *7*, 24333–24342.
27. Sharma, V.; Diba, A.; Tuytelaars, T.; Gool, L.V. Hyperspectral CNN for image classification & band selection, with application to face recognition. Technical report, KU Leuven, ESAT-PSI, 2016.

28. Zhihong Pan.; Healey, G.; Prasad, M.; Tromberg, B. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2003**, *25*, 1552–1560.
29. Di, W.; Zhang, L.; Zhang, D.; Pan, Q. Studies on Hyperspectral Face Recognition in Visible Spectrum With Feature Band Selection. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* **2010**, *40*, 1354–1361.
30. Becattini, F.; Song, X.; Baecchi, C.; Fang, S.T.; Ferrari, C.; Nie, L.; Del Bimbo, A. PLM-IPE: A Pixel-Landmark Mutual Enhanced Framework for Implicit Preference Estimation. ACM Multimedia Asia; Association for Computing Machinery: New York, NY, USA, 2021; MMAsia '21. doi:10.1145/3469877.3490621.
31. Graf, H.; Cosatto, E.; Bottou, L.; Dourdanovic, I.; Vapnik, V. Parallel Support Vector Machines: The Cascade SVM. Advances in Neural Information Processing Systems; Saul, L.; Weiss, Y.; Bottou, L., Eds. MIT Press, 2005, Vol. 17, pp. 1–8.
32. Chen, G.; Li, C.; Sun, W. Hyperspectral face recognition via feature extraction and CRC-based classifier. *IET Image Processing* **2017**, *11*, 266–272.
33. Tsai, A.C.; Ou, Y.Y.; Wu, W.C.; Wang, J.F. Integrated Single Shot Multi-Box Detector and Efficient Pre-Trained Deep Convolutional Neural Network for Partially Occluded Face Recognition System. *IEEE Access* **2021**, *9*, 164148–164158. doi:10.1109/ACCESS.2021.3133446.
34. Almabdy, S.; Elrefaei, L. Deep Convolutional Neural Network-Based Approaches for Face Recognition. *Applied Sciences* **2019**, *9*.
35. LUO, Y.; ming WU, C.; ZHANG, Y. Facial expression feature extraction using hybrid PCA and LBP. *The Journal of China Universities of Posts and Telecommunications* **2013**, *20*, 120–124. doi:https://doi.org/10.1016/S1005-8885(13)60038-2.
36. ZHU Ming-Han, LI Shu-Tao, Y.H. An Occluded Facial Expression Recognition Method Based on Sparse Representation. *Public Relations & Artificial Intelligence* **2014**, *27*, 708.
37. Yeh, R.A.; Chen, C.; Lim, T.Y.; Hasegawa-Johnson, M.; Do, M.N. Semantic Image Inpainting with Perceptual and Contextual Losses. *CoRR* **2016**.