Article

A Cognitively-Motivated Framework for Partial Face Recognition in Unconstrained Scenarios

João C. Monteiro * and Jaime S. Cardoso

INESC TEC and Faculdade de Engenharia, Universidade do Porto, Campus da FEUP,
Rua Dr. Roberto Frias, n 378, 4200-465 Porto, Portugal; E-Mail: jaime.cardoso@inescporto.pt

* Author to whom correspondence should be addressed; E-Mail: jcmonteiro89@gmail.com;
Tel.: +351-22-209-4299.

Academic Editor: Vittorio M.N. Passaro

Abstract: Humans perform and rely on face recognition routinely and effortlessly throughout their daily lives. Multiple works in recent years have sought to replicate this process in a robust and automatic way. However, it is known that the performance of face recognition algorithms is severely compromised in non-ideal image acquisition scenarios. In an attempt to deal with conditions, such as occlusion and heterogeneous illumination, we propose a new approach motivated by the global precedent hypothesis of the human brain's cognitive mechanisms of perception. An automatic modeling of SIFT keypoint descriptors using a Gaussian mixture model (GMM)-based universal background model method is proposed. A decision is, then, made in an innovative hierarchical sense, with holistic information gaining precedence over a more detailed local analysis. The algorithm was tested on the ORL, ARand Extended Yale B Face databases and presented state-of-the-art performance for a variety of experimental setups.

Keywords: biometrics; face recognition; partial data; universal background model; Gaussian mixture models

---

1. Introduction

Personal identification plays an important role in almost everyone's daily activities. Knowledge-based and token-based automatic personal identification are the most used techniques to tackle this problem. Token-based approaches take advantage of a personal item to distinguish between individuals, whereas

knowledge-based approaches are based on something the user knows to which, theoretically, nobody else has access [**?**].

Both of these approaches present obvious disadvantages: tokens may be lost, stolen or forgotten, while passwords can easily be forgotten by a valid user or guessed by an unauthorized one [**?**]. In fact, one can summarize the problem of these approaches by pointing that any piece of material or knowledge can be fraudulently acquired.

Biometrics can be seen as a return to a more natural way of identification. By attempting identification based on physiological and behavioral traits, we are testing someone by who (s)he is, instead of relying on something (s)he owns or knows. Such an approach seems likely to be the way forward [**?**].

Over the past few years, the issue of face recognition has been in the spotlight of many research works in pattern recognition, due to its wide array of real-world applications. The face is a natural, easily acquirable and usable trait with a high degree of uniqueness, representing one of the main sources of information during human interaction [**?**]. These marked advantages, however, fall short when images of limited quality, acquired under unconstrained environments, are presented to the system.

It has been noted that the performance of face recognition algorithms is severely compromised when dealing with non-ideal scenarios, such as non-uniform illumination, pose variations, occlusions, expression changes and radical appearance changes [**?**]. Whereas technological improvements in image capturing and transmitting equipment managed to attenuate most noise factors, partial face occlusions still pose a genuine challenge to automated face recognition [**?**].

Facial occlusions may occur due to a multiplicity of deliberate or unintentional reasons. Whereas accessories, such as sunglasses and scarves, and facial hair represent quite common sources of occlusion in daily life, they can also be explored by bank robbers and shop thieves in an attempt to avoid recognition. Furthermore, the use of some accessories might be enforced in restricted environments (such as medical masks in hospitals and protection helmets in construction areas) or by religious or cultural constraints [**?**]. The fact that humans perform and rely on face recognition routinely and effortlessly throughout their daily lives leads to an increased interest in replicating this process in an automated way, even when such limitations are known to frequently occur [**?**].

Even though there is no consensus in the cognitive science field as to how the human brain recognizes faces, either based on their individual local features or, more holistically, on the basis of their overall shape [**?**], several works have shown that both levels of information play a non-negligible role in human face perception [**?** **?**]. Whereas holistic representations provide a global summary of the spatial arrangement of contours and textures in an image, local features provide a more detailed regional description of the parts that compose it [**?**].

In the present work, we propose a robust alternative to face recognition under partial occlusions and variable illumination. An innovative hierarchical decision framework, incorporating both holistic and local descriptions, is proposed. The global precedent hypothesis for human perception [**?**] is the basis of this new decision strategy. Such a hypothesis claims that face recognition is performed by the human brain in a global-to-local flow, with holistic information gaining precedence over a more detailed local analysis. By following this rationale, we aim to replicate the cognitive process of face recognition by the human brain in an automated way. We evaluate the proposed algorithm on three widely-studied

databases—the ORL, ARand the Extended Yale B databases—characterized by a variety of occlusions, small pose variations, facial expressions and illumination conditions.

The rest of the paper is organized as follows: Section **??** summarizes the most recent trends of research in the field of unconstrained face recognition. Section **??** outlines and motivates the proposed algorithm. Section **??** summarizes the most relevant experimental results obtained, as well as a detailed analysis and comparison with the state-of-the-art, and finally, in Section **??**, we present the conclusions of the present work, as well as some suggestions for future improvements.

2. Related Work

Face recognition has been a widely-studied research topic in the last few decades. Some traditional approaches, like eigenfaces [**?**], Fisherfaces [**?**] and active appearance models [**?**], have become highly popular and laid out the foundations for a variety of commercial off-the-shelf systems. However, all of the previously mentioned techniques stumble upon the limitations presented in the last section: when non-ideal conditions are present during the acquisition step, recognition performance is severely compromised. The need to improve the state-of-the-art in face recognition to encompass a set of more realistic applications has, thus, been catalyzing research in the area to a set of new directions.

The study of invariant features to diminish the effects of occlusion, illumination and other nefarious sources of noise represents one of the most significant focuses of recent research. Liao *et al.* [**?**] tried to overcome the need for face alignment that characterizes most holistic approaches employing a multi-keypoint descriptor representation. In this work, any type of face image, either holistic or partial, can be probed for recognition, regardless of the global content. Nallammal and Radha [**?**] propose a non-negative matrix factorization (NMF) variation to explore the potential of the eyes and the bottom face regions for recognition when the probe images present a high degree of occlusion. An alternative approach is followed by Oh *et al.* [**?**], where random horizontal and vertical patches of face images are used as templates for cancelable identity verification. Such a technique intentionally distorts biometric information, in a repeatable, but non-reversible manner, to better deal with the compromising of biometric templates. Karande and Talbar [**?**] address the problem of face recognition with large rotation angles and variable illumination conditions through the use of edge information for independent component analysis (ICA). The work by Geng and Jiang [**?**] explores some of the known limitations of the widely-studied SIFT approach, for a specialized application in face image description. An alternative keypoint detection and a partial descriptor are both proposed in an attempt to adapt the traditional algorithm to non-rigid and smooth objects. Cho *et al.* [**?**] propose a two-step approach to face recognition, with principal component analysis (PCA) used at a coarser level and Gabor filtering at a finer level. This finer analysis in only carried out if the coarser recognition results do not present a high degree of reliability. Recently, Facebook's Deep Face Project claimed 97.25% accuracy, where humans achieve an accuracy of 97.53%. Their approach, published by Taigman *et al.* [**?**], was based on deep neural networks, allowing the effective use of highly complex statistical models trained for large volumes of data.

Recently, approaches based on sparse representation classification (SRC) have shown impressive performance in unconstrained face recognition and became one of the hot research topics in the area.

The first reported use of SRC for face recognition, by Wright *et al.* [**?**], approached the problem of partial occlusion by representing face images as a linear combination of the whole face gallery and a vector of residuals at the pixel level. Classification was then achieved by l_1 minimization of the vector of residuals for each possible identity. Zhou *et al.* [**?**] further improved this methodology by enforcing spatial coherence of occluded pixels through the use of Markov random fields (MRF). The spatial continuity of occlusions in face images was also explored by Qian *et al.* [**?**]. Their methodology takes advantage of the low-rank error images that are originated in occluded images by traditional SRC methodologies, to perform effective and robust recognition when such conditions are observed. Besides spatial coherence of occlusions, some other topics have been explored in recent works as possible improvements to the original SRC proposal. Wang *et al.* [**?**] propose an Adaptive SRC (ASRC) approach capable of selecting the most discriminative samples for each representation, using joint information from both sparsity (l_1 minimization), as well as correlation (l_2 minimization). Shen *et al.* [**?**] propose a variation of SRC for implementation in Android and iOS mobile devices. Their proposal optimizes the computation of residual values with significant gain in computational efficiency and no considerable losses in recognition accuracy. Jian *et al.* [**?**] also center their attention on the computational speed limitations of the SRC approach. Their proposal, based on the orthogonal matching pursuit (OMP) algorithm, achieves fast and robust face recognition, even though the best results are only achieved through a preliminary occlusion detection block. Even though considerable work has been performed in the area, the main drawback regarding the SRC approach is still posed by the need for an extensive and diverse library of well-aligned face examples.

Another focus of research in recent years concerns the use of prior knowledge regarding occlusions in face images. Zhang *et al.* [**?**] proposed an estimation of the probability distribution of occlusions in feature space using the Kullback–Leibler divergence (KLD). In a mixed approach regarding both previous detection of occlusions and SRC face recognition, Li *et al.* [**?**] present a two-step SRC approach, where SRC is both used to first discriminate occluded pixels from unoccluded regions and then to perform face recognition. The use of downsampled images allows a significant improvement in processing speed. Min *et al.* [**?**], on the other hand, perform occlusion detection using MRF to promote the spatial coherence of the detected occlusion regions. The recognition step is then carried out solely on the non-occluded regions. Even though the *a priori* detection of occlusions may significantly improve the accuracy of local face recognition, the introduction of a new block in the recognition system may bring about a new set of problems. An increase in the computational cost of the process, as well as the creation of a new source for errors that may condition the recognition process from its earlier steps can be counted among such challenges.

In the present work, we propose a robust approach to face recognition when non-ideal conditions, such as partial occlusions and severe illumination variations, affect the acquisition environment of the system. In an attempt to tackle most of the limitations presented in the works outlined in the last paragraphs, we designed a new hierarchical recognition framework. This innovative approach allows a considerable reduction in the computational cost of the whole recognition process, while also allowing an intuitive integration of multiple region-based details. The proposed algorithm is able to achieve accurate face recognition, even when a limited set of images with small variations is used for model training.

2.1. Algorithm Overview

The proposed algorithm is schematically represented in Figure **??**. Figure 1a,b depicts the enrollment process in the proposed approach. During enrollment, a new individual's biometric data are inserted into a previously existent database of individuals. In the present work, a hierarchical ensemble of M partial face models is trained for each enrolled subject. The M individual-specific models are built by maximum *a posteriori* (MAP) adaptation of the corresponding set of M universal background models (UBM) using individual-specific data. The UBM is a representation of the distribution that a biometric trait presents in the universe of all individuals. MAP adaptation works as a specialization of the UBM based on each individual's biometric data. The idea of MAP adaptation of the UBM was first proposed by Reynolds [**?**], for speaker verification, and will be further motivated in the following sections.
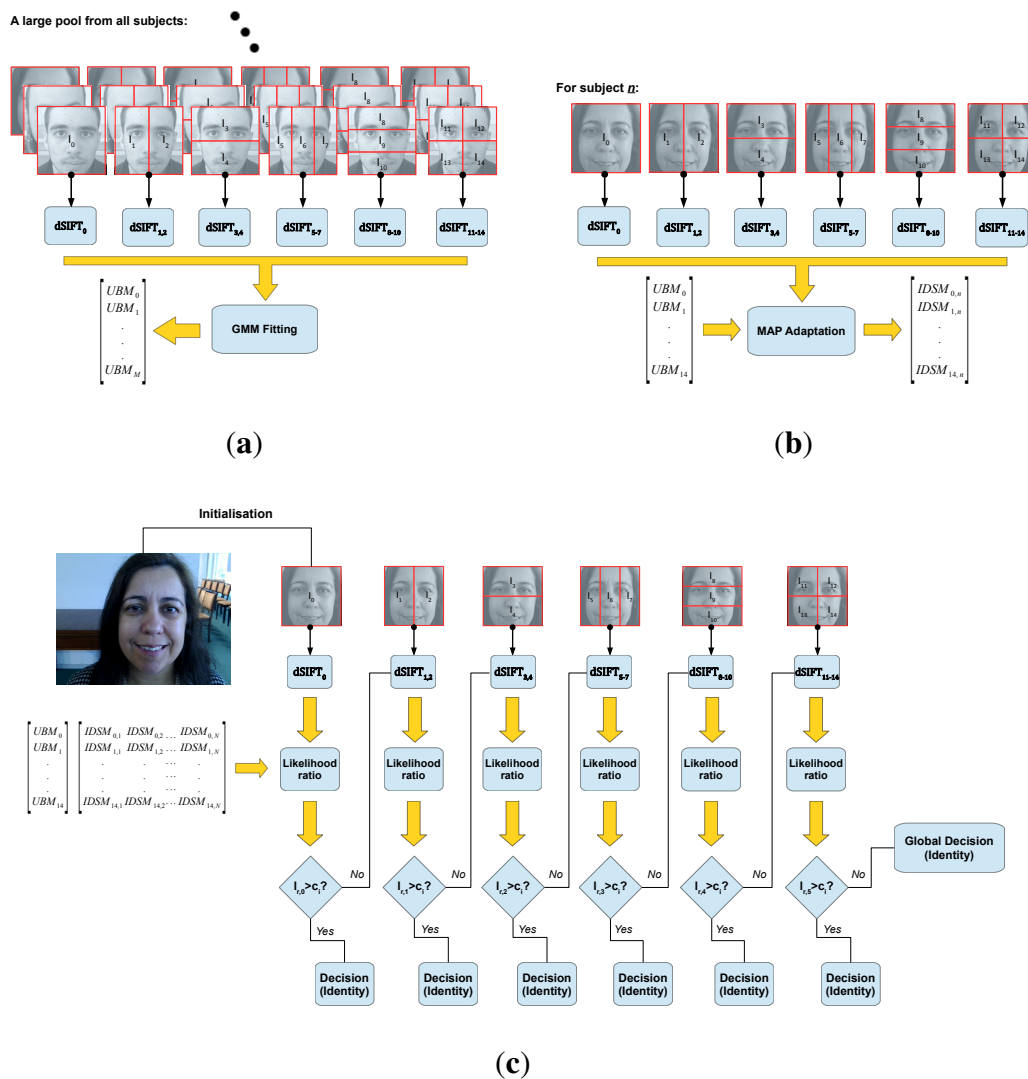

(a)   **(b)**

(c)

Figure 1. Schematic representation of the proposed algorithm and its main blocks: (**a**) training of the universal background models using data from multiple individuals; (**b**) maximum *a posteriori* (MAP) adaptation of the universal background models (UBM) to generate individual specific models; and (**c**) testing with new data from unknown sources.

3. Proposed Methodology

The database is probed during the recognition process to assess either the validity of an identity claim (verification) or the k most probable identities (identification) given an unknown sample of biometric data. In the present work, we propose an innovative approach to the recognition process based on the global precedence hypothesis of face perception by the human brain. Recognition is performed hierarchically, as depicted in Figure 1c, with global models taking precedence over more detailed ones. Partial models are hierarchically organized into levels. Each level is composed by a set of non-superimposing subregions, I_l, of equal size (Levels 2–3 and 4–5 were hierarchically ordered in an arbitrary order, even though their composing regions are of equal size. Previous knowledge of expected types of occlusion could be explored when specifying this order.). Subregions at the same level sum to the full-face image, I_0. During recognition, a test image from an unknown source follows the hierarchical flow depicted in Figure 1c, until a decision can be made with a significant degree of certainty. The significance of a decision carried out at a single level is defined through the analysis of the likelihood-ratio values obtained for each possible identity claim. Decisions are made independently for each subregion at the same level, and only the most significant one is kept. The following sections will further detail the process of model training and recognition score computation for a single generic subregion, while also motivating the proposed UBM framework.

3.1. Universal Background Model

The universal background modeling strategy was initially proposed in the field of voice biometrics [**?**]. Its framework can be easily understood if the problem of biometric verification is interpreted as a basic hypothesis test. Given a biometric sample Y and a claimed ID, S, we define:

$$H_0\text{: } Y \text{ belongs to } S$$
$$H_1\text{: } Y \text{ does not belong to } S$$

as the null and alternative hypothesis, respectively. The optimal decision is made by a likelihood-ratio test:

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} \geq \theta & \text{accept } H_0 \\ < \theta & \text{accept } H_1 \end{cases} \tag{1}$$

where θ is the decision threshold for accepting or rejecting H_0 and $p(Y|H_i)$ is the likelihood of observing Y knowing that H_i is true. The goal of a biometric verification system can, thus, be accomplished by the computation of the likelihood values $p(Y|H_0)$ and $p(Y|H_1)$ for a given sample. It is intuitive to note that H_0 will correspond to a model λ_{hyp} that characterizes the hypothesized individual, whereas H_1 will represent the alternative hypothesis, that is the model of all of the alternatives to the hypothesized individual, $\lambda_{\overline{hyp}}$.

The computation of $p(Y|H_i)$ depends on the specific strategy chosen for data modeling. If H_0 and H_1 correspond to a pair of generative models, trained on sets of genuine and impostor data, respectively, then the $p(Y|H_i)$ values can be roughly expressed as the projections of biometric sample Y onto each of these models. This formulation motivates the need for a model that successfully covers the space of

alternatives to the hypothesized identity. The most common designation in the literature for such a model is the universal background model or UBM [**?**]. Such a model must be trained on a rather large set of data, so as to faithfully cover a representative user space.

Even though the UBM approach was initially proposed for verification mode, we extrapolate its rationale for identification systems. Instead of performing a single one *vs.* one likelihood-ratio test and checking the validity of the condition presented in Equation (**??**), a one *vs.* all approach may alternatively be considered. Given an unknown sample, the most likely identity, Id_{max}, will correspond to the highest likelihood-ratio value, amongst all enrolled users:

$$Id_{max} = \underset{i}{\operatorname{argmax}} \left(\frac{p(Y|H_0^{(i)})}{p(Y|H_1)} \right) \tag{2}$$

where $H_0^{(i)}$ represents the model describing user i.

Defining an objective way of quantifying $p(Y|H_0)$ and $p(Y|H_1)$ becomes, thus, the true challenge when following this approach. In the following sections, we analyze in detail the strategies chosen to model both λ_{hyp} and $\lambda_{\overline{hyp}}$.

3.2. Hypothesis Modeling

In the present work, we chose Gaussian mixture models (GMM) to model both the UBM, *i.e.*, $\lambda_{\overline{hyp}}$, and the individual-specific models (IDSM), *i.e.*, λ_{hyp}. Such models are capable of capturing the empirical probability density function of a given set of feature vectors, so as to faithfully model their intrinsic statistical properties [**?**]. The choice of GMM to model feature distributions in biometric data is extensively motivated in many works of related areas. From the most common interpretations, GMMs are seen as capable of representing broad "hidden" classes, reflective of the unique structural arrangements observed in the analyzed biometric traits [**?**]. Besides this assumption, Gaussian mixtures display both the robustness of parametric unimodal Gaussian density estimates, as well as the ability of non-parametric models to fit non-Gaussian data [**?**]. This duality, alongside the fact that GMM have the noteworthy strength of generating smooth parametric densities, confers such models a strong advantage as generative models of choice. For computational efficiency, GMM models are often trained using diagonal covariance matrices. This approximation is often found in the biometrics literature, with no significant accuracy loss associated [**?**].

All models are trained on densely-sampled sets of scale-invariant feature transform (dSIFT) keypoint descriptors, extracted from previously normalized facial sub-regions. Illumination normalization is performed using the Weber-face approach [**?**]. Even though traditional SIFT descriptors present invariance to a set of common undesirable factors (image scaling, translation, rotation), thus conferring them a strong appeal in unconstrained biometrics, the fact that they fail to adapt to heterogeneous illumination conditions severely hinders their practical use in real-life applications. The normalization step is, therefore, of the utmost importance.

Dense SIFT is a variation of the traditional SIFT methodology [**?**], where keypoint descriptors are extracted in a roughly equivalent manner to running SIFT on a dense grid of locations at a fixed scale and orientation [**?**]. Dense sampling mitigates the potential errors introduced by the detection of an

unreliable set of interest points in its sparse counterpart [**?**]. In the present work, we train GMMs on the set of all densely-sampled keypoint descriptors from all individuals (UBM) and adapt individual models using data from specific subjects alone (IDSM), achieving a stable summary of the image content for every enrolled user.

Originally, SIFT descriptors were defined in 128 dimensions. However, we chose to perform a PCA, as suggested in [**?**], reducing the dimensionality to 32. Such a reduction allows not only a significant reduction in the computational complexity of the training phase, but also an improved distinctiveness and robustness to the extracted feature vectors, especially as far as image deformation is concerned [**?**]. We computed the principle components from the same set of keypoint descriptors used to train the UBM.

3.2.1. UBM Parameter Estimation

To train the universal background model, a large amount of "impostor" data, *i.e.*, a set composed of data from all the enrolled individuals, is used, so as to cover a wide range of possibilities in the individual search space [**?**]. The training process of the UBM is simply performed by fitting a k-mixture GMM to the set of PCA-reduced feature vectors extracted from all of the "impostors".

If we interpret the UBM as an "impostor" model, its "genuine" counterpart can be obtained by adaptation of the UBM's parameters, $\lambda_{\overline{hyp}}$, using individual specific data. For each enrolled user, n, an IDSM, defined by parameters λ_{hyp_n}, is therefore obtained. The adaptation process will be outlined in the following section.

3.2.2. MAP Adaptation

IDSMs are generated by the tuning of the UBM parameters, in a maximum *a posteriori* (MAP) sense, using individual-specific biometric data. This approach provides a tight coupling between the IDSM and the UBM, resulting in better performance and faster scoring than uncoupled methods, as well as a robust and precise parameter estimation, even when only a small amount of data is available [**?**]. The adaptation process consists of two main estimation steps. First, for each component of the UBM, a set of sufficient statistics is computed from a set of M individual-specific feature vectors, $X = \{\mathbf{x}_1, ..., \mathbf{x}_M\}$:

$$n_i = \sum_{m=1}^{M} p(i|\mathbf{x}_m) \tag{3}$$

$$E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{m=1}^{M} p(i|\mathbf{x}_m)\mathbf{x}_m \tag{4}$$

$$E_i(\mathbf{x}\mathbf{x}^t) = \frac{1}{n_i} \sum_{m=1}^{M} p(i|\mathbf{x}_m)\mathbf{x}_m\mathbf{x}_m^t \tag{5}$$

where $p(i|\mathbf{x}_m)$ represents the probabilistic alignment of $\mathbf{x}_m$ into each UBM component. Each UBM component is then adapted using the newly-computed sufficient statistics and considering diagonal covariance matrices. The update process can be formally expressed as:

$$\hat{w}_i = [\alpha_i n_i/M + (1 - \alpha_i)w_i]\,\xi \tag{6}$$

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{i}} = \alpha_i E_i(\mathbf{x}) + (1 - \alpha_i)\boldsymbol{\mu}_{\boldsymbol{i}} \tag{7}$$

$$\hat{\Sigma}_i = \alpha_i E_i(\mathbf{x}\mathbf{x}^t) + (1 - \alpha_i)(\boldsymbol{\sigma}_{\boldsymbol{i}}\boldsymbol{\sigma}_{\boldsymbol{i}}^t + \boldsymbol{\mu}_{\boldsymbol{i}}\boldsymbol{\mu}_{\boldsymbol{i}}^t) - \hat{\boldsymbol{\mu}}_{\boldsymbol{i}}\hat{\boldsymbol{\mu}}_{\boldsymbol{i}}^t \tag{8}$$

$$\boldsymbol{\sigma}_{\boldsymbol{i}} = \mathrm{diag}(\Sigma_i) \tag{9}$$

where $\{w_i, \boldsymbol{\mu}_{\boldsymbol{i}}, \boldsymbol{\sigma}_{\boldsymbol{i}}\}$ are the original UBM parameters and $\{\hat{w}_i, \hat{\boldsymbol{\mu}}_{\boldsymbol{i}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{i}}\}$ represent their adaptation to a specific speaker. To assure that $\sum_i w_i = 1$, a weighting parameter ξ is introduced. The α parameter is a data-dependent adaptation coefficient. Formally, it can be defined as:

$$\alpha_i = \frac{n_i}{r + n_i} \tag{10}$$

The relevance factor r weights the relative importance of the original values and the new sufficient statistics. In the present, work we set $r = 16$.

3.3. Hierarchical Decision

The whole training process is repeated at most M times for each of M facial subregions, defined *a priori*. In the present work, we used the $M = 14$ regions previously exemplified in Figure **??**.

Traditionally, the recognition phase with new data from an unknown source is a fairly simple process. The new test data, $X_{test} = \{\mathbf{x}_{t,1}, \ldots, \mathbf{x}_{t,N}\}$, where $\mathbf{x}_{t,i}$ is the i-th PCA-reduced SIFT vector extracted from a given subregion m of test subject t, is projected onto both the UBM and either the claimed IDSM (in verification mode) or all such models (in identification mode). The recognition score, $s_{t,m}$, is obtained as the average likelihood-ratio of all keypoint descriptors $\mathbf{x}_{t,i}$, $s_{t,m} = \frac{1}{N}\sum_{i=1}^{N} s_{t,m}^{(i)}$. The decision is then carried out by checking the condition presented in Equation (**??**), in the case of verification, or by detecting the maximum likelihood-ratio value for all enrolled IDs (Equation (**??**)), in the case of identification.

Such a decision step represents the second big advantage of the UBM approach. The ratio between the IDSM and the UBM probabilities of the observed data is a more robust decision criterion than relying solely on the IDSM probability. This results from the fact that some subjects are more prone to generate high likelihood values than others, *i.e.*, some people have a more "generic" look than others. The use of a likelihood ratio with a universal reference works as a normalization step, mapping the likelihood values in accordance with their global projection. Without such a step, finding a global optimal value for the decision threshold, θ, presented in Equation (**??**), would be a far more complex process.

In an attempt to integrate meaningful information from the M facial subregions and to deal better with partial or missing data situations originated by occlusions, an hierarchical recognition framework is proposed. The rationale behind the methodology described below can be easily understood if some studies of the human brain's cognitive mechanisms of perception are taken into consideration. One such work, proposed in 1977 by David Navon [**?**], describes the aforementioned mechanism as a hierarchical process, where holistic representations precede more detailed local features. If we interpret this perception mechanism in the scope of human face recognition, we can conclude that an attempt to classify the face at a global level is the starting point to the whole recognition process, whereas increasingly detailed descriptions are only taken subsequently if necessary. This conceptual description

of the human perception mechanism serves as the basis for the proposed hierarchical recognition algorithm, whose flowchart is presented in Figure 1c.

The main steps of the proposed hierarchical identification algorithm are as described below:

(1) Initialization: Starting with the full-face image from an unknown user, I_0, a densely-sampled grid of SIFT keypoint descriptors is extracted; the likelihood values, $l_{t,0}^{IDSM_t}$ and $l_{t,0}^{UBM_0}$, are then computed for the IDSM of every enrolled user $t \in \{1..T\}$ and the UBM of the tested region; the recognition scores for every possible identity, $\boldsymbol{s_{t,0}} = \{s_{1,0}, ..., s_{T,0}\}$, are then computed through Equation (**??**).

(2) Certainty index computation: The certainty index of a given region, c_m, measures how likely it is that the obtained vector of likelihood ratios, $\boldsymbol{s_{t,m}}$, corresponds with the ideal case of a single correct identity match. If no false positives corrupt the vector of likelihood ratios obtained for a single image, a significant difference will be observed between the highest value, $\boldsymbol{s_{t^*,m}}$, (true identity) and the average of all other values, $\frac{1}{T-1} \sum_{t=1,t\neq t^*}^{T} \boldsymbol{s_{t,m}}$, (average impostor). The certainty index can thus be interpreted as a degree of separability between these two quantities:

$$c_m = \boldsymbol{s_{t^*,m}} - \frac{1}{T-1} \sum_{t=1,t\neq t^*}^{T} \boldsymbol{s_{t,m}} \tag{11}$$

(3) Decision to go to next level: If the c_m value exceeds a previously optimized threshold, θ_l, the maximum likelihood-ratio decision is accepted. When $c_m < \theta_l$, however, the algorithm will consider that an analysis at a more detailed level is necessary to achieve a decision with a higher degree of confidence. At this point, the algorithm proceeds to the next level, working on subregions I_{1-2}, the second in the hierarchical chain depicted in Figure **??**. When one level is composed by multiple subregions, each one of them is treated independently, and only the maximum c_m value among them is considered for the decision criterion:

$$c_m = \begin{cases} c_0 & \text{if } l = 1 \\ \max\{c_1, c_2\} & \text{if } l = 2 \\ \max\{c_3, c_4\} & \text{if } l = 3 \\ \max\{c_5, c_6, c_7\} & \text{if } l = 4 \\ \max\{c_8, c_9, c_{10}\} & \text{if } l = 5 \\ \max\{c_{11}, c_{12}, c_{13}, c_{14}\} & \text{if } l = 6 \end{cases} \tag{12}$$

(4) Repeat: Steps 1 to 3 are hierarchically repeated for every level until $c_m > \theta_l$. If all L levels are considered and none is able to achieve a significant decision, the decision corresponding to the highest value of c_m, $\max\limits_{m}(c_m)$, amongst all L levels is considered.

As the proposed algorithm is capable of performing recognition without the need of processing all subregions, computational speed is significantly improved over simpler approaches that explore the fusion of all local recognition scores. Furthermore, the proposed algorithm is capable of automatically deciding if a more detailed exploration of local features is necessary or if the information obtained up to

a certain point is enough to make a decision. This autonomy alongside the intuitive notion of the global precedence hypothesis are the most notorious strengths of the proposed recognition algorithm.

An alternative to deciding non-classified images at the end of the hierarchical chain was also considered. In this new setting, if no level is capable of making a decision according to the aforementioned criteria, images are kept as "doubtful", and no decision is made, in a process similar to classification with a reject option [**?**]. This approach may be thought of as a viable alternative for real-life applications, where feedback to the user can be explored to adjust the environmental conditions in severely unconstrained scenarios.

4. Results and Discussion

4.1. Experimental Setups

The proposed algorithm was tested on the Database of Faces (formerly "The ORL Database of Faces"), the Extended Yale Face Database B and the AR face database. All tested databases are widely known for their diversity of pose, illumination and occlusion conditions, respectively. The next sections outline the main features of each database, as well as the experimental setup chosen for training and testing with each of them in the scope of the present work.

4.1.1. ORL

The Database of Faces (formerly "The ORL Database of Faces") [**?**] contains 400 images from 40 subjects, divided equitably with a total of 10 images per individual. Images were taken at different points in time with variable lighting, expression and pose conditions. For performance assessment, we use a single sample from each individual for training and the remaining nine images for testing. This process is repeated for each possible training image, and the performance is computed as the average of the 10 runs. An alternative approach is also explored by using multiple templates per subject instead. In this new setup, the first five samples per subject are used for training, while the remaining five samples are used for testing. Example images from two subjects may be observed in Figure **??**.
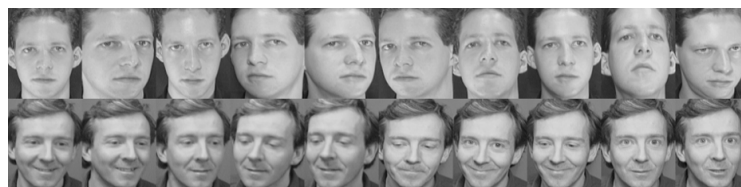

Figure 2. Example images from two subjects of the ORLdatabase.

4.1.2. Extended Yale B

The Extended Yale Face Database B [**?**] is composed of 2432 images corresponding to a total of 38 individuals. All images are frontal faces acquired under varying illumination conditions. The database is divided into five subsets, numbered 1 to 5, according to the ranges of angles between the light source

direction and the camera axis. An example of all images from a single subject may be observed in Figure **??**.
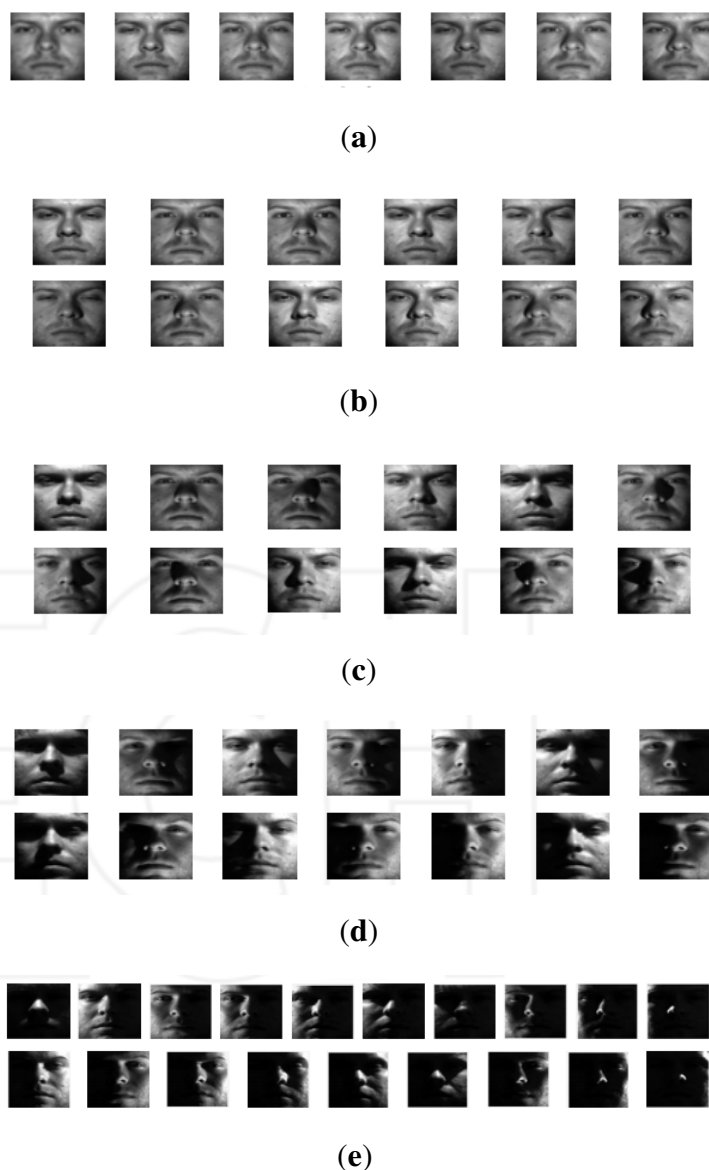

(a)

(b)

(c)

(d)

(e)

Figure 3. All images from a single subject enrolled in the Extended Yale B database. Images (**a**) to (**e**) correspond to Subsets 1 to 5, respectively.

All images from Subset 1 were used for model training, while all other subsets were tested independently, so as to better assess the robustness of the proposed algorithm to a variety of illumination conditions.

4.1.3. AR

The AR database [**?**] contains over 4000 frontal face images from 126 individuals, acquired under variable illumination, expression and occlusion. Occlusions can be divided into two main categories: sunglasses and scarf. An example of all images from a single individual is presented in Figure **??**.

Figure 4. Example images from one subject of the ARdatabase.

All unoccluded images from every individual are chosen to train both the local UBMs and IDSMs. The remaining scarf and sunglasses occluded images are tested separately, so as to better analyze the consistency of the proposed algorithm when exposed to variable types of occlusion.

4.1.4. Performance Analysis

Figures **??**–**??** depict the most relevant results obtained by the proposed algorithm for the ORL, Extended Yale B and AR databases, using the previously mentioned training and testing setups. Furthermore, Tables **??**–**??** present a comparison between the proposed work and the reported performance in some recent works, performed under similar experimental conditions. We chose to assess the rate of correctly identified individuals, by checking if the true identity is present among the N highest ranked identities. The N parameter is generally referred to as rank. This allows us to define the Rank-1 recognition rate, r_1, as the recognition rate at $N = 1$.
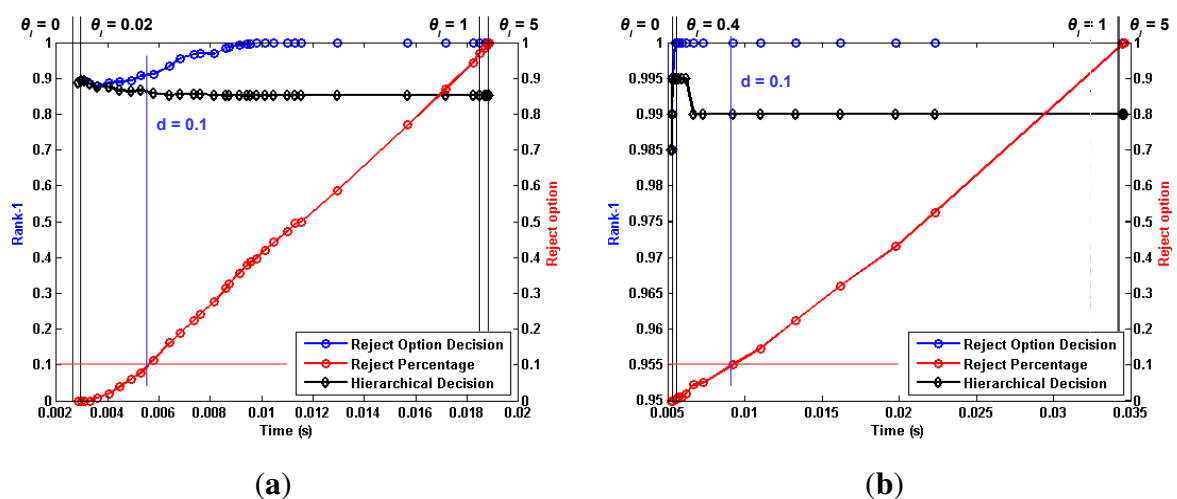

Figure 5. Main results obtained with the proposed methodology for the sunglasses and scarf images of the ORL face database. Each plotted point represent a specific value of parameter θ_l, ranging from $[0, \infty]$.
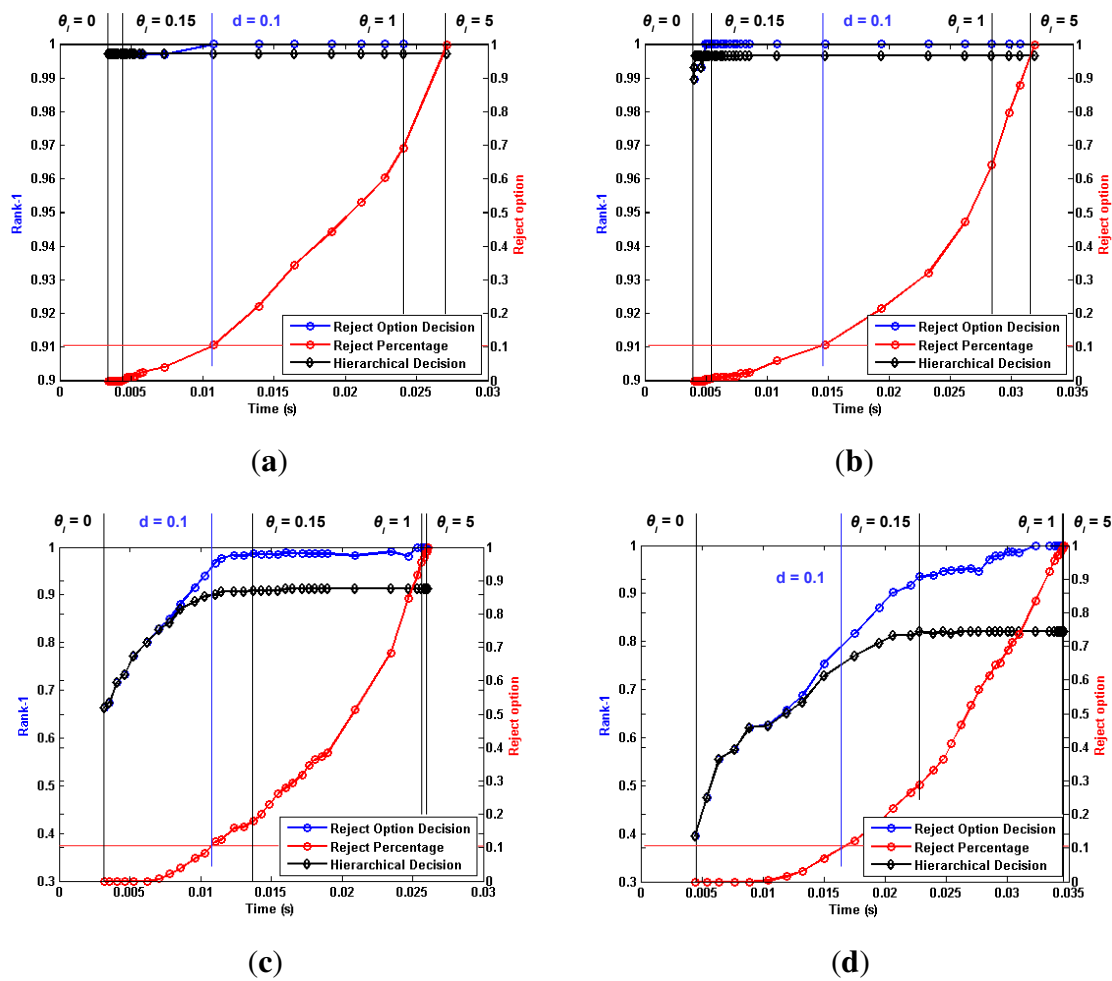
Figure 6. Main results obtained with the proposed methodology for Subsets 2–5 of the Extended Yale B Face database.
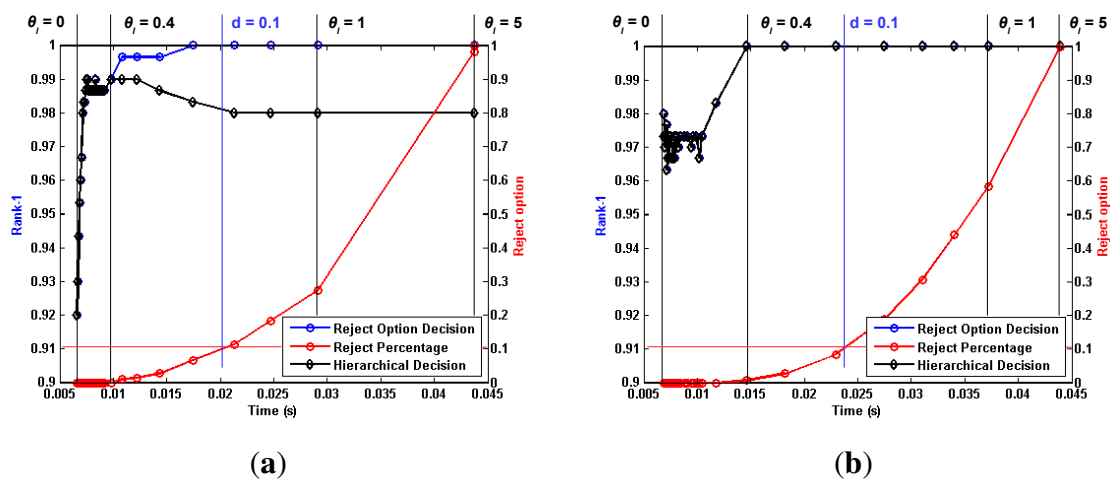

Figure 7. Main results obtained with the proposed methodology for the sunglasses and scarf images of the AR face database. Each plotted point represent a specific value of parameter θ_l, ranging from $[0, \infty]$.

Table 1. Comparison between the obtained r_1 values for the ORL database and some state-of-the-art algorithms.

Work	Single Template $\mu \pm \sigma$	Multiple Templates
Proposed$_{\theta=0.02|0.15}$	**89.45%** $\pm$ **1.24%**	**99.5%**
Proposed$_{\theta\to\infty}$	85.79% $\pm$ 1.85%	99.00%
Proposed$_{d<0.1}$	*93.60%* $\pm$ *1.85%*	*100.0%*
Geng and Jiang [**?**]	89.0%	**99.5%**
Zhang *et al.* [**?**]	–	85.0%
Xiao *et al.* [**?**]	–	97.50
Xu *et al.* [**?**]	–	97.10

Table 2. Comparison between the obtained r_1 values for the AR database and some state-of-the-art algorithms.

Work	AR Sunglasses	AR Scarf	AR Average
Proposed$_{\theta=0.4}$	**100.0%**	99.00%	**99.50%**
Proposed$_{\theta\to\infty}$	**100.0%**	98.00%	99.00%
Proposed$_{d<0.1}$	*100.0%*	*100.0%*	*100.0%*
Min *et al.* [**?**]	75.00%	92.08%	83.54%
Li *et al.* [**?**]	97.00%	52.00%	74.50%
Morelli Andrés *et al.* [**?**]	97.46%	99.15%	98.31%
Qian *et al.* [**?**]	99.00%	**100.0%**	**99.50%**

Table 3. Comparison between the obtained r_1 values for the Extended Yale B database and some state-of-the-art algorithms.

Work	Yale B SS_2	Yale B SS_3	Yale B SS_4	Yale B SS_5
Proposed$_{\theta=0.15}$	99.71%	99.66%	90.91%	81.99%
Proposed$_{\theta\to\infty}$	99.71%	99.66%	91.20%	81.99%
Proposed$_{d<0.1}$	*100.0%*	*100.0%*	*96.68%*	*81.66%*
Cho *et al.* [**?**]	100%	100%	99.2%	—
Yang *et al.* [**?**]	—	—	90.2%	47.9%
Jiang *et al.* [**?**]	100%	99.12%	83.46%	57.84%
Jia *et al.* [**?**]	100%	100%	**97.15%**	**96.13%**

Each plotted point refers to a single θ_l value, ranging from $[0, \infty]$. For each tested θ_l value, we plotted a series of performance metrics against the corresponding average processing time per image. Using the $\theta_l = 0.1$ vertical line from Figure 7a as a reference, we can distinguish three metrics that are common to all tested setups. The black line represents the evolution of r_1 when "doubtful" images

are classified through the aforementioned criterion, $\max_{m}(c_m)$. When no such *a posteriori* classification is performed, a set of images is left unclassified, as no level from the hierarchical chain was shown to present enough detail to perform an accurate recognition. The red line represents the evolution of the "doubtful" image percentage with increasing θ_l values. It is intuitive to note that lower θ_l values generate lower reject-option classification ratios, as a smaller dissimilarity between individuals will still trigger a "certain" decision. If no "doubtful" images are considered in the performance evaluation of the proposed algorithm, then the r_1 computation is performed only with respect to the non-rejected images. The blue line represents this alternative evaluation, and the corresponding reject percentage can be easily traced to the equivalent θ_l point in the red line. Taking, once again, the $\theta_l = 0.1$ vertical line from Figure 7a as a reference, we can see that the proposed methodology yields a r_1 value of roughly 98.00%. If the approximately 30% "doubtful" images are not considered for evaluation, the r_1 value increases to 100.00%, with respect to the non-rejected images.

For a better comprehension and deeper analysis of the method, we chose to assess performance in three specific points:

(1) Optimal $\boldsymbol{\theta_l}$ value: We consider the optimal θ_l value as the point where a visible performance plateau is achieved in the time *vs.* performance plot. For the AR database, this value was set to $\theta_l = 0.4$, whereas for the Extended Yale B Face database, it was set to $\theta_l = 0.15$. For the ORL database, the θ_l value was optimized for each of the aforementioned experimental setups. For the single template approach, the value was set to $\theta_l = 0.02$, while for multiple templates, it was set to $\theta_l = 0.15$.

(2) $\boldsymbol{\theta_l \to \infty}$: extreme behavior when the θ_l parameter is set to high values.

(3) $\boldsymbol{d < 0.1}$: point where the ratio of non-classified images at the end of the hierarchical chain reaches 10% of all tested images.

A thorough analysis of Figures **??**–**??** leads to some interesting conclusions regarding the behavior of the proposed algorithm under variable image acquisition conditions. The performance metric r_1 was computed for a set of θ_l values, ranging from $[0, \infty]$. It is trivial to understand that lower values of θ_l will lead the proposed algorithm to an extreme case where all images are classified in the first level of information, *i.e.*, the full face, thus reducing the computational complexity and average processing time. On the other hand, when $\theta \to \infty$, all images will reach the end of the hierarchical chain without a certain decision having been made. In this opposite extreme behavior, all images are classified *a posteriori* using information from every level, with a significant increase in both computational complexity and performance. With such a wide variety of possibilities, we chose to analyze the global behavior of the proposed work by plotting the evolution of r_1 values against the average processing time, when the θ_l parameter goes from $[0, \infty]$.

Regarding the evolution of r_1 for variable θ_l values, we might point out the considerable drop in performance under less ideal conditions, when $\theta \to 0$. Whereas Subsets 2 and 3 from the Extended Yale B database show excellent performance from both aforementioned approaches, rivaling those obtained for $\theta \to \infty$, such behavior is highly compromised for the more challenging scenarios of Subsets 4 and 5. In such cases, the observed drop in performance is less significant for higher θ_l values. Furthermore, the results obtained for the AR database stress that the proposed methodology

is capable of consistently presenting high performance regardless of the acquisition conditions and noise factors. Whereas the Extended Yale B database presented the challenge of heterogeneous illumination, which might be understood as a "natural" source of occlusion, the AR database presents the challenge of spatially coherent occlusion regions. For both cases, performance observed for the proposed hierarchical methodology rivals the state-of-the-art. Regarding the $\theta \to \infty$ case, the expectation that the best results would be consistently observed for this extreme scenario was fulfilled, at the expense of higher computation complexity and average processing time. Nevertheless, if no time constraints exist in the application scenario in which the proposed methodology is implemented, higher values of θ_l seem the ideal choice. The ORL database maintains a somewhat stable performance regardless of the chosen value for the θ_l parameter. As the proposed hierarchical methodology is contextually motivated for occlusion scenarios and no such conditions are present in this database, the observed behavior fits the expectations. Regardless of the absence of occlusions in the ORL images, the reject option classifier still proves a useful tool at discriminating "doubtful" images and increasing the reliability of the decisions obtained for the "non-doubtful" ones.

Tables **??**–**??** present a comparative analysis between the proposed methodologies and some state-of-the-art works, in similar experimental setups, for the AR and Extended Yale B databases, respectively. It is readily observed that the proposed algorithm presents the most consistent and robust behavior, regardless of the nature of the present occlusions, for the AR face database. While the work by Li *et al.* [**?**] presents higher performance for sunglasses occlusion, their performance regarding the alternative scarf occlusion is considerably lower than the one obtained with the proposed algorithm. Min *et al.* [**?**] present a work that suffers from exactly the opposite problem, with good performance observed for scarf occlusion, but lower results for sunglasses. The aforementioned works seem to suffer from overfitting to certain classes of images and lack the robustness to adapt to new cases. Such robustness to the nature and location of the occlusion can be observed for both the proposed methodology, as well as the works from Morelli *et al.* [**?**] and Qian *et al.* [**?**]. Both of these works show a similar trend to the proposed algorithm and high performance for the whole database. Alongside our work, and to the best of our knowledge, they represent the state-of-the-art performance for the AR database in unconstrained face recognition. A similar analysis can be made for both setups under which experiments were carried out with the ORL database. The best performance obtained with the proposed algorithm is on par with the best value found in the literature, in the work by Geng and Jiang [**?**]. This observation leads to the conclusion that the proposed methodology is capable of achieving state-of-the-art performance for a wider variety of scenarios besides occlusion. The aforementioned observation that performance seems to be independent of the chosen value for the θ_l parameter may, however, indicate that a simpler approach than the whole hierarchical chain might also output good results. Using simply the first level, *i.e.*, the full-face images, the r_1 values observed for the single and multiple template setups were 88.10% $\pm$ 1.05% and 98.5%, respectively. These results differ very little from the optimal results presented in Table **??**, thus corroborating the expectation of good performance for a less computationally complex approach.

Regarding the Extended Yale B Face database, our algorithm is shown to perform similar to state-of-the-art performance for all subsets, except Subset 5. Even for this subset, the obtained

performance is still higher than a few recently published works, even though no direct comparison can be performed with the work by Cho *et al.* [**?**], whose authors claim that its images are "excluded from the experiment because the illumination conditions are so severe that some images are difficult to recognize even to the naked eye". Observing the behavior of the blue and red lines for high threshold values, we note that a significant amount of images in this subset exceeds the recognition capability of the proposed methodology. The fact that the blue and black lines diverge greatly, when compared to the results observed in other subsets, shows that posterior classification by the maximum c_m value may not be the most beneficial approach when working with such severely degraded images. In such cases, the reject option classification yields considerably higher performance than the posterior classification of doubtful cases, even for low rejection ratios. It is easily noted how the blue and black lines start to diverge for a value of $d << 0.1$, as a result of this behavior.

4.2. Implementation Details

The proposed algorithm was developed in MATLAB R2012a and tested on a PC with 3.40 GHz Intel(R) Core(TM) i7-2600 processor and 8 GB RAM. To train the GMMs, we used the Netlab toolbox [**?**], whereas dSIFT keypoint description was performed using the VLFeat toolbox [**?**].

5. Conclusions and Future Work

In the present work, we propose an algorithm for face recognition under unconstrained settings, such as heterogeneous illumination and severe occlusions. By training models that only describe features of limited regions of the face, we confer our algorithm a robustness to events where all other regions are occluded. In an attempt to replicate the global precedent hypothesis of the human brain's cognitive mechanisms of perception, we designed an innovative hierarchical recognition algorithm, where face recognition is performed locally only if a more global representation is not capable of achieving a decision with a high degree of certainty. Even though good performance was observed for a wide variety of non-ideal conditions, some ideas are worth noting for future research in the area.

First of all, the choice of SIFT keypoint descriptors distributed in a dense grid is based on a series of assumptions, like constant scale and orientation, that might not always hold. Besides these assumptions, we performed no comparative analysis with other descriptors that might prove more appropriate for face description in less controlled acquisition scenarios. The feature extraction algorithm proposed by Miao and Jiang [**?**], for example, is reported to be more efficient at extracting interest points from human face images than the SIFT approach followed in the present work and might prove to be an interesting alternative. Still, regarding face representation, no alternative pre-processing methods were considered besides the aforementioned Weber-faces approach. Given that the most significant performance drop of the proposed algorithm was observed for images acquired under significantly low illumination, alternative normalization techniques might deserve further research. A different approach might be considered, as previously mentioned, by introducing a feedback mechanism to the recognition system, with the user being notified when the reject option is triggered in the end of the hierarchical chain. This could allow a user to actively adapt the environmental conditions so as to facilitate a more certain decision.

The choice of the fittest θ_l value for each specific application might also motivate a more detailed study. The hypothesis that some users are easier to identify than others inside a given population, an effect known as the Doddington zoo effect [**?**], suggests that parameter optimization in biometric applications might benefit from a more individual-specific approach. Besides this observation, we fixed the θ_l value for all levels of the hierarchical chain. We assumed that due to the normalization of the likelihood-ratio provided by the UBM, recognition scores at different levels share a similar range of values. Even though this assumption seems valid enough in the proposed framework, analyzing the effect of level-specific threshold values might bring about slight increases in performance.

In a final consideration, the proposed work was carried out fully on frontal face images. Most real-life applications in unconstrained scenarios should not enforce such a rigid constraint, and thus, the proposed algorithm should be capable of coping with pose variations. Two alternatives to approach this problem could be the training of multiple models describing individual poses or the training of a single model built on images acquired under multiple poses. Further research into both alternatives would be needed before either of them can be considered as the most fit to expand the proposed framework.

Acknowledgments

Author Contributions

João C. Monteiro and Jaime S. Cardoso conceived and designed the experiments; João C. Monteiro performed the experiments; João C. Monteiro and Jaime S. Cardoso analyzed the data; João C. Monteiro wrote the paper under the guidance of Jaime S. Cardoso.

Conflicts of Interest

The authors declare no conflict of interest.

References

. Prabhakar, S.; Pankanti, S.; Jain, A.K. Biometric recognition: Security and privacy concerns. *IEEE Secur. Priv.* **2003**, *1*, 33–42.
. Jain, A.; Hong, L.; Pankanti, S. Biometric identification. *Commun. ACM* **2000**, *43*, 90–98.
. Monteiro, J.C.; Sequeira, A.F.; Oliveira, H.P.; Cardoso, J.S. Robust iris localisation in challenging scenarios. *Commun. Comput. Inform. Sci.* **2014**, *458*, 146–162.
. Castrillón, M.; Déniz, O.; Hernández, D.; Lorenzo, J. A comparison of face and facial feature detectors based on the Viola–Jones general object detection framework. *Mach. Vis. Appl.* **2011**, *22*, 481–494.
. Min, R.; Hadid, A.; Dugelay, J.L. Efficient detection of occlusion prior to robust face recognition. *Sci. World J.* **2014**, *2014*, doi:10.1155/2014/519158.

. Li, Y.; Meng, L.; Feng, J.; Wu, J. Downsampling sparse representation and discriminant information aided occluded face recognition. *Sci. China Inf. Sci.* **2014**, *57*, 1–8.

. Karande, K.J.; Talbar, S.N. *Independent Component Analysis of Edge Information for Face Recognition*; Springer: New Delhi, India, 2014.

. Tanaka, J.W.; Farah, M.J. Parts and wholes in face recognition. *Q. J. Exp. Psychol.* **1993**, *46*, 225–245.

. Schwaninger, A.; Schumacher, S.; Bülthoff, H.; Wallraven, C. Using 3D computer graphics for perception: The role of local and global information in face processing. In Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization, Türbingen, Germany, 25–28 July 2007; pp. 19–26.

. Gold, J.M.; Mundy, P.J.; Tjan, B.S. The perception of a face is no more than the sum of its parts. *Psychol. Sci.* **2012**, *23*, 427–434.

. Cho, H.; Roberts, R.; Jung, B.; Choi, O.; Moon, S. An Efficient Hybrid Face Recognition Algorithm Using PCA and GABOR Wavelets. *Int. J. Adv. Robot. Syst.* **2014**, *11*, 1–8.

. Navon, D. Forest before trees: The precedence of global features in visual perception. *Cogn. Psychol.* **1977**, *9*, 353–383.

. Turk, M.; Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* **1991**, *3*, 71–86.

. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D. Eigenfaces *vs.* fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720.

. Cootes, T.F.; Edwards, G.J.; Taylor, C.J. Active appearance models. *Lect. Notes Comput. Sci.* **1998**, *1407*, 484–498.

. Liao, S.; Jain, A.K. Partial face recognition: An alignment free approach. In Proceedings of the 2011 International Joint Conference on Biometrics (IJCB), Washington, DC, USA, 11–13 October 2011; pp. 1–8.

. Nallammal, N.; Radha, V. Comparative Analysis of Partial Occlusion Using Face Recognition Techniques. *Int. J. Image Proc. (IJIP)* **2013**, *7*, 132–139.

. Oh, B.S.; Toh, K.A.; Choi, K.; Beng Jin Teoh, A.; Kim, J. Extraction and fusion of partial face features for cancelable identity verification. *Pattern Recognit.* **2012**, *45*, 3288–3303.

. Geng, C.; Jiang, X. Face recognition based on the multi-scale local image structures. *Pattern Recognit.* **2011**, *44*, 2565–2575.

. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colombus, OH, USA, 24–27 June 2014; pp. 1701–1708.

. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.

. Zhou, Z.; Wagner, A.; Mobahi, H.; Wright, J.; Ma, Y. Face recognition with contiguous occlusion using markov random fields. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 1050–1057.

. Qian, J.; Yang, J.; Zhang, F.; Lin, Z. Robust Low-Rank Regularized Regression for Face Recognition with Occlusion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Colombus, OH, USA, 23–28 June 2014; pp. 21–26.

. Wang, J.; Lu, C.; Wang, M.; Li, P.; Yan, S.; Hu, X. Robust Face Recognition via Adaptive Sparse Representation. *IEEE Trans. Cybern.* **2014**, *44*, 2368–2378.

. Shen, Y.; Hu, W.; Yang, M.; Wei, B.; Lucey, S.; Chou, C.T. Face recognition on smartphones via optimised sparse representation classification. In Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, Berlin, Germany, 15–17 April 2014; pp. 237–248.

. Jian, Z.; Luxi, H.; Jian, J.; Yu, X. A Fast Iterative Pursuit Algorithm in Robust Face Recognition Based on Sparse Representation. *Math. Probl. Eng.* **2014**, *2014*, doi:10.1155/2014/683494.

. Zhang, W.; Shan, S.; Chen, X.; Gao, W. Local Gabor binary patterns based on Kullback–Leibler divergence for partially occluded face recognition. *IEEE Signal Process. Lett.* **2007**, *14*, 875–878.

. Reynolds, D.; Quatieri, T.; Dunn, R. Speaker verification using adapted Gaussian mixture models. *Digit. Signal Process.* **2000**, *10*, 19–41.

. Povey, D.; Chu, S.M.; Varadarajan, B. Universal background model based speech recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 30 March–4 April 2008; pp. 4561–4564.

. Reynolds, D.A. An overview of automatic speaker recognition technology. In Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, 13–17 May 2002; Volume 4, pp. 4072–4075.

. Reynolds, D. Gaussian mixture models. *Encyclopedia of Biometric Recognition*; Springer: New York, NY, USA, 2008; pp. 12–17.

. Xiong, Z.; Zheng, T.; Song, Z.; Soong, F.; Wu, W. A tree-based kernel selection approach to efficient Gaussian mixture model–universal background model based speaker identification. *Speech Commun.* **2006**, *48*, 1273–1282.

. Wang, B.; Li, W.; Yang, W.; Liao, Q. Illumination normalization based on weber's law with application to face recognition. *IEEE Signal Process. Lett.* **2011**, *18*, 462–465.

. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

. Bosch, A.; Zisserman, A.; Muñoz, X. Scene classification via pLSA. *Lect. Notes Comput. Sci.* **2006**, *3954*, 517–530.

. Simonyan, K.; Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Fisher vector faces in the wild. In Proceedings of the British Machine Vision Conference, Bristol, UK, 9–13 September 2013; Volume 1, p. 7.

. Shinoda, K.; Inoue, N. Reusing Speech Techniques for Video Semantic Indexing [Applications Corner]. *IEEE Signal Process. Mag.* **2013**, *30*, 118–122.

. Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 2, pp. 506–513.

. Herbei, R.; Wegkamp, M.H. Classification with reject option. *Can. J. Stat.* **2006**, *34*, 709–721.

. Samaria, F.S.; Harter, A.C. Parameterisation of a stochastic model for human face identification. In Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; pp. 138–142.

. Georghiades, A.; Belhumeur, P.; Kriegman, D. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660.

. Martinez, A.M. *The AR Face Database*; CVC Technical Report; The Ohio State University: Colombus, OH, USA, 1998; Volume 24.

. Zhang, D.L.; Qiao, J.; Li, J.B.; Qiao, L.Y.; Chu, S.C.; Roddick, J.F. Optimizing Matrix Mapping with Data Dependent Kernel for Image Classification. *J. Inf. Hiding Multimed. Signal Process.* **2014**, *5*, 72–79.

. Xia, S.; Zheng, G.; Ma, Y.; Ma, X. A Novel Super-resolution Approach Based on Supervised Canonical Correlation Analysis. In Proceedings of the 2014 International Conference on Computer, Communications and Information Technology (CCIT 2014), Beijing, China, 16–17 January 2014.

. Xu, W.; Lee, E.J. Face Recognition Using Wavelets Transform and 2D PCA by SVM Classifier. *Int. J. Multimed. Ubiquitous Eng.* **2014**, *9*, 281–290.

. Morelli Andrés, A.; Padovani, S.; Tepper, M.; Jacobo-Berlles, J. Face recognition on partially occluded images using compressed sensing. *Pattern Recognit. Lett.* **2014**, *36*, 235–242.

. Yang, J.; Qian, J.; Luo, L.; Zhang, F.; Gao, Y. Nuclear Norm based Matrix Regression with Applications to Face Recognition with Occlusion and Illumination Changes. **2014**, arXiv preprint arXiv:1405.1207.

. Jiang, Y.; Wu, Y.; Li, W.; Wang, L.; Liao, Q. Log-domain polynomial filters for illumination-robust face recognition. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Firenze, Italy, 4–9 May 2014; pp. 504–508.

. Jia, Q.; Fang, C.; Wen, D.; Ding, X. Generating face images under multiple illuminations based on a single front-lighted sample without 3D models. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2014; pp. 1–6.

. Nabney, I.T. *NETLAB: Algorithms for Pattern Recognition*; Springer: London, UK, 2004.

. Vedaldi, A.; Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. In Proceedings of the International Conference on Multimedia, Firenze, Italy, 25–29 October 2010; pp. 1469–1472.

. Miao, Z.; Jiang, X. Interest point detection using rank order LoG filter. *Pattern Recognit.* **2013**, *46*, 2890–2901.

. Ross, A.; Rattani, A.; Tistarelli, M. Exploiting the "doddington zoo" effect in biometric fusion. In Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, 2009, Washington, DC, USA, 28–30 September 2009; pp. 1–7.