

Robust Periocular Recognition by Fusing Sparse Representations of Color and Geometry Information

Juan C. Moreno¹ · V. B. Surya Prasath² · Gil Santos¹ · Hugo Proença¹

Received: 6 October 2014 / Revised: 11 March 2015 / Accepted: 2 July 2015 / Published online: 21 July 2015
© Springer Science+Business Media New York 2015

Abstract In this paper, we propose a re-weighted elastic net (REN) model for biometric recognition. The new model is applied to data separated into geometric and color spatial components. The geometric information is extracted using a fast cartoon - texture decomposition model based on a dual formulation of the total variation norm allowing us to carry information about the overall geometry of images. Color components are defined using linear and nonlinear color spaces, namely the red-green-blue (RGB), chromaticity-brightness (CB) and hue-saturation-value (HSV). Next, according to a Bayesian fusion-scheme, sparse representations for classification purposes are obtained. The scheme is numerically solved using a gradient projection (GP) algorithm. In the empirical validation of the proposed model, we have chosen the *periocular region*, which is an emerging trait known for its robustness against low quality data. Our results were obtained in the publicly available FRGC and UBIRIS.v2 data sets and show consistent improvements

in recognition effectiveness when compared to related state-of-the-art techniques.

Keywords Sparse representation · Periocular recognition · Total variation · Elastic net regularization · Color · Texture decomposition

1 Introduction

Biometrics attempts to recognize human beings according to their physical or behavioral features [18]. In the past, various traits were used for biometric recognition, out of which *iris* and *face* are the most popular [20, 33, 39, 45]. The use of the periocular region was found to be useful on unconstrained scenarios [42]. The exploration of the periocular region as a biometric trait started with Park et al.'s pioneering approach [35], who performed local and global feature extraction. Images were aligned to take advantage of iris location, in order to define a 7×5 region of interest (ROI) grid. Patches were encoded by applying two well known distribution-based descriptors, local binary patterns (LBP) [30] and histogram of oriented gradients (HOG) [10], quantized into 8-bin histograms. Finally, they merged all histograms into a single-dimension array containing both texture and shape information, and matching was carried out based on the Euclidean distance. For the local analysis, authors employed Scale-Invariant Feature Transform (SIFT) [25]. The reported performance was fairly good, showing periocular fitness for recognition purposes, and further analysis was held on noise factors impact on performance [34].

Recently, various extensions and improvements based on Park et al. work [35] have been carried out. Miller et al. [26] presented an analysis which focused on periocular

✉ Juan C. Moreno
jcmb@ubi.pt

V. B. Surya Prasath
prasaths@missouri.edu

Gil Santos
gmelfe@ubi.pt

Hugo Proença
hugomcp@di.ubi.pt

¹ IT-Instituto de Telecomunicações, Department of Computer Science, University of Beira Interior, 6201-001 Covilha, Portugal

² Department of Computer Science, University of Missouri-Columbia, MO 65211, USA

skin texture, taking advantage of uniform local binary patterns (ULBP) [31] to achieve improved rotation invariance with uniform patterns and finer quantization of the angular space. Their work was extended by Adams et al. [1], who proposed using Genetic & Evolutionary Computing (GEC) to optimize feature set. Juefei-Xu et al. [21] used multiple local and global feature extraction techniques such as Walsh transforms and Laws’ masks, discrete cosine transform (DCT), discrete wavelet transform (DWT), Force Fields, Speed Up Robust Transform (SURF), Gabor filters and Laplacian of Gaussian (LoG). In their later work [22] efforts were made to compensate aging degradation effects on periocular performance. The possibility of score level fusion with other biometric traits was also addressed, for example in iris recognition [50]. Bharadwaj et al. [2] proposed the fusion of uniform local binary patterns (ULBP) with five perceptual dimensions, usually applied as scene descriptors: naturalness, openness, roughness, expansion and ruggedness – GIST [32]. In their approach the images were pre-processed with with Fourier transform for local contrast normalization, and then a spacial envelope computed with a set of Gabor filters (4 scales × 8 orientations). On the final stage, χ^2 distance was used to match the feature arrays, and results fused with a weighted sum. Woodard *et al.* [50] assessed how periocular texture information could improve iris data reliability, thus overcoming the difficulties associated with non-ideal imaging. To handle periocular information the LBP was computed over a ROI grid (on both datasets), and at the matching stage the LBP histograms were matched using Manhattan distance. Iris processing was achieved following Daugman’s approach [11], except with manual segmentation. Information from both traits was combined at score-level using a simple weighted sum after min-max normalization.

Based on the pioneering work of Wright et al. [51], the sparse representation theory is emerging as a popular method in the biometrics field and is considered specially suitable to handle degraded data acquired under uncontrolled acquisition protocols [28, 37, 44]. A query image is first sparsely coded over the template images, and then the classification is performed. Sparse Representation based Classification (SRC) is robust to occlusion, illumination and noise, and achieves excellent performance.

1.1 Sparse Representation

Model selection in high-dimensional problems has been gaining interest in the statistical signal processing community [4, 12]. Using convex optimization models, the main problem is recovering a sparse solution $\hat{\mathbf{x}} \in \mathbb{R}^n$ of an under-determined system of the form $\mathbf{y} = \mathbf{Ax}^*$, given a vector $\mathbf{y} \in \mathbb{R}^m$ and a matrix $A \in \mathbb{R}^{m \times n}$. There is a special interest in signal recovery when the number of predictors is much

larger than the number of observations ($n \gg m$). A direct solution to the problem is to select a signal whose measurements are equal to those of \mathbf{x}^* , with smaller sparsity by solving a minimization problem based on the ℓ^0 -norm:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subj. to } \mathbf{Ax} = \mathbf{y}, \tag{1}$$

($\|\mathbf{x}\|_0 = \#\{i : x_i \neq 0\}$), being a direct approach to seek the sparsest solution. Problem (1) is proved to be NP-hard and difficult to approximate since it involves non-convex minimization [5]. An alternative method is to relax the problem (1) by means of the ℓ^1 -norm ($\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$). Hence problem (1) can be replaced by the following ℓ^1 -minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subj. to } \mathbf{Ax} = \mathbf{y},$$

which can be solved by standard linear programming methods [9]. In practice, signals are rarely exactly sparse, and may often be corrupted by noise. Under noise, the new problem is to reconstruct a sparse signal $\mathbf{y} = \mathbf{Ax}^* + \kappa$, where $\kappa \in \mathbb{R}^m$ is white Gaussian noise with zero mean and variance σ^2 . In this case the associated ℓ^1 -minimization problem adopts the form:

$$\min_{\mathbf{x}} \left\{ \tau \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 \right\}, \tag{2}$$

where τ is a nonnegative parameter and $\|\cdot\|_2$ denotes the ℓ^2 -norm ($\|\mathbf{x}\|_2 = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$). The convex minimization problem (2) is known as the least absolute value shrinkage and selection operator (LASSO) [47].

Although sparsity of representation seems to be well established by means of the LASSO approach, some limitations were remarked by Hastie et al. [54]. LASSO model tends to select at most m variables before it saturates and in case predictors are highly correlated, LASSO usually selects one variable from a group, ignoring others. In order to overcome these difficulties, Hastie et al. [54] proposed the elastic net (EN) model as a new regulation technique for outperforming LASSO in terms of prediction accuracy. The elastic net is characterized by the presence of ridge regression term (ℓ^2 -norm) and it is defined by the following convex minimization problem:

$$\min_{\mathbf{x}} \left\{ \tau_1 \|\mathbf{x}\|_1 + \tau_2 \|\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 \right\}, \tag{3}$$

where τ_1 and τ_2 are non-negative parameters. An improvement for the EN model was proposed in [55] where a combination of the ℓ^2 -penalty and an adaptive version of the ℓ^1 -norm have been implemented by considering the minimization problem

$$\min_{\mathbf{x}} \left\{ \tau_1 \sum_{i=1}^n \omega_i |x_i| + \tau_2 \|\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 \right\}, \tag{4}$$

where the adaptive weights are computed using a solution given by the EN minimization problem (3). If we let the solution of EN to be $\hat{\mathbf{x}}(EN)$, then the weights are given by the equation $\omega_i = 1/(|\hat{x}_i(EN)| + (1/m)^\vartheta)$ where ϑ is a positive constant. A variant of the above model was proposed in [17] by incorporating the adaptive weight matrix W in the ℓ^2 -penalty term:

$$\min_{\mathbf{x}} \left\{ \tau_1 \sum_{i=1}^n \omega_i |x_i| + \tau_2 \sum_{i=1}^n \omega_i^2 x_i^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\}. \quad (5)$$

In this paper we use a re-weighted elastic net regularization model for periocular recognition application.

1.2 Summary of Contributions

The main contribution of this paper is to propose a re-weighted elastic net (REN) regularization model that enhances the sparsity of the solutions found. The proposed REN model is a regularization and variable selection method that enjoys sparsity of representation, particularly when the number of predictors is much larger than the number of observations. The weights are computed such that larger weights will encourage small coordinates by means of the ℓ^1 -norm, and smaller weights will encourage large coordinates due to the ℓ^2 -norm. Our model differs from the schemes in [55] and [17] (see Eqs. 4 and 5 above), since the ℓ^1 and ℓ^2 terms are automatically balanced by weights which are continuously updated using $\omega_i = 1/(|\hat{x}_i| + \epsilon)$ with ϵ a positive parameter [7]. We also provide a concise proof of the existence of a solution for the proposed model as well as its accuracy property.

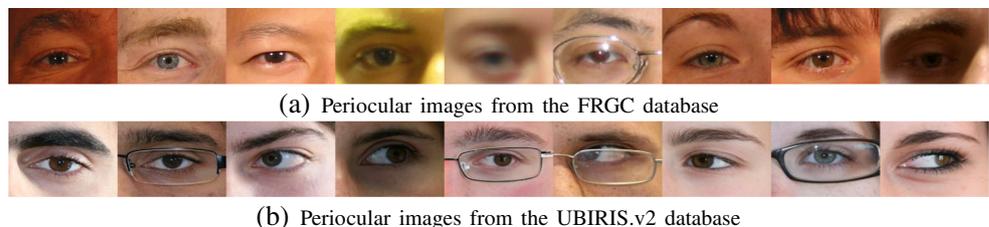
A complete presentation of the numerical implementation of the REN model using a gradient projection (GP) method [15], seeking sparse representations along certain gradient directions is described in this paper. We use a reformulation of the REN model as a quadratic programming (QP) problem. As a main application of our model, we consider the periocular recognition problem. The periocular region has been regarded as a trade-off between using the entire face or only the iris in biometrics. Periocular region is particularly suitable for recognition under visible wavelength light and uncontrolled acquisition conditions [34, 35, 50].

We enhance periocular recognition through the sparsity-seeking property of our REN model over different periocular sectors, which are then fused according to a Bayesian decision based scheme. The main idea is to benefit from the information from each sector, which should contribute in overall recognition robustness. Two different domains are considered for this purpose: (1) *geometry* and (2) *color*. Full geometry information is accessed by decomposing a given image into their cartoon - texture components by means of a dual formulation of the weighted total variation (TV) scheme [3, 29, 41]. For color, a key contribution is the use of nonlinear features such as chromaticity and hue components, which are thought to improve image geometry information according to human perception and previously used for image processing tasks [23, 27, 38]. Our methodology is inspired by two related works: 1) Wright et al. [51], which introduced the concept of *sparse representation for classification* (SRC) purposes; and 2) Pillai et al. [37], that used a SRC model for disjoint sectors of the iris and fused results at the score level, according to a confidence score estimated from each sector.

Our experiments are carried out in periocular images of two different data sets: (1) FRGC data set [36] which contains neutral and smiling frontal faces captured in scenarios under uniform and uncontrolled illumination conditions as shown in Fig. 1a. Noise effects such as eye blink, motion blur, occlusion and reflections presented in this database leads to difficulties in performing high recognition accuracy. (2) UBIRIS.v2 data set [40] in which images were acquired at visible wavelengths, from 4 to 8 meters away from the subjects and uncontrolled acquisition conditions. Varying gazes, poses and amounts of occlusions (due to glasses and reflections) are evident in this data set and makes the recognition task harder, see Fig. 1b. The results obtained using our model allowed us to conclude about consistent increase in performance when compared to the classical SRC model and other important approaches (e.g., Wright et al. [51] and Pillai et al. [37]). Also, it should be stressed that such improvements were obtained without a significant overload in the computational burden of the recognition process.

The rest of the paper is organized as follows. Section 2 summarizes the most relevant approaches in the scope of this work concerning penalized feature selection for sparse

Figure 1 Examples of periocular images of different subjects with and different illumination conditions (FRGC) and varying gazes (UBIRIS.v2), containing the *corneal*, *eyebrows* and *skin* regions.



representation. The re-weighted elastic net (REN) model is introduced together with statistical motivation ensuring high prediction rates. An algorithm based on gradient projection (GP) for the REN model is also introduced. Section 3 describes the different geometrical information extracted from periocular images for performing recognition based on cartoon - texture and chromaticity features in a total variation framework. Section 4 describes the experimental validation procedure carried out together with remarkable comparisons. Finally, Section 5 concludes the paper.

2 The Reweighted Elastic net Model for Classification Model

2.1 The LASSO Model for Recognition

We first briefly describe the sparse representation based classification framework which is a precursor to our REN based approach. Having a set of labeled training samples (n_i samples from the i^{th} subject), they are arranged as columns of a matrix $A(i) = [\mathbf{v}_{i,1}, \dots, \mathbf{v}_{i,n_i}] \in \mathbb{R}^{m \times n_i}$. A dictionary results from the concatenation of all samples of all classes:

$$A = [A(1), \dots, A(k)] = [\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,n_1} | \dots | \mathbf{v}_{k,1}, \dots, \mathbf{v}_{k,n_k}].$$

The key insight is that any probe \mathbf{y} can be expressed as a linear combination of elements of A . As the data acquisition process often induces noisy samples, it turns out to be practical to make use of the LASSO model. In this case it is assumed that the observation model has the form $\mathbf{y} = A\mathbf{x}^* + \boldsymbol{\kappa}$.

Classification is based on the observation that high values of the coefficients in the solution $\hat{\mathbf{x}}$ are associated with the columns of A of a single class, corresponding to the identity of the probe. A residual score per class $\mathbb{1}_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined: $\hat{\mathbf{x}} \rightarrow \mathbb{1}_i(\hat{\mathbf{x}})$, where $\mathbb{1}_i$ is a indicator function that set the values of all coefficients to 0, except those associated to the i^{th} class. Over this setting, the probe \mathbf{y} is then reconstructed by $\hat{\mathbf{y}}_i = A\mathbb{1}_i(\hat{\mathbf{x}})$, and the minimal reconstruction error deemed to correspond to the identity of the probe, between \mathbf{y} and $\hat{\mathbf{y}}_i$ given by $\text{id}(\mathbf{y}) = \arg \min_i r_i(\mathbf{y})$, with $r_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2$.

In [51] a sparsity concentration index (SCI) is used to accept/reject the response given by the LASSO model. The SCI of a coefficient vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ corresponds to:

$$SCI(\hat{\mathbf{x}}) = \frac{k \max_i \|\mathbb{1}_i(\hat{\mathbf{x}})\|_1}{\|\hat{\mathbf{x}}\|_1} - 1 \in [0, 1].$$

If $SCI(\hat{\mathbf{x}}) \approx 1$, the computed signal $\hat{\mathbf{x}}$ is considered to be acceptably represented by samples from a single class. Otherwise, if $SCI(\hat{\mathbf{x}}) \approx 0$ the sparse coefficients spread evenly

across all classes and a reliable identity for that probe cannot be given.

The recognition model proposed by Pillai et al. [37] obtains separate sparse representations from disjoint regions of an image and fusing them by considering a quality index from each region. Let L be the number of classes with labels $\{c_i\}_{i=1}^L$. A probe \mathbf{y} is divided into sectors, each one described by the SRC algorithm. SCI values are obtained over each sector, allowing to reject those with quality below a threshold. Let $\{d_i\}$ represent the class labels of the retained sectors, and $\mathbb{P}(d_i|c)$ be the probability that the i -th sector returns a label d_i , when the true class is c :

$$\mathbb{P}(d_i|c) = \begin{cases} \frac{t_1^{SCI(d_i)}}{t_1^{SCI(d_i)} + (L-1)t_2^{SCI(d_i)}} & \text{if } d_i = c, \\ \frac{t_2^{SCI(d_i)}}{t_1^{SCI(d_i)} + (L-1)t_2^{SCI(d_i)}} & \text{if } d_i \neq c, \end{cases}$$

being t_1 and t_2 constants such that $0 > t_1 > t_2 > 1$. According to a maximum a posteriori (MAP) estimate of the class label, the response corresponds to the class having the highest accumulated SCI:

$$\tilde{c} = \arg \max_{c \in \mathbf{C}} \frac{\sum_{j=1}^L SCI(d_j)\delta(d_j = c)}{\sum_{j=1}^L SCI(d_j)}. \tag{6}$$

2.2 The Re-weighted Elastic net (REN) Method

The proposed REN model is a sparse representation scheme balancing the LASSO shrinkage term (ℓ^1 -norm) and the strengths of the quadratic regularization (ℓ^2 -norm) coefficients by the following minimization problem:

$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^n \omega_i |x_i| + \sum_{i=1}^n (1 - \omega_i)^2 x_i^2 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}, \tag{7}$$

where $\omega_1, \dots, \omega_n$ are positive weights taking values in $(0, 1)$. The REN-penalty $\sum_{i=1}^n \omega_i |x_i| + \sum_{i=1}^n (1 - \omega_i)^2 x_i^2$ is strictly convex and it is a compromise between the ridge regression penalty and the LASSO. The convex combination in the REN-penalty term is natural in the sense that both the ℓ^1 and ℓ^2 norms are balanced by weights controlling the amount of sparsity versus smoothness expected from the minimization scheme. As in [7], the weights are chosen such that they are inversely related to the computed signal according to the equation $\omega_i = 1/(|\hat{x}_i| + \epsilon)$ with ϵ a positive parameter. Under this setting, large weights w_i will encourage small coordinates with respect to the REN-penalty term, whereas small weights imply big coordinates with respect to the REN-penalty term, respectively. Then,

it is straightforward that the new model combines simultaneously a continuous shrinkage and an automatic variable selection approach. We next consider the existence of solution and the sign recovery property of the REN model.

Next we describe an algorithm for the REN model allowing us to directly deal with the case $n \gg m$. It turns out that our REN model can be expressed as a quadratic program (QP), thus allowing us to apply a gradient projection approach to perform the sparse reconstruction.

2.3 Numerical Implementation

The algorithm that alternates between the computed signal and redefining the weights is as follows:

1. Choose initial weights $w_i = 1/2, i = 1, \dots, n$.
2. Find the solution $\hat{\mathbf{x}}$ of the problem

$$\min_{\mathbf{x}} \|\mathbf{W}\mathbf{x}\|_1 + \|(1 - \mathbf{W})\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2, \tag{8}$$

3. Update the weights: for each $i = 1, \dots, n, w_i = 1/(|\hat{x}_i| + \epsilon)$, where ϵ is a positive stability parameter.
4. Terminate on convergence or when a specific number of iterations is reached. Otherwise, go to step 2.

Note that our REN problem in Eq. 8 can also be expressed as a quadratic program [16], by splitting the variable \mathbf{x} into its positive and negative parts. That is, $\mathbf{x} = \mathbf{x}_+ - \mathbf{x}_-$, where \mathbf{x}_+ and \mathbf{x}_- are the vectors that collect the positive and negative coefficients of \mathbf{x} , respectively. Then, we handle the minimization problem,

$$\min_{\mathbf{z}} \{ Q(\mathbf{z}) = \mathbf{c}^T \mathbf{z} + \mathbf{z}^T \mathbf{B}\mathbf{z} \}, \tag{9}$$

where $\mathbf{z} = [\mathbf{x}_+, \mathbf{x}_-]^T, \mathbf{w}_n = [\omega_1, \dots, \omega_n]^T, \mathbf{c} = \mathbf{w}_{2n} + [-A^T \mathbf{y}; A^T \mathbf{y}]^T$ and $B = \frac{1}{2} B_1 + B_2$ with

$$B_1 = \begin{pmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{pmatrix}, \quad B_2 = \begin{pmatrix} (1 - W)^2 & -(1 - W)^2 \\ -(1 - W)^2 & (1 - W)^2 \end{pmatrix}.$$

The minimization problem (9) can then be solve using the Barzilai-Borwein Gradient Projection Algorithm [43]. Under this approach the iterative equation is given by, $\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \zeta^{(k)} \nu^{(k)}$, where $\zeta^{(k)}$ is the step size computed as $\zeta^{(k)} = (\mathbf{z}^{(k)} - \alpha^{(k)} \nabla Q(\mathbf{z}^{(k)}))_+ - \mathbf{z}^{(k)}$, with

$$\alpha^{(k+1)} = \begin{cases} \text{mid} \left\{ \alpha_{min}, \frac{\|\zeta^{(k)}\|^2}{(\zeta^{(k)})^T B \zeta^{(k)}}, \alpha_{max} \right\}, & \text{if } (\zeta^{(k)})^T B \zeta^{(k)} \neq 0 \\ \alpha_{max}, & \text{otherwise.} \end{cases}$$

The operator *mid* is defined as the middle value of three scalar arguments and α_{min} and α_{max} are two given parameters. The parameter ν take the form

$$\nu^{(k)} = \begin{cases} \text{mid} \left\{ 0, \frac{(\zeta^{(k)})^T \nabla Q(\mathbf{z}^{(k)})}{(\zeta^{(k)})^T B \zeta^{(k)}}, 1 \right\}, & \text{if } (\zeta^{(k)})^T B \zeta^{(k)} \neq 0, \\ 1, & \text{otherwise.} \end{cases}$$

The performance of the REN minimization along with comparisons is shown in Fig. 2 for a sparse signal. We want to reconstruct a length- n sparse signal (in the canonical basis) from m observations, with $m \ll n$. The matrix $A_{m \times n}$ is build with independent samples of a standard Gaussian distribution and by ortho-normalizing the rows, while the original signal \mathbf{x}^* contains 160 randomly placed \pm spikes and the observation is defined as $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \kappa$ with κ a Gaussian noise of variance $\sigma^2 = 10^{-4}$. The reconstruction of the original signal over the REN minimization problem produces a much lower mean squared error (MSE = $(1/n)\|\hat{\mathbf{x}} - \mathbf{x}^*\|^2$ with $\hat{\mathbf{x}}$ been an estimate of \mathbf{x}^*) equal to 3.499×10^{-05} , while the MSE given by the adaptive elastic model proposed in [17, 55] and LASSO are 5.194×10^{-05} , 4.791×10^{-05} and 1.445×10^{-04} respectively. Therefore, the proposed REN approach does an excellent job at locating the spikes.

Remark 1 The iterative reweighted algorithm falls in the general class of Majorization - Minimization (MM) algorithms [24]. An interesting example of separable iterative reweighing for sparse solutions is presented in [7] where the selection

$$w_i^{(k+1)} \rightarrow \frac{1}{|x_i^{(k+1)}| + \epsilon}$$

is suggested. Here ϵ is generally chosen as a fixed, application-dependent constant. In the noiseless case, it is demonstrated based on [14] that this amounts to iteratively solving

$$\min_{\mathbf{x}} \sum_{i=1}^n \log(x_i + \epsilon), \quad \text{subj. to } \mathbf{A}\mathbf{x} = \mathbf{y},$$

and convergence to a local minimum or saddle point is guaranteed. In [49] the reweighing

$$w_i^{(k+1)} \rightarrow \frac{1}{(x_i^{(k+1)})^2 + \epsilon |x_i^{(k+1)}|},$$

is also considered together with the case $\epsilon \rightarrow 0$. Related with the sparse solution of the model

$$\min_{\mathbf{x}} \left\{ \tau \|\mathbf{x}\|^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\},$$

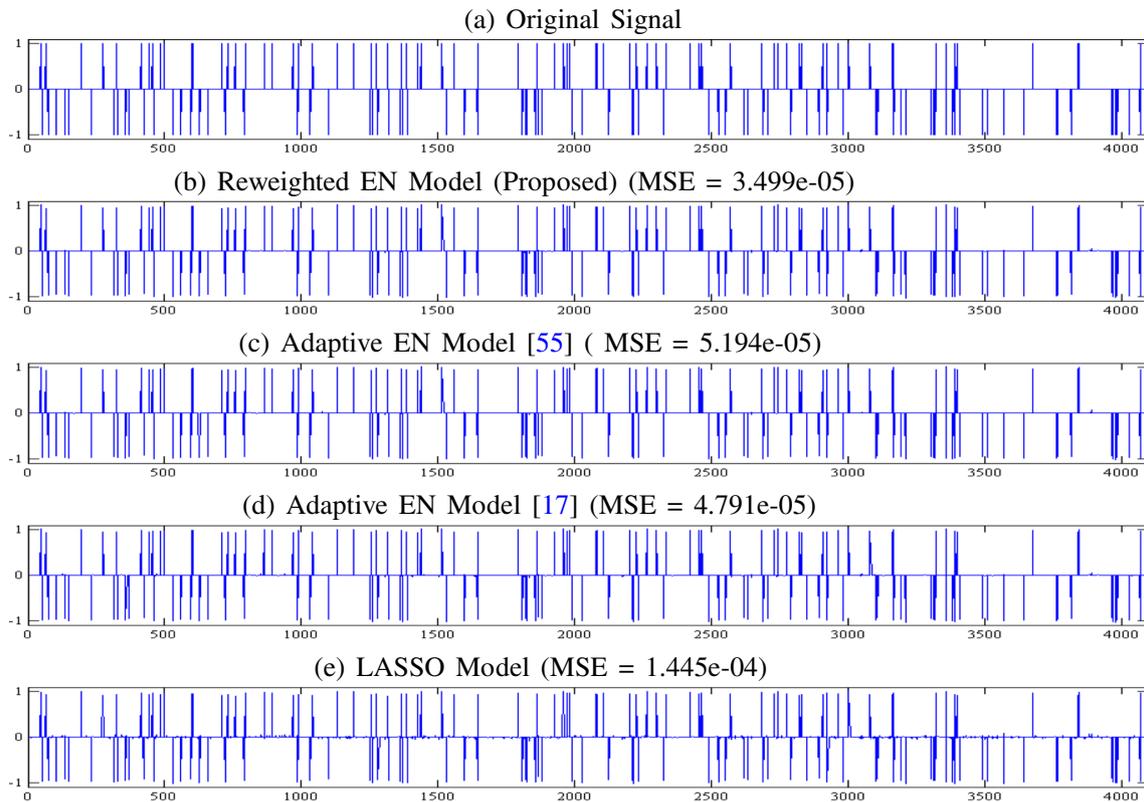


Figure 2 Sparse signal reconstruction with EN and LASSO models. (a) Sparse signal of Length $n = 4096$ with $k = 1024$ observations. (b)–(e) Response signals computed with the proposed reweighted elastic net, [17, 55] and LASSO, respectively.

the reweighting

$$w_i^{(k+1)} \rightarrow \frac{1}{\left(x_i^{(k+1)}\right)^2 + \epsilon^{(k+1)}}$$

is implemented in [8], where $\epsilon^{(k+1)} \geq 0$ is regularization factor the is reduced to zero as k becomes large.

3 Geometric and Color Spaces for Image Decomposition

Periocular features can be extracted using cartoon, texture and color components, which are then fused to improve periocular recognition according to Eq. 6. For this purpose, we next detail a variational cartoon + texture decomposition using the well known total variation regularization and different color spaces including a chromaticity-brightness decomposition based space.

3.1 Cartoon + Texture (CT) Space

The periocular images contain cartoon (smooth) and texture parts (small scale oscillations) which can be obtained

using the total variation (TV) [41] model. In this setting, the grayscale version of a periocular image is divided into two components representing the geometrical and texture parts. The TV based decomposition model is defined as an energy minimization problem,

$$\min_u \left\{ E_{TV}^{L^1}(u) = \int_{\Omega} g(\mathbf{x})|\nabla u| \, d\mathbf{x} + \lambda \int_{\Omega} |u - I| \, d\mathbf{x} \right\} \quad (10)$$

where I is the input grayscale image, and $g(\mathbf{x}) = \frac{1}{1+K|\nabla I|^2}$ is an edge indicator type function. Following [3] we use a splitting with an auxiliary variable v to obtain the following relaxed minimization,

$$\min_{u,v} \left\{ \tilde{E}_{TV}^{L^1}(u,v) = \int_{\Omega} g(\mathbf{x})|\nabla u| \, d\mathbf{x} + \frac{1}{2\theta} \int_{\Omega} (u+v-I)^2 \, d\mathbf{x} + \lambda \int_{\Omega} |v| \, d\mathbf{x} \right\}. \quad (11)$$

After a solution u is computed, it is expected to get the representation $I \approx u + v$, where the function u represents the geometric cartoon part, the function v contains texture information, and the function g represent edges. The minimization (11) is achieved by solving the following alternating sub-problems based on the dual minimization technique:

1. Fixing v , the minimization problem in u is:

$$\min_u \left\{ \int_{\Omega} g(\mathbf{x}) |\nabla u| \, d\mathbf{x} + \frac{1}{2\theta} \|u + v - I\|_{L^2(\Omega)}^2 \right\}. \quad (12)$$

The solution of (12) is given by $u = v - \theta \operatorname{div} \mathbf{p}$ where $\mathbf{p} = (p_1, p_2)$ satisfies $g(\mathbf{x}) \nabla(\theta \operatorname{div} \mathbf{p} - (I - v)) - |\nabla(\theta \operatorname{div} \mathbf{p} - (I - v))| \mathbf{p} = 0$, solved using a fixed point method: $\mathbf{p}^0 = 0$ and iteratively

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \delta t \nabla(\operatorname{div}(\mathbf{p}^n) - (I - v)/\theta)}{1 + \frac{\delta t}{g(\mathbf{x})} |\nabla(\operatorname{div}(\mathbf{p}^n) - (I - v)/\theta)|}.$$

2. Fixing u , the minimization problem in v is:

$$\min_v \left\{ \frac{1}{2\theta} \|u + v - I\|_{L^2(\Omega)}^2 + \lambda \|v\|_{L^1(\Omega)} \right\},$$

and the solution is found as

$$v = \begin{cases} I - u - \theta\lambda & \text{if } I - u \geq \theta\lambda, \\ I - u + \theta\lambda & \text{if } I - u \leq -\theta\lambda, \\ 0 & \text{if } |I - u| \leq \theta\lambda. \end{cases}$$

Figure 3 illustrates cartoon - texture decomposition of three grayscale periocular images for different iterations. With the increase in the number of iterations we notice that the cartoon component becomes smoother and texture component picks up more oscillations.

3.2 Color Spaces

For color periocular images we can obtain intensity and chromaticity decomposition which exploits color information. In computer vision there has been increasing interest in non flat image features that reside on curved manifolds which are well suited for edge detection and enhancement in color and multichannel images [46]. The flatness concept is related to functions taking all possible values in an open set in a linear space. The chromaticity feature of color images is an example of non-flat features. Given a color periocular image $\mathbf{I} : \Omega \rightarrow \mathbb{R}^3$, the RGB representation is defined by

a vector with three components $\mathbf{I} = (I_1, I_2, I_3)$. From the RGB color space, the chromaticity-brightness (CB) model arises by decomposing into the brightness component $\mathbf{B} : \Omega \rightarrow \mathbb{R}$ computed as $\mathbf{B} = |\mathbf{I}|$ and chromaticity components $\mathbf{C} = (C_1, C_2, C_3) : \Omega \rightarrow \mathbb{S}^2$ (where \mathbb{S}^2 is the unit sphere in \mathbb{R}^3) is computed by $C_i = I_i/\mathbf{B}$. We also make use of the Hue-Saturation-Value (HSV) color space since it is believed to be more natural and is related to human perception [52]. Figure 4 illustrates CB decomposition, and HSV color space conversions of a given RGB periocular image. In our experiments we compare grayscale CT decomposition and CB, RGB and HSV color space based decompositions along with the proposed REN model.

4 Experiments and Discussion

4.1 Performance Measures

REN model’s performance was quantitatively analyzed by using the Receiver Operating Characteristic (ROC) curves. In this case, given a signal $\hat{\mathbf{x}}$, if $\text{SCI}(\hat{\mathbf{x}}) > \beta$, the classifier outputs a positive response (P), otherwise a negative (N) result. For a fixed β , the sensitivity corresponds to the proportion of signals correctly detected by the SRC algorithm, whereas specificity counts the proportion for which the corresponding SCI values are below β , where β is an accepted threshold value.

$$\text{sensitivity} = \frac{\#TP}{\#TP + \#FN} \quad \text{and} \quad \text{specificity} = \frac{\#TN}{\#TN + \#FP}$$

where TP, FP, TN and FN correspond to the True Positive, False Positive, True Negative and False Negative, respectively. Table 1 summarizes these notions, combining the different classes of periocular signals and their relation with

Figure 3 Cartoon - Texture components for grayscale periocular images using a weighted TV model given in Eq. (10). (a) Grayscale periocular images. (b)-(c) Cartoon - Texture decompositions with 80 iterations. (d)-(e) Cartoon - Texture decompositions with 400 iterations.

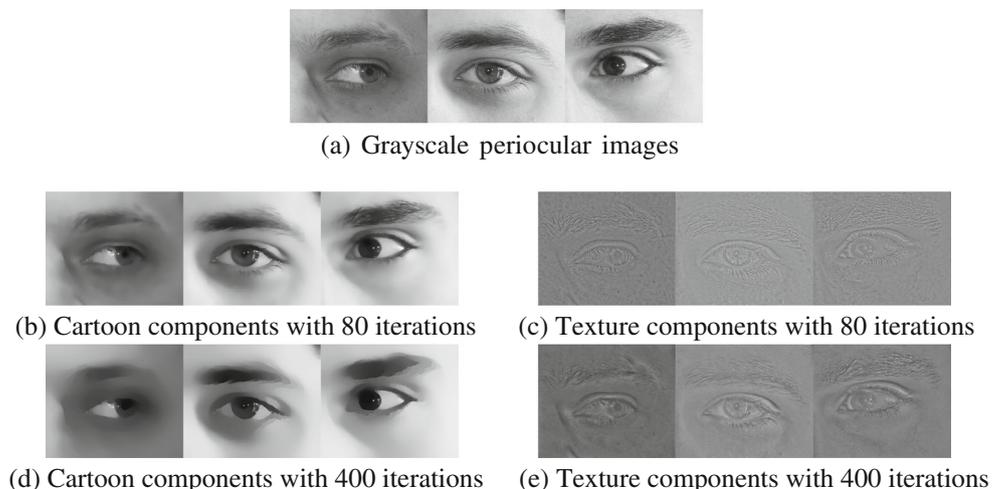
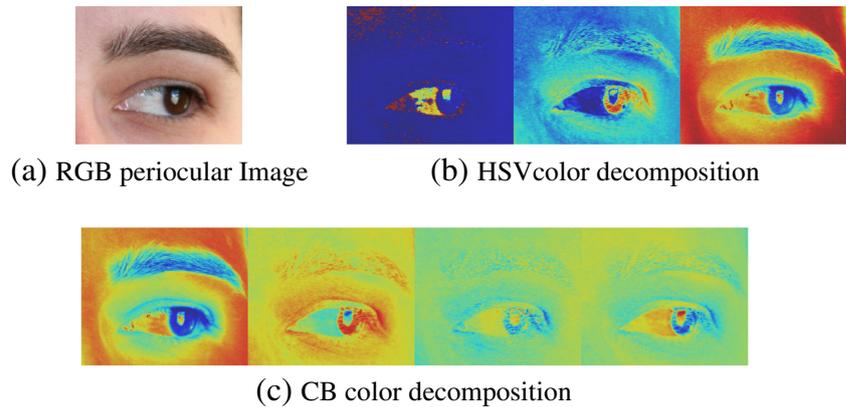


Figure 4 Different color decomposition for a given periocular image. **a** RGB color periocular image. **b** HSV color decomposition. **c** CB color decomposition.



the classifier induced by the minimal reconstruction error and the accumulated SCI value. The overall accuracy is given by:

$$\text{accuracy} = \frac{\# \text{TN} + \# \text{TP}}{\# \text{TN} + \# \text{FP} + \# \text{TP} + \# \text{FN}}.$$

In a ROC plot, the optimal recognition method would yield a point in the upper-left corner, corresponding to full sensitivity (no false negatives) and full specificity (no false positives). The statistical correlation between the outputs given by each channels considered in our method was also assessed. Considering that eventual dependences will be linear, the Pearson's sample correlation was used for that purpose. Given a pair of samples, the correlation coefficient is given by:

$$r(\hat{\mathbf{x}}^{(1)}, \hat{\mathbf{x}}^{(2)}) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\hat{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)}}{\sigma_{\hat{\mathbf{x}}^{(1)}}} \right) \left(\frac{\hat{x}_i^{(2)} - \bar{\mathbf{x}}^{(2)}}{\sigma_{\hat{\mathbf{x}}^{(2)}}} \right),$$

where $\hat{x}_i^{(1)}, \hat{x}_i^{(2)}$ denote the systems outputs, $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ are the sample means and $\sigma_{\hat{\mathbf{x}}^{(1)}}, \sigma_{\hat{\mathbf{x}}^{(2)}}$ the standard deviations.

4.2 Results for the FRGC Database

Among all the images in this dataset, we have selected those individuals with at least 14 images, making a total of 163 different classes. We randomly selected for each person half of the right periocular images for training (7 images) and the other half (7 images) for testing. For measuring the recognition performance of the REN model, periocular images are converted to grayscale, and they are also down-sampled according to the downsampling ratios of 1/24, 1/18, 1/12

and 1/6 which correspond to feature space dimensions of 30, 70, 140 and 560, respectively. The ability for the REN model to correctly classify a test subject is depicted in Fig. 5a by means of ROC curve for different downsampling ratios, and it is compared with the periocular recognition algorithms of Park et al. [35] and Woodard et al. [50] shown in Fig. 5b-c respectively.

Table 2 shows better recognition performance for the proposed REN model in comparison of the models implemented by Park et al. [35] and Woodard et al. [50], in terms of the sensitivity (sens.) and corresponding accuracy (acc.) for far (=1-specificity) $\leq 10\%$, the area under the curve (AUC) and the equal error rate (EER) measure. For the proposed REN model using grayscale images, the highest AUC and the lowest EER values are equal to 0.9218 and 0.0682 for downsampling ratio equal to 1/12. We noticed that the AUC and EER values are increased when downsampling ratio are decreased.

4.3 Results for the UBIRIS.v2 Database

For the UBIRIS.v2 data set, left side periocular images were considered for the biometric performance. We have randomly chosen the feature space dimension equal to 90 which correspond to images down-sampled to 10×9 pixels (other downsampling ratios can also be utilized as in FRGC dataset, see Section 4.2) and stored in "png" format. We took 6 samples from 150 different individuals such that one image per class is randomly chosen for the test and the remaining five samples are included in the dictionary. Experiments are repeated, changing the image used

Table 1 Types of errors, according to the SCI value and the sparse signal reconstruction following Wright et al. [51] and Pillai et al. [37] models.

Signal $\hat{\mathbf{x}}$	Optimal Signal Recovery	Non-Optimal Signal Recovery
$\text{SCI}(\hat{\mathbf{x}}) > \beta \rightarrow \text{Positive}$	True Positive (TP)	False Positive (FP)
$\text{SCI}(\hat{\mathbf{x}}) \leq \beta \rightarrow \text{Negative}$	False Negative (FN)	True Negative (TN)

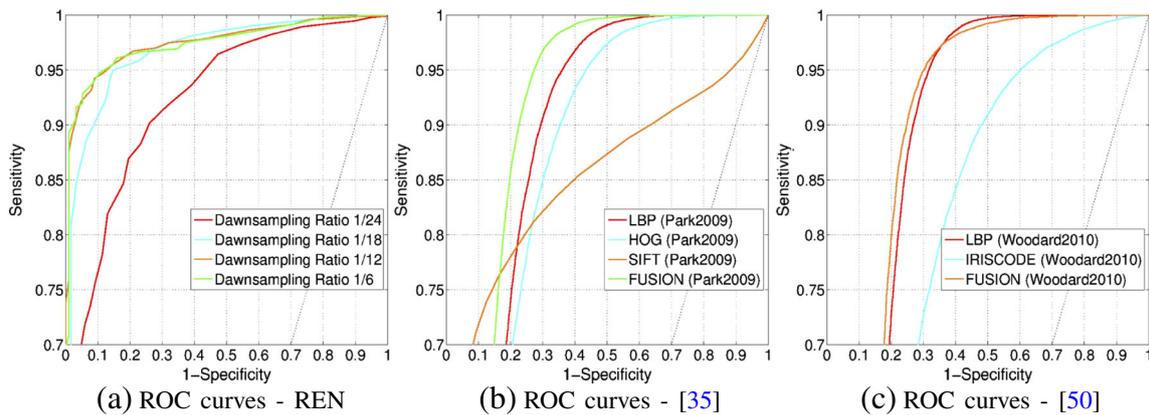


Figure 5 ROC curves for periocular recognition using the FRGC data set. **a** ROC curves for the proposed REN model with different downsampling ratios. **b-c** ROC curves of the models implemented by Park et al. [35] and Woodard et al. [50], respectively.

as probe (per subject). Hence, 100 dictionaries with dimension 90×750 are considered, each one tested in 150 probe samples.

Results are summarized in Table 3 in terms of true and false positive rates where the best sensitivity (sens.) and corresponding accuracy (acc.) for far ($=1$ -specificity) $\leq 10\%$ have been computed for various schemes and models studied here. The proposed reweighed elastic net demonstrates to be superior than the original SRC approach over grayscale impulses. In this case the area under the curve (AUC) and the equal error rate (EER) are equal to 0.9643 and 0.0904 for our model, against 0.9307 and 0.1529 produced by the original SRC model. The proposed models approximates more to the *optimal performance* point (complement of specificity = 0, sensitivity = 1). For the REN approach applied to the grayscale and the texture components alone the minimal distance from the ROC values to the (0, 1) point was of 0.1511 and 0.0812 respectively, while the value 0.1805 was observed for the classical SRC model. In relation to other

image representation components, the minimal distance from the ROC values to the (0, 1) point was of 0.0022, 0.0106, 0.0023, 0.0019 for the CT, CB, RGB and HSV spaces.

Comparisons have been carried out by implementing the models of Park et al. [35] and Bharadwaj et al. [2]. Even though both models make full use of local and global periocular information to perform recognition, they have shown not to improve better than our approach. In our experiments, a comparative analysis between the REN model with texture information and the works of Park et al. [35] and Bharadwaj et al. [2] was performed with respect to AUC and EER. The highest AUC for [35] and [2] is equal to 0.9564 and lowest EER is equal to 0.0954 when applying their fusion techniques. Meanwhile, using the texture information provided by the cartoon - texture space, our model got the values 0.9756 and 0.0589 for the AUC and EER, respectively. Our fusion method using different spaces completely describing the geometry and color periocular feature have also shown to reach great statistical values in comparison to

Table 2 AUC and EER values, as well as the best sensitivity for far $\leq 10\%$ using FRGC’s periocular images. The underline fonts indicate the best model observed in terms of AUC and EER.

Mehtod	Feature	Ratio	Sens.	Far	Acc.	Thres.	AUC	EER
REN (Proposed)	Grayscale (SRC)	1/24	75.65	9.23	78.09	0.2200	0.9218	0.1536
		1/18	90.19	9.01	90.27	0.2100	0.9673	0.09814
		1/12	94.24	9.18	93.95	0.1800	<u>0.9748</u>	<u>0.0682</u>
		1/6	93.79	8.42	93.60	0.1600	0.9741	0.0717
Park et al. [35]	LBP	–	70.61	10.00	83.54	0.7904	0.8534	0.2173
	HOG	–	64.59	10.00	81.53	0.6530	0.8375	0.2386
	SIFT	–	71.08	9.20	83.17	0.1731	0.8496	0.2138
	Fusion	–	77.39	10.00	85.80	0.2908	0.8777	0.1841
Woodard et al. [50]	GIST	–	73.13	10.00	84.37	0.7505	0.8392	0.2170
	Iriscode	–	52.02	10.00	77.34	0.4908	0.7703	0.2900
	Fusion	–	75.20	10.00	85.07	0.6092	0.8450	0.2012

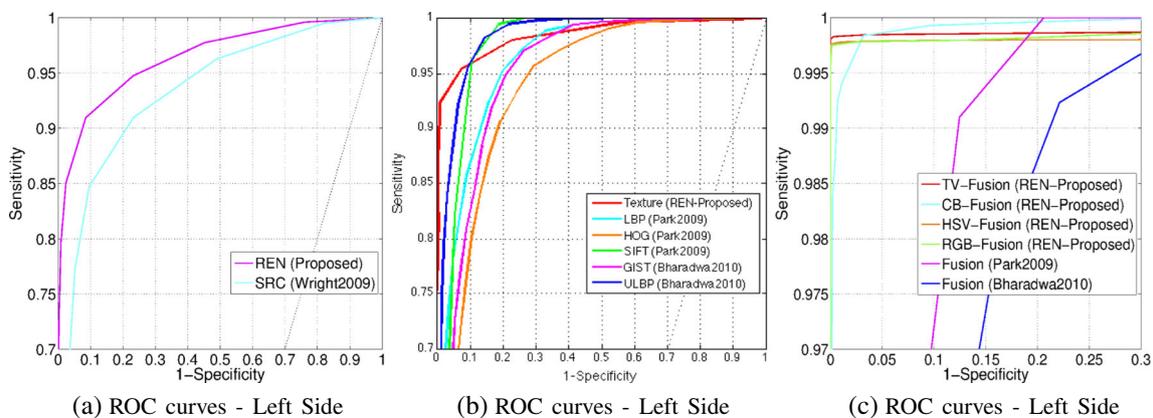
Table 3 AUC and EER values, as well as the best sensitivity for far $\leq 10\%$ using the UBIRIS.v2 database. The underline fonts indicate the best model observed in terms of the AUC and EER.

Method	Feature	sens.	far	acc.	thres.	AUC	EER
REN (Proposed)	Grayscale (SRC)	90.05	8.55	90.99	0.1553	0.9643	0.0904
	Texture (SRC)	92.10	1.89	92.40	0.0756	0.9756	0.0589
	CT (Fusion)	99.90	7.18	98.77	0.1641	<u>0.9994</u>	<u>0.0018</u>
	CB (Fusion)	99.82	7.37	98.49	0.2333	0.9992	0.0061
	RGB (Fusion)	99.83	4.11	99.31	0.1670	0.9990	0.0020
	HSV (Fusion)	99.83	2.13	99.57	0.1832	0.9991	0.0019
Wright et al. [51]	Grayscale (SRC)	84.70	9.59	85.14	0.05642	0.9307	0.1529
	LBP	80.70	9.99	86.90	0.7468	0.9189	0.1553
Park et al. [35]	HOG	69.29	9.99	83.11	0.6421	0.8656	0.2088
	SIFT	86.00	9.36	88.96	0.0477	0.9453	0.1232
	Fusion	90.58	9.99	90.21	0.1052	0.9564	0.0954
	LBP	75.56	9.99	85.20	0.7623	0.8927	0.1846
Bharadwaj et al. [2]	ULBP	85.82	9.99	88.61	0.8673	0.9259	0.1311
	Fusion	83.96	9.99	88.00	0.8008	0.9235	0.1386

those values attained by Park et al. [35] and Bharadwaj et al. [2]. In this case, the highest AUC and the lowest EER values are given by the CT space with values 0.9994 and 0.0018, see Fig. 6.

As it can be observed from Table 4, the proposed REN model applied to grayscale in texture setting is highly correlated when compared to the signals recovered in the CT, RGB and HSV spaces. The result is due to the high accuracy rates achieved over these image representations. Similarly, the signals recovered in the CB space are in low correlation with the signals lying in the grayscale setting

and the CT space, and in high correlation with the signals computed over texture domain alone. The chromaticity components lying in the unit sphere S^2 are the primary reason for these results since they have the advantage of depicting nonlinear features in different directions and therefore both strong and weak edges are distributed and represented along chromaticity components. Also, it should be noted the strong correlation between the outputs given by the fusion model when using exclusively color components. This is explained by the fact that the skin region comprises a large majority of the periocular region (see Fig. 4). It is

**Figure 6** ROC curves for periocular recognition using the UBIRIS.v2 data set. (a) ROC curves for the original REN approach and the SRC model. (b) ROC curves for the REN approach applied to the texture components together with different features extracted by Park et al. [35]

and Bharadwaj et al. [2]. (c) ROC curves for the REN model applied to the proposed fusion over the different geometry and color spaces, as well as the fusion implemented in Park et al. [35] and Bharadwaj et al. [2].

Table 4 Pearson’s sample correlation coefficients between the left side responses given by the recognition algorithms using the REN model and the UBIRIS.v2 data set with various components studied here.

	Grayscale	Texture	CT	CB	RGB	HSV
Grayscale	1	0.7173	0.5331	0.4272	0.7139	0.6230
Texture	–	1	0.6041	0.6285	0.9776	0.9043
	–	–	1	0.2206	0.6134	0.6146
CB	–	–	–	1	0.6146	0.7213
RGB	–	–	–	–	1	0.9180
HSV	–	–	–	–	–	1

particularly interesting to observe that the positive (and small) correlation values between the signals are obtained when using different color spaces representation, pointing for a complementarity that might contribute for the out-performing results of the method proposed in this paper. Although the CT space produces good recognition rates, its computed signals are in low correlation with respect to other signals over different domains, owing to the fact that CT space is given by geometric information in case of cartoon component, whereas weak and strong edges describe texture components, see Fig. 3.

5 Conclusions

This paper describes a novel re-weighted elastic net (REN) model that improves the sparsity of representations in periocular regions, which is an emerging biometric trait with high potential to handle data acquired under uncontrolled conditions. From this perspective, we have fused multiple sparse representations, associated with various spaces from different domains in geometry and color, which allow us to faithfully handle distortions in periocular images such as blur and occlusions. Our experiments were carried out in the highly challenging images of the FRGC and UBIRIS.v2 and data sets, and allowed us to observe consistent improvements in performance, when compared to the classical sparse representation model, and state-of-the-art periocular recognition algorithms. In addition, theoretical existence results have been proved for the REN minimization problem, mainly emphasizing our approach is good in the sense it performs as well as if the true underlying model were given in advance. As far as numerical approximation is concerned, the REN model is expressed as a quadratic programming (QP) expediting the implementation of the proposed gradient projection (GP) algorithm and

providing good results. Evaluating the proposed methodology on other biometric traits (iris, face, etc.) and on different databases are our future works.

Appendix A: Existence of Solution

We state necessary and sufficient conditions for the existence of a solution for the proposed model (7). We follow the notations and similar arguments to those used in [19, 48]. Suppose that $A_i = (A_{1i}, \dots, A_{mi})^T, i = 1, \dots, n$ are the linear independent predictors and $\mathbf{y} = (y_1, \dots, y_m)^T$ is the response vector. Let $A = [A_1, \dots, A_n]$ be the predictor matrix. In terms of ℓ^1 and ℓ^2 norms, we rewrite the minimization problem in Eq. (7) as,

$$\min_{\mathbf{x}} \left\{ m \|\mathbf{W}\mathbf{x}\|_1 + \frac{m}{2} \|(1 - \mathbf{W})\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\}. \quad (13)$$

Let us denote by \mathbf{x}^* and $\hat{\mathbf{x}}$ the real and estimated solution of Eq. 13 respectively. Given $\mathcal{I} = \text{supp}(\mathbf{x}^*) = \{i : x_i^* \neq 0\}$, we define the block-wise form matrix

$$A_{\mathcal{I}, \mathcal{I}^c} = \frac{1}{m} \begin{pmatrix} A_{\mathcal{I}}^T A_{\mathcal{I}} & A_{\mathcal{I}}^T A_{\mathcal{I}^c} \\ A_{\mathcal{I}^c}^T A_{\mathcal{I}} & A_{\mathcal{I}^c}^T A_{\mathcal{I}^c} \end{pmatrix},$$

where $A_{\mathcal{I}} (A_{\mathcal{I}^c})$ is a $m \times \#\mathcal{I} (m \times \#\mathcal{I}^c)$ matrix formed by concatenating the columns $\{A_i : i \in \mathcal{I}\} (\{A_i : i \in \mathcal{I}^c\})$ and $A_{\mathcal{I}}^T A_{\mathcal{I}}$ is assumed to be invertible.

First we assume that there exist $\mathbf{v}\mathbf{x} \in \mathbb{R}^n$ satisfying (13) and $\text{sign}(\hat{\mathbf{x}}) = \text{sign}(\mathbf{x}^*)$. Lets define $\mathbf{b} = \mathbf{W}_{\mathcal{I}} \text{sign}(\mathbf{x}_{\mathcal{I}}^*)$ together with the set,

$$\mathcal{D} = \left\{ \mathbf{d} \in \mathbb{R}^n : \begin{cases} d_i = b_i, & \text{for } \hat{x}_i \neq 0 \\ |d_i| \leq w_i, & \text{otherwise} \end{cases} \right\}.$$

From the Kauush-Kuhn-Tucker (KKT) conditions we obtain

$$\begin{cases} A_i^T (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) - m(1 - w_i)^2 \hat{x}_i = m w_i \text{sign}(x_i^*), & \text{if } \hat{x}_i \neq 0 \\ |A_i^T (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})| \leq m w_i, & \text{otherwise} \end{cases}$$

which can be rewritten as,

$$A_i^T A (\hat{\mathbf{x}} - \mathbf{x}^*) - A_i^T \boldsymbol{\kappa} + m(1 - w_i)^2 \hat{x}_i + m d_i = 0, \quad (14)$$

for some $\mathbf{d} \in \mathcal{D}$ with components $d_i, i = 1, \dots, n$. By substituting the equality $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \boldsymbol{\kappa}$. From the above (14) the following two equations arise:

$$A_{\mathcal{I}}^T A_{\mathcal{I}}(\hat{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}^*) - \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} + (1 - W)^2 \hat{\mathbf{x}}_{\mathcal{I}} = -\mathbf{b}, \tag{15}$$

$$A_{\mathcal{I}^c}^T A_{\mathcal{I}}(\hat{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}^*) - \frac{A_{\mathcal{I}^c}^T(\boldsymbol{\kappa})}{m} = -\mathbf{d}_{\mathcal{I}^c}. \tag{16}$$

Solving for $\mathbf{x}_{\mathcal{I}}$ in Eq. 15 and replacing in Eq. 16 to get \mathbf{b} in terms of $\mathbf{x}_{\mathcal{I}}$ leave us with

$$\hat{\mathbf{x}}_{\mathcal{I}} = \left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1 - W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}^* + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right), \tag{17}$$

$$A_{\mathcal{I}^c}^T A_{\mathcal{I}} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1 - W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}^* + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) - \mathbf{x}_{\mathcal{I}}^* \right) - \frac{A_{\mathcal{I}^c}^T \boldsymbol{\kappa}}{m} = -\mathbf{b}. \tag{18}$$

From Eqs. 17 and 18, we finally get the next two equations:

$$\text{sign} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1 - W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}^* + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) \right) = \text{sign}(\mathbf{x}_{\mathcal{I}}^*) \tag{19}$$

and

$$\left| A_{\mathcal{I}^c}^T A_{\mathcal{I}} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1 - W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}^* + \frac{A_{\mathcal{I}}^T(\boldsymbol{\kappa})}{m} - \mathbf{b} \right) - \mathbf{x}_{\mathcal{I}}^* \right) - \frac{A_{\mathcal{I}^c}^T \boldsymbol{\kappa}}{m} \right| \leq w_i, \tag{20}$$

for $i \in \mathcal{I}^c$.

Now, let us assume that Eqs. 19 and 20 both hold. It will be proved there exist $\hat{\mathbf{x}} \in \mathbb{R}^n$ satisfying $\text{sing}(\hat{\mathbf{x}}) = \text{sign}(\mathbf{x}^*)$. Setting $\hat{\mathbf{x}} \in \mathbb{R}^n$ satisfying $\hat{\mathbf{x}}_{\mathcal{I}^c} = \mathbf{x}_{\mathcal{I}^c}^* = 0$ and

$$\mathbf{x}_{\mathcal{I}} = \left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1 - W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}^* + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right),$$

which guarantees the equality $\text{sign}(\hat{\mathbf{x}}_{\mathcal{I}}) = \text{sign}(\mathbf{x}_{\mathcal{I}}^*)$ due to Eq. 19. In the same manner, we define $\mathbf{d} \in \mathbb{R}^n$ satisfying $\mathbf{d}_{\mathcal{I}} = \mathbf{b}$ and

$$\mathbf{d}_{\mathcal{I}^c} = - \left(A_{\mathcal{I}^c}^T A_{\mathcal{I}} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1 - W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}^* + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) - \mathbf{x}_{\mathcal{I}}^* \right) - \frac{A_{\mathcal{I}^c}^T \boldsymbol{\kappa}}{m} \right),$$

implying from Eq. 20 the inequality $|d_i| \leq w_i$ for $i \in \mathcal{I}^c$ and therefore $\mathbf{d} \in \mathcal{D}$. From previous, we have found a point a point $\text{mathbf{x}} \in \mathbb{R}^n$ and $\mathbf{d} \in \mathcal{D}$ satisfying (15) and (16) respectively or equivalently (14). Moreover, we also have the equality $\text{sign}(\hat{\mathbf{x}}) = \text{sign}(\mathbf{x}^*)$. Under these assertions we can prove the sign recovery property of our model as illustrated next.

Appendix B: Sign Recovery Property

Under some regularity conditions on the proposed REN model, we intend to give an estimation for which the event $\text{sign}(\hat{\mathbf{x}}) = \text{sign}(\mathbf{x}^*)$ is true. Following similar notations and arguments to those used in [53, 55], we intend to prove that our model enjoys the following probabilistic property:

$$\Pr \left(\min_{i \in \mathcal{I}} |\hat{x}_i| > 0 \right) \rightarrow 1. \tag{21}$$

For theoretical analysis purposes, the problem (7) is written as

$$\min_{\mathbf{x}} \left\{ \|W\mathbf{x}\|_1 + \|(1 - W)\mathbf{x}\|_2^2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\}.$$

The following regularity conditions are also assumed:

1. Denoting with $\Lambda_{\min}(S)$ and $\Lambda_{\max}(S)$ the minimum and maximum eigenvalues of a symmetric matrix S , we assume the following inequalities hold:

$$\theta_1 \leq \Lambda_{\min} \left(\frac{1}{m} A^T A \right) \leq \Lambda_{\max} \left(\frac{1}{m} A^T A \right) \leq \theta_2,$$

where θ_1 and θ_2 are two positive constants.

2. $\lim_{m \rightarrow \infty} \frac{\log(n)}{\log(m)} = \nu$ for some $0 \leq \nu < 1$
3. $\lim_{m \rightarrow \infty} \sqrt{\frac{m}{n}} \frac{1}{\max_{i \in \mathcal{I}} w_i} = \infty$.

Let

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|(1 - W)\mathbf{x}\|_2^2 \right\}. \tag{22}$$

By using the definitions of $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$, the next two inequalities arise

$$\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 + \|(1 - W)\hat{\mathbf{x}}\|_2^2 \geq \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2^2 + \|(1 - W)\tilde{\mathbf{x}}\|_2^2 \tag{23}$$

and

$$\begin{aligned} & \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2^2 + \|(1 - W)\tilde{\mathbf{x}}\|_2^2 + \sum_{i=1}^n w_i |\tilde{x}_i| \\ & \geq \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 + \|(1 - W)\hat{\mathbf{x}}\|_2^2 + \sum_{i=1}^n w_i |\hat{x}_i|. \end{aligned} \tag{24}$$

The combination of Eqs. 23 and 24 give

$$\begin{aligned} \sum_{i=1}^n w_i (|\tilde{x}_i| - |\hat{x}_i|) &\geq \|y - A\hat{\mathbf{x}}\|_2^2 + \|(1 - W)\tilde{\mathbf{x}}\|_2^2 - \|y - A\tilde{\mathbf{x}}\|_2^2 - \|(1 - W)\tilde{\mathbf{x}}\|_2^2 \\ &= (\hat{\mathbf{x}} - \tilde{\mathbf{x}})^T (A^T A + (1 - W)^2) (\hat{\mathbf{x}} - \tilde{\mathbf{x}}) \end{aligned} \tag{25}$$

On the other hand

$$\sum_{i=1}^n w_i (|\tilde{x}_i| - |\hat{x}_i|) \leq \sum_{i=1}^n w_i |\tilde{x}_i - \hat{x}_i| \leq \sqrt{\sum_{i=1}^n w_i^2} \|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2 \tag{26}$$

By combining Eqs. 25 and 26 we get

$$\begin{aligned} \Lambda_{\min} \left((A^T A) + (1 - W)^2 \right) \|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2 &\leq (\hat{\mathbf{x}} - \tilde{\mathbf{x}})^T (A^T A + (1 - W)^2) (\hat{\mathbf{x}} - \tilde{\mathbf{x}}) \\ &\leq \sqrt{\sum_{i=1}^n w_i^2} \|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2 \end{aligned}$$

which together with the identity

$$0 \leq \theta_1 \leq \Lambda_{\min} (A^T A) \leq \Lambda_{\min} \left((A^T A) + (1 - W)^2 \right)$$

allow us to prove

$$\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2 \leq \frac{\sqrt{\sum_{i=1}^n w_i^2}}{\Lambda_{\min} (A^T A)}, \tag{27}$$

Let us notice that

$$\begin{aligned} E \left(\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2^2 \right) &= E \left(- \left(A^T A + (1 - W)^2 \right)^{-1} (1 - W)^2 \mathbf{x}^* \right. \\ &\quad \left. + \left(A^T A + (1 - W)^2 \right)^{-1} A^T \kappa \right) \\ &\leq 2 \frac{\|(1 - W)\mathbf{x}^*\|_2^2 + n\Lambda_{\max} (A^T A) \sigma^2}{\Lambda_{\min} (A^T A)} \end{aligned} \tag{28}$$

From Eqs. 27 and 28 we conclude that

$$\begin{aligned} E \left(\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 \right) &\leq 2 \left(E \left(\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2^2 \right) - E \left(\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 \right) \right) \\ &\leq 4 \frac{\|(1 - W)\mathbf{x}^*\|_2^2 + n\Lambda_{\max} (A^T A) \sigma^2 + E \left(\sum_{i=1}^n w_i^2 \right)}{\Lambda_{\min} (A^T A)}. \end{aligned} \tag{29}$$

Let $\eta = \min_{i \in \mathcal{I}} |x_i^*|$ and $\hat{\eta} = \max_{i \in \mathcal{I}} w_i$. Because of Eq. 27,

$$\|\hat{\mathbf{x}}_{\mathcal{I}} - \tilde{\mathbf{x}}_{\mathcal{I}}\|_2^2 \leq \frac{\sqrt{n}\hat{\eta}}{\theta_1 m}.$$

Then

$$\min_{i \in \mathcal{I}} |x_i^*| > \min_{i \in \mathcal{I}} |\tilde{x}_i| - \frac{\sqrt{n}\hat{\eta}}{\theta_1 m} > \min_{i \in \mathcal{I}} |\hat{x}_i| - \|\tilde{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}_{\mathcal{I}}^*\|_2 - \frac{\sqrt{n}\hat{\eta}}{\theta_1 m}. \tag{30}$$

Now, we notice that

$$\frac{\sqrt{n}\hat{\eta}}{\theta_1 m} = O \left(\frac{1}{\sqrt{n}} \right) \left(\sqrt{\frac{n}{m}} \eta^{-1} \right) (\hat{\eta} \eta).$$

Since

$$\begin{aligned} E \left((\hat{\eta} \eta)^2 \right) &\leq 2\eta^2 + 2\eta^2 E \left((\hat{\eta} - \eta)^2 \right) \leq 2\eta^2 + 2\eta^2 E \left(\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 \right) \\ &\leq 2\eta^2 + 8\eta^2 \frac{\|(1 - W)\mathbf{x}^*\|_2^2 + \theta_2 n m \sigma^2 + E \left(\sum_{i=1}^n w_i^2 \right)}{\theta_1 m} \end{aligned}$$

and $\eta^2 m/n \rightarrow \infty$ as long as $m \rightarrow \infty$, it follows that

$$\frac{\sqrt{n}\hat{\eta}^{-1}}{\theta_1 m} = o \left(\frac{1}{\sqrt{n}} \right) O_{Pr}(1). \tag{31}$$

By using Eq. 29, we derive

$$E \left(\|\hat{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}_{\mathcal{I}}^*\|_2^2 \right) \leq 4 \frac{\|(1 - W)\mathbf{x}^*\|_2^2 + \theta_2 n m \sigma^2}{(\theta_1 m)^2} = \sqrt{\frac{n}{m}} O_{Pr}(1). \tag{32}$$

Substituting Eq. 31 and 32 in Eq. 30 allow us to conclude that

$$\min_{i \in \mathcal{I}} |x_i^*| > \eta - \sqrt{\frac{n}{m}} O_{Pr}(1) - o \left(\frac{1}{\sqrt{n}} \right) O_{Pr}(1).$$

Then Eq. 21 holds.

Remark 2 There is special interest in applying the REN model in the case the data satisfies the condition $n \gg m$. For the LASSO model it was suggested in [6] to make use of the Dantzig selector which can achieve the ideal estimation up to a $\log(n)$ factor. In [13] a performing of the Dantzig selector called the Sure Independence Screening (SIS) was introduced in order to reduce the ultra-high dimensionality. We remark that the SIS technique can be combined with the REN model (7) for dealing the case $n \gg m$. Then previous computations can be still applied to reach the sign recovery property.

References

- Adams, J., Woodard, D.L., Dozier, G., Miller, P., Bryant, K., & Glenn, G. (2010). Genetic-based type ii feature extraction for periorcular biometric recognition: Less is more. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 205–208).
- Bharadwaj, S., Bhatt, H.S., Vatsa, M., & Singh, R. (2010). Periocular biometrics: When iris recognition fails. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)* (pp. 1–6).
- Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J., & Osher, S. (2007). Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2), 151–167.
- Candès, E., Romberg, J., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8), 1207–1223.
- Candès, E., & Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12), 4203–4215.
- Candès, E., & Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n . *The Annals of Statistics*, 35(6), 2392–2404.

7. Candès, E., Wakin, M., & Boyd, S.P. (2008). Enhancing sparsity by reweighted ℓ^1 -minimization. *Journal of Fourier Analysis and Applications*, 14(5), 877–905.
8. Chartrand, R., & Yin, W. (2008). Iteratively reweighted algorithms for compressive sensing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 3869–3872).
9. Chen, S., Donoho, D., & Saunders, M. (1998). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1), 33–61.
10. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 886–893).
11. Daugman, J. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11), 1148–1161.
12. Donoho, D. (2006). For most large underdetermined systems of equations, the minimal ℓ^1 -norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, 59(7), 907–934.
13. Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849–911.
14. Fazel, M., Hindi, H., & Boyd, S. (2003). Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In *Proceedings of the American Control Conference* (pp. 2156–2162).
15. Figueiredo, M., Nowak, R., & Wright, S. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problem. *IEEE Journal of Selected Topics in Signal Processing*, 1(4), 586–597.
16. Fuchs, J.J. (1999). Multipath time-delay detection and estimation. *IEEE Transactions on Signal Processing*, 47(1), 237–243.
17. Hong, D., & Zhang, F. (2010). Weighted elastic net model for mass spectrometry image processing. *Mathematical Modelling of Natural Phenomena*, 5(3), 115–133.
18. Jain, A.K., Flynn, P., & Ross, A. (Eds.) (2007). *Handbook of biometrics*. New York, USA: Springer-Verlag.
19. Jia, J., & Yu, B. (2010). On model selection consistency of the elastic net when $p \gg n$. *Statistica Sinica*, 20, 595–611.
20. Jiang, R., Crookes, D., & Lie, N. (2010). Face recognition in global harmonic subspace. *IEEE Transactions on Information Forensics and Security*, 5(3), 416–424.
21. Juefei-Xu, F., Cha, M., Heyman, J.L., Venugopalan, S., Abiantun, R., & Savvides, M. (2010). Robust local binary pattern feature sets for periocular biometric identification. In *IEEE International Conference on Biometrics: Theory Applications and Systems* (pp. 1–8).
22. Juefei-Xu, F., Luu, K., Savvides, M., Bui, T.D., & Suen, C.Y. (2011). Investigating age invariant face recognition based on periocular biometrics. In *International Joint Conference on Biometrics* (pp. 1–7).
23. Kang, S., & March, R. (2007). Variational models for image colorization via chromaticity and brightness decomposition. *IEEE Transactions on Image Processing*, 16(9), 2251–2261.
24. Lange, K. (2004). *Optimization*. Springer Text in Statistics. New York: Springer.
25. Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
26. Miller, P.E., Rawls, A.W., Pundlik, S.J., & Woodard, D.L. (2010). Personal identification using periocular skin texture. In *ACM Symposium on Applied Computing, SAC'10* (pp. 1496–1500).
27. Moreno, J.C. (2014). Texture image segmentation by weighted image gradient norm terms based on local histogram and active contours. In Di Giambardino, P., Iacoviello, D., Jorge, R.N., & Tavares, J.M.R.S. (Eds.) *Computational Modeling of Objects Presented in Images* (pp. 225–243). New York: Springer.
28. Moreno, J.C., Prasath, V.B.S., & Proença, H. (2013). Robust periocular recognition by fusing local to holistic sparse representations. In *Sixth International Conference on Security of Information and Networks* (pp. 160–164). Aksaray, Turkey: Proceedings ACM Digital Library.
29. Moreno, J.C., Prasath, V.B.S., Vorotnikov, D., & Proença, H. (2013). Adaptive diffusion constrained total variation scheme with application to cartoon + texture + edge image decomposition. Technical Report 1354. Portugal: University of Coimbra.
30. Ojala, T., Pietikainen, M., & Harwood, D. (1994). Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In *International Conference on Pattern Recognition (ICPR)*, (Vol. 1 pp. 582–585).
31. Ojala, T., Pietikainen, M., & Maenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987.
32. Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145–175.
33. Park, U., & Jain, A.K. (2010). Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3), 406–415.
34. Park, U., Jillela, R.R., Ross, A., & Jain, A.K. (2011). Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1), 96–106.
35. Park, U., Ross, A., & Jain, A.K. (2009). Periocular biometrics in the visible spectrum: A feasibility study. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems* (pp. 153–158).
36. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview on the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 947–954).
37. Pillai, J.K., Patel, V.M., Chellappa, R., & Ratha, N.K. (2011). Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9), 1877–1893.
38. Prasath, V.B.S., Palaniappan, K., & Seetharaman, G. (2012). Multichannel texture image segmentation using weighted feature fitting based variational active contours. In *Eighth Indian Conference on Vision, Graphics and Image Processing (ICVGIP)* (p. 6).
39. Proença, H., & Alexandre, L. (2012). Toward covert iris biometric recognition: experimental results from the NICE contests. *IEEE Transactions on Information Forensics and Security*, 7(2), 798–808.
40. Proença, H., Filipe, S., Santos, R., Oliveira, J., & Alexandre, L.A. (2010). The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1529–1535.
41. Rudin, L., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1–4), 259–268.
42. Santos, G., & Proença, H. (2013). Periocular biometrics: An emerging technology for unconstrained scenarios. In *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)* (pp. 14–21).
43. Serafini, T., Zanghirati, G., & Zanni, L. (2003). Gradient projection methods for large quadratic programs and applications

in training support vector machines. *Optimization Methods and Software*, 20(2-3), 353–378.

44. Shekhar, S., Patel, V.M., Nasrabadi, N.M., & Chellappa, R. (2013). Joint sparse representation for robust multimodal biometrics recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1), 113–126.
45. Sznitman, R., & Jedynek, B. (2010). Active testing for face detection and localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10), 1914–1920.
46. Tang, B., Sapiro, G., & Caselles, V. (2001). Color image enhancement via chromaticity diffusion. *IEEE Transactions on Image Processing*, 10(5), 701–707.
47. Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society B*, 58(1), 267–288.
48. Wainwright, M. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ^1 -constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory*, 55(5), 2183–2202.
49. Wipf, D., & Nagarajan, S. (2010). Iterative reweighted ℓ^1 and ℓ^2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*, 4(2), 317–329.
50. Woodard, D.L., Pundlik, S., Miller, P., Jillela, R., & Ross, A. (2010). On the fusion of periocular and iris biometrics in non-ideal imagery. In *IEEE International Conference on Pattern Recognition* (pp. 201–204). Istanbul, Turkey.
51. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 210–227.
52. Wyszecki, G., & Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulas*. New York: Wiley.
53. Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.
54. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301–320.
55. Zou, H., & Zhang, H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, 37(4), 1733–175.



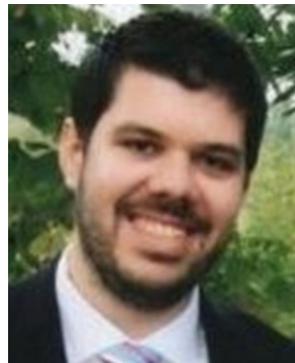
Juan C. Moreno is a research fellow at the Soft Computing and Image Analysis Lab, part of the Pattern and Image Analysis Group of the IT - Instituto de Telecomunicações, at the University of Beira Interior in Covilhã, Portugal. He was also an Invited Assistant Professor at the Computer Science Department at the University of Beira Interior. He received his B.Sc. and Master's Degree in Mathematics from the Central University of Venezuela (2004) and from the

Simón Bolívar University (2008), respectively. In 2013 he completed his PhD in Mathematics at the University of Coimbra. His interest revolve around mathematical based methods for image processing, computer vision and pattern recognition problems.



V. B. Surya Prasath is a research fellow in the Computer Science Department at the University of Missouri, USA. He received his PhD in Mathematics from the Indian Institute of Technology Madras, India in 2010. He has been a postdoctoral fellow at the Department of Mathematics, University of Coimbra, Portugal, for two years. Since 2012 he is with the Computational Imaging and Visualization Analysis Lab at the University of

Missouri, USA working on various mathematical image processing and computer vision problems. He had summer fellowships/visits at Kitware Inc. NY, USA, The Fields Institute, Canada, and IPAM, University of California Los Angeles (UCLA), USA. His main research interests include nonlinear PDEs, regularization methods, inverse & ill-posed problems, variational, PDE based image processing, and computer vision with applications in remote sensing, medical imaging, biometrics domains.



Gil Santos received the B.Sc. degree in 2007, the M.Sc. degree in 2009, and more recently the Ph.D. degree from the University of Beira Interior, in 2015, all in the field of Computer Science and Engineering. His research interest are focused on unconstrained biometrics, pattern recognition and signal processing. He is currently a researcher at the Soft Computing and Image Analysis Lab, part of the Pattern and Image Analysis Group of the IT - Instituto

de Telecomunicações. He is also a member of IEEE and the IEEE Biometric Council.



Hugo Proença received the B.Sc. degree from the University of Beira Interior, Portugal, in 2001, the M.Sc. degree from the Faculty of Engineering, University of Oporto, in 2004, and the Ph.D. degree from the University of Beira Interior, in 2007. His research interests are focused in the artificial intelligence, pattern recognition and biometrics. Currently, he serves an Associate Professor with the Department of Computer Science, University of Beira

Interior. He is the area editor (ocular biometrics) of the IEEE Biometrics Compendium Journal and member of the Editorial Board of the International Journal of Biometrics. Also, he served as Guest Editor of special issues of the Pattern Recognition Letters, Image and Vision Computing and Signal, Image and Video Processing journals.