MANCHESTER
1824

The University of Manchester

Face Recognition Based on Texture Discrimination by Using Geodesic Distance Approximations Between Multivariate Normal Distributions

Soldera, J and Dodson, CTJ and Scharcanski, J

2018

MIMS EPrint: **2018.15**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Face Recognition Based on Texture Information and Geodesic Distance Approximations Between Multivariate Normal Distributions

John Soldera[1,2], CTJ Dodson[3], Jacob Scharcanski[1,4]

[1] Graduate Programme on Electrical Engineering, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[2] Federal Institute of Education, Science and Technology Farroupilha, Santo Ângelo, Brazil
[3] School of Mathematics, University of Manchester, Manchester, UK
[4] Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

E-mail: jsoldera@inf.ufrgs.br, ctdodson@manchester.ac.uk, jacobs@inf.ufrgs.br

Abstract. Geodesic distances are a natural dissimilarity measure between probability distributions of a fixed type, and are used to discriminate texture in several image-based measurements. Besides, since there is no known closed-form solution for the geodesic distance between general multivariate normal distributions, we propose two efficient approximations to discriminate textures in the context of face recognition. Unlike the typical appearance-based approach that uses low-resolution grayscale face images, we propose a novel generative approach for face recognition based on texture discrimination. In the proposed approach, sparse facial features are extracted from high-resolution color face images using predefined landmark topologies, in which landmarks are in discriminative locations of face images. By adopting a common landmark topology, the dissimilarity between distinct face images can be scored in terms of the dissimilarities between the texture in their corresponding landmark vicinities. The proposed multivariate normal distributions represent the color intensities around each landmark location. The classification of new face samples occurs by determining the face image sample in the training set which minimizes the dissimilarity score. The proposed face recognition method was compared to methods representative of the state-of-the-art using color and grayscale face images, and presented higher recognition rates. Moreover, the proposed measures to discriminate textures tend to be efficient in face recognition and in general texture discrimination (e.g., texture recognition of material images), as our experiments suggest.

Keywords: geodesic distance, approximations, multivariate normal distribution, face recognition, sparse features.

1. Introduction

The instrumentation and measurement fields are associated to measure, detect, record and monitor a certain phenomenon (i.e., a measurand) in applications that usually involve uncertainty and/or probability distributions. Some measurands are invisible like the electromagnetic field, and other are visible like the light reflection on a surface.

In order to measure signals in the visible light spectrum, imaging sensors (i.e., cameras) are commonly used to record images or videos which tend to present higher resolutions due to the technological advancement. In such recorded data, colors are usually represented as basic color intensity combinations (i.e., red, green and blue), leading to an inherent high-dimensional multivariate feature representation.

Moreover, the image processing and computer vision fields may be used to help to extract reliable features for several instrumentation-related applications which use texture information, such as face recognition [1] [2] [3] [4], brain image recognition [5] [6], texture recognition of material images [7], food image recognition [8] [9], character recognition [10] [11], yawning detection [12], etc. In this work, we are mainly interested in face recognition by using efficient texture dissimilarity metrics based on geodesic distance approximations between probability distributions.

Face recognition is an instrumentation-related application which uses computer vision and pattern recognition techniques to identify individuals. Moreover, there are several emerging applications based in face recognition in augmented reality, gaming, security, and so on [3] [4] [13] [14]. Face recognition is also studied by neuroscientists and psychologists to provide useful insights in how the human brain works [15]. In such applications, features extracted from images or videos present high dimensionality and the sample availability for machine learning is scarce, potentially leading to the known *curse of dimensionality* [16].

In order to compact face features while discriminative characteristics of the original data are preserved, the *Eigenfaces* method [17] creates a linear projection to a lower dimensional space, where new face samples are recognized. It uses Principal Component Analysis (PCA) [18] to create a linear orthogonal projection which preserves most of the face data variability.

However, *Eigenfaces* not always leads to a reliable face class separability since it only preserves the global structure of the data. On other hand, the *Laplacianfaces* method [19] tries to preserve the local structure of the data in the lower dimensional space by creating a locality graph. However, the underlying class structure may be distorted since the final linear projection is non-orthogonal.

The Orthogonal Locality Preserving Projections method (OLPP) [20] was proposed as an extension to the *Laplacianfaces* method, adding the orthogonality property to the resulting linear transformation, causing a better preservation of the underlying class structure. Similarly, another method named Orthogonal Neighborhood Preserving Projections (ONPP) [21] tries to preserve the local and the global face data geometry by creating a neighborhood graph, which is used to determine the final orthogonal linear transformation.

However, preserving the face data structure in the lower dimensional space not always leads to a good face class separation. On the other hand, if the class labels of each training sample are previously known, it is possible to preserve better the class structure by using supervised dimensionality reduction approaches like Linear Discriminant Analysis (LDA) [22], which determines a linear projection (*Fisherfaces*) that moves samples from different classes away while approximates samples in the same classes in the lower dimensional space.

There are methods based in LDA, i.e., the Discriminative Orthogonal Neighborhood-Preserving Projection (DONPP) [23], which analyzes and preserves the intra-class and inter-class geometry in the lower dimensional space. Another LDA-based method is the *Multi-view Discriminant Analysis* (MvDA) [24], which creates projections of input face features in different perspectives and combine them obtaining the final linear transformation.

However, LDA may present inaccuracies for non-linear separation problems. Therefore, some LDA kernel extensions [25] [26] [27] were proposed to map the input data to a higher dimensional space through a non-linear mapping, where the inner product in this space can be computed by a kernel function without knowing the non-linear mapping explicitly [28]. The Spectral Regression Kernel Discriminant Analysis method (SRKDA) uses regularization techniques to provide an efficient computation of kernel LDA using large datasets [27], obtaining as result a linear projection that preserves the original non-linear class

structure.

Other methods try to learn non-linear transformations in order to preserve better the non-linear data structure of high dimensional data, and that is the case of the Isomap method [29], which creates a neighborhood graph in which the data manifold is approximated by calculating approximations for geodesic distances by determining the shortest path between samples. The final low dimensional representation of the original samples is obtained by the typical multidimensional scaling algorithm (MDS) [30].

The non-linear dimensionality reduction method called Locally Linear Embedding (LLE) [31] tries to model data samples as linear combinations of its neighboring samples and uses this information to determine the lower dimensional representation of the original samples, preserving the local data geometry existing in the original high dimensional space.

Although there are available in the literature several techniques to reduce efficiently the face data dimensionality and preserve the underlying face class structure, there are common issues which affect face representation such as variations of illumination, changes in the head pose, change of appearance, and others, demanding a high availability of distinct training samples in order properly to represent the face variability for machine learning as in the appearance-based approach methods [17] [19] [20] [21]. Since these methods concatenate all image pixels to create representative feature vectors, they need to downsample grayscale face images to reduce the computational complexity. The obtained face feature representation still presents high dimensionality and suffers from the aforementioned issues.

On the other hand, it is possible to extract sparse face features directly from high-resolution color face images by using face representations based in landmarks associated to key points on face images at important and discriminative locations, leading to an enhanced face representation [3] [32]. Landmarks can be automatically determined by using approaches like Active Shape Models (ASMs) [33] and several methods have been proposed to extract sparse features from face images using trained ASMs [34] [35].

However, not all extracted facial features are equally relevant from the point of view of face class discrimination. Therefore, some methods have been proposed to estimate the most discriminative landmarks, like the Enhanced ASM [32], where mutual information is used to rank the landmarks in terms of their discrimination capability using color high-resolution face images. In the Enhanced ASM method, face features are represented by Gaussian mixtures and new face samples are recognized by maximizing the class likelihood. However, this classification scheme can be adversely affected by outliers and noisy data.

As result, such statistical face representations demand fewer face samples for machine learning since relevant features are extracted from key points on the face images (e.g. the eyes, eyebrows and nose), preserving discriminative details from high-resolution color images unlike typical appearance-based methods [17] [19] [20] [21], which demand low-resolution grayscale face images.

Another method that extracts features from vicinities of landmark locations is the Customized OLPP (COLPP) [3], in which landmark topologies are used to mark important and discriminative information on face images. The pixels in the landmark vicinities are concatenated to form high dimensional feature vectors which are mapped into a lower dimensional space where the class structure of the original features is preserved. In this discriminative linear space, classification occurs by employing a linear soft margin Support Vector Machine (SVM) [36].

Most aforementioned methods usually extract feature vectors by concatenating pixel information from whole face images or from landmarks on the face, leading to high-dimensional feature vectors which are usually undersampled due to the low availability of training samples (i.e., face images). On other hand, probability distributions can be learnt from the texture in the vicinities of landmarks in high-resolution color face images, leading to more accurate and lower dimensional feature representations [32]. As a consequence, dissimilarities between distinct textures can be obtained as geodesic distances between probability distributions [4] [13].

In information geometry [37], the geodesic distance is defined as the length of the shortest path between probability distributions lying on a Riemannian manifold induced by the Fisher information metric applied to a parametric family of probability distributions [38] [37]. As result, the geodesic distance is a natural dissimilarity metric between probability distributions, and is used to discriminate texture in several image-based-applications, e.g., face recognition [4]. Moreover, the normal distribution is widely used in several applications, however, there is no known closed-form solution to the geodesic distance between general multivariate normal distributions. Therefore, we propose two efficient approximations with applications in face recognition.

Moreover, we propose a novel generative approach for face recognition which uses information geometry techniques [38] [37] to discriminate face textures, in which sparse facial features are extracted from high-resolution color face images by using predefined landmark topologies, unlike the appearance-based

approach, in which low-resolution grayscale face images are used to reduce the computational complexity [17] [19] [20] [21]. By adopting a common landmark topology, the dissimilarity between distinct face images can be scored in terms of the dissimilarities between the texture in their corresponding landmark vicinities, which are obtained by the proposed geodesic distance approximations between multivariate normal distributions which represent the color intensities around each landmark location.

The classification process of new face samples occurs by the determination of the face image sample present in the training set which minimizes the dissimilarity score. The proposed face recognition method was compared to methods representative in the state-of-the-art using color or grayscale face images and presented the higher recognition rates, supporting a trend in which color information is relevant on face recognition [3] [32]. Moreover, the proposed metrics to discriminate texture tend to be efficient in face recognition since they are considered effective in general texture discrimination (e.g., texture recognition of material images) according to an additional set of experiments that we provide in texture recognition, also overcoming state-of-the-art methods.

This paper is organized as follows. Section 2 proposes geodesic distance approximations between multivariate normal distributions to be used as a texture dissimilarity metric in face recognition. Section 3 presents the proposed face recognition method, where section 3.1 discusses how sparse features and probability distributions are obtained from face images, and section 3.2 presents how dissimilarities between distinct face images are scored in terms of the dissimilarities between textures in their corresponding landmark vicinities by using the proposed geodesic distance approximations. The experimental results are presented and discussed in section 4 and the final conclusions and ideas for future works are presented in section 5.

2. Geodesic Distance Approximations Between Multivariate Normal Distributions

In many face recognition methods, face features are represented as vectors [17] [19] [20] [21] [3]. However, those feature representations are highly affected by natural image issues such as variations in illumination, pose and scale. Moreover, usually there aren't enough samples (face images) to properly sample such high-dimensional feature spaces.

On other hand, multivariate probability distributions of color image pixels tend to preserve the original image characteristics in a lower dimensionality representation, which are useful for texture discrimination. Moreover, such feature representations are robust to scale and pose variations. Therefore, we choose to represent image features as multivariate normal distributions which are defined as follows:

$$F(x|\mu,\Sigma) = \frac{e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}{\sqrt{(2\pi)^C|\Sigma|}}, \quad (1)$$

where x is a C-dimensional vector, μ is the C-dimensional mean and Σ is the $C \times C$ covariance matrix, for images with C color channels.

Since geodesic distances are the natural distance measure for families of probability distributions [37], and assuming that the texture in the landmark vicinities is normally distributed, we use geodesic distances between normal distributions in order to measure dissimilarities between the textures of corresponding landmarks of distinct face images.

Considering the case when there are two univariate normal distributions $F_1(x|\mu_1,\sigma_1)$ and $F_2(x|\mu_2,\sigma_2)$, the geodesic distance $G_e(F_1,F_2)$ between both distributions is given in a closed-form [38] by:

$$G_e(F_1,F_2) = \sqrt{2}\ln\frac{1+\delta}{1-\delta} = 2\sqrt{2}\tanh^{-1}\delta, \quad (2)$$

where

$$\delta \equiv \left[\frac{(\mu_1-\mu_2)^2 + 2(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2 + 2(\sigma_1+\sigma_2)^2}\right]^{1/2}. \quad (3)$$

However, for the proposed method, a univariate normal distribution is not suitable since it supports only monochromatic images (i.e., grayscale images). On the other hand, we use color-based feature representations since color features tend to improve image class discrimination [3] [32]. Therefore, multivariate normal distributions are more adequate to represent face image features.

One special case of multivariate normal distribution is when the covariance matrix $\Sigma = \mathrm{diag}(\sigma_1^2,\sigma_2^2,...,\sigma_C^2)$ is a diagonal matrix (i.e., the color channels are independent features). Therefore, the geodesic distance $G_f(F_1,F_2)$ between multivariate normal distributions $F_1(x|\mu_1,\Sigma_1)$ and $F_2(x|\mu_2,\Sigma_2)$ given by [39] for diagonal covariance matrices can be used as a dissimilarity metric:

$$G_f(F_1,F_2) = \sqrt{\sum_{c=1}^{C} G_e(F_1^c,F_2^c)^2}, \quad (4)$$

where $F_1^c = F_1(x(c)|\mu_1(c),\Sigma_1(c,c))$ represents the c-th independent univariate normal distribution with mean $\mu_1(c)$ and variance $\Sigma_1(c,c)$, belonging to the multivariate distribution $F_1(x|\mu_1,\Sigma_1)$.

However, image color channels are usually not statistically independent, and using multivariate normal distributions with diagonal covariance matrices

may discard relevant and discriminative texture information which should be accounted by geodesic distances. Moreover, such distributions are not generally adequate for texture discrimination since they ignore the natural covariances between color channels inherent in the color images.

Therefore, in order to obtain more accurate geodesic distances, we can consider using geodesics for general multivariate normal distributions, where the covariances between color channels are also accounted. Unfortunately, there is no known closed-form solution for this case, but closed-form solutions for two specific multivariate normal distribution subcases are known [38] [39]:

(i) $\mu_1 \neq \mu_2, \Sigma_1 = \Sigma_2$:

$$G_\mu(F_1, F_2) = \sqrt{(\mu_1 - \mu_2)^T (\Sigma_1)^{-1} (\mu_1 - \mu_2)}, \quad (5)$$

(ii) $\mu_1 = \mu_2, \Sigma_1 \neq \Sigma_2$:

$$G_\Sigma(F_1, F_2) = \sqrt{\frac{1}{2}\sum_{j=1}^{C} \log^2(\lambda_j)}, \quad (6)$$

$$\text{with } \{\lambda_j\} = \text{Eig}((\Sigma_1)^{-1/2}\Sigma_2(\Sigma_1)^{-1/2}), \quad (7)$$

where Eig is a function that returns the eigenvalues of a given matrix and j indicates the j-th eigenvalue.

We intend to approximate the geodesic distance for the case of general multivariate normal distributions based on equations (5) and (6), however some adaptations are necessary due to the fact that distinct images often present different means and covariance matrices. As equation (6) does not consider means (μ_1 and μ_2), we can use it without changes since it is independent of the means. However, equation (5) requires a common covariance matrix Σ_1, but we have two different covariance matrices Σ_1 and Σ_2. Therefore, we propose the following two alternatives for computing G_μ for general multivariate normal distributions:

$$G^g_\mu(F_1, F_2) = 0.5\sqrt{(\mu_1 - \mu_2)^T (\Sigma_1)^{-1} (\mu_1 - \mu_2)} + 0.5\sqrt{(\mu_1 - \mu_2)^T (\Sigma_2)^{-1} (\mu_1 - \mu_2)}, \quad (8)$$

and,

$$G^h_\mu(F_1, F_2) = \sqrt{(\mu_1 - \mu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\mu_1 - \mu_2)}, \quad (9)$$

leading to two distinct ways to approximate the geodesic distance for general multivariate normal distributions:

$$G_g(F_1, F_2) = \frac{G^g_\mu(F_1, F_2) + G_\Sigma(F_1, F_2)}{2}, \quad \text{or} \quad (10)$$

$$G_h(F_1, F_2) = \frac{G^h_\mu(F_1, F_2) + G_\Sigma(F_1, F_2)}{2}. \quad (11)$$

Considering that the color channels of face images are statistically independent, we fall into the multivariate case with diagonal covariance matrix (G_f). Otherwise, the general multivariate case provides a more accurate geodesic distance approximation between multivariate normal distributions (G_g or G_h).

Next, we present the proposed approach for face recognition, which is based on the proposed geodesic distance approximations as a texture dissimilarity metric.

3. A Generative Method for Face Representation and Recognition

3.1. Sparse Face Feature Extraction

Typical appearance-based methods [17] [19] [20] [21] exploit the face data variability for machine learning. However, in order to reduce the computational complexity, these methods use low-resolution grayscale face images which are converted to the form of high-dimensional feature vectors. On the other hand, more discriminative features tend to be obtained from high-resolution color face images by extracting information from the texture in the vicinities of key points on the face images (i.e., landmarks) [3]. Therefore, we propose a feature extraction method based on the sparse approach, since this feature representation can be approached as a multivariate classification problem [32] [3].

Assuming a point distribution model to represent color face images, a predefined topology with Q landmarks can be used to represent the facial features at Q face image locations. These Q landmarks may be manually annotated or automatically identified in the face images. However, there is uncertainty about the correct location of manually annotated or automatically identified landmarks due to image artifacts (e.g. head pose, noise, illumination change, etc.). Therefore, given a landmark topology, we can introduce interpolated landmarks between each pair of consecutive landmarks on a face image, improving the reliability of the biometric information. The final landmark topology contains a set of L identified and interpolated landmarks, with $L > Q$. The adopted landmark topology used in this work is presented in figure 1, marking important information from the face images such as the eyes, eyebrows and nose.

Therefore, given a landmark topology with L landmarks (i.e., the landmark topology in figure 1), the texture in the squared vicinities with size $w \times w$ centered in each landmark l are extracted from each face image (i.e., head pose) b of face class a, considering face images with C color channels.

These features are the C-dimensional mean $\mu_{a,b,l} = \mathbb{E}[I_{a,b,l}(m,n)], \forall m,n$ and the $(C \times C)$-

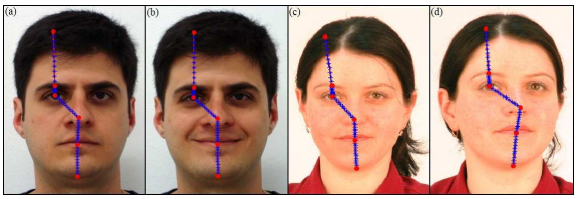

Figure 1. Adopted landmark topology exemplified in the PUT Face Database [40] and the FEI Face Database [41]. Face database landmarks are in red and interpolated landmarks are in blue.

dimensional covariance matrix $\Sigma_{a,b,l} = \mathbb{E}[(I_{a,b,l}(m,n) - \mu_{a,b,l})(I_{a,b,l}(m,n) - \mu_{a,b,l})^T], \forall m,n$, where $I_{a,b,l}(m,n)$ is the C-dimensional color vector representing the pixel with indexes (m,n) in the squared vicinity of the landmark l in the face image b of face class a, with $m,n = 1,2,...,w$.

For instance, considering color (RGB) face images, $\mu_{a,b,l}$ will be a 3-dimensional vector, but for grayscale face images $\mu_{a,b,l}$ will be a 1-dimensional vector. As result, each landmark l in the face image b of class a is represented by the mean $\mu_{a,b,l}$ and the covariance matrix $\Sigma_{a,b,l}$ computed from the vicinity of the same landmark.

Since the landmarks represent discriminative information on the face images, we propose to calculate dissimilarities between distinct face images in terms of the dissimilarities between the texture in their corresponding landmark vicinities by adopting a common landmark topology (i.e., the landmark topology in figure 1) as will be described in section 3.2.

3.2. Face Classification

Since geodesic distances are a natural dissimilarity metric for statistical distributions, we propose to calculate dissimilarities between distinct face images by summing dissimilarities between the texture of their corresponding landmarks which are given as geodesic distances approximations between multivariate normal distributions as presented in section 2. Considering L, the total number of landmarks in a landmark topology, a geodesic distance approximation between multivariate normal distributions can be adopted, i.e., G_f (equation (4)), G_g (equation (10)) or G_h (equation (11)).

Considering color (RGB) face images, the texture in the vicinity of each landmark can be considered as a multivariate normal distribution, since each pixel can be treated as a 3-dimensional sample within the landmark vicinity. Considering that the color channels are independent for each landmark, the geodesic distance approximation G_f for multivariate normal distributions with diagonal covariance matrices provides a suitable geodesic distance metric. In this case, the dissimilarity ${S_f}_{a,b}^{a',b'}$ between the face image (i.e., head pose) b of face class a with the face image b' of face class a' can be scored by using G_f as follows:

$${S_f}_{a,b}^{a',b'} = \sum_{l=1}^{L} G_f(F_{a,b,l}, F_{a',b',l}), \tag{12}$$

where $F_{a,b,l}$ represents a multivariate normal distribution with null covariances for the landmark l in the face image b of face class a with the C-dimensional mean $\mu_{a,b,l}$ and the $(C \times C)$-dimensional covariance matrix $\Sigma_{a,b,l}$. On the other hand, if the multivariate face data present relevant covariances between color channels, one of the proposed geodesic distance approximations for general multivariate normal distributions (G_g or G_h) should be more adequate for the score calculation:

$${S_g}_{a,b}^{a',b'} = \sum_{l=1}^{L} G_g(F_{a,b,l}, F_{a',b',l}), \tag{13}$$

or

$${S_h}_{a,b}^{a',b'} = \sum_{l=1}^{L} G_h(F_{a,b,l}, F_{a',b',l}), \tag{14}$$

As result, small scores indicate similar face images (which is the case of a sum of small dissimilarities between landmarks), and, similarly, bigger scores indicate dissimilar face images. Therefore, the classification of a new face sample image $I_{a',b'}$ occurs by determining the face image $I_{a,b}$ in the training set which is less dissimilar to $I_{a',b'}$ by using one of the three proposed score functions: ${S_f}_{a,b}^{a',b'}$, ${S_g}_{a,b}^{a',b'}$ or ${S_h}_{a,b}^{a',b'}$.

4. Experimental Results

Experiments were conducted to compare the proposed face recognition method presented in section 3 (which uses the geodesic distance approximations presented in section 2 to discriminate texture in the vicinities of the landmarks) to methods representative of the state-of-the-art using a face database commonly used in face recognition (i.e., the FERET face database [42]). This face database was created with the objective of providing credible data for the development of new techniques, technology, and algorithms for the automatic recognition of human faces. The database is used to develop, test, and evaluate face recognition algorithms. It presents color face images in high-resolution (512×768 pixels), organized in several subsets with specific head pose, expression, age, and illumination conditions. Experiments were performed with the color face images of the first 200 face classes of the subsets fa, fb, hl, hr, rb and rc, including all 6 head poses, totaling 1200 images (6 images for each class), as exemplified in figure 2.

Figure 2. Typical head poses in the FERET face database used in the experiments.

In all experiments with the proposed method, the feature extraction and representation method proposed in section 3.1 was applied to the face images in the database, and the means and covariance matrices were obtained using landmark vicinities of size 11×11 ($w = 11$) centered at each landmark location. In order to select consistent features (landmarks), only faces with no landmark occlusions were used. A common landmark topology was used in all experiments in table 1, which is reported in figure 1, collecting important and discriminative features from face images.

The methods used for comparison in the table 1 are the Customized OLPP method (COLPP) [3], Enhanced ASM method [32], Support Vector Machines (SVM) [43], Spectral Regression Kernel Discriminant Analysis (SRKDA) [27], Multi-view Discriminant Analysis (MvDA) [24], *Eigenfaces* [17], *Fisherfaces* [22], *Laplacianfaces* [19], Orthogonal Locality Preserving Projection (OLPP) [20], Locally Linear Embedding (LLE) [31] and Isomap [29]. The proposed method is compared using three distinct score functions (S_f, S_g and S_h) defined in section 3.2.

A set of experiments involving the proposed method and the aforementioned methods was conducted on the FERET face database, and 6 runs were executed on the entire test subset. In each run, a leave-one-out test strategy was adopted, and 5 head poses per class were randomly selected for training, and 1 head pose per class was randomly selected for testing. Table 1 shows the average face recognition rates for the proposed method and methods representative of the state-of-the-art. All methods in table 1 use the same selection of face images, in color (RGB) or in grayscale (color images were converted to grayscale).

For each method listed in table 1, the parameters obtaining the best experimental results were chosen by testing each method with several parameters configurations until the maximum recognition rate was reached. The parameter d used in *Eigenfaces*, *Laplacianfaces* and other methods is the dimensionality of the subspace, assuming k neighbors, and r is the PCA ratio [17] [19] [20], which also is used by the *Fisherfaces* and the MvDA methods. The adopted SVM implementation was the LIBSVM [44]. In SRKDA, the Gaussian kernel with standard deviation σ was used.

Table 1. Face recognition rates obtained for the FERET face database.

Methods	RGB	Grayscale
Proposed method with score S_h	95.3%	82.3%
Proposed method with score S_g	95.4%	83.2%
Proposed method with score S_f	83.1%	78.6%
COLPP (d=54,k=6,t=500,r=0.78)	93.8%	79.5%
Enhanced ASM	72.5%	53.8%
SVM	85.2%	76.4%
SRKDA (σ=20000)	64.9%	59.7%
MvDA (d=100)	76.1%	71.4%
Eigenfaces (d=51)	69.6%	64.1%
Fisherfaces (r=0.8)	67.1%	65.7%
LPP (d=50,k=1,t=500,r=0.34)	67.1%	65.5%
OLPP (d=54,k=1,t=500,r=0.34)	69.0%	66.2%
LLE	54.2%	46.2%
Isomap	69.1%	64.8%

In the iterative Boosting LDA method [45], 10 iterations were performed in each experiment, using half of the training samples for training and the other half for validation, and the Euclidean distance was used as the distance measure. In table 1, the method MvDA was trained to use head poses as views. In the Enhanced ASM method [32], the parameter α was set to 1 giving more importance to measurements in the local vicinity of the landmarks.

As shown in table 1, experiments with color images presented higher recognition rates than the experiments with the same images but converted to grayscale, confirming a trend that color face features tend to improve face class discrimination [32] [3]. Moreover, the proposed face recognition method with the score functions S_h or S_g presented higher recognition rates than with the score function S_f, pointing out that the covariance information between color channels is important to approximate better the geodesic distance between multivariate normal distributions. Finally, the proposed face recognition method presented higher recognition rates than comparable methods in the state-of-the-art.

The results obtained in table 1 show that the proposed metrics to discriminate texture are efficient in face recognition (G_g in (10) and G_h in (11)). Moreover, we provide an additional set of experiments in general texture discrimination (e.g., texture recognition of material images) in order to evaluate the efficiency of the proposed texture discrimination metric presented in section 2 applied to typical texture databases, i.e., the KTH-TIPS texture database [46] and the KTH-TIPS-2b texture database [47]. In these experiments, statistical feature descriptors (i.e., means and covariance matrices) were extracted from whole texture images by considering that each texture image is normally distributed. This is supported by the fact that the human face presents a well-defined structure

Table 2. Texture recognition rates obtained for the KTH-TIPS texture database.

Methods	Recognition Rates
Proposed method with score S_h	99.76%
Proposed method with score S_g	99.76%
Proposed method with score S_f	66.80%
SRP [7]	99.29%
PLS [48]	98.50%
SSLBP [49]	99.39%
LETRIST [50]	99.00%
DMD [51]	97.96%

which helps face recognition [3] [32], however, texture images have large stochastic variations of repeated patterns and vary in pose and scale, so they are not adequate for landmarking.

In order to evaluate the potential of the proposed method for texture recognition, additional tests were performed on the KTH-TIPS texture database [46] and on the KTH-TIPS-2b database [47]. The KTH-TIPS texture database [46] provides images of textured materials in color with size 200×200 organized in 10 texture classes, and each class consists of 81 samples which are captured under nine scales, three different poses and three distinct illumination directions. Experiments were run partitioning the database samples in 50 partitions of training and testing sets, in which half of the samples per class are randomly selected for training and the remaining half for testing [7]. Table 2 shows the average texture recognition rates for the proposed method and methods representative of the state-of-the-art.

The methods presented in table 2 used for comparison in the KTH-TIPS database are the Sorted Random Projections (SRP) [7], Pattern Lacunarity Spectrum (PLS) [48], Scale-Selective Local Binary Pattern (SSLBP) [49], Locally Encoded Transform Feature Histogram (LETRIST) [50] and Dense Microblock Difference (DMD) [51]. The proposed method was compared using three distinct score functions (S_f, S_g and S_h) defined in section 3.2.

Another challenging database used in texture recognition is the KTH-TIPS-2b database [47] which provides material images in color with size 200×200 organized in 11 texture classes, and each class consists of 432 samples which are captured under nine scales, three different poses and four distinct illuminants, as exemplified in figure 3. Fifty experiments were run partitioning the database samples in a ten-fold test strategy [52], in which 11 samples per class are randomly selected for testing and the remaining samples were selected for training in each experiment. Table 3 shows the average texture recognition rates for the proposed method and methods representative of the state-of-the-art.

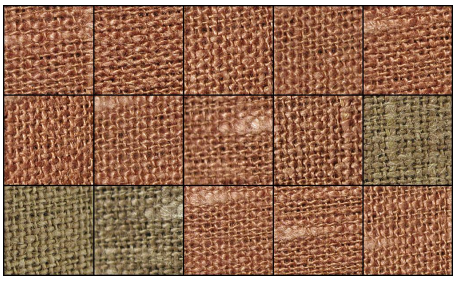

Figure 3. Sample texture images in the KTH-TIPS-2b texture database.

Table 3. Texture recognition rates obtained for the KTH-TIPS-2b texture database.

Methods	Recognition Rates
Proposed method with score S_h	99.24%
Proposed method with score S_g	99.23%
Proposed method with score S_f	93.70%
LBP [53]	92.53%
ILBP [54]	95.88%
SLBP [55]	95.54%
LTP [56]	96.61%
αLBP [52]	96.04%
IαLBP [52]	97.25%

The methods presented in table 3 used for comparison in the KTH-TIPS-2b database are the Local Binary Pattern (LBP) [53], Improved LBP (ILBP) [54], Shift LBP (SLBP) [55], Local Ternary Pattern (LTP) [56], α-Local Binary Pattern (αLBP) [52], and Improved αLBP (IαLBP) [52]. The proposed method was compared using three distinct score functions (S_f, S_g and S_h) defined in section 3.2.

In the experimental results presented in table 2 and 3, the proposed texture dissimilarity metric (i.e., geodesic distance approximations) applied to texture recognition with the score functions S_h or S_g presented higher recognition rates than with the score function S_f, also pointing out that the covariance information between color channels is important to better approximate the geodesic distance between multivariate normal distributions. Finally, the proposed method applied to texture recognition presented higher recognition rates than comparable methods in the state-of-the-art, also pointing out that the proposed texture discrimination metric is efficient not only to discriminate face texture, but to discriminate typical textures (i.e., material images).

5. Conclusions

In this work, geodesic distance approximations for multivariate normal distributions were proposed as texture dissimilarity metrics applied to face recognition. Also, a novel generative face recognition method that

uses information geometry [37] techniques to measure dissimilarities between the texture in the vicinities of corresponding landmarks of distinct face images was proposed, which uses a common landmark topology to mark discriminative locations on high-resolution color face images. The dissimilarity between distinct face images is scored in terms of the dissimilarities between their corresponding landmarks, which are obtained by the proposed geodesic distance approximations between multivariate normal distributions which represent the color intensities in the texture present in the vicinities of each landmark location.

Our proposed face recognition method tends to handle better common issues in face recognition, such as variations in illumination, changes in the head pose, change of appearance, and others, since the extracted pixel distributions sampled in the vicinities of the landmarks tend to be similar across different expressions and head poses. Moreover, the proposed method takes advantage of the natural redundancy that there exists in high-resolution color face images to more accurately measure dissimilarities between textures in the vicinities of corresponding landmarks.

The proposed face recognition method was compared to methods representative of the state-of-the-art using color or grayscale face images and it presented the higher recognition rates. Moreover, these results also support a trend in which color information is relevant on face recognition [32] [3].

Furthermore, the metrics proposed to discriminate texture tend to be efficient, since they can be considered effective in the general texture discrimination problem (e.g., texture recognition of material images), as our additional set of experiments suggest, and potentially can perform better than comparable state-of-the-art methods. The incorporation of the effect of different covariance matrices was found to be important.

Future work will deal with issues such as the identification of the best landmark topology for face recognition. Moreover, we also intent to use investigate texture feature representations for binary patterns applied to face recognition. We also intent to study alternative techniques to obtain other geodesic distance approximations for multivariate normal distributions, also considering Gaussian mixture models.

Acknowledgments

The authors thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, for financial support.

References

[1] Betta G, Capriglione D, Crenna F, Rossi G B, Gasparetto M, Zappa E, Liguori C, and Paolillo A. Face-based recognition techniques: proposals for the metrological characterization of global and feature-based approaches. *Meas. Sci. Techno*, 22(12):124005, 2011.

[2] Kashiko Kodate, Rieko Inaba, Eriko Watanabe, and Takeshi Kamiya. Facial recognition by a compact parallel optical correlator. *Meas. Sci. Techno*, 13(11):1756, 2002.

[3] Soldera J, Behaine C A R, and Scharcanski J. Customized orthogonal locality preserving projections with soft-margin maximization for face recognition. *IEEE Trans. Instrum. Meas.*, 64(9):2417–2426, 2015.

[4] Soldera J, Dodson K, and Scharcanski J. Face recognition based on geodesic distance approximations between multivariate normal distributions. In *Proc. IEEE Int. Conf. Imag. Syst. Techn*, pages 1–6, 2017.

[5] Grecchi E, Doyle O M, Bertoldo A, Pavese N, and Turkheimer F E. Brain shaving: adaptive detection for brain pet data. *Phys. Med. Biol.*, 59(10):2517, 2014.

[6] Arma nanzas R, Iglesias M, Morales D. A., and Alonso-Nanclares L. Voxel-based diagnosis of alzheimer's disease using classifier ensembles. *IEEE J. Biomed. Health Inform.*, 21(3):778–784, 2017.

[7] Liu L, Fieguth P, Kuang G, and Zha H. Sorted random projections for robust texture classification. In *Proc. ICCV*, pages 391–398, 2011.

[8] Pandey P, Deepthi A, Mandal B, and Puhan N B. Foodnet: recognizing foods using ensemble of deep networks. *IEEE Signal Process. Lett*, 24(12):1758–1762, 2017.

[9] Anthimopoulos M M, Gianola L, Scarnato L, Diem P, and Mougiakakou S G. A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE J. Biomed. Health Inform.*, 18(4):1261–1271, 2014.

[10] Gunawan D, Arisandi D, Ginting F M, Rahmat R F, and Amalia A. Russian character recognition using self-organizing map. *J. Phys. Conf. Ser*, 801(1):012040, 2017.

[11] Prashanth N R, Siddarth B, Ganesh A, and Kumar V N. Handwritten recognition of tamil vowels using deep learning. *IOP Conf. Ser. Mater. Sci. Eng.*, 263(5):052035, 2017.

[12] Omidyeganeh M, Shirmohammadi S, Abtahi S, Khurshid A, Farhan M, Scharcanski J, Hariri B, Laroche D, and Martel L. Yawning detection using embedded smart camera. *IEEE Trans. Instrum. Meas.*, 65(3):570–582, 2016.

[13] Verdoolaege G, Soldera J, Macedo T, and Scharcanski J. *Data and Information Dimensionality in Non-Cooperative Face Recognition*, pages 1–35. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[14] Scharcanski J. Bringing vision-based measurements into our daily life: a grand challenge for computer vision systems. *Front. ICT*, 3, 2016.

[15] Werner N-S, Kuhnel S, and Markowitsch H. The neuroscience of face processing and identification in eyewitnesses and offenders. *Front. Behav. Neurosci.*, 7:189, 2013.

[16] R E Bellman. *Dynamic Programming*. Dover Publications, Incorporated, 2003.

[17] Turk M and Pentland A. Eigenfaces for recognition. *J Cogn Neurosci.*, 3(1):71–86, 1991.

[18] Hotelling H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, 24, 1933.

[19] He X, Yan S, Hu Y, Niyogi P, and Zhang H-J. Face recognition using laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):328–340, 2005.

[20] Cai D, He X, Han J, and Zhang H-J. Orthogonal laplacianfaces for face recognition. *IEEE Trans. Image Process.*, 15(11):3608–3614, 2006.

[21] Kokiopoulou E and Saad Y. Orthogonal neighborhood preserving projections: a projection-based dimensionality reduction technique. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12):2143–2156, 2007.

[22] Belhumeur P N, Hespanha J P, and Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, 1997.

[23] Zhang T, Huang K, Li X, Yang J, and Tao D. Discriminative orthogonal neighborhood-preserving projections for classification. *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, 40(1):253–263, 2010.

[24] Kan M, Shan S, Zhang H, Lao S, and Chen X. Multi-view discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell*, 38(1):188–194, 2016.

[25] Mika S, Ratsch G, Weston J, Scholkopf B, and Mullers K R. Fisher discriminant analysis with kernels. In *Neural Netw. Signal Process*, volume 9, pages 41–48, 1999.

[26] Baudat G and Anouar F. Generalized discriminant analysis using a kernel approach. *Neural Comput.*, 12(10):2385–2404, 2000.

[27] Cai D, He X, and Han J. Speed up kernel discriminant analysis. *VLDB J.*, 20(1):21–33, 2011.

[28] Scholkopf B and Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[29] Tenenbaum J B, de Silva V, and Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[30] Cox M A A and Cox T F. *Multidimensional Scaling*, pages 315–347. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[31] Roweis S T and Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[32] Behaine C A R and Scharcanski J. Enhancing the performance of active shape models in face recognition applications. *IEEE Trans. Instrum. Meas.*, 61(8):2330–2333, 2012.

[33] Cootes T F, Taylor C J, Cooper D H, and Graham J. Active shape models - their training and application. *Comput Vis Image Underst.*, 61(1):38–59, 1995.

[34] Wan K-W, Lam K-M, and Ng K-C. An accurate active shape model for facial feature extraction. *Pattern Recognit. Lett.*, 26(15):2409–2423, 2005.

[35] Kim J, Çetin M, and Willsky A S. Nonparametric shape priors for active contour-based image segmentation. *Sign Process*, 87(12):3021–3044, 2007.

[36] Vapnik V N. *Statistical Learning Theory*. Wiley-Interscience, Ney York, 1998.

[37] Amari S-I and Nagaoka H. *Methods of Information Geometry*. American Mathematical Society, 2007.

[38] Atkinson C and Mitchell A F S. Rao's distance measure. *The Indian J. of Statistics*, 48(3):345–365, 1981.

[39] Costa S I R, Santos S A, and Strapasson J E. Fisher information distance: a geometrical reading. *Discrete Appl. Math.*, 197:59–69, 2015.

[40] Kasiński A, Florek A, and Schmidt A. The PUT face database. *Image Process. Commun.*, 13(3):59–64, 2008.

[41] Thomaz C E and Giraldi G A. A new ranking method for principal components analysis and its application to face image analysis. *Image Vis Comput.*, 28(6):902–913, 2010.

[42] Phillips P J, Wechsler H, Huang J, and Rauss P J. The FERET database and evaluation procedure for face recognition algorithms. *Image Vis Comput.*, 16(5):295–306, 1998.

[43] U H-G Kressel. Pairwise classification and support vector machines. In *Advances in Kernel Methods*, pages 255–268. MIT Press, Cambridge, MA, USA, 1999.

[44] Chang C-C and Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intel Sys Technol.*, 2:27:1–27:27, 2011.

[45] Lu J, Plataniotis K N, and Venetsanopoulos A N. Boosting linear discriminant analysis for face recognition. In *in Proc. IEEE Int. Conf. Image Proce.*, volume 1, pages I–657–660, 2003.

[46] Hayman E, Caputo B, Fritz M, and Eklundh J-O. On the significance of real-world conditions for material classification. In *in Proc. Eur. Conf. Comput. Vis*, pages 253–266, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[47] Caputo B, Hayman e, Fritz M, and Eklundh J-O. Classifying materials in the real world. *Image Vision Comput.*, 28(1):150–163, 2010.

[48] Quan Y, Xu Y, Sun Y, and Luo Y. Lacunarity analysis on image patterns for texture classification. In *Proc. CVPR*, pages 160–167, 2014.

[49] Guo Z, Wang X, Zhou J, and You J. Robust texture image representation by scale selective local binary patterns. *IEEE Trans. Image Process.*, 25(2):687–699, 2016.

[50] Song T, Li H, Meng F, Wu Q, and Cai J. Letrist: locally encoded transform feature histogram for rotation-invariant texture classification. *IEEE Trans. Circuits Syst. Video Technol*, PP(99):1–1, 2017.

[51] Mehta R and Eguiazarian K E. Texture classification using dense micro-block difference. *IEEE Trans. Image Process.*, 25(4):1604–1616, 2016.

[52] Delic M, Lindblad J, and Sladoje N. αlbp - a novel member of the local binary pattern family based on α-cutting. In *Proc. ISPA*, pages 13–18, 2015.

[53] Ojala T, Pietikainen M, and Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002.

[54] Jin H, Liu Q, Lu H, and Tong X. Face detection using improved lbp under bayesian framework. In *Proc. ICIG*, pages 306–309, 2004.

[55] Kylberg G and Sintorn I-M. Evaluation of noise robustness for local binary pattern descriptors in texture classification. *EURASIP J. Image Video Process*, 2013(1):17, 2013.

[56] Tan X and Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.*, 19(6):1635–1650, 2010.