

## Article

# Some Information Geometric Aspects of Cyber Security by Face Recognition

C. T. J. Dodson <sup>1,\*</sup>, John Soldera <sup>2</sup>  and Jacob Scharcanski <sup>3</sup><sup>1</sup> School of Mathematics, University of Manchester, Manchester M13 9PL, UK<sup>2</sup> Federal Institute of Education, Science and Technology Farroupilha, Santo Ângelo 98806-700, Brazil; John.Soldera@ifarroupilha.edu.br<sup>3</sup> Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre 91501-970, Brazil; jacobs@inf.ufrgs.br

\* Correspondence: ctdodson@manchester.ac.uk

**Abstract:** Secure user access to devices and datasets is widely enabled by fingerprint or face recognition. Organization of the necessarily large secure digital object datasets, with objects having content that may consist of images, text, video or audio, involves efficient classification and feature retrieval processing. This usually will require multidimensional methods applicable to data that is represented through a family of probability distributions. Then information geometry is an appropriate context in which to provide for such analytic work, whether with maximum likelihood fitted distributions or empirical frequency distributions. The important provision is of a natural geometric measure structure on families of probability distributions by representing them as Riemannian manifolds. Then the distributions are points lying in this geometrical manifold, different features can be identified and dissimilarities computed, so that neighbourhoods of objects nearby a given example object can be constructed. This can reveal clustering and projections onto smaller eigen-subspaces which can make comparisons easier to interpret. Geodesic distances can be used as a natural dissimilarity metric applied over data described by probability distributions. Exploring this property, we propose a new face recognition method which scores dissimilarities between face images by multiplying geodesic distance approximations between 3-variate RGB Gaussians representative of colour face images, and also obtaining joint probabilities. The experimental results show that this new method is more successful in recognition rates than published comparative state-of-the-art methods.

**Keywords:** entropy; information geometry; cyber security; classification; feature recognition; retrieval

**MSC:** 53B20; 62M40; 60D05



**Citation:** Dodson, C.T.J.; Soldera, J.; Scharcanski, J. Some Information Geometric Aspects of Cyber Security by Face Recognition. *Entropy* **2021**, *23*, 878. <https://doi.org/10.3390/e23070878>

Academic Editor: Tobias Oechtering

Received: 4 May 2021

Accepted: 29 June 2021

Published: 9 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It is probable that the widest use of cyber security software is in face and fingerprint recognition, with perhaps a billion or more users of phones, tablets and laptops thereby gaining daily access to their devices. The classification and searching of digital datasets for retrieving images or other objects usually will require multidimensional methods because the features used in classification depend on statistically distributed data. Information geometry provides a natural Riemannian metric structure on smooth spaces of probability density functions. This means that changing properties of a dataset or a subset thereof can be represented on a trajectory in the space of distributions with a natural distance function monitoring the changes. Very high dimensional datasets can be projected onto smaller spaces of features by dimensionality reduction, via eigenvalues of the positive definite symmetric matrices of inter-feature distances [1].

In the context of data represented via probability distributions, multivariate Gaussian distributions are a common choice in representing features in complex large datasets, in consequence of their maximal entropy for given mean and covariance; we outline their

geometry in Section 2. We described in [2,3] an efficient method for colour face image recognition using information geometry in such a way that each face image was represented by a set of 3-variate Gaussians, one for the vicinity of each landmark point in the face. The three variables are the RGB colours of pixels and we used sums of geodesic distance approximations between them at corresponding landmarks of distinct images to measure dissimilarities between face images. Such geodesic distance approximations between  $k$ -variate Gaussians are presented in Section 3. Here in Section 4 we describe a new face recognition method which represents face dissimilarities via the product of such geodesic distances and via joint probabilities. This new method proves to be better than comparable state-of-the-art other face recognition methods.

## 2. Multivariate Gaussian Distributions

In the classification of large sets of digital data objects, a common practical choice is the numerical representation of individual features through multivariate Gaussian distributions, which have a maximal entropy property among distributions with a given mean vector and covariance matrix. Then we have, as described below, an information metric on the space of such multivariate Gaussian probability density functions and we can retrieve all objects with a given feature near to that of a chosen object.

The  $k$ -variate Gaussian distributions have the parameter space

$$\mathbb{R}^k \oplus \mathbb{R}^{(k^2+k)/2}$$

with probability density functions  $f(x; \mu, \Sigma)$  given by:

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

where  $x \in \mathbb{R}^k$  is a possible value for the random variable,  $\mu \in \mathbb{R}^k$  a  $k$ -dimensional mean vector, and  $\Sigma \in \mathbb{R}^{(k^2+k)/2}$  is the  $k \times k$  positive definite symmetric covariance matrix, for features with  $k$ -dimensional representation [4]. In such cases the parameters are obtained using maximum likelihood estimation, as was the case for face recognition applications [3]. The Riemannian manifold of the family of  $k$ -variate Gaussians for a given  $k$  is well understood through information geometric study using the Fisher information metric. For an introduction to information geometry and a range of applications see [5–7].

The Fisher information metric is a Riemannian metric defined on a smooth statistical manifold whose points are probability measures from a probability density function. The Fisher metric determines the geodesic distance between between points in this Riemannian manifold. Given a statistical manifold with coordinates  $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ , and a probability density function  $p(x, \theta)$  as a function of  $\theta$ , the Fisher information metric is defined as:

$$g_{jk}(\theta) = \int_X \frac{\partial \log p(x, \theta)}{\partial \theta_j} \frac{\partial \log p(x, \theta)}{\partial \theta_k} p(x, \theta) dx, \quad (2)$$

which can be understood as the infinitesimal form of the relative entropy and it is also related to the Kullback-Leibler divergence [6,7]. Moreover, a closed-form solution for the Fisher information distance for  $k$ -variate Gaussian distributions is still unknown [8]. The entropy of the  $k$ -variate Gaussian (1) is maximal for a given covariance  $\Sigma$ , and mean  $\mu$ , and it is independent of translations of the mean

$$H(\mu, \Sigma) = - \int_0^\infty f(x) \log f(x) dx = \frac{1}{2} \log(2\pi e |\Sigma|)^k. \quad (3)$$

The natural norm on mean vectors is

$$\|\mu\| = \sqrt{\mu^T \Sigma^{-1} \mu} \quad (4)$$

and the eigenvalues  $\{\lambda_i\}_{i=1\dots k}$  of  $\Sigma$  yield a norm on covariances:

$$\|\Sigma\| = \sqrt{\sum_i^k (\lambda_i)^2} \tag{5}$$

The information distance, that is the length of a geodesic, between two  $k$ -variate Gaussians  $f^A$  and  $f^B$  is the infimum over the length of curves from  $f^A$  to  $f^B$ . It is known analytically in three particular cases:

**Diagonal covariance matrix:**  $\Sigma = \text{Diag}(\sigma_1, \dots, \sigma_k) : f^A = (k, \mu^A, \Sigma^A), f^B = (k, \mu^B, \Sigma^B)$

Here  $\Sigma$  is a diagonal covariance matrix with null covariances [8]:

$$D_\sigma(f^A, f^B) = \sqrt{2 \sum_{i=1}^k \left( \log s \frac{\left| \left( \frac{\mu_i^A}{\sqrt{2}}, \sigma_i^A \right) - \left( \frac{\mu_i^B}{\sqrt{2}}, -\sigma_i^B \right) \right| + \left| \left( \frac{\mu_i^A}{\sqrt{2}}, \sigma_i^A \right) - \left( \frac{\mu_i^B}{\sqrt{2}}, \sigma_i^B \right) \right|}{\left| \left( \frac{\mu_i^A}{\sqrt{2}}, \sigma_i^A \right) - \left( \frac{\mu_i^B}{\sqrt{2}}, -\sigma_i^B \right) \right| - \left| \left( \frac{\mu_i^A}{\sqrt{2}}, \sigma_i^A \right) - \left( \frac{\mu_i^B}{\sqrt{2}}, \sigma_i^B \right) \right|} \right)^2} \tag{6}$$

**Common covariance matrix:**  $\Sigma^A = \Sigma^B = \Sigma : f^A = (k, \mu^A, \Sigma), f^B = (k, \mu^B, \Sigma)$

Here  $\Sigma$  is a positive definite symmetric quadratic form and gives a norm on the difference vector of means:

$$D_\mu(f^A, f^B) = \sqrt{(\mu^A - \mu^B)^T \cdot \Sigma^{-1} \cdot (\mu^A - \mu^B)} \tag{7}$$

**Common mean vector:**  $\mu^A = \mu^B = \mu : f^A = (k, \mu, \Sigma^A), f^B = (k, \mu, \Sigma^B)$

In this case we need a positive definite symmetric matrix constructed from  $\Sigma^A$  and  $\Sigma^B$  to give a norm on the space of differences between covariances. The appropriate information metric is given by Atkinson and Mitchell [9] from a result attributed to S.T. Jensen, using

$$S^{AB} = \Sigma^{A-1/2} \cdot \Sigma^B \cdot \Sigma^{A-1/2}, \text{ with } \{\lambda_j^{AB}\} = \text{Eig}(S^{AB}) \text{ so}$$

$$D_\Sigma(f^A, f^B) = \sqrt{\frac{1}{2} \sum_{j=1}^k \log^2(\lambda_j^{AB})} \tag{8}$$

In principle, (8) yields all of the geodesic distances since the information metric is invariant under affine transformations of the mean [9] Appendix 1; see also the article of P. S. Eriksen [10].

In cases where we have only empirical frequency distributions, and empirical estimates of moments, we can use the Kullback-Leibler divergence, also called relative entropy, between two  $k$ -variate distributions

$$f^A = (x; \mu^A, \Sigma^A), f^B = (x; \mu^B, \Sigma^B)$$

with given mean and covariance matrices, its square root yields a separation measurement [11,12]:

$$\begin{aligned} KL(f^A, f^B) &= \frac{1}{2} \log\left(\frac{\det \Sigma^B}{\det \Sigma^A}\right) + \frac{1}{2} \text{Tr}[\Sigma^{B-1} \cdot \Sigma^A] \\ &+ \frac{1}{2} (\mu^A - \mu^B)^T \cdot \Sigma^{B-1} \cdot (\mu^A - \mu^B) - \frac{k}{2} \end{aligned} \tag{9}$$

This is not symmetric, so to obtain a distance we can take the average KL-distance in both directions:

$$D_{KL}(f^A, f^B) = \sqrt{\frac{|KL(f^A, f^B)| + |KL(f^B, f^A)|}{2}} \tag{10}$$

The Kullback-Leibler distance tends to the Fisher information distance as two distributions become closer together; conversely it becomes less accurate as they move apart. Using only the first and last term in (11) together with (10), we define a divergence  $DKL_{\Sigma}(f^A, f^B)$  by

$$DKL_{\Sigma}(f^A, f^B) = \frac{1}{2} \left( \sqrt{\frac{1}{2} \log \left( \frac{\det \Sigma^B}{\det \Sigma^A} \right) + \frac{1}{2} \text{Tr}[\Sigma^{-B} \cdot \Sigma^A] - \frac{k}{2}} \right. \\ \left. + \sqrt{\frac{1}{2} \log \left( \frac{\det \Sigma^A}{\det \Sigma^B} \right) + \frac{1}{2} \text{Tr}[\Sigma^{-A} \cdot \Sigma^B] - \frac{k}{2}} \right). \tag{11}$$

The Kullback-Leibler divergence does in fact induce the Fisher metric [5,6]. However, there are other geometries with known closed-form solutions for the geodesic distance between  $k$ -variate Gaussians such as the one defined by the  $L^2$ -Wasserstein metric which is derived by the optimal transport problem in which the mass of one distribution is moved to the other [13]. In this geometry, the space of Gaussian measures on a Euclidean space is geodesically convex and corresponds to a finite dimensional manifold since Gaussian measures are parameterized by means and covariance matrices. By restricting it to the space of Gaussian measures inside the  $L^2$ -Wasserstein space, giving a Riemannian manifold which is geodesically convex, several authors derived a closed-form solution for the distance between two such Gaussian measures  $A, B$ , for example Takatsu [13]:

$$W(f_A, f_B)^2 = |\mu_A - \mu_B| + \text{Tr}[\Sigma^A] + \text{Tr}[\Sigma^B] - 2\text{Tr}[\Sigma^{A\frac{1}{2}} \Sigma^B \Sigma^{A\frac{1}{2}}]^{\frac{1}{2}}. \tag{12}$$

Additionally, Bhatia et al. [14] used the Bures-Wasserstein distance on the space of  $k$ -variate Gaussian distributions with zero means in the form:

$$BW(f_A, f_B)^2 = \text{Tr}[\Sigma^A] + \text{Tr}[\Sigma^B] - 2\text{Tr}[\Sigma^{A\frac{1}{2}} \Sigma^B \Sigma^{A\frac{1}{2}}]^{\frac{1}{2}}. \tag{13}$$

### 3. Geodesic Separation between $k$ -Variate Gaussians

Using the results in Section 2 from [15], we investigated in [2,3] the following possible choices for approximating the geodesic distance between two  $k$ -variate Gaussians  $F_1, F_2$  with arbitrary means:

$$G_{\mu}^g(F_1, F_2) = 0.5 \sqrt{(\mu_1 - \mu_2)^T (\Sigma_1)^{-1} (\mu_1 - \mu_2)} \\ + 0.5 \sqrt{(\mu_1 - \mu_2)^T (\Sigma_2)^{-1} (\mu_1 - \mu_2)}, \tag{14}$$

and,

$$G_{\mu}^h(F_1, F_2) = \sqrt{(\mu_1 - \mu_2)^T \left( \frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2)}. \tag{15}$$

From (8),  $G_{\Sigma}(F_1, F_2)$  between the covariances at fixed mean is given by:

$$G_{\Sigma}(F_1, F_2) = \sqrt{\frac{1}{2} \sum_{j=1}^k \log^2(\lambda_j^{12})} \text{ with } S^{12} = \Sigma_1^{-1/2} \cdot \Sigma_2 \cdot \Sigma_1^{-1/2}, \tag{16} \\ \text{and } \{\lambda_j^{12}\} = \text{Eig}(S^{12}).$$

This led to two distinct ways to approximate the geodesic distance between  $k$ -variate Gaussians,

$$G_g(F_1, F_2) = \frac{G_\mu^g(F_1, F_2) + G_\Sigma(F_1, F_2)}{2}, \text{ or} \quad (17)$$

$$G_h(F_1, F_2) = \frac{G_\mu^h(F_1, F_2) + G_\Sigma(F_1, F_2)}{2}. \quad (18)$$

In the context of face recognition, dissimilarity metrics can be very useful to measure dissimilarities between face images or between patches of face images. Accordingly, geodesic distance approximations such as  $G_g$  and  $G_h$ , Equations (17) and (18), can be employed as a dissimilarity metric between probability distributions representative of face landmarks [2,3] as we show in the face recognition approach that we present next.

#### 4. Face Recognition Experiments

The distance between two Gaussian distributions lying in the Riemannian manifold of  $k$ -variate Gaussians is given by the arc length of a minimizing geodesic curve which connects both Gaussians. Moreover, geodesics are intrinsic geometric objects and they are invariant under smooth transformations of coordinates, so in particular the length of a segment is invariant under scale changes of the random variables, from which the mean vectors and covariances are computed.

Consequently, geodesic distances play the role of a natural dissimilarity metric in biometric applications which represent features by probability distributions such as face recognition [2,3]. In such applications, landmark topologies can be used to locate and extract compact biometric features from characteristic face locations in high resolution colour face images [16,17].

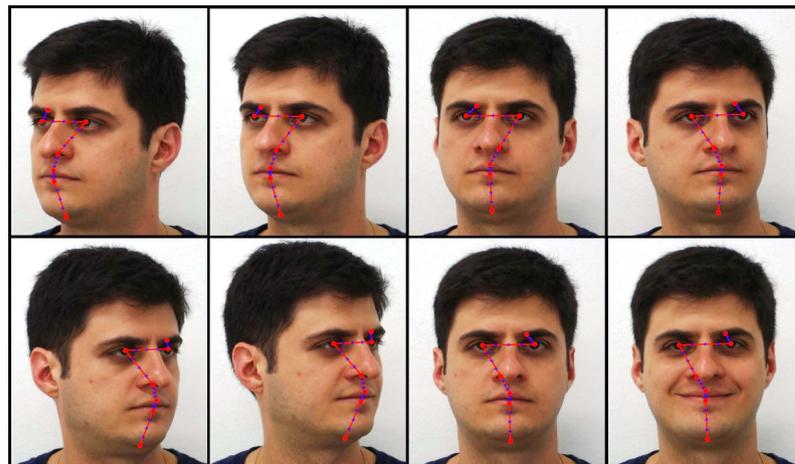
Since an analytic form for the geodesic distance in the Riemannian manifold of  $k$ -variate Gaussians is currently unknown, here we approximate it by constructing approximations applied in a set of face recognition experiments with features represented as  $k$ -variate Gaussians.

In order to extract efficient features for face recognition, we used the FEI Face database [18], which provides colour (RGB) face images with  $640 \times 480$  pixels. The database images were taken against a white homogenous background, with the head in the upright position, turning from left to right, and there are varying illumination conditions and face expressions. Since the images are 3-channeled (RGB), so here  $k = 3$ .

Also, we made use of another challenging database, namely the FERET Face Database [19], which provides colour (RGB) face images with  $512 \times 768$  pixels organized in several subsets with specific head pose, expression, age, and illumination conditions.

To extract meaningful features from face images of both databases, we adopted the landmark topology presented in Figure 1 with seven landmarks at characteristic face locations such as eyebrows, eyes, nose, mouth and chin (in red dots), together with three equally spaced interpolated landmarks between each pair of consecutive landmarks (in blue), leading to a total of  $L = 25$  landmarks for each face image. Next, all pixels inside squared patches with size  $11 \times 11$  centred at each landmark location are extracted, leading to a feature space dimensionality of 3025 pixels.

However, it is possible to reduce this high-dimensionality feature space and preserve its discriminative properties by representing each landmark  $\ell$  by the 3-dimensional mean  $\mu_\ell$  and the 3-variate covariance matrix  $\Sigma_\ell$  of each extracted face patch, using images with three colour channels (RGB). Accordingly, each landmark is represented by 9 dimensions (3 from the mean and 6 from the covariance matrix since it is symmetric). As result, the original feature space dimensionality is reduced to 225. Experimentally, the optimally small landmark topology, interpolated landmark number  $L$  and vicinity size were determined, leading to the landmark number  $L = 25$  and square patches with size  $11 \times 11$  pixels.



**Figure 1.** Adopted landmark topology in the FEI Face Database with varying face poses and expressions [18].

Therefore, by representing each face image as an ordered sequence of probability distributions as in previous approaches [2,3], dissimilarities between distinct face images were scored by summing geodesic distances between 3-variate Gaussians representative of corresponding landmarks of pairs of face images  $x$  and  $y$ . Differently here, we obtained an improved score function for dissimilarities between face images by multiplying the geodesics between corresponding landmarks as follows. We define the functions:

$$\text{Using } (\mu, \text{Diag}(\Sigma)) : S_d^{x,y} = \prod_{l=1}^L D_\sigma(F_x^\ell, F_y^\ell), \quad (19)$$

$$\text{Using } (\mu, \Sigma) : S_g^{x,y} = \prod_{l=1}^L G_g(F_x^\ell, F_y^\ell), \quad (20)$$

$$\text{Using } (\mu, \Sigma) : S_h^{x,y} = \prod_{l=1}^L G_h(F_x^\ell, F_y^\ell), \quad (21)$$

$$\text{Using } (\mu, \Sigma) : S_w^{x,y} = \prod_{l=1}^L W(F_x^\ell, F_y^\ell). \quad (22)$$

where  $F_x^\ell$  and  $F_y^\ell$  represent 3-variate Gaussians,  $F_x^\ell(\mu_x^\ell, \Sigma_x^\ell)$  and  $F_y^\ell(\mu_y^\ell, \Sigma_y^\ell)$ , respectively,  $\ell$  is the  $\ell^{\text{th}}$  landmark from a total of  $L$  landmarks, and  $S_d^{x,y}$ ,  $S_g^{x,y}$ ,  $S_h^{x,y}$ ,  $S_w^{x,y}$  are score functions applicable to images  $x$  and  $y$ . Clearly, in our experiments we cannot use the Bures-Wasserstein distance, Equation (13), since we measure varying means for our RGB variables, but the Wasserstein distance, Equation (12), is suitable and we tested it with the score in Equation (22). Equation (12) might be worth investigating further in future work, as might be a hybrid distance,  $G_{BW} + D_\mu$  using Equations (7) and (13).

All the aforementioned scores define face dissimilarities as products of individual landmark dissimilarities given by geodesic distances. However, by considering a face matching problem, it is possible to convert such dissimilarities between landmarks into probabilities of landmarks not matching, as follows:

$$P(x, y)^\ell = \frac{G(F_x^\ell, F_y^\ell)}{\sum_{m=1}^M G(F_x^\ell, F_m^\ell)}, \quad (23)$$

where  $m$  represents the  $m$ -th candidate face image from a total of  $M$  available face images, and  $G$  is a chosen dissimilarity metric. Then, the problem of finding the face image  $F_y$  which is more similar to  $F_x$  is converted into the problem of finding the face image  $F_y$  which has the least probability of not matching  $F_x$ . This probability is defined as the joint

probability of not matching for all landmarks, i.e., the product of the probabilities of not matching each landmark  $\ell$  as follows:

$$\text{Using } (\mu, \text{Diag}(\Sigma)) : P_d(x, y) = \prod_{l=1}^L \frac{D_\sigma(F_x^\ell, F_y^\ell)}{\sum_{m=1}^M D_\sigma(F_x^\ell, F_m^\ell)}, \quad (24)$$

$$\text{Using } (\mu, \Sigma) : P_g(x, y) = \prod_{l=1}^L \frac{G_g(F_x^\ell, F_y^\ell)}{\sum_{m=1}^M G_g(F_x^\ell, F_m^\ell)}, \quad (25)$$

$$\text{Using } (\mu, \Sigma) : P_h(x, y) = \prod_{l=1}^L \frac{G_h(F_x^\ell, F_y^\ell)}{\sum_{m=1}^M G_h(F_x^\ell, F_m^\ell)}, \quad (26)$$

$$\text{Using } (\mu, \Sigma) : P_w(x, y) = \prod_{l=1}^L \frac{W(F_x^\ell, F_y^\ell)}{\sum_{m=1}^M W(F_x^\ell, F_m^\ell)}. \quad (27)$$

We can also provide an informal interpretation of our three methods: joint probabilities, sums or products of geodesic distances over the set of  $L = 25$  landmarks. By defining the problem of matching one face to another in terms of corresponding landmark dissimilarities, such dissimilarities are converted into probabilities of landmarks not matching as previously presented. Then, by multiplying individual probabilities of landmarks not matching, we obtain the joint probability of all landmarks not matching together at the same time. However, the sum of such probabilities of distinct sequenced events does not have much statistical meaning in our case. Accordingly, by multiplying the landmark dissimilarities, the impact of very similar landmarks is greatly increased as well as very dissimilar landmarks, and the same occurs in the joint probability which also multiplies such landmark dissimilarities. Finally, the product of geodesics has a formulation very similar to the joint probability up to a normalizing factor unique for each test face image.

Finally, the classification procedure is according to the nearest neighbour rule, which means that a new test face sample is attributed to the database individual which presents the training sample that minimizes the chosen score function  $S_d, S_g, S_h, S_w$ , or joint probability  $P_d, P_g, P_h, P_w$ . Even with large datasets, this classification rule has presented low computational complexity due to the fact that we calculate geodesic distance approximations between  $k$ -variate Gaussians, with a small  $k$  value, i.e.,  $k = 3$ , allowing the proposed method to operate near real time as further detailed [2].

In order to validate these new score functions and our geodesic product distance approximations, face recognition experiments were performed to compare our methods with state-of-the-art comparative methods. In the experiments with the FEI face database [18], the first 100 individuals were selected considering the eight head poses indicated in Figure 1, which include the frontal neutral and smiling expressions. Ten runs were performed with the selected database images, and in each run, seven head poses per individual were randomly selected for training, and the remaining one was selected for testing. The averaged recognition rates for the proposed method and comparative methods are presented in Table 1, with all methods using features extracted from the landmark topology shown in Figure 1.

Additionally, an extended set of experiments was performed in the FERET face database [19] by using the first 150 individuals which present the subsets  $fa, fb, hl, hr, ql$  and  $qr$ , which are like the head poses and face expressions in Figure 1. Ten runs were performed with the selected database images, and in each run, five head poses per individual were randomly selected for training, and the remaining one was selected for testing. The averaged recognition rates for the proposed method and comparative methods are also presented in Table 1, with all methods using features extracted from the landmark topology shown in Figure 1. Some of the comparative methods presented in this Table have parameters, so the parameter values which maximized their recognition rates were experimentally determined to obtain their final recognition rates. Those methods are outlined briefly below.

The Eigenfaces method [20] linearly approximates the inherently non-linear face manifold by creating a orthogonal linear projection which best preserves the global feature geometry. On other hand, the Fisherfaces method [21] determines a linear projection which maximizes the between class covariance while minimizing the within class covariance, leading to a better class separation. Furthermore, the method Customized Orthogonal Laplacianfaces (COLPP) [17] obtains an orthogonal linear projection onto a discriminative linear space, which better preserves both the data and class geometry.

In another linear approach, the Multi-view Discriminant Analysis (MvDA) method [22] seeks for a single discriminant common space for multiple views in a non-pairwise manner by jointly learning multiple view-specific linear transforms. In the CCA method [23], multiple feature vectors are fused to produce a feature vector that is more robust to the weakness of each individual vector. And the Coupled Discriminant Multi-manifold Analysis (CDMMA) method [24] explores the neighbourhood information as well as the local geometric structure of the multi-manifold space.

Although the linear approach is simple and efficient, it is also possible to approximate the non-linear face manifold by using non-linear approaches like the Enhanced ASM method [16], which estimates the most discriminative landmarks and scores face similarities by summing probabilities associated to each landmark, taking advantage of this natural multi-modal feature representation. It turned out that the geodesic sum method [2,3] improves on this approach by more accurately scoring face dissimilarities by summing geodesic distances between corresponding landmarks of distinct face images. The experimental results presented in Table 1 include our new methods, geodesic products using the score functions  $S_g$ ,  $S_h$ ,  $S_d$ ,  $S_w$ , and joint probabilities using  $P_g$ ,  $P_h$ ,  $P_d$ ,  $P_w$ , which use our geodesic distance approximations between landmarks on face images. Finally, we performed experiments with the method CM (Continuous Model) [25], summing dissimilarities from corresponding landmarks by using Mahalanobis distance.

**Table 1.** Averaged recognition rates of comparative face recognition methods in the FEI Face database [18] and the FERET Face database [19] using colour (RGB) face images and the landmark topology presented in Figure 1.

Method	FEI	FERET
Joint Probabilities (with $P_g$ ) Equation (17)	99.50%	97.13%
Joint Probabilities (with $P_h$ ) Equation (18)	99.50%	96.86%
Joint Probabilities (with $P_d$ ) Equation (6)	90.00%	79.26%
Joint Probabilities (with $P_w$ ) Equation (12)	96.70%	87.06%
Geodesic Products (with $S_g$ ) Equation (17)	99.50%	97.13%
Geodesic Products (with $S_h$ ) Equation (18)	99.50%	96.86%
Geodesic Products (with $S_d$ ) Equation (6)	90.00%	79.26%
Geodesic Products (with $S_w$ ) Equation (12)	96.70%	87.06%
Geodesic Sums [3], uses Equation (17)	99.50%	96.80%
Geodesic Sums [3], uses Equation (18)	99.50%	96.73%
Geodesic Sums [3], uses Equation (6)	88.40%	75.93%
Geodesic Sums [3], uses Equation (12)	87.30%	81.40%
CM [25]	98.60%	92.33%
Enhanced ASM [16]	89.20%	69.20%
CCA [23]	70.90%	29.06%
CDMMA [24]	37.70%	12.26%
MvDA [22]	44.40%	20.13%
COLPP [17]	96.10%	88.66%
LDA [21]	87.20%	66.00%
Eigenfaces [20]	82.20%	52.00%

## 5. Conclusions

From the experiments reported in Table 1, the geodesic product distance approximation  $S_g$  Equation (17) for 3-variate Gaussians provided the best recognition rate in all experiments, overcoming comparative state-of-the-art methods and also confirming its efficiency as a dissimilarity metric applied in face recognition.

Another conclusion based on Table 1 is that recognition rates with the geodesic distance approximations  $S_g$  and  $S_h$  are better than with  $S_d$  (and  $P_g$  and  $P_h$  are better than with  $P_d$ ) mainly because they take account of local covariances among RGB values in the face images while  $S_d$  and  $P_d$  ignore all covariances, leading to the conclusion that covariances increase the reliability of geodesic distance approximations between 3-variate Gaussians.

Moreover, scores  $S_g$  and  $S_h$  and joint probabilities  $P_g$  and  $P_h$  based on our geodesic distance approximations applied in face recognition also achieved higher recognition rates than scores  $S_w$  and joint probabilities  $P_w$  based on the Wasserstein metric, helping to confirm the efficiency of the Fisher metric [6] over other geometries for such distributions in our case, since the Fisher metric better accounts the geometry of the  $k$ -variate Gaussian distributions because this metric measures the amount of information variation of probability distributions in relation to its parameters, in our case, individual means and covariances.

Finally, the results show that the product of geodesic distances (and joint probabilities) can more accurately score dissimilarities between 3-variate face feature representations than just summing such dissimilarities, since by multiplying landmark dissimilarities the impact of very similar landmarks is greatly increased as well as very dissimilar landmarks, increasing the reliability of face recognition.

**Author Contributions:** All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding; the authors used their own personal funds.

**Data Availability Statement:** Not Applicable.

**Acknowledgments:** The authors are grateful to all Referees whose comments and suggestions helped us improve the presentation.

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

## References

1. Dodson, C.T.J.; Mettanan, M.; Sampson, W.W. Dimensionality reduction for characterization of surface topographies. In *Computational Information Geometry: For Image And Signal Processing Springer Series in Signals and Communication Technology*; Nielsen, F., Critchley, F., Dodson, C.T.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 133–147.
2. Soldera, J.; Dodson, C.T.J.; Scharcanski, J. Face recognition based on geodesic distance approximations between multivariate normal distributions. In Proceedings of the 2017 IEEE International Conference on Imaging Systems and Techniques (IST), Beijing, China, 18–20 October 2017; pp. 1–6.
3. Soldera, J.; Dodson, C.T.J.; Scharcanski, J. Face recognition based on texture information and geodesic distance approximations between multivariate normal distributions. *Meas. Sci. Technol.* **2018**. Available online: <http://iopscience.iop.org/article/10.1088/1361-6501/aade18> (accessed on 31 August 2018).
4. Lenglet, C.; Rousson, M.; Deriche, R.; Faugeras, O. Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing. *J. Math. Imaging Vis.* **2006**, *25*, 423–444. [CrossRef]
5. Amari, S.-I. *Information Geometry and Its Applications*; Applied Mathematical Sciences; Springer: Berlin/Heidelberg, Germany, 2016; p. 194.
6. Amari, S.-I.; Nagaoka, H. *Methods of Information Geometry*; Oxford University Press: Oxford, UK, 2000.
7. Arwini, K.; Dodson, C.T.J. *Information Geometry Near Randomness and Near Independence*; Springer Lecture Notes in Mathematics; Berlin, Germany, 2008.
8. Costa, S.I.R.; Santos, S.A.; Strapasson, J.E. Fisher information distance: A geometrical reading. *Discret. Appl. Math.* **2015**, *197*, 59–69. [CrossRef]
9. Atkinson, C.; Mitchell, A.F.S. Rao's distance measure. *Sankhya Indian J. Stat.* **1981**, *48*, 345–365.
10. Eriksen, P.S. Geodesics connected with the Fisher metric on the multivariate normal manifold. In Proceedings of the GST Workshop, Lancaster, PA, USA, 28–31 October 1987; pp. 225–229. Available online: <http://trove.nla.gov.au/version/21925860> (accessed on 2 July 2021).
11. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

12. Nielsen, F.; Garcia, V.; Nock, R. Simplifying Gaussian mixture models via entropic quantization. In Proceedings of the 17th European Signal Processing Conference, Glasgow, UK, 24–28 August 2009; pp. 2012–2016. Available online: <https://ieeexplore.ieee.org/document/7077426/> (accessed on 2 July 2021).
13. Takatsu, A. On Wasserstein geometry of Gaussian measures. *Adv. Stud. Pure Math.* **2010**, *57*, 463–472.
14. Bhatia, R.; Jain, T.; Lim, Y. On the Bures–Wasserstein distance between positive definite matrices. *Expo. Math.* **2019**, *37*, 165–191. [[CrossRef](#)]
15. Dodson, C.T.J. Information distance estimation between mixtures of multivariate Gaussians. In Proceedings of the Workshop Computational Information Geometry: For Image And Signal Processing International Centre for Mathematical Sciences, Edinburgh, UK, 21–25 September 2015.
16. Behaine, C.A.R.; Scharcanski, J. Enhancing the performance of active shape models in face recognition applications. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 2330–2333. [[CrossRef](#)]
17. Soldera, J.; Behaine, C.A.R.; Scharcanski, J. Customized orthogonal locality preserving projections with soft-margin maximization for face recognition. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2417–2426. [[CrossRef](#)]
18. Thomaz, C.E.; Giraldo, G.A. A new ranking method for Principal Components Analysis and its application to face image analysis. *Image Vis. Comput.* **2010**, *28*, 902–913. [[CrossRef](#)]
19. Phillips, P.J.; Wechsler, H.; Huang, J.; Rauss, P. The FERET database and evaluation procedure for face recognition algorithms. *Image Vis. Comput.* **1998**, *16*, 295–306. [[CrossRef](#)]
20. Turk, M.; Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* **1991**, *3*, 71–86. [[CrossRef](#)] [[PubMed](#)]
21. Belhumeur, P.N.; Hespanha, J.A.P.; Kriegman, D.J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [[CrossRef](#)]
22. Kan, M.; Shan, S.; Zhang, H.; Lao, S.; Chen, X. Multi-view discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 188–194. [[CrossRef](#)] [[PubMed](#)]
23. Haghghat, M.; Abdel-Mottaleb, M.; Alhalabi, W. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Syst. Appl.* **2016**, *47*, 23–34. [[CrossRef](#)]
24. Jiang, J.; Hu, R.; Wang, Z.; Cai, Z. Cdmma: Coupled discriminant multi-manifold analysis for matching low-resolution face images. *Signal Process.* **2016**, *124*, 162–172. [[CrossRef](#)]
25. Farhan, H.R.; Al-Muifraje, M.H.; Saeed, T.R. A new model for pattern recognition. *Comput. Electr. Eng.* **2020**, *83*, 106602. [[CrossRef](#)]