Texture classification using Rao's distance: an EM algorithm on the Poincaré half plane

Salem Said, Lionel Bombrun, Yannick Berthoumieu

TEXTURE CLASSIFICATION USING RAO'S DISTANCE : AN EM ALGORITHM ON THE POINCARÉ HALF PLANE

Salem Said, Lionel Bombrun, Yannick Berthoumieu

Laboratoire IMS (CNRS – UMR 5218), Université de Bordeaux

ABSTRACT

This paper presents a new Bayesian approach to texture classification, yielding enhanced performance in the presence of intraclass diversity. From a mathematical point of view, it specifies an original EM algorithm for mixture estimation on Riemannian manifolds, generalising existing, non probabilistic, clustering analysis methods. For texture classification, the chosen feature space is the Riemannian manifold known as the Poincaré half plane, here denoted H, (this is the set of univariate normal distributions, equipped with Rao's distance). Classes are modelled as finite mixtures of *Riemannian priors*, (Riemannian priors are probability distributions, recently introduced by the authors, which represent clusters of points in H). During the training phase of classification, the EM algorithm, proposed in this paper, computes maximum likelihood estimates of the parameters of these mixtures. The algorithm combines the structure of an EM algorithm for mixture estimation, with a Riemannian gradient descent, for computing weighted Riemannian centres of mass.

Index Terms— Texture classification, Information geometry, Riemannian centre of mass, Mixture estimation, EM algorithm

1. INTRODUCTION

In information geometry [1], a parametric family of probability densities is considered as a Riemannian manifold. Precisely, the role of Riemannian metric is played by the Fisher metric, and that of Riemannian distance by Rao's distance, (see below, Sections 2 and 3). The use of Rao's distance has lead to several statistical applications being formulated as problems of *clustering analysis on Riemannian manifolds*. These include object detection and tracking, shape classification, and image segmentation [2–5].

The present paper considers the use of Rao's distance in texture classification. In doing so, it offers a new probabilistic approach to clustering analysis on Riemannian manifolds, taking advantage of the concept of Riemannian prior, recently introduced by the authors [6]. This allows for clustering analysis to be carried out using an expectation-maximisation, or EM, algorithm, instead of the essentially deterministic k-means approach of existing works, (*e.g.* [2–5]).

Riemannian priors are defined as follows [6], (see Section 4). Let Θ be the parameter space, of a given parametric family, and $d(\theta_1, \theta_2)$ denote Rao's distance, between $\theta_1, \theta_2 \in \Theta$. A Riemannian prior distribution $G(\bar{\theta}, \gamma)$ has two parameters, the centre of mass $\bar{\theta} \in \Theta$, and the dispersion $\gamma > 0$. Its density, with respect to the volume element defined by the Fisher metric, has the form

$$p(\theta|\bar{\theta}, \gamma) = \frac{1}{Z(\gamma)} \exp\left[-\frac{d^2(\theta, \bar{\theta})}{2\gamma^2}\right] \tag{1}$$

The most important property of Riemannian priors, as defined in (1), is the fact that the normalisation constant $Z(\gamma)$, does not depend on $\bar{\theta}$, (see Proposition 2, below). This property holds whenever Θ, equipped with Rao's distance, is a symmetric space of negative sectional curvature, (see definition in [7]). This is the case when Θ is the Poincaré half plane, considered in the present paper, or when Θ is the space of covariance matrices, as in [2–5], which will be the subject of a future journal submission.

The new approach to clustering analysis, introduced in this paper, is the following, (see Section 5). Assume a class of points $\mathcal{C} = \{\theta_1, \ldots, \theta_N\} \subset \Theta$ is expected to contain K clusters $C_1, \ldots, C_K$. Each cluster C_a is modelled as a sample from a Riemannian prior $G(\bar{\theta}_a, \gamma_a)$. Then, the class $\mathcal{C}$ is modelled as a sample from a mixture distribution, whose components are the Riemannian priors $G(\bar{\theta}_a, \gamma_a)$, (see equation (12)). Accordingly, the task of clustering analysis is reduced to maximum likelihood estimation of the parameters of a mixture of Riemannian priors. This is realised using an original EM algorithm, (the main steps are (14)–(16) and (17)–(18)).

The major advantage offered by the probabilistic approach of the present paper, over existing deterministic approaches, is the fact that a cluster C_a is described, in addition to its centre $\bar{\theta}_a$, by a dispersion parameter γ_a. This additional parameter serves as a measure of confidence, in the cluster C_a, with a more dispersed cluster being assigned a lower level of confidence, (see discussion after (8)).

In addition to its general contribution to clustering analysis, the present paper aims to pursue the application to texture classification, started in [6]. This was based on the description of textures using the outputs of a bank of Gabor filters, as in [8, 9]. These outputs are assumed to constitute univariate normal populations, (that is, they follow univariate normal distributions). Then, in [6], the idea is to use Riemannian priors on the Poincaré half plane, (as defined in (1)), as prior distributions for Bayesian classification of univariate normal populations. Recall, here, that the Poincaré half plane is the set of univariate normal distributions, equipped with Rao's distance. Geometrically, (see Section 3), this is the half plane $H \subset \mathbb{R}^2$, consisting of points $z = (x, y)$ where $y > 0$. If $\Pi = \{\pi_1, \ldots, \pi_n\} \subset \mathbb{R}$ is a univariate normal population, with empirical mean and variance $\hat{\mu}$ and $\hat{\sigma}^2$, then Π is represented by the point $z(\Pi) \in H$, where $z(\Pi) = (\hat{\mu}/\sqrt{2}, \hat{\sigma})$, (see (4) and (5)).

The present paper, in order to improve upon the performance obtained in [6], uses mixtures of Riemannian priors, rather than "single" Riemannian priors, as prior distributions for classification. This is motivated by the observation that Riemannian priors, mainly due to their relatively simple structure, fail to take into account the presence of intraclass diversity, which may arise from variations in illumination or viewing conditions [10]. Precisely, intraclass diversity leads to classes of textures having a more complex configuration, where each class may contain several distinct clusters. The gain in performance, achieved through the present use of mixtures of Riemannian priors, is demonstrated in Section 6.

2. FISHER INFORMATION AS A RIEMANNIAN METRIC

The Fisher information matrix is an object of considerable importance in parametric statistics. In finite sample statistics [11], it gives the Cramér-Rao lower bound, on the variance of unbiased estimators. In large sample statistics [12], its inverse matrix is the asymptotic covariance of the maximum likelihood estimator. In 1945 [13], Rao discovered that the Fisher matrix also has a geometric significance, as it can be used to define a Riemannian metric, (that is, a length element), on parametric models.

Recall the definition of the Fisher matrix. Let $\{P_\theta; \theta \in \Theta\}$ be a parametric family of positive probability densities, whose parameter space $\Theta \subset \mathbb{R}^p$ is an open set. For $\theta = (\theta^i; i = 1, \dots p)$, the Fisher matrix $I(\theta)$ has matrix elements,

$$I_{ij}(\theta) = \mathbb{E}_\theta \left[\frac{\partial \log P_\theta}{\partial \theta^i} \, \frac{\partial \log P_\theta}{\partial \theta^j} \right] \qquad i, j = 1, \dots, p \tag{2}$$

where E_θ denotes expectation under P_θ. Rao's main observation is that the expression

$$d\ell^2(\theta) = \sum_{i,j} I_{ij}(\theta) d\theta^i d\theta^j \tag{3}$$

is invariant under reparameterisation. Precisely, if a new parameter ρ is used instead of θ, then $d\ell^2(\rho) = d\ell^2(\theta)$, as long as the relation between ρ and θ is unique and differentiable. This invariance property makes it sensible to think of $d\ell^2(\theta)$, (indeed, a positive quantity), as a length element. In effect [1], expression (3) defines a Riemannian metric on Θ. This is often called the *Fisher metric*, although it is really due to Rao.

Throughout the following, the parametric model under consideration is the univariate normal model. That is, $\Theta \subset \mathbb{R}^2$ is the set of couples $\theta = (\mu, \sigma)$ where $\sigma > 0$, and P_θ the univariate normal distribution having mean μ and variance σ^2. To express the Fisher metric on this parametric model, one may take advantage of the invariance property just mentioned. Use the new parameter

$$z = (x, y) \qquad x = \mu/\sqrt{2} \text{ and } y = \sigma \tag{4}$$

Then [14], expression (3) reduces to

$$d\ell^2(z) = \frac{dx^2 + dy^2}{y^2} \tag{5}$$

This expression is the foundation of Rao's Riemannian geometry of the univariate normal model. As explained in the following section, it leads to identifying this parametric model with the Riemannian manifold known as the *Poincaré half plane*.

3. RAO'S DISTANCE AND CLASSIFICATION

Consider the univariate normal model with the parameter z, defined in (4) of the previous section. Clearly, z ranges over the set $H = \{(x, y); y > 0\} \subset \mathbb{R}^2$. As with any Riemannian metric, the metric (5) on H induces a Riemannian distance $d : H \times H \to \mathbb{R}_+$. In information geometry, this Riemannian distance, induced by the Fisher metric, is called *Rao's distance* [1, 14].

On the other hand, in Riemannian geometry [15, 16], the set H, (which is the upper half plane of $\mathbb{R}^2$), equipped with the Riemannian metric (5), is known as the Poincaré half plane. As such, the Fisher metric leads to identifying the univariate normal model with the Poincaré half plane.

Roughly [15], the Riemannian distance d is defined as follows. The length of a curve $c(t) \in H$, where $t \in [a, b]$, with $c(a) = z_1$ and $c(b) = z_2$, is $L(c) = \int_a^b d\ell(c(t))$. By definition, the Riemannian distance between z_1 and z_2 is the infemum of $L(c)$, taken over all such curves c. It has the analytic expression [16],

$$d(z_1, z_2) = \operatorname{acosh}\left(1 + \frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{2 y_1 y_2} \right) \tag{6}$$

In the present paper, Rao's distance is used in the classification of univariate normal populations, (recall every such population Π is represented by a point $z(\Pi) \in H$). For the purpose of classification, Rao's distance offers several advantages, which are discussed in the following. First, it yields the important property of *existence and uniqueness of Riemannian centre of mass*, given in Proposition 1 below, (an up-to-date reference on the concept of Riemannian centre of mass is [17]).

Let $C = \{z_1, \dots z_N\}$ be a cluster of points in H. Also, let $\omega = \{\omega_1, \dots, \omega_N\}$, be positive weights, with $\omega_1 + \dots + \omega_N = 1$. Consider the problem of minimising the weighted sum of squared distances,

$$V : H \to \mathbb{R}_+ \quad \text{where} \quad V(z) = \frac{1}{2} \sum_{j=1}^{N} \omega_j \times d^2(z_j, z) \tag{7}$$

Here, a global minimiser of V will be called a weighted Riemannian centre of mass of the cluster C, for the given set of weights ω. In the case of uniform weighting, $\omega_1 = \dots = \omega_N = 1/N$, such a global minimiser is called a Riemannian centre of mass of C.

Proposition 1 (Uniqueness of centre of mass) *For any cluster $C = \{z_1, \dots z_N\} \subset H$ and any weights $\omega = \{\omega_1, \dots, \omega_N\}$, as above, the function V of (7) has a unique global minimiser $\hat{z}_\omega$. In particular, any cluster C of points in H has a unique Riemannian centre of mass $\hat{z}$.*

Proof: This follows from a general result in [17], (Theorem 2.1, Page 659), which implies existence and uniqueness of weighted Riemannian centres of mass, in Riemannian manifolds with sectional curvature bounded above by a negative number. But the Poincaré half plane is a Riemannian manifold of constant negative sectional curvature, so the result of [17] applies immediately. ∎

A Riemannian gradient descent algorithm, for computation of weighted Riemannian centres of mass in H is given in Section 5.

The importance of Proposition 1 for classification is illustrated by the approach taken in [18]. Assume within a learning sequence of objects, each represented by a univariate normal population, clusters $C_1, \dots, C_K$ have been identified. These may be thought of as clusters of points in the Poincaré half plane H. To each cluster C_a, $(a = 1, \dots, K)$, one may associate its Riemannian centre of mass $\hat{z}_a$. Then, given a new test object, described by a point $z_t \in H$, one decides to associate it to the cluster C_* where $d(z_t, \hat{z}_*) = \min_a \{d(z_t, \hat{z}_a)\}$, (minimum over $a = 1, \dots, K$).

4. RIEMANNIAN PRIORS FOR BAYESIAN CLASSIFICATION

Recently [6], the authors of the present paper have proposed a new Bayesian approach, generalising the "nearest neighbour" approach of [18], (described at the end of the previous section), and improving upon its performance.

The idea of this Bayesian approach is to model a cluster $C = \{z_1, \dots z_N\} \subset H$, as a sample of size N drawn from a probability distribution $G(\bar{z}, \gamma)$ on H, called a *Riemannian prior*. The distribution $G(\bar{z}, \gamma)$ has two parameters, the centre of mass $\bar{z} \in H$, and the

dispersion $\gamma > 0$. It is designed to have the property, (see Proposition 2 below), that the maximum likelihood estimate of $\bar{z}$ is non other than the Riemannian centre of mass $\hat{z}$ of C, (which was defined after (7)). The maximum likelihood estimate of γ will be denoted $\hat{\gamma}$.

In [6], Riemannian priors were used to classify univariate normal populations, based on a maximum-marginal-likelihood decision rule. Precisely, assume that clusters C_a ($a = 1, \ldots, K$) lead to maximum likelihood estimates $(\hat{z}_a, \hat{\gamma}_a)$. It was shown, in [6], there exists an increasing function $Z : \mathbb{R}_+ \to \mathbb{R}_+$, such that a test object, described by $z_t \in H$, is associated to the cluster C_*, which realises the following minimum over a

$$\min_a \{\log Z(\hat{\gamma}_a) + \frac{1}{2\hat{\gamma}_a^2} d^2(z_t, \hat{z}_a)\} \tag{8}$$

This decision rule reduces to the "nearest neighbour" rule, used in [18], when all classes have the same dispersion, *i.e.* γ_a is the same for all a. By taking into account the role of dispersion, the decision rule (8) ascribes a smaller likelihood to classes with a larger dispersion.

The exact definition of Riemannian priors is the following. Recall the Fisher metric (5) can be used to define a Riemannian area element $dA(z)$. This is given by $dA(z) = dxdy/y^2$. The Riemannian prior distribution $G(\bar{z}, \gamma)$ has probability density with respect to $dA(z)$,

$$p(z|\bar{z}, \gamma) = \frac{1}{Z(\gamma)} \exp\left[-\frac{d^2(z, \bar{z})}{2\gamma^2}\right] \tag{9}$$

Where $d^2(z, \bar{z})$ is the square of Rao's distance, defined in (6), and the normalisation constant $Z(\gamma)$ is given by

$$Z(\gamma) = 2\pi \times \left(\pi\gamma^2/2\right)^{\frac{1}{2}} \times e^{\gamma^2/2} \times \mathrm{erf}(\gamma/\sqrt{2}) \tag{10}$$

where erf is the error function [19]. This last expression was found in [6], by ensuring the integral of $p(z|\bar{z}, \gamma)$ with respect to $dA(z)$ is equal to 1. It is remarkable, as it shows that the normalisation constant does not depend on $\bar{z}$. Mathematically [16], this results from the invariance of Rao's distance under all Moebius transformations of the Poincaré half plane. This yields the second advantage of using Rao's distance in classification, which is encapsulated in the following Proposition 2.

Proposition 2 (Maximum likelihood is Centre of mass) *For the Riemannian prior $G(\bar{z}, \gamma)$, the maximum likelihood estimate of $\bar{z}$, based on $C = \{z_1, \ldots, z_N\}$, is the Riemannian centre of mass $\hat{z}$ of C, defined according to (7).*

Proof: Using (9), as in [6], the log-likelihood function is found to be

$$L(\bar{z}, \gamma) = -N \log Z(\gamma) - \frac{1}{2\gamma^2} \sum_{j=1}^{N} d^2(z_j, \bar{z}) \tag{11}$$

Since $Z(\gamma)$ does not depend on $\bar{z}$, the first term on the right hand side may be overlooked, in seeking the maximum of $L(\bar{z}, \gamma)$ over $\bar{z}$. To conclude, it is enough to compare the second term to (7). ∎

5. EM ALGORITHM FOR MIXTURE ESTIMATION

This section presents the main original contribution of the present paper. Namely, a new EM algorithm, for the estimation of mixtures of Riemannian priors on the Poincaré half plane. The need for this task, of estimating mixtures of Riemannian priors, stems from the following observation.

While successful, in application to specific data sets, the Bayesian approach of [6], summarised in the previous section, fails to take into account the presence of *intraclass diversity*. Precisely, this approach assumes that the given learning sequence is immediately subdivided into clusters, whose members display "homogeneous" properties, in the sense that they can be faithfully modelled as belonging to the same Riemannian prior.

Clearly, this is a restrictive assumption. In the presence of intraclass diversity, a learning sequence should be subdivided into classes, whose members display "heterogeneous" properties, in the sense that they may belong, within the same class, to different clusters, each corresponding to a different Riemannian prior.

Here, this situation is formulated as follows. If a class $\mathcal{C}$, whose members are points $z_1, \ldots, z_N \in H$, is expected to contain K clusters, respectively corresponding to Riemannian priors $G(\bar{z}_a, \gamma_a)$, where $a = 1, \ldots, K$, then $\mathcal{C}$ is modelled as a sample of size N, drawn from the mixture of Riemannian priors

$$p(z|\varpi_a, \bar{z}_a, \gamma_a) = \sum_{a=1}^{K} \varpi_a \, p(z|\bar{z}_a, \gamma_a) \tag{12}$$

where $\varpi_1, \ldots, \varpi_K$ are positive weights, with $\varpi_1 + \ldots + \varpi_K = 1$, and each density $p(z|\bar{z}_a, \gamma_a)$ is given by (9).

Now, assume a training sequence is subdivided into classes, each containing a known numbers of clusters. In order to implement a decision rule, generalising rule (8), which associates any test object, described by $z_t \in H$, to the most likely cluster within the training sequence, it is necessary, for each class $\mathcal{C}$, modelled by (12), to find maximum likelihood estimates of the mixture parameters $\vartheta = (\varpi_a, \bar{z}, \gamma_a)$.

Here, this task is realised using an expectation-maximisation, or EM, algorithm. This algorithm proceeds along the general lines of EM algorithms for mixture estimation, see [20, 21], with necessary adaptation to the use of Riemannian priors of the form (9).

In order to respect the number of pages allocated for this paper, the derivation of the algorithm from first principles, exhibiting the "expectation" and the "maximisation" steps, is here omitted, in favour of a more detailed description of the algorithm itself, at a computational level.

Following [20], the starting point, for the EM algorithm, is the introduction of the following quantities

$$\omega_a(z_j) \propto \varpi_a \times p(z_j|\bar{z}_a, \gamma_a) \qquad n_a = \sum_{j=1}^{N} \omega_a(z_j) \tag{13}$$

where, $\propto$ denotes proportionality, so that $\sum_a \omega_a(z_j) = 1$. To emphasise the fact that $\omega_a(z_j)$ and n_a are computed for a given value of $\vartheta = (\varpi_a, \bar{z}, \gamma_a)$, they shall be denoted $\omega_a(z_j, \vartheta)$ and $n_a(\vartheta)$.

The algorithm iteratively updates $\hat{\vartheta} = (\hat{\varpi}_a, \hat{z}_a, \hat{\gamma}_a)$, an approximation of the maximum likelihood estimate of $\vartheta = (\varpi_a, \bar{z}_a, \gamma_a)$. Precisely, the update rules for $\hat{\varpi}_a$, $\hat{z}_a$, and $\hat{\gamma}_a$ are repeated, for as long as this introduces a sensible change in the values of $\hat{\varpi}_a$, $\hat{z}_a$, and $\hat{\gamma}_a$. Other stopping criteria, involving the amount of increase in the joint likelihood function of $\hat{\varpi}_a$, $\hat{z}_a$, and $\hat{\gamma}_a$, may be used. Also, it is useful to run the algorithm several times, with different initialisations. The update rules are the following,

▸ **Update for $\hat{\varpi}_a$:** Based on the current value of $\hat{\vartheta}$, assign to $\hat{\varpi}_a$ the new value

$$\hat{\varpi}_a^{\mathrm{new}} = \frac{n_a(\hat{\vartheta})}{\sum_{a=1}^{K} n_a(\hat{\vartheta})} \tag{14}$$

▸ **Update for $\hat{z}_a$:** Based on the current value of $\hat{\vartheta}$, compute $\hat{z}_a$ to be the global minimiser of the following function, (compare to (7)),

$$V(z|\hat{\vartheta}) = \frac{1}{2}\sum_{j=1}^{N} \omega_a(z_j, \hat{\vartheta}) \times d^2(z_j, z) \tag{15}$$

▸ **Update for $\hat{\gamma}_a$:** Based on the current value of $\hat{\vartheta}$, compute $\hat{\gamma}_a$ to be the solution of the following equation, for unknown γ,

$$F(\gamma) = \frac{1}{2n_a(\hat{\vartheta})}\, V(\hat{z}_a|\hat{\vartheta}) \tag{16}$$

where $F(\gamma) = \gamma^3 \times \frac{d}{d\gamma} \log Z(\gamma)$.

These three update rules should be performed in the above given order. Therefore, the "current value of $\hat{\vartheta} = (\hat{\varpi}_a, \hat{z}_a, \hat{\gamma}_a)$" is different, in each one of them. For instance, in the update rule of $\hat{\gamma}_a$, the current value of $\hat{z}_a$ is found from the minimisation of (15), just before.

Realisation of the update rules for $\hat{\varpi}_a$ and $\hat{\gamma}_a$ is rather straightforward. On the other hand, the update rule for $\hat{z}_a$ requires minimisation of the function $V(z|\hat{\vartheta})$, which is a function defined on H.

Since this is a function of the form (7), Proposition 1, of Section 3, guarantees it has a unique global minimiser $\hat{z}_a$. Moreover, by a further result of [17], $\hat{z}_a$ is the unique stationary point of $V(z|\hat{\vartheta})$. It is therefore justified to compute $\hat{z}_a$ using a Riemannian gradient descent algorithm, (for the definition of such algorithms, see [22]).

To state this algorithm, denote, for simplicity, $\omega_a(z_j, \hat{\vartheta}) \equiv \omega_j$ and $\hat{z}_a \equiv \hat{z}$. This is possible since a and ϑ are fixed. Also, denote the Riemannian exponential and logarithm mappings $\exp_z : \mathbb{C} \to H$ and $\log_z : H \to \mathbb{C}$, respectively, for $z \in H$. The exact expressions of these mappings are given in [6].

At the n-th iteration of the algorithm, an approximation $\hat{z}_n$ of $\hat{z}$ is available. Then, one computes,

$$\mu = \sum_{j=1}^{N} \omega_j \log_{\hat{z}_n}(z_j) \tag{17}$$

And, for the new approximation $\hat{z}_{n+1}$,

$$\hat{z}_{n+1} = \exp_{\hat{z}_n}(\tau_n \times \mu) \tag{18}$$

where $\tau_n > 0$ is a step-size parameter, which may be chosen manually, or according to a backtracking procedure.

This algorithm is indeed a Riemannian gradient descent algorithm, for the function $V(z|\hat{\vartheta})$ defined in (15). The Riemannian gradient of $V(z|\hat{\vartheta})$, computed at $\hat{z}_n$, is equal to $-\mu$, given by (17). Therefore, one may repeat the iteration (17)–(18) for as long as $|\mu| > \epsilon$, where ϵ is a suitable precision parameter.

6. APPLICATION TO THE VISTEX DATABASE

The present section proposes a new decision rule, for the classification of univariate normal populations, and applies it to texture classification, using the VisTex texture database [23]. The new decision rule, given by (19) below, is a generalisation of decision rule (8), taking into account the new mixture model (12), defined in the previous section.

To begin, the decision rule (19) is presented from the general point of view of classification of univariate normal populations. Assume a training sequence of univariate normal populations is available. Each univariate normal population Π is represented by a point $z(\Pi) \in H$. Thus, the training sequence can be thought of as a training sequence of points in H.

Within this training sequence, classes $\mathcal{C} = \{z_1, \ldots, z_N\} \subset H$ are identified, (the number N depends on the given class $\mathcal{C}$). Each class is assumed to contain the same number K of clusters, and is modelled as a sample drawn from a mixture distribution (12). First, the EM algorithm of Section 5 is applied to each class, leading to maximum likelihood estimates $(\hat{\varpi}_a, \hat{z}_a, \hat{\gamma}_a)$, for $a = 1, \ldots, K$. Each triple of such estimates defines a cluster within the training sequence. Denote the total number of clusters defined in this way L, and the corresponding maximum likelihood estimates $(\hat{\varpi}_c, \hat{z}_c, \hat{\gamma}_c)$, for $c = 1, \ldots, L$. Then, a test population represented by $z_t \in H$ is associated to the cluster C_*, realising the minimum over c,

$$\min_c \{-\log \hat{\varpi}_c + \log Z(\hat{\gamma}_c) + \frac{1}{2\hat{\gamma}_c^2} d^2(z_t, \hat{z}_c)\} \tag{19}$$

This is the new decision rule, proposed for use with the mixture model (12). It reduces to rule (8) of Section (4), in the case where $K = 1$ — in this case, $\varpi_c = 1$ identically, according to (12).

In application to the VisTex database, the following numerical experiment was carried out, (for a detailed presentation, the reader should refer to [6, 8]).

Half the database, 20 images, was used for training, and other half for testing. Each of the 20 training images was subdivided into 169 subimages of 128×128 pixels, with a 32 pixel overlap. For each training subimage, 24 Gabor energy subbands were computed. Each subband s of subimage j gives a univariate normal population Π_{sj}, represented by a point $z_{sj} \in H$.

Each training image $\mathcal{I}$, thought of as a class, in the sense of Section 5, gives $N = 169$ "arrays", $z_j \equiv (z_{sj}; s = 1, \ldots, 24)$, corresponding to subimages $j = 1, \ldots, N$. These arrays z_j are considered as multivariate realisations of a mixture distribution (12), with independent components $(z_{sj}; s = 1, \ldots, 24)$, (for this step, see [6]).

Decision rules (8) and (19) were applied to subimages from the 20 test images. The following table shows their performance, as well as that of a benchmark decision rule, here called the "conjugate prior rule", based on the use of a conjugate normal-inverse gamma prior, in place of the Riemannian prior (9), (again, this is detailed in [6]).

Table 1 : Classification performance with the VisTex database

Prior	Decision rule	Overall Accuracy
Riemannian prior (9)	Rule (8)	$83.29 \pm 0.51\%$
Mixture prior (12)	Rule (19) with K=3	$\mathbf{88.50 \pm 0.88\%}$
Conjugate prior	Conjugate prior rule	$83.48 \pm 0.53\%$

The table shows that decision rule (8) has sensibly the same performance as the conjugate prior rule. On the other hand, decision rule (19), based on the mixture model (12), and implementing the EM algorithm of Section 5, displays much better performance than the two other rules, (a gain of 5 points in overall accuracy).

It may be surprising, to the reader, that decision rule (8) does not perform significantly better than the benchmark conjugate prior rule. It should be noted this is due to the presence of overlap among the textures in the training sequence, (as mentioned above). In [6], each one of the 20 training images was subdivided into 16 non overlapping subimages. Then, decision rule (8) lead to an overall accuracy of $71.88 \pm 2.61\%$, while the conjugate prior rule lead to an overall accuracy of $68.73 \pm 2.92\%$.

7. ACKNOWLEDGMENT

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02).

8. REFERENCES

[1] S.I. Amari and H. Nagaoka, *Methods of information geometry*, American Mathematical Society, 2000.

[2] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, 2008.

[3] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based means on Riemannian manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 728–735.

[4] S. Kurtek, E. Klassen, Z. Ding, M. Avison, and A. Srivastava, "Parameterization invariant shape statistics and probabilistic classification of anatomical surfaces," in *IPMI*, G. Szkely and H.K. Hahn, Eds. 2011, vol. 6801 of *Lecture Notes in Computer Science*, pp. 147–158, Springer.

[5] X. Gu, J.D. Deng, and M.K. Purvis, "Improving superpixel-based image segmentation by incorporating color covariance matrix manifolds," in *International Conference on Image Processing (ICIP)*, 2014, pp. ??–??

[6] S. Said, L. Bombrun, and Y. Berthoumieu, "New Riemannian priors on the univariate normal model," *Entropy*, vol. 16, no. 7, pp. 4015–4031, 2014.

[7] A. Terras, *Harmonic analysis on symmetric spaces and application, II.*, Springer-Verlag, 1988.

[8] S.E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Trans. Im. Proc.*, vol. 11, no. 10, pp. 1160–1167, 2002.

[9] "Rotation-invariant and scale-invariant gabor features for texture image retrieval," *Image and Vision Computing*, vol. 25, no. 9, pp. 1474 – 1481, 2007.

[10] A. Schutz, L. Bombrun, and Y. Berthoumieu, "K-centroids-based supervised classification of texture images using the SIRV modeling," in *Geometric Science of Information (GSI)*, 2031, vol. 6801, pp. 1–8.

[11] E.L. Lehmann and G. Casella, *Theory of point estimation*, Springer-Verlag, second edition, 1998.

[12] I.A. Ibragimov and R.Z. Hasminskii, *Statistical estimation: asymptotic theory*, Springer-Verlag, 1981.

[13] C.R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.

[14] C. Atkinson and A.F.S. Mitchell, "Rao's distance measure," *Sankhya Ser. A*, vol. 43, pp. 345–365, 1981.

[15] M. Berger, *A panoramic view of Riemannian geometry*, Springer-Verlag, 2007.

[16] A. Ramsay, *Introduction to hyperbolic geometry*, Springer-Verlag, 1995.

[17] B. Afsari, "Riemannian L^p center of mass : existence, uniqueness and convexity," *Proc. Amer. Math. Soc.*, vol. 139, no. 2, pp. 655–673, 2011.

[18] F. Nielsen and R. Nock, "Hyperbolic voronoi diagrams made easy," in *International Conference on Computational Science and its Applications (ICCSA)*, 2010, pp. 74–80.

[19] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, ninth edition, 1964.

[20] K. Mengersen, C. Robert, and M. Titterington, *Mixtures : estimation and applications*, Wiley, 2011.

[21] S. Fruhwirth-Schnatter, *Finite mixture and Markov switching models*, Springer, 2006.

[22] P.A. Absil, R. Mahoney, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2008.

[23] MIT Media Lab Vision and Modeling group, "VisTex : Vision texture database, `http://vismod.media.mit.edu/pub/vistex`," 1995.