A Gaussian mixture model with multiple tangent planes

Sara Akodad, Lionel Bombrun, Christian Germain, Yannick Berthoumieu

A Gaussian mixture model with multiple tangent planes

Sara Akodad*, Lionel Bombrun†‡, Christian Germain†‡ and Yannick Berthoumieu†

* CNES, Centre National d'Etudes Spatiales, 18 Avenue Edouard Belin, F-31400 Toulouse, France
e-mail: sara.akodad@cnes.fr
† Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France
e-mail: {lionel.bombrun, christian.germain, yannick.berthoumieu}@ims-bordeaux.fr
‡ Bordeaux Sciences Agro, F-33175 Gradignan, France

***Abstract*—Second-order descriptors play an increasingly important role in many signal and image processing applications including for example remote sensing and medical imaging. In order to develop machine learning models adapted to this kind of descriptors, different probabilistic models based on the Gaussian assumption have been proposed. This includes intrinsic models on the manifold of symmetric positive definite (SPD) matrices such as the Riemannian Gaussian distribution, but also conventional Gaussian models defined on a tangent plane at a reference point. Even if the former is defined on the manifold, it suffers from a practical point of view (no close form expression for the normalization factor, scalar dispersion parameter). The Gaussian model on the tangent plane does not have these drawbacks but is limited by a fixed reference point which might lead to some distortions. To overcome these difficulties, we propose to define a Gaussian model on the tangent plane where the reference point is learned. We found that the maximum likelihood estimator of this reference point is the Karcher/Fréchet mean. Based on this, we introduce a Gaussian mixture model (GMM) with multiple reference points and derive the maximum likelihood estimator. Experimental results on synthetic dataset show that the proposed approach allows to limit the distortion while having an anisotropic dispersion matrix. Finally, an experiment on remote sensing image scene classification is performed to illustrate the potential of the proposed GMM model with a Fisher vector encoding of second-order descriptors computed on the feature maps of a convolutional neural network.**

***Index Terms*—Symmetric positive definite matrices, Gaussian distribution, Gaussian mixture model.**

I. Introduction

The goal of a supervised classification algorithm is to assign an image to the appropriate class depending on its content. The basic technique involves extracting discriminative information within image data, called features. Then, a suitable classification method is applied to categorize the image into defined groups or classes. During the first stage, various kinds of features can be extracted such as color, gradient, shape, edge or textural information. Therefore, the major challenge is to consider image features which are highly distinctive and robust to different nuisances such as photometric or geometrical transformations. To this end, characterizing local image properties attracted a great research interest. Standard approaches are based on computing first-order statistics to model the information behind each image. Later, some authors have dedicated their works in exploiting the information behind second-order statistics using covariance matrix features. These statistics have proved to be highly effective in diverse classification tasks, including person re-identification, texture recognition, material categorization or electroencephalogram (EEG) signals classification in brain-computer interfaces to cite a few of them [1]–[3]. Due to their specific properties, covariance matrices lie on a Riemannian manifold. In fact, conventional Euclidean tools are not adapted for covariance matrix manipulation since they are symmetric positive definite (SPD) matrices. For that, and in response to the need for effective methods to process data lying in the space $\mathcal{P}_d$ of $d \times d$ SPD matrices, attention has been given to metrics and distance measurement on the Riemannian manifold to propose statistical models by considering the specific geometry of covariance matrices [4], [5]. Two Riemannian metrics are generally employed: the log-Euclidean (LE) and the affine-invariant (AI) Riemannian metrics. Even if the AI metric space is endowed with stronger invariance properties compared to the LE metric, estimating the parameters of the statistical models relies on recursive estimation algorithms [5], [6], inducing thus high computational expenses. In contrast, as the LE mapping allows the transformation to a vector-form representation of covariance matrices, the complexity and computational expenses of the algorithms on the LE metric space are significantly reduced. However, when considering the LE metric, the reference point of the tangent plane is imposed and generally set at the identity matrix. This may lead to distortions when covariance matrices are located far from it. To overcome these difficulties, the main contribution of this paper focuses on a proposition of a trade-off between the two statistical models, for instance a Gaussian mixture model with multiple reference points, one for each cluster. This allows a better modeling by limiting the distortion when projecting the set of covariance matrices in the tangent space while maintaining a simple formulation of the model. Section II introduces a brief description of the stochastic modeling approaches on the SPD matrix space. Section III constitutes the main contribution of the paper, it focuses on a proposition of a Gaussian mixture model with multiple reference points. A comparison between different cases is

then assessed. Section IV illustrates the conducted experiments to evaluate the potential of the proposed model in terms of distortion and classification performances. Finally, Section V gives some conclusions.

II. Related work on Gaussian modeling

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of multiple Gaussian component densities. GMMs parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm [7]. For the following, a sample $\mathbf{M} = \{\mathbf{M}_1, ..., \mathbf{M}_N\}$ of N independent and identically distributed (i.i.d) observations is considered.

A. Riemannian affine-invariant Gaussian distribution

A Riemannian Gaussian distribution (RGD) depends on two parameters, a centroid $\bar{\mathbf{M}} \in \mathcal{P}_d$ and a scalar dispersion parameter $\sigma > 0$. It is defined by its probability density function with respect to the Riemannian volume element:

$$p(\mathbf{M}_n|\bar{\mathbf{M}}, \sigma) = \frac{1}{Z(\sigma)} \exp\left[-\frac{d^2(\mathbf{M}_n, \bar{\mathbf{M}})}{2\sigma^2}\right], \quad (1)$$

where $d(.)$ is the Riemannian geodesic distance [8]. The statistical models defined using the affine-invariant metric space are well suited to the Riemannian geometry and provide precise characterisation of SPD matrices. Nevertheless, they suffer from some drawbacks. A first important issue for a complete description of (1) is the definition of the normalization factor $Z(\sigma)$ which does not have a close form expression but can only be computed numerically (see [9] for more details). In addition, a recursive algorithm (Karcher mean) is necessary to estimate the centroid. Moreover, the Riemannian Gaussian is an isotropic model, *i.e.* σ is a scalar dispersion parameter. As a consequence, many authors oriented their researches to a less complex framework while preserving accurate practical results as is the case for the log-Euclidean framework detailed in the following.

B. Log-Euclidean Gaussian distribution

The probability density function of the Log-Euclidean Gaussian distribution is given by (2). It depends on two parameters, the mean $\text{Vec}(\text{Log}_{\mathbf{M}_{ref}}(\bar{\mathbf{M}})) \in \mathbb{R}^{d(d+1)/2}$, computed on the tangent space $\mathcal{T}_{\mathbf{M}_{ref}}$ at a reference point $\mathbf{M}_{ref}$ and the covariance matrix $\Sigma \in \mathcal{P}_{d(d+1)/2}$ which ensures model anisotropy[1]. For the LE metric, an anisotropic model is considered. All the computations are made on the tangent space at a defined reference point $\mathbf{M}_{ref}$. As such, calculations are simplified compared to the AI metric. Nevertheless, this may not preserve the specific geometry of SPD matrices while considering a unique tangent plane. In fact, the observed SPD matrices can be far away from this reference point and some distortions might appear. In order to capture more accurately the structure of the observed SPD matrices, a GMM model defined at different reference points is introduced in the following section, one per component k of the GMM.

[1]Log() is the log-map operator and Vec() corresponds to the vectorization of a matrix where upper triangular elements are multiplied by $\sqrt{2}$

III. GMM with multiple tangent planes

In this section, we introduce a GMM with multiple reference points where those reference points are automatically learned. In order to have a reference point close to the SPD matrices which belong to the cluster, we propose to define it equal to the centroid, *i.e.* $\mathbf{M}_{ref} = \bar{\mathbf{M}}_k$. It yields that (2) vanishes to:

$$p(\mathbf{M}_n|\bar{\mathbf{M}}_k, \Sigma_k) = \frac{\exp\{-\frac{1}{2}(\mathbf{m}_n^{\mathcal{T}_{\bar{\mathbf{M}}_k}})^T \Sigma_k^{-1} (\mathbf{m}_n^{\mathcal{T}_{\bar{\mathbf{M}}_k}})\}}{(2\pi)^{\frac{d(d+1)}{4}} |\Sigma_k|^{1/2}} \quad (3)$$

where $\mathbf{m}_n^{\mathcal{T}_{\bar{\mathbf{M}}_k}} = \text{Vec}(\text{Log}_{\bar{\mathbf{M}}_k}(\mathbf{M}))$. As observed, the k^{th} component of this GMM model corresponds to a zero-mean multivariate Gaussian distribution for the vectors computed at the reference point $\bar{\mathbf{M}}_k$. Interestingly, the mean is zero since it has been transferred to the reference point $\bar{\mathbf{M}}_k$.

In the following, Σ_k is assumed to be a diagonal matrix, and σ_k^2 will represent a vector containing the diagonal elements of Σ_k.

1) Parameter estimation: Let $\mathbf{M} = \{\mathbf{M}_1, \ldots, \mathbf{M}_N\}$ be a set of N i.i.d SPD matrices issued from the GMM model where each component k is defined in (3). We propose to define an EM algorithm to estimate the GMM parameters. First the Log-likelihood function is defined as follows:

$$\mathbb{L}(\mathbf{M}|\bar{\mathbf{M}}, \Sigma) = \log \prod_{n=1}^{N} \sum_{k=1}^{K} \omega_k p(\mathbf{M}_n|\bar{\mathbf{M}}_k, \Sigma_k). \quad (4)$$

- **The center of mass $\bar{\mathbf{M}}_k$:**

The estimation of $\bar{\mathbf{M}}_k$ is found by deriving the log-likelihood (4) with respect to $\bar{\mathbf{M}}_k$ as follows:

$$\frac{\partial}{\partial \bar{\mathbf{M}}_k} \mathbb{L}(\mathbf{M}_n|\bar{\mathbf{M}}_k, \Sigma_k) = \sum_{n=1}^{N} \frac{\omega_k \frac{\partial}{\partial \bar{\mathbf{M}}_k} p(\mathbf{M}_n|\bar{\mathbf{M}}_k, \Sigma_k)}{\sum_{j=1}^{K} \omega_j p(\mathbf{M}_n|\bar{\mathbf{M}}_j, \Sigma_j)}. \quad (5)$$

Under the assumption that observations which belong to cluster k are close to $\bar{\mathbf{M}}_k$, we obtain after some straightforward computations:

$$\hat{\bar{\mathbf{M}}}_k = \arg\min_{\bar{\mathbf{M}}} \sum_{n=1}^{N} \frac{\gamma_k(\mathbf{M}_n)}{N_k} d^2(\bar{\mathbf{M}}, \mathbf{M}_n), \quad (6)$$

where $\gamma_k(\mathbf{M}_n)$ is the posterior probability that a SPD matrix $\mathbf{M}_n$ belongs to cluster k:

$$\gamma_k(\mathbf{M}_n) = \frac{\omega_k \ p(\mathbf{M}_n|\bar{\mathbf{M}}_k, \Sigma_k)}{\sum_{j=1}^{K} \omega_j \ p(\mathbf{M}_n|\bar{\mathbf{M}}_j, \Sigma_j)} \quad (7)$$

and $N_k = \sum_{n=1}^{N} \gamma_k(\mathbf{M}_n)$. Note that d(.) is the Rao's geodesic distance induced by the affine-invariant Riemannian metric. In practice, (6) can be solved by a Karcher mean algorithm [10].

$$p(\mathbf{M}_n|\bar{\mathbf{M}},\Sigma)=\frac{\exp\left\{-\frac{1}{2}\left(\text{Vec}\left(\text{Log}_{\mathbf{M}_{ref}}(\mathbf{M}_n)\right)-\text{Vec}\left(\text{Log}_{\mathbf{M}_{ref}}(\bar{\mathbf{M}})\right)\right)^T\Sigma^{-1}\left(\text{Vec}\left(\text{Log}_{\mathbf{M}_{ref}}(\mathbf{M}_n)\right)-\text{Vec}\left(\text{Log}_{\mathbf{M}_{ref}}(\bar{\mathbf{M}})\right)\right)\right\}}{(2\pi)^{\frac{d(d+1)}{4}}|\Sigma|^{1/2}}. \quad (2)$$

- **The dispersion σ_k^2 :**

Similarly, the maximum likelihood estimator of the i^{th} component of σ_k^2 is the weighted sample variance defined as:

$$\hat{\sigma}_k^2(i)=\frac{1}{N_k}\sum_{n=1}^{N}\gamma_k(\mathbf{M}_n)\left[\mathbf{m}_n^{\mathcal{T}_{\bar{\mathbf{M}}_k}}(i)\right]^2. \quad (8)$$

- **The weight ω_k:**

The maximum likelihood estimator of ω_k is given by:

$$\hat{\omega}_k=\frac{N_k}{N}. \quad (9)$$

2) Parallel transport: In practice, it is often desirable to express data on the same reference plane. For that, we propose to use the parallel transport [4] in order to transport the set of covariance matrices around the identity matrix by applying the following operation:

$$\mathbf{Z}_{(n,k)}=\bar{\mathbf{M}}_k^{-\frac{1}{2}}\ \mathbf{M}_n\ (\bar{\mathbf{M}}_k^{-\frac{1}{2}})^T. \quad (10)$$

It is now possible to estimate the variance $\eta_k^2(j)$ for the transported set according to:

$$\hat{\eta}_k^2(j)=\frac{1}{N_k}\sum_{n=1}^{N}\gamma_k(\mathbf{M}_n)\left[\mathbf{z}_{(n,k)}^{\mathcal{T}_{\mathbf{I}_d}}(j)\right]^2, \quad (11)$$

where $\mathbf{z}_{(n,k)}^{\mathcal{T}_{\mathbf{I}_d}}$ is the LE vector representation of $\mathbf{Z}_{(n,k)}$ computed at the identity matrix $\mathbf{I_d}$. The proposed approach of parallel transport of data around the identity matrix allows hence to reduce the dispersion and avoid the numerical instabilities. Now, to derive the EM algorithm, the posterior probability $\gamma_k(\mathbf{M}_n)$ should be computed with the shifted SPD matrix $\mathbf{Z}_{(n,k)}$. It yields:

$$\gamma_k(\mathbf{M}_n)=\frac{\omega_k\ p(\mathbf{Z}_{(n,k)}|\mathbf{I}_d,\mathbf{H}_k)}{\sum_{j=1}^{K}\omega_j\ p(\mathbf{Z}_{(n,j)}|\mathbf{I}_d,\mathbf{H}_j)}, \quad (12)$$

where $\mathbf{H}_k$ is a diagonal matrix containing the variance vector elements η_k^2 on its diagonal. To summarize, the EM algorithm for the GMM with K reference points $\bar{\mathbf{M}}_k$ is given in Algorithm 1.

3) Comparison between the two GMM models: Table I draws an overview of the two considered GMM models. The first one consists in a classical GMM model where SPD matrices are projected on the tangent plane at the identity matrix while the second one considers projections with multiple tangent planes where, each mixture component has its own tangent space. Even if these two models have many similarities (GMM models, projection on a tangent plane), they differ in some aspects:

Algorithm 1 EM algorithm for a GMM model with different reference points

Input: $\mathbf{M}=\{\mathbf{M}_1,\ \ldots,\mathbf{M}_N\}$ a set of SPD matrices where $\mathbf{M}_n\in\mathcal{P}_d$, K number of components of the GMM, N_{iter} maximum number of iterations.

Initialize: $\gamma_k(\mathbf{M}_n)$ via the EM algorithm for the GMM defined at the tangent plane of the identity matrix.

```
it ← 1
while (it ≤ N_iter) do
    Update M̄_k by solving (6) with Karcher/Fréchet mean algorithm.
    Compute Z_(n,k) with (10) to transport M_n
    Update η_k with (11).
    Update ω_k with (9).
    Update the posterior probability γ_k(M_n) with (12).
    it ← it + 1
end while
```

Output: $\omega_k\in[0,1]$, $\bar{\mathbf{M}}_k\in\mathcal{P}_d$ and $\eta_k\in\mathbb{R}^{\frac{d(d+1)}{2}}$.

- For the second model, there is no offset parameter μ_k since the mean has been transferred to the reference point $\bar{\mathbf{M}}_k$.
- The maximum likelihood estimator of the centroid for the GMM model defined at a unique reference point is the weighted log-Euclidean mean vector while for the proposed model, it is the centroid on the manifold, *i.e.* the Karcher/Fréchet mean [10].

IV. Experiments

In this section, we propose two experiments in order to compare the proposed model which involves multiple tangent planes with a GMM model defined in a unique tangent plane at the identity matrix. With synthetic data, we first discuss in terms of distortions induced by matrix projection onto the tangent plane. Then, the next subsection introduces an application on remote sensing scene classification.

A. Distortion measurement

From a practical point of view, the difference between the two models can be discussed in terms of distortions induced by matrix projection onto the tangent plane. To ease the comparison, we consider that the number of clusters is equal to 1. Let's consider that on the tangent space of $\bar{\mathbf{M}}$, a set of N observations $\{\mathbf{m}_1^{\mathcal{T}_{\bar{\mathbf{M}}}},\ldots,\mathbf{m}_N^{\mathcal{T}_{\bar{\mathbf{M}}}}\}$ are generated from a zero-mean Gaussian distribution. With the exponential map at $\bar{\mathbf{M}}$, a set of N SPD matrices is obtained. We want to compare the modeling of this set by using two models: a Gaussian with a unique tangent space where the reference point is the identity

TABLE I: Comparison between the two considered GMM models: one defined at a unique reference point and one with K reference points, one per cluster.

	Unique reference point at identity	**K reference points**
Gaussian mixture model	$p(\mathbf{m}_n^{\mathcal{T}_{\mathbf{I}_d}}\|\lambda) = \sum_{k=1}^{K} \omega_k p\left(\mathbf{m}_n^{\mathcal{T}_{\mathbf{I}_d}}\|\mu_k, \sigma_k\right)$ where $p\left(\mathbf{m}_n^{\mathcal{T}_{\mathbf{I}_d}}\|\mu_k, \sigma_k\right) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{m}_n^{\mathcal{T}_{\mathbf{I}_d}}-\mu_k)^T\Sigma_k^{-1}(\mathbf{m}_n^{\mathcal{T}_{\mathbf{I}_d}}-\mu_k)\right\}}{(2\pi)^{\frac{d(d+1)}{4}}\|\Sigma_k\|^{1/2}}$ with $\mu_k \in \mathbb{R}^{\frac{d(d+1)}{2}}$, $\sigma_k^2 = diag(\Sigma_k) \in \mathbb{R}^{\frac{d(d+1)}{2}}$ and $\omega_k \in [0,1]$.	$p(\mathbf{M}_n\|\lambda) = \sum_{k=1}^{K} \omega_k p\left(\mathbf{Z}_{(n,k)}\|\mathbf{I}_d, \eta_k\right)$ where $p\left(\mathbf{Z}_{(n,k)}\|\mathbf{I}_d, \eta_k\right) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{z}_{(n,k)}^{\mathcal{T}_{\mathbf{I}_d}})^T\mathbf{H}_k^{-1}(\mathbf{z}_{(n,k)}^{\mathcal{T}_{\mathbf{I}_d}})\right\}}{(2\pi)^{\frac{d(d+1)}{4}}\|\mathbf{H}_k\|^{1/2}}$ with $\mathbf{Z}_{(n,k)} = (\bar{\mathbf{M}}_k^{-\frac{1}{2}})^T\,\mathbf{M}_n\,\bar{\mathbf{M}}_k^{-\frac{1}{2}}$, $\bar{\mathbf{M}}_k \in \mathcal{P}_d$, $\eta_k^2 = diag(\mathbf{H}_k) \in \mathbb{R}^{\frac{d(d+1)}{2}}$ and $\omega_k \in [0,1]$.
Characteristics	Projection onto a unique tangent plane at $\mathbf{I}_d$ Centroïd : Log-Euclidean mean vector	Projection on multiple tangent spaces Centroïd : Karcher/Fréchet mean

matrix and a zero-mean Gaussian model where the reference point is automatically learned as proposed in Section III. For both approaches, we study the distortion induced by the projection on the tangent space. To quantify this distortion, a similarity measure between two set of points is employed. Here, we propose to use the Hausdorff distance. It has been used in [11] for similar purpose and permits measuring the similarity between two sets. Two sets are considered close to each other in the sense of the Hausdorff distance if every point of the first set is close to some point of the second set. Let be $\mathbf{X}$ and $\mathbf{Y}$ two non-empty subsets of a metric space, their Hausdorff distance $d_H(\mathbf{X}, \mathbf{Y})$ is defined by:

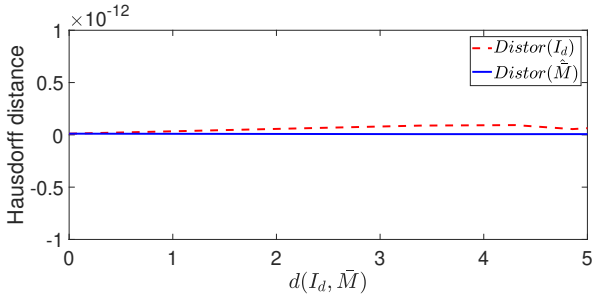
$$d_H(\mathbf{X}, \mathbf{Y}) = \max\left\{\sup_{x\in\mathbf{X}} d(x, \mathbf{Y}), \sup_{y\in\mathbf{Y}} d(y, \mathbf{X})\right\}, \qquad (13)$$

with $d(x, \mathbf{Y}) = \min_{y\in\mathbf{Y}} d(x, y)$ and $d(.)$ a distance. Since the computation is made on the tangent plane, the Euclidean distance is considered, where $d(x, y) = ||x - y||_2$.

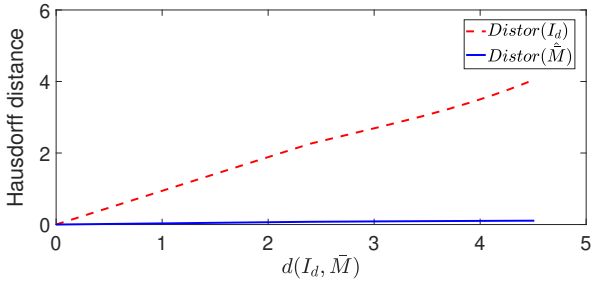
For this experiment, two models are considered for comparison as a function of the geodesic distance between the identity matrix $\mathbf{I}_d$ and $\bar{\mathbf{M}}$:

- **Distor($\mathbf{I_d}$)**: The Hausdorff distance is computed between scatter plot of SPD matrices projected at the identity matrix, and matrices transported to the identity matrix from $\bar{\mathbf{M}}$ according to the parallel transport in (10). The mean is removed to center at the same point with the purpose of comparing the scatter plot shapes.
- **Distor($\hat{\bar{\mathbf{M}}}$)**: Distance is computed between scatter plot of matrices projected at the identity matrix after applying the parallel transport operation (10) from $\bar{\mathbf{M}}$ and $\hat{\bar{\mathbf{M}}}$ to the identity matrix $\mathbf{I}_d$.

Hence, Distor($\mathbf{I_d}$) in red quantifies the distortion observed for a Gaussian model where the tangent plane is defined at the identity matrix, and Distor($\hat{\bar{\mathbf{M}}}$) in blue measures the distortion for the proposed model. Fig. 1 draws the evolution of these distortions as a function of the geodesic distance between the identity matrix and the true reference point $\bar{\mathbf{M}}$. For that, two



(a) when $\bar{\mathbf{M}} = K\mathbf{I}_d$



(b) when $\bar{\mathbf{M}} \neq K\mathbf{I}_d$

Fig. 1: Hausdorff distance comparison as a function of geodesic distance between $\mathbf{I_d}$ and $\bar{\mathbf{M}}$.

cases are considered: when $\bar{\mathbf{M}}$ is proportional to $\mathbf{I_d}$ (Fig. 1(a)) and when it is not the case (Fig. 1(b)). As observed in Fig. 1(a), no distortion is observed between both models when $\bar{\mathbf{M}} = K\mathbf{I}_d$. This can be explained by the fact that in this case $\text{Log}_{\bar{\mathbf{M}}}(\mathbf{M}_n) = K\ \text{Log}_{\mathbf{I}_d}(\mathbf{M}_n)$. However, when $\bar{\mathbf{M}} \neq K\mathbf{I}_d$, the proposed model (in blue) allows to significantly reduce the distortions since the estimated reference point $\hat{\bar{\mathbf{M}}}$ is close to the true one ($\bar{\mathbf{M}}$). This clearly illustrates the potential of the proposed model especially when the reference point is far from the identity matrix.

B. Experiments on remote sensing classification

As experimental results on synthetic data have demonstrated the benefit of using adapted tangent planes to preserve the

correct fitting of Gaussian models and avoid projection distortions, especially when data are located far from the reference point. The objective here is to illustrate the added value of the proposed GMM model in an image processing application on remote sensing scene classification. We propose here to extend a classification architecture introduced in [12] to a GMM modeling with multiple tangent planes. It consists on the Fisher vector encoding of covariance matrix descriptors computed locally on the features maps of a convolutional neural networks. For that, the EM algorithm is used to learn the codebook on the training set, and this latter one is used to encode a set $\boldsymbol{\mathcal{M}} = \{\mathbf{M}_1, \ \ldots, \mathbf{M}_N\}$ of N covariance matrices with Fisher vectors (FV). The FV associated to a GMM model with K reference points are obtained as:

$$\mathscr{G}^{\boldsymbol{\mathcal{M}}}_{\bar{\mathbf{M}}_k(j)} = \frac{1}{\sqrt{\omega_k}} \sum_{n=1}^{N} \gamma_k(\mathbf{M}_n) \left(\frac{\mathbf{z}^{\mathcal{T}_{\mathbf{I}_d}}_{(n,k)}(j)}{\eta_k(j)} \right), \tag{14}$$

$$\mathscr{G}^{\boldsymbol{\mathcal{M}}}_{\eta_k(j)} = \frac{1}{\sqrt{2\ \omega_k}} \sum_{n=1}^{N} \gamma_k(\mathbf{M}_n) \left(\frac{\left[\mathbf{z}^{\mathcal{T}_{\mathbf{I}_d}}_{(n,k)}(j)\right]^2}{\eta^2_k(j)} - 1 \right), \tag{15}$$

where $\mathbf{z}^{\mathcal{T}_{\mathbf{I}_d}}_{(n,k)}$ is the vector version of $\mathbf{Z}_{(n,k)}$ computed at the identity matrix $\mathbf{I_d}$. As detailed in Section IV-A, $\mathbf{Z}_{(n,k)}$ is the result of parallel transport of covariance matrices $\mathbf{M}_n$ from $\bar{\mathbf{M}}_k$ to the identity matrix $\mathbf{I_d}$ and $\eta^2_k(j)$ is the variance of the transported set. For more details, the interested reader is referred to [13].

Table II highlights the classification results obtained on a single convolutional layer of the VGG-16 model, for instance the second layer (conv 2). Each layer is represented by a set of region covariance matrices which are further encoded with Hybrid LE $FV_{\mathcal{T}_{\mathbf{I}_d}}$ when using a single tangent plane at the identity matrix whereas for $FV_{\mathcal{T}_{\bar{\mathbf{M}}_k}}$, multiple tangent planes are considered.

TABLE II: Classification results on the UC Merced dataset for the second convolutional layer of the VGG-16 model (10% of the images are used for training).

Method	Conv 2
Hybrid LE $FV_{\mathcal{T}_{\mathbf{I}_d}}$	**65.1 $\pm$ 1.6 %**
Hybrid LE $FV_{\mathcal{T}_{\bar{\mathbf{M}}_k}}$	**66.7 $\pm$ 0.9 %**

As observed, the proposed model Hybrid LE $FV_{\mathcal{T}_{\bar{\mathbf{M}}_k}}$ involving an adapted tangent plane yields to a slight but interesting gain in terms of overall accuracy compared to the approach with a tangent plane at the identity matrix (Hybrid LE $FV_{\mathcal{T}_{\mathbf{I}_d}}$). However, while working with the proposed model, it is important to note the trade-off between complexity and accuracy. In fact, for this case, the results are achieved at the cost of a higher model complexity.

V. Conclusion

This paper has introduced a Gaussian mixture model for manifold valued data. By exploiting multiple tangent planes (one per cluster), we have shown that the mean vector for each cluster is null. Actually, the mean is transferred to the choice of the reference point for each tangent plane where we have shown that the maximum likelihood estimator of it coincides with a Karcher mean. Compared to an intrinsic model on the manifold, the proposed model has many advantages (close form expression for the normalization factor, anisotropic dispersion matrix). And compared to the classical approach where data are projected on a unique tangent plane, experiments on synthetic data have shown that distortions are significantly reduced for the proposed model. Finally, an image processing application has been proposed for remote sensing scene classification. We have proposed a Fisher vector encoding of convolutional neural network feature maps. Experimental results have confirmed the potential of the proposed model. Future works will include the validation of this GMM model for other signal and image processing applications.

References

[1] M. Faraki, M. T. Harandi, and F. Porikli, "More about VLAD: A leap from Euclidean to Riemannian manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4951–4960.

[2] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *NeuroComputing*, vol. 112, pp. 172–178, 2013.

[3] S. Said, L. Bombrun, and Y. Berthoumieu, "Texture classification using Rao's distance on the space of covariance matrices," in *Geometric Science of Information*, vol. 9389, Oct 2015, pp. 371–378.

[4] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.

[5] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras, "Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor mri processing," *Journal of Mathematical Imaging and Vision*, vol. 25, pp. 423–444, Oct 2006.

[6] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977. [Online]. Available: http://www.jstor.org/stable/2984875

[8] A. Terras, *Harmonic analysis on symmetric spaces and applications*, ser. Harmonic Analysis on Symmetric Spaces and Applications. Springer-Verlag, 1988, no. vol. 1.

[9] S. Said, L. Bombrun, Y. Berthoumieu, and J. H. Manton, "Riemannian gaussian distributions on the space of symmetric positive definite matrices," *IEEE Transactions on Information Theory*, vol. 63, pp. 2153–2170, 2015.

[10] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[11] S. Labsir, "Méthodes statistiques fondées sur les groupes de Lie pour le suivi d'un amas de débris spatiaux." Thesis, Université de Bordeaux, Dec. 2020. [Online]. Available: https://tel.archives-ouvertes.fr/tel-03103892

[12] S. Akodad, "Ensemble learning methods on the space of covariance matrices : application to remote sensing scene and multivariate time series classification," Theses, Université de Bordeaux, Dec. 2021. [Online]. Available: https://theses.hal.science/tel-03484011

[13] S. Akodad, L. Bombrun, J. Xia, Y. Berthoumieu, and C. Germain, "Ensemble learning approaches based on covariance pooling of cnn features for high resolution remote sensing scene classification," *Remote Sensing*, vol. 12, no. 20, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/20/3292