An M-estimator for robust centroid estimation on the manifold of covariance matrices: performance analysis and application to image classification

Ioana Ilea*†, Hatem Hajri*, Salem Said*, Lionel Bombrun*, Christian Germain* and Yannick Berthoumieu*
*Université de Bordeaux, Laboratoire IMS, Signal and Image Group, France
†Technical University of Cluj-Napoca, Romania

***Abstract*—Many signal and image processing applications, including texture analysis, radar detection or EEG signal classification, require the computation of a centroid from a set of covariance matrices. The most popular approach consists in considering the center of mass. While efficient, this estimator is not robust to outliers arising from the inherent variability of the data or from faulty measurements. To overcome this, some authors have proposed to use the median as a more robust estimator. Here, we propose an estimator which takes advantage of both efficiency and robustness by combining the concepts of Riemannian center of mass and median. Based on the theory of M-estimators, this robust centroid estimator is issued from the so-called Huber's function. We present a gradient descent algorithm to estimate it. In addition, an experiment on both simulated and real data is carried out to evaluate the influence of outliers on the estimation and classification performances.**

I. Introduction

Covariance matrices are used in a wide variety of applications in signal and image processing, including array processing [1], radar detection [2]–[4], object detection [5], [6], image segmentation [7] or classification [8]–[11], etc.

Recently, covariance matrices have been modeled as realizations of Riemannian Gaussian distributions [12] (RGDs) and further used in classification algorithms. Mixture models have also been proposed, requiring clustering approaches such as k-means or expectation maximization (EM) to estimate their parameters. These clustering procedures are based on regrouping the dataset's elements into clusters characterized by their central values, called centroids. The most widely used centroid estimator is the Riemannian center of mass [13] which corresponds to the maximum likelihood estimate of the central element of an RGD [12]. While being efficient, the main disadvantage of the center of mass is its non-robust behavior to outliers that can exist in the dataset [14], [15]. To overcome this problem, the concept of median has been extended to Riemannian manifolds, as a robust alternative for the centroid computation [16]–[19].

The main contribution of this paper is to propose a centroid estimator which is a good trade-off between efficiency and robustness. Based on the theory of M-estimators, the present work introduces a novel class of robust centroid estimators, issued from the so-called Huber's function [20], [21]. This novel Huber's centroid extends both concepts of center of mass and median. Some experiments are proposed to evaluate the influence of outliers on estimation and classification performances. The second contribution of this paper is to draw an analogy between conventional covariance matrix estimators and the considered centroid estimation technique.

The paper is structured as follows. Section II presents a parallel between the covariance matrix robust estimation and the estimation of cluster centroids. Section III introduces the proposed Huber's estimator, along with a gradient descent algorithm for its computation. The center of mass and the median are presented as special cases and their behavior with respect to outliers is analyzed. In Section IV, the proposed Huber's centroid is used for texture image classification. Finally, Section V reports some conclusions and perspectives on this work.

II. From Covariance Matrix Estimation to Centroids Estimation

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a set of N independent and identically distributed (i.i.d.) random variables. In this case, the M-estimator $\hat{\mathbf{M}}$ of the covariance matrix $\mathbf{M}$ characterizing the dataset is defined as the solution of [20], [21]:

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{i=1}^{N} u(\mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i)\ \mathbf{x}_i \mathbf{x}_i^T, \tag{1}$$

where $u(\cdot)$ is a positive-valued function, which gives a weight to each observation $\mathbf{x}_i$. The purpose of this weight function is to control the influence of aberrant values in the estimation process. Thus, to ensure a small contribution of outliers, $u(\cdot)$ has to be a decreasing function.

Depending on the weight function, various covariance matrix estimators can be defined. For example, if $u(t) = 1$, all the observations have the same weight, resulting in the sample covariance matrix (SCM) estimator. Moreover, if $u(t) = 1/t$ a robust estimator called the fixed point (FP) estimator [21], also known as the Tyler's estimator, is obtained. In addition, the Huber's estimator [20], a trade-off between the SCM and the FP, can be considered for the specific Huber function defined as:

$$u(t) = \min\left(1, \frac{T}{t}\right). \tag{2}$$

This behavior is illustrated in Fig. 1. The influence of outliers is controlled by the fixed threshold T. If the quadratic term $t = \mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i$ is smaller than T, the Huber's function is constant, otherwise $u(\cdot)$ will start to decrease.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

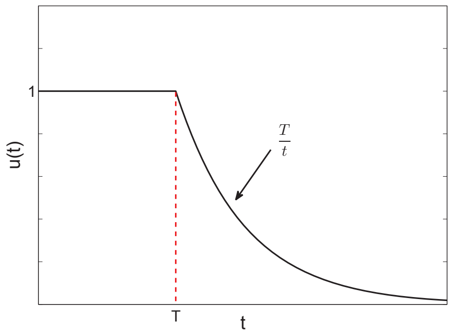

Fig. 1. Behavior of the Huber's function u with respect to the fixed threshold T, as a function of t.

The estimated covariance matrices can be further used in clustering algorithms, like k-means, or EM, and therefore the centroids of subsets of covariance matrices have to be determined. Let $\{\mathbf{M}_1, \ldots, \mathbf{M}_N\}$ be a random sample of N covariance matrices. The estimated centroid is the covariance matrix $\widehat{\bar{\mathbf{M}}}$ that minimizes the following cost function:

$$\widehat{\bar{\mathbf{M}}} = \underset{\bar{\mathbf{M}}}{\operatorname{argmin}}\, f(\bar{\mathbf{M}}). \quad (3)$$

In order to obtain an accurate representation of the dataset by the clusters' central values, an estimation algorithm robust to outliers has to be considered. Thus, the choice of the cost function $f(\bar{\mathbf{M}})$ needs special care. Starting from (3), different estimators can be defined, including the center of mass (CM) [13], [22] and the median (Med) [2], [17].

By drawing a parallel between covariance matrix estimation and centroid estimation, several similarities can be highlighted. First, the concept of center of mass is close to the one of SCM. More precisely, the SCM is the maximum likelihood estimate (MLE) for multivariate Gaussian distributions, while the center of mass is the MLE for Riemannian Gaussian distributions [12]. Second, even though it is a popular estimator, the CM, like the SCM, is easily influenced by the presence of outliers [16], [17]. Third, the problem of outliers can be solved by using robust estimators: the median [17] for centroids and its analogous, the FP estimator, for covariance matrices.

Considering all the previous common points and inspired by the theory of M-estimators [20], [21], [23], the next section introduces a novel centroid estimator on the manifold of covariance matrices, called the Huber's centroid.

III. The Huber's Estimator for Centroids Estimation

A. Definition of Huber's centroid

Starting from the M-estimator of covariance matrices given in (1), the M-estimator of a centroid is obtained by minimizing the following cost function:

$$f_u(\bar{\mathbf{M}}) = \frac{1}{N} \sum_{i=1}^{N} u\big(d(\bar{\mathbf{M}}, \mathbf{M}_i)\big) d^2(\bar{\mathbf{M}}, \mathbf{M}_i), \quad (4)$$

where $u(\cdot)$ is a positive-valued weight function, and d represents the Rao's Riemannian distance defined as [24] $d(\mathbf{M}_1, \mathbf{M}_2) = \left[\sum_i (\ln \lambda_i)^2\right]^{\frac{1}{2}}$, with λ_i, $i = 1 \ldots m$ being the eigenvalues of $\mathbf{M}_2^{-1}\mathbf{M}_1$.

Similar to the covariance matrix estimation problem, the weight function $u(\cdot)$ has to decrease towards zero. This condition is required in order to ensure that the outliers have a small contribution to the centroid estimation. Interestingly, note that the weight function u depends on the Riemannian distance for the centroid estimation problem, while u depends on the Mahalanobis distance for the covariance matrix estimation problem.

Based on the so-called Huber's function, recalled in (2), and on the cost function given in (4), the proposed Huber's centroid is the covariance matrix $\widehat{\bar{\mathbf{M}}}$ that minimizes:

$$\begin{aligned} f_H(\bar{\mathbf{M}}) = &\frac{1}{N} \sum_{i=1}^{N} d^2(\bar{\mathbf{M}}, \mathbf{M}_i)\ \mathbb{1}_{\{d(\bar{\mathbf{M}}, \mathbf{M}_i) \leq T\}} \\ &+ \frac{T}{N} \sum_{i=1}^{N} d(\bar{\mathbf{M}}, \mathbf{M}_i)\ \mathbb{1}_{\{d(\bar{\mathbf{M}}, \mathbf{M}_i\} > T\}}, \end{aligned} \quad (5)$$

where $\mathbb{1}_{\{a \leq b\}}$ is the indicator function, which equals 1 if $a \leq b$ and 0 otherwise. In addition, the threshold T discriminates between outliers and normal data.

In order to find the Huber's centroid, which is the minimum of (5), a gradient descent algorithm [25] is proposed. Then, the gradient of $f_H(\bar{\mathbf{M}})$ with respect to $\bar{\mathbf{M}}$, denoted by $\nabla f_H(\bar{\mathbf{M}})$, has to be computed as:

$$\begin{aligned} \nabla f_H(\bar{\mathbf{M}}) = &-\frac{2}{N} \sum_{i=1}^{N} \mathrm{Log}_{\bar{\mathbf{M}}}(\mathbf{M}_i)\ \mathbb{1}_{\{d(\bar{\mathbf{M}}, \mathbf{M}_i) \leq T\}} \\ &- \frac{T}{N} \sum_{i=1}^{N} \frac{\mathrm{Log}_{\bar{\mathbf{M}}}(\mathbf{M}_i)}{d(\bar{\mathbf{M}}, \mathbf{M}_i)}\ \mathbb{1}_{\{d(\bar{\mathbf{M}}, \mathbf{M}_i\} > T\}}, \end{aligned} \quad (6)$$

where $\mathrm{Log}_{\bar{\mathbf{M}}}(\cdot)$ is the Riemannian logarithm mapping. Once that this value is obtained, the centroid is recursively computed by:

$$\bar{\mathbf{M}}_{it+1} = \mathrm{Exp}_{\bar{\mathbf{M}}_{it}}(-s_{it}\ \nabla f_H(\bar{\mathbf{M}}_{it})), \quad (7)$$

with s_{it} being the descent step and $\mathrm{Exp}_{\bar{\mathbf{M}}}(\cdot)$ the Riemannian exponential mapping[1]. For each iteration $it + 1$, the value of s_{it} is determined by using the Armijo's backtracking procedure [27].

The computation procedure stops when a fixed number of iterations is reached, or when the gradient's norm is smaller than a predefined value. Thus, the Huber's centroid estimate is found.

B. Special cases of Huber's centroid

The value chosen for the threshold T in (5) may lead to the expression of two well-known centroid estimators, that are the

[1]Due to the page restriction length, the interested reader is referred to [12], [26] for a definition of the Riemannian exponential and logarithm mappings.

center of mass and the median. More precisely, if $T = \infty$ the cost function giving the center of mass is obtained by:

$$f_{CM}(\bar{\mathbf{M}}) = \frac{1}{N} \sum_{i=1}^{N} d^2(\bar{\mathbf{M}}, \mathbf{M}_i), \quad (8)$$

while $T = 0$ yields to the cost function of the median:

$$f_{Med}(\bar{\mathbf{M}}) = \frac{1}{N} \sum_{i=1}^{N} d(\bar{\mathbf{M}}, \mathbf{M}_i). \quad (9)$$

In other words, for a small number of outliers, the Huber's centroid behaves as the center of mass, while it is similar to the median in the presence of outliers.

From the computational point of view, it can be noticed that the first term in (6) corresponds to the cost function of the center of mass, while the second term represents the median.

Note that the median computation procedure may yield numerical instabilities due to the division by the distance $d(\bar{\mathbf{M}}, \mathbf{M}_i)$ in (6). If the estimated centroid $\bar{\mathbf{M}}$ is close to the sample $\mathbf{M}_i$, then $d(\bar{\mathbf{M}}, \mathbf{M}_i)$ becomes close to zero and the median cannot be defined. A possible solution for this situation is proposed in [16]: at each iteration, the observations that are too close from the estimated centroid are excluded. An important advantage of the Huber's centroid is the fact that this problem is automatically solved by choosing an appropriate value for the threshold T. In [28], a method for automatically compute the threshold's value is introduced, based on the concept of median absolute deviation.

C. Performance analysis

In this section, the estimation performance of the proposed method is analyzed on simulated data. The performed experiment investigates the influence of outliers on the Huber's centroid. The results are also compared to the center of mass and the median.

For this purpose, a set of N covariance matrices $\{\mathbf{M}_1, \ldots, \mathbf{M}_N\}$ of size $m \times m$ is considered. These matrices are i.i.d. samples from the Riemannian Gaussian distribution [12] of central value $\bar{\mathbf{M}}$ and dispersion σ. The probability density function of the RGD with respect to the Riemannian volume element is [12]:

$$p(\mathbf{M}|\bar{\mathbf{M}}, \sigma) = \frac{1}{Z(\sigma)} \exp\left\{ -\frac{d^2(\mathbf{M}, \bar{\mathbf{M}})}{2\sigma^2} \right\}, \quad (10)$$

where $Z(\sigma)$ is a normalising factor independent of the centroid $\bar{\mathbf{M}}$, and $d(\mathbf{M}, \bar{\mathbf{M}})$ is the Riemannian distance. The interested reader is referred to [12] to generate samples from an RGD.

In practice, a set of $N = 1000$ matrices of size 2×2 is created, knowing that its elements are obtained for a central value $\bar{\mathbf{M}}$ having the form:

$$\bar{\mathbf{M}}(i, j) = \rho^{|i-j|} \text{ for } i, j \in [\![1, m]\!]. \quad (11)$$

In this case, $\rho = 0.7$ and the dispersion parameter σ is set to 0.1. Further on, the dataset is corrupted by including aberrant data. These outliers are i.i.d. samples generated according to an RGD, having the dispersion parameter $\sigma_o = 0.1$ and the centroid $10 \times \mathbf{M}_o$, $\mathbf{M}_o$ being given by (11) with $\rho_o = 0.1$.

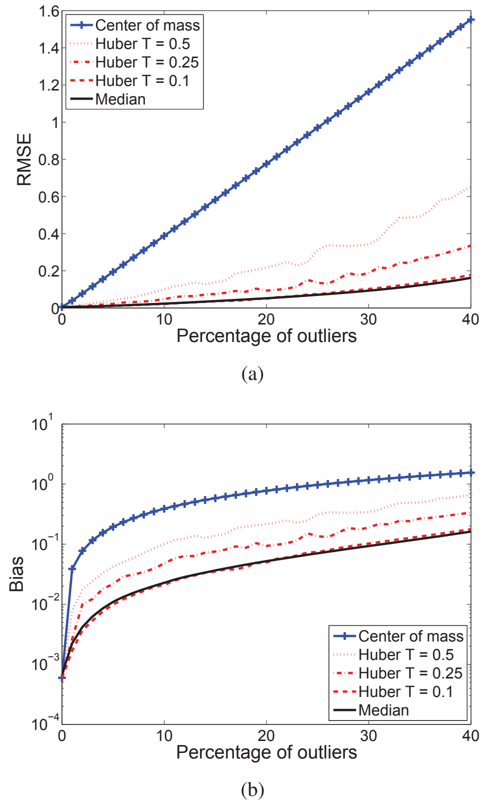

Fig. 2. Influence of the percentage of outliers on the estimation performance: (a) RMSE, (b) norm of the bias vector field.

In order to evaluate the centroid estimation methods, the concept of intrinsic analysis for statistical estimation [29]–[31] is used. Therefore, the definitions of intrinsic root-mean square error (RMSE) and intrinsic bias vector field are recalled next.

Let $\widehat{\mathbf{M}}$ be the estimated centroid of the real central value $\bar{\mathbf{M}}$. The intrinsic RMSE is given by [29]–[31]:

$$RMSE = \sqrt{E\left[d^2(\widehat{\mathbf{M}}, \bar{\mathbf{M}})\right]}, \quad (12)$$

where $d(\cdot)$ is the Riemannian distance. In addition, the bias vector field $\mathbf{b}(\bar{\mathbf{M}})$ of $\widehat{\mathbf{M}}$ is defined as [29]–[31]:

$$\mathbf{b}(\bar{\mathbf{M}}) = \text{Log}_{\bar{\mathbf{M}}} E_{\bar{\mathbf{M}}}\left[\widehat{\mathbf{M}}\right] = E\left[\text{Log}_{\bar{\mathbf{M}}} \widehat{\mathbf{M}}\right], \quad (13)$$

with $E_{\bar{\mathbf{M}}}\left[\widehat{\mathbf{M}}\right] = \text{Exp}_{\bar{\mathbf{M}}} E\left[\text{Log}_{\bar{\mathbf{M}}} \widehat{\mathbf{M}}\right]$. Since the bias vector field $\mathbf{b}(\bar{\mathbf{M}})$ in (13) is a covariance matrix, its norm is computed as:

$$||\mathbf{b}(\bar{\mathbf{M}})|| = \text{tr}\left(\left(\bar{\mathbf{M}}^{-1}\mathbf{b}(\bar{\mathbf{M}})\right)^2 \right). \quad (14)$$

The RMSE and the norm of the bias vector field are displayed in Figs. 2(a) and 2(b) respectively. In these figures, the influence of the percentage of outliers on the estimation methods is shown. The center of mass, the median and the Huber's estimator with $T = 0.1$, 0.25 and 0.5 are considered, knowing that the percentage of outliers varies from 0 to 40%. By analyzing these curves, it can be observed that the center of mass (in blue) is the estimator the most influenced by the presence of outliers. On the other hand, the median (in black) and the Huber's centroid (in red) demonstrate their robust behavior. In addition, the Huber's centroid confirms the fact that it can be interpreted as a trade-off between the center of mass and the median.

IV. Application to Texture Image Classification

In this section, the performances of the Huber's estimator for centroids computation are analyzed in the context of texture image classification, by using the MIT Vision Texture (VisTex) database [32].

The experiment is designed in order to evaluate the influence of outliers on the correct classification rate. Therefore, the original VisTex database is modified to contain aberrant data. First, each image in the database is divided into 169 patches of 128×128 pixels, with an overlap of 32 pixels. Next, abnormal data are introduced. For each image, between 0 and 60 patches have their intensity modified, by applying a gradient of luminosity. The corrupted patches are further considered as being outlier patches.

The new database is used for supervised classification, assuming that it contains 40 classes. Thus, for each patch a feature vector has to be extracted. In this case, the spatial dependence of the wavelet coefficients is considered. Therefore, the patches are filtered by using the Daubechies' db4 wavelet, with 2 scales and 3 orientations. In addition, to capture the textural information, two neighborhoods (2×1 and 1×2) are extracted for each pixel in the 6 wavelet subbands. These pixels are modeled by zero-mean multivariate Gaussian distributions, characterized by their sample covariance matrix. In the end, each patch is represented by a feature vector F containing 12 covariance matrices of size 2×2.

The patches (the set of feature vectors) are equally and randomly divided into training and testing subsets. This division is iterated 100 times and for each training class, the centroid of the f^{th} feature of class c, denoted $\widehat{\mathbf{M}}_{c,f}$, is computed. The classification is performed next, based on the Bayes rule. More precisely, a test patch t is affected to the class c representing the minimum over c of [9], [10]:

$$\sum_{f=1}^{F} d^2(\mathbf{M}_{t,f}, \widehat{\mathbf{M}}_{c,f}), \qquad (15)$$

where $\mathbf{M}_{t,f}$ is the SCM of the f^{th} feature of the test patch t.

The centroid computation is carried out by using the Huber's centroid estimator with $T = 0.1, 0.25, 0.5$ and its two special cases: the center of mass ($T = \infty$) and the median ($T = 0$). For all these methods, the classification results are reported in Fig. 3. The correct classification rate is represented as a function of the number of outlier patches per class. By analyzing this figure, several remarks can be retained. First, for no aberrant data, all the methods perform identically. Second, the center of mass (in blue) is strongly influenced by the presence of outliers and its performance decreases rapidly. Third, the median (in black) and the Huber's centroid (in red) demonstrate their robust behavior, being less sensitive to aberrant data. And finally, the Huber's centroid gives classification rates that are between those obtained with the center of mass and the median.

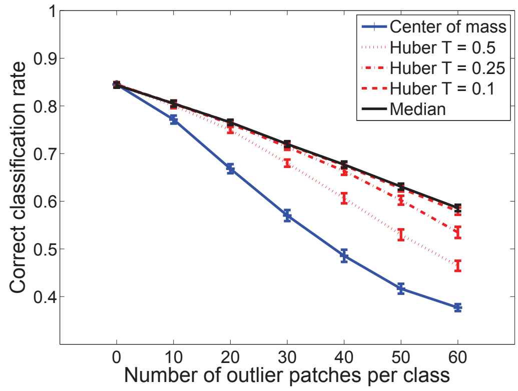

Fig. 3. Correct classification rate when centroids are estimated by using the center of mass, the median and the Huber's centroid with $T = 0.1, 0.25, 0.5$.

V. Conclusion

In this article, a novel robust centroid estimator is proposed. Based on the theory of M-estimators, this estimator, called the Huber's centroid, is defined by using the Huber's function. A gradient descent algorithm is also proposed to estimate it. Since this estimator generalizes both the center of mass and the median, the proposed Huber's centroid is a good trade-off between efficiency and robustness. In addition, by carefully choosing its unique parameter T, the numerical instabilities that may occur for the median computation are avoided.

The properties of the proposed estimator have been confirmed by the experiments. Its robustness to outliers has been investigated first on simulated data. Next, it has been applied to texture image classification on a modified VisTex database. For both experiments, for no outliers, all the three methods perform identically, but in the presence of aberrant data, the Huber's centroid and the median yield the best results.

In the considered experiments, no convergence problems have been encountered. Nevertheless, further works will investigate the convergence of the gradient descent. In addition, a special interest of the proposed centroid estimator will be dedicated to the construction of codebooks in patch-based classification algorithms [11].

Acknowledgment

This study has been carried out with financial support from the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02). It has also been supported by the French Foreign Affairs and International Development Ministry and by the Executive Agency for Higher Education, Research, Development and Innovation Funding Romania, under the projects 32619VL and PNII Capacitati 779/27.06.2014.

References

[1] E. Ollila and V. Koivunen, "Robust antenna array processing using M-estimators of pseudo-covariance," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, vol. 3, Sept 2003, pp. 2659–2663.

[2] L. Yang, M. Arnaudon, and F. Barbaresco, "Riemannian median, geometry of covariance matrices and radar target detection," *European Radar Conference (EuRAD)*, pp. 415–418, 2010.

[3] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions," *Signal Process.*, vol. 104, pp. 381–386, Nov. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.sigpro.2014.04.002

[4] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, Sept 2011.

[5] J. Robinson, "Covariance matrix estimation for appearance-based face image processing," *Proceedings of the British Machine Vision Conference 2005*, pp. 389–398, 2005. [Online]. Available: http://www.intuac.com/userport/john/pubs/covestbmvc.pdf

[6] K. Mader and G. Reese, "Using covariance matrices as feature descriptors for vehicle detection from a fixed camera," *ArXiv e-prints*, Feb. 2012.

[7] R. de Luis-García, C.-F. Westin, and C. Alberola-López, "Gaussian mixtures on tensor fields for segmentation: Applications to medical imaging," *Computerized Medical Imaging and Graphics*, vol. 35, no. 1, pp. 16–30, 01 2011.

[8] P. Formont, F. Pascal, G. Vasile, J. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric SAR images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 567–576, June 2011.

[9] 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.

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

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

[12] S. Said, L. Bombrun, Y. Berthoumieu, and J. Manton, "Riemannian Gaussian distributions on the space of symmetric positive definite matrices," *available on arxiv via http://arxiv.org/abs/1507.01760*, 2015.

[13] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977. [Online]. Available: http://dx.doi.org/10.1002/cpa.3160300502

[14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. Springer, Oct. 2007.

[15] B. Afsari, "Riemannian lp center of mass: existence, uniqueness and convexity," *Proceedings of the American Mathematical Society*, vol. 139, no. 2, pp. 655–673, 2011.

[16] L. Yang, "Riemannian median and its estimation," *LMS Journal of Computation and Mathematics*, vol. 13, pp. 461–479, 2010.

[17] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "The geometric median on riemannian manifolds with application to robust atlas estimation," *Neuroimage*, vol. 45, no. 1, pp. S143–S152, March 2009.

[18] M. Arnaudon, F. Barbaresco, and L. Yang, "Riemannian medians and means with applications to radar signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 4,, pp. 595–604, 2013.

[19] H. Hajri, I. Ilea, S. Said, L. Bombrun, and Y. Berthoumieu, "Riemannian laplace distribution on the space of symmetric positive definite matrices," *Entropy*, vol. 18, no. 3, p. 98, 2016. [Online]. Available: http://www.mdpi.com/1099-4300/18/3/98

[20] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[21] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 03 1987. [Online]. Available: http://dx.doi.org/10.1214/aos/1176350263

[22] F. Nielsen and R. Bhatia, *Matrix Information Geometry*. Springer Berlin Heidelberg, 2012.

[23] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, Jan. 1976.

[24] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 3, pp. 180–193, 2012.

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

[26] 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. [Online]. Available: http://dx.doi.org/10.1007/s11263-005-3222-z

[27] L. Armijo, "Minimization of functions having lipschitz continuous first partial derivatives." *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966. [Online]. Available: http://projecteuclid.org/euclid.pjm/1102995080

[28] I. Ilea, L. Bombrun, R. Terebes, M. Borda, and C. Germain, "An M-estimator for robust centroid estimation on the manifold of covariance matrices," *submitted to IEEE Signal Processing Letters*, 2016.

[29] J. M. Oller and J. M. Corcuera, "Intrinsic analysis of statistical estimation," *The Annals of Statistics*, vol. 23, no. 5, pp. 1562–1581, 1995. [Online]. Available: http://dx.doi.org/10.2307/2242534

[30] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, May 2005.

[31] G. Garcia and J. M. Oller, "What does intrinsic mean in statistical estimation?" *Statistics and Operations Research Transactions*, vol. 30, no. 2, pp. 125–170, 2006.

[32] "Vision Texture Database," MIT Vision and Modeling Group. Available: http://vismod.media.mit.edu/pub/VisTex.