Beyond Euclidean Gaussian : Application To The Classification Task Based On Riemannian Feature Coding

Yannick Berthoumieu

Co-Workers: Salem Said, Lionel Bombrun, Jialun Zhou, Sara Akodad

IMS Lab - Signal and Image processing Group
Université de Bordeaux / France

e-mail : yannick.berthoumieu@ims-bordeaux.fr

IPAM Workshop I: Geometric Processing 1-5 April 2019

I. Context : Video, Image and Signal Classification

Multivariate Data Analysis

$$X_n \in \mathbb{R}^m$$

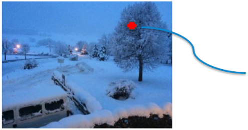
Problems

Unsupervised or supervised clustering, dictionary learning, etc.

Feature-based image classification

Brain interface using EEG signals

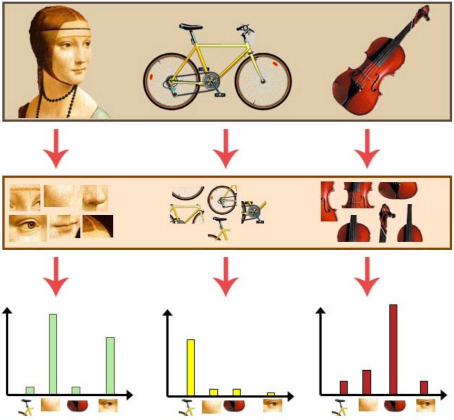
Feature Space X_n



$$X(x,y) = \left[I(x,y), \left|\frac{\partial I(x,y)}{\partial x}\right|, \left|\frac{\partial I(x,y)}{\partial y}\right|, \left|\frac{\partial^2 I(x,y)}{\partial x^2}\right|, \left|\frac{\partial^2 I(x,y)}{\partial y^2}\right| \right]^T$$

- Multimodality (RGB, IR, Radar, etc.)
- Filter banc (Wavelet, Gabor, etc.)
- SIFT, SURF, etc.

Dictionary Approches

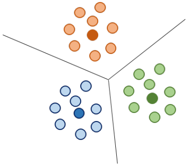


@Jan Kundrac

Inference Regarding on Multivariate Data Analysis

Mixture modeling

$$p(X|(\varpi_k, \bar{X}_k, \sigma_k)_{1 \leq k \leq K}) = \sum_{k=1}^{K} \varpi_k \times p(X|\,\bar{X}_k, \sigma_k)$$



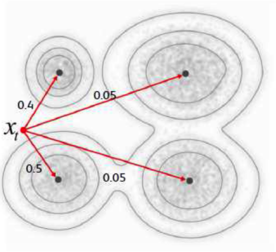
- Gaussian density or other kernels
- Universal modelling
- Expectation/Maximization algorithm [Dempster 1977]

Coding and Classification

Bag of words [F. Moosmann 2006], Fisher vectors [F. Peronnin 2006], Vector of Locally Aggregated Descriptors [H. Jegou 2012]

Fisher Vectors and Information Kernel

$$\mathcal{G}_{\theta}^{\mathcal{X}} = I_{\theta}^{-1/2} \mathcal{U}_{\mathcal{X}}^{\theta}$$



- Parameter set : $\theta = \left(\bar{X}_k, \sigma_k\right)$
- Soft assignment

$$\gamma^k(x) = \frac{\varpi_k u_{\theta_k}(x_n)}{\sum_{l=1}^{K} \varpi_L u_{\theta_l}(x_n)}$$

- Score function :

$$\mathcal{U}_{\mathcal{X}}^{\theta} = \sum_{n=1}^{N} \nabla_{\theta} log\left[u_{\theta}(x_n)\right]$$

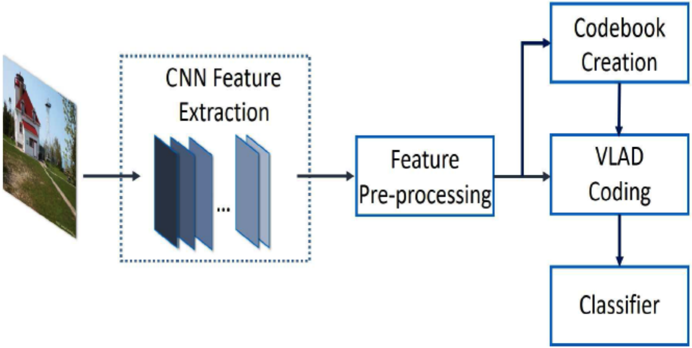
- Fisher Information Matrix (FIM) :
$I_{\theta} = \mathbf{E}_{\mathcal{X}}[\nabla_{\theta} log\left[u_{\theta}(x_n)\right]^T \nabla_{\theta} log\left[u_{\theta}(x_n)\right]]$

CNN and Dictionary Coding

Mixture modeling

Coding and Classification

Deep Learning + Dictionary Coding [F. Perronnin 2015-J. Y.-H. Ng 2015, Q. Li 2017]



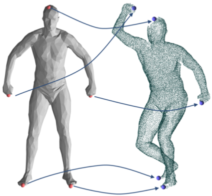
@Li, IEEE CSE, 2017.

II. Focus : Covariance Matrix
Symmetric Definite Matrix Manifold

Inference Regarding on Manifold-valued Data Analysis

Multivariate data

$$X_n \in \mathcal{M}_m$$



Human Posture
Spatial Transformation

Image Classification
Feature Covariance Matrix

Applications

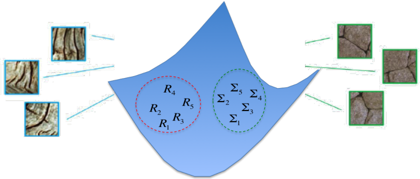
Brain interface, Radar detection, Computer vision, Texture classification, MRI imaging, ...

Manifold-valued Data Analysis

Samples of covariance matrices

$$M_n = \mathbf{E}[(X-\mu)(X-\mu)^*] \in \mathcal{H}_m \text{ where } X = \begin{bmatrix} f_0 & f_1 & \cdots & f_m \end{bmatrix}^*$$

$\mathcal{H}_\mathrm{m}$: manifold of the Hermitian symmetric positive definite (HSPD) matrices



- Various levels of invariance (linear transform, illumination, etc.)
- Fusion of multiple features
- Less sensitive to noisy data

Problems

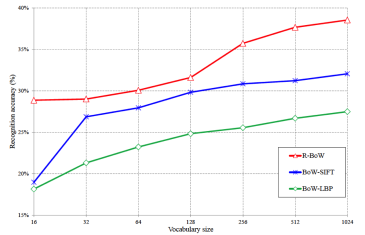
Unsupervised or supervised clustering, dictionary learning, signal denoising, image superresolution ...

Manifold-valued Data Analysis

Samples of covariance matrices

$$M_n = \mathbf{E}[(Y-\mu)(Y-\mu)^*] \in \mathcal{H}_m \text{ where } Y = \begin{bmatrix} f_0 & f_1 & \cdots & f_m \end{bmatrix}^*$$

$\mathcal{H}_m$: manifold of the Hermitian symmetric positive definite (HSPD) matrices



- Various levels of invariance (linear transform, illumination, etc.)
- Fusion of multiple features
- Less sensitive to noisy data

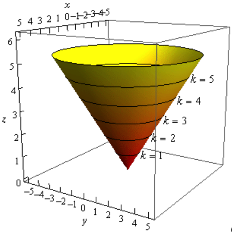
Problems

Unsupervised or supervised clustering, dictionary learning, signal denoising, image superresolution ...

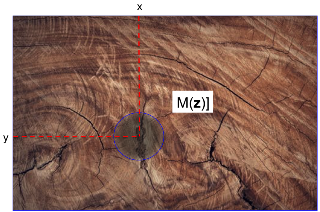
SPD Manifold

Case for m = 2 : $\mathcal{H}_2 = \{M = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, a > 0, ab - c^2 > 0\}$

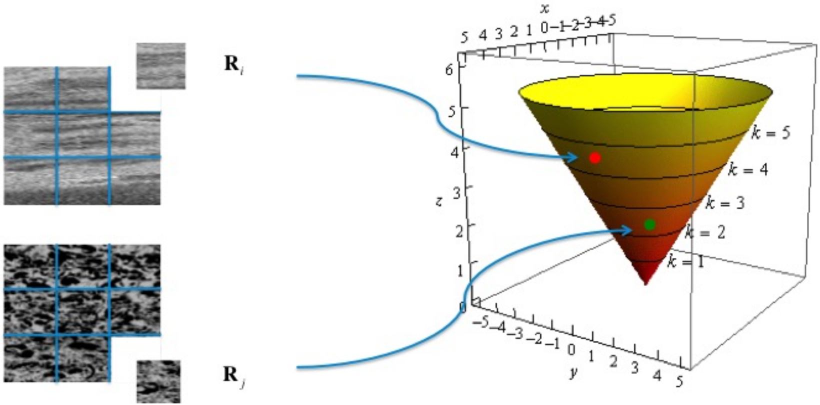
With the change of variables $u = \frac{1}{2}(a + b), v = \frac{1}{2}(a - b)$, the conditions of positive definiteness become :



$c^2 + v^2 < u^2$ with $u > 0$



Classification task from SPD set



Problems

Fisher coding extension to Riemannian case

III. Riemannian Geometry : SPD Gaussian Density

Manifold-Valued Samples and PDF

$$\mathcal{H}_m = \{M / x^T M x > 0 \; \forall x \in \mathbb{R}^m, M = M^*\}$$

$$M \in \mathcal{H}_m \text{ and } M \sim Wish_m(S, \nu)$$

Pdf for M WISHART modeling

$$Wish_m(M|\; S, \nu) = \frac{|M|^{\frac{\nu-m-1}{2}}}{2^{\frac{\nu m}{2}} \; |S|^{\frac{\nu}{2}} \; Gam_p\left(\frac{\nu}{2}\right)} \exp\left[-\frac{1}{2} tr\left(S^{-1}M\right)\right]$$

- Multivariate Generalization of the χ^2 Distribution
- Expected Value $\overline{M} = \mathbf{E}[M] = \nu S$

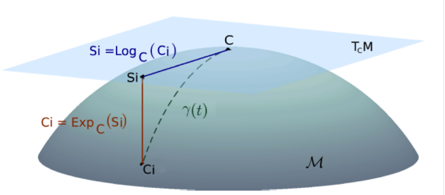
Problems

Dictonnary Coding versus Gaussian Mixture Modeling on the SPD manifold

Probability Density Function and Manifold

[Radon-Nikodyn theorem] : Let $(\mathbb{M}, \mathcal{F}(\mathbb{M}), P)$ a probability space, $f : \mathbb{M} \mapsto [0, +\infty[$ and $P(\mathcal{F}(\mathbb{M})) = \int_{\mathcal{F}(\mathbb{M})} f dv$ then $f = \frac{dP}{dv}$ is called the density of P w.r.t. the measure μ.

- Consider $\mathbb{R}^n$ and the coordinate set $\{x_i\}_{1 \leq i \leq n}$ the volume measure is $dv = \prod_{i=1}^{n} x_i$.
- Consider $\mathbb{M}(n)$ and the coordinate set $\{x_i\}_{1 \leq i \leq n}$ the volume measure is $dv = det(G)^{1/2} \prod_{i=1}^{n} dx_i$.



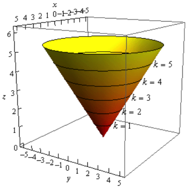
Metric expression : $ds^2 = \sum_{i,j=1}^{n} g_{ij} dx_i dx_j$

Volume Definition for SPD Manifold

$$\mathcal{H}_m = \{M / x^T M x > 0 \; \forall x \in \mathbb{R}^m, M = M^*\}$$

[Amari, Bhatia, Bhatia & Holbrook, Calvo & Oller, Lawson & Lim, Petz, Zerai, Terras, Moakher, Pennec, Vemuri etc.]

Information Geometry : Hessian of SHANNON ENTROPIE function $H(x) = -ln(det(M))$



- $ds^2(M) = trace(M^{-1}dM \; M^{-1}dM)$
- $dv(M) = det(M)^{-m} \prod_{i \leqslant j} \Re(dM_{i,j}) \prod_{i<j} \Im(dM_{i,j})$

$$dv(M) = c_1 \times \prod_{i=1}^{m} \lambda_i^{-m} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{m} d\lambda_i \prod_{i<j} \Re(\theta_{i,j}) \prod_{i<j} \Im(\theta_{i,j})$$

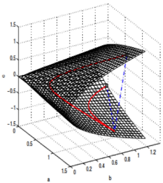
with $\theta_{i,j} = [U^*(dU)]_{i,j}$ and $U \in \mathcal{U}(N)$

Riemannian Gaussian Law

$$d^2_{\mathcal{H}_m}(A, B) = trace(Log^2\ (A^{-1}B))$$

$$p(M|\overline{M}, \sigma) = \frac{1}{Z(\overline{M}, \sigma)} \exp\left[-\frac{d^2_{\mathcal{H}_m}(M, \overline{M})}{2\sigma^2}\right]$$

where $Z(\overline{M}, \sigma) = \int_{\mathcal{H}_m} \exp\left[-\frac{d^2_{\mathcal{H}_m}(M,\overline{M})}{2\sigma^2}\right] dv(M)$



- $d^2_{\mathcal{H}_m}(A, B)$ Riemannian distance or divergence
- $dv(A)$ Riemannian volume element on $\mathcal{H}_m$
- $\overline{M}$ barycenter and σ dispersion

2017 Riemannian Gaussian Distributions on the SPD Matrices. IEEE Trans. Information Theory
2018 Fisher Vector Coding for Covariance Matrix Descriptors, J. Imaging

Normalization coefficient (1/2)

$$Z(\sigma) = c_1 \times \int_{\mathcal{U}(m)} \int_{(0,+\infty)^m} \prod_{i=1}^{m} \lambda_i^{-m} e^{-\frac{log^2 \lambda_i}{2\sigma^2}} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{m} d\lambda_i U^*(dU)$$

Haar measure

$$Vol[\mathcal{U}(m)] = \int_{\mathcal{U}(N)} U^*(dU) = \frac{2^m \pi^{m(m+1)/2}}{\prod_{i=1}^{m} Gam(i)}$$

$$Z(\sigma) = C \times \int_{(0,+\infty)^m} \prod_{i=1}^{m} e^{-\frac{log^2 x_i}{2\sigma^2}} \prod_{i<j} (x_i - x_j)^2 \prod_{i=1}^{m} dx_i$$

with $x_i = e^{m\sigma^2} \lambda_i$

Normalization coefficient (2/2)

Option 1

Algorithm 1 Monte Carlo integration for tabuled $Z(\sigma)$

1: Generate U from a uniform distribution on the group of unitary matrices of size $m \times m$
2: Generate $(r_1, \cdots, r_m) \in \mathbb{R}^m$ from the distribution $\propto e^{-|r|^2/2\sigma^2} \prod_{i<j} \sinh^2(|r_i - r_j|/2)$
3: $X \leftarrow U \cdot \text{diag}(e^{r_1}, \cdots, e^{r_m})$
4: $Y \leftarrow \bar{Y}^{1/2} \cdot X$

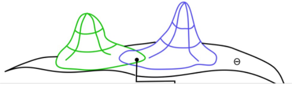
Option 2

$$\int \cdots \int \prod_{i=1}^{m} \omega(x_i) \prod_{i<j} (x_i - x_j)^2 dx_1 \cdots dx_m$$

$$=$$

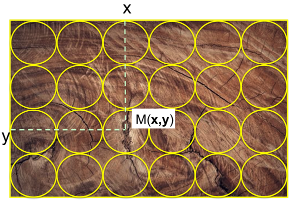
$$m! a_0^{-2m} \prod_{i=1}^{m-1} \left[\frac{a_{i-1}}{a_i}\right]^{2(m-i)}$$

SPD Gaussian Mixture



$$p(M|(\varpi_\mu, \overline{M}_\mu, \sigma_\mu)_{1\leq\mu\leq K}) = \sum_{\mu=1}^{K} \varpi_\mu \times p(M|\,\overline{M}_\mu, \sigma_\mu)$$

with $\sum_{\mu=1}^{K} \varpi_\mu = 1$



- $\{\hat{M}(\mathbf{i})\}_{i=1..N} \sim p(M|(\varpi_\mu, \overline{M}_\mu, \sigma_\mu)_{1\leq\mu\leq K})$
- $\omega_\mu(M_n) = \frac{\varpi_\mu \times p(M_n|\,\overline{M}_\mu, \sigma_\mu)}{\sum_{s=1}^{K} \varpi_s \times p(M_n|\,\overline{M}_s, \sigma_s)}$
- $N_\mu(\vartheta) = \sum_{n=1}^{N} \varpi\mu(M_n)$

Parameter estimation

EM algorithm

- Assign to $\hat{\varpi}_\mu$ the value $\hat{\varpi}_\mu = N_\mu(\hat{\vartheta})/N$
- Assign to $\hat{\overline{M}}_\mu$ the value

$$\hat{\overline{M}}_\mu = \arg\min_{M} \sum_{n=1}^{N} \varpi_\mu(M_n)\, d^2_{\mathcal{H}_m}(M, M_n)$$

  Gradient descent or Newton Riemannian methods [S.T. Smith 1993, M. Moakher 2003, etc.]

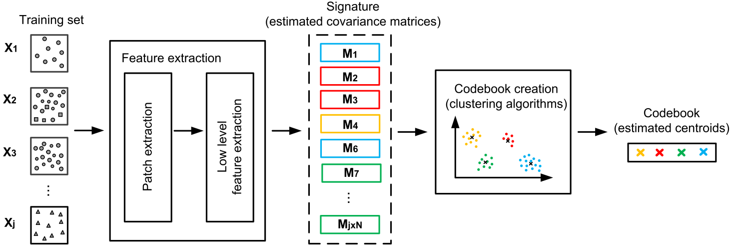
- Assign to $\hat{\sigma}_\mu$ the value

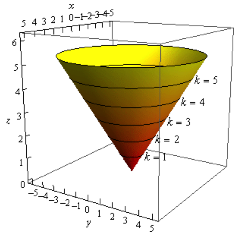
$$\hat{\sigma}_\mu = \phi\left(N_\mu^{-1}(\hat{\vartheta}) \times \sum_{n=1}^{N} \varpi_\mu(M_n)\, d^2_{\mathcal{H}_m}(\overline{M}_\mu, M_n)\right)$$

  where the function ϕ is the inverse of $\sigma \mapsto \sigma^3 \times \frac{d}{d\sigma} \log \zeta(\sigma)$.

IV. Coding

Coding what and where ?





$$\begin{cases} \left\| vec(log(M_i)) - vec(log(\overline{M}_k)) \right\|_2^2 \\ \left\| log(\overline{M}_k^{-1} M) \right\|_F^2 \end{cases}$$

Log-vectorization Euclidean Distance - Isotropic Case

- Vectorization

$$M = \begin{bmatrix} x & & x \\ x & \ddots & x \\ x & & x \end{bmatrix} \xrightarrow{\text{vec(M)}} m = \begin{bmatrix} x \\ \vdots \\ x \end{bmatrix}$$

[m x m] [m(m+1)/2,1]

- Euclidean Mixture for vectors

$$p(m|(\varpi_\mu, \overline{m}_\mu, \sigma_\mu)_{1\leq\mu\leq m}) = \sum_\mu \varpi_\mu \times p(m|\,\overline{m}_\mu, \sigma_\mu)$$

- Fisher scores

$$U_{\boldsymbol{\mathcal{X}}} = \nabla_\theta \log p(\boldsymbol{\mathcal{X}}|\theta) = \nabla_\theta \sum_{n=1}^{N} \log p(\mathbf{X}_n|\theta)$$

- Fisher vectors

$$\mathcal{G}_\theta^{\boldsymbol{\mathcal{X}}} = I_\theta^{-1/2} \nabla_\theta \log p(\boldsymbol{\mathcal{X}}|\theta)$$

Riemannian Distance - Isotropic Case

- Riemannian Mixture for SPD matrices

$$p(M|(\varpi_\mu, \overline{M}_\mu, \sigma_\mu)_{1\leq\mu\leq M}) = \sum_\mu \varpi_\mu \times p(M|\overline{M}_\mu, \sigma_\mu)$$

- Fisher scores

$$U_{\boldsymbol{\mathcal{X}}} = \nabla_\theta \log p(\boldsymbol{\mathcal{X}}|\theta) = \nabla_\theta \sum_{n=1}^{N} \log p(\mathbf{X}_n|\theta)$$

- Fisher vectors

$$\mathcal{G}_\theta^{\boldsymbol{\mathcal{X}}} = I_\theta^{-1/2} \nabla_\theta \log p(\boldsymbol{\mathcal{X}}|\theta)$$

Zanini et al. "Riemannian online algorithms for estimating mixture model parameters". In Geometric Science of Information, Proceedings of the Third International Conference, GSI 2017

Fisher Information Matrix

$$I_F(\Theta) = -E_\chi \left[\frac{\partial^2 \log p(M;\Theta)}{\partial\Theta\partial\Theta^T} \right]$$

with $\Theta = \{s_\mu, \overline{M}_\mu, \eta_\mu\}_{\mu=1}^{K}$

- $s_\mu^2 = \varpi_\mu \rightarrow s = [s_1, ..., s_K] \in \mathbb{S}^{K-1}$ (Sphere)
- $\overline{M}_\mu \in \mathcal{SPD}_m$
- $\eta_\mu = -\frac{1}{2\sigma_\mu^2} \in \mathbb{M}^{*-}$

$$\mathcal{M}(\Theta) = \mathcal{M}(s) \times \mathcal{M}(\overline{X}_c) \times \mathcal{M}(\eta)$$

$$I_F(\Theta) = \begin{pmatrix} I(s) & & \\ & I(\overline{X}) & \\ & & I(\eta) \end{pmatrix}$$

Symmetric Spaces and Irreducible Space

Globally Symmetric Space

A simply connected riemannian space $\mathcal{M}(m)$ is said to be a globally symmetric if we have

- geodesic symmetry $f(\gamma(-t)) = \gamma(t)$ for a fixed point X such as $\gamma(0) = X$
- geodesic symmetries are isometric

Irreducible spaces

- A symmetry space is irreducible if it is not the product of two or more symmetric spaces
- A simply connected symmetric space is a product of irreducible spaces

Schur's lemma

Let ψ linear map applied on a vector field over a symmetric irreducible space then ψ is scalar i.e. $\psi = \alpha * \mathbf{I}(m)$ with $\mathbf{I}(m)$ the Identity matrix.

Fisher Information Matrix

Sphere

$$I_{kl}(s) = E_{\chi}[v_k(z,s)v_l(z,s)] = 4(\delta_{kl} - s_k s_l) \implies$$

$$\boxed{I_F(s) = 4 * \mathbf{I}(K)}$$

$$SPD(m) \implies \mathbb{R} \times SP(m)$$

$$X \implies (M_1, M_2)$$

- $M_1 = \log\det M \implies \boxed{I_F(\overline{M}_{1\mu}) = \frac{\varpi_\mu}{\sigma_\mu^3}}$
- $M_2 = e^{-M_1/m} M \implies \boxed{I_F(\overline{M}_{2\mu}) = \frac{\varpi_\mu}{m\sigma_\mu^4}[log' Z(\sigma_\mu) - \sigma_\mu]\mathbf{I}(m)}$

Update Rule for weight coefficients

$$\widehat{\Theta}^{(n+1)} = \text{Exp}_{\widehat{\Theta}^{(n)}}(\xi^{(n+1)})$$

with $\xi^{(n+1)} = \gamma^{(n+1)} I_F^{-1}(\widehat{\Theta}^{(n)}) u((M_n)_{n=1..N}; \widehat{\Theta}^{(n)})$

- $F(.)$ Exponential map
- $\{\gamma^{(n)}\}_N$ a decreasing sequence
- $u((M_n)_{n=1..N}; \widehat{\Theta}^{(n)}) = \nabla_{\widehat{\Theta}^{(n)}} \log p((M_n)_{n=1..N}; \widehat{\Theta}^{(n)})$

Sphere

$$\widehat{s}_\mu^{(n+1)} = \widehat{s}_\mu^{(n)} \cos(\|\xi^{(n+1)}\|) + \frac{\xi^{(n+1)}}{\|\xi^{(n+1)}\|} \sin(\|\xi^{(n+1)}\|)$$

$$\xi^{(n+1)} = \frac{\gamma^{(n+1)}}{2} \left(\frac{h_k((M_n)_{n=1..N}; \widehat{\Theta}^{(n)})}{\widehat{s}_k} - \widehat{s}_k \right)$$

Update Rule for SPD barycenter components

$$\widehat{\Theta}^{(n+1)} = \mathrm{Exp}_{\widehat{\Theta}^{(n)}}(\xi^{(n+1)})$$

SPD(m)

$$\widehat{\overline{X}}_{1k}^{(n+1)} = \widehat{\overline{X}}_{1k}^{(n)} + \xi_1{}^{(n+1)}$$

$$\widehat{\overline{X}}_{2k}^{(n+1)} = \left(\widehat{\overline{X}}_{2k}^{(n)}\right)^{1/2} \exp\left(\left(\widehat{\overline{X}}_{2k}^{(n)}\right)^{-1/2} \xi_2{}^{(n+1)} \left(\widehat{\overline{X}}_{2k}^{(n)}\right)^{-1/2}\right)\left(\widehat{\overline{X}}_{2k}^{(n)}\right)^{1/2}$$

$$\xi_1 = \gamma^{(n+1)} I_F^{-1}(\widehat{\overline{X}}_{1k}^{(n)}) \frac{h_k(X_{n+1};\widehat{\Theta}^{(n)})}{\widehat{\sigma}_k^{2(n)}} \left(\widehat{\overline{X}}_{2k}^{(n)} - X_{1,n+1}\right)$$

$$\xi_2 = \gamma^{(n+1)} I_F^{-1}(\widehat{\overline{X}}_{2k}^{(n)}) \frac{h_k(X_{n+1};\widehat{\Theta}^{(n)})}{\widehat{\sigma}_k^{2(n)}} \widehat{\overline{X}}_{2k}^{(n)} \log\left(\left(\widehat{\overline{X}}_{2k}^{(n)}\right)^{-1} X_{2,n+1}\right)$$

V. Results

Results Texture Database VISTEX(40 classes)

model	number of modes (K)	on $\mathcal{H}_m$
RGD	K=1	78.24 ± 0.36
	K=3	90.24 ± 0.40
	BIC	91.91 ± 0.45
Wishart	K=1	81.93 ± 0.51
	K=3	86.07 ± 0.61
	BIC	86.14 ± 0.44

TABLE – Classification performance (κ accuracy) on the VisTex database.

Results

Dataset IIa of BCI competition IV

- 9 subjects who performed 576 trials of right-hand (RH), left-hand (LH), tongue (TO) and both feet (BF) motor imagery
- 22 electrodes are used (Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, POz).

Brain interface using EEG signals

Results

2018 Zanini et al. Transfer learning : a Riemannian geometry framework with applications to brain computer interfaces, IEEE Transactions on Biomedical Engineering.

	Adaptive RK-SVM		RK-SVM			SVM vec	CSP + LDA
	$\mathbf{C}_{\text{ref}} = \mathfrak{G}$	$\mathbf{C}_{\text{ref}} = \mathfrak{A}$	$\mathbf{C}_{\text{ref}} = \mathfrak{G}$	$\mathbf{C}_{\text{ref}} = \mathfrak{A}$	$\mathbf{C}_{\text{ref}} = \mathbf{I}_E$		
LH/RH	**82.7 (14.3)**	81.6 (14.8)	79.9 (13.4)	80.6 (13.2)	80.6 (12.8)	73 (14.8)	75.9 (19.2)
LH/BF	**89.5 (10.7)**	88.3 (10.2)	87.3 (11.7)	85.8 (14)	85 (12.9)	78.2 (11.2)	80.8 (16.1)
LH/TO	**88.7 (11.8)**	86.7 (11.9)	86.9 (11.9)	85.6 (12)	85 (12)	81 (13)	82.9 (16.8)
RH/BF	**87.3 (11)**	87.1 (10.2)	85.9 (10.4)	83.6 (12.6)	80.5 (14.4)	77 (11.3)	84.2 (11.8)
RH/TO	**88.3 (12.1)**	86.9 (11.6)	86 (12.1)	83.5 (12.3)	83.6 (12.6)	77.5 (14.1)	81.9 (15.6)
BF/TO	**79.5 (9.6)**	77.6 (8.7)	77.2 (7.9)	74.2 (6.5)	75.5 (9.1)	68.5 (9.1)	73.4 (8.1)
mean	**86** (4)	84.8 (4.2)	83.9 (4.2)	82.2 (4.4)	81.7 (3.6)	75.9 (4.4)	79.9 (4.3)

Table 1: Average classification accuracy (and standard deviation) across the 9 subjects for 6 pairs of mental tasks.

Results

BoRW [36]	85.9 ± 0.01
RVLAD [11]	82.8 ± 0.02
RFS [40]:$\bar{M}, \sigma, \omega$	91.2 ± 0.01
RFV:$\bar{M}, \sigma, \omega$	93.2 ± 0.01

Riemannian Fisher Score (RFS) and Fisher Vector (RFV) with SVM classifier.

Results

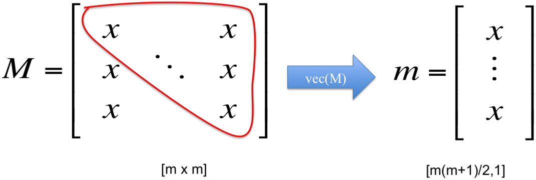
Database	Number of Classes	Number of Images per Class	Total Number of Images	Dimension
VisTex	40	64	2560	64 × 64 pixels
Brodatz	112	25	2800	128 × 128 pixels
Outex	68	20	1380	128 × 128 pixels
USPtex	191	12	2292	128 × 128 pixels

Database	Isotropic Model, Log-Euclidean Metric	Isotropic Model, Affine Invariant Riemannian Metric
VisTex	88.7 ± 0.01	91.3 ± 0.01
Brodatz	87.1 ± 0.01	92.9 ± 0.01
Outex	83.2 ± 0.01	85.4 ± 0.01
USPtex	81.5 ± 0.01	87.0 ± 0.01

Considering respectively isotropic Log-Euclidean and Riemannian Fisher Vector with SVM classifier.

VI. Perspectives

Log-vectorization Euclidean Versus Full Covariance Matrix



- Riemannien Mixture for SPD matrices

$$p(m|(\varpi_\mu, \overline{m}_\mu, \sigma_\mu)_{1\leq\mu\leq m}) = \sum_\mu \varpi_\mu \times p(m|\,\overline{m}_\mu, R_\mu)$$

- Fisher scores

$$U_{\boldsymbol{\mathcal{X}}} = \nabla_\theta \log p(\boldsymbol{\mathcal{X}}|\theta) = \nabla_\theta \sum_{n=1}^{N} \log p(\mathbf{x}_n|\theta)$$

- Fisher vectors

$$\mathcal{G}_\theta^{\boldsymbol{\mathcal{X}}} = I_\theta^{-1/2} \nabla_\theta \log p(\boldsymbol{\mathcal{X}}|\theta)$$

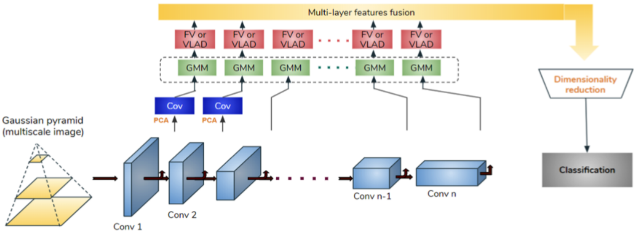
Results

Full covariance matrix model for LE distance

Coding Method	Log-Euclidean Metric	Affine Invariant Riemannian Metric
LE BoW [35]/BoRW [36]	86.4 ± 0.01	85.9 ± 0.01
LE VLAD [11]/RVLAD [11]	91.3 ± 0.01	82.8 ± 0.02
E-VLAD [37]	91.6 ± 0.01	
LE FS/RFS [40]:$\bar{\mathbf{M}}$	95.3 ± 0.01	88.9 ± 0.01
LE FS/RFS [40]:$\bar{\mathbf{M}}, \omega$	95.1 ± 0.01	90.0 ± 0.01
LE FS/RFS [40]:$\bar{\mathbf{M}}, \sigma$	95.2 ± 0.01	91.2 ± 0.01
LE FS/RFS [40]:$\bar{\mathbf{M}}, \sigma, \omega$	95.1 ± 0.01	91.2 ± 0.01
LE FV/RFV:$\bar{\mathbf{M}}$	95.5 ± 0.01	91.3 ± 0.01
LE FV/RFV:$\bar{\mathbf{M}}, \omega$	$\mathbf{95.7 \pm 0.01}$	92.6 ± 0.01
LE FV/RFV:$\bar{\mathbf{M}}, \sigma$	95.6 ± 0.01	92.7 ± 0.01
LE FV/RFV:$\bar{\mathbf{M}}, \sigma, \omega$	95.4 ± 0.01	93.2 ± 0.01

$$trace(\sigma_\mu^{-1} log^2(\overline{M}_\mu^{-1} M)) \Longrightarrow trace(\Sigma_\mu^{-1} log^2(\overline{M}_\mu^{-1} M))$$

CNN and Transfert Learning based on SPD



UC Merced Land Use

AID

Dataset	Methods	Accuracy
UCmerced (p=0.5)	CNN + FV	96,2
	CNN + RFV	96,7
AID (p=0.1)	CNN + FV	85,8
	CNN + RFV	87,9

Sara Akodad et Al : Image classification based on log-Euclidean Fisher Vectors for covariance matrix descriptors. IPTA 2018.