JOINT KERNEL COLLABORATIVE REPRESENTATION ON TENSOR MANIFOLD FOR FACE RECOGNITION

Yeong Khang Lee, Andrew Beng Jin Teoh, Kar-Ann Toh

School of Electrical and Electronic Engineering, Yonsei University, South Korea.
Sindrosa.24@gmail.com, bjteoh@yonsei.ac.kr, katoh@yonsei.ac.kr

Abstract

Gabor-based region covariance matrix (GRCM) is an emerging face feature descriptor, which has been shown promising for face recognition. The GRCM lies on Tensor manifold is inherently non-Euclidean, hence a disconnect exists between GRCM descriptor and vector-based classifiers, such as collaborative representation-based classifier (CRC). CRC is a strong alternative to sparse representation-based classifier yet enjoys high efficiency. In this paper, we bridge GRCM and CRC with kernel learning method. We investigate several geodesic distances on Tensor manifold that satisfy the Mercer's condition for kernel CRC construction as well as for speedy computation. Apart from that, we also devise two strategies to jointly combine the regionalized GRCMs with Tensor kernel CRC. Extensive experiments on the ORL and FERET datasets are conducted to verify the efficacy of the proposed method.

Index Terms— Face Recognition, Tensor manifold, kernel trick, collaborative representation classifier.

1. Introduction

Gabor-based region covariance matrix (GRCM) [1][2], was introduced as a face image descriptor followed by the popularity of regional covariance matrix (RCM) [3][4]. RCM is a powerful mean to fuse multiple correlated features or descriptors of generic objects yet invariant to illumination, scaling and rotation. Unlike the original RCM for generic object tracking, GRCM embodies Gabor filters was proved promising in face recognition [1][2].

However, GRCM does not lie on Euclidean space as a point, instead on Tensor manifold - a special type of differentiable Riemannian manifold [5]. The points on Tensor manifold are not connected in straight line thus the calculation of distance between two GRCMs, known as *geodesic distance*, has to take the geometry characteristic of the manifold into account [6]. Due to the inherent non-vector based (non-Euclidean) representation of GRCM, classification using well-developed vector-based classifiers is reckoned to be suboptimal. Prior works [4][7][8] attempted to generalize vector-based classifiers to Tensor manifold by flattening the manifold through tangent space to obtain a Euclidean approximation of the manifold. However only distances between points surrounding the tangent pole are equal to the true geodesic distances, which may lead to inaccurate modeling.

Sparse representation-based classifier (SRC) [9] proposed by Wright et al has spark a new research interest on face recognition. SRC suggests that the l_1-norm sparsity provides good discrimination ability for the classification task. However, the computation complexity of l_1-minimization is high and time consuming. Zhang et al [10] have demonstrated that collaborative representation (CR) based classification with regularized least squares required significantly less computation complexity and can even improve the face recognition performance.

Lately, Tahir et al. [11] and Biao Wang et al [12] applied the kernel trick to CRC in order to better group and separate the data with non-linear mapping. The latter capitalized KPCA and showed better performance compare to CRC and SRC as well as the kernel CRC treatment proposed by [11]. However, all of them are inherently vector-based classifiers.

In this paper, we attempt to bridge the gap between GRCM and vector-based CRC in order to harness the potential from both methods, by adopting a kernel trick to embed the Tensor manifold into the Reproducing Kernel Hilbert Space (RKHS) [13]. The embedding relies on the kernel function that satisfies the Mercer's conditions. We investigate and compare three geodesic distances on Tensor manifold, namely Affine Invariant Riemannian Metric (AIRM), Log-Euclidean Riemann Metric (LERM) and Cholesky distance to induce symmetric and positive definite kernel matrix. Due to the regionalized nature of GRCM descriptors, which was shown positive to alleviate the occlusion problem in face recognition [1], we devise two strategies to jointly combine multiple GRCMs for our Tensor kernel CRC. In a nutshell, the contribution of this paper is three-fold: (1) we link GRCM and CRC with a kernel method; (2) we compare the three geodesic distances on Tensor manifold for kernel construction and note that Cholesky distance is the best in terms of the performance and time complexity. (3) We propose two strategies to cater multiple regionalized GRCMs in Tensor KCRC for classification.

Here, we highlight that RKHS embedding for Tensor manifold data using AIRM and LERM were studied in [13][14] very recently but the exploration on bridging the GRCM and CRC as well as using Cholesky distance for kernel is absent.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

2. Preliminary

This section gives a brief account for GRCM and CRC. We refer the readers to [1], [5] and [10] for greater details regarding these techniques.

2.1 Gabor-based Region Covariance Matrix (GRCM)

Given a gray-scale face image with size h x w, thus total number of pixels, $p=hw$. By applying d number of image descriptors or extractors on the image, a three dimensional array, say $\mathbf{F}\in\mathbb{R}^{h\times w\times d}$ can be formed. $\mathbf{F}$ is then reshaped to a matrix, $\mathbf{Z} = \{\mathbf{z}_i\in\mathbb{R}^d \,|i=1,\ldots,p\}$. The covariance matrix, $\mathbf{C}$ can be computed as a $d\times d$ dimension matrix as follows:

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{u})(\mathbf{z}_i - \mathbf{u})^T \quad (1)$$

where $\mathbf{u}$ is the mean of $\mathbf{z}_i$

Gabor kernel is the product of elliptical Gaussian and a complex plane wave, defined as [15]:

$$\Psi_{u,v}(x) = \frac{||k_{u,v}||^2}{\sigma^2} e^{\frac{||k_{u,v}||^2||x||^2}{2\sigma^2}} (e^{ik_{u,v}x} - e^{-\sigma^2/2}) \quad (2)$$

such that u and v are the scale and orientation of the Gabor kernel. The wave vector $k_{u,v}$ is defined as $k_{u,v} = k_v e^{i\phi_u}$ where $k_v = k_{max}/f_v$ and $\phi_u = \pi u/8$.

In this paper we opt $v \in \{0, \ldots, 4\}$ and $u \in \{0, \ldots, 7\}$, and hence $8x5 = 40$ Gabor feature matrices can be sought by convoluting the face image and Gabor kernels as depicted in eq. (2). Together with the pixels location x and y, we obtain $\mathbf{F}\in\mathbb{R}^{h\times w\times 42}$. A Gabor-based covariance matrix, $\mathbf{G}\in\mathbb{R}^{42X42}$ is then computed according to eq. (1). In order to increase the robustness toward occlusion, each face image is segmented into 5 different parts – whole face, left, right, half up and half bottom, and thus correspond to five Gabor-based covariance matrices $\{\mathbf{G}^k\in\mathbb{R}^{42X42} | k=1,\ldots,5\}$, which collectively named as Gabor-based region covariance matrix (GRCM).

2.2 GRCM on Tensor Manifold

Tensor manifold is a special type differentiable manifold that accommodates symmetric positive definite (SPD) matrix. Formally, differentiable manifold, $\mathcal{M}$ is a topological space that is locally resembles Euclidean space and has a globally defined differential structure. $\mathcal{M}$ allows for defining tangent vectors and tangent spaces and have a metric on the tangent space that allows measurements of distances, angles, etc. Geometrically, GRCM, a type of SPD matrix can be perceived as a point on Tensor manifold, $G\in\mathcal{M}$. The geodesic distance between two points on $\mathcal{M}$ is defined as the length of the shortest curve connecting the two points.

To measure the distance between any two single GRCMs, $\mathbf{G}_i \in\mathcal{M}$ and $\mathbf{G}_j \in\mathcal{M}$, there are several geodesic distances available [5][16]. Here, we opt for three instances as follows:

1. *Affine Invariant Riemannian Metric* (AIRM)

$$p_A(\mathbf{G}_i, \mathbf{G}_j) = \sqrt{\sum_{i=1}^{d}\ln^2\lambda_i(\mathbf{G}_i, \mathbf{G}_j)} \quad (3)$$

where λ_i are the eigenvalues that can be solved through $\lambda_i\mathbf{G}_i\ \mathbf{u}_i = \mathbf{G}_j\ \mathbf{u}_i$, $i = 1,\ldots,d$.

2. *Log-Euclidean Riemann Metric* (LERM)

$$p_L(\mathbf{G}_i, \mathbf{G}_j) = ||\,\mathrm{Log}(\mathbf{G}_i) - \mathrm{Log}(\mathbf{G}_j)\,||_F \quad (4)$$

where $\mathrm{Log}(\cdot)$ is the matrix logarithm. LERM is closely related to AIRM and it is also a true geodesic distance.

3. *Cholesky distance* (CHOL)

$$p_C(\mathbf{G}_i, \mathbf{G}_j) = ||L_i - L_j||_F \quad (5)$$

CHOL is a re-parameterization measure that decomposes the matrices as $\mathbf{G}_i = L_i L_i^T$ and $\mathbf{G}_j = L_j L_j^T$. CHOL is not a true geodesic distance but it does maintain positive definite property

2.3 Collaborative Representation-based Classifier (CRC)

Suppose a training dataset $\mathbf{X}$ with c number of classes having m samples in each class. Let $\mathbf{X} = [\mathbf{X}_1,\ldots,\mathbf{X}_c]$ and $\mathbf{X}_i = [\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,m_i}] \in\mathbb{R}^{pxm}$ where $i=1,\ldots,c$, p is the dimension of image vector and $N = \sum_{i=1}^{c} m_i$ is the total number of samples in the data set. A test sample $\mathbf{y}\in\mathbb{R}^p$ can be coded by the following l_2-minimization problem.

$$\hat{\boldsymbol{a}} = \arg min_a \|\mathbf{y} - \mathbf{X}\boldsymbol{a}\|_2^2 + \lambda\|\boldsymbol{a}\|_2^2 \quad (6)$$

The recognition is made in according to the class with minimum reconstruction error:

$$identity(\mathbf{y}) = \arg min_i \{\|\mathbf{y} - \mathbf{X}_i\delta_i(\hat{\boldsymbol{a}})\|/\|\delta_i(\hat{\boldsymbol{a}})\|_2\} \quad (7)$$

i is the label associate with i^{th} class and $\delta_i(\hat{\boldsymbol{a}})$ is the coefficient vector of $\mathbf{X}$ that related to class i.

3. Kernel Collaborative representation-based Classifier on Tensor Manifold

Kernel leaning methods have been widely studied and proved versatile in machine learning and computer vision to explore non-linear structure of data. The key idea of kernel methods is to map the input data to a high dimensional feature space to harness a richer representation of the data distribution.

This idea can be generalized to differentiable manifolds such as Tensor manifold [13], Grassmann manifold [17] etc as follows: Each point $x\in\mathcal{M}$ on manifold is mapped to a feature vector $\phi(x)$ in a Hilbert space. A kernel function k : $(\mathcal{M}x\mathcal{M}) \rightarrow \mathbb{R}$ is used to define the inner product on Hilbert space, thus forming a Reproducing Kernel Hilbert Space (RKHS). According to Mercer's theorem, however, only positive definite and symmetric kernels delineate valid RKHS.

For GRCM on Tensor manifold, a positive definite and symmetric kernel function can be induced through [13]:

$$k(\mathbf{G}_i,\mathbf{G}_j) = exp(-\gamma^{-1}p(\mathbf{G}_i,\mathbf{G}_j)^2) \quad (8)$$

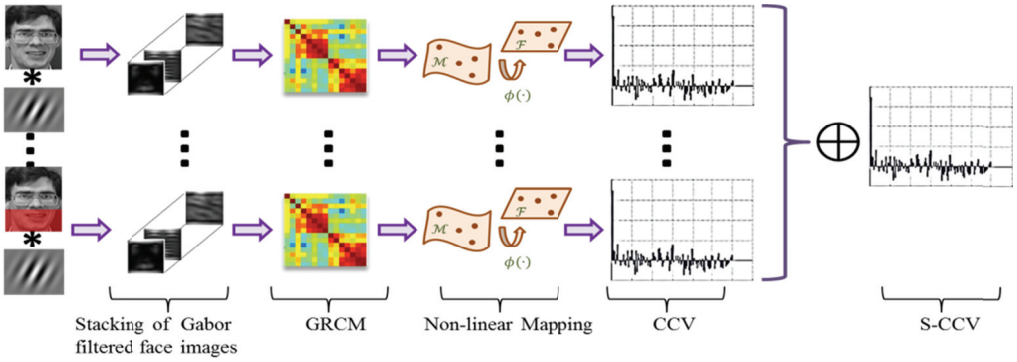

Fig. 1: Joint Tensor kernel Collaborative Representation Classification based on strategy 2

where $\mathbf{G}_i, \mathbf{G}_j \in \mathcal{M}$, $\gamma > 0$ and $p(,)$ is a distance measure defined in eq. (3) (4) or (5). In [13], the LERM and CHOL were proved to generate positive definite Gaussian Kernels while AIRM is regarded as a pseudo-kernel due to its defiance to satisfy the Mercer's condition [14]. However, the experimental result shows AIRM can be nevertheless useful.

By and large, a major advantage of being able to compute positive definite kernels on the Tensor Manifold is that it permits to utilize the vector-based classifiers, while still accounting for the geometry of manifolds.

We now show the use of CRC for GRCM classification on Tensor manifold via kernel method. By applying the mapping function into GRCM for CRC with eq. (6), we have:

$$\hat{\boldsymbol{a}} = \arg min_a \|\phi(\mathbf{G}_t) - \phi(\mathcal{G})\boldsymbol{a}\|_2^2 + \lambda\|\boldsymbol{a}\|_2^2 \quad (9)$$

where $\mathcal{G}=[\mathbf{G'}_1, \ldots, \mathbf{G'}_c] \in \mathcal{M}$ and $\mathbf{G'}_i = [\mathbf{G}_{i,1}, \ldots, \mathbf{G}_{i,m_i}] \in \mathcal{M}$ are the training set with c classes and having m samples in each class and $N = \sum_{i=1}^{c} m_i$ total training samples while $\mathbf{G}_t \in \mathcal{M}$ is a testing sample.

Unfortunately, eq. (9) cannot be solved directly due to the unknown mapping function $\phi(\cdot)$. As suggested by [12], the optimization problem in eq. (9) can be solved by resorting to kernel based dimensionality reduction methods, such as kernel principle component analysis (KPCA) [18]. $\phi(\mathbf{G}_t) = \phi(\mathcal{G})a$ is the l_2-norm regularized solution to Eq. (9). Since kernel feature space $\mathcal{F} \in \mathbb{R}^q$ tends to be high and even infinity in dimension, we take a projection matrix, $\mathbf{P}=[\mathbf{p}_1, \ldots, \mathbf{p}_r] \in \mathbb{R}^{q \times r}$ where $q>>r$ and project the sample points in $\mathcal{F}$ into a lower dimension subspace with dimension r.

In KPCA, projection vector $\mathbf{p}_i$ is the linear combination of samples in $\mathcal{F}$ such that

$$\mathbf{p}_i = \phi(\mathcal{G})\boldsymbol{\beta}_i \quad (10)$$

where each $\boldsymbol{\beta}_i \in \mathbb{R}^N, i = 1, \ldots, r$ $(r \leq N)$ is a normalized eigenvector corresponds to the first r largest non-zero eigenvalues. Denote $\mathbf{B} = [\boldsymbol{\beta}_1 \ldots \boldsymbol{\beta}_r] \in \mathbb{R}^{r \times N}$, then eq. (10) can be expressed as $\mathbf{P} = \phi(\mathcal{G})\mathbf{B}$. With similar algebraic manipulation as shown in [12], we have:

$$\mathbf{B}^T k(\cdot, \mathbf{G}_t) = \mathbf{B}^T \mathbf{K} a \quad (11)$$

where $\mathbf{K}$ is a kernel gram matrix with the elements $k_{ij} = k(\mathbf{G}_i, \mathbf{G}_j)$ as defined in eq. (8) and $k(\cdot, \mathbf{G}_t) = [k(\mathbf{G}_1, \mathbf{G}_t), \ldots, k(\mathbf{G}_N, \mathbf{G}_t)]$. Note that eq. (11) is a typical generalized eigenvalues problem and can be solved easily. Hence, Eq. (9) can be rewritten as:

$$\hat{\boldsymbol{a}} = \arg min_a \|\mathbf{B}^T k(\cdot, \mathbf{G}_t) - \mathbf{B}^T \mathbf{K}\boldsymbol{a}\|_2^2 + \lambda\|\boldsymbol{a}\|_2^2 \quad (12)$$

and can be solved analytically as follows [10]:

$$\hat{\boldsymbol{a}} = ((\mathbf{B}^T\mathbf{K})^T(\mathbf{B}^T\mathbf{K}) + \lambda\mathbf{I})^{-1} \cdot (\mathbf{B}^T\mathbf{K})^T(\mathbf{B}^T k(\cdot, \mathbf{G}_t)) \quad (13)$$

The residue of test image is then computed as:

$$r_i = \|\mathbf{B}^T k(\cdot, \mathbf{G}_t) - (\mathbf{B}^T\mathbf{K})\delta_i(\hat{\boldsymbol{a}})\|_2 / \|\delta_i(\hat{\boldsymbol{a}})\|_2 \quad (14)$$

Identity of $\mathbf{G}_t$ is determined using $identity(\mathbf{G}_t) = \arg min_i r_i(\mathbf{G}_t)$.

4. MULTIPLE REGIONALIZED GRCM

Since five GRCMs, $\{\mathbf{G}^k \in \mathcal{M} \mid k=1,\ldots,5\}$ are used to represent a face image, we propose two strategies to joinly combine five GRCMs for Tensor KCRC. First strategy is by fusing the distance measures that described in eq. (3), (4) or (5) in equal weights manner as follows [1]:

$$D(\mathbf{G}_i^k, \mathbf{G}_j^k) = \sum_{k=1}^{5} p(\mathbf{G}_i^k, \mathbf{G}_j^k) - \max_k(p(\mathbf{G}_i^k, \mathbf{G}_j^k)) \quad (15)$$

eq.(8) is then used to induce kernel value for subsequent computation. Second strategy, which was inspired by [19], combines five coefficient vectors into a single Super Collaborative Coefficient Vector (S-CCV), as shown in Figure 1. S-CCV is computed as $\hat{\boldsymbol{s}} = \sum_{k=1}^{5} \hat{\boldsymbol{a}}_k$, where k is the index of GRCM. Let the $\hat{\boldsymbol{s}}$ of $\mathbf{G}_t$ denoted as $\hat{\boldsymbol{s}}_t(i,j)$ where $i \in [1, c]$, c is the number of classes and $j \in [1, m_j]$, m_j is the number of images per class. A function is defined as:

$$\omega_i(\mathbf{G}_t) = \sum_{j=1}^{m_j} \hat{\boldsymbol{s}}_t(i,j) / m_j \quad (16)$$

which returns the normalized summation of l_2 coefficients within the same class. Then the identity is assigned by $identity(\mathbf{G}_t) = \arg max_i \omega_i(\mathbf{G}_t)$.

5. EXPERIMENTS AND DISCUSSION

The experiments are conducted on two face databases: AT&T[20] and FERET[21]. We use the full set of data in AT&T, ie. 40 subjects and 10 samples of each subject, and randomly select 292 subjects with 11 samples per subject for FERET. Regularized parameter λ (eq.(13)) is set as $1e^{-4}$ and $r=N$, number of training images (eq.(10)). The

parameters of Gabor kernel are set as $k_{max} = \pi/2$, $f_v = \sqrt{2}$ and $\sigma = 2\pi$ [15]. AT&T images are convoluted with size 21×21 Gabor kernel while 25×25 size of Gabor kernel is used for FERET. We use $\gamma = 100$ and $\gamma = 10$ for Strategy 1 and Strategy 2, respectively (eq. (8)). Both databases are partitioned into 2 distinct sets: training and testing set. Training image is randomly selected from size of 1 to 5 and multiple-fold cross-validation is performed and the average of correct recognition rate (CRR) is taken. Total computation time is also recorded.

The proposed methods are compared with vector-based CRC [10] and KCRC [12]. We follow the exact account given in [10] and [12], respectively. However, the best parameters are chosen based on the AT&T and FERET databases adopted in this paper. We also include a kernel that based on Frobenius distance (FROB), $p_F(\mathbf{G}_i, \mathbf{G}_j) = ||\mathbf{G}_i - \mathbf{G}_j||_F$ for comparison. Note that this kernel is not catered for Tensor manifold, instead assume that the GRCM resides on Euclidean space.

	Number of training samples				
AT&T	**1**	**2**	**3**	**4**	**5**
CRC[10]	64.66	78.75	87.31	91.76	94.01
KCRC[12].	74.25	86.19	91.12	93.87	95.44
Strategy 1					
FROB	69.13	83.25	89.72	93.15	95.11
CHOL	83.27	92.52	96.24	**98.02**	98.09
AIRM	84.57	93.13	96.36	97.90	98.52
LERM	84.75	93.12	96.21	97.72	98.48
Strategy 2					
FROB	69.22	82.25	88.94	92.67	95.00
CHOL	**85.19**	93.09	96.35	97.93	98.40
AIRM	85.61	**93.26**	**96.38**	97.77	**98.64**
LERM	84.72	92.93	96.00	97.60	98.40
FERET	**1**	**2**	**3**	**4**	**5**
CRC[10]	39.37	61.99	74.74	82.15	87.17
KCRC[12].	45.00	63.34	73.43	79.99	84.55
Strategy 1					
FROB	65.23	79.23	96.06	89.94	92.43
CHOL	76.46	88.18	92.88	95.24	96.62
AIRM	77.26	88.18	92.69	95.00	96.39
LERM	79.01	90.03	93.86	**96.43**	97.22
Strategy 2					
FROB	63.97	78.65	85.58	89.55	92.21
CHOL	**79.15**	90.28	**94.30**	96.26	**97.39**
AIRM	78.49	89.36	93.35	95.55	96.80
LERM	78.40	**90.40**	93.80	96.30	97.00

Table 1: CRR (%) for AT&T and FERRET

From Table 1, we observe that Tensor KCRC outperform consistently vector-based CRC and KCRC. This is especially glaring when small training size is used. The increment of the training samples improve the overall CRR as anticipated. This could be attributed to the use of GRCM descriptor and its dedicated classifier using Tensor KCRC. It is interesting to notice that the performance of FROB is inferior compare with that of geodesic distances explicit for Tensor manifold. This attests GRCM does not reside on Euclidean space.

We also observe that different geodesic distances on Tensor manifold do not bring significant impact to Tensor KCRC. LERM performs slightly better than that of CHOL and AIRM with less than 1% difference. As noted in section 2.2 and section 3 in which CHOL is not a true geodesic distance [13] while AIRM is a pseudo-kernel [14], they do perform well in Tensor KCRC nevertheless. In addition, despite [13] favors LERM for Tensor manifold kernel learning, it is impractical due to prohibitively high computational cost spent on logarithm matrix calculation, as shown in Fig. 2. Among the three geodesic distances considered in this paper, CHOL stands out to be the best candidate for Tensor kernel CRC in terms of accuracy performance and computation time, thanks to its efficient Cholesky decomposition algorithm.

Strategy 2, however has better CRR than strategy 1 in general with a few exception cases. Strategy 1 was devised based on the intuitive to discard the distance in eq. (15) that corresponded to the most unreliable region due to occlusion, illumination variation etc. This may affect classification ability of KCRC as CRC works jointly across face regions to identify an individual. On the other hand, Strategy 2 allows KCRC to collaboratively represent each face region, and thus face region that corrupted simply reduce its contribution to S-CCV.

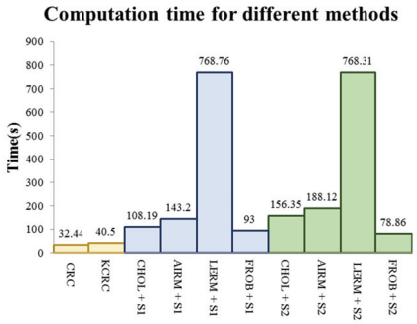

Fig. 2: Computation time(s) on AT&T database with 3 training image

6. CONCLUSION

In this paper, we bridge the gap between GRCM descriptor and the vector-based CRC using Tensor kernel method for face recognition. We also proposed two strategies to handle different regions of faces and combining coefficient vectors of Tensor Kernel CRC. We showed that the proposed method outperformed the conventional vector-based CRC and KCRC significantly on two benchmark datasets. For future works, we shall investigate into more geodesic distances on Tensor manifold towards better discriminant applications.

7. ACKNOWLEDGEMENT

This research was supported by the MSIP(The Ministry of Science, ICT and Future Planning), Korea and Microsoft Research, under IT/SW Creative research program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2013-(H0503-13-1001))

References

[1] Y. Pang, Y. Yuan, and X. Li, "Gabor-Based Region Covariance Matrices for Face Recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 7, pp. 989–993, 2008.

[2] H. Qin, L. Qin, L. Xue, and Y. Li, "A Kernel Gabor-Based Weighted Region Covariance Matrix for Face Recognition," Sensors, vol. 12, no. 12, pp. 7410–7422, May 2012.

[3] O. Tuzel, F. Porikli, and P. Meer, "Region Covariance: A Fast Descriptor for Detection and Classification," in Computer Vision – ECCV 2006, vol. 3952, 2006, pp. 589–600.

[4] D. Tosato, M. Spera, M. Cristani, and V. Murino, "Characterizing Humans on Riemannian Manifolds," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, no. 1, p. 1, 2013.

[5] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian Framework for Tensor Computing," Int J Computer Vision, vol. 66, no. 1, pp. 41–66, Jan. 2006.

[6] Cherian, A.; Sra, S.; Banerjee, A.; Papanikolopoulos, N., "Jensen-Bregman LogDet Divergence with Application to Efficient Similarity Search for Covariance Matrices," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 9, pp. 2161-2174, 2013

[7] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian Detection via Classification on Riemannian Manifolds," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 10, pp. 1713 –1727, Oct. 2008.

[8] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Common Spatial Pattern revisited by Riemannian geometry," in 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP), 2010, pp. 472–476.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust Face Recognition via Sparse Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[10] L. Zhang, M. Yang, and Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," CVPR, 2011, pp. 471–478.

[11] M. A. Tahir and A. Bouridane, "Face recognition using kernel collaborative representation and multiscale local binary patterns," in IET Conference on Image Processing (IPR 2012), 2012, pp. 1–4.

[12] Biao Wang, Weifeng Li, Norman Poh, and Qingmin Liao, "Kernel collaborative representation-based classifier for face recognition," ICASSP2013, 2013.

[13] Sadeep Jayasumana, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi, "Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices," CVPR, 2013.

[14] M. T. Harandi, C. Sanderson, A. Wiliem, and B. C. Lovell, "Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures," in Applications of Computer Vision, IEEE Workshop on, Los Alamitos, CA, USA, 2012, vol. 0, pp. 433–439.

[15] T. S. Lee, "Image representation using 2-D Gabor wavelets," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 10, pp. 959–971, Oct. 1996.

[16] X. Pennec, "Statistical Computing on Manifolds: From Riemannian Geometry to Computational Anatomy," in LNCS, vol. 5416, F. Nielsen, Ed. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 347–386.

[17] Tee Connie, Goh Kah Ong Michael & Andrew Teoh Beng Jin, 2013, "Gait Recognition using Sparse Grassmannian Locality Preserving Discriminant Analysis", ICASSP 2013,

[18] B. Scholkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in ADVANCES IN KERNEL METHODS - SUPPORT VECTOR LEARNING, 1999, pp. 327–352.

[19] R. Min and J.-L. Dugelay, "Improved combination of LBP and sparse representation based classification (SRC) for face recognition," in 2011 IEEE International Conference on Multimedia and Expo (ICME), 2011, pp. 1–6.

[20] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in Proc. Second IEEE Workshop Applications of Computer Vision, 1994.

[21] P. J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The ferret evaluation method for face recognition algorithms," IEEE Trans. on Patt. Analysis and Machine Intell., vol. 22, no. 10, pp. 1090–1104, 2000.