LEARNING THE DISCRIMINATIVE DICTIONARY FOR SPARSE REPRESENTATION BY A GENERAL FISHER REGULARIZED MODEL

Qingfeng Liu, Ajit Puthenputhussery, Chengjun Liu

Department of Computer Science
New Jersey Institute of Technology

ABSTRACT

This paper presents two novel discriminative dictionary learning models for sparse representation, namely the Fisher discriminative sparse model (FDSM) and the marginal Fisher discriminative sparse model (MFDSM). To learn the FDSM and the MFDSM efficiently and homogeneously, a general Fisher regularized model is further derived so that both of them can be learned without much modification. Experimental results on four popular databases, namely the extended Yale face database B, the AR face database, the 15 scenes dataset and the MIT-67 indoor scenes dataset show that the proposed method can improve upon other popular methods.

1. INTRODUCTION

The sparse representation methods [1] based on dictionary learning (sometimes called sparse coding, sparse modelling, etc.) have been applied to many areas, for example, face recognition. However, sparse representation method is not directly related to classification, which means the discriminative information of the data is not utilized. Meanwhile Fisher linear discriminant analysis [2] and its variants [3] are widely used for learning discriminative feature representations. Thus, the sparse representation is combined with the Fisher linear discriminant analysis to achieve the discriminative ability [4]. This method, however, assumes a fixed dictionary and uses the ratio of the within-class scatter matrix and the between-class scatter matrix for regularization, which introduces difficulty in derivation.

To avoid the disadvantages, this paper presents two novel discriminative sparse models and provides a simple and an efficient solution for the above models by learning a general Fisher regularized model. In particular, two novel discriminative dictionary learning models, namely the Fisher discriminative sparse model (FDSM) and the marginal Fisher discriminative sparse model (MFDSM) are proposed. The FDSM method constrains the sparse representation by using the traditional Fisher criterion [2] so that the within-class scatter matrix of the sparse representations of the training samples is minimized while the between-class scatter matrix is maximized. The MFDSM method regularizes the sparse representation by using the marginal Fisher criterion [3] which considers the marginal information so as to minimize the intra-class compactness and maximize the inter-class separability.

Generally, it is intractable or troublesome to derive the two proposed models directly due to the intricate structure of two Fisher terms for derivation. Thus, a general Fisher regularized sparse model is derived from the two proposed models for learning the discriminative dictionary efficiently and homogeneously by iteratively applying the Fisher regularized FISTA [5] method and the Lagrange dual method [6]. To guarantee that the proposed methods converge to an optimal solution quickly, the initialization, step size and the convergence issues are also discussed.

The performance of the two proposed models are assessed on four representative databases, namely the extended Yale face database B [7], the AR face database [8], the 15 scenes dataset [9] and the MIT-67 indoor scenes dataset [10]. The experimental results show the feasibility of the proposed methods.

The proposed FDSM and MFDSM have several advantages over the previous methods (see section 2). First, the two proposed methods can learn the discriminative dictionary by using two discriminative Fisher criteria to constrain the sparse representation. Thus, they can take advantage of the non-linear classifiers since their discriminative ability are not restricted to some specific linear classifiers. Second, the two proposed methods are capable of learning a global shared dictionary efficiently to avoid the high computational cost of learning the sub-dictionaries and the inferior performance caused by small training sample size in each class. Above all, a general Fisher regularized model can be derived from the two proposed models so that both of them can be derived homogeneously by the same algorithm without much modification.

2. RELATED WORK

Most of the previous methods of sparse representation focus on either developing efficient learning algorithms [5], [6] or exploring the data manifold structures for representation [11], [12], [13]. These methods, however, disregard the discriminative label information of the training data and thus are sub-

978-1-4799-8339-1/15/$31.00 ©2015 IEEE

optimal for classification. Note that our proposed methods result in a general model that has a similar form as the methods proposed in [11], [13]. However, their methods are unsupervised and only aim at exploring the data manifold structures for representation.

Recently, two categories of discriminative dictionary learning methods were proposed for sparse representation. The first category co-trains the discriminative dictionary, sparse representation and the linear classifier together. While, the second category combines the sub-dictionaries by exploiting their discriminative power. Zhang et al. [14] proposed a method to co-train the discriminative dictionary, sparse representation as well as the linear classifier using a combined objective function that is optimized by the discriminative KSVD (D-KSVD). Jiang et al. [15] improved upon the method introduced in [14] by introducing a label consistent regularization term. Their methods are closely tied to the linear classifiers, which discourages the use of non-linear classifiers that often obtain better results. Zhou et al. [16] presented a Joint Dictionary Learning (JDL) method that jointly learns both a commonly shared dictionary and class-specific sub-dictionaries to enhance the discrimination of the dictionaries. Yang et al. [17] proposed a Fisher Discrimination Dictionary Learning (FDDL) method, which learns a structured dictionary that consists of a set of class-specific sub-dictionaries. Their methods are time-consuming when the number of classes is large and may harm the performance when the number of the training images for each class is small.

3. TWO DISCRIMINATIVE SPARSE MODELS

Given the sample data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m] \in \mathbb{R}^{n \times m}$ that consists of m samples each with dimension n, the dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_k] \in \mathbb{R}^{n \times k}$ that represents k basis vectors and the sparse representation matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m] \in \mathbb{R}^{k \times m}$ which denotes the sparse representations for the m samples, the Fisher discriminative sparse model (FDSM) applies the traditional Fisher criterion [2] on the sparse representations to guarantee the discriminative ability. Mathematically, the FDSM method attempts to learn the dictionary $\mathbf{D}$ and the sparse representations $\mathbf{A}$ of training samples so that the within-class scatter matrix $\mathbf{S}_w$ of all the $\mathbf{a}_i (i = 1, 2, ..., m)$ is minimized while the between-class scatter matrix $\mathbf{S}_b$ is maximized. As a result, the following optimization problem has to be solved.

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{A}} \sum_{i=1}^{m} ||\mathbf{x}_i - \mathbf{D}\mathbf{a}_i||^2 + \lambda ||\mathbf{a}_i||_1 + \alpha \mathbf{tr}(\beta \mathbf{S}_w - (1-\beta)\mathbf{S}_b) \\ &s.t. \quad ||\mathbf{d}_j|| \leq 1, (j = 1, 2, ..., k) \end{aligned} \tag{1}$$

where the parameter λ controls the sparsity term, the parameter α controls the discriminative Fisher term ($\mathbf{tr}(\cdot)$ is the trace of the matrix), and the parameter β balances the contributions of the within-class scatter matrix $\mathbf{S}_w$ and the between-class scatter matrix $\mathbf{S}_b$.

The marginal Fisher discriminative sparse model (MFDSM) is proposed to extend the traditional Fisher criterion by considering the marginal information. The marginal Fisher criterion [3] seeks to minimize the intra-class compactness $\mathbf{S}_c = \sum_{i=1}^{m} \sum_{(i,j) \in N_k^c(i,j)} ||\mathbf{a}_i - \mathbf{a}_j||^2$ and maximize the inter-class separability $\mathbf{S}_p = \sum_{i=1}^{m} \sum_{(i,j) \in N_k^p(i,j)} ||\mathbf{a}_i - \mathbf{a}_j||^2$, where $(i, j) \in N_k^c(i, j)$ means all the (i, j) pairs where either the sample $\mathbf{x}_i$ is among the k nearest neighbours of sample $\mathbf{x}_j$ in the same class or the sample $\mathbf{x}_j$ is among the k nearest neighbours of sample $\mathbf{x}_i$ in the same class. And $(i, j) \in N_k^p(i, j)$ means that the k nearest (i, j) pairs among all the (i, j) pairs between sample $\mathbf{x}_i$ and sample $\mathbf{x}_j$ are from different classes. Then the MFDSM method becomes the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{A}} \sum_{i=1}^{m} ||\mathbf{x}_i - \mathbf{D}\mathbf{a}_i||^2 + \lambda ||\mathbf{a}_i||_1 + \alpha \mathbf{tr}(\beta \mathbf{S}_c - (1-\beta)\mathbf{S}_p) \\ &s.t. \quad ||\mathbf{d}_j|| \leq 1, (j = 1, 2, ..., k) \end{aligned} \tag{2}$$

where the parameters λ, α and β act the same as the ones defined in equation 1. The difference between the FDSM method and the MFDSM method is the Fisher criterion they used. The marginal Fisher criterion is more general than the traditional Fisher criterion [3] as the inter-class margin can better characterize the separability. In practice, the MFDSM method performs similar to the FDSM method for most of the cases.

4. A GENERAL FISHER REGULARIZED SPARSE MODEL

To learn the FDSM and the MFDSM efficiently and homogeneously, a general model of learning the discriminative dictionary can be derived from the two proposed models as shown in Theorem 4.1. Some notations are as follows: $\mathbf{A}$ is the sparse representation matrix defined above; $\mathbf{e} = [1, 1, ..., 1]^t \in \mathbb{R}^{m \times 1}$ and $\mathbf{e}_i = [0, 0, ..., 1, 1, ..., 1, 0, ..., 0]^t \in \mathbb{R}^{m \times 1}$ where only the indices of the training samples from the i-th $(i = 1, 2, ..., c)$ class are 1, otherwise 0; $\mathbf{D}_c$ and $\mathbf{D}_p$ are diagonal matrices whose diagonal values $\mathbf{D}_c(i, i) = \sum_{j \neq i} \mathbf{W}_c(i, j)$ and $\mathbf{D}_p(i, i) = \sum_{j \neq i} \mathbf{W}_p(i, j)$, $\mathbf{W}_c(i, j) = 1$ if sample $\mathbf{x}_i$ is among the k nearest neighbours of sample $\mathbf{x}_j$ in the same class or sample $\mathbf{x}_j$ is among the k nearest neighbours of sample $\mathbf{x}_i$ in the same class, otherwise 0, and similarly $\mathbf{W}_p(i, j) = 1$ if the pair (i, j) is among the k nearest pairs from all the pairs between samples of different classes, otherwise 0.

Theorem 4.1 *Both the FDSM method and the MFDSM method belong to a general model of learning the discrimi-*

native dictionary, which has the following form:

$$\min_{\boldsymbol{D},\boldsymbol{A}} \sum_{i=1}^{m} ||\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{a}_i||^2 + \lambda||\boldsymbol{a}_i||_1 + \alpha \boldsymbol{tr}(\boldsymbol{A}\boldsymbol{L}\boldsymbol{A}^t)$$
$$s.t. \quad ||\boldsymbol{d}_j|| \leq 1, (j = 1, 2, ..., k) \tag{3}$$

where $\boldsymbol{L} = \boldsymbol{I} - \sum_{i=1}^{c} \frac{1}{m_i} \boldsymbol{e}_i \boldsymbol{e}_i^t - (1-\beta)(\boldsymbol{I} - \frac{1}{m}\boldsymbol{e}\boldsymbol{e}^t)$ *for the FDSM method and* $\boldsymbol{L} = 2\beta(\boldsymbol{D}_c - \boldsymbol{W}_c) - 2(1-\beta)(\boldsymbol{D}_p - \boldsymbol{W}_p)$ *for the MFDSM method.*

Theorem 4.1 states a favorable property that the two proposed models can be derived by optimizing the equation 3 with the only difference of the matrix **L**.

The algorithm of optimizing the equation 3 iteratively learns the sparse representations and the discriminative dictionary. The Fisher regularized FISTA (FRFISTA) algorithm [5], is proposed to learn the Fisher regularized sparse representation. And the Lagrange dual method is used to learn the dictionary. In practice, the dictionary and the sparse representation are initialized by using the traditional sparse representation method. Then the algorithm is able to converge to the optimal solution very quickly.

When the dictionary **D** is given, the Fisher regularized FISTA (FRFISTA) algorithm attempts to derive each $\mathbf{a}_i (i = 1, 2, ..., m)$ separately by optimizing $\min_{\mathbf{a}_i} ||\mathbf{x}_i - \mathbf{D}\mathbf{a}_i||^2 + \alpha \mathbf{L}_{ii} \mathbf{a}_i^t \mathbf{a}_i + \alpha \mathbf{a}_i^t (\sum_{j \neq i} \mathbf{L}_{ij} \mathbf{a}_j) + \lambda ||\mathbf{a}_i||_1$, where $\mathbf{L}_{ij} (i, j = 1, 2, ..., m)$ is the value in the i-th row and the j-th column of the matrix **L**. Following the structure of the FISTA algorithm, the objective function can be decomposed into $f(\mathbf{a}_i) + g(\mathbf{a}_i)$, where $f(\mathbf{a}_i) = ||\mathbf{x}_i - \mathbf{D}\mathbf{a}_i||^2 + \alpha \mathbf{L}_{ii} \mathbf{a}_i^t \mathbf{a}_i + \alpha \mathbf{a}_i^t (\sum_{j \neq i} \mathbf{L}_{ij} \mathbf{a}_j)$ and $g(\mathbf{a}_i) = \lambda ||\mathbf{a}_i||_1$. Then the FISTA algorithm [5] optimizes the equation by using an adaptive step size. For each function $f(\mathbf{a}_i)$, the smallest Lipschitz constant of the gradient ∇f is $L(f) = 2\lambda_{max}(\mathbf{D}^t\mathbf{D} + \alpha \mathbf{L}_{ii} \mathbf{I})$, which means twice of the largest eigenvalue of the matrix $(\mathbf{D}^t\mathbf{D} + \alpha \mathbf{L}_{ii}\mathbf{I})$. Then the largest step size that guarantees convergence of the FRFISTA algorithm is $\frac{1}{L(f)}$. Note that the step size is adaptive for each iteration for different data.

When **A** is given, the dictionary **D** is updated using the Lagrange dual method [12]. The dual problem is formulated as $\Lambda^* = \min_\Lambda \mathbf{tr}(\mathbf{X}\mathbf{A}^t(\mathbf{A}\mathbf{A}^t + \Lambda)^{-1}\mathbf{A}\mathbf{X}^t + \Lambda - \mathbf{X}^t\mathbf{X})$, where Λ is a diagonal matrix whose diagonal values are the dual parameters of the primal optimization problem. The dual problem can be solved using the gradient descent method. Then the dictionary D is updated using $\mathbf{D} = \mathbf{X}\mathbf{A}^t(\mathbf{A}\mathbf{A}^t + \Lambda^*)^{-1}$.

5. EXPERIMENTS

To evaluate the performance of the two proposed methods, empirical studies are conducted on four representative databases, namely the extended Yale face database B [7], the AR face database [8], the 15 scenes dataset [9] and the MIT-67 indoor scenes dataset [10]. We use the principal component analysis and the improved marginal Fisher analysis [18] to reduce the dimension of the data and extract features. The dictionary size is 1024 for MIT-67 indoor scenes dataset and 512 for the others.

Methods	Accuracy %
D-KSVD [14]	75.30
SRC [1]	90.00
FDDL [19]	91.90
The FDSM method	**95.19**
The MFDSM method	**95.44**

Table 1. Comparisons between the two proposed methods and the other popular methods on the extended Yale face database B.

5.1. Extended Yale face database B

The extended Yale face database B consists of 2414 frontal view face images from 38 individuals each with around 64 images taken under various lightening conditions. A cropped version [7] of the database is used. As per the experimental settings defined in [19], for 10 iterations, 20 images are randomly selected for training, and the remaining images for testing each subject. The image is first scaled to 42×48 and represented as the concatenation of the column pixels. Then the dimension is reduced to 300. For the FDSM method, the parameters are selected as $\lambda = 0.1$, $\alpha = 0.5$, and $\beta = 0.5$. Then the RBF kernel based SVM, which is parameterized as $C = 4$ and $\gamma = 0.0003$, is employed for classification. Similarly, as for the MFDSM method, the model parameters are selected as $\lambda = 0.1$, $\alpha = 0.1$, and $\beta = 0.6$. And the RBF kernel based SVM is parameterized as $C = 6$ and $\gamma = 0.0002$. The final results are shown in table 1. The average accuracy of 10 iterations of the FDSM method and the MFDSM method are reported in table 1.

5.2. AR face database

The AR face database is composed of over 4000 frontal view images for 126 individuals each with 26 pictures taken in two separate sessions. A subset of the data [8] is chosen. The images are cropped to dimension 165*120. According to the experimental settings [1], [19] and [17], 14 images with only illumination change and expressions are selected for each person: the seven images from session 1 for training and the other seven from session 2 for testing. Each image is represented as the concatenation of the column pixels and the dimension is reduced to 320 for FDSM. The parameters $\lambda = 0.1$, $\alpha = 0.5$, and $\beta = 0.5$ are selected for FDSM. Then the parameters $C = 4$ and $\gamma = 0.0007$ are selected for the RBF kernel based SVM. As for the MFDSM method, the dimension is reduced to 300 and the parameters $\lambda = 0.1$, $\alpha = 0.2$, and $\beta = 0.6$ are selected. The RBF kernel based

Experimental setting 2	Accuracy %
D-KSVD [14]	85.40
LC-KSVD [15]	89.7
JDL [16]	91.7
FDDL [17]	92.00
The FDSM method	**94.71**
The MFDSM method	**95.00**

Table 2. Comparisons between the proposed methods and the other popular methods on AR face database.

Methods	Accuracy %
KSPM [9]	81.40 ± 0.50
LLC [12]	80.57
D-KSVD [14]	89.10
LC-KSVD [15]	90.40
LaplacianSC [11]	89.70
DHVFC [20]	86.40
The FDSM method	**97.90** ± 0.20
The MFDSM method	**97.94** ± 0.22

Table 3. Comparisons between the proposed methods and the other popular methods on the 15 scenes dataset

SVM selects the parameters $C = 4$ and $\gamma = 0.0004$ for classification. The results that are presented in table 2 show that the proposed methods are able to improve upon the other popular methods.

5.3. The 15 Scenes Dataset

The 15 scenes dataset [9] contains 4485 images from 15 scene categories, each with the number of images ranging from 200 to 400. Following the experimental protocol defined in [9], 100 images per class are randomly selected for training and the remaining for testing for 10 iterations. First, the spatial pyramid features provided by [15] are used to represent the image as a vector with the dimension of 3000 for fair comparison. Then the image vector is reduced to dimension 500 for both FDSM and MFDSM. Parameters $\lambda = 0.05$, $\alpha = 0.1$, and $\beta = 0.1$ are selected for the FDSM method. With respect to the MFDSM method, $\lambda = 0.05$, $\alpha = 0.2$, and $\beta = 0.4$ are selected. The parameters $C = 4$ and $\gamma = 0.0005$ are selected for the RBF kernel based SVM for both FDSM and MFDSM. The results are shown in table 3, it is seen that the two proposed methods are able to achieve significantly better results.

5.4. The MIT-67 Indoor Scenes Dataset

The MIT-67 indoor scenes dataset [10] is a very challenging indoor scene recognition dataset, which contains 67 indoor categories with 15620 images. Commonly used experimental setting [10] are followed, wherein 80*67 images are used for training and 20*67 images for testing. The performance is measured as the average classification accuracy over all the categories. Fisher vector features [21] are considered for representing the image. The SIFT feature is first projected to dimension 80 and a codebook with 256 visual words is learned, then the dimension of the Fisher vector is 2*256*80 = 40960, which is further reduced to 2000 for both FDSM and MFDSM. Parameters $\lambda = 0.05$, $\alpha = 0.1$, and $\beta = 0.1$ are selected for the FDSM method. With respect to the MFDSM method, $\lambda = 0.05$, $\alpha = 0.2$, and $\beta = 0.4$ are selected. The parameters $C = 4$ and $\gamma = 0.0001$ are selected for the RBF kernel based SVM for both FDSM and MFDSM. The results are shown in table 4, it is seen that the two proposed methods are able to achieve comparable results to the state-of-the-art methods [26]. Note that the Fisher vector is directly learned from the SIFT features for each image instead of learning the part detectors used in [26], which is time-consuming. Moreover, we reduce the dimension of Fisher vector from 40960 to 2000, which saves much storage space.

Methods	Mean Accuracy %
ROI + Gist [10]	26.1
DPM [22]	30.4
Object Bank [23]	37.6
miSVM [24]	46.4
D-Parts [25]	51.4
DP + IFV [26]	60.8
The FDSM method	58.13
The MFDSM method	57.24

Table 4. Comparisons between the proposed methods and the other popular methods on the MIT-67 indoor scenes dataset.

6. CONCLUSION

This paper proposes two discriminative dictionary learning methods by using two Fisher criteria: the traditional Fisher criterion and the marginal Fisher criterion, to overcome the disadvantages of the previous discriminative dictionary learning methods for sparse representation. It is further proved that the two proposed methods belong to a general model of learning the Fisher regularized sparse representation, which can be efficiently learned. The final experimental results on several popular databases show that the proposed methods are effective.

7. REFERENCES

[1] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma, "Robust face recognition via sparse

representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[2] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[3] Shuicheng Yan, Dong Xu, Benyu Zhang, HongJiang Zhang, Qiang Yang, and Stephen Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, 2007.

[4] Ke Huang and Selin Aviyente, "Sparse representation for signal classification," in *In Adv. NIPS*, 2006.

[5] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[6] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng, "Efficient sparse coding algorithms," in *NIPS*, 2007, pp. 801–808.

[7] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[8] Aleix M. Martínez and Avinash C. Kak, "Pca versus lda," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2001.

[9] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, 2006, pp. 2169–2178.

[10] Ariadna Quattoni and Antonio Torralba, "Recognizing indoor scenes," in *CVPR*, 2009, pp. 413–420.

[11] Shenghua Gao, Ivor Wai-Hung Tsang, and Liang-Tien Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, 2013.

[12] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas S. Huang, and Yihong Gong, "Locality-constrained linear coding for image classification.," in *CVPR*, 2010, pp. 3360–3367.

[13] Miao Zheng, Jiajun Bu, Chun Chen, Can Wang, Lijun Zhang, Guang Qiu, and Deng Cai, "Graph regularized sparse coding for image representation," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1327–1336, 2011.

[14] Qiang Zhang and Baoxin Li, "Discriminative k-svd for dictionary learning in face recognition.," in *CVPR*, 2010, pp. 2691–2698.

[15] Zhuolin Jiang, Zhe Lin, and Larry S. Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013.

[16] Ning Zhou, Yi Shen, Jinye Peng, and Jianping Fan, "Learning inter-related visual dictionary for object recognition," in *CVPR*, 2012, pp. 3490–3497.

[17] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, pp. 1–24, 2014.

[18] Qingfeng Liu and Chengjun Liu, "A new locally linear knn method with an improved marginal fisher analysis for image classification.," in *International Joint Conference on Biometrics*, 2014.

[19] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Fisher discrimination dictionary learning for sparse representation," in *ICCV*, 2011, pp. 543–550.

[20] Hanlin Goh, Nicolas Thome, Matthieu Cord, and Joo-Hwee Lim, "Learning deep hierarchical visual feature coding," *IEEE Transactions on Neural Networks and Learning Systems*, 2014.

[21] Jorge Sanchez, Florent Perronnin, Thomas Mensink, and Jakob J. Verbeek, "Image classification with the fisher vector: Theory and practice.," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013.

[22] Megha Pandey and Svetlana Lazebnik, "Scene recognition and weakly supervised object localization with deformable part-based models," in *ICCV*, 2011, pp. 1307–1314.

[23] Li-Jia Li, Hao Su, Eric P. Xing, and Fei-Fei Li, "Object bank: A high-level image representation for scene classification & semantic feature sparsification," in *NIPS*, 2010, pp. 1378–1386.

[24] Quannan Li, Jiajun Wu, and Zhuowen Tu, "Harvesting mid-level visual concepts from large-scale internet images," in *CVPR*, 2013, pp. 851–858.

[25] Jian Sun and Jean Ponce, "Learning discriminative part detectors for image classification and cosegmentation," in *ICCV*, 2013, pp. 3400–3407.

[26] Mayank Juneja, Andrea Vedaldi, C. V. Jawahar, and Andrew Zisserman, "Blocks that shout: Distinctive parts for scene classification," in *CVPR*, 2013, pp. 923–930.