LOCALITY-CONSTRAINED GROUP SPARSE REPRESENTATION FOR ROBUST FACE RECOGNITION

Yu-Wei Chao[1], *Yi-Ren Yeh*[1], *Yu-Wen Chen*[1,2], *Yuh-Jye Lee*[3], *and Yu-Chiang Frank Wang*[1]

[1]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
[2]Dept. Electrical Engineering, National Taiwan University, Taipei, Taiwan
[3]Dept. Computer Science & Info. Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

ABSTRACT

This paper presents a novel sparse representation for robust face recognition. We advance both group sparsity and data locality and formulate a unified optimization framework, which produces a locality and group sensitive sparse representation (LGSR) for improved recognition. Empirical results confirm that our LGSR not only outperforms state-of-the-art sparse coding based image classification methods, our approach is robust to variations such as lighting, pose, and facial details (glasses or not), which are typically seen in real-world face recognition problems.

Index Terms— Face recognition, sparse representation, group Lasso, data locality

1. INTRODUCTION

Face recognition is among the active research topics in pattern recognition and computer vision due to its wide applications in human-computer interaction, automatic photo-tagging, and information security [1]. Since real-world face images contain noisy background clutter and are typically with significant lighting, expression, pose, etc. variations, robust face recognition remains a very challenging task.

Numerous methods have been proposed to transform face image data to a lower dimensional feature space for recognition, e.g. Eigenfaces [2], Fisherfaces [2], and Laplacianfaces [3]. Nearest neighbor (NN) type classifiers are commonly used to recognize the projected data. Receiving growing attention from researchers, sparse coding (SC) [7] is a technique which reconstructs a target instance by a sparse linear combination of an over-complete dictionary (codebook). SC has been successfully applied for face recognition problems. For example, Wright *et al.* [4] proposed to convert the query face image into a sparse linear combination of training images with illumination, etc. variations. By imposing an ℓ_1-norm constraint on the resulting coefficients, their sparse representation classification (SRC) method achieved very promising results on face recognition. Yang *et al.* [6] further presented a metaface learning (MFL) approach, which aimed to construct a modified dictionary from training data for sparse representation and recognition. In [5], Yuan and Yan considered the group structure of training images (i.e. those from the same subject) and added an $\ell_{1,2}$ mixed-norm (group Lasso) constraint [8] upon the formulation. Their multi-task joint sparse representation was designed to produce a sparse solution at the group (class) level.

In general pattern recognition problems such as clustering, dimension reduction, data coding, etc., data locality has been observed to be critical [9, 10]. In prior SC-based approaches (including [4]), a test input might be reconstructed by training images (codewords), which are far from the test sample and thus produce unsatisfying classification results. In [10], the authors extended SC and proposed a locality-constrained linear coding (LLC) scheme, which learned a data representation using nearest codeword and achieved improved classification performances than the standard SC did. To the best of our knowledge, data locality and the aforementioned group sparsity constraints have not been jointly considered to address general image classification problems. Inspired by [4], we present a further extension of sparse representation for robust face recognition, called locality and group sensitive sparse representation (LGSR). Our LGSR aims to recover data sparse representation and achieve improved classification by integrating both group (class) sparsity and data locality structure into a unified formulation.

2. SPARSE CODING FOR IMAGE REPRESENTATION AND RECOGNITION

2.1. Image Classification via Sparse Representation

Sparse coding (SC) utilizes an over-complete dictionary to linearly reconstruct a data instance. For this instance, only a few weight coefficients will be non-zero, and thus the resulting coefficient vector is sparse, as shown in Figure 1a. In [4], a sparse representation classification (SRC) was proposed for face recognition. Using all training images from all subjects as the dictionary, SRC determines the sparse representation of a query input $\mathbf{x}_t$, and it classifies this input to the class if its associated reconstruction error is minimum.

More specifically, suppose that there are n training im-

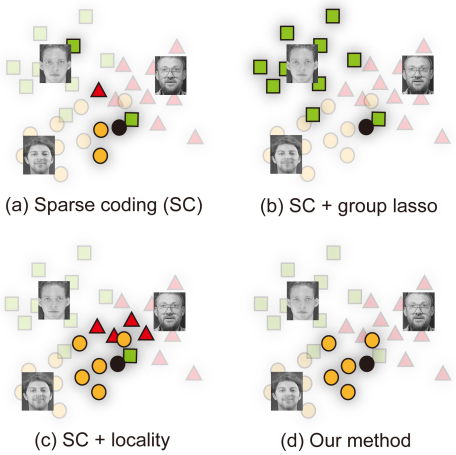

Fig. 1: Different image representation strategies: (a) sparse coding (SC) [4]. (b) SC with group Lasso [5], (c) SC with locality [10], and (d) our approach. The black circle is the target instance. Orange circles, red triangles, and green rectangles represent data from three different subjects, respectively.

ages from c different subjects, and each class j has n_j images available. We have $\mathbf{x}_{ji} \in \mathbb{R}^{m\times 1}$ as the image feature vector of the ith image in the jth class, and m is the dimensionality of the feature. Let $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_c] \in \mathbb{R}^{m\times n}$ be the entire training set, where $\mathbf{A}_j = [\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jn_j}] \in \mathbb{R}^{m\times n_j}$ contains training images of the jth class. The SRC minimizes the image reconstruction error with a ℓ_1-norm regularizer, i.e.

$$\min_{\mathbf{w}} \|\mathbf{x}_t - \mathbf{A}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1 \,, \tag{1}$$

where $\mathbf{w} = [\mathbf{w}_1; \mathbf{w}_2; \cdots; \mathbf{w}_c] = [w_{11}, w_{12}, \ldots, w_{cn_c}]^\top \in \mathbb{R}^{n\times 1}$, and $\mathbf{w}_j = [w_{j1}, w_{j2}, \ldots, w_{jn_j}]^\top \in \mathbb{R}^{n_j\times 1}$ are the weight coefficients of $\mathbf{A}_j$. Note that λ in (1) weights the ℓ_1-norm regularizer, which controls the sparsity of $\mathbf{w}$. All columns of $\mathbf{A}$ are normalized to unit length before solving (1). The formulation of (1) is also referred to as *Lasso* [11] in machine learning and statistics literatures.

The above $\mathbf{w}$ is considered as the image sparse representation of the input $\mathbf{x}_t$. While one can design a classifier such as SVM using the resulting vectors $\mathbf{w}$ for classification, SRC utilizes the reconstruction error to classify the query image. In other words, using the training samples from each class and the associated w_{ji}, the label of the query image will be assigned to the class with the minimum reconstruction error.

2.2. SC with Group Lasso for Classification

Although standard SC with ℓ_1-norm regularization in (1) produces a sparse coefficient vector $\mathbf{w}$, it has no control over which attributes to be zeroes (or non-zeroes); in other words, SC might reconstruct a query face image by training data from distinct subjects, and thus is not preferable for the task of classification. In [5], the authors proposed a joint sparse representation formulation. While all training samples were still used as codewords, those with the same class label were further defined as a group. A $\ell_{1,2}$ mixed-norm regularization term as the group sparsity constraint, also known as group Lasso [8], was further imposed on the reconstruction formulation. This constraint enforces non-zero coefficients to occur at few specific groups, while those within the same group can be non-sparse once that group is selected, as shown in Figure 1b.

Recall that $\mathbf{w}_j$ is the coefficient vector of $\mathbf{A}_j$, the modified formulation of (1) with group sparsity is now formulated as:

$$\min_{\mathbf{w}} \|\mathbf{x}_t - \mathbf{A}\mathbf{w}\|_2^2 + \lambda \sum_{j=1}^{c} \|\mathbf{w}_j\|_2. \tag{2}$$

From the above equation, one can see that a group-sensitive sparse solution $\mathbf{w}$ will be produced, while it tends to reconstruct the query image by training images from few correlated categories. Similar to [4], the query face image will be recognized as the class with the minimum reconstruction error.

2.3. Locality-Constrained SC for Classification

A major limitation of SC-based approaches for classification is that similar data instances do not guarantee to produce similar coding results (i.e. similar $\mathbf{w}$). Wang *et al.* [10] recently proposed a locality-constrained linear coding (LLC) for SC, and introduced a distance regularization when minimizing the reconstruction error:

$$\min_{\mathbf{w}} \|\mathbf{x} - \mathbf{A}\mathbf{w}\|_2^2 + \lambda\|\mathbf{d} \odot \mathbf{w}\|_2^2 \,, \tag{3}$$

where $\mathbf{A}$ is the codebook, $\mathbf{w}$ is the coefficient vector of $\mathbf{x}$, and $\mathbf{d} \in \mathbb{R}^{n\times 1}$ is the measurement of distance between $\mathbf{x}$ and each visual word in $\mathbf{A}$. Note that the symbol $\odot$ denotes element-wise multiplication. As shown in Figure 1c, minimizing (3) tends to encode the input using its nearby visual words, while the resulting $\mathbf{w}$ will still satisfy the sparsity constraint (a large d_i would make the corresponding w_{ji} shrink to zero).

It is worth noting that LLC and SRC share the same idea of sparsity in performing data reconstruction. Inspired by both group-sensitive representation in [5] and LLC [10], we propose a novel image sparse representation by imposing this locality constraint on the group Lasso regularized sparsity reconstruction problem, as we discuss in the next section.

3. LOCALITY AND GROUP-SENSITIVE SPARSE REPRESENTATION FOR IMAGE CLASSIFICATION

3.1. LGSR Algorithm

Our locality and group sensitive sparse representation (LGSR) algorithm advances the sparse coding technique, and takes advantages from both group sparsity and data locality structure in determining the optimal image representation for image classification. As shown in Figure 1, adopting the group lasso constraint may result in misclassification due to large within-group variations (e.g. pose). Integrating both group sparsity and locality constraints, our LGSR representation preserves

the similarity between the test input and its neighboring training data while seeking the optimal sparse representation.

Given a target instance $\mathbf{x}_t$, our LGSR formulates $\mathbf{x}_t$ as a compact linear combination of grouped training data (i.e. training samples with the same label). Similar to [5], we use the entire training set as our over-complete dictionary. We thus integrates the $\ell_{1,2}$ mixed-norm regularization and the locality constraint into a unified sparse representation formulation, and solve the following optimization problem:

$$\min_{\mathbf{w}} \|\mathbf{x}_t - \mathbf{A}\mathbf{w}\|_2^2 + \lambda_1 \sum_{j=1}^{c} \|\mathbf{w}_j\|_2 + \lambda_2 \|\mathbf{d} \odot \mathbf{w}\|_2^2 , \quad (4)$$

where λ_1 and λ_2 weight the group sparsity and locality constraints, respectively. In (4), the vector $\mathbf{d} \in \mathbb{R}^{n \times 1}$ penalizes the distance between $\mathbf{x}_t$ and each codeword (recall that n is the number of training instances from all classes). To measure the distance between x_t and each codeword (training sample) $\mathbf{x}_{ji}$, the distance metric is determined as:

$$d_{ji} = \exp\left(\frac{\|\mathbf{x}_t - \mathbf{x}_{ji}\|_2}{\sigma}\right). \quad (5)$$

A larger d_{ji} indicates a farther distance between $\mathbf{x}_t$ and $\mathbf{x}_{ji}$ (the ith training vector of class j). We note that this vector $\mathbf{d}$ is considered as a dissimilarity vector, and is used to suppress the corresponding weight coefficient w_{ji} in (4).

3.2. Optimization for LGSR and the Classification Rule

We note that the first and the third terms in (4), i.e. the reconstruction error and the data locality constraint, are both differentiable with respect to $\mathbf{w}$. Several methods such as [8] and [12] exist to solve such problems with the group Lasso constraint ($\ell_{1,2}$ mixed-norm regularization term). We apply the gradient-projection method proposed in [12], and use the software package [1] to solve our LGSR optimization problem.

Once (4) is minimized, the resulting coefficient vector $\mathbf{w}$ is used as the feature vector of the target instance $\mathbf{x}_t$. For classification, we first calculate the LGSR of the query input, and we recognize this input as the class with the lowest reconstruction error using *only* the associated coefficient attributes in $\mathbf{w}_j$. The decision process is shown as follows:

$$j^* = \arg\min_j \|\mathbf{x}_t - \mathbf{A}_j \mathbf{w}_j\|_2^2, \quad (6)$$

where $\mathbf{A}_j = [\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jn_j}] \in \mathbb{R}^{m \times n_j}$ contains training samples from the jth class.

It is clear that our LGSR produces a compact feature representation for a data instance, and this representation contains both group (class) and locality information for improved recognition. The group information in the LGSR coefficient vector implies the data reconstruction using the training samples from the specific group (class). Together with data locality information, the LGSR can be viewed as an extension of the k nearest neighbor (kNN) classifier. While our LGSR is a training-free (or lazy learning) algorithm (similar to kNN), the use of both group sparsity and locality constraints is expected to produce improved recognition performance.

[1] Available at http://www.cs.ubc.ca/~murphyk/Software

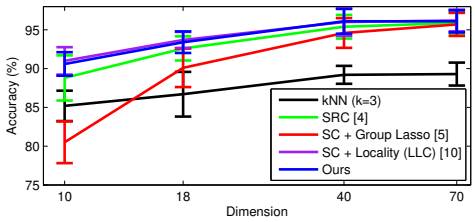

Fig. 2: Recognition accuracy (%) of different methods using Eigenface on the ORL database.

Table 1: Recognition of the ORL database using Fisherface.

kNN	SRC [4]	SC+GL [5]	SC+LLC [10]	Ours
93.20±1.10	94.20±0.76	94.40±0.82	95.00±1.32	**95.40±1.43**

4. EXPERIMENTAL RESULTS

4.1. ORL Database

The ORL database [13] contains 400 face images of 40 people, and each of size 112 × 92 pixels. The face images were taken under different *lighting, pose, facial details* (glasses or not), etc. conditions. We randomly and equally split the data into training and test sets (i.e. five images for each set), as [4] did. We first extract Eigenfaces as features and perform recognition using different number of Eigenfaces. We perform five random trials, and the recognition performance of different SC-based methods are reported in Figure 2. We note that in our implementation of LGSR, we use the standard deviation of training data $\mathbf{A}$ projected into the dominant eigenspace as the σ in (5). For each SC-based method (including ours), we vary the value of λ and present the best results for fair comparisons. We also consider the performance of kNN, since it is a training-free classifier utilizing data locality, and can be considered as a baseline approach. We set k = 3 for kNN because we observe that the average number of non-zero coefficients over four SC-based methods is 3.

From Figure 2, we see that our LGSR and locality constrained SC achieved comparable performances, and both outperformed other SC or kNN based approaches. Such improvements become more significant in lower dimensional spaces. From our experiments, we observed that the LGSR required a larger λ_2 in penalizing the data locality regularization, when a lower (dominant) dimensional space is considered. For example, we have $\lambda_1 = 0.1$ and $\lambda_2 = 0.5$ for our LGSR in the case using 10-dimensional Eigenface features. This implies that the data locality is better preserved in lower dimensional space, and a better recognition performance will be achieved if a stronger constraint on data locality is imposed (rather than the group sparsity one).

In addition to Eigenface, we also perform Fisherface to extract the image features and repeat the above experiments

Table 2: Comparisons of recognition performance on Extended Yale B. N_T indicates the number of training images.

N_T	8	16	32
kNN (k=3)	76.70 ± 1.14	86.99 ± 1.36	93.29 ± 0.42
SRC [4]	84.38 ± 1.21	92.95 ± 0.42	97.83 ± 0.14
SC+GL [5]	84.39 ± 1.21	92.95 ± 0.42	97.83 ± 0.17
SC+LLC [10]	84.74 ± 1.24	93.17 ± 0.19	97.89 ± 0.13
Ours	$\mathbf{85.17 \pm 1.15}$	$\mathbf{93.54 \pm 0.40}$	$\mathbf{98.14 \pm 0.21}$

(we use all 39 eigenspaces for Fisherface). Table 1 shows the recognition accuracy of the methods considered. Compared to the results using Eigenfaces with dimension 40, kNN has a remarked improvement from 89.2% to 93.2%. The is because the within-class variation is suppressed using Fisherface, while the separation of projected data from different classes is improved. This also explains why the SC with the group Lasso constraint performed better than the standard SRC using Fisherface, but not in the case of Eigenface.

4.2. Extended Yale B Database

The cropped Extended Yale [14] consists of 2414 frontal images of 38 subjects, each image has up to 64 illumination variations. We extract Fisherfaces as the features, and consider different number of training images (N_T) per class for evaluation. Once the training images are extracted, the remaining will be test images. We perform three random trials, and the recognition performances are shown in Table 2.

From Table 2, we see that our LGSR consistently outperforms other SC-based methods. Figure 3a shows an example query image which was correctly recognized by our LGSR but not by others, and Figure 3b shows the training images selected by different methods. The values below each image in Figure 3b are the associated weights. From the first three rows in Figure 3b, we see that the query image was reconstructed by different sets of training images, but those with the largest weights were not from the correct class to be recognized. In the third row of Figure 3b, large coefficients were assigned to images with smaller distances to the test image due to noise (i.e. those with similar illumination variations). In the last row of Figure 3b, our method identified two groups of training data for LGSR representation. The training images in the correct group were assigned larger weights, and thus the query input was able to be successfully recognized.

5. CONCLUSION

A locality and group sensitive sparse representation (LGSR) was presented for robust face recognition. Our LGSR balances data group sparsity and locality, and produces an improved image representation for classification by solving a unified optimization problem. Comparing to prior sparse coding based approaches, our LGSR is robust to lighting, pose, facial details, etc. variations in face recognition, and achieved very promising recognition results.

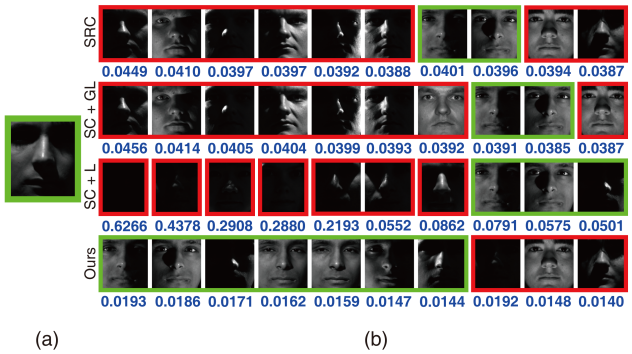

Fig. 3: (a) An input test image. (b) Selected training images with non-zero weights (the numbers below each image) using SC, SC+group Lasso, SC+locality, and our method. Images bounded by green rectangles are the correct class, and the red rectangle ones are faces from other identities.

Acknowledgements This work is supported in part by National Science Council of Taiwan via NSC 99-2221-E-001-020 and NSC 100-2631-H-001-013.

6. REFERENCES

[1] W. Zhao et al., "Face recognition: A literature survey," *ACM Computing Surveys*, 2003. 1

[2] P. N. Belhumeur et al., "Eigenfaces vs. fisherfaces recognition using class specific linear projection," *IEEE PAMI*, 1997. 1

[3] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE PAMI*, 2005. 1

[4] J. Wright et al., "Robust face recognition via sparse representation," *IEEE PAMI*, 2009. 1, 2, 3, 4

[5] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *IEEE CVPR*, 2010. 1, 2, 3, 4

[6] M. Yang et al., "Metaface learning for sparse representation based face recognition," in *IEEE ICIP*, 2010. 1

[7] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, 2006. 1

[8] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B*, 2006. 1, 2, 3

[9] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, 2000. 1

[10] J. Wang et al., "Locality-constrained linear coding for image classification," in *IEEE CVPR*, 2010. 1, 2, 3, 4

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1996. 2

[12] M. Schmidt et al., "Structure learning in random fields for heart motion abnormality detection," in *IEEE CVPR*, 2008. 3

[13] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *IEEE Workshop on Applications of Computer Vision*, 1994. 3

[14] K.-C. Lee et al., "Acquiring linear subspaces for face recognition under variable lighting," *IEEE PAMI*, 2005. 4