COUPLED DICTIONARY LEARNING FOR MULTIMODAL IMAGE SUPER-RESOLUTION

Pingfan Song[*]   *João F. C. Mota*[*]   *Nikos Deligiannis*[†‡]   *Miguel R. D. Rodrigues*[*]

[*] Department of Electronic and Electrical Engineering, University College London, UK
[†] Department of Electronics and Informatics, Vrije Universiteit Brussel, Belgium
[‡] iMinds vzw, Ghent, Belgium

ABSTRACT

Real-world data processing problems often involve multiple data modalities, e.g., panchromatic and multispectral images, positron emission tomography (PET) and magnetic resonance imaging (MRI) images. As these modalities capture information associated with the same phenomenon, they must necessarily be correlated, although the precise relation is rarely known. In this paper, we propose a coupled dictionary learning (CDL) framework to automatically learn these relations. In particular, we propose a new data model to characterize both similarities and discrepancies between multimodal signals in terms of common and unique sparse representations with respect to a group of coupled dictionaries. However, learning these coupled dictionaries involves solving a highly non-convex structural dictionary learning problem. To address this problem, we design a coupled dictionary learning algorithm, referred to sequential recursive optimization (SRO) algorithm, to sequentially learn these dictionaries in a recursive manner. By capitalizing on our model and algorithm, we conceive a CDL based multimodal image super-resolution (SR) approach. Practical multispectral image SR experiments demonstrate that our SR approach outperforms the bicubic interpolation and the state-of-the-art dictionary learning based image SR approach, with Peak-SNR (PSNR) gains of up to 8.2 dB and 5.1 dB, respectively.

Index Terms— coupled dictionary learning, multimodal data, sparse representation, sequential recursive optimization, multispectral image super-resolution

1. INTRODUCTION

Image super-resolution (SR) is a set of techniques to enhance pixel-based image resolution, while minimizing visual artifacts. Due to insufficient number of observations, image SR is a severely ill-conditioned problem that needs to be regularized via employing various image models and prior knowledge [1–13]. Among these models, image SR based on sparse representation over learned adaptive dictionaries has received

an increased interest [5–8]. Dictionary learning (DL) techniques [14–17] are playing a key role in such image SR techniques, leading to state-of-the-art performance results [8].

However, in some practical application scenarios, the observed images for a certain scene are often captured by different sensors. For example, in the medical imaging domain, simultaneous PET-MRI scanning provides both positron emission tomography (PET) and magnetic resonance imaging (MRI) data for the same underlying anatomy [18–20]. In the field of remote sensing, the earth observation for the same geographical region is commonly composed of multiple heterogeneous images, e.g., panchromatic band version, multispectral bands version, infrared (IR) band version, etc. [21–26]. In order to balance the cost, bandwidth and complexity, these multimodal images are usually acquired with different resolution [26]. This calls for multimodal image super-resolution approaches in order to enhance the resolution of heterogeneous images. Since these heterogeneous images originate from the same phenomenon, it is realistic to assume that they admit some common inherent characteristics. This belief motivates us to develop a new coupled dictionary learning (CDL) framework to capture the relations between multimodal signals. Then, based on the proposed CDL framework, we conceive a multimodal image super-resolution approach to exploit the learned prior knowledge of the relation to enhance image resolution.

2. RELATION TO PRIOR WORK

Most dictionary learning based image SR approaches [5–8] have two phases: learning some prior knowledge from training data and performing image SR with the aid of the learned prior knowledge. Yang et. al. [4, 5] proposed a sparse representation invariance assumption that a pair of low-resolution/high-resolution (LR/HR) images share the same sparse representation with respect to their corresponding dictionaries. Based on this assumption, [4, 5, 7] propose to jointly learn a pair of LR/HR dictionaries to capture the relation between LR/HR image pairs characterized by the common sparse representation. Image SR approaches based on coupled dictionary learning (CDL) [6] and semi-coupled dictionary learning (SCDL) [27, 28] allow more flexible mapping relations between sparse representations of LR/HR im-

age pairs. However, they still focus on images associated with a single modality and hence only involve training one LR/HR dictionary pair.

In contrast to existing work, our proposed multimodal data model expresses multimodal signals in terms of common sparse components (which capture the correlation between the distinct data modalities) and unique components (which capture the particular aspects of each individual data modality). These common and unique sparse components are discovered via our proposed sequential recursive optimization (SRO) algorithm tailored for the coupled dictionary learning problem, which differs from convectional dictionary learning ones.

3. COUPLED DICTIONARY LEARNING FOR MULTIMODAL DATA

We first introduce our data model and explain how it captures correlation between multimodal data. Next, based on this model, we propose a coupled dictionary learning problem and an associated algorithm. Finally, we show how to use the framework to perform multimodal image SR.

3.1. Multimodal Data Model

Given two classes of data of different modalities, $\mathbf{X}$ and $\mathbf{Y}\left(\in \mathbb{R}^{N \times T}\right)$, we assume that the set of data can be decomposed as

$$
\begin{aligned}
\mathbf{X} &= \boldsymbol{\Psi}_c \mathbf{Z} + \boldsymbol{\Psi}\, \mathbf{U}\,, \\
\mathbf{Y} &= \boldsymbol{\Phi}_c \mathbf{Z} + \boldsymbol{\Phi}\, \mathbf{V}\,,
\end{aligned} \tag{1}
$$

where, $\boldsymbol{\Psi}_c$, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_c$ and $\boldsymbol{\Phi}\left(\in \mathbb{R}^{N \times K}\right)$ are the dictionaries to be found, and $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{V}\left(\in \mathbb{R}^{K \times T}\right)$ are the corresponding sparse representations, also to be found. T corresponds to the number of training examples, N denotes the ambient data dimensionality (which we take to be the same for both data modalities without loss of generality) and K corresponds to the number of atoms in the various dictionaries (which we also take to be the same without loss of generality). The common sparse components $\mathbf{Z}$, with respect to $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Phi}_c$, express common characteristics underlying both $\mathbf{X}$ and $\mathbf{Y}$. In contrast, the unique sparse components $\mathbf{U}$ and $\mathbf{V}$, w.r.t $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ respectively, express unique characteristics associated with the data modalities $\mathbf{X}$ and $\mathbf{Y}$.

3.2. Coupled Dictionary Learning

Based on the data model (1), the coupled dictionary learning problem can be cast as

$$
\begin{aligned}
& \underset{\substack{\{\boldsymbol{\Psi}_c, \boldsymbol{\Psi}, \boldsymbol{\Phi}_c, \boldsymbol{\Phi}\} \\ \{\mathbf{Z}, \mathbf{U}, \mathbf{V}\}}}{\text{minimize}} \quad \left\| \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Psi}_c & \boldsymbol{\Psi} & 0 \\ \boldsymbol{\Phi}_c & 0 & \boldsymbol{\Phi} \end{bmatrix} \begin{bmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{V} \end{bmatrix} \right\|_F^2 \\
& \text{subject to} \quad \|\mathbf{z}_i\|_0 + \|\mathbf{u}_i\|_0 + \|\mathbf{v}_i\|_0 \le s,\ \forall i.
\end{aligned} \tag{2}
$$

where $\mathbf{z}_i$, $\mathbf{u}_i$ and $\mathbf{v}_i$ denote the i-th column of matrix $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{V}$, respectively. $\|\cdot\|_F$ and $\|\cdot\|_0$ denote the Frobenius norm and ℓ_0 norm, respectively.
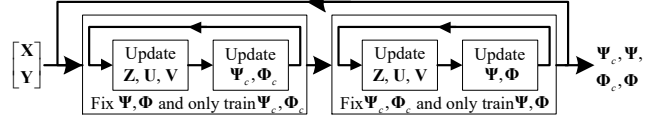

Fig. 1: Flowchart of the proposed SRO algorithm for CDL.

Algorithm 1 Sequential Recursive Optimization (SRO)

Input: Training data matrices $\mathbf{X}$ and $\mathbf{Y}\left(\in \mathbb{R}^{N \times T}\right)$.
Output: Coupled dictionaries $\boldsymbol{\Psi}_c$, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_c$, $\boldsymbol{\Phi}\left(\in \mathbb{R}^{N \times K}\right)$.
Initialization:
1: Initialize $\boldsymbol{\Psi}_c$, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_c$ and $\boldsymbol{\Phi}$.
2: Set sparsity constraints s.
3: Set main iteration $MainIter$ and sub-iteration $SubIter$.
Optimization:
4: **for** $p = 1$ to $MainIter$ **do**
5:     **for** $q = 1$ to $SubIter$ **do**
6:       // Fix $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$ and only train $\boldsymbol{\Psi}_c$, $\boldsymbol{\Phi}_c$ in the loop.
7:       **Global Sparse Coding**. Fix all the dictionaries, then use any off-the-shelf sparse coding algorithms to solve (3) to obtain updated sparse representations $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{V}$.

$$
\begin{aligned}
& \underset{\mathbf{Z}, \mathbf{U}, \mathbf{V}}{\min} \quad \left\| \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Psi}_c & \boldsymbol{\Psi} & 0 \\ \boldsymbol{\Phi}_c & 0 & \boldsymbol{\Phi} \end{bmatrix} \begin{bmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{V} \end{bmatrix} \right\|_F^2 \\
& \text{s.t.} \quad \|\mathbf{z}_i\|_0 + \|\mathbf{u}_i\|_0 + \|\mathbf{v}_i\|_0 \le s,\ \forall i.
\end{aligned} \tag{3}
$$

8:       **Partial Dictionary Update**. Fix $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$, $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{V}$, then only update $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Phi}_c$ via solving

$$
\underset{\boldsymbol{\Psi}_c, \boldsymbol{\Phi}_c}{\min} \quad \left\| \begin{bmatrix} \mathbf{X} - \boldsymbol{\Psi}\mathbf{U} \\ \mathbf{Y} - \boldsymbol{\Phi}\mathbf{V} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Psi}_c \\ \boldsymbol{\Phi}_c \end{bmatrix} \mathbf{Z} \right\|_F^2 \tag{4}
$$

9:     **end for**
10:     **for** $q = 1$ to $SubIter$ **do**
11:       // Fix $\boldsymbol{\Psi}_c$, $\boldsymbol{\Phi}_c$ and only train $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$ in the loop.
12:       **Global Sparse Coding**. The same as step 7.
13:       **Partial Dictionary Update**. Fix $\boldsymbol{\Psi}_c$, $\boldsymbol{\Phi}_c$, $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{V}$, then only update $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ via solving

$$
\underset{\boldsymbol{\Psi}}{\min} \quad \|(\mathbf{X} - \boldsymbol{\Psi}_c\mathbf{Z}) - \boldsymbol{\Psi}\mathbf{U}\|_F^2 \tag{5}
$$

$$
\underset{\boldsymbol{\Phi}}{\min} \quad \|(\mathbf{Y} - \boldsymbol{\Phi}_c\mathbf{Z}) - \boldsymbol{\Phi}\mathbf{V}\|_F^2 \tag{6}
$$

14:     **end for**
15: **end for**

Problem (2) is a highly non-convex dictionary learning problem with particular requirement on the structure, which can not be solved directly using conventional DL algorithms. To address this structural dictionary learning problem, we propose the sequential recursive optimization (SRO) algorithm to sequentially learn these coupled dictionaries part by part, as described in Algorithm 1 (see also Fig.1). Note that, (3), (4),(5) and (6) are convex. As problem (2) is non-convex,

our SRO algorithm can not guarantee the convergence on a global optimum. But, given good initialization for dictionaries, SRO will converge to a reasonable local optimum, leading to a satisfactory approximate solution for (2).

3.3. Multimodal Image Super-resolution

By capitalizing on the previous coupled dictionary learning framework, which learns dictionaries that couple a pair of data modalities, it is now possible to conceive a multimodal image super-resolution approach.

Let $\mathbf{X}^h \in \mathbb{R}^{N \times T}$ and $\mathbf{Y}^h \in \mathbb{R}^{N \times T}$ denote two classes of HR images with different modalities. Let $\mathbf{X}^l \in \mathbb{R}^{M \times T}$ ($M < N$) denote the LR image derived from $\mathbf{X}^h$ as follows:

$$\mathbf{X}^l = \mathbf{A}\,\mathbf{X}^h, \tag{7}$$

where the measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ denotes an observation operator that extracts a low-resolution version from $\mathbf{X}^h$. Then, according to the data model (1), $\mathbf{X}^l$, $\mathbf{X}^h$, $\mathbf{Y}^h$ can be expressed as follows:

$$\mathbf{X}^h = \boldsymbol{\Psi}_c^h\,\mathbf{Z} + \boldsymbol{\Psi}^h\,\mathbf{U}, \tag{8}$$

$$\mathbf{X}^l = \boldsymbol{\Psi}_c^l\,\mathbf{Z} + \boldsymbol{\Psi}^l\,\mathbf{U}, \tag{9}$$

$$\mathbf{Y}^h = \boldsymbol{\Phi}_c^h\,\mathbf{Z} + \boldsymbol{\Phi}^h\,\mathbf{V}, \tag{10}$$

where $\boldsymbol{\Psi}_c^l = \mathbf{A}\boldsymbol{\Psi}_c^h$ and $\boldsymbol{\Psi}^l = \mathbf{A}\boldsymbol{\Psi}^h$. This suggests immediately an approach to super-resolve a single LR image modality aided by another HR data modality. In particular, let us assume that the coupled dictionaries have been learnt for both data modalities as a priori using the proposed SRO algorithm. Let us also assume that we are given a new LR image $\mathbf{x}_{test}^l$ and the corresponding HR image $\mathbf{y}_{test}^h$. Then, we can super-resolve the LR image modality $\mathbf{x}_{test}^l$ via solving the optimization problem (11) and then obtaining the estimation $\mathbf{x}_{test}^h$ immediately from $\mathbf{x}_{test}^h = \boldsymbol{\Psi}_c^h\mathbf{z} + \boldsymbol{\Psi}^h\mathbf{u}$.

$$\begin{aligned} \min_{\mathbf{z},\mathbf{u},\mathbf{v}} \quad & \|\mathbf{z}\|_0 + \|\mathbf{u}\|_0 + \|\mathbf{v}\|_0 \\ \text{s.t.} \quad & \mathbf{x}_{test}^l = \boldsymbol{\Psi}_c^l\mathbf{z} + \boldsymbol{\Psi}^l\mathbf{u}. \\ & \mathbf{y}_{test}^h = \boldsymbol{\Phi}_c^h\mathbf{z} + \boldsymbol{\Phi}^h\mathbf{v}. \end{aligned} \tag{11}$$

4. PERFORMANCE RESULTS

We now conduct a series of simulation and experiments to verify the effectiveness of the proposed SRO algorithm and CDL based multimodal image SR approach. In our SRO algorithm, orthogonal matching pursuit (OMP) algorithm [29] is used for the global sparse coding and K-SVD [16] is modified for the partial dictionary update.

4.1. Simulation with Synthetic Data

Here, we show a certain simulation setting. Four Gaussian random matrices $\boldsymbol{\Psi}_c$, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_c$ and $\boldsymbol{\Phi}$ $\left(\in \mathbb{R}^{N \times K}\right)$ are generated as the true dictionaries, where $N = 64$ and $K = 256$. Three sparse random matrices $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{V}$ $\left(\in \mathbb{R}^{K \times T}\right)$ are generated as sparse representations, in which the sparsity of each column is set to be $s_z = 4$, $s_u = 2$ and $s_v = 2$, respectively, thus $s = s_z + s_u + s_v = 8$. T is set as 10000. The training dataset is synthesized via our model (1). $MainIter$ and

(a) Atoms recovery ratio   (b) Training error convergence
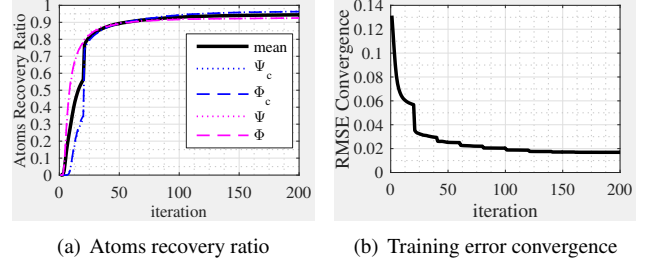
Fig. 2: Simulation results for Coupled Dictionary Learning using the SRO algorithm. 50 trials were conducted and their results were averaged. The "mean" in sub-figure (a) denotes the average ratio of the four dictionaries.

$SubIter$ are set to be 10 and 20, respectively. Dictionaries are initialized with random matrices. We adopt the atoms recovery ratio[1] and root-mean-square error (RMSE) as the criterion to evaluate the performance of the proposed algorithm. They are defined as

$$r = \frac{\sharp\left(d \in \mathbf{D} \cap \widehat{\mathbf{D}}\right)}{\sharp\left(d \in \mathbf{D}\right)}, \quad RMSE = \sqrt{\frac{\|\mathbf{X} - \widehat{\mathbf{D}}\boldsymbol{\alpha}\|_F^2}{\sharp(\mathbf{X})}},$$

where $\mathbf{D}$ and $\widehat{\mathbf{D}}$ denote the true and the learned dictionary, respectively, $\mathbf{X}$ and $\boldsymbol{\alpha}$ are the data and sparse representation, respectively, d denotes any one atom, $\sharp(\cdot)$ counts the number of elements in the set. Two atoms from $\mathbf{D}$ and $\widehat{\mathbf{D}}$ are considered identical when their Euclidean distance is less than a defined threshold ϵ with $\epsilon = 0.01$ by default.

Fig. 2 shows the results averaged on 50 trials. It illustrates that the proposed SRO algorithm achieves satisfactory atoms recovery ratio ($> 94\%$) with reasonable RMSE (< 0.02). This indicates that the coupled dictionaries learned by our algorithm are very good approximations to the true coupled dictionaries.

4.2. Multimodal image SR Experiments

Setup. In the multimodal image SR experiments, multispectral images are the signals of interest and corresponding RGB images serve as the side information. Our CDL based SR approach is also compared to the bicubic interpolation and the DL based SR approach. The magnification factor is set to be 2. The practical experiment data is obtained from the Columbia multispectral database[2]. In the experiments, 28 scenes are separated into two groups: the training image group consisting of 20 scenes and the testing image group for the remaining 8 scenes. For each scene, both modalities are already aligned with each other. The multispectral images of given wavelength and corresponding RGB images from the training image group are subdivided into many 8×8 patches which are then vectorized to form two training matrices $\mathbf{X}$ and $\mathbf{Y}$ with dimension 64×13000. DCT matrices are used as the initialization of dictionaries during the training.

---

[1]Please refer to Ron Rubinstein's K-SVD package for more details.
[2]http://www.cs.columbia.edu/CAVE/databases/multispectral/

Table 1: PSNR results of multispectral image super-resolution via Bicubic interpolation, DL based SR, and CDL based SR.

| wave	Img No.1			Img No.2			Img No.3			Img No.4			Img No.5			Img No.6			Img No.7			Img No.8		
/nm	Bic	DL	CDL	Bic	DL	CDL	Bic	DL	CDL	Bic	DL	CDL	Bic	DL	CDL	Bic	DL	CDL	Bic	DL	CDL	Bic	DL	CDL
440	30.0	32.7	37.1	31.2	33.6	37.8	38.2	41.2	45.0	31.0	34.1	37.6	28.0	31.5	35.0	29.8	31.8	35.9	32.2	36.5	38.8	31.9	35.0	40.4
490	30.9	33.6	39.8	31.7	33.2	41.4	41.6	44.4	49.9	33.9	35.1	42.2	30.5	32.3	38.9	30.5	32.3	38.1	30.4	34.5	38.2	30.4	34.1	39.5
540	30.7	33.4	39.1	32.0	32.4	41.8	40.1	43.6	49.7	33.3	37.1	42.5	30.3	32.1	40.3	30.5	30.7	38.6	31.7	35.5	41.6	29.8	33.9	39.2
590	30.2	32.8	38.8	30.2	32.8	40.1	38.7	42.5	48.1	33.1	37.3	42.6	29.5	33.1	39.0	29.6	33.3	38.4	31.3	35.2	41.0	31.2	33.0	41.4
640	30.8	36.3	39.3	29.2	31.6	35.7	37.7	41.7	45.3	32.2	34.9	40.6	28.4	32.1	37.0	29.0	32.1	37.0	32.0	36.4	39.3	33.3	35.5	43.7
690	31.5	36.9	38.9	29.8	33.2	34.0	37.0	39.7	43.8	31.5	35.2	38.8	27.8	32.2	35.9	29.1	33.8	35.9	33.8	37.6	41.6	34.6	40.8	43.7
mean	30.7	34.3	38.8	30.7	32.8	38.4	38.9	42.2	47.0	32.5	35.6	40.7	29.1	32.2	37.7	29.7	32.3	37.3	31.9	36.0	40.1	31.9	35.4	41.3

During the image SR phase, the testing dataset $\mathbf{X}_{test}$ and $\mathbf{Y}_{test}$ are constructed in a similar manner. Images from the testing group are subdivided into overlapping patches with overlap stride equal to 1 pixel[3]. In the experiments, the measurement matrix $\mathbf{A}$ is a down-sampling matrix which is assumed to be known in advance for both dictionary learning and coupled dictionary learning. Accordingly, the LR dictionaries can be exactly derived from the corresponding HR dictionaries. Thus, there exists no mismatch between LR and HR dictionaries which may interfere the fair comparison between coupled dictionaries and uncoupled ones.

Results. Fig. 3 shows the learned coupled dictionaries and overall training error convergence for the multispectral images of wavelength 590 nm and the corresponding RGB images. We can find that there exists resemblance between learned coupled HR dictionaries $\mathbf{\Psi_c}$ and $\mathbf{\Phi_c}$, which indicate that they capture the similarities between the two image modalities. In contrast, $\mathbf{\Psi}$ and $\mathbf{\Phi}$ represent the discrepancies and therefore rarely exhibit resemblance. Fig. 4 gives the multispectral image SR results for a certain scene with different wavelength versions. As we can see, proposed CDL based SR approach is able to reliably recover more image details, e.g., edges, textures, and in the meanwhile substantially suppress ringing artifacts than the counterparts. The overall comparison of the image SR performance in terms of Peak-SNR (PSNR) is shown in Table 1 which demonstrates that our SR approach outperforms the bicubic interpolation and the DL based image SR approach with PSNR gains of average 8.2 dB and 5.1 dB, respectively. The good performance of our approach is attributed to the beneficial information extracted from the RGB images by the learned coupled dictionaries. Limited to the space, more detailed results can be found in our website[4].

On the other hand, despite of good performance, our algorithm pays a price on the computational complexity, due to the time-consuming sparse coding for more dictionaries. On the same 3.4 GHZ PC, DL and CDL take average 3.82 and 34.88 minutes to learn a group of dictionaries, respectively. During the image SR phase, bicubic interpolation, DL based SR and CDL based SR take average 0.0035s, 59.89s, 259.5s

---

[3]The overlap stride denotes the distance between corresponding pixel locations in adjacent image patches.
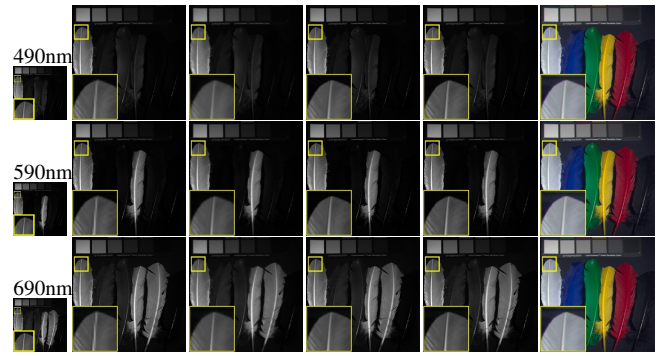[4]http://www.ee.ucl.ac.uk/~uceeong/
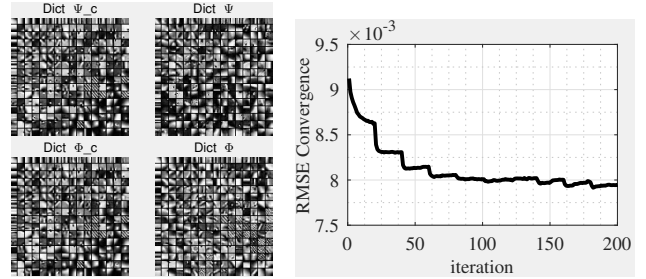
(a) Learned coupled HR dictionaries

(b) Training error convergence

Fig. 3: Coupled dictionary learning for multispectral images of wavelength 590nm and RGB images.

(a)LR (b) Bicubic (c) DL (d) CDL (e) True HR (e) Side info

Fig. 4: Visual comparison of multispectral image SR performance of 3 methods: bicubic interpolation, DL based SR approach and proposed CDL based SR approach. RGB images serve as the side information.

for processing one 512×512 image, respectively.

5. CONCLUSION

In this paper, we propose a new data model to characterize the correlation between multimodal data via the sparse representations with respect to a group of coupled dictionaries. In order to learn these coupled dictionaries, we propose the sequential recursive optimization (SRO) algorithm to solve a non-convex structural dictionary learning problem. Then, we establish a multimodal image SR scheme. Simulations with synthetic data and practical experiments with multispectral images demonstrate the good performance of our approach over conventional image SR approaches.

6. REFERENCES

[1] Hsieh S Hou and H Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. on Acoustics, Speech and Signal Process.*, vol. 26, no. 6, pp. 508–517, 1978.

[2] Richard R Schultz and Robert L Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, 1994.

[3] William T Freeman, Egon C Pasztor, and Owen T Carmichael, "Learning low-level vision," *Intl. J. computer vision*, vol. 40, no. 1, pp. 25–47, 2000.

[4] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma, "Image super-resolution as sparse representation of raw image patches," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[5] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

[6] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Tingwen Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, 2012.

[7] Roman Zeyde, Michael Elad, and Matan Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, pp. 711–730. Springer, 2010.

[8] Radu Timofte, Vincent De Smet, and Luc Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Computer Vision–ACCV 2014*, pp. 111–126. Springer, 2014.

[9] Michael Elad and Dmitry Datsenko, "Example-based regularization deployed to super-resolution reconstruction of a single image," *The Computer Journal*, vol. 52, no. 1, pp. 15–30, 2009.

[10] Jian Sun, Jian Sun, Zongben Xu, and Heung-Yeung Shum, "Image super-resolution using gradient profile prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[11] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.

[12] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *arXiv preprint arXiv:1501.00092*, 2014.

[13] Gungor Polatkan, MengChu Zhou, Lawrence Carin, David Blei, and Ingrid Daubechies, "A Bayesian nonparametric approach to image super-resolution," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 37, no. 2, pp. 346–358, 2015.

[14] Bruno A Olshausen et al., "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[15] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy, "Method of optimal directions for frame design," in *IEEE International Conference on Acoustics, Speech, and Signal Process.*, 1999, vol. 5, pp. 2443–2446.

[16] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[17] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[18] Simon R Cherry, "Multimodality in vivo imaging systems: twice the power or double the trouble?," *Annu. Rev. Biomed. Eng.*, vol. 8, pp. 35–62, 2006.

[19] DW Townsend, "Multimodality imaging of structure and function," *Physics in medicine and biology*, vol. 53, no. 4, pp. R1, 2008.

[20] Ciprian Catana, Alexander R Guimaraes, and Bruce R Rosen, "Pet and mr imaging: the odd couple or a match made in heaven?," *Journal of Nuclear Medicine*, vol. 54, no. 5, pp. 815–824, 2013.

[21] Luis Gomez-Chova, Devis Tuia, Gabriele Moser, and Gustau Camps-Valls, "Multimodal classification of remote sensing images: a review and future directions," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.

[22] Leyuan Fang, Shutao Li, Xudong Kang, and Jon Atli Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. on Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, 2014.

[23] Adriana Romero, Carlo Gatta, and Gustau Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *arXiv preprint arXiv:1511.08131*, 2015.

[24] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu, "Deep learning-based classification of hyperspectral data," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 2094–2107, 2014.

[25] Shashi Shekhar, Vishal M Patel, Nasser M Nasrabadi, and Rama Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 113–126, 2014.

[26] Laetitia Loncan, Sophie Fabre, Luis B Almeida, Jose M Bioucas-Dias, Wenzhi Liao, Xavier Briottet, Giorgio A Licciardi, Jocelyn Chanussot, Miguel SimoÌ Čes, Nicolas Dobigeon, et al., "Hyperspectral pansharpening: a review," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 3, no. 3, pp. 27–46, 2015.

[27] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2216–2223.

[28] Kui Jia, Xiaogang Wang, and Xiaoou Tang, "Image transformation based on learning dictionaries across image spaces," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 367–380, 2013.

[29] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.