ADAPTIVE IMAGE DECOMPOSITION VIA DICTIONARY LEARNING WITH STUCTURAL INCOHERENCE

Qiegen Liu[1,2], *Jianbo Liu*[1] *and Dong Liang*[1,*]

[1]Paul C. Lauterbur Research Centre for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen, China
[2]Department of Electronic Information Engineering, Nanchang University, Nanchang, China

ABSTRACT

Initialization sensitivity usually occurs in dictionary learning algorithm for image decomposition. In this paper, we propose an adaptive dictionary learning algorithm by promoting structural incoherence at the stage of dictionary updating. The structural incoherence based dictionary learning (SIDL) method guides the cartoon and texture parts to be more properly represented by two incoherent dictionaries. The resulting minimization is approximately addressed by majorization-minimization (MM) technique. Experimental results demonstrate that the dictionaries generated by SIDL can better describe different morphological contents and subsequently the cartoon and texture components are better separated, in terms of visual comparisons and quantitative measures.

Index Terms—Image decomposition, sparse representation, dictionary learning, structural incoherence, majorizaion-minimization

1. INTRODUCTION

Image decomposition is usually the first step to the solution of many image processing tasks. Suppose an image f is separated as the sum of two independent components $f = u_1 + u_2$: where the piecewise smooth function u_1 with quasi-flat intensity plateaus and jump discontinuities is usually called "cartoon", that contains main large-scale structure features of the image; u_2 is a small-scale oscillatory function capturing texture and possibly noise, and usually has some periodicity and oscillatory nature. The structure component u_1 can be used for feature detection, segmentation and object recognition, while the texture component u_2 for solving various texture-depended problems e.g. classification, surface analysis, shape/orientation from texture. Recently publications in a wide applications like inpainting [1], demosaicing [2] and registration [3] have shown that this "decompose and solve one by one" strategy is very successful, by adapting algorithms for two different components.

Variational approaches are the most popular approaches to address the image decomposition problem. The common methods are often related to total variation (TV) minimization. They perform decomposition by modeling the cartoon u_1 with TV semi-norm and using other appropriate norms for oscillating features u_2. The original formulation of such approaches was due to Meyer [4], who suggested starting from the ROF model [5]:

$$u_1 = \arg\min_{u_1} \{\lambda \|f - u_1\|_2^2 + \|TV(u_1)\|_1\} \tag{1}$$

where the fidelity term is measured by L^2-norm. Since this model rejects the textures, Meyer used a new function space G, and replaced the L^2-norm by the G-norm. It was proved that G corresponds to a space of oscillating functions, and thus is useful to model textures. Some approximated norm like $div(L^p)$-norm [6] and H^{-1}-norm [7] were developed by following this idea.

In recent years, applying sparse representation with general transforms/dictionaries to image decomposition has received an increasing amount of interest, under the name of morphological component analysis (MCA) [8-11]. MCA is an extension of TV-based model. In MCA, the given signal/image is decomposed into different morphological components, subject to the sparsity of each component in a known basis (or dictionary). One possible objective function can be formulated as follows:

$$\{\hat{u}_1, \hat{u}_2, \hat{\alpha}_i, \hat{\beta}_i\} = \arg\min_{u_1,\alpha_i,u_2,\beta_i} \left[\begin{array}{l} \lambda \|f - u_1 - u_2\|_2^2 + \sum_{i\in I} \|R_i u_1 - D_1\alpha_i\|_2^2 + \mu\|\alpha_i\|_0 \\ \qquad + \sum_{i\in I} \|R_i u_2 - D_2\beta_i\|_2^2 + \eta\|\beta_i\|_0 \end{array} \right] \tag{2}$$

* Correspondence to: Dong Liang (dong.liang@siat.ac.cn). This work was supported in part by the National Natural Science Foundation of China under Nos. 61102043, 61262084, 81120108012 and the Basic Research Program of Shenzhen JC201104220219A.

where D_1 and D_2 are assumed to be two mutually incoherent dictionaries. R_i denotes the operator that extracts the i -th patch $R_i u_1$ from image u_1 . Inspired by the success of dictionary learning (DL) in sparse representation of signal/image [12], Peyre et al. [10] presented an adaptive MCA scheme by learning the morphologies of image layers. They used both adaptive local dictionaries and fixed global transforms (e.g. wavelet, curvelet) for separating image from a single mixture. Li et al. [11] further developed an algorithm to adaptively learn both cartoon and textural dictionaries from the data. Examples empirically indicated that the method alleviated some deficiencies caused by using fixed dictionaries (e.g. capturing complex texture patterns).

A main deficiency of these models in refs. [10, 11] is that two dictionaries corresponding to cartoon and texture parts may exhibit coherence, i.e. the similar atoms existed in both dictionaries. Therefore, a proper user-defined initialization is usually needed to drive the algorithm to some desired solutions. Motivated by the success of sparse representation with structured incoherence for image classification and clustering [13, 14], we propose a structural incoherence based dictionary learning (SIDL) algorithm to alleviate this drawback for image decomposition. By promoting the incoherence between the cartoon and texture associated dictionaries, the resulting decomposition components would be as independent as possible, thereby high quality decomposition performance will be achieved.

2. ALGORITHM SIDL

Consider that both $D_1 \in \Upsilon_1$ and $D_2 \in \Upsilon_2$ are adaptively learned to represent two different components u_1 , u_2 under the sparse L_0-norm constraint [10, 11], the objective function is as follows:

$$\begin{Bmatrix} \hat{u}_1, \hat{D}_1, \hat{\alpha}_i \\ \hat{u}_2, \hat{D}_2, \hat{\beta}_i \end{Bmatrix} = \underset{\substack{u_1, D_1, \alpha_i \\ u_2, D_2, \beta_i}}{\arg\min} \begin{bmatrix} \lambda \|f - u_1 - u_2\|_2^2 + \sum_{i \in I} \|R_i u_1 - D_1 \alpha_i\|_2^2 + \mu \|\alpha_i\|_0 \\ + \sum_{i \in I} \|R_i u_2 - D_2 \beta_i\|_2^2 + \eta \|\beta_i\|_0 \end{bmatrix} \quad (3)$$
$$s.t. \quad D_1 \in \Upsilon_1, \quad D_2 \in \Upsilon_2$$

where Υ_1 and Υ_2 are two incoherence dictionary sets.

As analyzed in ref. [10, 11], the penalty function in Eq. (3) is non-convex jointly in all its arguments. The selections of initialization and optimization strategy are very important to achieve satisfying results. It is worth noting that, most of the existing methods use empirically proper initial dictionaries to approximately satisfy the constrained incoherence conditions in Eq. (3) [10, 11].

2.1. Dictionary Learning with Structured Incoherence

Recently, forcing structured incoherence between different dictionaries for classification and clustering has gained promising performance. An incoherence-promoting term encourages dictionaries associated to different classes to be as independent as possible, meanwhile allowing for different classes to share features. Since the cartoon and texture parts in Eq. (3) can be seen as two different classes, we borrow the idea from ref. [13] and add a regularization term to the objective function in Eq. (3) to enforce the incoherence between the two dictionaries, which yields

$$\begin{Bmatrix} \hat{u}_1, \hat{D}_1, \hat{\alpha}_i \\ \hat{u}_2, \hat{D}_2, \hat{\beta}_i \end{Bmatrix} = \underset{\substack{u_1, D_1, \alpha_i \\ u_2, D_2, \beta_i}}{\arg\min} \begin{bmatrix} \lambda \|f - u_1 - u_2\|_2^2 + \sum_{i \in I} \|R_i u_1 - D_1 \alpha_i\|_2^2 + \mu \|\alpha_i\|_0 \\ + \zeta \|D_2^T D_1\|_F^2 + \sum_{i \in I} \|R_i u_2 - D_2 \beta_i\|_2^2 + \eta \|\beta_i\|_0 \end{bmatrix} \quad (4)$$

where the regularization term for structural incoherence $\|D_2^T D_1\|_F^2$ sums up the Frobenius norms between pair of geometrical dictionary D_1 and textural dictionary D_2 . The associated incoherence weight parameter ζ balances the sparse approximation and dictionary incoherence.

It can be easily seen from Eq. (4) that, by incorporating the structured-incoherence induced regularizer, the cartoon and texture components associated to the two incoherent dictionaries will be more properly separated. i.e. the energy in Eq. (3) will lead to the learning of dictionary optimized to properly represent the corresponding component, meanwhile the new added term will ensure the coherence to be weak for each other.

2.2. Algorithm Outline

The minimization problem in Eq. (4) is highly non-convex. Usually iterative block relaxation coordinate descent minimization scheme was employed to address it. On one hand, under some approximation, the pursuit of dictionaries and coefficients can be achieved as follows:

$$\underset{\substack{D_1, \alpha_i \\ D_2, \beta_i}}{Min} \left[\sum_{i \in I} \left\| R_i f - [D_1, D_2] \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \right\|_2^2 + \zeta \|D_2^T D_1\|_F^2 + \mu \|\alpha_i\|_0 + \eta \|\beta_i\|_0 \right] \quad (5)$$

A two-step alternative procedure is used to solve Eq. (5): the sparse coding stage and the dictionary updating stage. The formal stage can be solved by the usual methods as that in [12, 15]. When updating the dictionaries D_1 , D_2 , the incoherence regularization term provides the coupling between them, and an efficient method will be proposed to address it in the next subsection.

On the other hand, once the sparse approximations of the image patches are obtained, image reconstruction can be conducted respectively for different morphological contents as following:

$$\underset{u_1, u_2}{Min} \left[\lambda \|f - u_1 - u_2\|_2^2 + \sum_{i \in I} \|R_i u_1 - D_1 \alpha_i\|_2^2 + \sum_{i \in I} \|R_i u_2 - D_2 \beta_i\|_2^2 \right] \quad (6)$$

Clearly, the minimization with respect to variable u_1 and u_2 can be attained by least square solution [10, 11].

The SIDL method is summarized as follows:

Algorithm SIDL: Structural incoherence based dictionary learning

1: For $n=0$ to $N-1$ do

2: Sparse coding and update dictionary D_2

$$\{D_2^{(n+1)},\alpha_i^{(n+1/2)},\beta_i^{(n+1/2)}\} = \arg\min_{D_2,\alpha_i,\beta_i}\sum_{i\in I}\left\|R_i f-[D_1^{(n)},D_2]\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}\right\|_2^2+\zeta\left\|D_2^T D_1^{(n)}\right\|_F^2+\mu\|\alpha_i\|_0+\eta\|\beta_i\|_0$$

3: $$u_2^{(n+1)}=\left(\lambda I+\sum_{i\in I}R_i^T R_i\right)^{-1}\left(\lambda(f-u_1^{(n)})+\sum_{i\in I}R_i^T D_2^{(n+1)}\beta_i^{(n+1/2)}\right)$$

4: Sparse coding and update dictionary D_1

$$\{D_1^{(n+1)},\alpha_i^{(n+1)},\beta_i^{(n+1)}\} = \arg\min_{D_1,\alpha_i,\beta_i}\sum_{i\in I}\left\|R_i f-[D_1,D_2^{(n+1)}]\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}\right\|_2^2+\zeta\left\|D_1^T D_2^{(n+1)}\right\|_F^2+\mu\|\alpha_i\|_0+\eta\|\beta_i\|_0$$

5: $$u_1^{(n+1)}=\left(\lambda I+\sum_{i\in I}R_i^T R_i\right)^{-1}\left(\lambda(f-u_2^{(n+1)})+\sum_{i\in I}R_i^T D_1^{(n+1)}\alpha_i^{(n+1)}\right)$$

6: End (For)

2.3. Dictionary Update

For the minimization problem in Eq. (5), we adopt a relaxation strategy which alternatively updates D_1 and D_2. In the following, we will take the updating of D_1 for instance. By calculating the objective function Eq. (5) with respect to D_1, it yields

$$\arg\min_{D_1}\sum_{i\in I}\left\|R_i f-[D_1,D_2]\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}\right\|_2^2+\zeta\left\|D_2^T D_1\right\|_F^2 \quad (7)$$

Let $x_i=R_i f-D_2\beta_i$, then it yields

$$\sum_{i\in I}\left\|R_i f-[D_1,D_2]\begin{bmatrix}\alpha_i\\ \beta_i\end{bmatrix}\right\|_2^2=\sum_{i\in I}\|x_i-D_1\alpha_i\|_2^2=\|X-D_1\Lambda\|_F^2 \quad (8)$$

For convenience, we rewrite matrix Λ as $\Lambda=[\gamma_1;\gamma_2;\cdots;\gamma_K]$, where $\gamma_k, k=1,2,\cdots,K$ is the row vector of Λ. The atoms of $D_1=[d_1,d_2,\cdots,d_K]$ are updated sequentially at a time to minimize the fidelity term. When updating d_k, all the other columns $d_l, l\neq k$ are fixed. Then the minimization of Eq. (8) with respect to d_k is reduced to

$$\arg\min_{d_k}\left\|X-\sum_{l\neq k}d_l\gamma_l-d_k\gamma_k\right\|_F^2+\zeta\left\|D_2^T d_k\right\|_2^2 \quad (9)$$

Majorization-minimization (MM) technique is employed to add an additional proximal-like penalty at each inner step, to cancel out the term $\left\|D_2^T d_k\right\|_2^2$ (for more details of MM technique, please refer to [16, 17, 18]). Assuming that index m denotes the iterative number of the inner iteration of MM, the optimal of Eq. (9) can be approximately found by iteratively solving the following sub-problem.

$$\arg\min_{d_k}\left\|X-\sum_{l\neq k}d_l\gamma_l-d_k\gamma_k\right\|_F^2+\zeta\left(\left\|D_2^T d_k\right\|_2^2+(d_k-d_k^m)^T(\kappa I-D_2D_2^T)(d_k-d_k^m)\right) \quad (10)$$
$$s.t.\ \ d_k^T d_k=1$$

where $\kappa\geq eig(D_2D_2^T)$.

Let $Y=X-\sum_{l\neq k}d_l\gamma_l$, Eq. (10) can be written as

$$\arg\min_{d_k}\|Y-d_k\gamma_k\|_F^2+\zeta\left(\left\|D_2^T d_k\right\|_2^2+(d_k-d_k^m)^T(\kappa I-D_2D_2^T)(d_k-d_k^m)\right)+\omega(d_k^T d_k-1) \quad (11)$$

Differentiating it with respect to d_k, and let it be 0, we have

$$d_k\gamma_k\gamma_k^T+\omega d_k-Y\gamma_k^T+\zeta(\kappa d_k-(\kappa I-D_2D_2^T)d_k^m)=0$$
$$d_k=[Y\gamma_k^T+\zeta(\kappa I-D_2D_2^T)d_k^m]/[\gamma_k\gamma_k^T+\zeta\kappa+\omega]$$

Thus the solution under constrain $d_k^T d_k=1$ is

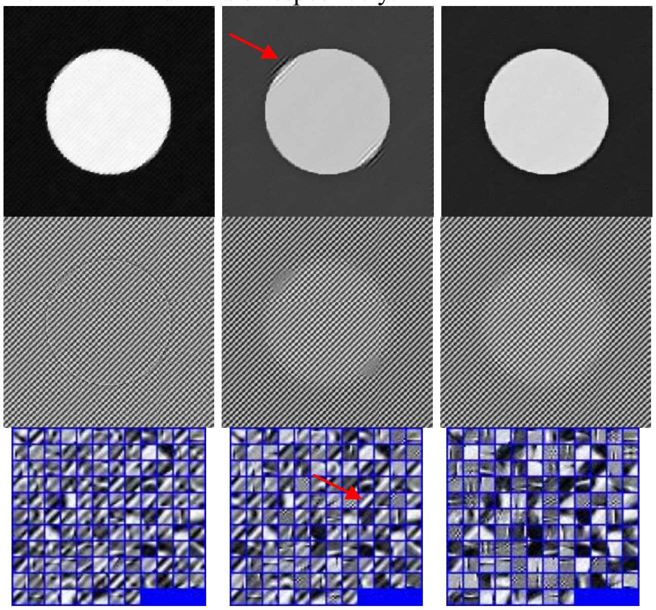
$$d_k=\frac{[Y\gamma_k^T+\zeta(\kappa I-D_2D_2^T)d_k^m]}{\left\|[Y\gamma_k^T+\zeta(\kappa I-D_2D_2^T)d_k^m]\right\|_2} \quad (12)$$

where $\|\bullet\|_2$ is the L^2-norm.

3. EXPERIMENTAL RESULTS

The performance of SIDL was evaluated on several experiments, where the proposed method was initialized by the decomposition components obtained by OSV method [7]. All the test images are normalized to have a maximum magnitude of 1 and Gaussian noise with $\sigma=0.02$ is added to the reference images. It is worth noting that DL method is the special instance of the SIDL with $\zeta=0$. The parameter settings of the sparse coding stage in both DL and SIDL follow the default values in refs. [12, 15].

Fig. 1 displays the decomposition for a synthesis image. The results by the OSV method (i.e. the initialization for SIDL) are given in Fig. 1(a), where we deliberately set the regularization parameter to be a bigger value($\lambda=0.2$ in Eq. (1)), such that some texture details still exist in cartoon part. Since the initialized cartoon part (subsequently the dictionary) contains few texture contents, the traditional DL method exhibits some artifacts, indicated by the red arrow in Fig. 1(b). With our proposed SIDL, the sparse representation that takes into account the dictionary incoherence achieves the best performance. As a result, commonly shared features across different components are suppressed, while the independent ones are preserved. As seen from the third line of Fig. 1, there are some oriental textures prototypes/atoms remain in the resulting cartoon-component dictionary by DL, while almost no such elements in that with SIDL algorithm. Additionally, the values of structural incoherence $\left\| D_2^T D_1 \right\|_F^2$ of methods DL and SIDL are 1637.16 and 246.97 respectively.



(a) (b) (c)

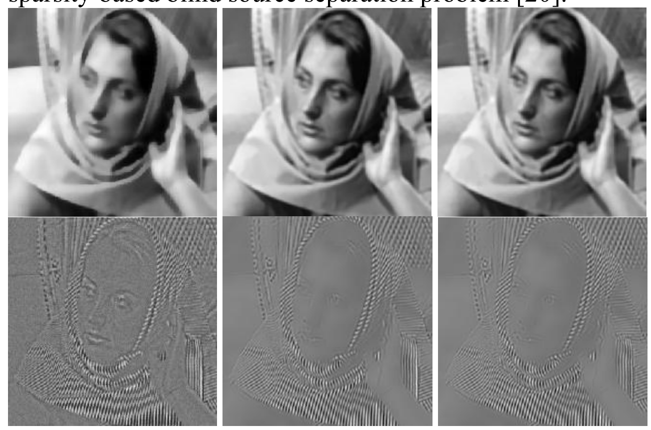
Figure 1. The cartoon-texture decomposition obtained by method (a) OSV, (b) DL and (c) SIDL with $\zeta = 0.1$.

The effect of imposing dictionary incoherence is also investigated on real-life image "Barb", which exhibits subtle cartoon parts and complex texture details. Figs. 2 and 3 display the decomposition results for the three methods, where the OSV estimations with $\lambda = 0.2$ and $\lambda = 1$ are chosen as the initial image respectively. As expected, when the initial dictionary is properly provided, the decompositions in DL and SIDL as shown in Fig. 2 do not show much difference in visual concept. However, as shown in Fig. 3 that when an inappropriate initial dictionary is chosen, promoting dictionary incoherence actually improves the decomposition. In Fig. 3(c), the geometric dictionary of SIDL algorithm describes the cartoon content better.

The angle derivation error (ADE) which is based on the orthogonality of two image partitions [19], was used as the criteria to measure the independence in image decomposition. Fig. 4 plots the curve of the ADE versus iteration n when different ζ in SIDL were used. It can be observed that encouraging incoherence between dictionaries during the iterative procedure really improves the performance. Additionally, $\zeta = 0.1$ is usually a good choice.

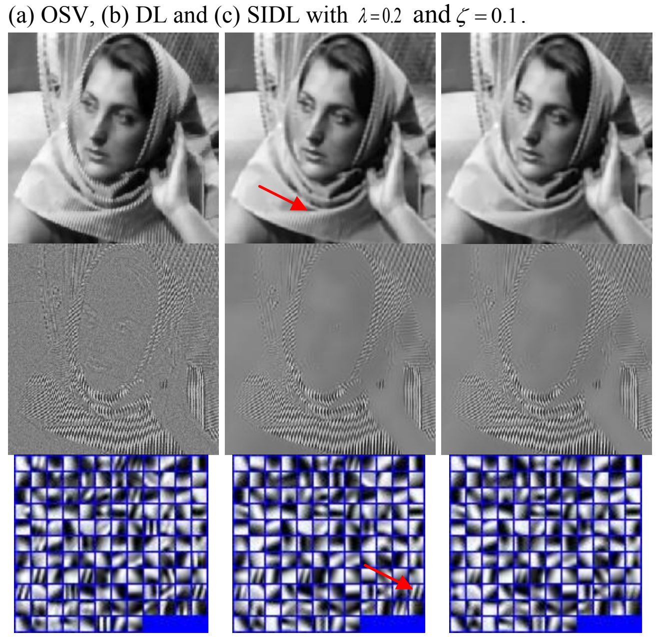
4. CONCLUSION

In this paper, an adaptive dictionary learning algorithm with structural incoherence for image decomposition was proposed. The decomposition results were improved by forcing the structured incoherence between the cartoon and textural dictionaries. Experimental results demonstrate the effectiveness and robustness of the proposed algorithm, in terms of visual concepts and quantitative measures. Further studies will focus on extending the SIDL to address the sparsity-based blind source separation problem [20].



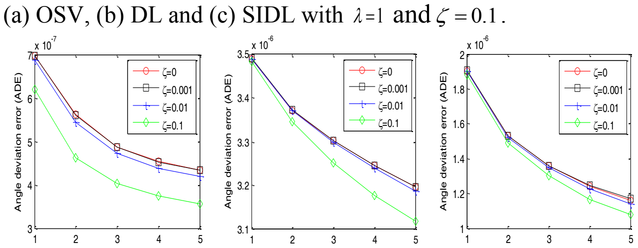
(a) (b) (c)

Figure 2. The cartoon-texture decomposition obtained by method (a) OSV, (b) DL and (c) SIDL with $\lambda = 0.2$ and $\zeta = 0.1$.



(a) (b) (c)

Figure 3. The cartoon-texture decomposition obtained by method (a) OSV, (b) DL and (c) SIDL with $\lambda = 1$ and $\zeta = 0.1$.



(a) (b) (c)

Figure 4. ADE versus iteration n of SIDL with $\zeta = 0, 0.001, 0.01, 0.1$. The initial image of (a) is in Fig. 1, (b) is in Fig. 2 and (c) is in Fig. 3. Noting that DL method is equal to SIDL with $\zeta = 0$.

5. REFERENCES

[1] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. on Image Processing*, vol. 12, no. 8, pp. 882–889, 2003.

[2] T. Saito, Y. Ishii, H. Aizawa, D. Yamada, and T. Komatsu, "Image processing approach via nonlinear image decomposition for a digital color camera," in *Proc. ICIP*, pp. 905–908, 2008.

[3] D. Paquin, D. Levy, E. Schreibmann, and L. Xing, "Multiscale image registration," *Math. Biosci. Eng.*, vol. 3, pp. 389–418, 2006.

[4] Y. Meyer, "Oscillating patterns in image processing and nonlinear evolution equations," Volume 22 of University Lecture Series. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[5] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D.*, vol. 60, pp. 259–268, 1992.

[6] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, vol. 19, no. 1, pp. 553–572, 2003.

[7] S.J. Osher, A. Sole, and L.A. Vese, "Image decomposition and restoration using total variation minimization and the H-1 norm," *SIAM Multiscale Modeling and Simulation*, vol. 1, no. 3, pp. 349-370, 2003.

[8] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[9] Shoham, N., Elad M., "Alternating KSVD denoising for texture separation," in: *The IEEE 25th Convention of Electrical and Electronics Engineers* in Israel, Eilat, Israel, vol. 3–5. 2008.

[10] G. Peyre, J. Fadili, and J. Starck, "Learning the morphological diversity," *SIAM Journal of Imaging Science*, vol. 3, no. 3, pp. 646–669, 2010.

[11] Yafeng Li, Xiangchu Feng, "Image decomposition via learning the morphological diversity," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 111–120, 2012.

[12] Elad, M., Aharon, M., "Image denoising via sparse and redundant representations over learned dictionaries,'' *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[13] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. CVPR*, pp. 3501–3508, 2010.

[14] Chih-Fan Chen, Chia-Po Wei, Wang, Y.-C.F, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. CVPR*, pp. 2618–2625, 2012.

[15] R Rubinstein, M Zibulevsky, M Elad, Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. Technical Report, CS Technion, 2008.

[16] D Hunter, K Lange, "A tutorial on MM algorithms," *Am Statist*, vol. 58, pp. 30–37, 2004.

[17] I Daubechies, M De Friese, C De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun Pure Appl Math*, vol. 57, pp. 3601–3608, 2004.

[18] J Oliveira, J Bioucas-Dias, MAT Figueiredo, "Adaptive total variation image deblurring: a majorization-minimization approach," *Signal Process*, vol. 89, no. 9, pp. 1683–1693, 2009.

[19] D. Szolgay and T. Sziranyi, "Adaptive image decomposition into cartoon and texture parts optimized by the orthogonality criterion," *IEEE Trans. on Image Processing*, vol. 21, no. 8, pp. 3405–3415, 2012.

[20] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Blind separation of image sources via adaptive dictionary learning," *IEEE Trans. on Image Processing*, vol. 21, no. 6, pp. 2921–2930, 2012.