ALGORITHMS FOR SIGNAL SEPARATION EXPLOITING SPARSE REPRESENTATIONS, WITH APPLICATION TO TEXTURE IMAGE SEPARATION

Neta Shoham and Michael Elad *

The Computer Science Department – The Technion, Haifa 32000, Israel
email: [shohamn,elad]@tx.technion.ac.il

ABSTRACT

We consider the problem of signal separation, where the observation $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{v}$ is composed of the signals $\mathbf{x}_1$ and $\mathbf{x}_2$ to be separated, along with additive noise $\mathbf{v}$. We further assume that $\mathbf{x}_1$ and $\mathbf{x}_2$ have sparse representations with respect to two different dictionaries $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. Exploiting sparse representations for signal separation has been introduced recently under the name *Morphological Component Analysis* (MCA).

In this work we consider several greedy MCA algorithms. First, we show that the alternated denoising, practiced extensively in past work on signal separation, is inferior to a far simpler direct approach. Secondly, we introduce the Minimum-Mean-Squared-Error (MMSE) estimator into the separation algorithm, in the form of a randomized average of several representations, and show the benefit it provides. Turning to the task of separating a mixture of texture images, we fuse the above-mentioned direct greedy algorithm as a local operation in a global system, and demonstrate the successful separation it leads to.

***Index Terms*—** Sparse representation, learned dictionary, separation, denoising, Maximum A-posteriori Probability (MAP), Minimum mean squared error (MMSE).

1. INTRODUCTION

A classic problem in signal processing is the recovery of a signal $\mathbf{x} \in \mathbb{R}^n$ from its noisy observation $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where $\mathbf{v}$ is an additive white Gaussian noise with known variance σ_v^2. Different techniques assume different priors on $\mathbf{x}$ in order to handle the denoising problem. Recent work on signal modeling promote a sparsity prior on $\mathbf{x}$, meaning that it is assumed to have a sparse representation $\alpha \in \mathbb{R}^k$ over some redundant dictionary $\mathbf{D} \in \mathbb{R}^{[n \times k]}$, with $n \le k$ [1]. Denoising is achieved by solving

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|\mathbf{D}\alpha - \mathbf{y}\|_2^2 \le n\sigma_v^2. \tag{1}$$

The error bound $n\sigma_v^2$ is a direct consequence of the knowledge of the noise power. The notation $\|\alpha\|_0$ is the ℓ_0-norm, counting the number of non-zeros in the vector α. Thus, among all feasible solutions satisfying $\|\mathbf{D}\alpha - y\|_2^2 \le n\sigma_v^2$, we seek the sparesest. Solving (1) is impossible as this is a combinatorial task of exponential complexity with k, and thus approximation methods, such greedy [2] and relaxation [3] methods are employed in practice.

In this paper we consider the more complicated and extended problem of signal separation, where the observation $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{v}$ is composed of two signals, $\mathbf{x}_1$ and $\mathbf{x}_2$, to be separated, along with additive noise $\mathbf{v}$. As before, we assume that $\mathbf{x}_1$ and $\mathbf{x}_2$ have sparse representations with respect to two known and different dictionaries $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. Exploiting sparse representations for signal separation has been introduced recently under the name *Morphological Component Analysis* (MCA) [4]. Following the above rationale, separation can be achieved by finding the two sparsest representations $\hat{\alpha}_1$ and $\hat{\alpha}_2$ that can explain $\mathbf{y}$. Put formally, this reads

$$\begin{aligned} \{\hat{\alpha}_1, \hat{\alpha}_2\} \quad = \quad & \arg\min_{\alpha_1,\alpha_2} \|\alpha_1\|_0 + \|\alpha_2\|_0 \\ & \text{s.t.} \quad \|\mathbf{D}_1\alpha_1 + \mathbf{D}_2\alpha_2 - \mathbf{y}\|_2^2 \le n\sigma_v^2. \end{aligned} \tag{2}$$

Again, as the above is impossible to solve directly, approximation methods are called for. The work reported in [4] considered an image separation, where the dimension of the portions to be separated are very high, and this was handled using an iterated shrinkage technique. A later work, reported in [5], proposed a mixture of the above with Orthogonal Matching Pursuit (OMP), such that local patch-based operations are using the OMP, while the global part still employs iterated shrinkage.

In this work we return to the basics of MCA, assuming at first a low-dimensional problem, and exploring several greedy signal separation algorithms based on the OMP. The first junction we cross in this quest is whether to use a direct approximate solution to the problem posed in Equation (2), or an iterated one. Past work on MCA [4, 5] recommended the iterated path. In this method, the solution of (2) is obtained by considering an *alternated denoising* problem, where $\hat{\alpha}_1$ is estimated as the denoising of the signal $\mathbf{y} + \mathbf{x}_2$, while $\hat{\alpha}_2$

*This research was supported by the Israel Science Foundation grant No. 796/05 and by the United States Israel Binational Science Foundation grant No. 2004199.

is obtained when denoising $\mathbf{y} + \mathbf{x}_1$, where in both cases we solve a problem of the structure posed in Equation (1). We show that while this iterated method is far more complicated than the direct solver, it provides hardly any benefit over the simpler direct method in terms of the separation quality.

A second junction crossed is the actual estimator to use in solving either (1) or (2). The classic work that promotes a quest for the sparsest representation can be shown to align with the MAP estimator [1, 6]. Recent work show that a far better estimation result can be obtained by averaging several randomly generated sparse representations that approximate the solution of the above problems [6]. This average approximates the MMSE estimator, explaining the improved performance. We harness this idea into the separation task, combined with both the direct and the alternate denoising algorithms, and we show the benefit it brings for both.

When turning to handle a mixture of texture images, one must start with the choice of dictionaries. While [4] assumed that those are known and fixed as global dictionaries (curvelet and overlapping-DCT), this is hardly the case with general context texture images. In [5], learned dictionaries based on the K-SVD are used for the texture portion of the images [7]. A second consideration is whether to operate globally or locally. When using trained dictionaries, those are necessarily treating small images patches. Following [7], the work reported in [5] fuses the local processing pieces into a global system. We follow the same route and deploy the various local algorithms presented above in this setting, demonstrating their performance on the image separation task. In particular, we show that a simple local and direct separation algorithm performs very well, being competitive in quality to the alternated denoising, while requiring much less computations.

The rest of this paper is organized as follows. In the next section we discuss the direct and the alternated denoising modes of separation. In Section 3 we expand on the MAP and the MMSE sparse coding options and their use for separation. Section 4 is dedicated to the modifications required for handling image mixtures. We conclude in Section 5 with a summary of this work and its contribution.

2. DIRECT VERSUS ALTERNATED DENOISING MCA ALGORITHMS

2.1. The Direct Algorithm

In the MCA original work [4], it was shown that the separation task as posed in Equation (2) can be reformulated with a concatenated dictionary $\mathbf{D}_T = [\mathbf{D}_1, \mathbf{D}_2]$ and a concatenated representation $\alpha_T^T = [\alpha_1^T, \alpha_2^T]$. Since $\|\alpha_T\|_0 = \|\alpha_1\|_0 + \|\alpha_2\|_0$, (2) converts with these definitions to

$$\{\hat{\alpha}_T\} = \arg\min_{\alpha_T} \|\alpha_T\|_0 \text{ s.t. } \|\mathbf{D}_T\alpha_T - \mathbf{y}\|_2^2 \leq n\sigma_v^2. \quad (3)$$

Clearly, this problem has the same structure as the denoising problem posed in Equation (1), and thus all the knowledge we have on solving (1) is relevant here as well. In particular, OMP can be used to approximate the solution of this problem by greedily accumulating the atoms of the solution such that the error $\|\mathbf{D}_T\alpha_T - y\|_2^2$ is reduced maximally at each stage.

Once the two sparse representations have been found, a proposed separation is readily available by $\hat{\mathbf{x}}_1 = \mathbf{D}_1\hat{\alpha}_1$ and $\hat{\mathbf{x}}_2 = \mathbf{D}_2\hat{\alpha}_2$. This is the direct approach, and its simplicity makes it appealing.

2.2. The Alternated Denoising Algorithm

The original work on MCA [4] did not use the direct method mentioned above, and preferred instead an iterative algorithm that peels the signals from $\mathbf{y}$. One possible reason for this choice of algorithm is the high-dimensions of the original image separation problem treated, that made greedy methods impractical, thus requiring alternative numerical scheme. Followup work in [5] kept this iterative paradigm, even though some of the operations in their method use OMP on low-dimensional signals. We refer to this iterative algorithm as the *Alternated Denoising*, and we now turn to describe this method.

Referring to Equation (2), suppose that at the i-th iteration we have a candidate solution for the representations, denoted as $\{\hat{\alpha}_1^i, \hat{\alpha}_2^i\}$. We can update these two by the following tow concurrent problems

$$\hat{\alpha}_1^{i+1} = \arg\min_{\alpha} \|\alpha\|_0 \ \text{ s.t. } \left\|\mathbf{D}_1\alpha - \left(\mathbf{y} - \mathbf{D}_2\hat{\alpha}_2^i\right)\right\|_2^2 \leq T$$
$$\hat{\alpha}_2^{i+1} = \arg\min_{\alpha} \|\alpha\|_0 \ \text{ s.t. } \left\|\mathbf{D}_2\alpha - \left(\mathbf{y} - \mathbf{D}_1\hat{\alpha}_1^i\right)\right\|_2^2 \leq T.$$

Both these problems are posed as denoising of residual signals (e.g., in the first problem, this signal is $\tilde{\mathbf{y}}_1^i = \mathbf{y} - \mathbf{D}_2\hat{\alpha}_2^i$). This process is initialized with $\{\hat{\alpha}_1^0, \hat{\alpha}_2^0\} = \{0, 0\}$, implying the first iteration is not likely to provide good result, and thus these denoising steps are to be performed repeatedly.

There are two important improvements over the above algorithm. First, rather than fixing $T = n\sigma_v^2$, it is recommended in [4, 5] to set it higher at first, and decrease it as the iterative process proceeds. This makes sense, since at first the "noise" in $\tilde{\mathbf{y}}_1^i$ is higher due to the presence of portions of the signal $\mathbf{x}_2$ in it. The second improvement is a relaxation that can be added to the algorithm in the form

$$\tilde{\mathbf{y}}_1^i = A \cdot \left(\mathbf{y} - \mathbf{D}_2\hat{\alpha}_2^i\right) + (1 - A)\mathbf{D}_1\hat{\alpha}_1^i, \quad (4)$$

with A in the range $[0, 1]$.

As can be clearly seen, this algorithm is far more complicated compared to the direct algorithm. For an overall of J iterations, this algorithm applies $2J$ denoising steps with k unknown in each, whereas the direct method applies only one denoising with $2k$ unknowns. The alternated denoising requires various parameter settings (decrease of the threshold, relaxation parameter), while the direct approach has hardly any.

2.3. Performance Comparison

In order to test the above two methods, we formed the following synthetic experiment. Two dictionaries $\mathbf{D}_1, \mathbf{D}_2$ of size 30×60 were created with iid normal entries, and the columns of these matrices were ℓ_2-normalized. Two sets of $N = 500$ test signals, $\{\mathbf{x}_1^j, \mathbf{x}_2^j\}_{j=1}^N$, were created, each being a random linear combination of 3 atoms from its corresponding dictionary. These signals were ℓ_2-normalized as well. We created 500 mixtures $\mathbf{y}^j = \mathbf{x}_1^j + \mathbf{x}_2^j + \sigma_v \mathbf{v}$, where $\mathbf{v}$ is a normalized random Gaussian vector with iid entries, and σ_v is varied in the range $[0.01, 1]$.

We tested the direct method with one small modification - the error threshold in (3) was set to $(1 + \sigma_v/6)n\sigma_v^2$, so as to increase the threshold for higher noise levels, when the constraint becomes less reliable. We also tested the alternated denoising method with $J = 8$ iterations, and with $A = 0.8$ (which was found empirically to give the best results). The initial threshold was set to be $T = n(1 + \sigma_v^2)$, and then decreased linearly towards $n\sigma_v^2$. Figure 1 shows the separation error as a function of σ_v for both methods. The separation error is defined as

$$\frac{1}{2N}\left[\sum_{j=1}^{N}\left\|\mathbf{x}_1^j - \hat{\mathbf{x}}_1^j\right\|^2 + \sum_{j=1}^{N}\left\|\mathbf{x}_2^j - \hat{\mathbf{x}}_2^j\right\|^2\right]. \quad (5)$$

As can be seen, the two methods perform comparably. However, the direct method is one order of magnitude faster.

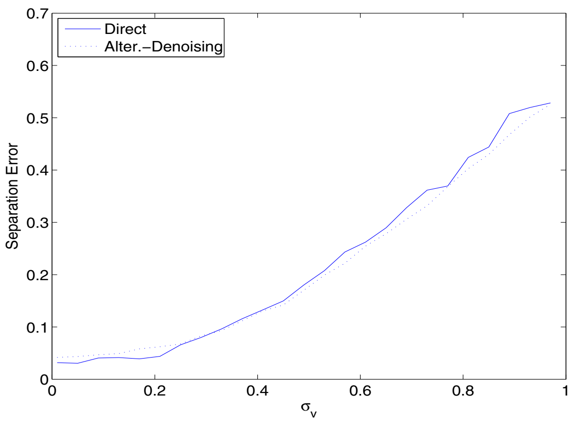

Fig. 1. The separation error for the direct and the alternated-denoising algorithms, as a function of the additive noise power.

3. MAP VERSUS MMSE MCA ALGORITHMS

3.1. MMSE Instead of MAP

Assuming that the direct method is to be used for the separation task, it is well known that solving (3) amounts to a MAP estimation of the representation. Can we do better? Based on [6], the answer is positive. Since we evaluate the performance of the separation result using the mean-squared-error quality measure, it would be better to use the MMSE estimator. While this is impossible to obtain in general, the work reported in [6] suggests a simple approximation of it based on a randomization of the OMP (termed *Random-OMP*) that leads to a set of (possibly) different representations. Their plain mean of these representations is shown to be a very good approximation of the MMSE estimator. Clearly, this idea can be adopted here quite easily, with the hope to obtain better separation results.

We must note that using the MMSE estimator as in [6] for the separation task does not necessarily mean that the separation MSE is the best it could be. The reason is that in the MMSE estimator used, we aim to minimize the expected error $\|\mathbf{x}_1 + \mathbf{x}_2 - \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2\|^2$, whereas our true desire is the minimization of the expectation of $\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\|^2 + \|\mathbf{x}_2 - \hat{\mathbf{x}}_2\|^2$, and the two might differ in general. We leave this matter for future work, and use here the plain MMSE estimator available to us.

3.2. Performance Comparison

Proceeding with the experiment described in Section 2.3, we simulated a third separation algorithm based on the Random-OMP. We used $J = 8$ iterations, so as to align its complexity with the alternated-denoising algorithm.

When using the Random-OMP, one parameter must be fixed - this is the factor c with which the projections are scaled in the randomization process [6]. Since in this work we have deviated form the model developed in [6], in this work we fix this heuristically to be $c = 100$ for all tests, with the understanding that better tuning of it may lead to better results.

Figure 2 presents the Random-OMP results on the same results shown in the previous section. The MMSE approximation improves over the direct method separation by 1.4dB on average.

4. TREATING IMAGES

4.1. Going to High-Dimensions

When turning to handle signals $\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ that are of high dimensions, the OMP is no longer appealing, or even possible to use. Suppose that $n = 100,000$ and the dictionaries $\mathbf{D}_1, \mathbf{D}_2$ are of size $10^5 \times 10^7$. Note that for these sizes, the dictionary is never stored nor used explicitly, as it stands for a fast transform, such as wavelet, Fourier, and the like. A representation of such signals that includes $5,000 - 10,000$ atoms is considered as very sparse.

The OMP requires an application of the inverse transform[1] for computing the projection of the signal on the atoms.

[1] Actually, it is the adjoint that is required, but for tight frames the two are the same.

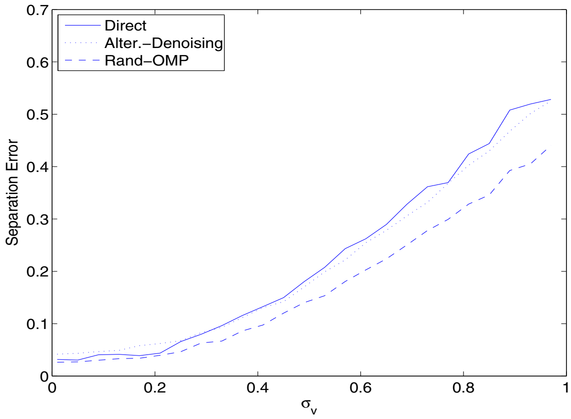

Fig. 2. The separation error for the direct, the alternated-denoising, and the Random-OMP algorithms, as a function of the additive noise power.

Taking the largest projection, we need then to solve a Least-Squares (LS) problem over the accumulated atoms. Accumulating $5,000 - 10,000$ atoms one by one is tedious and non-efficient. Furthermore, the above-mentioned LS problems can be either solved by accumulating the chosen atoms explicitly, or by using multiplications by the complete dictionary and its adjoint and masking the non-chosen atoms. Both options are very expensive, to the point of making the OMP impossible to use. This is why prior art on MCA that worked on images avoided the OMP on the global transforms [4, 5].

Can we handle high-dimensional signals and still enjoy the simplicity of the OMP? The answer is positive, if we adopt the approach taken in [7]. The idea is to operate on small image patches locally and fuse the results to a global solution. A side benefit to this local processing is the ability to depart from using pre-assigned global dictionaries, and replace those with trained ones. Indeed, learning sparsifying dictionaries for the signals to be separated using an algorithm such as the K-SVD is relevant only for low-dimensional signals, and this stands as yet another incentive for going local [7].

We consider here the problem of separating a mixture of texture images, where high-dimensions for the involved signals is a natural feature. In [5], a separation of an image to texture and cartoon layers is suggested. A learned dictionary based on the K-SVD is used for the texture portion of the image, and the operations on this part are performed locally. As the cartoon part is modeled globally, their overall algorithm still uses an iterated solver, and the overall structure of their algorithm is the alternated denoising, which requires many applications of the two sparse coding pieces.

4.2. The Proposed Algorithm

In contrast to the method in [5], we aim to propose a fully local separation algorithm that can use the direct OMP, and still gives competitive results to the alternated denoising scheme, with a small fraction of the original required complexity. We start by formulating the problem, following the denoising algorithm developed in [7]. We denote the full high-dimensional signals by capital letters, and remain with lower-case letters for the low-dimensional patches. The operator $\mathbf{R}_{pq}$ extracts a patch of a fixed and known size from the location $[p,q]$ in the image.

We form the following energy function $\epsilon(\mathbf{X}_1,\mathbf{X}_2)$, such that its minimizers $\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2$ define the separation proposed:

$$\begin{aligned}\epsilon(\mathbf{X}_1,\mathbf{X}_2) \quad = \quad & \lambda\,\|\mathbf{Y}-\mathbf{X}_1-\mathbf{X}_2\|^2 \\ & + \sum_{pq\in\Omega} \|\mathbf{R}_{pq}\mathbf{X}_1-\mathbf{D}_1\alpha_1^{pq}\|^2 \\ & + \sum_{pq\in\Omega} \|\mathbf{R}_{pq}\mathbf{X}_2-\mathbf{D}_2\alpha_2^{pq}\|^2 \\ & +\mu \sum_{pq\in\Omega} \left(\|\alpha_1^{pq}\|_0+\|\alpha_2^{pq}\|_0\right).\end{aligned} \tag{6}$$

The first term expresses our knowledge about the additive noise, the difference between $\mathbf{Y}$ and the combination $\mathbf{X}_1+\mathbf{X}_2$, being iid white and Gaussian. The second and the third terms express our belief that patches in locations $[p,q]$ extracted from the images $\mathbf{X}_1$ and $\mathbf{X}_2$ are expected to have sparse representations with respect to the two dictionaries $\mathbf{D}_1,\mathbf{D}_2$, respectively. The last term forces sparsity on the proposed representations. These representations are part of the unknowns, but since the objective is $\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2$, we do not list them as such in the definition of the energy function.

In order to solve the above minimization task, we propose a degenerated block-coordinate-descent algorithm, just like in [7]. We initialize the algorithm by $\mathbf{X}_1+\mathbf{X}_2=\mathbf{Y}$ and turn to approximate the sparse representations $\hat{\alpha}_{1,2}^{pq}$ of all the patches in the two images. Since we do not have initial values for $\mathbf{X}_1$ and $\mathbf{X}_2$, we use the following approximation:

$$\begin{aligned}&\|\mathbf{R}_{pq}\mathbf{X}_1-\mathbf{D}_1\alpha_1^{pq}\|^2+\|\mathbf{R}_{pq}\mathbf{X}_2-\mathbf{D}_2\alpha_2^{pq}\|^2 \\ &\quad\approx \|\mathbf{R}_{pq}(\mathbf{X}_1+\mathbf{X}_2)-\mathbf{D}_1\alpha_1^{pq}-\mathbf{D}_2\alpha_2^{pq}\|^2,\end{aligned} \tag{7}$$

which becomes correct if

$$(\mathbf{R}_{pq}\mathbf{X}_1-\mathbf{D}_1\alpha_1^{pq})^T(\mathbf{R}_{pq}\mathbf{X}_2-\mathbf{D}_2\alpha_2^{pq})=0, \tag{8}$$

i.e., when the separation errors of the two patches are uncorrelated. Thus, the problem obtained amounts to the very same problem as posed in Equation (3), applied on each patch in location $[p,q]$ separately,

$$\begin{aligned}\{\hat{\alpha}_1^{pq},\hat{\alpha}_2^{pq}\} \quad = \quad & \arg\min_{\alpha_1,\alpha_2} \|\alpha_1\|_0+\|\alpha_2\|_0 \\ & \text{s.t.} \quad \|\mathbf{D}_1\alpha_1+\mathbf{D}_2\alpha_2-\mathbf{R}_{pq}\mathbf{Y}\|_2^2\leq n\sigma_v^2.\end{aligned}$$

where we have used the fact that $\mathbf{R}_{pq}\mathbf{X}_1+\mathbf{R}_{pq}\mathbf{X}_2=\mathbf{R}_{pq}\mathbf{Y}$, and the knowledge about the noise power between $\mathbf{X}_1+\mathbf{X}_2$ and $\mathbf{Y}$. This solver can be replaced with the Random-OMP alternative. Nevertheless, if the patches are extracted with full overlap, the benefit is expected to be small, since the overlapping itself has an MMSE flavor to it.

Once we have all the sparse representations $\hat{\alpha}_1^{pq}$, $\hat{\alpha}_2^{pq}$ for all locations $[p, q]$, we minimize the energy in Equation (6) w.r.t. $\mathbf{X}_1$, and $\mathbf{X}_2$. The relevant parts for this minimization are only the ℓ_2 terms in Equation (6), and thus the solution is obtained easily as the linear equation of the form:

$$\begin{bmatrix} \lambda\mathbf{I} + \sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{R}_{pq} & \lambda\mathbf{I} \\ \lambda\mathbf{I} & \lambda\mathbf{I} + \sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{R}_{pq} \end{bmatrix} \cdot \qquad (9)$$

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \lambda\mathbf{Y} + \sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{D}_1\hat{\alpha}_1^{pq} \\ \lambda\mathbf{Y} + \sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{D}_2\hat{\alpha}_2^{pq} \end{bmatrix}.$$

For cyclic patch extraction, $\sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{R}_{pq} = c\mathbf{I}$, where c is the size of the patch. Thus the solution of (9) becomes

$$\mathbf{X}_1 = \frac{A}{c}(c\mathbf{Y} - \sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{D}_2\alpha_2^{pq}) + \frac{1-A}{c}\sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{D}_1\alpha_1^{pq}.$$

$$\mathbf{X}_2 = \frac{A}{c}(c\mathbf{Y} - \sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{D}_1\alpha_1^{pq}) + \frac{1-A}{c}\sum_{pq\in\Omega} \mathbf{R}_{pq}^T\mathbf{D}_2\alpha_2^{pq}.$$

for $A = \frac{\lambda}{2\lambda+c} \in [0, 0.5]$. Once the solution of the above linear system is computed, we conclude the separation algorithm. Thus, the complexity remains one joint OMP per patch, and an additional small overhead beyond that, making this algorithm very simple and fast.

4.3. Image Separation Results

We take two texture images of size 256×256 pixels and use the 128×128 left-upper corners for training. We extract 10×10 patches from the training regions and use the K-SVD [7] to get two trained dictionaries of size 100×200. We take the sum of the two lower corners, which are not used in the training stage, and use the trained dictionaries to separate the mixture. We perform this experiment twice, one time with the direct algorithm proposed above, and second time with the the alternated version as in [5]. We use a fixed number of atoms, 4, for each patch in both the training and the separation stages. The results are shown in Figure 3.

5. SUMMARY AND CONCLUSIONS

The knowledge that a signal can be represented sparsely with respect to a known dictionary is powerful and can serve its effective processing. In this paper we consider signal separation based on this model, following the work on MCA. We show that the long-practiced alternated denoising method can be replaced with a far simpler direct algorithm, with hardly any deterioration in performance. An MMSE version of the sparse coding method is shown to lead to benefit in separation results. The migration from the alternated-denoising to the direct methods is valid for low-dimensional signals for which

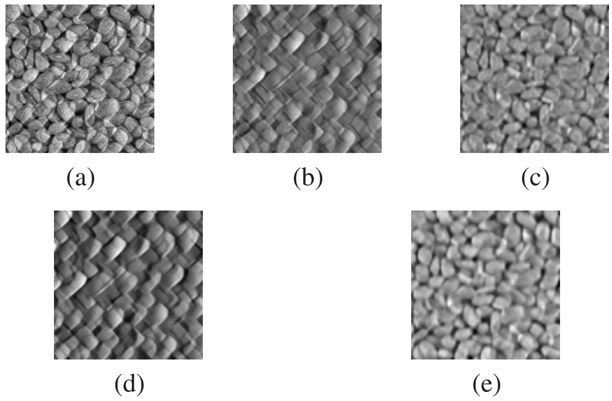

Fig. 3. (a) Mixed texture; (b,c) Separation results with the direct method (PSNR= 16.84dB); (d,e) Separation results with the alternated method (PSNR= 16.65dB).

OMP is efficient. We thus show how to keep this benefit when turning to handle images, by operating locally on patches, and fusing the results to a global separation algorithm.

6. REFERENCES

[1] A.M. Bruckstein, D.L. Donoho, and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, to appear in *SIAM Review*.

[2] S. Mallat and Z. Zhang, Matching pursuit in a time-frequency dictionary, *IEEE Transactions on Signal Processing*, Vol. 41, pp. 3397–3415, 1993.

[3] S.S. Chen, D.L. Donoho, and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM Review*, Volume 43, pp. 129–59, 2001.

[4] M. Elad, J-L. Starck, P. Querre, and D.L. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA), *Journal on Applied and Computational Harmonic Analysis*, Vol. 19, pp. 340-358, 2005.

[5] G. Peyre, J. Fadili, and J-L. Starck, Learning adapted dictionaries for geometry and texture separation, *Proceedings of SPIE Wavelet XII*, Augst 2007.

[6] M. Elad and I. Yavneh, A weighted average of sparse representations is better than the sparsest one alone, submitted to *IEEE Transactions on Information Theory*.

[7] M. Elad and M. Aharon, Image denoising via sparse and redundant representation over learned dictionaries, *IEEE Trans. on Image Processing*, Vol. 15, pp. 3736–3745, December 2006.