Robust joint reconstruction of misaligned images using semi-parametric dictionaries

Gilles Puy GILLES.PUY@EPFL.CH
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne

Pierre Vandergheynst PIERRE.VANDERGHEYNST@EPFL.CH
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne

Abstract

We propose a method for signal reconstruction in semi-parametric dictionaries. The proposed algorithm estimates both the signal decomposition and the intrinsic parameters of the dictionary during the reconstruction process. Theoretical results about the convergence of the algorithm are presented. The method is used here for joint reconstruction of misaligned images.

1. Introduction

Sparse representations are today very popular in signal processing to address problems such as, e.g., denoising, deconvolution, or signal reconstruction in compressed sensing. In the present work, we address the problem of signal reconstruction in semi-parametric dictionaries with unknown parameters. In addition to the estimation of the signal decomposition in the dictionary, the intrinsic parameters of this dictionary should also be estimated. For brevity, we concentrate on the following scenario.

Suppose that we have in hand a set of l observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l \in \mathbb{R}^m$ of the same scene $\boldsymbol{x}_0 \in \mathbb{R}^n$, $m \leqslant n$, taken from different places. We model the observation system as a linear operator $\mathsf{A} \in \mathbb{R}^{m \times n}$. First, the observations being done at different positions, the scene does not appear the same to the observer and undergoes some geometric transformations. Second, the scene is not always entirely visible, e.g., objects might create occlusions. We consequently model the observation system as follows:

$$\boldsymbol{y}_j = \mathsf{A}\left(\boldsymbol{x}_0 \circ \tau_j + \boldsymbol{x}_j\right) + \boldsymbol{n}_j, \ \ \forall j \in \{1, \ldots, l\},$$

Appearing in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.

where $\boldsymbol{x}_0 \circ \tau_j$ is the image $\boldsymbol{x}_0$ transformed by τ_j, $\boldsymbol{x}_j \in \mathbb{R}^n$ represents innovations in the j^{th} observations, and $\boldsymbol{n}_j \in \mathbb{R}^m$ represents additive noise.

Recovering the images $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_l$, and the transformations $\tau_1, \ldots, \tau_l$, from the observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l$, is obviously an ill-posed inverse problem. To restrict the set of admissible solutions, we assume that the transformations belong to some transformation group represented by $\boldsymbol{p}$ parameters, and that the signals are sparse in a dictionary $\mathsf{D} \in \mathbb{R}^{n \times N}$ with $N \geqslant n$. In the following, the vectors $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_l \in \mathbb{R}^p$ contain the parameter values corresponding to $\tau_1, \ldots, \tau_l$. Under these assumptions, the observation model becomes

$$\boldsymbol{y}_j = \mathsf{A}\mathsf{D}_{\tau_j}\boldsymbol{\alpha}_0 + \mathsf{A}\mathsf{D}\boldsymbol{\alpha}_j + \boldsymbol{n}_j, \ \ \forall j \in \{1, \ldots, l\}, \quad (1)$$

where $\boldsymbol{\alpha}_j \in \mathbb{R}^N$ is the sparsest decomposition of $\boldsymbol{x}_j$ in the corresponding dictionary, and D_{τ_j} is the dictionary D whose atoms are transformed by τ_j.

Note that the problem of multi-view compressive imaging is also considered in, e.g., (Park & Wakin, 2012). However, the convergence of the proposed algorithm is not studied. Furthermore, the authors consider a specific measurements matrix A and do not deal with the problem of occlusions.

2. Proposed approach

2.1. Motivation

To simplify the notations, we stack the vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l$, and $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_l$, in $\boldsymbol{y} \in \mathbb{R}^{lm}$ and $\boldsymbol{\alpha} \in \mathbb{R}^{(l+1)N}$, respectively: $\boldsymbol{y}^\mathsf{T} = (\boldsymbol{y}_1^\mathsf{T}, \ldots, \boldsymbol{y}_l^\mathsf{T})$ and $\boldsymbol{\alpha}^\mathsf{T} = (\boldsymbol{\alpha}_0^\mathsf{T}, \ldots, \boldsymbol{\alpha}_l^\mathsf{T})$, with $\cdot^\mathsf{T}$ denoting the transpose operator. We also define the observation matrix $\mathsf{A}(\boldsymbol{\tau}) \in \mathbb{R}^{lm \times (l+1)N}$, function of the parameters $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_l) \in \mathbb{R}^{p \times l}$, as follows:

$$\mathsf{A}(\boldsymbol{\tau}) := \begin{bmatrix} \mathsf{A}\mathsf{D}_{\tau_1} & \mathsf{A}\mathsf{D} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{A}\mathsf{D}_{\tau_l} & 0 & \ldots & \mathsf{A}\mathsf{D} \end{bmatrix}.$$

To recover a sparse vector $\boldsymbol{\alpha}^\star$ and a set of parameters $\boldsymbol{\tau}^\star$ that satisfy (1), we are first tempted to solve the following optimization problem:

$$(\boldsymbol{\alpha}^\star, \boldsymbol{\tau}^\star) = \underset{\boldsymbol{\alpha}, \boldsymbol{\tau}}{\operatorname{argmin}} \left\{ \eta \|\boldsymbol{\alpha}\|_1 + \frac{1}{2} \|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau})\boldsymbol{\alpha}\|_2^2 \right\}, \quad (2)$$

where $\eta > 0$ is a regularizing parameter.

Let us analyze the above minimization problem. We note that problem (2) is obviously non convex with respect to $\boldsymbol{\tau}$. Finding a global minimum is thus nearly impossible to guarantee. Nevertheless, we may still find a local minima by using a block coordinate descend method that, starting from an initial point $\boldsymbol{\alpha}^0 \in \mathbb{R}^{(l+1)N}$ and transformation $\boldsymbol{\tau}^0 \in \mathbb{R}^{p \times l}$, minimizes the objective function alternatively with respect to $\boldsymbol{\alpha}$ for $\boldsymbol{\tau}$ fixed and vice-versa. However, we note that such a method does not offer any control over the final value of the data error $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^\star)\boldsymbol{\alpha}^\star\|_2^2$. Furthermore, it is difficult to fix *a priori* the value of the regularizing parameter η. This value affects the final reconstruction quality and should be adapted with the number of signals, the energy of each signal, the level of the noise, etc. Formulation (2) is thus not entirely satisfying.

2.2. Bregman iterative regularization

Instead of solving problem (2) only once, the Bregman iterative regularization (Osher et al., 2005) solves a sequence of convex problem similar to (2) but involving the Bregman distance (Bregman, 1967) based on the ℓ_1 norm instead of the ℓ_1 norm itself.

Definition 1. *The Bregman distance based on the ℓ_1 norm between a point $\boldsymbol{\alpha} \in \mathbb{R}^{(l+1)N}$ and the reference point $\boldsymbol{\alpha}^0 \in \mathbb{R}^{(l+1)N}$ is defined as*

$$D_{\ell_1}^{\boldsymbol{p}}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^0) := \|\boldsymbol{\alpha}\|_1 - \|\boldsymbol{\alpha}^0\|_1 - \boldsymbol{p}^\mathsf{T}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0),$$

where $\boldsymbol{p} \in \mathbb{R}^{(l+1)N}$ is in the sub-differential of the ℓ_1 norm at $\boldsymbol{\alpha}^0$.

In the following, the sub-differential of the ℓ_1 norm at $\boldsymbol{\alpha}^0$ is denoted $\partial\|\boldsymbol{\alpha}^0\|_1$ and is the set $\{\boldsymbol{p} \in \mathbb{R}^{(l+1)N} : \|\boldsymbol{p}\|_\infty \leqslant 1 \text{ and } \boldsymbol{p}^\mathsf{T}\boldsymbol{\alpha}^0 = \|\boldsymbol{\alpha}^0\|_1\}$ (Bauschke & Combettes, 2011).

The Bregman iterative regularization was introduced in (Osher et al., 2005) for image denoising and deblurring. This procedure was then used in many applications including compressed sensing (Yin et al., 2008). We will see below that this procedure permits to control easily the ℓ_2 data error and that we can derive a simple rule to fix the regularizing parameter. Furthermore, we will also see that, for the problem considered, the algorithm has the tendency to work from coarse to fine scales, making it robust to large misalignments.

Algorithm 1

Inputs: observations $\boldsymbol{y} \in \mathbb{R}^{lm}$, initial transformation parameters $\boldsymbol{\tau}^0 = (\boldsymbol{\tau}_1^0, \ldots, \boldsymbol{\tau}_l^0) \in \mathbb{R}^{p \times l}$, $\gamma > 0$.
Initializations: Set $k = 0$, $\boldsymbol{\alpha}^0 = \mathbf{0} \in \mathbb{R}^{(l+1)N}$, $\boldsymbol{p}^0 = \mathbf{0} \in \mathbb{R}^{(l+1)N}$.
repeat
1) $\boldsymbol{\alpha}^{k+1} \leftarrow \operatorname{argmin}_{\boldsymbol{\alpha}} \left\{ \gamma D_{\ell_1}^{\boldsymbol{p}^k}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^k) + \frac{\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^k)\,\boldsymbol{\alpha}\|_2^2}{2} \right\}$.
2) $\boldsymbol{p}^{k+1} \leftarrow \boldsymbol{p}^k + \mathsf{A}^\mathsf{T}(\boldsymbol{\tau}^k)\left(\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^k)\,\boldsymbol{\alpha}^{k+1}\right)/\gamma$.
3) Find $\boldsymbol{\tau}^{k+1}$ such that $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^{k+1})\,\boldsymbol{\alpha}^{k+1}\|_2^2 \leqslant \|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^k)\,\boldsymbol{\alpha}^{k+1}\|_2^2$ (see Section 2.3).
until $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^{k+1})\,\boldsymbol{\alpha}^{k+1}\|_2^2 \leqslant \text{tol}$
Outputs: $\boldsymbol{\alpha}^\star = (\boldsymbol{\alpha}_0^\star, \ldots, \boldsymbol{\alpha}_l^\star)$, $\boldsymbol{\tau}^\star = (\boldsymbol{\tau}_1^\star, \ldots, \boldsymbol{\tau}_l^\star)$.

With a modification of the original Bregman iterative regularization procedure to incorporate an update of the optimization parameters, we propose Algorithm 1 to solve our signal recovery problem. With other few modifications of the proofs in (Osher et al., 2005), we can show that the procedure is well defined and state the following theorem.

Theorem 1. *The sequence $\{(\boldsymbol{\alpha}^k, \boldsymbol{\tau}^k)\}$ generated by Algorithm 1 satisfies the followings.*

1. ***Monotonic decrease of the ℓ_2 data error.*** $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^{k+1})\,\boldsymbol{\alpha}^{k+1}\|_2^2 \leqslant \|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^k)\,\boldsymbol{\alpha}^k\|_2^2$ *for all* $k \geqslant 0$.
2. ***Decrease of the ℓ_2 data error to zero.*** *Suppose there exist vectors $\tilde{\boldsymbol{\alpha}}_j \in \mathbb{R}^N$ such that $\boldsymbol{y}_j = \mathsf{AD}\tilde{\boldsymbol{\alpha}}_j$ for all $j \in \{1, \ldots, l\}$. Then, for all $k \geqslant 0$, $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^k)\,\boldsymbol{\alpha}^k\|_2^2 \leqslant 2\gamma\|\tilde{\boldsymbol{\alpha}}\|_1/k$, where $\tilde{\boldsymbol{\alpha}}^\mathsf{T} = (\mathbf{0}^\mathsf{T}, \tilde{\boldsymbol{\alpha}}_1^\mathsf{T}, \ldots, \tilde{\boldsymbol{\alpha}}_l^\mathsf{T})$.*
3. ***Convergence in D_{ℓ_1} in presence of noisy observations.*** *Suppose that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l$, are noisy and denote $\tilde{\boldsymbol{y}}_1, \ldots, \tilde{\boldsymbol{y}}_l$, the underlying noiseless observations. Suppose that there exist vectors $\tilde{\boldsymbol{\beta}}_j \in \mathbb{R}^N$ such that $\tilde{\boldsymbol{y}}_j = \mathsf{AD}\tilde{\boldsymbol{\beta}}_j$ for all $j \in \{1, \ldots, l\}$. Let us denote $\tilde{\boldsymbol{\beta}}^\mathsf{T} = (\mathbf{0}^\mathsf{T}, \tilde{\boldsymbol{\beta}}_1^\mathsf{T}, \ldots, \tilde{\boldsymbol{\beta}}_l^\mathsf{T})$ and let $\epsilon > 0$ represent the noise level*[1]*, i.e., $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau})\,\tilde{\boldsymbol{\beta}}\|_2^2 \leqslant \epsilon^2$. Then, $D_{\ell_1}^{\boldsymbol{p}^{k+1}}(\tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}^{k+1}) \leqslant D_{\ell_1}^{\boldsymbol{p}^k}(\tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}^k)$, as long as $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^{k+1})\,\boldsymbol{\alpha}^{k+1}\|_2^2 \geqslant \epsilon^2$.*

Let us comment briefly on Theorem 1. Point 1 shows that Algorithm 1 is working by iterative refinements. It updates the signal decomposition and the transformations to successively decrease the ℓ_2 data error. Point 2 shows that Algorithm 1 will eventually provide a solution at which the data error is zero. Note that it does not show that $\boldsymbol{\alpha}^k$ convergences to the trivial solution $\tilde{\boldsymbol{\alpha}}$. Hopefully, the solution provided by Algorithm

[1] Note that this value is independent of $\boldsymbol{\tau}$.

1 is more meaningful than the trivial solution $\tilde{\boldsymbol{\alpha}}$. This will be confirmed in Section 3. Finally, in presence of noisy observations, point 3 shows that the Bregman distance between the estimate $\boldsymbol{\alpha}^k$ and the trivial denoised solution $\tilde{\boldsymbol{\beta}}$ decreases. Point 3 also provides a natural stopping criterion for the algorithm. Indeed, if we have an estimate on the noise level, stopping as soon as $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^k)\,\boldsymbol{\alpha}^k\|_2^2 \leqslant \alpha\epsilon^2$, for some $\alpha > 1$, ensures that the Bregman distance between $\boldsymbol{\alpha}^k$ and $\tilde{\boldsymbol{\beta}}$ always decreases. Note that point 3 does not show that this distance decreases to zero.

2.3. Updating the transformation parameters

We discuss now the update of the transformation parameters at Step 3 of Algorithm 1.

First, we remark that the data error term is separable in a sum of l non-negative terms: $\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau})\boldsymbol{\alpha}^{k+1}\|_2^2 = \sum_{j=1}^{l} \|\boldsymbol{y}_j - \mathsf{AD}_{\tau_j}\boldsymbol{\alpha}_0^{k+1} - \mathsf{AD}\boldsymbol{\alpha}_j^{k+1}\|_2^2$, and minimizing the data error term with respect to $\boldsymbol{\tau}$ can be done by minimizing independently each summand. Second, we remark that the signal $\mathsf{D}_{\tau_j}\boldsymbol{\alpha}_0^{k+1}$ in the transformed dictionary D_{τ_j} corresponds to the signal $\boldsymbol{x}_0^{k+1} = \mathsf{D}\boldsymbol{\alpha}_0^{k+1}$ in the canonical dictionary D to which we apply the transformation τ_j: $\mathsf{D}_{\tau_j}\boldsymbol{\alpha}_0^{k+1} = \boldsymbol{x}_0^{k+1} \circ \tau_j^k$. Consequently, denoting $\tau_j^{k+1} = \tau_j^k + \Delta\tau_j$, Step 3 can be restated as follows. For all $j \in \{1, \ldots, l\}$, find $\Delta\tau_j \in \mathbb{R}^p$ such that

$$\|\boldsymbol{y}_j - \mathsf{A}\left(\boldsymbol{x}_0^{k+1} \circ (\tau_j^k + \Delta\tau_j) - \mathsf{D}\boldsymbol{\alpha}_j^{k+1}\right)\|_2^2 \leqslant \|\boldsymbol{y}_j - \mathsf{A}(\boldsymbol{x}_0^{k+1} \circ \tau_j^k - \mathsf{D}\,\boldsymbol{\alpha}_j^{k+1})\|_2^2.$$

For small $\Delta\tau_j$, we can linearize $\boldsymbol{x}_0^{k+1} \circ (\tau_j^k + \Delta\tau_j)$ around τ_j^k: $\boldsymbol{x}_0^{k+1} \circ (\tau_j^k + \Delta\tau_j) \approx \boldsymbol{x}_0^{k+1} \circ \tau_j^k + \mathsf{J}\,\Delta\tau_j$ where $\mathsf{J} \in \mathbb{R}^{n\times p}$ is the Jacobian of $\boldsymbol{x}_0^{k+1} \circ \tau_j^k$ with respect to the transformation parameters. Therefore, we have

$$\|\boldsymbol{y}_j - \mathsf{A}\left(\boldsymbol{x}_0^{k+1} \circ (\tau_j^k + \Delta\tau_j) - \mathsf{D}\boldsymbol{\alpha}_j^{k+1}\right)\|_2^2 \approx \|\boldsymbol{y}_j - \mathsf{A}(\boldsymbol{x}_0^{k+1} \circ \tau_j^k) - \mathsf{AJ}\Delta\tau_j - \mathsf{D}\,\boldsymbol{\alpha}_j^{k+1}\|_2^2. \quad (3)$$

The term on the right hand side of (3) is convex and quadratic with respect to $\Delta\tau_j$. Its minimum is attained at $\Delta\tau_j^\star = (\mathsf{AJ})^\dagger(\boldsymbol{y}_j - \boldsymbol{x}_0^{k+1} \circ \tau_j^k - \mathsf{D}\boldsymbol{\alpha}_j^{k+1})$, where $(\mathsf{AJ})^\dagger$ denotes the pseudo-inverse of AJ. If $\Delta\tau_j^\star$ is not too big, the approximation (3) is still valid and we can hope that $\Delta\tau_j^\star$ also minimize the left hand side of (3). In practice, we numerically compute the left hand side of (3) with $\Delta\tau_j = \Delta\tau_j^\star$ and check that it is smaller than with $\Delta\tau_j = 0$. In the positive case, we set $\tau_j^{k+1} = \tau_j^k + \Delta\tau_j^\star$, otherwise we keep $\tau_j^{k+1} = \tau_j^k$.

Figure 1. Top panel: first 3 ground truth images. Middle panel: first 3 reconstructed images when reconstructed independently. Bottom panel: first 3 recovered images $\boldsymbol{x}_0^\star \circ \tau_j^\star + \boldsymbol{x}_j^\star$ with Algorithm 1.

2.4. Setting the regularizing parameter

Even though Algorithm 1 converges for all $\gamma > 0$, the choice of this value is essential as, the problem being non convex, it can lead to very different solutions.

From point 1 of Theorem 1, we see that Algorithm 1 is working by iterative refinements. It was actually observed that the Bregman iterative regularization has the tendency to recover the coefficients in $\boldsymbol{\alpha}$ in decreasing order of magnitude (Yin et al., 2008). Let us consider that D is a wavelet basis and that the data under scrutiny are natural images. The largest wavelet coefficients of natural images usually live at the coarsest scales. Therefore, if we set γ so that only a few large coefficients are recovered at the first iteration, the algorithm will have the tendency to work from coarse to fine scales. Such a behavior is interesting as multi-scale approaches in image registration are usually robust to large misalignments.

At $k = 0$, the minimization problem at Step 1 of Algorithm 1 can be replaced by

$$\boldsymbol{\alpha}^1 \leftarrow \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\, \gamma\|\boldsymbol{\alpha}\|_1 + \frac{1}{2}\|\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^0)\,\boldsymbol{\alpha}\|_2^2.$$

Applying the first order optimality condition to the above problem, we have $\mathsf{A}^\mathsf{T}(\boldsymbol{\tau}^0)\left[\left(\boldsymbol{y} - \mathsf{A}(\boldsymbol{\tau}^0)\,\boldsymbol{\alpha}^1\right)/\gamma\right] \in$

$\partial\|\boldsymbol{\alpha}^1\|_1$. Therefore, $\boldsymbol{\alpha}^1 = \mathbf{0}$ if $\mathsf{A}^\mathsf{T}(\boldsymbol{\tau}^0)(\boldsymbol{y}/\gamma) \in \partial\|\mathbf{0}\|_1$, i.e., $\|\mathsf{A}^\mathsf{T}(\boldsymbol{\tau}^0)\boldsymbol{y}\|_\infty \leqslant \gamma$. Choosing a value of γ slightly smaller that $\|\mathsf{A}^\mathsf{T}(\boldsymbol{\tau}^0)\boldsymbol{y}\|_\infty$ ensures that only a few wavelet coefficients at the coarsest scales are recovered at the first iteration.

3. Experiment

In the following experiment, 5 different images[2] of the same scene are used to generate 5 different measurement vectors. These images contain $n = 256 \times 256$ pixels and we set $m = 0.1n$. Three of these images are shown in Fig. 1. The images are taken from different point of views and thus undergo geometric transformations. We assume that these transformations are homographies modeled by 8 unknown parameters. Let us remark that parts of scene are sometimes occluded. The measurements are obtained using the spread spectrum technique (Puy et al., 2012). The matrix A thus models a random modulation with a ± 1 sequence, followed by a random selection of m Fourier coefficients.

For comparison, these images are reconstructed independently by solving the Basis Pursuit problem (Candès, 2006) and jointly with Algorithm 1. In both cases, the Haar wavelet basis is used for the dictionary D. The results are presented in Fig. 1. One can notice that the reconstructions obtained with our method exhibit much finer details. To highlight the accuracy of the estimated transformations, Fig. 2 shows the ground truth images superposed before and after registration with the estimated parameters $\boldsymbol{\tau}^\star$. Algorithm 1 is able to correctly estimate the large geometric transformations between the images with only 10% of measurements and in presence of occlusions. Finally, Fig. 2 also shows the reconstructed background image $\boldsymbol{x}_0^\star$. One can notice that this image does not contain any object occluding parts of the scene. Noticing that these objects actually appears in the recovered images $\boldsymbol{x}_0^\star \circ \tau^\star + \boldsymbol{x}_j^\star$ in Fig. 1, we conclude that the separation between the background-foreground image is also very accurate.

4. Conclusion

We have presented a method for joint reconstruction of misaligned images. Theoretical results about the convergence of the algorithm have been presented. Experiments show that we accurately reconstruct a set of misaligned images in the presence of occlusions and large misalignments. The presented method may have interests in, e.g., non-dynamic cardiac MR imaging where one has access to only subsampled images of the heart at different positions.

[2]castle-R20 dataset available at `cvlab.epfl.ch/~strecha/multiview/rawMVS.html` (Strecha et al., 2008).

Figure 2. From left to right: sum of the 5 ground truth images before registration; sum of the 5 ground truth images after registration with the estimated transformations; recovered background image $\boldsymbol{x}_0^\star$.

References

Bauschke, H. H. and Combettes, P. L. *Convex analysis and monotone operator theory in Hilbert spaces.* Springer New York, 2011.

Bregman, L.M. The relaxation method of finding the commonpoint of convex sets and its application to the solution of problems in convex programming. *Comput. Math. Math. Phys.*, 7:200–217, 1967.

Candès, E. J. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, pp. 1433–1452, Madrid, Spain, 2006.

Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4:460–489, 2005.

Park, J. Y. and Wakin, M. B. A geometric approach to multi-view compressive imaging. *EURASIP J. Adv. Sig. Pr.*, 2012(37), 2012.

Puy, G., Vandergheynst, P., Gribonval, R., and Wiaux, Y. Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques. *EURASIP J. Adv. Sig. Pr.*, 2012(6), 2012.

Strecha, C., von Hansen, W., Gool, L. Van, Fua, P., and Thoennessen, U. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, 2008.

Yin, W., Osher, S., Goldfarb, D., and Darbon, J. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM J. Imaging Sci.*, 1:143–168, 2008.