# JOINT SPARSE RECOVERY OF MISALIGNED MULTIMODAL IMAGES VIA ADAPTIVE LOCAL AND NONLOCAL CROSS-MODAL REGULARIZATION

*Nasser Eslahi and Alessandro Foi*

Tampere University, Finland

## ABSTRACT

Given few noisy linear measurements of distinct misaligned modalities, we aim at recovering the underlying multimodal image using a sparsity promoting algorithm. Unlike previous multimodal sparse recovery approaches employing side information under the naive assumption of perfect calibration of modalities or of known deformation parameters, we adaptively estimate the deformation parameters from the images separately recovered from the incomplete measurements. We develop a multiscale dense registration method that proceeds alternately by finding block-wise intensity mapping models and a shift vector field which is used to obtain and refine the deformation parameters through a weighted least-squares approximation. The co-registered images are then jointly recovered in a plug-and-play framework where a collaborative filter leverages the local and nonlocal cross-modal correlations inherent to the multimodal image. Our experiments with this fully automatic registration and joint recovery pipeline show a better detection and sharper recovery of fine details which could not be separately recovered.

## 1. INTRODUCTION

Multimodal imaging methods acquire multiple measurements of an object using different acquisition techniques or distinct sensors, providing various aspects of a phenomenon of interest. The mutual and complementary information found across different modalities can be used synergistically, to enable exploration, insight, analysis, and diagnostics which would not be possible using the individual modalities alone. Direct capture of a multimodal image is often not possible, with some modalities requiring expensive indirect measurements. Multimodal sparse recovery aims to recover an underlying multimodal image from its few, possibly noisy, linear measurements. Such computational imaging technique is being widely used in diverse applications, including biomedical imaging [1–3], joint depth-intensity imaging [4], multispectral imaging [5], and beyond. The typical multimodal sparse recovery approaches either individually treat each modality, thus separately reconstructing the images, or pursue a joint recovery under the assumption that different modalities share structural similarities. The images reconstructed in the former approach may then be fused to reveal more informative data; however, it has been shown that the latter approach can remarkably improve the quality of recovery [1–5]. Joint multimodal sparse recovery commonly leverages additional guidance data, referred to as *side information* [4, 5], or considers much higher sampling rates for some of the modalities [2, 3] in order to enhance the recovery of the most sparsely sampled images.

In joint multimodal imaging, it is commonly assumed that modalities are perfectly co-registered, or that the deformation parameters are known beforehand and obtained by a separate offline calibration procedure. In a realistic scenario, however, the acquisition devices may not be perfectly aligned making the image registration a crucial component of multimodality imaging. Moreover, due to hardware constraints of acquisition devices or different acquisition setups, the acquired images may go under some types of deformation (e.g., geometric transformations, or optical deformations and aberrations) which along with the intensity variation and structural difference across modalities make the image registration especially challenging. Accurate registration is crucial, as joint recovery or fusion under imprecise registration can be detrimental [6, 7].

The objective of this paper is to exploit the inherent local and nonlocal cross-modal correlation in the sparse recovery from severely underdetermined measurements of misaligned modalities under unknown registration information. To this end, we resort to a hybrid recovery procedure consisting of sequential separate and joint recovery phases (Section 2). The intermediate step between these recovery phases is to estimate the deformation parameters from the separately recovered images, providing also the co-registered images used to initialize the joint recovery phase. Our key contributions are summarized as follows:

- We develop a dense registration method that proceeds alternately, in a coarse-to-fine multiscale fashion, by finding block-wise intensity mapping models and a shift vector field which is then used to obtain and refine the deformation parameters through a weighted least-squares approximation (Section 3.1).
- We propose a joint multimodal sparse recovery approach that proceeds iteratively by refining the estimation of the underlying signal using a stationary correlated noise denoiser, thus extending our work [8] to the multimodal case (Section 3.2).
- We validate our proposed approach in the context of multi-contrast magnetic resonance (MR) imaging and multichannel sparse recovery, showing the significant improvement by the joint recovery in artifact reduction and sharper recovery of finer details that could not be resolved from separate recovery (Section 4).

In this preliminary paper, we implemented and tested the approach in its simplest form, where the registration is performed in full over the separately recovered images, and the joint recovery is done at the finest scale after registration (see Fig. 1). However, since the the adopted collaborative filter can also be operated in a coarse-to-fine manner, all our developed elements can be combined so that the joint recovery is executed progressively within the coarse-to-fine multiscale registration.

♡ This work was supported by ECSEL JU (project no. 783162 FitOptiVis), and by the Academy of Finland (project no. 310779).

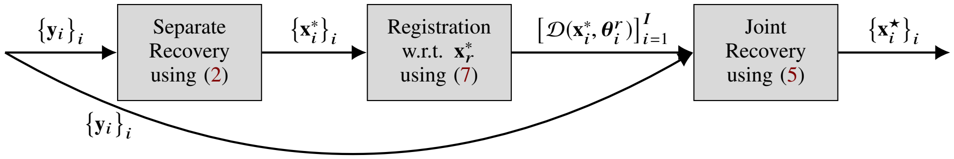
$\{\mathbf{y}_i\}_i$
Separate Recovery using (2)
$\{\mathbf{x}_i^*\}_i$
Registration w.r.t. $\mathbf{x}_r^*$ using (7)
$\big[\mathcal{D}(\mathbf{x}_i^*, \boldsymbol{\theta}_i^r)\big]_{i=1}^I$
Joint Recovery using (5)
$\{\mathbf{x}_i^\star\}_i$
$\{\mathbf{y}_i\}_i$

**Fig. 1**. The overall schematic of the proposed multimodal recovery pipeline.

## 2. PROBLEM FORMULATION

Let $\mathbf{x}_i \in \mathbb{R}^{n_i}$ represent an image of a same physical object viewed in the i-th modality, $i \in [1, \dots, I]$. We consider multimodal imaging as an underdetermined linear inverse problem

$$\mathbf{y}_i = \mathbf{A}_i\mathbf{x}_i + \boldsymbol{\epsilon}_i\,, \tag{1}$$

where for each modality $i$, $\mathbf{A}_i \in \mathbb{C}^{m_i \times n_i}$ is the measurement matrix with $m_i \ll n_i$, $\boldsymbol{\epsilon}_i$ represents the measurement error, and $\mathbf{y}_i$ is the measurement vector. The statistics of $\boldsymbol{\epsilon}_i$ (e.g., the noise distribution and level) can be different at each modality.

The individual estimation of $\mathbf{x}_i$ in (1) is an ill-posed linear inverse problem which can be obtained through a regularization-based optimization problem of the form

$$\mathbf{x}_i^* = \underset{\mathbf{x}_i \in \mathbb{R}^{n_i}}{\operatorname{argmin}}\, \mathcal{R}_i(\mathbf{x}_i) + \gamma_i Q_i(\mathbf{x}_i, \mathbf{y}_i),\ \forall i \in [1, \dots, I], \tag{2}$$

where $Q_i$ and $\mathcal{R}_i$ are respectively the data-fidelity with respect to the $i$-th observation model and the regularization terms, and $\gamma_i > 0$ the regularization parameter balancing the contribution of both terms.

Many optimization algorithms have been proposed to solve (2), with $\mathcal{R}_i$ modeling the sparsity of $\mathbf{x}_i$ with respect to a transform. Among those algorithms, proximal splitting methods [9–12] can be applied for large-scale optimization problems while handling the nonsmoothness of regularizers by using the proximal operator of $\mathcal{R}_i$. Plug-and-Play (PnP) is a non-convex framework that integrates advanced denoising filters $\Phi_i$ instead of proximal operators within proximal algorithms, adopting these filters as implicit prior models for model-based inversion. In this work, we employ the PnP-based FBS (*forward-backward splitting*) algorithm [11] to solve (2), as it does not require to perform an inversion on the forward model (1), following an iterative procedure of the form

$$\mathbf{b}_{i,k-1} = -\rho_i \nabla Q_i(\mathbf{x}_{i,k-1}, \mathbf{y}_i), \tag{3a}$$

$$\mathbf{u}_{i,k} = \Phi_i\big(\mathbf{x}_{i,k-1} + \mathbf{b}_{i,k-1}\big), \tag{3b}$$

$$\mathbf{x}_{i,k} = \mathbf{u}_{i,k} + t_k\big(\mathbf{u}_{i,k} - \mathbf{u}_{i,k-1}\big), \tag{3c}$$

where $\rho_i > 0$ is the step size, $\nabla$ is the gradient operator, $t_k \in [0, 1)$ is the prediction parameter at iteration $k \geq 1$, $t_1 = 0$, and $\mathbf{x}_{i,0} = \mathbf{u}_{i,0} \in \mathbb{R}^{n_i}$.

In (3b), the action of $\Phi_i$ on its input can be regarded as a denoiser seeking to recover $\mathbf{x}_i$ from the noisy observation [8]

$$\mathbf{z}_{i,k} = \mathbf{x} + \boldsymbol{\nu}_{i,k} = \mathbf{x}_{i,k-1} + \mathbf{b}_{i,k-1}, \tag{4}$$

where $\mathbf{z}_{i,k}$ and $\boldsymbol{\nu}_{i,k}$ respectively represent the *noisy signal* to be filtered and the *effective noise* at each iteration of the algorithm. We tacitly assume that the input of $\Phi_i$ is reshaped so to reconstitute the multidimensional representation of the data and that its output is vectorized back.

Now, let consider the image of one modality $\mathbf{x}_{r \in [1,\dots,I]}$ as the *reference* image so that the other images $\{\mathbf{x}_i\}_{i \neq r}$, which we call them *moving* images, are being aligned with respect to $\mathbf{x}_r$. We denote by $\boldsymbol{\theta}_i^r$ the deformation parameter whereby the moving image in the system $i$ is transformed to the reference system $r$, and represent this deformation by $\mathcal{D}(\mathbf{x}_i, \boldsymbol{\theta}_i^r)$, with $\mathcal{D}(\mathbf{x}_r, \boldsymbol{\theta}_r^r) = \mathbf{x}_r$. Given $\boldsymbol{\theta}_i^r$, the joint estimation of $\mathbf{x}_i$ in (1) can be obtained through solving

$$\{\mathbf{x}_i^\star\}_{i=1}^I = \underset{\{\mathbf{x}_i\}_I \in \mathbb{R}^{n_i}}{\operatorname{argmin}}\ \mathcal{R}\Big(\big[\mathcal{D}(\mathbf{x}_i, \boldsymbol{\theta}_i^r)\big]_{i=1}^I\Big) + \sum_{i=1}^I \gamma_i Q_i(\mathbf{x}_i, \mathbf{y}_i), \tag{5}$$

where the operator $[\cdot]_{i=1}^I$ stacks the overlapped portions of the registered images into a 3D array, and $\mathcal{R}$ is the joint regularizer. We employ PnP-FBS to solve (5) as well.

The following principal questions are addressed in the next section:
*1.* How are the deformation parameters $\{\boldsymbol{\theta}_i^r\}_{i \neq r}$ estimated?
*2.* How is $\boldsymbol{\nu}_{i,k}$ modeled and estimated in either separate and joint recovery procedures?

## 3. ADAPTIVE LOCAL AND NONLOCAL CROSS-MODAL REGULARIZATION

### 3.1. Automatic Multiscale Multimodal Image Registration

Given two modalities $\mathbf{x}_r$ and $\mathbf{x}_t$, $\forall r, t \in [1, \dots I]$, respectively as the reference and moving images, one may seek a functional relationship between the intensity values of them. Therefore, the registration can be obtained through maximizing a similarity measure, e.g.,

$$\underset{\boldsymbol{\theta}_t^r,\, \mathcal{P}}{\operatorname{argmin}} \left\| \mathbf{x}_r - \mathcal{P}\big(\mathcal{D}(\mathbf{x}_t, \boldsymbol{\theta}_t^r)\big) \right\|_2^2 \tag{6}$$

which yields deformation parameters $\boldsymbol{\theta}_t^r$ and an intensity mapping $\mathcal{P}$ such that after deformation of $\mathbf{x}_t$ and mapping of its intensities, $\mathcal{P}\big(\mathcal{D}(\mathbf{x}_t, \boldsymbol{\theta}_t^r)\big)$ becomes as close as possible to $\mathbf{x}_r$. The optimization (6) can be solved using the alternate minimization strategy along $\mathcal{P}$ and $\boldsymbol{\theta}_t^r$, e.g., as in [13, 14], where $\mathcal{P}$ was fitted using a global polynomial function.

However, the intensity relationship is often weaker and more complex than can be explained by a global functional form as described by (6). Instead, we seek for this relationship only locally in a block-wise fashion, by fitting low-order polynomial models independently for each pair of co-located blocks from the moving and reference images. We then treat the block extracted from the intensity-mapped $\mathbf{x}_t$ as a shifted version of the corresponding block in $\mathbf{x}_r$, and estimate the corresponding shift vector through least-squares fit of the block differences by the corresponding block of the horizontal and vertical gradients of $\mathbf{x}_r$. This implicitly relies on a locally affine model of the image intensities, and therefore we estimate the shift map in a progressive coarse-to-fine multiscale fashion (e.g., [15]), where only small incremental shifts are resolved at the finer scales. At each scale, the global deformation parameters $\boldsymbol{\theta}_t^r$ are estimated from a weighted least-squares fit of the shift vector field, with weights depending on how well the gradients of $\mathbf{x}_r$ can locally approximate the intensity-mapped shifted $\mathbf{x}_t$.

At every scale, one may interpret the above algorithm as loosely tackling the following optimization:

$$\underset{\boldsymbol{\theta}_t^r,\, \{\mathcal{P}_j\}_j}{\operatorname{argmin}} \sum_j \mathbf{w}_j^2 \left\| \mathcal{B}_j(\mathbf{x}_r^s) - \mathcal{P}_j(\mathcal{B}_j(\mathcal{D}(\mathbf{x}_t^s, \boldsymbol{\theta}_t^r))) \right\|_2^2, \tag{7}$$

where $\mathcal{B}_j$ denotes the extraction of the $j$-th block from an image, $\mathcal{P}_j$ the local intensity-mapping polynomial model, $\mathbf{w}_j$ are weights promoting blocks with higher contrast where $\mathcal{P}_j$ is able to locally explain the intensities of $\mathbf{x}_r$ from $\mathbf{x}_t$, and the superscript $s$ indicates different scales of the image.

### 3.2. Adaptive Stationary Correlated Noise Modeling

In [16] and [8], we tacitly modeled the degradations as hybrid noise comprised of a *nonstationary* and a *stationary* noise component, and have shown that employing an additive stationary correlated noise model within this modeling results in a better and faster signal recovery over the additive white Gaussian noise (AWGN) one. We extend our work [8] to the multimodal case by modeling the stationary component of $\nu_{i,k}$ as additive correlated noise in both individual and joint multimodal sparse recovery. In particular, leveraging this noise model in a sparsity-promoting collaborative denoiser $\Phi$ (3b) for solving (5) in the PnP framework, corresponds to installing an adaptive local and nonlocal cross-modal regularization.

The overall proposed multimodal sparse recovery pipeline is illustrated in Fig. 1.

## 4. PERFORMANCE EVALUATION

To evaluate our proposed multimodal imaging method, we first focus on muticontrast MR imaging and then on joint multichannel image recovery. In both experiments, for the separate (resp. joint) recovery, we employ the PnP-FBS framework with BM3D [17] (resp. BM4D [18]) as the denoiser $\Phi$.

### 4.1. Multicontrast MR Imaging

We consider 217×181 T1 and T2 BrainWeb transverse 2-D slices [19]. To produce a misaligned pair, the T2 image is first rotated by 6° clockwise, then magnified 1.2 times, translated by 4.4 pixels horizontally and 5.5 pixels vertically, and finally cropped to 217×181 pixels (see Fig. 2). We treat the T1 image as the moving $\mathbf{x}_1$ and the deformed T2 image as the reference $\mathbf{x}_2$. The matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ correspond to sampling the 2D FFT over 20 ($m_1/n_1 = 0.097$) and 30 ($m_2/n_2 = 0.144$) radial lines, respectively. We consider the problem of separate and joint recovery of $\mathbf{x}_1$ and $\mathbf{x}_2$ from incomplete noisy measurements (1) $\mathbf{y}_1$ (AWGN, SNR = 30 dB) and $\mathbf{y}_2$ (AWGN, SNR = 20 dB).

In the separate recovery phase, we did experiments for several values of step sizes $\rho_1$ and $\rho_2$ to obtain the estimates with the highest peak signal-to-noise ratio (PSNR, i.e. $20\log_{10}(\sqrt{n_i}\max(\mathbf{x}_i)\|\mathbf{x}_i - \mathbf{x}_{i,k}\|_2^{-1})$) over 100 iterations, i.e. $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$. We then obtain the deformation parameter $\boldsymbol{\theta}_1^2$ using the proposed automatic multiscale multimodal registration method. The partial overlapping parts of the co-registered $\mathcal{D}(\mathbf{x}_1^*, \boldsymbol{\theta}_1^2)$ and $\mathbf{x}_2^*$ are then extracted and stacked in a 3D array for the initialization in the joint recovery phase.

Fig. 3 shows the overlapped portion of the co-registered recovered MR images; the jointly recovered images are registered by the automatically obtained deformation parameter, whereas the rest are registered by ground-truth parameter. As can be seen, joint recovery helps in recovering details which could not be separately recovered.

### 4.2. Multichannel Image Recovery

We consider the red and blue channels of the 512×512 *Toy* RGB image from the CAVE dataset [20]. To misalign the channels, the blue channel is first rotated by 10° counterclockwise, then magnified 1.25 times, translated by 5.3 pixels horizontally and 2.8 pixels vertically, and finally cropped to 512×512 (see Fig. 4). We take the red channel as the moving $\mathbf{x}_1$ and the deformed blue channel as the reference $\mathbf{x}_2$. The matrix $\mathbf{A}_1$ is associated to sampling 35 radial lines of the 2D-FFT ($m_1/n_1 = 0.067$), whereas $\mathbf{A}_2$ to 15% pseudo-random sampling of the 2D FFT. Noisy measurements (1) are then produced as $\mathbf{y}_1$ (AWGN, SNR = 30 dB) and $\mathbf{y}_2$ (AWGN, SNR = 20 dB).

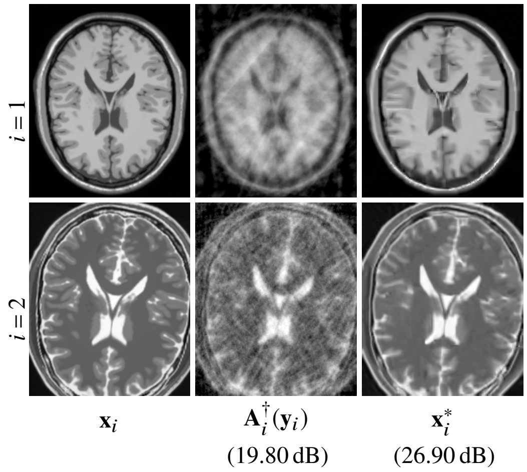

**Fig. 2**. From left to right: original, back-projected, and separately recovered multicontrast MR images in T1 (top) and T2 (bottom) modality. The intensity values of the recovered images are clipped to the intensity range of the original images, for a better visualization. The reported values are obtained by averaging the PSNRs values of the individual modalities.

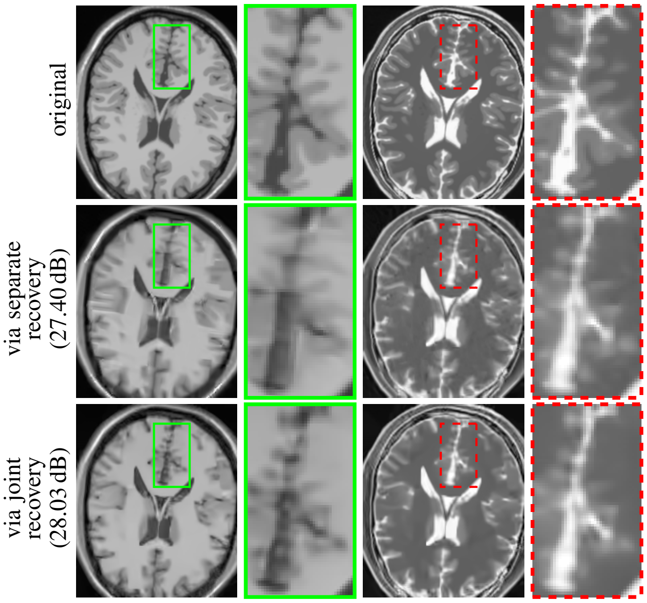
original
via separate recovery (27.40 dB)
via joint recovery (28.03 dB)

**Fig. 3**. Co-registered recovered MR T1 (left) and T2 (right) images. The reported values are obtained by averaging the PSNR values computed over the overlapping supports of the recovered modalities.

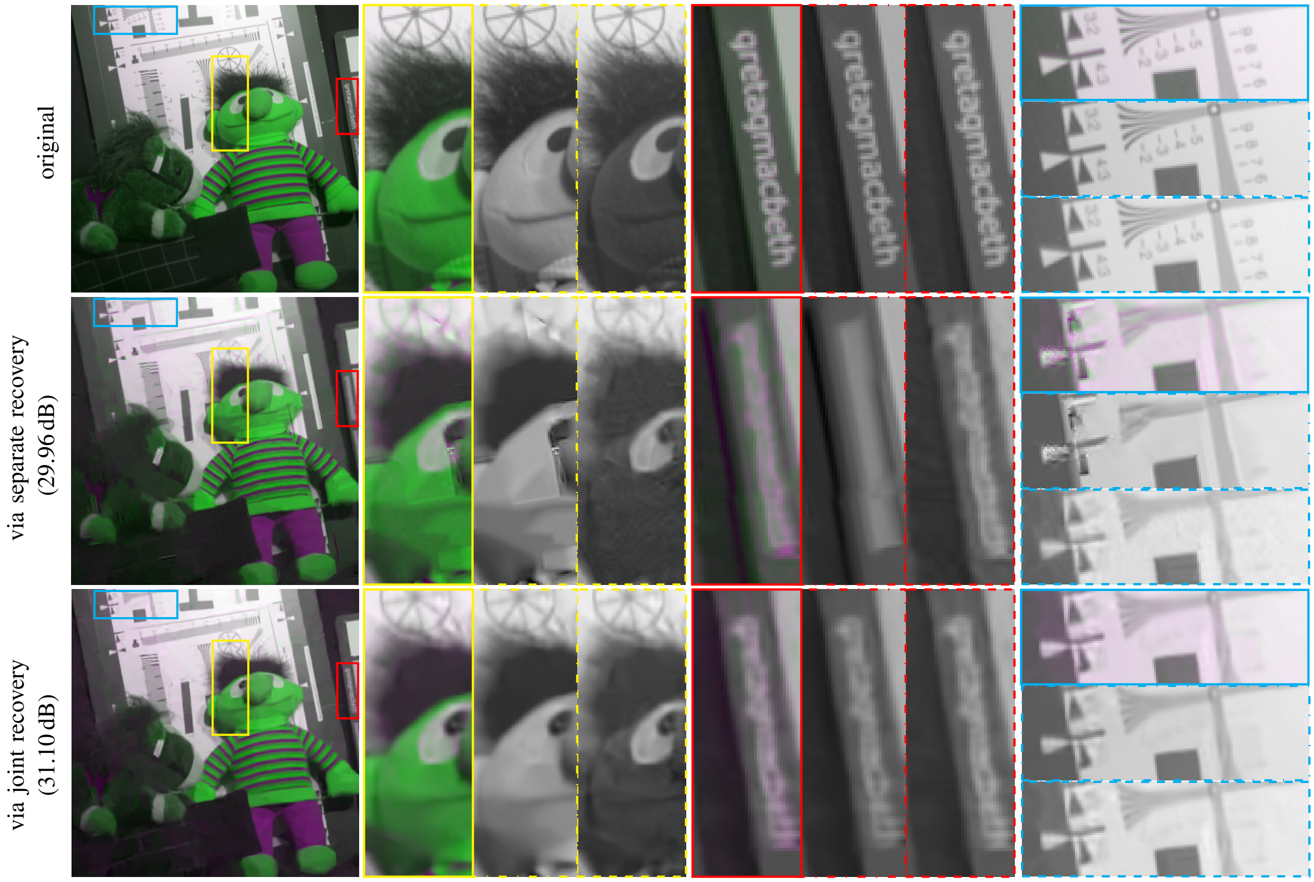

**Fig. 5**. Visual comparison of the registered recovered modalities. The images on the leftmost column are green-magenta pseudo-color representation of the two modalities. The magnifications shown by solid line, dashed-dotted, and dashed borders correspond respectively to pseudo-color, first modality, and second modality. The reported values are obtained by averaging the PSNR values computed over the overlapping supports of the recovered modalities.

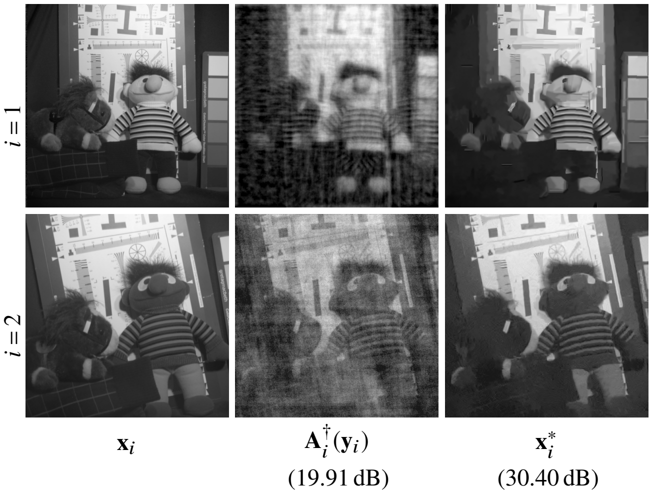

**Fig. 4**. From left to right: original, back-projected, and separately recovered multichannel images. The intensity values of the recovered images are clipped to the intensity range of the original images, for a better visualization. The reported values are obtained by averaging the PSNRs values of the individual modalities.

Fig. 5 shows the green-magenta pseudo-color representation of the overlapped portion of the co-registered recovered images. As can be seen from Fig. 5, the proposed joint recovery yields a better detection and sharper recovery of fine structures with less artifacts around the data compared to the separate recovery results.

MATLAB software used for the experiments of this section are available on the authors' institutional homepage at http://www.cs.tut.fi/~foi/multimodal .

## 5. CONCLUSIONS

We estimate deformation parameters in a coarse-to-fine multiscale fashion, mapping the intensities of each block by a localized polynomial model. These parameters are embedded in a PnP recovery approach promoting sparsity via a collaborative filter that exploits the local and nonlocal cross-modal correlations. Experimental results demonstrate the superior subjective and objective performance of the proposed joint recovery approach over the separate one.

In this paper, the registration is performed over the separately recovered images, and the joint recovery is done at the finest scale after registration. However, since the adopted collaborative filter can also be operated in a coarse-to-fine manner, all our developed elements can be combined so that the joint recovery is executed progressively within the coarse-to-fine multiscale registration, i.e. integrating (5) and (7) into a unique optimization. This is the subject of ongoing work which we will report in an extended version of this paper.

## 6. REFERENCES

[1] L. Martí-Bonmatí, R. Sopena, P. Bartumeus, and P. Sopena, "Multimodality imaging techniques," *Contrast Media & Mol. I.*, vol. 5, no. 4, pp. 180–189, 2010.

[2] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, "Joint reconstruction of PET-MRI by exploiting structural similarity," *Inverse Probl.*, vol. 31, no. 1, p. 015001, 2014.

[3] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. K. Sodickson, "Joint MR-PET reconstruction using a multi-channel image regularizer," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 1–16, 2016.

[4] K. Degraux, U. S. Kamilov, P. T. Boufounos, and D. Liu, "Online convolutional dictionary learning for multimodal imaging," in *Proc. IEEE ICIP*, 2017, pp. 1617–1621.

[5] X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 964–976, 2015.

[6] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, 2015.

[7] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[8] N. Eslahi and A. Foi, "Anisotropic spatiotemporal regularization in compressive video recovery by adaptively modeling the residual errors as correlated noise," in *Proc. IEEE IVMSP*, 2018.

[9] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, 1992.

[10] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[11] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[13] A. Roche, X. Pennec, G. Malandain, and N. Ayache, "Rigid registration of 3-D ultrasound with MR images: a new approach combining intensity and gradient information," *IEEE Trans. Med. Imag.*, vol. 20, no. 10, pp. 1038–1049, 2001.

[14] W. Ou and C. Chefd'Hotel, "Polynomial intensity correction for multimodal image registration," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2009, pp. 939–942.

[15] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. ECCV'92*, G. Sandini, Ed., 1992, pp. 237–252.

[16] N. Eslahi, V. Ramakrishnan, K. Wiik, and A. Foi, "Sparse signal recovery via correlated degradation modeling," in *Proc. SPARS*, 2017.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[18] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, 2013.

[19] R. Vincent. (2006) Brainweb: Simulated brain database. [Online]. Available: https://brainweb.bic.mni.mcgill.ca/

[20] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "CAVE multispectral image database," http://www.cs.columbia.edu/CAVE/databases/multispectral/, 2008.