Deep Sparse Representation for Robust Image Registration

Yeqing Li*[1], Chen Chen*[1], Fei Yang[2], and Junzhou Huang[1]
[1]University of Texas at Arlington
[2]Facebook Inc.

Abstract

The definition of the similarity measure is an essential component in image registration. In this paper, we propose a novel similarity measure for registration of two or more images. The proposed method is motivated by that the optimally registered images can be deeply sparsified in the gradient domain and frequency domain, with the separation of a sparse tensor of errors. One of the key advantages of the proposed similarity measure is its robustness to severe intensity distortions, which widely exist on medical images, remotely sensed images and natural photos due to the difference of acquisition modalities or illumination conditions. Two efficient algorithms are proposed to solve the batch image registration and pair registration problems in a unified framework. We validate our method on extensive challenging datasets. The experimental results demonstrate the robustness, accuracy and efficiency of our method over 9 traditional and state-of-the-art algorithms on synthetic images and a wide range of real-world applications.

1. Introduction

Image registration is a fundamental task in image processing and computer vision [29, 23, 20]. It aims to align two or more images into the same coordinate system, and then these images can be processed or compared. Accuracy and robustness are two of the most important metrics to evaluate a registration method. It has been shown that a mean geometric distortion of only 0.3 pixel will result in noticeable effect on a pixel-to-pixel image fusion process [3]. Robustness is defined as the ability to get close to the accurate results on different trials under diverse conditions. Based on the feature used in registration, existing methods can be classified into feature-based registration (e.g., [28, 16, 15]) and pixel-based registration (e.g., [10, 6, 26, 25]). Feature-based methods rely on the landmarks extracted from the images. However, extracting reliable features is still an open problem and an active topic of research [20]. In this paper, we are interested in image registration by directly using their pixel values. In addition, we wish to successfully register the images from a variety of applications in subpixel-level accuracy, as precisely as possible.

One key component for image registration is the energy function to measure (dis)similarity. The optimized similarity should lead to the correct spatial alignment. However, finding a reliable similarity measure is quite challenging due to the unpredicted variations of the input images. In many real-world applications, the images to be registered may be acquired at different times and locations, under various illumination conditions and occlusions, or by different acquisition modalities. As a result, the intensity fields of the images may vary significantly. For instance, slow-varying intensity bias fields often exist in brain magnetic resonance images [22]; the remotely sensed images may even have inverse contrast for the same land objects, as multiple sensors have different sensitivities to wavelength spectrum [24]. Unfortunately, many existing pixel-based similarity measures are not robust to these intensity variations, e.g., the widely used sum-of-squared-difference (SSD) [23].

Recently, the sparsity-inducing similarity measures have been repeatedly successful in overcoming such registration difficulties [17, 19, 27, 9]. In RASL [19] (robust alignment by sparse and low-rank decomposition), the images are vectorized to form a data matrix. The transformations are estimated to seek a low rank and sparse representation of the aligned images. Two online alignment methods, ORIA [27] (online robust image alignment) and t-GRASTA [9] (transformed Grassmannian robust adaptive subspace tracking algorithm), are proposed to improve the scalability of RASL. All of these methods assume that the large errors among the images are sparse (e.g., caused by shadows, partial occlusions) and separable. However, as we will show later, many real-world images contain severe spatially-varying intensity distortions. These intensity variations are not sparse and therefore difficult to be separated by these methods. As a result, the above measures may fail to find the correct alignment and thus are less robust in these challenging tasks.

*indicates equal contributions. Corresponding author: Junzhou Huang. Email: jzhuang@uta.edu. This work was partially supported by NSF IIS-1423056, CMMI-1434401, CNS-1405985.

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

The residual complexity (RC) [17] is one of the best measures for registering two images corrupted by severe intensity distortion [8], which uses the discrete cosine transform (DCT) to sparsify the residual of two images. For a batch of images, RC has to register them pair-by-pair and the solution may be sub-optimal. In addition, DCT and inverse DCT are required in each iteration, which slows down the overall speed of registration. Finally, although RC is robust to intensity distortions, the ability of RC to handle partial occlusions is unknown.

Unlike previous works that vectorize each image into a vector [19, 27, 9], we arrange the input images into a 3D tensor to keep their spatial structure. With this arrangement, the optimally registered image tensor can be deeply sparsified into a sparse frequency tensor and a sparse error tensor (see Fig. 1 for more details). Severe intensity distortions and partial occlusions will be sparsified and separated out in the first and second layers, while any misalignment will increase the sparseness of the frequency tensor (third layer). We propose a novel similarity measure based on such deep sparse representation of the natural images. Compared with the low rank similarity measure which requires a batch of input images, the proposed similarity measure still works even when there are only two input images. An efficient algorithm based on the Augmented Lagrange Multiplier (ALM) method is proposed for the batch mode, while the gradient descent method with backtracking is presented to solve the pair registration problem. Both algorithms have very low computational complexity in each iteration. We compare our method with 9 traditional and state-of-the-art algorithms on a wide range of natural image datasets, including medical images, remotely sensed images and photos. Extensive results demonstrate that our method is more robust to different types of intensity variations and always achieves higher sub-pixel accuracy over all the tested methods.

2. Image registration via deep sparse representation

In this paper, we use bold letters denote multi-dimensional data. For example, $\mathbf{x}$ denotes a vector, $\mathbf{X}$ denotes a matrix and $\boldsymbol{\mathcal{X}}$ is a 3D or third-order tensor. $\boldsymbol{\mathcal{X}}_{(i,j,t)}$ denotes the entry in the i-th row, j-th column and t-th slice. $\boldsymbol{\mathcal{X}}_{(:,:,t)}$ denotes the whole t-th slice, which is therefore a matrix. The ℓ_1 norm is the summation of absolute values of all entries, which applies to vector, matrix and tensor.

2.1. Batch mode

We introduce our deep sparsity architecture in the inverse order for easy understanding. Suppose we have a batch of grayscale images $\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_N \in \mathbb{R}^{w \times h}$ to be registered, where N denotes the total number of images.

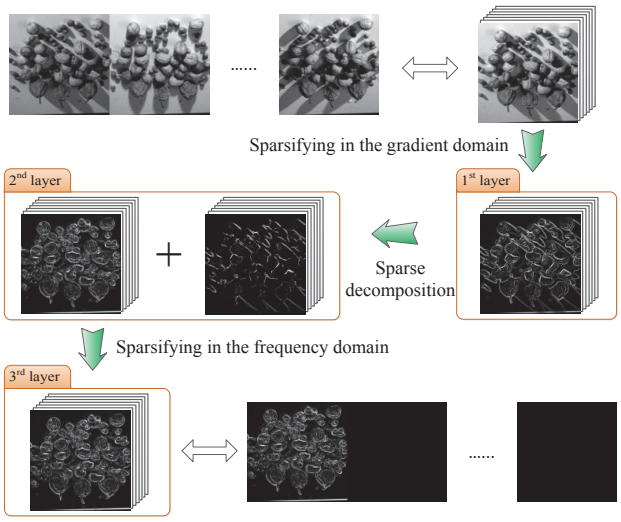

Figure 1. Deep sparse representation of the optimally registered images. First we sparsify the image tensor into the gradient tensor (1st layer). The sparse error tensor is then separated out in the 2nd layer. The gradient tensor with repetitive patterns are sparsified in the frequency domain. Finally we obtain an extremely sparse frequency tensor (composed of Fourier coefficients) in the 3rd layer.

First, we consider the simplest case that all the input images are identical and perturbed from a set of transformations $\tau = \{\tau_1, \tau_2, ..., \tau_N\}$.

We arrange the input images into a 3D tensor $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{w \times h \times N}$, with

$$\boldsymbol{\mathcal{D}}_{(:,:,t)} = \mathbf{I}_t, \quad t = 1, 2, ..., N, \tag{1}$$

After removing the transformation perturbations, the slices show repetitive patterns. Such periodic signals are extremely sparse in the frequency domain. Ideally the Fourier coefficients from the second slice to the last slice should be all zeros. We can minimize the ℓ_1 norm of the Fourier coefficients to seek the optimal transformations:

$$\min_{\boldsymbol{\mathcal{A}}, \tau} \ ||\mathcal{F}_N \boldsymbol{\mathcal{A}}||_1, \ s.t. \ \boldsymbol{\mathcal{D}} \circ \tau = \boldsymbol{\mathcal{A}}, \tag{2}$$

where $\mathcal{F}_N$ denotes the Fourier transform in the third direction.

The above model can be hardly used on practical cases, due to the corruptions and partial occlusions in the images. Similar as previous work [19], we assume the noise is negligible in magnitude as compared to the error caused by occlusions. Let $\boldsymbol{\mathcal{E}}$ be the error tensor. We can separate it from the image tensor if it is sparse enough. Similar, we use the ℓ_1 norm to induce sparseness:

$$\min_{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \tau} \ ||\mathcal{F}_N \boldsymbol{\mathcal{A}}||_1 + \lambda ||\boldsymbol{\mathcal{E}}||_1, \ s.t. \ \boldsymbol{\mathcal{D}} \circ \tau = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{E}}, \tag{3}$$

where $\lambda > 0$ is a regularization parameter.

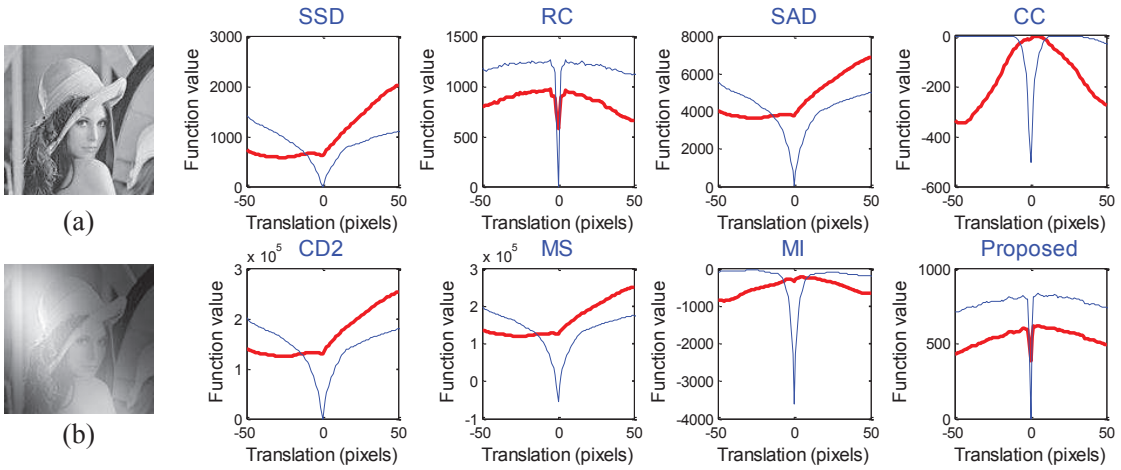

Figure 2. A toy registration example with respect to horizontal translation using different similarity measures (SSD [23], RC [17], SAD [23], CC [13], CD2 [6], MS [18], MI [26] and the proposed pair mode). (a) The Lena image (128 × 128). (b) A toy Lena image under a severe intensity distortion. Blue curves: registration between (a) and (a); red curves: registration between (b) and (a).

The above approach requires that the error $\boldsymbol{\mathcal{E}}$ is sparse. However, in many real-world applications, the images are corrupted with spatially-varying intensity distortions. Existing methods such as RASL [19] and t-GRASTA [9] may fail to separate these non-sparse errors. The last stage of our method comes from the intuition that the locations of the image gradients (edges) should almost keep the same, even under severe intensity distortions. Therefore, we register the images in the gradient domain:

$$\min_{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{E}},\tau} ||\mathcal{F}_N\boldsymbol{\mathcal{A}}||_1 + \lambda||\boldsymbol{\mathcal{E}}||_1,\ s.t.\ \nabla\boldsymbol{\mathcal{D}}\circ\tau = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{E}}, \quad (4)$$

where $\nabla\boldsymbol{\mathcal{D}} = \sqrt{(\nabla_x\boldsymbol{\mathcal{D}})^2 + (\nabla_y\boldsymbol{\mathcal{D}})^2}$ denotes the gradient tensor along the two spatial directions. This is based on a mild assumption that the intensity distortion fields of natural images often change smoothly.

With this rationale, the input images can be sparsely represented in a three layer architecture, which is shown in Fig. 1. We call it deep sparse representation of images. Comparing with existing popular low rank representation [19], our modeling has two major advantages. First, the low rank representation treats each image as a 1D signal, while our modeling exploits the spatial prior information (piece-wise smoothness) of natural images. Second, when the number of input images is not sufficient to form a low rank matrix, our method is still effective. Next, we will demonstrate how does our method register only two input images.

2.2. Pair mode

For registering a pair of images, our model can be simplified and the registration can be accelerated. After two-point discrete Fourier transform (DFT), the first entry is the sum and the second entry is the difference. The difference term is much sparser than the sum term when the two images have been registered. We can discard the sum term to seek a sparser representation. Let $\mathbf{I}_1$ be the reference image, and $\mathbf{I}_2$ be the source image to be registered. The problem (4) can be simplified to

$$\min_{\mathbf{A}_1,\mathbf{A}_2,\mathbf{E},\tau} ||\mathbf{A}_1 - \mathbf{A}_2||_1 + \lambda||\mathbf{E}||_1,$$
$$s.t.\ \nabla\mathbf{I}_1 = \mathbf{A}_1, \nabla\mathbf{I}_2\circ\tau = \mathbf{A}_2 + \mathbf{E}. \quad (5)$$

Both ℓ_1 norms in (5) implies the same property, i.e., sparseness of the residual image $\mathbf{E}$. Therefore, we can further simplify the above energy function:

$$\min_{\tau} ||\nabla\mathbf{I}_1 - \nabla\mathbf{I}_2\circ\tau||_1. \quad (6)$$

It's interesting that (6) is equivalent to minimizing the total variation (TV) of the residual image. The TV has been successfully utilized in many image reconstruction [12, 11] and non-rigid registration [14] problems.

We compare the proposed similarity measure with SSD [23], RC [17], sum-of-absolute value (SAD) [23], correlation coefficient (CC) [13], CD2 [6], MS [18] and mutual information (MI) [26] on a toy example. The Lena image is registered with itself with respect to the horizontal translations. The blue curves in Fig. 2 show the responses of different measures, all of which can find the optimal alignment at the zero translation. After adding intensity distortions and rescaling, the appearance of source image shown in Fig. 2(b) is not consistent with that of the original Lena image. The results denoted by the red curves show that only RC and the proposed pair mode can handle this intensity distortion while other methods fail.

3. Algorithms

3.1. Batch mode

Problem (4) is difficult to solve directly due to the non-linearity of the transformations τ. We use the local first order Taylor approximation for each image:

$$\nabla \mathbf{I}_t \circ (\tau_t + \triangle\tau_t) \approx \nabla \mathbf{I}_t \circ \tau_t + \boldsymbol{\mathcal{J}}_t \otimes \triangle\tau_t \tag{7}$$

for $t = 1, 2, ..., N$, where $\boldsymbol{\mathcal{J}}_t = \frac{\partial}{\partial \zeta}(\nabla \mathbf{I}_t \circ \zeta)|_{\zeta=\tau_t} \in \mathbb{R}^{w\times h\times p}$ when τ_t is defined by p parameters. The *Tensor-Vector Product* of the last term is defined by:

Definition 1. *Tensor-Vector Product*. The product of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and a vector $\mathbf{b} \in \mathbb{R}^{n_3}$ is a matrix $\mathbf{C} \in \mathbb{R}^{n_1\times n_2}$. It is given by $\mathbf{C} = \boldsymbol{\mathcal{A}} \otimes \mathbf{b}$, where $\mathbf{C}_{(i,j)} = \sum_{t=1}^{n_3} \boldsymbol{\mathcal{A}}_{(i,j,t)}\mathbf{b}_{(t)}$, for $i = 1, 2, ..., n_1$ and $j = 1, 2, ..., n_2$.

Based on this, the batch mode (4) can be rewritten as:

$$\begin{aligned} &\min_{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{E}},\triangle\tau} ||\mathcal{F}_N\boldsymbol{\mathcal{A}}||_1 + \lambda||\boldsymbol{\mathcal{E}}||_1, \\ &s.t.\ \nabla\boldsymbol{\mathcal{D}} \circ \tau + \boldsymbol{\mathcal{J}} \otimes \triangle\tau = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{E}}, \end{aligned} \tag{8}$$

This constrained problem can be solved by the augmented Lagrange multiplier (ALM) algorithm [19, 4]. The augmented Lagrangian problem is to iteratively update $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau$ and $\boldsymbol{\mathcal{Y}}$ by

$$\begin{aligned} (\boldsymbol{\mathcal{A}}^{k+1}, \boldsymbol{\mathcal{E}}^{k+1}, \triangle\tau^{k+1}) &= \arg\min_{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{E}},\triangle\tau} \mathcal{L}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau, \boldsymbol{\mathcal{Y}}), \\ \boldsymbol{\mathcal{Y}}^{k+1} &= \boldsymbol{\mathcal{Y}}^k + \mu^k h(\boldsymbol{\mathcal{A}}^k, \boldsymbol{\mathcal{E}}^k, \triangle\tau^k), \end{aligned} \tag{9}$$

where k is the iteration counter and

$$\begin{aligned} \mathcal{L}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau, \boldsymbol{\mathcal{Y}}) =& < \boldsymbol{\mathcal{Y}}, h(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau) > + ||\mathcal{F}_N\boldsymbol{\mathcal{A}}||_1 \\ &+ \lambda||\boldsymbol{\mathcal{E}}||_1 + \frac{\mu}{2}||h(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau)||_F^2, \end{aligned} \tag{10}$$

where the inner product of two tensors is the sum of all the element-wise products and

$$h(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau) = \nabla\boldsymbol{\mathcal{D}} \circ \tau + \boldsymbol{\mathcal{J}} \otimes \triangle\tau - \boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{E}}. \tag{11}$$

A common strategy to solve (9) is to minimize the function against one unknown at one time. Each of the subproblem has a closed form solution:

$$\begin{aligned} \boldsymbol{\mathcal{A}}^{k+1} &= \mathcal{T}_{1/\mu^k}(\nabla\boldsymbol{\mathcal{D}} \circ \tau + \boldsymbol{\mathcal{J}} \otimes \triangle\tau + \frac{1}{\mu^k}\boldsymbol{\mathcal{Y}}^k - \boldsymbol{\mathcal{E}}^k) \\ \boldsymbol{\mathcal{E}}^{k+1} &= \mathcal{T}_{\lambda/\mu^k}(\nabla\boldsymbol{\mathcal{D}} \circ \tau + \boldsymbol{\mathcal{J}} \otimes \triangle\tau + \frac{1}{\mu^k}\boldsymbol{\mathcal{Y}}^k - \boldsymbol{\mathcal{A}}^{k+1}) \\ \triangle\tau_t^{k+1} &= \boldsymbol{\mathcal{J}}_t^T \otimes (\boldsymbol{\mathcal{A}}_{(:,:,t)}^{k+1} + \boldsymbol{\mathcal{E}}_{(:,:,t)}^{k+1} - \nabla\boldsymbol{\mathcal{D}}_{(:,:,t)} \circ \tau \\ &\quad - \frac{1}{\mu^k}\boldsymbol{\mathcal{Y}}_{(:,:,t)}^k), \quad \text{for } t = 1, 2, ..., N \end{aligned} \tag{12}$$

where the $\mathcal{T}_\alpha()$ denotes the soft thresholding operation with threshold value α. In the third equation of (12), we use the *Tensor-Matrix Product* and *Tensor Transpose* defined as follows:

Definition 2. *Tensor-Matrix Product*. The product of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and a matrix $\mathbf{B} \in \mathbb{R}^{n_2\times n_3}$ is a vector $\mathbf{c} \in \mathbb{R}^{n_1}$. It is given by $\mathbf{c} = \boldsymbol{\mathcal{A}} \otimes \mathbf{B}$, where $\mathbf{c}_{(i)} = \sum_{j=1}^{n_2}\sum_{t=1}^{n_3} \boldsymbol{\mathcal{A}}_{(i,j,t)}\mathbf{B}_{(j,t)}$, for $i = 1, 2, ..., n_1$.

Definition 3. *Tensor Transpose*. The transpose of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the tensor $\boldsymbol{\mathcal{A}}^T \in \mathbb{R}^{n_3\times n_1\times n_2}$.

The registration algorithm for the batch mode is summarized in Algorithm 1. Let $M = w \times h$ be the number of pixels of each image. We set $\lambda = 1/\sqrt{M}$ and $\mu_k = 1.25^k\mu_0$ in the experiments, where $\mu_0 = 1.25/||\nabla D||_2$. For the inner loop, applying the fast Fourier transform (FFT) costs $\mathcal{O}(N \log N)$. All the other steps cost $\mathcal{O}(MN)$. Therefore, the total computation complexity of our method is $\mathcal{O}(N \log N + MN)$, which is significantly faster than $\mathcal{O}(N^2M)$ when applying SVD decomposition in RASL (if $M \gg N$).

Algorithm 1 Image registration via DSR - batch mode

input: Images $\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_N$, initial transformations $\tau_1, \tau_2, ..., \tau_N$, regularization parameter λ.

repeat

1) Compute $\boldsymbol{\mathcal{J}}_t = \frac{\partial}{\partial \zeta}(\nabla \mathbf{I}_t \circ \zeta)|_{\zeta=\tau_t}, t = 1, 2, ..., N$;

2) Warp and normalize the gradient images: $\nabla\boldsymbol{\mathcal{D}} \circ \tau = [\frac{\nabla \mathbf{I}_1\circ\tau_1}{||\nabla \mathbf{I}_1\circ\tau_1||_F}; ...; \frac{\nabla \mathbf{I}_N\circ\tau_N}{||\nabla \mathbf{I}_N\circ\tau_N||_F}]$;

3) Use (12) to iteratively solve the minimization problem of ALM: $\boldsymbol{\mathcal{A}}^*, \boldsymbol{\mathcal{E}}^*, \triangle\tau^* = \arg\min \mathcal{L}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{E}}, \triangle\tau, \boldsymbol{\mathcal{Y}})$;

4) Update transformations: $\tau = \tau + \triangle\tau^*$;

until Stop criteria

3.2. Pair mode

Similar as that in the batch mode, we have:

$$\nabla \mathbf{I}_2 \circ (\tau + \triangle\tau) \approx \nabla \mathbf{I}_2 \circ \tau + \boldsymbol{\mathcal{J}} \otimes \triangle\tau \tag{13}$$

where $\boldsymbol{\mathcal{J}} \in \mathbb{R}^{w\times h\times p}$ denotes the Jacobian. Thus, the pair mode (6) is to minimize the energy function with respect to $\triangle\tau$:

$$E(\triangle\tau) = ||\nabla \mathbf{I}_1 - \nabla \mathbf{I}_2 \circ \tau - \boldsymbol{\mathcal{J}} \otimes \triangle\tau||_1 \tag{14}$$

The ℓ_1 norm in (14) is not smooth. We can have a tight approximation for the absolute value: $|x| = \sqrt{x^2 + \epsilon}$, where ϵ is a small constant (e.g. 10^{-10}). Let $\mathbf{r} = \nabla \mathbf{I}_1 - \nabla \mathbf{I}_2 \circ \tau - \boldsymbol{\mathcal{J}} \otimes \triangle\tau$, and we can obtain the gradient of the energy function by the chain rule:

$$\nabla E(\triangle\tau) = \boldsymbol{\mathcal{J}}^T \otimes \frac{\mathbf{r}}{\sqrt{\mathbf{r} \circ \mathbf{r} + \epsilon}} \tag{15}$$

where $\circ$ denotes the Hadamard product. Note that the division in (15) is element-wise.

Gradient descent with backtracking is used to minimize the energy function (14), which is summarized in Algorithm 2. We set the initial step size $\mu^0 = 1$ and $\eta = 0.8$. The computational complexity of each iteration is $\mathcal{O}(M)$, which is much faster than $\mathcal{O}(M \log M)$ in RC when fast cosine transform (FCT) is applied [17]. Similar as the batch mode, we use the normalized images to rule out the trivial solutions. We use a coarse-to-fine hierarchical registration architecture for both the batch mode and pair mode.

Algorithm 2 Image registration via DSR - pair mode

input: $\mathbf{I}_1$, $\mathbf{I}_2$, $\eta < 1$, τ, μ^0.
repeat
1) Warp and normalize $\mathbf{I}_2$ with τ;
2) $\mu = \mu^0$;
3) Compute $\triangle\tau = -\mu\nabla E(\mathbf{0})$;
4) If $E(\triangle\tau) > E(\mathbf{0})$,
set $\mu = \eta\mu$ and go back to 3);
5) Update transformation: $\tau = \tau + \triangle\tau$;
until Stop criteria

4. Experimental results

In this section, we validate our method on a wide range of applications. We compare our batch mode with RASL [19] and t-GRASTA [9], and compare our pair mode with RC [17] and SSD [23]. One of the most important advantages of our method is its robustness and accuracy on natural images under spatially-varying intensity distortions. As shown in [17] and Fig. 2, SAD [23], CC [13], CD2 [6], MS [18], MI [26] are easy to fail in such cases. We do not include them in the following experiments. All experiments are conducted on a desktop computer with Intel i7-3770 CPU with 12GB RAM.

4.1. Batch image registration

To evaluate the performance of our batch mode, we use a popular database of naturally captured images [1]. We choose the four datasets with the largest lighting variations: "NUTS", "MOVI", "FRUITS" and "TOY". These datasets are very challenging to register, as they have up to 20 different lighting conditions and are occluded by varying shadows. Random translations on both directions are applied on the four datasets, which are drawn from a uniform distribution in a range of 10 pixels.

After registration on the "NUTS" dataset, the two components of each algorithm is shown in Fig. 3. RASL [19] and t-GRASTA [9] fail to separate the shadows and large errors, while we can successfully find the deep sparse representation of the optimally registered images. The average

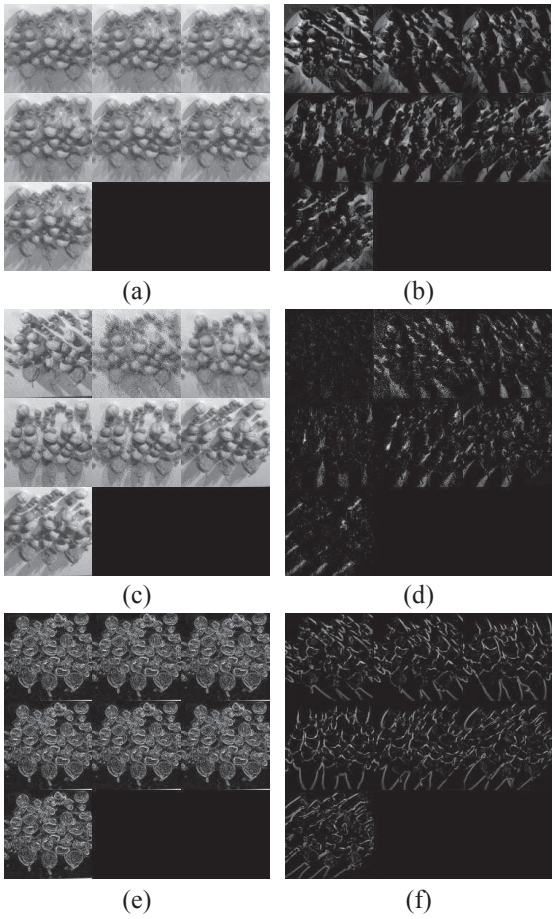

Figure 3. Batch image registration on the NUTS datasets. (a) The low rank component by RASL. (b) The sparse errors by RASL. (c) The subspace representation by t-GRASTA. (d) The sparse errors by t-GRASTA. (e) The visualization of $\boldsymbol{\mathcal{A}}$ by our method. (f) The sparse error $\boldsymbol{\mathcal{E}}$ by our method.

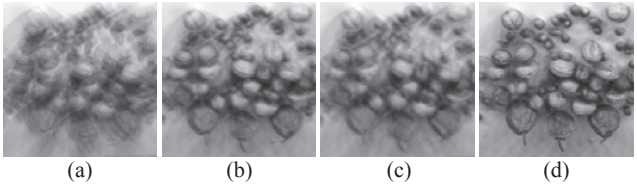

Figure 4. Registration results on the "NUTS" dataset. (a) The average image of perturbed images. (b) The average image by RASL. (c) The average image by t-GRASTA. (d) The average image by our method.

of perturbed images and results are shown in Fig. 4, where the average image by the proposed method has significantly sharper edges than those by the two existing methods. The quantitative comparisons on the four datasets are listed in Table 1 over 20 random runs. The overall average errors of our method are consistently lower than those of RASL

and t-GRASTA. More importantly, only our method can always achieve subpixel accuracy. For 20 images with size 128×128 pixels, the registration time is around 7 seconds for both RASL and our method, while t-GRAST costs around 27 seconds. RASL should be much slower on larger datasets due to the higher complexity of SVD, although we did not test.

	RASL	t-GRASTA	Proposed
NUTS	0.670/2.443	1.153/3.842	**0.061/0.488**
MOVI	0.029/ 0.097	0.568/ 2.965	**0.007/0.024**
FRUITS	0.050/0.107	1.094/4.495	**0.031/0.076**
TOY	0.105/ 0.373	0.405/2.395	**0.038/0.076**

Table 1. The mean/max registration errors in pixels of RASL, t-GRASTA and our method on the four lighting datasets. The first image is fixed to evaluate the errors.

We evaluate these three methods on the Multi-PIE face database [7]. This database contains 20 images of each subject captured at different illumination conditions. We add random artificial rotations in a range of 10° and translations in 10 pixels on the first 100 subjects from the Session 1. As the optimal alignment is not unique (e.g, all images shift by 1 pixel), we compare the standard derivation (STD) of the transformations after registration. Ideally, the STD should be zero when all the perturbations have been exactly removed. Fig. 5 shows the average registration results over 20 runs for each subject. Our method is more accurate than RASL and t-GRASTA for almost every subject.

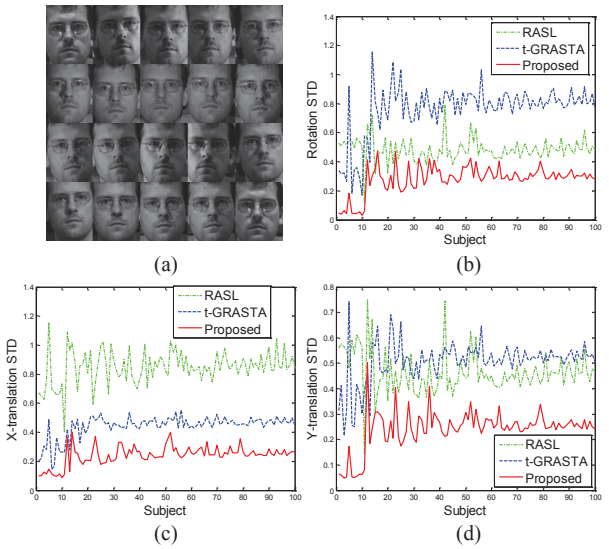

Figure 5. (a) An example input of the Multi-PIE image database. (b) The STD (in degrees) of rotations after registration. (c) The STD (in pixels) of X-translation after registration. (d) The STD (in pixels) of Y-translation after registration.

4.2. Pair image registration

4.2.1 Simulations

For quantitative comparisons, we evaluate SSD, RC and the proposed method on the Lena image with random intensity distortions (Fig. 2) and random affine transformations (with a similar range as the previous settings). The number of Gaussian intensity fields K is from 1 to 6. The reference image without intensity distortions is used as ground-truth. The root-mean-square error (RMSE) is used as the metric for error evaluation of both image intensities and transformations. We run this experiment 50 times and the results are plotted in Fig. 6. It can be observed that the proposed method is consistently more accurate than SSD and RC, with different intensity distortions.

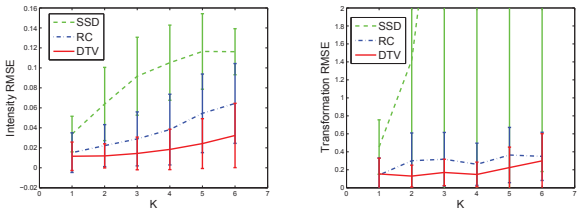

Figure 6. Registration performance comparisons with random transformation perturbations and random intensity distortions. (a) Intensity RMSE on the Lena image. (b) Transformation (affine) RMSE on the Lena image.

4.2.2 Multisensor remotely sensed image registration

Multisensor image registration is a key preprocessing operation in remote sensing, e.g., for image fusion [5], change detection. The same land objects may be acquired at different times, under various illumination conditions by different sensors. Therefore, it is very possible that the input images have significant dissimilarity in terms of intensity values. Here, we register a panchromatic image to a multispectral image acquired by IKONOS multispectral imaging satellite [21], which have been pre-registered at their capture resolutions. The multispectral image has four bands: blue, green, red and near-infrared, with 4 meter resolution (Fig. 7 (a)). The Pan image has 1 meter resolution (Fig. 7 (b)). The different image resolutions make this problem more difficult. From the difference image in Fig. 7 (c), we can observe that there exists misalignment in the northwest direction.

We compare our method with SSD [23] and RC [17], and the results are shown in Fig. 7 (d)-(f). It is assumed that the true transformation is formed by pure translation. Although we do not have the ground-truth, from the difference image, it can be clearly observed that our method can reduce the

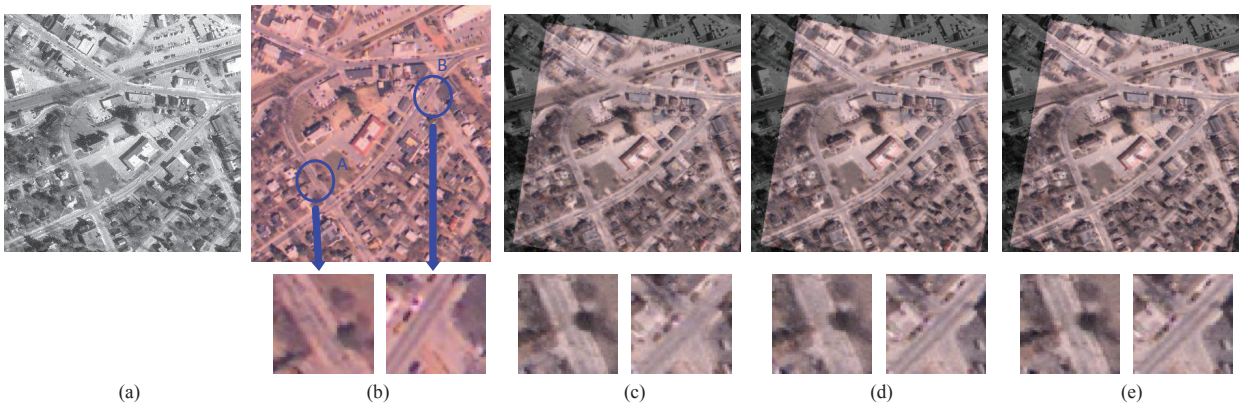

Figure 8. Registration of an aerial photograph and a digital orthophoto. From left to right, the images are: the reference image, the source image, the overlay by MATLAB, the overlay by RC, the overlay by our method. The second row shows the zoomed-in areas of streets A and B. Best viewed in $\times 2$ sized color pdf file.

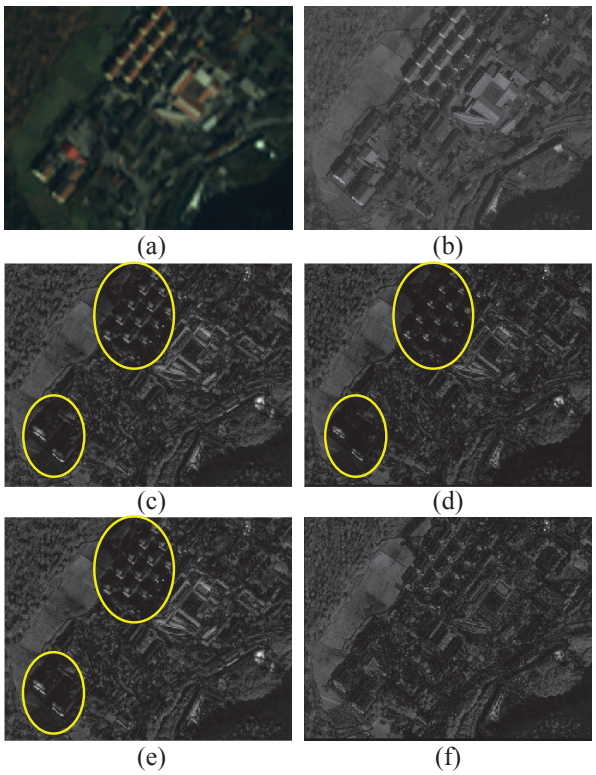

Figure 7. Registration of a multispectral image and a panchromatic image. (a) The reference image. (b) The source image. (c) The difference image before registration. (d) The difference image by SSD. (e) The difference image by RC. (f) The difference image by our method. Visible misalignments are highlighted by the yellow circles. Best viewed in $\times 2$ sized color pdf file.

misalignment. In contrast, SSD and RC are not able to find better alignments than the preregistration method.

We register an aerial photograph to a digital orthophoto. The reference image is the orthorectified MassGIS georegistered orthophoto [2]. The source image is a digital aerial photograph, which does not have any particular alignment or registration with respect to the earth. The input images and the results are shown in Fig. 8. MATLAB uses manually selected control points for registration, while RC and our registrations are automatic. At the first glance, all the methods obtain registration with good quality. A closer look shows that our method has higher accuracy than the others. In the source image, two lanes can be obviously observed in streets A and B. After registration and composition, street B in the result by MATLAB and street A in the result by RC are blurry due to the misalignment. Our method is robust to the local mismatches of vehicles.

5. Conclusion and discussion

In this paper, we have proposed a novel similarity measure for robust and accurate image registration. It is motivated by the deep sparse representation of the optimally registered images. The benefit of the proposed method is three fold: (1) compared with existing approaches, it can handle severe intensity distortions and partial occlusions simultaneously; (2) it can be used for registration of two images or a batch of images, with various types of transformations; (3) its low computational complexity makes it scalable to large datasets. We have conducted extensive experiments to test our method on multiple challenging datasets. The promising results demonstrate the robustness and accuracy of our method over the state-of-the-art batch registration methods and pair registration methods, respectively. We also show that our method can be used to reduce the registration errors in many real-world applications.

Due to the local linearization in the optimization, our method as well as all the compared methods cannot handle large transformations. However, this is not a big issue for

many real-world applications. For example, the remotely sensed images can be coarsely georegistered by their geographical coordinates. For images with large transformations, we can use the FFT-based algorithm [25] to coarsely register the images and then apply our method as a refinement. Therefore, we did not test the maximum amount of transformations that our method can handle. So far, the proposed method can only be used for offline registration. How to extend this method to the online mode is an interesting topic of future research.

References

[1] http://www.robots.ox.ac.uk/∼vgg/research/affine/. 5

[2] http://www.mathworks.com/help/images/register-an-aerial-photograph-to-a-digital-orthophoto.html. 7

[3] P. Blanc, L. Wald, T. Ranchin, et al. Importance and effect of co-registration quality in an example of pixel to pixel fusion process. In *2nd International Conference "Fusion of Earth Data: merging point measurements, raster maps and remotely sensed images"*, pages 67–74, 1998. 1

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011. 4

[5] C. Chen, Y. Li, W. Liu, and J. Huang. Image fusion with local spectral consistency and dynamic gradient sparsity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2760–2765, 2014. 6

[6] B. Cohen and I. Dinstein. New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition*, 35(2):455–463, 2002. 1, 3, 5

[7] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010. 6

[8] V. Hamy, N. Dikaios, S. Punwani, A. Melbourne, A. Latifoltojar, J. Makanyanga, M. Chouhan, E. Helbren, A. Menys, S. Taylor, et al. Respiratory motion correction in dynamic mri using robust data decomposition registration–application to dce-mri. *Medical image analysis*, 18(2):301–313, 2014. 2

[9] J. He, D. Zhang, L. Balzano, and T. Tao. Iterative grassmannian optimization for robust image alignment. *Image and Vision Computing*, 32(10):800–813, 2014. 1, 2, 3, 5

[10] J. Huang, X. Huang, and D. Metaxas. Simultaneous image transformation and sparse representation recovery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. 1

[11] J. Huang, S. Zhang, H. Li, and D. Metaxas. Composite splitting algorithms for convex optimization. *Computer Vision and Image Understanding*, 115(12):1610–1622, 2011. 3

[12] J. Huang, S. Zhang, and D. Metaxas. Efficient MR image reconstruction for compressed MR imaging. *Medical Image Analysis*, 15(5):670–679, 2011. 3

[13] J. Kim and J. A. Fessler. Intensity-based image registration using robust correlation coefficients. *IEEE Transactions on Medical Imaging*, 23(11):1430–1444, 2004. 3, 5

[14] Y. Li, C. Chen, J. Zhou, and J. Huang. Robust image registration in the gradient domain. In *Proceedings of the International Symposium on Biomedical Imaging (ISBI)*. 2015. 3

[15] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. Yuille, and Z. Tu. Robust L2E estimation of transformation for non-rigid registration. *IEEE Transactions on Signal Processing*, 63(5):1115–1129, 2015. 1

[16] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu. Robust point matching via vector field consensus. *IEEE Transactions on Image Processing*, 23(4):1706–1721, 2014. 1

[17] A. Myronenko and X. Song. Intensity-based image registration by minimizing residual complexity. *IEEE Transactions on Medical Imaging*, 29(11):1882–1891, 2010. 1, 2, 3, 5, 6

[18] A. Myronenko, X. Song, and D. J. Sahn. Maximum likelihood motion estimation in 3d echocardiography through non-rigid registration in spherical coordinates. In *Functional Imaging and Modeling of the Heart*, pages 427–436. 2009. 3, 5

[19] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2233–2246, 2012. 1, 2, 3, 4, 5

[20] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013. 1

[21] Space-Imaging. IKONOS scene po-37836. *Geoeye IKONOS Scene Data*, 2000. 6

[22] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas. Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Transactions on Medical Imaging*, 25(5):626–639, 2006. 1

[23] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006. 1, 3, 5, 6

[24] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot. Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1301–1312, 2008. 1

[25] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki. Robust FFT-based scale-invariant image registration with image gradients. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1899–1906, 2010. 1, 8

[26] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997. 1, 3, 5

[27] Y. Wu, B. Shen, and H. Ling. Online robust image alignment via iterative convex optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1808–1814, 2012. 1, 2

[28] Y. Zheng, E. Daniel, A. A. Hunter III, R. Xiao, J. Gao, H. Li, M. G. Maguire, D. H. Brainard, and J. C. Gee. Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix. *Medical image analysis*, 18(6):903–913, 2014. 1

[29] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003. 1