Computers in Biology and Medicine 73 (2016) 56–70

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Medical image registration using sparse coding of image patches

Maryam Afzali [a], Aboozar Ghaffari [a], Emad Fatemizadeh [a,*], Hamid Soltanian-Zadeh [b,c,d]

[a] Department of Electrical Engineering, Biomedical Signal and Image Processing Laboratory (BiSIPL), Sharif University of Technology, Tehran, Iran
[b] Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[c] School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[d] Image Analysis Laboratory, Departments of Radiology and Research Administration, Henry Ford Health System, Detroit, MI, USA

ARTICLE INFO

Article history:
Received 6 January 2016
Received in revised form
27 February 2016
Accepted 28 March 2016

Keywords:
Non-rigid image registration
Analysis K-SVD algorithm
Sparse coding

ABSTRACT

Image registration is a basic task in medical image processing applications like group analysis and atlas construction. Similarity measure is a critical ingredient of image registration. Intensity distortion of medical images is not considered in most previous similarity measures. Therefore, in the presence of bias field distortions, they do not generate an acceptable registration. In this paper, we propose a sparse based similarity measure for mono-modal images that considers non-stationary intensity and spatially-varying distortions. The main idea behind this measure is that the aligned image is constructed by an analysis dictionary trained using the image patches. For this purpose, we use "Analysis K-SVD" to train the dictionary and find the sparse coefficients. We utilize image patches to construct the analysis dictionary and then we employ the proposed sparse similarity measure to find a non-rigid transformation using free form deformation (FFD). Experimental results show that the proposed approach is able to robustly register 2D and 3D images in both simulated and real cases. The proposed method outperforms other state-of-the-art similarity measures and decreases the transformation error compared to the previous methods. Even in the presence of bias field distortion, the proposed method aligns images without any preprocessing.

1. Introduction

Image registration is a fundamental image processing approach that has many applications in medical imaging [1]. It is an inevitable step in medical image analysis tasks [2] like group analysis and atlas construction [3]. Its goal is to align two or more images into the same coordinate system so that the aligned images can be directly compared or combined. For instance, in surgical planning, it is useful to do image registration to precisely locate a region of interest [4–9]. There are different methods for image registration, which can be categorized into two major groups: feature-based and intensity based. Feature based methods rely on the landmarks extracted from the floating and reference images [10–15], [16]. The accuracy of this class of methods depends on the accuracy of feature extraction. Accurate feature extraction helps in applications like object tracking in video sequences [15]. On the other hand, intensity based approaches [17–20] directly use the intensities of the two images to find the transformation.

A key component of image registration algorithms is the similarity measure they use. An appropriate similarity measure is expected to generate a close spatial alignment. Due to the variations in the input images, finding an appropriate similarity measure is challenging. In many real-world applications, the images to be registered are acquired at different times under various illumination conditions. As a result, their intensity field may vary significantly. For instance, slow varying intensity bias fields often exist in brain magnetic resonance images (MRI) [21]. However, many existing similarity measures do not consider these intensity variations. Well-known intensity-based similarity measures include: Sum of Absolute Differences (SAD) [22], Correlation Coefficient (CC) [23], Sum of Squared Differences (SSD) [24] and Mutual Information (MI) [21,25]. The main advantage of these measures is their easy implementation. However, they consider the intensity relationship of the two voxels independent of the other voxels, while they are related. Consequently, in the presence of intensity distortions, these measures are not robust. To solve this problem, different techniques have been proposed.

In some researches, [26–28], local similarity measures are used. In these measures, the intensity distortion in the neighborhood of a voxel is considered constant. Although the results of these local methods are much better than the global methods, these methods do not lead to appropriate results in the presence of intensity distortion and outlier.

Some probabilistic models are used in [29–31]. The key point of these methods is the definition of local intensity interactions.

* Corresponding author.
E-mail addresses: afzali@ee.sharif.edu (M. Afzali), aboozar412@gmail.com (A. Ghaffari), fatemizadeh@sharif.edu (E. Fatemizadeh), hszadeh@ut.ac.ir, hsoltan1@hfhs.org (H. Soltanian-Zadeh).

http://dx.doi.org/10.1016/j.compbiomed.2016.03.022
0010-4825/

Another strategy to solve the problem of intensity distortion is to do the intensity correction and image registration simultaneously [32,33]. Friston et al. [32] align images using SSD and correct the intensity distortions with a convolution filter and non-linear intensity transformation at the same time. Modersitzki and Wirtz [33] define the multiplicative intensity correction function with a total variation norm (TV) regularization. The limitation of these methods is that they need an accurate model for intensity correction. Myronenko and Song in [34] propose residual complexity (RC) that analytically solves the intensity correction field. RC is one of the best measures for registering two images corrupted by intensity distortion which uses the Discrete Cosine Transform (DCT) to sparsify the residual of the two images. Farnia et al. [35] propose a wavelet-based RC for multi-modal image registration.

A Sparse Induced Similarity Measure (SISM) is proposed in [36], which considers spatial dependences between pixels. SISM is a sparse measure whose dictionary is made of DCT and wavelet bases. Also, Ghaffari et al. [37] propose the Robust Huber Similarity Measure (RHSM) for image registration. Also, they proposed the Rank-Induced Similarity Measure (RISM) based on nonlinear and low-rank matrix decomposition [38]. The distance measure using ℓ_2-norm was proposed by Zhang et al. [39] and ℓ_1, ℓ_2 mixed norms by Yuan et al. [40]. Also, there are some works on metric learning or similarity learning in the sparse domain [41–43].

We propose a new similarity measure based on the definition of the sparse coding with the "Analysis K-SVD" algorithm [44,45]. We know that a spatially varying intensity distortion has a sparse representation in a transform domain [36]. We utilize an algorithm that learns dictionary from image patches. Indeed, it uses the image patches for dictionary learning. Using the image patches instead of image voxels will consider the correlation between image intensities in the proposed method and will lead to a more sparse representation compared to the fix dictionary methods. The main reason for using the sparse coding in this work for image registration is the presence of spatially-varying intensity distortions in the images.

The rest of the paper is organized as follows. Section 2 describes a background on sparse coding. Section 3 presents the proposed method. Experimental results are provided in Section 4 while discussions and conclusions are given in Section 5.

2. Background: Sparse representations and analysis dictionary learning

In sparse representation, our goal is to find a mapping from signal space $x_i \in R^p$ to a representation space $\alpha_i \in R^m$ where $m > p$ such that the representation in the new space is the sparsest possible. Using this mapping, we can describe a signal x_i as $x_i = \mathbf{D}\alpha_i$ where $\mathbf{D}$ is an underdetermined $p \times m$ matrix called dictionary or design matrix. To find the sparse representation, we solve the following problem:

$$\hat{\alpha}_i = \arg\min_{\alpha_i} \|\alpha_i\|_0 \quad s.\ t. \quad \|\mathbf{D}\alpha_i - x_i\|_2^2 \le \varepsilon^2 \tag{1}$$

where $\|\alpha_i\|_0$ counts the non-zero elements in α, and ε is an error threshold. Clearly, in the case of $\varepsilon = 0$, the exact equation $\mathbf{D}\alpha_i = x_i$ holds. For noisy observation of x_i, ε is set as the noise level. Since problem (1) is an NP-hard problem, the representation is approximated by a pursuit algorithm such as Orthogonal Matching Pursuit (OMP) [45–48]. The aforementioned learning method is based on the synthesis model while there is another method named analysis model [44]. In this method the signal is multiplied by analysis dictionary, to create sparse representation of the signal. The optimized backward greedy algorithm is used for finding the analysis dictionary [44].

There are many methods which use a pre-defined dictionary to represent a signal in the sparse domain. For example, DCT or wavelet bases can be used to construct the dictionary in some applications (SISM).

On the other hand, learning the dictionary $\mathbf{D}$ from a set of signals $\{x_i\}_{i=1}^N$ results in a sparser representation than the one obtained using a pre-defined dictionary.

To use sparse representation in image processing, we divide the image into overlapping patches. Then, we represent each patch sparsely using a learned analysis dictionary. In this paper, we use the idea of the analysis K-SVD algorithm [44,45], which divides the image of size $N \times N$ into overlapping patches of size $\sqrt{N} \times \sqrt{N}$.

3. Method

Our goal is to develop a registration method that works in the presence of spatially varying intensity distortion. We assume the following intensity relationship between two images:

$$\mathbf{R} = \mathbf{F}(\mathbf{T}) + \mathbf{S} + \mathbf{n} \tag{2}$$

where $\mathbf{S}$ is an intensity correction field, $\mathbf{R}$ is the reference image (corrupted by intensity distortion), $\mathbf{F}$ is the floating image (corrupted by intensity distortion), $\mathbf{n}$ is a zero-mean white Gaussian noise with the covariance matrix $\Sigma_n = \sigma^2\mathbf{I}$, and $\mathbf{T}$ is a spatial transform.

Our proposed method for sparse registration contains three steps:

1. The reference and float images that have intensity distortion (one of them or both) are used to learn the analysis dictionary and find sparse coefficients using analysis K-SVD.
2. The proposed similarity measure, based on the sparse representation of the references and floating images, is calculated.
3. The proposed similarity measure is used to register the images based on the free form deformation (FFD) method [49].

In the rest of this section, we explain the above steps.

3.1. Analysis dictionary learning and sparse coding

The analysis model [44] for the signal $x_i \in R^p$ uses the possibly redundant analysis dictionary $\mathbf{\Omega} \in R^{m \times p}$ (redundancy here implies $m \ge p$) and assumes that the analysis representation vector $\mathbf{\Omega}x_i$ should be sparse. The co-sparsity ℓ of the analysis model is defined as the number of zero elements in the vector $\mathbf{\Omega}x_i$, $\|\mathbf{\Omega}x_i\|_0 = p - \ell$.

In the synthesis model [45], the representation α_i is obtained by a complex and non-linear pursuit process that seeks (or approximates) the sparsest solution to the linear system of equations $x_i = \mathbf{D}\alpha_i$. This representation is sparse, $\|\alpha_i\|_0 = k < < m$. The signal x_i is characterized by the k non-zero elements in the representation vector α_i and their associated atoms define the subspace this signal belongs to. The dimension of this subspace equals k and as mentioned before, it is small in comparison to the signal dimension m.

In contrast, in the analysis model, the computation of the representation is trivial and obtained by the multiplication $\mathbf{\Omega}$ by x_i, $\mathbf{\Omega}x_i$. In this model, we put an emphasis on the zeros of $\mathbf{\Omega}x_i$ and define the co-support Λ of x_i as the set of rows that are orthogonal to it. For a given analysis dictionary $\mathbf{\Omega}$, we define the co-rank of a signal x_i with co-support Λ as the rank of $\mathbf{\Omega}_\Lambda$.