Computers in Biology and Medicine 76 (2016) 69–79

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Local structure orientation descriptor based on intra-image similarity for multimodal registration of liver ultrasound and MR images

Minglei Yang [a], Hui Ding [a], Jingang Kang [b], Longfei Cong [b], Lei Zhu [b], Guangzhi Wang [a,*]

[a] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Room C249, Beijing 100084, China
[b] Beijing Shen Mindray Medical Electronics Technology Research Institute Co. Ltd, 3 F, Building5, No.8 Chuangye Road, Shangdi Haidian District, Beijing 100085, China

ARTICLE INFO

Article history:
Received 21 April 2016
Received in revised form
11 June 2016
Accepted 24 June 2016

Keywords:
Local structure orientation
Multimodal registration
US and MR
Fusion imaging
Liver

ABSTRACT

Purpose: Ultrasound (US)-magnetic resonance (MR) fusion imaging is a profitable tool for image-guided abdominal diagnosis and biopsy. However, the automatic registration of liver US and MR images remains a challenging task. An effective local structure orientation descriptor (LSOD) for use in registering multimodal images is proposed in this study.

Methods: LSOD utilizes a normalized similarity distance vector of intra-image patch pairs to extract intensity change orientations from intensity value changes in a local area. The multimodal similarity measure is then derived using the LSOD vector difference. Experiments were performed on simulated US and liver 2D US–3D MR images from a phantom, two healthy volunteers, and seven patients.

Results: Using the LSOD-based method, the root-mean-square target registration errors (RMS-TREs) were 1.76 ± 1.90 mm/2.03 ± 0.84 mm in phantom/clinical experiments. All of the results outperformed those obtained using modality independent neighborhood descriptor (MIND)- and linear correlation of linear combination (LC^2)-based methods (phantom/clinical: 5.23 ± 3.35 mm/4.32 ± 3.63 mm and 9.79 ± 5.03 mm/6.29 ± 3.85 mm, respectively). The registration cover range for all subjects of the LSOD-based method was 9.16 mm, which was larger than those of the MIND- and LC^2-based methods (5.06 and 5.12 mm, respectively).

Conclusions: The results demonstrated that the LSOD-based registration method could robustly register 2D US and 3D MR images of different liver sections with acceptable accuracy for clinical requirements. This approach is useful for the practical clinical application of the US–MR fusion imaging technique.

1. Introduction

The ultrasound (US)-magnetic resonance (MR) fusion imaging technique aligns and then fuses a pre-acquired three-dimensional (3D) MR dataset and real-time two dimensional (2D) US images obtained from a single patient. The fused images containing complementary information from both modalities are used to guide clinicians in diagnosis, intervention, and post-operative evaluation [1]. This method improves organ lesion identifiability and locating precision by integrating the high lesion visibility and rich spatial information of the 3D MR images with the real-time imaging ability of 2D US imaging. It has been employed in minimally invasive surgeries in the liver, prostate, and kidney [2]. A crucial procedure for US-MR fusion imaging is the online registration of the 2D US and 3D MR images. This registration in current clinical applications is usually achieved through the tedious manual alignment of anatomical landmarks in cross-modal images [3–6]. The main challenges in the automatic registration of liver US and MR images include the absence of a functional relationship between the intensities of the corresponding landmarks of the two modalities, and the poor quality of the multimodal images. This absence of an intensity relationship results from the different physical principles that these two imaging methods employ. The quality of the multimodal images would be adversely affected by noise or pathologies. Registering US images of the liver is especially challenging, as the images are often degraded by speckle noise and acoustic scattering.

Given the absence of a functional relationship between the intensity values, the corresponding images of different modalities often do not correlate well. This problem could be addressed to a certain extent using geometric feature-based registration approaches [7]. The main advantage of feature-based approaches is their robustness against intensity distortions. However, these approaches require the segmentation of organs and/or vessels in

* Corresponding author.
E-mail addresses: yangml122@gmail.com (M. Yang), dinghui@tsinghua.edu.cn (H. Ding), kangjingang@mindray.com (J. Kang), conglongfei@mindray.com (L. Cong), zhulei@mindray.com (L. Zhu), wgz-dea@tsinghua.edu.cn (G. Wang).

http://dx.doi.org/10.1016/j.compbiomed.2016.06.025
0010-4825/

both the US and MR images. They are thus not practicable in our fusion imaging scenario because of the difficulties in automatic liver/vessel segmentation, especially in poor-quality US images.

In intensity-based registration approaches, mutual information (MI) and correlation ratio (CR) were commonly utilized to measure the similarity between US and MR images. Roche et al. [8] integrated the intensity and gradient of MR images to compute a CR-based metric for brain 3D US and 3D MR images. Chan et al. [9] registered 3D US and 3D MR images of the carotid using MI. These approaches were prone to fail due to US image noise and local intensity distortion in cross-modal images. Penney et al. [10] mapped multimodal images to intermediate vessel possibility images and then used MI to measure their similarity. However, this method was dependent on manual vessel segmentation to establish a training dataset. Zhang et al. [11] used local phases of cardiovascular US and MR images to measure their similarity. Fuerst et al. [12] combined the intensity and gradient of MR images adaptively to make them comparable with real US images and then adopted the linear correlation of linear combination (LC^2) metric to register brain 3D US and 3D MR images for neurosurgery.

All of the abovementioned intensity-based methods derive a scalar characteristic from the image intensity, but designing a robust and distinctive similarity metric that is based only on the scalar characteristic for US and MR images might be difficult [13]. To address this difficulty, some researchers attempted to extract high-dimensional structure information from intensities in local regions. Lange et al. [14] used the normalized gradient field (NGF) to describe the local image structure. Zhuang et al. [15] used normal vector information (NVI) to register brain 3D MR images. The NVI was also computed from the normalized gradient vector of gray values. The gradient vector might be affected by severe noise. De Nigris et al. [16] used gradient orientation alignment (GOA) to register brain 3D US and 3D MR images, while Fuerst et al. [12] found that LC^2 outperformed GOA in brain US and MR image registration experiments. Hu et al. [17] adopted vector alignment (Ch. 4.2) to register 3D US and 3D MR images of the prostate, which still required manually identifying point features in several US slices and extracting the prostate surface from 3D MR images. Heinrich et al. [13] proposed a self-similarity-based modality-independent neighborhood descriptor (MIND) to describe the local structure, however this approach might not sufficiently consider the cross-modal local image intensity distortion (e.g., acoustic shadow-induced feature weakening).

Just as the local macro structure is degraded in poor-quality images, the subtle local texture is also sensitive to local noise. The pixel/voxel intensity value alone cannot adequately represent the local anatomic structure. Fortunately, these local structures can be described in a larger view with the help of their neighboring elements. This phenomenon is similar to human vision: the profile of one object can be observed from a distance, whereas its details can be observed from a close-up. In our case, there exist no direct correlations between corresponding tissue intensities in US–MR images, and variable noises are present in both modalities. However, anatomical structures, such as the capsules and vessels of organs (liver and kidney), still share similar appearances (e.g., outlines) when local intensity changes are observed in a relatively macroscopic view. What is more, these appearances can be expressed in comparable form even in cross-modal images. We focus, based on these understandings, on the intensity change orientation (i.e., structure orientation) instead of the intensity value change to describe the modality-independent similarity in appearance. This concept is a combination of feature- and intensity-based methods.

To this end, this paper proposes a local structure orientation descriptor (LSOD) based on intra-image similarity to depict multimodal image local structure information to overcome the considerable differences in cross-modal images. A similarity metric based on LSOD vectors is then derived to achieve the automatic registration of liver 2D US and 3D MR images. Experiments were performed on simulated US images, a liver phantom, healthy volunteers, and patients to evaluate the performance of the proposed method.

2. Methods

2.1. Local structure orientation descriptor

The intensity change orientation of a landmark in a medical image can be described by the intensity value changes of its neighboring region. Considering that the gradient orientation field cannot steadily describe the orientation of the local structure in poor-quality noisy images, we consider a greater number of neighboring elements of a current point.

To extract the change orientation, a high-dimensional structure vector is first constructed. For an arbitrary pixel/voxel x in an image I (Fig. 1), its $D-$ dimensional structure vector can be derived as:

$$\overrightarrow{f(x)} = \left[d(x, x+1), \cdots, d(x, x+i), \cdots, d(x, x+D) \right]^T \quad (1)$$

where $d(x, x+i)$ is the similarity distance between x and its ith neighboring pixel/voxel. The neighboring region is defined as centered at x and extending to $-R$ and $+R$ in each direction. $D = (2R+1)^n$, and n is the image dimension ($R=2$, $n=2$ in Fig. 1).

To unify the form of the identical anatomical structure that displays different degrees of intensity changes in multimodal images (e.g., dark- and bright-colored blood), the intensity value changes of a pixel/voxel and its neighbors were adopted on the image patch-level rather than merely on the pixel-/voxel-level. Usually, it can be assumed that the direction of the intensity changes (decrease or increase) at corresponding locations in multimodal images may point to either the same or the opposite

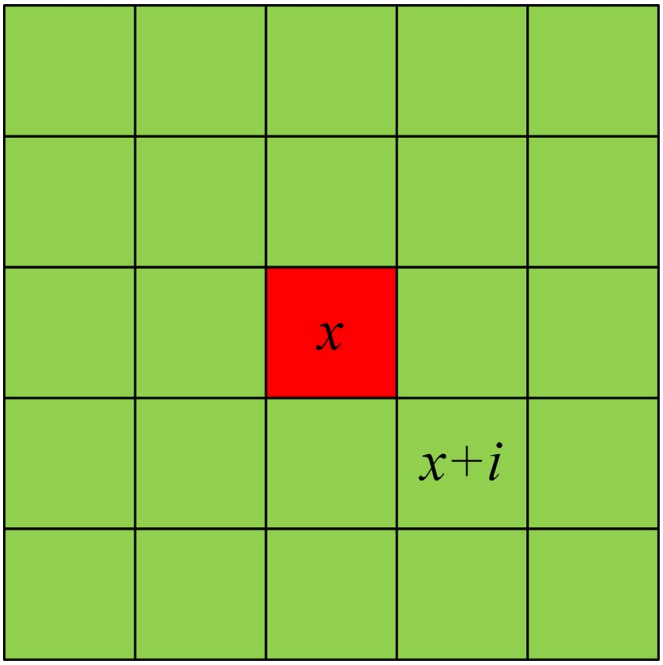

Fig. 1. Schematic of the structure vector of x. The similarity distance between x and its ith neighboring pixel/voxel $x+i$ acts as the ith feature. All elements in the neighboring region together construct the D-dimensional structure vector. The neighboring region is defined as centered at x and extending to $-R$ and $+R$ in each direction. ($R=2$, $D=25$ herein).