Biomedical Signal Processing and Control 19 (2015) 68–76

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Robust point matching method for multimodal retinal image registration

Gang Wang, Zhicheng Wang*, Yufei Chen, Weidong Zhao

CAD Research Center, Tongji University, No. 4800, Cao'an Highway, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 15 December 2014
Received in revised form 27 February 2015
Accepted 10 March 2015
Available online 11 April 2015

Keywords:
Image registration
Multimodal retinal image
Robust point matching
PIIFD
SURF

ABSTRACT

In this paper, motivated by the problem of multimodal retinal image registration, we introduce and improve the robust registration framework based on partial intensity invariant feature descriptor (PIIFD), then present a registration framework based on speed up robust feature (SURF) detector, PIIFD and robust point matching, called SURF–PIIFD–RPM. Existing retinal image registration algorithms are unadaptable to any case, such as complex multimodal images, poor quality, and nonvascular images. Harris-PIIFD framework usually fails in correctly aligning color retinal images with other modalities when faced large content changes. Our proposed registration framework mainly solves the problem robustly. Firstly, SURF detector is useful to extract more repeatable and scale-invariant interest points than Harris. Secondly, a single Gaussian robust point matching model is based on the kernel method of reproducing kernel Hilbert space to estimate mapping function in the presence of outliers. Most importantly, our improved registration framework performs well even when confronted a large number of outliers in the initial correspondence set. Finally, multiple experiments on our 142 multimodal retinal image pairs demonstrate that our SURF–PIIFD–RPM outperforms existing algorithms, and it is quite robust to outliers.

1. Introduction

Image registration is an important element in the fields of computer vision, pattern recognition, and medical image analysis. In this problem, two or more images are aligned together in a same spatial axis to receive a comprehensive understanding. In this paper, we focus on digital retinal images which are widely used to diagnose varieties of diseases, such as diabetic retinopathy, glaucoma, and age-related macular degeneration [1,2]. Then, using the computer-assisted retinal image registration technique is helpful to assist doctors to diagnose diseases and make treatment planning. There are four main retinal registration applications: mono-modal registration, multimodal registration, temporal registration, and multi-images fusion. Mono-modal and multimodal retinal images are captured by the same sensor (e.g. fundus camera) and different sensors (e.g. red-free and fluorescein angiography) at the same time, respectively, while temporal retinal images are captured at different times. These applications align images to create a wider view, and integrated data information.

* Corresponding author. Tel.: +86 13636524116; fax: +86 02165983989.
E-mail addresses: gwang.cv@gmail.com (G. Wang), zhichengwang@tongji.edu.cn (Z. Wang), yufeichen@tongji.edu.cn (Y. Chen), wd@tongji.edu.cn (W. Zhao).

http://dx.doi.org/10.1016/j.bspc.2015.03.004
1746-8094/

Recently, many related registration approaches have been proposed for retinal image registration, can be classified into three classes: area-based, feature-based, and hybrid approaches, typically.

The area-based approaches are widely used in image registration, they mainly use a certain similarity metric, such as mutual information (MI) [3–5], cross correlation (CC) [6], entropy correlation coefficient (ECC) [7], and phase correlation [8,9], to match the intensity difference of image pairs. In order to minimize the measures of match, some optimizations are applied, such as simulated annealing [10] and genetic algorithms. However, there are several shortcomings. (1) The huge searching space depends upon the complication of the transformation models. (2) The metric of similarity is always disturbed by nonoverlapping areas when faced with too small overlaps. (3) The optimization often meets local minima when handling with high order transformations, and it may have huge searching space to fall in a computational bottleneck. What is more, the performance of area-based approaches degrades when confronted with illumination, content, and texture changes.

Feature-based approaches extract several features, such as the bifurcations of retinal vasculature, fovea, optic disc, and corners, whose number is much less than the number of pixels, and they are more appropriate for retinal image registration. Feature extraction and transformation estimation are two key components of these approaches. Typically, bifurcations [11–15], fovea, and optic disc [16,17] are common features in retinal image registration. Bifurcations are invariant feature to intensity, scale, rotation, and illumination variations, and dependent of vasculature detection [11]. The vascular tree is detected and bifurcations are labeled with surrounding vessel orientations. Then angle-based invariant is used to give a probability for every matching bifurcation pairs. However, it is difficult to extract bifurcations, fovea, and optic disc in poor quality, and unhealthy retinal images [18]. Thus, feature-based approaches based on an assumption that certain features are easy to be extracted. The bifurcations are used as landmarks, but some bifurcations with more than one correspondence, then the transformation can be estimated by a hierarchical strategy which makes the registration approach robust to unmatchable features and mismatches between image pairs [12]. Local features, such as Harris corner [19], scale invariant feature transform (SIFT) [20–22], speed up robust feature (SURF) [23], are also widely used general features and easier to extract than bifurcations, however these feature descriptors are not appropriate for multimodal registration. More precisely, SIFT and SURF descriptors are designed for mono-modal retinal image registration [24]. Shape context (SC) [25] only uses the locations of feature points to describe point set in log-polar histogram bins, it is rotation invariant, scale invariant, and affine invariant, but it is highly sensitive to outliers.

Hybrid approaches integrate area-based with feature-based approaches to improve the registration performance. For instance, [7] combines mutual information technique and bifurcations to register retinal images. Chen et al. [18] presented a partial intensity invariant feature descriptor (PIIFD) for multimodal registration, even for poor quality images. It is a hybrid area-feature descriptor due to the surrounding area of each corner point is used to extract structural outline. However, the Harris-PIIFD registration framework cannot detect more repeatable and scale invariant key points, and its sensitive to large amount of mismatches. Ghassabi et al. [26] analyzed the problems related to SIFT, and proposed an uniform and robust scale invariant transform feature extraction (UR-SIFT) to instead Harris detector, and obtained an efficient UR-SIFT-PIIFD registration approach.

From the angle of point matching, iterative closest point (ICP) [27] is widely used to register retinal images. Stewart et al. [28] proposed dual-bootstrap ICP. There are three steps in each bootstrap region, such as refining the transformation estimation, expanding the bootstrap region, and using higher order transformation model. Yang et al. [29] used the dual-bootstrap ICP algorithm to refine each estimate, and proposed the generalized dual-bootstrap ICP (GDB-ICP) algorithm. Edge-driven dual-bootstrap ICP (ED-DB-ICP) [30] combines SIFT key points and vascular features to register multimodal retinal images. All of those approaches only need one initial correct match to run iterative registering process successfully. However, their performance degrades when faced with poor quality images. Deng et al. [31] introduce graph matching method for retinal image registration, and they also combined graph-based matching and ICP to generate a registration framework called GM-ICP. The methods require sufficient feature points to obtain efficient performance. Although there are many approaches in retinal image registration, several challenges still exist in retinal image registration. Firstly, how to extract reliable, repeatable, and distinctive features in different modal retinal images. Secondly, how to find correspondences between multimodal retinal pairs, i.e. how to design robust descriptor for matching control point candidates. Thirdly, how to remove outliers, due to the initial matching correspondences are contaminated by outliers which highly impact the registration accuracy. As mentioned earlier, the Harris-PIIFD is designed for multimodal images, and poor quality images. It can register multimodal images successfully, but the fusion images contain some degree of dislocation and ghost, two main corresponding problems are features repeatable and outlier removal, due to the limitation of the Harris corner detector, and outlier rejection strategy, respectively.

In this paper, we improve the Harris-PIIFD registration framework using SURF key points to solve the features repeatability, and propose a novel robust point matching algorithm to reject outliers and estimate transformation robustly. The improved SURF–PIIFD is useful to extract repeatable, rotation, scale invariant, intensity and affine partial invariant local features. In outliers rejection process, we assume that the inliers satisfy a single Gaussian distribution, then we search for the optimal mapping function in reproduced kernel Hilbert space with a Gaussian radial basis kernel function. A novel low-rank Gram matrix approximation is proposed to construct control points to speed up our algorithm. Thus, the described robust automatic multimodal retinal image registration framework named SURF–PIIFD–RPM.

The rest of the paper is organized as follows: In Section 2, we introduce the improved retinal image registration framework. In Section 3, we present the improved SURF–PIIFD of feature descriptor. In Section 4, we devote the proposed robust point matching for outliers removing and transformation estimation. In Section 5, we describe the experimental settings and report the results. In Section 6, we give a discussion and conclusion.

2. Retinal image registration framework

In this paper, we concentrate on the hybrid area-feature based registration method for multimodal retinal images. Our improved multimodal retinal image registration framework, as showed in Fig. 1, based on SURF–PIIFD and RPM contains the following four main parts:

(1) Locate local feature points by SURF detector.
(2) Extract feature descriptor based on PIIFD.
(3) Feature matching and mismatches removing using RPM.
(4) Estimate the transformation using weighted least-squares.

Note that image preprocessing is applied before extracting local feature descriptor, since the SURF detector can detect key points based on color images directly. Then we select green component from the input RGB image format, and scale the intensities of them to the full eight bits intensity range [0, 255]. In order to reduce the sensitivity of algorithm parameters and the different scale, the Harris-PIIFD framework suggests zooming out the input image to a fixed size. In our improved framework, we use the original image size for lossless intensities and repeatable key points.

Following the matching algorithm of the Harris-PIIFD, we use the bilateral matching based on the unilateral best-bin-first (BBF) method [32]. In this way, we can obtain more accurate matching point pairs, though losing some pairs. In terms of our following robust point matching method, the threshold of the nearest neighbor criterion is set to 0.96 in this paper for getting more candidate matching pairs. Note that the larger the threshold, the more outliers may be obtained. Another important operation is to tune corresponding control point locations using cross correlation (e.g. function 'cpcorr' in Matlab), then refined matching point locations are used to estimate transformation parameters.

Transformation models, such as rigid, affine, and second order polynomial (quadratic), are adaptively chosen to register image pairs according to the matches number. The reason why we use the quadratic transformation is that the surface of retina is approximately spherical [12]. Nonetheless, the difference between [18] and ours is that the former uses a hierarchical style from linear conformal to affine or second order polynomial transformation iteratively, our framework registers images only depending on the