Computerized Medical Imaging and Graphics 55 (2017) 95–105

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Retinal image registration under the assumption of a spherical eye

Carlos Hernandez-Matas[a,b,*], Xenophon Zabulis[a], Areti Triantafyllou[c], Panagiota Anyfanti[c], Antonis A. Argyros[a,b]

[a] Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece
[b] Computer Science Department, University of Crete, Heraklion, Greece
[c] Department of Internal Medicine, Papageorgiou Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form 23 May 2016
Accepted 21 June 2016

Keywords:
Retinal image registration
Medical imaging
Particle Swarm Optimization
Pose estimation

ABSTRACT

We propose a method for registering a pair of retinal images. The proposed approach employs point correspondences and assumes that the human eye has a spherical shape. The image registration problem is formulated as a 3D pose estimation problem, solved by estimating the rigid transformation that relates the views from which the two images were acquired. Given this estimate, each image can be warped upon the other so that pixels with the same coordinates image the same retinal point. Extensive experimental evaluation shows improved accuracy over state of the art methods, as well as robustness to noise and spurious keypoint matches. Experiments also indicate the method's applicability to the comparative analysis of images from different examinations that may exhibit changes and its applicability to diagnostic support.

1. Introduction

Functional and structural assessment of small vessels in vivo can promote accurate diagnosis and monitor progression of diseases with a strong vasculopathy, such as hypertension and diabetes (Grosso et al., 2005). Of all the organs within the human body, the eye, particularly the retina, provides an easily accessible way to non-invasively estimate the microvascular status via fundoscopy (Abramoff et al., 2010). The analysis of the retinal structures and particularly the microvascular network is important for the diagnosis of illnesses that affect the eyesight, such as macular edema, age-related macular degeneration or glaucoma (Abramoff et al., 2010). Retinal images can be acquired with either a fundus camera or a Scanning Laser Ophthalmoscope (SLO) (Abramoff et al., 2010). Fundus camera, which is most widely applied, is essentially a low power microscope with an attached photographic camera. A SLO uses a laser beam and a sensor to scan the fundus using a raster pattern.

The analysis of fundus images can be greatly facilitated by retinal image registration. In general, the issue of image registration involves a pair of images, the reference and the test one. Its solution seeks the spatial warping of the target image so that its points are imaged at the 2D coordinates of the corresponding points in the reference image. The reference and test images may differ with respect to the viewpoint, the time and the image acquisition device.

There are several applications of retinal image registration. Images acquired during the same examination are not expected to have significant anatomic changes. If the image pair presents significant overlap, images can be combined to generate images of higher resolution and definition (Meitav and Ribak, 2011; Molodij et al., 2014; Hernandez-Matas and Zabulis, 2014) enabling more accurate measurements of the vessel structure such as Arteriolar-to-Venular diameter Ratio, which is important for the early diagnosis of hypertensive retinopathy (Hubbard et al., 1999). In contrast, images with minor overlap can be combined into mosaics that image larger retinal areas (Can et al., 2002; Ryan et al., 2004; Cattin et al., 2006). The smaller the overlap, the lower the number of images needed to image a large area of the retina, thus increasing examination efficiency. Additionally, image pairs can be utilized for reconstructing the surface of the retina (Lin and Medioni, 2008; Choe et al., 2006; Chanwimaluang et al., 2009; Tang et al., 2011). More importantly, images acquired during different periods of time can be registered and employed for performing longitudinal studies of the retina (Narasimha-Iyer et al., 2007; Troglio et al., 2010). This allows to monitor health status and disease progression of the patient over different time-points and may represent an alternative method of assessment of the effectiveness of a treatment and patient's response.

Retinal image registration may evolve to a promising potential clinical tool, but is a challenging problem as well and several issues

* Corresponding author at: Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece.

http://dx.doi.org/10.1016/j.compmedimag.2016.06.006

need to be addressed. As the images to be registered may differ in the acquisition viewpoint and time they may exhibit considerable illumination, color and contrast changes as well as a limited joint field of view. Images obtained by different modalities might even capture complementary information. As an example, microaneurysms are depicted more prominently in angiograms than in regular color images (Abramoff et al., 2010). At the same time, in order to be able to support medical treatment, the requirements on registration accuracy are very high.

In the following section, we provide an overview of existing methods that address this interesting and challenging problem.

1.1. Related work

Image registration methods can be classified according to whether the approach is global or local, on the basis of the transformation model utilized to align the images, or on whether the method can register images from different modalities.

1.1.1. Global vs local methods

Image registration is performed by exploiting information in the common regions of the two images. This information can be retrieved either in the frequency domain (Cideciyan et al., 1992) or in the spatial domain. In the field of retinal image registration, most of the methods fall in the category of spatial methods. These are further categorized into methods that utilize similarity of intensities, methods that utilize image features or methods that combine both.

Methods based on similarity of intensities are referred to as global methods, as they utilize the entirety of image pixels. For retinal images, usually methods based on mutual information (Pluim et al., 2003; Legg et al., 2013) have been proposed. Instead of employing all image pixels, certain feature-based methods rely on carefully selected, localized features. Feature-based approaches are known as local methods. Local methods are the most popular, usually utilizing keypoint feature correspondences (Tsai et al., 2010; Chen et al., 2010; Perez-Rovira et al., 2011; Zheng et al., 2011; Lin and Medioni, 2008; Tang et al., 2011; Hernandez-Matas and Zabulis, 2014; Hernandez-Matas et al., 2015) or retinal features such as vessel trees (Matsopoulos et al., 1999) or vessel bifurcations (Stewart et al., 2003; Chaudhry and Klein, 2008; Ryan et al., 2004; Matsopoulos et al., 2004). Recently, hybrid methods that combine both global and local cues (Reel et al., 2013; Gharabaghi et al., 2012; Adal et al., 2014) are becoming increasingly popular.

Feature based methods are preferred for registering image pairs with a small overlap. These pairs exhibit an increased registration difficulty, due to the small amount of commonly available information. They are also preferred for registering images with anatomical changes. Features provide stronger cues to perform registration between images, are robust to local image differences and, in general, they require less processing power, leading to faster registration.

1.1.2. 2D vs 3D transformation models

Registration of retinal images has been performed using both 2D and 3D transformation models. While 2D transformations do not account for perspective explicitly (Cideciyan et al., 1992; Matsopoulos et al., 1999, 2004), they overcome this by employing non-linear transformations (Stewart et al., 2003; Tsai et al., 2010; Adal et al., 2014; Lin and Medioni, 2008; Ryan et al., 2004; Pluim et al., 2003; Chen et al., 2010; Perez-Rovira et al., 2011; Zheng et al., 2011). However, these transformations do not necessarily include consideration of the 3D shape and size of the eye. In cases of weak feature matching, such as in weakly textured images or image pairs with small overlap, this may cause the registration method to calculate parameters for the transformation that deviate from any meaningful approximation of the eye's shape. In turn, this can lead to inaccurate registration. Conversely, utilizing an eye model safeguards for unreasonable parameter model estimates and is, in this work, shown to provide a more accurate registration.

Moreover, considering the problem in 3D is useful because it enables metric, 3D measurements that are devoid of perspective distortion. While 3D models account for perspective, they require a knowledge of the shape of the imaged surface, either via modeling or via reconstruction. Even simple eye shape models have shown to improve registration accuracy of retinal images (Hernandez-Matas et al., 2015).

1.1.3. Intra-modal vs cross-modal image registration

Methods capable of registering images acquired from different imaging techniques or sensors are known as cross-modal or multi-modal (Ryan et al., 2004; Tsai et al., 2010; Chen et al., 2010; Matsopoulos et al., 2004). Different modalities offer different information, and registration enables the combination of such information. Generally, featured based methods can perform cross-modal registration if cross-modal features are utilized.

1.2. Contributions of this work

We propose a retinal image registration method that is a more accurate, robust and computationally efficient alternative to the method presented in Hernandez-Matas et al. (2015). In terms of the previously mentioned classification, the proposed method is a local, intra-modal registration method employing a 3D transformation model. More specifically, the method assumes that the human eye has a spherical shape. Retinal image registration is then based on first recovering the relative pose of the cameras that acquired the images to be registered. This is performed by a 3D model-based initialization step, followed by a pose refinement that is achieved by solving an optimization problem. The objective function that is optimized involves quantities depending on keypoint correspondences. However, by design, any alternative local or even global cue may be accommodated.

Improvements over the prior work in Hernandez-Matas et al. (2015) are the following. SIFT keypoints are found to be more valuable for fundus image registration than SURF which were previously utilized, thus resulting to an increase of registration accuracy. An initialization of pose estimation is introduced, yielding more accurate results due to the avoidance of local optimization minima. Moreover, this initialization enables better utilization of computational resources resulting at a reduced computational cost, besides increased registration accuracy. An extensive and elaborate experimental evaluation demonstrates the benefits of these improvements quantitatively and, also, employs the proposed method in a super-resolution application.

2. Method

The proposed method (Fig. 1) performs the registration of the reference (F_0) and the test (F_t) images by first estimating the relative pose of the cameras that acquired those images. To do that, point correspondences between F_0 and F_t are established. The relative pose $\{\mathbf{R}, \mathbf{t}\}$ of the two views consists of a rotation $\mathbf{R}$ and a translation $\mathbf{t}$. A spherical eye model $\mathcal{S}$ centered at $\mathbf{c}_s = [0, 0, 0]^T$ is assumed, with a calibrated camera at a distance δ, located at $\mathbf{c}_c = [0, 0, -\delta]^T$. K_c and K_t are respectively the intrinsic camera matrices for F_0 and F_t. Equivalently, this pose estimate can be also calculated as the pose transformation of the retina between the two frames if a stationary camera is assumed. Hypothesis, or "candidate pose", with id h regards camera motion $\{\mathbf{R}_h, \mathbf{t}_h\}$ and can be considered as a point in a 6D space.