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
0895-6111/

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.

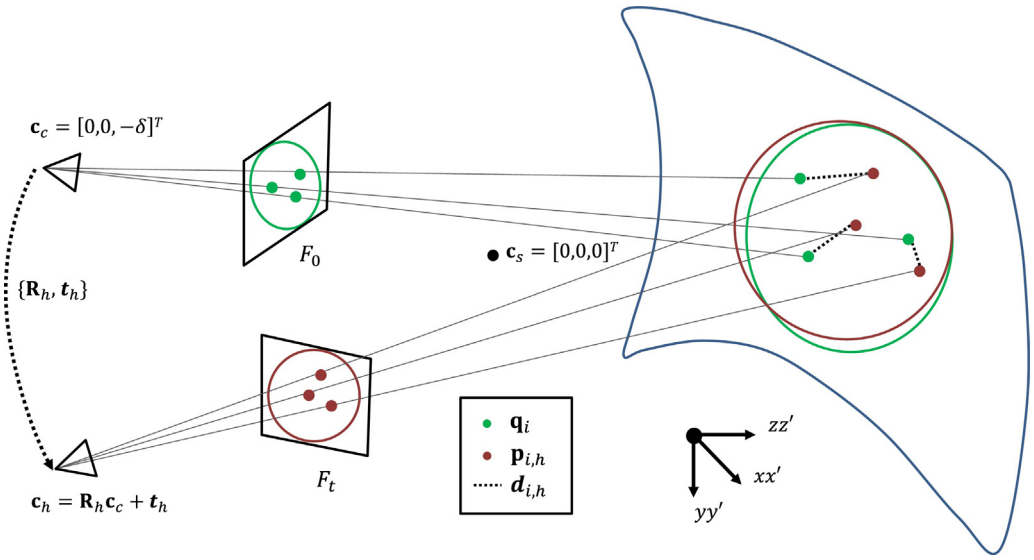

Fig. 1. Geometry of the proposed registration method.

The method evaluates hypotheses around an initial pose estimate $\{\mathbf{R}_0, \mathbf{t}_0\}$. If such an estimate is not available, the identity transformation $\{\mathbf{I}, \mathbf{0}\}$ is utilized, encoding no prior knowledge about the relative pose of the camera that acquired the test image. Utilizing an initial pose estimation $\{\mathbf{R}_0, \mathbf{t}_0\}$ allows for a reduction of the search space of the subsequent optimization method. Further hypotheses are formed and evaluated within the stochastic optimization framework of Particle Swarm Optimization (PSO) (Poli et al., 2007). Several variants of this baseline approach have been formulated and experimentally validated. Having estimated the relative 3D motion of the cameras that acquired F_0 and F_t, it is then straightforward to warp F_t and register it to F_0, based also on the hypothesis of the spherical shape of the eye.

The following sections describe the above algorithmic steps in more detail.

2.1. Feature correspondences

The proposed method utilizes SIFT (Lowe, 2004) keypoint features to establish correspondences between points in the reference image F_0 and the test image F_t. SIFT are general purpose features that exhibit considerable invariance to image translation, scaling, rotation and illumination. In contrast to Y-features (Ryan et al., 2004), SIFT features are not specifically oriented for use in retinal images. SIFT keypoints were preferred over Y-features as they provide more corresponding points. Moreover, they were observed to be more evenly distributed over the imaged surface, enabling the proposed method to utilize information from all areas of the image ensuring a more robust registration result. Additionally, they have a strong descriptor that provides more reliable feature matching. Corresponding keypoints are matched conventionally as described in Lowe (2004), i.e., by first identifying the nearest neighbor of each keypoint on the basis of the Euclidean distance of the descriptor vectors. At a second step, and in order to increase matching confidence, matches are rejected if the distance to the first and second closest neighbors are similar. SIFT features were comparatively evaluated to the, also general-purpose, SURF features (Bay et al., 2008), which were utilized in Hernandez-Matas et al. (2015). The results presented in Section 3.3 lead to the conclusion that the proposed method is more accurate when utilizing SIFT features, due to the higher amount of correspondences (approximately 20% more) established in the weakly textured image areas.

2.2. Eye model

The 3D eye model utilized is commonly known as the Navarro model (Navarro et al., 1985), which describes a spherical eye with radius $\rho = 12$ mm. The Navarro model provides an approximation to the eye's real shape, which in reality may be better approximated by an ellipsoid. This is true not only for eyes with pathologies, such as myopia and hypermetropia, but also for healthy eyes. The consideration of departures from the spherical model and the implications of more accurate eye shape modeling on retinal image registration accuracy are left as a topic of future work.

2.3. Initialization

A first hypothesis on the relative pose of the two views is provided by employing Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981). SIFT correspondences are utilized to solve the Perspective-n-Point (PnP) problem (Fischler and Bolles, 1981). This method estimates the 3D pose of an object given a set of 2D–3D correspondences and the camera projection matrix P and is robust to the presence of outliers. The 3D pose is calculated by minimizing the reprojection error between the 2D points and the projected 3D points. It should be noted that in F_0 and F_t we retrieve 2D–2D and not 2D–3D correspondences. However, the 3D location of the keypoints of F_0 can easily be retrieved given the calibration of the cameras as well as the spherical shape of the eye (see Eq. (1)).

2.4. Optimization

The estimation of the relative pose of two cameras or the refinement of an initial estimation of it (see Section 2.3) can be formulated as the solution of an optimization problem. More specifically, a candidate pose $\{\mathbf{R}_h, \mathbf{t}_h\}$ is evaluated utilizing an objective function. Given the reference or candidate camera pose, keypoints can be geometrically traced to 3D locations on the eye model surface $\mathcal{S}$. This is calculated by finding the line that passes through both the camera optical center and the keypoint in the respective image. The intersection between this line and $\mathcal{S}$ indicates the 3D position of the point.

In the proposed approach, the origin of coordinates is located at the center of $\mathcal{S}$. This provides the following sphere equation: $\|\mathbf{x} - \mathbf{c}_s\|^2 = \rho^2$ where $\mathbf{x}$ is a point on $\mathcal{S}$ and $\mathbf{c}_s$ is the center of $\mathcal{S}$. To

find $\mathbf{x}$, the line equation from the camera center $\mathbf{c}$ ($\mathbf{c}_c$ or $\mathbf{c}_h$) and through pixel $\mathbf{u}$ is solved for λ:

$$\mathbf{x} = P^{+}\mathbf{u} + \lambda\mathbf{c}, \tag{1}$$

where $P^{+} = P^{T}(PP^{T})^{-1}$ (see Eq. (6.13), Hartley and Zisserman (2004, p. 162)).

Let $\mathbf{q}_i$ be the locations on $\mathcal{S}$ of the keypoints from reference image F_0. The 3D locations of keypoints from the test image are determined by pose $\{\mathbf{R}_h, \mathbf{t}_h\}$ (see Fig. 1). Let $\mathbf{p}_{i,h}$ be the locations on $\mathcal{S}$ of the keypoints from test image F_t for pose hypothesis h. The 3D distances of corresponding keypoints on $\mathcal{S}$ are $d_{i,h} = |\mathbf{q}_i - \mathbf{p}_{i,h}|$. To increase robustness to spurious matches on $o(\cdot)$, a percentile of accumulation of distances $d_{i,h}$ is used:

$$o(\{R_h, \mathbf{t}_h\}) = \sum_j d_{j,h}, \tag{2}$$

where j enumerates the smallest 80% values of $d_{i,h}$, found after sorting them. To combine more cues, Eq. (2) can be amended with further terms.

Despite RANSAC separating keypoint correspondences into inliers and outliers, all correspondences are taken into consideration by the objective function. We chose to pursue robustness through the subsequent percentile filtering of $d_{j,h}$ instead. The reason is the approximation of eye shape by $\mathcal{S}$, which may potentially attribute an unintentional large distance to a correct correspondence.

2.5. Particle Swarm Optimization

The optimization of Eq. (2) cannot be performed based on an exhaustive, grid-based search of the 6D space of poses because this is too computationally prohibitive. Additionally, such an approach creates a discretized version of the problem, as the accuracy in pose estimation is bounded by the size of the grid. For this reason, we optimize the objective function of Eq. (2) by employing a stochastic, derivative-free optimization method called Particle Swarm Optimization (PSO) (Poli et al., 2007). PSO was successfully employed not only for this particular problem (Hernandez-Matas et al., 2015) but also in other pose estimation problems (Oikonomidis et al., 2011; Zhang et al., 2008; Padeleris et al., 2012; Kyriazis and Argyros, 2014; Panteleris and Argyros, 2014).

PSO achieves optimization utilizing a set of n_p particles that evolve through n_g generations. PSO depends on few parameters, does not require knowledge of the derivatives of the objective function and can handle multimodal and possibly discontinuous objective functions. Additionally, it requires a relatively low number ($n_p \cdot n_g$) of objective function evaluations, which is called "budget" of the optimization. A small budget will terminate the process prematurely with a poor pose estimate, while a too large budget will lead to extra processing time without leading to noticeable improvements in accuracy. Beyond these extreme conditions, the selection of budget offers a trade-off between the accuracy and the speed of the method. Additionally, for a given budget, the distribution n_p/n_g of n_p particles and n_g generations is relevant to the final performance of the method.

In our problem formulation, the 6D search space is a hypercube centered around $\{\mathbf{R}_0, \mathbf{t}_0\}$. The six parameters to optimize correspond to the relative rotations and translations of $\mathbf{c}_h$ with relation to $\mathbf{c}_c$. If there is no initial pose, $\{\mathbf{I}, \mathbf{0}\}$ is utilized. Poses are parameterized as translations with Euclidean coordinates and rotations using Euler angles. Thus $\mathbf{t}_h = [t_{x,h}\ t_{y,h}\ t_{z,h}]^T$, while $\mathbf{R}_h = \mathbf{R}_x(r_{\theta,h}) \cdot \mathbf{R}_y(r_{\phi,h}) \cdot \mathbf{R}_z(r_{\omega,h})$. The search space around the initial pose is denoted as $[\mu_x, \mu_y, \mu_z, \gamma_\theta, \gamma_\phi, \gamma_\omega]$, meaning that $t_{x,h} \in [-\mu_x/2, \mu_x/2]$ and correspondingly for the rest of dimensions.

In the first generation, $2^6 = 64$ particles are initialized at the corners of the hypercube-shaped search space, with an additional particle in the middle of each edge, ensuring a spread particle allocation. The rest of particles are randomly and uniformly distributed within the search space.

If several PSO stages are executed in a cascaded manner, the result of the preceding execution is pipelined as the initial pose estimate.

2.6. Method variants

Aiming at accuracy and robustness, several method variants were considered.

- **Coarse (C):** Baseline method described in Section 2.4.
- **Coarse-to-Fine (CF):** PSO is executed two times. In the second execution, the search space is a smaller hypercube centered at the estimate provided by the first execution. The same budget is utilized, but the hypercube is reduced in all 6 dimensions thus performing a denser search. This second hypercube range is $\left[\frac{\mu_x}{\alpha}, \frac{\mu_y}{\alpha}, \frac{\mu_z}{\alpha}, \frac{\gamma_\theta}{\alpha}, \frac{\gamma_\phi}{\alpha}, \frac{\gamma_\omega}{\alpha}\right]$ with $\alpha = 3$.
- **Pairing of conflicting Dimensions (PD):** PSO is executed three times. This is due to the observation that in the 6D neighborhood of the correct pose, local minima might occur. The 6D locations of said minima correspond to poses in which points $\mathbf{p}_{i,h}$ occur at close locations. Small changes in r_ϕ and in t_x can have similar effects on $\mathbf{p}_{i,h}$, as they induce very similar displacements when performed in small scale. This might lead to different locations in the search space returning similar objective function scores. This effect is also observed for t_y and r_θ. To overcome this, the first execution is variant C. In the second execution, the range is $\left[\frac{\mu_x}{\beta}, \frac{\mu_y}{\alpha}, \frac{\mu_z}{\beta}, \frac{\gamma_\theta}{\alpha}, \frac{\gamma_\phi}{\beta}, \frac{\gamma_\omega}{\beta}\right]$, centered around the first estimate. In the third execution, the range is $\left[\frac{\mu_x}{\alpha}, \frac{\mu_y}{\beta}, \frac{\mu_z}{\beta}, \frac{\gamma_\theta}{\beta}, \frac{\gamma_\phi}{\alpha}, \frac{\gamma_\omega}{\beta}\right]$ centered around the second estimate. Here $\alpha \ll \beta$, implying that in the second execution the search will be mostly along μ_y and γ_θ and in the third execution along μ_x and γ_ϕ. Values of $\alpha = 3$, $\beta = 30$ were chosen as in Hernandez-Matas et al. (2015).
- **RANSAC (R):** In this approach, $\{\mathbf{R}_0, \mathbf{t}_0\}$ constitutes the solution, i.e., the RANSAC-based initialization described in Section 2.3.
- **R-C:** Coarse variant with initialization.
- **R-F:** Fine variant with initialization. Similar to variant R-C, but with search range $\left[\frac{\mu_x}{\alpha}, \frac{\mu_y}{\alpha}, \frac{\mu_z}{\alpha}, \frac{\gamma_\theta}{\beta}, \frac{\gamma_\phi}{\beta}, \frac{\gamma_\omega}{\beta}\right]$. Values of $\alpha = 3$, $\beta = 10$ were chosen empirically.
- **R-CF:** Coarse-to-fine with initialization.
- **R-PD:** Pairing of dimensions with initialization.

Thus, variants C, CF and PD are initialized at $\{\mathbf{I}, \mathbf{0}\}$. Variants R-C, R-CF and R-PD are identical to them, but initialized at $\{\mathbf{R}_0, \mathbf{t}_0\}$.

2.7. Multiple swarm execution

Given the stochastic nature of both RANSAC initialization and PSO, the results of the proposed method are non-deterministic. In certain cases, this can lead to entrapment of the optimization process in local minima, providing suboptimal registration results. To avoid this, n_s independent swarms are run and the result with the best score is chosen as solution. In this work, we adopt $n_s = 10$, as suggested by the results in our previous work Hernandez-Matas et al. (2015). Despite the increase in the computational cost with relation to running the method just one time, this solution offers increased robustness and higher reliability. This is particularly noticeable in variants without initialization or without an efficient cascaded PSO approach, where the search space is large and an

equivalent budget with a single run is likely to converge to a less accurate solution.

2.8. Image registration

Once a solution $\{\mathbf{R}, \mathbf{t}\}$ is estimated, it is used to warp the test image F_t so that it is registered to the reference image F_0. In the resulting image F_w, each pixel coordinate $\mathbf{u}$ in F_0 corresponds to a 3D location $\mathbf{x}$ on the retina (Eq. (1)). The intensity of the pixels in F_w is calculated as:

$$F_w(\mathbf{u}) = F_t(P \cdot \mathbf{x}), \tag{3}$$

where $P = K \cdot [\mathbf{R}, \mathbf{t}]$ denotes the projection matrix of the camera at the reference pose. Points $P \cdot \mathbf{x}$ do not occur at integer pixel coordinates, thus intensity values $F_t(P \cdot \mathbf{x})$ are obtained using bilinear interpolation of the four intensities of the nearby points.

2.9. Parallelization

The proposed variants are amenable to parallelization in two levels. The first is at particle level, as particles communicate through generations, but are independent inside each generation. The second is at swarm level, given that each of the n_s swarm is independent of the rest. In the current implementation, both CPU and GPU acceleration are available and the execution of individual particles is performed in parallel on the available cores of either the CPU or the GPU. In this way particles are run each time in parallel.

3. Experiments

The goal of the conducted experiments was twofold. First, to configure the parameters of the proposed approaches and, second, to compare the registration accuracy of the proposed approaches to the current state of the art.

Experimental evaluation compares the proposed method with three retinal image registration methods. These are RANSAC, the method in Hernandez-Matas et al. (2015) and GDB-ICP (Yang et al., 2007). RANSAC is evaluated as it is a standard approach in correspondence-based pose estimation. The approach in Hernandez-Matas et al. (2015) was chosen as an earlier version of the proposed method and was already compared to GDB-ICP. Finally, GDB-ICP is included in this study as it is widely employed in retinal (Stewart et al., 2003; Tsai et al., 2010; Chen et al., 2010; Zheng et al., 2011; Hernandez-Matas et al., 2015) and general (Yang et al., 2007) image registration.

Quantitative evaluation requires ground truth information. To the best of our knowledge, there is currently no publicly available dataset which includes high-precision retinal registration ground truth or pose estimation data. The dataset in Tang et al. (2011) focuses on stereo retinal images, with image pairs taken during the same examination session, having large overlapping area and no anatomical changes between them, providing little diversity for the purpose of our experiments. More importantly, the dataset features images of very narrow field of view that image a very small portion of the retina, where the imaged surface can be accurately approximated by a plane. The proposed work is intended for registration of retinal images of wide extent, where retinal curvature cannot be neglected. Thus, datasets with synthetic and with real data were created; the data and their collection process are presented in Section 3.1. The evaluation process is described in Section 3.2.

The evaluation of the variants presented in Section 2.6 is performed in Section 3.3. The best of the variants is selected and computational budget issues are investigated in Section 3.4. The variant and budget combination is then evaluated against methods in 3D retinal registration utilizing synthetic data in Section 3.5. Accuracy and robustness comparison with 2D retinal registration methods is performed utilizing real data in Section 3.6. Their time performances are evaluated in Section 3.7. The proposed method is also tested in the task of constructing super resolved images in Section 3.8. In all experiments, δ (≈57.7 mm) is known from device specifications due to the employed chin-rest and head-placement apparatus.

3.1. Datasets

Datasets with real and synthetic images are utilized. Real images were acquired utilizing a Nidek AFC-210 fundus camera, with a resolution of 2912×2912 pixels and a Field of View (FOV) of 45° in both dimensions. Synthetic images were generated utilizing a virtual camera with the same characteristics. Representative image pairs are shown in Fig. 2.

In real data, ground truth regarding eye poses are difficult to obtain because of a number of unknown factors, such as the exact eye size, the imaging distance, and other optical approximations (distortions caused by cornea and the vitreous humor, etc). Thus, ground truth annotations are restricted to point correspondences between images.

3.1.1. Synthetic data

Synthetic data annotated by ground truth were generated by conventional 3D rendering and by recording of the corresponding poses. The data consist of 100 image pairs rendering a spherical eye of $\rho = 12$ mm, using texture from a real fundus image, at random poses in the ranges of $\{-3 \leq t \leq 3\}$ mm for translation and $\{-10 \leq r \leq 10\}$ degrees for rotation, which produce large displacements of the imaged retina.

3.1.2. Real data

A collection of 116 real image pairs have been classified into three datasets, depending on the similarities and differences between the images of each pair; each pair is a member of only one group. The properties of each dataset are shown in Table 1. In Dataset 3, anatomical changes are due to the progression, or remission due to appropriate treatment, of retinopathy such as increased vessel tortuosity, microaneurysms, cotton-wool spots, etc. All three datasets include pathological cases such as myopia and hypermetropia that mostly affect the shape of the eye.

As 3D eye motion and image acquisition parameters are not precisely known, ground truth is provided in the form of 2D correspondences. Two ways of establishing these correspondences were investigated. Conventionally, the first was to ask medical experts to annotate the images, marking the same physical points in pairs of images. The second was to automate a part of the process, by letting a computational method establish correspondences and then letting a human supervisor verify the results by, e.g., rejecting spurious matches.

Table 1
The characteristics of the three employed real image pairs datasets.

Dataset	1	2	3
# Image pairs	57	45	14
Examination session	Same	Same	Different
Overlap	>75%	<75%	>75%
Anatomical changes	No	No	Yes
Indicative application	Super resolution	Mosaicing	Longitudinal study

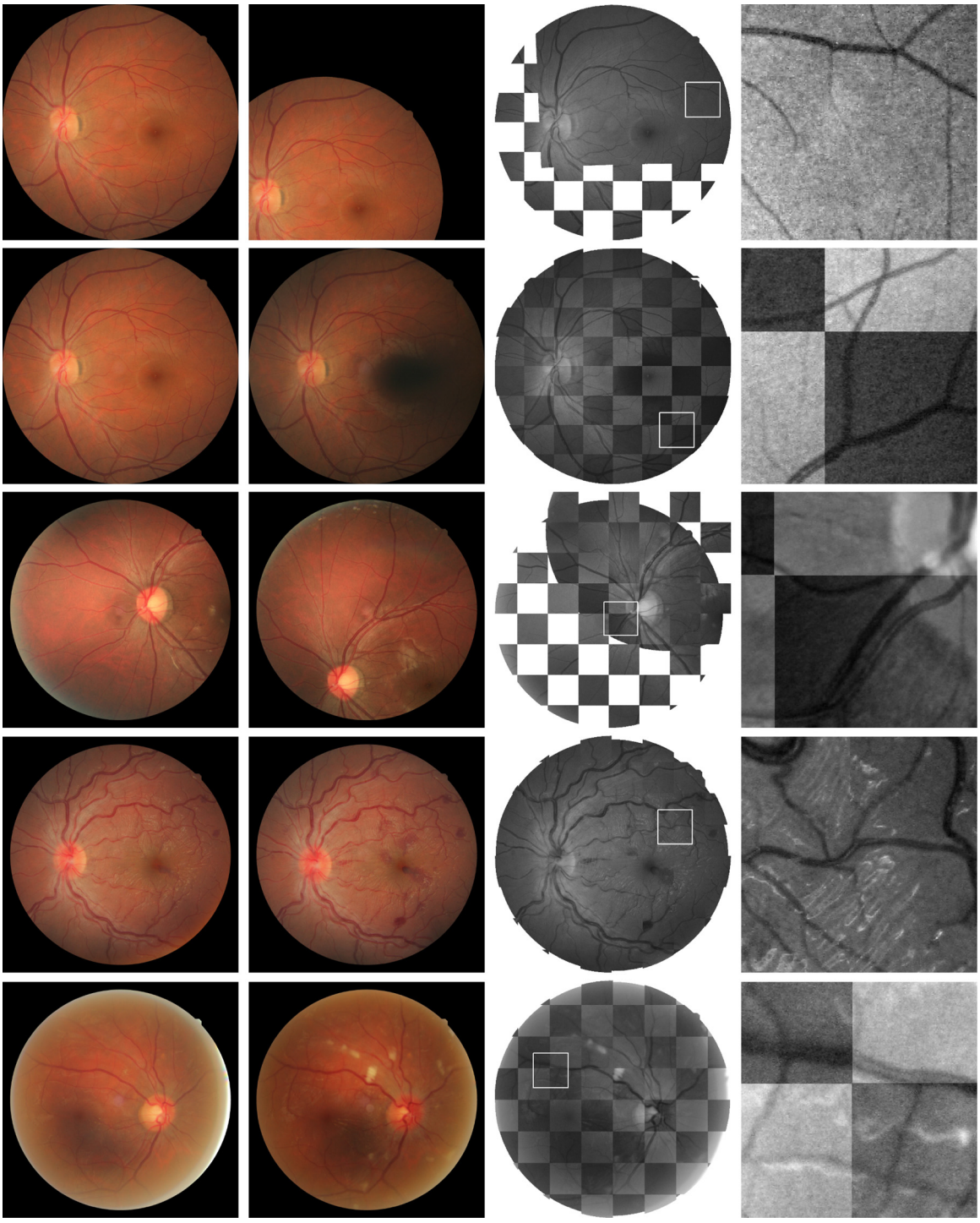

Fig. 2. Evaluation images and registration results for 5 image pairs from the datasets. Row 1 shows a pair from the synthetic dataset. Row 2 for real Dataset 1. Row 3 for real Dataset 2. Rows 4 and 5 for real Dataset 3. The collages show, alternatively, the green channels of the reference and registered image. The marked region with solid white line indicates the magnified image detail shown on the right, which has been normalized for increased visibility.

The two different approaches were evaluated in synthetic images, where the ground truth is known. Two medical experts were asked to annotate two pairs of images, one with small and the other with large overlap. The results were then compared against matches calculated automatically and supervised by a non-medical expert. As the proposed method relies on SIFT features, alternative features were selected to maintain independence from the registration method. SURF features were selected due to their proven reliability and accuracy when utilized in retinal images (Hernandez-Matas and Zabulis, 2014; Hernandez-Matas et al., 2015).

Let $\mathbf{q}_i$ be the 2D locations of control points in F_0. Let $\mathbf{r}_i$ be the 2D locations of control points in F_i. Loci $\mathbf{r}_i$ are traced to $\mathcal{S}$ from pose $\{\mathbf{R}_h, \mathbf{t}_h\}$ and projected into F_0, resulting in $\mathbf{p}_i$. The 2D distances of corresponding 2D locations are $d_i = |\mathbf{q}_i - \mathbf{p}_{i,h}|$. The error E is the average of the distances d_i as formulated in Eq. (4). If the control points are annotated manually, all of them are used. If they are selected automatically, j enumerates the manually

Table 2
Error indicates the average error of the control points after registration, in pixels.

	Expert A	Expert B	SURF
Large overlap >75%	1.27 (10 points)	1.36 (10 points)	0.22 (581 points)
Small overlap <75%	1.09 (10 points)	1.31 (10 points)	0.24 (305 points)

supervised points from the smallest 25% values of d_i, found after sorting them.

$$E = \frac{1}{j}\sum_{j} d_j \tag{4}$$

Table 2 shows that the automatic keypoint selection with human supervision provides a larger selection of points, while presenting a smaller error and performing faster. This presents a fast and reliable alternative to having medical experts manually performing these annotations. For these reasons, semi-automatic SURF features are selected for the ground truth.

3.2. Evaluation criteria

As each dataset offers different types of ground truth, the evaluation of registration is susceptible to different criteria.

For synthetic data, we quantify and evaluate the 3D pose estimation error. Error is quantified as in Hinterstoisser et al. (2012), utilizing the average distance between corresponding model points $\mathbf{x}_\kappa$ at the ground truth and the estimated pose. Given the ground truth pose $\{\mathbf{R}_g, \mathbf{t}_g\}$ and an estimate pose $\{\mathbf{R}_h, \mathbf{t}_h\}$, the error is expressed as:

$$E = \frac{1}{\nu}\sum_{\kappa=1}^{\nu} |\mathbf{g}_\kappa - \mathbf{e}_\kappa|, \tag{5}$$

where $\mathbf{g}_\kappa = \mathbf{R}_g \cdot \mathbf{x}_\kappa + \mathbf{t}_g$, $\mathbf{e}_\kappa = \mathbf{R}_h \cdot \mathbf{x}_\kappa + \mathbf{t}_h$, and ν is the number of model points.

For real data, we estimate the reprojection error (provided by Eq. (4)) for the annotated points. Additionally, the success of a method relative to a competitor method is measured for every individual image pair. A method is considered as successful in registering an image pair if the registration error was at least 5% lower compared to the competitor. Otherwise, the result is considered a tie.

3.3. Comparison of the proposed variants

The eight variants described in Section 2.6 are compared, using both SURF and SIFT features, to find the most suitable variant and feature type for registration. Registration is evaluated upon synthetic data, enabling the comparison in terms of 3D pose accuracy. In the evaluation, every variant utilizes $n_s = 10$ independent swarms. Variants C, R-C and R-F run 7500/200 in a single PSO stage. Variants CF and R-CF run 3750/200 in each of their two PSO stages. Variants PD and R-PD run 2500/200 in each of their three PSO stages. Thus, all variants, except R, have a budget of 1500k particles per swarm. This budget is chosen so that it is identical to the budget utilized in Hernandez-Matas et al. (2015).

Table 3 shows that, in general, utilizing SIFT features leads to increased accuracy over SURF. This is particularly noticeable in variants with initialization, for which the accuracy limits of the method are reached. Additionally, in our data SIFT provides about 20% more correspondences than SURF. Thus, SIFT features are selected for the proposed approach.

Regarding variant selection, among variants without initialization (C, CF and PD), variant PD offers the best result, showing that pairing of conflicting dimensions helps avoiding local minima when no initialization is present, providing a result that is better than RANSAC (R). This result is in line with findings in Hernandez-Matas et al. (2015).

Among variants that utilize initialization (R-C, R-F, R-CF, R-PD), R-C is less accurate than the rest. R-F, R-CF and R-PD perform similarly with only subtle differences. This experiment shows that with a good initialization, focusing the budget around the initial estimation provides the most accurate solution. As all 3 short-listed variants perform equivalently, R-F is selected due to its simplicity of implementation and its greater potential for parallelization.

A statistical significance analysis is performed on the results summarized by Table 3. Each method is compared with R-F (SIFT) utilizing Student's t-test (Gosset, 1908). Results are presented in Table 4. Results are shown to be statistically highly significant for all the variants when utilizing SURF features, as well as variants that utilize SIFT features and do not combine RANSAC and PSO. When comparing with R-C, R-F results are statistically significant. However, there is no statistical significance in the results when comparing with R-CF and R-PD. These results support the utilization of R-F over R-CF and R-PD due to similarity of results but simpler implementation and greater parallelization potential.

Table 3
Average 3D registration error (and std) per point, in μm, for method variants A-H (Section 2.6) utilizing both SIFT and SURF features.

	SURF	SIFT
C	11.06 (10.22)	12.45 (12.00)
CF	3.92 (5.99)	5.47 (8.98)
PD	1.07 (1.74)	0.68 (0.71)
R	7.43 (5.93)	7.20 (6.05)
R-C	0.74 (0.41)	0.58 (0.72)
R-F	0.71 (0.31)	0.43 (0.18)
R-CF	0.72 (0.37)	0.41 (0.17)
R-PD	0.70 (0.38)	0.42 (0.18)

Table 4
p-values of experimental results from Table 3 when compared to R-F (SIFT features).

	SURF	SIFT
C	>0.001	>0.001
CF	>0.001	>0.001
PD	>0.001	0.001
R	>0.001	>0.001
R-C	>0.001	0.0466
R-F	>0.001	–
R-CF	>0.001	0.0711
R-PD	>0.001	0.2427

3.4. Impact of PSO budget

The aim of this experiment is to investigate the impact of key PSO parameters on the accuracy and performance of the registration of the R-F variant that has been selected in the experiment of Section 3.3. A certain computational budget can be achieved with different combinations of n_g/n_p. However, the accuracy of the solutions reached by these combinations may differ considerably. As previously, registration is performed upon synthetic data, with 3D pose ground truth. The proposed method is utilized to register the images from the dataset with varying combinations of n_p/n_g for 5 different budgets.

The obtained results are illustrated in Fig. 3. For any budget, a configuration with fewer generations leads to larger errors and unstable results (i.e. $n_g = 150$). On the other hand, using more generations and fewer particles (i.e. $n_g = 400$), does not necessarily result in improved accuracy, as the particle density in 6D space becomes small. As a result, it becomes more likely to converge to local minima of the objective function the area close to the optimal solution largely unexplored.

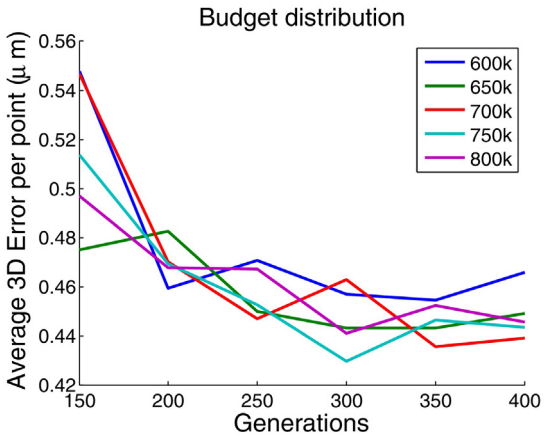

Fig. 3. Registration error of budgets as a function of their distribution across generations.

Table 5
Average 3D registration error (and std) per point, in μm.

	Proposed	(Hernandez-Matas et al., 2015)	RANSAC
Error	0.43 (0.18)	0.68 (0.71)	7.20 (6.05)

The combination of 2500 particles and 300 generations appears to be the best choice. This is equivalent to a budget of 750k particles per swarm. In this case, the error is close to the minimum error found among all the variants in the experiment of Section 3.3, yet requiring only half of that budget.

3.5. 3D pose estimation accuracy

This experiment compares the R-F with two pose refinement methods utilized in retinal image registration. The opposing methods are RANSAC, as described in Section 2.3, and the method in Hernandez-Matas et al. (2015). As 3D ground truth of the pose is required, synthetic data is utilized.

The results of this experiment are shown in Table 5. The proposed method clearly outperforms the competing methods. It is worth noting that the proposed method is a combination of the two opposing methods, as it employs RANSAC for initializing the search space used by a simplified variant of Hernandez-Matas et al. (2015). Besides being more accurate, this combination approximately halves the PSO budget and overall execution time.

An additional 3D registration method is Lin and Medioni (2008). Given a large enough baseline, it reconstructs the retinal shape. Aside from this difference, comparison with this method is not possible quantitatively, as their data is multimodal and no ground truth is provided. Thus, this method is not included in Table 5. The proposed method displays registration results with higher accuracy than (Lin and Medioni, 2008) when performing qualitative comparison,[1] as for (Lin and Medioni, 2008) discontinuities are observed in vessels when comparing registration results as in Fig. 2.

3.6. 2D registration accuracy

This experiment compares the proposed method with three retinal image registration methods. These are RANSAC, the method in Hernandez-Matas et al. (2015) and GDB-ICP (Yang et al., 2007).

[1] The results at http://iris.usc.edu/people/yupingli/research.html were utilized for the comparison.

Table 6
Proposed method vs RANSAC. Top: Mean (and std) registration error (in pixels) of the proposed method and (Hernandez-Matas et al., 2015). Bottom: percentage of image pairs in which the proposed method performed more accurately than RANSAC.

Dataset	1	2	3	Total
Error R-F (Proposed)	0.53 (0.25)	1.17 (1.17)	1.25 (0.51)	0.61 (0.50)
Error RANSAC	0.55 (0.27)	7.64 (11.77)	1.35 (0.64)	1.29 (4.31)
Proposed is better	50.87%	71.11%	42.85%	57.75%
RANSAC is better	17.54%	0.00%	7.14%	9.48%
Tie	31.57%	28.88%	50.00%	32.75%

Table 7
Proposed method vs the method in Hernandez-Matas et al. (2015). Table interpretation is in Table 6.

Dataset	1	2	3	Total
Error R-F (Proposed)	0.53 (0.25)	1.17 (1.17)	1.25 (0.51)	0.61 (0.50)
Error (Hernandez-Matas et al., 2015)	0.54 (0.27)	1.51 (1.28)	1.33 (0.62)	0.65 (0.58)
Proposed is better	28.07%	55.55%	21.42%	37.93%
(Hernandez-Matas et al., 2015) is better	10.52%	13.33%	7.14%	11.20%
Tie	61.40%	31.11%	71.42%	50.86%

These experiments are performed upon the three datasets of the real data, so as to evaluate the performance of the proposed method in the real setting of its application.

Table 6 summarizes the results of the comparison of the proposed method with RANSAC. The proposed method provides a lower average registration error in all cases, clearly outperforming RANSAC. This is most pronounced in Dataset 2, which contains large pose differences. In this dataset, the proposed method performs better than RANSAC in 71.11% of the image pairs. This is attributed to the sensitivity of RANSAC to local minima, in case of small image overlap.

In Table 7 the proposed method is compared with (Hernandez-Matas et al., 2015). The proposed method is more accurate and robust than (Hernandez-Matas et al., 2015). A lower average registration error is observed in all cases. The largest improvement is in Dataset 2 where the proposed method performs better than (Hernandez-Matas et al., 2015) in 55.55% of the image pairs. The increased accuracy and robustness are due to the efficiency of devoting the PSO budget to performing a narrow search around $\{\mathbf{R}_0, \mathbf{t}_0\}$.

Table 8
Proposed method vs GDB-ICP. Table interpretation is in Table 6; see text for bottom, additional row. Note that comparison is based on a subset where GDB-ICP provided a solution. GDB-ICP fails in a large percentage of cases.

Dataset	1	2	3	Total
Error R-F (Proposed)	0.52 (0.24)	0.97 (0.42)	1.21 (0.41)	0.56 (0.30)
Error GDB-ICP	0.53 (0.24)	0.88 (0.42)	1.24 (0.46)	0.57 (0.29)
Proposed is better	43.85%	11.11%	14.28%	27.58%
GDB-ICP is better	21.05%	13.33%	7.14%	16.37%
Tie	17.54%	0.00%	7.14%	9.48%
No GDB-ICP Solution	17.54%	75.55%	71.42%	46.55%

Table 9
Average registration time per image pair, in seconds.

	RANSAC	(Hernandez-Matas et al., 2015)	GDB-ICP	Proposed CPU	Proposed GPU
Time (s)	12	146	73	71	51

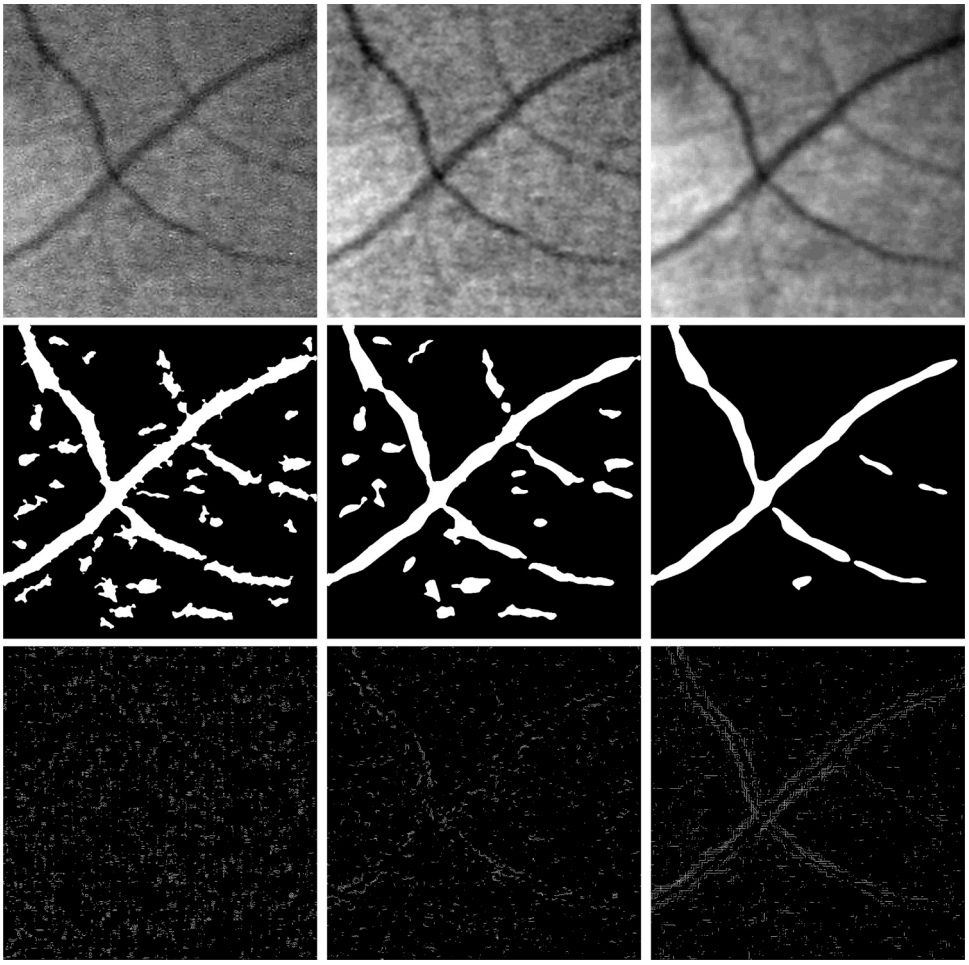

Fig. 4. Normalized green channels (top), segmentation results (middle) and Sobel edge detection response (bottom). Left original, center bilateral, right super resolution. F_0 is left image of row 2 from Fig. 2.

In Table 8 the proposed method is compared with GDB-ICP. The employed implementation[2] may result in no registration result for a given image pair. This is due to the algorithm having a weak initialization, based on a single keypoint match per initialization. It finds 25 different initializations for the parameters, and iterates through them performing the registration process until one of them performs a successful registration. If no initialization leads to a satisfactory result, no solution is provided. This is the case in 46.55% of the proposed image pairs, as opposed to the proposed method which always yields a result. Comparing the results of the proposed method between Tables 7 and 8 it is observed that when the error of the proposed method is calculated in the whole dataset, it increases, thus showing that the pairs missed by Yang et al. (2007) present increased registration difficulty. For the registered pairs, the proposed method provides a lower average registration error for Datasets 1 and 3, and higher for Dataset 2. The competing method may show less error in Dataset 2, but it is computed over the fewer pairs for which it provided a result, indicating that the proposed method is more robust. The proposed method copes with these challenging cases, with just a marginal increase of error compared to less demanding cases.

The statistical significance of the results is presented. We compare RPE with the competing methods utilizing Student's *t*-test (Gosset, 1908). This test indicates that the experimental results of RPE compared to RANSAC are statistically highly significant ($P < 0.001$). Compared to Hernandez-Matas et al. (2015) ($P = 0.0129$) and GDB-ICP ($P = 0.0039$) they are statistically significant.

3.7. Time performance

This experiment evaluates the time performance of the proposed method both with CPU and GPU acceleration, as well as RANSAC, the method in Hernandez-Matas et al. (2015) and GDB-ICP (Yang et al., 2007). No proper analysis of the paralellization effect on GDB-ICP was performed, due to the employed implementation.[2] All experiments were performed on a desktop computer with an i7-4770 CPU, at 3.40 GHz and 16 GB of RAM with an NVIDIA GeForce GTX 750 Ti GPU.

[2] We utilized the implementation provided by the authors at http://www.vision.cs.rpi.edu/gdbicp/exec/.

The results presented in Table 9 show that RANSAC is the most computationally efficient of the compared registration methods. The second fastest method is the proposed method with GPU acceleration, then the proposed method with CPU acceleration, followed closely by GDB-ICP and, doubling the computational cost, the method in Hernandez-Matas et al. (2015).

3.8. Super resolution

The problem of super resolution states that an image acquired by a device is, by definition, a noisy, blurred and downsampled representation of a scene. A better image of the scene can be obtained by combining the information contained in multiple images of the same scene (Farsiu et al., 2004). For this purpose, the images need to be registered to the same reference frame. The resulting image has a higher spatial resolution compared to the original images prior to registration.

In a final experiment, we demonstrated the suitability of the proposed method as the base for performing super resolution. A subset of Dataset 1 containing 9 images of the same eye is registered with the proposed method. This is the maximum amount of images of the same eye present in Dataset 1. A most naive super resolution approach is employed that combines images through bilinear interpolation to obtain an image with a scale factor of 3. Using a naive super resolution approach indicates registration accuracy, as the result is not favored by improvements due to a sophisticated SR method. The resulting super resolved image is compared with F_0 scaled by the same factor, utilizing nearest neighbor interpolation. It is also compared with the scaled F_0 after applying a bilateral filter, a state of the art non-linear filtering method that preserves edges and reduces image noise (Tomasi and Manduchi, 1998).

The results are shown in Fig. 4. F_0 is shown in Fig. 2, left image, 2nd row. The details are extracted from the top-right periphery of the image. The proposed method allows for a more consistent vessel segmentation utilizing (Frangi et al., 1998), producing also a less noisy response. It also performs better Sobel edge detection. This indicates its suitability for enhanced vessel analysis via automated tools.

4. Conclusions

A method for retinal image registration for fundoscopy images is proposed. The proposed method is based on a 3D approach that involves a spherical model of the eye and 3D pose estimation across the two images to be registered. Several experiments demonstrated the increased accuracy and robustness of the proposed approach compared to the relevant state of the art.

Comparative experimental evaluation showed that the proposed method increases considerably the accuracy of registration compared to competitive approaches at decreased computational costs.

In Section 3.3 it was demonstrated that for the particular problem, SIFT allows for more accurate registration results than SURF. It was also shown that combining a RANSAC-based initialization with a fine PSO search around the yielded result produces the most accurate registration while using the least computational resources. Additionally, as a single PSO stage is performed, it presents the most simple parameter configuration of all the considered variants. This is due to RANSAC providing a fast, robust and accurate initial estimation, around which PSO performs a refined search for increased accuracy.

In Section 3.4 the effects of diverse budget distributions were studied. For a fixed budget, a high amount of particles with a low amount of generations may not able to converge to the solution. Also a low amount of particles with a high amount of generations results in a low density of particles, thus leaving unexplored large regions of the search space, which may contain the optimal solution. In both cases, less accurate results were obtained compared to the case where a given budget is achieved with a balanced number of particles and generations.

In Section 3.5 it was shown that the proposed method outperforms both RANSAC and our previous work in Hernandez-Matas et al. (2015). This is due to the proposed method being a combination of the two that takes advantage of the strengths of both methods.

In Section 3.6 it was shown that the proposed method is more robust and accurate than GDB-ICP (Yang et al., 2007), a widely employed method in retinal registration. While GDB-ICP has a weak initialization based on a single keypoint match, the proposed method obtains an initial pose estimation by performing RANSAC. This leads to an increased robustness and a higher amount of image pairs successfully registered. Additionally, the proposed method utilizes a rigid 3D model that approximates the shape of the eye, instead of a general quadratic kernel like the one utilized by GDB-ICP. Even though the utilization of a rigid model might seem to be a drawback, it ultimately proves to provide more accurate registration results.

In Section 3.7 it is shown that the addition of RANSAC for pose initialization in the proposed method allows to greatly reduce its computational cost compared to our previous work (Hernandez-Matas et al., 2015), while also making it less computationally expensive than GDB-ICP.

Finally, in Section 3.8 it was shown that the proposed method is suitable for tasks requiring high accuracy registration. Such a task is super resolution, a technique that allows for producing large, high resolution images from a combination of lower resolution ones, together with its suitability for enhanced vessel analysis via automated tools.

Indicative experiments also showed the applicability of the method to diagnostic support. Moreover, future work aims to incorporate more sources of information in the objective function of Eq. (2). More specifically, this can be amended with further terms, i.e. accumulating edge distances, intensity or gradient vector angle differences, etc. Other extensions regard extending the optimization to more dimensions so as to include the shape and size of the retina as well as the camera parameters.

Acknowledgments

This research was made possible by a Marie Curie grant from the European Commission in the framework of the REVAMMAD ITN (Initial Training Research Network), Project 316990. It was also supported by the FORTH-ICS internal RTD Programme “Ambient Intelligence and Smart Environments”. Authors thank Ruth Ibán-Arias (University of Crete, Greece) for assistance with the statistical significance analysis and the anonymous reviewers for their invaluable feedback.

References

Abramoff, M., Garvin, M., Sonka, M., 2010. Retinal imaging and image analysis. IEEE Rev. Biomed. Eng. 3, 169–208, http://dx.doi.org/10.1109/RBME.2010.2084567.

Adal, K., Ensing, R., Couvert, R., van Etten, P., Martinez, J., Vermeer, K., van Vliet, L., 2014. A hierarchical coarse-to-fine approach for fundus image registration. Biomed. Image Regist., 93–102.

Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110 (3), 346–359.

Can, A., Stewart, C., Roysam, B., Tanenbaum, H., 2002. A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina. IEEE Trans. Pattern Anal. Mach. Intell. 24 (3), 412–419, http://dx.doi.org/10.1109/34.990145.

Cattin, P., Bay, H., Van Gool, L., Szkely, G., 2006. Retina mosaicing using local features. In: Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Vol. 4191, pp. 185–192.

Chanwimaluang, T., Fan, G., Yen, G., Fransen, S.R., 2009. 3-D retinal curvature estimation. IEEE Trans. Inf. Technol. Biomed. 13 (6), 997–1005, http://dx.doi.org/10.1109/TITB.2009.2027014.
Chaudhry, A.R., Klein, J.C., 2008. Ophthalmologic image registration based on shape-context: application to fundus autofluorescence (FAF) images. Vis. Imaging Image Process., 1–7.
Chen, J., Tian, J., Lee, N., Zheng, J., Smith, T., Laine, A., 2010. A partial intensity invariant feature descriptor for multimodal retinal image registration. IEEE Trans. Biomed. Eng. 57 (7), 1707–1718, http://dx.doi.org/10.1109/TBME.2010.2042169.
Choe, T., Cohen, I., Medioni, G., Walsh, A., Sadda, S., 2006. Evaluation of 3-D shape reconstruction of retinal fundus. In: Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, pp. 134–141.
Cideciyan, A., Jacobson, S., Kemp, C., Knighton, R., Nagel, J., 1992. Registration of high resolution images of the retina. SPIE Med. Imaging 1652, 310–322, http://dx.doi.org/10.1117/12.59439.
Farsiu, S., Robinson, M.D., Elad, M., Milanfar, P., 2004. Fast and robust multiframe super resolution. IEEE Trans. Image Process. 13 (10), 1327–1344.
Fischler, M., Bolles, R., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. Assoc. Comput. Mach. 24 (6), 381–395, http://dx.doi.org/10.1145/358669.358692.
Frangi, A., Niessen, W., Vincken, K., Viergever, M., 1998. Multiscale vessel enhancement filtering. In: Medical Image Computing and Computer-Assisted Intervention, pp. 130–137.
Gharabaghi, S., Daneshvar, S., Sedaaghi, M.H., 2012. Retinal image registration using geometrical features. J. Digit. Imaging 26 (2), 248–258.
Gosset, W.S., 1908. The probable error of a mean. Biometrika 6 (1), 1–25, http://dx.doi.org/10.2307/2331554, originally published under the pseudonym "Student".
Grosso, A., Veglio, F., Porta, M., Grignolo, F.M., Wong, T.Y., 2005. Hypertensive retinopathy revisited: some answers, more questions. Br. J. Ophthalmol. 89 (12), 1646–1654, http://dx.doi.org/10.1136/bjo.2005.072546.
Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, 2nd ed. Cambridge University Press.
Hernandez-Matas, C., Zabulis, X., 2014. Super resolution for fundoscopy based on 3D image registration. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, USA, pp. 6332–6338, http://dx.doi.org/10.1109/EMBC.2014.6945077.
Hernandez-Matas, C., Zabulis, X., Argyros, A.A., 2015. Retinal image registration based on keypoint correspondences, spherical eye modeling and camera pose estimation. In: IEEE Engineering in Medicine and Biology Society Conf., Milan, Italy, pp. 5650–5654, http://dx.doi.org/10.1109/EMBC.2015.7319674.
Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N., 2012. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Asian Conf. on Computer Vision, pp. 548–562.
Hubbard, R., Brothers, L., King, W., Clegg, L., Klein, R., Cooper, L., Sharrett, A., Davis, M., Cai, J., 1999. Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study. Ophthalmology 106, 2269–2280.
Kyriazis, N., Argyros, A., 2014. Scalable 3D tracking of multiple interacting objects. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 3430–3437.
Legg, P., Rosin, P., Marshall, D., Morgan, J., 2013. Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation. Comput. Med. Imaging Graph. 37.
Lin, Y., Medioni, G., 2008. Retinal image registration from 2D to 3D. In: IEEE Conf. on Computer Vision and Pattern Recognition, Chicago, USA.
Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.
Matsopoulos, G., Mouravliansky, N., Delibasis, K., Nikita, K., 1999. Automatic retinal image registration scheme using global optimization techniques. IEEE Trans. Inf. Technol. Biomed. 3 (1), 47–60.
Matsopoulos, G., Asvestas, P., Mouravliansky, N., Delibasis, K., 2004. Multimodal registration of retinal images using self organizing maps. IEEE Trans. Med. Imaging 23 (12), 1557–1563, http://dx.doi.org/10.1109/TMI.2004.836547.
Meitav, N., Ribak, E.N., 2011. Improving retinal image resolution with iterative weighted shift-and-add. J. Opt. Soc. Am. A 28 (7), 1395–1402.
Molodij, G., Ribak, E., Glanc, M., Chenegros, G., 2014. Enhancing retinal images by extracting structural information. Optics Commun. 313, 321–328, http://dx.doi.org/10.1016/j.optcom.2013.10.011.
Narasimha-Iyer, H., Can, A., Roysam, B., Tanenbaum, H.L., Majerovics, A., 2007. Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences. IEEE Trans. Biomed. Eng. 54 (8), 1436–1445, http://dx.doi.org/10.1109/TBME.2007.900807.
Navarro, R., Santamaría, J., Bescós, J., 1985. Accommodation-dependent model of the human eye with aspherics. J. Opt. Soc. Am. A 2 (8), 1273–1281.
Oikonomidis, I., Kyriazis, N., Argyros, A., 2011. Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints. In: Int. Conf. on Computer Vision, pp. 2088–2095.
Padeleris, P., Zabulis, X., Argyros, A., 2012. Head pose estimation on depth data based on particle swarm optimization. In: IEEE Conf. Computer Vision and Pattern Recognition Workshops, pp. 42–49.
Panteleris, P., Argyros, A.A., 2014. Vision-based slam and moving objects tracking for the perceptual support of a smart walker platform. In: European Conf. on Computer Vision Workshops, pp. 407–423.
Perez-Rovira, A., Cabido, R., Trucco, E., Mckenna, S., Hubschman, J.P., 2011. RERBEE: Robust Efficient Registration via Bifurcations and Elongated Elements applied to retinal fluorescein angiogram sequences. IEEE Trans. Med. Imaging 31, 1–11.
Pluim, J., Maintz, A., Viergever, M., 2003. Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imaging 22 (8), 986–1004, http://dx.doi.org/10.1109/TMI.2003.815867.
Poli, R., Kennedy, J., Blackwell, T., 2007. Particle swarm optimization. Swarm Intell. 1 (1), 33–57.
Reel, P., Dooley, L., Wong, K., Börner, A., 2013. Multimodal retinal image registration using a fast principal component analysis hybrid-based similarity measure. In: Int. Conf. on Image Processing.
Ryan, N., Heneghan, C., de Chazal, P., 2004. Registration of digital retinal images using landmark correspondence by expectation maximization. Image Vis. Comput. 22 (11), 883–898, http://dx.doi.org/10.1016/j.imavis.2004.04.004.
Stewart, C., Tsai, C., Roysam, B., 2003. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Trans. Med. Imaging 22 (11), 1379–1394, http://dx.doi.org/10.1109/TMI.2003.819276.
Tang, Li, Garvin, M.K., Lee, Kyungmoo, Alward, W.L.W., Kwon, Y.H., Abramoff, M.D., 2011. Robust multiscale stereo matching from fundus images with radiometric differences. IEEE Trans. Pattern Anal. Mach. Intell. 33 (11), 2245–2258, http://dx.doi.org/10.1109/TPAMI.2011.69.
Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images. In: International Conf. on Computer Vision, p. 839.
Troglio, G., Alberti, M., Benediksson, A., Moser, G., 2010. Unsupervised change-detection in retinal images by a multiple-classifier approach. In: Multiple Classifier Systems, pp. 94–103.
Tsai, C., Li, C., Yang, G., Lin, K., 2010. The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. IEEE Trans. Med. Imaging 29 (3), 636–649, http://dx.doi.org/10.1109/TMI.2009.2030324.
Yang, G., Stewart, C., Sofka, M., Tsai, C., 2007. Registration of challenging image pairs: initialization, estimation, and decision. IEEE Trans. Pattern Anal. Mach. Intell. 29 (11), 1973–1989, http://dx.doi.org/10.1109/TPAMI.2007.1116.
Zhang, X., Hu, W., Maybank, S., Li, X., Zhu, M., 2008. Sequential particle swarm optimization for visual tracking. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.
Zheng, J., Tian, J., Deng, K., Dai, X., Zhang, X., Xu, M., 2011. Salient feature region: a new method for retinal image registration. IEEE Trans. Inf. Technol. Biomed. 15 (2), 221–232, http://dx.doi.org/10.1109/TITB.2010.2091145.