

A Comparative Study of Transformation Models for the Sequential Mosaicing of Long Retinal Sequences of Slit-Lamp Images Obtained in a Closed Loop Motion

Kristina Prokopets^{1,2} · Adrien Bartoli¹

Received: / Accepted:

Abstract

Purpose. Navigated Pan Retinal Photocoagulation (PRP) is a standard care for proliferative diabetic retinopathy. Slit-lamp based systems used for this treatment provide a narrow view of the retina. Retinal mosaics are used for view expansion and treatment planning. Mosaicing slit-lamp images is a hard task due to the absence of a physical model of the imaging process, large textureless regions and imaging artifacts, mostly reflections.

Methods. We present a comparative study of various geometric transformation models applied to retinal image mosaicing in computer-assisted slit-lamp imaging. We propose an efficient point correspondence based framework for transformation model evaluation in a typical closed loop motion scenario. We compare the performance of multiple linear and non-linear models of different complexity and assess the effect of number of points used for parameter estimation. We use a Local Fitting Error (LFE) metric to estimate the models' performance in pairwise registration. Because LFE alone is not conclusive regarding the problem of accumulated drift, we propose a Loop Closure Error (LCE) metric to quantify the effect of accumulated local registration errors. We also provide a new normalization procedure for the quadratic transformation model, widely used in retinal image registration.

Results. In total seven transformation models were evaluated on three datasets of long image sequences. LFE decreases with increasing complexity of the model while LCE, in contrast, shows superior performance of simple models. Varying the number of point correspondences did not reveal a common trend for the LCE metric, showing an increase of the error for simple models and an unstable behavior of the complex models.

Conclusion. Our results show that simple models are less sensitive to drift and preferable for sequential mosaicing in slit-lamp imaging, while more complex models are the best choice for short-term registration.

Keywords transformation · model · drift · mosaicing · retina · image registration

¹ALCoV-ISIT, UMR 6284 CNRS / Université d'Auvergne
28 place Henri Dunant, 63001 Clermont-Ferrand, France

²QuantelMedical, 11 Rue du Bois Joli, 63808 Cournon-d'Auvergne, France
E-mail: prokopets.kristina@mail.ru, adrien.bartoli@gmail.com

1 Introduction

Proliferative diabetic retinopathy is a major cause of visual impairment and blindness among people with diabetes. Early detection and treatment can reduce the risk of blindness [11]. The currently used standard treatment is Pan Retinal Photocoagulation (PRP) where a therapeutic laser is used to make tiny burns on the affected areas of the retina. Conventional PRP through the slit-lamp with manual navigation has been the standard for laser delivery for many decades. However, sequential improvements in the past ten years led to the development of computer guided photocoagulation systems with integrated imaging and automatic navigation [6,9,10]. Although the visible area of the retina is smaller with the slit-lamp, compared with images acquired by non-mydratic fundus photography [9,4], magnification, expansion and control offered by the slit-lamp makes it a very popular choice for laser delivery in the clinical environment [2]. The imaging set-up is based on the eyepiece and microscope optics of the slit-lamp and the magnifying contact lens attached to the eye such that slit illumination is projected onto the retina (Figure 1a). This set-up is used to perform retinal examination and treatment where the ophthalmologist typically explores the retina in closed loop manner starting from the optic nerve. Images obtained with this device have a narrow field of view (FOV). Thus, only small thin portions of the retina can be visualized (Figure 1b). Obtaining the larger view can greatly facilitate the delivery of PRP and treatment planning.

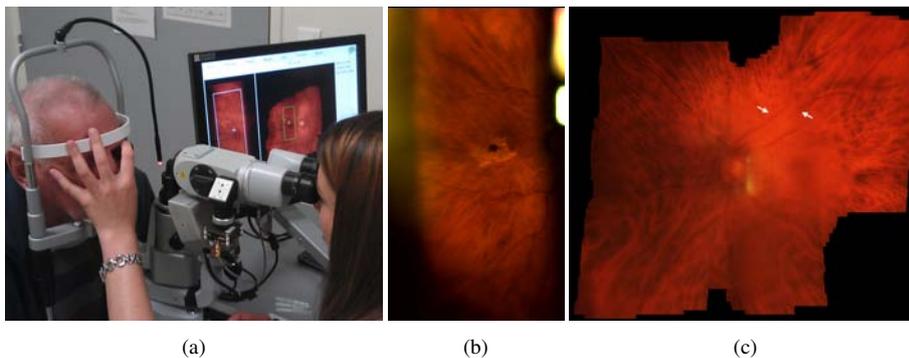


Fig. 1: Retinal image mosaicing with a slit-lamp. (a) slit-lamp system developed in QuantelMedical, France, (b) typical slit-lamp image, (c) example of drift on the mosaic: the white arrows show a misaligned vessel.

Computer-assisted retina mosaicing for view expansion using a slit-lamp device has been recently proposed [18,17] and integrated in the slit-lamp system of QuantelMedical, France (Figure 1a). A map of the retina is built in real-time using a combination of direct and feature-based tracking. Direct visual tracking uses color discrepancy to infer motion without the need to extract image features [8,21]. The objective is to find the parameters of a transformation model that aligns the reference image of a target object to its current image so that a chosen similarity measure between the two images is maximized. It is combined with feature-based tracking to perform re-initialization if the target is lost [17]. Retinal mosaics built with this method suffer from accumulated drift induced by sequential mapping (Figure 1c). This causes misalignment artifacts and degrades the visual quality. The influen-

tial factors are (i) the small number of features away from the optic nerve, (ii) the geometric transformation model, (iii) the distortion induced by the geometry of the eye and the contact lens and (iv) the fitting algorithm.

The majority of existing works on retinal image registration and mosaicing uses images obtained from a fundus camera [1,23,19,5,22]. The quality of this type of images is higher. They have fewer specular reflections, good contrast and almost no blur. The transformation models applied in these works include translation, rigid (translation and rotation), similarity, affine and quadratic. On the other hand, slit-lamp images were covered only in [2,18]. This type of data is degraded by an uneven illumination which comes from outside the eye, especially from the contact lens. It creates viewpoint dependent artifacts, glare and specular reflections. The results recently presented in [18] demonstrate the use of the rigid transformation model. Despite the variety of works which report on different transformation models for retinal image registration, only a few address their comparison and evaluation. Four transformation models (translation, affine, quadratic and radial distortion correction) were evaluated on different image modalities in [12,14]. These works, however, do not consider the mosaicing of long image sequences obtained in a closed loop motion which is exemplary for retinal examination with the slit-lamp. Thus, they do not address the problem of accumulated registration errors and drift.

We present a comparative study of seven geometric transformation models on the subject of their performance in sequential retinal image mosaicing of slit-lamp images. We evaluate multiple models from existing works on retina image mosaicing as well as the homography and the Thin-Plate Spline (TPS). We independently investigate the effects of model's complexity and the number of point correspondences on drift accumulation. We propose a new evaluation framework and error metric to compute the amount of drift. We also derive a normalization procedure for the quadratic model [5]. The normalization of the homography is well known [7] but for the quadratic model, however, it was not. Our normalization method improves the quadratic model fitting.

2 Materials and Methods

2.1 Slit-lamp Imaging and Geometric Assumptions

We used image sequences of retinal examination performed on volunteers in a University Hospital of Saint-Etienne, France. The navigated PRP system developed at QuantelMedical was used. The images were captured with a CCD camera at 60fps. Typical videos are between 2-3 minutes long. The retina is illuminated with a narrow light beam focused using a direct contact lens. The standard way of retinal examination is to perform a closed loop motion starting from the optic nerve, moving to the periphery and coming back. The camera is fixed on the moving base controlled by the ophthalmologist and undergoes translation only. Small rotations caused by head tilts occasionally occur. The spherical curvature of the retina has relatively low depth variation. The system's optics include several parts moving independently, namely the contact lens and the camera. Therefore, *the imaging device cannot be calibrated* (the relationship between a pixel's position in an image and the corresponding line of sight varies in time). Thus, there is no simple physically valid transformation to relate the images geometrically. This makes mosaicing tremendously difficult.

2.2 Transformation Models

Previous works do not conclude on which model can best approximate the image transformation in retinal image mosaicing. Thus, we have specifically chosen to evaluate the following seven transformation models: **T** - translation, as an intuitive choice reflecting the lateral motion of the camera; **RG** - rigid, is currently integrated in the mosaicing algorithm used in the slit-lamp device of QuantelMedical; **SM** - similarity and **AF** - affine models were chosen to check whether the modeling of slight eye movements during procedure improves accuracy; **H** - homography, as the widely used model in mosaicing [7, 20]; **QD** - quadratic, as a popular choice in retinal image registration [5]; and finally the **TPS** - Thin-Plate Spline with adaptive parameter smoothing [3] which might have a great potential of success due to its elastic properties. The properties of these models are summarized in Table 1.

	T	RG	SM	AF	H	QD	TPS
DoF	2	3	4	6	8	12	$2k$
Linear w.r.t. source points	yes	yes	yes	yes	no	no	no
Linear w.r.t. parameters	yes	no	yes	yes	no	yes	yes

Table 1: Summary of the transformation models’ characteristics. The Degrees of Freedom (DoF) define the number of estimated parameters. We label each model according to whether it is linear w.r.t. its parameters or the source point. k indicates the number of control points of the TPS.

3 Evaluation Framework

To evaluate the models’ accumulated drift we propose a point correspondence based framework. The principle is to provide a noisy but outlier-free set of correspondences to minimize the effect of the fitting algorithm and evaluate the drift with an independent set of points transferred through a closed loop motion. We evaluate pairwise fitting and quantify how the model is able to connect the last and first frames in long-term image registration without using the closed loop constraint. Our framework consists of four main steps: (1) data acquisition and processing, (2) point correspondence selection, (3) transformation parameter estimation and (4) model accuracy evaluation through a number of tests. The details of each step are given below.

3.1 Data Acquisition

Three datasets were used in this study (Figure 2). Each dataset consists of an image sequence obtained from a retinal examination video where every 5th frame was taken. Two datasets (Figures 2a-b) were obtained from retinal examination videos of patients in the hospital. One dataset of a phantom eye was included as a simplified case where the phantom was fixed on a holder and the procedure did not involve a contact lens (Figure 2c). The length of the datasets is 254, 242 and 326 images respectively. The image size is 720×1280 pixels. The numbers

of point correspondences for each dataset were not the same, resulting in as minimum 100 points per pair of frames and as maximum 400. Frames containing the minimum number of points were mostly on the periphery of the retina while frames containing more points were closer to the optic nerve. The size of the illumination slit was fixed according to the patients' comfort for the first two datasets. The visible part of the retina excluding regions of strong specularities covers at least 50% of the image.

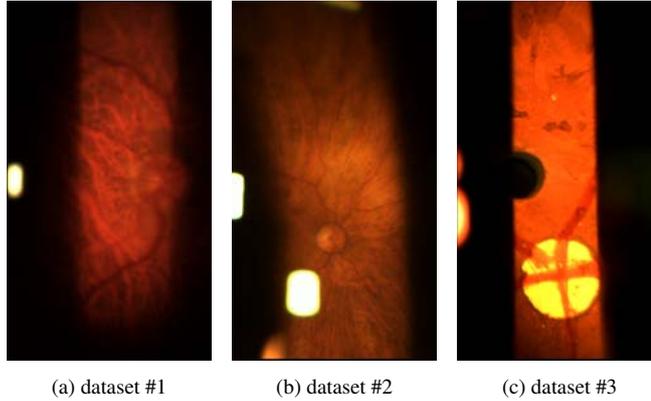


Fig. 2: Sample image from each dataset.

3.2 Selection of Pairwise Point Correspondences

We segment the visible part of the retina and filter out strong specularities using intensity thresholding and morphological operations [17]. We then detect and extract key-points with Scale Invariant Feature Transform (SIFT) [15] and match them between consecutive frames. Matching is performed by measuring the L_2 norm of the difference between key-point descriptors within a pair of frames, and the basic matching algorithm suggested by [15] to reject matches that are too ambiguous. A combination of automated and manual refinement steps are incorporated to exclude the remaining outliers. First we use a threshold on the points' relative displacement. The threshold is defined by summing the median and the median of absolute deviation of the points' displacement. Points which moved more than the computed threshold are discarded. Second, the manual checkup is performed with every set of point correspondences visualized on the associated images. The position of erroneous points is adjusted manually using a specifically developed Graphical User Interface (GUI) in Matlab. Thus, each dataset contains between 100 and 400 correspondences $\mathbf{p} \longleftrightarrow \mathbf{q}$ obtained from f frame pairs in a closed loop $I_1 \longleftrightarrow I_2 \longleftrightarrow \dots \longleftrightarrow I_{f-1} \longleftrightarrow I_f \longleftrightarrow I_1$.

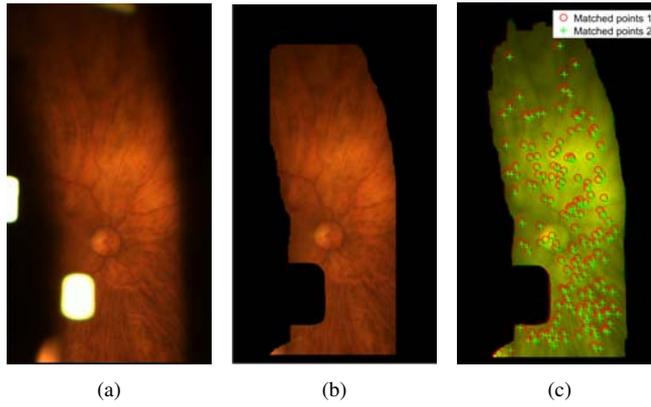


Fig. 3: Selection of point correspondences. (a) - original image, (b) - segmented image, (c) - point correspondences (subsamped for legibility).

3.3 Transformation Parameter Estimation

3.3.1 General Points

A transformation function has the form $w(\mathbf{p}, \theta)$ where θ is a vector of transformation parameters. All transformations are estimated by minimizing the sum of squared transfer discrepancies. Due to the numerical instability of models containing cross terms and/or squared terms, as the homography and the quadratic models, the estimates might not be stable. This may be improved by data normalization, which has been well-studied for the homography [7] but not for the quadratic model. Typical image points may have various orders of magnitude. Their increase in squared and cross terms may cause the pixel coordinates to become very large. Normalization converts pixel coordinates e.g. $\mathbf{p} \in [1; 1000]$ to normalized coordinates e.g. $\mathbf{p} \in [-1; 1]$. This is done by a simple affine transformation. The detailed explanation and substantiation of the normalization procedure for the homography can be found in [7] (Ch. 4.4.4). The question is whether it is possible to normalize the quadratic model with an affine transform; the answer is *yes*.

3.3.2 Quadratic Model Normalization Rules

We derive normalization w.r.t. the rules of function compositions. Let $N(\mathbf{p}) = \mathbf{S}\mathbf{p} + \mathbf{c}$, where $\mathbf{S} \in \mathbb{R}^{2 \times 2}$, $\mathbf{c} \in \mathbb{R}^2$ be the normalization transform applied to the point correspondences from two consecutive frames. Let \tilde{Q} be the quadratic model estimated from normalized data, using [13] for instance. Thus, to compute Q the quadratic transform in pixel coordinates we write:

$$Q(\mathbf{p}) = (D' \circ \tilde{Q} \circ N)(\mathbf{p}) = D'(\tilde{Q}(N(\mathbf{p}))) \quad (1)$$

where $D'(\mathbf{p}) = \mathbf{E}\mathbf{p} + \mathbf{k}$, where $\mathbf{E} \in \mathbb{R}^{2 \times 2}$, $\mathbf{k} \in \mathbb{R}^2$ is the denormalization transform from the second image such that $N' \circ D' = D' \circ N' = I$.

The quadratic model is the second order Taylor series expansion of the general transformation [5]:

$$Q(\mathbf{p}) = [\mathbf{B}_{2 \times 3} | \mathbf{A}_{2 \times 2} | \mathbf{t}_{2 \times 1}] X(\mathbf{p}) \quad (2)$$

where $\mathbf{B} \in \mathbb{R}^{2 \times 3}$, $\mathbf{A} \in \mathbb{R}^{2 \times 2}$, $\mathbf{t} \in \mathbb{R}^{2 \times 1}$ are the 2^{nd} , 1^{st} and 0^{th} order terms of the transformation, $X(\mathbf{p}) = [x^2, xy, y^2, x, y, 1]^T$. We define a symmetric matrix $\hat{\mathbf{B}}_x \in \mathbb{R}^{2 \times 2}$ to represent the quadratic and cross terms as:

$$\boldsymbol{\mu}(\mathbf{b}_x) \stackrel{\text{def}}{=} \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} \\ \frac{1}{2}b_{12} & b_{13} \end{bmatrix} = \hat{\mathbf{B}}_x, \quad \mathbf{v}(\hat{\mathbf{B}}_x) \stackrel{\text{def}}{=} \begin{bmatrix} \hat{b}_{11} \\ 2\hat{b}_{12} \\ \hat{b}_{22} \end{bmatrix} = \mathbf{b}_x \quad (3)$$

where $\mathbf{b}_x^T \in \mathbb{R}^{1 \times 3}$ is the first row of \mathbf{B} and $\boldsymbol{\mu}(\mathbf{b}_x)$ is the ‘packing’ vector to matrix form and $\mathbf{v}(\hat{\mathbf{B}}_x)$ its ‘unpacking’. This is a simple reorganization of model’s entries. Thus, with $\boldsymbol{\mu} \circ \mathbf{v} = id$ and $\mathbf{v} \circ \boldsymbol{\mu} = id$, we have:

$$\mathbf{v}(\hat{\mathbf{B}}_x)^T \begin{bmatrix} x^2 \\ xy \\ y^2 \end{bmatrix} = \mathbf{p}^T \hat{\mathbf{B}}_x \mathbf{p}, \quad \mathbf{b}_x^T \begin{bmatrix} x^2 \\ xy \\ y^2 \end{bmatrix} = \mathbf{p}^T \boldsymbol{\mu}(\mathbf{b}_x) \mathbf{p} \quad (4)$$

Each dimension of \tilde{Q} can then be written as:

$$\tilde{Q}_x(\mathbf{p}) = \mathbf{p}^T \hat{\mathbf{B}}_x \mathbf{p} + \mathbf{a}_x^T \mathbf{p} + t_x \quad (5)$$

where $\mathbf{a}_x^T \in \mathbb{R}^{1 \times 2}$ is the first row of \mathbf{A} and t_x is the first element of \mathbf{t} .

First, to compose the quadratic model with a normalization transform N we use composition rules expressed in equation (1) and equation (5). We write the composition as follows:

$$\begin{aligned} (\tilde{Q}_x \circ N)(\mathbf{p}) &= \frac{1}{2}(\mathbf{S}\mathbf{p} + \mathbf{c})^T \hat{\mathbf{B}}_x (\mathbf{S}\mathbf{p} + \mathbf{c}) + \mathbf{a}_x^T (\mathbf{S}\mathbf{p} + \mathbf{c}) + t_x \\ &= \frac{1}{2} \mathbf{p}^T \mathbf{S}^T \hat{\mathbf{B}}_x \mathbf{S} \mathbf{p} + (\mathbf{c}^T \hat{\mathbf{B}}_x + \mathbf{a}_x^T) \mathbf{S} \mathbf{p} + \left(\frac{1}{2} \mathbf{c}^T \hat{\mathbf{B}}_x + \mathbf{a}_x^T \right) \mathbf{c} + t_x \\ &= \left[\frac{1}{2} \mathbf{v}(\mathbf{S}^T \hat{\mathbf{B}}_x \mathbf{S})^T (\mathbf{c}^T \hat{\mathbf{B}}_x + \mathbf{a}_x^T) \mathbf{S} \left(\frac{1}{2} \mathbf{c}^T \hat{\mathbf{B}}_x + \mathbf{a}_x^T \right) \mathbf{c} + t_x \right] X(\mathbf{p}) \end{aligned} \quad (6)$$

which shows that $\tilde{Q}_x \circ N$ is a quadratic transformation which follows that $\tilde{Q}_y \circ N$ is a quadratic transformation too.

To compose the denormalization transform D' with the quadratic model resulting from equation (6), we follow the previous derivation and write the composition as follows:

$$(D' \circ \tilde{Q})(\mathbf{p}) = \mathbf{E}(\mathbf{B}[\mathbf{A}|\mathbf{t}] X(\mathbf{p})) + \mathbf{k} = [\mathbf{EB}|\mathbf{EA}|\mathbf{Et} + \mathbf{k}] X(\mathbf{p}) \quad (7)$$

which shows that $D' \circ \tilde{Q}$ is a quadratic transformation. Consequently, this establishes that, $D' \circ \tilde{Q} \circ N$ is a quadratic transformation too and that normalized estimation of the quadratic transformation is possible.

3.3.3 Normalized Estimation of the Quadratic Transformation

The following steps summarize the normalization procedure and parameter estimation for the quadratic model:

1. **Normalize:** compute N and D' using point correspondences from two consecutive images $\mathbf{p} \longleftrightarrow \mathbf{q}$.
2. **Fit \tilde{Q} :** apply the LLS algorithm [13] to the point correspondences $\mathbf{p} \longleftrightarrow \mathbf{q}$ to obtain \tilde{Q} .
3. **Find Q :** compute $(\tilde{Q}_x \circ N)(\mathbf{p})$ using equation (6) and apply equation (7) to obtain the final Q .

We denote the normalized quadratic model as **QDn** and include it for evaluation. The effect of this normalization is also discussed and illustrated in §4.1.

3.4 Evaluation

To independently evaluate the effect of the model complexity and of the number of point correspondences we analyze two types of error metrics (Figure 4 and 5). We compute the Local Fitting Error (LFE) - the discrepancy of data point and corresponding model estimate in pixels, to evaluate model fitting in pairwise registration as follows:

$$\xi_{LFE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \| \mathbf{q}_i - w(\mathbf{p}_i, \theta) \|^2} \quad (8)$$

where $\mathbf{p}_i \longleftrightarrow \mathbf{q}_i, i = 1, \dots, n$ are all the point correspondences.

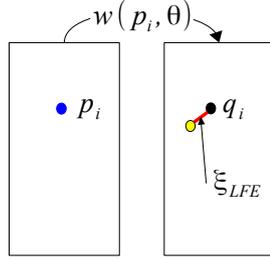


Fig. 4: The Local Fitting Error quantifies pairwise registration error.

We propose a Loop Closure Error (LCE) metric. This shows how the composition of estimated transformations affects the global registration and accumulated drift. The idea is to initialize a uniform grid of points g_1, \dots, g_l at the first frame of the sequence and use the set of pairwise estimated transformations applied sequentially to transfer the grid throughout the sequence. The discrepancy between the initial and resulting sets of points is then measured in pixels as follows:

$$\xi_{LCE} = \sqrt{\frac{1}{l} \sum_{i=1}^l \| g_i - \zeta \|^2} \quad (9)$$

where $\zeta = w(\dots(w(g_i, \theta_{1,2})), \dots, \theta_{f,1})$.

4 Results and Discussion

Our evaluation has two parts. In Part I the ξ_{LFE} and ξ_{LCE} metrics were computed for every model on three datasets where all the pairwise point correspondences were used for parameter estimation. This is to analyze how the model complexity affects the local registration error and accumulated drift. The narrow FOV, poorly textured regions of the retina and small amount of landmarks sometimes complicates the automatic detection of a sufficient number of point correspondences. A suitable transformation model has to cope with this limitation. Thus, in Part II we study the effect of varying the number of point correspondences.

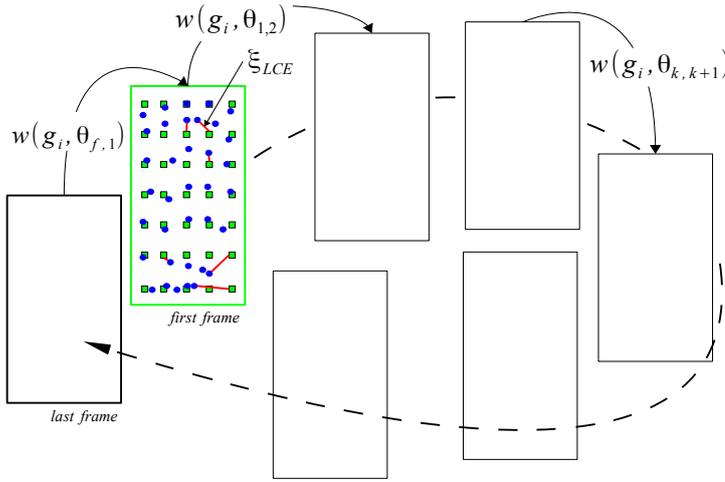


Fig. 5: The proposed Loop Closure Error quantifies how well the model captures the closed loop motion.

4.1 Part I: Effect of Model Complexity

The results show that ξ_{LFE} decreases with increasing complexity of the model (Table 2). We found that T, RG and SM provide similar results. The difference between RG and SM is negligible. This is because isotropic scaling is almost minimal in slit-lamp imaging. H, despite its complexity over AF, generally gives similar results to AF and even inferior in datasets #2 and #3. This raises the question of whether perspective matters. The answer would be *no*. Modeling perspective is not useful for curved retina and purely lateral motion of the camera. Finally, the TPS provides the smallest ξ_{LFE} in datasets #1 and #2 and QD gives the smallest ξ_{LFE} in dataset #3.

	dataset # 1		dataset # 2		dataset # 3	
	ξ_{LFE}	ξ_{LCE}	ξ_{LFE}	ξ_{LCE}	ξ_{LFE}	ξ_{LCE}
T	3.185	62.906	3.165	18.441	3.179	59.892
RG	3.164	57.262	3.066	50.288	3.161	75.162
SM	3.162	72.473	3.064	49.743	3.158	76.175
AF	3.105	102.150	2.986	78.785	3.056	221.050
H	3.073	201.950	3.000	333.650	3.066	351.390
TPS	3.019	125.150	2.864	275.920	2.971	191.790
QD	$F(28)$	$F(28)$	$F(149)$	$F(149)$	2.762	478.070
QDn	$F(56)$	$F(56)$	2.866	254.870	2.886	236.330

Table 2: Average ξ_{LFE} and ξ_{LCE} across the different datasets.

ξ_{LCE} , in contrast, shows the superior performance of simpler models. RG gives the smallest ξ_{LCE} for dataset #1, while T is best in datasets #2 and #3. Following the same pattern as for ξ_{LFE} , RG and SM have errors with difference close to 1 pixel for datasets #2

and #3. However this does not hold for dataset #1. As one can see the difference in ξ_{LCE} between T and RG in dataset #1 is small (only 5.644 pixels) while for datasets #2 and #3 it is much larger (31.847 and 15.27 pixels respectively). This indicates that the rotation component of the model was completely redundant when the patient froze during examination (dataset #2) and the phantom eye was fixed to the holder (dataset #3). AF and TPS showed close results in datasets #1 and #3 while for dataset #2 ξ_{LCE} differs considerably. Additionally, ξ_{LFE} was similar between AF and TPS for dataset #2. This means a small impact of affine deformations in datasets #1 and #3. H appeared to be the worst model.

QD was derived specifically to fit the curved retina [5]. However, it turned out that its estimation from our data is not stable. As one can see this model gives the smallest ξ_{LFE} for dataset #3 only. This, somehow, correlates with results described in the literature [1, 23, 19, 5]. However, this model completely fails in ξ_{LCE} as indicated with $F(x)$ where x is a number of the frame where failure occurred. Indeed, the accumulated drift causes some models to prematurely stop registration before the end of the sequence. In such case, the model contains numerically unstable parameter combinations (quadratic and cross terms) which force point coordinates to become very large if a ‘faulty estimate occurs in the process of chaining for ξ_{LCE} computation. Therefore, when the points tend to be in a degenerate configuration it is the most sensitive model. Thus, we rule out QD from the next experiment. Our normalization method improves the fitting of the quadratic model. Results for dataset #1 showed that failure has been delayed by QDn for 28 frames. The failure was completely eliminated in datasets #2 and #3. We illustrate this improvement with graph plots in Figure 6. We show the ‘spread’ of the points from the uniform grid defined for ξ_{LCE} computation. This demonstrates the model response to scene geometry at central and peripheral portions of the retina. The dashed black lines indicate the frame when failure occurred. One can observe that QDn provides the smallest ξ_{LFE} for dataset #3 and nearly the same ξ_{LFE} as TPS for dataset #2. One can see that normalization suppressed the effect of quadratic part making QDn fit similar to AF in dataset #3. Examples of registered image pairs highlighting areas in which the output of the evaluated transformation models differ are shown in Figure 7.

4.2 Part II: Effect of the Number of Points

We defined the minimum and maximum number of samples as 20 and 100 respectively and computed the ξ_{LFE} and ξ_{LCE} by selecting points randomly with steps of 2 samples. We made 50 trials and averaged the results. Results for this evaluation test are similar among the three datasets. The example of dataset #2 is shown in Figure 8. All transformations show an increase in ξ_{LFE} approximately 0.5 pixels with an increase in the number of point correspondences. This happens because more data brings more constraints to the estimated parameters. However, there is no common trend among results on ξ_{LCE} . Varying subsets of point from 20 to 100 lead ξ_{LCE} to decrease approximately 1.5 times for T, RG, SM and AF. It also decreased approximately 2.5 times for H. One can see that H shows high variance when the number of points is not sufficient and stabilizes only when more than 50 points are supplied. TPS showed a decreasing trend between 20 and 45 points followed by unstable behavior in 45-78 points and starts increasing between 78 to 100 points. This instability is due to the number of control points used to define the deformation grid in TPS, it was constant despite of changing the number of point correspondences. QDn started to give meaningful results only when 68 points were supplied for estimation. It showed unstable behavior with varying ξ_{LCE} from 256 to 263 between 70 and 100 points. This indicates that this model is very sensitive to the number of points.

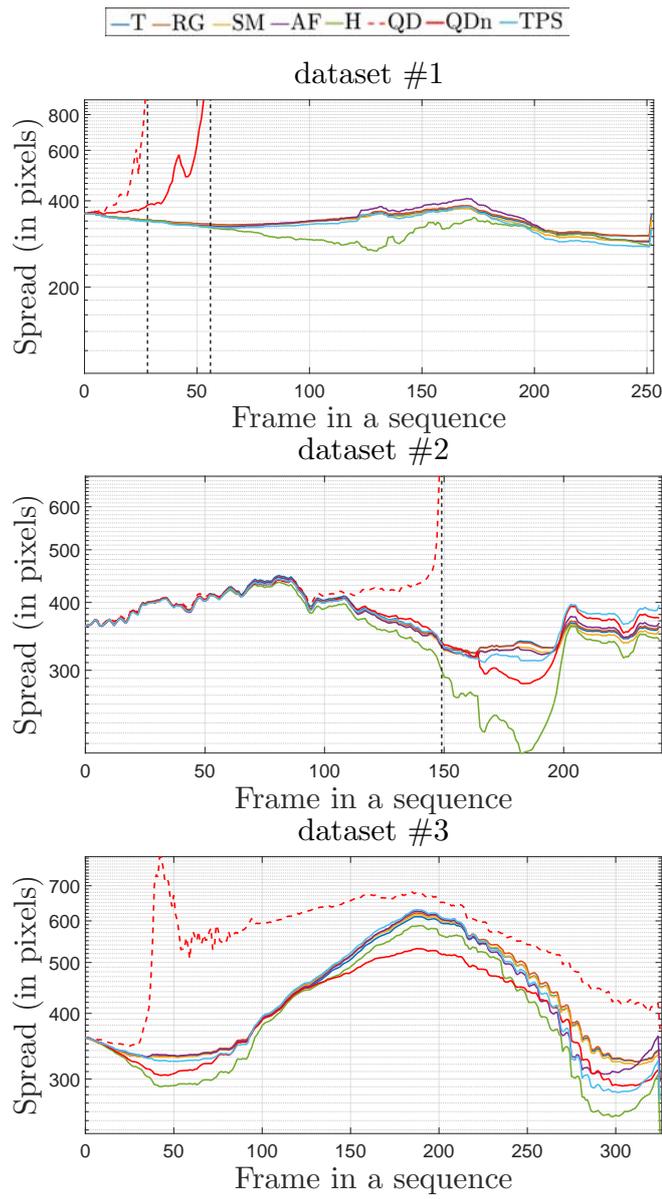


Fig. 6: The ‘spread’ evaluation results over different datasets without subsampling.

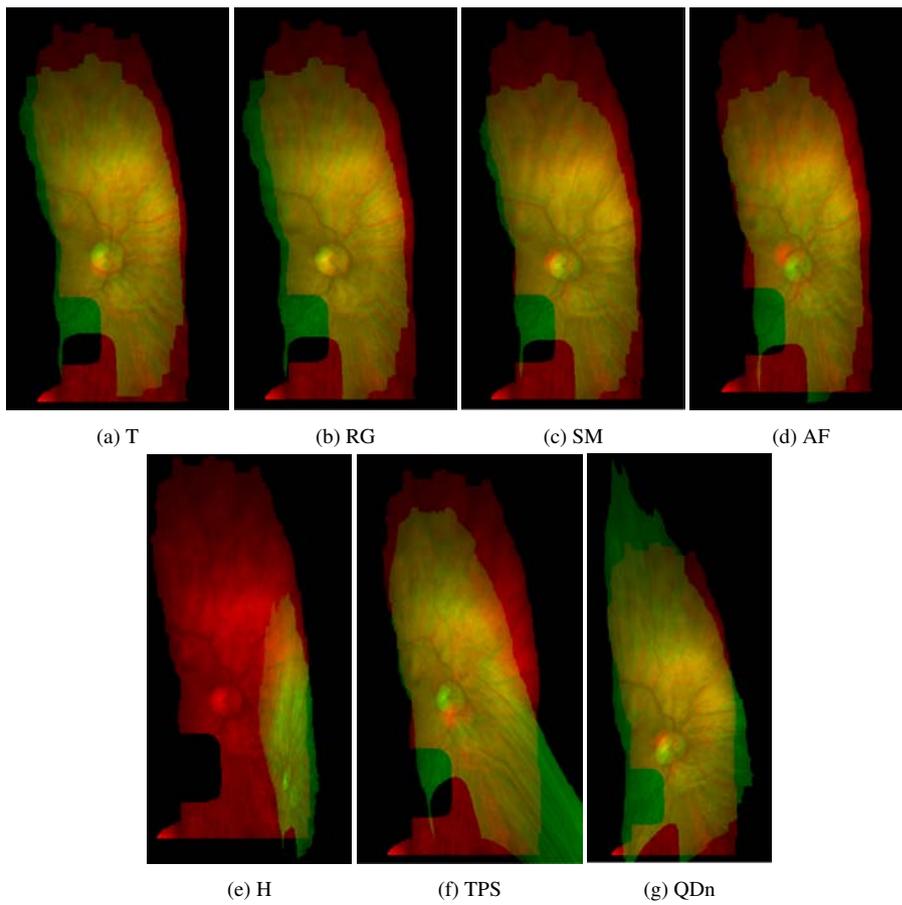


Fig. 7: Examples of registered image pairs with different transformation models. The images are taken from dataset #2. The first image of the sequence is registered with the last image by applying the set of 241 pairwise estimated transformations sequentially.

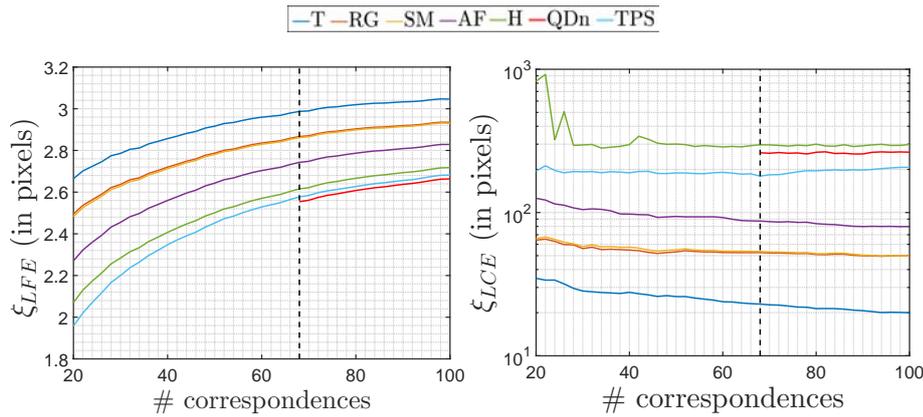


Fig. 8: Effect of the number of points. Example of dataset #2.

5 Conclusion

We have presented a comparative study of transformation models applied to sequential retinal image mosaicing in computer-assisted slit-lamp imaging. We evaluated different models on the subject of drift accumulation. We proposed the point correspondence based evaluation framework and the LCE metric to quantify the drift. We also derived a new normalization procedure to improve the quadratic model's fitting.

The results obtained on three datasets have shown that local registration error decreases with increasing complexity of the transformation model while simple models appeared to produce less accumulated drift. The homography turned out to be irrelevant as perspective deformations might be considered absent. Despite its popularity in the applications of retinal image registration the quadratic model turned out to be completely unstable on our data even after improvement by our proposed normalization procedure. Thus, the choice is meant to be done from the remaining models, namely translation, similarity, rigid, affine and TPS.

The translation, despite of the lowest accumulated drift, is too simple for the majority of clinical cases where the patients are normally very photosensitive and cannot completely freeze during the procedure. The rigid transformation, which is currently used in the system (Figure 1a) and can model rotations, is not sufficiently flexible. The similarity covers isotropic scaling which sometimes occur during the examination. The affine model represents a superset for translation, rigid and similarity models. It covers more deformation types and provides better results. The TPS is complex but the adaptive smoothing makes it always stiff causing, however, a large drift. Therefore, in sequential mosaicing with long slit-lamp image sequences the simple models, specifically translation, rigid, similarity and affine can be the choice among others. However, an affine model is the best possible compromise between ability to model pairwise transformation and simplicity in dealing with drift. The models with higher complexity are best for short-term registration on different types of data.

This work is planned to be extended to replace the rigid model currently used in QuantelMedical's slit-lamp device. The idea is to establish long-term inter-frame point correspondences with the affine transformation model. This will allow us to proceed with Bundle

Adjustment as a global refinement step [16]. Using the affine model is fundamental to provide a sensible estimate to initialize Bundle Adjustment.

6 Disclosures

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Adal, K. M. and Ensing, R. M. and Couvert, R. and Van Etten, P. and Martinez, J. P. and Vermeer, K. A. and Van Vliet, L.J.: A hierarchical coarse-to-fine approach for fundus image registration. In: *Biomedical Image Registration*, pp. 93–102. Springer International Publishing (2014)
2. Asmuth, J., Madjarov, B., Sajda, P., Berger, J.W.: Mosaicking and enhancement of slit lamp biomicroscopic fundus images. *British journal of ophthalmology* **85**(5), 563–565 (2001)
3. Bartoli, A.: Maximizing the predictivity of smooth deformable image warps through cross-validation. *Journal of Mathematical Imaging and Vision* **31**(2-3), 133–145 (2008)
4. Broehan, A.M., Rudolph, T., Amstutz, C.A., Kowal, J.H.: Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system. *IEEE Transactions on Biomedical Engineering* **58**(10), 2816–2824 (2011)
5. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicking the curved human retina. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(3), 412–419 (2002)
6. Chhablani, J., Mathai, A., Rani, P., Gupta, V., Arevalo, J.F., Kozak, I.: Comparison of Conventional Pattern and Novel Navigated Panretinal Photocoagulation in Proliferative Diabetic Retinopathy Comparison of PASCAL and NAVILAS for PRP. *Investigative ophthalmology & visual science* **55**(6), 3432–3438 (2014)
7. Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision*. Cambridge university press (2003)
8. Irani, Michal and Anandan, P.: About direct methods. In: *Vision Algorithms: Theory and Practice*, pp. 267–277. Springer Berlin Heidelberg (1999)
9. Kernt, M., Cheuteu, R., Vounotrypidis, E., Haritoglou, C., Kampik, A., Ulbig, M.W., Neubauer, A.S.: Focal and panretinal photocoagulation with a navigated laser NAVILAS®. *Acta ophthalmologica* **89**(8), 662–664 (2011)
10. Kernt, M., Cheuteu, R.E., Cserhati, S., Seidensticker, F., Liegl, R.G., Lang, J., Haritoglou, C., Kampik, A., Ulbig, M.W., Neubauer, A.S.: Pain and accuracy of focal laser treatment for diabetic macular edema using a retinal navigated laser (Navilas). *Clin Ophthalmol* **6**(1), 289–296 (2012)
11. Klein, R.: Retinopathy in a population-based study. *Transactions of the American Ophthalmological Society* **90**, 561 (1992)
12. Laliberté, F., Gagnon, L., Sheng, Y.: Registration and fusion of retinal images-an evaluation study. *Medical Imaging, IEEE Transactions on* **22**(5), 661–673 (2003)
13. Lawson, Charles L and Hanson, Richard J: *Solving least squares problems*, vol. 161. Prentice-hall (1974)
14. Lee, S., Abràmoff, M.D., Reinhardt, J.M.: Validation of retinal image registration algorithms by a projective imaging distortion model. In: *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 6471–6474. IEEE (2007)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **60**(2), 91–110 (2004)
16. McLauchlan, P.F., Jaenicke, A.: Image mosaicking using sequential bundle adjustment. *Image and Vision computing* **20**(9), 751–759 (2002)
17. Richa, R., Linhares, R., Comunello, E., Von Wangenheim, A., Schnitzler, J.Y., Wassmer, B., Guillemot, C., Thuret, G., Gain, P., Hager, G.: Fundus image mosaicking for information augmentation in computer-assisted slit-lamp imaging. *Medical Imaging, IEEE Transactions on* **33**(6), 1304–1312 (2014)
18. Souza, M., Richa, R., Puel, A., Caetano, J., Comunello, E., von Wangenheim, A.: Robust visual tracking for retinal mapping in computer-assisted slit-lamp imaging. In: *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, pp. 132–137. IEEE (2014)
19. Stewart, C.V., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *Medical Imaging, IEEE Transactions on* **22**(11), 1379–1394 (2003)
20. Szeliski, R.: Image alignment and stitching: A tutorial. *Found. Trends. Comput. Graph. Vis.* **2**(1), 1–104 (2006)
21. Torr, P.H., Zisserman, A.: Feature based methods for structure and motion estimation. In: *Vision Algorithms: Theory and Practice*, pp. 278–294. Springer Berlin Heidelberg (1999)
22. Yang, G., Stewart, C.V.: Covariance-driven mosaic formation from sparsely-overlapping image sets with application to retinal image mosaicking. In: *Computer Vision and Pattern Recognition (CVPR), Proceedings of the IEEE Computer Society Conference on*, vol. 1, pp. I–804. IEEE (2004)
23. Zheng, Y., Daniel, E., Hunter, A.A., Xiao, R., Gao, J., Li, H., Maguire, M.G., Brainard, D.H., Gee, J.C.: Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix. *Medical image analysis* **18**(6), 903–913 (2014)