

# MULTIMODAL MACULA MAPPING: STUDY FOR RIGID, PERSPECTIVE AND DEFORMABLE IMAGE REGISTRATION

Pedro Baptista\*, Rui Bernardes\*, João Ferreira\*,\*\*, Jorge Dias\*\* and José Cunha-Vaz\*,\*\*\*

\* Center of New Technologies for Medicine, Association for Innovation and Biomedical Research on Light and Image, Coimbra, Portugal

\*\* Institute of Systems and Robotics, Coimbra, Portugal

\*\*\* Institute of Biomedical Research on Light and Image, Faculty of Medicine of the University of Coimbra, Coimbra, Portugal

pbaptista@aibili.pt

**Abstract:** In this work, thirteen image-pairs, composed of a color fundus photograph and a scanning laser ophthalmoscope fluorescein angiography (SLO.FA), were registered by rigid, non-rigid and deformable image registration procedures to assess the need for deformable image registration methods when using SLO.FA modalities. Each registration was quantitatively assessed by three different metrics that showed improved performance of the deformable approach. Moreover, the computed deformation grids demonstrated the infeasibility of global transformation approaches to deal with saccadics, as expected. Also, in the registration of medical images, some knowledge of the acquisition mode and its particularities was shown necessary to successfully achieve good registration results with a reduced search space for the parameters involved.

## Introduction

The fusion of functional and morphological information from *in vivo* imaging modalities of the human eye fundus is a major need, both on the assessment of treatment efficacy and on the monitoring of disease progression on the initial stages of diabetic retinopathy. The potential for multimodal macula mapping was demonstrated in [1].

Published work on multimodal image registration deal mostly with images acquired using area sensors, either of color or monochromatic type, with which the whole image is acquired at once [2, 3, 4], while available retinal imaging modalities encompass a number of different sensors and sensor types. As the use of confocal scanning laser ophthalmoscopes (CSLO) [5] becomes more commonplace, a new problem arises for the registration of this sort of images, more precisely the intrinsic image distortions due to saccadic eye movements, which in turn can be of two kinds: voluntary (e.g., when the eyes move from one fixation point to another as in reading) and involuntary (e.g., due to following the flying spot that illuminates the eye fundus). These effects in the image will obviously depend on the amount of time for the full scan, ranging from tenths of seconds to seconds.

These facts lead us to study the comparative performance between the rigid, non-rigid and deformable registration approaches when using imaging modalities involving scan-based acquisition mode systems. Non-rigid transformation parameters are generally of the perspective type and not of the affine or simple projective type, hence the wording *perspective* instead of *non-rigid* in the title, although throughout the text the *non-rigid* wording will be used.

## Materials and Methods

In this work we will focus on two different types of light detectors — area sensors and single-point sensors — to acquire *in vivo* 2D morphology and/or functional information of the human retina, aiming to assess the need for the deformable image registration when using this sort of single-point sensors by comparing the performance of the rigid, non-rigid and deformable image registration approaches.

The fundus references herewith considered are digital RGB color fundus photographs (retinographies) of 50° field-of-view (FOV), centered on the fovea, with a resolution of 768 × 576 pixels and were taken using a 3CCD Sony camera mounted on a Zeiss Fundus Camera system model FF450 (Zeiss, Germany). Retinographies represent an area sensor modality since all pixels are imaged simultaneously, thus being virtually distortion-free.

Imaging modalities based on scanning systems, on the other hand, which suffer from intrinsic distortions, are represented by 20° FOV high-speed fluorescein angiographies (SLO.FA) centered on the fovea and performed by a CSLO after the administration of a dye. In this work, a Heidelberg Retina Angiograph (HRA, Heidelberg Engineering, Germany) was used to acquire SLO.FA images of 256 × 256 pixels (maximum).

Since the aim of this work is to assess the need for the deformable registration when using assembled images from single-point sensors due to saccadics, we decided to eliminate the scaling factor by decimating the highest resolution image to the lowest resolution one, thus having each pixel representing the same area in the eye fundus. The retinography, originally having a res-

olution of 14.5 pixels/degree was therefore decimated to 12.8 pixels/degree (corresponding to the resolution of the SLO.FA image) at the preprocessing stage.

The set of features available for each modality herein considered are the fovea, optic disc and the vascular network for retinographies, and the fovea and a subset of the vascular network for the SLO.FA modality.

Retinographies were preprocessed to remove background artifacts and to be converted into gray-scale images. The fovea appears therefore as a dark region, the optic disc appears as a bright area and the vessels, veins and arteries, are darker than the surrounding tissue, i.e., the background.

Vascular network expression in SLO.FA images is dependent on the time after injection, i.e., the amount of time elapsed since the dye was injected into the patient's blood-stream and the scanning of the eye. After intravenous administration, arteries become brighter as fluorescein starts appearing in the retinal circulation and veins are still dark. Thereafter, veins also brighten as fluorescein is collected after microcirculation. The foveal area is normally kept darker than the background. For the time frame of interest for our application, 10 to 20 minutes after dye injection, the fovea is characterized by a dark region and the vascular network by bright vessels, as opposed to their expression in retinographies.

For rigid registration, one needs to compute  $\mathbf{P}$  so that

$$x^r = \mathbf{P}x^i, \quad (1)$$

where  $x^r$  are the coordinates of  $x^i$ , respectively in the retinography ( $r$ ) and SLO.FA ( $i$ ) and  $\mathbf{P}$  ( $3 \times 3$ ) is the transformation matrix (of the rigid type) from the SLO.FA coordinate system to the retinography coordinate system.  $\mathbf{P}$  is therefore defined by three parameters: a translation in the  $x$ -axis ( $t_x$ ); a translation in the  $y$ -axis ( $t_y$ ), and; a rotation ( $\theta$ ). So as to reduce the search space for these three parameters ( $t_x, t_y, \theta$ ), we made use of the common features available for the modalities at hand, namely the fovea and the vascular network. Whilst both features were used to compute  $t_x, t_y$  and  $\theta$ , the fovea provided us with the initial estimate for the translation and the vascular network helped us on fine-tuning the initial translation estimate and on computing the rotation, as will be shown later on.

For the initial estimation of the translation ( $t_{x0}^i, t_{y0}^i$ ), computed fovea locations on both modalities ( $\Phi_x^r, \Phi_y^r$ ) and ( $\Phi_x^i, \Phi_y^i$ ), respectively for the retinography and SLO.FA, were used and the translation computed as in (2).

$$\begin{cases} t_{x0}^i = \Phi_x^r - \Phi_x^i \\ t_{y0}^i = \Phi_y^r - \Phi_y^i \end{cases}. \quad (2)$$

To compute the fovea location in each of the modalities, a mathematical model for the fovea as the one described in [6] was used, namely a 2D inverted Gaussian

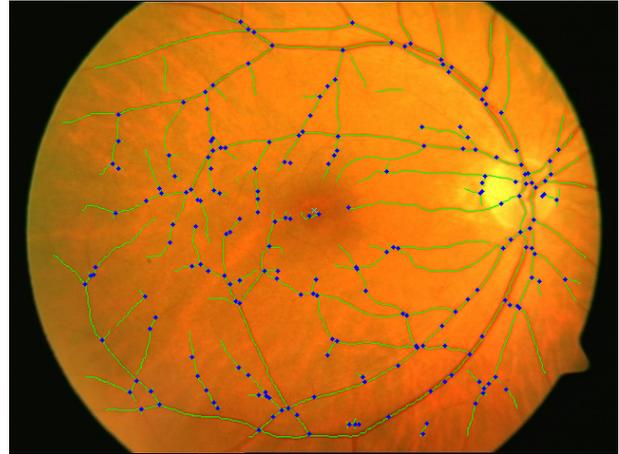


Figure 1: Vascular network (vessel centerlines) extracted from the color fundus photograph (retinography) shown on green. Blue dots represent vessel bifurcations/crossings candidate points.

function as defined in (3)

$$g(i, j) = k \left[ 1 - \frac{1}{2} \exp \left( \frac{-(i^2 + j^2)}{2\sigma^2} \right) \right], \quad (3)$$

where  $k > 0$  is a real, arbitrary constant (made equal to 1 in this work) and  $\sigma$  was experimentally determined to be equal to half an average reference value for the optic disc diameter in each modality. Each image modality is therefore cross-correlated with (3) to determine the fovea location ( $\Phi_x, \Phi_y$ ).

In order to compute the rotation between modalities, the vascular network is detected using a geometric differential approach. Images can be considered as 3D surfaces with image intensity converted to height and vessel centerlines as ridges/gorges of the surfaces  $S$  defined by  $S = \{x, y, \mathbf{I}(x, y)\}$ . For each pixel with coordinates  $(x_0, y_0)$  the Hessian matrix ( $\mathbf{H}(x_0, y_0)$ ) is computed as well as respective eigenvectors. Being  $\mathbf{u}$  the eigenvector associated to the maximum/minimum eigenvalue of  $\mathbf{H}(x_0, y_0)$ , the pixel at  $(x_0, y_0)$  is a ridge/gorge if the first directional derivative across the direction  $\mathbf{u}$  at  $(x_0, y_0)$  is null [4], i.e.,

$$\langle \nabla \mathbf{I}(x_0, y_0), \mathbf{u} \rangle = 0. \quad (4)$$

Vessel centerlines are then segmented by determining the skeleton of the newly computed binary images, after the application of morphological filters — erosion, dilation, skeletonizing and pruning. Vessel centerline can be seen plotted in green in figures 1 and 2 over the respective fundus images along with blue dots which represent vessel bifurcations/crossings candidate points.

Having determined the vascular network for both modalities, in fact vessel centerlines networks, the rotation ( $\theta$ ) can be simply estimated by computing the translation along the *angle-axis* ( $\alpha$ ) on their polar representations, where each pixel of the vessel centerlines network has coordinates  $(\rho, \alpha)$  computed relatively to the center of the fovea, i.e.,  $(\Phi_x, \Phi_y)$ .

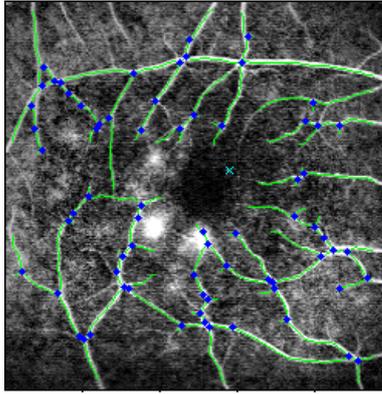


Figure 2: Vascular network (vessel centerlines) extracted from the SLO.FA image shown on green. Blue dots represent vessel bifurcations/crossings candidate points.

Since the computed location of the fovea on both modalities might not correspond to the same retinal (physical) location, e.g., due to noise in any or both images, this corresponds to having a different representation of both vascular networks. This effect is demonstrated on figure 3 where the same point will be mapped by  $\vec{v}_1 = (\rho_1, \alpha_1)$  or  $\vec{v}_2 = (\rho_2, \alpha_2)$  depending on the viewpoint.

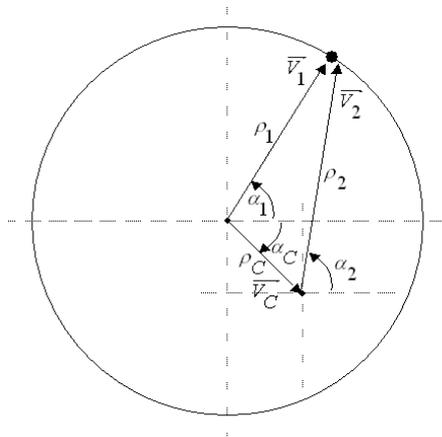


Figure 3: Schematic of the same location seen from two different viewpoints.

The effect of these different viewpoints can be better seen by mapping two artificial objects as shown in figure 4. The shift in the angle can be seen by the two representations of the same line, being this shift dependent on the radius. On the other hand, the representation of the circle shows the modulation effect on the radius, which is dependent on the polar representation angle and given by

$$\rho_2 = \sqrt{\rho_1^2 + \rho_C^2 - 2\rho_1\rho_C \cos(\alpha_1 - \alpha_C)} . \quad (5)$$

A set of windows  $W$  of size  $M \times M$  were considered on the polar representation of the vascular network and the best match sought by a phase cross-correlation

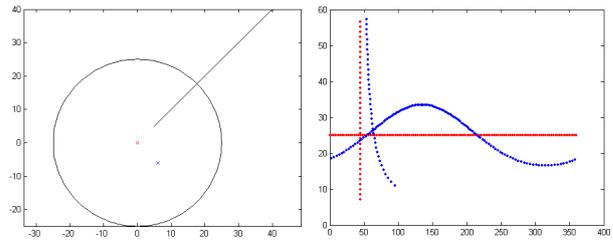


Figure 4: Two objects, a circle and a line, from two viewpoints, on the left, and respective polar representation, on the right (the vertical axis in pixels (radius) and the horizontal axis in degrees (angle)).

method [7], both on the radius ( $\rho$ ) and the angle ( $\alpha$ ). By fitting through an optimization procedure (5) to the translation corresponding to each window,  $\vec{v}_C = (\rho_C, \alpha_C)$  is computed. In fact this corresponds to updating the initial estimate for translation between image modalities, i.e.,

$$\begin{cases} t_x^i = t_{x0}^i + \langle \vec{v}_C, \hat{x} \rangle \\ t_y^i = t_{y0}^i + \langle \vec{v}_C, \hat{y} \rangle \end{cases} , \quad (6)$$

where  $\hat{x}$  and  $\hat{y}$  are respectively the unit vectors on the  $x$ - and  $y$ -axis, and  $\langle \cdot \rangle$  represents the inner product operation. This update of the initial estimated translation corresponds to forcing the same viewpoint on both image modalities and to correct for the aforementioned distortions on the polar representation. By doing so, it is now possible to estimate the rotation ( $\theta$ ) between image modalities by computing the translation of the vascular network in polar coordinates using the same phase cross-correlation method.  $\mathbf{P}$  can now be computed as

$$\mathbf{P} = \mathbf{T}' \mathbf{R}_\theta \mathbf{T}^i , \quad (7)$$

where  $\mathbf{T}^i$  translates the *corrected* fovea coordinates of the SLO.FA to the origin ( $\mathbf{0}$ ),  $\mathbf{R}_\theta$  rotates the SLO.FA by the computed rotation  $\theta$  and  $\mathbf{T}'$  translates the origin to the coordinates of the fovea detected in the retinography ( $\Phi'_x, \Phi'_y$ ).

Our new task is now to compute a non-rigid transformation over the same images defined by a transformation matrix  $\mathbf{P}'$  given by

$$\mathbf{P}' = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & 1 \end{bmatrix} , \quad (8)$$

where  $p_{11}$  to  $p_{32}$  are the parameters to be determined and  $\mathbf{P}'$  is of the rigid, affine or perspective type depending on the set of estimated parameters  $p_{ij}$ .

In order to compute  $\mathbf{P}'$ , a set of point-pairs need to be determined. Considering the data at hand, the point-pairs need to be established using the vessel bifurcations/crossings. These are determined by computing the cross-correlation of the binary image containing vessel centerlines with a unit matrix ( $J_3$ ) and earmarking as a candidate point for vessel bifurcations/crossings any location where the cross-correlation is over 4. Repeating

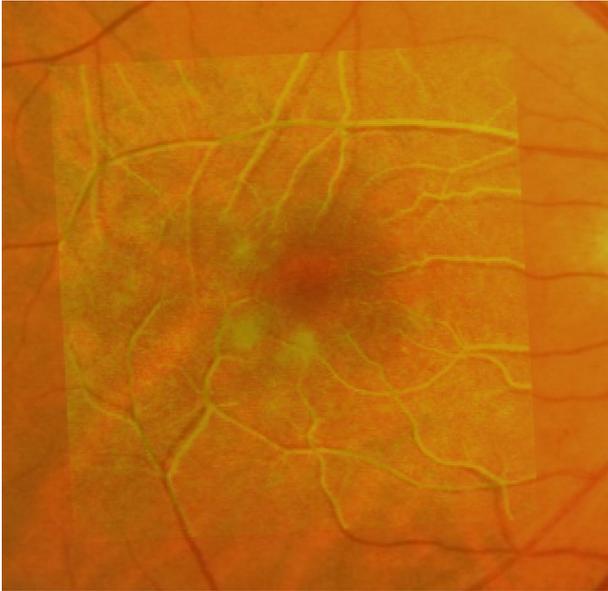


Figure 5: Result of rigid registration. Only the common region between both modalities is shown to enhance details. The SLO.FA is shown as a transparent layer over the retinography modality.

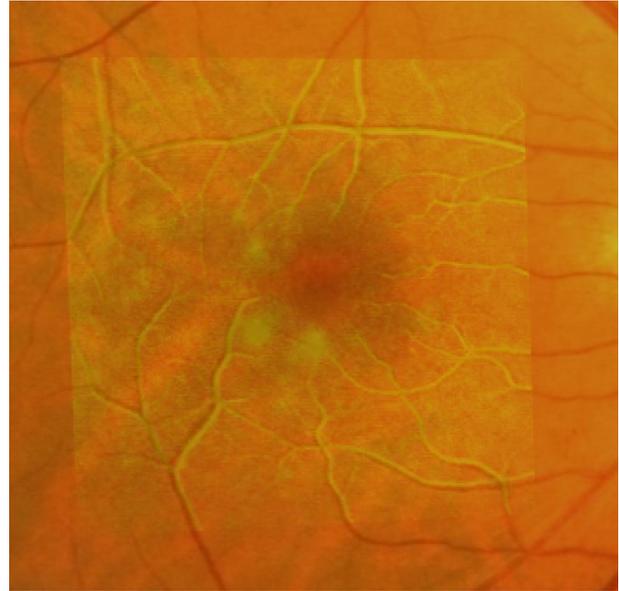


Figure 6: Result of non-rigid registration. Although still showing space for improvement, better registration is achieved compared to the rigid approach. Only the common region between both modalities is shown to enhance details. The SLO.FA is shown as a transparent layer over the retinography modality.

this procedure for both image modalities, outliers can be eliminated by using the rigid registration performed previously to identify locations where candidate bifurcations/crossings in one image do not have candidate bifurcations/crossings in the other image modality in a window centered on  $x' = \mathbf{P}x^i$ , with  $x^i$  being the coordinates of a bifurcations/crossings in one modality and  $\mathbf{P}$  the rigid matrix previously computed. Having removed outliers, remaining candidate points are paired using a modified version of the *softassign* algorithm described in [8] using the *corrected* fovea positions to play the role of the geometrical center candidate points as used in the original algorithm.

Having paired candidate bifurcations/crossings points, their coordinates are fine-tuned by computing the best-fit in the point neighborhood area using the PIU similarity metric (to be defined later on). This fine-adjustment yields a better result for the non-rigid registration process by diminishing the error associated with the relative point location of each pair of points.

Finally, the non-rigid transformation matrix  $\mathbf{P}'$  can be determined by computing the Moore-Penrose pseudo-inverse matrix of the overdetermined system, as a surplus of candidate pairs of points is available. The registration achieved in this fashion is particularly sensitive to the distribution of the paired of points over the image area. Better results are achieved when having an even compared to an uneven distribution of paired points.

Our final task is now to compute a deformable registration over the same images. To accomplish this, a *local* transformation needs to be computed. In this work we defined the local transformation ( $\mathbf{P}_{j,k}$ ) as the transformation

at coordinates  $(j, k)$  defined by

$$\mathbf{P}_{j,k} = \mathbf{S}_{j,k} \mathbf{P}_{global} \quad (9)$$

where  $\mathbf{S}_{j,k}$  represent a local translation (due to saccadics) on top of a global (either of rigid or non-rigid type) transformation ( $\mathbf{P}_{global}$ ). In our particular case,  $\mathbf{P}_{global}$  was chosen to be made equal to  $\mathbf{P}$ , i.e., the rigid transformation result. Therefore, (9) becomes

$$\mathbf{P}_{j,k} = \mathbf{S}_{j,k} \mathbf{P} \quad (10)$$

In order to compute  $\mathbf{S}_{j,k}$ , the SLO.FA image modality was considered as a set of overlapping windows ( $W^i$ ) of size  $M \times M$ . Each  $W^i$  window was projected into the retinography by  $\mathbf{P}$  where a best-fit was sought on a window  $W^r$  of size  $N \times N$  ( $N > M$ ), sharing both, the projected  $W^i$  and  $W^r$  windows, the same center. This allows to reduce the computing time by restricting the search area.

For the similarity metric, the partitioned intensity uniformity (PIU) was used (11). For details see [9].

$$PIU = \sum_b \frac{n_b}{N} \frac{\sigma_A(b)}{\mu_A(b)} \quad (11)$$

The PIU similarity metric decreases as image similarity increases, as it measures the ratio between the variance and the average of the projection of areas considered as having a uniform intensity distribution.

So as not to compute  $\mathbf{S}_{j,k}$  for every pixel, pixels at a grid spaced  $\Delta_{xy}$  were considered and  $\mathbf{S}_{j,k}$  determined for their coordinates, i.e., for each of these pixel coordinates

a translation on the  $x$ - and  $y$ -axis, respectively,  $dx$  and  $dy$ , was computed for a total of over 200 reference points.

To compute the entire deformation map, two surfaces might be considered where respective heights represent either the  $dx$  or  $dy$  translations that constitute  $\mathbf{S}_{j,k}$ . By fitting a spline surface to each of these two surfaces, one can estimate  $\mathbf{S}_{j,k}$  for every pixel in the SLO.FA image modality.

The result achieved can be seen in figure 7, where  $\mathbf{P}_{j,k}$ , as defined in (10), was computed for pixels at a grid spaced of  $\Delta$  pixels.

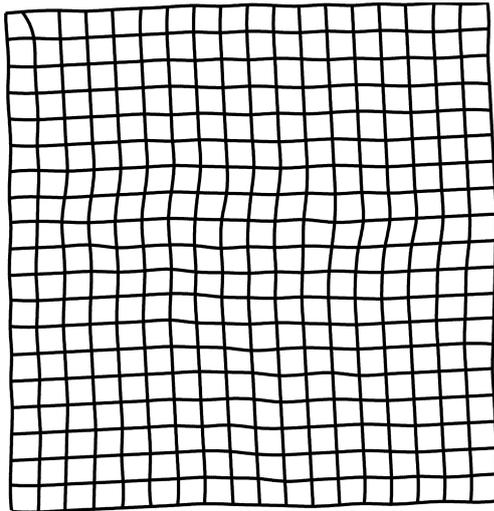


Figure 7: Grid representing the sampled deformation computed for the SLO.FA image modality.

The computed deformation is then applied to the entire SLO.FA image modality to bring it into registration with the retinography image. This final result can be seen in figure 8.

Having performed the three types of registration proposed to be compared, the rigid, non-rigid and the deformable registration, three metrics were used to assess their relative performances: the correlation (CORR), the mean square of differences (MSD), and the partitioned intensity uniformity (PIU).

## Results

Thirteen pairs of retinography/SLO.FA images were registered by the rigid, non-rigid and deformable approaches explained at length throughout the text. For each of the thirteen pairs, each registration was assessed by the three different registration metrics selected: CORR, MSD, and PIU.

The relative performance of each registration method can be checked in table 1. Considering the different metrics at hand, one should pay attention to the fact that different metrics yield different interpretations: values of MSD and PIU decrease as similarity increases, and the correlation (CORR) values increase as the similarity also increases.

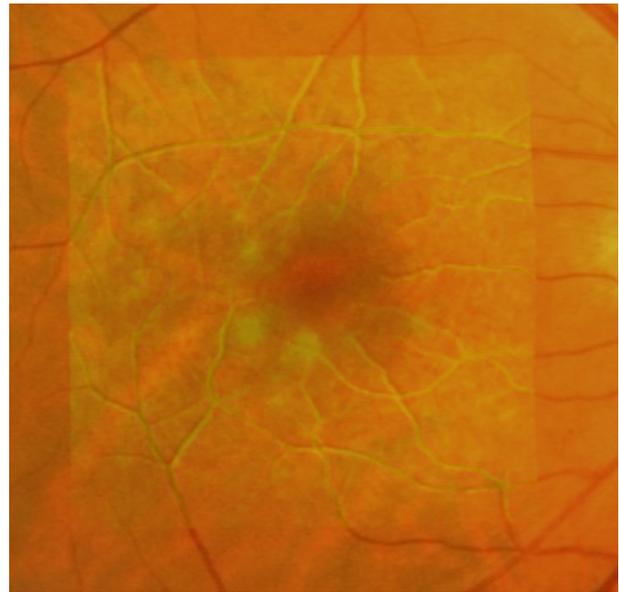


Figure 8: The result of the deformable registration, presenting improved registration accuracy relatively to both the rigid and non-rigid registration. Only the common region between both modalities is shown to enhance details. The SLO.FA is shown as transparent layer over the retinography modality.

These results clearly show that registration performance using the deformable approach is better than either the rigid or non-rigid ones. Moreover, neither the rigid or non-rigid procedures can deal with local distortions due to saccadics as the ones seen on figure 7.

While both the rigid and non-rigid registration methods used in this work can be applied to the entire SLO.FA image area, the deformable registration method herewith presented uses windows  $W^i$  of size  $M \times M$  to compute translations on top of a global registration. Therefore there exists a strip of  $(M-1)/2$  pixels width on the outer region of the SLO.FA image where  $\mathbf{S}_{j,k}$  values are extrapolated while for the remaining image they are interpolated. This explains the non-uniform performance of the deformable registration seen in figure 8.

## Discussion

The results presented on table 1 evidence the improvement achieved throughout the three registration methods, in a universe of thirteen pairs of image modalities (retinography and SLO.FA). Analyzing these data it is perceptible that the performance of the registration processes becomes better as the order of the estimated transformation increases. The nature of the three transformations can be described in two different ways: analytically and non-analytically. The rigid and non-rigid registration can be defined by a set of parameters and/or a transformation matrix. The fact that the deformable registration was decided not to be defined analytically was to avoid using high-order functions that either propagate

	CORR			MSD			PIU		
	Rigid	Non-Rigid	Deformable	Rigid	Non-Rigid	Deformable	Rigid	Non-Rigid	Deformable
Avg	0,311	0,409	0,487	5,53E+03	5,40E+03	5,07E+03	0,131	0,127	0,113
Std	0,118	0,103	0,084	2,20E+03	2,19E+03	2,30E+03	0,010	0,010	0,010

Table 1: Comparative performance results for the three registration methods assessed by three different metrics. The different metrics yield different interpretations: values of MSD and PIU decrease as similarity increases, and the average correlation values increase as the similarity also increases. All metrics indicate that non-rigid registration performs better than the rigid registration and that the deformable registration performs better than both of the former.

oscillations from the location of saccadics or do not completely follow these saccadics.

The deformable registration method proved to be most useful when at least one of the images is acquired in a raster scan acquisition mode, while this approach also simultaneously produces the better performance ratings.

### Conclusions

The results presented in this work prove the usefulness of a deformable image registration procedure when modalities resulting in images with local deformations are involved, such as those obtained with a raster scan acquisition mode. Also, in the registration of medical images, some knowledge of the acquisition mode and its particularities is necessary to successfully achieve good registration results with a reduced search space for the parameters involved.

### Acknowledgment

The work herewith presented was funded by the POSI/SRI/45151/2002 project, sponsored by the Fundação para a Ciência e a Tecnologia, Portugal.

### References

[1] RUI BERNARDES, CONCEIÇÃO LOBO, and JOSÉ CUNHA-VAZ. Multimodal macula mapping: A new approach to study diseases of the macula. *Surv Ophthalmol*, 47(6):580–589, November–December 2002.

[2] FRANCE LALIBERTE and LANGIS GAGNON. Registration and fusion of retinal images—an evaluation study. *IEEE T Med Imaging*, 22(5):661–673, May 2003.

[3] F. ZANA and J. C. KLEIN. A multimodal registration algorithm of eye fundus images using vessels detection and hough transform. *IEEE T Med Imaging*, 18(5):419–428, May 1999.

[4] GEORGE MATSOPOULOS, PANTELIS ASVESTAS, NIKOLAOS MOURAVLIANSKY, and KONSTANTINOS DELIBASIS. Multimodal registration of retinal images using self organizing maps. *IEEE T Med Imaging*, 23(12):1557–1563, December 2004.

[5] ROBERT WEBB, G. HUGHES, and F. DELORI. Confocal scanning laser ophthalmoscope. *Appl. Opt.*, 26(8):1492–1499, April 1987.

[6] CHANJIRA SINTHANAYOTHIN, JAMES F BOYCE, HELEN L COOK, and THOMAS H WILLIAMSON. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol*, 83:902–910, 1999.

[7] J. PEARSON, D. HINES, S. GOLOSMAN, and C. KUGLIN. Video-rate image correlation processor. In *Proceedings of SPIE: applications of Digital Image Processing*, volume 119, pages 197–205, 1977.

[8] L. RANGARAJAN, H. CHUI, and F. BOOKSTEIN. The softassign procrustes matching algorithm. In James Duncan and Gene Gindi, editors, *Information Processing in Medical Imaging*, pages 29–42. Springer, 1997.

[9] DEREK HILL and PHILIPPE BATCHELOR. Registration methodology: Concepts and algorithms. In Joseph Hajnal, Derek Hill, and David Hawkes, editors, *Medical Image Registration*, The Biomedical Engineering Series, chapter 3, pages 39–70. CRC Press, 2001.