Robust Matching of Multi-Modal Retinal Images using Radon Transform based Local Descriptor

Yogesh Babu Bathina
Center for Visual Information Technology,
International Institute of Information Technology,
Hyderabad, India.
yogesh_b@research.iiit.ac.in

N.V Kartheek Medathati
Center for Visual Information Technology,
International Institute of Information Technology,
Hyderabad, India.
nvkartheek@research.iiit.ac.in

Jayanthi Sivaswamy
Center for Visual Information Technology,
International Institute of Information Technology,
Hyderabad, India.
jsivaswamy@iiit.ac.in

ABSTRACT

Multi-Modal image registration is the primary step in fusing complementary information contained in different imaging modalities for diagnostic purposes. We focus on two specific retinal imaging modalities namely, Color Fundus Image(CFI) and Fluroscein Fundus Angiogram(FFA). In this paper we investigate a projection based method using Radon transform for accurate matching in multi-modal retinal images. A novel Radon Transform based descriptor is proposed, which is invariant to illumination, rotation and partially to scale. Our results show that our descriptor is well suited for retinal images as it is robust to lesions, and works well even in poor quality images. The descriptor has been tested on a dataset of 126 images and compared for matching application against gradient based descriptors. The results show that Radon based descriptor outperforms the gradient based ones in both being able to discriminate between true and false matches and under presence of lesions.

Categories and Subject Descriptors

I.4 [**Image Processing and Computer Vision**]: Applications, Miscellaneous.

General Terms

Algorithms.

Keywords

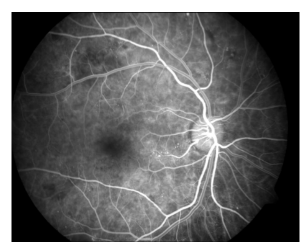
Registration, Multimodal, Transformation estimation, Image mosaic, Retinal imaging, Feature extraction, Ophthalmic image processing, Biomedical image processing, Radon descriptor, Vessel enhancement, Curvature orientations histogram, Multimodal feature descriptor.

IHI'10, November 11–12, 2010, Arlington, Virginia, USA.

1. INTRODUCTION

Multi-modal image registration is the task of spatially aligning a pair of images of the same scene acquired from different sources, viewpoints and time. It is the fundamental step in fusing complementary information contained in different modalities. In this paper we focus on two specific retinal optical imaging modalities, namely Color Fundus Image(CFI) and Fluroscein Fundus Angiography (FFA) as shown in Figure 1. CFI captures optical range information and hence reveals the overall condition of the fundus(retina) while FFA captures the blood flow information and hence reveals only structures such as vessels involved in blood flow. Information obtained through registration of these two imaging modalities aids in the diagnosis of various kinds of retinal diseases such as glaucoma, diabetic retinopathy and age related macular degeneration [1]. A key application area for registration of FFA and CFI is surgery, both in the planning stage and during surgery at which time only optical range information is available. Fusion of these modalities also helps increase the anatomical range of visual inspection and provides a means for early detection of potentially serious pathologies [2] and reveals the relationship between the blood flow and the diseases occurring on the surface of the retina [3].

(a) Color Fundus Images



(b) Fluroscein Angiogram

Figure 1: Multi Modal Retinal Images.

Existing work on registration of retinal images can be broadly classified into two types: area based methods [4] [5] and feature based methods [6][7][2][8]. Area based methods operate directly on the intensity values at a global level and choose a suitable similarity measure to drive the registration. These methods typically use an optimization framework with the objective of maximizing the similarity measure while estimating the transformation between the images. Feature based methods on the other hand, follow typically a three step approach - detection of significant landmarks across images, establishing correspondence using features extracted around landmarks and estimation of the transformation function using correspondences. Feature based methods have gained popularity over the area based methods as they are more robust to illumination changes, partial overlap between the images, occlusion, changes in background and viewpoint [1].

The success of a feature based registration scheme depends on the number of accurate correspondences established [9]. Finding accurate correspondence in retinal image is challenging due to the following reasons:

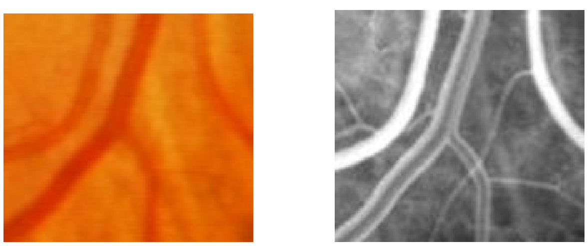
1. Non-uniform illumination in CFI images and the varying appearance of vessels, with time of acquisition, in the case of FFA. Example of such occurrence is shown in Figure 2.

2. Complementary nature of multimodal images also means that some of the pathologies are visible only in one of the modalities. Examples of such occurrences are shown in Figure 3.

3. Variability in contrast and resolution is another common problem that affects matching.

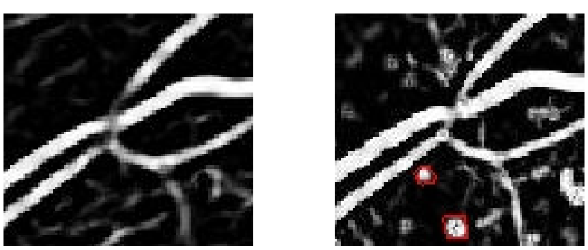
4. Varying degree of overlap between CFI and FFA due to wide range of view angles.

5. Poor image quality due to other reasons such as cataract, poor dilation, etc.



(a) Color Fundus Images (b) Fluroscein Angiogram

Figure 2: Corresponding Patches of CFI and FFA.



(a) Enhanced CFI (b) Enhanced FFA

Figure 3: lesions perceived in a single modality are highlighted in red

Feature based methods can be broadly divided into two categories, ones which rely on weak descriptors (rough correspondences) coupled with a strong registration framework (such as robust outlier rejection schemes to establish accurate correspondence) or those which use strong descriptors (single shot accurate correspondence) with a simpler framework. Both these techniques usually start with vessel junctions as landmarks. An example in the former category is [10] where entropy based similarity measure is employed for establishing weak correspondence in a hierarchical registration frame work. Alternatively, orientation information of skeletonized vascular structures has also been chosen as features to establish weak correspondence [2],[11]. Robust outlier rejection is used to simultaneously fit a transformation and to detect accurate correspondences. Another similar approach [3] does simultaneous region growing and transformation model estimation by using vessel orientation and width information.

General feature descriptors found in computer vision such as Scale Invariant Feature Transform (SIFT)[12] and Speed Up Robust Features (SURF) [13] poses several desirable properties such as scale, illumination and noise invariance but are designed for monomodal images. These rely on gradient information which is not consistent in multimodal medical images. An attempt to address this problem is Gradient Mirror based SIFT (GM - SIFT)[9] which is able to handle nonlinear changes in intensity found in generic multimodal images. A descriptor designed specific to the registration of poor quality retinal images is called Partially intensity invariant feature descriptor (PIIFD) [14]. Here, starting with Harris corners as landmarks, the orientation information of the local neighborhood is extracted to construct the descriptor. PIIFD is not only partially invariant to affine, viewpoint and intensity changes but is also reported to perform better than the SIFT [1]. Both the GM-SIFT and PIIFD use gradient orientation information to construct their descriptors. Hence, presence of lesions in the neighborhood of a landmark will adversely affect their performance. These lesions typically appear only in one modality image.

In this paper, we investigate the use of projection based local shape descriptors since the projection operation has potential to render the descriptor less sensitive to lesions and noise. This should help in establishing accurate correspondence even when the lesions are in the proximity of the landmarks. We choose the Radon transform (RT) for deriving the projections as RT based representations have been shown to be effective for shape description and object recognition [15] [16] [17]. However, they have been applied only to binary or greyscale images consisting of a single object. We propose using RT based descriptors for capturing local structures around landmarks instead of objects as a whole.

2. METHOD

Our primary objective is to establish one to one correspondence between CFI-FFA image pair for the estimation of the transformation function between them. The pipeline of the algorithm for establishing correspondence is

1. Landmark detection and Vessel enhancement.

2. Deriving a Radon Transform based descriptor.

3. Matching based on Descriptors.

2.1 Landmark detection and Vessel Enhancement

Landmarks are anatomically significant, visually salient, distinct features in an image that are identifiable and comparable across images. In our previous work [18] we illustrated a method for vessel enhancement using a Hessian filter and obtaining vessel junctions as landmarks based on changes in the local curvature computed at multiple scales. Specifically, a curvature orientation histogram over a patch around every vessel point is computed and the entropy of this histogram is used to determine vessel junctions (as they are high entropy points). We use this method to derive the desired landmarks in our current work. The results of landmark detection and enhanced vessels are shown in Figure 4 and Figure 5 respectively.

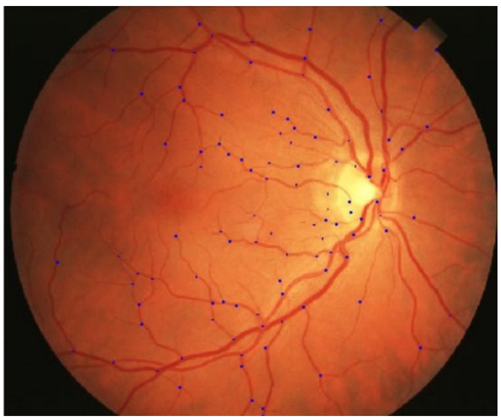
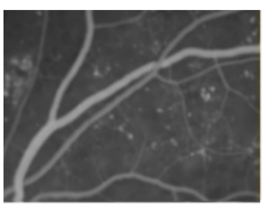
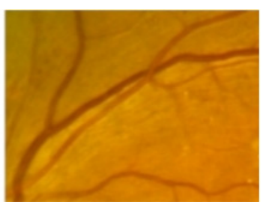


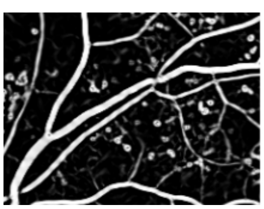
Figure 4: Landmark Detection



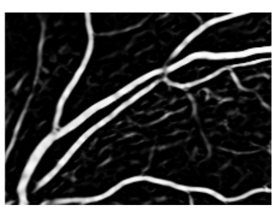
(a) FFA



(b) CFI



(c) Enhanced FFA



(d) Enhanced CFI

Figure 5: Boosted vessel structure through hessian computation

2.2 Radon Transform Based Descriptor

Computation of a descriptor in general can be seen as an attempt to represent a signal or image data in a compact format while retaining relevant information. In addition to compact representation, it is also important for a descriptor to be robust to geometric and photometric changes that occur across images. Common geometric distortions of interest are rotation, translation and scaling. In the case of local descriptors, the translation problem might arise because of localization issues of the landmark detectors. Even if the compactness criteria of a descriptor is set aside, unless we carefully retain the relevant information, the descriptor becomes over sensitive and matching accuracy falls. In the case of retinal images, if vessel junctions are used as landmarks the nearly (locally) linear vessel structure around these landmarks contain most relevant information for matching. Radon transform (RT) is a well known shape descriptor [15] which we use to capture the boosted local structures around each landmark point. The RT of a function f(x,y) denoted by $g(s,\theta)$ is given by

$$g(s,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta\left(xcos\theta + ysin\theta - s\right) dxdy, \quad (1)$$

with $-\infty < s < \infty$, $0 \leq \theta < \pi$. In the above, s is the distance from the origin and θ is the angle of projection. Rotating a function by an angle θ_0 results in a shift in the projection (Radon) domain by θ_0.

$$\tilde{R}f_\rho(r,\phi+\theta_0) = g(s,\theta+\theta_0)) \quad (2)$$

where $\tilde{R}$ is the transform operator and $f_\rho(r,\theta)$, is the representation of the function f in polar co-ordinate frame. Since we wish to derive a descriptor that is rotation invariant, we achieve it by aligning the axis of the RT to the dominant local direction at a landmark point. We next explain this in detail.

For each landmark point, a patch of size WxW is extracted and multiscale Hessian filtering is done to boost the vessels as explained in the above section. The largest eigenvector of the Hessian for each pixel within the patch is binned into 18 equally spaced bins. The center of the bin with highest peak in this histogram is taken as the dominant orientation. The patch is then rotated to align to this dominant direction. We use the local curvature direction to find the dominant local orientation as this information has already been computed in the landmark detection stage.

The next important step is to address the issue of translation or localization.

Let the RT of a function f(x,y) translated by a vector $[x_0, y_0]$ be denoted by $g'(s,\theta)$. The is related to the RT of the original function as follows.

$$g'(s,\theta) = g(s - x_0cos\theta - y_0sin\theta, \theta) \quad (3)$$

where $g(s,\theta)$ is the Radon Transform of the function f(x,y).

Thus the translation of the signal in 2D spatial domain results only in a shift along the radial axis of the transformation. The amount of shift changes according to the angle of rotation. In order to make the descriptor shift invariant, we exploit the translation property of the Discrete Fourier Transform (DFT). Specifically, we compute 1D DFT of each of the projections and retain only the amplitude information which is unaffected by shifts in the signal. The radon descriptor($\tilde{RD}$) thus obtained is,

$$\tilde{RD} = \{\hat{g}_{\theta_1}\ \hat{g}_{\theta_2}\ \hat{g}_{\theta_3}\ .....\hat{g}_{\theta_n}\} \quad (4)$$

where $\hat{g}_{\theta_i}$ is the amplitude spectrum of the 1D Discrete Fourier Transform of the projection $g(s,\theta_i)$.

An alternate method to compute $\tilde{RD}$ is by using the Fourier slice theorem which establishes the equivalence between 1-D Fourier Transform (FT) of Radon transform projections and oriented slices of the 2-D FT. Thus to derive the descriptor, one needs to compute the DFT, convert the result to polar coordinates and then select the required slices. Though this method has been adopted by some [19], it is inefficient as a full DFT has to be computed first and errors also creep in the interpolation process involved in the coordinate conversion stage. The proposed method in contrast, requires no interpolation and permits the desired slices to be extracted naturally by selecting the appropriate projection orientation.

Finally, we turn to compactness. The symmetry property of the DFT permits retaining of the transform coefficients thus reducing the length of the descriptor.

The descriptor computation can be summarized as

- Use Multi-Scale Hessian filter to boost the vessel structures.
- Extract local patches for each landmark.
- Align the patch to maximum curvature orientation direction.
- Compute the Radon transform.
- For each projection calculate 1D DFT.
- Discard phase information.
- Append the first half coefficients of magnitude information in FFT to form the descriptor.

2.3 Matching based on Descriptors

After the computation of descriptors, the next important stage for feature based registration is to obtain accurate correspondences by using appropriate matching strategy. In our approach, Bilateral matching technique[1] is used to ensure one to one correspondence. For two sets of landmarks A_i and B_j , D_{Ai} and D_{Bj} being their respective descriptors, first set of correspondence C_1 is obtained by finding the best matches using distance ratio between D_{Ai} to D_{Bj1} and D_{Bj2} (D_{Bj1} and D_{Bj2} being the first and the second nearest neighbors of D_{Ai}) and the second set C_2 obtained by distance ratio between D_{Bj} to D_{Ai1} and D_{Ai2}. The final correspondence is given by $C = C_1 \cap C_2$. If matching is Unilateral (one way) ie D_{Ai} to D_{Bj} or D_{Bj} to D_{Ai} one to one correspondence cannot be guaranteed.

The nearest neighbor in descriptor space is obtained based on Euclidean distance (in $[0, \infty)$, with 0 closest) between points. This distance has been converted into a similarity measure (in $[0, 1]$, with 1 closest) by a monotonic decreasing function. In this paper we choose to use the following similarity measure known as Euclidean-normalized similarity [20] given by

$$Similarity|(\mathbf{D}_{x1}, \mathbf{D}_{x2})| = e^{-\|\mathbf{D}_{x1}-\mathbf{D}_{x2}\|_{2}{}^{2}} \qquad (5)$$

Where D_{x1} and D_{x2} are two descriptors.

3. EXPERIMENTS AND RESULTS

The proposed method was evaluated on 126 CFI-FFA image pairs obtained from the same number of patients. The dataset was collected from a local collaborating hospital. The resolution of the CFI images is 576x768 and that of FFA images is 615x768. The angular resolution is fixed at 50°, which implies that the ratio of the image size to the width of the largest vessel in the image is always constant. We exploit this fact to bring both the images to a similar resolution by resizing. Our dataset has a wide variety of pathologies and the images exhibit various kinds of aquisition artifacts like non-uniform illumination, motion blur etc. Hence, it can be considered as a good dataset for validation of the proposed method.In our implementation, the RT is computed at an angular interval of 15 degrees. For a typical patch of size 41x41, the length of the descriptor is 360 and the time taken to compute each descriptor is 0.21 Sec. Sample matched pair of CFI(green channel) and FFA patches found using the Radon based descriptor are shown in Figure 6. Also matching between a challenging image pair is shown in Figure 7. These indicate that our descriptor is able to handle a wide variety of variations such as illumination, presence of lesion and transformations. On an average about 30 accurate matches are established per image on the above dataset. Next, we evaluate the proposed descriptor quantitatively and compare it with SIFT and PIIFD.

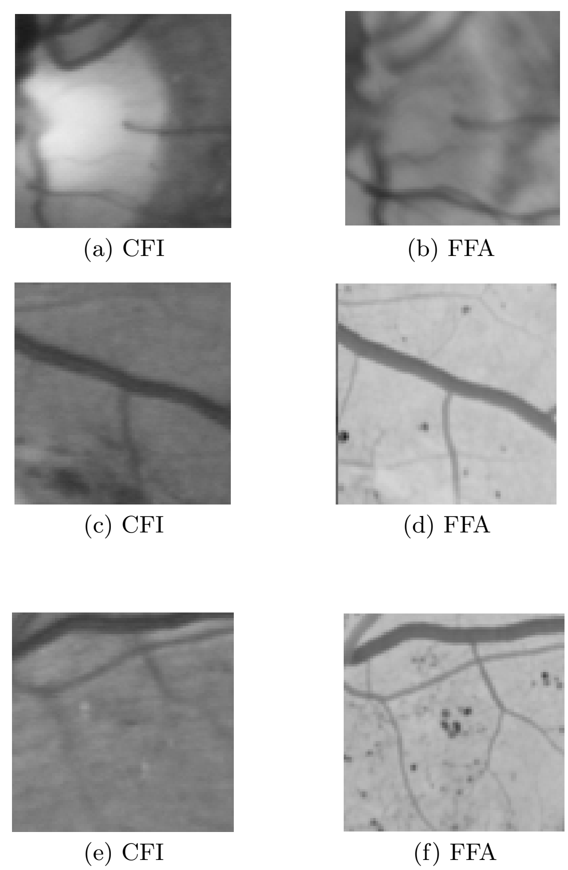


Figure 6: Shows successfully matched pairs: (a)-(b) on optic disk (c)-(f) in the presence of lesions

3.1 Comparative evaluation

Appropriate evaluation methodology is critical for unbiased and accurate quantification of the performance of various descriptors. The most popular evaluation scheme for descriptors in computer vision is the one proposed by [21].

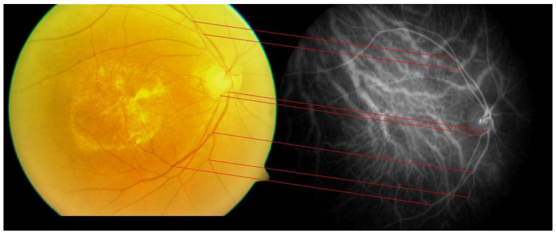


Figure 7: Matching between Challenging pair of Retinal Images

In this scheme, a benchmark dataset and a set of predetermined (ground truth) homographies are made available. These are used to benchmark descriptors based on precision and recall figures. Unfortunately, in the case of fundus images, no such benchmark dataset exists. Ground truth generation by selecting manual correspondences is tedious and error prone. This has prompted researchers to develop alternative comparison schemes such as test of discrimination (Multi-modality test in [1]).

We conducted a set of experiments to evaluate the proposed descriptor against PIIFD as well as SIFT. The objective of the experiments was to test the effect on the performance of a) shift, possibly due to localization errors in landmark detection b) the presence of lesions and c) discrimination margin between corresponding and non corresponding pairs.

A good descriptor should have on average, high similarity for correct pairs and low similarity for non-corresponding pairs. It is also equally important for the detector to exhibit low variance in case of correct matches as it indicates its consistency. Hence, in all our experiments, mean and standard deviation of a similarity measure are used as performance indicators. [1] report performance in the multimodal case based on a similar criteria.

3.1.1 *Shift Test*

The proposed Radon descriptor is designed to handle shifts efficiently. A corresponding pair is chosen to demonstrate this property. Translations of known amounts are applied to the landmarks of one image. The performance has been shown against the PIIFD descriptor in Figure 8.

It can be inferred that the Radon transform descriptor shows better shift tolerance than the other descriptor. This shows that our descriptor does not require well localized landmarks.

3.1.2 *Performance in the presence of lesions*

One of the challenges in Multi-Modal registration is the presence of lesions and structures like small vessels in either CFI or FFA. Lesions such as Exudates and Hemorrhages are visible in CFI images where as microanuerysms are better spotted (as bright dots) in FFA. In this test, we wish to compare the lesion handling capabilities of projection based method against popular gradient based descriptors. Thirty corresponding image pairs were manually picked with significant lesions present around the landmark points to evaluate the performance.

Table 1 shows the Euclidean-normalized similarity measure (Eq. 4) computed between each of these pairs for all the descriptors. The high mean similarity coupled with low

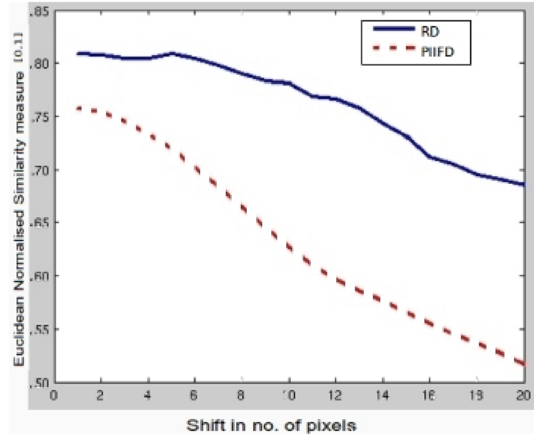


Figure 8: The average similarity measure of 100 corresponding pairs with shift in X direction

Table 1: Performance in the presence of the lesions

	Radon Descriptor	PIIFD	SIFT
Mean Similarity	0.73	0.66	0.64
Variance	0.015	0.066	0.047

variance of the Radon descriptor indicates that it has better lesion handling capabilities.

3.1.3 *Test of Discrimination*

This test was proposed by [1] to assess the discrimination power of their descriptor. Let corresponding pairs of CFI and FFA be denoted as CPRs and non-corresponding pairs NCPRs. The discrimination margin is taken to be the difference in mean similarity value between CPRs and NCPRS.The higher the margin the better is the over all matching accuracy between the regions. For this test, 100 corresponding and 100 non corresponding pairs were hand picked to assemble a dataset with variety of structures, textures and pathologies. The mean similarity scores for the Radon descriptor, PIIFD and SIFT are shown in Table 2. It can be seen that the difference in mean similarities between CPRS and NCPRs is highest for the proposed method.

To give a completeness to the current work, registration of a pair of images is shown in Figure 9. Once the correspondence is established, MLESAC [22] robust estimator is used to refine the correspondence and reject the outliers for estimating the transformation between the pair of given retinal images.

4. CONCLUSION

A local descriptor based on Radon transform has been proposed for robust matching in multi-modal retinal im-

Table 2: Mean Similarity : Corresponding and Non-Corresponding Pairs

	Radon Descriptor	PIIFD	SIFT
CPRs	0.69	0.63	0.53
NCPRs	0.38	0.35	0.28
Difference	0.31	0.28	0.24

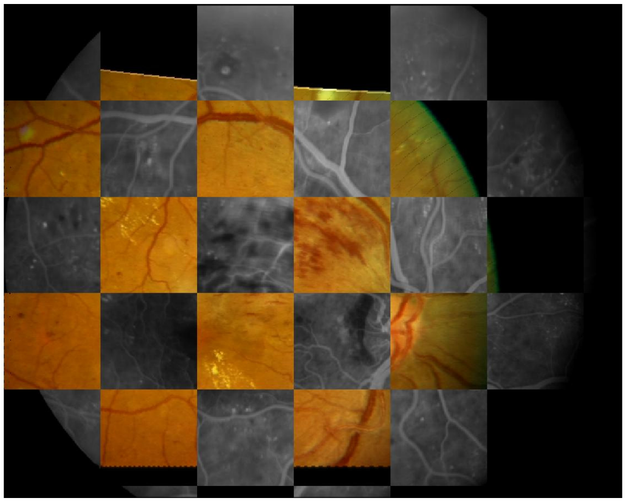


Figure 9: Registration of challenging image pair

ages. The proposed method uses curvature(Hessian) based enhancement to boost the vessel structures and Radon transform based representation to make it invariant to geometric changes. The attractive feature of this descriptor is its robustness to the presence of lesions while still retaining the required structural information. We show superior performance of our descriptor over PIIFD and SIFT. In future we aim to extend this operator for use as a generic descriptor for multimodal matching in medical images. We also intend to build a complete registration frame work for the retinal images. Better evaluation scheme for the framework would be the prime objective of this part of the research.

5. REFERENCES

[1] J. Chen, J. Tian, N. Lee, J. Zheng, R. Smith, and A. Laine. A partial intensity invariant feature descriptor for multimodal retinal image registration. *IEEE Trans Biomed Eng*, 2010.

[2] Neil Ryan, Conor Heneghan, and Philip de Chazal. Registration of digital retinal images using landmark correspondence by expectation maximization. *Image and Vision Computing*, 22(11):883 – 898, 2004.

[3] Charles V. Stewart, Chia ling Tsai, and Badrinath Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans. Med. Img*, 22, 2003.

[4] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita. Automatic retinal image registration scheme using global optimization techniques. *IEEE Trans. Inform. Technol*, 1999.

[5] Frederik Maes, AndrÃľ Collignon, Dirk V, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16:187–198, 1997.

[6] Neil Ryan, Conor Heneghan, and Philip De Chazal. Registration of digital retinal images using landmark, 2004.

[7] William E. Hart and Michael H. Goldbaum. Registering retinal images using automatically selected control point pairs. In *Proceedings of IEEE International Conference on Image Processing 3*, pages 576–581, 1994.

[8] LalibertÃľ F Gagnon L and Sheng Y. Registration and fusion of retinal images-an evaluation study. *Medical Imaging, IEEE Transactions on*, 22(5):661–673, May 2003.

[9] Avi Kelman, Michal Sofka, and Charles V. Stewart. Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. In *CVPR,2007. IEEE Conference on*, pages 1–7, June 2007.

[10] Fransen SR. Chanwimaluang T, Fan G. Hybrid retinal image registration. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):129–142, Jan. 2006.

[11] Ali Can, Charles V. Stewart, Badrinath Roysam, and Howard L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *PAMI, IEEE Transactions on*, 24(3):347–364, Mar 2002.

[12] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[13] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008.

[14] Chen J, Smith R, Tian J, and Laine AF. A novel registration method for retinal images based on local features. In *EMBS,International Conference of the IEEE*, pages 2242–2245, Aug. 2008.

[15] Salvatore Tabbone, Oriol Ramos Terrades, and Sabine Barrat. Histogram of Radon Transform. A useful descriptor for shape retrieval. In *ICPR*, Tampa United States, 2008.

[16] S. Tabbone, L. Wendling, and J. P. Salmon. A new shape descriptor defined on the radon transform. *Computer Vision and Image Understanding*, 102(1):42–51, 2006.

[17] Xuan Wang, Bin Xiao, Jian-Feng Ma, and Xiu-Li Bi. Scaling and rotation invariant analysis approach to object recognition based on radon and fourier-mellin transforms. *Pattern Recogn.*, 40(12):3503–3508, 2007.

[18] Keerthi Ram, Yogesh Babu, and Jayanth Sivaswamy. Curvature orientation histograms for detection and matching of vascular landmarks in retinal images. In *Proceedings of SPIE-Medical Imaging*, Feb. 2009.

[19] K. Jafari-Khouzani and H. Soltanian-Zadeh. Radon transform orientation estimation for rotation invariant texture analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):1004 –1008, jun. 2005.

[20] Alexander Strehl. Conversion of distance metric to similarity.

[21] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, 2005.

[22] P. H. S. Torr and A. Zisserman. Mlesac: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.*, 78(1):138–156, 2000.