

# MULTIMODAL RETINAL IMAGE REGISTRATION AND FUSION BASED ON SPARSE REGULARIZATION VIA A GENERALIZED MINIMAX-CONCAVE PENALTY

Xin Tian, Rencheng Zheng, Colin J Chu, Oliver H Bell, Lindsay B Nicholson, Alin Achim

University of Bristol, UK

## ABSTRACT

We introduce a novel framework for the fusion of retinal OCT and confocal images of mice with uveitis. Input images are semi-automatically registered and then fused to provide more informative retinal images for analysis by ophthalmologists and clinicians. The proposed feature-based registration approach extracts vessels through the use of the ISO-DATA algorithm and morphological operations, in order to match confocal images with OCT images. Image fusion is formulated as an inverse problem, with the corresponding cost function containing two data attachment terms and a non-convex penalty function (the Generalized Minimax-Concave function) that maintains the overall convexity of the problem. The minimization of the cost function is thus tackled by convex optimization. Objective assessment results on image fusion show that this novel image fusion method has competitive performance when compared to existing image fusion methods. Some features of retina that cannot be observed directly in the original images are shown to be enhanced in the fused representations.

**Index Terms**— Retinal image analysis, multimodal image registration, non-convex penalty, GMC, image fusion

## 1. INTRODUCTION

Multimodal image fusion can provide more informative images especially for medical imaging in which image registration is often the first critical step. Accurate image registration is crucial as fusing misaligned images together will not only provide less information but also wrong information [1]. Retinal images are usually registered by featured-based methods instead of intensity-based methods. The challenge of registering OCT image and confocal images is that confocal images captured *ex vivo* are rare in retinal image analysis, and OCT images of rodent eyes have very low quality. There are no existing methods that can be applied directly to these datasets. Thus, a multimodal retinal image registration method for this special dataset is needed.

Recently, numerous signal and image processing techniques including image fusion based on sparse approximation were proposed by researchers [2]. It is common to use L1-Norm regularized least squares method to find solutions

of sparse approximations. Instead of using L1-Norm regularization, Selesnick [3] introduced a novel non-convex penalty function, the Generalized Minimax-Concave (GMC), which can avoid the systematic underestimation effect of the L1-Norm, but is able to maintain the overall convexity of the cost function. The GMC penalty has been proved to be superior to L1-Norm regularization and some other non-convex regularization in 1-D signal restoration. The cost function with GMC penalty is formulated as:

$$F(x) = \frac{1}{2} \|y - Ax\|_2^2 + \lambda \Psi_B(x), \lambda > 0, \quad (1)$$

where  $\Psi_B(x)$  is the non-convex GMC penalty, defined as:

$$\Psi_B(x) = \|x\|_1 - S_B(x), \quad (2)$$

where  $S_B(x)$  is the generalized Huber function defined in the form of an infimal convolution:

$$S_B(x) = \min_{v \in \mathbb{R}^N} (\|v\|_1 + \frac{1}{2} \|B(x - v)\|_2^2). \quad (3)$$

The cost function  $F(x)$  is convex when:

$$B^T B \leq \frac{1}{\lambda} A^T A, \quad (4)$$

with suitable  $B$  matrix chosen, the convexity of the cost function  $F(x)$  is maintained and the solution can be found through convex optimization.

In this paper, a feature-based semiautomatic image registration method for mouse retinal OCT and confocal images is proposed. The GMC penalty is applied to 2-D signal field for image fusion. The paper presents the steps of image registration, and introduces the novel image fusion model based on sparse approximation and GMC penalty. The results of the proposed image registration and fusion are also presented and discussed.

## 2. METHODOLOGY

### 2.1. Multimodal image registration

The proposed registration method is feature-based and semi-automatic. One OCT image with less noise and clearer vessel

contour is selected to be the reference image and all confocal images are set as moving images. Features used to create point-by-point correspondence between OCT and confocal images are vessels, the centroid and the perimeter of the optic disc. Steps for registration include vessel segmentation, optic disc localization, and similarity measurement.

As OCT images display only main vessels and the red channel in confocal images contains capillaries, while the blue channel does not show the shape of any vessel, the green channel is segmented to achieve more accurate similarity measurement result. The unsupervised ISDODATA (Iterative Self-Organizing Data) classification method [4] is used to classify vessel elements in images. In the binary image, pixels considered as vessel are with value one and background with zero (Fig 1 c, d).

As optic disc is where vessels converge in retina, the coordinate of the optic disc centroid is used as a reference point for initial translation, and the size of the optic disc is used for initial scaling. Optic disc localization is implemented through morphological operations, which involve erosion, which achieves image shrinkage and dilation, which scales up the image [5]. A square region where optic disc lies is constrained and eroded to separate the vessels from the optic disc. Separated vessels are eliminated and the image is then dilated to restore pixels of optic disc removed by erosion. Then, the centroid and the area of the optic disc can be calculated.

To find the best geometric transformation criterion for registration, the binary confocal images are rotated from 0 to 360 and scaled within a certain range at each angle after the initial scaling. The XOR (exclusive-OR) based similarity measurement is executed simultaneously at each rotation angle and each scaling size for every image. The sum of the numeric results from every XOR operation is defined as misalignment value, the smaller the misalignment value, the better the matching. When the lowest misalignment value is reached for the first confocal image, the best transformation criterion for this confocal image is found and the corresponding rotation angle and scaling size are stored. A new confocal image is imported and all the steps above are repeated to find its transformation criterion. The new transformation criterion is compared to the former transformation criterion, if the new misalignment value is lower, the best transformation criterion will be replaced by the new one; if not, the best transformation criterion remains to be the previous criterion. Images registered according to the best transformation criterion are then ready for fusion.

## 2.2. Multimodal image fusion

To allow 2-D image fusion, the general cost function (1) in 1-D signal processing is extended as:

$$J(x) = \frac{1}{2} \|y_1 - \beta_1 H_1 W x\|_2^2 + \frac{1}{2} \|y_2 - \beta_2 H_2 W x\|_2^2 + \lambda \Psi_B(x), \lambda > 0, \quad (5)$$

where  $y_1$  and  $y_2$  are the two input images,  $\beta_1$  and  $\beta_2$  are two sensor selectivity coefficients (sensor gain) to be estimated through principal component analysis (PCA) [6],  $H_1$  and  $H_2$  are the point spread functions, which represent convolution operators,  $W$  represents the discrete wavelet transform,  $\lambda$  is the regularization parameter,  $\hat{x}$  is the fused result. The cost function is proved to be convex when:

$$B = \sqrt{\frac{\gamma}{\lambda} (W^T H_1^T \beta_1^2 H_1 W + W^T H_2^T \beta_2^2 H_2 W)}, 0 \leq \gamma \leq 1. \quad (6)$$

**Proof:** Let  $v \in \mathbb{R}^N$ , and

$$Z_1 = \frac{1}{2} \|y_1 - \beta_1 H_1 W x\|_2^2, \quad (7)$$

$$Z_2 = \frac{1}{2} \|y_2 - \beta_2 H_2 W x\|_2^2. \quad (8)$$

Then the cost function

$$\begin{aligned} J(x) &= Z_1 + Z_2 + \lambda \Psi_B(x) \\ &= Z_1 + Z_2 + \lambda \|x\|_1 - \min_{v \in \mathbb{R}^N} (\lambda \|v\|_1 + \frac{\lambda}{2} \|B(x-v)\|_2^2) \\ &= \max_{v \in \mathbb{R}^N} (Z_1 + Z_2 + \lambda \|x\|_1 - \lambda \|v\|_1 - \frac{\lambda}{2} \|B(x-v)\|_2^2) \\ &= \max_{v \in \mathbb{R}^N} \left( \frac{1}{2} x^T Z_3 x + \lambda \|x\|_1 + g(x, v) \right) \\ &= \frac{1}{2} x^T Z_3 x + \lambda \|x\|_1 + \max_{v \in \mathbb{R}^N} g(x, v), \end{aligned} \quad (9)$$

where

$$Z_3 = W^T H_1^T \beta_1^2 H_1 W + W^T H_2^T \beta_2^2 H_2 W - \lambda B^T B, \quad (10)$$

where  $g(x, v)$  is affine in  $x$ ,  $\max_{v \in \mathbb{R}^N} g(x, v)$  is convex since it is the pointwise maximum of a set of convex functions. Therefore, if  $Z_3$  in (9) is positive semidefinite, the cost function  $J(x)$  can be proved to be convex. Hence the cost function (5) is convex when the suitable matrix  $B$  meets (6).

The image fusion model for single channel images can be derived from the model above by setting the operators as the identity matrix ( $H_1 = H_2 = I$ ), so that the cost function for image fusion is:

$$J(x) = \frac{1}{2} \|y_1 - \beta_1 W x\|_2^2 + \frac{1}{2} \|y_2 - \beta_2 W x\|_2^2 + \lambda \Psi_B(x), \lambda > 0. \quad (11)$$

The condition for convexity is:

$$B = \sqrt{\frac{\gamma}{\lambda} W^T (\beta_1^2 + \beta_2^2) W}. \quad (12)$$

As the sensor gain is the normalized eigenvalue, it satisfies:

$$\beta_1^2 + \beta_2^2 = 1. \quad (13)$$

Thus, the convexity condition can be simplified as:

$$B = \sqrt{\frac{\gamma}{\lambda} W}, \lambda > 0. \quad (14)$$

While the cost function has met the condition of convexity, the minimization problem of the cost function can be regarded as a saddle-point problem [3] which can be rewritten as:

$$(x^{\text{opt}}, v^{\text{opt}}) = \arg \min_{x \in \mathbb{R}^N} \max_{v \in \mathbb{R}^N} F(x, v), \quad (15)$$

where

$$F(x, v) = \frac{1}{2} \|y_1 - \beta_1 W x\|_2^2 + \frac{1}{2} \|y_2 - \beta_2 W x\|_2^2 + \lambda \|x\|_1 - \lambda \|v\|_1 - \frac{\gamma}{2} \|W(x - v)\|_2^2, \lambda > 0. \quad (16)$$

As saddle-point problems are instances of monotone inclusion problems, they can be solved by an iterative algorithm, such as the forward-backward splitting (FBS) [7]. The iterative thresholding algorithm is based on the soft thresholding function, which is defined as:

$$\text{soft}(x, y) = (\max(|x| - y, 0)) \text{sign}(x). \quad (17)$$

Let  $0 < \mu < \frac{2}{\varepsilon}$ , where  $\varepsilon = \max\left(1, \frac{\gamma}{1 - \gamma}\right)$ . Then the sequences  $x^{(i)}$  and  $v^{(i)}$ ,  $i \in N$ , generated by the iteration:

$$w^{(i)} = x^{(i)} - \mu [W^T \beta_1 (\beta_1 W x^{(i)} - y_1) + W^T \beta_2 (\beta_2 W x^{(i)} - y_2) + \gamma W^T (W(v^{(i)} - x^{(i)}))], \quad (18)$$

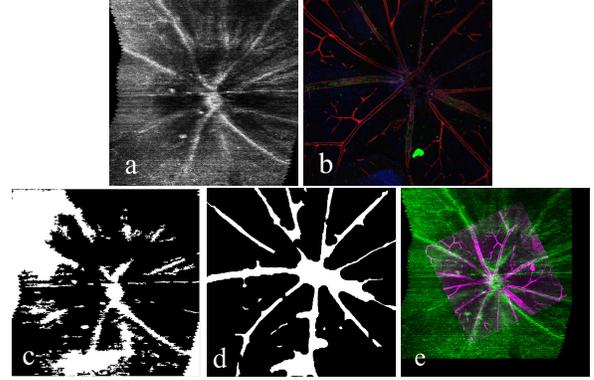
$$u^{(i)} = v^{(i)} - \mu \gamma W^T (W(v^{(i)} - x^{(i)})), \quad (19)$$

$$x^{(i+1)} = \text{soft}(w^{(i)}, \mu \lambda), \quad (20)$$

$$v^{(i+1)} = \text{soft}(u^{(i)}, \mu \lambda), \quad (21)$$

converge to a minimizer of the objective function defined by function (16). After the solution of  $(x^{\text{opt}}, v^{\text{opt}})$  is found, the fused image is restored by taking an inverse DWT of the sparse approximation of the fused image  $\hat{x}$ . In this work, the number of iterations  $i$  is set to 25,  $\lambda$  is 0.001,  $\gamma$  is 0.8, and the number of decomposition level is level 2.

To keep colour information in confocal images, duplications of the greyscale OCT images are concatenated to create a three channel OCT image. Same channel in OCT and confocal images are fused separately and concatenated after. The image fusion algorithm is applied to both images in RGB format and HSV format where  $\lambda$  and  $\gamma$  remain the same. Before fusion, intensities of confocal images are adjusted to achieve better fusion results.



**Fig. 1.** Multimodal retina image registration a, original OCT image; b, original confocal image; c, segmented OCT image; d, segmented green channel confocal image; e, registered OCT (green) and confocal (purple) image

### 3. RESULTS AND DISCUSSION

The images used in this paper are from mice with autoimmune uveitis induced by immunizing with peptide derived from retinol binding protein-3, peptide 1-20. This provokes an immune response that causes inflammation in the retina. The process is controlled by CD4+ T-lymphocytes.

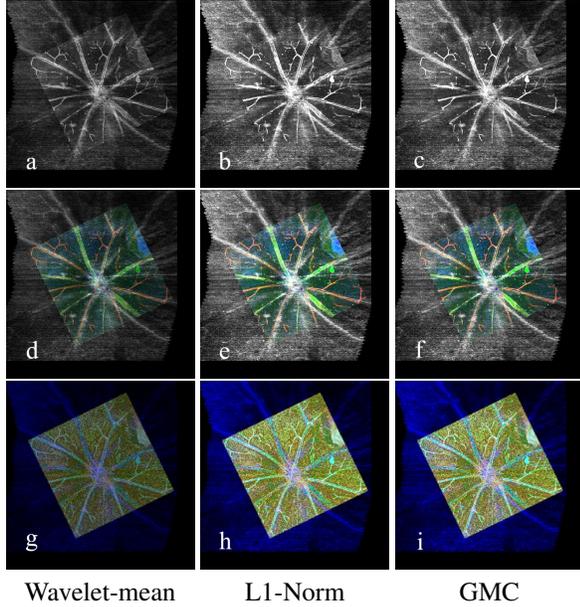
OCT scans were acquired using the Micron IV fundus camera and an OCT scan head equipped with a mouse objective lens (Phoenix Technologies, California) at different time points (days 10, 14, 17, 24) *in vivo*. B scans are preprocessed [8] to build the B scan cube. Enface of the OCT B scans (day 24) used for registration and fusion are extracted from the B scan cube (Fig 1 a).

The mice were killed on day 25 to carry out confocal images *ex vivo*. Four incisions around the outside of the retina were made to allow it to be flat-mounted. Retinas were stained with antibodies attached to four different fluorochromes. Images were processed to four colour channels: blue (DAPI) stains cell nuclei, green (Alexa-488) stains CD4+ T cells which mainly lie in veins, red (rhodamine red, staining isolectin IB4) stains endothelial cells lining blood vessels and white (Alexa-633, staining Iba1) stains microglia and macrophages. Serial images were obtained through the thickness of the retina using a Leica SP5-AOBS confocal laser scanning microscope attached to a Leica DM I6000 inverted epifluorescence microscope (Fig 1 b).

#### 3.1. Multimodal image registration

Registration results are shown in Fig 1 e. The optimal rotation angle derived from the semiautomatic registration algorithm for confocal images is  $116^\circ$ , and confocal images shrink to 58.6% of the original size.

Most vessels are well registered but there are still some misalignments mainly because images captured by different imaging techniques naturally have distortions, especially



**Fig. 2.** Retina images fused through wavelet-mean, L1-Norm, and GMC method. a, b and c are grey images; d, e and f are grey-RGB images; g, h and i are HSV images

when OCT images are not flattened but confocal images are naturally flattened as they are captured *ex vivo*. Thus, non-rigid transformation and proper OCT flatten algorithm are needed.

Furthermore, the poor quality of the OCT images makes vessels in OCT images to not be perfectly segmented. Although denoising methods such as wavelet hard/soft thresholding filter, bilateral filter, etc. are applied, they only provide very limited denoising effect but more of complicating the algorithm. Therefore, more efficient denoising methods should be applied in future work.

### 3.2. Multimodal image fusion

Fusion results are presented in Fig 2. As a comparison, images are also fused by classical wavelet domain image fusion algorithm with average coefficients (wavelet-mean) [9], and inverse modal with L1-Norm regularization [10]. Compared to original images, fused grey images visually present clearer vessel edges. Grey-RGB images preserved color information in confocal images which can help clinicians learn more about vessel types and distribution in retina. HSV images have strong visual impact compared to RGB images. No matter under which image fusion format, the main difference between images gained through wavelet-mean method and other two methods is the luminance.

Features that cannot be observed directly from original images appeared after image fusion. It can be observed from the grey-RGB images (Fig 2 d, e, and f) that the neighbors of a green vessel are two red vessels and the neighbours of a red vessel are two green vessels. The alternately distributed

red vessels and green vessels indicate that arteries and veins in mouse retina are alternately distributed which is consistent with known anatomical facts.

Three different assessment metrics, Piella’s metric [11], Cvejic’s metric [12], and Xydeas’s metric [13], are used for objective performance evaluation (Table 1). Quality metrics range between 0 and 1 with 0 as worst and 1 as best results. For grey image fusion, GMC and L1-Norm method present similar performance and are better than wavelet-mean method. For grey-RGB image fusion, GMC shows its advantages over both L1-Norm and wavelet-mean methods. However, L1-Norm shows its advantage over the other two methods in HSV image fusion, especially with Xydeas’s metric and Cvejic’s metric. The overall numeric results indicate that the GMC regularization and the L1-regularization have competitive performance in image fusion and they all perform better than the classical wavelet-mean method.

**Table 1.** Objective assessment results of grey image, grey-RGB image, and HSV image fusion

	Metric	Wavelet-Mean	L1-Norm	GMC
Grey	Piella	0.5248	0.7203	<b>0.7257</b>
	Cvejic	0.3906	0.5419	<b>0.5454</b>
	Xydeas	0.2380	<b>0.5306</b>	0.5275
Grey-RGB	Piella	0.5491	0.7520	<b>0.7589</b>
	Cvejic	0.4062	0.5706	<b>0.5982</b>
	Xydeas	0.2562	0.5653	<b>0.5810</b>
HSV	Piella	0.3291	<b>0.7991</b>	0.7672
	Cvejic	0.3154	<b>0.6453</b>	0.5810
	Xydeas	0.2962	<b>0.6867</b>	0.5982

## 4. CONCLUSION AND FUTURE WORK

This paper provides a featured-based semiautomatic multimodal retinal image registration approach for registering mouse OCT retinal images with confocal retinal images. This registration method efficiently registers mouse OCT retina images with confocal retinal images and can be applied not only to the dataset used in this work but also similar dataset with only a few parameters (erosion and dilation level, etc.) changing. Nevertheless, improvements are needed in vessel segmentation in OCT image and spatial misalignment elimination between multimodal images to deliver more accurate image registration.

The GMC penalty is extended to 2-D signal processing to build a novel image fusion method which is based on sparse representation and GMC penalty. This method allows fusing not only greyscale images but also colour images. With multimodal retina images fused, some features of retina are clearer, e.g. the fact that the structure of vessel distribution in retina is brought to the surface after image fusion. The objective quality metrics suggest that GMC regularization offers competitive performance when compared to L1-Norm regularization and is better than the classical wavelet domain image fusion using average coefficients in image fusion.

## 5. REFERENCES

- [1] F. Oliveira and J. Tavares, "Medical image registration: a review," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, pp. 73–93, 2012.
- [2] J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*, 2nd ed. Cambridge University Press, 2015.
- [3] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Transactions on Signal Processing*, vol. 65, pp. 4481–4494, 2017.
- [4] A. El-Zaart, "Images thresholding using isodata technique with gamma distribution," *Pattern Recognition and Image Analysis*, vol. 20, pp. 29–41, 2010.
- [5] A. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1200–1213, 2006.
- [6] M. Kumar and S. Dass, "A total variation-based algorithm for pixel-level image fusion," *IEEE Transactions on Image Processing*, vol. 18, pp. 2137–2143, 2009.
- [7] Y. Huang and Y. Dong, "New properties of forward-backward splitting and a practical proximal-descent algorithm," *Applied Mathematics and Computation*, vol. 237, pp. 60–68, 2014.
- [8] N. Anantrasirichai, L. Nicholson, J. E. Morgan, I. Erchova, and A. Achim, "Adaptive-weighted bilateral filtering and other pre-processing techniques for optical coherence tomography," *Computerized Medical Imaging and Graphics*, vol. 38, pp. 526–239, 2014.
- [9] G. PAJARES, "A wavelet-based image fusion tutorial," *Pattern Recognition*, 2004.
- [10] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *Applied Mathematics and Computation*, vol. 237, pp. 60–68, 2014.
- [11] G. Piella and H. Heijmans, "A new quality metric for image fusion," *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, 2003.
- [12] N. Cvejic, C. Canagarajah, and D. Bull, "Image fusion metric based on mutual information and tsallis entropy," *Electronics Letters*, vol. 42, p. 626, 2006.
- [13] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, p. 308, 2000.