

# Joint alignment of multispectral images via semidefinite programming

YUANJIE ZHENG,<sup>1,2,3,\*</sup> YU WANG,<sup>1</sup> WANZHEN JIAO,<sup>4</sup> SUJUAN HOU,<sup>1</sup> YANJU REN,<sup>5</sup> MAOLING QIN,<sup>1</sup> DEWEN HOU,<sup>1</sup> CHAO LUO,<sup>1</sup> HONG WANG,<sup>1,2,3</sup> JAMES GEE,<sup>6</sup> AND BOJUN ZHAO<sup>4</sup>

<sup>1</sup>School of Information Science & Engineering, Shandong Normal University, Jinan, China

<sup>2</sup>Institute of Life Sciences at Shandong Normal University, Jinan, China

<sup>3</sup>Key Lab of Intelligent Information Processing at Shandong Normal University, Jinan, China

<sup>4</sup>Dept. of Ophthalmology, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China

<sup>5</sup>School of Psychology, Shandong Normal University, Jinan, China

<sup>6</sup>Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

\*yjzheng@sdu.edu.cn

**Abstract:** In this paper, we introduce a novel feature-point-matching based framework for achieving an optimized joint-alignment of sequential images from multispectral imaging (MSI). It solves a low-rank and semidefinite matrix that stores all pairwise-image feature-mappings by minimizing the total amount of point-to-point matching cost via a convex optimization of a semidefinite programming formulation. This unique strategy takes a complete consideration of the information aggregated by all point-matching costs and enables the entire set of pairwise-image feature-mappings to be solved simultaneously and near-optimally. Our framework is capable of running in an automatic or interactive fashion, offering an effective tool for eliminating spatial misalignments introduced into sequential MSI images during the imaging process. Our experimental results obtained from a database of 28 sequences of MSI images of human eye demonstrate the superior performances of our approach to the state-of-the-art techniques. Our framework is potentially invaluable in a large variety of practical applications of MSI images.

© 2017 Optical Society of America

**OCIS codes:** (110.4234) Multispectral and hyperspectral imaging; (100.4145) Motion, hyperspectral image processing; (100.0100) Image processing; (100.2960) Image analysis; (170.4470) Ophthalmology.

## References and links

1. P. Ghassemi, T. E. Travis, L. T. Moffatt, J. W. Shupp, and J. C. Ramella-Roman, "A polarized multispectral imaging system for quantitative assessment of hypertrophic scars," *Biomed. Opt. Express* **5**(10), 3337–3354 (2014).
2. N. T. Clancy, S. Arya, D. Stoyanov, M. Singh, G. B. Hanna, and D. S. Elson, "Intraoperative measurement of bowel oxygen saturation using a multispectral imaging laparoscope," *Biomed. Opt. Express* **6**(10), 4179–4190 (2015).
3. M. B. Bouchard, B. R. Chen, S. A. Burgess, and E. M. Hillman, "Ultra-fast multispectral optical imaging of cortical oxygenation, blood flow, and intracellular calcium dynamics," *Opt. Express* **17**(18), 15670–15678 (2009).
4. C. Zimmer, D. Kahn, R. Clayton, P. Dugel, and K. Freund, "Innovation in diagnostic retinal imaging: multispectral imaging," *Retina Today* **9**(7), 94–99 (2014).
5. D. Hitchmoth, "Multispectral Imaging: A revolution in retinal diagnosis and health assessment," *Adv. Ocul. Care* **4**(4), 76–79 (2013).
6. D. L. Shechtman and P. M. Karpecki, "A look at MSI: multispectral imaging may help eye care providers diagnose retinal conditions earlier than conventional funduscopy," *Rev. Opt.* **149**(1), 88–90 (2012).
7. R. Maharaj, "The clinical applications of multispectral imaging," *Rev. Opt.* **148**(11), SS19–SS19 (2011).
8. N. Everdell, I. Styles, E. Claridge, J. Hebden, and A. Calcagni, "Multispectral imaging of the ocular fundus using LED illumination," in *European Conferences on Biomedical Optics*, pp. 73,711C–73,711C (International Society for Optics and Photonics, 2009).
9. Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," *Medical image analysis* **15**(4), 622–639 (2011).
10. F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review," *Comput. Method. Biomec.* **17**(2), 73–93 (2014).
11. A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath, "Brain functional localization: a survey of image registration techniques," *IEEE Trans. Med. Imaging* **26**, 427–451 (2007).
12. J. Modersitzki, *Numerical Methods for Image Registration* (Oxford University, New York, NY, USA, 2004).

13. L. Zöllei, E. Learned-Miller, E. Grimson, and W. Wells, "Efficient population registration of 3D data," in *International Workshop on Computer Vision for Biomedical Image Applications*, pp. 291–301 (Springer, 2005).
14. C. Wachinger and N. Navab, "Simultaneous registration of multiple images: Similarity metrics and efficient optimization," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(5), 1221–1233 (2013).
15. J. Kim, C. Liu, F. Sha, and K. Grauman, "Deformable spatial pyramid matching for fast dense correspondences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2307–2314 (2013).
16. T. Zhou, Y. J. Lee, X. Y. Stella, and A. A. Efros, "Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1191–1200 (IEEE, 2015).
17. R. Estrada, C. Tomasi, M. T. Cabrera, D. K. Wallace, S. F. Freedman, and S. Farsiu, "Enhanced video indirect ophthalmoscopy (VIO) via robust mosaicing," *Biomedical optics express* **2**(10), 2871–2887 (2011).
18. H. Rabbani, M. J. Allingham, P. S. Mettu, S. W. Cousins, and S. Farsiu, "Fully Automatic Segmentation of Fluorescein Leakage in Subjects With Diabetic Macular Edema Automatic Leakage Segmentation in DME," *Investigative Ophthalmol. Vis. Sci.* **56**(3), 1482–1492 (2015).
19. C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. A. Argyros, "Retinal image registration under the assumption of a spherical eye," *Computerized Medical Imaging and Graphics* (2016).
20. E. G. Learned-Miller, "Data driven image models through continuous joint alignment," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(2), 236–250 (2006).
21. G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8 (IEEE, 2007).
22. S. K. Balci, P. Golland, and W. Wells, "Non-rigid groupwise registration using B-spline deformation model," *Open source and open data for MICCAI* pp. 105–121 (2007).
23. A. Rasouli, R. Rohling, and P. Abolmaesumi, "Group-wise registration of point sets for statistical shape models," *IEEE Trans. Med. Imaging* **31**(11), 2025–2034 (2012).
24. P. Ghosh and B. Manjunath, "Robust simultaneous registration and segmentation with sparse error reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(2), 425–436 (2013).
25. A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(2), 372–386 (2012).
26. J. Kutarnia and P. Pedersen, "A Markov random field approach to group-wise registration/mosaicing with application to ultrasound," *Med. Image Anal.* **24**(1), 106–124 (2015).
27. O. Arandjelovic, D.-S. Pham, and S. Venkatesh, "Groupwise registration of aerial images," presented on the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015) (2015).
28. X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," in *European Conference on Computer Vision*, pp. 309–324 (Springer, 2014).
29. B. Li, W. Wang, and H. Ye, "Multi-sensor image registration based on algebraic projective invariants," *Optics express* **21**(8), 9824–9838 (2013).
30. J. Lin, Y. Zheng, W. Jiao, B. Zhao, S. Zhang, J. Gee, and R. Xiao, "Groupwise registration of sequential images from multispectral imaging (MSI) of the retina and choroid," *Optics Express* **24**(22), 25,277–25,290 (2016).
31. Q.-X. Huang and L. Guibas, "Consistent shape maps via semidefinite programming," in *Computer Graphics Forum*, vol. 32, pp. 177–186 (Wiley Online Library, 2013).
32. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning* **3**(1), 1–122 (2011).
33. D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157 (IEEE, 1999).
34. J. Lewis, "Fast normalized cross-correlation," in *Vision Inter.*, vol. 10, pp. 120–123 (1995).
35. C. V. Stewart, "Robust parameter estimation in computer vision," *SIAM Rev.* **41**(3), 513–537 (1999).
36. C. Rorden and M. Brett, "Stereotaxic display of brain lesions," *Behav. Neurol.* **12**(4), 191–200 (2000).
37. Y. Zheng, P. Parikh, B. L. VanderBeek, and J. Gee, "Automated Quantification of Morphologic Features and Vasculature of Choroid on Enhanced Depth Imaging Optical Coherence Tomography (EDI-OCT)," *Investigative Ophthalmology & Visual Science* **57**(12), 4652–4652 (2016).
38. S. Park, S.-K. Park, and M. Hebert, "Fast and scalable approximate spectral matching for higher order graph matching," *IEEE transactions on pattern analysis and machine intelligence* **36**(3), 479–492 (2014).
39. Y. Li, C. Chen, F. Yang, and J. Huang, "Deep sparse representation for robust image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4894–4901 (2015).
40. J. Yao, Z. Xu, X. Huang, and J. Huang, "Accelerated dynamic MRI reconstruction with total variation and nuclear norm regularization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 635–642 (Springer, 2015).
41. Y. Zheng, B. Wei, H. Liu, R. Xiao, and J. C. Gee, "Measuring sparse temporal-variation for accurate registration of dynamic contrast-enhanced breast MR images," *Computerized Medical Imaging and Graphics* **46**, 73–80 (2015).
42. X. Peng, S. Zhang, Y. Yang, and D. N. Metaxas, "Pief: Personalized incremental and ensemble face alignment," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3880–3888 (2015).
43. Y. Zheng, S. Zhang, J. Huang, and W. Cai, "Guest Editorial: Special issue on advances in computing techniques for

big medical image data,” Neurocomputing (2016).

## 1. Introduction

Multispectral imaging (MSI) refers to imaging systems that penetrate certain features of physical object within the field of view by using a number of spectral bands [1–3]. It has been used in various applications such as but not limited to airborne mapping, astronomical imaging, dentistry, dermatology, histopathology and ophthalmology [4]. Taking MSI in ocular imaging for example, the inner limiting membrane all the way to the choroid can be dissected and visualized by taking a sequence of spectral images via MSI systems like the Retinal Health Assessment (RHA, Annidis Health System Corp.) [5–7]. MSI provides a collection of noninvasive *en face* views of the retina and choroid, that collect and combine spectral information to highlight anatomic and metabolic signatures for diagnosing and managing diseases. Interpretation of MSI images based on human visual inspection remains the reference standard in MSI applications. However, computer based algorithms are attracting more and more research interests due to the ability to provide objective and automatic image assessments.

One grand challenge in analyzing MSI images comes from the image-to-image spatial misalignment which is caused by object movement during the imaging process. For example, we need an analysis of multiple MSI slices from different spectral bands to assess the retinal pigment epithelium (RPE) disruption because it is characterized by a slight structural variation of pigment pattern in red and infrared spectral slices in contrast to normal structures in shorter wavelengths [4]. However, it is not a trivial in practice due to the misalignment between these spectral slices mainly because of the fact that the acquisition time of eye MSI images is commonly longer than the natural time scale of eye’s saccadic movement [8]. Similar to the research in other fields [9–16], this spatial misalignment may introduce troubles into not only automatic but also manual interpretation and quantification of the anatomic/metabolic variations between spectral slices.

There exist a variety of techniques for retinal image registration [17–19] in the field of eye and generic image registration [10] in a broader field. A detailed review of all these techniques is beyond the scope of this paper and we only provide a summary on current group image registration methods as they are closely related to the proposed technique in this paper. There are various group image registration approaches, including the well-known congealing framework and its extensions [20–22] which focus on aligning a number of binary images of handwritten digits, 2D face images or 3D medical images via a sequential optimization process to gradually reduce the entropy of image intensity’s distribution. Other approaches try to improve alignment performances by taking advantage of shape matching [23], sparse learning [24, 25], image’s Markov property [14, 26], pairwise registration based on normalized cross-correlation [27], robust matching based on gradient descent optimization [28], image contours [29] and joint alignment by enforcing temporal smoothness on motion [30].

Theoretically speaking, most of the existing image registration techniques can be used for estimating and eliminating the spatial misalignment between MSI images of eye with certain adaptations. However, they are limited according to our experiments because of several basic reasons. First, many of them rely on independent processes of matching pairwise images, in which the joint information of MSI images can’t be taken into account. Second, the point matches may become unreliable when the predefined matching cost is corrupted by the anatomic/metabolic variations of retina across spectral slices. Third, not all feature-mappings of pairwise images are solvable considering the fact that different spectral slices (penetrating different light-absorbing species/chromophores) may be associated with different anatomical structures. Finally, it is not a trivial to accomplish a semi-automated alignment which is of great importance in practice especially for difficult images.

In this paper, we propose a novel framework for achieving a joint discrete alignment of sequential MSI images. It generates near-optimal feature matches between images by jointly aligning all image pairs via a convex optimization process. Given a collection of sequential MSI images, we extract a set of salient feature points from each of them. Then, the proposed framework concerns a recovery of a globally consistent feature matches by taking advantage of the aggregated information from the point-to-point matching cost of all image pairs. Specifically, we encode the feature mapping across each pair of images via a binary matrix and leverage a stacked matrix to store all these binary matrices. Inspired by a recent finding that enforcing global consistency (composite feature maps along cycles are identity maps) on the resulted maps is equivalent to imposing a low-rank constraint upon this stacked matrix [31], we formulate the joint alignment problem as a semidefinite program (SDP) and solve it via a convex optimization to obtain a semidefinite and low-rank stacked matrix.

The proposed approach is effective in dealing with the above challenges in MSI image registration due to several of its advantages over the existing image alignment techniques. First, pairwise image matching and joint refinement of these matches are solved all at once, resulting in a near-optimal solution. Second, expert-specified feature-point correspondences can be seamlessly incorporated into the framework without any change of the optimization technique, which allows human expert to use his knowledge to guide matching in difficult cases. Third, our approach is robust to not only the corruptions in point matches but also the incompleteness of pairwise maps. Experiments of this paper are carried out by using a database consisting of 28 MSI image sequences of human eye and the results not only demonstrate our approach's superior performances to state-of-the-art but also confirm its practical applicability.

## 2. Our approach

### 2.1. Problem formulation

Given a collection of MSI images  $\{I_1, I_2, \dots, I_N\}$ , we extract  $m_i$  feature points from image  $I_i$  and then obtain  $N$  feature sets denoted by  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_N\}$ . For any two feature point sets  $\mathcal{S}_i$  and  $\mathcal{S}_j$ , we define an attributed graph  $\mathcal{G}_{ij} = \{\mathcal{S}_i \cup \mathcal{S}_j, E_{ij}, C_{ij}\}$  where any  $k \in \mathcal{S}_i$  and  $l \in \mathcal{S}_j$  correspond to an edge  $e = kl \in E_{ij} = \mathcal{S}_i \times \mathcal{S}_j$  and are attributed to a cost  $c \in C_{ij} \subset \mathbb{R}^{m_i \times m_j}$ . For  $\mathcal{S}_i$  and  $\mathcal{S}_j$ , we define a partial map  $\phi_{ij} \subset \mathcal{S}_i \times \mathcal{S}_j$  to encode point-to-point maps between these two sets of feature points, with which each element of  $\mathcal{S}_i$  is paired with at most one element of  $\mathcal{S}_j$  and *vice versa*. The ultimate goal of this paper is to propose a tractable algorithm for computing optimal pairwise partial maps  $\Phi = \{\phi_{ij} | 1 \leq i, j \leq N\}$  by taking advantage of the aggregated information contained in all attributed graphs  $\mathcal{G} = \{\mathcal{G}_{ij} | 1 \leq i, j \leq N\}$ .

### 2.2. Matrix representation

We encode each partial map  $\phi_{ij}$  as a binary matrix  $\mathbf{X}_{ij} \in \{0, 1\}^{m_i \times m_j}$  in which the elements associated with  $\phi_{ij}$  take value 1 and the left are 0. When  $\phi_{ij}$  is a one-to-one map,  $\mathbf{X}_{ij}$  satisfies the following constraints

$$\begin{cases} \mathbf{X}_{ij}\mathbf{1} = \mathbf{1}, & \text{if } |\mathcal{S}_i| \leq |\mathcal{S}_j| \\ \mathbf{X}_{ij}^T\mathbf{1} = \mathbf{1}, & \text{otherwise} \end{cases} \quad (1)$$

where  $\mathbf{1}$  represents a vertical vector for which all elements are ones.

We leverage an  $M \times M$  block matrix  $\mathbf{X} \in \{0, 1\}^{M \times M}$  to store the entire collection of  $\{\mathbf{X}_{ij} | 1 \leq i, j \leq N\}$ , where  $M = \sum_{i=1}^N m_i$  denoting the total number of features points extracted

from the entire collection of MSI images:

$$\mathbf{X} = \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1N} \\ \mathbf{X}_{21} & \mathbf{I}_{m_2} & \cdots & \mathbf{X}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{N1} & \mathbf{X}_{N2} & \cdots & \mathbf{I}_{m_N} \end{bmatrix} \quad (2)$$

where the diagonal blocks are identity matrices representing self-maps. Obviously,  $\mathbf{X}_{ij} = \mathbf{X}_{ji}^T$  when  $\phi_{ij}$  and  $\phi_{ji}$  are one-to-one maps and  $\mathbf{X}$  in Eq. (2) is hence a symmetric matrix. In addition, when  $\Phi$  is cycle-consistent (composite maps along cycles, e.g.  $\phi_{ki} \circ \phi_{jk} \circ \phi_{ij}$  for a 3-circle, are identity maps),  $\mathbf{X}$  is low-rank and positive semidefinite (denoted by  $\mathbf{X} \geq 0$ ), as detailed in [31].

Similarly, we build a  $M \times M$  block matrix  $\mathbf{C}$  by stacking all cost matrices  $\{C_{ij} | 1 \leq i, j \leq N\}$  as shown below

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1N} \\ C_{21} & C_{22} & \cdots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \cdots & C_{NN} \end{bmatrix}. \quad (3)$$

### 2.3. Semidefinite programming objective function

Joint MSI image matching can be treated as searching combinatorially maps  $\{\phi_{ij} | 1 \leq i, j \leq N\}$  by minimizing the total matching cost computed by summing the edge cost of all matched points, which amounts to a minimization of the following objective function

$$\min \text{Tr}[\mathbf{CX}] \quad (4)$$

with the following constraints

$$\mathbf{X} \in \{0, 1\}^{M \times M}, \quad (5)$$

$$\mathbf{X} \geq 0 \quad (6)$$

and

$$\mathbf{X}_{ii} = \mathbf{I}_{m_i}, \quad (7)$$

together with the condition in Eq. (1). In Eq. (4),  $\text{Tr}[\cdot]$  denotes trace of matrix.

In difficult cases, expert-specified feature correspondences between MSI images are required to be incorporated into the algorithm for producing accurate results. In this way, human's knowledge can be exploited to guide the matching process. Suppose the manually-provided feature correspondences are associated with a set of elements in  $\mathbf{X}$ , denoted by their rows and columns:  $\mathbf{U} = \{(u, v), 1 \leq u, v \leq M\}$ . Then, we have

$$\mathbf{X}(u, v) = 1, \quad \forall (u, v) \in \mathbf{U}, \quad (8)$$

where  $\mathbf{X}(u, v)$  means the element of  $\mathbf{X}$  as specified by  $(u, v)$ .

### 2.4. Optimization via convex relaxation

Eq. (4) with the constraints in Eqs. (1) and (5)-(8) formulates a standard form of semidefinite programming (SDP). A direct optimization of Eq. (4) is NP-hard because the variables contained in  $\mathbf{X}$  are binary. To deal with this, we relax these variables to take real values in  $[0, 1]$ . Then, Eq. (4) becomes a convex optimization problem which can offer a global optimum within polynomial time. We employ the alternating direction of multiplier method (ADMM) algorithm [32] which solves a convex optimization problem by breaking it into smaller easier-to-handle pieces.

The solutions of ADMM are continuous values within  $[0, 1]$  and we round them by using the simple greedy algorithm in [31]. The basic idea is to resolve a binary version of  $\mathbf{X}$  by minimizing its differences with the continuous version via a linear programming.

### 2.5. Matching cost

Feature points used for matching MSI images are specified as the SIFT [33] key locations due to its prestige of being stable. SIFT detects maxima and minima of the image convoluted with a Gaussian at different image scale followed by a postprocessing of discarding low contrast candidate points and edge response points. A robust cross correlation is used to measure the matching cost between two points [28,30], which is defined as below when the  $k$ th point of  $I_i$  is matched to the  $l$ th point of  $I_j$ :

$$C_{i,j}(k,l) = w \cdot \rho \left( \frac{1}{\gamma(I_i(k), I_j(l))} \right) + (1-w) \cdot \rho \left( \frac{1}{\gamma(\nabla I_i(k), \nabla I_j(l))} \right) \quad (9)$$

where  $w \in [0, 1]$  adjusts the contributions of the two terms,  $I_i(k)$  and  $\nabla I_i(k)$  represent a square patch (e.g.  $15 \times 15$ ) surrounding point  $k$  in  $I_i$  and its gradient space  $\nabla I_i$ , respectively. In Eq. (9),  $I_j(l)$  and  $\nabla I_j(l)$  have a similar meaning with  $I_i(k)$  and  $\nabla I_i(k)$ , respectively, but are defined on point  $l$  in  $I_j$ .  $\gamma(\cdot, \cdot)$  in Eq. (9) is the normalized cross correlation [34]. The robust function  $\rho()$  is defined as the *Geman-McClure* function

$$\rho(t) = \frac{t^2}{1 + t^2/a^2} \quad (10)$$

where  $a$  is a parameter which can be determined automatically from the given image and is used for adjusting the stringency in rejecting outliers [35].

## 3. Experiment

### 3.1. Data

We evaluate our algorithm with a validation database comprising 28 MSI image sequences acquired by using an Annidis RHA™ instrument (Annidis Health Systems Corp., Ottawa, Canada). These images are of oculus dexter (OD) and oculus sinister (OS) from 9 patients diagnosed with hypertensive retinopathy and 8 healthy subjects. They are provided in the format of dicom with a bit depth of 16 and in size of  $2048 \times 2048$ . Each sequence bears at least 8 images captured with wavelengths of amber, green, infrared, red and yellow, as shown in an exempling sequence in Fig. 1.

An ophthalmologist manually pick 15 points for each sequence and marked them in all the MSI images of this sequence by using MRICron [36]. These manual marks are treated as the ground truth for validating our algorithm.

### 3.2. Evaluation

We evaluate several unique characteristics of our framework, including the power of joint alignment in contrast to the pairwise fashion, its noise resistance, the benefits from the introduced interactive mechanism and its superiority of performances to the state-of-the-art techniques, respectively. In our experiments, we set empirically the only parameter  $w$  (in Eq. (9)) of our framework as  $w = 0.8$ .

#### 3.2.1. Power of joint alignment

To validate the power of joint alignment in our approach, we exploit the performance curves used in [31] for assessing the matching errors. Specifically, we calculate for a manually-marked point in a MSI image the distance between its matched point in the other image and the ground-truth correspondence. We treat a match as correct if this distance divided by an approximated value (1035 pixels in our experiments) of the radius of retina is below a threshold. Then, we plot a curve (analogous to the precision-recall curve) showing the ratio (a value in  $[0, 1]$ ) of the

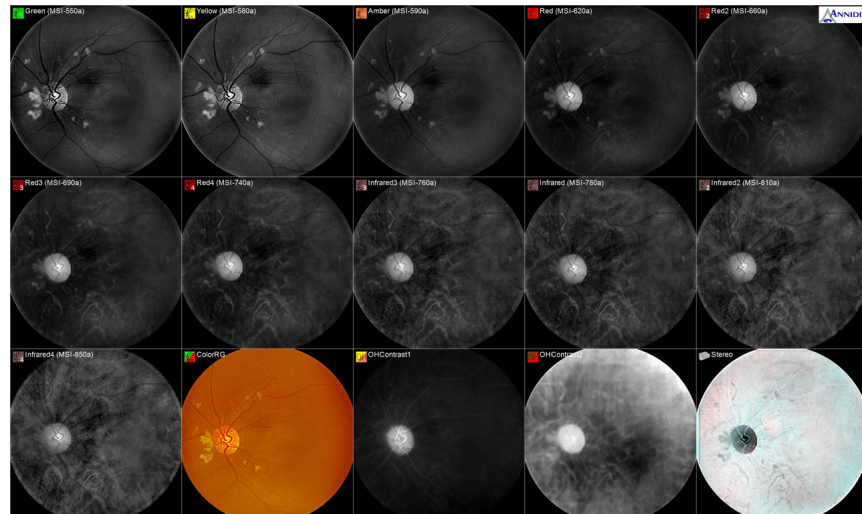


Fig. 1. A sequence of MSI images acquired by Annidis RHA from a patient diagnosed with hypertensive retinopathy, in which from left to right and from top to bottom, the first 11 images are captured with short wavelengths of “Green” (MSI-550), “Yellow” (MSI-580) and “Amber” (MSI-590), respectively, followed by 4 wavelengths of “Red” (MSI-620, MSI-660, MSI-690 and MSI-740) and 4 wavelengths of “Infrared” (MSI-760, MSI-780, MSI-810 and MSI-850), respectively.

manually-marked points with correct matches over the threshold values. For points bearing no corresponded point in the other image, interpolation is conducted to estimate the corresponding location. A curve closer to the upper-left corner indicates that the algorithm performs better. In our experiment, we mix the manually specified marks with the detected SIFT key locations and use them by computing their matching cost as defined in Sec. 2.5.

For each image sequence, we choose randomly a portion (with a percentage 0%, 25%, 50%, 75% and 100%, respectively) of the images and process them with our joint alignment scheme. For the matches of each of the left images to other images, we exploit pairwise alignment which can be carried out by setting  $N = 2$  in our framework as explained in Sec. 2. For the percentages of 25%, 50% and 75%, the random portion choosing process is repeated for 10 times and the resulted ratios are averaged. Final results for each test are computed by averaging over correct-correspondence ratios of all image pairs.

As shown in Fig. 2, we can see that our algorithm produces improved accuracies when more images are aligned in a joint way. It means that more images should be included in joint alignment when using the proposed approach.

### 3.2.2. Noise resistance and benefits from manual matches

To validate the noise resistance property of our approach and the benefit of providing manual matches, we test our approach only with the feature points marked manually by the ophthalmologist. We add Gaussian noise with zero mean and various variances (0, 0001, 0.05, 0.1, 0.15 and 0.2, respectively) in the matching costs computed for these points. Also, we randomly choose different percentages (0%, 25%, 50%, 75% and 95%, respectively) of the pairs of manually-matched points and fix them (via Eq. (8)) when resolving the joint alignment by our approach. The randomly choosing process is repeated for 10 times and the ratios of correct matches are averaged as the final results. The threshold of matching distance is set to 0.03 in our experiments.

We have two findings from the results in Fig. 3. First, the algorithm performs better when

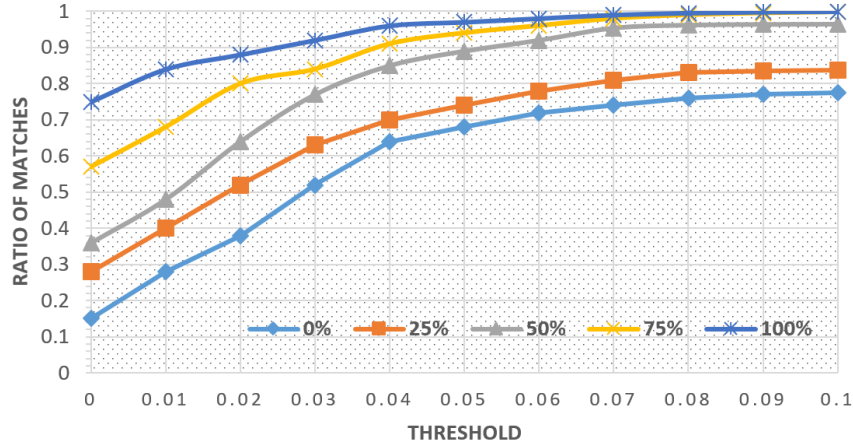


Fig. 2. Performance curves of our approach when different percentage (0%, 25%, 50%, 75% and 100%, respectively) of the images in each sequence are matched jointly while the left ones are matched in a pairwise fashion. The y-axis shows the ratio of correct matches while the x-axis denotes the threshold of matching distance divided by the radius of retina.

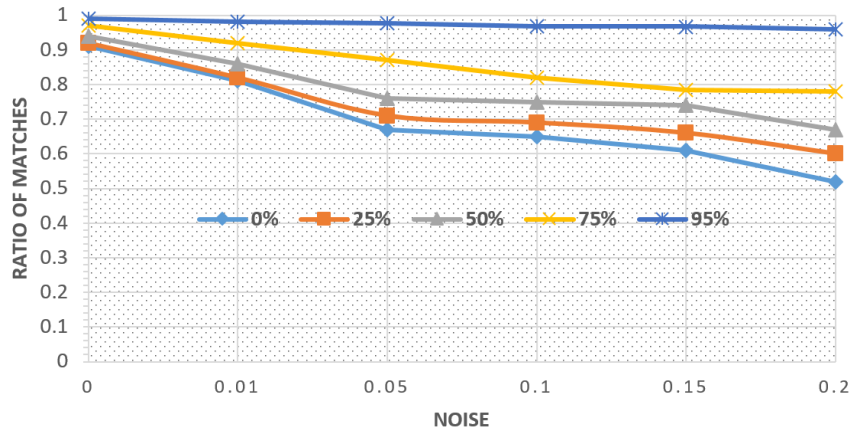


Fig. 3. Performance curve of our approach when Gaussian noise with zero mean and various variances (0, 0001, 0.05, 0.1, 0.15 and 0.2, respectively) is added in the matching cost and when different percentage (0%, 25%, 50%, 75% and 95%, respectively) of manual correspondences are fixed (via Eq. (8)) when solving the joint alignment.



more point correspondences are added into our algorithm as a constraint known *a priori*. For example, the average improvement over all noise levels of the case using 75% of manual correspondences compared with no manual correspondences is 26.2% for all images. We also find that this improvement is even larger (with a value of 37.4%) for the images from patients. They demonstrate the benefit of manual correspondences in improving matching performances especially for difficult images (e.g. images from patients). Second, our algorithm deteriorates with corruptions added in the matching cost. However, we find that more than half of the features points are still matched correctly even when heavy noise (with a Gaussian variance of 0.2) is added and no manually-set correspondences are provided.

### 3.2.3. Comparisons with state-of-the-art techniques

We run our approach in both a pairwise and joint alignment fashion and compare the performance with two state-of-the-art techniques in sense of matching sequential MSI images in a fully-automatic mode, including the quadratic programming based group registration in [30] and the pairwise robust matching in [28]. In more detail, the automatically-detected SIFT key locations are used for matching the images and the manually-specified marks are employed for validating the matching accuracy. We estimate the corresponding location in one image of a manual key location in the other image by an interpolation process which is carried out using the matched SIFT key locations. Correct-match ratios are then computed over all manual marks (with a matching distance threshold of 0.03 as explained above).

From the values of correct-match ratio on the healthy subjects and patients in Table 1 and from the displays of matching results in Fig. 4, we have at least two findings. First, joint alignment outperforms pairwise matching and at the same time the joint alignment function of the proposed SDP based scheme is superior to the one of the quadratic programming (QP) [30]. As a major reason, the proposed approach exploits a binary matrix representation (as explained in Sec. 2.2) to resolve all point correspondences jointly and near-optimally in contrast to the temporal smoothness constraint applied to the resulted matches by QP [30]. Second, more accurate matches are obtained on healthy subjects than patients for all algorithms probably because of the affects of image deterioration caused by retinal degeneration.

Table 1. Ratios of correct matches (with a distance threshold 0.03) of our approach in a pairwise way (“SDP-Pairwise”), our approach of joint alignment (“SDP-Joint”), the quadratic programming based group registration in [30] (“QP-Group”) and the pairwise robust matching in [28] (“Robust”), performed on the healthy subjects and patients, respectively.

Ratio	SDP-Pairwise	SDP-Joint	QP-Group	Robust
Healthy	0.63	0.95	0.86	0.69
Patient	0.46	0.88	0.79	0.57

We use a Dell PC with 2.39 GHz Intel Core 2 CPU and the optimization process can be finished within 120 seconds when matching simultaneously 10 MSI images for each of which about 420 landmark points are detected. In our experiments, the stoppage criterion of ADMM is set as the condition that the Frobenius norm of the difference between  $\mathbf{X}$ 's estimations in neighboring iterations exceeds  $10^{-4}$ , which costs about 350-500 iterations for a typical MSI sequence.

## 4. Discussion and future work

Misalignment of sequential images from multispectral imaging (MSI) needs to be eliminated in order for manual/automatic assessment of them in practical application of MSI. To accomplish this, image alignment (registration/matching) plays an important role by providing spatial

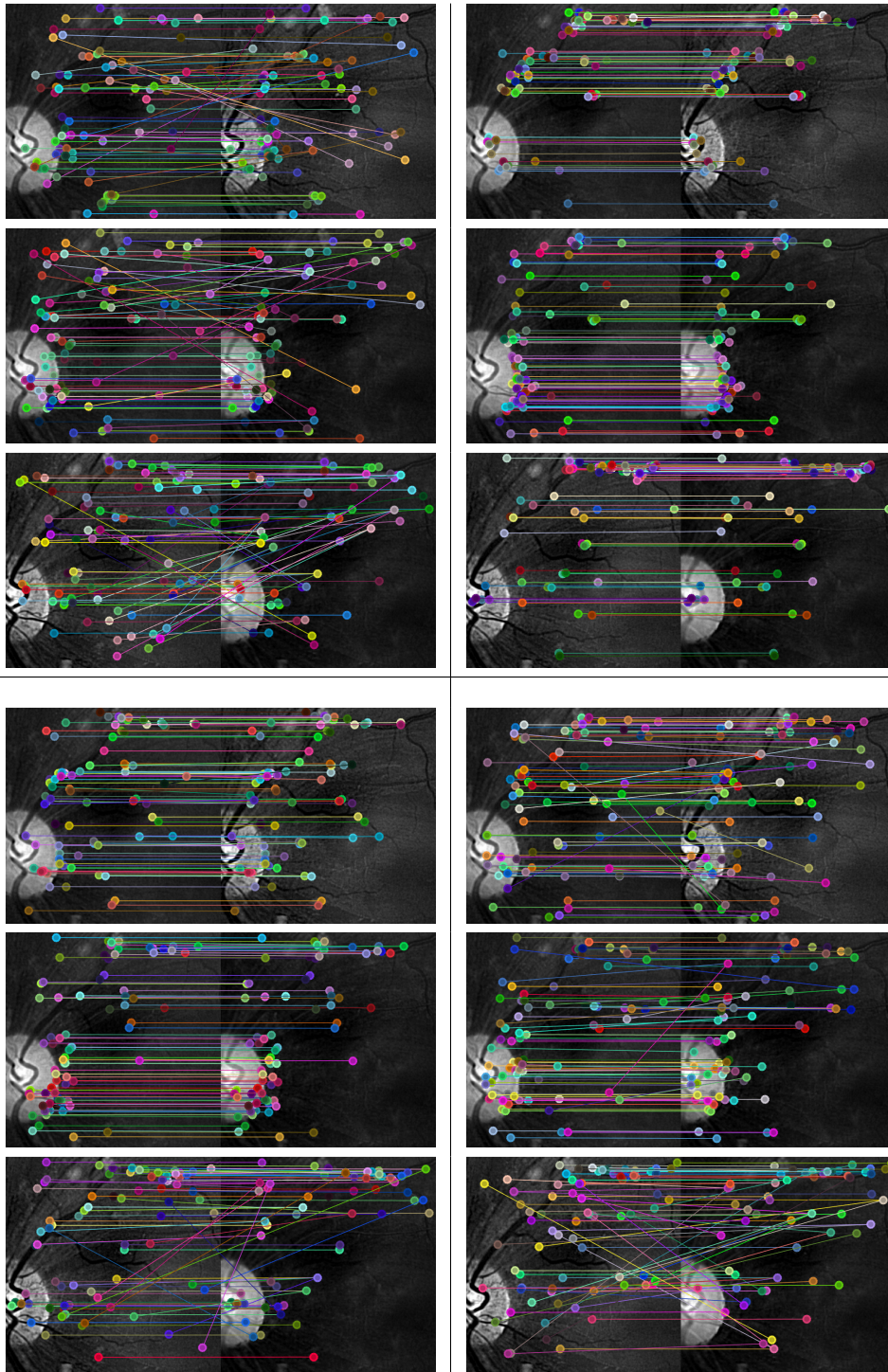


Fig. 4. Comparisons of matched feature points on the images in Fig. 1 by our algorithm in a pairwise-matching fashion (upper-left), our algorithm in a joint alignment way (upper-right), the technique in [30] (lower-left) and the one in [28] (lower-right). From top to down of each panel, a local area of the “Amber” (MSI=590), “Green” (MSI=550) and “Red” (MSI=620) images is shown, respectively.

correspondences between images. A naive approach is to compute pairwise matching across all image pairs in isolation, i.e. computing feature point correspondences between a pair of images independently from the operations on other image pairs. Many off-the-shelf algorithms can accomplish this with minor adaptations to MSI images. However, a joint alignment is capable of generating better pairwise matches by resolving all pairwise matches simultaneously, as shown by our previous work which is carried out by enforcing spatial smoothness on transformations fields between temporally neighboring image pairs [30].

In this paper, we have presented a novel approach for joint alignment of sequential MSI images by formulating the problem as a convex optimization framework. We encode pairwise image matching map via a binary matrix and store the binary matrices of all image pairs in a stacked matrix. As shown by recent work in [31], enforcing consistency on the resulted pairwise matching maps is equivalent to imposing low-rankness upon this stacked matrix. By leveraging this important finding, we express joint alignment of sequential images as a semidefinite programming (SDP) and solve a low-rank semidefinite matrix via convex optimization together with two additional processes of relaxation and rounding. This new approach recovers globally consistent feature matches by taking advantage of the aggregated information from the point-to-point matching cost. The power of our approach in contrast to state-of-the-art (e.g. our previous work in [30]) can be reflected by its four characteristics. First, pairwise matching and joint refinement of pairwise matches are solved all at once by minimizing the total feature point matching cost and at the same time exploiting the mutual relations between pairwise maps. Second, expert-specified feature point correspondences can be easily incorporated in our framework as additional equal constraints of the SDP. This allows an interactive aligning function while keeping the complete optimization process unchanged. Third, our approach performs better in resisting both the corruptions in pairwise matches and the incompleteness of pairwise maps. Finally, our main part of our algorithm is parameter free (the only parameter in our framework is  $w$  in Eq. (9)) and computationally feasible.

We have validated the proposed approach with a database of 28 sequences of MSI images of human eyes with and without retinal degenerations. Experimental results show our approach's advantages over pairwise alignments and the state-of-the-art techniques in sense of accuracy and robustness to both map's corruptions and incompleteness.

Our approach can run in both a fully-automatic way and an interactive fashion, which is invaluable in ophthalmologist's clinical practice. Fully-automatic algorithm is extremely useful in ophthalmologist's clinical practice because it can reduce labor cost and generate objective results. However, its performance may deteriorate significantly in practice especially for various difficult cases where the automatically-detected feature points become unstable. In this situation, manual intervention becomes necessary and our approach provides an easy solution for that.

In addition to MSI image alignment, our joint alignment framework can find applications in other domains, e.g. aligning MSI images used in other fields such as dentistry, astronomy and dermatology (as explained in Sec. 1), or conducting object recognition, segmentation [37], image retrieval and structure from motion [38].

We have several works to address in future. First, in this paper, we assume that each feature point of an image with fewer points detected can be matched to one point in the other image. However, it doesn't always hold in practice especially when certain contents in one image are missing in the other. We would deal with this in our future work. Second, we expect  $\mathbf{X}$  in our objective function to be of low rank, however, we don't include an explicit low rank constraint in the optimization process. Although it is proved that the enforcement of the positive-semidefinite property on  $\mathbf{X}$  (by Eq. (6)) is equivalent to a constraint of low rank [31], we think it is very interesting to investigate the performance of an explicit constraint. Third, we will also validate our approach by using other format of images, e.g. RGB vs near-infrared images, flash vs no-flash images and depth vs color images. Finally, we would like to try other matching cost measurements [39–41] or optimization

techniques [40,42,43] to get better performances.

### **Funding**

Natural Science Foundation of China (NSFC) (61572300, 61672329, 61402267); Natural Science Foundation of Shandong Province in China (ZR2014FM001, ZR2015HM026, ZR2016FQ20, ZR2014FQ004); Taishan Scholar Program of Shandong Province in China (TSHW201502038); National Institutes of Health (NIH) (P30 EY001583); Project of Humanities and Social Sciences in Universities of Shandong Province (J13WH07); Shandong Provincial Project for Science and Technology Development (2014GGX101026).