



# DSR: Direct Simultaneous Registration for Multiple 3D Images

Zhehua Mao<sup>1</sup>, Liang Zhao<sup>1</sup>(✉), Shoudong Huang<sup>1</sup>, Yiting Fan<sup>2</sup>,  
and Alex P. W. Lee<sup>3</sup>

<sup>1</sup> Robotics Institute, Faculty of Engineering and Information Technology,  
University of Technology Sydney, Ultimo, NSW 2007, Australia  
[Liang.Zhao@uts.edu.au](mailto:Liang.Zhao@uts.edu.au)

<sup>2</sup> Department of Cardiology, Shanghai Chest Hospital,  
Shanghai Jiao Tong University, Shanghai, China

<sup>3</sup> Division of Cardiology, Department of Medicine and Therapeutics, Prince of Wales  
Hospital and Laboratory of Cardiac Imaging and 3D Printing, Li Ka Shing Institute  
of Health Science, Faculty of Medicine, The Chinese University of Hong Kong,  
Hong Kong, China

**Abstract.** This paper presents a novel algorithm named Direct Simultaneous Registration (DSR) that registers a collection of 3D images in a simultaneous fashion without specifying any reference image, feature extraction and matching, or information loss or reuse. The algorithm optimizes the global poses of local image frames by maximizing the similarity between a predefined panoramic image and local images. Although we formulate the problem as a Direct Bundle Adjustment (DBA) that jointly optimizes the poses of local frames and the intensities of the panoramic image, by investigating the independence of pose estimation from the panoramic image in the solving process, DSR is proposed to solve the poses only and proved to be able to obtain the same optimal poses as DBA. The proposed method is particularly suitable for the scenarios where distinct features are not available, such as Transesophageal Echocardiography (TEE) images. DSR is evaluated by comparing it with four widely used methods via simulated and in-vivo 3D TEE images. It is shown that the proposed method outperforms these four methods in terms of accuracy and requires much fewer computational resources than the state-of-the-art accumulated pairwise estimates (APE). Codes of DSR are available at <https://github.com/ZH-Mao/DSR>.

## 1 Introduction

Image registration is a fundamental task for many medical image analysis problems where valuable information conveyed by two or more images needs to be combined and examined [14, 20]. In recent decades, mainstream medical imaging techniques, such as CT, MRI, and Ultrasound (US), have evolved from 2D to

---

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-16446-0\\_10](https://doi.org/10.1007/978-3-031-16446-0_10).

3D, which proposes new challenges to medical image registration, such as feature extraction and high computational complexity [15]. Compared to feature-based methods [13], direct (intensity-based) methods [14, 16] have occupied a dominant position in the field of medical image registration [7, 18] because of their avoidance of feature extraction and high accuracy, especially when handling images that lack distinct features, such as 3D Transesophageal Echocardiography (3D TEE) images. Direct methods estimate the frame poses by maximizing the similarity between the images. Widely used similarity metrics include sum-of-squared differences (SSD), correlation ratio (CR), and mutual information (MI) [14].

Registration of a collection of images is much more complex than pairwise registration [20]. One solution to this problem is to deduce the global poses from the results of pairwise registration [2, 11, 16, 21]. This strategy, although intuitive, is usually biased to the selected reference image and inevitably brings in accumulating errors. In comparison, a better strategy is to optimize the poses of all local frames simultaneously to avoid biases. In [9], a framework called congealing is proposed, which uses underlying entropic information of images for alignment. A large number of images are necessary for congealing because the estimation is done with the information at one location at a time [20]. And as [3] pointed out, employing entropy for congealing is problematic due to its poor optimization characteristics. Recently in [20], an accumulated pairwise estimates (APE) method is proposed for simultaneous registration. The method considers overlapping areas of images in the objective function multiple times, thus may have the information reuse issue and bring in extra complexity in the optimization.

In this paper, we propose a novel direct simultaneous registration (DSR) method which optimizes global poses of a collection of 3D images directly based on image intensity. The novelties of the paper include: 1) simultaneous registration is formulated as a direct bundle adjustment (DBA) problem, which redefines classical bundle adjustment (BA) [17] by jointly optimizing the poses of local frames and the intensities of the predefined panorama; 2) DBA uses intensity information directly instead of the extracted and matched feature points of the local 2D images in classical BA. Therefore, our method can deal with images lacking distinct features such as 3D TEE images; 3) importantly, we prove in DBA, the pose estimation is independent of the intensities of the panorama during the optimization process; 4) based on 3), we derive DSR that *only* solves the poses *without* solving the intensities of the panorama but obtaining the same poses as DBA. Simulated and in-vivo 3D TEE images are used to evaluate the proposed DSR method compared with pairwise [2], Lie normalization [21], sequential [11], and APE [20] methods. Running in a simultaneous fashion, DSR is an unbiased method that can employ all intensity information of images without information reuse, which is an elegant way to obtain the optimal poses of local frames with high accuracy.

## 2 Methodology

### 2.1 Direct Bundle Adjustment

Suppose there are  $m$  frames of 3D images taken from different viewpoints denoted as  $\mathbf{I} = \{I_1, \dots, I_i, \dots, I_m\}$ . Correspondingly, the rigid transformation for

each frame is parameterized in Lie algebra space [6] with the pose parameters  $\mathbf{x}_\xi = [\boldsymbol{\xi}_1^\top, \dots, \boldsymbol{\xi}_i^\top, \dots, \boldsymbol{\xi}_m^\top]^\top \in \mathbb{R}^{6m}$ . Simultaneous registration is the process of estimating the optimal pose parameters of all local frames  $\hat{\mathbf{x}}_\xi$  simultaneously in order to align all the images in one global coordinate frame.

Assume  $M$  is defined as a 3D panoramic image in the global frame.  $M$  consists of  $n$  voxels  $\{\mathbf{p}_1, \dots, \mathbf{p}_j, \dots, \mathbf{p}_n\}$  where  $\mathbf{p}_j \in \mathbb{R}^3$ , which fuses all the local frame images. The intensity of voxel  $\mathbf{p}_j$  in  $M$  is obtained from fusing different points' intensities in local images. Denote the intensity of  $\mathbf{p}_j$  in  $M$  and  $\mathbf{p}_j$ 's corresponding point  $\mathbf{p}_{ij}$  in local frame  $I_i$  as  $M(\mathbf{p}_j)$  and  $I_i(\mathbf{p}_{ij})$ , respectively. The intensity difference between  $M(\mathbf{p}_j)$  and  $I_i(\mathbf{p}_{ij})$  is

$$e_{ij}(\boldsymbol{\xi}_i, M(\mathbf{p}_j)) = M(\mathbf{p}_j) - I_i(\omega(\boldsymbol{\xi}_i, \mathbf{p}_j)) = M(\mathbf{p}_j) - I_i(\mathbf{p}_{ij}), \quad (1)$$

where  $\mathbf{p}_{ij} = \omega(\boldsymbol{\xi}_i, \mathbf{p}_j) = T(\boldsymbol{\xi}_i)\mathbf{p}_j$  transforms  $\mathbf{p}_j$  to  $\mathbf{p}_{ij}$ , and  $T(\cdot) \in SE(3)$  maps the pose parameters  $\boldsymbol{\xi}_i$  to a 3D Euclidean transformation. When calculating the intensity difference  $e_{ij}$  in (1), the intensity of  $\mathbf{p}_{ij}$  in  $I_i$  is obtained using trilinear interpolation to reduce the error of the intensity difference computation.

Inspired by the conventional BA framework that considers both the 3D point positions and camera poses in the optimization [1, 17], we propose the DBA framework that jointly optimizes the poses of local frames and the intensities of the panoramic image (instead of 3D point positions in BA). The overall state parameters considered in DBA are  $\mathbf{x} = [\mathbf{x}_\xi^\top, \mathbf{x}_M^\top]^\top$ , where  $\mathbf{x}_M = [M(\mathbf{p}_1), \dots, M(\mathbf{p}_j), \dots, M(\mathbf{p}_n)]^\top$  are the intensities of voxels in  $M$ . Then, we seek to obtain the optimal solution  $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_\xi^\top, \hat{\mathbf{x}}_M^\top]^\top$  that minimizes the sum-of-squared intensity differences between the panoramic image and the local images, i.e.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}_\xi, \mathbf{x}_M}{\operatorname{argmin}} \sum_{j=1}^n \sum_{i=1}^m \sigma(\mathbf{p}_{ij})(e_{ij}(\boldsymbol{\xi}_i, M(\mathbf{p}_j)))^2, \quad (2)$$

where  $\sigma(\mathbf{p}_{ij}) = 1$  if the transformed point  $\mathbf{p}_{ij}$  is within Image  $I_i$ , i.e.  $\mathbf{p}_j$  is observed in  $I_i$ , otherwise  $\sigma(\mathbf{p}_{ij}) = 0$ . Such a formulation of DBA circumvents the process of feature extraction and matching in most BA problems. Additionally, since we take the intensities of the panoramic image into account, DBA can also obtain the optimal panoramic image besides the global poses of local frames.

Gauss-Newton (GN) method is commonly used to solve nonlinear least-squares (NLLS) problems like (2). The method obtains the solution by starting with parameters initialization and then updating the parameters using the step changes calculated from GN equation in each iteration until the algorithm converges. If we write the overall observed intensity differences as a concatenation vector  $e(\mathbf{x}) = [\dots, e_{ij}, \dots]^\top$ , the objective function of (2) can be rewritten as  $f(\mathbf{x}) = e(\mathbf{x})^\top e(\mathbf{x})$ . And step changes  $\Delta\mathbf{x}$  in each iteration can be calculated from the GN equation:

$$J(\mathbf{x})^\top J(\mathbf{x})\Delta\mathbf{x} = -J(\mathbf{x})^\top e(\mathbf{x}), \quad (3)$$

where  $J(\mathbf{x})$  is the Jacobian matrix of  $e(\mathbf{x})$  w.r.t.  $\mathbf{x}$ .

Let  $J_{ij}(\mathbf{x})$  denote one row of  $J(\mathbf{x})$ , which is the gradient of one intensity difference  $e_{ij}$  w.r.t.  $\mathbf{x} = [\mathbf{x}_\xi^\top, \mathbf{x}_M^\top]^\top$ . It is shown in (1) that  $e_{ij}$  is only dependent on  $\xi_i$  and  $M(\mathbf{p}_j)$ , thus only two blocks in  $J_{ij}(\mathbf{x})$  are nonzero, i.e.

$$\frac{\partial e_{ij}(\xi_i, M(\mathbf{p}_j))}{\partial \xi_i} = -\frac{\partial I_i}{\partial \omega(\xi_i, \mathbf{p}_j)} \frac{\partial \omega(\xi_i, \mathbf{p}_j)}{\partial \xi_i}, \text{ and } \frac{\partial e_{ij}(\xi_i, M(\mathbf{p}_j))}{\partial (M(\mathbf{p}_j))} = 1,$$

which indicates  $J_{ij}(\mathbf{x})$  is very sparse.

Although the optimal poses and panoramic image can be obtained simultaneously, DBA seems more difficult to solve than traditional multi-image registration problems since a much higher order state vector is involved. However, we can further prove that the pose optimization is actually independent of the panoramic image in the GN iterations (see Sect. 2.2), which means we do not need to solve the intensities of the panoramic image but can obtain exactly the same optimal poses as solving the complete DBA. This is also an important property that conventional BA frameworks do not have.

## 2.2 Simultaneous Registration Without Intensity Optimization

**Theorem:** *When solving (2) with GN iterations, the optimization of poses is independent of the intensities of the panoramic image.*

**Proof:** If we write Jacobian matrix of  $e(\mathbf{x})$  w.r.t.  $\mathbf{x}_\xi$  and  $\mathbf{x}_M$  separately as  $J(\mathbf{x}) = [J_\xi, J_M]$ , then (3) can be rewritten as the following format:

$$\begin{bmatrix} H_{\xi\xi} & H_{\xi M} \\ H_{M\xi} & H_{MM} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_\xi \\ \Delta \mathbf{x}_M \end{bmatrix} = \begin{bmatrix} b_\xi \\ b_M \end{bmatrix}, \quad (4)$$

where  $H_{\xi\xi} = J_\xi^\top J_\xi$ ,  $H_{\xi M} = H_{M\xi}^\top = J_\xi^\top J_M$ ,  $H_{MM} = J_M^\top J_M$ ,  $b_\xi = -J_\xi^\top e(\mathbf{x})$ , and  $b_M = -J_M^\top e(\mathbf{x})$ . Then, through Schur complement [23], the step changes of poses and intensities of voxels can be computed sequentially as:

$$(H_{\xi\xi} - H_{\xi M} H_{MM}^{-1} H_{M\xi}^\top) \Delta \mathbf{x}_\xi = (b_\xi - H_{\xi M} H_{MM}^{-1} b_M), \quad (5)$$

$$H_{MM} \Delta \mathbf{x}_M = b_M - H_{M\xi}^\top \Delta \mathbf{x}_\xi. \quad (6)$$

Suppose intensity differences  $e(\mathbf{x})$  are decomposed into two components  $e(\mathbf{x}) = \mathbf{A} - \mathbf{B}$ , where  $\mathbf{A} = [\dots, M(\mathbf{p}_j), \dots]^\top$  and  $\mathbf{B} = [\dots, I_i(\mathbf{p}_{ij}), \dots]^\top$  represent observed intensities of the panoramic image and their corresponding intensities in local frames, respectively. The right side of (5) becomes:

$$\begin{aligned} b_\xi - H_{\xi M} H_{MM}^{-1} b_M &= -J_\xi^\top (\mathbf{A} - \mathbf{B}) + J_\xi^\top J_M (J_M^\top J_M)^{-1} J_M^\top (\mathbf{A} - \mathbf{B}) \\ &= -J_\xi^\top (\mathbf{A} - J_M (J_M^\top J_M)^{-1} J_M^\top \mathbf{A}) - J_\xi^\top (J_M (J_M^\top J_M)^{-1} J_M^\top \mathbf{B} - \mathbf{B}). \end{aligned} \quad (7)$$

It is shown from  $J_{ij}(\mathbf{x})$  that there is one and only one nonzero element 1 in each row of  $J_M$ . The nonzero element means the voxel  $\mathbf{p}_j$  is observed in the local frame  $i$  and corresponds to the intensity difference  $e_{ij} = M(\mathbf{p}_j) - I_i(\mathbf{p}_{ij})$ . Therefore, according to the observed status of the panoramic image in the local

frames which is indicated by the structure of  $J_M$ , it can be easily deduced that  $\mathbf{A} = J_M \mathbf{x}_M$ . Substituting  $\mathbf{A}$  to the first term on the right side of (7), we have:

$$-J_\xi^\top (J_M \mathbf{x}_M - J_M (J_M^\top J_M)^{-1} J_M^\top J_M \mathbf{x}_M) = \mathbf{0}. \quad (8)$$

Then, (5) becomes:

$$(H_{\xi\xi} - H_{\xi M} H_{MM}^{-1} H_{\xi M}^\top) \Delta \mathbf{x}_\xi = -J_\xi^\top (J_M H_{MM}^{-1} J_M^\top \mathbf{B} - \mathbf{B}), \quad (9)$$

which indicates that the step change  $\Delta \mathbf{x}_\xi$  is independent of intensities  $\mathbf{x}_M$  in every GN iteration. Therefore, obtaining the optimal poses is independent of the intensities of panoramic image  $M$  during the optimization process. **Q.E.D.**

Although  $H_{MM}$  in (9) has a huge dimension due to the large number of intensities in  $M$ ,  $H_{MM}$  is sparse and diagonal because of the sparse structure of  $J_M$ . The value of a diagonal element represents the number of times that the corresponding voxel of  $M$  has been observed in the local frames in the current iteration. So, the inverse of diagonal matrix  $H_{MM}$  can be easily computed by finding the inverse of each diagonal element, which makes solving (9) efficient.

Such an independent property of DBA is very attractive since it allows us to optimize the poses only using (9), which is equivalent to solving the complete DBA problem using (4). A 3D image typically contains millions of voxels but we only need six parameters to represent its pose. Therefore, the independence of optimizing poses to intensities can greatly help us reduce the dimension of the solution space. For distinction with DBA, we call this method DSR.

In addition, after  $\hat{\mathbf{x}}_\xi$  is obtained, the NLLS problem in (2) becomes a linear least-squares problem. Therefore, if required, we can calculate the optimal panoramic image easily in only one step from the following closed-form formula:

$$\hat{\mathbf{x}}_M = -H_{MM}^{-1} J_M^\top \mathbf{B}. \quad (10)$$

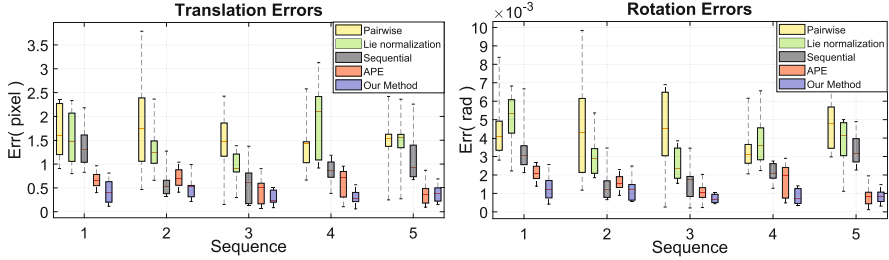
The implementation process of DSR is summarized in Algorithm 1 in the supplementary materials.

### 3 Experiments and Results

Registration of US images is usually more challenging than other modalities like CT and MRI due to the relatively low signal-to-noise ratio. Additionally, registration of a collection of 3D TEE images is especially valuable to overcome the drawback of small field of view (FoV) of 3D TEE probes. Thus, in this section, 3D TEE images are used as examples to evaluate the proposed DSR algorithm compared with the pairwise [2], Lie normalization [21], sequential [11], and APE [20] methods. Both simulated and in-vivo experiments are performed.

#### 3.1 Simulated Experiments

Five sequences of 3D images (in grayscale ranging from 0 to 255) are simulated by transforming a 3D TEE volume to a real 3D CT scan of the heart with different



**Fig. 1.** Accuracy of DSR method compared to pairwise [2], Lie normalization [21], sequential [11], and APE [20] methods using 5 sequences of simulated 3D TEE images.

poses and cropping the corresponding image area. The FoV of the simulated TEE images is the same as the 3D TEE image to get similar imagery information as real TEE volumes. Each sequence contains 11 frames of 3D images. The magnitude of these transformations varies between  $\pm 12$  degrees for rotations and  $\pm 15$  pixels for translations, which are typical ranges of poses in our obtained in-vivo 3D TEE images. Multiplicative speckle noise [10] in the original US signal can be transformed into a kind of additive noise close to Gaussian distribution in the obtained US images after logarithmic compression [10, 24]. Thus, Gaussian noise with a standard deviation of 25 is generated randomly and added to the intensities of these five sequences of images.

The accuracy of the proposed DSR method is evaluated by comparing it with pairwise, Lie normalization, sequential, and APE methods via simulated datasets. For fair comparisons, pairwise, sequential, and APE methods use the SSD as the similarity metric, GN method as optimization solver, and the initial pose parameters in each method are the same. Lie normalization optimizes poses obtained from pairwise methods and does not directly involve images [21]. Thus, we use the results from the pairwise registration as the input to Lie normalization. The mean absolute errors (MAE) of translation and Euler angles obtained from the proposed DSR and other four methods are compared in Fig. 1.

It is shown from Fig. 1 that MAE of the results obtained from DSR, sequential, and APE methods are much smaller than the pairwise and Lie normalization methods in most of the cases, which indicates the better accuracy of these three methods. Additionally, among DSR, sequential, and APE methods, the accuracy of DSR is within 0.5 pixels for translations in most of the cases and is within  $1 \times 10^{-3}$  rad for rotations in more than half cases. Although the accuracy of APE is the closest to DSR among four competing methods, it still has larger errors than DSR. The errors of APE are greater than 0.5 pixels for translations and greater than  $1 \times 10^{-3}$  rad for rotations in more than half of experiments. And the results show that the accuracy of both translations and rotations from the sequential method is lower than DSR and APE. Furthermore, it is seen from Fig. 1 that the distribution of errors from DSR is more concentrated than the others, which indicates it also has better robustness than the other four methods.

The above comparisons indicate that the proposed method has the highest accuracy, followed by APE. Both these two simultaneous registration methods

are more accurate than the other three which are deduced from pairwise registration. In addition, compared with the proposed DSR method, one apparent drawback of APE is its much higher computational complexity. Since both DSR and APE use the sum-of-squared intensity differences as the objective function, the computational complexities for both methods are closely related to the number of intensity differences. Suppose there are  $m$  images, each image has  $h$  pixels, and every two images have around  $\alpha\%$  overlapping area, the computational complexity of APE is around  $\mathcal{O}(m(m-1)/2 \times \alpha\% \times h)$  [19] since APE considers all the combinations of images, while that of DSR is only  $\mathcal{O}(m \times h)$ . In the simulated experiments for each sequence, it is found that APE calculates around four times as many intensity differences as DSR and needs around 4–5 times longer time than DSR for each iteration. Theoretically, the more images involved, the higher the computational complexity of APE is, and the more time it takes than DSR.

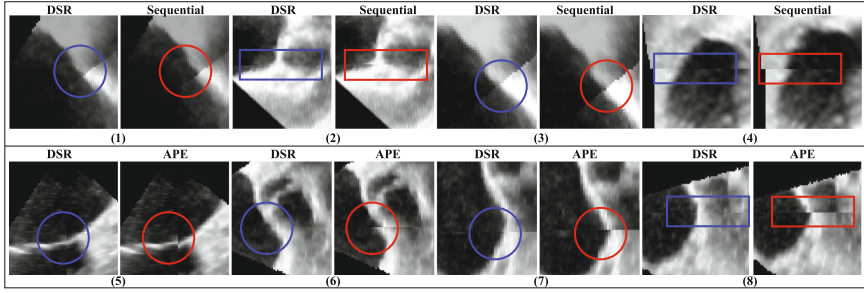
### 3.2 In-Vivo Experiments

In the in-vivo experiments, forty-six 3D TEE images from six patients (Patient #1 to #6) are collected using an iE33 ultrasound system (Philips Medical Systems) equipped with an X7-2 real-time 3D transducer. ECG-gating technique [5] is used to assist capture ECG-gated 3D TEE images so that registration of these images can be considered as rigid. The number of images in each dataset varies from 6 to 11. Each 3D TEE image contains around six million voxels. Since the proposed DSR, sequential, and APE methods outperformed the other two methods in terms of accuracy, in this section, the proposed method is compared with the sequential and APE methods only. Pose parameters of three methods are initialized using the results from the pairwise method. It is found that such an initialization method is enough for DSR to converge to the correct results.

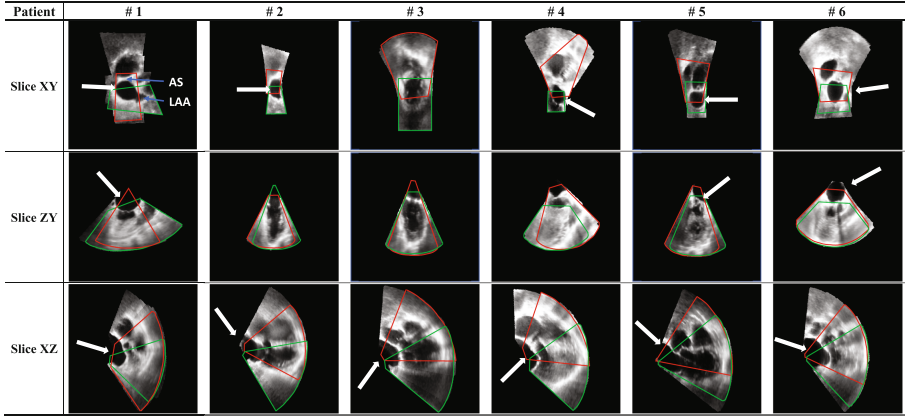
Analyzing the accuracy of a registration algorithm based on in-vivo datasets is complex because the ground-truth poses are usually not available. If images are aligned using the estimated poses, visually we can confirm that the stitching areas of the aligned images should be smooth and without misalignment if the poses are accurate. Therefore, to evaluate the accuracy of the proposed DSR method, pairwise images are aligned using the poses obtained by the sequential, APE, and DSR methods, respectively.

Similar to the results in our simulated experiments, aligned images obtained from DSR have the best quality, followed by APE and then the sequential method. In all the aligned images, there is no misalignment found from the proposed DSR in the experiments, while some apparent misalignment is found from the results of APE and sequential methods. Several examples are displayed in Fig. 2. Additionally, although APE can obtain results that are closer to those from the proposed method than the sequential method, it requires a much longer time for each iteration. In the experiments, APE usually needs around 2–5 times longer time than the proposed method for each iteration.

To further evaluate the accuracy of the proposed method, in-vivo 3D TEE images in each dataset are fused using (10) and the estimated poses from DSR. We manually select areas which contain sharp boundaries like the left atrium



**Fig. 2.** Comparisons of the aligned images using poses from sequential, APE, and DSR.



**Fig. 3.** Fused 3D TEE images using registration results from DSR for six in-vivo datasets. LA walls which have sharp structures in the images are indicated by white arrows in selected areas. Colored frames are the boundaries of two registered volumes. (Color figure online)

(LA) wall in the fused images for evaluation since generally, misalignment caused by poses with low accuracy can be easily found in these areas. The selected regions are shown in Fig. 3 with three orthogonal slices and two of the registered images in the fused images are highlighted in color boundaries. By observing the LA walls which are indicated by white arrows in Fig. 3, it is shown that the stitching areas have smooth transition and no misalignment is found in the images, which suggests good quality of alignments have been obtained by DSR.

The motivation for our current study is to enlarge the FoV of 3D TEE to assist transcatheter left atrial appendage (LAA) occlusion [4]. First, the enlarged FoV of 3D TEE allows the LAA to be observed completely (see Patient # 1 in Fig. 3) to facilitate device size selection for LAA occlusion [12]. Secondly, a complete structure of the left atrium in the enlarged 3D TEE image allows measuring the relative position and orientation of LAA w.r.t. atrial septum (AS) to facilitate the planning for LAA occlusion [4]. By counting the number of



voxels, it is found that the FoV of the fused image is enlarged to 2.18, 2.10, 2.02, 2.01, 1.80, and 2.06 times as compared with the original single TEE volume of Patient # 1 to # 6, respectively.

## 4 Conclusion

Starting from the framework of direct bundle adjustment, a novel direct simultaneous registration algorithm for 3D images is proposed in the paper. The method can optimize the poses of a collection of local images simultaneously without any information loss or reuse. Results from the simulated and in-vivo experiments demonstrate that the proposed method outperforms the other four competing methods in terms of accuracy and is more efficient than the state-of-the-art APE method. From the results of simulated experiments, it is evident that our method improved the accuracy of registration by more than 50% compared to the other four methods for most cases. In-vivo experiments also show accurate structures and extended field of view of the fused images, indicating a good quality of registration and a significant potential clinical value of the proposed method.

The proposed method can be very useful in practice when real-time performance is not required, e.g. using the enlarged FoV of 3D TEE for surgical planning of LAA occlusion. Our current focus is more on accuracy than efficiency, thus we implemented DSR in MATLAB on CPU. Since the linear system (4) has a special sparse structure that is similar to other bundle adjustment problems, it is very promising for us to use techniques in [22] to achieve parallel implementation of DSR on GPU. Additionally, optimization techniques used in g2o [8] and parallax BA [25] could also help us achieve the fast implementation. Our future work will focus on the efficient implementation of DSR.

## References

1. Alismail, H., Browning, B., Lucey, S.: Photometric bundle adjustment for vision-based SLAM. In: Lai, S.-H., Lepetit, V., Nishino, K., Sato, Y. (eds.) ACCV 2016. LNCS, vol. 10114, pp. 324–341. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-54190-7\\_20](https://doi.org/10.1007/978-3-319-54190-7_20)
2. Carminati, M.C., et al.: Reconstruction of the descending thoracic aorta by multiview compounding of 3-d transesophageal echocardiographic aortic data sets for improved examination and quantification of atheroma burden. *Ultrasound Med. Biol.* **41**(5), 1263–1276 (2015)
3. Cox, M., Sridharan, S., Lucey, S., Cohn, J.: Least-squares congealing for large numbers of images. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1949–1956. IEEE (2009)
4. Fan, Y., et al.: Device sizing guided by echocardiography-based three-dimensional printing is associated with superior outcome after percutaneous left atrial appendage occlusion. *J. Am. Soc. Echocardiogr.* **32**(6), 708–719 (2019)
5. Fenster, A., Downey, D.B., Cardinal, H.N.: Three-dimensional ultrasound imaging. *Phys. Med. Biol.* **46**(5), R67 (2001)

6. Hall, B.C.: Lie Groups, Lie Algebras, and Representations. GTM, vol. 222. Springer, Cham (2015). <https://doi.org/10.1007/978-3-319-13467-3>
7. Hill, D.L., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. *Phys. Med. Biol.* **46**(3), R1 (2001)
8. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g2o: a general framework for graph optimization. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3607–3613. IEEE (2011)
9. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(2), 236–250 (2005)
10. Loizou, C.P.: Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. *IEEE Trans. Ultrason. Ferroelect. Freq. Control* **52**(10), 1653–1669 (2005)
11. Mao, Z., Zhao, L., Huang, S., Fan, Y., Lee, A.P.W.: Direct 3d ultrasound fusion for transesophageal echocardiography. *Comput. Biol. Med.* **134**, 104502 (2021)
12. Morais, P., et al.: Semiautomatic estimation of device size for left atrial appendage occlusion in 3-d tee images. *IEEE Trans. Ultrason. Ferroelect. Freq. Control* **66**(5), 922–929 (2019)
13. Ni, D., et al.: Reconstruction of volumetric ultrasound panorama based on improved 3d sift. *Comput. Med. Imaging Graph.* **33**(7), 559–566 (2009)
14. Oliveira, F.P., Tavares, J.M.R.: Medical image registration: a review. *Comput. Methods Biomech. Biomed. Eng.* **17**(2), 73–93 (2014)
15. Schneider, R.J., et al.: Real-time image-based rigid registration of three-dimensional ultrasound. *Med. Image Anal.* **16**(2), 402–414 (2012)
16. Szeliski, R.: Image alignment and stitching a tutorial. *Found. Trends<sup>®</sup> Comput. Graph. Vision.* **2**(1), 1–104 (2006)
17. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) *IWVA 1999*. LNCS, vol. 1883, pp. 298–372. Springer, Heidelberg (2000). <https://doi.org/10.1007/3-540-44480-7-21>
18. Viergever, M.A., Maintz, J.A., Klein, S., Murphy, K., Staring, M., Pluim, J.P.: A survey of medical image registration - under review. *Med. Image Anal.* **33**, 140–144 (2016)
19. Wachinger, C., Navab, N.: Structural image representation for image registration. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 23–30. IEEE (2010)
20. Wachinger, C., Navab, N.: Simultaneous registration of multiple images: similarity metrics and efficient optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(5), 1221–1233 (2012)
21. Wachinger, C., Wein, W., Navab, N.: Three-dimensional ultrasound mosaicing. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007*. LNCS, vol. 4792, pp. 327–335. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-75759-7\\_40](https://doi.org/10.1007/978-3-540-75759-7_40)
22. Wu, C., Agarwal, S., Curless, B., Seitz, S.M.: Multicore bundle adjustment. In: *CVPR 2011*, pp. 3057–3064. IEEE (2011)
23. Zhang, F.: *The Schur complement and its applications*, vol. 4. Springer, New York (2006). <https://doi.org/10.1007/b105056>
24. Zhang, J., Cheng, Y.: *Despeckling Methods for Medical Ultrasound Images*. Springer, Singapore (2020). <https://doi.org/10.1007/978-981-15-0516-4>
25. Zhao, L., Huang, S., Sun, Y., Yan, L., Dissanayake, G.: Parallaxba: bundle adjustment using parallax angle feature parametrization. *Int. J. Robot. Res.* **34**(4–5), 493–516 (2015)