Multi-View Depth-Aware Rigid 2-D/3-D Registration

Roman Schaffert, Jian Wang, Peter Fischer, Anja Borsdorf and Andreas Maier

***Abstract*—Overlays of CT images onto X-ray images are used in minimally invasive procedures to provide navigation assistance. Image registration is usually required to align the images. Registration using a single view is often not sufficient, especially for a limited field of view. Therefore, multi-view registration is used. Recently, a depth-aware 2-D/3-D registration framework was proposed and shown to achieve high accuracy and robustness for single view registration. In this paper, we extend this method to multiple views by reformulating the point-to-plane correspondence model. We perform experiments on a publicly available cerebral angiography dataset as well as on clinical spine data. Furthermore, we investigate the effect of the viewing direction and the angle between the views to achieve optimal registration results while minimizing the obstruction to the clinical workflow. For the native 2-D images of the angiography dataset, our method achieves a success rate (SR) of 99.22 % and an accuracy of 0.21 mm, outperforming the baseline method (SR of 90.73 % mm and accuracy of 0.23 mm). For registration of single vertebrae, our method shows increased robustness (SR of up to 95.44 %) compared to a "naive" multi-view extension. We demonstrate that registration robustness reaches a plateau for 30° between views for single vertebra registration. Our evaluation also shows a dependency of registration accuracy and robustness on the actual viewing direction.**

I. Introduction

In minimally invasive interventions, live X-ray images are used to guide the physicians. To display additional information or enhance depth perception, 3-D images can be overlaid to the X-ray images. To achieve a high accuracy of the overlay, 2-D/3-D registration methods can be used. Single-view registration is desirable to save dose and not to obstruct the clinical workflow. Recently, a depth-aware 2-D/3-D registration framework using a point-to-plane correspondence (PPC) model [1] was proposed in [2], which shows a high accuracy and robustness in single-view scenarios. However, single-view registration is challenging especially in the cases of small registration targets. In these cases, multi-view registration is required to improve the registration outcome. However, the PPC model is view-dependent and cannot be directly applied to multi-view registration. In this work, we present a multi-view registration method by generalizing the PPC model to view-independent coordinate systems. By simultaneously optimizing the alignment over multiple views, the registration accuracy and robustness can be improved.

R. Schaffert (e-mail: roman.schrom.schaffert@fau.de) and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany; J. Wang, P. Fischer and A. Borsdorf are with Siemens Healthcare GmbH, 91301 Forchheim, Germany; A. Maier is also with Erlangen Graduate School in Advanced Optical Technologies (SAOT), 91058, Erlangen, Germany.

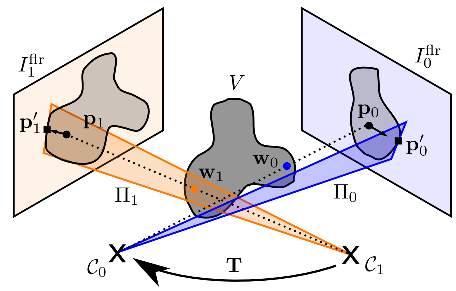

Fig. 1. Correspondences for two views. For the volume V, each view $\mathcal{C}_k$ and the corresponding image I_k^{flr}, one correspondence is shown (consisting of point $\mathbf{w}_k$, projected point $\mathbf{p}_k$, found correspondence $\mathbf{p}_k'$ and the corresponding plane Π_k). The views are related by $\mathbf{T}$.

In the clinical practice, orthogonal views are typically used for multi-view registration. For more widespread monoplanar C-arm systems, a 90° rotation of the C-arm can obstruct the ongoing workflow. Here, the question arises which view combinations can maximize the registration performance while minimizing the obstruction. In order to optimize the registration setup, we systematically evaluate the accuracy and robustness of our method for single vertebra registration depending on the used views.

II. Method

The registration framework proposed in [2] works by measuring the local misalignment between the 3-D image V and the 2-D image I^{flr}, followed by a computation of a 3-D motion using the PPC model to compensate the misalignment. The method as well as the extension to multi-view registration are described as follows.

To measure the local misalignment, 3-D apparent contour points $\mathbf{w}$ are projected onto the image plane. Additionally, gradient rendering of the 3-D image V is performed for different depth intervals $d \in \{1; D\}$, resulting in the images $\{\nabla I_d^{\text{proj}}\}$. For the projected point $\mathbf{p}$, a correspondence $\mathbf{p}'$ is searched using patch matching between the gradient images ∇I^{flr} and ∇I_d^{proj} for the depth interval d which contains $\mathbf{w}$. Based on the assumption that a motion along the contour is not observable, the correspondence is only searched perpendicular to the contour, i.e. in the direction of the 2-D gradient $\nabla I_d^{\text{proj}}(\mathbf{p})$. Given the correspondence, the motion components along the contour and the viewing ray, defined by $\mathbf{p}'$, can still not be determined. Because the 2-D contour is perpendicular to the 3-D gradient $\mathbf{g}$ and the position vector $\mathbf{w}$, the contour direction can be expressed as $\mathbf{w} \times \mathbf{g}$. In combination, the new point position $\mathbf{w} + d\mathbf{w}$ (where $d\mathbf{w}$ is the point displacement caused by the registration) is on the plane Π with the normal $\mathbf{n}$ spanned by the contour direction and the backprojection direction, i.e. $\mathbf{n} = (\mathbf{w} \times \mathbf{g}) \times \mathbf{p}'$.

The aim of the PPC model is to align the point after transformation to the plane Π. This can be expressed as the Euclidean distance $d(\Pi, \mathbf{w} + d\mathbf{w}) = 0$ or as $\mathbf{n}^\intercal(\mathbf{w}+d\mathbf{w}) = 0$. The method proposed in [2] consists of iterative search for correspondences and minimization of the distance over all correspondences over multiple resolution levels. Figure 1 shows point-to-plane correspondences for two views. In [1], the PPC model is formulated in the camera coordinate system $\mathcal{C}$ of the view, prohibiting simultaneous optimization over multiple views. In order to perform multi-view registration, the PPC model needs to be reformulated in a coordinate system $\mathcal{D}$ independent of $\mathcal{C}$. As distances are invariant to a rigid transformation T, the following holds:

$$T(d(\Pi, \mathbf{w} + d\mathbf{w})) = d(T(\Pi), T(\mathbf{w}) + T(d\mathbf{w})) = 0. \tag{1}$$

With $\mathbf{T}$ as the matrix form of T, the transformed points $\mathbf{w}_T = \mathbf{T}\mathbf{w}$, plane normals $\mathbf{n}_T = \mathbf{T}\mathbf{n}$ and the origin of $\mathcal{C}$ in $\mathcal{D}$, $\mathbf{o}_T = \mathbf{T}\mathbf{0}$, Eq. 1 is reformulated as

$$\mathbf{n}_T^\intercal(\mathbf{w}_T + T(d\mathbf{w})) - \mathbf{n}_T^\intercal\mathbf{o}_T = 0. \tag{2}$$

As $T(d\mathbf{w}) = d\mathbf{w}_T$ denotes the motion of a point $\mathbf{w}_T$ in $\mathcal{D}$, it is related directly to the motion estimate $d\tilde{\boldsymbol{\nu}}$ in $\mathcal{D}$. Using the Rodrigues rotation formula and the small angle assumption, the relationship is expressed as

$$d\mathbf{w}_T = d\tilde{\boldsymbol{\omega}} \times \mathbf{w}_T + d\tilde{\boldsymbol{\nu}}, \tag{3}$$

where $\delta\tilde{\mathbf{v}} = \begin{pmatrix} d\tilde{\boldsymbol{\omega}}^\intercal & d\tilde{\boldsymbol{\nu}}^\intercal \end{pmatrix}^\intercal$ and $d\tilde{\boldsymbol{\omega}}$ denotes the rotational and $d\tilde{\boldsymbol{\nu}}$ the translational motion components. Using Eq. 2 and Eq. 3, the relation between the correspondences and the motion in the multi-view point-to-plane correspondence model (PPC-M) can be expressed as

$$\begin{pmatrix} (\mathbf{n}_T \times \mathbf{w}_T)^\intercal & -\mathbf{n}_T^\intercal \end{pmatrix} \delta\tilde{\mathbf{v}} = \mathbf{n}_T^\intercal(\mathbf{w}_T - \mathbf{o}_T). \tag{4}$$

Combining Eq. 4 for all found correspondences, a system of linear equations $\mathbf{A}\delta\tilde{\mathbf{v}} = \mathbf{b}$ is created and solved for $\delta\tilde{\mathbf{v}}$. As $\delta\tilde{\mathbf{v}}$ is independent of the view, different views can be combined in the equation system. In this work, the first view $\mathcal{C}_0$ is used as the reference $\mathcal{D}$, meaning $\mathbf{T} = \mathbf{C}_0 C_k^{-1}$.

For each correspondence, a weight w_k according to the estimated alignment quality of the corresponding view k is assigned as

$$w_k = \min(0.1, \overline{s}_k), \tag{5}$$

where $\overline{s}_k$ is the local gradient correlation for all feature points for the current pose [2]. This is done in order to emphasize well-aligned views and prevent the registration from failing due to high in-plane misalignment in one view.

As in [2], the maximum correntropy criterion for regression (MCCR) is used to increase the robustness of the estimation. In this work, the weights w_k are additionally applied multiplicatively to all correspondences.

In addition to the proposed multi-view extension, an alternative "naive" multi-view registration scheme is used which does not depend on the reformulated PPC model. Here, single-view registration is performed for both views independently. The residual point-to-plane misalignment is used to assess the quality of the registration result and determine which view to select. To increase the registration accuracy, a refinement step is performed where only the in-plane parameters are updated for each view iteratively (PPC-I).

III. Experimental Setup

To evaluate both PPC-M and PPC-I, a publicly available "gold standard" cerebral angiography dataset [3] (ten 3-D images and corresponding 2-D image pairs) and standardized evaluation methodology [3], [4] is used. Registration is performed using native 2-D images (i.e. images with contrast agent, but no subtraction of surrounding anatomy), as this case is more challenging and suited to assess the robustness of a registration method. We compare our methods to MGP + BGB [3], which achieves the best results in [3].

To evaluate the effect of different viewing directions, varying angular distances between the views as well as the effect of the size of registered structures, experiments on a spine dataset are performed. The dataset consist of four clinical 3-D acquisitions of the abdominal region from different patients. The registration is performed between the reconstructed volume and X-ray images selected from the 3-D run. Poisson-distributed noise and a log-normalization are applied to the X-ray images to simulate fluoroscopy images. The 3-D calibration provides the projection geometry with clinically relevant accuracy. Regions of interest (ROIs) around the spine are defined in the volumes and applied during the registration. Uniformly distributed points within the ROIs are used to compute the mean target registration error (mTRE). To compare single-view with multi-view registration, standardized evaluation [3], [4] is performed on all four volumes. Four angular distances are evaluated for the two-view scenario, i.e. $30°$, $45°$, $60°$ and $90°$.

To evaluate the effect of the structure size on the registration performance, an evaluation using a single vertebra per volume is also performed. The L4 vertebra was used in all cases. An additional ROI is defined around the vertebra and only points $\mathbf{w}$ from this ROI are used during registration. We observed a highly decreased registration performance in one of the acquisitions for single-vertebra registration due to a high intestine contrast in the 2-D images. Therefore, we exclude this acquisition and perform our evaluation on three volumes.

To further investigate the effect of viewing angles, a systematic evaluation for different combinations of viewing angles of the first view α_0 and angular distances between the views $d\alpha$ is performed for the single vertebra case. For single vertebrae, single-view registration is more challenging and the effect from utilizing information from a second view is expected to be larger. α_0 is varied from $0°$ to $90°$ in steps of $10°$ and $d\alpha$ is varied from $0°$ (single-view registration) to $90°$ in steps of $15°$. Instead of evaluating on one start position as in [5], we

TABLE I
EVALUATION RESULTS FOR THE CEREBRAL ANGIOGRAPHY DATASET [3] USING NATIVE 2-D IMAGES FOR THE MULTI-VIEW SCENARIO.

Method	mTRE (mean±std.) [mm]	SR [%]	CR [mm]
PPC-M	**0.21**±0.06	99.22	19
PPC-I	0.23±0.07	**99.68**	**20**
MGP + BGB [3]	0.23±0.10	90.73	12

TABLE II
EVALUATION RESULTS FOR THE SPINE DATASET FOR DIFFERENT VIEW COMBINATIONS (ANGLES ARE RELATIVE TO AP). RESULTS FOR MTRE INCLUDE MEAN AND STANDARD DEVIATION.

Views	Method	mTRE [mm]	SR [%]	CR [mm]
AP	PPC	0.89±0.45	87.67	7
LAT		0.96±0.24	90.58	7
AP + LAT	PPC-M	**0.79**±0.20	97.75	25
	PPC-I	1.00±0.23	**99.63**	**30**
AP + 60°	PPC-M	**0.87**±0.22	**96.75**	21
	PPC-I	0.89±0.24	95.92	**22**
AP + 45°	PPC-M	**0.96**±0.25	**97.63**	**25**
	PPC-I	0.99±0.26	95.92	18
AP + 30°	PPC-M	**0.70**±0.26	95.38	**18**
	PPC-I	0.77±0.24	**97.29**	**18**

use 90 random start positions with initial mTREs uniformly distributed within 0 mm - 30 mm. This allows the evaluation on robustness, i.e. the success rate (SR), of the registration for each view combination. To evaluate the precision of the method, the average standard deviation is computed over all used acquisitions.

IV. RESULTS

A. Cerebral Angiography Data

As can be seen in Tab. I, both PPC-M and PPC-I outperform the MGP + BGB [3] method on the angiography dataset [3]. While the capture range (CR) and the SR are slightly larger for the PPC-I method, the mean target registration error (mTRE) is slightly smaller for PPC-M. Overall, PPC-M and PPC-I perform similarly on the angiography dataset. Compared to the single-view scenario, where the depth-aware registration framework achieves a mTRE of around 0.7 mm and a CR of 11 mm - 13 mm (depending on the used view) [2], the performance is improved considerably.

TABLE III
EVALUATION RESULTS FOR SINGLE VERTEBRA REGISTRATION FOR DIFFERENT VIEW COMBINATIONS (ANGLES ARE RELATIVE TO AP). RESULTS FOR MTRE INCLUDE MEAN AND STANDARD DEVIATION. ONE ACQUISITION EXCLUDED DUE TO HIGH INTESTINE CONTRAST.

Views	Method	mTRE [mm]	SR [%]	CR [mm]
AP	PPC	1.35±0.34	26.61	0
LAT		1.05±0.24	71.17	6
AP + LAT	PPC-M	**0.90**±0.12	**94.00**	**16**
	PPC-I	1.09±0.18	89.67	15
AP + 60°	PPC-M	**1.01**±0.17	**92.06**	**16**
	PPC-I	1.15±0.23	87.44	**16**
AP + 45°	PPC-M	**1.12**±0.21	**95.44**	**18**
	PPC-I	1.15±0.24	91.22	16
AP + 30°	PPC-M	**0.90**±0.19	**89.61**	**10**
	PPC-I	0.97±0.20	84.78	7

B. Spine Data

For the spine data, an increased robustness is observed when using a multi-view setup (see Tab. II). Using either only the AP or only the LAT view, a SR of 87.67 % and 90.58 % is achieved. The SR increases to roughly 94 % - 99 % for multi-view registration, depending on the used method and views. While the highest SR is achieved for PPC-I and 90° between the views, both PPC-I and PPC-M lead to comparable SR values, where the difference is within 2 % for all cases. The CR is increased from 7 mm for single-view registration to 18 mm - 30 mm for multi-view registration. While the CR varies considerably for the multi-view scenario, no clear tendency towards better results for higher angles is observable. Compared over all cases, PPC-M and PPC-I lead to similar robustness. However, the PPC-M method leads to consistently higher accuracy. While the highest difference in mTRE of 0.21 mm is observed for 90° between the views, the smallest difference of 0.02 mm is observed for the 60° case.

C. Single Vertebra Registration

The results for the single vertebra evaluation are shown in Fig. III. In general, a decreased robustness and accuracy can be observed compared to the previous experiment. The decrease in robustness is strongest in the single-view registration, where the SR drops to 71.17 % for the lateral view and even to 26.61 % for the AP view. Contrary to the results using all visible vertebrae, the PCC-M method leads to a higher robustness compared to PCC-I for all cases. While the highest SR and CR for PPC-M are 95.44 % and 18 mm, the highest SR and CR for PPC-I are 91.22 % and 16 mm, both for the 60° case. Similar to the evaluation using the whole visible spine, the PPC-M method achieves a higher accuracy compared to PPC-I. The mTRE for PPC-M compared to PPC-I is smaller with a maximum difference of 0.19 mm for 90° between the views and a minimum difference of of 0.03 mm for the 60° case. Again, no correspondence between the angle between the views and registration performance can be observed.

D. Systematic Viewing Angle Evaluation

Results for the systematic viewing angle evaluation are shown in Fig. 2.

Regarding the robustness (see Fig. 2c), we observe that different α_0 lead to large differences in the success rate, especially for $d\alpha = 0°$ (9.25 % - 73.33 %) and $d\alpha = 15°$ (51.48 % - 96.67 %). Averaging the SR over all α_0 (see Fig. 2e), we observe an increase of the SR when increasing $d\alpha$ from 0° (mean SR of 49.41 %) to 30° (mean SR of 95.56 %). For larger $d\alpha$, the SR does not change considerably. This shows that relatively small angles suffice to achieve a high robustness and no further improvement can be achieved by increasing $d\alpha$.

For the mTRE, no general improvement for large angles can be observed (see Fig. 2a). While the highest average error is observed for $d\alpha = 0°$ (1.27 mm), $d\alpha = 90°$ leads to the second largest average error of 1.17 mm. Regions of higher error (e.g. around $\alpha_0 = 40°$ and $d\alpha = 90°$) and lower error

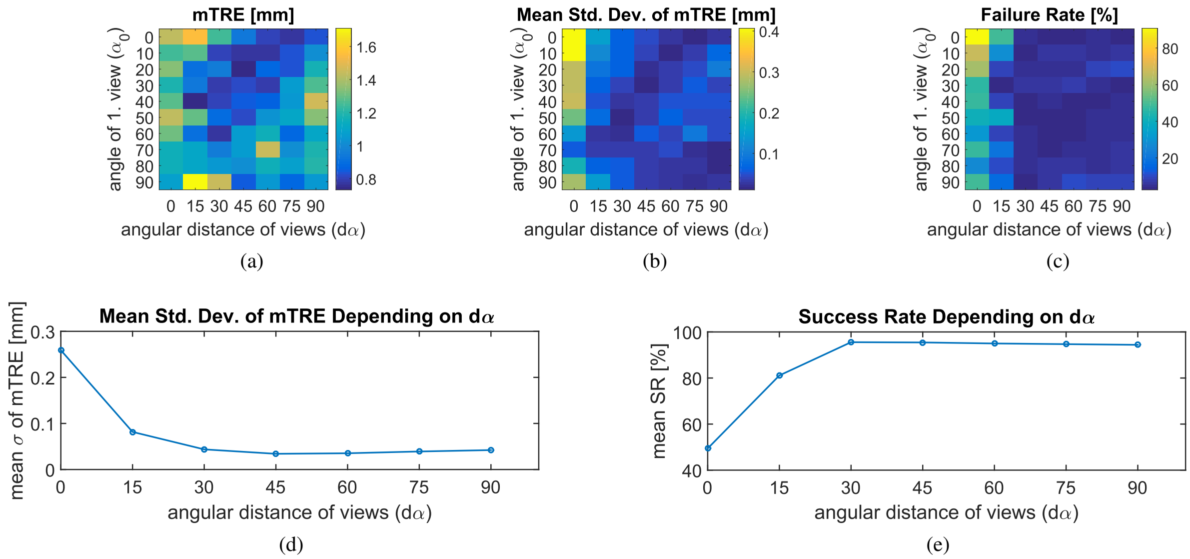

Fig. 2. Registration results: a) mTRE , b) mean standard deviation of the mTRE, c) failure rate (1 - success rate) for different combinations of viewing angle of the first view α_0 and the angular distance between views $d\alpha$, d) mean standard deviation of the mTRE averaged over all α_0 and e) mean success rate averaged over all α_0.

(e.g. around $\alpha_0 = 20°$ and $d\alpha = 45°$) exist, i.e. the accuracy depends on the actual used views and not primary on $d\alpha$.

As calibration errors are present, absolute errors have to be considered with care. Therefore, we also evaluate the standard deviation of the mTRE (which indicates the precision of the method). Similar to the mTRE, regions of higher and lower precision exist (see Fig. 2b), indicating a dependence on the used views. The deviation for the single-view registration is high (mean of 0.26 mm over all cases with $d\alpha = 0$) compared to the multi-view cases (mean of 0.046 mm over all cases with $d\alpha \neq 0$). A high precision of 0.043 mm is reached for $d\alpha = 30°$ and varies only slightly (0.034 mm - 0.042 mm) for higher angles between the views (see Fig. 2d). Again, no trend towards better performance for high $d\alpha > 30°$ is observed.

Overall, our results show that the registration performance is dependent on the actual viewing angles and that a relatively small $d\alpha \geq 30°$ is sufficient to achieve optimal performance for the evaluated data. A similar effect for the accuracy of a registration on spine data is demonstrated in [5], indicating that relatively small angles are sufficient for optimal registration performance regardless of the used method.

V. Discussion and Conclusion

The results show that the PPC-M method is capable of achieving high accuracy and robustness. On the used “gold standard” dataset [3], it outperforms the (MGP + BGB) method, which leads to the best results in [3]. Compared to the PPC-I method, it consistently achieves a lower mTRE. The robustness is similar for PPC-M and PPC-I for large structures. However, the PPC-M method outperforms PPC-I for the more challenging single-vertebra registration.

The systematic viewing angle evaluation shows that an angle of 30° between views is sufficient to achieve good registration results and larger angles do not improve the performance. Moreover, the registration performance depends on the actual viewing direction, especially for small angles between the views or single-view registration.

In the future, we plan to further investigate the dependency of the registration on the viewing direction. Another direction is the registration of small structures other than single vertebrae, e.g. instruments used during an intervention.

In this paper, we present a depth-aware multi-view registration method by extending the method in [2] and by evaluating our method on a publicly available “gold standard” dataset [3] we show that our method achieves accurate and robust registration. We furthermore show that robust registration can be achieved with our method for relatively small angles of 30° between the views, and large angles do not improve the results for the tested data. We furthermore demonstrate a dependency of the registration performance on the actual viewing angles.

References

[1] J. Wang, A. Borsdorf, B. Heigl, T. Köhler, and J. Hornegger, “Gradient-based differential approach for 3-D motion compensation in interventional 2-D/3-D image fusion,” in *2014 2nd International Conference on 3D Vision*, vol. 1, Dec 2014, pp. 293–300.

[2] J. Wang, R. Schaffert, A. Borsdorf, B. Heigl, X. Huang, J. Hornegger, and A. Maier, “Dynamic 2-D/3-D rigid registration framework using point-to-plane correspondence model,” *IEEE Trans Med Imaging*, vol. 36, no. 9, pp. 1939–1954, Sept 2017.

[3] U. Mitrovic, Z. Spiclin, B. Likar, and F. Pernus, “3D-2D registration of cerebral angiograms: A method and evaluation on clinical images,” *IEEE Trans Med Imaging*, vol. 32, no. 8, pp. 1550–1563, Aug 2013.

[4] E. B. van de Kraats, G. P. Penney, D. Tomazevic, T. van Walsum, and W. J. Niessen, “Standardized evaluation methodology for 2-D-3-D registration,” *IEEE Trans Med Imaging*, vol. 24, no. 9, pp. 1177–1189, Sept 2005.

[5] A. Uneri, Y. Otake, A. S. Wang, G. Kleinszig, S. Vogt, A. J. Khanna, and J. H. Siewerdsen, “3D-2D registration for surgical guidance: Effect of projection view angles on registration accuracy,” *Phys. Med. Biol.*, vol. 59, no. 2, pp. 271–287, Nov 2014.