Depth-Aware Template Tracking for Robust Patient Motion Compensation for Interventional 2-D/3-D Image Fusion

Jian Wang, *Student Member, IEEE,* Anja Borsdorf, Jürgen Endres, and Joachim Hornegger *Member, IEEE*

***Abstract*—In interventional radiology, pre-operative three dimensional (3-D) images (e.g. CT and MRI) can be fused to the real-time 2-D fluoroscopic images, known as 2-D/3-D image fusion or overlay. 2-D/3-D registration is usually performed at the beginning to ensure an accurate overlay. However, the misalignment caused by patient motion has to be corrected during the procedure. In this paper, we propose an adapted template-based tracking approach that fits well into the novel depth-layer-based framework, where 3-D motion can be estimated by depth-aware tracking in 2-D X-ray images. Templates are generated according to the 2-D/3-D matching procedure, where the 2-D features are matched with the depth layer images generated from the 3-D volume. Our template update strategy takes both initial and current frame into account, which effectively enhances the accuracy of the template tracking. Two strategies are applied to enhance robustness against external disturbances: template lock and motion estimation feedback. The experiment results show that our new approach is capable of estimating 3-D motion and much more robust against external disturbances compared to the Kanade-Lucas-Tomasi tracker that is previously used in the depth-layer-based framework.**

***Index Terms*—2D/3D image fusion, motion compensation, depth-aware tracking, template tracking.**

I. Introduction

In INTERVENTIONAL radiology, 2-D fluoroscopic images are taken in real-time by the interventional C-arm system. Fluoroscopy provides guidance information of the intravascular devices, such as catheter and guide wire. Nowadays, pre-operative 3-D images (e.g. CT and MR) can be fused to the 2-D fluoroscopic view, known as 2-D/3-D image fusion or overlay. The image fusion provides 1) additional information of structures that are not directly visible in fluoroscopic view (e.g. 3-D road-mapping [1]) and 2) planning information in the 3-D image.

The accuracy of image fusion is crucial for interventional procedures. The starting accuracy is usually ensured by performing the initial 2-D/3-D registration. A review of the state-of-the-art 2-D/3-D registration methods for interventional guidance is found in [2]. However, the misalignment caused by patient motion needs to be corrected during the procedure. Cardiac and breathing motion are mostly discussed in the literature, e.g. [3] and [4]. Most of the algorithms are model-based due to the periodic characteristics of the motion. However, the patient movement leads to both inaccuracy in the overlay and failure of the model-based methods. Therefore, the general patient movement needs to be corrected on the fly. In state-of-the-art applications, patient movement is usually detected by clinicians and the correction is triggered via user interaction [5]. Automatic motion compensation is challenging in this scenario. Firstly, no certain pattern is available for the patient movement. Secondly, standard 2-D/3-D registration algorithms are not real-time capable. Moreover, the external disturbances caused by interventional devices or contrast media injection are significant challenges to the robustness of the framework.

Recently, a novel depth-layer-based patient motion compensation framework has been proposed [5], where 3-D rigid motion can be estimated directly from the depth-aware 2-D tracking by introducing the concept of depth layer (see Section II-A). This tracking-based approach has the following advantages: 1) since no frame-by-frame iterative computation of digitally reconstructed radiograph (DRR) is involved, it is computationally efficient; 2) 3-D rigid motion can be automatically detected by the tracker and estimated from the depth-aware 2-D correspondences from the single-view fluoroscopic image sequences.

Finding the 2-D correspondences over frames is one of the most important steps in the depth-aware tracking framework. In [5], standard Kanade-Lucas-Tomasi (KLT) tracker [6] was employed to find 2-D correspondences between neighboring frames. However, the standard KLT tracker is not optimal for the framework. On the one hand, KLT tracker detects feature points independently from 2-D/3-D matching step (see Section II-A). On the other hand, optical flow is used in KLT tracker, so the tracking suffers from the severe "drifting" problem in the condition under external disturbances, such as contrast media injection (see Section III).

Template tracking is a popular topic in computer vision. It has been used for post estimation in [7], where the 3-D model of the object is known. In this paper, an adapted template tracking approach is presented. In the new approach, templates are extracted from the intermediate matching results in the depth-layer-based framework. The depth-aware templates are tracked along frames for 3-D motion estimation. Two strategies are employed to enhance the robustness of motion compensation against external disturbances: template lock and feedback from motion estimation.

J. Wang (e-mail: jian.wang@cs.fau.de), J. Endres and J. Hornegger are with the Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany; A. Borsdorf is with Siemens AG, Healthcare Sector, 91301 Forchheim, Germany; J. Hornegger is also with Erlangen Graduate School in Advanced Optical Technologies (SAOT), 91058, Erlangen, Germany.

978-1-4799-0534-8/13/$31.00 ©2013 IEEE

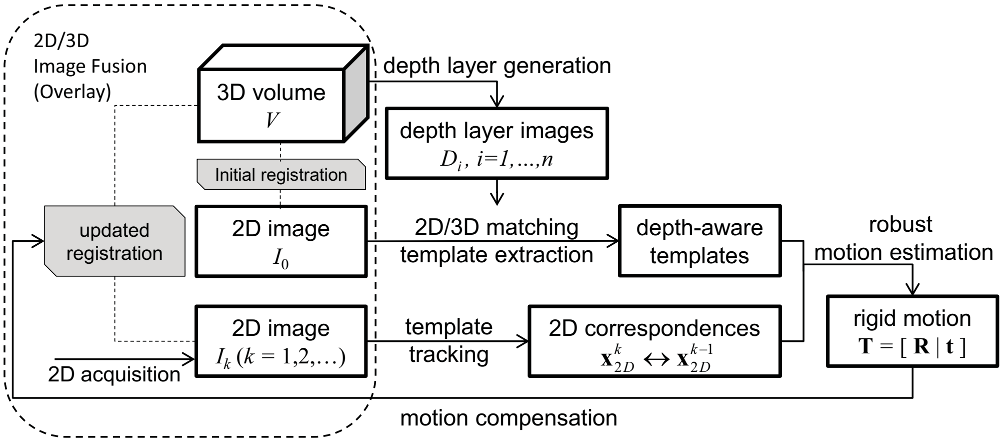


Fig. 1: Depth-aware tracking framework.

II. Methods

A. The Depth-Layer-Based Motion Compensation Framework

The key idea of the depth-layer-based framework in [5] is to recover depth information of 2-D features using depth layers generated from 3-D image. Depth-aware 2-D tracking is then performed for motion estimation. It consists of the following steps:

1) Depth layer generation: given the projection geometry of the initial registration, the 3-D volume V can be divided into a stack of sub-volumes $\{V_i, i = 1, ..., N\}$ along the viewing direction. The sub-volumes have the corresponding center depths $\{d_i, i = 1, ..., N\}$ away from the X-ray source along the viewing direction, where N is the number of depth layers. Instead of performing the whole volume rendering, depth layers $\{D_i, i = 1, ..., N\}$ are generated by rendering the sub-volumes separately. Non-photorealistic volume rendering technique [8] is applied: occluding contours are generated by gradient-based volume rendering, which usually corresponds to the high gradient region in the projection image; windowing is applied for bone structures, because bones usually represent well the patient movement;

2) 2-D/3-D matching: the matching is performed patch-wise between the gradient map of the initial frame $|\nabla I_0|$ (2-D) and each depth layer D_i (3-D), where a 2-D image is divided into K patches. For each patch location k ($k = 1, ..., K$), similarity (e.g. normalized cross correlation (NCC)) between the patch pair $\left(p^k_{|\nabla I_0|}, p^k_{D_i}\right)$ is measured. It is used as the matching weight w^k_i between the k-th 2-D image patch and a certain depth d_i of the corresponding layer D_i;

3) $2D+$ reconstruction: thresholding is performed to select the 2-D patches with high matching weights (w.r.t. the threshold w_δ). Given the corresponding depth d_i, the 2-D patch locations are back-projected to 3-D space by using the projection geometry of the C-arm system [5];

4) Depth-aware 2-D tracking and motion estimation: 2-D tracking is performed over time to acquire 2-D correspondences from neighboring frames. 3-D rigid motion is then estimated using the tracking results and the reconstruction from steps 3. The 3-D motion is applied to compensate the motion in the fused image.

B. Depth-Aware Templates Tracking for Motion Compensation

Finding 2-D correspondences in step 4) is crucial for motion estimation. However, the standard KLT tracker [6] detects and tracks feature points independently from the 2-D/3-D matching results in step 2) in Section II-A. Therefore, we improve the tracking procedure by adapted template tracking (Fig. 1). Although template tracking in optical photographs for pose estimation is not new, we have established a new way of fitting this technique into the novel depth-layer-based motion compensation framework, such that robust 3-D motion estimation in the 2-D/3-D image fusion is achieved. The adapted template tracking consists of the following steps:

1) Depth-aware template extraction: after the 2-D/3-D matching procedure (Section II-A), a set of image patches $\left\{p^k_{I_0}, k \in [1, K]\right\}$ is mapped to a set of matching weights $\left\{w^k_i, i \in [1, N] \text{ and } k \in [1, K]\right\}$. Template extraction is performed through all depths $\{d_i\}$. For each d_i, the connected patches with matching weights $w^k_i > w_\delta$ are clustered into one group. An 8-neighbor region growing segmentation is performed for clustering. A group of the patches contains more structural information as a individual patch. Rectangular templates $\{\mathcal{T}^j_i, j = 1, ..., J_i\}$ are created from the bounding boxes of the patch groups, where J_i is the number of templates corresponding to the i-th depth interval (Fig.2). The center of each template is considered its 2-D position. Then, the 3-D position of each template $\mathbf{x}_{2D+}(\mathcal{T}^j_i)$ is calculated using the depth d_i.

2) Template tracking: the extracted image regions are tracked in the next frames by template matching using fast normalized cross-correlation [9]. Instead of maintaining one set of templates $\{\mathcal{T}^j_i\}$ from the initial frame, another set of templates $\{\mathcal{T}'^j_i\}$ is also maintained by updating from the tracking results in every frame. The NCCs ($\mathrm{NCC}(\mathcal{T}^j_i)$ and $\mathrm{NCC}(\mathcal{T}'^j_i)$) are calculated in each frame using the two sets of templates in the searching windows (neighborhood regions of previous targets). The target position is the one that maximizes

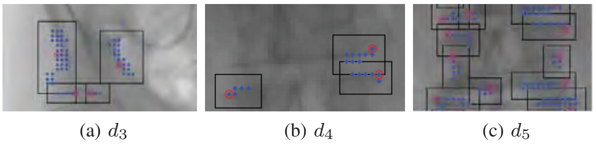

(a) d_3 (b) d_4 (c) d_5

Fig. 2: Examples of extracted templates of different depths d_i. Blue points indicate the image patches with high matching weights w_i^k, red points indicate the seed points for region growing. (See Section II-B)

$\exp\left(\text{NCC}(\mathcal{T}_i^j) + \text{NCC}(\mathcal{T}_i^{\prime j})\right)$. This avoids losing the target due to the accumulated appearance change after several frames and the "drifting" problem by updating the templates in every frame.

3) Robust motion estimation: the tracking procedure gives the new 2-D position of each tracked template $\mathbf{x}'_{2D}(\mathcal{T}_i^j)$ (center of the rectangle). The maximum depth matching weight of the sub-patches belonging to each template group is used as the matching weight of the template $w(\mathcal{T}_i^j)$. Rigid 3-D motion $\hat{\mathbf{T}} \in \mathbb{R}^{3\times 4}$ is estimated by non-linear optimization

$$\hat{\mathbf{T}} = \underset{\mathbf{T}}{argmin}\left(\sum_{(\mathcal{T}_i^j)} w \cdot \|\mathbf{x}'_{2D} - \mathbf{K}\cdot\mathbf{T}\cdot\mathbf{x}_{2D+}\|\right), \qquad (1)$$

where $\mathbf{K} \in \mathbb{R}^{3\times 3}$ is the camera matrix [5]. RANdom SAmple Consensus (RANSAC) is used to neglect outliers.

4) Strategies against external disturbances: during the interventional procedures, disturbances from external devices (e.g. catheter motion and contrast media injection) can affect the tracking results of some individual templates. However, the disturbances affect only limited regions in the fluoroscopic image in most of the cases. By neglecting the affected templates, motion can still be estimated using the remaining templates. Two strategies are applied to maintain the robustness against external disturbances:

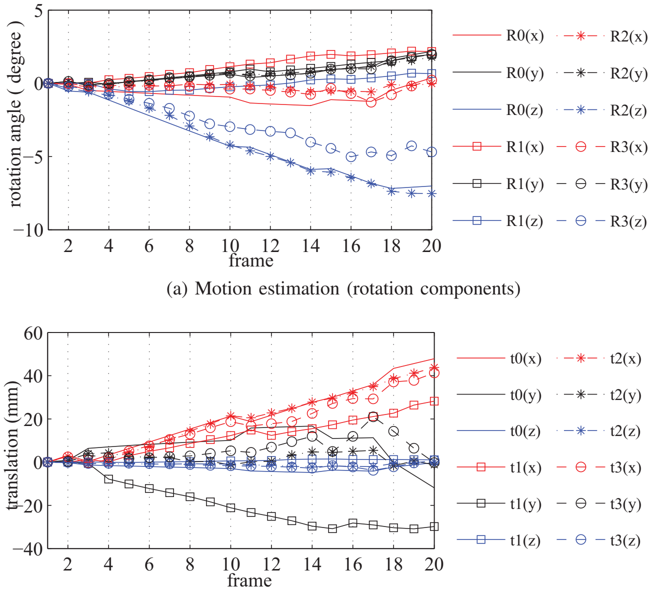
Template lock: template tracking can suffer from a significant large offset due to external disturbances, which can be detected by measuring the target offsets between neighboring frames. Since tracking is performed in every frame, valid tracking results give only small offsets. When extreme offsets detected (e.g. 5 times greater than the mean value), the corresponding templates are treated as outliers for motion estimation and the templates are "locked" from updating in the current frame.

Motion feedback: the estimated 3-D motion is applied to the $2D+$ model of the "locked" templates to predict the new 2-D positions of search windows in the next frame. This allows to find the targets back when external disturbances are out of the region, i.e. catheter moved away or contrast media washed out.

III. Experiments & Results

A. Experiments Setup

A sequence of rigid 3-D motion $\{\mathbf{T}_{GT}\}$ is generated as ground truth, which is then applied to a clinical CT dataset (with 512^3 voxels) to generate a sequence of simulated X-ray images (DRRs). Projection parameters of a commercially available C-arm system is used for DRR generation. Using DRRs is an established way of algorithm evaluation [10][5]. In this case, arbitrary motions can be tested in known conditions. For the purpose of robustness test, a virtually generated image sequence showing contrast media (CM) injection based on computational fluid dynamics (CFD) is added to the image sequence (Fig. 4a). Our CFD-based blood flow simulation is proved clinically valid in [11]. CM is a more challenging task for robustness test compared to catheter motion. 10 depth layers are generated from the 3-D volume. High depth resolution is not applied here, which can lead to severe truncation of structural information in the templates.



(a) Motion estimation (rotation components)

(b) Motion estimation (translation components)

Fig. 3: Estimated motions are decomposed into rotation angles (a) and translation components (b) over 20 frames (three colors): four series (0 - 3) of curves (different markers) show results of respectively: 0) the ground truth, 1) the standard KLT-tracking approach in [5] with contrast media (CM), 2) the new approach without and 3) with CM.

B. Quantitative Results of Motion Estimation

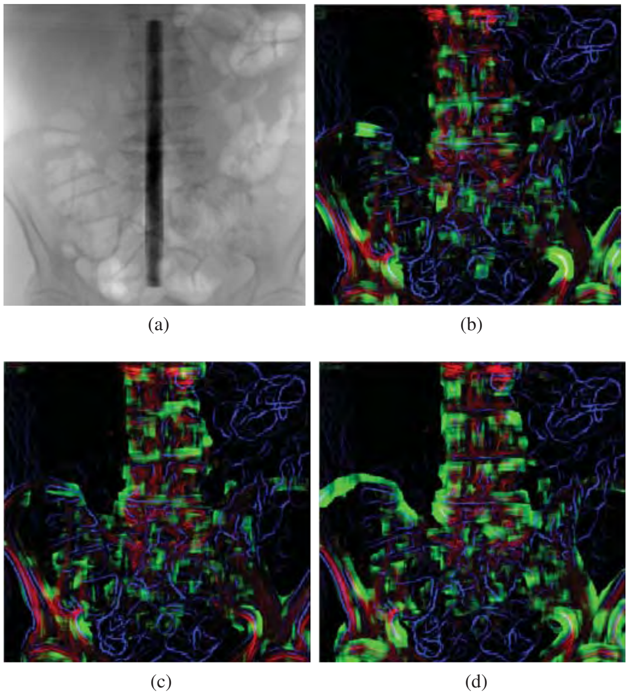
3-D rigid motion $\mathbf{T}$ is decomposed into rotation components (in degree) and translation components (in mm). In Fig.3, the motion estimation results along the frames are shown for the following cases: 1) ground truth (R0 and t0); 2) the standard KLT-based approach ("*K-approach*" later on) with CM (R1 and t1); 3) the template-based approach ("*T-approach*" later on) without CM (R2 and t2); 4) the T-approach with CM (R3 and t3).

In our setup, the rotation around y-axis and translation along $x-$ and $z-$axes are in-plane motions. Without CM, the T-approach was able to estimate the 3-D motion well with average errors of (0.57, 0.10, 0.23)° in rotation and

(1.48, 7.32, 1.52) mm in translation. Under the condition of CM, the K-approach only gave reasonable in-plane motion components and failed in off-plane motion components, which gave average errors of (1.91, 0.18, 4.03)° and (9.0, 26.6, 3.2) mm. Compared to the K-approach, the T-approach managed to keep the off-plane motion in a reasonable range and estimated better the in-plane motion with average errors of (0.48, 0.06, 1.28)° and (3.3, 6.6, 1.2) mm.

C. *Qualitative Results using Similarity Maps*

In order to show the motion compensation results qualitatively, we use similarity map (Fig.4) to give a visual impression of the overlay accuracy. It shows the similarity response (green) between gradient magnitude of the X-ray image (blue) and the gradient-based rendering of the 3-D volume (red), where NCC is used as the similarity measure. For the purpose of visualization, the 2-D gradient magnitude is calculated from native X-ray images without CM. Fig.4 shows the similarity maps at frame 20 of respectively the cases without motion correction (4b), with motion correction using the K-approach (4c) and using the T-approach(4d). The motion in X-ray mainly causes large offsets of the vertebra column (Fig.4b). The K-approach managed to maintain only a few regions aligned (high similarity response in green). However, the new T-approach managed to maintain a globally high alignment of the overlay, even after 19 frames of continuous motion with external disturbance (CM).



(a) (b) (c) (d)

Fig. 4: (a) shows an example frame with CM; (b), (c) and (d) show the similarity maps of respectively no motion correction, KLT-based and template-based motion correction at frame 20 (with CM).

IV. Discussion & Conclusion

The results show that the adapted template tracking approach achieves higher accuracy in 3-D motion estimation. As expected, the in-plane motion is better estimated than off-plane motion. The robustness against external disturbances is enhanced by combining the template lock and motion feedback.

As future extensions, depth correction and improved template tracking approach will be considered. The extracted templates are supposed to contain reasonable structural information, sparse depth sampling is preferred but introduces significant systematic error. A depth correction strategy in this scenario is discussed in [12]. Extended template matching is needed to achieve scale and rotation invariance, such that the template tracking is more robust to different motions.

In this paper, we present an adapted template tracking approach that fits well in the novel depth-layer-based framework for patient motion compensation [5]. The approach is evaluated under the condition of external disturbances (simulated contrast media injection). The results show that 3-D rigid motion estimation is more accurate and robust against external disturbance using the improved approach.

References

[1] S. Rossitti and M. Pfister, “3d road-mapping in the endovascular treatment of cerebral aneurysms and arteriovenous malformations,” *Interventional Neuroradiology*, vol. 15, no. 3, p. 283, 2009.

[2] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, “A review of 3d/2d registration methods for image-guided interventions,” *Med Image Anal*, vol. 16, no. 3, pp. 642 – 661, April 2012.

[3] J. Hadida, C. Desrosiers, and L. Duong, “Stochastic 3d motion compensation of coronary arteries from monoplane angiograms,” in *Proceedings of the 15th international conference on Medical Image Computing and Computer-Assisted Intervention, Part I*, 2012, pp. 651–658.

[4] A. Brost, R. Liao, N. Strobel, and J. Hornegger, “Respiratory motion compensation by model-based catheter tracking during ep procedures,” *Medical Image Analysis*, vol. 14, no. 5, pp. 695–706, 2010.

[5] J. Wang, A. Borsdorf, and J. Hornegger, “Depth-layer based patient motion compensation for the overlay of 3d volumes onto x-ray sequences,” in *Proceedings Bildverarbeitung für die Medizin*, 2013, pp. 128–133.

[6] T. Carlo and T. Kanade, “Detection and tracking of point features,” 1991.

[7] S. Ravela, B. Draper, J. Lim, and R. Weiss, “Adaptive tracking and model registration across distinct aspects,” in *Proceedings of 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1. IEEE, 1995, pp. 174–180.

[8] B. Csébfalvi, L. Mroz, H. Hauser, A. König, and E. Gröller, “Fast visualization of object contours by non-photorealistic volume rendering,” in *Computer Graphics Forum*, vol. 20, no. 3. Wiley Online Library, 2001, pp. 452–460.

[9] J. Lewis, “Fast normalized cross-correlation,” in *Vision interface*, vol. 10, no. 1, 1995, pp. 120–123.

[10] W. Wein and A. Ladikos, “Detecting patient motion in projection space for cone-beam computed tomography,” in *MICCAI 2011 Proceedings*, ser. Lecture Notes in Computer Science. Springer, Sep. 2011.

[11] J. Endres, M. Kowarschik, T. Redel, P. Sharma, V. Mihalef, J. Hornegger, and A. Dörfler, “A workflow for patient-individualized virtual angiogram generation based on CFD simulation,” *Computational and Mathematical Methods in Medicine*, vol. 2012, no. 306765, pp. 1–24, 2012.

[12] J. Wang, C. Riess, A. Borsdorf, B. Heigl, and J. Hornegger, “Sparse Depth Sampling for Interventional 2-D/3-D Overlay: Theoretical Error Analysis and Enhanced Motion Estimation,” in *Computer Analysis of Images and Patterns*, 2013, pp. 86–93.