

Depth-Layer-Based Patient Motion Compensation for the Overlay of 3D Volumes onto X-Ray Sequences

Jian Wang^{1,2}, Anja Borsdorf², Joachim Hornegger^{1,3}

¹Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg

²Healthcare Sector, Siemens AG, Forchheim

³Erlangen Graduate School in Advanced Optical Technologies (SAOT)

jian.wang@cs.fau.de

Abstract. A novel depth-layer based patient motion compensation approach for 2D/3D overlay applications is introduced. Depth-aware tracking enables automatic detection and correction of patient motion without the iterative computation of digitally reconstructed radiographs (DRR) frame by frame. Depth layer images are computed to match and reconstruct 2D features into 2D+ space. Using standard 2D tracking and the additional depth information, we directly estimate the 3D rigid motion. The experimental results show that with about 30 depth layers a 3D motion can be recovered with a projection error below 2 mm.

1 Introduction

In interventional radiology, a typical scenario is that real-time information (e.g. projection position of a catheter tip) is provided by 2D X-ray sequences acquired by the C-arm system. The pre-interventional 3D volume is overlaid to augment the 2D images with additional spatial information, noted as 2D/3D overlay. Over decades, different 2D/3D registration methods were developed [1]. In nowadays clinical practice, 2D/3D overlay can achieve high accuracy at the beginning of a procedure as well as after the correction of misalignments due to patient motion during the procedure. However, patient motion needs to be detected by the physician and the correction is manually triggered in most of state-of-the-art applications. Our goal is to achieve automatic patient motion detection and real-time compensation. State-of-the-art 2D/3D registration methods [1] commonly depend on the expensive iterative computation of DRR images of the 3D volume. Meanwhile, standard tracking based motion detection/compensation methods are not optimal for X-ray sequences, due to the fact that structures from different depths overlap each other in the X-ray projection sequences.

In this paper, a novel approach of 2D/3D registration for rigid patient motion compensation is introduced. Instead of iteratively computing DRRs frame by frame, we introduce the concept of depth layer images for recovering depths of the 2D features. Then a standard tracking method is employed to find 2D correspondences between neighboring frames. 3D rigid motion is directly estimated from the depth-aware correspondences.

2 Materials and methods

The proposed approach takes the advantage of depth information from the initially registered 3D volume, and estimates the rigid 3D motion from depth-aware tracking. As illustrated in Fig. 1, it starts with an initially registered 2D/3D overlay, based on which depth layer images are generated (2.1) and compared to the initial frame for 2D+ reconstruction of 2D patches (2.2). A standard tracker is then employed to track patches over time. The rigid 3D motion is estimated taking into account depth information (2.3), which is then applied to the 3D volume for motion compensation.

2.1 Depth layer generation

Depth information is lost in the X-ray projections, which is one of the reasons of using 2D/3D overlay. Given a registered 2D/3D overlay, depth of the overlaid 3D can be encoded in color to improve the depth perception [2]. This leads to the idea of recovering depth information of 2D X-ray features from the initially registered 3D volume. We divide the 3D volume into a stack of depth layer volumes $V_{d_i}, i = 1, \dots, n$ (n is the number of depth layers), in the way that all subvolumes are uniformly divided along the principal ray direction (Fig. 2), which can be rendered independently to generate depth layer images D_i .

One advantage of this concept is that depth layers decompose the overlapping structures into certain depth intervals. Another advantage is that the windowing and volume rendering technique can be specifically selected according to the registration content. For example, in our work bone structures are rendered as the structure of interest due to the fact that bones are almost rigid during the intervention and can represent the rigid patient motion. Contour enhanced volume rendering is used for the purpose of 2D/3D matching. Since the patient motion during the procedure is relative small, the depth layer generation is done only once as an initialization for a specific working position.

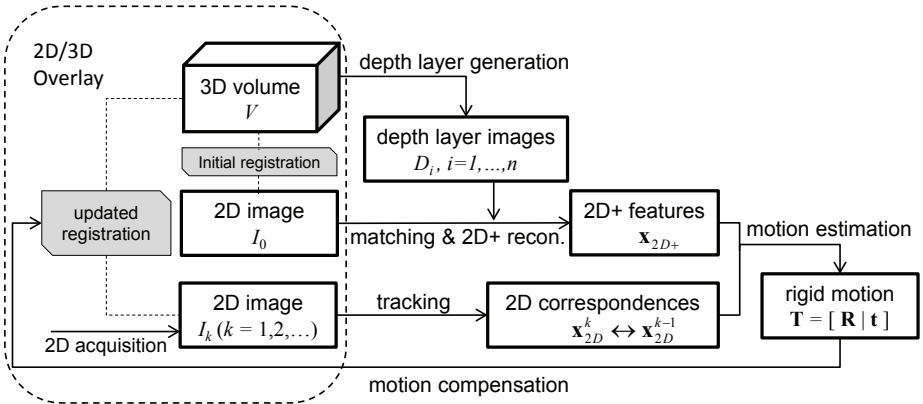


Fig. 1. Flow chart of depth layer based 2D/3D registration for motion compensation.

2.2 2D/3D matching and 2D+ reconstruction

Given the depth layers generated from the initially registered 3D volume, the 2D/3D matching procedure is performed between the X-ray image I_0 and the depth layer image D_i ($i = 1, \dots, n$). As distinctive boundaries of structures in 3D image correspond to strong intensity gradients in the 2D image, our matching strategy is gradient-based: volumetric contour rendering technique [3] is employed to generate the depth layers, and the 2D gradient magnitude map ($|\nabla I_0|$) is used.

The matching is done by patchwise similarity measurement. 2D grids are applied to $|\nabla I_0|$ and D_i to generate 2D patches $p_{|\nabla I_0|}^k$ and $p_{D_i}^k$ from the images ($k = 1, \dots, K$, where K is the number of patches in a 2D image). In our case, $|\nabla I_0|$ and D_i with the size of 800×800 pixels have the patch size of 8×8 pixels. Normalized cross correlation (NCC) [4] is employed as the similarity measure, and the similarity between $p_{|\nabla I_0|}^k$ and $p_{D_i}^k$ is weighted by

$$w_i^k = NCC(p_{D_i}^k, p_{|\nabla I_0|}^k) \cdot \left(\sum_{j=1}^n NCC(p_{D_j}^k, p_{|\nabla I_0|}^k) \right)^{-1} \quad (1)$$

The weight w_i^k indicates the matching probability of $p_{I_0}^k$ (the k th patch of the X-ray image I_0) with a certain depth d_i . The weight w_i^k is normalized over all depths d_i ($i = 1, \dots, n$) for the same patch position k .

Then the 3D position of the 2D feature patches $\{p_{I_0}^k, k = 1, \dots, K\}$ with high matching weight (e.g. for $w_i^k \geq 0.5$) are estimated by 2D+ reconstruction. Since the depth values of the 2D patches associated with D_i are all estimated as the center depth d_i of V_{d_i} , we call the procedure 2D+ reconstruction.

The reconstruction procedure is based on the back-projection of points to rays [5]. In homogeneous coordinates, a 2D point $\mathbf{x}_{2D} \in \mathbb{R}^3$ can be back projected to a ray $\mathbf{r}(\lambda) = \mathbf{P}^+ \mathbf{x}_{2D} + \lambda \mathbf{c}$, that passes through the X-ray source \mathbf{c} and the point

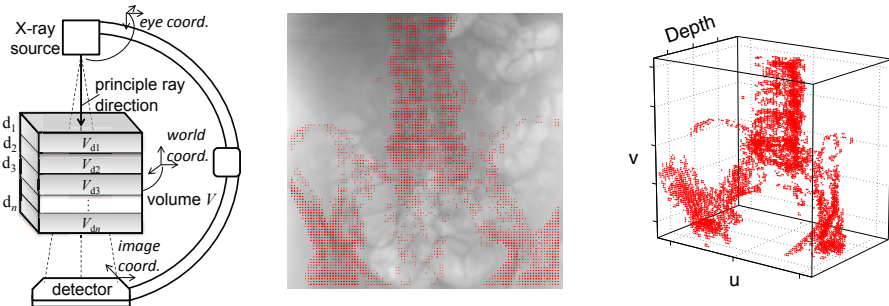


Fig. 2. Matching & reconstruction. Left: Depth layer volumes in C-arm geometry; middle: Matching patches with high weights; right: 2D+ reconstructed matching patches.

$\mathbf{x}_r = \mathbf{P}^+ \mathbf{x}_{2D}$ (\mathbf{P}^+ is the pseudo-inverse of the projection matrix $\mathbf{P} \in \mathbb{R}^{3 \times 4}$). And the 2D+ reconstruction \mathbf{x}_{2D+} of \mathbf{x}_{2D} in depth d_i can be determined by

$$\mathbf{x}_{2D+} = \mathbf{x}_r + \lambda(d_i)\mathbf{c} \quad (2)$$

In the eye coordinate system (Fig. 2), the origin is at the X-ray source \mathbf{c} and the z -axis is aligned with the principal ray direction. Therefore, we have $\mathbf{c} = (0, 0, 0, 1)^T$ and $z_{2D+}/\omega_{2D+} = d_i$, where $\mathbf{x}_{2D+} = (x_{2D+}, y_{2D+}, z_{2D+}, \omega_{2D+})^T$. Together with 2, we have $\lambda(d_i) = (z_{\mathbf{x}_r} - d_i \omega_{\mathbf{x}_r})/d_i$, which is used to determine entries of \mathbf{x}_{2D+} . The results can be transformed to world coordinate system using the projection parameters of C-arm system. The center points of 2D patches with high weights (above) are reconstructed in 2D+ space (below).

2.3 Tracking and motion estimation

After the reconstruction procedure, motion in 3D is estimated by depth-aware tracking. Kanade-Lucas-Tomasi (KLT) Feature Tracker [6] is employed to find 2D correspondences $\mathbf{x}_{2D} \leftrightarrow \mathbf{x}'_{2D}$ between neighboring frames. Our goal is to estimate the motion of the patient between two frames. This motion can be expressed as the inverse of the relative motion of the C-arm, which is considered as a perspective camera. The projection matrix \mathbf{P} can be represented as $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, where $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ contains the intrinsic parameters, rotation $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ and translation $\mathbf{t} \in \mathbb{R}^3$ give the rigid camera motion $\mathbf{T} = [\mathbf{R}|\mathbf{t}] \in \mathbb{R}^{3 \times 4}$.

Non-linear optimization is applied to recover the parameters of rotation and translation. The following error function is minimized between two frames

$$\arg \min_{\mathbf{T}} \left(\sum w \cdot \text{dist}(\hat{\mathbf{x}}_{2D}, \mathbf{x}'_{2D}) \right) \quad (3)$$

where $\text{dist}(\cdot, \cdot)$ is the Euclidean distance and $\hat{\mathbf{x}}_{2D} = \mathbf{K} \cdot \mathbf{T} \cdot \mathbf{x}_{2D+}$. The projection errors are weighted by the matching weights w (calculated in (1)) so that points with higher w contributes more to the results and vice versa.

3 Results

3.1 Point-based simulation experiment (quantitative)

The aim of the experiment is to answer the following questions: (i) how accurate the algorithm can be, (ii) how many depth layers are needed and (iii) how robust is the method with respect to noise. The projection parameters of a real C-arm system (detector pixel size 0.308 mm) is used in the experiment. 3D point sets (20 sets, 50 points per set) were randomly generated in 3D space where the patient is usually laid. Two projections of the 3D points (without and with rigid motion) were generated as two neighboring frames.

The experiment results of 5 examples are shown in Tab. 1. Four motions (pure translation(motion 1), in-plane motion(motion 2), pure rotation(motion 3) and general motion(motion 4)) were tested without 2D corresponding noise. 2D

Table 1. Experiment results of projection errors. The general case refers to motion 4 with 2D noise.

	motion 1	motion 2	motion 3	motion 4	general
translation (mm)	(6, 4, 4)	(6, 0, 4)	(0, 0, 0)	(6, 4, 4)	(6, 4, 4)
rotation (°)	(0, 0, 0)	(0, 9, 0)	(-9, 9, 4.5)	(-9, 9, 4.5)	(-9, 9, 4.5)
ϵ_{ref} (mm)	2.78	4.74	6.15	6.45	6.45
# depth layers	projection errors after registration ϵ_{reg} (mm)				
15	0.12 ± 0.05	0.11 ± 0.05	3.39 ± 1.58	3.20 ± 1.48	3.22 ± 2.01
25	0.08 ± 0.03	0.06 ± 0.03	1.47 ± 1.46	1.90 ± 0.93	2.70 ± 0.79
35	0.06 ± 0.03	0.05 ± 0.02	1.69 ± 0.92	1.56 ± 0.85	2.08 ± 1.23

correspondence noise (± 2 pixels or ± 0.62 mm) was added in the general motion case. The mean 2D offsets (ϵ_{ref}) caused by the motions vary from 2.78 mm to 6.45 mm. The projection errors after registration (ϵ_{reg}) with 15, 25 and 35 depth layers are shown. In the non-noise case, the projection error ϵ_{reg} decreased below 2 mm using 25 depth layers. In the pure translation and the in-plane motion cases, the motions were better corrected. Using 35 depth layers, ϵ_{reg} was around 2 mm in the general motion case with noise.

In Fig. 3, the estimation errors of the motion components are plotted with all tested depth resolutions (motion 4 with noise). The rotation errors were bounded within $\pm 1^\circ$ after 10 depth layers. The in-plane rotation (ΔR_y) was better estimated than off-plane rotation (ΔR_x and ΔR_z). The translation errors also stabilized after 20 to 30 depth layers. These results show that our approach is capable of recovering 3D motion by using depth-aware 2D correspondences, even for the stronger motion (motion 4) under 2D correspondence noise.

3.2 Preliminary tracking-based experiment (qualitative)

In the tracking-based experiment, the X-ray sequence was simulated by DRR computation from a clinical CT volume with a rigid motion sequence. KLT tracking method was employed for 2D tracking. An example of the results is shown in Fig. 4. The initial registered 2D/3D overlay (left), which is the starting point of our approach. Using our depth-layer based motion compensation, the overlay was corrected using the estimated motion from frame to frame. The

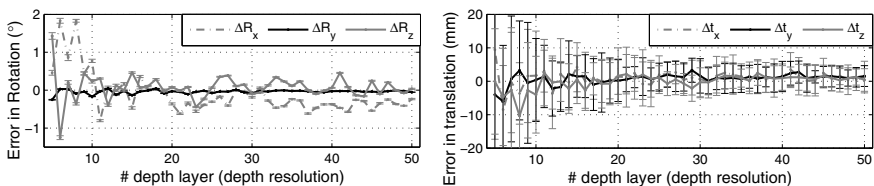
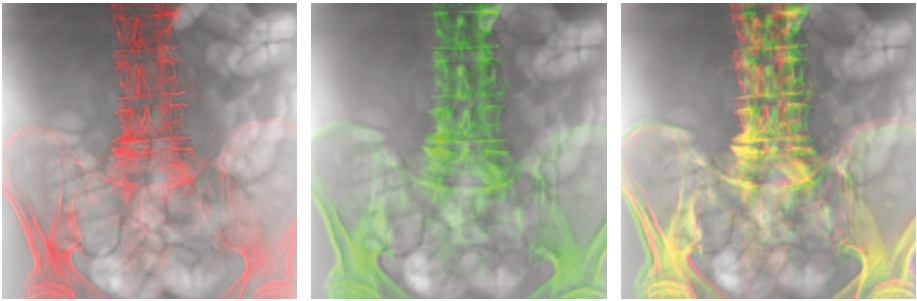
**Fig. 3.** Simulation experiment: general motion case (motion 4, with 2D noise).

Fig. 4. Tracking based experiment. Left: Initially registered 2D/3D overlay (red); middle: Depth layer based motion compensation (green); right: Overlay with (green) and without (red) motion compensation.



preliminary result shows the potential of our depth-aware tracking based approach towards real-time rigid motion compensation.

4 Discussion

We have presented a novel depth-layer based 2D/3D registration approach. The experimental results show that our approach is capable to estimate the rigid 3D motion by 2D tracking and depth-aware motion estimation. With depth layers ($n \in [25, 35]$) the 3D motion can be recovered with a projection error below 2 mm. The method was tested for its robustness against noise in 2D tracking. Furthermore, the method is computationally very efficient, because we do not rely on frame-by-frame iterative DRR computation. This shows the high potential of our approach for robust and real-time compensation of patient motion for 2D/3D overlay.

References

1. Markelj P, Tomaževič D, Likar B, et al. A review of 3D/2D registration methods for image-guided interventions. *Med Image Anal.* 2012;16(3):642–61.
2. Wang J, Fallavollita P, Wang L, et al. Augmented reality during angiography: integration of a virtual mirror for improved 2D/3D visualization. *Proc IEEE Int Symp Mixed Augment Real.* 2012; p. 257–64.
3. Csébfalvi B, Mroz L, Hauser H, et al.; Wiley Online Library. Fast visualization of object contours by non-photorealistic volume rendering. *Proc Comput Graph Forum.* 2002;20(3):452–60.
4. Penney G, Weese J, Little J, et al. A comparison of similarity measure for use in 2D–3D medical image registration. *IEEE Trans Med Imaging.* 1998;17(4):586–95.
5. Hartley R, Zisserman A. *Multiple view geometry in computer vision.* 2nd ed. Cambridge University Press; 2003.
6. Tomasi C, Kanade T. Detection and tracking of point features. Tech Report: CMU-CS; 1991.