Sparse Depth Sampling for Interventional 2-D/3-D Overlay: Theoretical Error Analysis and Enhanced Motion Estimation

Jian Wang[1,2], Christian Riess[1], Anja Borsdorf[2], Benno Heigl[2], and Joachim Hornegger[1,3]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2] Healthcare Sector, Siemens AG, Forchheim
[3] Erlangen Graduate School in Advanced Optical Technologies (SAOT)
`jian.wang@cs.fau.de`

Abstract. Patient motion compensation is challenging for dynamic 2-D/3-D overlay in interventional procedures. A first motion compensation approach based on depth-layers has been recently proposed, where 3-D motion can be estimated by tracking feature points on 2-D X-ray images. However, the sparse depth estimation introduces a systematic error. In this paper, we present a theoretical analysis on the systematic error and propose an enhanced motion estimation strategy accordingly. The simulation experiments show that the proposed approach yields a reduced 3-D correction error that is consistently below 2 mm, in comparison to a mean of 6 mm with high variance using the previous approach.

Keywords: interventional 2-D/3-D overlay, error analysis, sparse depth sampling, 3-D motion estimation.

1   Introduction

In interventional radiology, pre-operative three-dimensional (3-D) images (e.g. computed tomography (CT) or magnetic resonance angiography (MRA)) can be fused with interventional two-dimensional (2-D) X-ray images (fluoroscopy), which is known as 2-D/3-D overlay. This yields several advantages: 1) the pre-operative planning information in the 3-D images can be displayed on the fluoroscopic images; 2) additional information that is not visible in the fluoroscopic images (e.g. vascular structure and spatial information) can be seen in the overlaid 3-D images. A good 2-D/3-D overlay can shorten the time of the procedure and reduce the radiation dose [1]. Accuracy is the most critical factor for the quality of 2-D/3-D overlay. The proper spatial alignment of a 2-D projection to a 3-D image (e.g. volume) is typically referred to as 2-D/3-D image registration.

2-D/3-D registration is usually performed before the intervention to ensure an accurate overlay at the starting point. However, patient motion during the intervention makes it necessary to correct the registration on the fly. In state-of-the-art applications, the patient motion is usually detected by clinicians and

the correction is triggered by user interaction. However, clinicians have limited time and attention for computer interaction during the treatment [2].

Recently, research work for real-time motion compensation can be found in literature [3–5]. All the approaches either are application specific or rely on specific devices or particular motion models.

Recently, we proposed in [6] a depth-layer-based tracking approach for patient motion compensation. The key innovative contribution of this approach is that the depth information is transferred from 3-D image to 2-D feature points using depth layers, which are the images rendered separately from sub-volumes of different depth intervals. Fig. 1(a) shows how the sub-volumes are generated. Based on the initial registration, the 2-D feature points can be mapped to certain depth intervals by matching them to the depth layers. To the knowledge of the authors, this is the first approach that is capable of estimating real 3-D motion by only tracking 2-D feature points from single-view X-ray images. Since this approach does not rely on a particular device or motion model, and no iterative computation of digitally reconstructed radiographs (DRRs) is involved, it yields a high potential for real-time motion compensation in dynamic 2-D/3-D overlay.

However, depth sampling (quantization) introduces a systematic error in motion estimation. Using fine depth sampling can of course reduce the error, but 3-D structures are rather truncated into several small sub-volumes, and this leads to bad 2-D/3-D matching results; In contrast, the 3-D structures are more likely to be preserved in sub-volumes using coarse depth intervals, i.e. using sparse depth sampling. Therefore, we see a requirement to extend the method to be able to handle the depth error caused by sparse depth sampling.

In this paper, we present a mathematical model of the systematic error introduced by sparse depth sampling. Based on this analysis, we propose a depth correction strategy for motion estimation, which handles the systematic error together with random noise. Quantitative simulation experiments are performed to evaluate the new approach. Qualitative results are shown by an example of motion compensated 2-D/3-D overlay using our approach.

2   Theoretical Error Analysis of Sparse Depth Sampling

In this part, we analyze the systematic error of sparse depth sampling, and set it into relation with the random noise coming from other noise sources.

2.1   The Systematic Error Introduced by Depth Sampling

The principle of 2-D/3-D overlay is to virtually place the 3-D volume at the corresponding position of the patient, so that the volume is rendered as imaged from the X-ray source and fused with the live fluoroscopic image [1]. The projection geometry of a C-arm system is described by a pinhole camera model, as shown in Fig. 1(a). The projection procedure is described by the projection matrix

$\mathbf{P} \in \mathbb{R}^{3 \times 4}$, which can be represented as $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, where $\mathbf{K} = \begin{bmatrix} a & & u \\ & a & v \\ & & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3}$

contains the intrinsic parameters, rotation $\mathbf{R} \in \mathbb{R}^{3\times3}$ and translation $\mathbf{t} \in \mathbb{R}^3$ are known as extrinsic parameters [7]. All the parameters are known during the 2-D acquisition from a calibrated C-arm system.

To simplify the problem, the motion estimation is done in the camera coordinate system, where the origin is located at the camera center $\mathbf{c}$ and the z-axis is aligned with the principal ray direction ($\mathbf{L}_0$ in Fig.1(a)). So that the z component of a 3-D point represents its depth, and the projection matrix in the camera coordinate system is simplified as $\mathbf{P}_c = \mathbf{K}[\mathbf{I}|\mathbf{0}]$.

To analyze the systematic error by depth sampling, we start with a 3-D point $\mathbf{x} = (x, y, z, 1)^{\mathrm{T}}$ (homogeneous coordinates) and its projection $\mathbf{p} = (u_p, v_p, 1)^{\mathrm{T}}$ on the detector plane D (Fig. 1(b)). Given the projection matrix $\mathbf{P}_c$, the 2-D projection point $\mathbf{p}$ can be back-projected to a ray in 3-D [7], denoted as $\mathbf{r}(\mathbf{p}) = \mathbf{v}_r(\mathbf{p}) + \lambda\mathbf{c}$, where $\mathbf{v}_r(\mathbf{p}) = \mathbf{P}_c^{+}\mathbf{p} = \left(\frac{u_p - u}{a}, \frac{v_p - v}{a}, 1, 0\right)^{\mathrm{T}}$ and λ is a scalar related to the depth of the 3-D point on the ray [7] [6]. In the camera coordinate system, where $\mathbf{c} = (0, 0, 0, 1)$, $\mathbf{r}(\mathbf{p})$ can be further simplified as

$$\mathbf{r}(\mathbf{p}) = \left((\mathbf{v}_r^{xyz}(\mathbf{p}))^{\mathrm{T}}, \lambda\right)^{\mathrm{T}}, \quad \text{with } \mathbf{v}_r^{xyz} = \left(\frac{u_p - u}{a}, \frac{v_p - v}{a}, 1\right)^{\mathrm{T}} \tag{1}$$

Since the 3-D point $\mathbf{x}$ with depth d and the point $\mathbf{x}_E$ with sparsely estimated depth d_E are both on $\mathbf{r}(\mathbf{p})$, it yields $\lambda(\mathbf{x}) = 1/z = 1/d$ and $\lambda(\mathbf{x}_E) = 1/d_E$. The points can be then reformulated as

$$\mathbf{x} \doteq \left((\mathbf{v}_r^{xyz}(\mathbf{p}))^{\mathrm{T}}, 1/d\right)^{\mathrm{T}} \quad \text{and} \quad \mathbf{x}_E \doteq \left((\mathbf{v}_r^{xyz}(\mathbf{p}))^{\mathrm{T}}, 1/d_E\right)^{\mathrm{T}}. \tag{2}$$

Since $\mathbf{v}_r^{xyz}$ is determined by the 2D projection, the representations in Eq. 2 show the geometric relationship between $\mathbf{x}$ and $\mathbf{x}_E$ (as in Fig. 1(b)): they share the same projection but with a shift of $\Delta d = d_E - d$ in depth.

After a rigid motion (rotation $\mathbf{R}_0$ and translation $\mathbf{t}_0$), the new projections of $\mathbf{x}$ and $\mathbf{x}_E$ are $\mathbf{p}'$ and $\mathbf{p}'_E$, respectively, as shown in Fig. 1(b). In this scenario,

(a) Depth sampling along viewing direction   (b) The systematic error
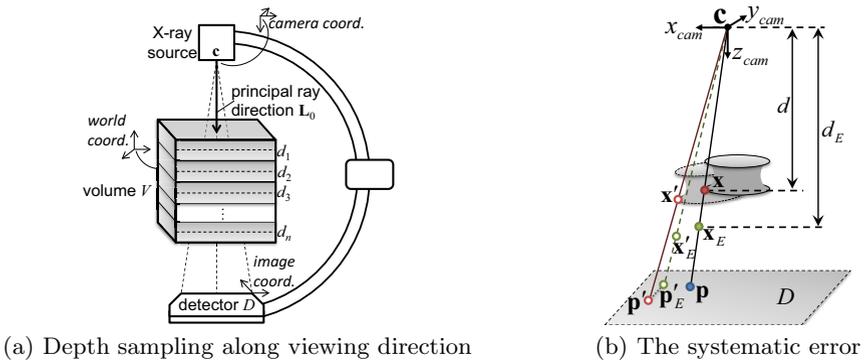
Fig. 1. Illustration of depth sampling and the systematic error

the points $\mathbf{p}$ and $\mathbf{p}'$ are observations of $\mathbf{x}$ on the 2-D image before and after the motion. Since the estimated 3-D point $\mathbf{x}_E$ and $\mathbf{p}'$ (instead of $\mathbf{p}'_E$) are used in motion estimation [6], the systematic error of one point is introduced by the difference vector between $\mathbf{p}'$ and $\mathbf{p}'_E$, as follows:

$$\mathbf{p}'^{xy} - \mathbf{p}'^{xy}_E = a \cdot \frac{d - d_E}{(d \cdot \mathbf{r}_3 \mathbf{v}_r^{xyz} + t_0^z)(d_E \cdot \mathbf{r}_3 \mathbf{v}_r^{xyz} + t_0^z)} \begin{bmatrix} t_0^z & -t_0^x \\ t_0^z & -t_0^y \end{bmatrix} \mathbf{R}_0 \mathbf{v}_r^{xyz} \ , \quad (3)$$

where $\mathbf{r}_3 \in \mathbb{R}^{1 \times 3}$ is the third row of $\mathbf{R}_0$ and $\mathbf{t}_0 = (t_0^x, t_0^y, t_0^z)^{\mathrm{T}}$. The above 2-D vector corresponds to a line segment $\mathbf{l}_\varepsilon$ connecting $\mathbf{p}'$ and $\mathbf{p}'_E$, which is exactly a segment of the epipolar line of $\mathbf{p}$ under the motion of $[\mathbf{R}_0|\mathbf{t}_0]$ [7]. As Eq. 3 shows, the direction of the vector is only determined by the motion $[\mathbf{R}_0|\mathbf{t}_0]$ and the 2-D projection $\mathbf{p}$. The depth error Δd together with the off-plane motion ($\mathbf{r}_3$ and t_0^z) affects the length of $\mathbf{l}_\varepsilon$. Therefore, the systematic error by sparse depth sampling is not only influenced by the estimation error Δd in depth.

2.2   The Systematic Error in Relation to the Random Noise

In the last section, the mathematical representation of the systematic error is derived. However, all the measurements in the real world are subject not only to systematic error but also to random noise [8]. Therefore, we model the noise from other steps in the whole procedure in [6] (e.g. tracking error) as random noise, which is defined by a uniform distribution. In this section, we analyze the systematic error in Eq. 3 together with random noise, in order to treat them differently to achieve a better motion estimation.

Fig. 2(a) illustrates the systematic error together with random noise in our scenario. After an initial motion estimation using sparse depth sampling as in [6], we can compute the systematic error vector using Eq. 3, which corresponds to the line segment $\mathbf{l}_\varepsilon$ through $\mathbf{p}'_E$ (Fig. 2(a)). The possible maximum length of $\mathbf{l}_\varepsilon$ determined by the depth bounds can be used as an explicit measurement of the the systematic error, denoted as $\mathcal{S}^i$ for $\mathbf{p}'^i_E$ (Fig. 2(b) and 2(c)).

Furthermore, we assume that the random noise can shift a 2-D point towards an arbitrary direction with maximum distance $\delta_{\max}$, where direction and distance are uniformly distributed. Therefore, the observation of the projection $\mathbf{p}'$ (denoted as $\widetilde{\mathbf{p}}'$) is positioned within a disk-like region with a radius of $\delta_{\max}$
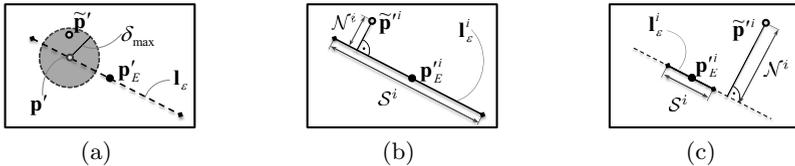
Fig. 2. (a) Illustration of projection and the errors; (b) the metrics for the "influences" the errors; (c) a case with more significant random noise

(Fig. 2(a)), which reflects the accuracy of the tracking method applied in our procedure. However, $\mathbf{p}'$ and $\delta_{\max}$ are unknowns in practice, there is no explicit measurement for the random noise. Therefore, we again make use of Eq. 3 for the metric of the random noise. Since the in-plane motion can be well estimated initially [6] and the depth error as well as the off-plane motion affects more on the length of the systematic error vector in Eq. 3, the true projection $\mathbf{p}'$ appears near to or on the line $\mathbf{l}_\varepsilon$. Thus, we introduce here the point-to-line distance $\mathcal{N}^i$ (the distance between $\widetilde{\mathbf{p}}'^i$ and $\mathbf{l}_E^i$ in Fig. 2(b)) as the metric for the "influence" of the random noise.

Fig. 2(b) and 2(c) show two examples of error conditions. In Fig.2(b), $\mathcal{N}^i$ is obviously smaller then $\mathcal{S}^i$. If it is mostly the case for other points, we can draw the conclusion that the systematic error is more dominant than the random noise. Contrarily, if $\mathcal{N}^i$ is bigger then $\mathcal{S}^i$ (Fig. 2(c)) for most of the points, the random noise appears more dominant.

3   The Error-Dependent Motion Estimation Strategy

In order to reduce the depth error, we now consider a depth correction step after the initial motion estimation (denoted as $[\hat{\mathbf{R}}_0|\hat{\mathbf{t}}_0]$). It can be performed by solving a least-squares optimization problem as

$$\left\{\hat{d}_E^i\right\} = \underset{\{d_E'\}}{argmin} \left(\sum_i^n \text{dist}\left(\widetilde{\mathbf{p}}'^i, \mathbf{p}_E'(d_E'^i)\right) \right) \ , \tag{4}$$

where n is the number of points, $\text{dist}(\cdot, \cdot)$ is the Euclidean distance of the two points and $\mathbf{p}_E' = \mathbf{K}[\hat{\mathbf{R}}_0|\hat{\mathbf{t}}_0]\mathbf{x}_E$. This least-squares optimization helps to find the best fitting corrected depth values based on the estimated motion. We then refine the motion by a follow-up motion estimation using the corrected depths.

However, if the random error is about the same level of or more dominant than the systematic error (e.g. under a small or specific motions causing small systematic error), depth correction can even introduce more error in the motion estimation results (see section 4). The reason is that minimizing $\sum_i^n \text{dist}\left(\widetilde{\mathbf{p}}', \mathbf{p}_E'(d_E'^i)\right)$ in Eq. 4 does not lead to proper fitting depth values 2(c). In contrast, if the random error has an acceptable range (i.e. with reasonable $\delta_{\max}$), it's better to include more points in the motion estimation procedure, so that a globally consistent solution of the motion can be estimated while the effect of the random error is averaged to a minimum.

An error-based motion estimation strategy is therefore proposed according to the influence metrics proposed in section 2.2 and a dominance factor f (Tab. 1). We consider as a strong depth correction criterion if $\bar{\mathcal{S}} > f \cdot \bar{\mathcal{N}}$, we perform depth correction and motion estimation on all points. For the cases not satisfying the strong criterion, we consider as a weak criterion if still some points $\left\{\mathbf{x}_E^i\right\}^{\text{weak}}$ contain dominant systematic error ($\mathcal{S}^i > f \cdot \bar{\mathcal{N}}$) and perform depth correction on $\left\{\mathbf{x}_E^i\right\}^{\text{weak}}$, but still refine the motion using all points. If neither of the criteria are satisfied, we consider all points as random noise dominant (no further correction).

Table 1. The error-dependent motion estimation via dynamic depth correction

---

Inputs 3-D point set with sparse depth estimation $\{\mathbf{x}_E^i\}$, 2-D projections before motion $\{\mathbf{p}^i\}$ and the observed projections after motion $\{\widetilde{\mathbf{p}}'^i\}$.

Initialization Initial estimation of the rigid motion $[\hat{\mathbf{R}}_0|\hat{\mathbf{t}}_0]$ using $\{\mathbf{x}_E^i\}$ and $\{\widetilde{\mathbf{p}}'^i\}$ [6].

Error analysis Based on $[\hat{\mathbf{R}}_0|\hat{\mathbf{t}}_0]$, compute the systematic error vectors for all $\{\mathbf{x}_E^i\}$ using Eq. 3, the systematic error influences $\{\mathcal{S}^i\}$ and the mean influence $\overline{\mathcal{S}}$. Compute the random noise influences $\{\mathcal{N}^i\}$ and the mean influence $\overline{\mathcal{N}}$.

Optimization determination criteria
1. Strong criterion: if $\overline{\mathcal{S}} > f \cdot \overline{\mathcal{N}}$, perform depth correction and motion refinement on all points $\{\mathbf{x}^i\}$;
2. Weak criterion: else if $\max(\mathcal{S}^i) > f \cdot \overline{\mathcal{N}}$, perform depth correction on $\{\mathbf{x}^i\}^{\mathrm{weak}}$ and motion refinement on all points $\{\mathbf{x}^i\}$;
3. Otherwise: no further correction.

---

4   Experiments and Discussion

In this section, we quantitatively evaluate our new approach using point-based simulation experiments. Furthermore, we show the qualitative motion estimation results using a real clinical CT volume with simulated X-ray images.

4.1   Point-Based Simulation Experiments

Point-based simulation is a convenient and established way of evaluating the theoretical-analysis-based algorithms (as in [6]). It allows to neglect the external influences and gives better insights how things behave in the scope of interest. Our point-based simulation set-up is similar as in [6]. The projection parameters of a real C-arm system are applied. 3-D point sets are randomly generated within a bounding box $(20\,\mathrm{cm} \times 20\,\mathrm{cm} \times 30\,\mathrm{cm})$. Random 3-D motions of different scales are generated, which cause 2-D projection errors from about $1\,\mathrm{mm}$ to $13\,\mathrm{mm}$ on the detector plane. Random, uniformly distributed noise with $\delta_{\mathrm{max}} = 4$ pixels (see Sec. 2.2) is added to the 2-D correspondences.

In Fig. 3, results of the test cases using 5- and 10-interval depth sampling are shown. The plots show how the errors caused by a motion, which represent the scale of the 3-D motion, are corrected in 2-D and 3-D. In each plot, the horizontal and vertical axes show respectively the error before and after motion correction. We discuss three important properties of the proposed algorithm.

1. *Error reduction in 2-D and 3-D* – In clinical practice, an error of $2\,\mathrm{mm}$ can be considered acceptable [3]. As shown in Fig. 3, the baseline algorithm [6] often fails to achieve this requirement in 3-D error correction. Conversely, with our proposed depth correction scheme, we consistently yield a 3-D correction error below $2\,\mathrm{mm}$. This shows the fact that the 3D motion (and even the off-plane motion part) can be well estimated using the our proposed approach.

2. *Effects of depth quantization* – Since both examples show sparse depth sampling, where the systematic error causes a significant quantization effect.

(a) Results of using 5-interval depth sampling;

(b) Results of using 10-interval depth sampling
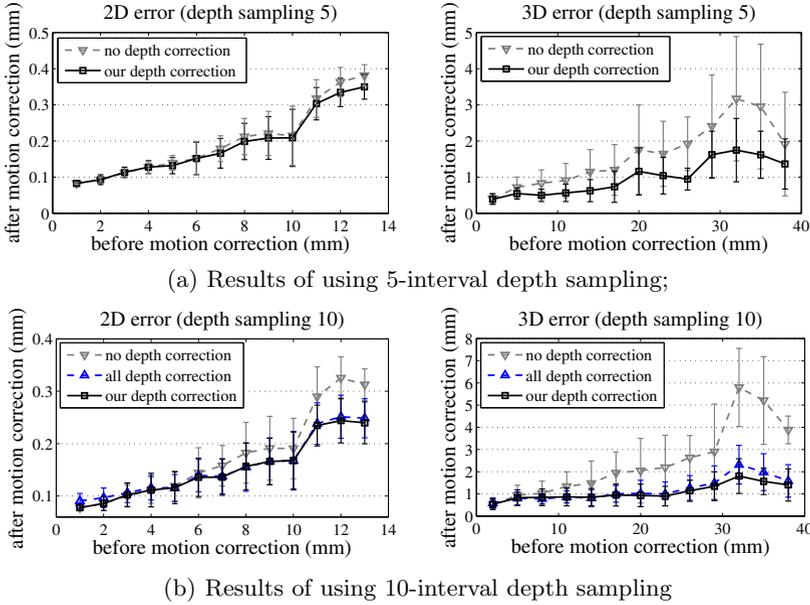
Fig. 3. Plots of the results of point-based experiments

The quantization effect leads to an uncertainty of the motion estimation. As we can see by comparing the results in Fig. 3(a) and 3(b), 10-interval depth sampling even yields worse results then 5-interval sampling. The results are much more stable (consistently under 2/ts mm) by using our proposed depth correction.

3. *Performance gained by the point selection criteria* – In Fig. 3(b), the estimation results using all points for depth correction, where none of the criteria in Tab. 1 are considered, are shown together with the results using our motion estimation strategy (dominance factor $f = 3$). Obviously, more computation is involved if all points are considered for depth correction. Nevertheless, we can observe better results by using points selection criteria for depth correction.

4.2   Image-Based Experiments

Similar as in [6], a sequence of DRRs is generated from a clinical CT volume under a sequence of rigid motion, where 2-D/3-D overlay is initially registered. 10-layer depth sampling is used for motion compensation. Normalized cross correlation (NCC) based similarity map between the 2-D projection (gradient magnitude) and the 3-D volume (gradient-based rendering [6]), as shown respectively in green, blue and red in Fig. 4 together with the overlays. Fig. 4(a) and 4(b) show the 2-D/3-D overlay without and with our motion estimation approach at frame 19. Due to the patient motion in 2-D projection, the overlay loses 2-D/3-D similarity (green) along the frames. However, our motion compensation approach maintains the high 2-D/3-D similarity. The results clearly show that the 2-D/3-D overlay accuracy is strongly improved using our proposed approach.

(a) Without motion compensation   (b) With motion compensation
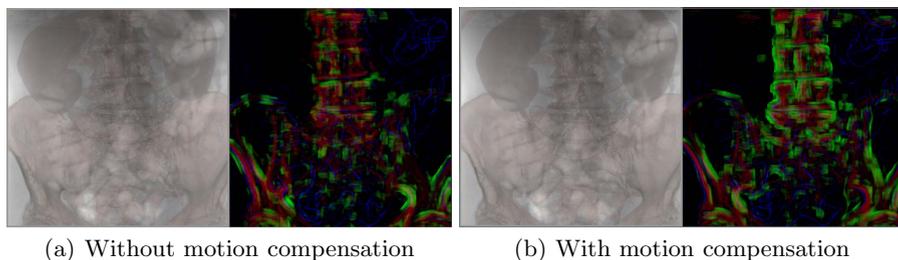
Fig. 4. Motion compensation using sparse depth estimation

5   Conclusion and Future Work

In this paper, a theoretical analysis of the systematic error introduced by sparse depth sampling for motion estimation is presented. An improved motion estimation strategy which handles the depth error is proposed. The experimental results show the improved estimation of 3-D motion in cases of very sparse depth sampling, the 3-D errors are below 2 mm after motion correction.

As an outlook, we will evaluate different tracking methods using the presented error analysis. The theoretical analysis can further help to adapt the tracking methods to X-ray images for our motion compensation framework.

References

1. Rossitti, S., Pfister, M.: 3D road-mapping in the endovascular treatment of cerebral aneurysms and arteriovenous malformations. Interventional Neuroradiology 15(3), 283 (2009)
2. Ruijters, D.: Multi-modal image fusion during minimally invasive treatment. PhD thesis, Katholieke Universiteit Leuven and the University of Technology Eindhoven, TU/e (2010)
3. Brost, A., Liao, R., Strobel, N., Hornegger, J.: Respiratory motion compensation by model-based catheter tracking during ep procedures. Medical Image Analysis 14(5), 695–706 (2010)
4. Ma, Y., King, A.P., Gogin, N., Rinaldi, C.A., Gill, J., Razavi, R., Rhode, K.S.: Real-time respiratory motion correction for cardiac electrophysiology procedures using image-based coronary sinus catheter tracking. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 391–399. Springer, Heidelberg (2010)
5. Wang, P., Marcus, P., Chen, T., Comaniciu, D.: Using needle detection and tracking for motion compensation in abdominal interventions. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 612–615. IEEE (2010)
6. Wang, J., Borsdorf, A., Hornegger, J.: 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 (2013)
7. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge Univsersity Press (2003)
8. Taylor, J.R.: An Introduction Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books (1997)