Preliminary Study on Optical Feature Detection for Head Tracking in Radiation Therapy

Tobias Wissel, Patrick Stüber, Benjamin Wagner, Ralf Bruder, Achim Schweikard, and Floris Ernst

Abstract—**Marker-less tracking provides a non-invasive as well as comfortable approach to compensate for head motion in high precision radiotherapy. However, it suffers from a lack of point-to-point correspondences, typically requiring characteristic spatial landmarks to match point clouds. In this study, we show that cutaneous and subcutaneous structures can be uncovered using an 850 nm laser setup. For three subjects, we compare features extracted from camera images with MR scans serving as an anatomical ground truth. The results confirm the validity of the optically detected structures. The negative correlation between skin thickness and reflected light energy is likewise predicted by Monte-Carlo simulations and can be used to improve spatial point cloud matching. Tissue thickness and its facial structure can be predicted with sub-millimeter accuracy using a Support Vector regression machine. In addition, the optical measurements reveal the location of vessels that are not immediately visible in the MR scan. These promising findings highly encourage its application for a marker-less tracking system.**

I. Introduction

Compensating for patient motion constitutes a challenging as well as important problem for high precision radiotherapy. Common approaches rely on patient immobilization using stereotactic frames, masks or bite bars and mainly employ cone-beam computed tomographies (CTs) for validation purposes [1-3]. To increase patient comfort and reduce the radiation dose, a patient is exposed to, marker-less optical tracking establishes a non-invasive and more convenient approach to compensate for motion or to position the patient [4].

While promising results have already been achieved using commercial systems, the accuracy still lags behind more common approaches of immobilization [5]. One major drawback of marker-less tracking is given by a lack of point-to-point correspondences. Finding these, for objects which are not strictly rigid, is challenging and prone to registration errors. Point cloud matching algorithms are used which rely, if at all, on a few spatial landmarks only [6]. This includes tracking using Microsoft's Kinect camera [7] or Time-of-Flight approaches [8].

In this study we investigate the feasibility of optically detecting subcutaneous as well as cutaneous structures that might be used as additional landmarks for improving the tracking accuracy of the head.

Monte-Carlo simulations of light-tissue interactions recently gave evidence, that soft tissue thickness can be computed from skin reflections in a laser scanner setup. Best results were obtained in the near-infrared (NIR) range [9]. To practically evaluate these results we generated a ground truth for the soft tissue thickness from high resolution magnetic resonance (MR) scans in three subjects. Here we carefully compare it to the NIR results obtained from an experimental setup and model a functional link between features and tissue thickness.

The next section describes the acquisition of the MRI ground truth, the optical setup and the data processing chain. In sections three and four we qualitatively demonstrate and discuss the identification of subcutaneous tissue structures. Additionally, a quantitative evaluation is presented using spatial correlation and Support Vector regression (SVR). Particular focus is directed to the impact of the matching accuracy between the MR and NIR point cloud. In this context the sections report on improving the performance of the Iterative Closest Point algorithm (ICP) by using a correlation based refinement (rCor). Finally, conclusions with respect to marker-less tracking are provided.

II. Materials and Methods

A. Ground Truth - Skin Segmentation in MRI Scans

The soft tissue thickness corresponding to an optical measurement is virtually not known beforehand. We hence employed MR imaging to generate a precise ground truth for the thickness distribution across the forehead.

We recorded MR images using T1 weighting from three male subjects. The images were acquired and reconstructed from a 3D k-space. Using a 3D region-of-interest (ROI) of 240 mm × 240 mm × 90 mm (1600 × 1600 × 90 voxels) at the forehead, this measurement sequence is capable of quickly acquiring a highly resolved volume (reconstructed resolution of 0.15 mm × 0.15 mm × 1 mm). The ROI was aligned to the anterior commissure – posterior commissure line (AC-PC line) of the subject to set one of the coordinate axes of the volume orthogonal to the forehead. After restricting the MR volume to the subject's forehead, the skin and bone surfaces have been extracted in four main steps. First, a subject-specific intensity threshold was applied to the

Manuscript received June 28, 2013. This work was supported by Varian Medical Systems, Inc. and the Graduate School for Computing in Medicine and Life Sciences, University of Lübeck under Grant DFG GSC 235/1.

T. Wissel, P.Stüber and B. Wagner are with the Institute for Robotics and Cognitive Systems, University of Lübeck, 23538 Lübeck, Germany and the Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, 23538 Lübeck, Germany (corresponding author phone: +49 (0) 451 500 5690; fax: +49 (0) 451 500 3364; e-mail: wissel@rob.uni-luebeck.de). All other authors are with the Institute for Robotics and Cognitive Systems, University of Lübeck, 23538 Lübeck, Germany.

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

smoothed volume. This smoothing was achieved with a Gaussian lowpass filter. It cuts off noise in higher spatial frequency ranges. After binarizing, opening and closing operators merge subregions in the images to obtain more compact main regions. These morphological operators were applied several times using disk-shaped structural elements of different sizes. Third, a region growing algorithm was used to select and finally extract the skin region from the image. Finally, the Canny algorithm [10] was used to find the edges corresponding to the skin and bone surface irrespective of their orientation.

B. Experimental Setup

The optical setup contained a beam shaped 850 nm laser (Thorlabs L850P010) diode. The beam was split with a pellicle beamsplitter (Thorlabs CM1-BP145B2) and deflected onto the surface of the subject's forehead. A high dynamic range (HDR) camera (Andor Zyla), connected to the beamsplitter, served as detector for backscatter and reflections from the tissue. The camera was equipped with an infrared filter to block most external light sources. The focal length was fixed to 30 cm.

The entire measurement setup was mounted onto a robot (Adept Viper s850) in order to sample a point cloud across the head. Therefore, a grid of 3 mm resolution was defined by three points spanning the reference plane on the forehead. A second infrared-sensitive camera (IDS UI-3240CP-NIR-GL) was used for laser triangulation to obtain spatial information about the point. With that the laser-to-skin distance was fixed at 30 cm. While scanning, the robot halts at each sample point and adjusts the laser-to-skin distance. Then an image is acquired with the HDR camera and the spatial coordinates of the triangulated laser spot center on the skin surface are stored using coordinates within the robot coordinate system. Details about the laser scanner setup and the technical specification of the experimental setup can be found in earlier work of our group [9, 11].

C. Data Processing

Data analysis focused on a subset of the grid where the laser penetrates the surface approximately perpendicularly. The camera images were recorded as frames of 1000 × 1000 pixels (ca. 1.95 cm × 1.95 cm) and are processed analogously to the simulated case [9].

First, the spot center is detected as the center of gravity. Concentric circles around the spot center define seven regions-of-interest (ROIs) as shown for an example image in Fig. 1a. Second, the light energy within each ROI is integrated to obtain robust features and to reduce the dimensionality of the feature space. Using the recorded point clouds, spatial information was assigned to each feature, serving as a basis for interpolation between the grid vertices (cf. Fig. 1b-h).

In order to compare the MRI surface to the triangulated surface, a transformation matrix was computed using a four-step procedure. First, the initial matching was obtained by manual pre-alignment with respect to spatial landmarks, e.g. the subject's eyes. Second, a point-to-plane Iterative-Closest-Point algorithm was used to minimize the spatial matching error [6]. Third, a correlation refinement (rCor) is proposed which maximizes the Pearson correlation coefficient between subcutaneous structures extracted from the MR image and those from the NIR measurements. A correction was based on the mean correlation of ROIs 2-7. The refinement was performed on 2D projections of the 3D surface onto the x-y-plane (cf. Fig. 2). Out of all 6 spatial dimensions, this refinement only adjusted the translation in x- and y-directions as well as the rotation around the z-axis. Finally, a brute-force algorithm was applied to again reduce the 3D matching error between the surfaces. This exhaustive spatial search (rSpa) only covered the remaining three dimensions (z-translation, x- and y-rotation), while the others were kept fixed.

Based on the resulting transformation matrix, one MR skin thickness label was mapped to each seven-dimensional NIR feature vector of the experimental data set. To model a nonlinear relationship such as given in eq. 1, we trained an SVR model with the generated data set.

$$d_s = f(\mathbf{b}); \text{ with } \mathbf{b} = [b_1, \quad \ldots \quad b_7], f : \mathbb{R}^7 \mapsto \mathbb{R}^1 \qquad (1)$$

This SVR machine builds a functional model based on training data by optimizing the error between true skin thickness and the output of a formulated model hypothesis using an ε-insensitive loss. To allow for nonlinearities, the model hypothesis is approximated using a radial basis function (RBF) kernel. A full account of the algorithm is given by Smola et al. [12]. Relevant model parameters have been determined by grid search. Finally, the generalization error was estimated using a 20-times 10-fold cross-validation procedure.

III. Results

A. Identification of Subcutaneous Structures

Fig. 1 illustrates the entire feature space spatially overlaid on the triangulated point clouds. Prominent structures correspond to changes in reflected light energy. Darker regions indicate patches of less and brighter regions of higher backscatter, i.e. light detected by the camera. Each ROI has been extracted at a certain radius from the spot center. The further a ROI is located from this center, the less light is detected in an absolute sense. Nevertheless, most prominent structures are obtained from ROI 3 onwards.

Fig. 2 relates the optical features with the MRI reference images and hence the soft tissue thickness. The triangulated point cloud was transformed into MR space and again overlaid with color-coded optical features.

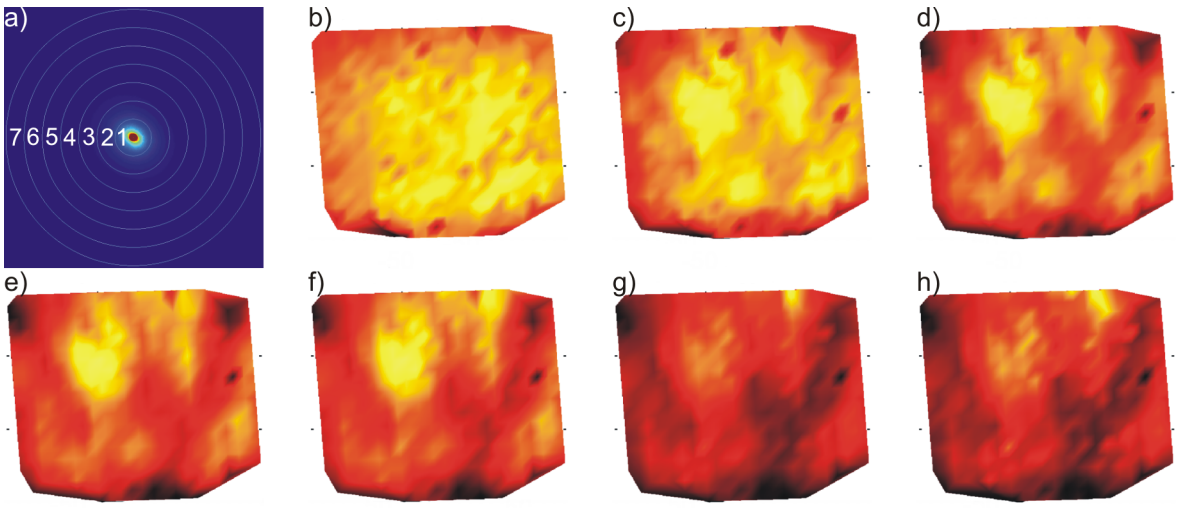

Fig. 1. Feature extraction from NIR tissue reflections. (a) Each feature is extracted as the cumulative light energy integrated across all pixels within one of the concentric ROIs around the spot center (dark = few counts, bright = many counts). The color scale denotes the cumulative ADC counts at locations across the forehead. (b) ROI 1: 0-1.365 mm. (c) ROI 2: 1.365 - 2.73 mm. (d) ROI 3: 2.73-4.095 mm. (e) ROI 4: 4.095-5.46 mm. (f) ROI 5: 5.46-6.825 mm. (g) ROI 6: 6.825-8.19 mm. (h) ROI 7: 8.19-9.555 mm.

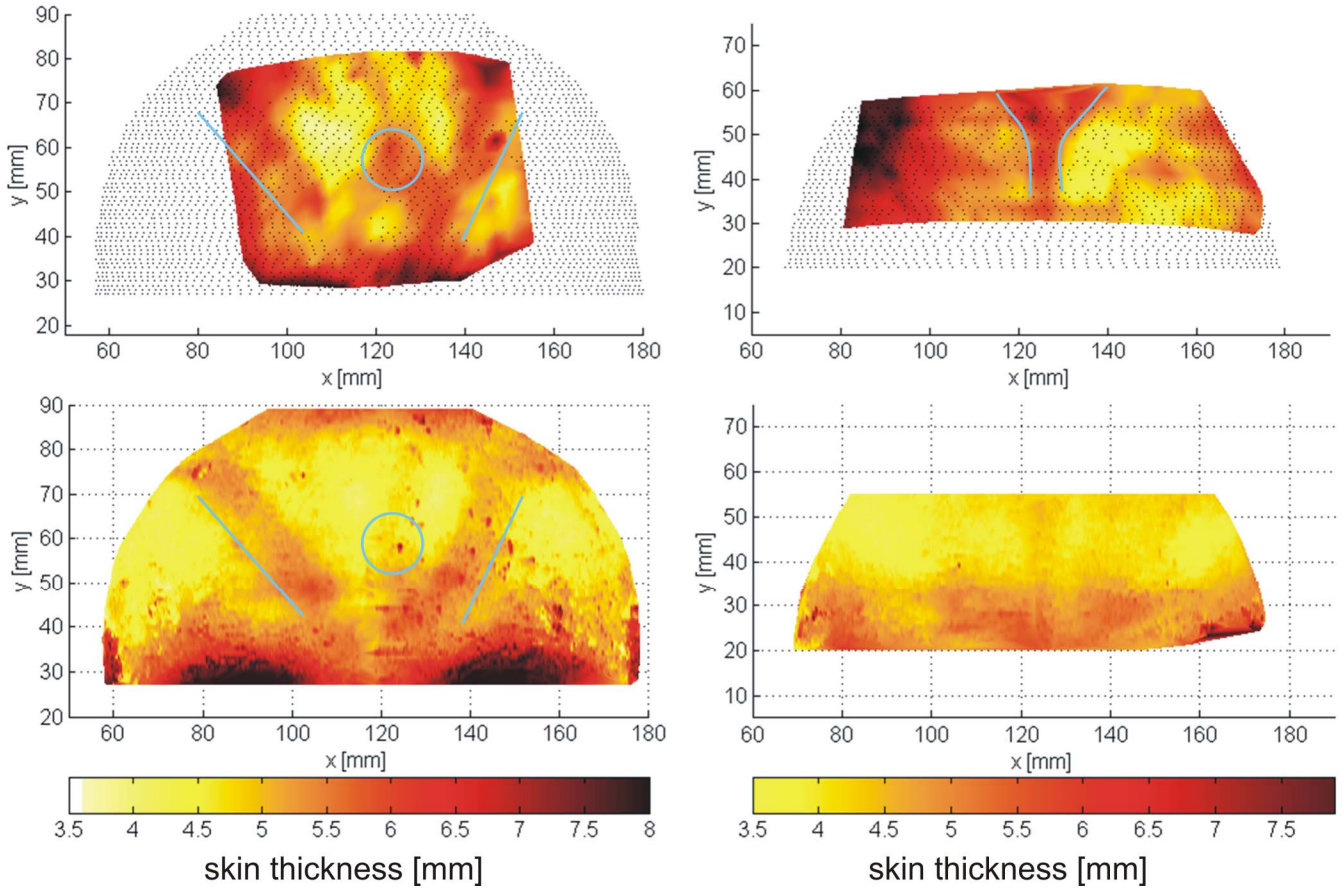

Fig. 2. NIR features projected onto the triangulated point cloud and transformed from robot to MRI space. NIR patches are shown for subjects one (left) and three (right) with the MR extracted point cloud denoted in the background (top). The MR extracted point cloud is overlaid with color coded skin thickness information (bottom, cf. color bar). Characteristic landmarks are indicated using colored marks. View in z-direction, orthogonal to the x-y-plane of the MRI space.

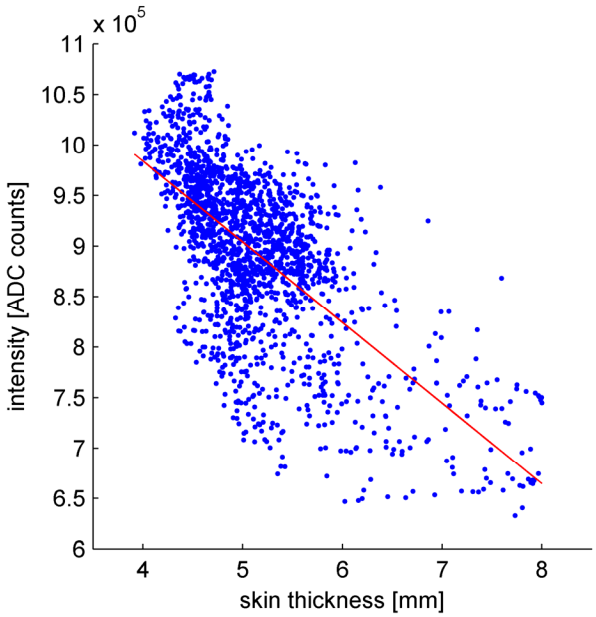

Fig. 3. Feature subspace spanned by ROI 3 for subject 1. The red line denotes a linear fit described by the negative correlation coefficient.

The plots for subjects one and three exemplarily illustrate the results for ROI 3. To compare both modalities, characteristic landmarks such as sharp transitions from thick to thin skin have been emphasized by colored marks. The described matching procedure was successfully carried out for subjects 1 and 2. Tissue structures from MR and NIR imaging were highly correlated with each other and led to a well-aligned matching (cf. Fig. 2, left). For subject 3, the NIR imaging uncovered a large vessel across the forehead (Fig. 2, right, marked in blue), that is not immediately visible in the MR. This vessel, having a substantial impact on the correlation coefficient, corrupted the performance of the correlation refinement converging into wrong local maxima. The matching for subject 3 was thus restricted to the point-to-plane ICP only.

B. Quantitative Data Analysis

Each ROI feature was correlated with the corresponding thickness labels to quantify the matching performance. Fig. 3 shows the feature subspace spanned only by ROI 3. The red line fitted into the relationship is described by the correlation coefficient. The correlation for all ROIs and the first two subjects is plotted in Fig. 4. The coefficients are generally negative and - in an absolute sense - tend to be higher for higher ROIs. The latter tendency is revealed only after the matching refinement as illustrated by the red lines. The blue lines contrast the situation before refining the matching. For subject two this line has no clear trend in an absolute sense and moves around zero at low positive or negative correlations. The red line yields stable correlation around 0.5 with slight superiority for the two highest ROIs.

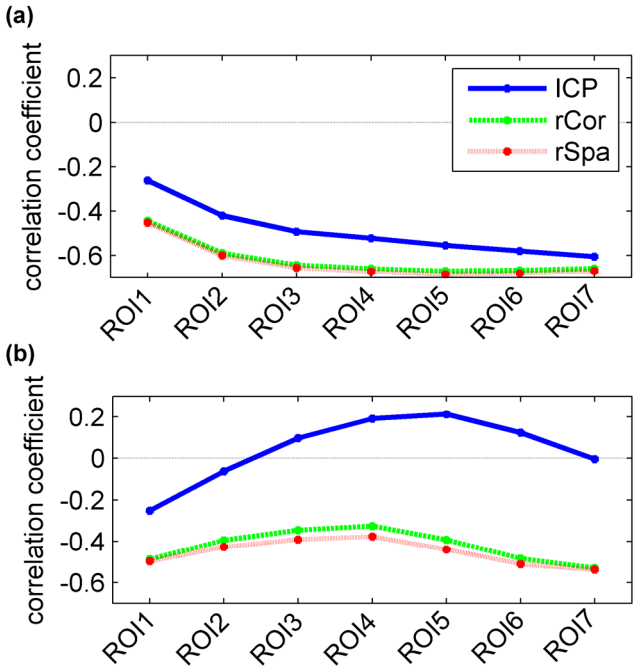

Fig. 4. Correlation coefficients for all ROIs after each matching step: (a) subject 1, (b) subject 2.

TABLE I
MATCHING ERRORS AND PREDICTION ACCURACY

Error[a] [*mm*]	Subject 1	Subject 2	Subject 3
ICP RMS	0.4683	0.8402	0.3016
rCor RMS	1.5830	2.4289	-
rSpa RMS	0.5707	1.0797	-
SVR RMS (after ICP)	0.5643	0.6126	0.2175
SVR RMS (after rCor)	0.4344	0.4910	-
SVR RMS (after rSpa)	0.4221	0.4982	-

[a] ICP = Iterative Closet Point algorithm, rCor = Pearson correlation coefficient-based refinement, rSpa = spatial search-based refinement, SVR = Support Vector regression, RMS = root mean square

The correlation refinement substantially improves the coefficient for all features. The spatial search further enhances the correspondence between the structures recorded with both modalities. A similar trend has been found for the RMS SVR generalization error in Table I. In contrast to the correlation coefficient, the SVR evaluates the nonlinear and multi-dimensional functional relationship between NIR features and skin thickness. On average, the SVR can determine the skin thickness with sub-millimeter accuracy. The error falls below 0.5 mm for all subjects. In contrast the mean absolute error falls below 0.36 mm (not shown) for all subjects. With respect to the conservative RMS measure, this suggest that most thickness samples can

be obtained with low error and that there are some outliers with a higher error. These are mainly located in regions around the eye brows. After refining the initial ICP transformation, the correlation refinement slightly worsens the spatial matching. The subsequent spatial search, however, recovers an RMS matching error that is in a similar range as the initial ICP error (cf. Table I).

IV. Discussion

Comparing the shaded regions (denoting structures in the optical features) to the MRI analysis in Fig. 2 confirms the qualitative correspondence between both plots. The correlation between both plots is clearly visible and also indicated by high linear correlation coefficients of up to 0.7 in Fig. 4. Distinct attention should be directed to the landmarks for subject one marked in Fig. 2. Passing down the head on both sides of the forehead, muscles are forming a "V" shape. This leads to thicker tissue, represented by darker regions in the optical measurements (left and right muscle denoted by lines). Bright yellow areas spreading across the head correspond to regions of lower soft tissue thickness in the MRI plot. This is in line with the simulations: thinner tissue (being less subject to absorption effects) remits more photons than thicker tissue and exhibits a bright color in the top plot.

As predicted by the simulations earlier on, the correlation is generally negative and increases in strength for ROIs at the outer margins of the spot [9]. This particular result is remarkable, since simulations have also shown that the informative effects are on a small scale and correlated in a nonlinear way to the skin thickness. These nonlinearities are modeled by SVR. The model yields generalization errors of sub-millimeter accuracy (cf. Table 1). These are highly promising with respect to the total range covered by the facial tissue thickness (cf. Fig. 2). Similarly, the univariate mapping in Fig. 3 exhibits noisy variations that go far beyond the mean absolute generalization error and introduce a high degree of uncertainty. The substantial improvement suggests that the nonlinear and multivariate SVR model is well suited to accurately map from optical NIR features to the skin thickness scalar.

Both the correlation coefficients as well as the SVR model confirm that precise spatial matching between robot and MR space is vital to accurately estimate the tissue thickness from optical features. A correlation refinement followed by another spatial correction step led to substantial improvement for structural correspondence (cf. Fig. 4 and Table I), but only minor loss in spatial matching error. This emphasizes two points. First, due to few prominent landmarks on the forehead, the ICP algorithm tends to miss the optimal alignment. This weakness establishes a problem for high precision tracking aiming at sub-millimeter accuracy. Second, elastic properties of the tissue further impede a precise matching by merely spatial means. Computing the location of rigid bony structures or uncovering subcutaneous structures by estimating the soft tissue thickness hence provides valuable information to approach the optimal matching. Our results suggest using a fixed marker geometry for the initial ground truth matching between MR and NIR features. This is supported by the optical sensitivity with respect to vessels, which impedes the correlation refinement (cf. subject 3). Nevertheless, this additional sensitivity seems not to impact the SVR performance and provides even more structural information to match NIR to NIR data as necessary for optical tracking.

V. Conclusions

In this study we confirmed earlier simulative results and demonstrated in a practical proof of concept, that soft tissue thickness can be predicted using optical features. We showed that spatial matching is improved by exploiting structural information obtained from NIR features. These results highly encourage a more comprehensive study with a more diverse subject pool and a marker-based ground truth registration.

References

[1] Fuss, M., Salter, B. J., Cheek, D., Sadeghi, A., Hevezi, J. M., and Herman, T. S., "Repositioning accuracy of a commercially available thermoplastic mask system," Radiother. Oncol., vol. 71, pp. 339-345, 2004.

[2] Robar, J., Clark, B., Schella, J., and Kim, C. S., "Analysis of patient repositioning accuracy in precision radiation therapy using automated image fusion," J. Appl. Clin. Med. Phys., vol. 6, pp. 71-83, 2005.

[3] Minniti, G., Clarke, E., Cavallo, L., Osti, M.F., Esposito, V., Cantore, G., Cappabianca, and P., Enrici, R.M, "Fractionated stereotactic conformal radiotherapy for large benign skull base meningiomas," Radiat. Oncol., vol. 6, p. 1, 2011.

[4] Schaerer, J., Fassi, A., Riboldi, M., Cerveri, P., Baroni, G., and Sarrut, D, "Multidimensional respiratory motion tracking from markerless optical surface imaging based on deformable mesh registration," Phys. Med. Biol., vol. 57, pp. 357-373, 2012.

[5] Xia, J., and Siochi, R.A., "A real-time respiratory motion monitoring system using Kinect: Proof of concept," Med. Phys., vol. 39, pp. 2682–2685, 2012.

[6] Besl, P.J. and Keil, N. D., "A method for registration of 3-D shapes," IEEE T. Pattern Anal., vol. 14, pp. 239-256, 1992.

[7] Santhanam, A., Low, D., and Kupelian, P., "3D tracking of interfraction and intrafraction head and neck anatomy during radiotherapy using multiple Kinect sensors," 53rd Annual Meeting of the AAPM, Med. Phys., vol. 38, p. 3858, Aug 2011.

[8] Placht, S., Stancanello, J., Schaller, C., Balda, M., and Angelopoulou, E., "Fast time-of-flight camera based surface registration for radiotherapy patient positioning," Med. Phys., vol. 39, pp. 4–17, 2012.

[9] Wissel, T., Bruder, R., Schweikard, A., and Ernst, F., "Estimating soft tissue thickness from light-tissue interactions — a simulation study," Biomed. Opt. Express, vol. 4, pp. 1176-1187, 2013.

[10] Canny, J., "A computational approach to edge detection," IEEE T. Pattern Anal., vol. 8, pp. 679–698, November 1986.

[11] Stüber, P., Wissel, T., Wagner B., Bruder, R., Schweikard, A., and Ernst, F., "Design and evaluation of a highly accurate optical setup for backscatter analysis," 44. Jahrestagung der Deutschen Gesellschaft für Medizinische Physik, Köln, Germany, September 2013.

[12] Smola, A., and Schölkopf, B., "A tutorial on Support Vector regression," Stat. Comput., vol. 14, pp. 199-222, 2004.