Tissue Thickness Estimation for High Precision Head-Tracking using a Galvanometric Laser Scanner - A Case Study

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

Abstract— **Marker-less optical head-tracking constitutes a comfortable alternative with no exposure to radiation for real-time monitoring in radiation therapy. Supporting information such as tissue thickness has the potential to improve spatial tracking accuracy. Here we study how accurate tissue thickness can be estimated from the near-infrared (NIR) backscatter obtained from laser scans. In a case study, optical data was recorded with a galvanometric laser scanner from three subjects. A tissue ground truth from MRI was robustly matched via customized bite blocks. We show that Gaussian Processes accurately model the relationship between NIR features and tissue thickness. They were able to predict the tissue thickness with less than 0.5 mm root mean square error. Individual scaling factors for all features and an additional incident angle feature had positive effects on this performance.**

I. Introduction

Precise treatment of tumors in radiotherapy directly depends on the accuracy of patient positioning and motion compensation during the treatment. Therefore, state-of-the art intracranial radiotherapy either employs stereotactic masks for patient immobilization [1] or X-ray based monitoring and motion compensation such as the 6D skull tracking used by the CyberKnife® [2]. These entail several drawbacks: First, the mask systems are rather uncomfortable and are not tolerated by all patients. They further provide no motion monitoring during the treatment - the patient is assumed to be fixed. This is only true up to errors in millimeter range [1]. Second, X-ray based imaging exposes the subject to an unnecessary amount of additional radiation. This exposure also limits the head-tracking speed to about 1 Hz.

Therefore, recent studies propose marker-less optical head-tracking as a promising alternative, where a laser constantly scans the patient's forehead. This approach would require only light patient fixation and provides a basis for fast real-time monitoring [3, 4]. To tackle the lack of prominent landmarks at the forehead and to increase robustness, spatial information is combined with optical backscatter features for each laser spot. Monte-Carlo simulations [3] and an experimental study [4] gave evidence that, for light in the near-infrared (NIR) range, these features can be used to reconstruct the tissue thickness across the forehead, revealing subcutaneous structures. These may act as supportive landmarks for a tracking algorithm. This reconstruction can be achieved using regression techniques such as Support vector regression (SVR).

Despite general feasibility, [4] states four unsolved challenges. First, the matching between the NIR scans and the MR image, which is used as a ground truth for the tissue thickness, is not reliable. So far there was no satisfying way to judge the accuracy of the matching approach consisting of a combination of an Iterative-Closest-Point (ICP) algorithm [5] and a correlation-based refinement. For some subjects this approach was not fully applicable. Second, the laser scanning process was performed by a robot, which results in a slow scanning speed. Third, training an SVR model [6] required time-consuming grid search and could not readily incorporate scaling parameters for individual features such as the incident angle [7]. Fourth, the preliminary study was limited to only three subjects.

In this study we tackle the first three challenges. All of them require careful attention before a more containing study can be conducted. We therefore propose the following: First, a marker-based ground truth for seeding the ICP is used. Patients will be outfitted with a marker visible during MR imaging and the NIR scan. This provides the spatial transformation between the two spaces given a known marker geometry. Second, a faster mirror-based deflection system will replace the robot. We show how tissue thickness estimation is influenced by this galvanometric laser scanner, which naturally implies stronger non-informative impact of changing incident angles on the data [7]. Third, Gaussian Processes (GPs) will be used to compute the tissue thickness from the recorded optical features [8]. They employ Bayesian inference to determine the most likely model parameters given the data and avoid time-consuming grid search on the training set. They implicitly provide the possibility to optimize scaling parameters for each feature by incorporating them into the inference. This renders automatic relevance detection (ARD) practically feasible. For comparison with the study conducted in [4], we will present a case study on new data recorded from the same three volunteering subjects.

The article will first review materials and methods. Particular attention will be directed to the new contributions. Then the results and subsequent discussion will present and debate matching and thickness estimation accuracies. A final conclusion will be provided.

Manuscript received March 17, 2014. 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.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

II. Materials and Methods

A. Experimental Setup - NIR Laser Scans

The robotized setup described in [4], including an 850 nm laser source, a triangulation camera (IDS UI-3340CP-NIR-GL) and a 16 bit High Dynamic Range (HDR) camera (ANDOR Zyla) was extended by a galvanometric deflection unit (ASX-V20, laserwinkel.nl) [9]. NIR features were extracted with the HDR camera. A tracking camera (Polaris Spectra, NDI, 0.25-0.3 mm RMS tracking error) was used to track optical marker spheres attached to each customized subject marker (cf. Fig. 1).

The marker was constructed out of Poly(methyl methacrylate) (PMMA) and can be rigidly attached to the subject's upper jaw using a customized dental cast. The liquid capsules at the upper end of the marker are MR-visible.

The triangulation camera was calibrated to the tracking camera using the standard direct linear transformation (DLT) algorithm [10] and recordings of the passive marker sphere of the marker. The scanner was configured to raster a grid of 32 x 32 = 1024 points. For each of the corresponding rays a galvanometer-to-triangulation-camera calibration was carried out and stored in a look-up table manner as described in [9]. Using the mentioned calibrations and the marker geometry, the NIR scans were transformed into the MR coordinate space. NIR and MR scans were recorded from three subjects (male, aged 25-34), who's head was immobilized using a vacuum cushion.

B. Ground Truth - Skin Segmentation in MRI Scans

MR imaging was used to generate a precise ground truth for the thickness distribution across the forehead.

The MR images were acquired with T1 weighting and were reconstructed from a 3D k-space. Using a 3D volume-of-interest (VOI) of 210 mm × 210 mm × 60 mm (2048 × 2048 × 60 voxels) at the forehead, this measurement sequence is capable of quickly acquiring a highly resolved volume (reconstructed resolution of 0.103 mm × 0.103 mm × 1 mm). The VOI 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. The z-axis passes parallel to the forehead. After restricting the MR volume to the subject's forehead, the skin and bone surfaces have been extracted as described in [4].

Each of the markers as shown in Fig. 1 had MR-visible spherical capsules and was used by every subject during data acquisition. The marker geometry was extracted from high resolution CT images (0.33 mm × 0.33 mm × 0.2 mm resolution, 512 × 512 × 1000 voxels). Both the liquid capsules and the reflective marker spheres were visible in the CT image. Their relative position with respect to each other is hence given within this marker coordinate system.

All three recordings (MR-, CT- and NIR-scans) were acquired in temporal proximity. For any acquisition, the marker was used in the same pose (shown in Fig. 1). Preliminary tests with the tracking camera revealed no measureable deformation errors in any of the marker axes. They can hence be assumed to lie below the tracking accuracy.

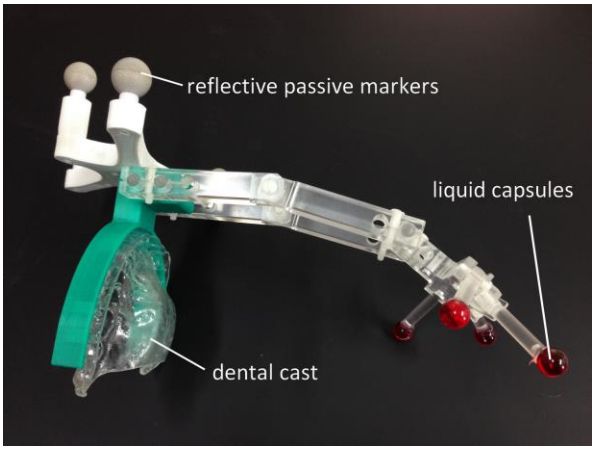

Figure 1. Marker geometry used to match NIR to MRI data. The marker can be rigidly linked to the cranial bone using a customized dental cast. It is visible for both MRI and the tracking camera via liquid capsule and reflective marker spheres, respectively.

C. Data Pre-processing and Feature Extraction

The images from the HDR camera were processed as described in [4]. After detecting the centroid of the spot, the pixel intensities were accumulated within 5 concentric regions-of-interest (ROIs) around the spot center. To speed up the data transfer rate between HDR camera and host computer, a smaller image size (1000 x 1000 pixels) than in [4] was chosen. This resulted in 5 instead of 7 ROIs. The incident angle for each spot was obtained using the triangulated forehead surface and the stored ray information from the galvanometer-to-triangulation-camera calibration. No angle compensation was applied to the spots, but the angle was used as a sixth feature [7].

After subtracting the minimum, each feature and the tissue thickness label were scaled to range [0,1] and the mean was removed. Sample-wise correspondences between features and tissue labels were derived after coordinate transformation into the same space. As described earlier, this was done using the transformation defined by the marker and the other calibration steps. Finally, this transformation was taken as a seed for an ICP refinement step to compensate for remaining minor displacement, if necessary.

D. Gaussian Process Models

Similar to SVR [6], Gaussian Processes (GPs) use a general data model with white noise as illustrated in (2) [7]. By assuming that each skin thickness sample d_s follows a Gaussian distribution, the GP corresponds to a distribution over functions f in the feature space:

$$f \sim \mathcal{GP}(m,k). \qquad (1)$$

The process is fully described by second-order statistics (with feature vector $\mathbf{b}$, mean and covariance function, $m(\mathbf{b})$ and $k(\mathbf{b},\mathbf{b}')$, respectively):

$$d_s = f(\mathbf{b}) + \varepsilon; \text{ with } \mathbf{b} = [b_1, \ldots b_6],\ f\colon \mathbb{R}^6 \mapsto \mathbb{R}^1 \quad (2)$$

Each function is assigned a probability given the training data. This is achieved by Bayesian inference which yields a posterior distribution $p(f|D)$ for the function parameters given the data $D = \{\mathbf{B}=[\mathbf{b_1}, \ldots \mathbf{b_N}], \mathbf{d_s}=[d_{s1}, \ldots d_{sN}]\}$. An appropriate parameter set is obtained by minimizing the negative log marginal likelihood (NLML) using several runs of randomly initialized gradient descent. This replaces time-consuming grid search based on cross-validation as it was used to optimize the SVR parameters in [3,4]. It finds the most likely parameters for the model given a single training data set.

Since the mean is removed from the data during preprocessing, the mean function is always set to zero. The entire function is hence described by the covariance $k(\mathbf{b},\mathbf{b}')$, which models the mutual statistical dependencies between different regions in the feature space. It thus learns similarity between data points. This similarity is expressed by kernel functions. A comparable analog to the SVR radial basis function (RBF) kernel is given by the squared exponential (SE) kernel described by Rasmussen [7]. However, Stein argues that the Matérn kernel function includes the SE kernel and offers higher local flexibility [11]. Therefore the Matérn class was used:

$$k(\mathbf{r}) = s^2(1+\mathbf{r}\sqrt{3})\cdot\exp(-\mathbf{r}\sqrt{3}). \quad (3)$$

$$k_{iso}(\mathbf{r}) = k(\mathbf{r}/\gamma_0) \quad (4)$$

$$k_{ard}(\mathbf{r}) = k(\mathrm{diag}(\gamma_1^{-1}, \ldots \gamma_6^{-1})\mathbf{r}). \quad (5)$$

with the scaling factors s and γ_i, and $\mathbf{r} = |\mathbf{b}-\mathbf{b}'|$. Eq. (4) represents the isotropic and (5) the automatic relevance detection (ARD) Matérn kernel. The latter weighs features according to relevance by introducing individual scaling factors γ_i for each feature. A new sample d_s^* for a feature vector $\mathbf{b}^*$ is predicted using the conditional distribution $p(d_s^*|\mathbf{b}^*,D) \sim N(\text{đ}_s^*, \mathrm{var}[d_s^*])$ [8], with

$$\text{đ}_s^* = k(\mathbf{B}, \mathbf{b}^*)^T\, k(\mathbf{B}, \mathbf{b}^*)^{-1}\, \mathbf{d_s} \quad (6)$$

$$\mathrm{var}[d_s^*] = k(\mathbf{b}^*,\mathbf{b}^*) - k(\mathbf{B},\mathbf{b}^*)^T\, k(\mathbf{B},\mathbf{B})^{-1}\, k(\mathbf{B},\mathbf{b}^*). \quad (7)$$

The evaluation of the prediction error on each data set has been carried out using 5-times-10-fold cross-validation in terms of root-mean-square (RMS) and mean absolute (MA) error.

III. Results

A. Matching Results and Angle Influence

Table 1 summarizes the results from all three subjects. For the data analysis, only a subset of the in total 1024 points was used. Data sample rejection was carried out when the laser hit hair or when the triangulation was not possible due to very steep incident angles. The mean incident angle ranged between 32° and 37° and was higher than with the robotized scanner in [4] or [7], where approximately orthogonal irradiation (0°) or a mean angle of about 20° was found, respectively. The head motion in all three coordinate axes for all subjects was on average below 1 mm for almost all cases.

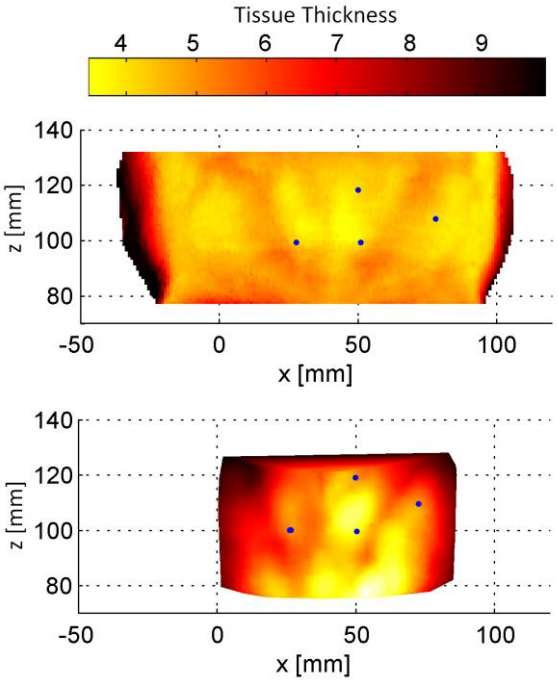

Figure 2. Matching result using a seeded ICP algorithm for subject 1. The plot contrasts skin thickness from an MR scan (top) with the NIR feature of ROI number 2 (bottom). Both are plotted in the same coordinate space and four blue markers show selected correspondences for better orientation. The NIR feature is measured in accumulated ADC counts of the camera sensor (from yellow (high intensity) to dark red (low intensity).

This motion was compensated for each spot individually using the corresponding tracking pose of the marker at that time. Considering this and the matching procedure described earlier, a very low matching ICP error of less than 0.7 mm was achieved for all subjects.

In addition to this quantitative measure, Fig. 2 illustrates the matching result for subject one. The subcutaneous structure in the MR segmentation (top) can be partially identified in the NIR scan (bottom) as well. The blue spots provide landmarks for better orientation and confirm a good alignment between the scans. Although the relationship between NIR intensity and tissue thickness is nonlinear, a negative correlation is clearly visible (thick skin corresponds to low intensities). This relationship is perturbed by angle influences, which are small in the center of the patch and increase to the sides. They are visible as darker red regions at the outer margins (cf. Fig 2, bottom), which seem to indicate a higher thickness where there is none.

B. Tissue Thickness Estimation

All proposed kernel functions yielded tissue thickness estimation errors below 0.5 mm. The resulting RMS as well as MA errors are in a similar range for all subjects. Table 1 clearly shows that the ARD kernel gave superior accuracy compared to the isotropic kernel version (11.2% higher for S1, 3.4 % for S2 and 9.4% for S3). Dropping the angle features led to a small decrease in performance. For ARD the scaling factors γ_i were negative for all NIR features and

TABLE I
DATA CHARACTERISTICS, MATCHING ERRORS AND PREDICTION ACCURACY

	Subject 1 (S1)	Subject 2 (S2)	Subject 3 (S3)
Data Samples	888	945	790
$\alpha_{mean} \pm \alpha_{std}$ [°]	37.1±10.6	39.0±10.6	32.4±10.4
$x_{std}/y_{std}/z_{std}$ [mm]	0.19/0.20/0.46	0.23/1.67/0.86	0.06/0.05/0.05
ICP: RMS [mm]	0.648	0.465	0.500
RMS (MA) error [mm]			
GP: $k_{iso,\ with\ angle}$	0.366 (0.270)	0.344 (0.268)	0.308 (0.232)
GP: $k_{ard,\ w/o\ angle}$	0.336 (0.235)	0.346 (0.266)	0.285 (0.211)
GP: $k_{ard,\ with\ angle}$	0.325 (0.225)	0.332 (0.255)	0.280 (0.205)
$I_{90\%}$ [mm]	0.499	0.559	0.442

positive for the angle feature (except for S1, where it took small values around zero).

The last row of Table 1 lists the upper bound of the interval [0, $I_{90\%}$]. For each subject 90% of all absolute errors fell into this interval. The upper bound falls below 0.6 mm for all subjects.

IV. DISCUSSION

All results presented in the previous section suggest that the customized marker geometry delivers a reliable seed for the subsequent ICP refinement. This is essential since the algorithm is sensitive to run into local error minima, if the initial alignment of the two surfaces is not close to the optimum already. This is particularly the case for surfaces lacking prominent landmarks such as the forehead. The evidences are threefold: The ICP *and* the GPs yielded small errors, both below 0.7 mm (cf. Table 1). Finally, visual inspection (cf. Fig. 2) of subcutaneous landmarks also confirmed accurate alignment. Possible errors introduced by the marker were kept as small as possible: segmentation errors from MR and CT range around half the voxel resolution, and tracking and deformation accuracy are below 0.3 mm. In contrast, it was difficult to judge the accuracy of the ground truth matching in [4] and [7]. A supporting correlation refinement of the ICP based on the features seems helpful, but as discussed in [4], is not applicable for all subjects. Reasons include prominent vessels across the forehead, which hardly affect the tissue thickness, or increased angle influences of the incident laser beam. The latter play an even bigger role for the galvanometric scanner used here (cf. angles in Table 1) and superpose the informative features. A non-optimal ground truth matching may hence have caused higher and more heterogeneous estimation errors across subjects in [4]. Here, we demonstrate for the first time very low estimation errors for data from a galvanometric laser scanner. The errors are homogeneous across subjects (all around 0.3 RMS). Nevertheless, the results confirm the feasibility of submillimeter accuracy in [4] as well as the beneficial effect of an additional angle feature found in [7]. However, the latter has been found to be small. Automatic relevance detection (ARD) was able to outperform the isotropic kernel in every case and found individual scaling factors for all features by the means of Bayesian inference instead of the time-consuming grid search applied in [4]. The isotropic kernel scales up the angle effects within the NIR features as well. It cannot fully exploit the heterogeneity of informative content among different features. By introducing a separate factor in the ARD version, the angle feature can compensate for these angle dependent effects within the NIR features.

In fact, the overall accuracy is bounded by the ground truth, i.e. measurement resolution. But, the presented results already meet the project goal of sub-millimeter accuracy and are below fixation errors achieved with common mask systems.

V. CONCLUSION

Using an enhanced ground truth matching, Gaussian Processes were able to predict tissue thickness with high accuracy from NIR data recorded with a galvanometric laser scanner. Automatic relevance detection as well as the angle feature yielded promising results for all subjects studied. The results form a solid basis for a more containing study investigating more subjects and the impact on a tracking algorithm in future research.

ACKNOWLEDGMENT

This work was supported by Varian Medical Systems, Inc. The authors further acknowledge the support of Dr. C.-S. Schwarz and Dr. Jacobsen from the clinic of oral and maxillofacial surgery, University Hospital Schleswig-Holstein.

REFERENCES

[1] E. Tryggestad, M.Christian, E. Ford, C. Kut, Y. Le, Y. G. Sanguineti, D.Y. Song, and L. Kleinberg, "Inter- and intrafraction patient positioning uncertainties for intracranial radiotherapy: A study of four frameless, thermoplastic mask-based immobilization strategies using daily cone-beam CT," *Int. J. Rad. Oncol.*Biol.*Phys,* vol. 80, pp. 281 - 290, 2011.

[2] G. Kurup, "CyberKnife: A new paradigm in radiotherapy," *J. Med. Phys.,* vol. *35*, pp. 63-64, 2010.

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

[4] T. Wissel, P. Stüber, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst, "Preliminary Study on Optical Feature Detection for Head Tracking in Radiation Therapy," *13th IEEE International Conference on BioInformatics and BioEngineering (BIBE*), IEEE, Chania, Greece, pp. 1-5, 2013.

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

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

[7] T. Wissel, P. Stüber, B. Wagner, R. Bruder, A. Schweikard, and F. Ernst, "Angle influence and compensation for marker-less head tracking based on laser scanners," *Proceedings of the 28th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS'14),* Fukuoka, Japan, accepted, 2014.

[8] C.E. Rasmussen, and C.K.I. Williams, "Gaussian processes for machine learning", *MIT Press*, 2006, www.gaussianprocess.org/gpml

[9] B. Wagner, P. Stüber, T. Wissel, R. Bruder, A. Schweikard, and F. Ernst. "Accuracy analysis for triangulation and tracking based on time-multiplexed structured light", *Med Phys*, submitted, 2014.

[10] Y.I. Abdel-Aziz, and H.M. Karara, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Proceedings of the Symposium on Close-Range Photogrammetry, Falls Church, VA,* pp. 1-18, 1971.

[11] M.L. Stein, "Interpolation of spatial data: some theory for kriging," *Springer, New York*, 1999.