Daniel Wulff*, Ivo Kuhlemann, Floris Ernst, Achim Schweikard, and Svenja Ipsen

Robust motion tracking of deformable targets in the liver using binary feature libraries in 4D ultrasound

<https://doi.org/10.1515/cdbme-2019-0151>

Abstract: In radiation therapy of abdominal targets, optimal tumor irradiation can be challenging due to intrafractional motion. Current target localization methods are mainly indirect, surrogate-based and the patient is exposed to additional radiation due to X-ray imaging. In contrast, 4D ultrasound (4DUS) imaging provides volumetric images of soft tissue tumors in real-time without ionizing radiation, facilitating a non-invasive, direct tracking method. In this study, the target was defined by features located in its local neighborhood. Features were extracted using the FAST detector and the BRISK descriptor, which were extended to 3D. To account for anatomical variability, a feature library was generated that contains manually annotated target information and relative locations of the features. During tracking, features were extracted from the current 4DUS volume and compared to the feature library. Recognized features are used to estimate feature position and shape. The developed method was evaluated in 4DUS sequences of the liver of three healthy subjects. For each dataset, a target was defined and manually contoured in a training and a test sequence. Training was used for library creation, the test sequence for target tracking. The target estimations are compared to the annotations to quantify a tracking error. The results show that binary feature libraries can be used for robust target localization in 4DUS data of the liver and could potentially serve as a tracking method less sensitive to target deformation.

Keywords: Image guidance, motion compensation, radiation therapy, deformation, feature detection

1 Introduction

In radiation therapy, ionizing radiation is used to destroy tumor cells while healthy tissue is preserved. During treatment, the tumor moves due to intrafractional motion like breathing. Translation, rotation and deformation of the tumor pose chal-

lenges for the optimal dose delivery to the tumor [1]. Current tracking methods to detect the target motion mainly rely on X-ray imaging and are invasive because of the use of implanted markers. The indirect methods depend on correlation models between the markers and the target [2]. Furthermore, the patient is exposed to additional radiation due to X-ray imaging. Using markers is a surrogate to approximate the target motion because the target cannot be detected directly. Depending on the marker location, relative motion between target and markers can be observed [3]. However, with a direct tracking method, target motion could be detected without a surrogate. 4D ultrasound (4DUS) provides volumetric images of soft-tissue tumors in real-time without ionizing radiation. This imaging method provides the possibility of a non-invasive, direct tracking method without using markers [4].

A segmentation process in an US volume can be time-consuming and inaccurate due to low resolution. Therefore finding methods for fast feature tracking is an important research question. Banerjee et al. [5] use block matching to locate a predefined target in 3D US volumes. In contrast to that, Royer et al. [6] use a tetrahedral mesh model that describes the target shape. This model is adapted to a moving target on the basis of inner and outer transformations. The developed tracking method in this study uses features in the local neighborhood of the target. Features are fast detectable points in a volume that contain local volume information. The use of features has the advantage that the target does not have to be segmented in each volume of a 4DUS sequence. Thus, a target tracking method based on 4DUS and features is developed and evaluated for three *in vivo* US sequences of the liver of healthy subjects.

2 Materials and Methods

The developed target tracking method is based on the FAST feature-detector by Rosten et al. [7] and on the BRISK feature-descriptor by Leutenegger et al. [8]. Due to the fact that FAST and BRISK were developed for 2D image data, these methods need to be extended to 3D volume data to make them usable for 4DUS. With the help of 3D FAST and BRISK and anno-

*Corresponding author: Daniel Wulff, Universität zu Lübeck, Ratzeburger Allee 160, Lübeck, Germany, e-mail: wulff@rob.uni-luebeck.de

Ivo Kuhlemann, Floris Ernst, Achim Schweikard, Svenja Ipsen, Institut für Robotik und Kognitive Systeme, Lübeck, Germany

Fig. 1: 3D FAST (a) and BRISK (b) masks define which voxels nearby the centered keypoint are used in the algorithms.

tated 4DUS training data, a feature library is created that is used in the tracking process.

2.1 Feature detection and description in 3D US data

The FAST feature-detector detects keypoints by means of a sampling-mask, which is illustrated in Fig. 1a, and intensity comparisons between the red marked central voxel p and nearby voxels defined by the sampling-mask. In a first step, keypoint candidates are selected by comparing the intensity of the central voxel I_p and the intensities $I_{p \rightarrow x}$ of the six voxels $x \in \{1..6\}$ on the mask that lie on the space axes in positive and negative direction relative to p . In Fig. 1a, these voxels are marked green. The comparison is done by

$$S_{p \rightarrow x} = \begin{cases} d, & I_{p \rightarrow x} \leq I_p - t & \text{(darker)} \\ s, & I_p - t < I_{p \rightarrow x} < I_p + t & \text{(equal)} \\ b, & I_p + t \leq I_{p \rightarrow x} & \text{(brighter)}, \end{cases} \quad (1)$$

where t is a threshold that defines up to which intensity difference voxels are considered as "equal". This parameter is determined empirically for validation and was chosen as $t = 30$. If at least five voxels are categorized as "darker" or "brighter", the central voxel is stored as a keypoint candidate. In a second step, the keypoint candidates are analyzed by executing the comparison for all nearby voxels defined by the mask. A candidate is defined as a keypoint if the majority of nearby voxels are categorized as "brighter" or "darker" respectively.

BRISK is a binary feature-descriptor that builds a feature vector by intensity comparisons. The voxels to be compared are defined by a mask, illustrated in Fig. 1b, and a predefined pair sequence $\mathcal{A}$. This mask consists of three nested bullets with the keypoint in the center. The nearby voxels defined by the mask are numbered. On the basis of these numbered voxels, pairs (p_j, p_i) are built and split into shortpairs $\mathcal{S}$ and longpairs $\mathcal{L}$. For this, thresholds define up to which Euclidean Distance a pair is a shortpair and from which distance a pair is a

Fig. 2: Slices of the annotated 3D targets of all three subjects.

longpair. For validation, these thresholds are determined empirically and were chosen as $\delta_{max} = 13$ and $\delta_{min} = 15$. Furthermore, the pair-sequence is generated randomly so that 68 longpairs and 198 shortpairs are created. The longpairs are used to determine a mean orientation by

$$g = \begin{pmatrix} g_x \\ g_y \\ g_z \end{pmatrix} = \frac{1}{L} \cdot \sum_{(p_j, p_i) \in \mathcal{L}} g(p_j, p_i), \quad (2)$$

where L is the number of longpairs and $g(p_j, p_i)$ is the local gradient of a longpair that is computed as

$$g(p_j, p_i) = (p_j - p_i) \cdot \frac{I(p_j, \sigma_j) - I(p_i, \sigma_i)}{\|p_j - p_i\|^2}. \quad (3)$$

Here, $I(p_j, \sigma_j)$ is the Gauss-filtered voxel intensity of p_j . To obtain rotational invariance, the mask is rotated corresponding to this orientation by determining an angle α using $\arctan2$ function. The Gauss-filtered voxel intensities of the rotated shortpairs $I(p_i^\alpha, \sigma_i)$ are used to build the feature vector by

$$b = \begin{cases} 1, & I(p_j^\alpha, \sigma_j) > I(p_i^\alpha, \sigma_i) \\ 0, & \text{otherwise} \end{cases} \quad \forall (p_j^\alpha, p_i^\alpha) \in \mathcal{S}. \quad (4)$$

2.2 US data acquisition and target annotation

Validation of the tracking method is executed with 4DUS sequences of the liver of three subjects. The sequences were acquired at an imaging frequency of 3.5 Hz, 4.0 Hz and 4.5 Hz, respectively, and have a length of 30 min. In each 4DUS sequence, a training sequence and a test sequence with a length of 1:30 min is selected such that a sequence contains about 20 breathing cycles. Training and validation sequence were 11:00 min apart.

The US sequences show parts of the liver of healthy subjects where vessel bifurcations are defined as the target that has to be tracked. The shape of the 3D targets are annotated manually in every fifth volume in the training and test sequences. In Fig. 2, a slice of each annotated target is illustrated. The annotation was done by setting points at the target boundary in all slices of the volume covering the target. The resulting point clouds are converted to binary volumes by defining the space that is bordered by the point cloud as the target volume.

Fig. 3: All binary volumes of a feature class are aligned to the feature and summed. Determining the mean in each voxel results in the mean-target-volume.

2.3 Feature library generation

The defined 3D target has to be tracked in 4DUS sequences. For this, features nearby the target are used. These features are detected in the training sequences and their positions relative to the annotated target are stored separately in a feature library for each subject. In all volumes of a training sequence, features are detected and compared to each other. If features from different volumes are similar, these features are sorted into the same feature class in the feature library. Because binary features are used, two features are compared by using the hamming distance. If the hamming distance between two features is less than an empirically selected threshold, the features lie in the same anatomical structure in different US volumes so that they are sorted into the same feature class. Thus, a feature class contains the same feature out of different volumes of a 4DUS sequence and the feature library contains several feature classes.

In each feature class, a mean target volume (MTV) is determined so a mean relation between the features and the target is computed. In all feature classes an annotated binary target volume for each detected feature is stored. As illustrated in Fig. 3, these target volumes are aligned on the basis of the corresponding features. In each voxel the mean is determined over all binary volumes. With the feature library, it is possible to estimate the target position and shape.

During tracking in a new US volume, features are detected and classified into the feature library. The classification is done by a basic random forest classifier (RFC) that is created with the feature vectors stored in the feature library. It consists of 10 decision trees and is implemented in Matlab. Due to the automatic feature library generation the numbers of feature classes and features per class are variable. However, a feature class contains several features so the RFC is generated by using the binary feature vectors for training. Features that are recognized by the feature library using the RFC are used to generate a tar-

get estimation. The mean feature target relations stored in the feature library are used to estimate the target by averaging the MTVs of all recognized features in the US volume.

2.4 Validation

The 4DUS training sequences are used to generate subject specific feature libraries. These are used to track the targets in the annotated test sequences. For each annotated volume of the test sequences, target estimations are determined and compared with the annotations. Both a 3D position error as the Euclidean Distance between the centers of the annotated volume and the target estimation and the dice coefficient are determined to quantify the accuracy of position and shape of the estimation.

3 Results

Before evaluating the tracking method by tracking the target in the test sequences of the three subjects, the motion of the annotations in these sequences are analyzed. The target center is defined as the current target position. The motion range of the annotated target in the test sequence as well as the mean shift of the target between the training and test sequence are presented in Tab. 1. In the test sequence, the target moves in

Tab. 1: Range of the target motion in the test sequence and the target position shift between training and test sequence.

subject	motion range (x,y,z) / mm			shift / mm
A	3.06	9.46	16.41	2.14
B	12.09	23.82	35.77	10.01
C	5.40	9.48	14.23	4.56

differing magnitudes, depending on the direction. Due to the temporal separation of training and test sequence the target had the potential to shift.

In each volume of the test sequence, the target is tracked and a target estimation is determined. For all volumes in the test sequence, a target estimation could be determined, so there were no tracking failures. As illustrated in Fig. 4, there are differences between the annotation and the estimation. The mean position error and dice coefficient for all subjects are determined and presented in Tab. 2. A smaller position error indicates a more accurate position estimation on average and a high dice coefficient indicates a higher similarity between annotation and estimation. The used feature library of subject A had 92 feature classes (B: 105, C: 46). On average, with this

Fig. 4: Comparison of the annotated boundary (green) and the estimated boundary (blue) of the target in a slice of a US volume.

Tab. 2: Mean position and volumetric error between target estimation and annotation of subjects A, B and C.

subject	position error / mm	dice coefficient / %
A	1.38 ± 0.67	82.7 ± 3.7
B	2.42 ± 0.87	78.1 ± 3.7
C	2.35 ± 0.95	75.9 ± 7.6
Average	2.05	78.9

library in the test sequence of subject A, 24 features were detected (B: 24, C: 14). The mean position error over all subjects is 2.05 mm, the mean dice coefficient is 78.9 %. The error in subject B is larger than the errors of A and C. However, the target shift and motion amplitude also are higher compared to A and C (Tab. 1).

4 Discussion

The aim of this study was to analyze if features in 4DUS can be used to track deformable targets in the liver during free breathing. Using a feature library, the target could be detected even if there is a time span of 11:00 min between the training and test phase. Relating to radiation therapy, this time span is realistic, because the acquisition of training data can be done days or weeks before treatment. Furthermore, a large target shift of 10 mm could be measured, so errors in patient positioning on the treatment table of this magnitude could be compensated. Though the target moves 35.77 mm at maximum in the test sequence, the target position could be estimated with an error between 1.35 mm and 1.42 mm. In comparison to the results of [5] and [6], the results in this study are in the same order of magnitude. Additionally, it was shown that it is possible to generate a 3D target estimation based on the detected features. A high correspondence between the estimation and annotation of the target was observed. In contrast, [5] and [6] only evalu-

ated the position of landmarks with position of the target estimation. The results in our study indicate that translation, rotation and deformation of a moving target can be detected by the use of features in 4DUS data, making it a promising method for robust target localization in 4DUS data.

However, there is still potential to improve the method. Parameters that were chosen empirically need to be analyzed to possibly determine them automatically. Furthermore, so far only the feature library is used to localize the target. Tissue variations that did not occur in the training sequence can lead to errors or detection loss during the tracking process. This could be addressed by inserting an online component that handles unknown tissue variations.

Author Statement

Research funding: Parts of this work were supported by the German Federal Ministry of Education and Research (grant no. 13GW0228B). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration. The data was obtained for another study and has been approved by the authors's review board.

References

- [1] Wannemacher M, Wenz F, Debus J. Strahlentherapie. 2nd ed. Berlin: Springer; 2013
- [2] Schweikard A, Glosser G, Bodduluri M, Murphy MJ, Adler JR. Robotic motion compensation for respiratory movement during radiosurgery. *Comput Aided Surg.* 2000;5(4):263-77
- [3] Wunderink W, Méndez Romero A, Seppenwoolde Y, de Boer H, Levendaq P, Heijmen B. Potentials and limitations of guiding liver stereotactic body radiation therapy set-up on liver-implanted fiducial markers. *Int J Radiat Oncol Biol Phys.* 2010;77(5):1573-83
- [4] Shen D, Zhan Y, Davatzikos C. Segmentation of prostate boundaries from ultrasound images using statistical shape model. *IEEE Transactions on Medical Imaging.* 2003;22(4):539-551
- [5] Banerjee J, Klink C, Peters ED, Niessen WJ, Moelker A, van Walsum T. Fast and robust 3D ultrasound registration - block and game theoretic matching. *Medical Image Analysis* 2015;20:173-183
- [6] Royer L, Marchal M, Bras AL, Dardenne G, Krupa A. Real-time target tracking of soft tissues in 3D ultrasound images based on robust visual information and mechanical simulation. *Medical Image Analysis* 2017;35:582-598
- [7] Rosten E, Drummond T. Machine learning for high speed corner detection. *ECCV;* 2006;430-443.
- [8] Leutenegger S, Chili M, Seigwart RY. BRISK: Binary robust invariant scalable keypoints. *ICCV* 2011;2548-2555