Markerless estimation of patient orientation, posture and pose using range and pressure imaging

Robert Grimm, Sebastian Bauer, Johann Sukkau, Joachim Hornegger & Günther Greiner

International Journal of Computer Assisted Radiology and Surgery
A journal for interdisciplinary research, development and applications of image guided diagnosis and therapy

ISSN 1861-6410

Int J CARS
DOI 10.1007/s11548-012-0694-5

Springer

Your article is protected by copyright and all rights are held exclusively by CARS. This e-offprint is for personal use only and shall not be self-archived in electronic repositories.
If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Springer

Author's personal copy

Int J CARS
DOI 10.1007/s11548-012-0694-5

ORIGINAL ARTICLE

Markerless estimation of patient orientation, posture and pose using range and pressure imaging

For automatic patient setup and scanner initialization in tomographic imaging

Robert Grimm · Sebastian Bauer · Johann Sukkau · Joachim Hornegger · Günther Greiner

Received: 8 February 2012 / Accepted: 18 April 2012
© CARS 2012

Abstract
Purpose In diagnostic tomographic imaging, patient setup and scanner initialization is a manual, tedious procedure in clinical practice. A fully-automatic detection of the patient's position, orientation, posture and pose on the patient table holds great potential for optimizing this part of the imaging workflow. We propose a markerless framework that is capable of extracting this information within seconds from either range imaging (RI) or pressure imaging (PI) data.
Methods The proposed method is composed of three stages: First, the position and orientation of the reclined patient are determined. Second, the patient's posture is classified. Third, based on the estimated orientation and posture, an approximate body pose is recovered by fitting an articulated model to the observed RI/PI data. Being a key issue for clinical application, our approach does not require an initialization pose.
Results In a case study on real data from 16 subjects, the performance of the proposed system was evaluated quantitatively with a 3-D time-of-flight RI camera and a pressure sensing mattress (PI). The patient orientation was successfully determined for all subjects, independent of the modality. At the posture recognition stage, our method achieved mean classification rates of 79.4 % for RI and 95.5 % for PI data, respectively. Concerning the approximate body pose estimation, anatomical body landmarks were localized with an accuracy of ±5.84 cm (RI) and ±5.53 cm (PI).
Conclusions The results indicate that an estimation of the patient's position, orientation, posture and pose using RI and PI sensors, respectively, is feasible, and beneficial for optimizing the workflow in diagnostic tomographic imaging. Both modalities achieved comparable pose estimation results using different models that account for modality-specific characteristics. PI outperforms RI in discriminating between prone and supine postures due to the distinctive pressure distribution of the human body.

Keywords Range imaging · Pressure imaging · Diagnostic tomographic imaging · Posture classification · Pose estimation

Introduction

In the past decade, much progress has been made in improving and automating the image acquisition process in computed tomography (CT) and magnetic resonance imaging (MRI). However, the permanent increase in complexity of medical imaging technologies is accompanied by a rise in costs for the device itself, its maintenance, and the staff needed for patient setup and image acquisition. In this regard, the automation and optimization of examination workflows holds great potential to compensate for the growing expenses. In diagnostic tomographic imaging, patient setup and scanner initialization including patient positioning, table adjustments, and the input of patient-specific examination parameters into the scanner software are still done manually. The automation of these steps would reduce both the examination time and the workload for highly

R. Grimm (✉) · S. Bauer · J. Hornegger
Pattern Recognition Lab, Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander-Universität Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, Germany
e-mail: robert.grimm@informatik.uni-erlangen.de; robert.grimm@cs.fau.de

J. Sukkau
Siemens AG, Healthcare Sector, Allee am Röthelheimpark 2, 91052 Erlangen, Germany

R. Grimm · G. Greiner
Chair of Computer Graphics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstr. 11, 91058 Erlangen, Germany

Published online: 15 May 2012

Author's personal copy

trained technical and radiological staff. As an example, the pre-imaging protocol for routine MRI examinations usually comprises the following steps. First, the patient orientation (head first, feet first) and posture (left, right, prone, supine) have to be specified along with additional biometric information such as body height and weight. Among others, this information is required to ensure patient safety and the correct labeling of the images to be acquired. These data are then entered by hand into the patient registration dialog of the scanner software. Second, the radiologist manually defines the center of the region of interest on the patient's body, typically using a laser crosshair mounted at the front panel of the scanner.

In this paper, we propose a markerless system that supersedes a good portion of the manual input of patient data that are required by the scanner software. Furthermore, it can be used to automatically localize a given examination site and set the appropriate region of interest. As a consequence, the tedious procedure of patient setup and scanner initialization may be reduced to a single mouse click, specifying the examination type and site.

This requires a fully automatic analysis of the patient's *orientation* and *posture* that can then be transferred seamlessly to the patient registration dialog of the scanner. In addition, in order to supersede the use of laser crosshairs for setting the examination site and region of interest, our method estimates the approximate body *pose* of the patient, yielding the coarse position of anatomical landmarks.

For the purpose of estimating the body pose of a reclined patient lying on the scanner table, we propose a model-based optimization framework that is generic in the sense that it can process data from two sensor modalities of fundamentally different nature using virtually the same algorithmic base. In an experimental study on real data from 16 subjects, we investigated the fitness of range imaging (RI) and pressure imaging (PI) devices for the automation of patient setup and scanner initialization. In particular, we compared the performance of a 3-D time-of-flight (ToF) range imaging camera to a pressure sensing mattress.

Among others, using a PI device is motivated by the fact that it can handle cases of loose clothing and visual occlusion of the patient (e.g., being covered by a blanket) that might impair recognition performance when using optical sensors.

Related work

A markerless system that is capable of detecting the patient's position, orientation, posture and pose in a fully automatic manner is of great value for the optimization of workflows and the improvement of patient safety in various clinical contexts, for example, in diagnostic and interventional imaging, radiation therapy, computer-assisted treatment, and sleep monitoring. As the topic of this paper is the automation of patient setup and scanner initialization in diagnostic tomographic imaging, we will focus our discussion of related work on this subfield.

The use of a low-resolution MR prescan in order to detect the patient's position and orientation was proposed by Keil et al. [1] for automatic optimization of the specific absorption rate. Here, a *move-during-scan* protocol was employed to acquire low-resolution MR data during the initial movement of the patient table into the scanner. Wachinger et al. [2] recently improved the original approach [1] by replacing principal component analysis-based dimensionality reduction with a manifold learning technique.

Fenchel et al. [3] employed MR localizer data acquired during continuous movement of the patient table for automatic MR examination planning. In particular, the authors proposed a method for fast and automatic labeling of anatomical structures and body landmarks using a statistical atlas. However, today's clinical workflow still requires patient-specific information such as the posture to be specified *before* the prescan.

Furthermore, we also address the automation of the MR prescan itself, and the workflow of radiographic tomographic imaging (e.g., CT), where a whole-body prescan is unfeasible with respect to the additional radiation dose.

Optical or range imaging sensors are a common choice for human pose estimation. In the past decade, a variety of approaches have been proposed for recovering human pose from visual observations, for a comprehensive survey see Moeslund et al. [4] and Poppe [5]. However, the corpus of related work is focused on situations where the subject stands upright.

Markerless solutions for clinical and biomechanical applications are addressed only rarely [6,7]. Here, in contrast to accurate marker-based approaches, no interaction with the patient is required. This is particularly important, since the placement of markers would impose an additional task in the workflow.

Articulated pose estimation on 3-D surface data from low-cost RI sensors has recently gained more interest [8–11]. A fundamental drawback of many model-based approaches is the necessity of a specific body pose for model initialization [9–11] which is usually an unacceptable interaction for clinical scenarios. Recently, a ToF-based system that coarsely detects anatomical regions with an ordinary golden ratio division was proposed by Schaller et al. [12]. However, the heuristic algorithm cannot cope with the variety of articulated real-world patient poses. In terms of accuracy, existing ToF-based pose estimation systems have reported errors of 1–9 cm [11] and 2–6 cm [8], respectively, for the upper body pose of a subject standing upright and facing the camera.

In the field of pressure imaging, Seo et al. [13] proposed to use an array of pressure sensors to classify the posture of

Author's personal copy

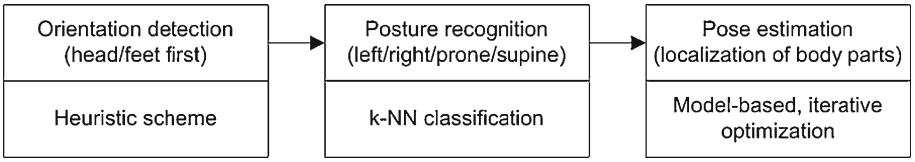

Fig. 1 Flowchart of the proposed framework, from *left* to *right*. The orientation, posture, and body pose of the patient are extracted in a consecutive manner. Note that this pipeline can be applied to both RI and PI data

a human reclining on a bed. Without specifying the dimension of the evaluation data set, the authors state an accuracy of 93.6 % for classification with a radial basis function neural network. Harada et al. [14] used a PI-based approach for posture and pose classification based on template matching with a training database that was established by estimating the pressure distribution of a synthetic human body model.

Later, the authors also proposed a motion tracking system with a pressure imaging device [15]. However, a manual pose initialization is required with these approaches.

Materials and methods

We propose a markerless framework that is capable of classifying the orientation and posture of a patient. Furthermore, it delivers an estimate of the articulated body pose. The system is generic in the sense that it can be fed with sensor data from both range and pressure imaging modalities. Our framework consists of three top-down algorithmic stages, see Fig. 1. In the first stage, the patient's orientation is detected using a heuristic scheme. Subsequently, in the second stage, we perform a k-nearest neighbor (k-NN) classification for determination of the patient posture. Third and based on the detected orientation and posture, the approximate patient pose is estimated by iteratively fitting an articulated body model to the observed RI and PI data, respectively. Note that each stage benefits from prior knowledge of the previous stage. Unlike many optical systems for human–computer interaction, our approach does not require the subject to adapt an initialization pose.

Sensor modalities

Time-of-flight range imaging measures 3-D surface information with a single sensor based on the phase shift between an actively emitted and its reflected optical signal [16]. The device is compact, easy to integrate and delivers metric information in real-time (25 Hz). The precision in z-direction is in the scale of 5 mm (1 standard deviation) for an operating distance of 1 m. Let us represent the measurement data of the sensor with a resolution of $w \times h$ pixels by a point cloud $\boldsymbol{D} \in \mathbb{R}^{w \times h \times 3}$,

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{d}_{1,1} & \dots & \boldsymbol{d}_{1,h} \\ \vdots & \ddots & \vdots \\ \boldsymbol{d}_{w,1} & \dots & \boldsymbol{d}_{w,h} \end{bmatrix}, \tag{1}$$

where $\boldsymbol{d}_{i,j}$ denotes the measured 3-D point $(d_x, d_y, d_z)^{\mathrm{T}}$ at the sensor element (i, j). The depth component d_z denotes the orthogonal distance between sensor and object. Subtracting the depth component of the measured data $\boldsymbol{D}$ from a previously acquired background model $\boldsymbol{B} = \{\boldsymbol{b}_{i,j}\}$ yields $\boldsymbol{I}_{\mathrm{D}} \in \mathbb{R}^{w \times h}$, representing the 2-D height profile of the patient above the table plane:

$$\boldsymbol{I}_{\mathrm{D}}(i, j) = b_z(i, j) - d_z(i, j) \tag{2}$$

Exemplary RI height maps from the data set used for evaluation are depicted in Fig. 4a.

Pressure sensing mattresses consist of an array of flexible pressure sensors embedded into a thin mattress. At each sensor node at coordinates (i, j), the applied force $\boldsymbol{I}_{\mathrm{F}}(i, j)$ is measured in metric units, examples are shown in Fig. 4b. To date, PI devices are commonly used for visualization and qualitative analysis of surface pressure distributions only. However, they also offer great potential to the computer vision and pattern recognition community, as they allow to capture information about the pressure exerted on the sensor surface in real time. Even though off-the-shelf products as the one used in our experiments ("Experiments and results" section) are not compatible with MR scanners, the required components to build suitable sensor systems are available.

Orientation detection

In order to determine the orientation of the patient (head first, feet first), we apply a heuristic scheme that analyzes the distribution of the RI and PI data, respectively. In particular, the algorithm evaluates the 1-D profiles of the patient's height and weight along the two orthogonal directions of the patient table. Here, we exploit the fact that the RI profiles along the body axes typically exhibit a high concentration of volume in the torso region, while PI profiles feature spots of high pressure about the shoulders and hips. Based on these indicators for localizing the position of the torso region relative

Author's personal copy

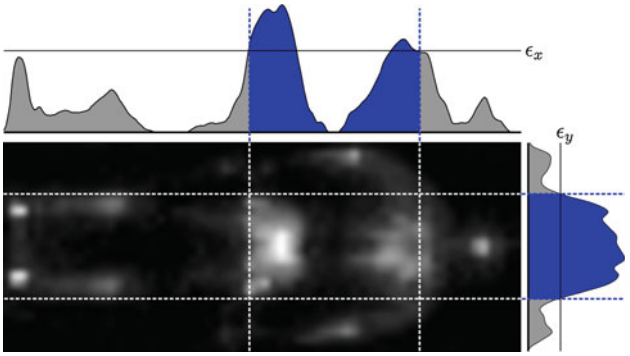

Fig. 2 Detection of the patient orientation (head first, feet first) on PI data. The weight distributions along the principal axes are shown as a function on *top* and on the *right* of the pressure image. The torso region (labeled in *blue*) is detected using dynamic thresholds (ϵ_x, ϵ_y)

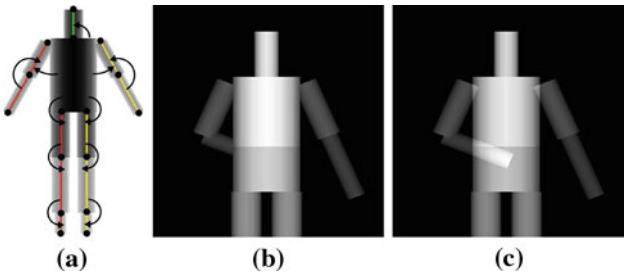

Fig. 3 **a** *Frontal view* of the articulated 3-D body model and joint angles $\boldsymbol{\delta}$ used for RI-based pose estimation. For application on PI data, this model is enhanced using additional geometry according to Table 1. **b** Conventional depth sampling based on z-buffer. **c** Our modified sampling scheme, accounting for self-occlusions that can occur due to the restriction of the 3-D body model to planar joint angles

to the patient body and position on the table, the orientation of the patient can be deduced using dynamic thresholds (ϵ_x, ϵ_y) that are determined in relation to the individual depth and weight distribution statistics of the patient, respectively, see Fig. 2.

Posture classification

A robust classification of the patient's posture on the examination table is of two-fold importance. First, it is a mandatory input during patient registration. Second, it is an essential prior information for initialization of the body model for pose estimation, see "Model-based pose estimation" section. In our framework, we distinguish between the four basic postures (prone, supine, left, right) with a k-nearest neighbor classification. In particular, the data $(\boldsymbol{I}_{\mathrm{D}}, \boldsymbol{I}_{\mathrm{F}})$ is normalized by aligning its center of mass to coincide with the origin. The normalized data then directly serves as a feature vector that is classified using the L2 norm as distance metric. A measurement is assigned the posture class of the k most similar training instances.

Table 1 List of anatomic regions where additional weights were used for augmenting the pressure model in order to obtain a realistic pressure distribution, for the four postures

Supine	Prone	Left	Right
Head	Head	Head	Head
Elbows	Upper arms	Upper arm (l)	Upper arm (r)
Shoulders	Shoulders	Upper leg (l)	Upper leg (r)
Lower back	Lower back	Calf (l)	Calf (r)
Calves	Knees	Foot (l)	Foot (r)
Heels	Feet		

Model-based pose estimation

Based on the known orientation and posture, our framework recovers the approximate body pose by fitting an articulated model to the observed data. Here, a 3-D kinematic model of the human body, composed of elliptic cylinders (see Fig. 3a), is used to generate synthetic depth maps and pressure distributions to match the RI or PI sensor measurements, respectively. In the PI case, the model geometry resembles the true shape of the body parts. Accounting for the characteristics in PI data, the pressure model is augmented with additional geometry of heuristically determined dimensions to simulate a realistic pressure distribution on the patient table, see Table 1 for details.

The 15 degrees of freedom of the model configuration are expressed by a vector $\boldsymbol{\theta} = (\boldsymbol{t}, \phi, s, \boldsymbol{\delta}) \in \mathbb{R}^{15}$, including a global translation $\boldsymbol{t} \in \mathbb{R}^2$, a global rotation angle ϕ, a scale factor s accounting for the individual body proportions, and a vector of Euler angles $\boldsymbol{\delta} \in \mathbb{R}^{11}$, representing the joint angles in the plane parallel to the patient table, see Fig. 3a. For an automatic localization of the treatment site and region of interest, the 2-D patient pose and articulation with respect to the table plane holds sufficient information. Hence, we restrict the articulation of our 3-D model to the planar domain. As a side-effect, this reduces the dimensionality of the pose search space. The effect of perspective foreshortening of extremities is negligible, since the non-planar angular component is usually not relevant to the use-cases considered here. Furthermore, the angles are limited to ranges that are likely to be encountered in the clinical context.

Model sampling In order to project the model configuration into the observation space, a depth map of the 3-D model representation is computed using OpenGL rendering. From a given model configuration, either type of sensor data can be emulated simply by using the corresponding model geometry. The model parameterization itself and the sampling scheme are identical for both RI- and PI-based pose estimation. With a conventional orthographic projection model (Fig. 3b), the z-buffer holds a map of the orthogonal distance to the surface of the closest limb. However, for a 3-D

Author's personal copy

model that is restricted to planar articulation, this sampling scheme would not account for cases where different parts of our cylindrical model intersect or coincide.

Hence, we use a modified sampling scheme that resolves self-occlusions into a realistic model surface, regardless of the restriction to planar joint angles, see Fig. 3c. For this purpose, individual body parts are rendered separately in a subsequent manner. Interpreting their z-buffer representation yields a thickness information $t_{\boldsymbol{\theta}}^{(l)}(i, j)$ for each body part l and position (i, j) with respect to the model configuration $\boldsymbol{\theta}$. Note that the thickness information corresponds to the path length through the body part model at the given location. In order to obtain the complete model sampling $\boldsymbol{M}_{\boldsymbol{\theta}} \in \mathbb{R}^{w \times h}$, the thickness information of the set of body parts is accumulated for each pixel position:

$$\boldsymbol{M}_{\boldsymbol{\theta}}(i, j) = \sum_{l} t_{\boldsymbol{\theta}}^{(l)}(i, j), \tag{3}$$

where the index $\boldsymbol{\theta}$ denotes the body model configuration.

The goal now is to find the model configuration $\hat{\boldsymbol{\theta}}$ that minimizes the difference $F_{\boldsymbol{I}}(\boldsymbol{\theta})$ between the generated model data $\boldsymbol{M}_{\boldsymbol{\theta}}$ and the actually observed data $\boldsymbol{I}$ of the corresponding sensor modality:

$$\hat{\boldsymbol{\theta}} = \arg_{\boldsymbol{\theta}} \min F_{\boldsymbol{I}}(\boldsymbol{\theta}) = \arg_{\boldsymbol{\theta}} \min \sum_{i,j} \left(\boldsymbol{M}_{\boldsymbol{\theta}}(i, j) - \boldsymbol{I}(i, j)\right)^2, \tag{4}$$

where $\boldsymbol{I}$ may be either the applied force $\boldsymbol{I}_{\mathrm{F}}$ with PI data or depth information $\boldsymbol{I}_{\mathrm{D}}$ with RI data.

Pose optimization Our pose optimization routine is based on a genetic algorithm [17] that holds of a pool of genotypes and repeatedly applies two stages: selection and recombination (or reproduction). In the selection stage, a fitness function determines which of the individuals in the pool are kept for the next iteration. Genetic recombination is performed by generating a set of offsprings from the current pool of genotypes. Let us note that genetic algorithms do not require an analytic derivative of the objective function for minimization.

In our case, a genotype corresponds to a model configuration vector $\boldsymbol{\theta}$, as outlined in Algorithm 1. Only the fittest single individual $\hat{\boldsymbol{\theta}}$ is selected for reproduction. The fitness of an individual $\boldsymbol{\theta}$ is inversely proportional to the corresponding objective function value $F_{\boldsymbol{I}}(\boldsymbol{\theta})$. $\hat{\boldsymbol{\theta}}$ is initialized with a default pose, depending on the posture estimate. As stated before, each individual degree of freedom j of the model is restricted to values within a reasonable range $[r_j^{\min}, r_j^{\max}]$. They are initialized at the maximal possible range for each degree of freedom, see Algorithm 1. During the optimization process, these search ranges will be narrowed increasingly.

The subroutine `genOffsprings`$(\hat{\boldsymbol{\theta}}, j, r_j^{\min}, r_j^{\max}, K)$ generates a fixed number of K instances of model configurations $\{\boldsymbol{\theta}^k, k = 1, \ldots, K\}$ to be considered for evaluation. Note that these offsprings differ only in the values of the j-th degree of freedom. In each of the N_{dofs} inner iterations, a different single degree of freedom is manipulated. The evolutionary process is repeated N_{iter} times over all degrees of freedom. In general, the degrees of freedom correspond to the degrees of freedom in $\boldsymbol{\theta}$. However, it proved to be more efficient to change two components at a time for the lateral postures, where the angles of hip and knees are adjusted simultaneously rather than individually.

Algorithm 1 Genetic Algorithm for pose estimation

Input: Initial pose $\hat{\boldsymbol{\theta}}$
Input: Maximal ranges $[r_j^{\min}, r_j^{\max}] \forall\, j = 1, \ldots, N_{\mathrm{dofs}}$
for $i = 1, \ldots, N_{\mathrm{iter}}$ **do**
  for $j = 1, \ldots, N_{\mathrm{dofs}}$ **do**
    $\{\boldsymbol{\theta}^k\} \Leftarrow$ `genOffsprings`$(\hat{\boldsymbol{\theta}}, j, r_j^{\min}, r_j^{\max}, K)$
    for $k = 1, \ldots, K$ **do**
      if $F_{\boldsymbol{I}}(\boldsymbol{\theta}^k) < F_{\boldsymbol{I}}(\hat{\boldsymbol{\theta}})$ **then**
        $\hat{\boldsymbol{\theta}} \Leftarrow \boldsymbol{\theta}^k$
      end if
    end for
    $r_j^{\min} \Leftarrow \hat{\theta}_j - 0.7 \cdot (\hat{\theta}_j - r_j^{\min})$
    $r_j^{\max} \Leftarrow \hat{\theta}_j + 0.7 \cdot (r_j^{\max} - \hat{\theta}_j)$
  end for
end for
Output: Estimated pose $\hat{\boldsymbol{\theta}}$

The optimization of configuration vector components is controlled by incorporating prior knowledge on the hierarchy of human kinematics: First, the global translation and rotation are optimized. Then, proceeding to the extremities, the position and orientation of the upper arms and legs, the head, the lower arms, and eventually the lower legs are estimated. An adaptive annealing sampling scheme in the parameter domain enforces convergence of the optimization: In the first iteration, the individuals are generated by uniformly sampling the maximum parameter range $[r_j^{\min}, r_j^{\max}]$. In the subsequent iterations, the evaluation is restricted to a dense sampling of 70 % of the local range around the last estimate for this parameter, $\hat{\theta}_j$, as the optimal value is expected to be close to the previous guess. No explicit collision detection or avoidance is required to penalize model configurations where body parts intersect, owing to our dedicated sampling scheme described in Eq. 3.

Experiments and results

RI/PI evaluation data set

The proposed method is evaluated on real data covering full body poses from 16 reclined subjects. For the experiments on RI data, a PMD S3 ToF sensor (PMD Technologies GmbH, Germany) with a resolution of 64 × 48 pixels was mounted on a linear motion slide and moved over the patient parallel to the examination table. The depth images were stitched

Author's personal copy

Table 2 Top: cumulative confusion matrix of k-NN posture classification for experiments on RI data, from four-fold inverse cross-validation (using data from four volunteers as training set and testing on data from the twelve remaining subjects), and bottom: cumulative confusion matrix for experiments on PI data

	Supine	Prone	Left	Right
RI classification				
Supine	**44**	13	0	0
Prone	19	**32**	0	0
Left	2	3	**42**	1
Right	2	2	0	**44**
PI classification				
Supine	**52**	0	2	0
Prone	2	**48**	0	1
Left	3	0	**44**	1
Right	0	0	0	**48**

In this table, rows denote the ground truth label and columns the respective classification result. Boldface values indicate fields where the classification matched the true label, i.e. these are numbers of correct classifications

with a cross-correlation scheme delivering images of about 64×190 pixels. Thus, the effective resolution lies in the range of state-of-the-art ToF sensors that feature a resolution of up to 200×200 pixels. At a distance of 2.5 m above the patient table, the spatial resolution is 1.1 cm in x- and y-direction. For evaluation of PI data, an XSensor X3 PX100:26.64.01 pressure sensing mattress (XSensor Technology Corporation, Canada) was placed between table and subject. Its array of 26×64 capacitive sensors provides a spatial resolution of 3.175 cm. Using the RI and PI sensor simultaneously, we have acquired a comprehensive database of 68 poses from 16 healthy volunteers. At least one ToF scan and a pressure measurement in each fundamental posture were acquired per subject. In one case, the PI data were not usable due to sensor mis-calibration.

Experimental results

For both types of sensor data, our heuristic scheme detected the correct orientation (head first, feet first) for all subjects.

For evaluation of the k-NN classification of the patient's posture, we performed a fourfold *inverse* cross-validation [18]. Here, first, the data are partitioned into four folds. Second, iterating over the folds, training is performed on one fold and predicting on data of the remaining three. Applying cross-validation in this inverse fashion accounts for the fact that in practice the training set is significantly smaller than the testing set.

Table 2 shows the cumulative confusion matrix after the four cross-validation runs, for a heuristically determined value of $k = 5$. With a mean classification rate of 95.5 %, the pressure sensing mattress outperformed the RI-based setup (79.4 %). Concerning the inferior performance of RI data, please note that several mismatches occurred between prone and supine postures. In particular, for the classification of prone and supine postures, we observed mean classification rates of 95.2 % for PI and 70.0 % for RI data. This is due to the fact that these postures exhibit a similar 3-D shape. In contrast, the superior performance of PI data results from the distinct pressure distributions in prone and supine position.

Exemplary qualitative results of our body pose estimation stage are shown in Fig. 4, on RI (a) and PI data (b), demonstrating the robustness of the algorithm on real data. After $N_{\text{iter}} = 5$ optimization cycles of pose estimation (approx. 600 objective function evaluations), reliable locations for the head and torso, as well as for the extremities in virtually all cases, are found. In each inner iteration in Algorithm 1, a set of $K = 8$ offsprings was generated and evaluated. The whole process takes less than 5 s for our single-threaded implementation on a 2.26 GHz CPU system with an ATI Radeon HD 3400 graphics card.

For quantitative analysis, the metric coordinates of the estimated anatomical landmarks are compared to ground truth data that were manually labeled in the measurements by two observers following a predefined convention based on the joint positions defined in the body model. To avoid bias of the root mean square (RMS) error, large outliers were not taken into account for the computation of the average localization errors. In particular, landmark estimations with a planar Euclidean distance exceeding 15 cm were rejected as outliers. The remainder was considered as correctly detected. Table 3 shows the resulting localization errors for clinically relevant landmarks. The average RMS errors over all landmarks are 5.84 and 5.53 cm for RI and PI data, respectively. The detection rates are listed in Table 4, with mean rates of 96.6 % (RI) and 95.0 % (PI). The characteristic outlines of head and shoulders in RI data result in high detection rates. Similarly, high pressure is exerted on the PI mattress by the head and shoulders in supine position, and by the hips and knees in all postures, also leading to almost perfect detection. In contrast, a lower accuracy was observed for the crest, neck and shoulder landmarks in prone and lateral posture in PI data. This is due to the fact that some of the subjects rested their head on their arms, leading to an ambiguous pressure distribution in these regions.

Overall, our results indicate that an estimation of the patient's approximate body pose is feasible using RI and PI sensors. The achieved accuracy of our method is in the same order as previously reported results [8,13,11], see "Related work" section, for both types of sensors. However, our approach does not require an initialization pose, being a key issue for clinical application.

Author's personal copy

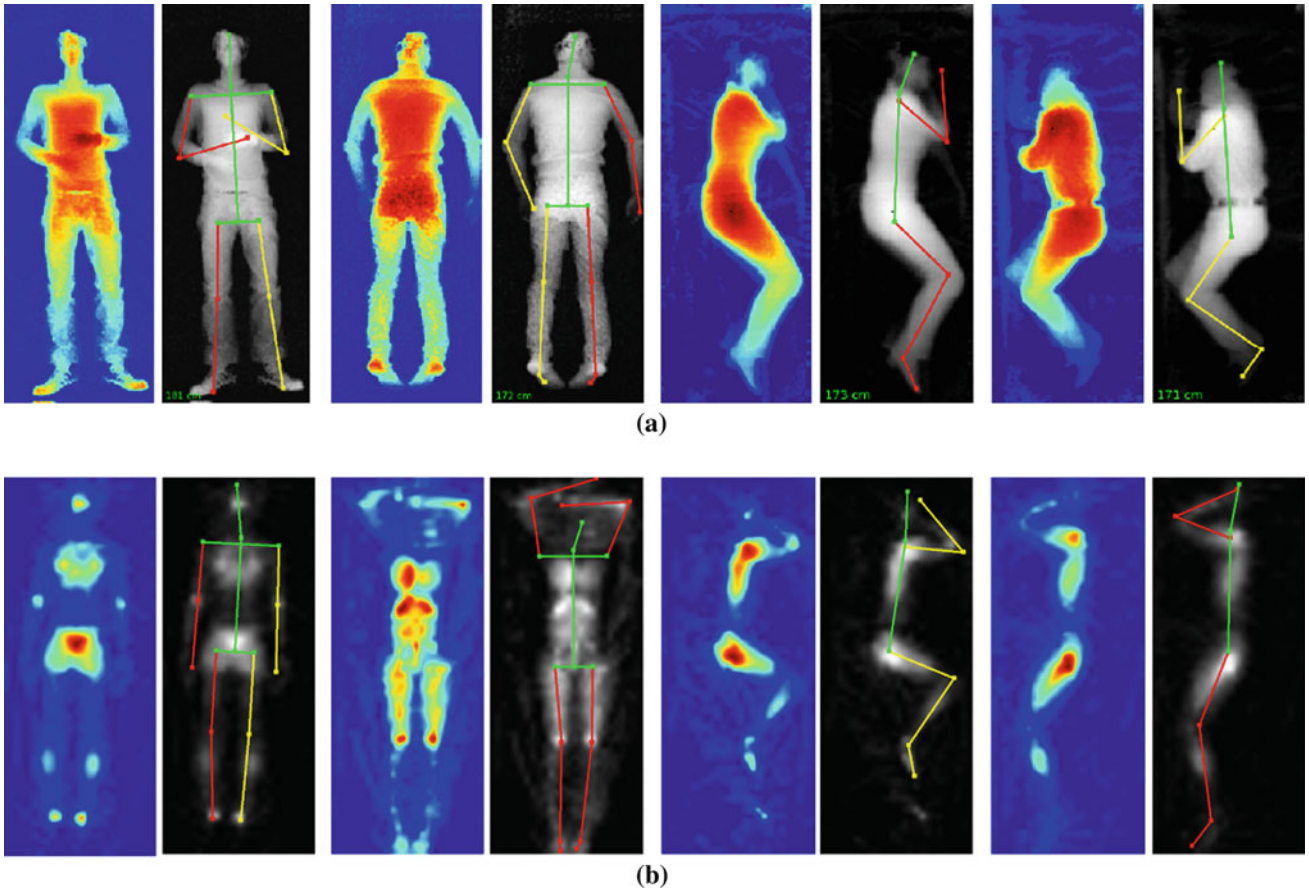

Fig. 4 Qualitative results of the body pose estimation stage. *Upper/lower row*: results for RI/PI data, overlaid with the skeleton of the estimated pose after five optimization cycles. From *left* to *right*: prone, supine, left, right postures. Right extremities are colored in *red*, left ones in *yellow*, head, shoulder, torso and hips in *green*. **a** RI-based pose estimation results, **b** PI-based pose estimation results

Table 3 Quantitative results of our approximate pose estimation algorithm

	Crest	Neck	Shoulder		Hip		Knee	
			L	R	L	R	L	R
RI								
Supine	3.8	3.3	3.7	4.6	6.2	6.3	4.9	4.3
Prone	5.5	4.4	7.3	5.3	6.7	5.9	7.6	6.4
Left	6.0	5.0	5.8	5.9	7.7	6.9	–	7.0
Right	6.1	5.3	6.5	6.5	7.1	7.6	5.6	–
PI								
Supine	3.2	4.5	5.1	5.2	4.0	3.9	4.2	4.4
Prone	8.6	6.7	6.6	7.0	6.4	6.7	2.2	2.4
Left	7.7	6.8	6.1	5.1	5.3	5.8	7.0	–
Right	6.2	6.6	4.7	4.6	6.7	5.6	–	6.5

Given is the RMS of the Euclidean distance [cm] (in the planar domain) between the estimated and ground truth location of clinically relevant body landmarks, left (L) and right (R), for the subset of correctly detected landmarks. Note that in lateral position, one knee is usually severely occluded (RI) or hardly discriminable from the other (PI), hence no results are given

Table 4 Body landmark detection rates in %

	Crest	Neck	Shoulder		Hip		Knee	
			L	R	L	R	L	R
RI								
Supine	1.00	1.00	1.00	1.00	0.97	0.95	1.00	1.00
Prone	1.00	1.00	0.94	0.91	0.97	1.00	0.97	0.97
Left	1.00	1.00	0.97	1.00	0.97	0.91	–	0.91
Right	0.84	0.97	1.00	0.97	0.91	0.88	0.97	–
PI								
Supine	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Prone	0.47	0.85	0.91	0.79	1.00	1.00	1.00	1.00
Left	0.88	0.94	1.00	0.97	1.00	1.00	0.94	–
Right	0.88	0.91	1.00	1.00	1.00	1.00	–	0.94

Landmarks for which the estimation error exceeded 15 cm were considered as outliers and excluded from the pose estimation accuracies given in Table 3

Author's personal copy

Furthermore, in contrast to methods focusing on dynamic motion tracking of the upper body [8,11], our methods aim at whole-body pose estimation from a single-shot measurement.

Discussion and conclusions

We have presented a novel markerless solution for the detection of a reclined patient's orientation, posture and pose on the patient table. Extracting this information in a fully automatic manner may supersede a good portion of the conventional manual input for patient registration in diagnostic tomographic imaging. Furthermore, it may be used to automatically localize a given examination site. In a case study on real data from 16 subjects, the performance of the proposed system was evaluated quantitatively for a 3-D time-of-flight RI camera and a pressure sensing mattress (PI). For both modalities, the proposed heuristic scheme successfully detected the patient orientation for all subjects. Using k-NN posture classification, PI outperformed RI with a mean classification rate of 95.5 % (RI: 79.4 %). Here, our experiments have shown that the distinctive pressure distribution of the human body is better suited for discriminating between supine and prone postures.

Concerning the proposed approximate body pose estimation, both modalities achieved comparable results. Anatomical body landmarks were localized with an accuracy of ± 5.84 cm for RI and ± 5.53 cm for PI data, respectively.

In clinical practice, the acceptable tolerance margins for orientation detection, posture classification and landmark localization depend on the type of examination and cannot be quantified in a generic manner. The impact on specific workflows and the requirements to gain clinical acceptance have to be validated in dedicated clinical studies. In general, note that the proposed system intends to improve the clinical workflow by supporting staff during patient setup and scanner initialization. It suggests appropriate settings based on the proposed framework, but the final decision and responsibility remains with the human operator.

In this discussion, as a concrete example, let us consider the case of MRI prescan positioning. Here, in order to gain clinical acceptance, we expect the following demands to be met. We consider a reliable orientation detection as mandatory and expect that the posture must be classified correctly in at least 90 % of the cases. Both requirements are met by the PI sensor. Posture classification based on RI does not fulfill the demands. Regarding the requirements in terms of localization accuracy, it should be noted that MR localizers usually cover the maximal field of view in the scanner bore (40–50 cm in all dimensions) and are centered around the target region of interest. Hence, deviations in the scale of 2–8 cm are acceptable for routine prescans. Lateral deviations are negligible as the prescan typically covers the full width of the patient table.

Regarding practical limitations of the proposed approach, we remark that optical sensors are sensitive with respect to visual occlusion. This limitation was not considered in our study. The use of cushions, blankets, loose clothing, and the presence of medical staff and equipment may therefore impose a risk for mis-detection and misclassification, especially for RI sensors. In contrast, we expect the PI-based setup to provide reliable results in these cases. Another important aspect beyond the scope of this work is the fact that in case additional contextual information is available, the search space for both posture and pose estimation can be highly reduced. For instance, prior knowledge about a head coil being plugged into an MR scanner is a reliable indicator for supine posture and head-first orientation. Also, the potential locations of crest and neck are severely restricted.

Let us conclude that the proposed framework is not restricted to application in diagnostic tomographic imaging but also holds potential for interventional applications, motion analysis, and sleep monitoring [7].

Even though we expect that a combination of both modalities could further improve robustness and accuracy of our method, the fusion of RI and PI sensors is debatable with respect to the growing total system cost. Further investigations concerning the optimal level of detail in the body model geometry, different clinical scenarios and pose tracking over the course of examination will be subject to future research.

Acknowledgments This research was partially funded by Siemens Healthcare MR, Erlangen, Germany, and Siemens Shenzhen Magnetic Resonance (SSMR), Shenzhen, P. R. China. S. Bauer gratefully acknowledges the support by the European Regional Development Fund (ERDF) and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (StMWIVT), in the context of the R&D program IuK Bayern under Grant No. IUK338.

Conflicts of interest Robert Grimm receives financial support of Siemens AG, Healthcare Sector, Erlangen, Germany. Johann Sukkau is employee of Siemens AG, Healthcare Sector, Erlangen, Germany. The studies were carried out with support from Siemens Healthcare MR, Erlangen, Germany, and Siemens Shenzhen Magnetic Resonance (SSMR), Shenzhen, P. R. China.

References

1. Keil A, Wachinger C, Brinker G, Thesen S, Navab N (2006) Patient position detection for SAR optimization in magnetic resonance imaging. In: Proceedings of international conference on medical image computing and computer assisted intervention, pp 49–57
2. Wachinger C, Mateus D, Keil A, Navab N (2010) Manifold learning for patient position detection in MRI. In: Proceedings of IEEE international symposium on biomedical imaging, pp 1353–1356
3. Fenchel M, Thesen S, Schilling A (2008) Automatic labeling of anatomical structures in MR FastView images using a statistical atlas. In: Proceedings of international conference on medical image computing and computer assisted intervention, pp 576–584

Author's personal copy

4. Moeslund T, Hilton A, Krüger V (2006) A survey of advances in vision-based human motion capture and analysis. Comput Vis Image Underst 104(2–3):90–126
5. Poppe R (2007) Vision-based human motion analysis: an overview. Comput Vis Image Underst 108(1–2):4–18
6. Mündermann L, Corazza S, Andriacchi TP (2006) The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. J Neuroeng Rehabil 3:6
7. Wang CW, Hunter A (2010) Robust pose recognition of the obscured human body. Int J Comput Vis 90:313–330
8. Haker M, Böhme M, Martinetz T, Barth E (2009) Self-organizing maps for pose estimation with a time-of-flight camera. In: Proceedings of DAGM dynamic 3D imaging workshop, pp 142–153
9. Jensen R, Paulsen R, Larsen R (2009) Analysis of gait using a treadmill and a time-of-flight camera. In: Proceedings of DAGM dynamic 3D imaging workshop, pp 154–166
10. Knoop S, Vacek S, Dillmann R (2009) Fusion of 2D and 3D sensor data for articulated body tracking. J Rob Auton Syst 57(3): 321–329
11. Zhu Y, Dariush B, Fujimura K (2008) Controlled human pose estimation from depth image streams. In: Proceedings of IEEE conference on computer vision and pattern recognition workshops, pp 1–8
12. Schaller C, Rohkohl C, Penne J, Stürmer M, Hornegger J (2009) Inverse C-arm positioning for interventional procedures using real-time body part detection. In: Proceedings of international conference on medical image computing and computer assisted intervention, pp 549–556
13. Seo KH, Oh C, Lee JJ (2004) Intelligent bed robot system: pose estimation using sensor distribution mattress. In: Proceedings of IEEE international conference on robotics and biomimetics, pp 828–832
14. Harada T, Mori T, Nishida Y, Yoshimi T, Sato T (1999) Body parts positions and posture estimation system based on pressure distribution image. In: Proceedings of IEEE international conference on robotics and automation, vol 2, pp 968–975
15. Harada T, Sato T, Mori T (2001) Pressure distribution image based human motion tracking system using skeleton and surface integration model. In: Proceedings IEEE international conference on robotics and automation, vol 4, pp 3201–3207
16. Kolb A, Barth E, Koch R, Larsen R (2009) Time-of-flight sensors in computer graphics. In: Eurographics, pp 119–134
17. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning, 1st edn. Addison-Wesley, Reading, MA
18. Brefeld U, Gärtner T, Scheffer T, Wrobel S (2006) Efficient coregularised least squares regression. In: Proceedings of international conference on machine learning, ACM, pp 137–144