

Tolga-Can Çallar*, Elmar Rueckert, and Sven Böttger

Efficient Body Registration Using Single-View Range Imaging and Generic Shape Templates

<https://doi.org/10.1515/cdbme-2018-XXXX>

Abstract: Computer-aided medical systems, e.g. in the fields of medical robotics or image-based assistance, are continuously investigated to overcome human limitations concerning perception, memory or dexterity. A common requirement of such systems is the availability of a digital model describing the patient's position and morphology during a procedure. Operational complexity and technical limitations of established 3D imaging methods leave clinical settings in need of a method for the fast acquisition of a three-dimensional body surface representation. For this purpose, we propose an unsupervised and efficient body registration pipeline based on the markerless elastic registration and completion of single-view stereo range images of the body surface with statistical parametric body shape templates. Initial results show a promising representative quality of the models generated through the registration process with submillimetric fitting accuracy and realistic surface morphology, indicating the general feasibility of our approach as an instant body registration method for automated medical and biometric applications.

Keywords: body registration, range imaging, 3D modeling, computer vision, medical robotics, automation

1 Introduction

The clinical success of surgical, diagnostic and therapeutic procedures exhibits a high degree of dependency upon the interactions and capabilities of medical personnel. Through the emergence of computer-aided medicine and medical automation, many conceptual solutions for negative effects caused by user-dependency and human limitations of perception, cognition, physical strength or dexterity have been developed [1]. Currently studied use cases within this research field including robotic medical ultrasound [2] indicate improvements in reliability and consistency of outcome when compared to conventional medical practice. A commonality of virtually all automated medical systems employed for physical patient in-

teraction is the requirement of a patient body model for the definition and localization of task-specific anatomic target regions. Conventionally, established 3D imaging modalities with varying image quality, acquisition speed, and nocuous potential, e.g. CT and MRI have been utilized in combination with range imaging techniques to perform on-site registration of body parts with previously acquired 3D image data [3]. However, the associated health risks, procedural complexity and technical limitations of the image acquisition and registration process render these approaches infeasible in cases where large body surface coverage is required, visualization of internal anatomy is not desired or no prior suitable imaging studies have been done. Furthermore, unavoidable positional differences between the subject in the source image and during registration may reduce the obtained model's representative capacity regarding body pose and physiognomy at the moment of a medical procedure. This encouraged us to investigate single-view range imaging with active infrared stereo cameras as a low-cost and stand-alone technique for this purpose, due to its inherent advantages effectively mitigating the mentioned drawbacks: The instant optical image acquisition increases the temporal representative quality of acquired image data while facilitating an efficient and innocuous registration process. When compared to other range imaging techniques, e.g. ToF-cameras, infrared stereo imaging provides a larger field of view and higher depth resolution at medium range, which is crucial for clinical applications. The major difficulty with this technology is the occurrence of optical occlusions restricting the retrievable body surface area from a given acquisition perspective. Proposed techniques to resolve this issue generally utilize multi-view acquisition settings to increase the amount of captured surface area. Classically, overlapping partial scans are acquired and rigidly registered [4] to reconstruct a body model. To recover unobserved regions, promising results were achieved through the non-rigid registration of densely acquired full-body scans with template models [5] or sparse surface data with individual shape priors in form of handcrafted templates [6] or shape representations learned from subject-specific multi-view data [7]. As these approaches involve significant user-interaction and technical complexities decreasing the registration efficiency, we present a method for the unsupervised elastic registration of single-view range images with generic templates allowing the rapid and accurate digitization of the surface anatomy in clinical settings.

*Corresponding author: **Tolga-Can Çallar**, Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Ratzeburger Allee 160, Lübeck, Germany, e-mail:

tolgacan.callar@student.uni-luebeck.de

Elmar Rueckert, Sven Böttger, Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Germany

2 Methods

Our proposed approach for the registration of the body surface consists of three main elements: First, we acquire one single-view range image of a subject on an examination couch with the low-cost Intel RealSense D415 active IR stereo camera (Intel Corporation, Santa Clara, USA). Afterward, we perform the segmentation of the subject’s body and the reduction of range imaging artifacts. In the last step, the segmented body surface is elastically registered with a parametric template body model, which has been modified in advance to resemble the subject’s anthropometric features. The segmentation and registration processes of our framework are implemented using the 3D-processing software VTK (Kitware, Inc., Clifton Park, USA).

2.1 Range Image Acquisition

A range image of the complete scenery is captured from which the sections belonging to the subject are extracted and used as a target input for the registration of the body surface. The stereo camera is placed lateral to the examination couch and oriented to align the x -axis of the camera frame with the expected longitudinal axis of the body of the subject to be registered. The camera’s height and vertical tilt are adjusted, until the field of view spans the entire space to be occupied by the body, with the image frame being centered around the center of the examination couch. In preparation for the subsequent segmentation, we first capture the scenery without the subject and then with the subject in the supine position. As visualized in Figure 1, we acquire a frontal and oblique image for every subject to generate source data for evaluation of our registration process under varying acquisition conditions. All of the obtained range images are converted to point clouds and exported as a polygonal mesh through the internal Delaunay-triangulation function of the RealSense Viewer application.

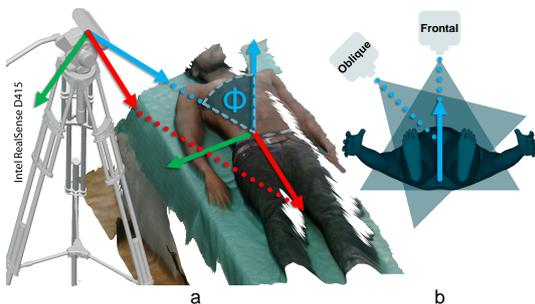


Fig. 1: **a** Acquisition setup and orientation of the camera and body frame. **b** Utilized image acquisition perspectives.

2.2 Body Segmentation

To define the area of the range scan to be registered with the template, we perform the segmentation of the subject in the triangulated point cloud of the scene. For this purpose we implemented a background subtraction variant utilizing the VTK-library: We generate k -d-tree-based space-partitionings of both the background scenery captured in advance and the scene including the subject. On these, we perform a nearest-neighbor search to identify those vertices in the subject’s scene without any corresponding vertex from the background mesh in their vicinity. Due to optical occlusions causing regions with sparsely imaged steep depth gradients around border regions of the body (see Figure 2), the segmentation result still contains artifacts visible in the form of local elongation of the triangular faces in these regions of the mesh. We exploit the large ratio between the perimeter of the triangular faces and their area to filter out the elongated faces.

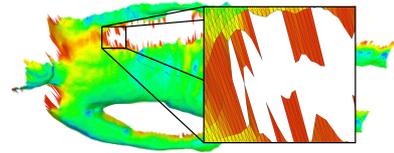


Fig. 2: Segmented body surface with elongated triangular faces caused by locally steep depth gradients.

2.3 Template-Based Registration

Regardless of optical occlusions, single-view range images always retrieve less than half of the body surface. This renders a solely segmentation-based approach insufficient for the spatial registration of complete 3D body surfaces. Therefore, we propose the elastic registration of a parameterized body shape template onto the segmented range image to augment the anatomically incomplete range image source while preserving the captured anatomic details. To ensure the template’s morphological resemblance of the subject, we employ adjustable polygonal body models originating from statistical full-body laser scans from UMTRI Human Shape [8]. This allows us to create individual base shapes accommodating for the respective anthropometric properties of a subject that may also allow for semantic pre-annotations, e.g. for task-specific target regions. We implemented an affine variation of the point set registration algorithm ICP derived from the method described by [5]. Our algorithm iteratively optimizes transformations of the template to minimize the distance to the nearest-neighbor correspondences between the template and target vertices while enforcing locally stiff deformations. In contrast to classical ICP we do not utilize global transformations on our template

model. Instead we define n individual affine transformations \mathbf{X}_i for every vertex $v_i \in \mathbf{V}$ of the template mesh \mathcal{T} . These are arranged in a matrix $\mathbf{X} := [\mathbf{X}_1 \dots \mathbf{X}_n]$ representing the entirety of the vertex-wise transformations regularized by the distance to closest points on the target surface \mathcal{T} as expressed by the energy term

$$E_d(\mathbf{X}) := \sum_{v_i \in \mathbf{V}} w_i \text{dist}^2(\mathcal{T}, \mathbf{X}_i v_i) \quad , \quad (1)$$

where w_i denotes a binary reliability weighting for the nearest-neighbour correspondences. Two reliability tests are performed on every candidate vertex pair $(v_i \in \mathbf{T}, v_j \in \mathcal{T})$ after finding the initial set of correspondences. A correspondence is rejected, if the template's vertex normal $\mathbf{n}_{v_i \in \mathbf{T}}$ exceeds an angular deviation threshold of $\theta = 90^\circ$ with respect to the closest target vertex normal $\mathbf{n}_{v_j \in \mathcal{T}}$, i.e.

$$\arccos \left(\frac{\mathbf{n}_{v_i} \cdot \mathbf{n}_{v_j}}{\|\mathbf{n}_{v_i}\| \cdot \|\mathbf{n}_{v_j}\|} \right) > \theta \quad , \quad (2)$$

or if the nearest template vertex is located on the border of the template mesh. This avoids deformations towards topologically incompatible vertices and the mapping of multiple template vertices onto the same target vertex. We penalize the dissimilarity of the transformations of vertices connected via an edge \mathcal{E} in the template mesh, based on the Frobenius norm $\|\cdot\|_F$ defined as the square root of the sum of the squares of the matrix elements, with a weighting matrix $\mathbf{G} := \text{diag}(1, 1, 1, \gamma)$, where γ can be controlled to emphasize rotational over translational similarity in the transformations:

$$E_s(\mathbf{X}) := \sum_{\{i,j\} \in \mathcal{E}} \|(\mathbf{X}_i - \mathbf{X}_j)\mathbf{G}\|_F^2 \text{dist}^2(\mathcal{T}, \mathbf{X}_i v_i) \quad (3)$$

This ensures the transformations to retain local stiffness, resulting in a globally smooth deformation and infilling of holes in the registered mesh to ensure robustness against missing shape topology while approximating the target surface. We deliberately omitted a penalty for the distance between landmarks on the mesh surfaces, to avoid complicating the registration process by setting landmark correspondences. The final cost function assumes the shape,

$$E(\mathbf{X}) := E_d(\mathbf{X}) + \alpha E_s(\mathbf{X}) \quad (4)$$

and constrains a conjugate gradient sparse solver (Eigen v3.3.7), which we employ to solve for an optimal set of transformations \mathbf{X} . As transformations for regions containing rejected correspondences are not subject to the distance constraint $E_d(\mathbf{X})$ and thus would remain unchanged, we enforce smooth transformation of these vertices by emphasizing the similarity constraint by through the parameter α . To initialize the registration with semantically correct point-to-point correspondences, we align the longitudinal axis of the template

mesh with the x-axis of the camera (see Figure 1), confining the rotational difference to this axis. Exploiting the approximate coincidence of the surface normal of the couch with the body's sagittal axis in the supine position, we infer the body orientation from the average direction of the vertex normals of the couch nearest to its intersection point with the camera's z-axis. The template is rotated about its longitudinal axis by the angle Φ that the subject's estimated sagittal axis deviates from the camera's optical axis, which is equal to the acquisition angle. The translational misalignment is resolved by

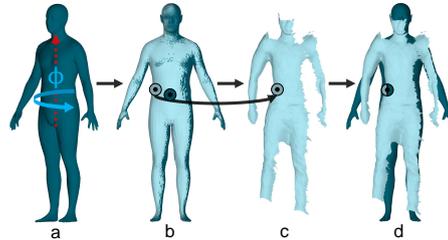


Fig. 3: **a** Rotational alignment of the template through the acquisition angle Φ . **b** Definition of a COM for the simulated region of the template visible at the acquisition angle, which semantically corresponds with **c** the target's COM. **d** Alignment after translation with the vector between the COMs.

finding the displacement between the centers of mass (COMs) of the template and target. Due to missing regions in the target mesh, the COMs of the template and target do not correspond to the same anatomical location. To compensate this, we recalculate the template's COM for the surface area being visible from the acquisition angle Φ (see Figure 3). The anatomically corresponding COMs are then used to define the translational displacement and obtain an initialization for the elastic registration.

3 Results

We evaluated the performance of our method by registering frontal and oblique range images of four male volunteers covering a spectrum of body morphologies with respect to age ($M = 36.5$ a, $SD = 12.69$ a), height ($M = 180.25$ cm, $SD = 5.89$ cm) and weight ($M = 100.75$ kg, $SD = 13.63$ kg). Using template models with approximately 10000 polygons, the average runtime per registration pass amounted to 30 seconds on a mobile 2.0 GHz quad-core CPU. The registration quality is assessed by the morphological realism and the residual RMS Euclidean distances $d_{\text{RMS}}(\mathcal{T}, \mathbf{T})$ between the vertices of the source meshes and their respective closest vertices in the deformed templates (see Table 1). Our method preserves anatomical details by accurately approximating overlapping regions

Tab. 1: Average residual RMS Euclidean distance between the target and template for the total meshes and the region of the torso for the respective acquisition views.

View	$d_{\text{RMS}}(\mathcal{T}, T)$	$d_{\text{RMS}}(\mathcal{T}_{\text{torso}}, T)$
Frontal	12.1357 mm	0.8015 mm
Oblique	13.3294 mm	0.6562 mm

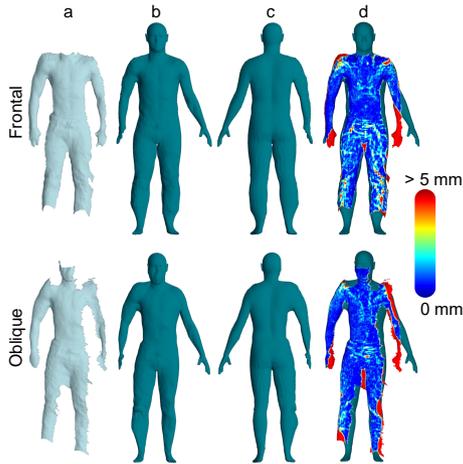


Fig. 4: Registration results obtained for frontal and oblique range images of one subject: **a** Segmented body surface. **b** Anterior and **c** posterior deformations of the template models. **d** Residual local minimal Euclidean distances to the respective target surface.

between template and target and completes missing regions in a statistically probable manner concerning principal anthropometric features of the subject (see Figure 4). With submillimetric deviations, the registration of the torso can be interpreted as a reparameterization of the source image. The registration fit of articulated regions suffers from positioning inconsistencies, which are partially recovered (see the right arm in Figure 4). As these regions often are not of direct medical interest, this is irrelevant for most applications. Regardless of similar numerical results, the overall realism achievable through frontal views appears to be superior, as the symmetric source data guides the deformation evenly. For practical applications, the viewing angle capturing most of the respective application-specific regions of interest still should be preferred. As missing regions in the target mesh are filled in based on the enforced similarity local transformations, surface models produced by the registration of oblique range images may be partially insufficient for high accuracy applications.

4 Conclusion

We presented an efficient 3D body registration framework based on the markerless elastic registration of single-view

range images with parametric shape templates. Our method is designed to be applicable within clinical environments, requiring minimal technical equipment, source data and user interaction for the registration process. Initial experiments indicate the general feasibility of this technique, validating its robustness and accuracy under variable acquisition settings. This work represents a contribution to alleviate the establishment of automated medical systems, by providing the ability to rapidly obtain complete digital surface anatomy representations to be used for targeting and motion planning purposes, e.g. within robotic systems. More generally, we demonstrate an alternative imaging technique, suitable for the assessment of the external morphology in humans. Based on these preliminary results we intend to further improve our method's shape completion accuracy and robustness against body pose variations by exploiting anatomic symmetries to enforce mirroring of transformations about the sagittal plane.

Author Statement

Research funding: Authors state no funding involved. 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 conducted research is not related to either human or animal use.

References

- [1] Marc T, Computer-aided medicine and surgery. In: Digital Medicine. 2.3;2016:81-87.
- [2] Al-Badri M, Ipsen S, Böttger S, Ernst F. Robotic 4D ultrasound solution for real-time visualization and teleoperation. Current Directions in Biomedical Engineering 2017;3(2):559-561.
- [3] Placht S, Stancanello J, Schaller C, Balda M, Angelopoulou E. Fast time-of-flight camera based surface registration for radiotherapy patient positioning. Medical physics 2012;39(1):4-17.
- [4] Chen C, Hung Y, Cheng JB. A fast automatic method for registration of partially-overlapping range images. IEEE Sixth International Conference on Computer Vision 1998;242-248.
- [5] Amberg B, Romdhani S, Vetter T. Optimal step nonrigid ICP algorithms for surface registration. IEEE Conference on Computer Vision and Pattern Recognition 2007;1-8.
- [6] Li H, Adams B, Guibas LJ, Pauly M. Robust single-view geometry and motion reconstruction. ACM Transactions on Graphics 2009;28(5):1-10.
- [7] Angelov D, Srinivasan P, Koller D, Thrun S, Rodgers, J, Davis J. SCAPE: shape completion and animation of people. ACM SIGGRAPH Papers 2005;408-416.
- [8] Reed M, Raschke U, Tirumali R, Parkinson M. Developing and implementing parametric human body shape models in ergonomics software. Proceedings of the 3rd international digital human modeling conference 2014.