Soft-Body Registration of Pre-operative 3D Models to Intra-operative RGBD Partial Body Scans

Richard Modrzejewski[1,2], Toby Collins [2], Adrien Bartoli [1], Alexandre Hostettler[2] and Jacques Marescaux[2]

[1] EnCoV, Institut Pascal, UMR 6602, CNRS/UBP/SIGMA, 63000 Clermont-Ferrand
[2] IRCAD and IHU-Strasbourg, 1 Place de l'Hopital, 67000 Strasbourg

Abstract. We present a novel solution to soft-body registration between a pre-operative 3D patient model and an intra-operative surface mesh of the patient lying on the operating table, acquired using an inexpensive and portable depth (RGBD) camera. The solution has several clinical applications, including skin dose mapping in interventional radiology and intra-operative image guidance. We propose to solve this with a robust non-rigid registration algorithm that handles partial surface data, significant posture modification and patient-table collisions. We investigate several unstudied and important aspects of this registration problem. These are the benefits of heterogeneous versus homogeneous biomechanical models and the benefits of modeling patient/table interaction as collision constraints. We also study how abdominal registration accuracy varies as a function of scan length in the caudal-cranial axis.

1 Introduction, Background and Contributions

An ongoing and major objective in computer-assisted abdominal interventional radiology and surgery and is to robustly register pre-operative 3D images such as MR or CT, or 3D models built from these images, to intra-operative data. There are two broad clinical objectives for this. The first is to facilitate automatic radiation dose mapping and monitoring in fluoroscopy-guided procedures [1, 2], using a pre-operative model as a reference. The most important aspect is registering the skin exposed to primary radiation. Good registration would enable dose exposure monitoring across the patient's skin, and across multiple treatments. The second clinical objective is to achieve interventional image guidance using pre-operative 3D image data if interventional 3D imaging is unavailable. Recently methods have been proposed to register a pre-operative 3D model using external color [3] or depth+color (RGBD) images [4–7], capturing the external intra-operative body shape and posture of the patient, operating table and surrounding structures. The advantages of registering with color or RGBD cameras is they are very low-cost, very safe, compact, and large regions of the patient's body can be imaged in real-time. They also facilitate 'body see-through' AR visualization using hand-held devices such as tablets or head-mounted displays

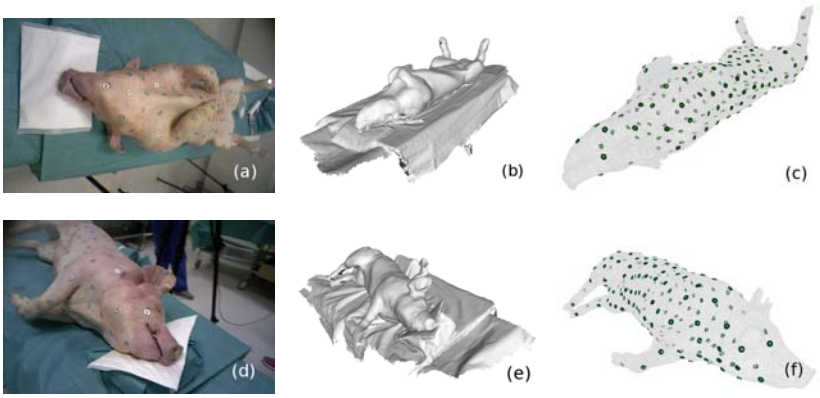

Fig. 1. Porcine dataset. (a) Example of supine position, (b) RGBD scan corresponding to (a), (c) CT model corresponding to (a) with segmented surface markers. (d-f) equivalent images with a right-lateral position.

Their disadvantage is that the internal anatomy cannot be imaged. Therefore we can only establish correspondence (data association) on the patient's visible surface, which can be difficult particularly in the abdominal region, where the skin has few distinguishing geometrical features. The second main difficulty are large occluded regions. For example, for patients in the supine position the posterior is never visible to the camera. Because of these difficulties, previous registration methods that use external cameras have been limited to rigid registration [3–7]. These methods cannot handle soft-body deformation, which is unavoidable and often difficult to precisely control. Such deformation can be significant, which is particularly true when the patient's lying position is different. For example, CT and MR are mainly acquired in the supine position, but the procedure may require the patient in lateral or prone positions. We show our solution can improve registration accuracy with strong posture changes.

The main contributions of this paper are both technical and scientific. Technically, this is the first solution to soft-body patient registration using a preoperative CT model and 3D surface meshes built from multiple external RGBD images. We build on much existing work on fast, soft-body registration using surface-based constraints. The approaches most robust to missing data, occlusions and outliers, are currently iterative methods based on robust Iterative Closest Point (ICP) [8–11]. These work by interleaving data association with deformable model fitting, while simultaneously detecting and rejecting false point matches. These methods have been applied to solve other medical registration problems including laparoscopic organ registration [9], and registering standing humans with RGBD surface meshes e.g. [11]. We extend these works in the following ways: 1) to model table-patient interaction via table reconstruction and collision reasoning, 2) outlier filtering to avoid false correspondences. Scientifically, it is well known that measuring soft-body registration accuracy with real

data is notoriously difficult, but essential. In related papers quantitative evaluation is performed using virtual simulations, with simplified and not always accurate modeling of the physics and data. We have designed a systematic series of experiments to quantitatively asses registration accuracy using real porcine models in different body postures. The ethically-approved dataset consists of a pig in 20 different postures, with 197 thin metal disc markers (10mm diameter, 2mm width) fixed over the pig's body. For each posture there is a CT image, an RGBD body scan and the marker centroids. The centroids were excluded from the biomechanical models, preventing them being exploited for registration. We could then answer important and unstudied questions:

- Does modeling patient/table collision improve registration results? Is this posture dependent? Are the improvements mainly at the contact regions?
- Does using a heterogeneous biomechanical model improve registration compared to a heterogeneous model? This tells us whether accurate biomechanical modeling of different tissue classes/bones are required.
- How much of the abdominal region is required in the CT image for good registration? We study this by varying image size in the caudal-cranial axis.

We also demonstrate qualitatively our registration algorithm on a human patient in real operating room conditions for CT-guided percutaneous tumor ablation. This result is the first of its kind.

2 Methods

2.1 Biomechanical Model Description

We take as input a generic biomechanical mesh model representing the patient's body (either partial or full-body). We denote the model's surface vertices corresponding to the patient's skin as $\mathcal{V}_s$, and the interior vertices as $\mathcal{V}_I$. We use $f(\mathbf{p};\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ to denote the transform of a 3D point $\mathbf{p}$ in 3D patient coordinates to patient scan coordinates, provided by the biomechanical model. This is parameterized by an unknown vector $\mathbf{x}$, and our task is to recover it. In our experiments we model $E_M(\mathbf{x})$ using a mass-spring model generated from segmentations provided by [12]. We used Tetgen to generate tetrahedral meshes from surface triangles, which formed the interior vertex set $\mathcal{V}_I$. We emphasize that the algorithm is compatible with any first order differentiable biomechanical model.

2.2 Intra-operative Patient Scanning and Segmentation

We scan the patient using a hand-held RGBD camera (Orbic Astra Pro), that is swept over the patient by an assistant. Another option for scanning would be to use ceiling-mounted RGBD cameras, however this has some limitations. The cost is higher, there may be line-of-sight problems, and closer-range scanning is not generally possible (<1m), which limits depthmap accuracy. We reconstruct 3D

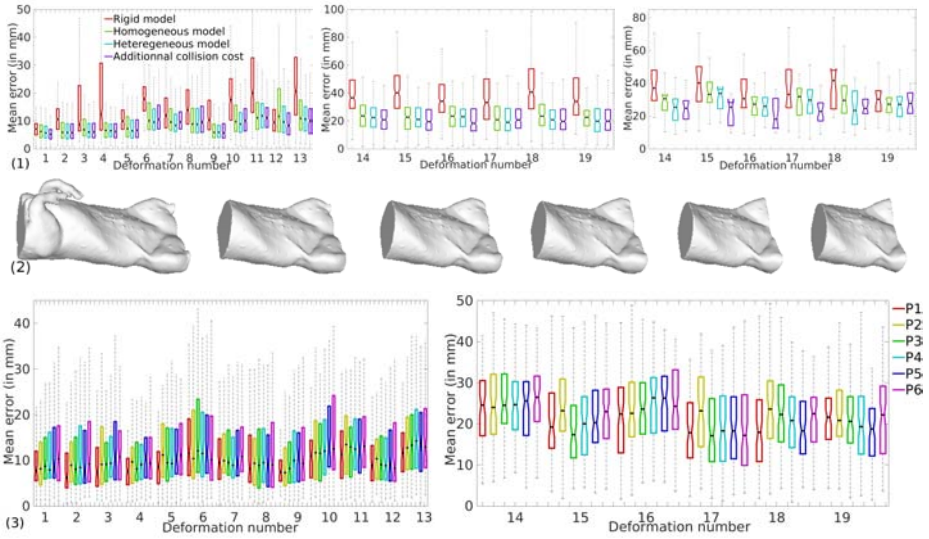

Fig. 2. NLD of markers for different configurations: row (1) small posture changes (left), large changes (center) and markers localized only on the table (right), row (3) different section sizes (corresponding to row(2)), from P1 (biggest section) to P6 (smallest section), with supine postures (left) and lateral postures (right).

surface meshes from the RGBD video using an existing reconstruction method [13] but any other method could be used. The scanning process can be performed in approximately 10 seconds, allowing it to be done during a single breath hold to circumvent breathing artifacts. To reduce occlusions, we perform scanning before surgical coverings are placed. We denote the fused output by the vertex set $\mathcal{Q}$. When using [13] this is typically of size $O(10^5)$. We model the table as a planar surface with a flat thin padded top layer, which is valid in most real conditions. We use the distance $t \in \mathbb{R}^+$ mm to denote the maximal compression the padded layer can undergo for any patient, which can be measured once *a priori*. We fit a plane in the form $ax + by + cz + d = 0$ to the points in $\mathcal{Q}$ corresponding to the table top. This is done using RANSAC, where the largest planar region is found within a maximal depth of 1m to the camera. This heuristic is used to eliminate the ground plane.

2.3 Registration Problem Formulation

We formulate the problem as iterative energy minimization based on [9, 8]. The energy has the form $E(\mathbf{x}) = E_M(\mathbf{x}) + \lambda_{ICP} E_{ICP}(\mathbf{x}; \mathcal{V}_s, \mathcal{Q}) + E_{collision}(\mathbf{x}; \mathcal{V}_s, \mathcal{V}_I)$, where $\lambda_{ICP} \in \mathbb{R}^+$ is the ICP weight. The term E_{ICP} is the *ICP energy*, which attracts the model's transformed surface vertices $\mathcal{V}_s$ to fit to their closest vertices in $\mathcal{Q}$, while simultaneously handling mismatches, which are in practice inevitable.

We compute the closest vertices using fast matching on the GPU, using the computer vision library OpenCV. We then filter out matches that are strongly likely to be outliers (e.g. matches to surrounding objects on the table, or the table itself). This filter is derived from [13], by measuring the angle between surface normals of the matched points, and rejecting each match which disagrees by a tolerance threshold τ_N. We set this to 45 degrees. We denote the remaining matches by the set $\{(\mathbf{p}_i, \mathbf{q}_i)\}$, $\mathbf{p}_i \in \mathcal{V}_s$, $\mathbf{q}_i \in \mathcal{Q}$. We define E_{ICP} as:

$$E_{ICP}(\mathbf{x}; \mathcal{V}_s, Q) = \sum_i \rho_h \left(\mathbf{n}_i^\top (f(\mathbf{p}_i) - \mathbf{q}_i) \right) \tag{1}$$

where $\mathbf{n}_i$ denotes the normal of point $\mathbf{q}_i$ and ρ_h denotes the Huber norm. This defines a *point-to-plane distance* function, which improves convergence in ICP by allowing surfaces to slice across each other during optimization [14]. The Huber norm is used to reduce the influence of any remaining outlier matches on the energy function. This is essential to achieve good robustness.

The term $E_{collision}$ is the table collision energy, which prevents model vertices penetrating the table. We assume that the patient's body is always above the table, which is valid is practically all cases. Collision is handled by forcing all model vertices to exist above the table at a state of maximum pad compression. We do this using a *signed distance-to-table* function $d(\mathbf{p})$. For a planar table this is $d(\mathbf{p}) = [a, b, c]\mathbf{p} + d - t$. The extension to non-planar tables can be handled by modifying $d(\mathbf{p})$ according to the table's shape. We define the penalty as:

$$E_{collision}(\mathbf{x}; \mathcal{V}_s, \mathcal{V}_I) = \sigma_k \sum_{\mathbf{p} \in \mathcal{V}_s \cup \mathcal{V}_I} \max(d(\mathbf{p}), 0) \tag{2}$$

where σ_k is the penalty coefficient. At each iteration σ_k is increased by a factor of 10. At convergence the collision constraint is exactly enforced. To avoid incorrect data association with the table, we eliminate all ICP matches that associate points on the scan that correspond to the table. This is done with a distance-to-table threshold τ with a default of 10mm.

2.4 Optimization and Initialization

We adopt the method in [15] to reduce the deformable model's deformation space, exploiting the fact that feasible deformations tend to be mostly smooth. This works by performing a modal analysis using the model's volumetric Laplacian $\mathbf{L}$. The eigenvectors of $\mathbf{L}^\top \mathbf{L}$ with lowest eigenvalues correspond to the smoothest modes of variation. The reduced model is parameterized by $\mathbf{x}$ with a default length of 200. We optimize $E(\mathbf{x})$ iteratively using a stiff-to-flexible strategy [9, 8], which is important to avoid local minima. Initially the model is kept rigid, by setting λ_{ICP} to small value (1.1 in our experiments). We then optimize $E(\mathbf{x})$ using a single Gauss-Newton iteration, and increase λ_{ICP} by a factor (we use a default of 1.2). We then repeat the process, truncating λ_{ICP} to a maximal value, which we set to 10 times the initial value. We continue until convergence is detected or a maximum number of iterations is reached (we use a limit of 50

iterations). We initialize using a roughly-estimated rigid transform. This is currently done by 6 manual landmark correspondences between the model's surface and the scanned surface, then running Horn's algorithm in OpenCV. In future works, automated landmarks extraction can also be used (computed using 2D features learnt [16] for instance).Landmark assocations can be added to Eq.(2) with an extra energy term, with a similar form as the ICP term but with point-to-point distances and fixed associations. In the experimental results we did not include this to validate the performance of ICP-only association.

3 Experimental Results

3.1 Quantitative Analysis with Porcine Datasets

We performed this using a euthanized 50Kg pig with 197 metal markers (1cm diameter) quasi-uniformly arranged over its skin (Fig. 1). A reference CT was made in the supine position, discretized using Tetgen into 19069 3D vertices corresponding to 87524 tetrahedron. We then moved the pig to 19 different positions, representing different intra-operative poses, on a CT table (13 supine (Fig. 1a), 4 left-lateral (Fig. 1d) and 2 right-lateral). For each position we took a CT image and a corresponding RGBD scan. We compared soft-body registration with rigid ICP registration. This allowed us to measure the impact of modeling soft-body deformation. Several conditions were tested: the first was where the full reference CT was used to build the biomechanical model, and we compared accuracy using either a homogeneous biomechanical model and a heterogeneous model. We implemented the latter with two classes (bone and other), where the stiffness of bone springs were 100 times stronger than the other class, to mimic rigid body motion). We further tested with and without table collision. The error metric we used is the Nearest Landmark distance (NLD). It is infeasible to determine real matches between all markers, so nearest neighbours were used as proxies. These were computed after registration by matching each marker in the reference CT to its nearest neighbour in the CT associated with a given RGBD scan. As the average spacing was large (approximately 50mm) compared to the observed errors, the NLD approximates well the true endpoint error.

In the first row of Fig. 2 left, we show the marker endpoint error for the 13 supine positions averaged over all 197 markers. Here the pig's deformation is smaller than in the lateral positions. We observe a considerable improvement using the soft-body models in most cases. There is no substantial improvement in the first case, the reason being that the body posture was more similar to the reference CT posture. We observe a general small improvement using the heterogeneous model and with the addition of table collision constraints. In the first row of Fig. 2 middle, we show similar results for the lateral positions. We generally observe a larger improvement with the heterogeneous versus homogeneous model, justifying its use for large deformations caused by posture change. However there is little to no improvement using the collision constraints. In the first row of Fig. 2 right, we show NLD for the lateral postures, measured only at the markers in contact with the table. Here we can see a general small improvement

Fig. 3. Results with a human patient in a live operating room. The RGBD scan (left), rigid registration (middle) and deformable registration (right).

using the collision constraints, indicating they can improve localized registration accuracy at the table/skin interface. The errors are larger than Fig. 2 middle because they represent completely hidden regions. We then studied the influence of the size of the abdominal region in the reference CT on registration accuracy. Five reference models were tested by cropping the CT in the caudal-cranial axis with six lengths (80 cm to 35 cm) centered on the abdomen, labeled P1 to P6. These are shown from left to right in Fig. 2, row 3. We measured registration accuracy using only the markers in the 35cm section (79 markers). This was to measure the influence of having more geometric information surrounding the abdominal region (hips, arms, shoulders...). For this the heterogeneous model was used with collision constraints. The results for the supine positions are given in Fig. 2, bottom left. The results for the lateral positions are given in Fig. 2, bottom right. We see a general reduction in registration accuracy with increased cropping of the reference CT. The problem is caused by the reduced geometrical information needed to 'anchor' the deformation at geometrically distinct regions. This is clearly present when we move from P1 to P2, where arms and shoulders regions are cropped out.

3.2 Qualitative analysis with a Human Patient

We tested our method in our local hospital for registering a human patient under general anesthetic, before undergoing a percutaneous radio frequency tumor ablation (Fig. 3). The idea was to check the reliability of our registration in computer-assisted percutaneous surgery by visually checking the position of the organs, and by assessing the surface fit. The patient was 58 years old. From her CT scans, a model containing 4383 vertices and 19362 tetrahedron was extracted and 13 organs were segmented and seen in the AR view. Results are presented using AR visualization, using a standard external calibrated RGB camera on a tripod. Both registration and AR visualization were performed live during the operation. We see an improved result through deformable registration, clearly demonstrated by inspecting the patient's silhouette contours. This is the first time soft-body registration has been performed using an intra-operative abdominal RGBD scan of a human patient in an operating theater.

4 Conclusion

We have presented a low-cost method to perform soft-body registration between partial or full-body pre-operative 3D models and intra-operative, partial RGBD scans. We have quantitatively evaluated using a challenging real porcine dataset, which is the first of its kind and will be made public. The achieved accuracy is sufficient for interventional radiation skin dose modeling [1]. It may also be sufficient for rough AR visualization of internal structures, as we have demonstrated with a real patient. In future work we will investigate the impact of more detailed biomechanical models and evaluate internal registration accuracy using segmented anatomical structures.

References

1. Johnson, P.B., Borrego, D., Balter, S., Johnson, K., Siragusa, D., Bolch, W.E.: Skin dose mapping for fluoroscopically guided interventions. Medical Physics (2011)
2. Rodas, N.L., Bert, J., Visvikis, D., De Mathelin, M., Padoy, N.: Pose optimization of a C-arm imaging device to reduce intraoperative radiation exposure of staff and patient during interventional procedures. In: ICRA. (2017)
3. Mahmoud, N., Grasa, .G., Nicolau, S.A., Doignon, C., Soler, L., Marescaux, J., Montiel, J.M.: On-patient see-through augmented reality based on visual SLAM. International Journal of Computer Assisted Radiology and Surgery (2017)
4. Macedo, M.C.F., Apolinário Jr, A.L., Souza, A.C.S., Giraldi, G.A.: High-Quality On-Patient Medical Data Visualization in a Markerless Augmented Reality Environment. Journal on Interactive Systems (2014)
5. Lee, J.D., Huang, C.H., Huang, T.C., Hsieh, H.Y., Lee, S.T.: Medical augmented reality using a markerless registration framework. ESA (2012)
6. Chen, X., Xu, L., Wang, Y., Wang, H., Wang, F., Zeng, X., Wang, Q., Egger, J.: Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. Biomedical Informatics (2015)
7. Tully, S.T.: BodySLAM: Localization and Mapping for Surgical Guidance. (2012)
8. Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid ICP algorithms for surface registration. In: CVPR. (2007)
9. Collins, T., Chauvet, P., Debize, C., Pizarro, D., Bartoli, A., Canis, M., Bourdel, N.: A System for Augmented Reality Guided Laparoscopic Tumour Resection with Ex-vivo User Evaluation. Computer-Assisted and Robotic Endoscopy (2016)
10. Petit, A., Lippiello, V., Siciliano, B.: Real-time tracking of 3D elastic objects with an RGB-D sensor. In: IROS. (2015)
11. Huang, Q.X., Adams, B., Wicke, M., Guibas, L.J.: Non-rigid registration under isometric deformations. Eurographics Symposium on Geometry Processing (2008)
12. IRCAD: Visiblepatient, https://www.visiblepatient.com/
13. Izadi, S., Davison, A., Fitzgibbon, A., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S.: KinectFusion. In: ISMAR. (2011)
14. Segal, A., Haehnel, D., Thrun, S.: Generalized-ICP. RSA (2009)
15. Collins, T., Bartoli, A., Bourdel, N., Canis, M.: Robust, real-time, dense and deformable 3D organ tracking in laparoscopic videos. In: MICCAI. (2016)
16. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: CVPR. (2017)