A 3D Camera-Based Approach to High-Density ECG Imaging

Nikhil Shenoy[1], Maryam Toloubidokhti[2], Linwei Wang[2], Vivek Singh[1], Ankur Kapoor[1]

[1] Siemens Healthineers, Princeton, NJ
[2] Rochester Institute of Technology, Rochester, NY

Abstract

Non-invasive electrocardiographic imaging (ECGI) is a technique used to reconstruct the electrical activity on the epicardial surface using measurements recorded from body-surface electrodes. A typical workflow requires heart and torso geometry from tomographic data, along with 3D electrode locations. This is not part of the routine arrhythmia workflow, and further acquisition burdens the patient. Our aim is to construct an accurate 3D torso model of a patient using a 3D camera, with the intent of eliminating the need for additional scanning. We compare against results obtained when using traditional imaging.

1. Introduction

Noninvasive electrocardiographic (ECG) imaging is a computational technique which reconstructs electrical activity on an individual's heart from the corresponding body-surface ECG data. It has the potential to image a variety of cardiac rhythm disorders such as premature ventricular contractions [1] and reentrant ventricular tachycardia [2].

A typical ECG-imaging workflow requires tomographic imaging to obtain the patient-specific heart and thorax geometry, from which the forward operator linking cardiac electrical activity and body-surface ECG measurements can be established. This often requires non-standard thorax imaging of a subject wearing a high-density electrode array such that these surface electrodes can be identified in the images. However, thorax imaging is outside the routine clinical workflow for arrhythmia patients, and poses a challenge to widespread adoption of ECG-imaging due to the logistical and financial burdens placed on the patient and hospital staff.

Image-less ECG-imaging proposes to circumvent the thorax imaging by using a low-cost 3D-camera to obtain the surface electrode positions [3]. However, existing image-less systems are designed for simplified ECG-imaging, which use the standard 12-lead ECG configuration and may not be able to image more complex arrhythmias due to its sparsity. In addition, none of these systems address how to align the camera-generated electrode positions to the heart geometry, which literature has shown to be critically important to the accuracy of ECG-imaging.

We present a modified ECG-imaging workflow that constructs the torso model using a color and depth (RGBD) camera instead of MRI/CT-based thorax imaging. It includes 1) a **sensing system** in which an RGBD camera records video of a patient wearing ECG electrode strips 2) **electrode localization** which generates the 3D torso model from 2D electrode positions in the video 3) **heart-torso registration** during which the camera-generated torso is aligned to the scan-generated heart geometry and 4) **body-surface deformation** in which the aligned torso is deformed to be close to the skin surface. The resulting torso is viable for full ECG-imaging, significantly reducing the cost and complexity of the existing ECG-imaging workflow without sacrificing accuracy.

2. Methodology

Construction of 3D Torso Model: The camera setup consists of an RGBD sensor mounted on a tripod. To capture the data, the patient stands with his entire body facing the camera, and then rotates in place for a full revolution. Registered color and depth video is captured while the patient rotates and is stored for processing.

The first step in constructing the 3D torso is to identify the 2D positions of each electrode in every frame of the registered RGBD video. Using the optical properties of the 3D camera, we generate an electrode point cloud P_i from the identified electrodes for each frame i of the N frames of video. Our goal in this step is, for all i, to find the optimal 3D rigid transformations R_i^* that align each P_i to a chosen reference coordinate system. Since the patient directly faces the camera at the start of the acquisition, we choose the world coordinate system of frame 0 as the reference coordinate system. Once all the R_i^* are found, the separate point clouds P_i can be aligned to the reference frame and combined to generate the full 3D torso model.

We perform an optimization via *pose graph* to find the transformations [4]. We construct the pose graph using the point clouds P_i as the nodes, and the pairwise rigid trans-

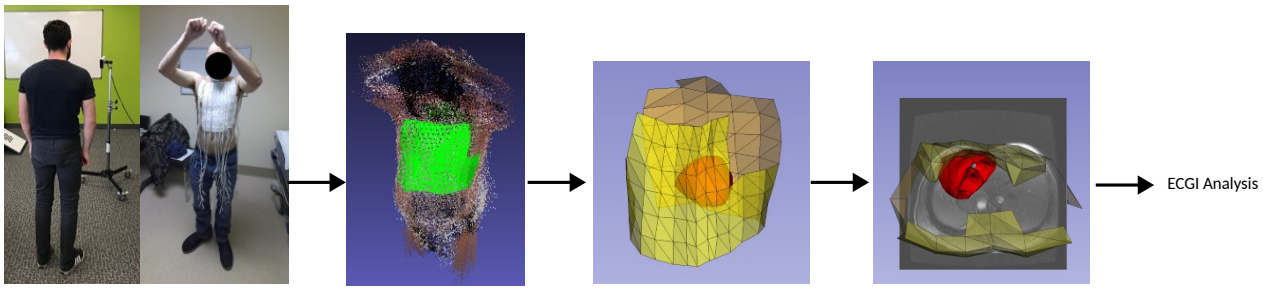

Figure 1: Pipeline Overview. In the first step, the patient wears a high-density array of ECG electrodes and rotates in front of the 3D camera to acquire data. The torso model is then constructed in the camera coordinate system using the video data. Third, the torso and heart are aligned using anatomical landmarks. Finally, the torso is deformed to the skin surface, after which it is ready for use with ECG imaging.

forms $R_{i,i-1}$ from the point clouds i to $i-1$ as the directed edges. P_0 is chosen as the reference point cloud. For each of the remaining point clouds, we calculate the pairwise transformation $R_{i,i-1}$ between point clouds P_i and P_{i-1}, then add the node P_i and the directed edge $R_{i,i-1}$ to the graph. We also estimate the global transform R_i by accumulating the pairwise transformations that were estimated up to node i, as shown in Equation 1.

$$R_i = \prod_{j=1}^{i} R_{j,j-1} \tag{1}$$

Finally, we add an additional edge between P_0 and P_{N-1} to generate a circular pose graph. The objective function that we aim to minimize describes the cost of aligning point clouds created from consecutive video frames:

$$E(\mathbf{R}, \mathbf{K}) = \sum_{i} \sum_{(\mathbf{p},\mathbf{q}) \in K_{ij}} ||R_i \mathbf{p} - R_{i-1}\mathbf{q}||^2 \tag{2}$$

where $\mathbf{R}$ is the set of global transformations R_i of point clouds P_i to the reference frame, and $\mathbf{K}$ is a set of sets, in which each element K_{ij} is a set containing the 3D point correspondences between two point clouds P_i and P_j. A least-squares optimization is performed to minimize $E(\mathbf{R}, \mathbf{K})$ and find all the $R_{i,i-1}$ which reduce the correspondence error between electrodes of the same index in the reference frame. Once all electrode point clouds are transformed to the reference frame, they are combined to generate the camera-based torso model.

Heart-Torso Registration: In order to perform ECG imaging, the torso model must be aligned to the patient-specific heart model, which we derive from cardiac scans. We align the camera and scan coordinate systems by estimating a rigid-body transformation between a set of anatomical landmarks observed in both coordinate systems. We choose the landmark set based on the members' proximity to the heart and their ability to constrain the orientation of the final transformation. The set consists of 1) centroids of Spine Vertebrae T1-T9 2) the Sternum Tip (point where the sternum ends in the center of the chest) and 3) the Left and Right Lung Tops (the highest points in each lung) for a total of 12 landmarks. We utilize a deep learning-based detector introduced in [5], which predicts the 3D location of internal anatomical landmarks from a 2D depth image of a patient, to get the camera coordinate system's landmarks. The corresponding landmarks are annotated in the scan data. Once the 3D positions of both sets of landmarks are identified, we compute the rigid transformation between the two point sets by employing the Procrustes Method [6]. The ensuing transformation can then be applied to align the torso model to the scan data.

Deformation to Skin Surface: After aligning the camera-based torso model to the scan data, we deform it to be close to the skin surface (as shown in the scan data) in order to account for deformations in the body shape when moving between the two data acquisitions. We iteratively move the electrodes close to the skin while preserving the general form of the torso by minimizing an energy functional, inspired by [7], that combines features of the scan data and the torso's shape. The full energy formulation is:

$$\begin{aligned} E_{electrodes} = &-\frac{1}{2}||G_\sigma(x,y) * \nabla I(x,y)||^2 \\ &-\frac{1}{2}||G_\sigma(y,z) * \nabla I(y,z)||^2 \\ &+\sum_{i=0}^{p-1}\sum_{j=0}^{n-1}(K_D[i,j] - K_R[i,j])^2 \end{aligned} \tag{3}$$

where $G_\sigma(x,y)$ and $G_\sigma(y,z)$ are Gaussian kernels parameterized by variance σ for the axial and sagittal views

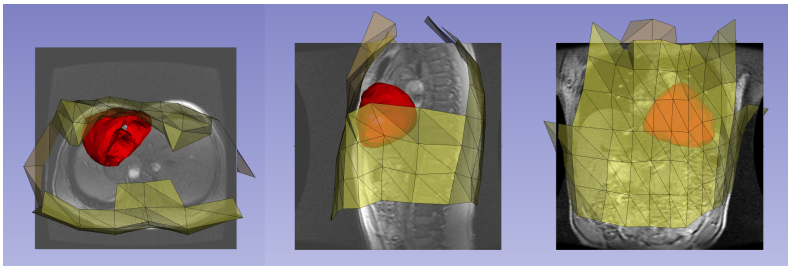

Figure 2: The camera-generated torso model, heart model, and scan data all aligned in the scan coordinate system. Axial, sagittal, and coronal views are shown.

respectively, $\nabla I(x, y)$ and $\nabla I(y, z)$ are the image gradients for the axial and sagittal planes, and $*$ is the convolution operator. p is the total number of electrodes and n is the number of neighboring electrodes in the space surrounding each central electrode. Finally, K_R and K_D are [p,n] matrices containing the distances of node i to its neighbors in 3D space.

The external energy, summarized by the first two terms, calculates the sum of the lead energies when they are orthographically projected onto the two Gaussian-smoothed image gradients. The skin surface in the image gradient is prominent due to the large difference between the anatomy and the background in the original image, so the minimizer draws the electrodes toward that feature. The internal energy, characterized by the last term, measures the inter-electrode ground truth distances before and after deformation and ensures that the torso shape is not unrealistically deformed.

The coordinate axes here are the left-right, back-front, and feet-head axes. We choose to model the deformation using a full similarity transform M, parameterized by three rotations ρ, ϕ and θ, three translations t_x, t_y, and t_z, and three scale parameters s_x, s_y, s_z. The optimal deformation transform is found by minimizing $E_{electrodes}(M(\mathbf{P}))$ using genetic algorithms, where $\mathbf{P}$ is the aligned torso model. After deforming the torso with the optimal transform, the electrodes are sufficiently close to the skin that ECG imaging can be performed.

3. Experiments and Results

We evaluate the accuracy of the camera-generated torso against the torso derived from thorax scans of five patients. We compute the Euclidean distances between corresponding electrodes in the camera and scan data, and plot the distribution of the localization errors for all patients in Figure 3. Four out of the five patients have mean errors of 50 millimeters or lower, which is the distance between consecutive electrodes on the same strip. This indicates that torsos generated through the camera-based method can expect to have approximately 1 electrode distance worth of error compared a scan-generated torso.

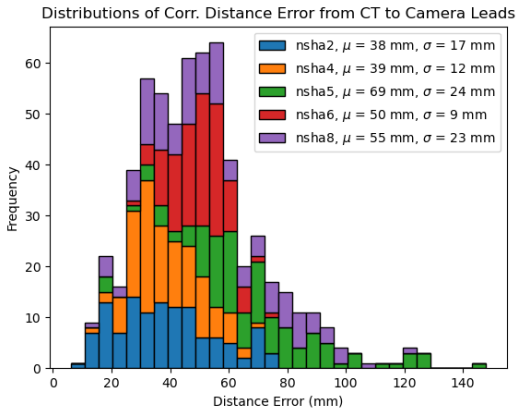

Figure 3: Correspondence Error Distribution

In latitudinal analysis, we analyze the correspondence error of electrodes using their location along the circumference of the patient's body. We group the electrode strips by the body region of that they are applied on, with the first region starting under the patient's right arm, and the remaining regions defined counter-clockwise along the patient's circumference. Table 1 shows the grouping and the average inter-lead distance error for each group. Areas where the ribs are more prominent, such as the front of the chest, have high rigidity and prevent the strips from deforming. The softer areas on the side and center of the torso allow for shape variation, which increases the electrode correspondence error. This is reflected in the data, which shows the Front Middle group having the least error, and the softer areas of the Back Right, Front Right and Back Left having the most error.

We also conducted a longitudinal analysis to observe how the electrode correspondence error varies from the top of the lead strip to the bottom. We align all strips using their respective bottom electrodes, and assign intra-strip electrode indexes in descending order from the bottom electrode to the top. For each strip in a group, we

Body Region	Electrode Strips	Avg. Error (mm)
Front Right	1,2,3	5.479
Front Middle	4,5,6,7,8	3.612
Front Left	9,10,11	4.004
Back Left	12,13,14	5.308
Back Middle	15,16,17	5.082
Back Right	18	6.643

Table 1: Electrode strip assignment and average latitudinal error.

calculate the correspondence error against the ground truth and average the value across leads with the same intra-strip index. Figure 4 shows curves of the average correspondence error in each group when the strips are traversed from top to bottom. All the curves have convexities that indicate minima near the center of the strip, which is the section of the strip closest to the heart. Since positional accuracy of the electrodes is crucial to the success of the procedure, having a minimum of error near the heart is a promising indication that the computed signals will be reliable. Errors occurring away from the central region are tolerable since those electrodes have a smaller contribution to the signal.

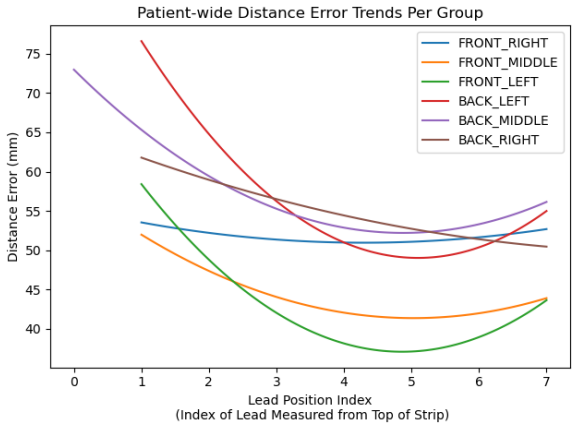

Figure 4: Longitudinal error variation of each strip group.

4. Conclusion

In our work, we presented a novel way to construct the torso geometry of a patient in preparation for an ECG imaging pipeline through the use of a 3D camera, and a way to register that torso to the heart geometry derived from the scan data. We showed that the electrode localization error is within an average range of 50 millimeters, which is less than the inter-lead distance. Latitudinal and longitudinal analyses show that the torso error is minimal in the region around the heart, indicating that the camera-generated torso can be a viable substitute to one generated from an additional thorax scan.

5. Acknowledgements

The authors would like to thank the NIH for their support of this work through grant R01HL145590. The concepts and information presented in this paper are based on research results that are not commercially available. Future commercial availability cannot be guaranteed.

References

[1] Van Dam PM, Tung R, Shivkumar K, Laks M. Quantitative localization of premature ventricular contractions using myocardial activation ECGI from the standard 12-lead electrocardiogram. Journal of Electrocardiology November 2013;46(6):574–579. ISSN 00220736.

[2] Wang L, Gharbia OA, Nazarian S, Horáček BM, Sapp JL. Non-invasive epicardial and endocardial electrocardiographic imaging for scar-related ventricular tachycardia. EP Europace September 2018; 20(FI2):f263–f272.

[3] Ghanem RN, Ramanathan C, Jia P, Rudy Y. Heart-surface reconstruction and ECG electrodes localization using fluoroscopy, epipolar geometry and stereovision: application to noninvasive imaging of cardiac electrical activity. IEEE transactions on medical imaging October 2003;22(10):1307–1318.

[4] Sungjoon Choi, Zhou QY, Koltun V. Robust reconstruction of indoor scenes. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, June 2015; 5556–5565.

[5] Singh V, Ma K, Tamersoy B, Chang YJ, Wimmer A, O'Donnell T, Chen T. DARWIN: Deformable patient avatar representation with deep image network. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017; .

[6] Gower JC, Dijksterhuis GB. Procrustes problems. Number 30 in Oxford statistical science series. Oxford ; New York: Oxford University Press, 2004. OCLC: ocm53156636.

[7] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. International Journal of Computer Vision January 1988;1(4):321–331.

Address for correspondence:

Nikhil Shenoy
755 College Road East, Princeton, NJ 08540
nikhil.shenoy@siemens-healthineers.com