Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Instance Selection for Estimation of Epicardial Activation Sequence from Venous Catheter Measurements

Bülent Yılmaz[1] and Robert S. MacLeod[2]
[1]Biomedical Engineering Department, Baskent University
Baglica Kampusu Eskisehir Yolu 20. km, 06530, Ankara, Turkey
[2]Nora Eccles Harrison Cardiovascular Research and Training Institute (CVRTI)
Scientific Computing and Imaging Institute (SCI)
Department of Bioengineering, University of Utah, Salt Lake City, Utah, U.S.A.

Abstract— **Catheter-based electrophysiological studies of the outer surface of the heart (epicardium) are limited to regions near the heart vessels or require transthoracic access. We have developed a statistical signal processing approach by which to estimate high-resolution epicardial activation maps from multi-electrode venous catheter measurements. This technique uses a linear minimum mean-squared Bayesian estimation model that derives a relationship between venous catheter measurements and unmeasured epicardial sites from a set of previously recorded, high-resolution epicardial activation-time maps used as a training data set. The training data set selection consisted of choosing a subset of epicardial activation-time maps from a database that could be used in all possible test cases with focal ectopic activity. In this study, our hypothesis was that the number of maps necessary for successful estimation could be reduced without a significant loss of performance. We developed three approaches for this purpose. Our results showed that 100 maps would be sufficient to obtain an estimation accuracy level that was better than all 470 maps paced from all over the epicardium. The results suggest that such an approach is feasible for providing accurate reconstruction of complete epicardial activation-time maps in a clinical setting and with fewer maps we can obtain similar reconstruction accuracy levels.**

Index Terms— **Instance selection, training set selection, linear estimation, ventricular arrhythmias, catheter mapping**

I. Introduction

Cardiac mapping is the collection of methods used to correctly localize the origin of heart rhythm disturbances (arrhythmias) and direct the ablation procedure, *i.e.,* burning of the tissue using high RF energy. A source of error in most of the existing cardiac mapping approaches is that they are not capable of acquiring data from the outer surface of the heart (epicardium) even though arrhythmic substrates involving epicardial and subepicardial layers account for about 15% of the ventricular tachycardias [1]. In this subgroup of patients, mapping techniques that are limited to the inner surface of the heart (endocardium) result in localization errors and failure in subsequent ablation procedures. One approach targeting the epicardium uses the multielectrode catheters placed in the heart veins [2]. However, signals from venous catheters are limited to nearby regions missing most of the epicardium.

To overcome this limitation, we developed a statistical estimation technique to reconstruct the activation pattern over the entire epicardium using sparse venous catheter electrode recordings. The technique was a linear least-squares estimator based on a set of previously recorded, high-resolution epicardial activation-time maps used as a training data set and therefore the training set becomes a key component in this approach.

We previously showed that signal morphology and activation time values from venous catheters recordings were highly correlated with those from nearby epicardial sites [3]. We selected a subset of electrodes that lay near the coronary veins from a high-resolution epicardial electrode array and treated them as surrogates for true catheter measurements.

The topic of training data set content has challenged investigators for many years in a large number of application areas, for example, pattern recognition, pattern classification, medical image segmentation, and remote sensing. The general goal of training data set selection is to determine the relevant data to employ in a prediction algorithm in the situation in which there is more data than necessary. The assumption is that in most cases all data are not equally useful in the training phase of a learning algorithm. Starting from a data set consisting of all the data, training set selection algorithm finds a suitable subset to be used in the learning algorithm.

In a previous study [4] with fewer maps in the database, we found that selection of the training data set had a bearing on the accuracy of the resulting estimation and set out to examine this behavior in more detail. The ultimate goal of this study was to determine the best number and location of pacing sites from which we should obtain the epicardial activation-time training maps so as to minimize the data acquisitions required. Our aim was to minimize the redundancy in the database and to be able to guide the eventual procedures required to obtain training data from patients.

In this study, our specific hypothesis was that there was a redundancy in the full training data set and the number of maps necessary for successful estimation could be reduced. For this purpose, we examined the best number of activation maps to be included in a general purpose training set.

II. METHODS

A. Experimental Setup

In all the experiments from this study we used a 490-electrode sock array (average inter-electrode distance was 4.3 mm) to record epicardial electrograms from dog hearts. The locations of the coronary vessels, from which we defined the 42-lead surrogate venous catheter subset, were also indicated on the mold, as shown in Figure 1.

We performed 14 dog experiments with various interventions, which were approved by our institution's animal care and use committee, to create an epicardial activation-time map database. Our database included 592 epicardial activation maps. These maps were separated into two sets; training and test data sets. The training data set was obtained by pacing a total of 470 different ventricular sites (maps with 239 right ventricular (RV) and 231 left ventricular (LV) pacing sites) in 12 dogs. The test set included data from two experiments (53 maps with RV and 69 with LV pacing sites).

The custom-built measurement system for the study was capable of sampling and saving continuously to magnetic disk up to 1024 channels with 1 kHz sampling rate and 12-bit resolution. Processing of the resulting recordings consisted of selecting one representative and relatively high-quality beat from each 3-second recording. Determination of activation times was by means of finding the time of the minimum slope during the QRS complex of each of the electrograms.

B. Linear MMSE estimation

Details of the linear minimum mean-squared Bayesian estimation algorithm applied to this problem have been reported elsewhere [4]. Briefly, we first defined the training database and selected the surrogate catheter leadset (L leads). We assumed that those L leads contained "known" values (surrogates for the venous catheter leads) and the remaining 490-L leads (for which we wished to estimate values) contained "unknown" activation values. We reordered the training set in such a way that the known values comprised the first L rows and then calculated the covariance matrix. The transformation matrix, $\mathbf{T}$, is formed by solving the simple matrix equation, $\mathbf{T} = \mathbf{C_{ku}}^T \ \mathbf{C_{kk}}^{-1}$, where $\mathbf{C_{kk}}$ is the auto-covariance of the known leads and $\mathbf{C_{uk}}$ is the cross-covariance of known and unknown leads.

Left multiplication by $\mathbf{T}$ of any measurement vector of the known leads yields an estimate of the values at all remaining sites and thus a complete, high-resolution map. In the computation of the inverse of C_{kk}, we used the truncated-singular value decomposition (SVD) technique. We employed the "information index" metric to determine a threshold for the number of eigenvectors and eigenvalues to be included in the inverse computation. We set the threshold at 99.1% that means we included the eigenvectors which had the largest eigenvalues whose total sum constituted 99.1% of the sum of all the eigenvalues.

C. Testing paradigms and error metrics

To evaluate the performance of the estimator, we used a "separate-test-set" protocol (STest), in which training data did not include any maps from the hearts that were used to obtain test data. Comparing the test map to the associated estimate for each of the maps in the database provided a means of computing overall statistics that included beats from a range of pacing sites.

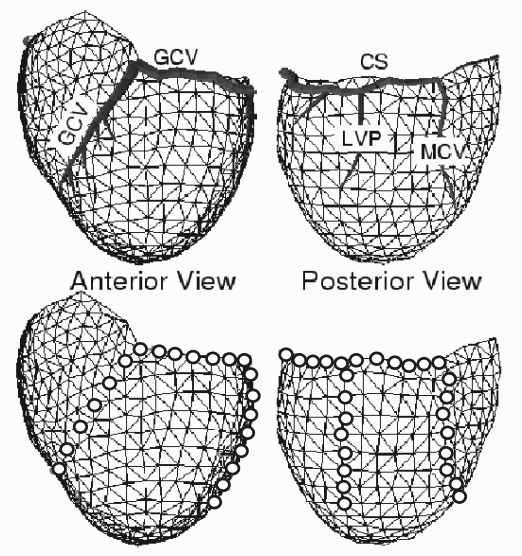

Fig. 1. A diagram representing the 490-electrode epicardial sock. The 490 electrodes are located in the nodes of the mesh; the 42 leads used as a surrogate catheter subset are indicated by larger dots. GCV: Great Cardiac Vein, CS: Coronary Sinus, LVP: Left-ventricular Posterior Vein, MCV: Middle Cardiac Vein.

The Euclidean distance between the actual and the estimated site of earliest activation, LDist, served as an index for error in each map estimation.

D. Generalized training set selection

To find a generalized training data set, we investigated three different selection methods, which were adopted from the data reduction literature in machine learning [4]. The first method, which we referred to as "single-map-addition", consisted of adding one map at a time to an initial training set with 10 maps whose pacing sites were evenly distributed over the epicardium. The addition metric was the mean LDist and the selection algorithm had four steps: (1) Add one map from the full database (one out of 460 maps in the first iteration), (2) apply estimation using the maps included in the current training set (11 maps in the first iteration) on the separate test set (122 maps used in STest testing paradigm), (3) compute mean LDist over the test maps, and (4) repeat steps 1 to 3 and determine the map whose addition resulted in the best performance (minimum mean LDist). We performed the addition until the training set contained all the maps in the full database.

The second method, which we referred to as "single-map-removal", consisted of removing one map at a time from the 470-map training set and training the transformation matrix with the remaining maps (469 in the first iteration). We then determined the map whose removal yielded the minimum mean LDist and removed it from the data set from which

additional selection occurred. The next iteration applied the same concept with the remaining maps and determined the map that would be removed. The removal process continued until the number of maps in the training data set was 10.

The third method, which we referred to as "spatial covariance-based removal," was based on the training data set itself and the spatial covariance of the measurement sites using the training set. In this approach, we first obtained the spatial covariance using all the maps in the training database and then removed each map from this set one-by-one and computed the spatial covariance of the one-map-removed set. We compared the rows of the all-maps covariance matrix and each covariance matrix from the one-map-removed set (490-by-490 matrices) using correlation coefficient. We could determine the least contributing map to the spatial covariance by finding the largest average correlated set and thus the removed map. In the next iteration we removed one map from the remaining 469 maps, computed the spatial covariance matrix and performed the correlation comparison to determine the least contributing map. This algorithm stopped when the number of maps were 10.

III. Results

The correlation between those maps in the training database (470 maps) that had nearby pacing sites was quite high. We computed the correlation coefficients ("CC") by comparing each map with its first order neighbors and determined the mean of CC for each map. The maximum of mean CC's was 0.99, minimum was 0.64, mean was 0.91. In addition, we found that 97% of the maps had a CC that was greater than 0.80. When we compared the maps from different experiments but paced from the same site (totally 83 comparisons), mean CC was 0.88 (max 0.98 min 0.65) which showed the high correlation of ectopic foci maps even when they came from different hearts. These results indicated the redundancy in this database and motivated the need for determining the set of maps in the training data that minimized redundancy yet still maintained the best performance.

Figure 2 shows the mean LDist values with respect to the number of added maps using the single-map-addition method. With only 150 maps we could obtain an LDist value (10.8 mm) similar to that when using all the maps in the database. After the further addition of approximately 100 maps, the mean LDist did not change significantly and oscillated between 10.5 and 12 mm. This general selection method was based on including all possible combinations of maps in the training set one-by-one and evaluating them on the separate test set, therefore, it was time-consuming (approximately 70 hours of computation on SGI workstations with two 300 MHz processors).

Figure 3 shows the estimation results using the single-map-removal method. We started with 470 maps and stopped at 10 maps. Removal of maps improved estimation until the number of maps was around 100 and estimation error stayed in a similar range until there were 70 maps in the training set, after which the mean LDist increased with the removal

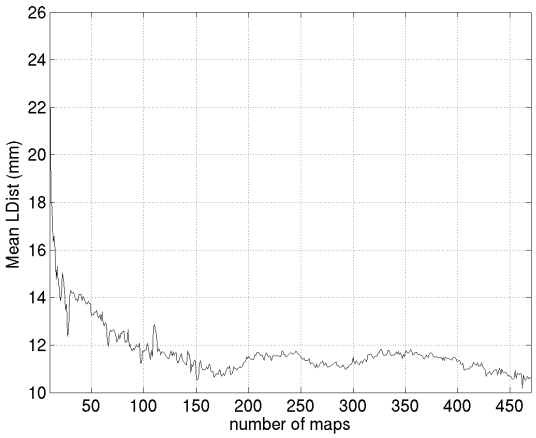

Fig. 2. Estimation performance for different number of selected maps in the training set using the single-map-addition method.

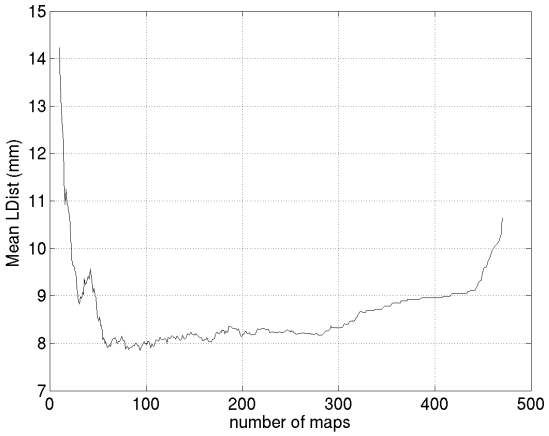

Fig. 3. Estimation performance for different number of selected maps in the training set using single-map-removal method.

of additional maps. This method resulted in a mean LDist of 7.9 mm with 100 maps compared to 10.6 mm with 470 maps (2.7 mm improvement on average). This method was also a highly time consuming (approximately 70 hours of computation).

Figure 4 shows the estimation performance with respect to the number of maps using the spatial covariance-based removal method. The figure depicts that the performance did not change significantly until the number of maps was around 150. A slight decrease in the mean LDist values occurred as we decreased the number of maps to 70. When the maps were fewer than 70, the estimation resulted in 12.5 mm and worse mean LDist. The computation took approximately 50 hours on the same processors.

Figure 5 contains a set of original and estimated maps using either 100 maps selected with the single-map-removal, single-map-addition, and spatial covariance-based removal methods or all the maps in the training database (470 maps). The test map was from the separate test-set. This figure shows the similar performance of the four training sets, one

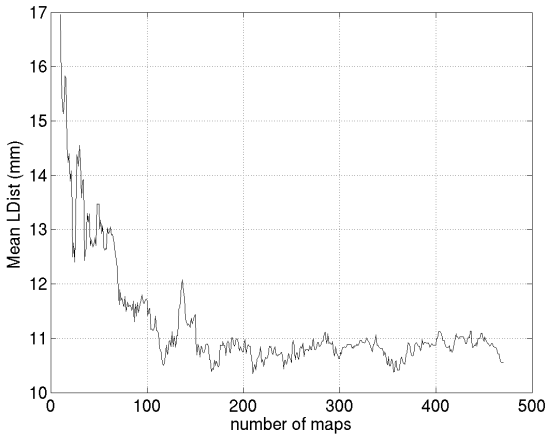

Fig. 4. Estimation performance for different number of selected maps in the training set using the spatial covariance-based removal method.

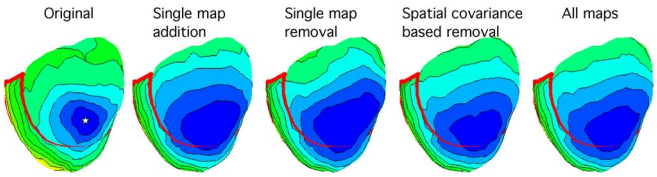

Fig. 5. Examples of estimation results using 100 maps selected with the single-map-removal, single-map-addition, and spatial covariance-based removal methods and all the maps in the training database.

of size 100 and the other 470 on a test map with a pacing site on the posterior RV.

IV. Discussion

The generalized training set selection study illustrated the redundancy in our database and suggested that by including an optimal subset of the full database the estimation technique was able to perform as well as and even in some cases better than including all the maps in the database. Specifically, we proposed three different approaches for determining the optimal number of maps with minimal redundancy and the best performance to be included in the generalized training set and consistently found that around 100 maps with unique pacing sites on the epicardium could be used in the training set to estimate the high-resolution activation-time maps for epicardially originating ectopic activity.

The results of this study encourage further investigation and provides adequate evidence that an epicardial mapping approach based on intravenous catheter measurements is feasible and can provide adequate accuracy for clinical applications. With the advances in transthoracic access to the pericardial space in order to apply catheter ablation of cardiac arrhythmias [6], such an estimation approach will complement this type of treatment as a minimally invasive diagnostic technique.

V. Acknowledgment

We gratefully acknowledge Jayne Davis, Matt Allison, Dr. Phil Ershler, and Bruce Steadman at the CVRTI, for their assistance in the experiments. We thank Dr. Jeroen Stinstra at the CVRTI for his thoughtful suggestions. The support for this work comes from the Whitaker Foundation, the Nora Eccles Treadwell Foundation, and the Richard A. and Nora Eccles Harrison Fund for Cardiovascular Research.

References

[1] L. Littmann, R. H. Svenson, J. J. Gallagher, J. G. Selle, S. H. Zimmern, J. M. Fedor, and P. G. Colavita, Functional role of the epicardium in postinfarction ventricular tachycardia. *Circ.*, vol. 83, pp. 15771591, 1991.

[2] A. de Paola, W. Melo, M. Tavora, and E. Martinez, Angiographic and electrophysiological substrates for ventricular tachycardia mapping through the coronary veins, Heart (CEN), vol. 79, no. 1, pp. 5963, 1998.

[3] R.O. Kuenzler, R.S. MacLeod, B. Taccardi, Q. Ni, and R.L Lux. "Estimation of epicardial activation maps from intravascular recordings." *J. Electrocardiol.*, 32(2):77-92, April 1999.

[4] B. Yilmaz, R.S. MacLeod, S. Shome, B.B. Punske, and B. Taccardi. "Minimally invasive epicardial activation mapping from multielectrode catheter." *Proceedings of the IEEE Engineering in Medicine and Biology Society 23th Annual International Conference*, 1(supp):290-293, 2001.

[5] J.R. Cano, F. Herrera, and M. Lozano. "Using evolutionary algorithms as instance selection for data reduction in KDD: Experimental study." *IEEE Trans. Evolutionary Comp.*, 7:561575, 2003.

[6] C. Vahlhaus, J. Stypmann H.J. Bruns, T.D. Tjan, F. Janssen, M. Schafers, H.H. Scheld, O. Schober, G. Breithardt, and T. Wichter. "Direct epicardial mapping predicts the recovery of left ventricular dysfunction in chronic ischaemic myocardium." *Eur Heart J.*, 25(2):151-157, January 2004.