A General Technique for Automatic Left Ventricle Boundary Validation: Relation Between Gray Scale Cardioangiograms and Observed Boundary Errors

Jasjit S. Suri, Robert M. Haralick, and Florence H. Sheehan

This article presents an automatic left ventricle boundary validation technique using the gray scale cardioangiograms, the observed boundary errors, and the left ventricle boundaries from any source that needs to be validated. This validation technique is based on the gray scale information near the boundary of the left ventricle in the cardioangiograms. Using a mutually exclusive window of fixed size, which is centered on the left ventricle boundary vertex and along the left ventricle contour, we compute a difference in contrast value for areas of the window both inside and outside the left ventricle region. These contrast values then are regressed against the observed boundary errors. The observed boundary errors are computed using the polyline distance measure [3], [4] by comparing two sets of boundaries: boundaries estimated from any boundary estimation algorithm, and the original ground truth boundaries as traced by the cardiologist. We performed our experiments on a database of 245 patient studies, each having two frames: end-diastole (ED) and end-systole (ES). The mean boundary error before running the validation system was 4.4 mm. Using our boundary validation system, by rejecting 5% to 10% of the patient studies, the validation system results in an error of 4.0 mm for the cross-validation case and 3.85 mm for the ideal case. We show the reliability curves of our validation system by computing the probability of false alarm, probability of mis-detection, and mean predicted errors when a total of *n* patients are rejected from the database of 245 studies.

KEY WORDS: boundary, validation, angiograms, rejection, reliability.

A VALIDATION SYSTEM for left ventricle (LV) cardioangiogram analysis becomes necessary as there are large variabilities in heart rates, sizes, shapes, and contrast values.[1] Researchers have tried to build a computer-based LV boundary estimation system,[2-5] and also have evaluated the performance of the estimated boundaries, but little attention has been given to its validation. For the LV boundary estimation system, the boundary validation is a crucial step because it determines those LV boundaries from the database (which could be coming from any source) whose boundary error is above a given threshold.

In this article we develop a general and automatic validation technique to detect the LV boundaries whose mean end frame (end-diastole [ED] and end-systole [ES]) boundary errors ((ED + ES)/2) are above a given threshold, R_{th}. This validation scheme has several features and advantages. It determines those LV boundaries from the database whose ((ED + ES)/2) error is above a given threshold error. The scheme provides feedback to the boundary estimation system so that the system knows which LV boundaries can be rejected. The validation technique estimates the overall performance of the boundary estimation system without taking the rejected boundaries into consideration. The validation technique provides a check for consistency and reliability of the output boundary estimation algorithms (eg, pixel classification, calibration, active contour, and statistical segmentation algorithms). For this test, we need three types of inputs: the boundary coordinates (x, y) of the LV boundaries that need to be validated, the gray scale cardioangiograms, and the binary indicator for the LV region (1 for inside the LV region and 0 for outside the LV region). This boundary rejection scheme is based on the gray scale information near the boundary of the LV. Using a mutually exclusive window of a fixed size centered on the LV boundary vertex and along the left ventricle contour (LVC), we compute the mean gray level value for the area of the window that is inside the LV region and the area of the window that is outside the LV region (Fig 1, *left*). We then associate this difference in the mean gray scale intensities (also referred to as contrast values) to the corresponding vertex of the LV boundary to be validated. Because we know the observed boundary error ((ED + ES)/2) for a patient study estimated using polyline

From the Intelligent Systems Laboratory and Cardiovascular Research and Training Center, University of Washington, Seattle, WA.

Address reprint requests to Jasjit S. Suri, PhD, Intelligent Systems Laboratory, Department of Electrical Engineering, Box 352500, University of Washington, Seattle, WA 98195.

0897-1889/97/1003-1070$3.00/0

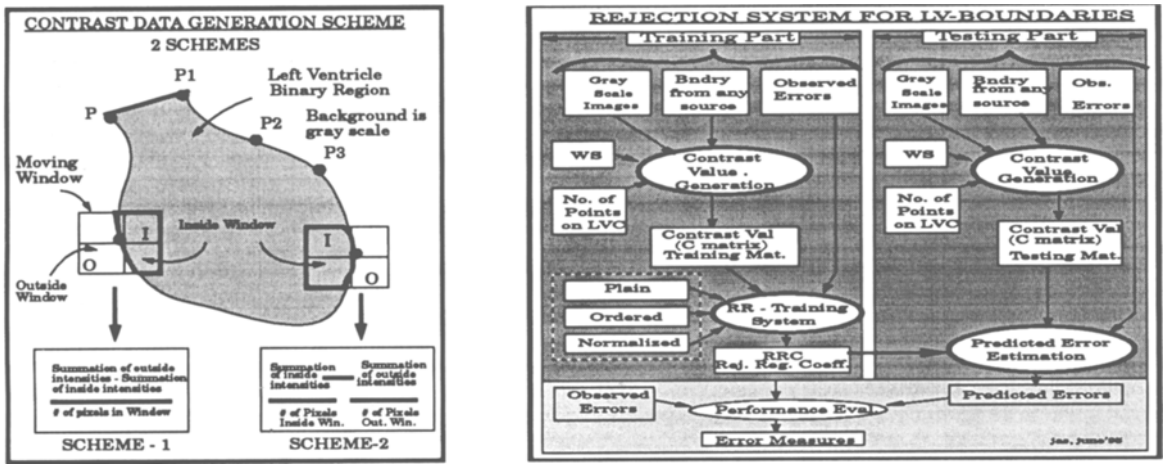

Fig 1. ***Left:*** **Contrast boundary data generation process showing two schemes: (1) non-separate contrast data and (2) separate contrast data.** ***Right:*** **Overall system for reliability and validation test. It has two parts. The first part is the generation of the training rejection coefficients from the contrast data set. The second part consists of applying the training coefficients on the test contrast data to generate the predicted errors. These predicted errors undergo the performance evaluation of the system. WS is the window size, which moves along the LVC. Three statistical techniques are used for the generation of the contrast data matrix (C).**

distance technique,[3,4] we can *regress* these contrast values against the observed boundary errors to compute the rejection training coefficients. These training coefficients then are used to compute the predicted boundary errors on the test contrast boundary data (CBD).

Our validation approach is based on a *cross-validation* procedure for estimating the predicted errors. This procedure takes a database of N patient studies and partitions this database into K equal-sized subsets. Then for all the K *choose* L combinations, we train the system using L subsets, and apply the estimated transformation on the remaining $(K - L)$ subsets. The mean predicted error of the contrast boundary data then is computed from these $(K - L)$ subsets coming from all K *choose* L combinations. The predicted errors that are above the threshold correspond to boundary delineations that are to be rejected. This rejection system is a reliability test because this system helps in determining the procedure as to how reliable the estimated boundaries are. In the following section, we give the mathematical statement for computing the predicted errors. We then discuss the reliability algorithm and the training system; followed by the mathematical formulae for computing the probability of false alarm, probability of mis-detection, and mean predicted errors when a total of n patients are rejected from the database N. We then discuss the resultant reliability curves from our experiments, and summarize our validation system.

PROBLEM STATEMENT: RELIABILITY EQUATION

We present here the mathematical statement for the estimation of the prediced errors for the LV boundaries given the contrast data matrix $\mathbf{C}$ and observed end frame error ((ED + ES)/2) vector $\mathbf{e}$. The predicted errors are used for spotting the LV boundaries that are above the given threshold.

Boundary contrast data refers to the gray scale data generated along the LV boundary by superimposing the LV boundary over the gray scale cardioangiograms. At each chosen vertex of the LVC, there is a corresponding contrast value. The contrast value for a given window is the difference between the mean gray scale intensities inside and outside the window. This superimposed boundary can be from any given boundary estimation algorithm whose output needs to be validated.

Observed boundary error is the mean end frame boundary error ((ED + ES)/2) for each patient study n. Let $c'_n = [c_1, \ldots, c_P]$ be the row vector of contrast values of dimension P for patient study n, where, $n = 1, \ldots, N$. Let $\mathbf{e} = [e_1, \ldots, e_N]$ be the vector of the observed end frame errors, ((ED + ES)/2) for N patient studies. For the pre-

dicted errors of the patient study n, we are

- **Given:** Corresponding pairs of contrast boundary data matrix $\mathbf{C}$ $[N \times (P+3)]$, and the observed error $\mathbf{e}$ $[N \times 1]$, respectively as:

$$\mathbf{C}^{N\times(P+3)} = \begin{pmatrix} c'_1 \underbrace{1_{s_1 m_1}} \\ \cdots \\ c'_N \underbrace{1_{s_N m_N}} \end{pmatrix} \mathbf{e}^{N\times 1} = \begin{pmatrix} e_1 \\ \cdots \\ e_N \end{pmatrix}$$

where s_n, m_n is the standard deviation and the mean of the P contrast values for the patient study n along the LVC.

Let $\mathbf{a}$ $[N \times 1]$ be the vector of unknown regression coefficients. The problem is to estimate the coefficient vector $\mathbf{a}$, to minimize $\|\mathbf{e} - \mathbf{Ca}\|^2$. Then for any boundary contrast data matrix $\mathbf{C}$, the predicted error for the boundary is given as: $\mathbf{C\hat{a}}$, where $\hat{\mathbf{a}}$ is the estimated coefficients.

RELIABILITY ALGORITHM: C-e RELATION

Following are algorithmic steps for estimating the predicted errors that form the basis for the reliability of the boundaries estimated from any boundary estimation algorithm.

1. Contrast boundary data generation ($\mathbf{C}$): The contrast matrix can be generated by first superimposing the LV boundaries over the gray scale cardioangiograms. We use two methods for generating the contrast data. The mathematical statements for expressing the contrast value at a vertex i are given as follows: Let g_p be the gray scale value for pixel p. Let I and O be the sets that contain the pixels inside and outside the mutually exclusive moving window. Let G_I and G_O be the sum of all the gray scale intensities for pixels that are inside and outside the window given by

$$G_I = \sum_{p\in I} g_p \;\&\; G_O = \sum_{p\in O} g_p \qquad (1)$$

Let f_I and f_O be the cardinality of the sets I and O. Using the above notations, we give the expression for the contrast value at vertex i using the two methods.

$$c_i = \left(\frac{G_I - G_O}{f_I + f_O}\right)_i (method - I),$$
$$c_i = \left(\frac{G_I}{f_I}\right)_i - \left(\frac{G_O}{f_O}\right)_i (method - II) \qquad (2)$$

Note, the difference between the methods lies in the way the sums of the gray scale intensities G_I and G_O are subtracted in the two cases. In method-I, we compute the gray scale difference between the total gray scale values inside and outside the window and then divide the difference by the total number of pixels in the window. In method-II, we find the mean gray values for inside and outside the window separately and then subtract it. The first one is named *non–separate contrast* data whereas the later is called *separate contrast* data. This is shown in Fig 1 (*left*). We also compute the mean and standard deviation of the contrast values for each patient study n. Now, using equations 1 and 2, we find the contrast values for P vertices on LVC, which yields the contrast vector c'_n for patient study n. Repeating this process for all the studies N, we generate the contrast data matrix $\mathbf{C}$ $[N \times (P+3)]$.

2. Estimation of the training coefficients ($\hat{\mathbf{a}}$): We compute the training coefficients using the standard least squares to minimize the error function ϵ^2_{rr} given as:

$$\epsilon^2_{rr} = \|\mathbf{C}_{tr}\mathbf{a} - \mathbf{e}\|^2 \;\&\; \hat{\mathbf{a}} = (\mathbf{C}^{\mathrm{T}}_{\mathrm{tr}}\mathbf{C}_{\mathrm{tr}})^{-1}\mathbf{C}^{\mathrm{T}}_{\mathrm{tr}}\mathbf{e} \qquad (3)$$

Note, $\mathbf{C_{tr}}$ has a dimension of $N_{\mathrm{tr}} \times (P+3)$, $\hat{\mathbf{a}}$ has a dimension $(P+3) \times 1$, and $\mathbf{e}$ has a dimension of $N_{\mathrm{tr}} \times 1$. N_{tr} are the number of training studies. Equation 3 is solved using singular value decomposition.[6]

3. Estimating the predicted errors $\hat{\mathbf{e}}$ on test contrast boundary data ($\mathbf{C}_{\mathrm{te}}$) with N_{te} studies is $\hat{\mathbf{e}} = \mathbf{C}_{\mathrm{te}}\hat{\mathbf{a}}$, where $\mathbf{C}_{\mathrm{te}}$ has dimension $N_{\mathrm{te}} \times (P+3)$.
4. Statistical techniques and window sizes: Here we repeat the rejection regression coefficient estimation process for three different sets of statistical techniques. They are (1) plain, (2) ordered, and (3) normalized and ordered. Plain, because we arrange the contrast values as per the vertex number of the LVC. Ordered, because we arrange the contrast values in increasing order and then use it for the validation system. Normalized, because we normalize the contrast values by their standard deviations, arrange them in ascending order, and then append them back to the sorted original contrast data vector. This makes the contrast vector of length $(2P+3)$.

For each of these techniques, we also change the window size for generation of the contrast data. Two sets of window sizes (WS) were taken into consideration namely, 11 × 11 square pixels and 22 × 22 square pixels.

PERFORMANCE EVALUATION

Probability of False Alarm and Mis-Detection

Let $(s_{(i)}, e_{(i)})$ be the pair of name and the error for a patient study from the list of ideal (observed) errors and the corresponding names. This list is the output of the boundary estimation algorithm. Similarly let $(q_{(j)}, p_{(j)})$ be the pair of name and the error for a patient study from the list of the predicted errors and the corresponding names. This list of errors results from the output of boundary rejection scheme. Note that the errors in both lists are sorted in increasing order and hence the names $s_{(i)}$ and $q_{(j)}$ of the patient boundaries are not in the same order in both the lists. These two lists will be used for finding the probabilities of mis-detection versus the probabilities of false alarm, mean errors for rejected and non-rejected patient boundaries.

Let us define mathematically the terms of the contingency table shown in the Table 1: m_{rr} is the number of patient studies truly rejected and was assigned to be rejected; m_{rs} is the number of patient studies truly rejected but was assigned to be non-rejected (selected); m_{sr} is the number of patient studies truly non-rejected (selected) but was assigned to be rejected; and m_{ss} is the number of patient studies truly non-rejected (selected) and assigned to be non-rejected (selected). These four terms can be expressed mathematically using our two input lists as

$$
\begin{aligned}
m_{rr}(n) &= \#[i \mid \exists j, (s_{(i)} = q_{(j)}),\ j > (N-n),\ i > (N-n)\\
m_{rs}(n) &= \#[i \mid \exists j, (s_{(i)} = q_{(j)}),\ j \geq (N-n),\ i > (N-n)\\
m_{sr}(n) &= \#[i \mid \exists j, (s_{(i)} = q_{(j)}),\ j > (N-n),\ i \leq (N-n)\\
m_{ss}(n) &= \#[i \mid \exists j, (s_{(i)} = q_{(j)}),\ j \leq (N-n),\ i \leq (N-n)
\end{aligned}
\tag{4}
$$

Table 1. Contingency Table

	Assigned	
True	m_{rr}	m_{rs}
True	m_{sr}	m_{ss}

Note that the elements of the contingency table are a function of each set of patient studies rejected, n. Using these definitions, we can express the probability of mis-detection $\mathbf{P}_{md}(n)$ and the probability of false alarm $\mathbf{P}_{fa}(n)$ as a function of the total number of patient studies rejected n:

$$
\begin{aligned}
\mathbf{P}_{md}(n) &= \frac{m_{rr}(n)}{m_{rr}(n) + m_{rs}(n)},\\
\mathbf{P}_{\mathrm{fa}}(n) &= \frac{m_{sr}(n)}{m_{sr}(n) + m_{ss}(n)}
\end{aligned}
\tag{5}
$$

Predicted Errors for Rejected and Non-Rejected Studies

Given the previous lists, we can express the mean error of the rejected and non-rejected patient boundaries for the ideal (observed) and cross-validation cases as follows: Let $\bar{e}_r(n)$ and $\bar{e}_{non}(n)$ be the errors for rejected and non-rejected (selected) patient boundaries for the ideal (observed) case. Let $\bar{p}_r(n)$ and $\bar{p}_{non}(n)$ be the errors for rejected and non-rejected (selected) patient boundaries for the cross-validation case. They are given as

$$
\begin{aligned}
\bar{e}_r(n) &= \frac{1}{n}\sum_{i=N-n+1}^{N} e_{(i)}\\
\bar{e}_{non}(n) &= \frac{1}{N-n}\sum_{i=1}^{N-n} \mathrm{e}_{(i)},\\
\bar{p}_r(n) &= \frac{1}{n}\sum_{i=N-n+1}^{N}\\
p_{(j)}\ \&\ \bar{p}_{non}(n) &= \frac{1}{N-n}\sum_{i=1}^{N-n} p_{(j)}
\end{aligned}
\tag{6}
$$

The rejection threshold for n patient studies is computed using the list of predicted errors and is given as $R_{\mathrm{th}}(\mathrm{n}) = p_{N-n}$.

RELATIONSHIPS AND RESULTS: RELIABILITY CURVES

The resultant reliability curves are shown in Fig 2 and are as follows:

1. $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ versus n: This is the relation between the mean error for the non-

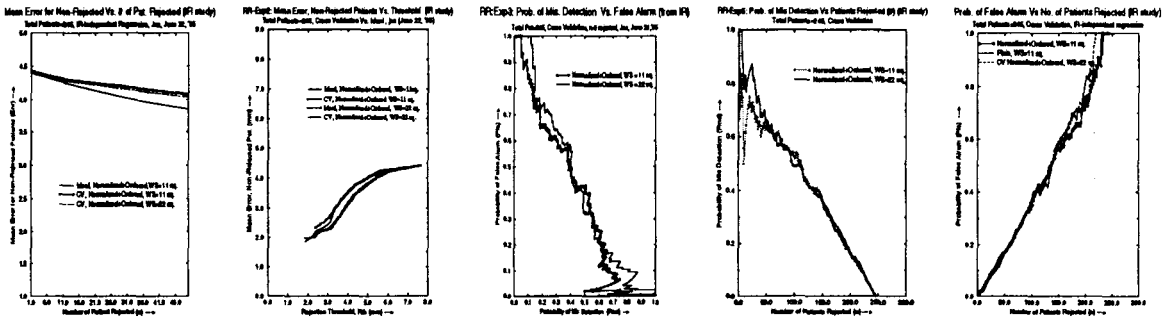

Fig 2. ***Left to right:*** **(1) Plot of $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ versus n; (2) plot of $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ versus R_{th}; (3) plot of $P_{md}(n)$ versus $P_{fa}(n)$; (4) plot of $P_{md}(n)$ versus n; (5) plot of $P_{fa}(\hat{n})$ versus n. Note that each point on the curve corresponds to the total number of patients rejected (n). There are two cases shown in these plots. First, when the window size is 11 × 11 square pixels, and second, when the window size is 22 × 22 square pixels. Calibration parameters: $N = 245$, $K = 245$, $L = 244$, $K - L = 1$, $P = 33$, $n = 5\%$ to 10% of 245.**

rejected patients for both ideal and cross-validation (CV) cases versus the total number of patients rejected n. $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ can be computed using equation 6. Our validation system inputs two quantities: Classifier boundaries,[2] which need to be validated, and the observed errors computed using polyline distance method,[3] which was the output of the left ventricle boundary calibration system[3]. Our database had 245 studies for experiments. In the first case, n is made to increase from 1 to 50 and, in the second case, we increased n from 1 to all the patients in the data base (N). We see as we rejected 10% of the patient boundaries that the mean error for $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ drops down from 4.4 mm to 4 mm (for CV case) and 3.85 mm (for Ideal case), respectively. For the plot when all the patients (N) are taken into account, there is a steep drop of the mean error to 1.98 mm. This experiment was done when the ideal errors were estimated from our boundary calibration system[3] (see plot 1, Fig 2).

2. $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ verus R_{th}: This is the relation between the mean error for the non-rejected patients for both ideal and cross-validation cases versus the rejection threshold R_{th}. $\bar{e}_{non}(n)$ and $\bar{p}_{non}(n)$ can be computed using equation 6 (see plot 2, Fig 2). With the reduction in the rejection threshold R_{th}, the mean error drops down gradually.
3. $\mathbf{P}_{md}(n)$ versus $\mathbf{P}_{fa}(n)$: This is the relation for the probability of mis-detection versus the probability of false alarm for the set of n studies rejected (see plot 3, Fig 2). With increase in $\mathbf{P}_{md}(n)$, $\mathbf{P}_{fa}(n)$ decreases. Note both scales are from 0 to 1 (see equation 5).
4. $\mathbf{P}_{md}(n)$ versus n: This is the relation for the probability of mis-detection versus the number of patients rejected n (see equation 5). As n increases, $\mathbf{P}_{md}(n)$ drops (see plot 4, Fig 2).
5. $\mathbf{P}_{fa}(n)$ versus n: This is the relation between the probability of false alarm versus the number of patients rejected n, using the relations expressed in equation 5 (see plot 5, Fig 2). With increase in n, the probability of false alarm $\mathbf{P}_{fa}(n)$ increases. We have shown two cases, first when the window size is 11 × 11 square pixels and second when the window size is 22 × 22 square pixels. When n is small, that is when a small number of patient boundaries are rejected, the $\mathbf{P}_{fa}(n)$ is more or less the same for both window size, but when large number of patients are rejected, then the $\mathbf{P}_{fa}(n)$ is higher for larger window size.

DISCUSSION AND CONCLUSIONS

We have discussed a rejection analysis approach for detecting the given boundaries that have large perturbation compared with the ground truth boundaries. We also have seen performance of the validation system when a certain set of boundaries are rejected. We discussed three statistical approaches for using the contrast value data from the gray scale cardioangiograms and we also saw that the best approach is one in which the contrast data are normalized with their standard deviation, sorted, and then appended to the original data. We also discussed the effect of window size variation on the

performance of the detection scheme, in which we saw that 22×22 performs better than 11×11. We observed that by rejecting 10% of the patient boundaries from the database, we can reduce the error by 0.4 mm. Furthermore, we validated our system by finding the probability of mis-detection versus probability of false alarm, probability of mis-detection versus the total number of patients rejected n, and probability of false alarm versus the total number of patients rejected n.

REFERENCES

1. Sheehan FH, et al: Variability in the measurement of regional left ventricular wall motion from contrast angiograms. Circulation, Ventriculography 68:550-559, 1983

2. Lee CK: Automated Boundary Tracing Using Temporal Information, [doctoral thesis]. Seattle, WA, Department of Electrical Engineering, University of Washington, 1994

3. Suri JS, Haralick RM, Sheehan FH: Correction of systematic errors in automatically produced boundaries from low contrast ventriculograms. Proceedings of ICPR Vienna, Austria, Track D, August 25-30, 1996, 4:361-365

4. Suri JS, Haralick RM, Sheehan FH: Accurate left ventricle apex position and boundary estimation from noisy ventriculograms. Proceedings of IEEE Computers in Cardiology (CinC), Indianapolis, September 8-11, 1996

5. Suri JS, Haralick RM, Sheehan FH: Greedy algorithm for error correction in automatically produced boundaries from low contrast ventriculograms. IEEE Transaction in Medical Imaging 1997 (in press)

6. Press WH, Flannery BP, Teukolsky SA, et al: Numerical Recipes in C. Cambridge, UK, Cambridge University Press, 1988