

Outlier Detection and Handling for Robust 3-D Active Shape Models Search

Karim Lekadir, Robert Merrifield, and Guang-Zhong Yang*

Abstract—This paper presents a new outlier handling method for volumetric segmentation with three-dimensional (3-D) active shape models. The method is based on a shape metric that is invariant to scaling, rotation and translation by using the ratio of inter-landmark distances as a local shape dissimilarity measure. Tolerance intervals for the descriptors are calculated from the training samples and used as a statistical tolerance model to infer the validity of the feature points. A replacement point is then suggested for each outlier based on the tolerance model and the position of the valid points. A geometrically weighted fitness measure is introduced for feature point detection, which limits the presence of outliers and improves the convergence of the proposed segmentation framework. The algorithm is immune to the extremity of the outliers and can handle a highly significant presence of erroneous feature points. The practical value of the technique is validated with 3-D magnetic resonance (MR) segmentation tasks of the carotid artery and myocardial borders of the left ventricle.

Index Terms—Active shape models, invariant shape metric, outlier handling, volumetric image segmentation.

I. INTRODUCTION

WITH the increasing capabilities of current imaging modalities, accurate quantification and visualization of three-dimensional (3-D) anatomical structures represent important tasks in clinical studies as well as imaging research. Obtaining statistically significant results often calls for accurate segmentation from increasingly large data sets. In such cases, manual segmentation is tedious and often involves significant interobserver and intraobserver variabilities.

In recent years, automatic segmentation based on statistical shape models has shown its unique strengths in exploiting *a priori* knowledge about the geometry of a class of anatomical shapes. Amongst these techniques, the active shape models (ASMs) [1] are used to capture principal shape variations across different individuals and over time from a set of labeled examples based on a point distribution model (PDM). ASMs can also be used as a deformable template to fit the model to unseen shapes in an iterative manner where pose and shape parameters are estimated using a least squares minimization approach. Specific constraints found in the model are usually

imposed to limit the shapes to valid instances. Model construction [2]–[4], as well as image segmentation [5], [6] using ASM, has been generalized to 3-D and higher dimensional cases, and the method is used extensively in clinical applications [7], [8].

In practice, the use of ASM for image segmentation is complicated by the presence of outliers, i.e., when a proportion of the generated feature points lie on incorrect boundary positions. This generally occurs when image data contains noisy, incomplete or confusing image features, which is the case for most biomedical imaging applications. Outliers can be seen as extreme values which make the least squares minimization procedures sensitive to the non-Gaussian distribution of the residuals introduced. In this case, the boundary is often dragged away from its real position and the error introduced is difficult to be corrected for at subsequent iterations. The effective handling of outliers is a major research issue for the practical applications of ASM as the accuracy and robustness of the segmentation are significantly affected in the presence of erroneous feature points.

To address this problem, a number of modifications to the original ASM have been proposed. Cootes *et al.* [9] replace the traditional least squares approach by weighted least squares fitting where varying weights are assigned to each landmark to penalize points that are further away from the current model instance than the average. In the work of Rogers and Graham [10], the weights are calculated using robust estimators applied to the shape residuals or by considering the image information at the landmark points. Duta and Sonka [11] identify outliers as points that represent a high degree of the total shape variation, and replace these points by the corresponding mean values in the shape model. In the work of Li and Chutape [12], the points that are at a distance to the current model instance beyond a predefined threshold are considered as outliers and subsequently rejected from the model fitting procedure.

Although the above methods introduce marked improvements to the ASM search, their actual performance is compromised as the number and the extent of the outliers increase, especially for volumetric segmentation. The purpose of this paper is to present a new outlier detection method that is robust to the number and extremity of the outliers. The proposed method uses a shape metric that is invariant to scaling, rotation and translation using the ratio of inter-landmark distances. This metric can be used as a local shape dissimilarity measure, and thus is ideal for outlier identification. For the proposed method, statistical tolerance intervals for the descriptors are calculated from the training samples and used as a tolerance model to infer the validity of the feature points. Instead of rejecting or replacing the identified outliers by their corresponding mean values, the configuration of the valid points (inliers) is used with the tolerance model to derive a replacement value for the

Manuscript received July 12, 2006; revised November 8, 2006. Asterisk indicates corresponding author.

K. Lekadir and R. Merrifield are with the Royal Society/Wolfson Foundation Medical Image Computing Laboratory, Department of Computing, Imperial College London, London SW7 2BZ, U.K.

*G.-Z. Yang is with the Royal Society/Wolfson Foundation Medical Image Computing Laboratory, Department of Computing, Imperial College London, London SW7 2BZ, U.K. (e-mail: g.z.yang@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2006.889726

outliers. After the correction is carried out, the shape residuals become normally distributed, thus allowing a robust shape instantiation by the ASM. The geometrical knowledge gathered using the invariant descriptors at each iteration is propagated to successive iterations for feature point detection. This limits the presence of outliers and improves the convergence of the segmentation process. The practical value of the technique is validated with different 3-D magnetic resonance (MR) segmentation tasks, including the inner vessel lumen and outer vessel wall of the carotid artery, and the endo-cardial and epi-cardial borders of the left ventricle (LV).

II. METHODS

A. Outliers in Statistical Shape Models

In ASMs, shapes in 3-D can be represented from a set of n landmarks by using a concatenated vector of the coordinates $x = (x_1, y_1, z_1, \dots, x_i, y_i, z_i, \dots, x_n, y_n, z_n)^T$. The main modes of variation are extracted from the training samples through eigen-decomposition by applying Principle Component Analysis (PCA). New instances of the shape can then be obtained by varying the shape parameters b of the following equation:

$$x = \bar{x} + Pb \quad (1)$$

where \bar{x} is the mean shape, P is a matrix containing the eigenvectors corresponding to the principle axes of variation, and b a weighting vector that varies within limits prescribed by the allowable shape domain to ensure new shapes to be valid.

ASMs can also be used as a deformable template for image segmentation, which involves feature point search, pose parameters calculation, and shape parameters estimation. The search for feature points is carried out for each landmark P_i in a local search window L_i , which usually corresponds to a set of discrete positions located perpendicularly to the current instance of the shape (h_s candidates from either side of the point). This involves the detection of the position that best corresponds to gray-level characteristics captured from the training samples. To this end, a number of different strategies have been proposed. Cootes *et al.* suggested choosing the points that minimize the Mahalanobis distance between a gray-level profile g calculated in the image and the mean profile built from the training samples [13], i.e.,

$$(g - \bar{g})^T S_g^{-1} (g - \bar{g}). \quad (2)$$

Other methods include the use of cross correlation between the mean profile derived from the model and the profile in the image [14]. Several improved strategies that involve more advanced feature point detection have also been proposed [15]–[18].

Once the feature points are detected, they must be aligned to the model by estimating the pose parameters. This is generally achieved by deriving the transformation matrix T that incorporates scaling s , rotations $(\theta_x, \theta_y, \theta_z)$ and translations (t_x, t_y, t_z) . For this purpose, the procrustes method [19] is a popular algorithm for finding the parameters that minimize the sum of

squared distances between the feature points Y and the current model shape x , i.e., by minimizing

$$(Y - T[x])^T (Y - T[x]). \quad (3)$$

Finally, the shape parameters b are calculated by minimizing the sum of squared distances between the current model shape x and the projection of the shape Y onto the model centered coordinate system

$$(T^{-1}[Y] - (\bar{x} + Pb))^T (T^{-1}[Y] - (\bar{x} + Pb)). \quad (4)$$

Both (3) and (4) involve least squares minimization, which is known to be sensitive to large residuals and performs well only under the assumption of normality [20]. Hence, when erroneous points are present amongst the feature points, significant errors in the estimation of both the pose and shape parameters can be introduced, thus leading to poor segmentation results. Fig. 1 shows typical examples illustrating the effect of the outliers on the ASM search. In Fig. 1(a), only one outlier is present amongst the feature points, in which case most of the boundaries were recovered by the ASM search, but noticeable error is introduced to the final shape near the outlier. In practice, multiple outliers can be generated during the feature point detection stage either due to artifacts or weak features [Fig. 1(b)]. In these cases, the corresponding ASM model fitting errors can increase significantly. Furthermore, these outliers can be clustered into a group [Fig. 1(c)], which can further hamper the ASM search procedure.

In practice, erroneous feature points are inevitable despite the steady improvement in image quality. The effect of the outliers on the segmentation accuracy is directly linked to their percentage as well as their extremity, i.e., the distance of the outlier to its correct position. As the influence of these two factors increases, the errors involved in pose and shape parameter estimation also accumulate. In this case, it is difficult to distinguish between the valid points and the outliers from the residual distribution as the shape is often dragged away from its real position. To tackle this problem, an invariant shape metric is proposed for outlier handling in both two-dimensional (2-D) and 3-D ASM search. The method allows the identification of outliers independently of the model fitting procedure.

B. Invariant Shape Metric

The proposed outlier handling method is based on the use of an invariant shape metric; the ratio of interlandmark distances, which can be calculated from any triplet of points (P_i, P_j, P_k) as follows:

$$r_{ijk} = \frac{\|P_i - P_j\|}{\|P_j - P_k\|} = \frac{d_{ij}}{d_{jk}}. \quad (5)$$

In addition to being fully invariant to scale, translation, and rotation, the ratio of distances has other properties that make it ideal for outlier analysis. First, it represents the relative geometrical configuration of landmark points, thus each point can be analyzed with regard to the position of other points in the

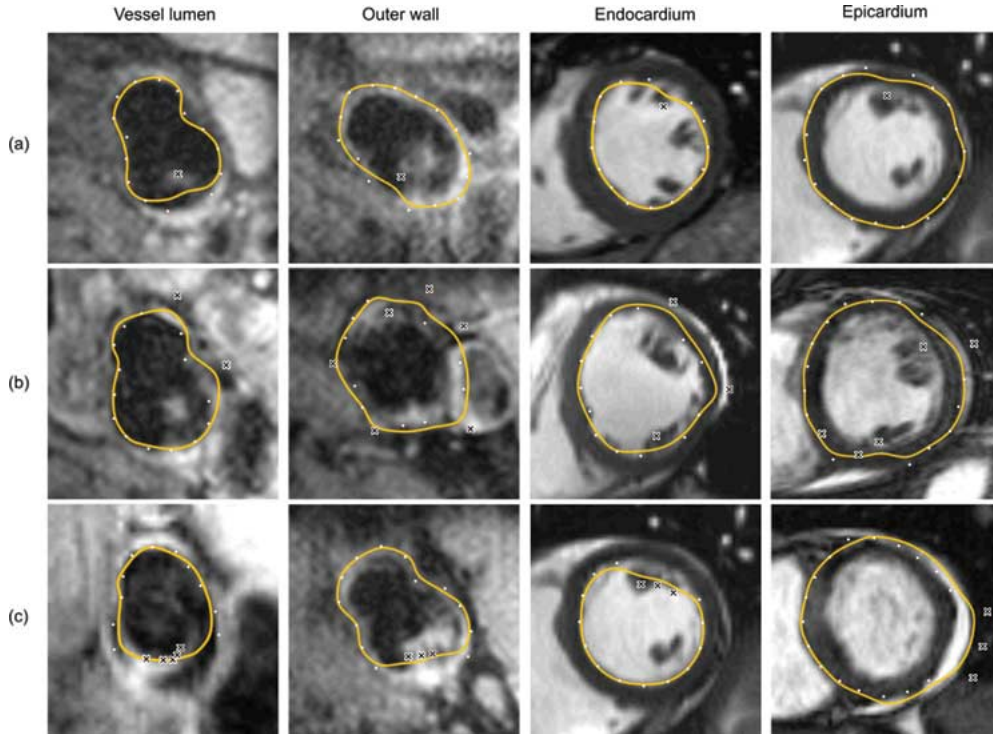


Fig. 1. Two-dimensional examples illustrating the effect of the outliers (shown as crosses) on the ASM search. Here, (a) shows examples with a single outlier, (b) with multiple outliers, and (c) with clustered outliers.

shape. Second, it allows a subdivision of the shape into subsets of points, with the aim of distinguishing the subset of erroneous points from the inliers. In each triplet of points, six different ratios can be derived, but by the use of symmetry only three are considered. With K triplets, the object can be represented by using the following vector:

$$R = (r_1, \dots, r_k, \dots, r_{3K}). \quad (6)$$

The main advantage of this descriptor is its ability to perform as a geometrical measure for shape comparison, and thus can be used as a local shape dissimilarity metric for detecting subsets of the shape that are inconsistent with the model. We show in the next section how to construct such a model and calculate the associated dissimilarity likelihood measures.

C. Tolerance Model for Outlier Detection

With the invariant shape metric presented above, each landmark is associated with a set of ratios. The idea behind the proposed method is that the outliers are inconsistent with the corresponding landmark points in the training samples and thus some of the associated ratios are invalid. The first step of the process is, therefore, to find a suitable definition for extreme or invalid ratios, which can be achieved by using tolerance analysis [21]. Statistical tolerance intervals are calculated from the training samples and any ratio found outside of the interval is considered as an extreme value, which indicates the presence of an outlier.

Let F be the cumulative distribution function of a ratio r , the corresponding statistical tolerance interval $T = [r_L, r_U]$ is a two-sided tolerance interval with β -content and γ -confidence if

$$P[F(r_U) - F(r_L) \geq \beta] = \gamma. \quad (7)$$

This equation indicates that at least a proportion β of the population will lie within the interval $[r_L, r_U]$ with confidence coefficient γ . In most applications, each ratio of the interlandmark distances can be assumed to follow a Gaussian distribution, in which case it can be shown [22] that the statistical tolerance interval, i.e., the solution of (7), can be calculated from the mean \bar{r} and standard deviation s in the following form

$$T = [\bar{r} - k_2 s, \bar{r} + k_2 s] \quad (8)$$

where k_2 is the two-sided tolerance factor, which can be approximated by [23]

$$k_2 = z_{(1-\beta)/2} \sqrt{\left(1 + \frac{1}{N}\right) \frac{N-1}{\chi_{\gamma, N-1}^2}} \quad (9)$$

where N is the sample size, $z_{(1-\beta)/2}$ is the upper $(1-\beta)/2$ quantile of the standard normal distribution, and $\chi_{\gamma, N-1}^2$ is the lower γ quantile of the chi-squared distribution with $N-1$ degrees-of-freedom.

In some cases, the distribution may not be Gaussian or known at all, depending on the datasets chosen for training and their corresponding type (normal, abnormal subjects). In such cases, nonparametric tolerance intervals should be used. They can be

calculated based on the smallest and largest observations [24] i.e.,

$$T = [\min(r), \max(r)]. \quad (10)$$

In this case, combinations of confidence and coverage coefficients that match the distribution-free tolerance interval can be calculated by using the following [25]:

$$\gamma = \sum_{k=0}^{N-2} \binom{N}{k} \beta^k (1-\beta)^{N-k}. \quad (11)$$

To use the vector of ratios in (6), some triplets must be selected amongst the $n(n-1)(n-2)/6$ possible combinations. The exhaustive use of all the ratios can be time consuming if the number of landmarks n is large. Moreover, due to covariance between landmarks, some of the ratios can have large tolerance intervals, and thus are not useful for, or even detrimental to, outlier detection. Therefore, it is more efficient to select a subset of the ratios. To this end, two possible approaches can be used. One is to group the landmarks into segments of smaller number of points with low covariance according to morphological or topological criteria. This option was used in [26] to tackle the over-fitting problem in ASM. In this case, the ratios used to construct the tolerance model are all within the selected groups. An alternative approach is to select for each landmark P_i at least K_0 ratios that have the smallest average intersection of the tolerance intervals with the local search window L_i based on the training samples. These selected ratios are ideal for outlier detection as they can capture movements of the point from its correct location. This approach was used in the experiments.

By using the tolerance interval, a likelihood measure f_r is calculated for each ratio r_{ijk} to detect extreme values due to potential outliers. Typically, this measure is equal to 1 if the ratio is within the tolerance interval and 0 otherwise

$$f_r(r_{ijk}) = \begin{cases} 1, & \text{if } r_{ijk} \in T_{ijk} \\ 0, & \text{elsewhere} \end{cases}. \quad (12)$$

Since three points are associated with each interlandmark ratio, it is not straightforward to identify which point is an outlier when a particular ratio falls outside of the tolerance interval. However, it is possible to estimate the likelihood f_p of a point to be an inlier by summing all the ratios, i.e.,

$$f_p(P_i) = \frac{1}{K_i} \sum_{j,k} f_r(r_{ijk}) f_r(r_{jki}) f_r(r_{kij}) \quad (13)$$

where K_i is the number of triplets that landmark P_i is associated with. After the calculation of f_p for all the feature points, an iterative procedure is used for the outlier detection. At each iteration, the point with the lowest likelihood measure is rejected as an outlier and the likelihood measures of the remaining points are updated by subtracting the contribution of the rejected point. The procedure is repeated until the lowest likelihood measure is

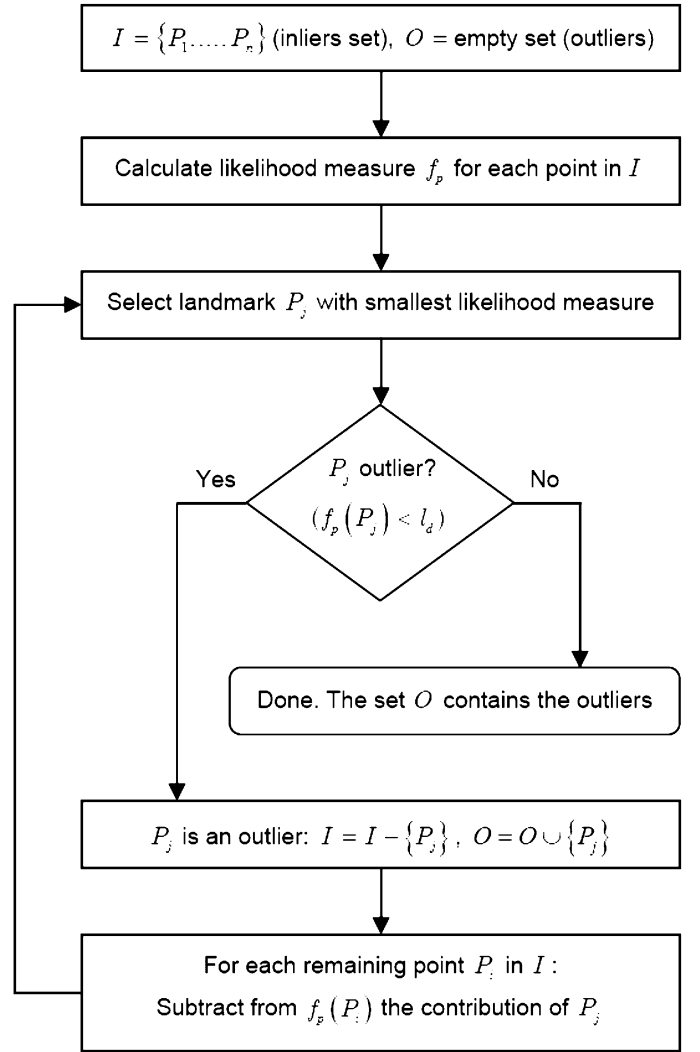


Fig. 2. Schematic diagram of the iterative outlier detection algorithm.

close to 1, suggesting all the remaining points are likely to be inliers. The output of this procedure results in two sets of points I and O representing inliers and outliers, respectively. This iterative outlier detection procedure is schematically illustrated in Fig. 2. The limit l_d used to terminate the iterative outlier detection procedure corresponds to the minimum proportion of valid ratios for each inlier. We found that choosing any limit close to the confidence coefficient γ achieves similar results, as the main outliers are detected well before reaching that limit, while the same level of uncertainty is given to both tolerance intervals calculation and outlier detection.

D. Outlier Correction

Once the outliers are detected, their positions must be adjusted so as to eliminate their effect on the ASM model fitting. A replacement point must be selected from the initial candidate points in the local search window L_i in the image space, such that its associated ratios with respect to the inliers are close or within the corresponding tolerance intervals. This is equivalent to maximizing the product of the probability density function

(p.d.f.) of the ratios as defined in (7). The function to maximize is, therefore, proportional to

$$\prod_{j,k \in I} \frac{1}{s_{ijk} \sqrt{2\pi}} \exp\left(-\frac{(r_{ijk} - \bar{r}_{ijk})^2}{2s_{ijk}^2}\right). \quad (14)$$

This can be reduced to selecting the point that minimizes the following least-squares function:

$$\sum_{j,k \in I} \left(\frac{r_{ijk} - \bar{r}_{ijk}}{s_{ijk}}\right)^2 = \sum_{j,k \in I} \left(\frac{d_{ij} - d_{jk} \bar{r}_{ijk}}{d_{jk} s_{ijk}}\right)^2. \quad (15)$$

The value suggested by the above step by using the geometrical information represents a good approximation of the true boundary position but may not always optimally fit the underlying image feature. Therefore, a final local search in the vicinity of that position is carried out using gray-level information. Only a few candidate positions around the current point are considered in this stage (h_c positions from either side of the point, $h_c = 2$ in the experiments). After this adjustment, the new plausible shape can be safely generated by the ASM model fitting procedure.

E. Geometrically Weighted Feature Detection

To prevent outliers from reappearing during subsequent feature point search, the gray-level based fitness measure is weighted by a function w that is calculated by using the geometric information gathered at the previous outlier detection procedure. The total fitness measure becomes

$$f_{\text{total}}(P_i) = d_g(P_i)w(P_i) \quad (16)$$

where d_g is the gray-level based distance measure to maximize. By using the set I of inliers, a geometric likelihood measure is drawn for each search position in the local search window L_i as follows:

$$f_g(P_i) = \frac{1}{K_i} \sum_{j,k \in I} f_r(r_{ijk})f_r(r_{jki})f_r(r_{kij}) \quad (17)$$

where K_i is the number of triplets associated with P_i . This function describes the degree of intersection of the different tolerance intervals and takes a value between 0 and 1. By setting $w = f_g$, the region around the true boundary is heavily weighted and the weight decreases for regions that are further away. Fig. 3 shows an example of a feature point search, where the normalized cross correlation was used as the gray-level fitness measure. In this example, local search based on gray-level information alone generates an outlier due to a local maximum located at an incorrect boundary position as shown in Fig. 3(a). By combining the geometric likelihood measure as shown in Fig. 3(b), the total weighted cost function of Fig. 3(c) permits the localization of the correct maximum position for the feature point.

It is also possible to limit the extent of the local search, by considering only positions in the search window with a high

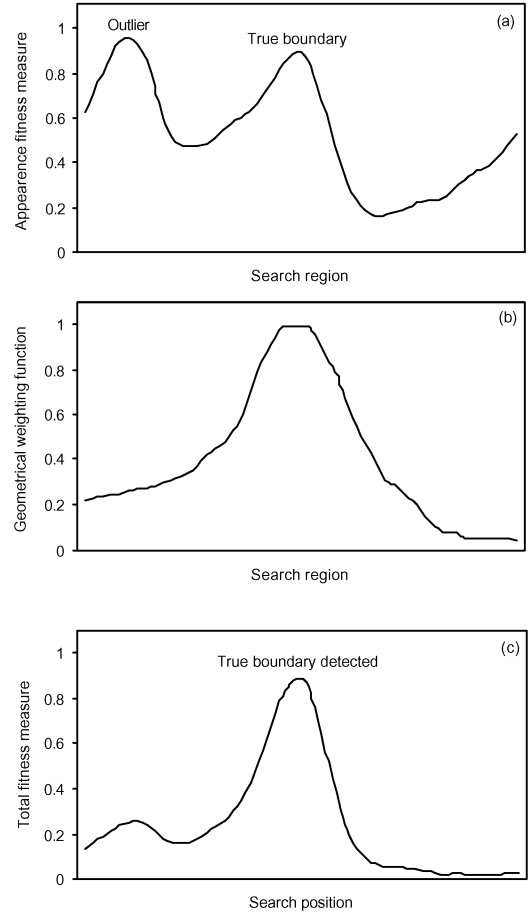


Fig. 3. Example of a feature point search where an outlier is avoided by using the geometrically weighted fitness measure.

geometric likelihood value, beyond a predefined threshold l_g . In this case, the weights are calculated as follows:

$$w(P_i) = \begin{cases} 1, & \text{if } f_g(P_i) > l_g \\ 0, & \text{otherwise} \end{cases}. \quad (18)$$

Setting $l_g = 0$ is equivalent to using the entire local search window. In general, the true boundary position lies in a region of geometric likelihood value that is close to the maximum value of f_g in the local search window. Therefore, choosing $l_g = \max(f_g) - \delta_g$ (with δ_g a small number, equal to 0.02 in the experiments) allows the window to be increasingly restricted after each iteration. This significantly limits the presence of outliers, thus allowing final convergence within a few iterations while improving the overall accuracy. This combined fitness measure is mainly required for difficult image search tasks with a high level of outliers (more than 20%), but can be discarded otherwise to decrease the time complexity of the segmentation procedure.

F. Validation

The validation of the technique was carried out on the segmentation of the endo-cardial and epi-cardial borders of the left ventricle and the luminal and outer wall borders of the carotid

TABLE I
PARAMETERS OF THE SEGMENTATION FRAMEWORK USED IN THE EXPERIMENTS

| Parameter | | Carotid artery | Left ventricle |
|------------|--|----------------|----------------|
| n | Number of landmarks | 272 | 136 |
| N | Number of datasets | 40 | 36 |
| | Part of shape variation explained by the statistical models | 0.99 | 0.99 |
| | Bounds on eigenvalues | 3 | 3 |
| h_p | Gray-level profile length on either side of the landmark point | 7 | 7 |
| h_s | Local search size on either side of the current point | 10 | 10 |
| h_c | Final search size for outlier correction on either side of the point | 2 | 2 |
| | Profiles step size (mm) | 0.2 | 1.5 |
| | Maximum number of iterations for the ASM search | 100 | 100 |
| β | Statistical tolerance intervals coverage coefficient | 0.99 | 0.99 |
| γ | Statistical tolerance intervals confidence coefficient | 0.95 | 0.95 |
| K_0 | Minimum number of ratios per point in the tolerance model | 200 | 200 |
| k_2 | Two-sided tolerance intervals factor (derived from Equation (7)) | 3.22 | 3.27 |
| l_d | Threshold for terminating the iterative outlier detection | 0.95 | 0.95 |
| δ_c | Threshold for the geometrically weighted fitness measure | 0.02 | 0.02 |

artery. These segmentation tasks have different appearance, topological, and geometrical characteristics. The lumen of the carotid artery generally has strong edges but can be affected by residual blood flow artifacts, which often introduce clustered outliers Fig. 1(c). The outer vessel wall, on the other hand, usually has poor contrast to surrounding structure and is particularly prone to noise. The same situation applies to the endo- and epi-cardial borders of the LV (Fig. 1).

In this study, the carotid artery datasets were collected from 40 subjects using a 1.5T MR scanner (Sonata, Siemens, Erlangen Germany), a purpose-built two element phased-array surface carotid coil and a specially designed head and neck cushion for immobilization. A 3-D volume-selective TSE sequence [27] was used with a pixel size of 0.47 mm and slice thickness of 2 mm. For each dataset, 20 slices around the bifurcation were selected as a region for measurement. For the LV datasets, 36 subjects were scanned using the same scanner and a TrueFISP sequence (TE = 1.5 ms, TR = 3 ms, slice thickness = 10 mm, pixel size of 1.5 to 2 mm) within a single breath-hold. The short axis stack began in the left atrium above the mitral valve ring, and was continued through the LV beyond the apex. The acquisition scheme provided a comprehensive anatomical coverage of the LV as well as its inflow and outflow tracts.

For all the datasets, manual delineation was carried out by an expert observer. The statistical shape model was built for each case with simple point correspondences using the parameters in Table I. The datasets used for gray-level appearance and tolerance model construction and evaluation were selected on a leave-one-out basis. For comparison, the proposed method, as well as existing ASM techniques (original ASM by Cootes *et al.* [1], Rogers and Graham [10], Duta and Sonka [11], and Li and Chutatape [12]) were applied to the *in vivo* datasets using the same ASM parameters (Table I) and initializations. The method in [10] does not require any additional parameter,

whereas the methods in [11] and [12] require a threshold value chosen empirically. The method in [9] is a less robust version of the method in [10] and therefore was not included in the performance comparison. For all methods, a suitable initialization was required to achieve satisfactory convergence. The mean shape was used for this purpose and placed close to the centre of the sample under investigation. The segmentation error was measured by calculating the absolute point-to-surface distance from each final point to the corresponding manual delineation.

III. RESULTS

A. Synthetic Outliers

The performance of the technique with respect to different levels of outliers, both in terms of extremity and percentage, was first assessed with synthetic outliers introduced to the manual delineation of the vessel lumen datasets. This was achieved by perturbation of randomly selected points using non-Gaussian noise. In the first experiment, the percentage of outliers was fixed to 25% of the total number of landmarks while their displacement varied between 0 and half the average vessel diameter. In the second experiment, the percentage of outliers generated from the manual annotations was varied between 0 and 60% and the standard deviation of the perturbation was fixed to 25% of the average vessel diameter. The generated points were used as feature points for a single iteration of the ASM search with the proposed outlier handling technique. The experiment was repeated 10 times, each with different random perturbation.

The reliability of the proposed outlier detection algorithm is measured by considering the number of false outlier/inlier identification using the following measure:

$$1 - 0.5 \left(\frac{\text{missed outliers}}{n_O} + \frac{\text{false outliers}}{n - n_O} \right) \quad (19)$$

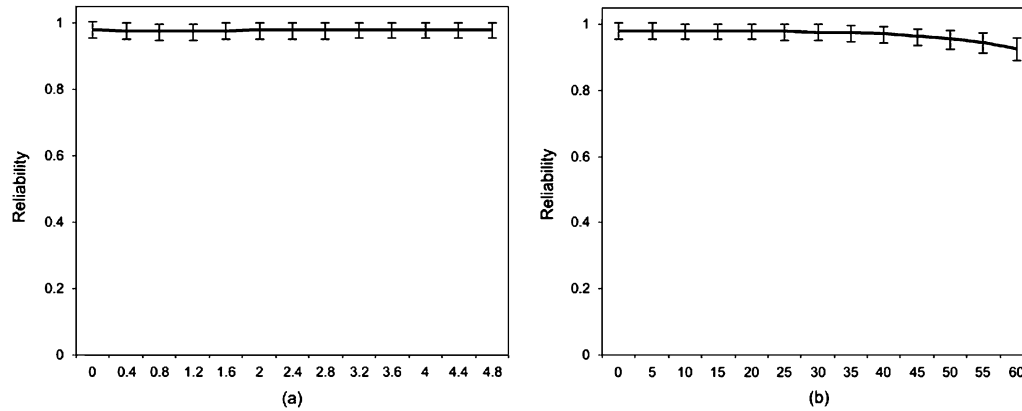


Fig. 4. Reliability measure of the proposed outlier detector in response to different (a) displacement (in mm) and (b) percentage (%).

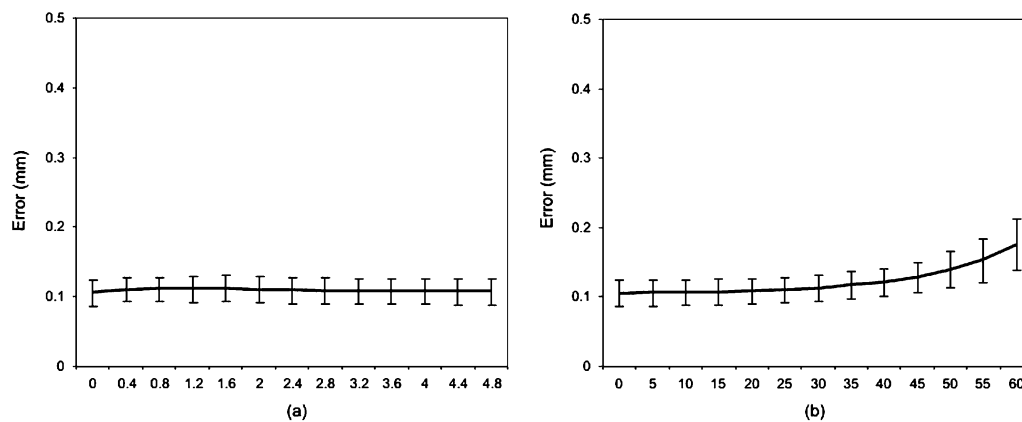


Fig. 5. Tolerance of the proposed technique against different (a) displacement (in MM) and (b) percentage (in %) of the outlier.

where n_O is the number of outliers introduced. The outlier detection step was applied to the synthetic feature points for each sample and the average reliability together with one standard deviation error bars are plotted against the displacement and the percentage of the outliers in Fig. 4. The relatively flat curves of the figure indicate that the proposed outlier detection algorithm can distinguish the outliers from the valid points with a consistently high reliability, independently of the displacements and up-to 50% outliers. This is very important as the performance of the subsequent ASM model fitting relies heavily on the correct identification of the outliers.

In Fig. 5, the segmentation errors after one iteration of the ASM search using the proposed outlier detection and correction algorithms are plotted against the displacement and the percentage of the outliers. The results show that the outlier handling method is accurate despite the extremity of the outliers and the accuracy is maintained for about up-to 50% of erroneous feature points, which shows the robustness of the technique. It can also be seen from Fig. 5(b) that the breakdown point of the algorithm is situated at about 50%, which suggests that the algorithm can handle a highly significant amount of misplaced feature points.

B. *In vivo* Data

For *in vivo* validation, the average and standard deviation of the segmentation errors are plotted in Fig. 6 for the carotids

and LV datasets. The corresponding results by using the existing ASM approaches are also provided for comparison. It can be seen from the graphs that the proposed technique outperforms the existing robust ASM methods used for comparison, particularly for the segmentation of the outer vessel walls and epicardial borders, as they are more prone to outliers, due to poorer contrast to surrounding structures and low intrinsic signal-to-noise ratio. Table II provides detailed assessment of each step involved in the proposed segmentation framework. In particular, it can be seen that the outlier correction stage improves significantly the position of the feature points before the application of the ASM model fitting procedure. It is also clear from the results presented in Table II that the use of the geometrically weighted feature detection significantly reduces the percentage of erroneous feature points, and therefore enhances the stability of the algorithm by minimizing the reoccurrence of the outliers during the iterative segmentation process.

Fig. 7 illustrates the boundary localization results for the images shown in Fig. 1, demonstrating the improvement achieved by using the proposed outlier handling algorithm and its ability in dealing with different outlier distributions. In particular, the proposed method performs well with clustered outliers [Fig. 7(c)] as the invariant metric used represents the global relative positions of the points, i.e., each point can be analyzed with respect to the position of other points in the entire shape (not only in a localized region). For detailed visual assessment in 3-D,

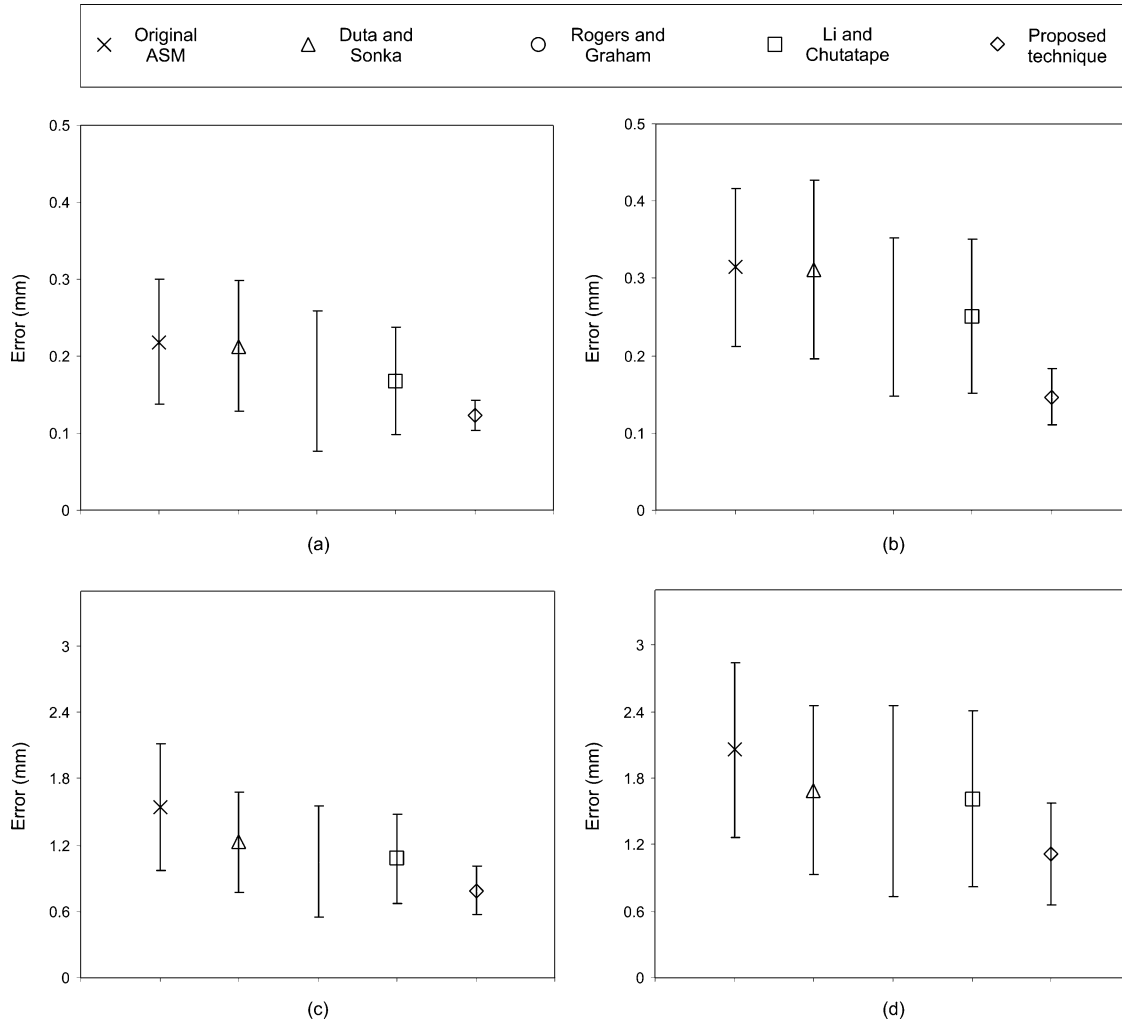


Fig. 6. Segmentation error of the proposed framework when applied to the different datasets and compared to existing robust ASM methods. The proposed technique outperforms the existing ASM approaches, especially for the outer wall of the carotid arteries and the epicardial border of the left ventricle. (a) Vessel lumen. (b) Outer wall. (c) Endocardium. (d) Epicardium.

TABLE II
DETAILED SEGMENTATION RESULTS WITH THE PROPOSED TECHNIQUE

| | Error (mm) after feature point search | Error (mm) after outlier correction | % outliers using gray-level fitness measure | % outliers using proposed fitness measure | Final segmentation error (mm) | |
|--------------|---------------------------------------|-------------------------------------|---|---|-------------------------------|--|
| | | | | | Using proposed method | Best result from existing robust ASM methods |
| Vessel lumen | 0.29 ± 0.09 | 0.16 ± 0.02 | 14.8 ± 5.8 | 4.1 ± 3.3 | 0.12 ± 0.01 | 0.16 ± 0.06 (Li and Chutatape) |
| Outer wall | 0.32 ± 0.06 | 0.19 ± 0.03 | 16.0 ± 5.7 | 3.2 ± 3.5 | 0.14 ± 0.03 | 0.25 ± 0.10 (Rogers and Graham) |
| LV endo. | 2.09 ± 0.62 | 1.25 ± 0.55 | 22.3 ± 10.6 | 6.4 ± 7.9 | 0.78 ± 0.21 | 1.05 ± 0.50 (Rogers and Graham) |
| LV epi. | 2.42 ± 0.82 | 1.52 ± 0.46 | 31.0 ± 9.7 | 9.5 ± 6.6 | 1.11 ± 0.46 | 1.59 ± 0.86 (Rogers and Graham) |

Fig. 8 illustrates how the outliers are detected and corrected for and the final errors after model fitting with and without outlier handling. It is evident that in these cases, the ASM search alone introduces significant errors, especially at regions with localized outliers, e.g., around the bifurcation point of the carotid artery.

Finally, Table III shows the convergence behaviour and time complexity of the proposed method as compared to existing ASM methods. It can be seen that with the proposed technique, convergence was reached within 25 iterations of the segmentation process for over 90% of the datasets and before reaching

the maximum number of iterations in all cases, which demonstrates the stability of the method. Although the time spent on each individual iteration is longer than the original ASM, the results also show that the overall time complexity of the proposed method is better than the existing ASM methods used for comparison.

IV. DISCUSSION AND CONCLUSION

Existing research has shown that the ASM is an effective tool for image segmentation but is susceptible to outliers. Due to the

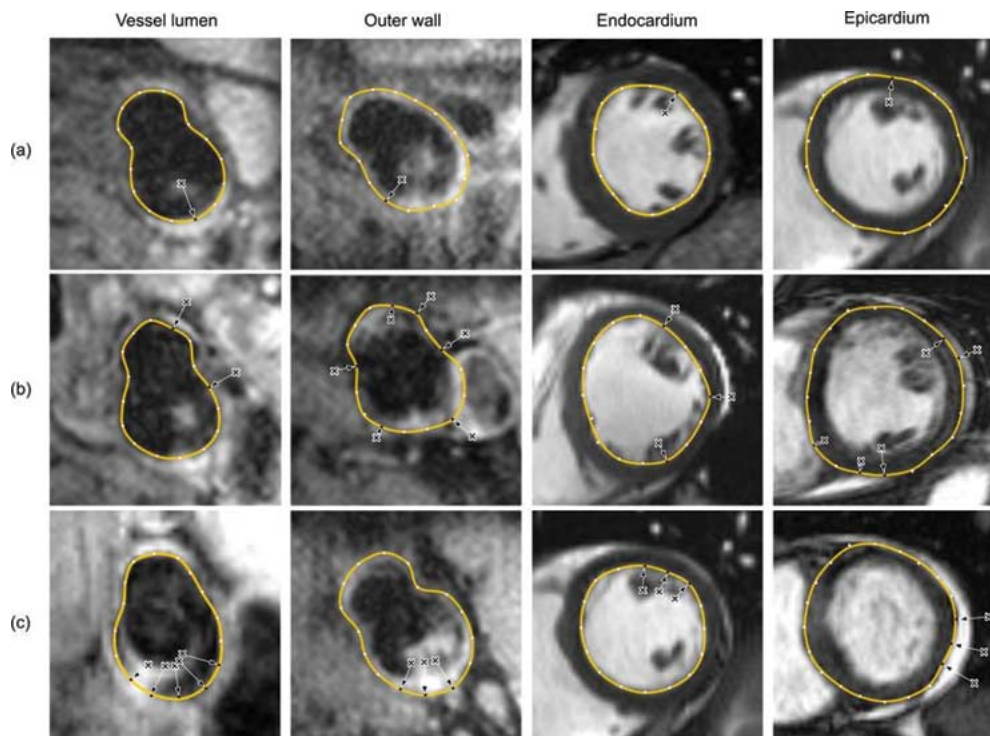


Fig. 7. Rectified contours for the examples illustrated in Fig. 1, demonstrating the improvement achieved by using the outlier handling algorithm.

relatively noisy and complex image structures as well as motion and flow artifacts, detailed 3-D segmentation in many medical imaging applications has been a challenging task.

The proposed technique presents a number of advantages over existing outlier handling methods. The use of an invariant shape metric allows outlier analysis to be based only on shape information and carried out prior to the least squares minimization procedures of the ASM. Furthermore, the identified outliers are not rejected or replaced by the corresponding mean values; a correction mechanism suggests a replacement point for each outlier before the model fitting procedure. It is conceivable to estimate the locations of the outliers based on the conditional probability conditioned on the location of the inliers. The difference between such a missing data approach and the proposed outlier correction method is that the latter suggests replacement points based on geometrical and statistical information as well as gray-level appearance information. Thus, the accuracy is increased by finding the true position of the point which best fits the image data rather than ignoring completely the variation induced by the landmark in the estimation of the shape parameters. The outlier correction approach is particularly useful when the number of outliers or the shape variability is high and also if the outliers form a local cluster.

Another important advantage of the method is that the proposed algorithm performs independently of the extent of the outliers and can handle a highly significant presence of erroneous points. The results derived from this study suggest that when the contrast of the local image features is good and the amount of outliers is relatively small, the performance of the proposed algorithm is similar to that of the existing ASM

methods, particularly to those presented in [10] and [12]. As the number of outliers increases, the advantage of the proposed method becomes more evident, as shown in Fig. 6 where existing techniques have resulted in large residual errors for the outer vessel wall of the carotids and epicardial borders of the LV. Finally, the technique uses a geometrically weighted fitness measure for feature point search which exploits outlier analysis results from successive iterations and prevents outliers from reappearing at subsequent iterations. This makes the technique more stable and allows achieving final convergence within a few iterations.

It must be noted that the performance of the proposed technique can be affected beyond 50% outliers. This is due to the fact that the algorithm detects the largest subset of points that is consistent with the statistical tolerance model. Therefore, when more than half of the feature points are outliers, it is possible that these points lie in a configuration such that they are identified as valid points. In this case, the technique can fail as it relies on a good identification of the outliers. This is the main limitation of the method but it can be argued that this situation is unlikely in practice.

In this paper, the construction of the statistical shape model and the model fitting procedure were achieved by using the ASM in its standard form. No changes were made to the ASM itself in order to show the general applicability of the method, but the proposed outlier detection technique can be applicable to other extensions of the ASM framework. Furthermore, the method was presented in its simplest form and some adaptations can be introduced. For example, a final outlier detection step could be added after the outlier correction to detect any remaining outliers, which are then rejected during the model fit-

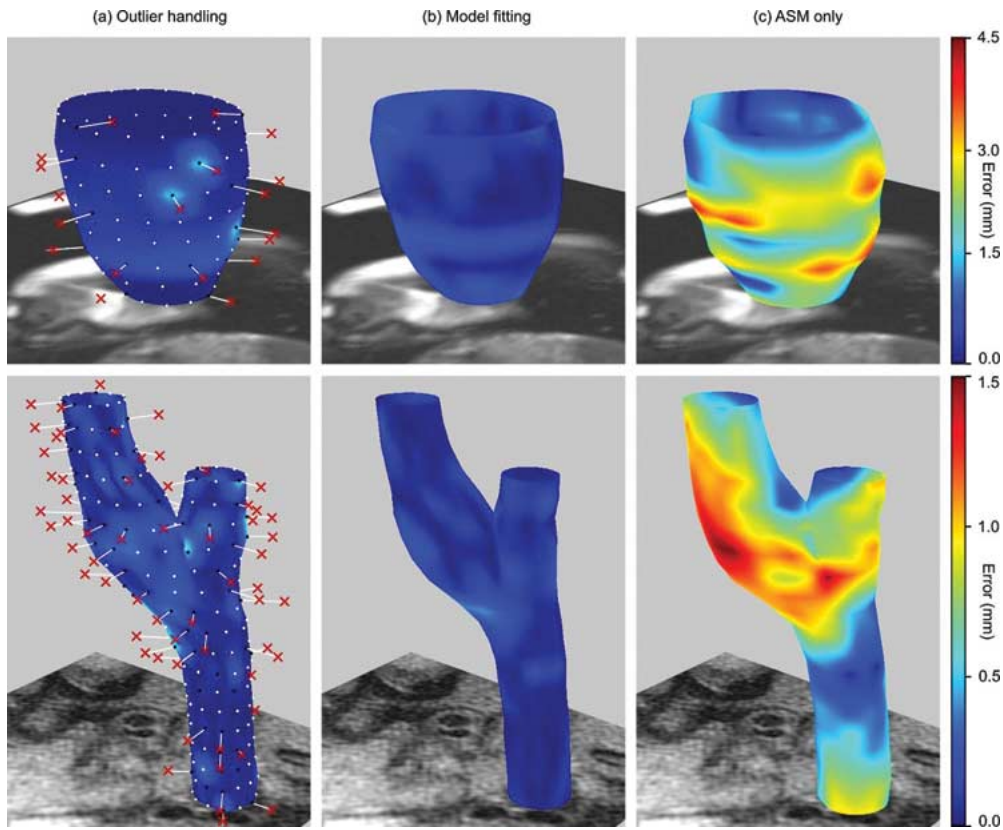


Fig. 8. Two examples illustrating the distribution and extent of 3-D surface localization errors after the application of (a) the outlier detection and correction algorithms and (b) the subsequent ASM model fitting. In (c), without the proposed outlier handling step, the ASM search introduces significant errors.

TABLE III
COMPARISON OF CONVERGENCE PROPERTIES AND TIME COMPLEXITY OF THE DIFFERENT ASM METHODS

| | Convergence for the outer wall datasets | | | Convergence for the LV epicardium datasets | | |
|--------------------|---|-----------------------------|------------------------|--|-----------------------------|------------------------|
| | within 25 iterations (%) | reached max. iterations (%) | Average time (seconds) | within 25 iterations (%) | reached max. iterations (%) | Average time (seconds) |
| Original ASM | 7.5 | 67.5 | 2.2 ± 0.6 | 8.3 | 88.8 | 1.1 ± 0.3 |
| Duta and Sonka | 20.0 | 55.0 | 2.4 ± 1.0 | 11.1 | 36.8 | 1.1 ± 0.4 |
| Rogers and Graham | 47.5 | 15.0 | 1.5 ± 1.1 | 41.6 | 11.1 | 0.9 ± 0.7 |
| Li and Chutatape | 42.5 | 7.5 | 1.4 ± 0.7 | 36.1 | 30.5 | 0.7 ± 0.5 |
| Proposed technique | 97.5 | 0 | 0.8 ± 0.3 | 91.6 | 0 | 0.5 ± 0.2 |

ting procedure. Alternatively, the outlier detection and correction stages can be iterated until no further outlier is detected. It is also worth noting that the geometrically weighted fitness measure is mainly used to limit the reoccurrence of the outliers in challenging segmentation tasks, with particularly high level of erroneous points (more than 20%).

The outlier handling algorithm presented in this paper introduces three main parameters (k_2 , l_d , and δ_g). The two-sided tolerance factor k_2 is calculated from the coverage and confidence coefficients, equal to 0.99 and 0.95 in the experiments, respectively. These are typical settings in tolerance analysis that allow a suitable coverage of most of the variation. The threshold l_d for terminating the iterative outlier detection algorithm should be chosen as equal or close to the confidence coefficient, to allow the same level of uncertainty in both tolerance intervals calcula-

tion and outlier detection. The choice of the third parameter δ_g is not critical as it is mainly used to restrict the local search window and any value around 0.05 gives fairly similar results. It can be seen from the experiments (Table I) that identical parameters are used for the carotids and LV datasets despite their differences in image characteristics and geometry. The proposed technique does not require case-specific or subject-specific tuning of these parameters.

In summary, we have presented a robust outlier handling strategy that is immune to the extremity of the outliers and performs well in the presence of a highly significant level of erroneous feature points. The experimental results on the MR volumetric segmentation of the carotid artery and the left ventricle demonstrate the robustness of the algorithm and its potential clinical value.

ACKNOWLEDGMENT

The authors would like to thank Dr. N. Keenan and Prof. D. Pennell from the CMR Unit of Royal Brompton Hospital London for providing the image data and the corresponding manual delineations.

REFERENCES

- [1] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [2] A. Hill, A. Thornham, and C. J. Taylor, "Model-based interpretation of 3-D medical images," in *Proc. Br. Mach. Vision Conf. (BMVC)*, Surrey, U.K., 1993, pp. 339–348.
- [3] P. Horkaew and G.-Z. Yang, "Optimal deformable surface models for 3-D medical image analysis," in *Information Processing in Medical Imaging*. Berlin, Germany: Springer, 2003, vol. 2732, Lecture Notes in Computer Science, pp. 13–24.
- [4] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "3-D statistical shape models using direct optimisation of description length," in *European Conference Computer Vision*. Berlin, Germany: Springer, 2002, vol. 2352, Lecture Notes in Computer Science, pp. 3–20.
- [5] B. Li and J. M. Reinhardt, "Automatic generation of 3-D shape models and their application to tomographic image segmentation," in *Proc. SPIE Conf. Med. Imag.*, San Diego, CA, 2001, vol. 4322, pp. 311–322.
- [6] M. Dickens, S. Gleason, and H. Sari-Sarraf, "Volumetric segmentation via 3-D active shape models," in *Proc. IEEE Symp. Image Anal. Interpretation*, Sante Fe, NM, 2002, pp. 248–252.
- [7] H. C. V. Assen, M. G. Danilouchkine, F. Behloul, H. J. Lamb, R. J. V. D. Geest, J. H. C. Reiber, and B. P. F. Lelieveldt, "Cardiac LV segmentation using a 3-D active shape model driven by fuzzy inference," in *Medical Image Computing and Computer Assisted Intervention*. Berlin, Germany: Springer, 2003, vol. 2878, Lecture Notes in Computer Science, pp. 533–540.
- [8] M. D. Bruijne, B. V. Ginneken, W. Bartels, M. J. V. D. Laan, J. D. Blankensteijn, W. J. Niessen, and M. A. Viergever, "Automated segmentation of abdominal aortic aneurysms in multi-spectral MR images," in *Medical Image Computing and Computer Assisted Intervention*. Berlin, Germany: Springer, 2003, vol. 2879, Lecture Notes in Computer Science, pp. 538–545.
- [9] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image Vision Comput.*, vol. 12, pp. 355–366, 1994.
- [10] M. Rogers and J. Graham, "Robust active shape model search," in *European Conference Computer Vision*. Berlin, Germany: Springer, 2002, vol. 2353, Lecture Notes in Computer Science.
- [11] N. Duta and M. Sonka, "Segmentation and interpretation of MR brain images: An improved active shape model," *IEEE Trans. Med. Imag.*, vol. 17, no. 6, pp. 1049–1067, Dec. 1998.
- [12] H. Li and O. Chutatape, "Boundary detection of optic disk by a modified ASM method," *Pattern Recognit.*, vol. 36, no. 9, pp. 2093–2104, 2003.
- [13] T. F. Cootes and C. J. Taylor, "Active shape model search using local grey-level models: A quantitative evaluation," in *Proc. Br. Mach. Vision Conf. (BMVC)*, Surrey, U.K., 1993, pp. 639–648.
- [14] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens, "Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models," *Med. Image Anal.*, vol. 6, no. 1, pp. 47–62, 2002.
- [15] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.
- [16] F. Jiao, S. Li, H. Shum, and D. Shuurmans, "Face alignment using statistical models and wavelet features," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, Jun. 2003, vol. 1, pp. 321–327.
- [17] S. Li, L. Zhu, and T. Jiang, *Medical Imaging and Augmented Reality*. Berlin, Germany: Springer, 2004, vol. 3150, Lecture Notes in Computer Science, pp. 121–128.
- [18] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, and P. Dewaele, "Active shape model-based segmentation of digital X-ray images," in *Medical Image Computing Computer Assisted Intervention*. Berlin, Germany: Springer, 1999, vol. 1679, Lecture Notes in Computer Science.
- [19] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.
- [20] V. Barnett and T. Lewis, *Outliers in Statistical Data*. New York: Wiley, 1994.
- [21] I. Guttman, *Statistical Tolerance Regions: Classical and Bayesian*. London, U.K.: Griffin, 1970.
- [22] A. Wald and J. Wolfowitz, "Tolerance limits for a normal distribution," *Ann. Math. Statist.*, vol. 17, pp. 208–215, 1946.
- [23] W. G. Howe, "Two-sided tolerance limits for normal populations—Some improvements," *J. Amer. Statist. Assoc.*, vol. 64, pp. 610–620, 1969.
- [24] J. W. Pratt and J. D. Gibbons, *Concepts of Nonparametric Theory*. New York: Springer-Verlag, 1981.
- [25] S. S. Wilks, "Determination of sample sizes for setting tolerance limits," *Ann. Math. Statist.*, vol. 12, pp. 91–96, 1941.
- [26] C. Davatzikos, X. Tao, and D. Shen, "Hierarchical active shape models, using the wavelet transform," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 414–423, Mar. 2003.
- [27] L. A. Crowe, P. Gatehouse, G.-Z. Yang, R. H. Mohiaddin, A. Varghese, C. Charrier, J. Keegan, and D. N. Firmin, "Volume selective 3-D TSE imaging for vascular wall imaging and distensibility measurement," *J. Magn. Reson. Imag.*, vol. 17, pp. 572–580, 2003.