Quantitative Evaluation and Comparison of Lung Contour Extraction Methods

José Silvestre Silva[1], Augusto Silva and Beatriz Sousa Santos

[1]*Departamento de Física*
Universidade de Coimbra
Rua Larga
3004-516 Coimbra
PORTUGAL
email: jsilva@ci.uc.pt

Abstract – **Pulmonary contours in chest CT images are closed curves that may, in some cases, exhibit multiple concave and convex irregularities. In this paper we assess the quality of pulmonary contours obtained by several fully automated methods. A comparison among these contours and corresponding manually drawn reference contours was performed and several figures of merit were assigned to each one of the automated contours extraction algorithms. Likeness between a computed contour and reference contour was measured through the Pratt figure of merit, similarity index, mean error and fraction of errors greater than 5 pixels. For each of these figures we present an exploratory statistical analysis and discuss its sensibility to the longitudinal location of the CT slice.**

I. INTRODUCTION

Contour extraction often occurs as a pre-processing step of a more global image analysis task. It happens to be the case of computer aided analysis of pulmonary X-ray tomograms [1] where many analysis algorithms start by correctly identifying each one of the pulmonary regions.

Most of the methods for pulmonary segmentation rely on intensity discrimination within the Hounsfield scale [2, 3]. This task can be complicated by the presence of spurious structures within the same scale range and/or by visual merging of the pulmonary regions themselves. In previous works [4, 5] the authors presented several methods designed to cope with these difficulties leading to lung contours with a variable degree of similarity to those provided by several radiologists.

In this paper we present several methods to quantitatively assess the similarity between the contours produced fully automatically as described in [4, 5] and those manually produced over the same set of images by human experts. Besides localized error distribution images, in order to achieve meaningful comparisons [6], each contour detection technique will be assigned several global figures of merit [7, 8]. Some aspects related with the sensitivity of each of the figures of merit will be discussed.

II. REVIEW OF COMPARED CONTOUR EXTRACTION METHODS

In previous works we developed three fully automated methods that perform pulmonary contours extraction.

The first two methods, which are from here on called *Fast Method* and *Lung's Split Method*, were based on judicious intensity discrimination [9] followed by several levels of morphologic processing [10]. The latter, besides intensity discrimination, performs a Radon Transform [11] based region identification by estimating the lung's center of mass. The algorithm is still able to split the pulmonary area when, eventually, the left and right lung appear as a unique visually merged region.

For the sake of comparison with a classical approach [12], a third method base on the active contour concept [13, 14] was also developed. As usually, this iterative approach is controlled by a energy minimization process that leads each contour point to an equilibrium state in terms of internal and external forces [4, 5].

III. METHODS

The relative performance of the contour extraction algorithms [15] was assessed by a twofold methodology based both on the localized distribution of absolute error (measured in pixels) and on global figures of merit. The reference contours were digitized thinned versions of hand drawn curves made my several radiologists.

A. Reference contours

To obtain the reference contours for the selected images we have asked expert radiologists to draw the contours on transparent sheets superimposed on hardcopies of the images. These hand drawn contours were scanned and subjected to some processing in order to identify the contours of the right and left lungs.

To register the reference curves as the right and left contours, the image Radon transforms for 0 and 90 degrees were calculated and from the maximum values of

these two transforms we obtain an estimate of the centre of each lung. Applying a morphologic filling [16] starting from the centre of one lung and a second filling starting from any point external to the lungs, we obtain an image containing the filled area of the other lung. The contour of the lung is then easily obtained. Erosion was applied to obtain a thinner version of each of the reference contours

B. Figures of merit

In order to compare two contours using the distances between them, the following steps are necessary:
1 -   find pairs of corresponding points (one from each contour);
2 -   determine distances between these points;
3 -   finally compute a global figure that characterizes the similarity of the contours.
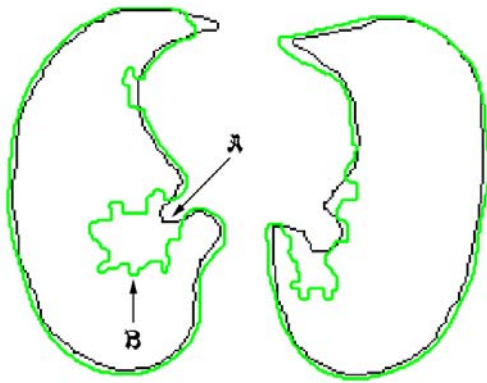

Figure 1 – Example of lung contours A and B.

Let A and B (figure 1) be two similar closed lines corresponding to the contour of a lung:

Contour A = {set of n points}
Contour B = {set of m points}

each point belonging to an image.

Let areas C, D and E be:

C = area enclosed by contour A
D = area enclosed by contour B
E = C XOR D

F is defined as a closed line existing inside area E in zones where contour A is not equal to B, coincident with these contours otherwise (figure 2).
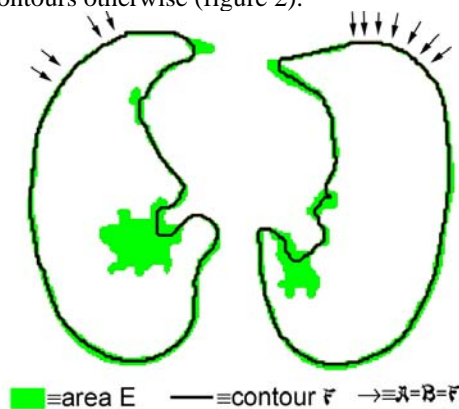

■ =area E   ── =contour F̄   → =Ā=B̄=F̄

Figure 2 – Line F (black) inside area E (grey) and a zone where F≡A≡B.

Lines perpendicular[1] to contour F are obtained for each point of this contour. The interception of each of these lines with contours A and B defines a pair of corresponding points on both contours (figure 3). This completes step 1.
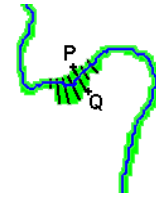

Figure 3 – Lines perpendicular to F are used to determine corresponding points (P, Q) on contours A and B.

Step 2 corresponds to the computation of Euclidean distances (in our case) between all the corresponding points.

Finally (as step 3), several figures of merit were used to assess the similarity between contours A and B:
- Pratt figure of merit,
- mean distance,
- percentage of distances > 5 pixels.
- similarity index

The Pratt figure of merit [17] is defined as:

$$F_{Pratt} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+\alpha \times d_i^2} \qquad (1)$$

where N is the number of distances used, d_i is the distance between two corresponding points and α is a parameter related to the size of the contours. The value of $\alpha = 1/9$ was chosen so that, if all the distances are equal to 3 pixels, F_{Pratt} will have the value 0.5.

The mean distance is used to obtain a global average view of the distance between contours.

The percentage of distances greater than 5 pixels (which corresponds to an error of 2% in images of 256×256) allows assessing the incidence of significant errors. Finally, the Pratt figure of merit also corresponds to a measure of the global behaviour of distances between contours, however it is a relative measure that varies in the interval [0,1].

Another possible way of comparing two contours consists in using the binary images defined by these contours and calculate their similarity.

Two binary images are defined having the pixels interior to areas C and D set to "1" and all the others set to "0". If these images have n pixels, two n dimensional vectors $\vec{H}$ and $\vec{I}$ are defined having as components the values of the pixels (1 or 0).

---

[1]Perpendicular lines are computed as orthogonal to the tangents to each point of F using neighbourhoods of 20 contour points. The size of this neighbourhood is a parameter that is related to the sensitivity to local characteristics of line F and had to be adjusted.

These binary images may be regarded as elements of a Hilbert vector space where, using the usual definitions of norm and inner product, one can obtain a morphological similarity measure based on the parameter:

$$\theta = \arccos\left(\frac{\vec{H}\cdot\vec{I}}{\left\|\vec{H}\right\|\left\|\vec{I}\right\|}\right) \qquad (2)$$

θ measures the similarity between the two contours. This value will vary from a value of 0 degrees, for coincident contours, to 90 degrees for completely different contours.

C. Visualization of differences between contours

Figures of merit provide information on the global behaviour of the differences between contours, however local information can be important in some conditions. When we compare a contour with a reference to gain insight on how to improve a detection method, analysing the values of all distances can be a valuable help. Thus we have developed a visualization technique that allows the analysis of all the distances in a simple but effective way (as shown in figure 4). The "XOR area" obtained from the two contours and the corresponding F line (described in III.B) convey an approximate idea of the zones where the two contours are more different and the amount of difference. This visual representation is complemented with a bar chart, for each lung, displaying the sequence of all the computed differences. The capital letters on both the visual representation and the bar chart provide context.

Additionally, a histogram showing the number of differences with a certain value is presented (as shown in figure 5).
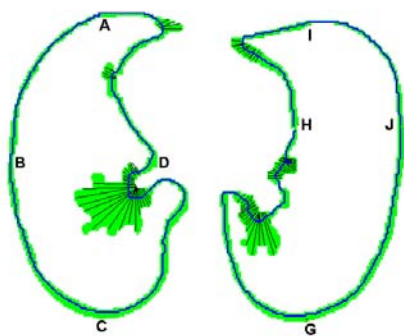

Figure 4 – Visualization of location of differences between contours including the F lines.
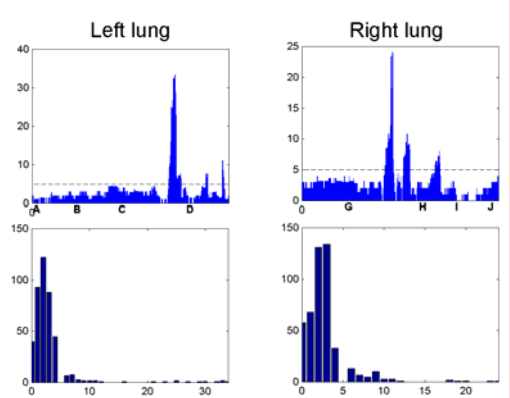

Figure 5 – Bar chart and histogram of the difference values for left and right lungs correspondi4ng to figure 4.

IV. RESULTS AND DISCUSSION

To compare the three pulmonary contours extraction techniques that we have developed [4, 5], it is necessary to quantify the accuracy of the contours obtained with these methods from chest CT images. To perform this quantification we have used the methods described previously.

The results presented in this section were obtained using a set of 32 CT images selected from two different patients with images from all regions of the lungs and as different as possible.

A. Comparison of the contour detection methods based on data visualization – worst cases

We present two examples as worst cases in lung contour comparison.

The first case, shown in figures 1, 2, 4 and 5, corresponds to a situation where the contours A and B are similar, but having a region where the difference is considerable (due to the presence of relevant vascular structures) displayed in figure 2 as a large grey stain. From figure 5 we can conclude that almost all errors are below 5 pixels with the exception of a small region near point D and between points G and H.

The histogram of figure 5 confirms that our contour detection method produces generally good results; we can see that almost all errors are below 5 pixels.

Another worst case is shown in figures 6 through 9. It corresponds to the result of the active contour method, applied to an image in the basal region, where the right lung has a very thin area not recognised as belonging to the lung.

The contour comparison technique handles this problem calculating the distances near point I, as shown in figure 8.

In the left histogram of figure 9 it is shown that the errors concerning the left lung are less or equal than 6 pixels, however most errors are in the range 3 to 5 pixels. For the

right lung, the results are identical with an exception near point I, where the error reaches 40 pixels.

These examples show that this contour comparison technique allows comparing similar contours, even when the two contours have significant differences in some regions.
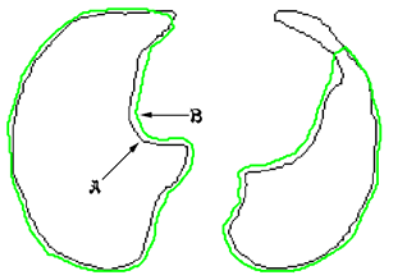

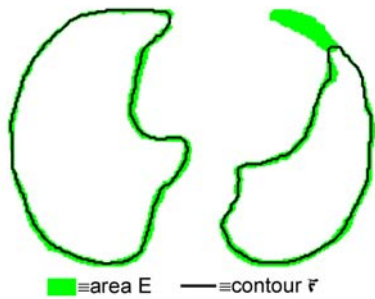
Figure 6 – Another example of lung contours A and B

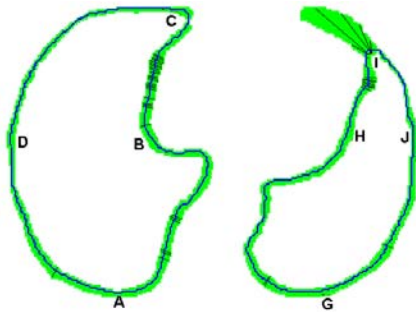Figure 7 – Line F (black) inside area E (grey).

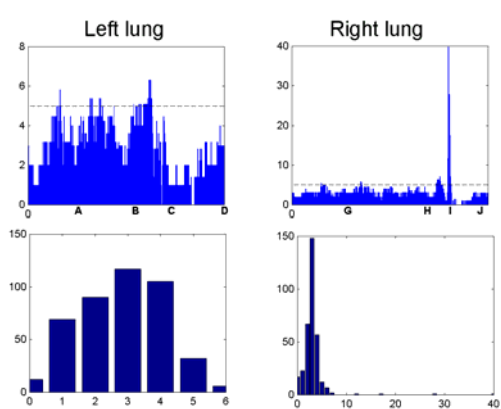Figure 8 – Visualization of the location and magnitude of differences between contours including the F lines.

Figure 9 – Bar chart and Histogram chart of the difference values for left and right lungs of figure 8.

B. Similarity index

The box-and-whiskers plot [18] of figure 10 seems to indicate that there is no significant difference among the three methods regarding the similarity index computed by equation (2), however the snake's method exhibits the larger range values, especially in the basal region of the lungs (figure 11). This is due to the difficulty in adapting the snake to a line that has multiple irregularities.
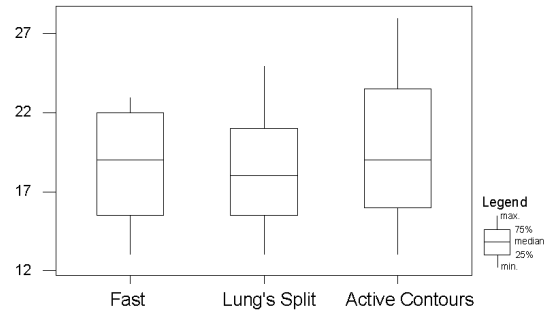

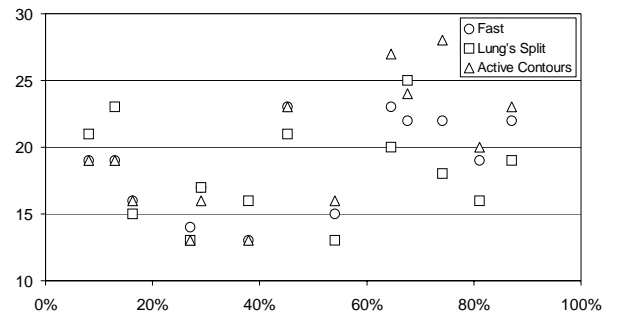Figure 10 – Box and whisker plot for similarity index.

Figure 11 – Similarity index along the pulmonary longitudinal axis (where apex region = 0% and basal region = 100%).

A value between 15 and 20 degrees for the similarity index seems to be a good result, having in mind the used scale (0 degrees = excellent; 90 degrees = very bad).

C. Pratt's figure of merit

Figure 12 shows the results obtained using the Pratt's figure of merit on the selected image set; the second method seems more consistent than the others.
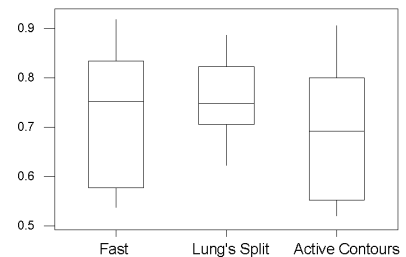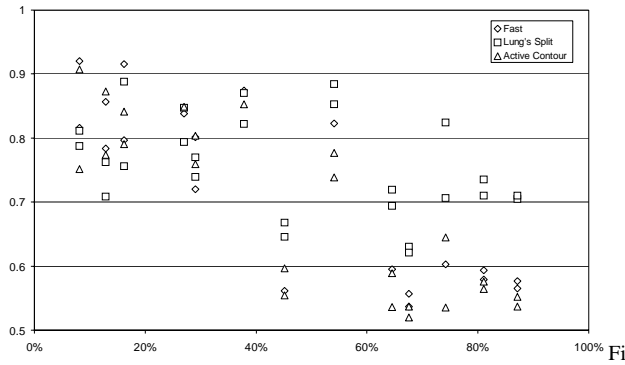

Fig. 12 - Box and whisker plot for Pratt's figure of merit.

g. 13 – Pratt's figure of merit, along the pulmonary longitudinal axis (where apex region = 0% and basal region = 100%).

This figure of merit also reveals worse performances of all the methods in slices closer to basal regions. For slices in the middle pulmonary region, the Pratt figure of merit may show some inconsistency. In fact, all our contour extraction algorithms are rather sensitive to border irregularities that often appear in this region as a consequence of the presence of relevant vascular structures.

D. Mean error

Again the lung's split method is more consistent than the others, having a smaller dispersion in the mean errors (figure 14); this is consistent with the results obtained for this method when evaluated by both other figures of merit.
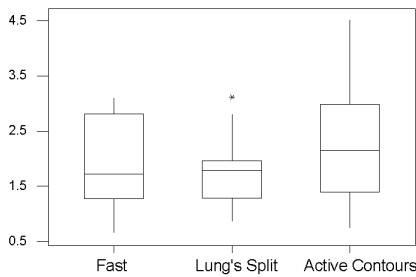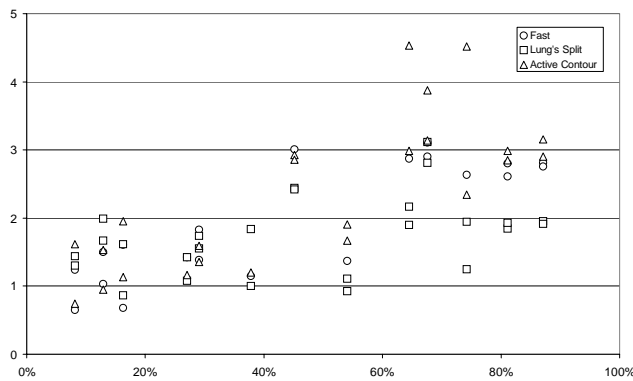

Fig 14 – Box and whisker plot for mean error.

Fig 15 - Mean error (distance) along the pulmonary longitudinal axis (where apex region = 0% and basal region = 100%).

In the apex region, the values of the mean errors are lower than the average, since that the contours of the lungs are nearly circular.

In the basal region, the errors are higher than the average, due to the concave and convex curvatures.

In the middle region once again the lack of contour smoothness and greater inter-patient variability induces a non-negligible inconsistency, as already concluded by analysing figure 13.

E. Percentage of points with error > 2% (5 pixels)

Using this figure of merit, the snakes' method gets the worst result as seen by the large dispersion in figure 16.

Figure 17 shows that the percentage of points with error higher than 2% of CT image size (5 pixels) is lower than 0.06%, which emphasises the effectiveness of the three contour detection methods.
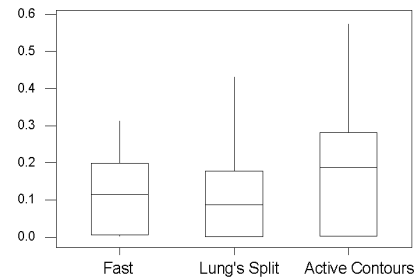

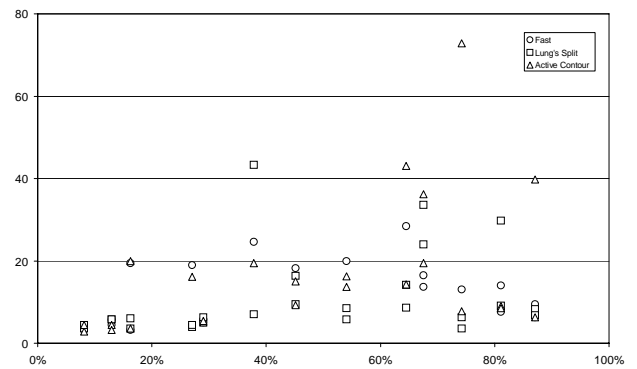Fig 16 – Box and whisker plot for percentage of points with error (distance) greater than 2% (5 pixels).

Fig. 17 - Percentage of points with error (distance) greater than 2% (5 pixels), along the pulmonary longitudinal axis (where apex region = 0% and basal region = 100%).

An overall analysis allows us to conclude that the lung's split method achieves the best results when compared with the other two methods.

V - CONCLUSIONS

We have quantitatively assessed the relative performance of three contour detection techniques previously developed for the extraction of lung contours.

The differences between the computed contours and the reference contours were evaluated by global quantifiers

such as the Pratt figure of merit, the similarity index, the mean error and percentage of points with error higher than 5 pixels.

These quantifiers were applied to all the contours obtained on 32 CT images and an overall analysis points to a global consistency among the results obtained with all these figures of merit.

The exploratory data analysis performed suggests that the lung separation technique is the best method for the task of lung contour extraction. However, in order to fully prove this claim, further statistical analysis using contours obtained on a larger image set is necessary.

ACKNOWLEDGEMENTS

REFERENCES

[1] J. Brink and et al., "Helical CT: Principles and Technical Considerations", *RadioGraphics*, vol. 14, no. 4, pp. 887 - 893, 1994.

[2] R.P. Parker, "Measurement of Basic CT Data", *Proceedings of Physical Aspects of Medical Imaging*, pp. 291-295, Manchester, U.K., 25th-27th June, 1980.

[3] A. Hasegawa, *et al.*, "A Shift-invariant Neural Network for the Lung Field Segmentation in Chest Radiography", *Journal of VLSI Signal Processing*, vol. 18, pp. 241-250, 1998.

[4] J.S. Silva, A. Silva, and B.S. Santos, "A Fast Approach to Lung Segmentation in X-ray CT Images", *Proceedings of the 11th Portuguese Conference on Pattern Recognition - RECPAD 2000*, pp. 415-418, Porto - Portugal, May 2000.

[5] J.S. Silva, A. Silva, and B.S. Santos, "Lung Segmentation Methods in X-ray CT Images", *accepted for publication in Proceedings of V Ibero-American Symposium On Pattern Recognition - SIARP'2000*, Lisbon - Portugal, September 2000.

[6] M.D. Heath, *et al.*, "A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1338-1359, 1997.

[7] A.J. Pinho and L.B. Almeida, "Figures of Merit for Quality Assessment of Binary Edge Maps", *Proceedings of the 3rd IEEE International Conference on Image Processing - ICIP'96*, vol. III, pp. 591-594, Lausanne - Switzerland, September 1996.

[8] A.J. Pinho and L.B. Almeida, "A new method for Quantitative Evaluation of Edge Maps", *Procedings of 8th Portuguese Conference on Pattern Recognition*, pp. 153-160, Guimarães - Portugal, March 1996.

[9] M.S. Brow, *et al.*, "Method for Segmentation Chest CT Image Data Using an Anatomical Model: Preliminary Results", *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 828-839, 1997.

[10] A. Watt and F. Policarpo, *The Computer Image.* Addison Wesley, 1998.

[11] S.R. Deans, *Radon and Abel Transforms.* Russ ed. Handbook of Image and Transforms, ed. Series Russ, CRC Press, pp. 631-717,7 1996.

[12] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models", *International Journal of Computer Vision*, pp. 321-331, 1988.

[13] A. Amini, T.E. Weymouth, and R.C. Jain, "Using Dynamic Programming for Solving Variational Problems in Vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 9, pp. 855-867, 1990.

[14] A. Blake and M. Isard, *Active Contours.* Springer Verlag London Limited, 1998.

[15] B. Shahraray and D.J. Anderson, "Optimal Estimation of Contour Properties by Cross-Validated Regularization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 600-610, 1989.

[16] C.R. Giardina and E.R. Dougherty, *Morphological Methods in Image and Signal Processing.* Prentice Hall, 1988.

[17] I.E. Abdou and W.K. Pratt, "Quantitative Design and Evaluation of Enhancement / Thresholding Edge Detectors", *Proceedings of the IEEE*, vol. 67, no. 5, pp. 753-763, 1979.

[18] Minitab, *User's Guide 2: Data Analysis and Quality Tools.* Minitab Inc., Statistical Software, 1998.