Classification of Pathological Shapes using Convexity Measures

Paul L. Rosin

Cardiff School of Computer Science, Cardiff University, Queen's Buildings, Newport Road, PO Box 916, Cardiff CF24 3XF, UK, Tel: +44 (0)29 2087 5585, Fax: +44 (0)29 2087 4598, Paul.Rosin@cs.cf.ac.uk

Abstract

Two new shape measures for quantifying the degree of convexity are described. When applied to assessment of skin lesions they are shown to be an effective indicator of malignancy, outperforming Lee *et al.*'s OII scale-space based irregularity measure. In addition, the new measures were applied to the classification of mammographic masses and lung field boundaries and were shown to perform well relative to a large set of common shape measures that appear in the literature such as moments, compactness, symmetry, etc.

Key words: shape, shape measure, polygon, convexity, convex hull, convexification, medical classification

1 Introduction

In image-based computer-aided diagnosis of suspected pathologies, classification is commonly determined by their colour, density, texture, morphology, etc. This paper focuses on the last characteristic, namely outline shape. Ideally, a shape measure should be non-parametric (i.e. free from tuning parameters), simple and efficient to implement and compute, robust, and invariant to transformations such as rotation, translation, and scaling. The starting point for the work described here was the paper by Lee *et al.* [16] on developing a measure of irregularity which they applied to skin lesions in order to differentiate benign melanocytic nevi from malignant melanomas. They worked with an extensively set of 40 lesion borders with extensive ground-truth. each of which was assessed by fourteen dermatologists on a four point scale. Figure 1 shows the skin lesion data as originally presented in [16], but reordered according to each lesion's mean ground-truth score. While Lee *et al.* demonstrated that

irregularity was a reasonable indicator of malignancy, examination of figure 1 also suggests that convexity is a strong factor.

Computing Lee *et al.*'s irregularity measure (OII) requires indentations and protrusions to be localised, which is a fairly involved process. It is a curvature scale-space filtering approach, and therefore smooths the boundary at multiple scales, identifying zeros and extrema of curvature at each scale (the latter an extension of the standard curvature scale-space). These points are tracked and connected over scale. Two separate collections of hierarchical data structures of indentation segments and protrusion segments are then generated in which the nesting of multiple fine scale structures within coarser scale segments is described. Since the smoothing process reduces the curvature values a threshold is required to identify and eliminate flat sections, which effectively provides the stopping condition for defining the roots of the segment trees. For each of the indentation/protrusion segments the area which is filled/removed by the smoothing process is determined. Either the maximum or the sum of these normalised areas is used as the irregularity measure.

In contrast, there are several standard convexity measures in the literature that are more straightforward, in particular two based on the convex hull of the boundary polygon P. Either the ratio of areas or perimeters can be used; we will denote the measures by $C_A = \mathbf{area}(P)/\,\mathbf{area}(\mathbf{CH}(P))$ and $C_L = \mathbf{perimeter}(\mathbf{CH}(P))/\,\mathbf{perimeter}(P)$, where $\mathbf{CH}(P)$ is the convex hull of P. Following on from this, we propose two new convexity measures in this paper: the first based on convexification: C_A^F, C_A^{FT} and the second on contained lines C_F. These measures are then evaluated as indicators of lesion malignancy alongside a large set of other shape measures from the literature as well on two other classification tasks involving mammographic masses and lung field boundaries.

2 Measuring Convexity by Convexification

In this section we describe a novel method for measuring convexity which has as its genesis a polygonal convexification process arising from a problem set by Paul Erdös [6]. Given a simple (non-intersecting) polygon, let its concavities ("pockets") be simultaneously reflected about their corresponding edges in the convex hull (their "lids") – this is the *flip* operation. Does repeating this process converge in a finite number of steps to a convex polygon? First, it was shown that to avoid self-intersection the pockets should only be flipped one at a time. Second, that only a finite number of flips are required for convexification, but that the number of flips required is not bounded by any function of n, the number of vertices. This led to a modification in which the pocket is flipped and also has the order of its vertices reversed (a *flipturn*) – whereas

flips preserve the *order* of edges around the polygon, flipturns preserve their *slopes*. In contrast to flips, Aichholzer *et al.* [1] show that any simple polygon can be convexified by at most $n^2 - 4n + 1$ flipturns. More historical details are given by Toussaint [29].

The basic steps of convexification are straightforward as illustrated in figure 2. The initial polygon is shown in figure 2a. The pocket is drawn in bold, and its lid as the dashed line. The results of applying a reflection of the pocket about the lid (i.e. a flip) is shown in figure 2b. When the order of the vertices is also reversed this is equivalent to rotating the pocket 180° about the midpoint of its lid, and produces a flipturn, see figure 2c.

It is possible for special situations to occur in which the lid is a proper subset of a convex hull edge, which extends beyond the lid, as illustrated in figure 3a (the complete edge of the convex hull is shown by bold dashes). The standard flipturn rotates the pocket 180° about the midpoint of the lid (figure 3c). Alternatively, the extended flipturn [1] treats the complete convex hull edge as an extended lid; rotation of the pocket 180° about the midpoint of this lead results in the polygon shown in figure 3d. In this paper we have used extended flipturns.

A simple implementation in which the convex hull is recomputed from scratch at each iteration would result in an algorithm that is linear per iteration. If an appropriate data structure for online updates is used each iteration be performed in $O(\log^4 n)$ amortised time [1]. However, if only the final convexified polygon using flipturns is required then this can be computed more efficiently. Flipturns do not change orientations or lengths of edges, so that the edges of the original polygon can be sorted by orientation in $O(n \log n)$ time and then reconnected to form the convexified polygon.

Whereas flips and flipturns have previously only been considered as an interesting computational geometry problem, in this paper we use the convexified polygon to measure convexity in the same way as the more traditional convex hull based method, namely the ratio of the areas of the original and convexified polygons (denoted C_A^F and C_A^{FT} when using flips or flipturns respectively). An alternative would be to use the number of flips or flipturns as an indication of convexity – however this would discard information relating to the size of the flips or flipturns. Since there are many possible different sequences of flips or flipturns that will convexify a polygon it is important to ensure that the result is stable. Aichholzer *et al.* [1] show that using flipturns all the sequences result in the same final polygon, but there is no such guarantee using flips. To ensure repeatability for similar shapes we standardise the order of flipping and flipturning. At each iteration the maximum deviation between each pocket and its lid is determined, and the pocket with the largest deviation is selected for flipping.

The convex hull based measurements are very asymmetric in that they are far less sensitive to intrusions than protrusions. This is demonstrated on a circle which has spikes added or subtracted from it; see figure 4. The solid and dotted lines in the graphs refer to the circles with protrusions and intrusions respectively. It can be seen that convexity based on the area of the convexified polygon behaves in a close to symmetric manner. The reason is that any intrusions are quickly converted into protrusions by the convexification process.

Another comparison between the measures is shown in figure 5. The rectangle in the left-hand column has the notch in different locations. This shift has no effect on the values returned by C_A and C_L, or by convexification using flipturning. However, when just flips are applied the different notch positions result in different convexified polygons. C_A^F is maximal when the notch is furthest from the centre of the rectangle. Since the maximum inscribed convex polygon in the latter rectangle is larger than the maximum inscribed convex polygons in the other notched rectangles then it could be argued that such a convexity measure is appropriate.

3 Measuring Convexity by Contained Lines

There are many definitions of (perfect) convexity [3] and many of these can be employed to generate measures of (approximate) convexity [18]. The one used here is based on the set of all straight line segments $\mathcal{L}$ formed from all pairs of points lying within a polygon P. Polygon P is considered to be convex if and only if all the lines in $\mathcal{L}$ are completely contained within P. Given the basic definition it is possible to adapt it in many ways to create more specific or general concepts of convexity. For instance, if the straight lines are *digital* (i.e. they are sampled on a grid) then digital convexity can be determined [12]. Another example would be to restrict the lines to lie in a single pre-specified orientation, the so called $\mathcal{O}$-convex set [7].

In this paper the requirement for $\mathcal{L}$ to be completely contained in P will be relaxed. Instead it will be sufficient for a sufficiently large fraction of each line in $\mathcal{L}$ to be contained. The motivation is to make the approach less sensitive to minor fluctuations of the boundary which are often not significant, being caused by noise, etc. Thus, for a desired fraction F, the new convexity measure C_F is simply given by the proportion of lines in $\mathcal{L}$ that have at least fraction F of their length inside P. Therefore, for an appropriate value of F, C_F will be insensitive to even deep intrusions into the region if they are sufficiently narrow.

In practice, this is computed by a Monte Carlo simulation in which a large

number of lines are sampled from $\mathcal{L}$. All our experiments use 10^5 lines as this was found to provide satisfactory convergence. The first step is to generate a random set of points in P. This is done by generating points inside the axis aligned, minimum area rectangle that bounds P, and then discarding points outside P. While the points could be generated by assigning their X and Y coordinates independently from a random number generator, it is known that this approach forms clusters in Euclidean space [27]. A more uniform sampling of the space is provided by quasi-random low-discrepancy sequences such as the Sobol sequence [20,28]. This technique is often used in Monte Carlo integration since avoiding wasteful samples due to clusters, and missing samples due to voids, improves the convergence of the Monte Carlo estimation [28].

The j'th dimension of the n'th term x_n in the Sobol sequence of points is computed by taking the bitwise XOR of a set of "direction numbers" V_i^j

$$x_n^j = a_1 V_1^j \oplus a_2 V_2^j \oplus \cdots \oplus a_w V_w^j$$

where the coefficients a_i are defined by $n = \sum_{i=0}^{w} a_i 2^i$, and the direction numbers are of length w bits. Full details of how the direction numbers are generated is beyond the scope of this paper. Briefly, they are chosen so that they satisfy a recurrence relation using the coefficients of a primitive (irreducible) polynomial (using modulo two integer arithmetic). In addition, to ensure low discrepancy, Sobol [28] defined two properties that the integers initialising the recurrence should satisfy. A different polynomial is used for each dimension.

The differences between random points and the low discrepancy sequence are demonstrated in figure 6. Points sampled inside P using the Sobol sequence are shown in figure 6a, and pairs are connected to form straight line segments (see figure 6b). It is evident that using successive points in the sequence results in a restricted set of generated line orientations. One solution is to add random jitter to the coordinate values as shown in figure 6c using uniformly distributed random offsets with maximum amplitude 10% of the bounding rectangle. This provides a better distribution of orientations, but a less uniform distribution over space. Alternatively, the required length of Sobol sequence can be generated first, and randomly reordered before points are paired to form line segments (see figure 6d). For comparison, randomly generating the X and Y coordinates produces a less uniform coverage of P (figures 6e&f). The computational complexity of generating and testing each line is $O(n)$ since both testing for the inclusion of a point inside a polygon and testing for the intersection of a line and polygon are linear.

The new measure is compared against C_A and C_L for a circle of radius R with a sinusoid of amplitude $f \times R$ superimposed (one example was shown in figure 6). The graph of convexity values in figure 7a shows that even requiring the complete line to lie inside the polygon (C_1) the measure is insensitive to small fluctuations of the circle. For low amplitude sinusoids the points inside

the sinusoid protrusions are visible to most of the circle, and so there are few intersections with the polygon, yielding high convexity values. This holds for values of $f < 0.1$, and then there is a drop-off in C_1 values, ultimately dropping below C_A and C_L which are seen to respond in a less linear manner (figure 7c). $C_{0.9}$ has the same behaviour as C_1, but drops-off more slowly. Testing the different sampling strategies in figure 7b indicates that they provide similar performance, at least for this test data.

4 Experiments

This section describes two experiments to evaluate the effectiveness of the proposed shape measures. For the classification of skin lesions they are compared against Lee *et al.*'s [16] results since their data was available, making direct comparison possible. Of course there has been much other work on automating the diagnosis of skin lesions. Much of this uses a large gamut of shape and colour/radiometric features, but the focus of the research is often on other aspects such as machine learning [4] or feature selection [9], although some papers discuss the importance of various aspects of shape, such as irregularity [2] and symmetry [26]. Although the new shape measures were developed with this application in mind they describe general aspects of shape, and their usefulness in other tasks was tested by applying them to another medical classification problem.

For comparison with the proposed convexity measures we have used the following shape measures from the literature, all of which are calculated from the boundary (ignoring density/intensity information): rotation, translation, and scale (RTS) geometric moment invariants, affine moment invariants [8], aspect ratio, compactness ($\mathbf{perimeter}^2(P)/\,\mathbf{area}(P)$), bilateral symmetry (the normalised area of overlap between the region and a rotated version of itself), circularity [10] and several other convexity measures (C_A, C_L, Žunić and Rosin's [31] measure C_Z, and Rosin and Mumford's [25] measures C_P, C_Q and C_S). In addition, various measures of ellipticity, rectangularity, and triangularity, described in [24] were applied. Only those giving the best overall performance over the experiments are reported here: ellipticity calculated using the least median squared error fitted ellipse, rectangularity calculated using the minimum bounding rectangle, and triangularity based on the first affine moment invariant.

The new measures were implemented as described above. In addition, in case the convexification method is sensitive to fine detail or quantisation effects it was run on both the raw boundary data as well as its polygonal approximation generated using Lowe's non-parametric algorithm [17]. Likewise, Žunić and Rosin's [31] convexity measure C_Z and Rosin and Mumford's [25] convexity

measures C_P, C_Q and C_S were applied after polygonal approximation using Ramer's [21] algorithm with a threshold of 2 pixels.

Table 1 shows the maximum and mean number of iterations for the convexification process to converge to a convex polygon. It can be seen that although there is no upper bound on the number of flips required, in practice several hundred are usually sufficient for all the data tested in this paper (102 boundaries).

4.1 Skin Lesions

This dataset consists of 40 melanomas were extracted from 512×486 RGB colour images selected from the image database collected from the Pigmented Lesion Clinics of the Division of Dermatology in Vancouver, Canada. The images were acquired from a hand-held camera placed directly on the patient's skin and lit by a ring of light [15]. Dark thick hairs were removed from the images [14]. The resulting images were then filtered to remove noise, and segmented to extract the lesion boundaries [15].

Following Lee *et. al* [16], the Spearman rank correlation values of the shape measures against the mean expert ground-truth score were computed, and were compared against Lee's OII which produced a value of 0.88. Several of the measures listed in table 2 provided an improvement over OII. The best performances were from Rosin and Mumford's [25] convex skull based convexity measure C_S (which uses the largest *inscribed* convex polygon rather than the *circumscribing* convex polygon – i.e. the convex hull) scoring 0.954, followed by their symmetric convexity measures C_P and C_Q.[1] Convexity measured by contained lines, specifically $C_{0.9}$ also performed well (scoring 0.933) which implies that it is advantageous to be insensitive to narrow intrusions in this instance. Surprisingly even the standard convexity measure C_A scored higher than OII. Although these convexity measures performed well, several others did not in this application. For instance, the convexification approach was found to be susceptible to fine details of the boundary, inflating the number of flips or flipturns when applied to the raw pixel data. Although polygonal approximation provided an improvement, the results for C_A^F and C_A^{FT} were still poor. Likewise, C_L and C_Z only scored 0.4. Although Schmid-Saugeon *et al.* [26] and others have emphasised the importance of symmetry in diagnosis of lesions, it only received a score of 0.880, equal to OII, but worse than many of those measures listed in table 2.

[1] Note that Rosin and Mumford [25] described the slighter better results of C_P and C_Q scoring 0.958 when the algorithm was applied directly to the complete boundary data set without polygonal simplification.

4.2 Mammographic Masses

As a second test, the shape measures are applied to a set of 54 masses from mammograms, combining images from the MIAS and Screen Test databases [22], see figure 8. Considerable analysis of this data has already been carried out by Rangayyan *et al.* [22], who considered a variety of new and old shape measures as well as several density based techniques (e.g. region based edge profile acutance). Since this paper is restricted to shape, the proposed techniques will only be compared against the former set of results. Rangayyan *et al.* assessed the measures by classifying them as circumscribed/spiculated, benign/malignant, and CB/CM/SB/SM, in two group and four group classification experiments. They used the BMDP "7M" discriminant analysis program to carry out classification, and published percentage total classification accuracy rates.

In our experiments classification was performed using a nearest neighbour classifier. Mahalanobis distances were used, and so when only a single shape property was used to build the classifier (as was the case of all our experiments except for the last) this reduces to Euclidean distances. In order to produce statistically correct and significant results the "632 bootstrap" [5,13] was employed to provide estimates of classification accuracy with minimal bias. This operates by first training and testing the classifier on all the data (the *design set*) of size N to produce an over-optimistic classification rate r^a. Next, B bootstrap samples of size N are drawn by sampling with replacement from the design set. The classifier is trained on each bootstrap and tested on the members of the design set *not* in the bootstrap. For the ith bootstrap sample let the number of members tested be m_i and the number correctly classified be n_i. Then the 632 bootstrap estimate is calculated as $0.368r^a + 0.632r^c$, where $r^c = \sum_{i=1}^{B} n_i / \sum_{i=1}^{B} m_i$. Furthermore, an estimation of the error rate of the estimate can be efficiently computed using the *jackknife-after-bootstrap* approach. For computational convenience we use instead an equivalent but less efficient two-deep nested approach for the jackknife in which multiple 632 bootstrap estimates are generated, and the standard deviation computed over this set.

In the three classification experiments our accuracy for compactness was similar to those obtained by Rangayyan *et al.* in the first instance, and substantially lower in the second two instances suggesting that the classification results obtained from the shape measures are not inflated in this paper relative to those of Rangayyan *et al.* [2]

[2] In a previous set of experiments we performed classification using Murthy *et al.*'s oblique decision tree [19] with leave-one-out cross validation. This provided better classification accuracies for compactness, that were identical to those obtained by Rangayyan *et al.* in the first two instances, and only slightly lower in the third

Rangayyan *et al.*'s [22] best shape measure results were achieved by: 1/ both compactness and a Fourier based shape factor, with a percentage total classification accuracy of 88.9% for circumscribed/spiculated discrimination, 2/ the Fourier based shape factor (75.9%) for benign/malignant discrimination, and 3/ both the Fourier based shape factor and compactness (64.8%) for the four-way discrimination. Accuracies of the new measures described in this paper are given in table 2, with $B = 100$ bootstrap samples, and 1000 iterations for the jackknife. The maximum standard deviation in any of the classification accuracies estimated by the jackknife over all the shape measures was 0.592, which demonstrates that a high confidence can be assigned to the correctness of the values in table 2. It can be seen that for two out of the three classification tasks some improvements have been made by the new measures compared to Rangayyan *et al.*'s results.

The convexification measures (C_A^F and C_A^{FT}) had performed poorly on the skin lesion data, but fared better on the mammographic data. In many instances, performing a polygonal approximation of the data boosted the results. Figure 9 demonstrates the convexified polygons for some of the masses from figure 8. Nevertheless, as table 2 shows, this measure is mostly outperformed by other shape measures.

The results for the convexity measure based on contained lines (C_F) are often good (i.e. better than those of Rangayyan *et al.*). There is not conclusive evidence as to which is the best method for generating line samples, although it is clear that the basic approach of connecting successive elements in the Sobol sequence is *not* a good approach.

Although the standard convexity measure C_A again performed reasonably it was still outperformed by some of the new measures. While the other standard convexity measure C_L had performed poorly on the skin lesion data, and only reasonably on the first two mammogram classification tasks it managed to outperform all the other measures on the four-way discrimination task.

Interestingly, given the effectiveness of our convexity measures, later work by Rangayyan *et al.* [23] concentrated on developing two new shape measures, for convexity and irregularity. The first is based on the concave and convex fractions of the boundary length. The second, so called spiculation index (SI), forms a polygonal approximation of the boundary, and modulates the length of each line segment by its narrowness (i.e. a function of the angle between two adjacent segments). On the same data (but with one mass excluded) they reported accuracies for the benign/malignant classification task of 74% for the fractional concavity or convexity, and 79% for SI, the latter outperforming our

(61.11% rather than 64.8%). However, it was found that the classification rate depended on the ordering of the data and included a probabilistic element to building the decision tree, causing high variance in different runs of the classifier.

methods. They did not report results for the other two classification tasks.

4.3 Chest X-rays

A final, small scale test is performed on eight pairs (left and right) of lung field boundaries taken from chest radiographs [30], see figure 10. Ground truth classification (normal or emphysematous) was provided by a radiologist. Tsai *et al.* [30] achieved perfect classification using the level set function as a shape descriptor, and using the EM algorithm to perform both classification as well as estimation of the most representative shape contours for each class.

Here, for simplicity, we have applied the shape measures to each contour separately. That is, left and right lung pairs are not processed as single units, as was done using the level set shape descriptor. Classification was performed in the same manner to the mammogram example in subsection 4.2. The maximum standard deviation in any of the classification accuracies estimated by the jackknife was 1.103, which again is small enough to indicate that the values in table 2 are meaningful despite the small sample size. None of the measures were able to produce 100% accurate classification, but nevertheless some of the results were reasonable. Among the convexity measures, the convexification method gave the best results, with C_A^F effectively misclassifying one lung pair out of the eight. Better accuracy was achieved by rectangularity, effectively misclassifying a single lung out of the sixteen.

5 Conclusions

The importance of convexity in medical diagnosis has been noted many times, which prompted the development of the new convexity measures described in this paper. While for clarity we have concentrated on applying them as individual shape measures, in practice their effectiveness would be enhanced by using combinations of shape, texture, and colour features. Different discrimination tasks are often sensitive to different features, and this was found to be true in this case. Examining the four sets of classification accuracies shows that there is no single shape measure that is uniformly the most effective – and this is not necessarily to be expected. However, it is clear that the highest ranking methods were generally convexity measures, which tended to outperform the standard feature descriptors (e.g. moment invariants, symmetry, compactness/form factor, etc.).

Nevertheless, as the paper shows, there are many ways to compute a single shape descriptor such as convexity [18], and these alternative measures have

their specific characteristics, and can be more effective in one instance rather than another. Unfortunately it is not possible to make a complete set of universal recommendations as to which of the methods and all their variations is most effective. However, based on the presented experiments the following conclusions can be drawn:

- Both of the two new convexity measures (convexification: C_A^F, C_A^{FT} and contained lines: C_F) were shown to be effective for some of the classification tasks. In particular, improvements were made over Lee *et al.*'s OII scale-space based irregularity measure [16] for the lesion data, and over Rangayyan et al.s results [22] for two out of the three mammographic classification tasks.
- For the convexity by contained lines measure it appears that $C_{0.9}$ is often superior to C_1. Despite testing several sampling methods it is not clear which, if any, is best. In particular, the simple generation of random points is not generally worse than the Sobol sequence based methods. This suggests that a future area for investigation would be to try alternative sampling strategies and to check on the convergence of the Monte Carlo simulation.
- The performance of the convexification method was inconsistent, and it is not possible to say from these experiments which of the options (flips versus flipturns; full pixel data versus polygonal simplification) are more effective.

Ideally the methods should be tested on more applications and on data sets with a much larger set of images. Unfortunately such data is generally difficult to acquire, especially with the desired ground truth necessary for evaluation. [3] Therefore, the approach taken in this paper has been to use the bootstrap estimator to maximise the reliability of the obtained results.

Acknowledgements

I would like to thank the following people who kindly provided the data used in this paper, and which was invaluable for evaluating the proposed measures: Tim K. Lee and Rangaraj M. Rangayyan.

References

[1] O. Aichholzer, C. Cortés, E.D. Demaine, V. Dujmović, J. Erickson, H. Meijer, M. Overmars, B. Palop, S. Ramaswami, and G.T. Toussaint. Flipturning

[3] For example, while the Digital Database for Screening Mammography (DDSM) [11] provides a large dataset, the boundaries of the abnormalities in the DDSM images are not publicly available.

polygons. *Discrete & Computational Geometry*, 28:231–253, 2002.

[2] E. Claridge, J.D. Morris Smith, and P.N. Hall. Evaluation of border irregularity in pigmented skin lesions against consensus of expert clinicians. In *Proc. Medical Image Understanding and Analysis*, pages 85–88, 1998.

[3] G. Cristescu and L. Lupsa. Classes of discrete convexity properties. *Discrete and Computational Geometry*, 31(3):461–490, 2004.

[4] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, and H. Billhardt. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J. Biomedical Informatics*, 34(1):28–36, 2001.

[5] B. Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *J. of American Statistical Association*, 78(382):316–331, 1983.

[6] P. Erdös. Problem number 3763. *Amer. Math. Monthly*, 42:627, 1935.

[7] E. Fink and D. Wood. Fundamentals of restricted-orientation convexity. *Informatics and Computer Science*, 92:175–196, 1996.

[8] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26:167–174, 1993.

[9] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, 20(3):233–239, March 2001.

[10] R.M. Haralick. A measure for circularity of digital figures. *IEEE Transactions on Systems, Man and Cybernetics*, 4:394–396, 1974.

[11] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer Jr. The digital database for screening mammography. In *Proc. 5th Int. Workshop on Digital Mammography*, pages 457–460, 2001.

[12] C.E. Kim and A. Rosenfeld. Digital straight lines and convexity of digital regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(2):149–153, 1982.

[13] W.J. Krzanowski. Data-based interval estimation of classification error rates. *J. Applied Statistics*, 28(5):585–595, 2001.

[14] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean. DullRazor: a software approach to hair removal from images. *Comput Biol Med.*, 27:533–543, 1997.

[15] T. Lee, V. Ng, D. McLean, A. Coldman, R. Gallagher, and J. Sale. A multi-stage segmentation method for images of skin lesions. In *Proc. IEEE Pacific Rim Conf. on Communications, Computers, and Signal Processing*, pages 602–605, 1995.

[16] T.K. Lee, D.McLean, and M.S. Atkins. Irregularity index: A new border irregularity measure for cutaneous lesions. *Medical Image Analysis*, 7(1):47–64, 2003.

[17] D.G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355–395, 1987.

[18] R.R. Martin and P.L. Rosin. Turning shape decision problems into measures. *Int. J. Shape Modelling*, 10(1):83–113, 2004.

[19] S.K. Murthy, S. Kasif, and S. Salzberg. System for induction of oblique decision trees. *J. of Artificial Intelligence Research*, 2:1–33, 1994.

[20] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vettering. *Numerical Recipes in C*. Cambridge University Press, 1990.

[21] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–256, 1972.

[22] R.M. Rangayyan, N.M. Elfaramawy, J.E.L. Desautels, and O.A. Alim. Measures of acutance and shape for classification of breast-tumors. *IEEE Transactions on Medical Imaging*, 16(6):799–810, 1997.

[23] R.M. Rangayyan, N.R. Mudigonda, and J.E.L. Desautels. Boundary modeling and shape analysis methods for classification of mammographic masses. *Medical and Biological Engineering and Computing*, 38(5):487–496, 2000.

[24] P.L. Rosin. Measuring shape: Ellipticity, rectangularity, and triangularity. *Machine Vision and Applications*, 14:172–184, 2003.

[25] P.L. Rosin and C.L. Mumford. A symmetric convexity measure. *Computer Vision and Image Understanding*, 103(2):101–111, 2006.

[26] P. Schmid-Saugeon, J. Guillod, and J.-P. Thiran. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*, 27(1):65–78, 2003.

[27] S.K. Sen, T. Samanta, and A. Reese. Quasi-versus pseudo-random generators: discrepancy, complexity and integration-error based comparison. *Int. J. Innovative Computing, Information and Control*, 2(3):349–4198, 2006.

[28] I.M. Sobol. *A Primer for the Monte Carlo Method*. CRC-Press, 1994.

[29] G. Toussaint. The Erdos-Nagy theorem and its ramifications. In *Proc. 11th Canadian Conference on Computational Geometry*, pages 9–12, 1999.

[30] A. Tsai, W.M. Wells, S.K. Warfield, and A.S. Willsky. An EM algorithm for shape classification based on level sets. *Medical Image Analysis*, 9:491502, 2005.

[31] J. Žunić and P.L. Rosin. A new convexity measurement for polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):923–934, 2004.

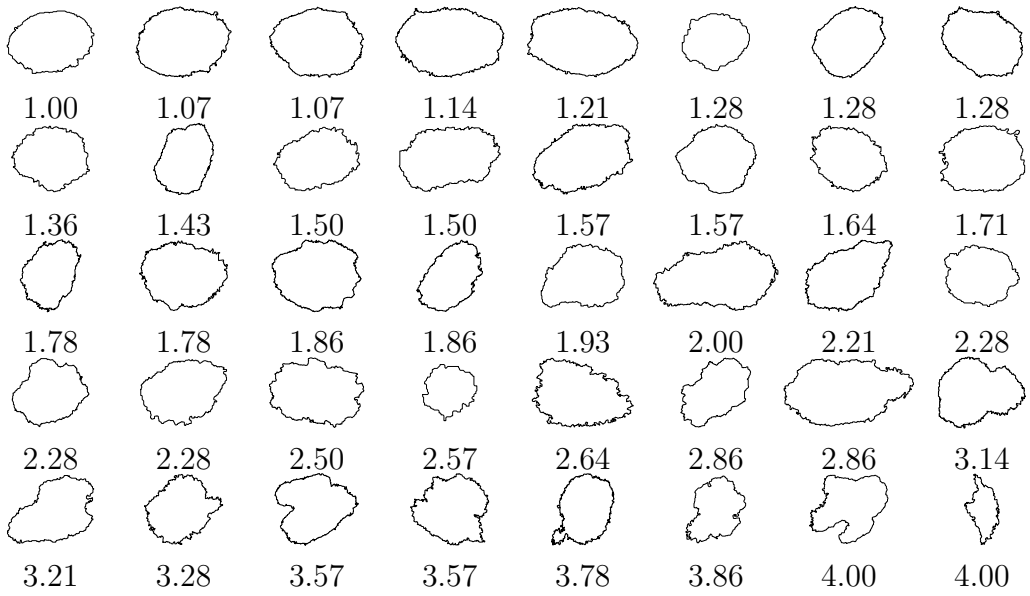

Fig. 1. The full set of 40 skin lesion outlines from Lee *et al.* [16] reordered according to the mean ground-truth score calculated from the individual scores provided by 14 dermatologists (drawn rescaled). A score of 1 corresponds to the healthiest lesion while 4 indicated the most severely malignant.

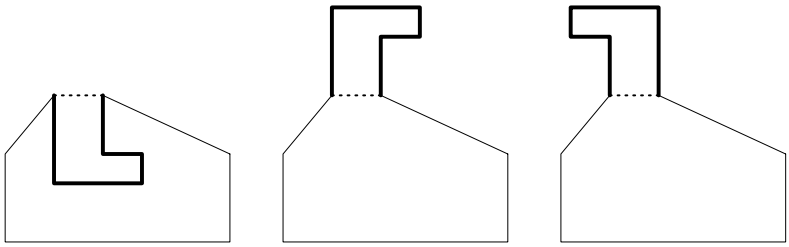

Fig. 2. One iteration of the convexification process: a) input polygon; b) after a flip; c) after a flipturn.

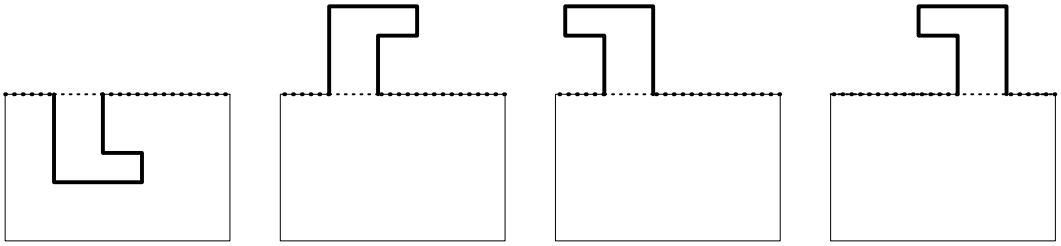

Fig. 3. One iteration of the convexification process: a) input polygon; b) after a flip; c) after a standard flipturn; d) after an extended flipturn.

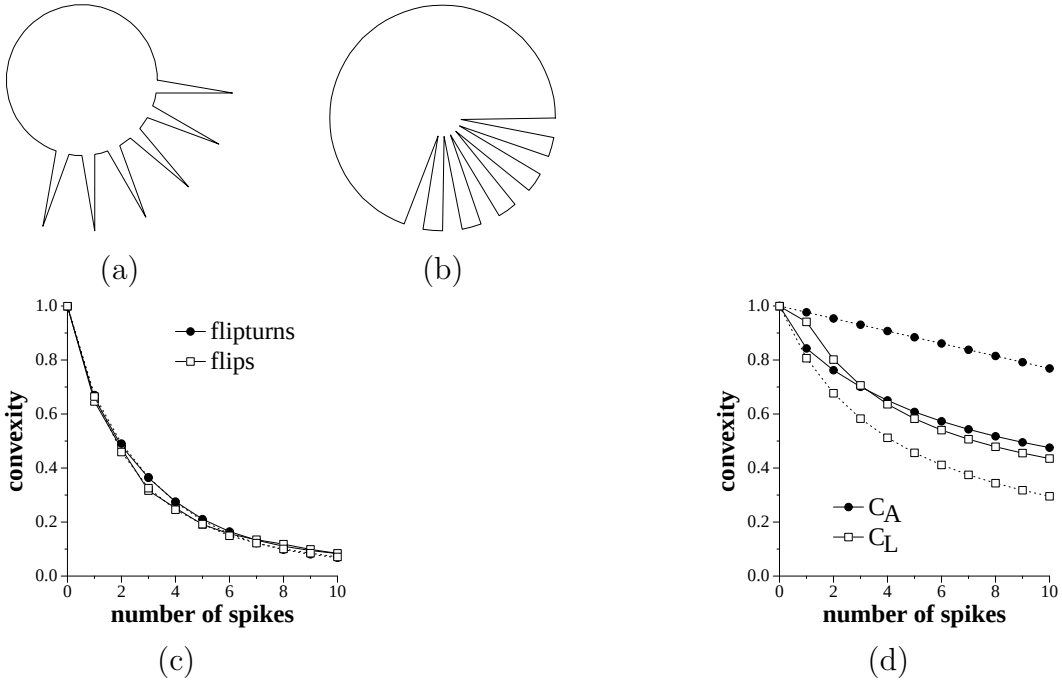

Fig. 4. Measured convexities of circles with increasing numbers of spikes inserted (solid lines in graphs) or removed (dotted lines). An example of such circles with six spikes is presented in (a) & (b). Results of using the new convexification approach are shown in (c), while the standard convex hull based measures are given in (d).

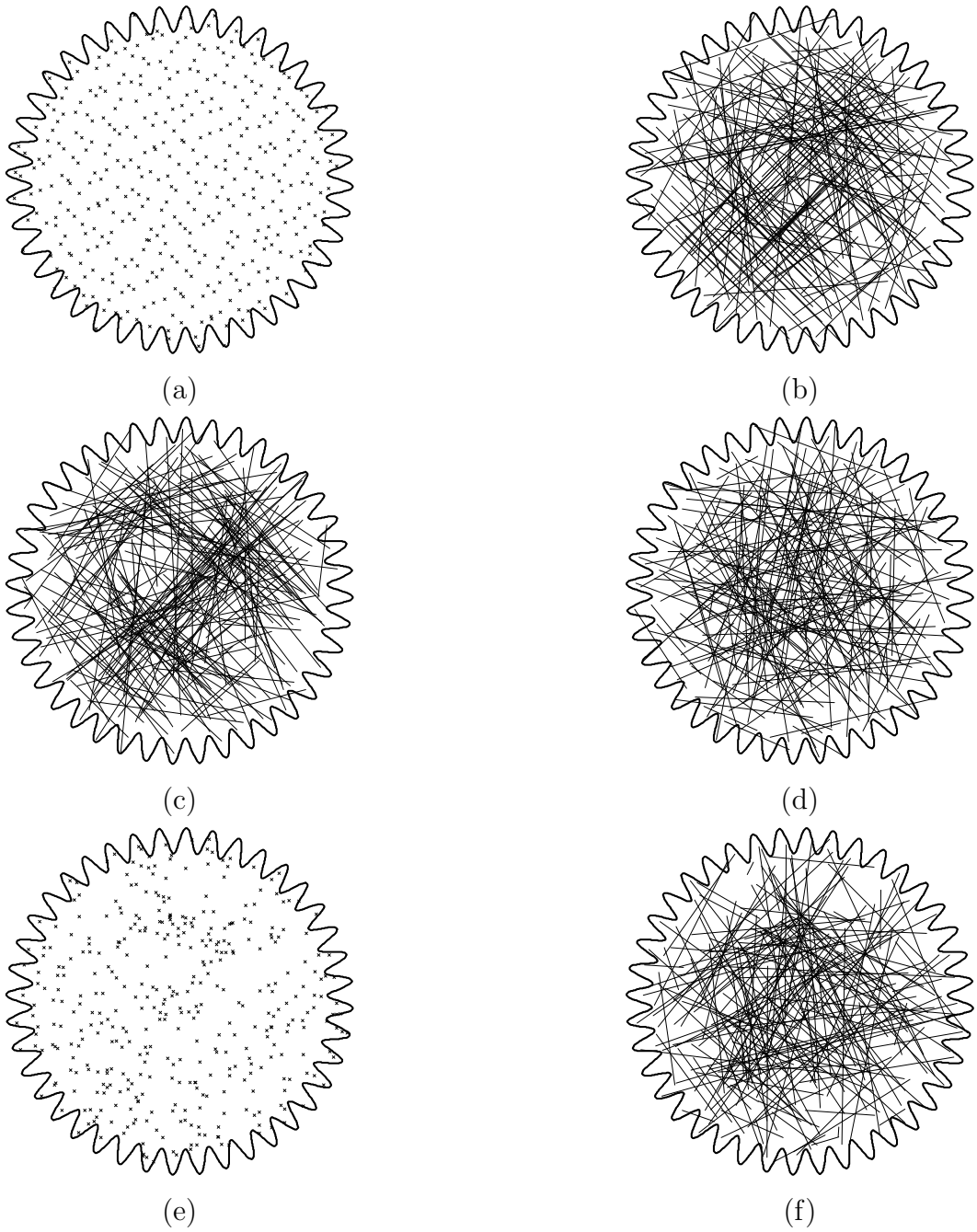

Fig. 6. Strategies for sampling: (a) 400 points generated using the Sobol sequence; (b) line segments formed from sequential pairs of points in (a); (c) line segments in (b) with added random jitter; (d) line segments formed from a randomised reordering of points in (a); (e) 400 points with coordinates obtained from a random number generator; (f) line segments formed from sequential pairs of points in (e).

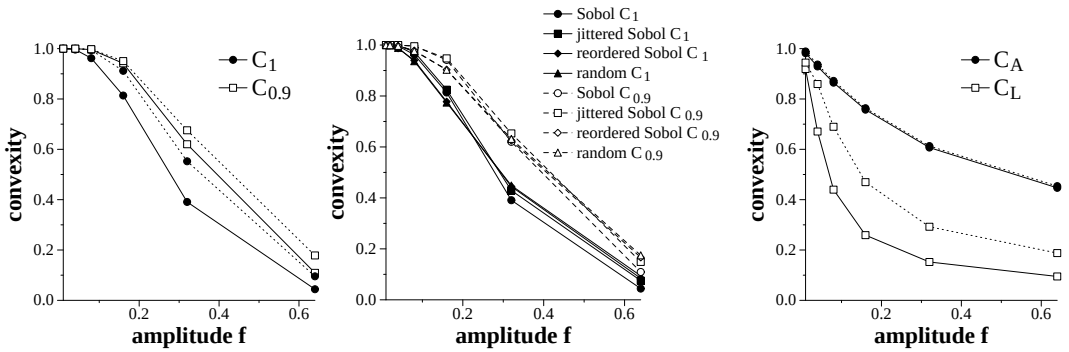

Fig. 7. Measured convexity values for a circle of radius R with a sinusoid of amplitude $f \times R$ superimposed. The solid line shows convexity values for polygons similar to the one in figure 6, the dotted line shows convexity values for sinusoids with half the frequency. In (a) are shown convexities computed using the basic Sobol sequence for C_1 and $C_{0.9}$. The effect of the different sampling strategies is demonstrated in (b). For comparison the standard convex hull based measures are given in (c).

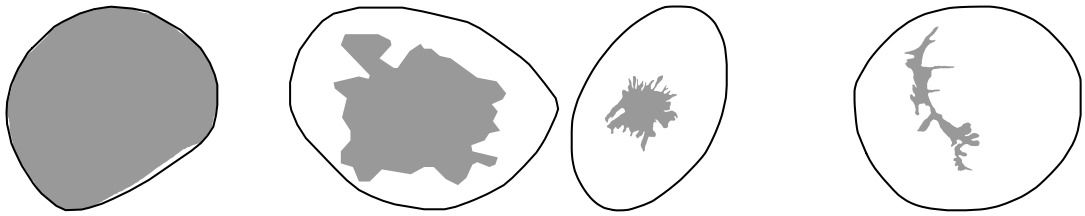

Fig. 9. Examples of mammographic masses after convexification. Increased undulation of the boundary results in a greater increase in convexified area relative to the original polygon.

method	skin lesions		mammographic masses		lung fields	
	max	mean	max	mean	max	mean
C^F all pixels	1404	674	1996	658	169	142
C^F polygonal approximation	462	191	644	125	9	5
C^{FT} all pixels	1983	500	1420	281	79	59
C^{FT} polygonal approximation	446	178	598	120	8	4

Table 1
The maximum and mean number of flips and flipturns required for convexification for the three data sets.

	method	skin	mammographic masses			lung
		lesions	circ/spic	mal/ben	4-way	fields
convexity	C_A^F all pixels	0.520	81.5	68.5	63.0	68.7
	C_A^F polygonal approximation	0.565	85.2	68.5	**64.8**	87.5
	C_A^{FT} all pixels	0.529	83.3	70.4	**64.8**	75.0
	C_A^{FT} polygonal approximation	0.565	**92.6**	53.7	50.0	75.0
	$C_{0.9}$ Sobol lines	**0.928**	83.3	48.1	42.6	50.0
	C_1 Sobol lines	0.786	81.5	57.4	48.2	43.7
	$C_{0.9}$ jittered Sobol lines	**0.906**	**88.9**	70.4	**66.7**	75.0
	C_1 jittered Sobol lines	0.778	87.0	64.8	59.3	50.0
	$C_{0.9}$ reordered Sobol lines	**0.923**	**90.7**	59.2	53.7	62.5
	C_1 reordered Sobol lines	0.829	87.1	51.9	46.3	43.7
	$C_{0.9}$ random	**0.933**	**94.4**	59.3	57.4	75.0
	C_1 random	0.863	87.0	51.9	50.0	25.0
	C_A	**0.888**	**90.8**	57.4	57.4	37.5
	C_L	0.453	85.2	74.1	**68.5**	25.0
	C_Z	0.462	79.6	70.4	**64.8**	50.0
	C_P	**0.939**	85.2	59.3	55.6	68.7
	C_Q	**0.936**	79.6	72.2	57.4	56.2
	C_S	**0.954**	83.3	55.6	46.3	68.7
miscellaneous	area	0.041	48.2	37.0	24.1	68.7
	compactness	0.532	87.0	59.3	57.4	68.7
	aspect ratio	0.076	61.1	57.4	20.4	0.0
	1st RTS moment invariant	0.309	68.5	68.5	51.8	56.2
	2nd RTS moment invariant	0.038	64.8	14.8	35.2	68.7
	1st affine moment invariant	**0.907**	87.0	64.8	55.6	25.0
	2nd affine moment invariant	0.435	51.8	50.0	25.9	62.5
	symmetry	**0.880**	79.6	53.7	44.4	25.0
	circularity	0.418	68.5	46.3	37.0	56.2
	ellipticity	**0.881**	77.8	74.1	59.3	56.2
	rectangularity	0.690	**90.8**	61.1	57.4	93.7
	triangularity	0.757	85.2	68.5	57.4	43.7

Table 2
Results of using both new and published shape measures on three medical image analysis tasks: 1/ the first column of figures gives absolute Spearman rank correlation scores for the skin lesion data against the mean expert ground-truth score, 2/ the next four columns give correct classification percentages, estimated using the bootstrap, obtained by the nearest neighbour classifier on i) three discrimination problems involving the mammographic data: circumscribed versus spiculated, benign versus malignant, and a four way discrimination: CB/CM/SB/SM, and ii) classification of lung boundaries as normal or diseased. In each column the top score is underlined, and scores better than or equal to Lee's OII or Rangayyan *et al.*'s [22] shape measures are highlighted in bold.

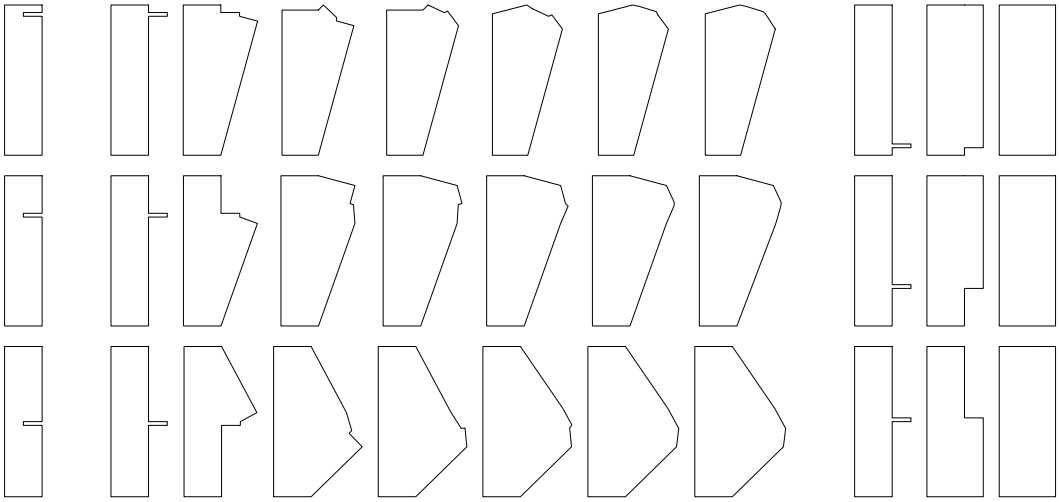

Fig. 5
Three versions of a rectangle with a notch are given in the left hand column. The middle and right sets of polygons show the convexification sequence using flips and flipturns respectively.

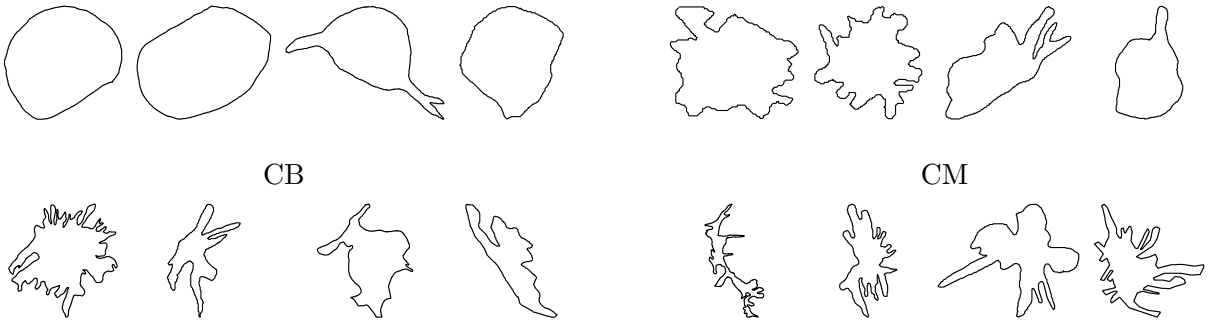

Fig. 8
Examples of the four classes of mammographic masses: circumscribed benign (CB), circumscribed malignant (CM), spiculated benign (SB), spiculated malignant (SM); they are drawn rescaled.

Fig. 10
Left and right lung field boundaries from chest radiographs classified as normal or with emphysema disease.