Pattern Recognition Letters 30 (2009) 570–578

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Classification of pathological shapes using convexity measures

Paul L. Rosin *

Cardiff School of Computer Science, Cardiff University, Queen's Buildings, Newport Road, P.O. Box 916, Cardiff CF24 3XF, UK

ARTICLE INFO

Article history:
Received 22 May 2008
Received in revised form 26 November 2008
Available online 13 December 2008

Communicated by G. Borgefors

Keywords:
Shape measure
Polygon
Convexity
Convex hull
Convexification
Medical classification

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.

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. (2003) 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 extensive set of 40 lesion borders with extensive ground-truth. each of which was assessed by fourteen dermatologists on a four point scale. Fig. 1 shows the skin lesion data as originally presented in (Lee et al., 2003), 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 Fig. 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 $C_A = \boldsymbol{area}(P)/\boldsymbol{area}(\boldsymbol{CH}(P))$ and $C_L = \boldsymbol{perimeter}(\boldsymbol{CH}(P))/\boldsymbol{perimeter}(P)$, where $\boldsymbol{CH}(P)$ 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 Erdös (1935). Given a simple

* Tel.: +44 029 2087 5585; fax: +44 029 2087 4598.
E-mail address: Paul.Rosin@cs.cf.ac.uk.

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.12.001

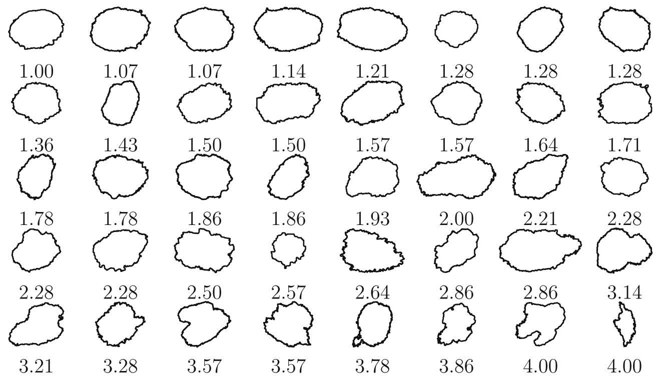

Fig. 1. The full set of 40 skin lesion outlines from Lee et al. (2003) reordered according to the mean ground-truth score calculated from the individual scores provided by 14 dermatologists (drawn rescaled). A score of 1 corresponds to the healthiest lesion while 4 indicated the most severely malignant.

(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. (2002) show that any simple polygon can be convexified by at most $n^2 - 4n + 1$ flipturns. More historical details are given by Toussaint (1999).

The basic steps of convexification are straightforward as illustrated in Fig. 2. The initial polygon is shown in Fig. 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 Fig. 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 Fig. 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 Fig. 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 (Fig. 3c). Alternatively, the extended flipturn (Aichholzer et al., 2002) 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 Fig. 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

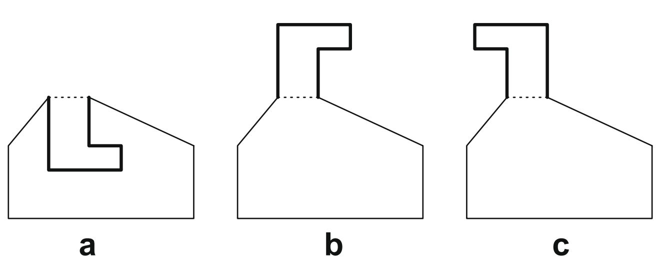

Fig. 2. One iteration of the convexification process: (a) input polygon; (b) after a flip; (c) after a flipturn.

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 (Aichholzer et al., 2002). 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. (2002) 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 Fig. 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 Fig. 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 (Cristescu and Lupsa, 2000) and many of these can be employed to generate measures of (approximate) convexity (Martin and Rosin, 2004). The one used here is based on the set of all straight line segments $\mathscr{L}$ 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 $\mathscr{L}$ 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 (Kim and Rosenfeld, 1982). Another example would be to restrict the lines to lie in a single pre-specified orientation, the so called $\mathscr{O}$-convex set (Fink and Wood, 1996).

In this paper the requirement for $\mathscr{L}$ to be completely contained in P will be relaxed. Instead it will be sufficient for a sufficiently large fraction of each line in $\mathscr{L}$ to be contained. The motivation is to make the approach less sensitive to minor fluctuations of