Statistics and Computing (2000) **10,** 105–120

Mathematical morphology: A useful set of tools for image analysis

EDMOND J. BREEN, RONALD JONES and HUGUES TALBOT

CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde NSW 2113
ronald.jones@cmis.csiro.au, hugues.talbot@cmis.csiro.au

Received April 1998 and accepted February 1999

In this paper we give an overview of both classical and more modern morphological techniques. We will demonstrate their utility through a range of practical examples. After discussing the fundamental morphological ideas, we show how the classic morphological opening and closing filters lead to measures of size via granulometries, and we will discuss briefly their implementation. We also present an overview of morphological segmentation techniques, and the use of connected openings and thinnings will be demonstrated. This then leads us into the more recent set-theoretic notions of graph based approaches to image analysis.

Keywords: filtering, graphs, region growing, segmentation

1. Introduction

Mathematical morphology is a branch of image analysis based on algebraic, set-theoretic and geometric principles. Initially developed by Matheron (1967, 1975) and Serra (1982) for performing sizing transforms on random sets, it has since been extended to functions (Serra 1982, Sternberg 1986) and to the more general framework of complete lattices (Serra 1988, Ronse and Heijmans 1991). More recently, attention has focussed on graph-based morphology (Heijmans *et al.* 1992, Vincent 1989) and connected component transforms (Serra and Salembier 1993, Vincent 1993a, Salembier and Serra 1995, Breen and Jones 1996a, b, Salembier and Oliveras 1996), which are closely related to the concept of morphological reconstruction (Klein 1976).

Mathematical morphology is particularly useful in providing basic building blocks to more sophisticated imaging applications. Using mathematical morphology, image data can be filtered to either preserve or remove features of interest, sizing transformations can be constructed, and information relating to shape, form and size can be easily applied. In this paper we present the fundamentals of image processing based on mathematical morphology. From these basics, we develop transforms for performing practical image analysis. We demonstrate the utility of these concepts for sizing image objects, for image segmentation of automobile registration plates, and for connected component filtering of aerial photographs for overlaid fiducial marks in aerial photographs. In order to segment more complex images, we develop a higher level image representation, which allows us to assign signatures to image regions and to apply set-theoretic graph notions to image analysis and filtering.

2. Basic notions

2.1. *Definition of an image*

An image is defined as a collection of pixels, where each pixel maps a coordinate $\vec{z}$ in the support of the image f to a single value or vector of values $f(\vec{z})$. Formally, we define an image f by $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, where $\mathbb{R}^n$ is the n-dimensional support of f and $\mathbb{R}^m$ is the m-dimensional range of values. In practice, the support is typically restricted to a discrete subset of the rectangular grid $\mathbb{Z}^2$ and the range of values to a subset of the set of integers $\mathbb{Z}$ called greyscales. Binary images are images that have only two pixel values (typically 0 and 1). Colour images are defined by using $\mathbb{Z}^3$ as the range of pixel values, representing the 3 components (red, green and blue) of a colour image.

An example is shown in Fig. 1, where the greyscale image in Fig. 1(a) is represented by the graph in Fig. 1(b). The support $\mathbb{Z}^2$ is the two dimensional rectangular grid shown shaded in the figure and the range $\mathbb{Z}$ is a one dimensional set of integers indicated by the vertical axis.

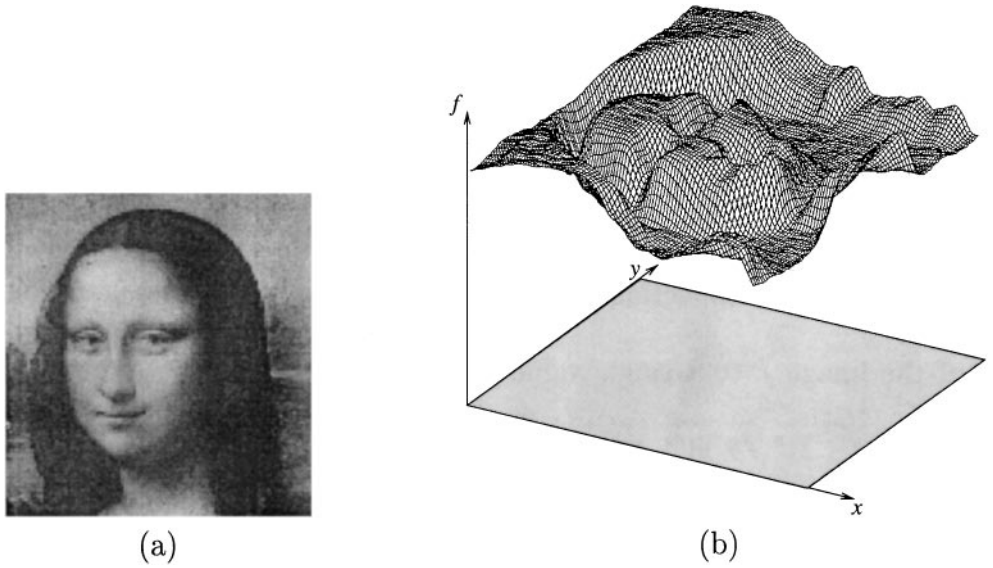

Fig. 1. *Definition of an image. (a) Example of a greyscale image. (b) Mathematical definition of the image as a collection of pixels, where each pixel maps a coordinate $\vec{z} = (x, y)$ in the support of the image f to a single value or vector of values $f(\vec{z})$*

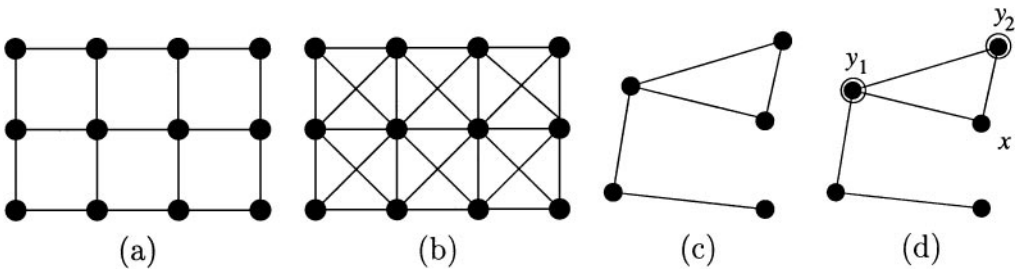

Fig. 2. *Graph representation of an image support. Vertices are considered neighbours if there is an edge linking them. (a) A 4 connected graph. (b) An 8 connected graph. (c) General graph. (d) Neighbourhood of a graph vertex, labelled x*

2.2. *Neighbourhoods in the image support*

A general framework for describing the neighbours of an image pixel is to consider the support of the image as an undirected graph $G = (V, E)$, which is composed of a set of vertices V connected by a set of edges E. For example, a 2D rectangular image support can be considered as either a 4 or 8 connected graph, as illustrated in Fig. 2(a) and (b) respectively. Figure 2(c) shows the graph representation of an image support of arbitrary structure.

The graph representation allows a generic definition of the neighbourhood of any image pixel coordinate x:

$$\mathcal{N}_G(x) = \{\, y \mid (x, y) \in E \}. \tag{1}$$

For example, Fig. 2(d) shows the neighbourhood $\{y_1, y_2\}$ for the vertex labelled x.

2.3. *Complete lattices, ordering and inclusion*

One of the fundamental ideas in mathematical morphology is that the set of images of a given type forms a complete lattice (denoted by '$\mathcal{L}$'), where every pair of lattice elements has a greatest lower bound and a least upper bound, themselves elements of the lattice (Kaufmann 1972, Serra 1998). The notion of a complete lattice can be demonstrated using the simple example of a small 3 by 1 binary image with pixel values **0** and **1**. In all, there exist a total of eight different 3 by 1 binary images, as shown in brackets in Fig. 3. These images can be ordered using the partial ordering relation of inclusion '$\leq$', as indicated by the arrows in this figure. For example, **1 0 0** $\leq$ **1 1 0**, or the image **1 0 0** is included within the image **1 1 0**. We can say this because each

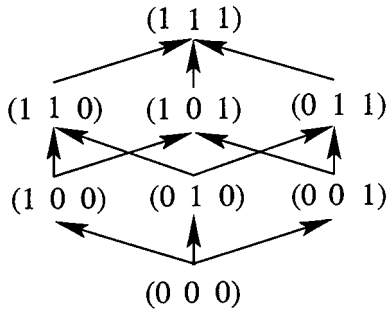

Fig. 3. *Example of a complete lattice: the set of all 3 by 1 binary images*

pixel in the image **1 0 0** is less than or equal to its corresponding pixel in the image **1 1 0**. The ordering '$\leq$' is only partial because some pairs of images cannot be ordered, for example **0 1 0** and **1 0 1**. Here we have neither **0 1 0** $\leq$ **1 0 1** nor **1 0 1** $\leq$ **0 1 0** and so no arrow has been drawn between these two images.

2.4. *The infimum and supremum operators*

The *infimum* (denoted by '$\wedge$') and the *supremum* (denoted by '$\vee$') are the fundamental operators in mathematical morphology and stem directly from the partial ordering relation that exists between images. The infimum of a set of images is defined as the greatest lower bound of the images. Because we deal with discrete single-variate images with finite support, this infimum is in fact a pixel-wise minimum. For example, the infimum of **1 0 0** and **1 1 0** is **1 0 0** and the infimum of **1 1 0** and **0 1 1** is **0 1 0**. In a similar way, we define the supremum of a set of images as the least upper bound, or the pixel-wise maximum of the images. If the supremum and infimum exist for any collection of images taken from the lattice of all images, then that lattice is called a complete lattice. The set of 3 by 1 binary images is such a lattice. More generally, the set of all binary images is a complete lattice, as is the set of all greyscale images and the set of all colour images (in this case, the supremum is obtained by taking the supremum in each channel).

3. Filtering and segmentation

From this somewhat abstract beginning one can define a range of image transforms that have proved to be enormously useful for practical applications in image analysis. In mathematical morphology, we consider an image transform as a mapping performed within a complete lattice $\mathcal{L}$; the transform maps one lattice element (the input image) to another lattice element (the output image). The measure of how well a transform performs can in part be formalised by describing what properties it has. For example, the property of *idempotence* is generally regarded as a useful filtering property because it guarantees that iterating the transform has no further effect than the first application of the transform on the image. If a transform is successful at removing all the unwanted noise in the image, then we can expect the transform to be idempotent. The property of idempotence together with some other basic properties are defined in Table 1.

Table 1. *Basic properties used to describe morphological image transforms. Here f and g are images, ψ is an image transform and t is a translation vector*

Property	Definition
Idempotence	$\psi[\psi(f)] = \psi(f)$
Increasingness	$f \leq g \Rightarrow \psi(f) \leq \psi(g)$
Translation-invariance	$\psi(f_t) = [\psi(f)]_t$
Extensivity	$\psi(f) \geq f$
Anti-extensivity	$\psi(f) \leq f$

3.1. *Morphological image transforms*

The key insight provided by mathematical morphology is to construct image transforms from the infimum and supremum operators. The simplest examples of morphological transforms are called *erosions* and *dilations*, both of which have the property of increasingness. There exist many definitions of erosion and dilation, of varying generality, but the classical definitions used most often in the literature are given respectively by:

$$\epsilon_B(f) = \bigwedge_{b \in B} f_{-b} \quad \text{and} \quad \delta_B(f) = \bigvee_{b \in B} f_b. \tag{2}$$

Here 'ϵ' denotes erosion, 'δ' denotes its dual dilation, f is the input image and B is a *flat structuring element* – typically a small convex set such as a square or a disc. It is also possible to define erosions and dilations with non-flat structuring elements (Serra 1982, 1988, Soille, Serra and Rivest 1992).

The process of erosion as defined in equation (2) is illustrated in Fig. 4, where the input image on the left is eroded by a circular structuring element B. An erosion translates the image f around through vectors $b \in B$ and takes the point-wise infimum of all the translated images. The result at any given point x in the output image is therefore given by an infimum of the set of input pixel values $f(x - b)$, taken over all $b \in B$. The circular region shows a close-up of the image values found around a given location in the input image. The result in the output image at this location

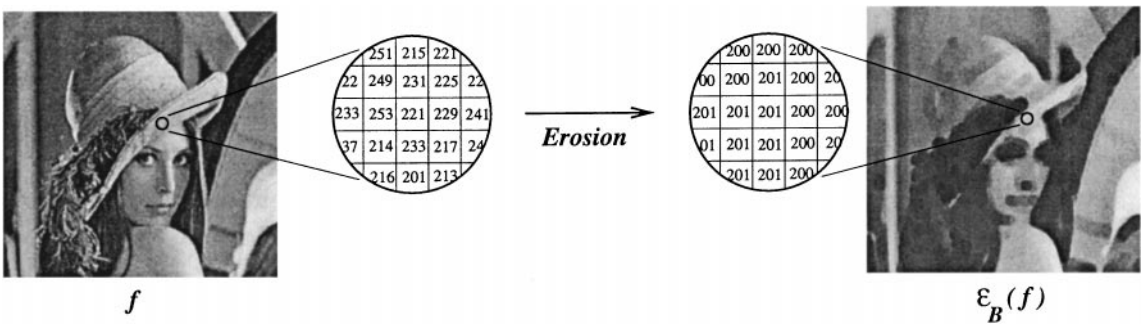

Fig. 4. *Erosion of a greyscale image using a circular structuring element. The result at any given point in the output image is given by an infimum of the input pixel values within the circular structuring element B translated at that position*

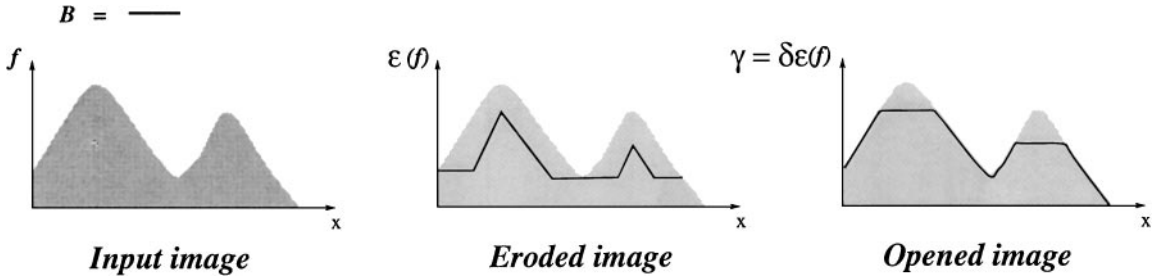

Fig. 5. *Example of a morphological opening of a 1D greyscale image using a line structuring element*

is the infimum of the values found in the circular region (or a digital approximation to it).

By replacing the set B with a function, volumetric structuring (or 3D) elements can be produced. However, we generally use a set because there exist efficient algorithms for the implementation of morphological transforms with such structuring elements (Borgefors 1986, Vincent 1990, 1992, Soille, Breen and Jones 1996, Vandroogenboeck and Talbot 1996) and they commute with any increasing image contrast changes (whereas volumetric structuring elements do not (Soille, Serra and Rivest 1992)).

Besides having the property of increasingness, the classical definitions of erosions and dilations in equation (2) also have the property of translation invariance. More generally however, an erosion is any increasing lattice mapping ϵ that obeys the relation $\epsilon(\wedge_i f_i) = \wedge_i \epsilon(f_i)$, where f_i are elements of the complete lattice of images $\mathcal{L}$. Similarly, a dilation is any increasing lattice mapping δ that satisfies: Justified $\delta(\vee_i f_i) = \vee_i \delta(f_i)$. The importance of such notions is discussed in detail in many theoretical treatments of mathematical morphology; see for instance Serra (1988).

3.2. *Openings and closings*

Erosions and dilations have proved to be very powerful as building blocks for effective image transforms with desirable filtering properties. For example, an idempotent image transform known as an *opening* can be constructed by cascading an erosion with its dual dilation, $\gamma = \delta\epsilon$. Idempotent and increasing image transforms are generally known in the field of mathematical morphology as *filters*. The dual *closing* filter is constructed by reversing the order of the erosion and dilation, $\phi = \epsilon\delta$. Openings and closings are perhaps the most widely-used examples of morphological image transforms. They have the properties of increasingness and idempotence, as defined in Table 1. In addition, openings are anti-extensive filters and closings are extensive filters.

Figure 5 shows an example of an opening constructed from the translation-invariant erosions and dilations defined in equation (2). Generally structuring elements are chosen judiciously to filter out or preserve particular image features. One may even use a set of structuring elements (Matheron 1988). Here, the 1D greyscale image on the left of Fig. 5 is opened using the short linear structuring element B shown at the top of the image. The image is first eroded by the line structuring element; the result is shown in the middle of the figure, where the eroded image $\epsilon(f)$ is indicated by the dark line. The subsequent dilation produces the opening on the right hand side of Fig. 5. As dictated by the idempotent property of the opening filter, if the opening were to be repeated on the image on the right hand side, the result would remain unchanged.

Figure 6(a) shows the opening of a binary image of Australia using a circular structuring element, where the boundary of the opened image is shown by the dark line. The notion of an

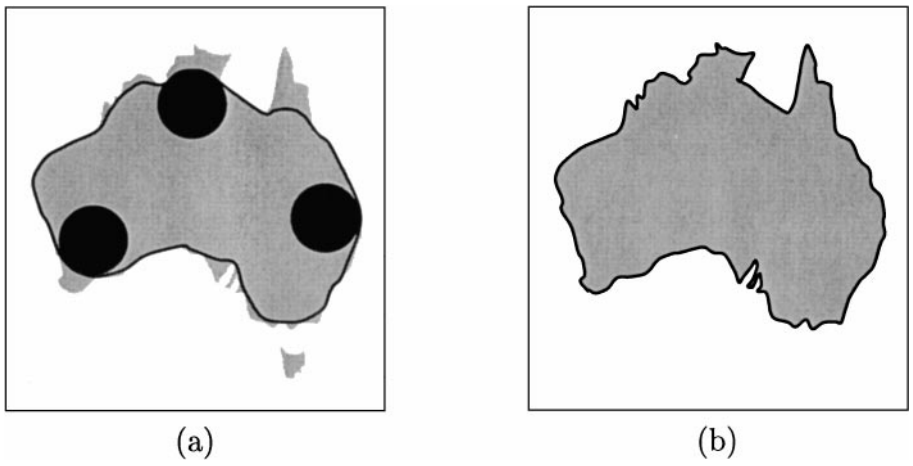

Fig. 6. *Examples of a morphological opening of a 2D binary image. (a) Using a disk structuring element. (b) An opening by reconstruction*

opening accommodates a somewhat more intuitive description than the formal definition based on equation (2). We can consider an opening as a filter that fits the given structuring element at every position in the image. Pixels are preserved wherever the structuring element fits, but are removed where it does not. The dark line in Fig. 6(a) shows the border of the region where the circular structuring element can fit into the map of Australia; the pixels outside this region are removed by the opening. In this way, the opening filter can be used to filter noise in many image analysis applications by choosing a suitable structuring element or set of structuring elements.

As is apparent from these examples, morphological filters such as openings are related to the notion of inclusion and size. Because openings are anti-extensive, an opened image is always included in the image itself, that is $\gamma(f) \leq f$. Also, only that part of the image that is of greater size than the given structuring element is preserved by the filter. We will be exploring the important notion of size in more detail in Section 3.3.

Although openings can be constructed as an erosion followed by a dilation, in general any image filter that satisfies the three properties of increasingness, idempotence and anti-extensivity is called an opening. A particularly useful example of an opening that cannot be constructed in the simple form $\gamma = \delta\epsilon$ is the *opening by reconstruction* (Klein 1976). In this case, an initial transform ψ is used to form a *marker* image $\psi(f)$ by removing unwanted features in the input image f. Only those features in f that are connected to the objects in the marker image are reconstructed. For example, Fig. 6(b) shows the reconstruction of the image in Fig. 6(a), using the opening by a circular structuring element as a marker image. Note that the only difference from the original map is that the island of Tasmania (off the south-east coast of the mainland) has been removed because it is too small to contain the circular structuring element and so cannot be reconstructed. The opening by reconstruction has the useful property of removing noise while preserving contour information in the image. We will return to this theme in the more general framework of *component transforms* in Section 3.6.

The closing filter is the dual of the opening filter and has the properties of increasingness, idempotence and extensivity. Whereas the anti-extensive opening removes light noise in the image, the extensive closing removes dark noise in the image.

3.3. *Granulometries*

Mathematical morphology provides several ways of obtaining size information. One important technique is the *granulometry* (Serra 1982) and is derived directly from the notion of morphological filters.

The *granulometric curve* of an image is a representation of the size distributions in the image. The idea is analogous to the physical process of sieving rocks through screens of increasing size: an image can be opened with structuring elements of increasing size. The *residue*, or difference between the original image and its opened version, is characteristic of the content of the image corresponding to the size of the opening.

3.3.1. Theory

It is possible to construct a granulometric curve by repeatedly opening an image with structuring elements of increasing size and by plotting the residues from one step to the next, via the formula:

$$G_f(r) = \sum f - \sum \gamma_r(f) \quad r \in [0, 1, 2, \ldots, R] \tag{3}$$

where f is the image, $\sum f$ is the sum of all the pixels of the image (i.e. the integral or volume of the image), r is a number characteristic of the dimension of the structuring element (for example the length of side of the structuring element if a square is used, or the radius if a disk is used), and R is the idempotent point, that is the structuring element dimension for which no further change occurs.

In order to guarantee a monotonic granulometry curve, one must use a family of convex structuring elements that obeys the *absorption law* (Matheron 1975), i.e. $\gamma_{r+t} \leq \gamma_r, \forall r, t \geq 0$. Such a family is called a *granulometry*.

The granulometry curve by openings can be supplemented by carrying out the same procedure using closings instead. Granulometries by openings size white structures (or domes) in images, while granulometries by closings size dark structures (or troughs). It is possible, and sometimes desirable, to substitute openings or closings by reconstruction instead of the traditional openings or closings with structuring elements.

3.3.2. A simple example

Figure 7 illustrates the concept of granulometries based on both openings and closings by reconstruction. The right-hand side of the curve is obtained by performing openings of increasing size on the image, while the left-hand side of the curve is obtained by performing closings of increasing size. This technique allows sizing of both the dark and the light structures in the image. Depending on the problem at hand, using a granulometry by either opening or closing, with or without reconstruction, might be sufficient. In this simple figure, when the size of the closing reaches the radius of the dark disk, a jump on the granulometry curve is seen. The second jump is caused by the two smaller white disks. The last jump is caused by the larger white disk. The amplitude of the jump is proportional to the volume of the image that is filtered out by the corresponding opening or closing. The general shape of the granulometry curve is representative of the size distributions within the image.

It should be noted that the jumps in the granulometry curve are clearly defined in this example because we have used openings and closings by reconstruction, which ensure that entire objects are removed at once. This effect may or may not be desirable depending on the problem at hand. For example, if two objects of different sizes are connected by even a thin line, neither of the objects will be removed by the opening until the bigger object is completely eroded. This does not occur in the present example because the objects in the image are clearly separated.

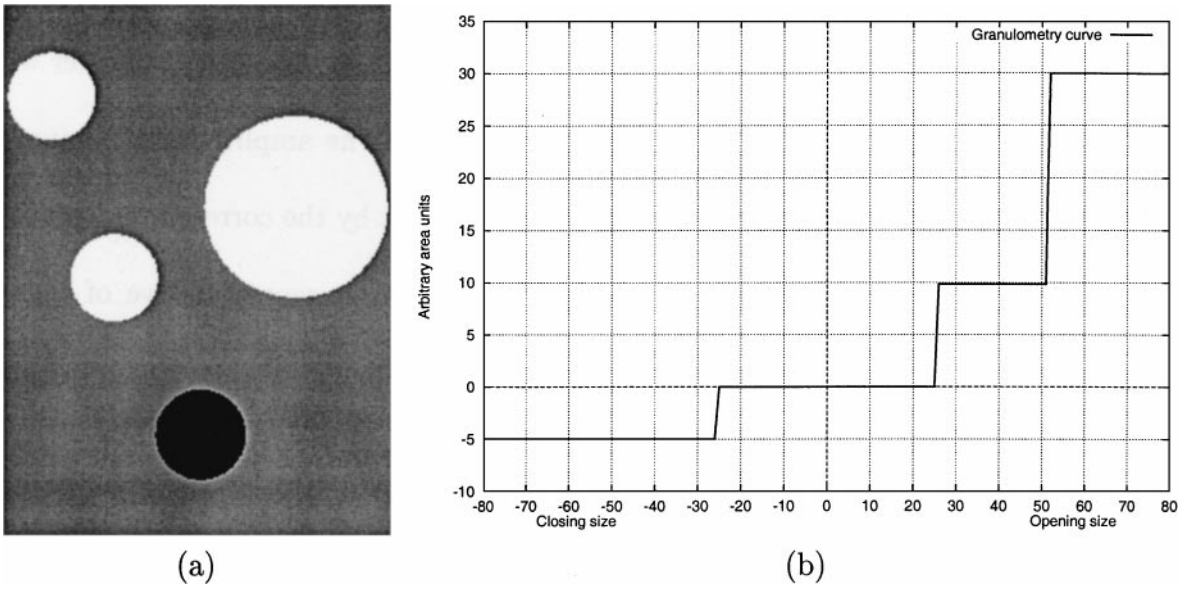

Fig. 7. *Granulometry curve of a simple grey-level image (a). When the size of the closing reaches the radius of the dark disk, a jump on the granulometry curve (b) is seen. The second jump is caused by the two smaller white disks. The last jump is caused by the larger white disk. The amplitude of the jump is proportional to the volume of the image that is filtered out by the corresponding opening or closing*

An application. In order to illustrate the usefulness of the granulometries, we present the following image analysis application: the Australian Electricity Industry has a need to have reliable estimates of likely safe remaining life of plant equipment, such as boiler tubing and headers, and steam piping. A commonly used measure of the remaining life, based on the temperature history of the particular plant item, is the Sherby-Dorn (SD) parameter (Coade 1993). Expert metallurgists have developed techniques for estimating the SD parameter based on visual interpretation of the microstructure of samples. Several years ago, CSIRO investigated the use of image analysis techniques to automate this process (Jones *et al.* 1995). An alternative approach is to use a granulometry. Without going into too much detail (see Breen, Jones and Talbot (To appear) for a complete description), samples were prepared following standard metallographic methods and observed in an optical microscope at 500× magnification. Figure 8 shows two such images with their associated granulometry curve by openings. Figure 8(a) is a sample from boiler tubing which has had very little use, while Fig. 8(b) is a sample from tubing nearing the end of its useful life.

From Fig. 8(c) and (d) we can see that as the boiler degrades, the associated granulometric curve increases to the idempotent point sooner. The radius at which 50% of the idempotent point was reached can be used as a measure of the steepness of each curve. This measure correlates quite well with the SD parameter, as illustrated in Fig. 9.

3.4. *Morphological segmentation*

Granulometries provide sizing information from whole images, but this is not always sufficient. Sometimes information must be obtained from a specific part of the image. To solve these problems, *image segmentation* must be used; see for example Gonzalez and Wintz (1987).

Given the scope of the topic it is not possible to give a complete overview of image segmentation techniques, even restricted to those using mathematical morphology (Meyer and Beucher 1990). In this section, we will only present the most generic methods, two examples and a number of references. The rest of the paper also presents some morphological techniques which can be used in a generic way.

3.4.1. Region-based segmentation

There are two main approaches to segmentation in image analysis. One can first segment an image in regions that are similar according to some measure. This segmentation will not be immediately useful: one must then use this partition and high-level knowledge to refine the segmentation, understand and analyse the content of the image (the bottom-up approach). Conversely, one can first use a-priori and high-level knowledge to devise a segmentation method that will partition the image in an immediately relevant way (the top-down approach). Both approaches are valid and in practice complementary.

Traditional approaches to segmentation are mostly bottom up. Most, but not all, morphological segmentation approaches are top-down. Of those, the region-based methods are the most generic, and we will present the basic ideas behind them now.

The watershed method. A 2-D grey-level image, or function, can be viewed as a 3-D terrain, with the grey-level as the z dimension (see Fig. 1(b)). Building the *watershed line* under such a representation involves "flooding" the image from the regional minima, and building "dams" where "catchment basins"

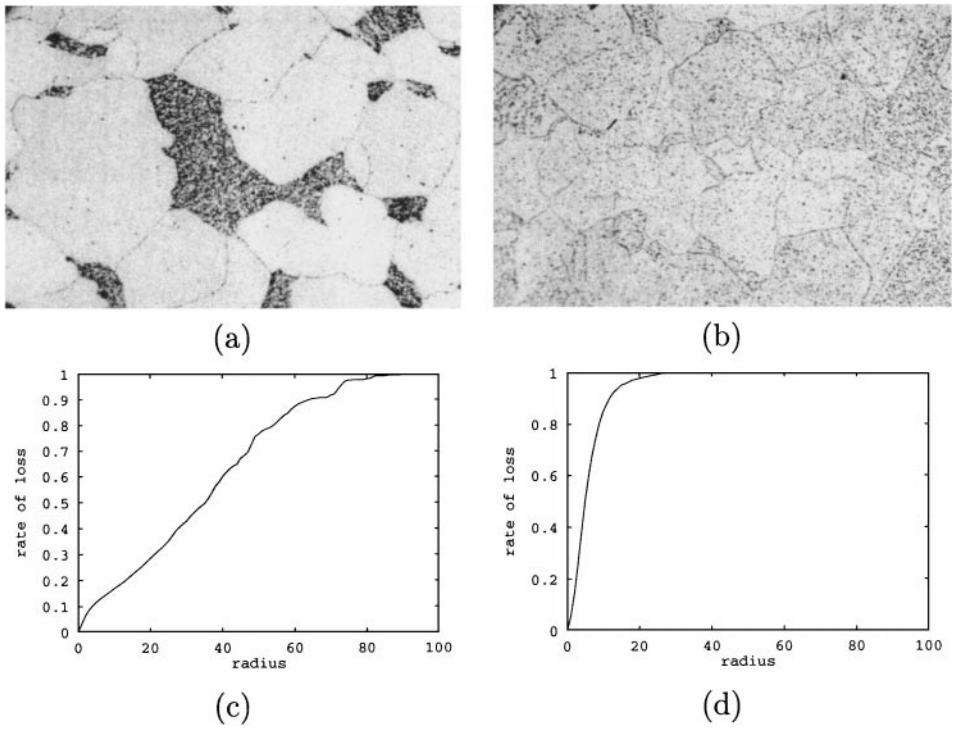

Fig. 8. *Images of temperature degraded boiler tubing, increasing from left-to-right. (a) sample y00; (b) sample y16; (c) and (d) are the granulometric curves for the images in (a) and (b) respectively*

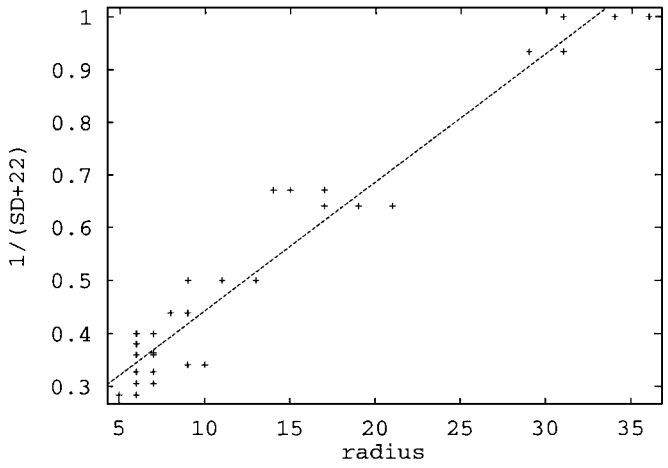

Fig. 9. *A linear representation of the data given in (a), modelled as* $SD = \frac{1}{a+bx} - 22$, *where* $a = 0.198769$, $b = 0.024406$ *and* x *represents the midpoint on the normalised granulometry curve. The sample correlation coefficient is 0.97*

meet. When the water level reaches the absolute maximum of the image, the process stops and the tessellation of all the catchment basins surrounded by dams constitutes the segmentation of the image. Figure 10 shows the "dam-building" process in action.

Such a tessellation is not usually useful, because each individual regional minimum in the image generates a region surrounded by a watershed line. To make this method more versatile, one can *impose* a smaller number of minima on the image and force the "water" to only come out of these imposed minima. This way, there will be only as many regions in the segmented image as there were imposed minima in the first place. The place where minima are imposed are usually called *markers* (or *seeds*). Although there are a number of algorithms for computing the watershed of an image, one of the most efficient and widely used is that of Vincent and Soille (1991).

To make use of the edge information, the watershed line is generally computed on a *gradient* of the original image. The art of the image analyst is to find a good combination of markers and gradient that will guarantee a robust segmentation.

Figure 11 shows an example of segmentation using the watershed method in a top-down manner. Figure 11(a) is an electron micrograph of man-made vitreous fibres (MMVF) on top of a porous polycarbonate filter. We wish to segment the fibres from the polycarbonate background. A single marker (Fig. 11(b)) is obtained from the pores of the fibres (which are the small, round, dark parts of the image). A second marker (Fig. 11(c)) is made of the internal, bright and elongated parts of the image. A simple morphological gradient is computed (Fig. 11(d)) by subtracting the erosion of the original image by a 3×3 structuring element from the original image. The watershed line is then computed (Fig. 11(e)) on the gradient image using both markers as imposed minima. Note how we have used a-priori knowledge to design our markers in order to obtain a segmentation of the image into two regions: the fibres and the background of the image. We will discuss the separation of the overlapping fibres in Section 3.4.2.

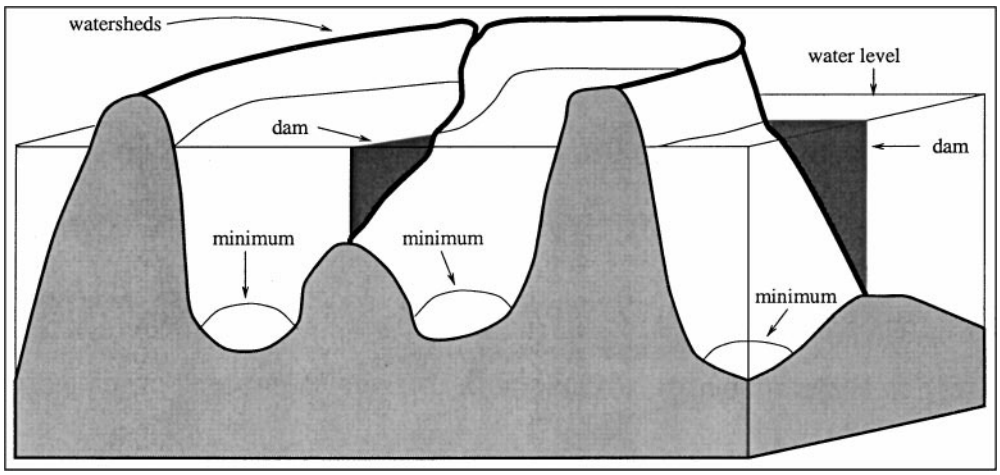

Fig. 10. *Building the watershed line*

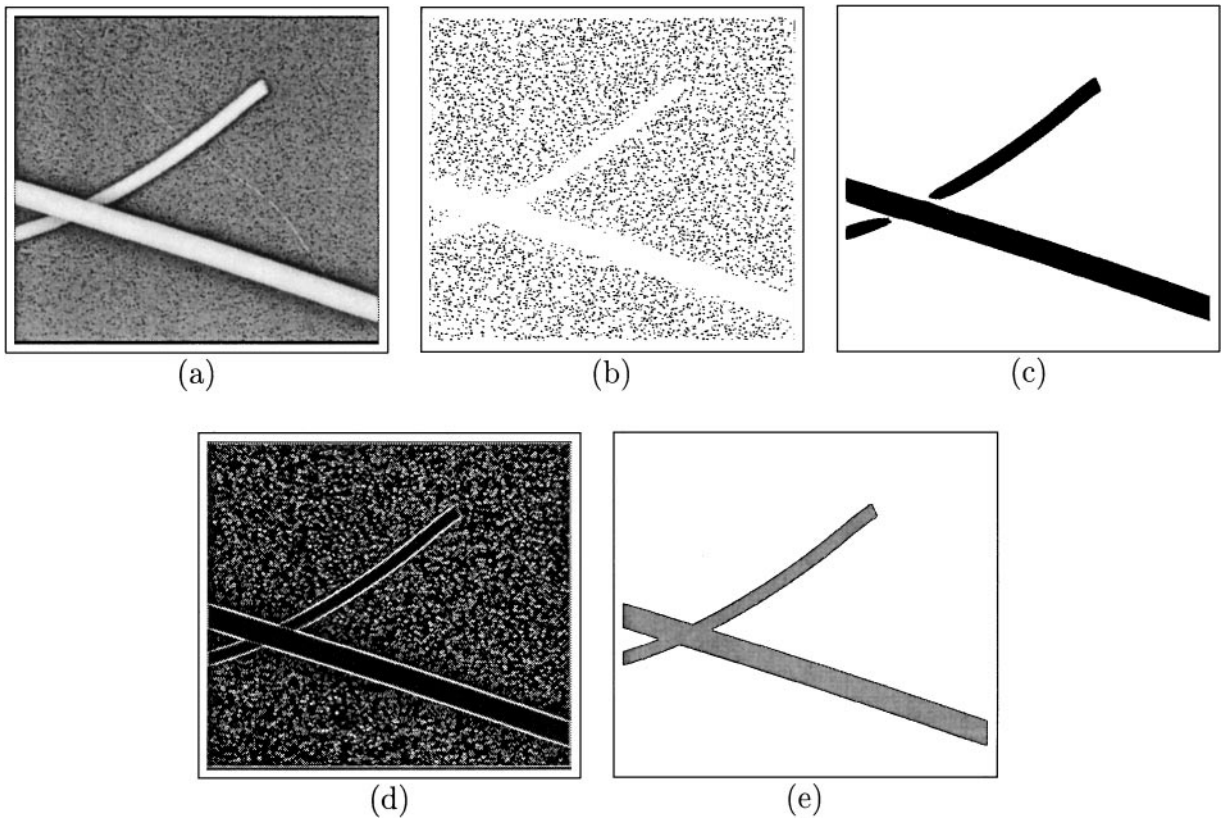

Fig. 11. *Example of segmentation using the watershed method*

Region growing methods. Region-growing methods (see for example Adams and Bischof (1994)) are similar in concept to the watershed algorithm. This time, however, no gradient is computed. Small starting regions of interest (markers, or seeds) are "planted" in the image at relevant locations. Statistics for the image data under the seeds are computed (for example, the mean value of all the pixels included in the seed), and neighbouring pixels at the boundary of each seed are considered in turn in a breadth-first manner for inclusion in each the growing region stemming from the seeds. The one that is closest to the mean (or other measure of central tendency) of its neighbouring region according to some measure (e.g. least squares) is added to that region. Statistics are re-computed each time a pixel is added to a region, and the boundary of the enlarged seeds is considered again. The algorithm terminates when all pixels in an image have been assigned to a region.

These methods extend naturally to multi-spectral images (for example colour images) and do not require the computation of a

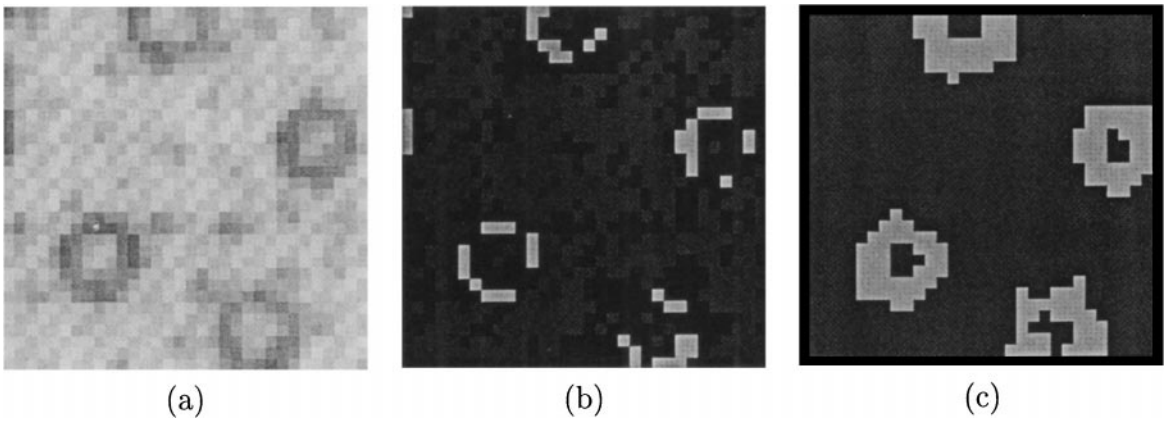

Fig. 12. *Segmentation of a colour image using a seeded region growing method. (a) is the first principal component of the original image, (b) is the original set of seeds (light grey for the foreground and dark grey for the background) and (c) shows the final segmentation*

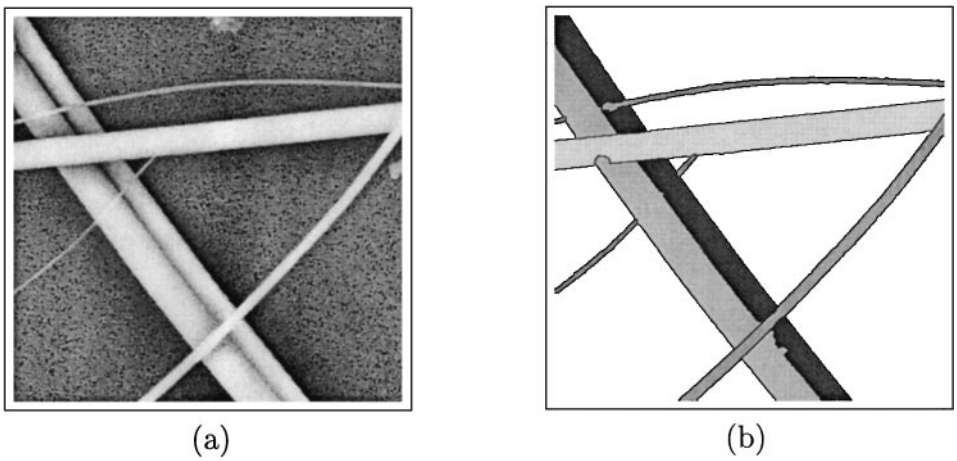

Fig. 13. *Sample segmentation: (a) original; (b) final segmentation result*

gradient, which can be difficult to calculate if regions are small or thin.

The art of the image analyst is again to find a good set of markers and relevant regional statistics for the problem at hand. Figure 12 shows an example of segmentation of a colour map using a region growing approach. The seeds were chosen by converting the original colour image to the principal component space, and performing two thresholdings on the first component: a threshold at a low value yielded the seeds for the foreground and a threshold at a high value yielded the seeds for the background.

3.4.2. Refining the segmentation and sizing

Segmenting an image usually doesn't stop at the region growing step. In order to obtain measurements, the initial segmentation might need to be refined. For example, Fig. 13 shows the continuation of the application shown in Fig. 11. The manufacturer of the MMVF required fibre diameter measurements for quality control purposes. Segmenting the fibres from the background goes part of the way, but in this case overlapping fibres need to be separated.

Figure 13(a) shows another typical electron micrograph of such a material: a number of overlapping fibres can be seen. Because a very precise measurement is needed for each fibre, a further segmentation of the image is required.

A watershed segmentation step similar to that described in Section 3.4.1 is performed, followed by a detection of crossing points, and a reconnection of well aligned objects across the crossing points. This step is described in detail in Talbot, Jeulin and Hanton (1996). Figure 13(b) is the result of the final segmentation.

Once the image has been segmented and each object identified, it is a simple task to measure the diameter of each individual object, and their distribution computed for quality control purposes.

3.5. *The marker problem*

Although region-growing methods present a versatile framework for image segmentation, they do not present a general method for image understanding. The task of finding appropriate gradient functions, a good set of statistics and relevant markers

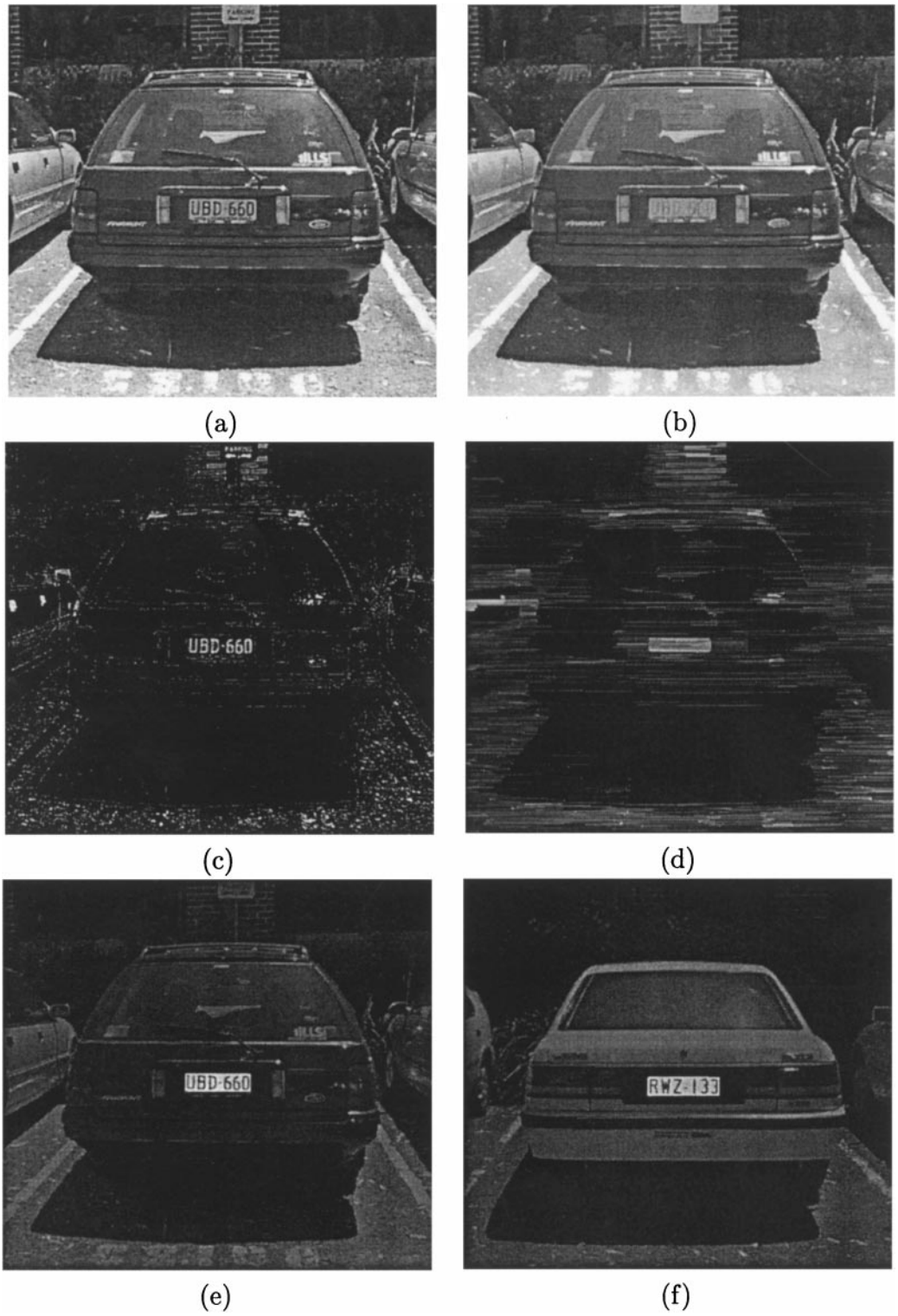

Fig. 14. *Extracting the characters in a car registration plate. See text for details*

is left to the image analyst. However, mathematical morphology provides a general framework for describing features present in images in terms of shape, orientation, size, etc., which themselves can be used as markers. Because the problem of finding markers is application-dependent, we can only present here a detailed application.

An application which uses the closing filter is illustrated in Fig. 14, where the goal is to segment the registration plate of a

car. An image of the rear end of a car is shown in Fig. 14(a). Note that the scene is fairly cluttered in this example, with bright areas of sunlight, shadows, lines in the brick-work of the walls, stickers on the car, a window wiper, etc. What distinguishes the registration plate is that it is a small white rectangular block parallel to the image frame, containing short dark lines. The use of size information is important because it allows one to distinguish the short dark lines in the lettering from the longer dark lines in the frame of the car. The goal of segmenting the registration plate can then be framed in the following way: find a cluster of short dark lines on a light background, forming a small rectangle parallel to the image frame. This leads to the following segmentation procedure for an input image f:

1. $g_1 = \bigwedge_i \phi_{L_i}(f)$
   In this first step, an infimum of closings (Matheron 1988, Soille, Breen and Jones 1996) using a set of line structuring elements $\{L_i\}$ of length 21 pixels orientated in 64 different directions is used to fill in short dark line segments in the image. It is necessary to apply the filter in many different directions to accommodate the many orientations of the line segments in the registration characters. The resulting image g is shown in Fig. 14(b).
2. $g_2 = g_1 - f$
   By subtracting the closed image from the original image (a *top-hat* transform (Serra 1982)), the short dark line segments can be extracted, as shown in Fig. 14(c).
3. $g_3 = \phi_L(g_2)$
   Any number of simple clustering techniques can now be used to separate the set of characters in the registration from all the other linear features detected in the image. Here we have applied a closing with a linear structuring element L of length 41 pixels and orientated in the horizontal direction. This joins the characters together into a single block parallel to the image frame; the result is shown in Fig. 14(d).
4. $g_4 = \gamma_R(g_3)$
   This subsequent opening using a rectangular structuring element R of dimensions 15×41 pixels removes everything but the largest block corresponding to the registration.
5. $g_5 = g_4 > 30$
   It is then a simple matter to threshold the image g_4 at an appropriate value to pull out markers for the registration plate. A region growing method can then be used to obtain a robust segmentation. The result is shown superimposed on the original image in Fig. 14(e). The result for a different car using Steps 1–5 is shown in Fig. 14(f).

3.6. *Component transforms*

Connected transforms have the property of simplifying the image while preserving contour information and have proved useful in a large number of applications, including image filtering and segmentation, multi-resolution decomposition and pattern recognition. They were first introduced in the form of *openings by reconstruction* (Klein 1976), which is an extension of the classical opening using structuring elements (Serra 1982), and subsequently generalised to gray-level images using the so-called *reconstruction process* (Vincent 1993b). Studies of connected operators and their relation to pyramids, flat zones and filters by reconstruction have been treated in Serra and Salembier (1993), Salembier and Serra (1995) and Crespo, Serra and Schafer (1995). Practical extensions of connected operators appear in Salembier and Oliveras (1996), Breen and Jones (1996b) and Salembier, Oliveras and Garrido (1998), while work on the efficient implementation of some connected operators has been presented in Vincent (1993a) and Breen and Jones (1996b).

A connected component $\mathcal{C}$ is defined as a set where every element can be connected by a path of elements contained within $\mathcal{C}$ (Gonzalez and Wintz 1987). We will refer to connected components simply as components. Within the 2D discrete domain $\mathbb{Z}^2$, the path used is typically either four or eight connected, and a component is called either 4-connected or 8-connected depending on the connectivity of the path used. For example, consider the objects in the binary image shown in Fig. 15(a). There are either two 8-connected components or four 4-connected components depending on the connectivity used, as indicated by the shaded objects in Fig. 15(b) and (c) respectively.

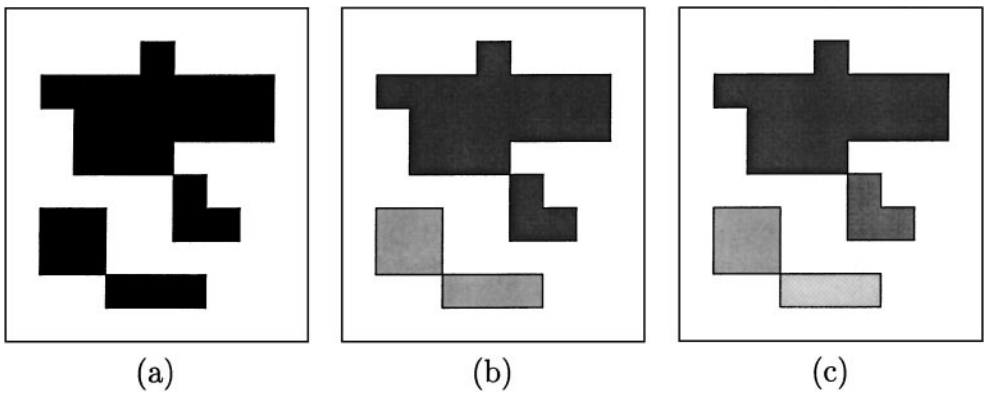

Fig. 15. *Components in a binary set. (a) Example set. (b) There are two 8-connected components. (c) There are four 4-connected components*

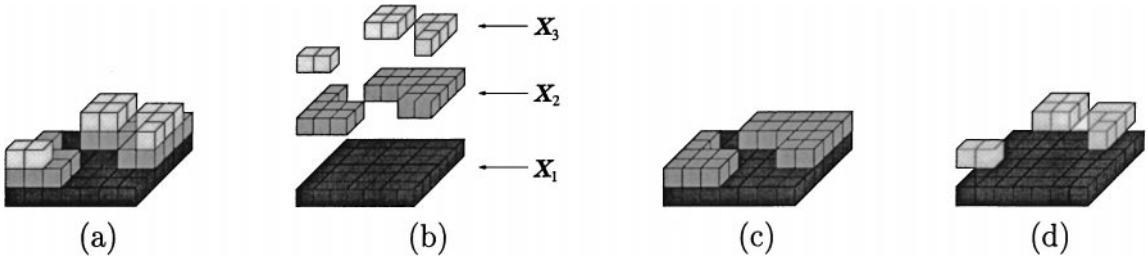

Fig. 16. *The component transform. (a) Perspective view of a greyscale image. (b) The components; there are 3 in the threshold set $\mathcal{X}_3$, 2 in the threshold set $\mathcal{X}_2$ and 1 in the threshold set $\mathcal{X}_1$. (c) Result from component transform with criterion:* Area must be greater than *4* pixels*. (d) Gaps between components in the filtered image can occur when non-increasing criteria are used*

Although the notion of components is traditionally applied to binary images, it can be extended to greyscale images using *threshold sets* (Wendt, Coyle and Gallagher 1986). A threshold set $\mathcal{X}$ at a threshold t is defined:

$$\mathcal{X}_t(f) = \{x \mid f(x) \geq t\}. \tag{4}$$

A greyscale component can then be defined as a component $\mathcal{C} \subseteq \mathcal{X}_t(f)$, for some threshold t. For example, there are in total 6 greyscale components (considered over all 3 thresholds) in the image shown in Fig. 16(a), as illustrated in Fig. 16(b).

A component transform removes components from the image that do not satisfy a given criterion, denoted by $\mathcal{T}$ (Serra and Vincent 1992, Breen and Jones 1996a, b). The criterion is based on one or more of the component's attributes, for example the area or perimeter of the component, and requires that the values of these attributes lie within certain given thresholds. Component transforms have the desirable property of simplifying the image while preserving contour information. In this way, component transforms are similar to the opening by reconstruction described in Section 3.2. Figure 16(c) shows the result when a component transform is applied to the image in Fig. 16(a), where the criterion used is: *area must be greater than* 4 *pixels*. All those components in the image that have an area of 4 pixels or less are removed by the transform; the remaining components are preserved in their entirety.

The component transform has the properties of anti-extensivity and idempotence and is therefore what is known in mathematical morphology as a *thinning* (Serra 1982). If in addition we consider only filtering criteria $\mathcal{T}$ that are increasing, it can then be described as an opening. We define an increasing criterion by the fact that if a component $\mathcal{C}$ satisfies the criterion, then so will any superset of $\mathcal{C}$. The criterion *area must be greater than* λ is an example of an increasing criterion. If a component $\mathcal{C}$ has an area greater than λ elements, then so will any superset of $\mathcal{C}$. Amongst many other attributes, the area or perimeter of the convex hull, the maximum Feret diameter and the area of the smallest enclosing circle are all attributes on which increasing criteria can be based. On the other hand, there are many interesting attributes that are non-increasing; the perimeter of the component, maximum geodesic distance of the component (Lantuéjoul and Beucher 1981); the length of the minimal skeleton (Vincent 1991); and the major and minor axes of the ellipse that best fits the region (according, say, to a method-of-moments criterion). Note that when a non-increasing criterion is used in the component transform, we have the interesting situation where it is possible to have gaps between components in the filtered image. Figure 16(d) shows an example using the non-increasing criterion: *area must be less than* 5 *pixels or greater than* 15 *pixels*. We treat this problem by imposing the constraint that is true for increasing criteria: If a component $\mathcal{C}$ satisfies the criterion, then so does any superset of $\mathcal{C}$. For the example shown in Fig. 16(d), the resulting filtered image would then be the original input image in Fig. 16(a). Other treatments of this problem can be found in Breen and Jones (1996b) and Salembier and Oliveras (1996).

A practical example of a component transform using a non-increasing criterion is shown in Fig. 17 (Sun and Wu 1997).

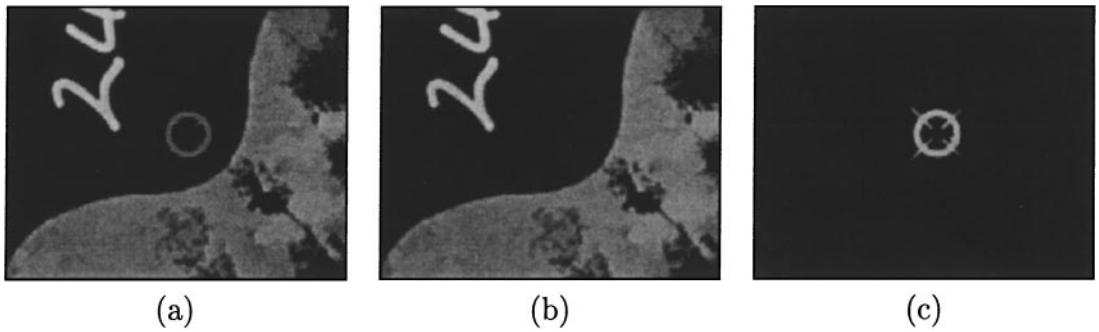

Fig. 17. *Automated segmentation of fiducial markers. (a) Input image. (b) Component thinning by eccentricity. (c) Subtraction of image in (b) from image in (a) (a top-hat filter)*

The input image in Fig. 17(a) is a small section of an aerial photographic image. The small circular shape in the centre is a fiducial mark which is used to define the frame of reference for spatial measurements on aerial photographs. By using a component transform that removes image components with eccentricity of the best-fit ellipse approximately equal to unity, the circular fiducial marks can be removed from the image, as shown in Fig. 17(b). Note that eccentricity is a non-increasing function, and thus the corresponding component transform is a thinning. By then subtracting this result from the original image, Fig. 17(c), the fiducial marks are extracted and easily segmented using a constant threshold.

4. Higher level representations

In order to efficiently describe and implement more complex image transforms, it is convenient to use a higher level image representation other than the basic notion of an image support mapping to a set of values (viz. $f : \mathbb{R}^n \rightarrow \mathbb{R}$). In this section, we discuss a graph representation of the image, called a *component graph*, which affords a suitable framework for the description and implementation of greyscale component transforms.

A graph is defined as a set of vertices connected by a set of edges. In the context of a greyscale image, each vertex represents a single component in the image. The vertex is an abstract representation of the component that may be anything from a list of all the pixels in the component to a single attribute such as the area or perimeter of the component. The edges between vertices represent some relation between components in the image; in a component graph they represent inclusion. That is, a directed edge $E(\mathcal{C}_1 \rightarrow \mathcal{C}_2)$ exists between two components if $\mathcal{C}_1 \subseteq \mathcal{C}_2$. In contrast, the edges in the undirected graph used to define image supports in Section 2.2 represent the neighbourhood relation between vertices.

A diagrammatic illustration of a component graph is shown in Fig. 18. The example image in Fig. 18(a) has a total of 6 components, each of which has a representative vertex labelled A to F in the component graph shown in Fig. 18(b). The arrows in the figure show pairs of components that are related by inclusion. Notice that, as any given component $\mathcal{C}$ satisfies $\mathcal{C} \subseteq \mathcal{C}$, each vertex has an arrow that points back to itself. The vertices at the top of the graph coincide with the three regional maxima of the image and thus have only one edge directed into them.

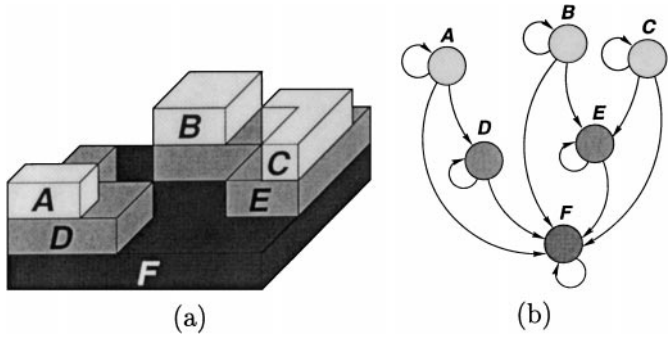

Fig. 18. *Component graph representation of a greyscale image. (a) Perspective view of greyscale image from Fig. 16. (b) Corresponding component graph*

Image filtering and segmentation using component graphs can be conveniently formalised using the notion of a *transitive closure* from graph theory (Kaufmann 1972). The transitive closure Γ of a vertex v is defined as the set of all vertices on the path in the graph originating at v. Referring to Fig. 18(b) for example, $\Gamma(A) = \{A, D, F\}$, $\Gamma(B) = \{B, E, F\}$ and $\Gamma(C) = \{C, E, F\}$. As vertices in a component graph represent components in a greyscale image, it is a simple matter to apply the notion of a transitive closure to greyscale images. Figure 19(a) shows an example on a small greyscale image with three of its regional maxima labelled A, B and C. The transitive closures for each of these regional maxima are shown respectively in Figs. 19(a)–(c).

Greyscale component transforms can be defined in terms of transitive closures. An example of such a transform is that introduced by Jones (1997), as follows:

$$\gamma(f) = \max_{r \in R(f)} \Gamma(r). \tag{5}$$

where the set $R(f)$ is a subset of the complete set of regional maxima $\mathcal{R}(f)$ in the image and $\Gamma(r)$ is the transitive closure of a given regional maximum $r \in R(f)$. This component transform will preserve any image features whose corresponding regional maxima are included in the set $R(f)$. In order to generate this set, it is necessary to employ a classification process which selects only those regional maxima that are within features that should be preserved by the filtering process. Regional maxima that are not included in this set will be filtered as much as possible while preserving the rest of the image.

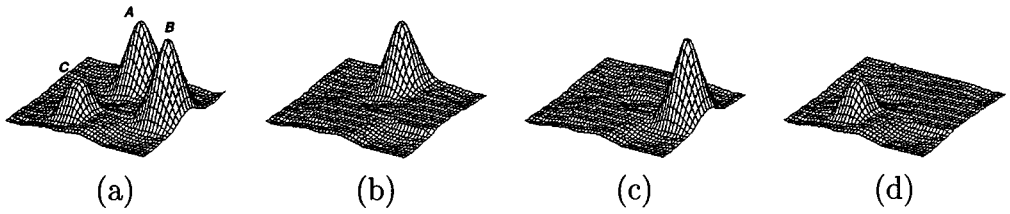

Fig. 19. *Transitive closures on a greyscale image. (a) Perspective view of input image with regional maxima labelled A, B and C. (b), (c) and (d) Transitive closures $\Gamma(A)$, $\Gamma(B)$ and $\Gamma(C)$ respectively*

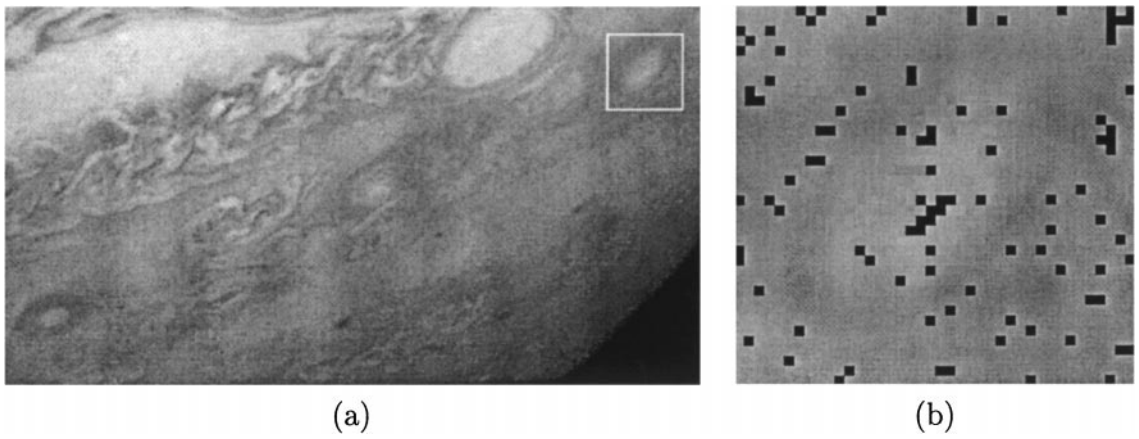

Fig. 20. *Elliptical objects in an image of part of the planet Jupiter. (a) Input image with an elliptical object indicated by the small white box. (b) Enlargement of this region with regional maxima superimposed in black*

The approach is illustrated on the image of part of the planet Jupiter shown in Fig. 20(a). The elliptical objects in this image, one of which is indicated by the small white box in Fig. 20(a), can be filtered and segmented using a greyscale component transform constructed from transitive closures. Figure 20(b) shows an enlargement of the small boxed region, with regional maxima superimposed in black. By defining $R(f)$ to be the set of all regional maxima outside the elliptical region, the ellipse can be filtered and segmented using equation (5). In order to automatically select the set of regional maxima in $R(f)$ we can employ the notion of an *attribute signature* (Jones 1997). Each vertex in the component graph represents a single attribute of the component, such as area or perimeter. We define an attribute signature of a vertex v as the sequence of vertex attributes in the transitive closure $\Gamma(v)$.

The regional maxima in the image are automatically classified on the basis of their attribute signatures. Figure 21(a) shows a typical attribute signature for a regional maximum that lies within the elliptical region, where greyscales are along the horizontal axis and attribute values are along the vertical axis. The attribute used here is eccentricity, which is defined as the length of the minor axis of the best fit ellipse to the component divided by the length of the major axis. Based on the shape of the elliptical regions in this image, we expect the eccentricity to be about 0.5 for components corresponding to the elliptical image features (a 2-1 ratio of major to minor axes). The attribute signature in Fig. 21(a) does indeed contain eccentricity values of about 0.5 at greyscales corresponding to those of the elliptical region (in the range 120 to 180). This eccentricity value also occurs at low image greyscales (less than around 80) because as the greyscale falls the shape of the component tends towards the actual image support and, for this image, the image support has an eccentricity of approximately 0.5. In contrast, an attribute signature for a regional maximum outside the elliptical region is shown in Fig. 21(b). This signature does not exhibit the required eccentricity except at low image values and so may be easily distinguished from the signature in Fig. 21(a). By combining other attributes, such as component area, it is relatively easy to devise a procedure based on attribute signatures that automatically classifies regional maxima into those that lie within elliptical image features and those that do not.

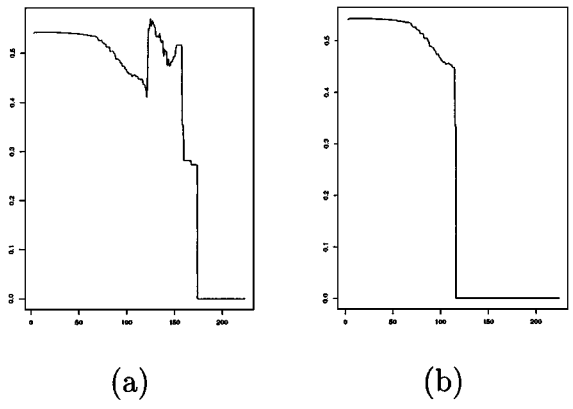

Fig. 21. *Attribute signatures using eccentricity as the attribute. (a) Attribute signature for a typical regional maximum within the elliptical feature. (b) Attribute signature for a typical regional maximum in the background of the image*

The results from the classification process are shown in Fig. 22(a), where the white points indicate regional maxima that have signatures classified as belonging to the elliptical region. The result from the component transform γ (equation (5)) is shown in Fig. 22(b). Notice that the background has been preserved completely, while the elliptical region has been removed as much as possible without modifying the components that are connected to the background. This is a useful feature because a segmentation of the elliptical region can be obtained simply by locating those pixels that have been changed by γ; see Fig. 22(c).

The result for the entire image is shown in Fig. 23, where the segmented elliptical regions are bordered in white.

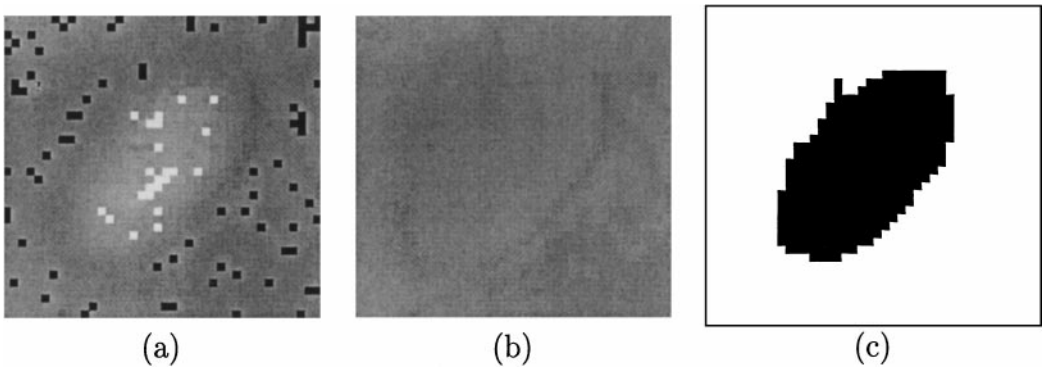

Fig. 22. *Filtering and segmentation results. (a) Results from the classification process, where the white points indicate regional maxima that have signatures classified as belonging to the elliptical region. The black points indicate regional maxima that will be included in the set R(f). (b) Result for filtered image γ; the elliptical feature has been removed because its regional maxima were not included in R(f). (c) Result for segmented image: the set of pixels that have been changed by γ*

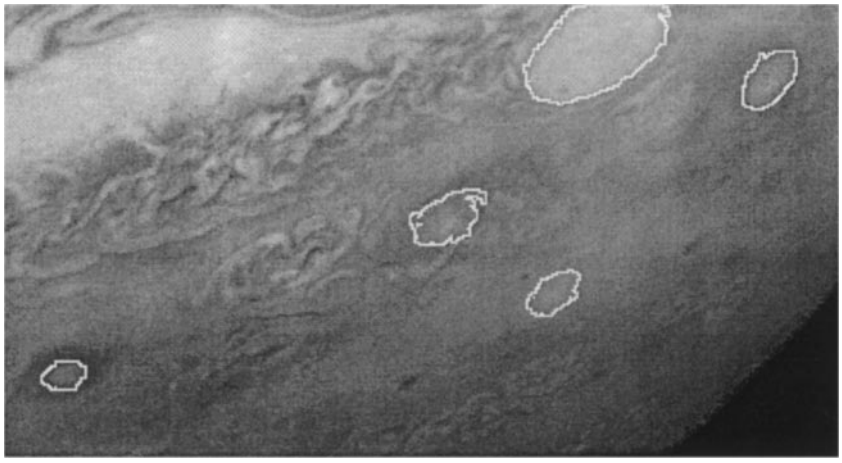

Fig. 23. *Segmentation result on full image*

5. Conclusion

In this paper we have demonstrated the utility of morphological image analysis for image filtering and image segmentation. We have looked at image reconstruction and sizing transforms, and have presented a procedure for segmenting automobile registration plates. We have seen how to use connected component transforms to isolate image regions and have developed a graph notation for image representation that allowed us to perform image segmentation on complex images in a very efficient manner.

We have explored the fundamental and representation concepts in mathematical morphology. Its rigorous framework based on set theory makes it a suitable branch of research for the theoretician. On the other hand, high-level concepts such as component graphs and transitive closures can prove useful to the image analysis practitioner as well.

References

Adams R. and Bischof L. 1994. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 16: 641–647.

Borgefors G. 1986. Distance transformations in digital images. Computer Vision, Graphics, and Image Processing 34: 344–371.

Breen E. and Jones R. 1996a. An attribute-based approach to mathematical morphology, In: Maragos P., Schafer R., and Butt M. (Eds.), Mathematical Morphology and its Applications to Image and Signal Processing. Kluwer Academic Press, Atlanta, pp. 41–48.

Breen E.J. and Jones R. 1996b. Attribute openings, thinnings and granulometries. Computer Vision and Image Understanding 64(3): 377–389.

Breen E., Jones R., and Talbot H. To appear. The morphological approach to industrial image analysis applications. Acta Stereologica.

Coade R. 1993. Determination of effective temperature for residual life, assessment of microstructural methods. Technical Report ESAA/1.7/MET/WP 005.

Crespo J., Serra J., and Schafer R. 1995. Theoretical aspects of morphological filters by reconstruction. *Signal Processing* 47(2): 201–225.

Gonzalez R. and Wintz P. 1987. Digital Image Processing, 2nd edn. Reading, Addison Wesley, Massachusetts.

Heijmans H., Nacken P., Toet A., and Vincent L. 1992. Graph morphology. Journal of Visual Communication and Image Representation 3(1): 24–38.

Jones R. 1997. Component trees for image filtering and segmentation. In: Coyle E. (Ed.), Proceedings of the 1997 IEEE Workshop on Nonlinear Signal and Image Processing. Mackinac Island.

Jones R., Berman M., Jiang Y., Buckley M., and Drew M. 1995. Image analysis of microstructural steels exposed to elevated temperatures. Technical Report E 95/20, CSIRO, Division of Mathematics and Statistics.

Kaufmann A. 1972. Points and Arrows: the Theory of Graphs. Transworld Student Library, Transworld Publishers Ltd., London.

Klein J.C. 1976. Conception et réalisation d'une unité logique pour l'analyse quantitative d'images. PhD Thesis, Nancy University, France.

Lantuéjoul C. and Beucher S. 1981. On the use of the geodesic metric in image analysis. Journal of Microscopy 121: 39–49.

Matheron G. 1967. Eléments pour une théorie des milieux poreux. Masson, Paris.

Matheron G. 1975. Random Sets and Integral Geometry. Wiley, New York.

Matheron G. 1988. Filters and lattices. In: Serra J. (Ed.), Image Analysis and Mathematical Morphology, Vol. II: Theoretical Advances. Academic Press, London, pp. 115–140.

Meyer F. and Beucher S. 1990. Morphological segmentation. Journal of Visual Communication and Image Representation 1(1): 21–46.

Ronse C. and Heijmans H. 1991. The algebraic basis of mathematical morphology-part 2: Openings and closings. Computer Vision, Graphics, and Image Processing: Image Understanding 54(1): 74–97.

Salembier P. and Oliveras A. 1996. Practical extensions of connected operators. In: Maragos P., Schafer R.W., and Butt M.A. (Eds.), Mathematical Morphology and its Application to Image and Signal Processing. Kluwer Academic Publishers, Atlanta, pp. 97–110.

Salembier P., Oliveras A., and Garrido L. 1998. Anti-extensive connected operators for image and sequence processing. IEEE Transactions on Image Processing 7(4): 555–570.

Salembier P. and Serra J. 1995. Flat zones filtering, connected operators and filters by reconstruction. IEEE Transactions on Image Processing 3(8): 1153–1160.

Serra J. 1982. Image Analysis and Mathematical Morphology. Academic Press, London.

Serra J. (Ed.) 1988. Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances. Academic Press, London.

Serra J. and Salembier P. 1993. Connected operators and pyramids. In: Image Algebra and Mathematical Morphology, Vol. 2030. SPIE, San Diego, pp. 65–76.

Serra J. and Vincent L. 1992. An overview of morphological filtering. Circuits, Systems and Signal Processing 11(1): 47–108.

Soille P., Breen E., and Jones R. 1996. Recursive implementation of erosions and dilation along discrete lines at arbitrary angles. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(5): 562–567.

Soille P., Serra J., and Rivest J.-F. 1992. Dimensional measurements and operators in mathematical morphology. In: Nonlinear Image Processing III. pp. 127–138.

Sternberg S. 1986. Grayscale morphology. Computer Graphics and Image Processing 35: 333–355.

Sun C. and Wu X. 1997. A method for automatic segmentation of fiducial markers. In: Pan H., Brooks M., McMichael D., and Newsam G. (Eds.), Image Analysis and Information Fusion. CSSIP, Adelaide, pp. 43–52.

Talbot H., Jeulin D., and Hanton D. 1996. Image analysis of insulation mineral fibres. Microscopy, Microanalysis and Microstructures 7: 361–368.

Vandroogenboeck M. and Talbot H. 1996. Fast computation of morphological operations with arbitrary structuring elements. Pattern Recognition Letters 17: 1451–1460.

Vincent L. 1989. Graphs and mathematical morphology. Signal Processing 16: 365–388.

Vincent L. 1990. Algorithmes morphologiques à base de files d'attente et de lacets. Extension aux graphes. PhD Thesis, Ecole des Mines de Paris.

Vincent L. 1991. Efficient computation of various types of skeletons. In: Medical Imaging V. SPIE, San Jose.

Vincent L. 1992. Morphological algorithms. In: Dougherty. E. (Ed.), Mathematical Morphology in Image Processing. Marcel-Dekker, New York, pp. 255–288.

Vincent L. 1993a. Grayscale area openings and closings, their efficient implementation and applications. In: Serra J. and Salembier P. (Eds.), Mathematical Morphology and its Applications to Signal Processing. UPC Publications, Barcelona, pp. 22–27.

Vincent L. 1993b. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. IEEE Transactions on Image Processing 22(2): 176–201.

Vincent L. and Soille P. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(6): 583–598.

Wendt P.D., Coyle E.J., and Gallagher N.C. 1986. Stack filters. IEEE Transactions on Acoustics, Speech and Signal Processing 34(4): 898–911.