

The use of gradient direction in pre-processing images from crystallization experiments

Julie Wilson^{a*} and Ian Berry^b

Received 26 October 2004

Accepted 8 March 2005

^aYork Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, UK, and ^bDivision of Structural Biology and the Oxford Protein Production Facility, The Henry Wellcome Building for Genomic Medicine, Roosevelt Drive, Oxford OX3 7BN, UK. Correspondence e-mail: julie@ysbl.york.ac.uk

Robots are now used routinely to perform crystallization experiments and many laboratories now have imaging systems to record the results. These images must be evaluated rapidly and the results fed back into optimization procedures. Software to analyse the images is being developed; described here are methods to restrict the area of the image to be analysed in order to speed up processing. Properties of the gradient of greyscale images are used to identify first the well and then the crystallization drop for various crystallization trays and different imaging systems. Methods are discussed to identify artefacts in the images that are not related to the experimental outcome, but can cause problems for the machine-learning algorithms used in classification and waste time during analysis. Gradient angles are exploited to eliminate faults in the crystallization trays, bubbles and splatter droplets prior to analysis.

© 2005 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

Protein crystallography can often provide the three-dimensional structures of macromolecules necessary for functional studies and drug design. However, identifying the conditions that will provide diffraction-quality crystals for structural biology is not straightforward. Numerous reagents and additives must be tested in combination with variation in concentration, pH and temperature, and often very many trials are necessary to determine suitable crystallization conditions. The results of these experiments must be assessed repeatedly over a period of time and integrated into optimization protocols. Robotic systems are routinely used to perform more automated crystallization experiments in smaller laboratories as well as large structural genomics centres, and a number of systems are now available for image acquisition and storage. The bar-coding of crystallization trays links experimental conditions to the results and the use of databases allows the information to be used for intelligent crystallization recipe prediction. With robots capable of producing many thousands of experiments a day in high-throughput mode, inspection of the results by eye is becoming increasingly impractical.

Software to analyse the images and classify the results is being developed (see Bern *et al.*, 2004; Cumbaa *et al.*, 2003; Wilson, 2002, 2004) and will ideally provide results that can be exploited in subsequent experiments. Obviously crystals must be identified, not only large single crystals but also thin needles and plates. In the absence of such success, the conditions closest to those required for crystallization must be recognized. Such conditions may be apparent by the occur-

rence of micro-crystals, spherulites (rounded but crystal-like objects) and so-called sea-urchins (spiky nucleation sites from which needle crystals often start to grow). Other phenomena, such as phase separation or crystalline precipitate, indicate conditions that may just need slight adjustment in order to promote crystal growth. However, experiments resulting in heavy amorphous precipitate or denatured protein show that the conditions are not suitable for crystal growth. In this way the results may be graded according to the experimental outcome. However, the images also exhibit other artefacts due to the experimental setup, such as dust and other foreign bodies, air bubbles and defects in the crystallization trays. These objects often cause problems in classification and valuable time is wasted on their analysis. The aim here is to remove commonly occurring but irrelevant objects prior to analysis.

The crystallization drop may be relatively small in relation to the well in which the robot deposits it and, even if centred originally, may migrate across the well as the tray is moved. Imaging systems must ensure that the crystallization drop is entirely within the image. This means capturing the entire well and thus much of the image is not of interest as far as the experimental result is concerned. It is therefore important to identify the region of interest early in the analysis.

2. Imaging systems

The Oxford Protein Production Facility (OPPF) at the University of Oxford supplied most of the images used in this paper. Crystallization experiments are performed in 96-well Greiner plates (micro-titre format) and the images are taken

using an automated Oasis 1700 imaging system (Veeco, Cambridge, UK). Native images are $1024 \times 1024 \times 8$ bit bitmap (BMP) images (~ 1 Mbyte in size), corresponding to a pixel width of about $3 \mu\text{m}$. Additionally, Fig. 1(a) shows an image of a $1 \mu\text{l}$ drop in an Art Robbins Intelliplate (Hampton

Research) acquired with the Tritek Crystal Pro imaging system at the Synchrotron Radiation Source in Daresbury. The original colour image ($1280 \times 1014 \times 8$ bit BMP) was converted to greyscale here. Fig. 1(e) shows a $1024 \times 768 \times 8$ bit BMP image of an experiment performed in a Greiner low-

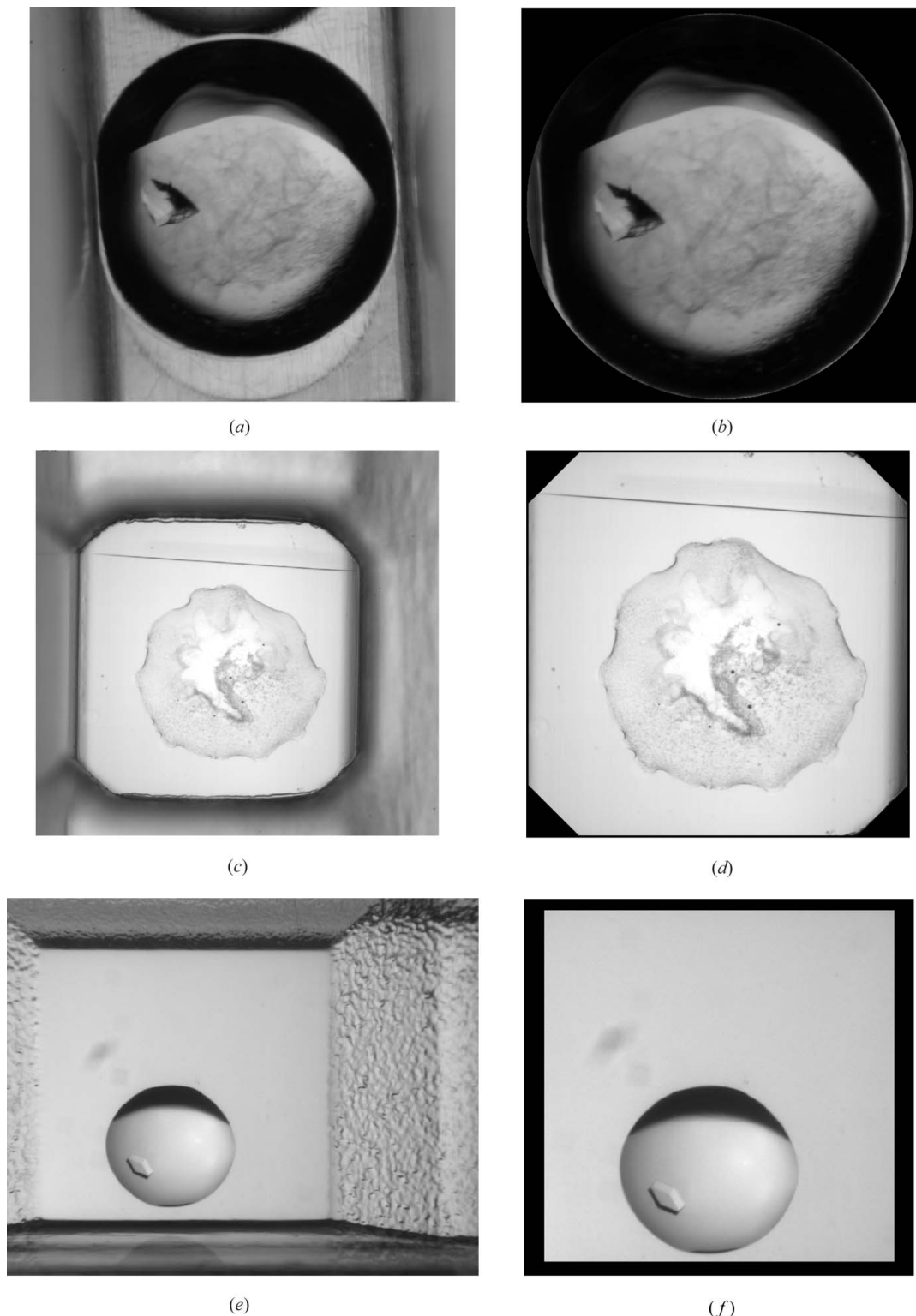
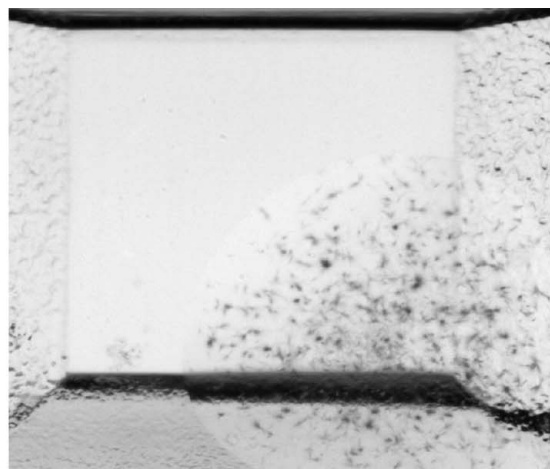


Figure 1 Three different types of crystallization trays are shown in (a), (c) and (e). The corresponding masked wells are shown in (b), (d) and (f), with the images reduced in size accordingly.

profile crystallization tray and was supplied by the National Kanker Institute, in Amsterdam. The image was acquired using the BioTom storage and visualization robot. The $1000 \times 880 \times 8$ bit JPEG image shown in Fig. 2, also a Greiner low-profile plate, was taken at the Protein Structure Factory in Berlin using their in-house system.

3. The gradient of a greyscale image

The edges of objects give rise to sudden changes in intensity and can be identified by analysing the gradient of the image intensity, or rate of change of the greyscale. Here specific patterns in the direction of the steepest gradient are used to locate particular types of object early in processing and therefore save time in analysis.



(a)



(b)

Figure 2
A well in which the crystallization drop is overlapping the sloping slides.

Mathematically, the gradient is calculated by differentiation, but simple operators can be used to approximate the gradient of an image (Sobel, Prewitt or Roberts for example). The Sobel operator used here (see Figs. 3 and 4) approximates the rate of change in x (the horizontal direction) at the pixel, x_0 , using the filter

$$G_x \simeq (x_7 - x_1) + 2(x_6 - x_2) + (x_5 - x_3).$$

Here x_1, \dots, x_8 are the pixel's immediate neighbours arranged as follows:

$$\begin{array}{ccc} x_1 & x_2 & x_3 \\ x_8 & x_0 & x_4 \\ x_7 & x_6 & x_5 \end{array}$$

Similarly, the rate of change in y (the vertical direction) is approximated by

$$G_y \simeq (x_3 - x_1) + 2(x_4 - x_8) + (x_5 - x_7).$$

The magnitude of the gradient, M , giving the steepest change at x_0 , is calculated from

$$M = (G_x^2 + G_y^2)^{1/2}$$

with the direction of this change determined by the angle

$$\alpha = \tan^{-1}(G_y/G_x),$$

measured with respect to the x axis.

To locate important details, some threshold must be specified for a significant gradient magnitude. Taking all pixels with magnitudes above this threshold allows objects to be defined as connected sets of pixels and analysed separately (Wilson, 2002). The local maxima of these pixels define the edges in an image and more complicated algorithms, such as that of Canny (Canny, 1986), can also allow weak edges to be followed where they are connected to strong edges. Spraggon *et al.* (2002) and Bern *et al.* (2004) use Canny edge detection in the analysis of crystallization images. Other methods use the rate of change of the gradient, or Laplacian operator, to detect edges (see *e.g.* Gonzalez & Woods, 2002). Cumbaa *et al.* (2003) use an approximation to the Laplacian to detect image discontinuities in crystallization images.

At any edge point, the direction of the steepest gradient along an edge is always perpendicular to the edge so that the angle, $\alpha(x, y)$, will be constant along a straight-line edge and change gradually around smooth curves, as can be seen in Fig. 3. The original image is shown in (a) and the gradient magnitudes are shown in (b). In (c) the gradient direction is plotted as an angle in degrees (also rescaled to values in $[0, 255]$). The horizontal line across the image is clearly visible in the gradient direction where the greyscale reflects the constant angle. The gradual change in angle around the edge of the crystallization drop is also evident in the similarity of the greyscale values.

4. The initial well mask

Many different crystallization trays are available and each brings new challenges (see Fig. 1 for some examples). In all

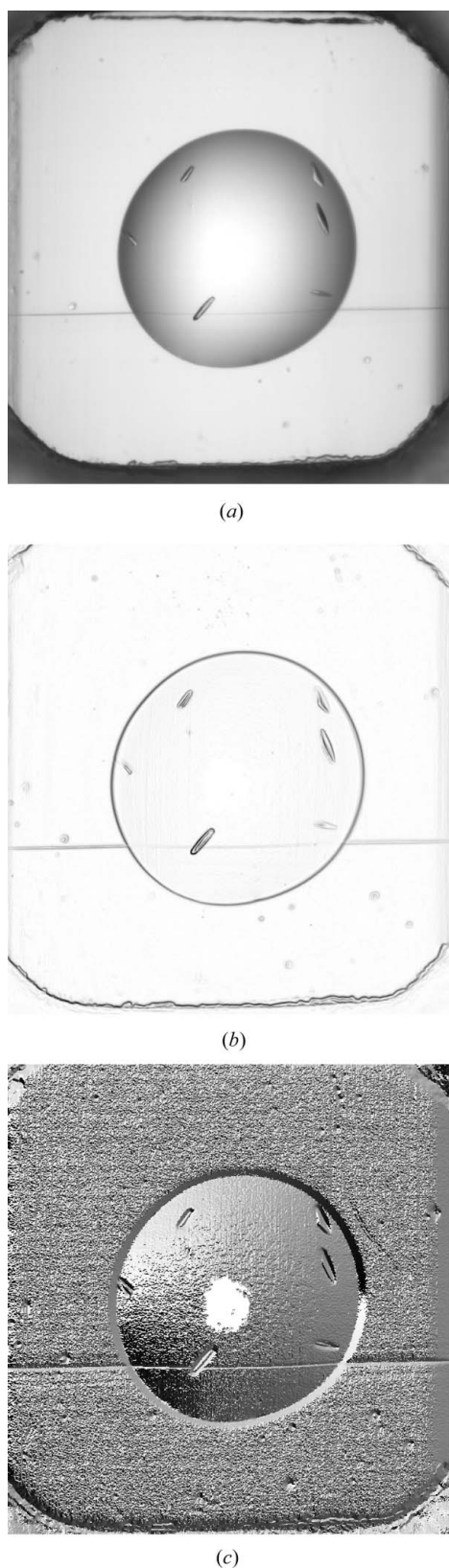


Figure 3
 The gradient magnitudes obtained using the Sobel operator for the image in (a) are shown in (b). The values have been rescaled to lie in [0, 255], the possible values for an 8 bit greyscale image, and darker pixels indicate higher values. In (c) the direction of the steepest gradient is plotted as an angle between 0 and 359° (also rescaled for plotting).

cases, however, the crystallization drop only covers a small area of the initial image and, in order to speed up the image processing, the size of the image should be reduced as soon as possible. As the drops are dispensed by robots, their exact position in the well is difficult to control and can be affected by movement of the trays during imaging and storage. Trays with curved wells have been developed to overcome this problem (see Fig. 1a), but these create severe shadows that are difficult to deal with, and even identifying the boundary of the crystallization drop can be very problematic. Before attempting to locate the crystallization drop, however, the well edges can be masked and this alone can reduce the image to $\sim 1/3$ of its original size. An initial mask can be obtained using a Hough transform, for example, to identify the circle (Hough, 1962) and, in wells with vertical sides (Fig. 1b), the gradient direction clearly shows the sharp change in greyscale in the horizontal and vertical directions. This is not as straightforward in wells with sloping sides (Fig. 1e). Again the horizontal boundaries of the well are easily identified, but the effects of the plastic mould make the vertical sides more difficult. After the top and bottom of the image have been masked, the variation in intensities across each row of pixels can be used to find the vertical sides. This is further complicated by the fact that the crystallization drop can overlap the edges of the well as seen in Fig. 2. Here only rows for which a certain percentage of pixels have very low variation (the bottom of the well) are used to identify the horizontal edges. This ensures that only rows that do not pass through the drop are used. The pattern caused by the plastic would make analysis of the overlapping part of the drop extremely difficult and, rather than compromise the classification algorithm, we mask off this area of the drop using the bottom of the well as the new boundary of the image and only the part of the drop that does not overlap the well sides is analysed (see Fig. 2b).

5. Removing artefacts due to the crystallization trays

Before attempting to identify the limits of the crystallization drop, other artefacts due to the experimental setup can be identified and eliminated. In Fig. 1, a horizontal line can be seen across the image in both the gradient magnitudes and the gradient direction. Such lines are found in many images and are effects of the manufacturing process for the crystallization trays. They are created where the molten plastic meets as it is injected into the mould from different points. Straight lines are characteristic of many crystals and so this can cause serious problems during classification. These lines are generally horizontal or vertical and can usually be detected through the gradient direction and eliminated prior to analysis. Fig. 1(c) shows the line clearly in the plot of the gradient angle. This is because the maximum gradient along the line is perpendicular to it and therefore we see angles of 270° closely followed by angles of 90°. In Fig. 5, only angles very close to 90° (light grey) or 270° (dark grey) are shown for this image. In this particular case the line is very straight, but in many cases it is distorted by the lens effect of the crystallization drop. This effect is dependent on the curvature of the drop and is more

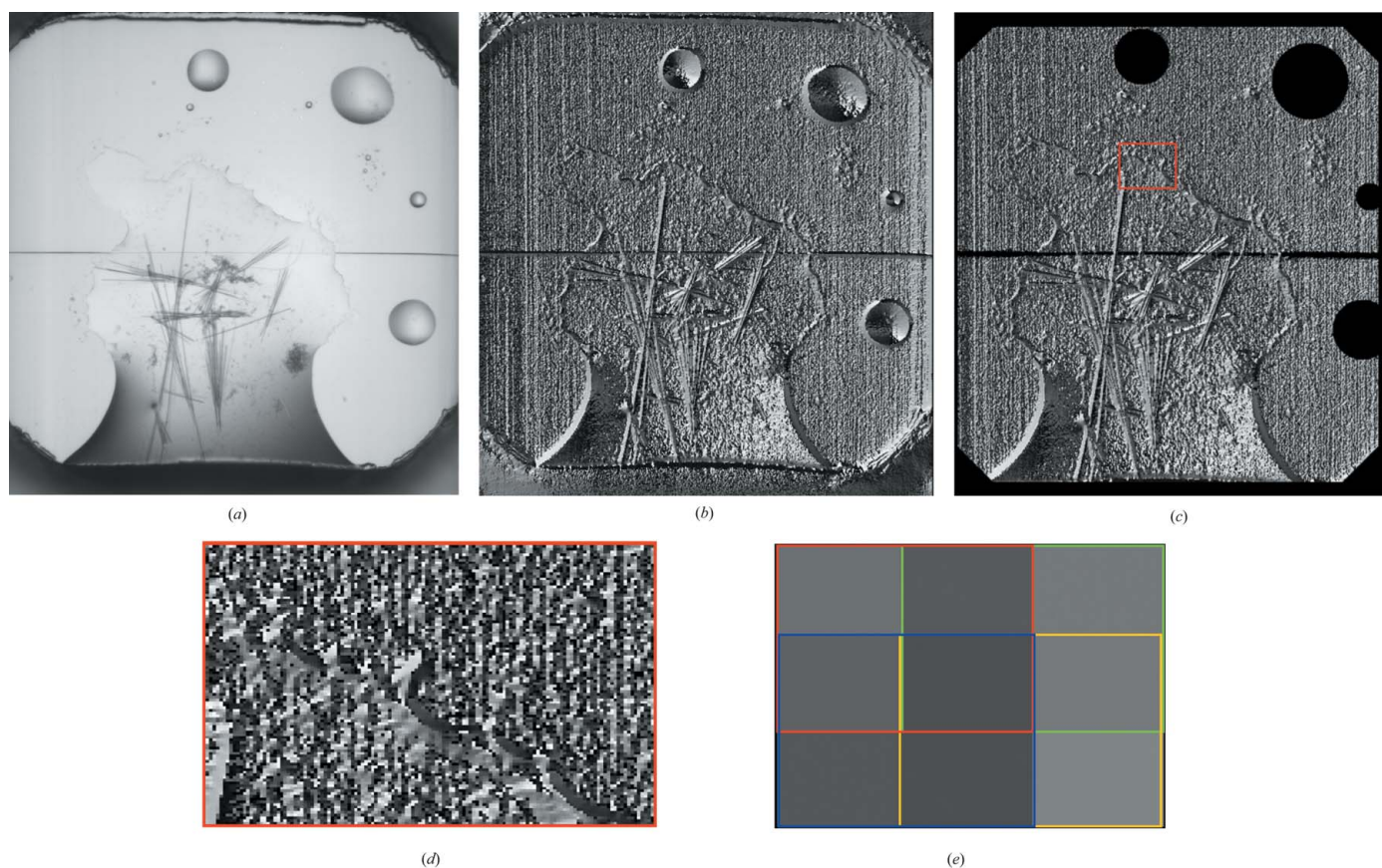


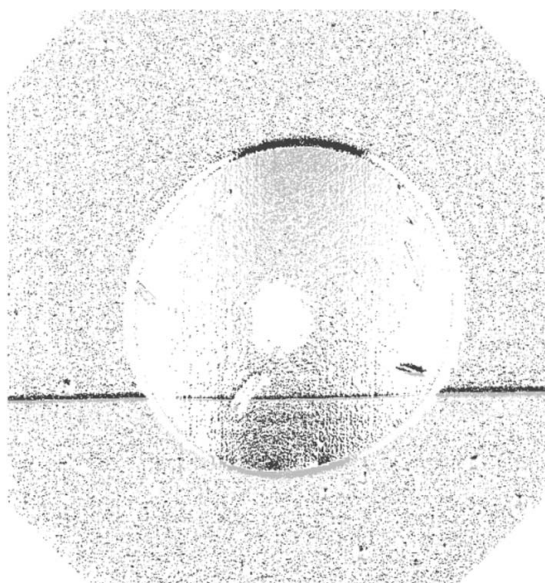
Figure 4

The gradient directions for the image in (a) are plotted in (b). In (c), a mask for the well has been applied and the small droplets and line in the plastic tray have been removed. The area indicated by the red rectangle is enlarged in (d). In (e), the possible choices of four pixels to average over, including the central pixel, are outlined in red, green, blue and yellow.

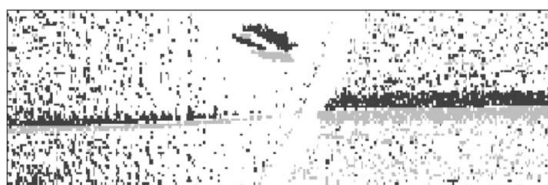
pronounced in drops containing PEG, for example, which allows the drops to hold their shape rather than spread out. Even when the line appears to be straight, the width of the line has an effect. The close-up view in Fig. 5(b) shows that there may be pixels ‘missing’ and these must be allowed for when identifying this artefact. The risk here is that of eliminating the edges of crystals that happen to lie horizontally or vertically across the drop, and for this reason only lines that are deemed ‘long enough’ are masked. Whilst long thin crystals can be expected to lie within the drop, the lines due to the crystallization trays usually extend beyond the crystallization drop and this fact can be exploited. However, as with all artefacts to be eliminated or attributes to be utilized for classification, there are always images displaying counter-examples. It cannot be assumed that the lines reach the edges of the well as this is not the case in many images. However, an initial search is performed independently from both sides of the image for horizontal lines and from the top and bottom of the image for vertical lines. If found, a line is then followed across the image with some deviation and missing segments allowed for in passing through the drop. If no line is found from the edges of the image, then a search is conducted further inside. Small disconnected sections will not be masked in order to reduce the risk of eliminating crystals; such sections rely on the classification algorithms for their identification.

6. Removing artefacts due to experimental setup

Often robotic dispensing creates small droplets as well as the main crystallization drop, as can be seen in Fig. 6. These make determination of a mask for the crystallization drop more difficult and can lead to background areas being analysed unnecessarily. In addition to the droplets, the speed of dispensing can give rise to air bubbles within the drops. This creates objects that provide no information about the experiment but, as with the line across the image, not only waste time in analysis but can also give rise to false positives in classification as lighting effects can make these objects look very interesting in terms of the classification variables. However, an obvious property that can be exploited is their circular nature and bubbles and droplets of a reasonable size can both be identified from this. The gradient direction is used for identifying these objects from the concentric circles of smoothly changing gradient angles. Firstly, a new binary image is created using any pixels lying in a straight line consisting of seven (this may vary with resolution) pixels of very similar gradient direction. Fig. 6 shows that this allows the larger bubbles (as well as the edge of the crystallization drop) to be located easily, but that there could be confusion between crystals and smaller bubbles. For this reason, only bubbles and droplets with a radius



(a)



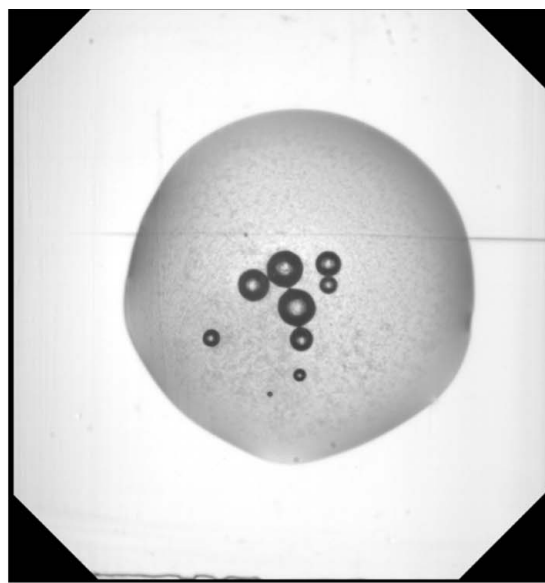
(b)

Figure 5
In (a) all pixels for which the gradient angle is close to 90° are shown in light grey and those close to 270° are shown in dark grey. In (b) a close-up view of the right-hand side close to the horizontal line is shown.

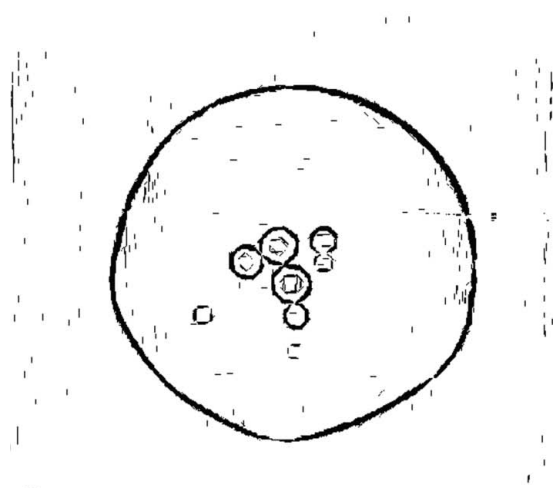
between pre-set minimum and maximum values are deleted prior to analysis.

7. Masking the crystallization drop

Previously, a circular mask was generated for the crystallization drop (Wilson, 2002). This is obviously unsuitable for drops that have been very disturbed and can create problems, even when the drops appear to be roughly circular. For the image in Fig. 7, a circular mask can be found that eliminates most of the background without losing much of the crystallization drop. However, even in this case where the best circular fit has been found, the edges of the drop can still cause problems. In the analysis, each connected set of pixels is considered as an object to be evaluated individually (Wilson, 2002). The objects obtained in this case are indicated by the grey pixels in Fig. 7(c), some of which are due to the edge of the crystallization drop. These objects often look more interesting than they really are, as can be seen initially from their shapes. The straight edges, as well as sharp changes in intensity due to light and shadow, are characteristics used to identify crystals. As well as creating objects that are likely to give rise to false positives in classification, there is always the danger of losing information with a badly fitting mask. Cumbaa *et al.*



(a)



(b)

Figure 6
An image in which air bubbles have occurred is shown in (a). In (b) those pixels lying in a straight line with six other pixels of very similar gradient angle are plotted in black. This shows that the bubbles as well as the edge of the crystallization drop are recognizable in the gradient direction.

(2003) use a probabilistic model to segment the well into three regions on a coarse grid. Each coarse pixel is assigned to be empty well (W), inside the drop (D) or on the edge of the drop (E). They then exclude both well pixels and edge pixels from the analysis at the risk of discarding crystals growing at the edge of the drop.

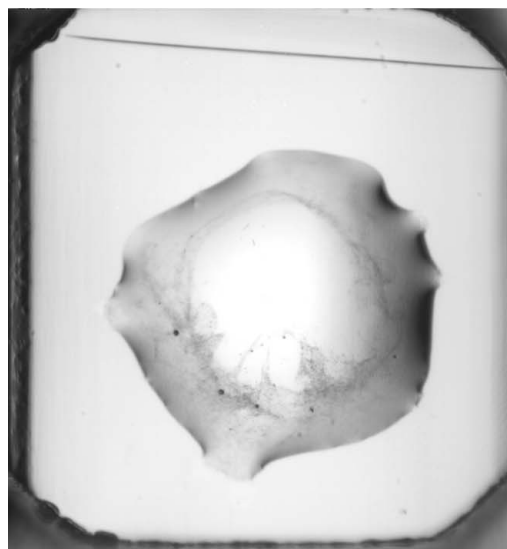
Here we also use a coarse grid (with each coarse pixel covering 20×20 of the original pixels) to provide an initial mask for the crystallization drop, but we use the gradient direction to determine the status of each coarse pixel. Fig. 4 shows an image in which the edges of the drop are not clear. As there is little change in the greyscale, the magnitude of the gradient around the edge of the drop will also be small.

However, the direction of the steepest gradient should still be perpendicular to the edge of the drop and, as seen in Fig. 4(b), the edge of the drop can be identified by the smooth change in gradient angle. The small rectangle indicated in Fig. 4(c) is enlarged in (d) and shows that, while the variation in angle is less along the drop boundary, comparing any pixel with all eight of its neighbours would not show this. We therefore compute the variation over four pixels rather than nine. As Fig. 4(e) shows, this means that, for any pixel, there are four possible choices for the four pixels. In fact we consider all four choices and take the smallest variation as the value for any particular pixel. Fig. 8(a) shows the map obtained from these values. For each coarse pixel the number of original pixels (out of a possible 400) with a standard deviation less than a pre-set threshold are counted. Taking into account the status of neighbouring pixels, each coarse pixel is assigned to be in the mask or in the drop. As Fig. 4(c) shows, the horizontal line and bigger droplets have been eliminated, as described in the previous sections, before attempting to identify the limits of the crystallization drop. Fig. 8(b) shows the edge pixels of the coarse mask in grey overlaid on the object pixels (black) for the image in Fig. 4. We do not eliminate these 'edge pixels' but do some further refinement of the mask within them. That is, the mask is increased until either an object pixel or the inner edge of the coarse pixel is reached.

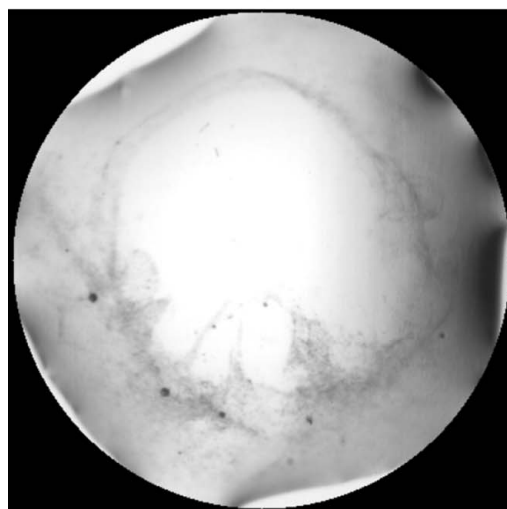
8. Conclusions

The Oxford Protein Production Facility has produced over fourteen million images and is now regularly generating in excess of seventy-five thousand images per day. These images must be accessible to crystallographers and this creates a huge storage problem. Methods of image compression are being considered, but have an adverse effect on classification (Berry *et al.*, 2004). Reducing the image to the size of the well can decrease the number of pixels by up to two-thirds with no loss of useful information. Identification of the crystallization drop allows the image to be cropped further. Whilst storing only this reduced image would certainly help address the problem, suitable compression is still needed for long-term storage. Cropping the image early in the image processing would also speed up the automated analysis. With structural genomics centres like the OPPF capable of producing an image every 2 s, it is vital that the on-line analysis of these images is computationally efficient. The removal of any artefacts not related to the experimental outcome prior to analysis facilitates the generation of a suitable mask for the crystallization drop. This also prevents valuable time being wasted on the evaluation of uninteresting objects and reduces the risk of false positive results in classification.

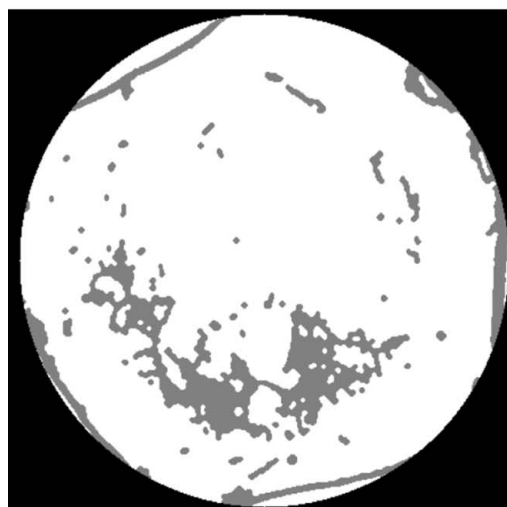
JW is a Royal Society University Research Fellow and IB is supported by the Wellcome Trust (grant H5RCZR0). Most images shown were supplied by the Oxford Protein Production Facility in Oxford. The OPPF is funded by the UK Medical Research Council. The authors would also like to



(a)



(b)



(c)

Figure 7

The circular mask for the crystallization drop in (a) is shown in (b). The object pixels for this image are shown in grey in (c). It can be seen how objects arising from the edge of the crystallization drop could cause problems in classification.

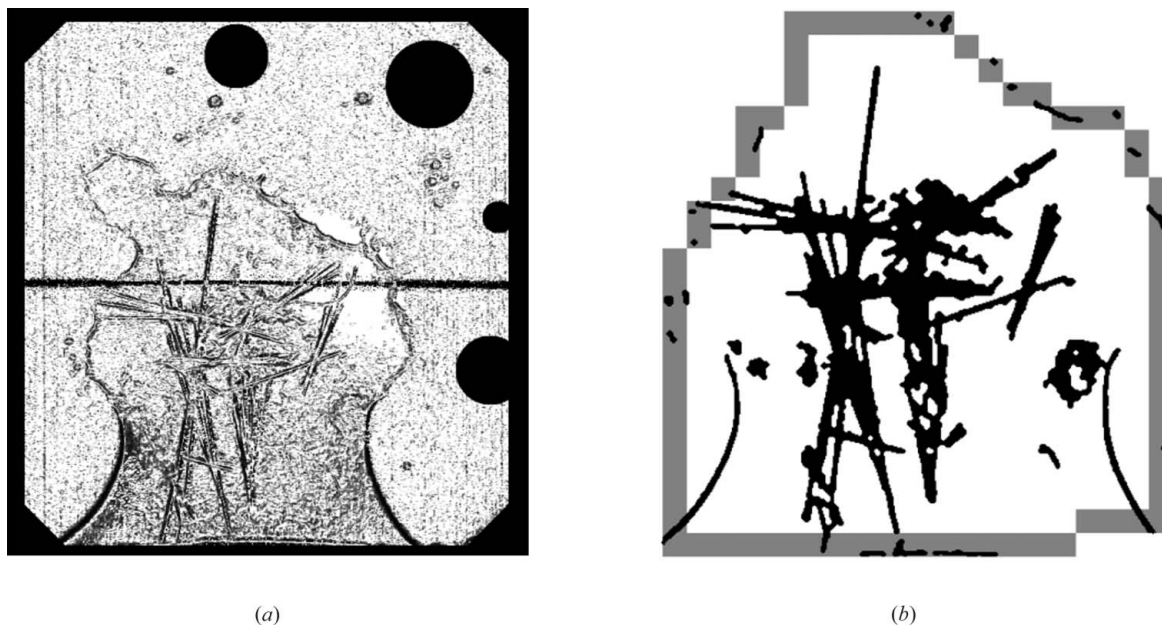


Figure 8
The variation in gradient angle taken over four pixel blocks is shown in (a) with darker pixels indicating lower values. In (b) the edge pixels for the coarse mask obtained from (a) are shown in grey with the object pixels in black.

thank Tony Fordham-Skelton and Miroslav Papiz at the Synchrotron Radiation Source in Daresbury, Anastassis Perrakis at the Nationaal Kanker Instituut in Amsterdam, and Professor Udo Heinemann at the Protein Structure Factory in Berlin, for supplying additional images.

References

Bern, M., Goldberg, D., Stevens, R. C. & Kuhn, P. (2004). *J. Appl. Cryst.* **37**, 279–287.
Berry, I., Wilson, J., Diprose, J., Fuller, S. & Esnouf, R. (2004). *Int. J. Neural Syst.* In the press.

Canny, J. (1986). *IEEE Trans. Pattern Anal. Machine Intell.* **8**, 679–698.
Cumbaa, C. A., Lauricella, A., Fehrman, N., Veatch, C., Collins, R., Luft, J., DeTitta, G. & Juristica, I. (2003). *Acta Cryst.* **D59**, 1619–1627.
Gonzalez, R. & Woods, R. (2002). *Digital Image Processing*. Second Edition. Prentice Hall.
Hough, P. V. C. (1962). *A Method and Means of Recognizing Complex Patterns*. US Patent 3069 654.
Spraggon, G., Lesley, S. A., Kreusch, A. & Priestle, J. P. (2002). *Acta Cryst.* **D58**, 1915–1923.
Wilson, V. (2002). *Acta Cryst.* **D58**, 1907–1914.
Wilson, V. (2004). *Crystallogr. Rev.* **10**, 73–84.