Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Automated detection and centring of cryocooled protein crystals

Sudhir Babu Pothineni, Tilo Strutz and Victor S. Lamzin*

European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

Correspondence e-mail: victor@embl-hamburg.de

Received 7 April 2006
Accepted 11 August 2006

A novel method is presented for the automated recognition of cryocooled macromolecular crystals. The method uses several texture-based image-processing algorithms for automated crystal centring, which are able to cope with a variety of crystal morphologies and illumination conditions. The results combined from different algorithms, together with their estimated standard uncertainties, provide a robust determination of the crystal location and allow an internal assessment of their reliability. The method was coded within the software *XREC* and showed a good performance on 104 sets of images from various beamlines.

1. Introduction

With increasing use of modern robotics equipment, macromolecular crystal structure determination now requires an automated software and hardware pipeline from biological samples to atomic models which can function in a high-throughput fashion. At many steps along such a pipeline, the recognition of crystalline samples is essential. Examples include monitoring crystal growth, scoring of crystallization conditions or crystal centring for data collection.

Since the early 1990s, the vast majority of X-ray experiments have been carried out on dedicated synchrotron beamlines, capitalizing on their superior properties, *e.g.* beam intensity, collimation and energy tuneability (Minor *et al.*, 2000). The diffraction data collection is the last experimental step in the whole process of structure determination and there have been impressive developments in the recent past that have allowed the replacement of many manual tasks by automated alternatives (Abola *et al.*, 2000; Jacquamet *et al.*, 2004; Snell *et al.*, 2004). Nevertheless, we are still too far from having a beamline that is able to provide truly automated high-throughput data collection. User interaction is generally required for sample mounting, centring and setting up the experiments. This not only precludes efficient usage of synchrotron beam time, but also influences the success of experiments owing to the human factor.

In this paper, we address the automation of crystal centring in X-ray crystallographic experiments, which should not only speed up the data collection, but primarily assure the reproducibility of experiments and allow automated screening of the samples taken from a carousel by a sample-changer robot.

A number of approaches have been proposed for the recognition and topological analysis of the loop and its content

in visual light (*e.g.* Karain *et al.*, 2002; Andrey *et al.*, 2004), infrared (Snell *et al.*, 2005) and ultraviolet illumination (Jacquamet *et al.*, 2004; Vernede *et al.*, 2006; Forsythe *et al.*, 2006). An interesting approach published by Karain *et al.* (2002) uses X-ray fluorescence from a crystalline sample containing anomalous scatterers. Andrey *et al.* (2004) showed that the use of a high-resolution zoomable camera for visualization combined with back-lighting through a condenser illuminated by optical fibre leads to good results for crystal centring. The usage of two cameras at the same time, a low-magnification camera to track the rotation of the loop and a high-magnification microscope for visualization of the crystal (*e.g.* BL38B1 at SPring8 and F2-Station at MacCHESS), seems attractive, but its wide applicability has probably still to be investigated. The image-processing algorithms proposed so far are relatively simple and robust, but assume special conditions (lighting, field of view *etc.*) that may be specific for particular hardware installations. A setup for semi-automated crystal centring, based on a user clicking a mouse or fingering a touch-screen to indicate the crystal location through a specially designed software interface, is employed at several synchrotron beamlines (Snell *et al.*, 2004).

Here, we present a novel approach for automatic crystal detection and centring that uses texture analysis, which is targeted towards direct identification of a cryocooled crystal rather than the loop or surrounding liquid. Because of the large variety in size and morphology of crystals, illumination and other hardware specifics that affect the visual appearance of macromolecular crystalline samples (Fig. 1), a number of different algorithms have been designed. In order to determine the location of the crystal in three-dimensional space, several images taken at different crystal orientations are processed simultaneously. The results from different algorithms are weighted dynamically. This increases the accuracy of the estimated coordinates of the crystal centre and, more importantly, serves as an internal validation criterion indicating the reliability of the result. All this brings the developed method to a qualitatively more advanced level and allows its use in a black-box manner.

2. Methods

Four different algorithms followed by a statistical combination of the results have been developed. Each algorithm is applied to images of the crystal in different orientations with respect to the camera. Prior to the application of the individual algorithms, the input images are checked for the presence of the loop containing the crystal and, in addition, the image background is modified if needed.

2.1. Preliminary image analysis

Dependent on the orientation of the loop (Figs. 1*a*–1*d*), the image is rotated so that the loop (alone or together with its stem and the pin) enters from the bottom of the picture. It may be the case that an input image $I(x, y)$ does not contain a loop with a crystal. This can be caused, for example, by inaccuracy of the sample-changing equipment, a too long or too short length of the pin, too high a magnification of the camera or the complete absence of the loop. An intermediate case could be that the loop is present within the image window only for certain angular positions of the goniometer system. In order to identify an empty image, its content is modelled using a two-dimensional second-order parabolic function. Such a parabolic fit is usually satisfactory for background regions or a featureless image, while for an image containing an object (with a loop and a crystal) the deviations from the fitted parabola are significantly high (Fig. 2). In the non-empty images the modelled background is subtracted. The crystal-detection algorithms are targeted towards the high-contrast regions in the image. Thus, the presence of a pin may have adverse effects. Therefore, the region containing the pin is detected and its content is removed. The pin has an almost

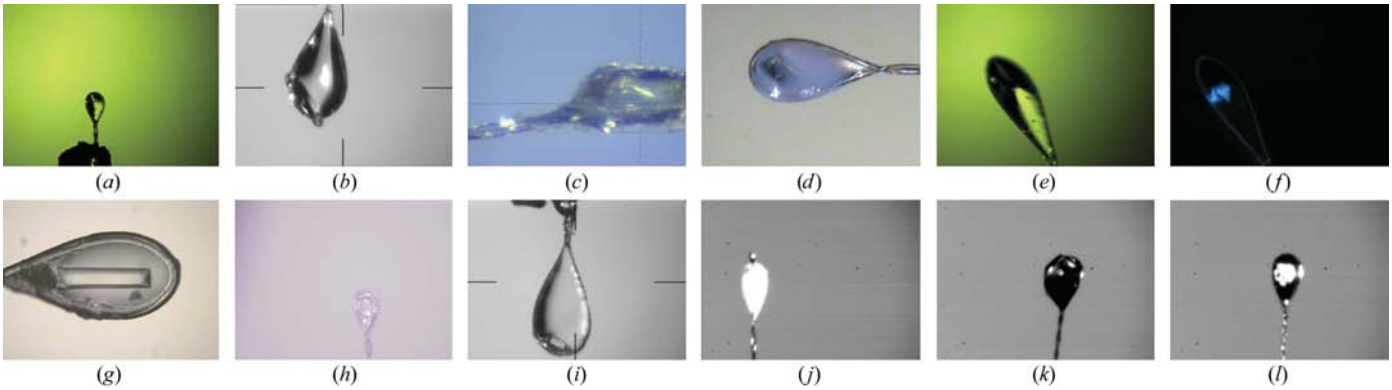

Figure 1
The large variety in the visual appearance of crystals mounted in loops. Orientation of the loop: from the bottom (*a*), top (*b*), left (*c*) and right (*d*); colour (*a*, *c*, *d*, *e*, *f*, *g*, *h*) and greyscale (*b*, *i*, *j*, *k*, *l*) images; use of ultraviolet (*f*) and visual (all other) light; back-lighting (*g*), front-lighting (*d*, *f*, *j*, *k*, *l*) and front light together with a reflection from the back (*a*, *b*, *c*, *e*, *h*, *i*); high (*b*, *c*, *d*, *g*, *i*) and low (*a*, *e*, *f*, *h*, *j*, *k*, *l*) magnification; reflections (*b*, *c*) and occlusions (*e*, *j*, *k*); special markers (*b*, *i*) on the images; low (*h*) and high (*a*, *b*, *e*, *g*) contrast of the picture; dark (*f*, *j*) and bright (all others) background; crystal appearing in dark (*b*, *d*, *e*, *j*, *k*, *l*), light (*c*, *f*) or transparent (*a*, *g*, *h*, *i*) colour; irregular (*c*) or well shaped (*b*, *d*, *g*, *i*) crystals; images containing a part of the pin (*a*); small crystals in large loops (*a*, *d*, *i*) and large crystals extending beyond the loop (*b*, *j*).

constant width, which is exploited in a row-by-row examination of the image.

Cases of an empty image or a pin crossing almost the entire image are reported to the beamline control. These images are not considered further.

2.2. Molecular replacement in crystal imaging (MoRCI)

MoRCI resembles a two-dimensional analogue of the molecular-replacement technique in X-ray crystallography. If the shape of a crystal is approximated by an ellipsoid, one rotational and two translational parameters have to be determined to locate it in the image. In our first implementation we constrained the lengths of the axes of the ellipsoid to be the same, so that the crystal in the loop can be considered to be spherical. This simplification results in the fact that the rotational component of the search vanishes. The translational parameters are found through sliding of a circular object with radius r through the image using the scoring function

$$\sigma^2(x_0, y_0) = \frac{1}{N}\sum_{(x,y)\in A} I_{\rm E}^2(x, y) - \left[\frac{1}{N}\sum_{(x,y)\in A} I_{\rm E}(x, y)\right]^2, \quad (1)$$

where (x_0, y_0) are the coordinates of the centre of the circular object, A is its area and N is the number of pixels inside A. $I_{\rm E}(x, y)$ refers to an equalized image, which is obtained by histogram matching of the content of $I(x, y)$ to a uniform distribution on a scale from 0 to 255. The translation search in the above formulation is equivalent to an evaluation of a local variance, analogous to the technique for identification of protein regions in an electron-density map (*e.g.* Terwilliger, 1999; Cowtan, 2001). The search for the highest local variance has been implemented in reciprocal space.

The resulting 'textured' image, for a given radius r of the circular object, is

$$I_{\rm T}(x, y|r) = \sigma(x, y). \quad (2)$$

We note that, given a reasonably large value of r, the value of the local variance $\sigma^2(x, y)$ is (i) zero when the greyscale of the points inside the mask is alike (*e.g.* all points are black), (ii) equal to $255^2/12$ for uniformly distributed pixel intensities inside the mask and (iii) is at its maximum of $3 \times (255^2/12)$ for the highest contrast (*e.g.* when half of the points are black and the other half are white). Image $I_{\rm T}$ is further transformed into a binary image $I_{\rm B}$, which in effect selects the areas with the most contrastive texture,

$$I_{\rm B} = \begin{cases} 1 & \text{if } I_{\rm T}(x, y) > 255/12^{1/2} \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

The optimum radius r of the circular object is determined as follows. The largest object in binary image $I_{\rm B}$ is chosen and the ratio of its area to the area of the circular search object is computed. For small values of the radius this ratio is high and it reduces as the radius increases. The optimum radius is taken when this ratio is about 1.0 (Fig. 3). The centroid of the largest object in image $I_{\rm B}$ is taken as the crystal centre. For radii that are either smaller or larger than the optimum, the results become biased towards the centre of the loop. The successive steps of MoRCI are depicted in Fig. 4.

2.3. Fuzzy categorization of image histogram (FuCHi)

In general, the distribution of the pixel intensities and, in particular, the texture type are different for the image background and the area containing a crystal. Histograms of typical images are dominated by the background (Fig. 5*a*). Since a crystal can be considered as a spatially localized collection of bright and dark points, the histogram for its area would primarily contain a contribution from dark (low grey values) and bright (high grey values) pixels. FuCHi upweights (categorizes) the pixels with those grey levels, which are at both ends of the equalized image histogram (Fig. 5*b*). The categorization is achieved using an empirical sigmoidal activation function

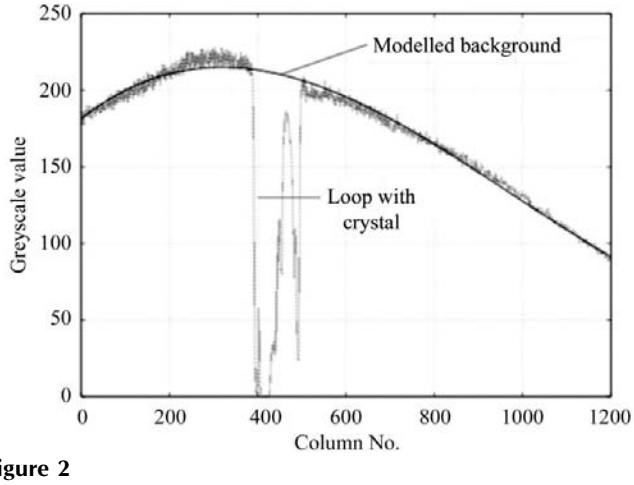

Figure 2
Modelling image background: an example of one row of an image containing a loop object. The greyscale values of the pixel intensities in the row are shown together with the modelled background.

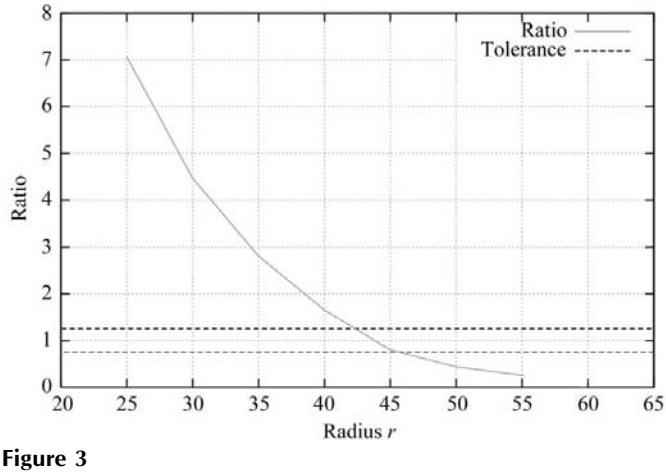

Figure 3
Search for the optimum value of r of the circular object within the MoRCI algorithm. See text for details.

$$I_A(x, y) = \frac{1}{2}\{1 + \tanh[a \cdot I_E(x, y) + b]\}. \quad (4)$$

The parameters a and b defining the shape of the function (Fig. 5*c*) are set to empirical values on the basis of the library of available images. Dependent on the histogram of the input image, either its dark, light or both parts are activated.

The resulting image I_A thus contains a fuzzy likelihood of pixels belonging to the crystal area. It is then further smoothed. In order to avoid selection of objects other than the crystal (*e.g.* the remainder of the mother liquor around the crystal or the loop), the centroid of 0.2% of points with the highest values of likelihood is taken as the crystal centre. Although the above assumptions do not entirely hold in general, for large crystals FuCHi produces satisfactory results.

2.4. Two-dimensional features (2DF)

Human vision is amazing in its cognitive capability to give different attention to the different patterns of the input picture. The features that our eyes concentrate on first are in computer vision called 'information points' (other terms are 'points of interest', 'corners' or 'two-dimensional features'; Harris & Stephens, 1988; Shi & Tomasi, 1994; Smith & Brady, 1997). Here, we apply a pattern-recognition technique based on the use of 2DF, which is similar to what Shi & Tomasi (1994) have suggested for the tracking of moving objects.

Within a square window around each point in image $I(x, y)$ the characteristic of the texture is determined using the gradients in the x and y direction, from which the matrix $\mathbf{G}$ is computed,

$$g_{x_i} = \frac{\partial I(x_i, y_i)}{\partial x},$$

$$g_{y_i} = \frac{\partial I(x_i, y_i)}{\partial y},$$

$$\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix},$$

where

$$g_{11} = \sum_{i=1}^{N} g_{x_i} g_{x_i}, \quad g_{12} = g_{21} = \sum_{i=1}^{N} g_{x_i} g_{y_i}, \quad g_{22} = \sum_{i=1}^{N} g_{y_i} g_{y_i}.$$

The summation is carried out over the total number of points inside the window, N. The size of the window is set to 5% of the estimated value of the loop width (described in §2.5 below).

The eigenvalues ($\lambda_1 > \lambda_2$) of $\mathbf{G}$ serve as good descriptors of the local texture. Both eigenvalues are low for a flat profile of the image region; both are high for a two-dimensional texture (*e.g.* a corner) and $\lambda_1 >> \lambda_2$ for a unidirectional texture pattern (*e.g.* an edge). The cases of edges and particularly the corners are of interest for crystal detection. For these, the value of λ_2 should be sufficiently high. 200 regions with the top values of λ_2 are selected and their centres are taken as 2DF points. The minimum distance between the 2DF points is set to the size of the window.

The most likely centre of the crystal is subsequently determined from the coordinates of 2DF points. The distribution of the 2DF points on the image (Fig. 6*a*) is smoothed with the Parzen-window technique using Gaussians as kernel functions (Fig. 6*b*). Applying an empirical threshold to the resulting image gives different objects, where the largest one is presumed to represent the crystal mask. In cases when there are no reflections from the incoming light, the mask represents the size and shape of the crystal well (Fig. 6*c*).

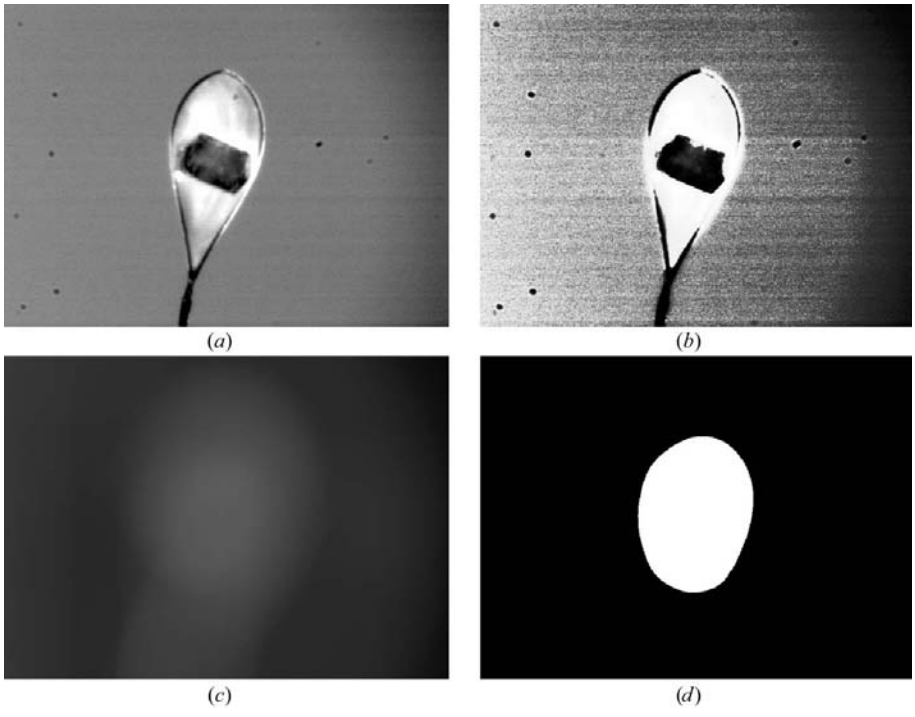

Figure 4
Image-processing steps in MoRCI: (*a*) the original image; (*b*) equalized image I_E; (*c*) textured image I_T; (*d*) binary image I_B.

2.5. Model-based automated crystal detection (MACyD)

The MACyD algorithm classifies the image content so that the points least likely to belong to the area of the crystal are successively removed. The removal is based on the assumption that the crystal is located more or less inside the loop, which may have a visible stem connected to a pin. The points that are left after the removal are analysed and assigned a score. The point with the highest score is taken as the crystal centre.

Rather than processing the entire image $I(x, y)$, MACyD operates with its subsampled versions $I_M(x, y)$ at different offset positions. The minimum

dimension of a subsampled image is chosen to be approximately 120 pixels. The required subsampling factor M (the ratio between the dimensions of the original and the subsample images) determines the number of subsampled images: $2 \times M - 1$ subsampled images are taken out of the total of $M \times M$.

For each subsampled image $I_M(x, y)$ the two first-derivative images ($\partial I/\partial x$ and $\partial I/\partial y$) and three second-derivative images ($\partial^2 I/\partial x^2$, $\partial^2 I/\partial y^2$ and $\partial^2 I/\partial x \partial y$) are computed. On the basis of these derivatives, the edges and the lines from the subsampled image are enhanced and low gradients and single peaks are suppressed, yielding a contrast image $I_C(x, y)$ (Fig. 7*a*). This compensates for the diversity of the input images (see Fig. 1) and provides a more general applicability for the recognition algorithm described below.

Since the input image is rotated when necessary so that the loop enters from the bottom of the picture (see §2.1), the rows of contrast images $I_C(x, y)$ are examined and in each row the horizontal width of the object is computed (Fig. 7*b*). The shape of this curve provides an indication where the pin and the loop stem are located and the content of the corresponding rows is deleted. The remaining rows constitute the loop object (with the crystal inside it). The width of the loop is determined from the widths of these rows and is subsequently used for estimation of the crystal size. In addition, the estimated boundaries of the loop object are removed. This reduces the contrast contribution from the loop wire, the thickness of which is assumed to be proportional to the loop width.

The loop region (after removal of its boundaries) in contrast image $I_C(x, y)$ (Fig. 7*c*) is smoothed by convolution with a Gaussian kernel. The full width of the Gaussian kernel is set to one third of the size of the loop object, which should be comparable to the average crystal size. The resulting global maximum is taken as the crystal centre (Fig. 7*d*). A use of a wider Gaussian kernel would bias the solution towards the centre of the loop object, while a convolution with too sharp Gaussians would pick up individual texture features rather than the centre of the crystalline object.

The set of estimated coordinates of the crystal centre derived from all subsampled images is checked for possible outliers, which are removed. The remaining coordinates are averaged.

3. Implementation

3.1. Combination of the results of the four detection algorithms

The problem of the visualization of the crystal and the determination of its coordinates is not only defined by the great diversity of the image content (Fig. 1), but is additionally complicated by the fact that a three-dimensional object is projected onto a two-dimensional picture. When the crystal is rotated on a goniostat axis, its projection to the camera view (and therefore the image) is changing. Therefore, each of the four algorithms described above is applied to a series of

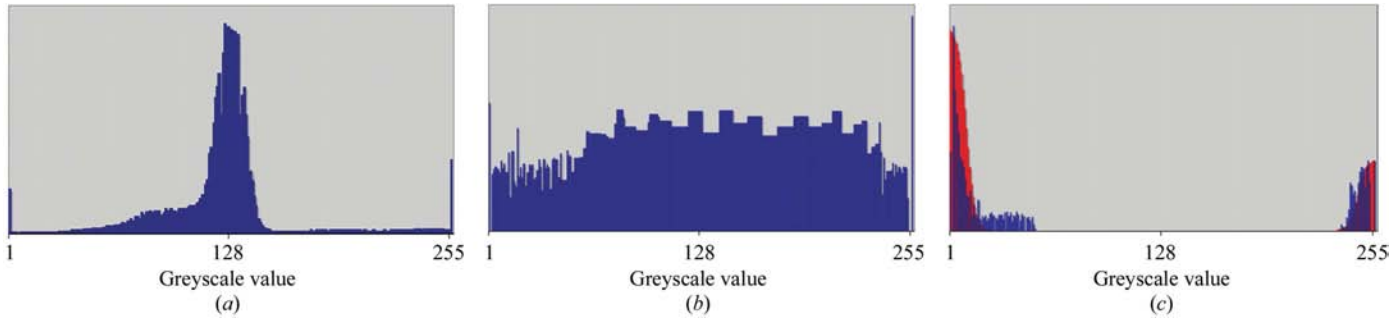

Figure 5
An example histogram (*a*) of an original image; (*b*) after histogram equalization; (*c*) blue, the region containing the crystal; red, the activation function used by the FuCHi algorithm.

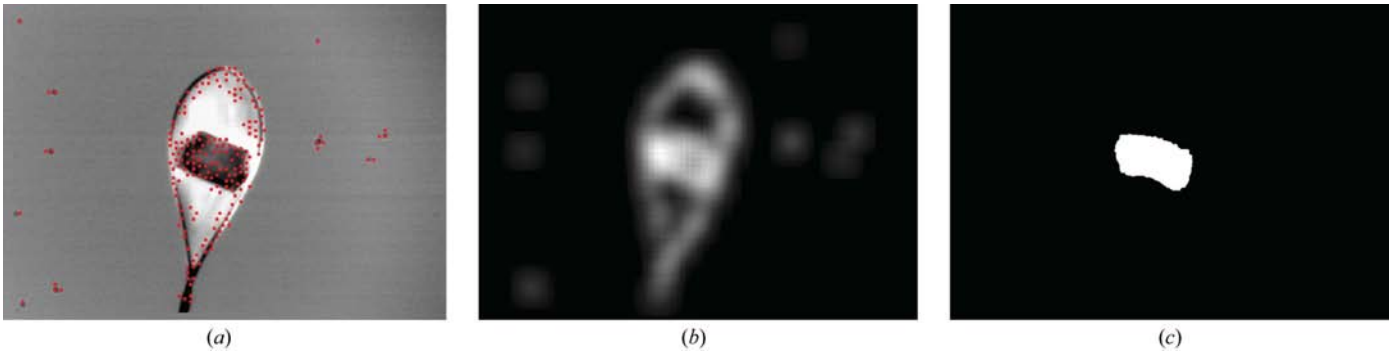

Figure 6
Use of two-dimensional features for the location of the crystal: (*a*) top 200 2DFs; (*b*) image after smoothing with Parzen-window kernel functions; (*c*) resultant binary image after selection of the largest object.

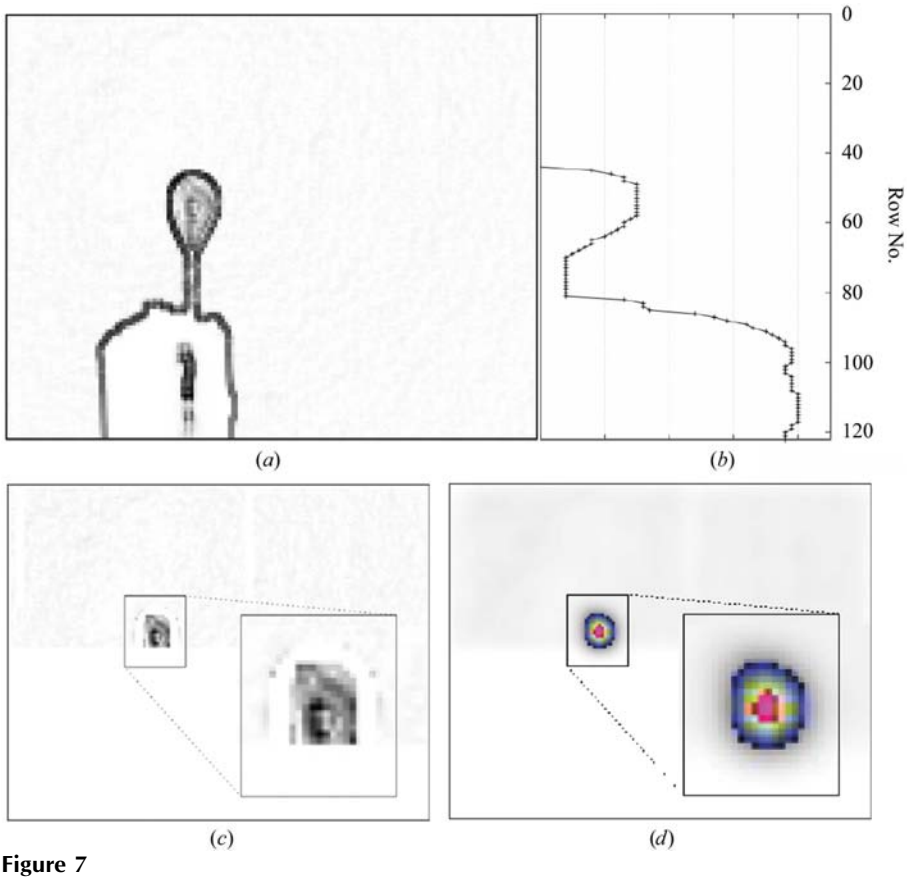

Figure 7
Image processing with the MACyD algorithm. (*a*) An example of a contrast image $I_C(x, y)$: the darker the point, the higher the contrast is. (*b*) The horizontal object width (in pixels) plotted for each row of the image. (*c*) Remaining pixels in image $I_C(x, y)$ after elimination of pin, stem and the contribution of the loop wire. (*d*) Smoothed version of the contrast image with the pixels scored for their likelihood to be the crystal centre, displayed in colours (red corresponds to the highest score). An enlarged picture is also shown in (*c*) and (*d*).

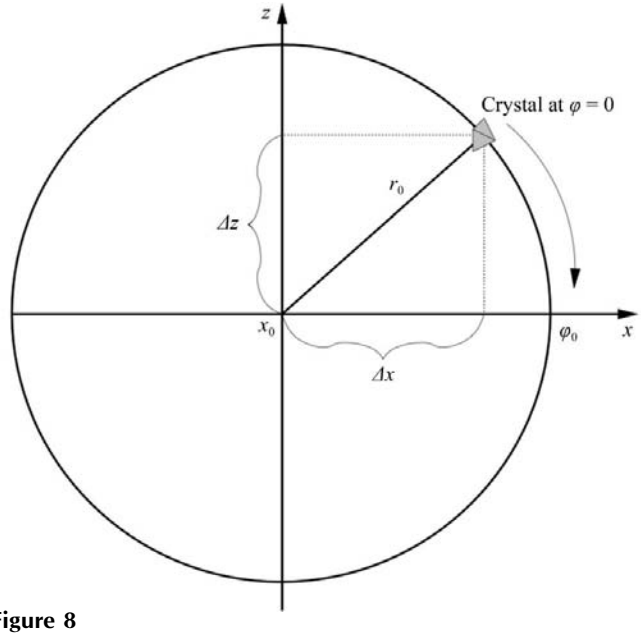

Figure 8
Modelling of the motion of the crystal with the camera view being orthogonal to the rotation axis. The rotation axis is along the *y* axis, whereas the camera view is along the *z* axis. See text for details.

images with different orientations of the crystal with respect to the camera. This results in L two-dimensional coordinates of the crystal centre ($L_{max} = 4N$), where N is the number of images in the series. A full angular range of 360° would give the best results, but 270° proved to be sufficient in almost all test cases. For typically six to nine images spanning the rotation range, the task of finding the crystal centre becomes highly overdetermined and the random errors affecting the results from individual images are largely alleviated. When the crystal is rotated on the beamline goniostat, its movement describes a circle in the plane perpendicular to the rotation axis (Fig. 8).

In the camera view the crystal moves along the x direction, while its y coordinate is independent of the spindle rotation. The motion of the crystal in the image plane can be described by

$$f(\varphi) = x_0 + r_0 \cos(\varphi - \varphi_0), \quad (6)$$

where r_0 and x_0 are the radius and the centre of the circle and φ_0 is the offset to the goniometer angular system. Setting $p_0 = r_0\cos(\varphi_0) = \Delta x$ and $q_0 = r_0\sin(\varphi_0) = \Delta z$ leads to a linear function with respect to the parameters x_0, p_0 and q_0,

$$f(\varphi) = x_0 + p_0 \cos(\varphi) + q_0 \sin(\varphi). \quad (7)$$

The parameters x_0, p_0 and q_0 are determined by an iteratively re-weighted least-squares minimization (Holland & Welsch, 1977) of the residual

$$\sum_{i=1}^{L} w_i \cdot [f(\varphi_i) - x_i]^2, \quad (8)$$

where the summation is carried out over results from all four algorithms, w_i is the weight for each observation and x_i are the observations provided by the four algorithms for each single image.

Given an initial set of equal weights w_i, the least-squares solution of (8) provides values of the refined parameters x_0, p_0 and q_0. The new weights are then estimated on the basis of the deviates Δ_i between the observations x_i and the modelled function $f(\varphi_i)$,

$$\tilde{w}_i = \begin{cases} 1/\sigma_x^2 & \text{if } \Delta_i^2 \le \sigma_x^2 \\ 1/\Delta_i^2 & \text{if } \sigma_x^2 < \Delta_i^2 \le \kappa \\ 0 & \text{if } \Delta_i^2 > \kappa \end{cases}, \quad (9)$$

where σ_x^2 is the weighted average of the squared deviates

$$\sigma_x^2 = \sum_{i=1}^{L} w_i \cdot [f(\varphi_i) - x_i]^2 = \sum_{i=1}^{L} w_i \cdot \Delta_i^2. \quad (10)$$

For convenience, the weights are normalized so that their sum is equal to 1,

$$w_i = \tilde{w}_i \Big/ \sum_{i=1}^{L} \tilde{w}_i \qquad (11)$$

The basic idea behind this is to downweight observations that are far from the modelled function and to upweight those that are close to the model using $\tilde{w}_i = 1/\Delta_i^2$. When the deviations from the modelled function are too small, $\Delta_i^2 \leq \sigma_x^2$, the observations are given equal weights. This avoids assignment of unjustifiably high weights. Observations that are too far away from the approximated function, $\Delta_i^2 > \kappa = 9\sigma_x^2$, are considered outliers and assigned zero weights. However, as soon as the first outlier is detected, this threshold is kept fixed for all subsequent iterations in order to avoid consecutive rejection of too many observations. The re-weighted least-squares procedure is iterated until convergence.

The observed values of y_i are treated in a similar manner but with their own weights,

$$\dot{y} = \sum_{i=1}^{L} w_i^{(y)} \cdot y_i. \qquad (12)$$

3.2. Uncertainty of coordinates

In order to estimate the reliability of the determined coordinates of the crystal centre, $(\dot{x}, \dot{y}, \dot{z})$, their estimated standard uncertainty, $\sigma_{\rm tot}$, is computed as follows. The estimated variance in the y direction is

$$\sigma_{\dot{y}}^2 = \frac{1}{L-1} \sum_i w_i^{(y)} \cdot (y_i - \dot{y})^2. \qquad (13)$$

The estimated variances $\sigma_{x_0}^2$, $\sigma_{p_0}^2$ and $\sigma_{q_0}^2$ for the refined parameters in (7) are the diagonal elements of the covariance matrix **C**,

$$\mathbf{C} = \frac{\chi^2}{L_{\rm free}} \cdot (\mathbf{J}^{\rm T} \cdot \mathbf{W} \cdot \mathbf{J})^{-1}, \qquad (14)$$

where $\chi^2/L_{\rm free}$ is the goodness of fit,

$$\frac{\chi^2}{L_{\rm free}} = \frac{\sum_{i=1}^{L} w_i \cdot [f(\varphi_i) - x_i]^2}{L-3}. \qquad (15)$$

W is the diagonal matrix of weights w_i and **J** is the Jacobian matrix containing the first derivatives of the model function (7) with respect to the parameters. Since the uncertainty of the final position in the plane of crystal movement is only a function of p_0 and q_0, which are orthogonal to each other and to the y axis, elementary derivation leads to

$$\sigma_{\rm tot} = (\sigma_{\dot{y}}^2 + \sigma_{p_0}^2 + \sigma_{q_0}^2)^{1/2}. \qquad (16)$$

3.3. Further statistical evaluation

In addition to the value of $\sigma_{\rm tot}$, it is of interest to estimate the probability that, after the crystal has been centred using the described method, the X-ray beam would hit the crystal throughout the data collection. To avoid discussion of the relative orientation of the crystal with respect to the beam, the crystal is assumed to be spherical with diameter d. We guess the crystal size using the estimated loop width (§2.5). The apparent width of the loop varies as the sample is rotated and its minimum value is taken as the lower estimate of the crystal size. This corresponds to the worst-case scenario for crystals with highly non-spherical shape, *e.g.* plates or needles. In addition, it is also assumed that the error in crystal centring follows an isotropic three-dimensional normal distribution centred on the values $(\dot{x}, \dot{y}, \dot{z})$ and having a standard deviation equal to the value of $\sigma_{\rm tot}$. Let us now assume that the centre of the beam has been aligned to the determined centre of the crystal $(\dot{x}, \dot{y}, \dot{z})$. The probability of such a beam centre hitting the crystal (hereafter called the estimated reliability of the method) is equal to the volume of the three-dimensional normal distribution that is covered by the crystal volume,

$$P = \mathrm{erf}\left(\frac{d}{2 \times 2^{1/2} \cdot \sigma_{\rm tot}}\right) - \frac{d}{\sigma_{\rm tot}(2\pi)^{1/2}} \cdot \exp\left(-\frac{d^2}{8 \cdot \sigma_{\rm tot}^2}\right). \qquad (17)$$

4. Results and discussion

The proposed approach was tested on 104 sets, mostly containing eight or nine images with 40° increments. These images were compiled from different beamlines worldwide (EMBL Hamburg, SRS Daresbury, ESRF Grenoble, SSRL Stanford, ALS Berkeley and APS Argonne) and are a part of the growing BIOXHIT annotated image database (Fig. 9). The crystals were mostly illuminated with visual light, but there were four cases with UV light. It seems that the use of UV light provides better visualization of fluorescent crystals and the success rate of the correct determination of their centre is higher.

Fig. 10 shows the comparison between the estimated reliability of the calculated coordinates (17) and the actual situation judged from visual inspection. If the value of estimated reliability was above 90%, the method indicated successful centring, otherwise not. 77 cases (red colour in Fig. 10) were correctly centred on the crystal. In 13 cases (green) the automated centring pointed towards the edge of (but was still within) the crystal. In 14 remaining cases (yellow and black) the centring failed. Indeed, the failure cases correspond to either very small crystals in large loops, poor illumination conditions or the presence of ice. Although not straightforwardly, they could nevertheless be centred manually. The success or failure of centring was appropriately indicated by the estimated reliability: there were only four false negatives and five false positives. Thus, the overall success rate of the presented method could be taken as $(77 + 13 - 4 - 5)/104 = 78\%$. Increasing this success rate will be the subject of further research.

The method was designed to be of general applicability regardless of the imaging system and lighting used. At the

same time, one may want to know the capability of the method given an 'optimal' lighting schema. Each beamline may obviously have its own optimum conditions and out of 104 sets of images we have subjectively taken 26 sets from beamline X12 at the EMBL Hamburg equipped with a MAR Research desktop beamline (examples of such images are shown in Figs. 1*b* and 1*i*). The success rate for these was 81%, practically the same as for the whole set of 104 crystals.

The success rate of the crystal centring depends on many factors and some of them are discussed below.

4.1. Success rate of different algorithms

MoRCI generally produces good results for crystals whose size is comparable to the size of the loop. However, it also works for crystals that are as small as the thickness of the loop wire, provided that there are no strong light reflections from surrounding liquid. The main assumption of FuCHi is that the crystal is seen as a collection of bright and/or dark pixel intensities. Therefore, it shows lower success rate for images with back lighting when crystals appear transparent. The 2DF algorithm also gives good results for crystals that are smaller than the loop (Figs. 1*b*, 1*d* and 1*i*). However, for large transparent crystals (Fig. 1*g*) the derived coordinates may be biased towards the centre of the loop. This may particularly occur at high magnification levels. MACyD models the loop–stem–pin object based on the contrast between pixel intensities. Apart from cases with misleading reflections at the surface of the surrounding liquid, it yields good detection results. Only MoRCI explicitly assumes that the crystalline object is spherical. However, the other algorithms also use the steps of image smoothing, thresholding and centroid determination. Therefore, they all perform better for solid crystals and poorer for thin needles.

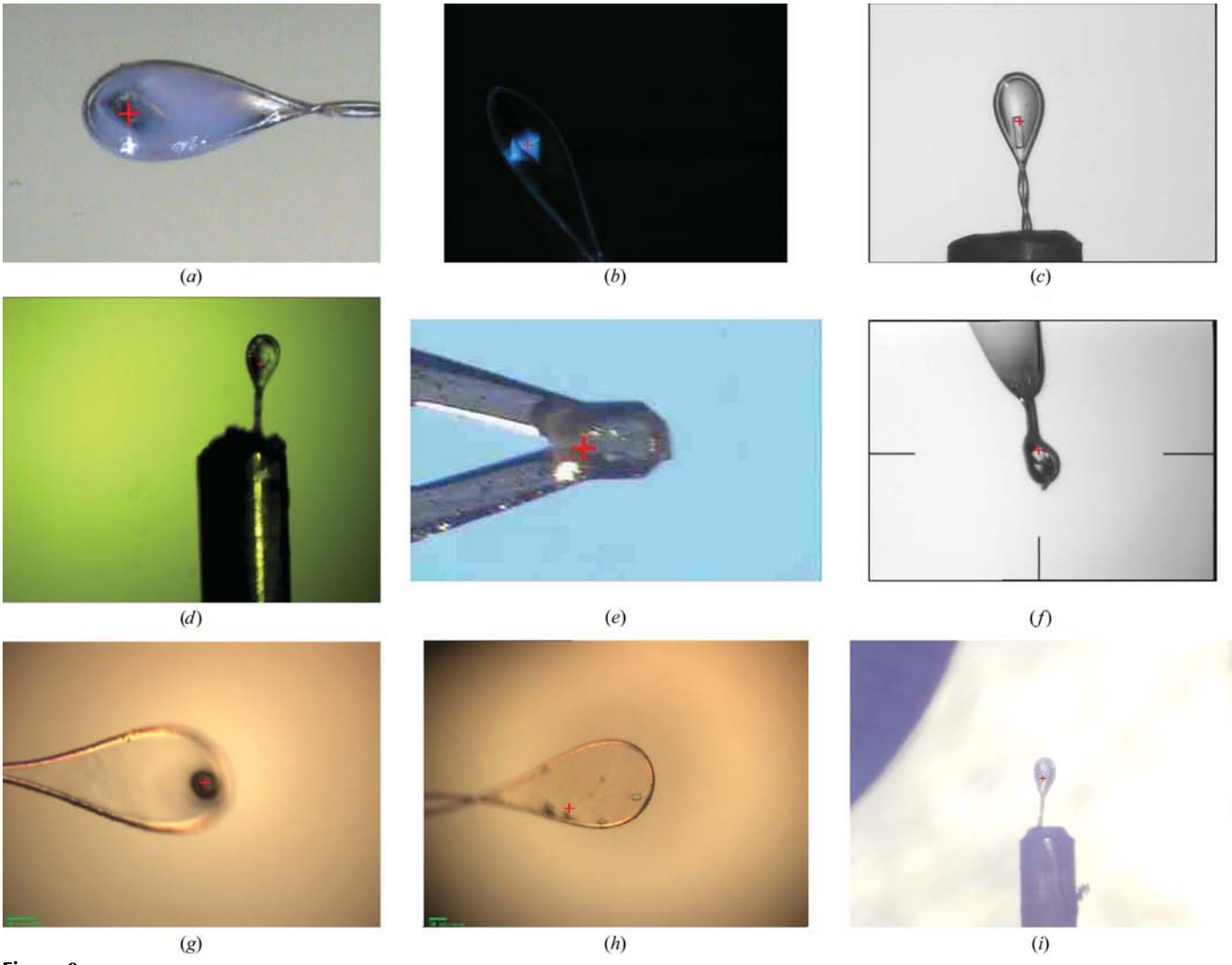

Figure 9
A gallery of crystals mounted in different loops. The results of crystal centring are shown by a red cross. (*a*) A well shaped crystal (courtesy of Cristine Trame, ALS, Berkeley); (*b*) a crystal in UV light (EMBL, Hamburg); (*c*) a small needle-like crystal centred on the edge (courtesy of Gleb Bourenkov, MPG, Hamburg); (*d*) a low-magnification image (EMBL, Hamburg); (*e*) MicroMounts (courtesy of Ana Gonzales, SSRL, Stanford); (*f*) LithoLoops (EMBL, Hamburg); (*g*) a blob-like crystal (courtesy of Bernard Lavault, EMBL, Grenoble); (*h*) a very small crystal, centred on an artifact (courtesy of Bernard Lavault); (*i*) a cryo shadow into the image (courtesy of Mylrajan Muthusamy, SRS, Daresbury),

Each of the four single algorithms may fail to find the crystal centre in some images, *e.g.* owing to an unfortunate projection, specific lighting effect or shadows. Also, the algorithms are based on different assumptions and not all of them can be valid for all cases. However, such failures are efficiently compensated by the weighted least-squares combination of the algorithms. Fig. 11 shows an example for the curve fitting of (6) for a data set containing nine images with 40° increments. Four algorithms produced 36 raw results. The majority of them correspond well to the fitted curve. However, in two cases FuCHi deviates considerably from the modelled function owing to non-compensated background effects. In one case algorithm 2DF also produced an outlier owing to an unfortunate combination of the loop orientation and occlusion. The contribution from these three cases was automatically down-weighted.

Detailed evaluation of each algorithm's performance will be the subject of future investigations and the derived information is expected to be helpful for 'knowledge-based' image interpretation.

4.2. Validation of the error estimates

Equation (16) provides an error estimate on the position of the identified crystal. Although the dynamic adjustment of weights is expected to compensate for various effects in processing two-dimensional projections, it may be thought that both the accuracy of the result and the error estimate could be particularly affected by the orientation of the loop with respect to the image projection. We have therefore carried out a specially designed experiment to address this issue.

For two crystals, one well and one poorly visible by eye (Fig. 12), 36 images with an angular increment of 10° have been taken. The crystal centre was determined 36 times using different series of seven images with 40° increment (0, 40, 80, 120, 160, 200, 240°; 10, 50, 90, 130, 170, 210, 250°; *etc.*). The locations of the determined crystal centre agree well with each other within an r.m.s. deviation of nine pixels for one crystal (15 µm with the used image magnification) and 13 pixels (25 µm) for the other. These values are also consistent with the estimated accuracy of crystal location ($\sigma_{\rm tot}$ in equation 16).

4.3. Type of loop

In most of the test cases the crystals were mounted using Hampton loops. In recent years, new types of loops have become available displaying new properties (*e.g.* less X-ray absorption and scattering, better holding of the crystal). For example, MicroMounts and LithoLoops have different geometries compared with conventional Hampton loops. Although the *XREC* software was not specifically trained to deal with these loops, the results are rather promising (Figs. 9*e* and 9*f*). Four out of six crystals in LithoLoops could be centred. We have not extensively tested the automated centring with MicroMounts. On one hand, one would expect that owing to the fixed position of the crystal relative to the loop, the centring should be more straightforward compared with other cases. However, it has also been reported that in practice the condition of a fixed crystal position relative to the stem is often not met (Lavault *et al.*, 2006).

4.4. Size and morphology of the crystal

The success rate of the proposed method is directly related to the size and the shape of the crystal. For crystals that are very small compared with the loop, the determination of the crystal centre becomes more difficult since there is a relatively higher fraction of non-crystalline objects that may attract the detection algorithms. In two cases only the loop could be centred and, accordingly, the reported reliability was low. Fig. 9(*h*) shows an example of unsuccessful centring on such small crystals. For plate-like crystals and thin needles a better model for the estimation of the 'reliability' (17) is obviously required. In such cases the method can give a false signal of low 'reliability' even if a centre of the crystal has been

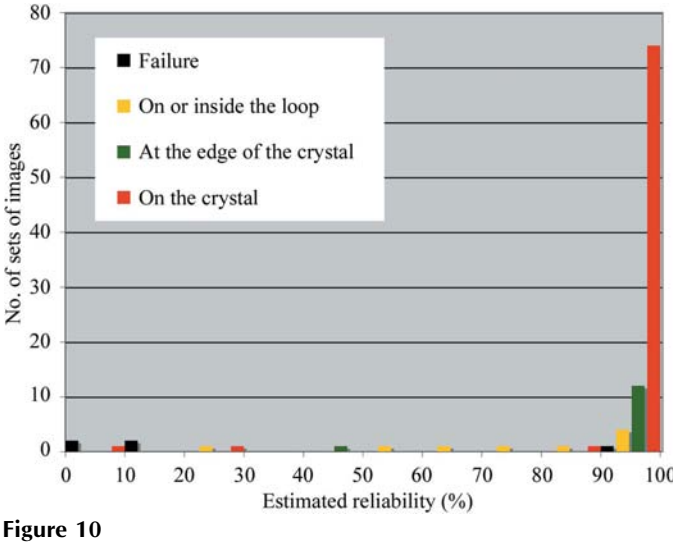

Figure 10
Success rate of crystal detection as evaluated on 104 data sets with a combined use of four algorithms. Cases with an estimated reliability of above 90% would be deemed to result in successful centring.

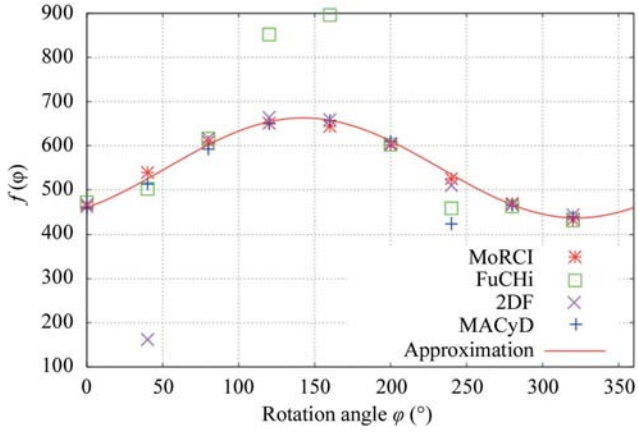

Figure 11
Example of processing of the results from four algorithms using (6) and (8).

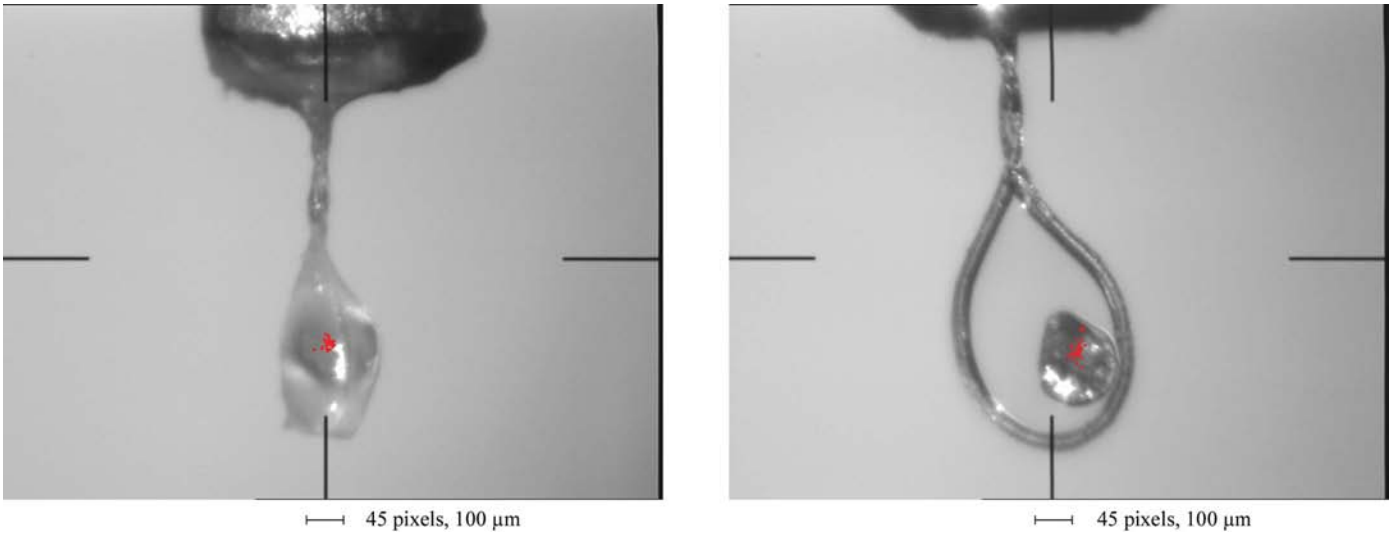

Figure 12
Dependence of the determined crystal centre on the relative orientation of the loop. For each crystal, 36 images with an angular increment of 10° have been taken. The crystal centre (marked by red dots) was determined 36 times using different series of seven images with 40° increment. The scale of the images, both in pixels and μm, is shown for reference.

successfully determined. At the same time, the shape of such crystals can be exploited for an enhancement of the weighting scheme (8).

Reflections from the surrounding liquid, occlusions from loop wire or icing effects can also mislead the crystal recognition. For transparent crystals the additional use of a line-detection algorithm as implemented in *C3D* (Lavault *et al.*, 2006) may be advantageous.

If there is more than one crystal present in the image, the method tries to centre on the largest one. However, in the case of equally sized closely located crystals the determined position may be in between, thus leading to a false positive.

There are a number of hardware-related factors that may adversely affect crystal centring. One is the magnification of the camera (zooming). If it is too high, so that the crystal covers a large area of the image, the proposed method may pick a point in the crystal which is not exactly at its centre. If, on the other hand, the magnification were too low, so that the crystal (or even a loop) could hardly be seen, the method would also be unable to find it. In addition, other objects (*e.g.* a cryocooling device or its shadow) may appear in the image as an additional complication. A rule of thumb is that the centring is successful if the crystal occupies 0.2–15% of the image area (*i.e.* the size of the crystal is about 4–40% of the dimensions of the image). Focusing may also be an important factor requiring further studies. The out-of-focus regions might be less attractive for the described texture-search algorithms, although in all three (out of 104) test sets with poor focus the crystals were centred correctly.

5. Conclusion

The presented algorithms have been implemented in a standalone software module *XREC*, which is available from http://www.embl-hamburg.de/XREC/ for Mac OSX, Linux and Windows platforms. Support for other platforms would be straightforward to add. The software has been written in C/C++, is reasonably fast and CPU requirements for a typical image are around 2 s on a modern workstation. We envision that the highest benefit from the proposed method would be achieved when it is combined with automated mounting of the crystals from a carousel and with software (*e.g. BEST* from Popov & Bourenkov, 2003) that would allow the estimation of crystal diffraction properties before the X-ray experiment commences. It could also be advantageous to make combined use of the method with other software for crystal centring, *e.g.* the *C3D* software from EMBL Grenoble (Lavault *et al.*, 2006). Last but not least, *XREC* has been developed with the aim that its underlying algorithms could also be applicable for automated classification of crystallization drops. This will be a focus of future research.

We thank EMBL staff and colleagues from other synchrotron beamlines and home laboratories for provision with the crystals and/or sets of images. We also would like to thank Uwe Ristau for integrating the recognition software into the BW7B beamline at EMBL-DESY, as well as Matthew Groves and Andrew Leslie for fruitful discussions and comments on this manuscript. This work was supported by the BIOXHIT project, EC FP6 contract No. LSHG-CT-2003-503420.

References

Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. (2000). *Nature Struct. Biol.* **7**, 973–977.
Andrey, P., Lavault, B., Cipriani, F. & Maurin, Y. (2004). *J. Appl. Cryst.* **37**, 265–269.
Cowtan, K. (2001). *Acta Cryst.* D**57**, 1435–1444.
Forsythe, E., Aniruddha, A. & Pusey, M. L. (2006). *Acta Cryst.* D**62**, 339–346.

Harris, C. & Stephens, N. (1988). *Proceedings of The Fourth Alvey Vision Conference*, pp. 147–151. Sheffield: University of Sheffield Printing Unit.
Holland, P. W. & Welsch, R. E. (1977). *Commun. Stat. Theor. Methods A*, **6**, 813–827.
Jacquamet, L., Ohana, J., Joly, J., Legrand, P., Kahn, P., Borel, F., Pirocchi, M., Charrault, P., Carpentier, P. & Ferrer, J.-L. (2004). *Acta Cryst.* D**60**, 888–894.
Karain, W. I., Bourenkov, G. P., Blume, H. & Bartunik, H. D. (2002). *Acta Cryst.* D**58**, 1519– 1522.
Lavault, B., Ravelli, R. B. G. & Cipriani, F. (2006). *Acta Cryst.* D**62**, 1348–1357.
Minor, W., Tomchick, D. & Otwinowski, Z. (2000). *Structure*, **8**, R105–R110.
Popov, A. & Bourenkov, G. P. (2003). *Acta Cryst.* D**59**, 1145–1153.
Shi, J. & Tomasi, C. (1994). *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR94)*, pp. 593–600. Piscataway, NJ, USA: IEEE.
Snell, E. H., van der Woerd, M. J., Miller, M. D. & Deacon, A. M. (2005). *J. Appl. Cryst.* **38**, 69–77.
Snell, G., Cork, C., Nordmeyer, R., Cornell, E., Meigs, G., Yegian, D., Jaklevic, J., Jin, J., Stevens, R. C. & Earnest, T. (2004). *Structure*, **12**, 537–545.
Smith, S. M. & Brady, J. M. (1997). *Int. J. Comput. Vis.* **23**, 45–78.
Terwilliger, T. C. (1999). *Acta Cryst.* D**55**, 1174–1178.
Vernede, X., Lavault, B., Ohana, J., Nurizzo, D., Joly, J., Jacquamet, L., Felisaz, F., Cipriani, F. & Bourgeois, D. (2006). *Acta Cryst.* D**62**, 253–261.