Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

C3D: a program for the automated centring of cryocooled crystals

Bernard Lavault,* Raimond B. G. Ravelli and Florent Cipriani

Instrumentation, EMBL, 10 Rue Jules Horowitz, Grenoble 38000, France

Correspondence e-mail: lavault@embl.fr

Received 29 March 2006
Accepted 8 August 2006

A method is presented to automatically locate a crystal and its holder for centring on a goniometer spindle and alignment with an X-ray beam. Here, a novel algorithm that has been developed and tested with the images of users' crystals saved in an annotated database is described. The algorithm improves on the difficult situations that are commonly observed and poorly handled by the first-generation crystal-centring algorithms. These include highly transparent crystals, bad cryocooling or lens effects arising from the geometry of the drop. Most crystals have polyhedral shapes and a number of straight edges, which yield useful information. In this method, crystal detection relies on a feature-scoring system in which line extraction has the highest weight. Here, the image processing and calculations implemented in the program *C3D* are described. This program is designed to operate with a client program that controls specific diffractometer hardware. In order to select the best detection conditions, *C3D* provides various functionalities adapted to various hardware configurations.

1. Introduction

Automatic crystal centring is a key step towards fully automated macromolecular crystallography (MX) experiments (Abola *et al.*, 2000). These experiments will allow the collection of data from a large number of crystals without user intervention. This requires reliable detection of the crystal and its holder, with criteria to evaluate the success of the centring. Several strategies can be used: mounting the crystal in a known position, a beam size encompassing the whole loop, software calculations of the loop or crystal position from sample images, detection of the crystal using X-rays or any combination of these techniques.

Ueno *et al.* (2004) demonstrated a sample holder designed at SPring-8 that can be mounted and removed in a highly reproducible manner. This allows the storage of reliable crystal positions. A beamline operator is still required to manually mark the positions, which are then used for overnight data collections. This is unlikely to work with European SPINE standard sample holders (http://www.spineurope.org), which lack a reference point for repositioning. Thorne *et al.* (2003) expect that microfabricated mounts made from polyimide film (MicroMounts) will facilitate the automation of crystal centring because a crystal of known dimensions can be accurately matched with the mount size. However, according to our observations, it is a delicate operation to place the crystal in the positioning hole while harvesting with MicroMounts. Other solutions to simplify crystal detection include the removal of the surrounding solvent, thus enhancing the contrast between crystal and support. This operation, conducted by Kitago *et al.* (2005), has already been tested for the purpose of improving data-collection quality, but the aspiration of the cryo-buffer through a micropipette has not yet been automated.

The diversity in size and morphology of protein crystals requires sophisticated detection algorithms, especially for the methods based on images taken with visible light. The solvent surrounding the crystal can form a lens that modifies the apparent crystal position, depending on the visualization angle. Methods using only loop recognition (Abad-Zapatero, 2005) require a careful matching of crystal, loop and beam size; otherwise, they must be complemented by other techniques for accurate centring. Karain *et al.* (2002) used automated loop centring, together with X-ray fluorescence or X-ray diffraction, for crystal detection. However, as S/C/O *K* fluorescence energies are too low to be measured in air, additional anomalous scatterers are required in the sample if X-ray fluorescence is to be used. Furthermore, the concentration of the anomalous scatterer in the cryosolution must be sufficiently low in order to obtain a good contrast between crystal and solution. Note that the use of X-ray fluorescence requires an energy-dispersive detector. For crystal-alignment methods based on X-ray diffraction, CCD readout time and X-ray damage are critical issues (Weik *et al.*, 2000; Ravelli & McSweeney, 2000).

Roth *et al.* (2002) developed a method in which the centres of mass of the brightest and darkest areas in the loop are determined separately and the results are combined to obtain an estimate of the position of the crystal. This method is simple and very fast. This generally works when the crystal is located in the middle of the loop, but we believe that the contribution of the mother liquor to the centre of mass is not negligible, limiting the success rate of actually centring inside the crystal.

Crystal detection using UV-induced fluorescence has recently been explored. The contrast between solvent and crystal is usually higher than in visible light. Another advantage of UV fluorescence is that the technique can still be used in the presence of ice. Pohl *et al.* (2004) used a non-monochromatic UV source for fluorescence-based crystal centring. A size-filtering procedure was applied to remove the nylon-loop fluorescence. The centre of mass of the remaining object gives the crystal position. This method works when there is a single crystal in the loop that is larger than the diameter of the nylon loop. UV-induced fluorescence-based setups have been pushed further with 266 nm laser sources: with standard visible-light optics and polyimide mounts by Vernede *et al.* (2006) or by Judge *et al.* (2005). These methods based on fluorescence should be fast and reliable, but are still not routinely used. Particular attention should be paid to crystals that are sensitive to radiation damage (Nanao & Ravelli, 2006) and to safety issues linked to UV-light sources.

Crystal detection in crystallization drops has both similarities to and differences from the detection of crystals in cryocooled loops. The automation of crystallization platforms requires software that predicts the probability of the presence of one or several crystals in a drop, but the automation of data collection requires the three-dimensional position in the crystal for illumination with the X-ray beam. The centring algorithm must handle loop-orientation changes and defocusing conditions. On the other hand, the automatic centring algorithm can use simplifying hypotheses such as 'there is only one crystal'. Nevertheless, work on crystal recognition in drops as described by Wilson (2002, 2004), Wilson & Berry (2005), Spraggon *et al.* (2002), Cumbaa *et al.* (2003), Bern *et al.* (2004) and Saitoh *et al.* (2005) could certainly have some positive impact on crystal-centring algorithms. For instance, classifiers used to discriminate between crystal and artifacts, such as the number of straight lines, are valuable for detection both in crystallization drops and in cryoloops.

In this issue, both *XREC* (Pothineni *et al.*, 2006) and *C3D* automatic centring methods are described. The *C3D* program described here has been trained from the start using only high-quality crystal images provided by a microdiffractometer (Perrakis *et al.*, 1999). Our work started in 2002 with the development of an algorithm (*Crystal Trace*; Andrey *et al.*, 2004) using the correlation between all the images of a sample rotated on the spindle axis. Extensive use of the algorithm on the beamline showed that difficult conditions such as icing, lens effects and poor contrast occur more often (Table 1) than initially anticipated. For example, the inspection of 881 images of crystals brought by users to the ESRF shows icing artifacts on 10% of the crystals. A new algorithm based on another principle has been written, with the aim of being more robust in these difficult situations and improving the precision of the crystal centre determination. The idea is to give a maximum weight to what humans use to detect a crystal at first glance: crystal edges and their geometric aspects. The implementation uses Canny edge-detection (Canny, 1986) and a scoring system based mainly on the Radon transform (Kak & Slaney, 1988). Additional classifiers complete the scoring of irregular objects. The program, integrated with an auto-centring module on MD2 microdiffractometers, has been extensively tested for more than 2 y, mainly on ID14-3 and BM14 at the ESRF. This development period has allowed us to collect many images and to compile statistics on our results. Automatic centring

Table 1
Statistics of user samples in nylon loops show the frequency of issues for crystal detection using visible light.

Users' crystal images have been saved before or after data collection on three diffractometers at ESRF. The diffractometers at BM14, ID14-EH3 and ID14-EH4 gave 881 images of different crystals during 186 d. 'Sparse icing' indicates the presence of ice artifacts on the drop surface. 'Bad freezing' leads to a non-vitreous drop; hence, the crystal is hidden or the contrast inside the loop is reduced. The 'lens effect' arises from the drop geometry. It occurs with backlit large drops and creates an opaque ring that potentially hides part of the crystal. 'Invisible crystal' refers to a condition where the crystal is not distinguishable by an expert eye on the majority of the images taken during rotation.

Issue	% of total
Sparse icing	10.0
Bad freezing	4.5
Invisible crystal	2.3
Lens effect	14.4
Multiple crystals	1.6

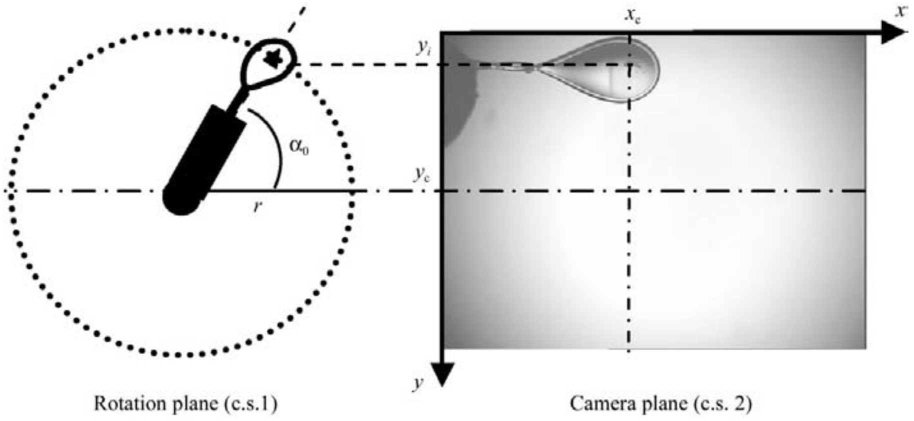


Figure 1
The viewing plane is perpendicular to the rotation plane. The three-dimensional position of the crystal corresponding to the first image is given by r and α_0 in the rotation plane (left). To align the crystal, two coordinates are given in the image plane (right): x_c is the average abscissa of the crystal and y_c is the rotation-axis ordinate and will be the crystal position after centring.

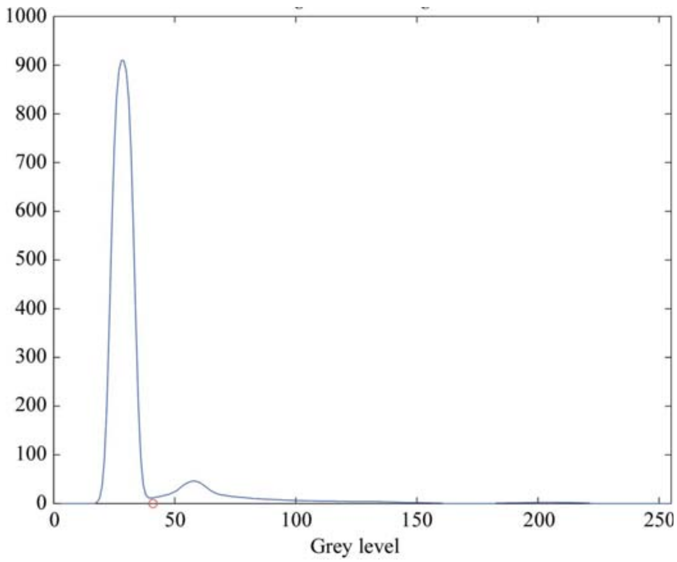


Figure 2
Histogram of pixel intensities in one greyscale working image once the original image has been subtracted from the background (inverted image). The first peak corresponds to the background pixels and the second (on the right) to the crystal. The circle shows the threshold.

scripts using *C3D* are now being installed at all the ESRF MX beamlines on MD2 and MD2M diffractometers. *C3D* comes with a set of features specific to the hardware of the MD2*x* diffractometers, but can also support other setups (§4.2). *C3D* is continuously being evaluated and optimized using a growing annotated database. It has been successfully installed and tested at a number of synchrotrons (CCLRC, ESRF, MAX-lab, NSLS, SLS). Several fully automated screening experiments have been carried out successfully at the ESRF using *C3D*.

2. Principles of crystal centring

Automatic crystal centring consists of several steps depending on the characteristics of the optical system used to collect the images: field of view, depth of field, camera resolution and lighting. The procedure on MD2*x* diffractometers starts by a loop search and a translation of the spindle to place the loop at a nominal viewing position. This is followed by a loop-centring routine, an optical zoom adjustment and a loop reorientation to collect the best images necessary to detect the crystal. The final centring consists of moving the crystal onto the rotation axis and translating the whole crystal in the X-ray beam. When the goniometer axis rotates, the sample describes a circle of radius r (Fig. 1). The centring of both the loop and the crystal rely on the same principle: the frame grabber takes a series of images at different spindle angles and their image analysis returns the position of the target. Each image corresponds to a 'viewing angle' (α_i) between the rotating object and the optical axis of the camera (Fig. 1). The method works independently of the angles, but a careful choice can improve the success rate of crystal detection. For instance, during loop centring the angles are equally distributed in order to calculate the loop orientation. This allows an optimal start angle for subsequent crystal centring to be determined. The centring script takes two types of images: two images with the loop face perpendicular to the visualization plane and a set of images with the loop parallel to the visualization plane (chosen for optimal crystal visibility). The loop or crystal three-dimensional position (r, α_0, x_c, y_c), as presented in §4.3, is obtained by minimizing the cost function (Andrey *et al.*, 2004)

$$f = \sigma_y^2 = \sum_{i=1}^{n} [y_i - y_c - r \cdot \sin(\alpha + \alpha_0)]^2. \tag{1}$$

σ_y represents the standard deviation of the distance between the rotation circle and the positions y_i of the crystal or loop tip as detected on each image; y_c is the rotation-axis ordinate and α_0 is the 'viewing angle' for the first image. x_c is calculated separately with the median of all abscissae x_i. In principle, only three two-dimensional images would be necessary to calculate the three-dimensional position of an object. Nevertheless, as the intensity gradients of the crystal edges change with the viewing angle, it is useful to increase the number of images in order to add more data into the least-squares function (1), minimizing the effect of detection errors on some images.

3. Image analysis

The novel crystal-detection algorithm includes several steps: histogram analysis, binarization, loop search, edge-detection, line detection, scoring and object selection. These steps are described below. Before any processing, images are converted to greyscale and inverted. If a background image is provided, images are subtracted from the background.

3.1. Binarization

This step is used to obtain the loop contour for removal from the images. It consists in converting the pixels into 'one' for the sample and 'zero' for the background. The binarization threshold is automatically determined by analysis of the histogram of the pixel intensities in the background and crystal regions. The histogram is strongly smoothed prior to analysis. One of the design goals is to keep the sample contour entirely, even if the loop signal is mixed into the background, which is a frequent condition when the optics has a small depth of field and show a fuzzy loop. The segmentation of the histogram into two parts, loop and background, is based on a threshold T. If a background image is provided, the working image has a dark background with a well identified peak (Fig. 2). In this case T is as close as possible to the end of the background peak, because the only visible object in the image is the loop and its content. Otherwise, if there is no background image provided, T is between the end of the background peak and the start of the crystal peak, at the intersection of the histogram gradient slopes at FWHM of the background and crystal peaks.

3.2. Loop search

Loop-tip detection is used for both loop centring and removal of the loop contour from the images prior to crystal detection. The loop removal depends on the loop type. Currently, only nylon loops are fully supported and Litholoops are being evaluated. As this detection is usually robust and accurate, the loop-tip position is used to precisely calculate the position of the rotation axis. The program scans objects from the right to the left of the binarized image. Depending on the hardware setup, the images are rotated before processing in order to obtain an adequate orientation, with the loop holder on the left. Once an object is found during the scan, its position is marked as the tip and the search continues until a maximum vertical thickness is found. The extension of the loop at this point and its abscissa are stored to later translate the loop in the middle of the screen and adjust the zoom level to match the loop size.

3.3. Crystal detection

The images are pre-processed prior to crystal detection: they are resized and smoothed and the loop contour is removed from the image. The detection is based on either histogram analysis or scoring of the shape and intensity of the object. In the first case a threshold is applied to select only the crystal during a binarization process. It occurs when the probability that the crystal signature is not mixed into the liquid signature is high. The estimation of this probability is based on a comparison of the distances between peaks in the histogram and their heights relative to empirical thresholds. If the crystal contribution is not obvious in the histogram, the scoring method is used. In this case, a modified Canny edge-detection (Canny, 1986) is applied (Fig. 3). The loop wire frequently contains strong gradients, so the maximum gradient for the normalization cannot be taken in the loop-wire area. Labelling the skeletons provides a list $L(o)$ $(1 \leq o \leq m)$ of all the objects inside the loop. The program eliminates the opaque areas (typically arising from the lens effects of large cryo-drops) and reshapes the drop to its visible content only. After labelling objects, the scoring is performed with a formula based on the weighting of different properties reg, sharp, close and extent,

$$s(o) = a \times \text{reg} + b \times \text{sharp} + c \times \text{close} + d \times \text{extent}. \quad (2)$$

Reg is an estimation of the regularity of the object. This score uses the Radon transform (Kak & Slaney, 1988), which can be used to detect all the straight lines. Here, the Radon transform is a projection of the local skeleton along lines parallel to y' (Fig. 4*a*). The projection for all angles $0 \leq \theta \leq 180^\circ$ gives a map $R(\rho, \theta)$ (Fig. 4*b*) containing the number of pixels in the skeleton with the same abscissa $x' = \rho$ in each rotated coordinate system. The maximum on the Radon map occurs when a line of pixels within the skeleton is parallel to the y' axis. All lines including more than four pixels are extracted in order to detect a wide range of crystal sizes. The number of lines extracted is limited by an empirical threshold, as the edges of a regular object usually fit a limited number of lines. These lines are used to sketch the object and the score is the proportion of pixels belonging to the sketched lines. We noticed that the Radon transform can only be applied to a local object (*i.e.* pixels with gradient continuity) because the loop contains many aligned pixels that are not correlated.

Sharp is a sharpness score based on the maximum gradient in the image around the object skeleton divided by the average gradient within the loop.

Close is a measure of the 'closedness' of the skeleton. Crystals are usually convex objects and their skeletons are closed, unlike the boundaries of many artifacts. We devised a fast way to estimate the 'closedness' of any kind of shape: the program counts the number of pixels of the skeleton that can be projected horizontally or vertically onto another pixel of the skeleton in the xy coordinate system; afterwards, the count is normalized by the number of pixels. Nevertheless, the closure of the boundary of large objects may not be detected, since the skeleton is usually cut into parts. This term does not

Figure 3
Canny edge-detection is a very sensitive method to extract crystal edges and gradient orientations. It is obvious that strong object selection must be applied to remove the numerous artifacts created by this method. Subsequent processing will detect the lines among all the selected skeletons.

include shape consideration, so it helps to catch crystals with no visual regularity.

Extent is a size score with a relatively small weight, since artifacts can be large. On the other hand, smaller objects should be given less importance.

The hard-coded weighting constants a, b, c and d are chosen manually to optimize the quality of the results against a database of reference images from users' crystals. The centre of the crystals, as marked by the users, has been stored and *C3D* compares these positions with its own results. This automated procedure helps to choose optimal constants for each *C3D* version. All the objects with a good score are compared to adjust the local scores: this is the global scoring step. Edges having a similar orientation are connected together in order to reconstruct the space inside the crystal and to boost the score of the newly reconstructed object. Finally, the best objects are selected if their score exceeds a threshold and an image score (S_i) is built. Its centre of mass gives the first estimation of the crystal position $M_i(x, y)$ for an image taken at angle φ_i.

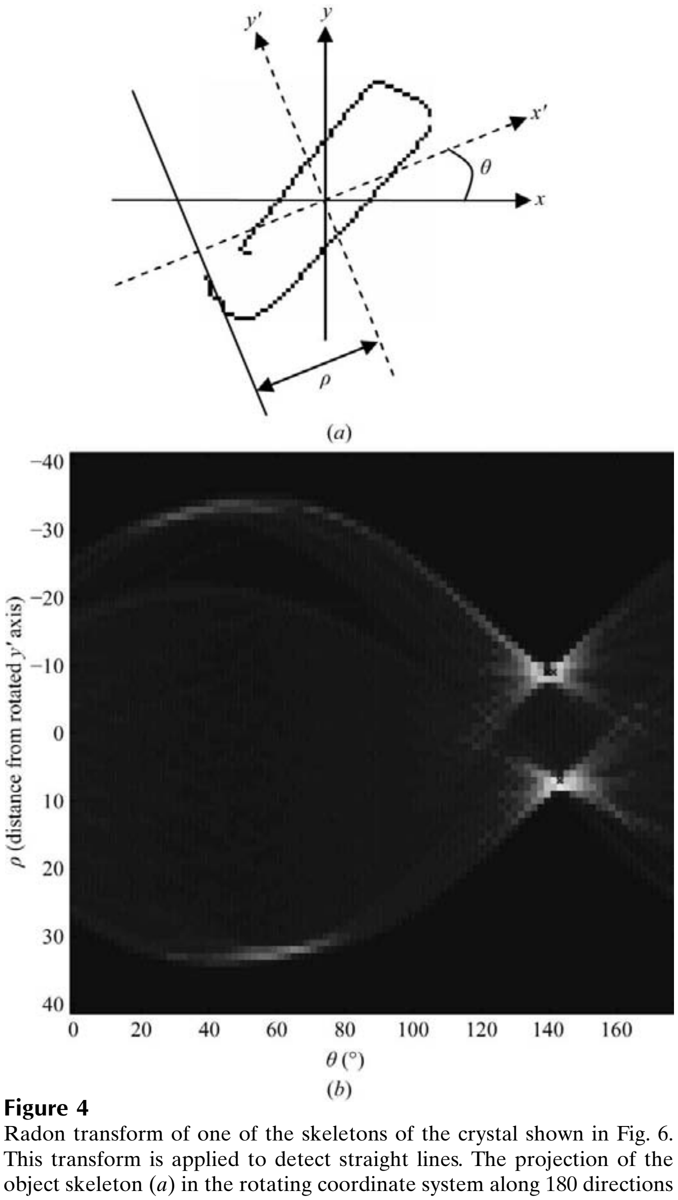


Figure 4
Radon transform of one of the skeletons of the crystal shown in Fig. 6. This transform is applied to detect straight lines. The projection of the object skeleton (*a*) in the rotating coordinate system along 180 directions gives a map (*b*) containing the number of pixels in the skeleton with the same abscissa $x' = \rho$ in each rotated coordinate system. The rotation centre is chosen to be in the middle of the skeleton bounding box. The maximum on the Radon map occurs when a line of pixels within the skeleton is parallel to the y' axis.

3.4. Crystal shape

The crystal segmentation leaves many sparse objects, particularly at high magnifications. Estimating if an object is complete or partial is an additional step required in the crystal-detection process. Detecting the probability of having crystals, as in high-throughput crystallization, is not sufficient, because the goal is to choose a single crystal and to determine its exact position. The knowledge of the crystal extension and even its exact shape helps to choose the correct target between sparse objects. This task could be performed by an algorithm based on 'active contours', directly on the raw image, as an estimate of the crystal centre is already available at this stage. In order to increase the image-processing speed, we have devised a fast 'active bounding box' method to estimate the crystal area. Objects are rejected or recombined in the score space S_i. The box starts with a rectangle $h \times l$, centred on $M_i(x, y)$. The procedure is iterative and involves a block of pixels. The best box size is accepted when the score density is maximized (Fig. 5). The centroid of the score space inside the bounding box is the final crystal centre estimate (Fig. 6). The bounding-box method is still at an experimental stage and is used to determine a region of interest rather than to obtain the crystal shape.

3.5. Quality estimation

Here, we report the current state of our attempts to *a priori* estimate the quality of the auto-centring. We evaluated five indicators: Q_1, detection dispersion, Q_2, the ratio of accepted images, Q_3, the average score of selected objects, and Q_4, the probability that the distance between the crystal centre and the beam centre does not exceed a fraction of the diameter of the larger (either the beam or the crystal). Table 2 contains the average values for these three categories, according to manual inspections. As the detected points reconstructed in three dimensions $P_i(x, y, z)$ must fit a circle in the yz plane perpendicular to the x axis, the standard deviation between the detected crystal centre on each image and this circle should indicate the uncertainty of the detection. We estimated the probability of a centring error by an analysis of the distance between the three-dimensional position detected and the corresponding position selected by a user. From our experiments with an annotated test database, we could see that the distribution of the errors follows a Gaussian curve. We can therefore calculate, *a priori*, the probability that the centring error is between $-e$ and $+e$. The cumulated deviation function (Abramowitz & Stegun, 1972) is $\Phi = \mathrm{cdf}(e) = \mathrm{erf}(e/\sigma_e 2^{1/2})$. The program provides the standard deviation σ_e of the distance between each detected point $P_i(x, y, z)$ and the matching circle given by (1).

Table 2
Average value of quality estimators for 171 series classified according to manual inspection: 'good', 'average' and 'bad'.

'*Q*' variables can be calculated automatically, while '*U*' variables are based on user input. U_1, standard deviation of distance between user click and *C3D* selection. U_2, average distance between user clicks and *C3D* selections. Q_1, detection dispersion (standard deviation of distance between two-dimensional detections and final selection). Q_2, fraction of accepted images. Q_3, average score of selected objects in all the images. Q_4, probability that the distance between the crystal centre and the beam centre does not exceed 30% of the diameter of the largest object. Two beam sizes were used: 100 µm for $e(100)$ and negligible for $e(0)$.

	Good centring	Average centring	Bad centring	Dispersion of quality variable
$U_1 = \sigma_u$ (µm)	11.7	12.5	12.9	7.3
$U_2 = \bar{d}_u$ (µm)	26.7	39.0	95.4	36.7
$Q_1 = \sigma e$ (µm)	10.7	13.5	32.5	14.4
$Q_2 = Q_{\rm im}$	83.9	82.0	82.1	10.9
$Q_3 = \bar{Q}_{\rm sel}$	170	158	170	48.8
$Q_4 = {\rm cdf}[e(100)]$ (%)	97.7	98.2	82.4	10.8
$Q_4 = {\rm cdf}[e(0)]$ (%)	97.1	95.2	75.8	13.9

The maximum acceptable deviation can be defined as an arbitrary function of the beam and the crystal size. Ideally, the crystal boundaries should not cross the beam boundaries and *vice versa* during a rotation. A simple criteria has been tested: $e(r_c, r_b) = 0.6 \times \max(r_c, r_b)$, where the beam is modelled as a cylinder of radius r_b and crystals are approximated as spheres (Pothineni *et al.*, 2006) with radius r_c. In this case, the quality variable Q_4 is the probability that the distance between the crystal centre and the beam centre does not exceed 30% of the diameter of the largest object (Fig. 7). In other words, it is the probability that the cross-section between the auto-centred crystal and the beam is between 63 and 100% of the optimal cross-section. It is possible to estimate this probability because the algorithm uses the three-dimensional bounding box around the crystal ($H \times L \times T$) to calculate the crystal's 'radius': $r_c = (H + L + T)/6$. The quality variables have been correlated against the average distance $\bar{d}_u$ between user and program selections. With *C3D*, none of the quality variables have a good correlation with the actual centring quality. The difference between the average Q_i values for successful centrings and for bad centrings is relatively small compared with their standard deviations, so the value ranges of the two categories overlap. The prediction of the success of the centring is not meaningful in this condition, even with the best estimate (Q_4). For 171 samples, Q_4 ($r_b = 100$) gives seven false negatives and eight false positives. The reason why it is difficult to estimate the quality is the averaging effect of using multiple images, the frequency of well centred artifacts (false positives) and the frequency of larger crystals, which are difficult to detect consistently (false negatives), but easy to centre.

4. *C3D*

4.1. Program description

C3D is a standalone program performing only image processing. It is designed to operate with a client program such as a centring module that controls specific diffractometer hardware. The client program provides a set of images and associated parameters to *C3D*, which returns all the information required to automatically execute the steps of a crystal alignment. *C3D* helps the client program to quickly search for the loop and to set the zoom and loop orientation to the optimal start conditions for crystal detection. For this purpose, it provides the loop's two-dimensional position, its extension, the contrast in the loop and the best angle to start the image acquisition for the crystal centring. The client program specifies the type of target (a loop or a crystal) and what kind of position to calculate (two-dimensional or three-dimensional). The two-dimensional position can be used before centring to pre-position the loop in the field of view of the camera or later as an input for three-dimensional calculation. *C3D* provides the loop tip and the loop centre as results to the client program. The position information is provided with quality estimators to allow several checks during the process and to abort the centring if necessary. These are noise level, contrast, intersections with the display borders, number of rejected images, RMS dispersion and a score. On request, the program can generate intermediate images to allow the evaluation of each detection step and to simulate what the centring position would be on each input image. An additional feature to aid in manual centring allows the user to provide a two-dimensional position on one of three images. *C3D* will then calculate a user-assisted three-dimensional position to enable single-mouse-click centring.

4.2. Program requirements

Although *C3D* is independent of the hardware, it needs images taken under good conditions and the calculation of the centring variables makes several assumptions regarding the hardware setup. The requirements for performing centring using *C3D* include a camera (colour or black and white), a rotation axis perpendicular to the camera sight and a centring device mounted on the rotation axis. The lighting system must be either a backlight in the visible range or front UV lighting to detect the crystal's fluorescence (Vernede *et al.*, 2006). A typical setup based on the design of the microdiffractometer (Perrakis *et al.*, 1999) is implemented on the MD2/MD2M diffractometers (http://www.embl-grenoble.fr/groups/instr/index.html). By default, the cryoloop pin is supposed to stand on the left of the images. The program accepts three additional spindle orientations, top, right and bottom, provided that it is specified as a command-line parameter. The input images can include some 'background objects' such as a cryocooler shadow or dark areas resulting from the viewing device. *C3D* can remove these by using an empty background image. This helps in histogram segmentation when the loop signature is mixed into the background signature. A region of interest can be also specified as an alternative solution.

The images can be of any size, but *C3D* performs an internal resizing if the resolution is larger than required for a specific calculation. For instance, in the loop-detection process resolutions higher than 500 pixels mm^{-1} are reduced to gain speed. The recommended formats are png, tiff, ppm or bmp for lossless compressed images. The jpg format is also accepted but not recommended for routine use because compression artifacts can perturb the detection.

4.3. Program use

The input is a set of n images (typically $6 \leq n \leq 10$) and several parameters in command-line arguments. The command line can include a reference to a file containing the input parameters. The mandatory parameters include the scale of the images (pixel-to-micrometre), the image-name template with the indices and an array of rotation angles for each image. When all the crystal positions have been calculated, (1) provides the three-dimensional position of the crystal for α_0 (corresponding to the first image) and the simulation of the beam position after the centring (Fig. 6).

The results are stored in ASCII result files. The two-dimensional positions, $M_i(x, y)$ for $1 \leq i \leq n$, are stored in 'two-dimensional' files and the centring variables are saved in one 'three-dimensional' file. The positions of the loop or crystal are given in two coordinate systems (Fig. 1): the image-plane coordinate system (c.s.1) and the rotation plane coordinate system (c.s.2). *C3D* provides the two-dimensional positions in c.s.1. It also provides a three-dimensional position, combining the alignment information (x_c, y_c) in c.s.1 and the polar coordinates (r, α_0) in c.s.2, which are orthogonal to c.s.1. The typical use of these output variables is to shift the sample on the rotation axis with a centring table device by using r and α_0. The method depends on the hardware, but for example one can rotate the spindle axis to a position where the crystal is viewed at the extreme distance from y_c (Fig. 1) and then vertically translate the sample towards the rotation axis by r. Afterwards, the centring script should translate the rotation axis and the sample onto the beam by ($x_{\rm beam} - x_c$, $y_{\rm beam} - y_c$).

Figure 5
The result of the scoring and the object selection is in grey. The grey level is linked to the score. The active bounding box result, marked in red, contributed to eliminating artifacts not rejected by the scoring. The blue point represents the initial centre of mass and the green point represents the final centre position.

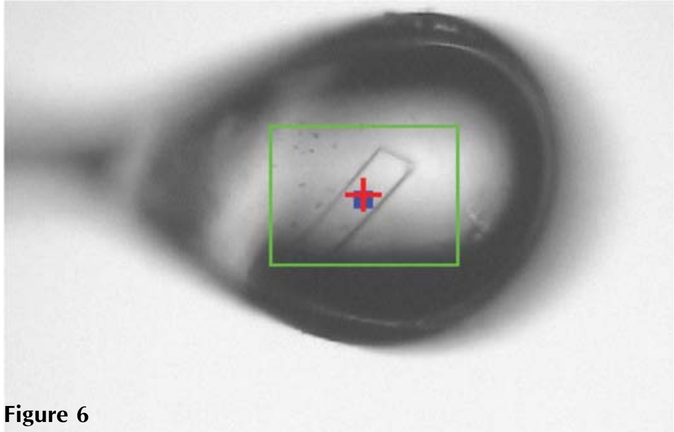


Figure 6
The red cross shows where the crystal is going to be centred according to the result values x_c, α_0, r. The green bounding box is given by the median of all the active bounding boxes of all images. The blue point is the crystal position detected in this particular image.

An optimized choice of rotation angles for image acquisition significantly increases the success rate. The 'best starting angle' is calculated by the program during a loop centring for this purpose. There are two categories of valuable images: when the loop face is not visible (profile of the loop) and when the crystal is best viewed, in other words when the loop plane is perpendicular or parallel to the viewing plane. The first category is handled in a special way because the program can calculate a very accurate vertical position for the crystals in these images. Two particular loop positions are useful in this category and can be obtained using the 'best start angle' plus or minus 90°. The second category of images, which includes a loop with its face nearly parallel to the camera viewing plane, limits the lens effect inherent to backlit cryocooled liquid drops (Fig. 6). This category of loop positions can be obtained around the 'best starting angle' and its opposite at 180°. *C3D* can be launched in server mode, where it waits for the input images and processes them as soon as they are written on the disk by the client program. In order to speed up the three-dimensional calculation, the client program can launch one *C3D* process per image or one process per CPU. In this case, the two-dimensional result file name should include a template. Another process can collect the results of the two-dimensional detections from several processes running on different processors delivering significantly faster results on multi-CPU machines.

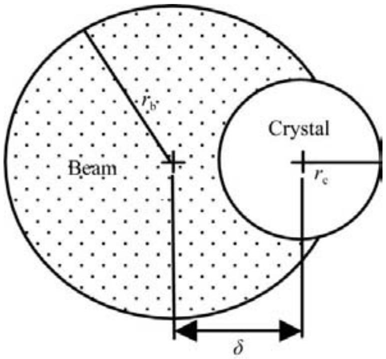
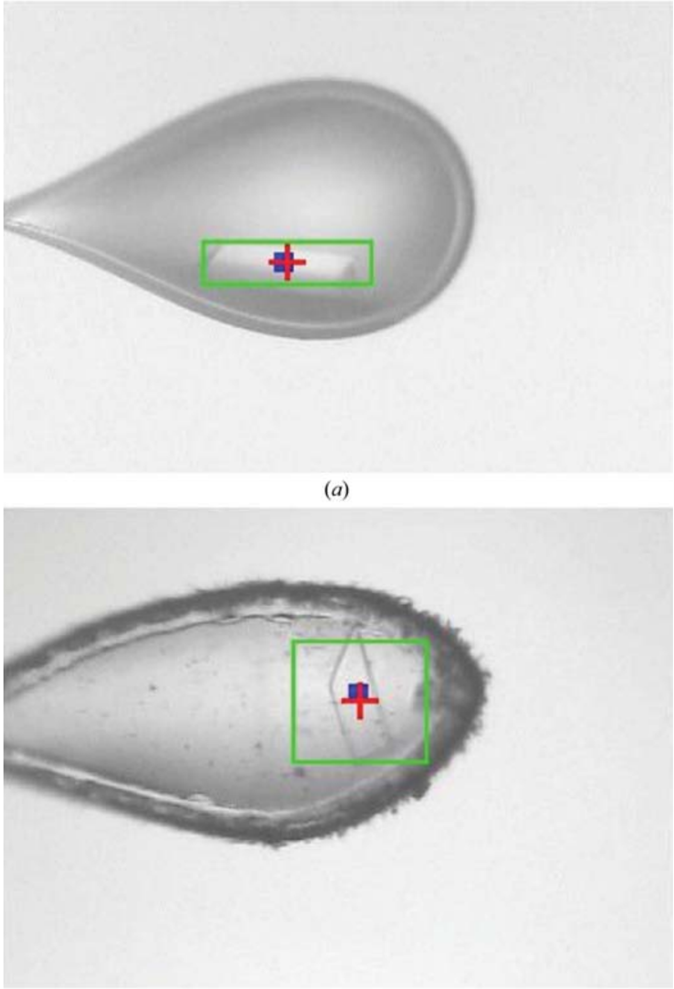


Figure 7
The reliability score can rely on the probability of a deviation δ between the beam centre and crystal centre.

5. Results and discussion

Here, we present a program that is suitable for centring crystals in cryoloops under visible light, with emphasis on the use of the Radon transform. The results of *C3D* have been compared with a human-annotated database containing 171 sets of images, mainly collected on microdiffractometers during routine user operation at the ESRF. The series are representative of the crystal diversity, according to the statistics in Table 1. The simulations can locate the correct centring position inside the crystal at different rotation angles for 84% of the series; 9% are failures and 5% are unclassified because the crystal could not be detected unambiguously by eye. The efficiency of the scoring on two challenging transparent crystals is demonstrated in Fig. 8. However, 20% of successful detections are very close to one edge, because *C3D* failed to identify the other edges of the crystal. Another estimation of the quality is the distance between the position detected by *C3D* and the position selected by the user (Fig. 9). *C3D* used with a microdiffractometer performs 50% of detections within 30 µm of the position chosen by the user and 80% of detection within 50 µm. Detection results against the EMBL Grenoble and the BIOXHIT image database can be found at http://www.embl-grenoble.fr/groups/instr/auto_centring/index.html. The first 100 sets of images were used to optimize the weighting of the scoring function. In principle, this might bias the overall success rate of the ranking for these images. To test this, we used the same weights for the subsequent 71-image series and this does not show any altered outcome when ranking the success of centring. Detection using visible light is quite demanding in terms of image quality, resolution and the absence of digitizing artifacts. The centring script also plays a role in the choice of the images. The result statistics are therefore only fully valid for one visualization system and one centring strategy. Thanks to the BIOXHIT EU Program, which encourages the dissemination of *C3D* and the growth of a test database with images from different setups, we will have a better idea of the flexibility of the presented method. Wilson (2002), Spraggon *et al.* (2002), Cumbaa *et al.* (2003) and Bern *et al.* (2004) have validated other classifiers in parallel to our work for the recognition of crystals in crystallization drops. Some of them could be applied within *C3D*. The recent work with *XREC* at EMBL Hamburg covers both the field of crystal centring (Pothineni *et al.*, 2006) and crystallization screening. The *C3D* method has proved to have sufficient detection quality to be usable for automated screening of samples. The reliability of detecting a point inside the crystal opens the possibility of finding the exact crystal shape using an active contour method (Cocquerez & Philipp, 1995). Moreover, the crystal's two-dimensional shape determination on each image is a step towards reconstructing the three-dimensional shape of the crystal. This would allow an accurate X-ray absorption to be calculated and different parts of the crystal to be selected automatically during data collection.


(*a*)

(*b*)

Figure 8
The efficiency of the scoring incorporating the Radon transform can be seen on two challenging crystals with low contrast (*a*) or transparency (*b*). The red cross shows where the crystal is going to be centred.

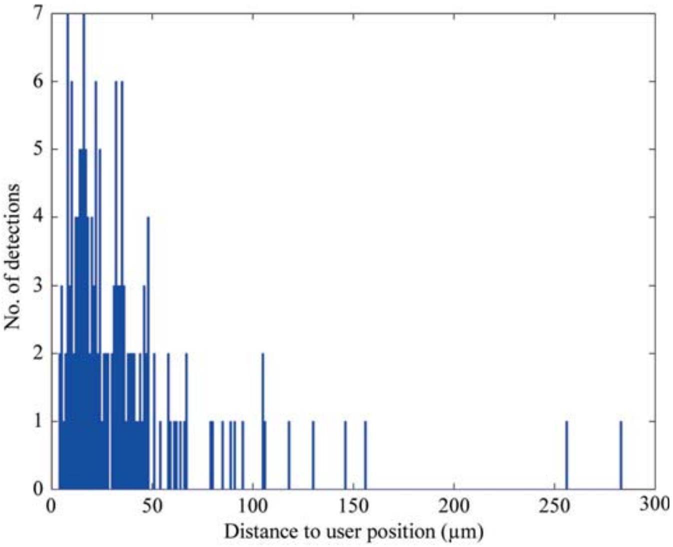


Figure 9
The histogram of differences between crystal positions estimated manually by the user and by the program *C3D*, based on images of 171 crystals. 50% of the *C3D* estimations are within 30 µm and 80% of *C3D* estimations are within 50 µm of the positions chosen by the user.

Table 3
Timing of a full centring procedure, from sample-changer mounting to centring a typical sample on the beam, on an MD2 microdiffractometer along with *C3D* on a Windows XP Dual 2.6 GHz PC.

In practice, *C3D* allows running the calculation in parallel to spindle rotation, thus reducing the centring duration from 76 to 62 s.

Task	Software (s)	Hardware (s)
Search and translate	2	3
Auto contrast	3	3
Scan loop†	12	18
Centring loop	0	3
Auto zoom	0	4
φ positioning	0	1
Scan crystal‡	12	14
Crystal centring	0	3
Total	29	49

† Loop scan, ten images. ‡ Crystal scan, eight images.

The program still has limited capabilities when strong artifacts are present or when the crystal is hardly visible. Hence, it is interesting to explore complementary centring methods, such as those using X-ray or UV fluorescence. As stated in §1, UV-based centring should be reliable and even faster. With visible light *C3D* does not use *Crystal Trace* (Andrey *et al.*, 2004), the algorithm originally implemented on the MD2. Nevertheless, with UV-induced fluorescence images the *Crystal Trace* algorithm will be re-evaluated. It is fast and may be a good candidate for identifying crystals that are obscured behind Kapton foils or competing with fluorescence from residual proteins in the crystallization drop. Algorithms based on centroid calculation are also valuable in this area. New sample holders such as Mesh Litholoops and Micro-Mounts should be implemented soon in *C3D*, whereas UV fluorescence can already be used (Vernede *et al.*, 2006). Kappa goniometers will bring different challenges, as the sample-holder pin might not longer coincide with the spindle axis. There are conditions under which the Kappa angle and loop-bending angle will lead to a wrong loop-tip determination using the current horizontal scan. We are evaluating this issue and some possible solutions.

There are other remaining challenges, such as the detection of multiple crystals in a cryoloop and the subsequent selection of the best one while avoiding the other crystals. We showed that very good 'reliability' estimation is challenging and more work is needed to make it of practical use. *C3D* will include the Q_4 variable as described in §3.5 and we expect to improve this indicator in the future.

The speed of the automatic centring largely depends on the hardware. On a 3 GHz 32-bit Pentium, crystal detection takes typically 1.5 s per image, independently of the operating system (Linux or Windows). On microdiffractometers, the total hardware moving time exceeds the *C3D* calculation time (Table 3). To gain speed, the focus should be on parallelizing the hardware actions with the software calculations as much as possible. Other optimizations are possible, such as taking images on the fly without stopping the spindle, which should divide the time to grab the images on on a microdiffractometer by a factor of three.

6. Program avaibility

The program is a compiled Matlab script. Information and documentation can be found at http://www.embl.fr/groups/instr/auto_centring/index.html. The program has been tested on Windows 2000, XP, Linux Suse 8.2 and RedHat 9.0. All Linux with kernel 2.2.*x* or 2.4.*x* should work with glibc (libc6) 2.1.2 or 2.2.5 or with emulation packages. The Linux and Windows versions are available upon request for non-commercial use.

We thank ESRF MX beamlines scientists and Martin Walsh from MRC France for enabling the tests on the automated beamlines, the ESRF BLISS support group and in particular Matias Guijaro for his help integrating *C3D* on the MD2M diffractometers and the BIOXHIT partners (http://www.bioxhit.org) and in particular Sudhir Babu Pothineni from EMBL Hamburg. The authors like to thank Andrew McCarthy, John McGeehan, Tilo Strutz and Victor Lamzin for critical reading of the manuscript. The work was supported by the EU grants BIOXHIT, under contract No. LHSG-CT-2003-503420, and SPINE, under contract No. QLG2-CT-2002-00988.

References

Abad-Zapatero, C. (2005). *Acta Cryst.* D**61**, 1432–1435.

Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. (2000). *Nature Struct. Biol.* **7**, *Suppl.*, 973–977.

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.

Andrey, P., Lavault, B., Cipriani, F. & Maurin, Y. (2004). *J. Appl. Cryst.* **37**, 265–269.

Bern, M., Goldberg, D., Stevens, R. C. & Kuhn, P. (2004). *J. Appl. Cryst.* **37**, 279–287.

Canny, J. (1986). *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698.

Cocquerez, J. P. & Philipp, S. (1995). *Analyse d'Images: Filtrage et Segmentation.* Paris: Éditions Masson.

Cumbaa, C. A., Lauricella, A., Fehrman, N., Veatch, C., Collins, R., Luft, J. R., DeTitta, G. & Jurisica, I. (2003). *Acta Cryst.* D**59**, 1619–1627.

Judge, R. A., Swift, K. & Gonzalez, C. (2005). *Acta Cryst.* D**61**, 60–66.

Kak, A. C. & Slaney, M. (1988). *Principles of Computerized Tomographic Imaging.* Piscataway, NJ, USA: IEEE Press.

Karain, W. I., Bourenkov, G. P., Blume, H. & Bartunik, H. D. (2002). *Acta Cryst.* D**58**, 1519–1522.

Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 1013–1021.

Nanao, M. H. & Ravelli, R. B. G. (2006). *Structure*, **14**, 791–800.

Perrakis, A., Cipriani, F., Castagna, J. C., Claustre, L., Burghammer, M., Riekel, C. & Cusack, S. (1999). *Acta Cryst.* D**55**, 1765–1770.

Pohl, E., Ristau, U., Gehrman, T., Yahn, D., Robrahn, D., Malthan, D., Dobler, H. & Hermes, C. (2004). *J. Synchrotron Rad.* **11**, 372–377.

Pothineni, S. B., Strutz, T. & Lamzin, V. S. (2006). *Acta Cryst.* D**62**, 1358–1368.

Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.

Roth, M., Carpentier, P., Kakati, O., Joly, J., Charrault, P., Pirrochi, M., Kahn, R., Fanchon, E., Jacquamet, L., Borel, F., Bertoni, A., Israel-Gouy, P. & Ferrer, J. L. (2002). *Acta Cryst.* D**58**, 805–814.

Saitoh, K., Kawabata, K., Asama, H., Mishima, T., Sugahara, M. & Miyano, M. (2005). *Acta Cryst.* D**61**, 873–880.

Spraggon, G., Lesley, S. A., Kreusch, A. & Priestle, J. P. (2002). *Acta Cryst.* D**58**, 1915–1923.

Thorne, R. E., Stum, Z., Kmetko, J., O'Neill, K. & Gillilan, R. (2003). *J. Appl. Cryst.* **36**, 1455–1460.

Ueno, G., Hirose, R., Ida, K., Takashi, K. & Yamamoto, M. (2004). *J. Appl. Cryst.* **37**, 867–873.

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.

Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.

Wilson, J. (2002). *Acta Cryst.* D**58**, 1907–1914.

Wilson, J. (2004). *Crystallogr. Rev.* **10**, 73–84.

Wilson, J. & Berry, I. (2005). *J. Appl. Cryst.* **38**, 493–500.