Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Evaluation of protein crystallization states based on texture information derived from greyscale images

Kanako Saitoh,[a,b]* Kuniaki Kawabata,[b] Hajime Asama,[c] Taketoshi Mishima,[a] Mitsuaki Sugahara[b] and Masashi Miyano[b]

[a]Saitama University, 255 Shimo-okubo, Sakura, Saitama, Saitama 338-0825, Japan, [b]RIKEN (The Institute of Physical and Chemical Research), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, and [c]The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-0041, Japan

Correspondence e-mail: kana@me.ics.saitama-u.ac.jp

Received 8 November 2004
Accepted 14 March 2005

In recent years, several projects have advanced research and development related to the automation of the protein crystallization process. However, evaluation of crystallization states has not yet been completely automated. In the usual crystallization process, researchers evaluate the protein crystallization growth states based on visual impressions and assign them a score over and over again. The method presented here automates this evaluation process. This method attempts to categorize the individual crystallization droplet images into five classes. The algorithm is comprised of pre-processing, feature extraction from images using texture analysis and a categorization process using linear discriminant analysis. The performance of this method has been evaluated by comparing the results obtained by using this method with the results from a human expert and the concordance rate was 90.6%.

1. Introduction

High-throughput protein structure determination has been advanced by many projects (*e.g.* Abola *et al.*, 2000) and the necessary time for determining protein structure has been greatly reduced. In contrast, although the crystallization process is the first requisite for protein structural determination by X-ray crystal structure analysis, it still requires a great deal of time. Therefore, there have been several attempts to achieve high-throughput crystallization; for example, Hauptman–Woodward Medical Research Institute Inc. have developed a robotic system which sets up 40 000 crystallization conditions per day and an image-processing system which images 28 plates (1536-well) in 10 h and Stevens (2000) has developed another robotic system which screens 10–100 000 crystallization conditions per day. RIKEN (The Institute of Physical and Chemical Research) in Japan also has been developing a robotic system to automate the process of crystallization. The details of the system will be described in §2.

Through the development of such robotic systems, the efficiency of crystallization work has seen rapid improvements. However, the systems have mainly realised the automation of setup and storage of the crystallization samples and the observation process has not yet been completely automated. The researchers therefore have to keep the samples under observation.

In previous work, methods related to automation of the evaluation of the crystallization droplet have been proposed. For example, Bodenstaff *et al.* (2002) and Echalier *et al.* (2004) used a rotating-polarizing filter to detect crystals in protein crystallization droplets containing precipitate. Their processing targets are not images but are the droplets themselves. Crystallization images have been used by Cumbaa *et al.*

(2003), who used the Radon transform to extract straight-edge features, Zuk & Ward (1991), who used the Hough transform to identify crystal edges, Rupp (2003), who used phase congruency to detect a large number of small crystals, and Gester *et al.* (2003), who automated counting the number of crystals to generate three-dimensional surface plots of the crystals and to determine crystal size based on the length of the perimeter of the crystals. All these methods put the emphasis on discriminating between the presence and absence of crystals. However, in addition to detecting crystals, it is also very important and valuable to observe and evaluate the crystallization growth states from start to finish.

Various methods for evaluating crystallization growth states have been proposed. Spraggon *et al.* (2002) used texture analysis and a self-organizing neural network to categorize individual crystal trials into six classes (0, experimental mistake; 1, clear drop; 2, homogenous precipitant; 3, inhomogenous precipitant; 4, microcrystals; 5, mountable crystals). Bern *et al.* (2004) used the Hough transform and curve tracking and classified into the classes 0, empty; 1, clear; 2, precipitate; 3, microcrystal; 4, crystal. Jurisica *et al.* (2001) used a two-dimensional Fourier transform and classified into classes 0, clear; 1, amorphous precipitate; 2, phase separation; 3, microcrystals; 4, crystals. Adams *et al.* (2002) extracted 11 features from images acquired by the RoboMicroscope II system and classified the droplets into four classes (0, clear; 1, precipitate; 2, crystals; 3, other). Wilson (2002) categorized the objects in the crystallization droplets into three classes (0, single crystals, overlapping crystals and clusters; 1, promising conditions such as interesting precipitate or microcrystals; 2, objects arising from skin on the drop as well as various lighting effects and other unfavourable outcomes) based on their size, shape, curvature of the boundary and variance in intensity *etc.*

We consider that there are two reasons why it is not easy to automate such observation processes. One is that there is no definition of the detailed criteria for the evaluation. The other is that it is unpredictable how the crystallization samples grow. In this article, we attempt to evaluate the incipient growth states of protein crystallization using image processing and statistical analysis and aim to achieve a categorizing accuracy of more than 80% in individual classes, whereas the evaluation accuracies in each class are not stabilized in the above-mentioned methods. To be specific, we extract texture information from greyscale crystallization images and derive numerical evaluation criteria using evaluation scores assigned by an expert's empirical knowledge.

2. Methods

2.1. Image acquisition

Protein crystallization images vary depending on the crystallization techniques and imaging devices. Fig. 1 shows some example images taken with the above-mentioned TERA crystallization system developed by RIKEN that were used in this article. The system employs the microbatch method as the crystallization technique and has already provided the following capabilities.

(i) Dispensing the crystallization reagents for the microbatch method (total drop size is 16 µl: 0.5 µl protein, 0.5 µl precipitant and 15 µl oil).

(ii) Taking images of each drop on an *xyz* stage using an inverted microscope from Olympus with a cooling CCD camera from Roper Scientific.

(iii) Supervising 2560 crystallization plates and 125 reagent plates (72-well plate).

(iv) Handling the plates between the appliances with manipulators.

The acquisition image size is 1392 × 1040 pixels and the pixel size is 4.65 × 4.65 µm.

2.2. Classification

The growth states of protein crystallization can take many forms; for example, precipitate, amorphous agglutinate, crystals with varied shapes and combinations of these. In order to respond to these variations, RIKEN sets the ten standard

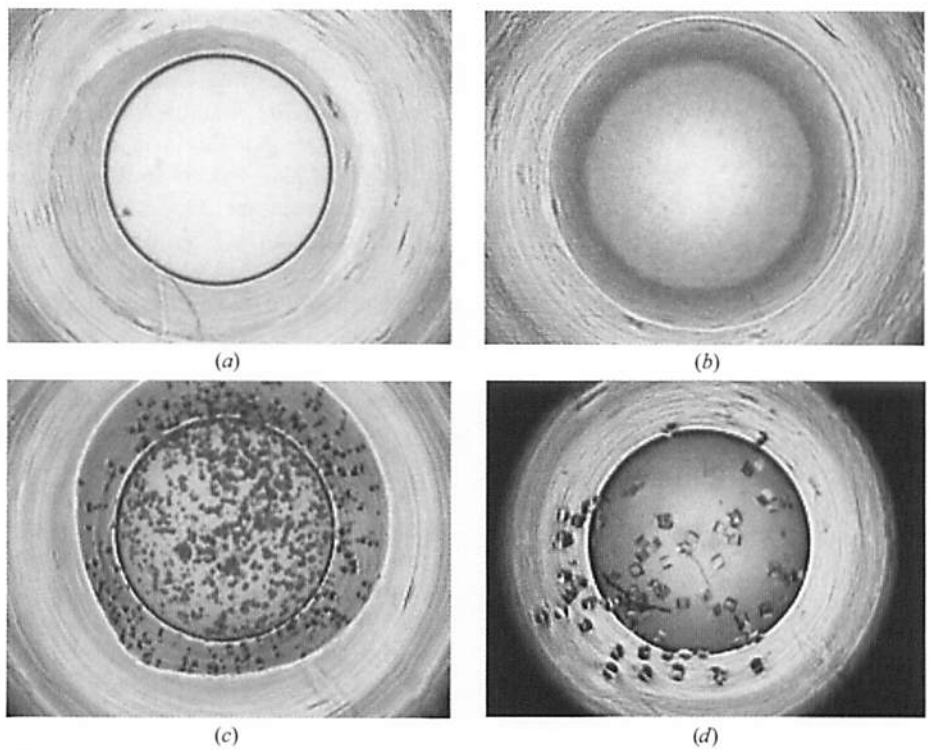

Figure 1
Examples of drop images taken with the TERA system. The size is 1392 × 1040 pixels and the well is visible somewhere in the centre. (*a*) Clear, (*b*) precipitate, (*c*) amorphous grain and (*d*) crystal.

categories for evaluation as shown in Fig. 2. A brief description of each category is given below.

0, Clear drop.

1, Precipitate (i): creamy and grainless precipitate.

2, Precipitate (ii): fine or granulated sugar-like precipitate.

3, Precipitate (iii): amorphous state (whether it will crystallize in the future or not is not known).

4, Amorphous circular grain (will not crystallize in the future).

5, Microcrystal (size 50 μm or less).

6, Crystal (i): needle crystal or plate crystal.

7, Crystal (ii): cluster of crystals.

8, Crystal (iii): single crystal (0.05–0.2 mm).

9, Crystal (iv): single crystal (>0.2 mm).

The reason the precipitate state is subdivided is that we want to monitor tendencies in the growth of precipitates. It is difficult even for experts to evaluate a good or bad precipitate. Therefore, we accumulate in-depth growth records and make a database of them so that trends can be shown. The reason the crystalline state is subdivided is because how they should be treated changes according to their shape. For example, single crystals can be used in X-ray analysis without any change, while a cluster of crystals needs to be separated out before analysis and very thin needle crystals and plate crystals may not be usable for analysis.

Although our ultimate goal is to realise a totally automated evaluation system for the ten categories, we attempt to evaluate categories 0–3 as the first step in this article. The states from class 0 to class 3 do not have shape, while those from 4 to 9 have some shape, for example; circular form, needle and plate *etc.* Therefore, to evaluate samples in categories 0–3 and 4–9, we think that features based on image pattern and based on shape should be effective, respectively. We attempt to evaluate the first of these two sets of categories in this article. Our target classification, which is a reclassification of the RIKEN system, is as shown in Table 1. The groups from 4 (amorphous grain) to 9 (crystal iv) are put together into one category, *E*, and we classify all samples into five categories from *A* to *E* [*A*, clear; *B*, precipiate (i); *C*, precipitate (ii); *D*, precipitate (iii); *E*, amorphous grain–crystal (iv)]. In the RIKEN system, the samples in clear and precipitate stages (0–3) occupy approximately 70% of all supervised samples, so we consider that our target categorization is very effective for categorization work. Group *E* includes non-crystalline objects (group 4 of the RIKEN categorization) with crystalline objects. The objects in group 4 have circular shape, so it may be possible to classify them by measuring circularity. This subject is not handled in this article and will be addressed in future work.

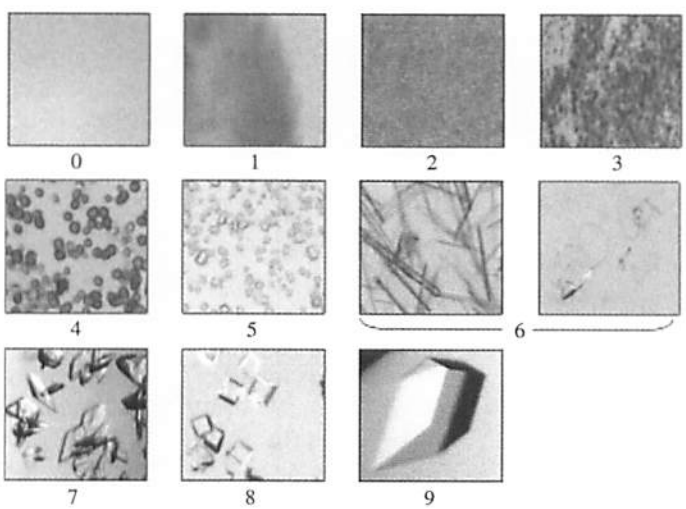

Figure 2
Ten categories for evaluation set by RIKEN. 0, clear; 1, precipitate (i); 2, precipitate (ii); 3, precipitate (iii); 4, amorphous grain; 5, microcrystal; 6, crystal (i); 7, crystal (ii); 8, crystal (iii); 9, crystal (iv). (These images are a part of the well in the full drop images.)

Table 1
Listing of evaluation categories.

Categorization in this paper is into five groups from *A* to *E*. Categories *A*–*D* correspond to from 0 to 3 of RIKEN's scoring, respectively. Groups 4–9 are put together as class *E*.

RIKEN evaluation categories	Evaluation categories in this paper
0, clear	*A*
1, precipitate (i)	*B*
2, precipitate (ii)	*C*
3, precipitate (iii)	*D*
4, amorphous grain	*E*
5, microcrystal	
6, crystal (i)	
7, crystal (ii)	
8, crystal (iii)	
9, crystal (iv)	

3. Algorithm

The method presented here for evaluating the crystallization growth states uses the processing flow shown in Fig. 3. This processing flow is of a general form in the field of pattern recognition. The method consists of 'pre-processing', 'feature extraction' and 'classification'. Generally, in the pre-processing phase, denoising, normalization and so on are performed on input images. In feature extraction, the features

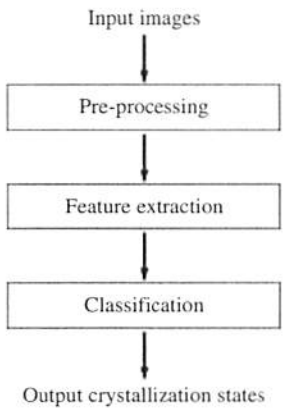

Figure 3
The processing flow of the method presented here. The method consists of pre-processing, feature extraction and classification.

Table 2
Listing of extracted feature values.

Haralick *et al.* (1973) defined the 14 feature values. These are calculated by using co-occurrence matrix P, where $P_x(i) = \sum_j P(i, j)$, $P_y(j) = \sum_i P(i, j)$, $P_{x+y}(k) = \sum_i \sum_j P(i, j)$, $P_{x-y}(k) = \sum_i \sum_j P(i, j)$, $\mu_x = \sum_i iP_x(i)$, $\mu_y = \sum_j jP_y(j)$, $\sigma_x^2 = \sum_i (i - \mu_x)^2 P_x(i)$, $\sigma_y^2 = \sum_j (j - \mu_y)^2 P_y(j)$, $HXY = -\sum_i \sum_j P(i, j) \times \log P(i, j)$, $HX = -\sum_i P_x(i) \log P_x(i)$, $HY = -\sum_j P_y(j) \log P_y(j)$, $HXY1 = -\sum_i \sum_j P(i, j) \log[P_x(i)P_y(j)]$, $HXY2 = -\sum_i \sum_j P_x(i)P_y(j) \log[P_x(i)P_y(j)]$ and $Q(i, j) = \sum_k P(i, k)P(k, j)/P_x(i)P_y(j)$.

No.	Texture-feature values	
1	Angular second moment	$\sum_i \sum_j P^2(i, j)$
2	Contrast	$\sum_k k^2 P_{x-y}(k)$
3	Correlation	$\sum_i \sum_j ijP(i, j) - \mu_x\mu_y/\sigma_x\sigma_y$
4	Sum of square variance	$\sum_i \sum_j (i - \mu_x)^2 P(i, j)$
5	Inverse difference moment	$\sum_i \sum_j P(i, j)/[1 + (i - j)^2]$
6	Sum average	$\sum_k kP_{x+y}(k)$
7	Sum variance	$\sum_k [k - \sum_l lP_{x+y}(l)]^2 P_{x+y}(k)$
8	Sum entropy	$-\sum_k P_{x+y}(k) \log P_{x+y}(k)$
9	Entropy	$-\sum_i \sum_j P(i, j) \log P(i, j)$
10	Difference variance	$\sum_k [k - \sum_k kP_{x-y}(k)]^2 P_{x-y}(k)$
11	Difference entropy	$-\sum_k P_{x-y}(k) \log P_{x-y}(k)$
12	Information measure of correlation 1	$HXY - HXY1/\max(HX, HY)$
13	Information measure of correlation 2	$\{1 - \exp[-2.0(HXY2 - HXY)]\}^{1/2}$
14	Maximal correlation coefficient	(second biggest eigenvalue of $Q)^{1/2}$

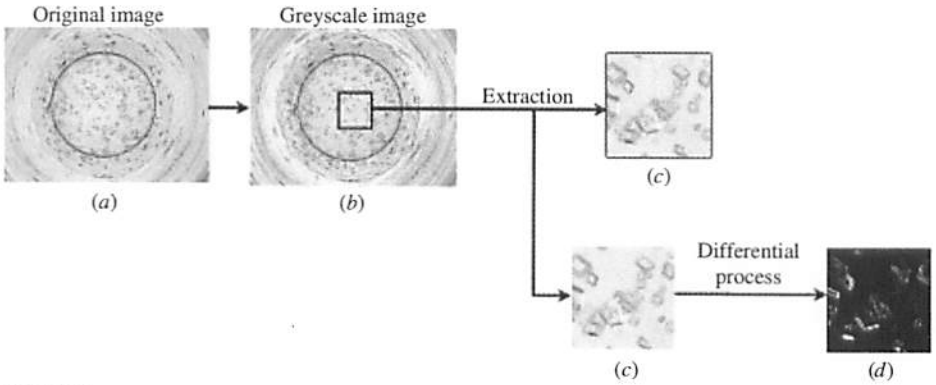

Figure 4
Pre-processing sequence. (*a*) Original image; the image size is 1392 × 1040 pixels. (*b*) 256-level greyscale image. (*c*) A portion of the original image; the image size is 150 × 150 pixels. (*d*) Differentiated image.

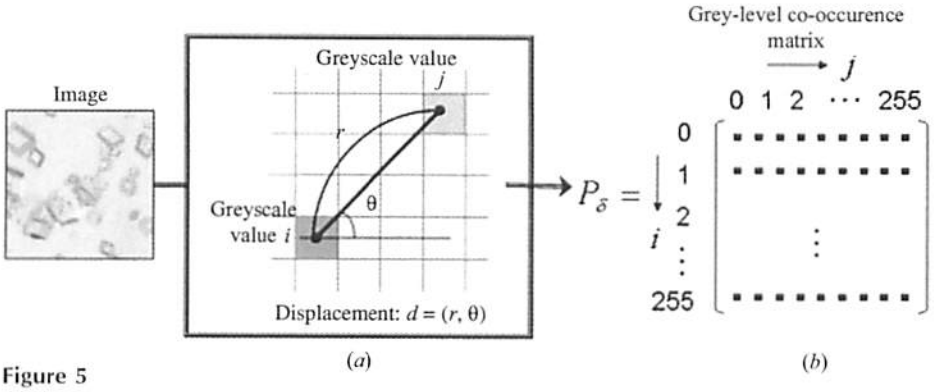

Figure 5
Algorithm of deriving the grey-level co-occurrence matrix. Texture-feature values are calculated by using a grey-level co-occurrence matrix [P_δ in (*b*)]. (*a*) Each element of the matrix expresses the probability that the greyscale value of one pixel is i and the greyscale value of another pixel located r pixels away in direction θ from the former pixel is j. In this article, $r = 1$ and $\theta = 0, 45, 90, 135°$ are used.

used for classification are extracted from the original images, which contain a great deal of information. In the classification process, the input images are classified using these features.

The details of the pre-processing, feature-extraction process and classification process of this method are described below.

3.1. Pre-processing

Before the feature-extraction process, some pre-processing steps are carried out. Fig. 4 indicates the schematic flow of the pre-processing. The original images (Fig. 4*a*) are photographed by TERA using a microscope at 40-fold magnification. The image size is 1392 × 1040 pixels. The number of utilized images totalled 874.

Initially, the original colour images are transformed into 256-level greyscale images (Fig. 4*b*), because colour information of images is not utilized in this method.

Next, a portion of the original image (Fig. 4*c*) is manually extracted from inside the well. The processing object in this study is assumed to be inside the well only. The extraction size is 150 × 150 pixels, which is determined by considering the approximate average size of the microcrystals and crystals in the original images.

Finally, the extracted image is differentiated with a Sobel first-order differential filter (Fig. 4*d*). This process highlights the characteristic pattern of the image. Both differentiated (Fig. 4*d*) and non-differentiated (Fig. 4*c*) images are utilized in this method.

3.2. Feature extraction using texture analysis

In computational image analysis, 'texture analysis', 'smooth', 'sharpen' and 'edge detect/enhance' are often utilized to quantify images (Takagi & Shimoda, 1991). In the method presented here, we decided to introduce texture analysis (Haralick *et al.*, 1973), which quantifies the array of greyscale values of each pixel in an image, to extract features from the crystallization images. Spraggon *et al.* (2002) have already used texture analysis to investigate crystallization trials. They used five texture features and six edge features, whereas we have used 14 texture features. Additionally, there are differences in target classification, the size of processed area and the classification method between the two methods.

Texture-feature values are calculated by using a grey-level co->occurrence matrix. Fig. 5 shows the algorithm used

Table 3
Results of classification.

The results by the method presented here ('automatic classification') are compared with the results by an expert ('manual classification'). The accuracy are class *A*, 98.1%; class *B*, 80.3%; class *C*, 82.8%; class *D*, 85.4%; class *E*, 93.9%. The overall accuracy is 90.6%.

		Automatic classification						
	Total	*A*	*B*	*C*	*D*	*E*	*F*	Accuracy (%)
Manual classification								
A	53	**52**	0	1 (1.9%)	0	0	0	98.10
B	71	3 (4.2%)	**57**	10 (14.1%)	0	1 (1.4%)	0	80.30
C	29	0	1 (3.4%)	**24**	2 (6.9%)	2 (6.9%)	0	82.80
D	41	0	0	1 (2.4%)	**35**	5 (12.2%)	0	85.40
E	245	0	0	2 (0.8%)	11 (4.5%)	**230**	2 (0.8%)	93.90

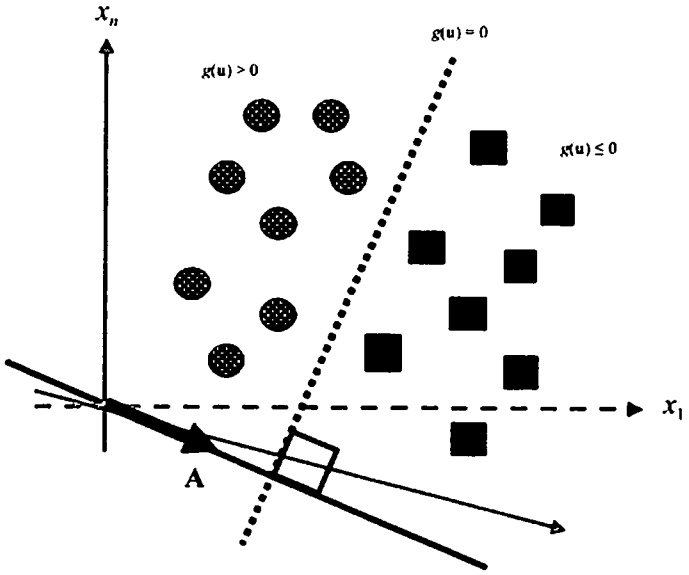

Figure 6
The linear discriminant function $g(\mathbf{u})$ divides the feature space into two groups. A discriminant space is constructed from the linear transformation $g(\mathbf{u}) = \mathbf{A}^T\mathbf{u} + a_0$. Input data are classified depending on their sign.

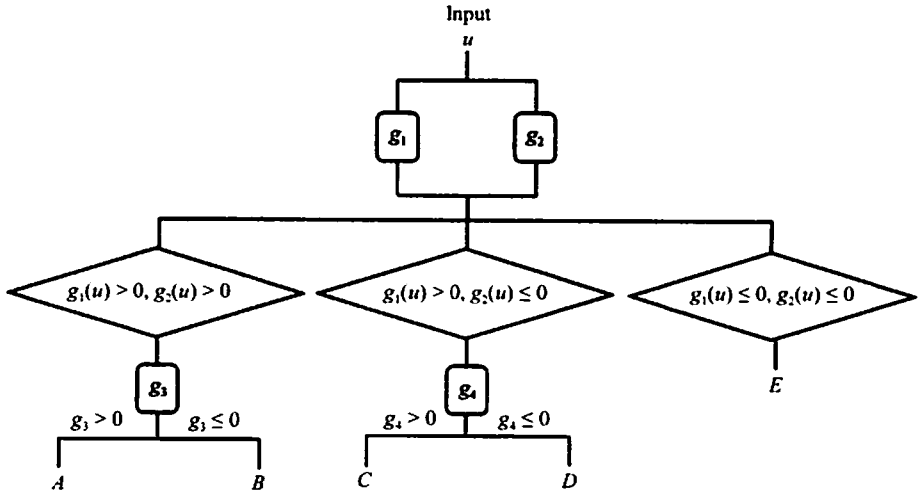

Figure 7
The discrimination procedure, consisting of two steps. In the first step, the input data is classified into *A*/*B*, *C*/*D* or *E* using the functions g_1 and g_2. In the second step, the data is classified into *A*, *B*, *C* or *D* using the functions g_3 and g_4.

to derive the matrix. Each element of the matrix $P(i, j)$ in Fig. 5(*b*) expresses the probability that the greyscale value of one pixel is i and the greyscale value of another pixel located r pixels away in direction θ from the former pixel is j (Fig. 5*a*), where the displacement between two pixels is denoted $\delta = (r, \theta)$. The distance between pixels r used to calculate the co-occurrence matrices is 1 and the direction θ is 0, 45, 90, 135°. The four directions correspond to horizontal, vertical and the two diagonals. If nothing is done, the features are calculated anisotropically. However, the crystallization growth states are not anisotropic, so we take the average of the results calculated using the four directions. By using the matrix P_δ, 14 texture features are calculated as shown in Table 2.

3.3. Classification using linear discriminant analysis

In previous studies on crystal image analysis, several classifiers have been used; for example, a self-organizing neural net (Spraggon *et al.*, 2002), C5.0 (Bern *et al.*, 2004), linear discriminant analysis (LDA; Cumbaa *et al.*, 2003) and Bayes theorem (Wilson, 2002). In this study, LDA, which is one of the standard techniques of multivariate analysis, is used to classify the feature vector.

LDA (Fisher, 1936) computes a linear discriminant function which divides the feature space into two groups. A discriminant space is constructed from the linear transformation

$$g(\mathbf{u}) = \mathbf{A}^T\mathbf{u} + a_0,$$

where $\mathbf{A}$ is a coefficient matrix, $\mathbf{u}$ is a texture-feature vector and a_0 is a constant involving $\mathbf{A}$. The coefficient matrix $\mathbf{A}$ is computed so that the discriminant criterion

$$J_\Sigma(\mathbf{A}) = \frac{\mathbf{A}^T\Sigma_B\mathbf{A}}{\mathbf{A}^T\Sigma_W\mathbf{A}}$$

may be maximized, where Σ_B and Σ_W are the between-class covariance matrix and within-class covariance matrix, respectively. To classify a new input vector, we compute $g(\mathbf{u})$ and determine which class it belongs to depending on its sign (Fig. 6).

Our target categorization is five classes from *A* to *E*. In order to realise this categorization, we apply the LDA for two classes step by step. We compute a total of four linear discriminant functions: g_1, g_2, g_3, g_4. The grouping and the sequence of the discrimination process were decided as shown in Fig. 7 by the result of repetition of basic experiments.

When the functions g_1, g_2 and g_4 are derived, the texture-feature values calculated from differentiated images are used. In case of the other function g_3, the feature values were calculated from non-differentiated images.

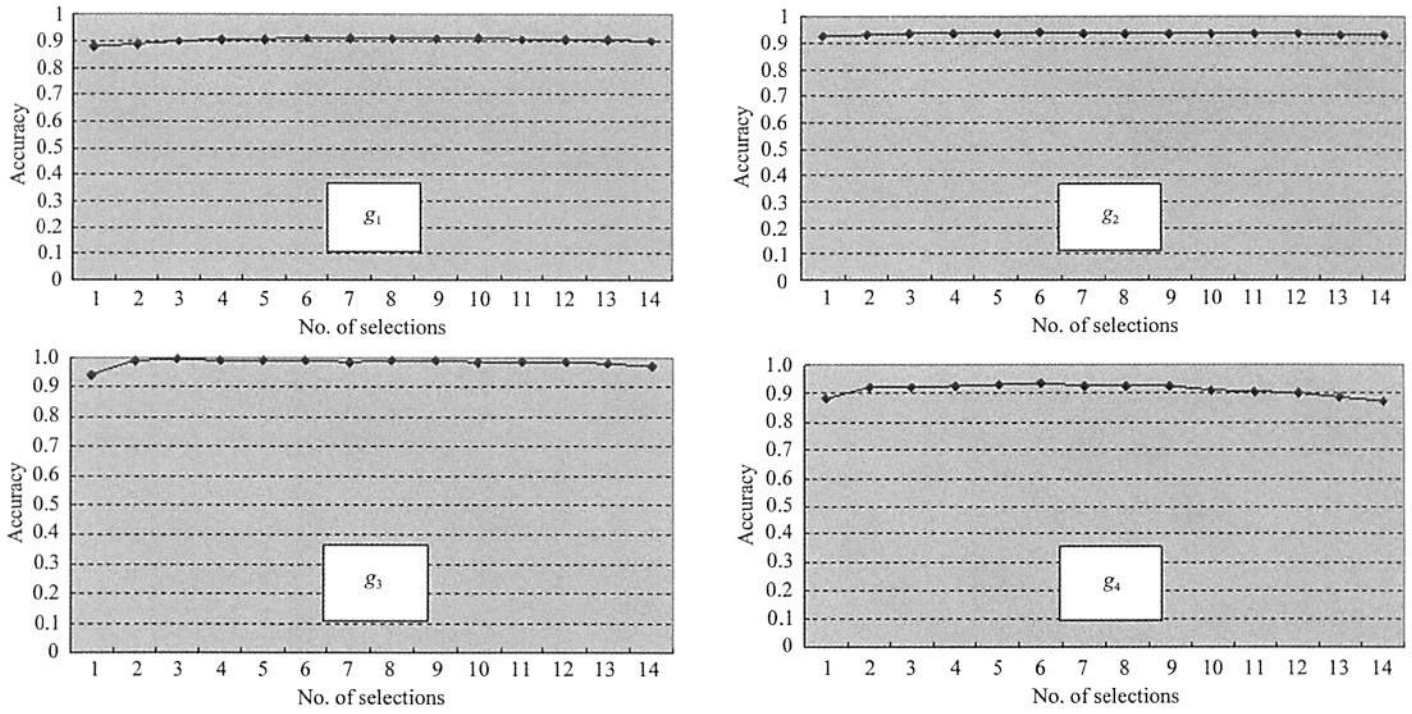

Figure 8
Results of evaluating the accuracy of discrimination for each feature combination. The number of all combination patterns is 16 383, but only the highest accuracy for each number of selections was plotted.

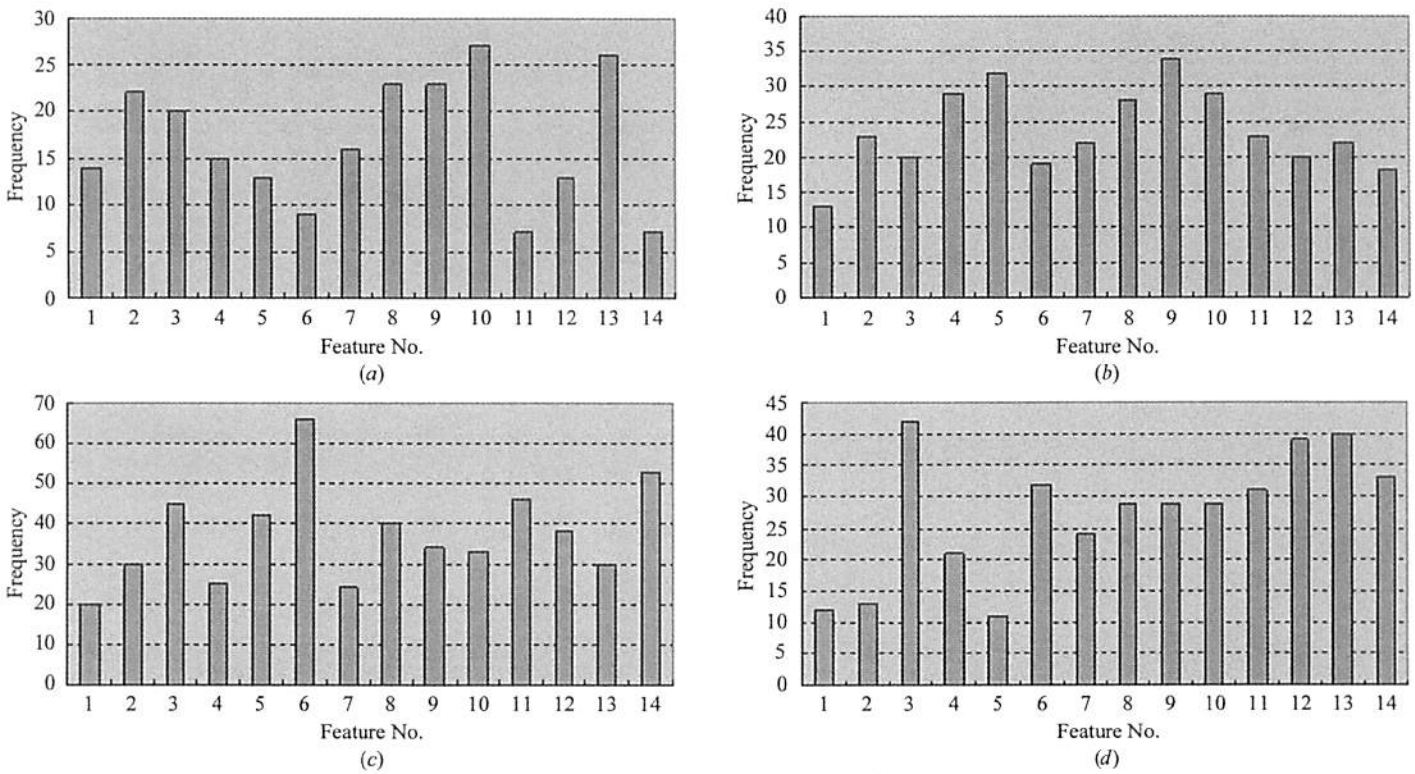

Figure 9
The frequencies of the feature No. being a constituent part of the combinations taking the highest accuracy were graphed by histograms. (*a*), (*b*), (*c*) and (*d*) are the results for the classification steps g_1, g_2, g_3 and g_4, respectively.

4. Results and discussion

The data set used in this study contains 874 images that were annotated by a human expert at RIKEN. The images were obtained with RIKEN's TERA system. The number of images in each category is *A* (clear), 102 (11.6%); *B* [precipitate (i)], 116 (13.2%); *C* [precipitate (ii)], 78 (8.9%), *D* [precipitate (iii)], 90 (10.2%) and *E* (amorphous grain, microcrystal, crystal), 488 (55.8%). Category *E* contains over five times more images than any other category because the images were acquired to maintain a roughly constant number from each of the ten categorizations set by RIKEN. Of the complete 874 images, 435 (*A*, 49; *B*, 45; *C*, 49; *D*, 49; *E*, 243) images were used as a training set and 439 (*A*, 53; *B*, 71; *C*, 29; *D*, 41; *E*, 245) were used as a test set. Training images in which artifacts appeared in the drop were obtained in advance. The evaluation result obtained by using the test set is shown in Table 3. We compared the results obtained using the classification method presented here ('automatic classification' in Table 3) with the results by a human expert at RIKEN ('manual classification' in Table 3) and calculated the concordance rates by the expression (accuracy) = (number of images classified correctly)/(total number) × 100, where class *F* in Table 3 was for the samples that were not categorized into any other classes (*A*–*E*). The accuracies were class *A*, 98.1%; class *B*, 80.3%; class *C*, 82.8%; class *D*, 85.4%; class E, 93.9%. 90.6% of the images were classified into the same category as manually classified by an expert.

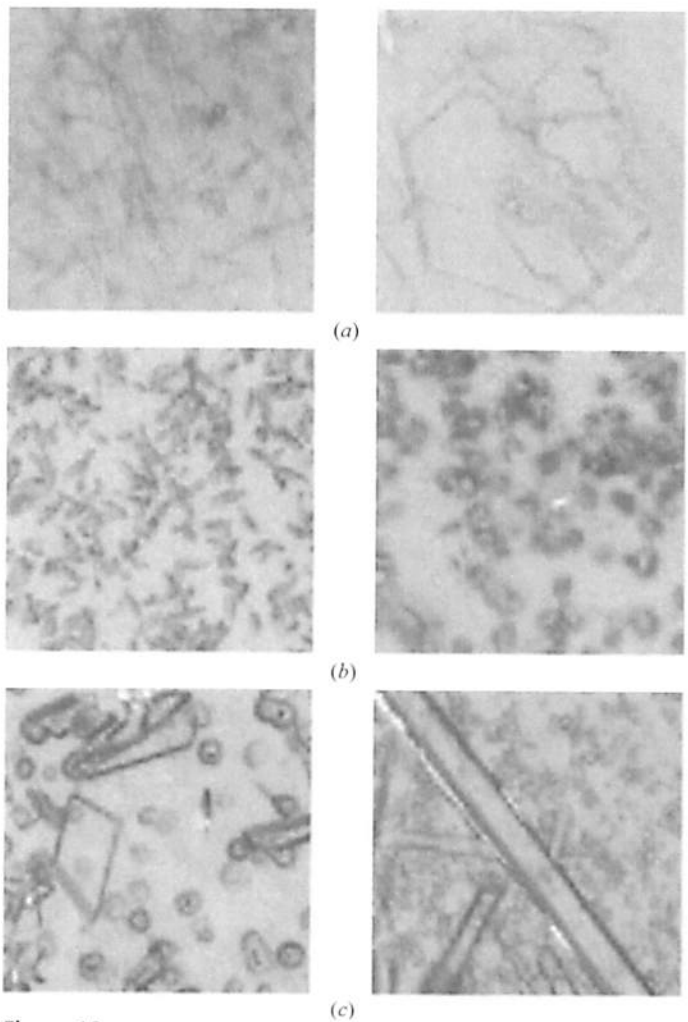

Figure 10
Examples of mis-categorized images belonging to class *E*. (*a*) The precipitated crystals are very thin. (*b*) There are microcrystals that have microscopic linear pattern and angles in the images. (*c*) The crystallization solution weaves precipitate with crystals.

We confirmed that our proposed method is useful for detailed classification of the samples into the categories clear and precipitate. At the present time, human experts judge the samples depending on their empirical knowledge. Individual experts may sometimes make different judgements. The method presented here will establish the numerical criteria between each category of clear and precipitate and overcome this problem.

Next, we discuss which features are most effective in the classification.

Although we used the all 14 kinds of texture features f_k ($k = 1, 2, \ldots, 14$) in this study, it is very important for future research to clarify which features are most effective in each classification step (g_1, g_2, g_3, g_4). We then evaluate the accuracy of discrimination for each combination of the features. The number of combinations when i features are selected from 14 features is denoted by C_i. The number of all combination patterns $\sum C_i$ ($i = 1, 2, 3, \ldots, 14$) is 16 383 (for instance, when selecting one or 13 features from the 14 features the combination numbers C_1 and C_{13} are 14 and when selecting two or 12 features the combination numbers C_2 and C_{12} are 91). We calculated the accuracies e_j^i for all patterns ($i = 1, 2, \ldots, 14$; $j = 1, \ldots, C_i$), but in order to make the graphs easily viewable, only the highest accuracy $\max_j(e_j^i)$ taken in each of the number of selections i were plotted (Fig. 8). The figures show that the evaluating accuracies are not dependent on the number of selections in any classification step. Next, only the combinations that took the highest accuracy $\max_j(e_j^i)$ for each selection number i (points on the graphs in Fig. 8) were covered and the frequencies of the features that were a constituent part of them were graphed by the kind of feature f_k ($k = 1, 2, \ldots, 14$) by histograms (Fig. 9). The histogram horizontal axis corresponds to feature No. k in Table 2. The combinations that took the highest accuracy for each selection number is not necessarily one and sometimes multiple combinations take the highest accuracy. As shown in Fig. 9, the features f_{10} (difference variance) and f_{13} (information measure of correlation 2) are most effective features in the classification step of g_1. In a similar way, the features f_5 (inverse difference moment) and f_9 (entropy), the feature f_6 (sum average) and the features f_3 (correlation), f_{12} (information measure of correlation 1) and f_{13} (information measure of correlation 2) are most effective in the discrimination of g_2, g_3 and g_4, respectively.

Finally, the mis-categorized images are described. From the experimental results, we thought that the mis-classification of the images belonging to classes *A*, *B*, *C* and *D* may decrease if the number of training data were increased. One of the images belonging to class *E* resulted in missing crystals, which would not be affected if more training data were added. Examples of

mis-categorized images belonging to class *E* are shown in Fig. 10. The main characteristics of these images can be summarized by the following three points.

(i) The precipitated crystals were very thin, which caused some crystals to be missed (Fig. 10*a*).

(ii) Despite the microcrystals having microscopic linear patterns and angles, these were overlooked. The method presented here extracts features based on only the array of pixel value, so cannot extract features related to their shape at this time (Fig. 10*b*).

(iii) The crystallization solution had precipitate mixed in with the crystals. We thought that the precipitate reflected strongly and caused the crystal to be missed (Fig. 10*c*).

Considering these points, the method should be refined in our future work.

5. Summary

In this paper, a method for classifying protein crystallization states based on texture information derived from greyscale images has been presented. The method presented here classifies the images into five groups which consisted of mainly clear or precipitate outcomes. The images taken by the automated crystallizing system TERA are used and a texture-analysis method is utilized to extract feature values from each image. Linear discriminant analysis is utilized step by step to classify the samples into five groups. To evaluate the performance of the method presented here, the results annotated by our method are compared with the results of a human expert in RIKEN. As a result, although there are a few mis-classifications, 90.6% of the images were automatically classified as manually classified by an expert.

In the method presented here we process part of the inside well in the original image, but we will try to extend our method in our future work by treating plural parts of the image and processing them comprehensively. If this capability can be realised, images including several crystallization categories will be able to be evaluated. Moreover, we will attempt more detailed classifications by applying contour-extraction methods and so on.

The authors thank Miss Maki Kumei, Mr Nobuo Okazaki, Mr Yuki Nakamura and Mr Tomoyuki Tanaka (RIKEN Harima Institute) for supplying the protein sample images and for their useful advice.

References

Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. (2000). *Nature Struct. Biol.* **7**, 973–977.
Adams, J. A., Jewell, D., Jorgensen, K., Mickly, M. & Newman, J. M. (2002). *JALA*, **7**, 36–40.
Bern, M., Goldberg, D., Stevens, R. C. & Kuhn, P. (2004). *J. Appl. Cryst.* **37**, 279–287.
Bodenstaff, E. R., Hoedemaeker, F. J., Kuil, M. E., de Vrind, H. P. M. & Abrahams, J. P. (2002). *Acta Cryst.* D**58**, 1901–1906.
Cumbaa, C. A., Lauricella, A., Fehrman, N., Veatch, C., Collins, R., Luft, J., DeTitta, G. & Jurisica, I. (2003). *Acta Cryst.* D**59**, 1619–1627.
Echalier, A., Glazer, R. L., Fülöp, V. & Geday, M. A. (2004). *Acta Cryst.* D**60**, 696–702.
Fisher, R. A. (1936). *Ann. Eugen.* **7**, 179–188.
Gester, T. E., Rosenblum, W. M., Christopher, G. K., Hamrick, D. T., DeLucas, L. J. & Tillotson, B. (2003). United States Patent 6 529 612.
Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973). *IEEE Trans. Syst. Man. Cybern.* **3**, 610–621.
Jurisica, I., Wolfley, J. R., Rogers, P., Bianca, M. A., Glasgow, J. I., Weeks, D. R., Fortier, S., DeTitta, G. & Luft, J. R. (2001). *IBM Syst. J.* **40**, 394–409.
Rupp, B. (2003). *Acc. Chem. Res.* **36**, 173–181.
Spraggon, G., Lesley, S. A., Kreusch, A. & Priestle, J. P. (2002). *Acta Cryst.* D**58**, 1915–1923.
Stevens, R. C. (2000). *Curr. Opin. Struct. Biol.* **10**, 558–563.
Takagi, M. & Shimoda, H. (1991). *Handbook of Image Analysis.* Tokyo: University of Tokyo Press.
Wilson, J. (2002). *Acta Cryst.* D**58**, 1907–1914.
Zuk, W. M. & Ward, K. B. (1991). *J. Cryst. Growth*, **110**, 148–155.