

タンパク質結晶化状態判定支援システムの開発

○ 齊藤 佳奈子^{[1][2]}, 川端 邦明^[2], 浅間 一^[3], 三島 健稔^[1]
^[1]埼玉大学, ^[2]理化学研究所, ^[3]東京大学

Development of the protein crystallization states evaluation system

○ Kanako SAITOH^{[1][2]}, Kuniaki KAWABATA^[2], Hajime ASAMA^[3], Taketoshi MISHIMA^[1]
^[1]Saitama University, ^[2]RIKEN, ^[3]The University of Tokyo

Abstract: This paper presents our protein crystallization states evaluation system. In this system, protein crystallization images are input and crystallization states of plural partial regions in target area for evaluation are output. We have done some studies on feature extraction from protein crystallization image and composition method of multi-classes classifier before. In this paper, the results of our previous developed methods, a method that extracts judgment target region from original image, and so on are integrated and experimented.

1. 緒言

近年, 生命科学分野においては効率的なタンパク質構造解析を実現するため, 結晶化プロセスの自動化を目的とした研究開発が進められている. しかし, タンパク質結晶化ドロップの成長状態をスコア付けする結晶化状態判定プロセスは未だ自動化が進んでおらず, 構造解析ハイスループット化のボトルネックとなっている. そこで我々は, 結晶化状態判定プロセスの自動化に関する研究開発を進めている.

我々はこれまでにタンパク質結晶化画像から抽出する特徴に関して¹⁾, また特徴空間中で複雑に分布する多クラス高次元データに対する識別器構築手法に関して研究を重ねてきた²⁾. しかし, 結晶化画像からの判定対象領域抽出など自動化に必須なその他の要素技術が未開発であったため自動化には至っていなかった.

本稿では, 新たにタンパク質結晶化画像からの判定対象領域抽出手法, および教師データの信頼度向上のための効率的な教師データ再評価手法を提案し, 更にこれらの研究成果とこれまでの成果を統合したタンパク質結晶化状態判定システムについて述べる.

2. 判定対象領域の抽出

2.1 判定対象領域

開発環境の異なる撮像装置から得られた結晶化画像はそれぞれ様相が大きく異なる. そのため各撮像環境に特化した領域抽出手法を適用していく必要がある.

本研究で用いる画像は理研播磨研究所で開発された自動結晶化観察ロボットシステム TERA³⁾により撮影されたものである. Fig.1 に画像内の領域と結晶化ウェルの部位との対応関係を示す. 本研究ではウェルの底面部分のみを判定対象領域とする.

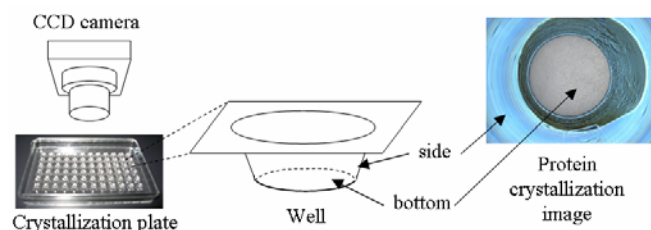


Fig.1 結晶化ウェルと結晶化画像の対応関係

2.2 判定対象領域抽出手法

画像中のウェル底面に該当する円の位置および大きさには多少のばらつきが生じる. そのため画像上の固定位置での対象領域抽出は適用できない. また, 領域内部では結晶化が時々刻々と進行しており, 一つとして同じパターンが存在しないため, 標準パターンを作成することも不可能である. そこで, 析出物の有無およびパターンに依存せず, 且つ画像ごとに対象領域の位置および大きさを逐次決定可能な領域抽出手法を提案する.

TERA で撮影された画像は, ウェルの底面部分は無彩色であるのに対し側面部分は青みがかった有彩色であることがわかる. これを特徴として利用することを考える. 有彩色と無彩色は彩度という属性により表現される. ここで結晶化画像から算出した彩度 S を $[0:255]$ の範囲で正規化した値を階調値とするグレースケール画像を Fig.2 に示す. Fig.2 より底面内部が黒く際立っていることが確認できる. そこで, 底面が十分におさまる範囲である $310-370[\text{pixel}]$ の各半径をもつ円領域を画像中で走査し, 円領域内の彩度分散値が最小となる半径および中心座標を導出することで対象領域に該当する円を決定する.

2.3 判定対象領域抽出実験

1369 枚の画像に提案手法を適用し判定対象抽出実験を行った. その結果, 正しく抽出した画像は実験した 1369 枚中 1269 枚 (92.7%), 抽出不能のエラーを出力した画像が 94 枚 (6.9%), 誤った円領域を抽出した画像が 6 枚 (0.4%) となり, 高い精度での領域抽出を可能とした.

3. 効率的な教師データの再評価手法

計算機によるタンパク質結晶化状態判定では熟練者により割り当てられたスコアを教師データとする. し

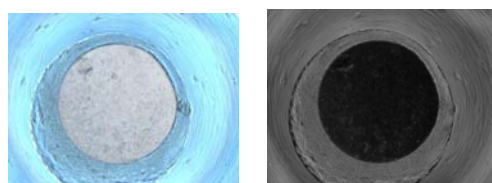


Fig.2 特徴量に彩度を用いた場合の効果

かし、結晶化状態判定では連続的に成長する対象に対して定性的に定義された離散的なスコアを割り当てるため個人による判断基準のばらつきが生じ、また瞬時の判断が求められる大変集中力を要する作業であるため誤判定が生じる場合がある。そのため教師データに含まれる曖昧性を除去できないという問題があった。

そこで、特徴空間中で異なるクラスのサンプルが混在する領域を *overlapping-zone* として特定し、該当領域内のサンプルのみを再評価するという効率的な再評価シーケンスを提案する。これにより、教師データの信頼性向上に貢献すると期待される。

3.1 Overlapping-zone 特定手法

ここでは、特徴空間中のサンプルの混合率を指標として *Overlapping-zone* を特定する。混合率の導出には、ある領域 A 内のサンプル混合率を求める尺度であるエントロピー不純度 $e(A)$ を用いた。

$$e(A) = -\sum_j P(\omega_j) \log_2 P(\omega_j)$$

更にノイズに強く、且つサンプルの分布に柔軟に対応可能な混合率の導出のため、粗い分割から細かい分割までの複数の分解能を用いて特徴空間を分割し、各分解能における部分領域内のエントロピーの和を特徴空間内の混合率とした。本研究では 2×2 から 15×15 までの分解能を用いた。

次に、導出した混合率を用いてその値が高い領域を *overlapping-zone* に特定する。この特定基準には、 k -最近傍決定則を適用した際の識別率を用いることとした。

3.2 Overlapping-zone 適用実験

ここでは一例として、人工2クラスデータを用いて *overlapping-zone* の特定実験を行った。

識別対象データは正規乱数を発生させて作成した2クラス的人工データである (Fig.3(a))。各クラスのサンプル数は500個である。

Fig.3(b),(c)は *overlapping-zone* の特定結果である。図よりサンプルの重なりが大きいクラス境界付近の領域のみを特定できていることが確認できる。また、本実験で特定されたサンプル数は全サンプル数1000個に対してわずか69個であり、効率的な再評価が可能となることがわかる。

4. タンパク質結晶化状態判定システム

前述の判定対象領域抽出手法と *overlapping-zone* 特定手法に加え、我々がこれまでに提案してきた特徴抽出、識別器構築手法を統合し、タンパク質結晶化状態判定システムを実現した。

最終的な状態判定は、判定対象領域内の内接四角形を 3×3 分割した9箇所の矩形領域および内接四角形の外側領域にあたる4箇所の部分領域の計13箇所 (Fig.4) に対してスコアを割り当てることで行った。Fig.4は本システムによる出力判定結果の一例である。スコアごとに色分けし、該当色で部分領域を縁取りして表示した。図中の色は、水色：沈殿の初期状態、青：沈殿の中期状態、赤：結晶状態をそれぞれ表わしている。沈殿と結晶が混在する複雑な対象に対しても詳細な判定

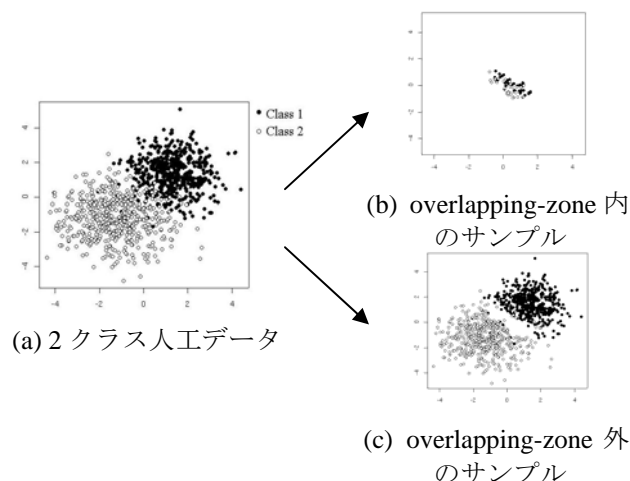


Fig.3 Overlapping-zone 特定結果

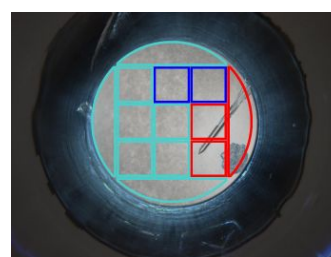


Fig.4 結晶化状態判定システムによる判定結果

結果を出力できていることがわかる。また、*overlapping-zone* 内に位置したサンプルに対しては、誤判定を防ぐために熟練者による評価付けを行うこととした。

5. 結言

本研究は、X線結晶構造解析に必須の処理である結晶化のボトルネックとなっているタンパク質結晶化状態判定処理の自動化を目標としたものである。本稿では判定対象領域抽出手法および教師データの効率的な再評価手法を提案し、更にこれまでの研究成果を統合した結晶化状態判定システムを構築した。

謝辞

理化学研究所播磨研究所宮野雅司主任研究員ならびに菅原光明研究員には、本研究に用いた画像を提供していただくと共に多くの有益なご助言を頂きました。記して感謝申し上げます。

参考文献

- 1) K.Saitoh *et al.* "Evaluation of protein crystallization states based on texture information derived from greyscale images," *Acta Cryst. D61*, pp. 873-880 (2005)
- 2) K.Saitoh *et al.* "Design of Classifier to Automate the Evaluation of Protein Crystallization States," *Proc. of the 2006 IEEE International Conference on Robotics and Automation, Florida, USA, May 15-19, 2006.*
- 3) 菅原光明ら "ハイスループット自動結晶化・観察システムの開発," *蛋白質核酸酵素*, Vol.47, No.8, pp.1026-1032 (2002)