

# ハイコンテンツ解析ソフトウェアのための画像処理技術

## Image-processing Technologies for High Content Analysis Software

大橋 功治 <sup>*1</sup>	伊奈 幸広 <sup>*1</sup>
Koji Ohashi	Yukihiro Ina
荒木 雅裕 <sup>*2</sup>	坂下 浩史 <sup>*1</sup>
Masahiro Araki	Hirofumi Sakashita

ハイコンテンツ解析とは、新薬開発の支援を目的に、細胞又は組織の形状、形態、個数、変動などを、顕微鏡を通して観察し、その様を画像として処理解析し、様々なデータ（例：細胞のサイズ、周囲長、蛍光強度、数量、局在位置、移動速度、総移動距離など）を数値化することである。横河電機は、この分野に特化して独自に開発した画像処理技術を保有しており、その代表例として、画像データ共通化技術、空間と時間の情報を持つ4次元画像解析技術、機械学習を利用した画像認識技術などがある。さらに、これらの技術を組み込んだ解析ソフトウェア（CellPathfinder）も開発している。本稿では、横河が保有する高度な画像処理、画像解析技術について報告する。

High content analysis is an ideal method for drug discovery. It processes and analyzes images of cells and tissues, quantifies the information contained in these images, and outputs useful data. These include the count, size, shape, behavior, circumference, fluorescence intensity, localization, and speed and total distance of movement of cells. For high content analysis, Yokogawa has developed original image processing technologies. These include image data platform technology, four-dimensional image processing technology containing time and space information, and image recognition technology based on machine learning. Yokogawa has also developed the CellPathfinder analysis software using these technologies. This paper describes Yokogawa's advanced image processing and analysis technologies.

### 1. はじめに

ハイコンテンツ解析ソフトウェア（High Content Analysis Software）とは、薬の候補となる化合物を細胞に投与した場合の細胞の変化を画像として捉え、それを処理、解析し、薬効などを分析して化合物のスクリーニングを行い、創薬を支援するツールである。新薬開発で用いられる細胞は形状が多様であり、また、細胞内の様々な小器官が密集しているため、その画像は極めて複雑である。また、顕微鏡システムの技術進歩により、細胞の深さ方向、即ち3次元画像や生きた細胞の時系列画像を容易に取得することができるようになった。

しかしその反面、光学的な要因のため、深さ方向の画質が異なっていたり、データ容量の制約に起因して、時間分解能が悪い画像になったりすることが多い。このような完璧ではない画像群から、創薬に役立つ細胞に関する

様々な数値データを高効率で高精度に引き出すためには、革新的な画像処理技術が必要である。画像処理技術次第で、同じ画像から様々な有用なデータを抽出することができ、その結果、定量的で客観的な評価を行うことが可能になる。

デジタル画像処理の進化と普及の結果、インターネットから汎用的な画像処理ライブラリを入手することができ、2次元画像（平面画像）に対するフィルタリング、モルフォロジー演算<sup>(1)</sup>、ラベリングなど基本的な画像処理を行うことができるようになった。単調な細胞画像に限れば、これらのツールによって、2次元でオブジェクト（細胞）を認識し、種々の特徴量を算出することができる。しかし、実際の新薬開発では、使用する画像群が複雑かつ不完全で、スクリーニングに求められる特徴量も多岐にわたる。市場に流通している画像解析ツールは、そういった要求を満たさないことが多く、また、競争などに対して、機能的に、性能的に差別化できる技術をもっていない。

本稿では、当社が独自に開発したソフトウェアの新製品であるCellPathfinderに搭載した画像処理の差別化技

\*1 計測事業本部 ライフサイエンスセンター 開発部

\*2 計測事業本部 ライフサイエンスセンター 品質保証・CS部

術の例を3つ紹介する。

## 2. 画像データ共通化技術

### 2.1 画像データ共通化技術の必要性

これまで当社では、細胞生物の基礎研究から創薬支援の応用開発まで、幅広い分野で用いられるバイオイメージング装置を製品化してきた。機器を制御するソフトも、取得した画像を解析するソフトも、それぞれ分野ごとの用途に合わせて、専用品を開発してきた。専用品には装置に特化した画面を作りやすいメリットがある。しかし、逆に機器間の使い勝手が異なるため、ユーザーにソフトウェアの操作を覚えることの負担をかけてしまうこと、システムとして運用する場合に他の装置間とのインターフェースを新規に開発しなければならないこと、というデメリットもあった。この問題を解決するための手段として、画像解析ソフトウェアを1本化することを考え、そのために画像データの共通化技術を開発した。

一般に JPEG, PNG フォーマットなどの画像ファイルを画像データと呼ぶが、ここではこれらに加えて、撮像に使用した顕微鏡の倍率、カメラの露光時間などの撮像条件も含めて、計測データと呼ぶ。計測データの共通化の流れを図1に示す。解析を行うためには、様々な装置ごとの計測データから解析に必要な情報を抽出するとともに、共通化を行う必要がある。共通化を行ったデータを、計測メタデータと呼ぶ。計測メタデータを生成する処理をプラグイン化することにより、既存の装置や今後の新しい装置などで簡単に拡張できるようにした。

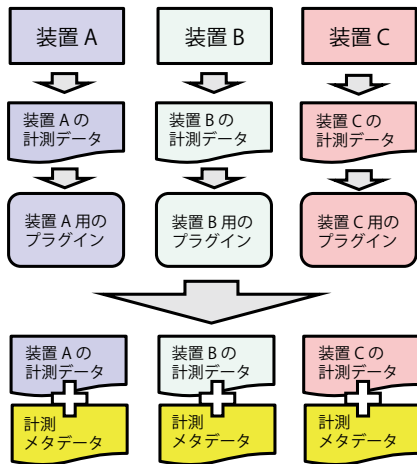


図1 共通化の流れ

### 2.2 共通化で抽出する主な情報

共通化では、撮像した画像から正しい解析結果を得るために、以下の情報を計測データより抽出し、計測メタデータとして記録する。

#### (1) 計測値の変換

(A) 時間変化：図4の撮像のタイムポイントを経過時間

である時間単位「秒」に変換する。

(B) X, Y 方向の実距離：カメラの画素単位「Pixel」から長さの単位「 $\mu\text{m}$ 」に変換する。この際、図2のように、対物レンズ倍率、カメラの画素の大きさ、ビニングの有無を用いる。ビニングとは付近の画素をまとめて、明るさを上げる代わりに解像度を下げる手法のことである。

$$\mu\text{m}/\text{Pixel} = (\text{カメラの1画素の大きさ} \times \text{ビニング}) \div \text{対物レンズ倍率}$$

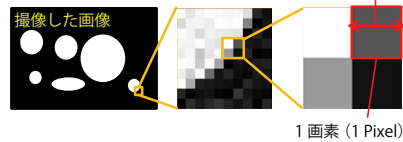


図2 X, Y 方向の実距離の計算

(C) Z 方向の実距離：対物レンズの移動距離から撮像面が細胞内を移動した距離を計算する。この際、図3のように、対物レンズのタイプによる液浸物質の屈折率（水や空気など）、ウェル内の媒質（培養液の屈折率）を用いる。

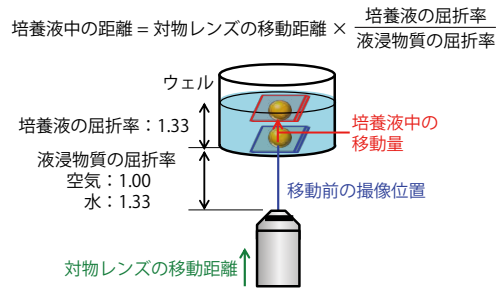
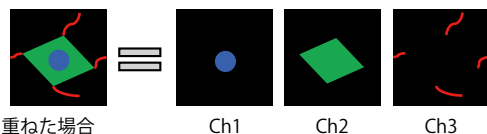
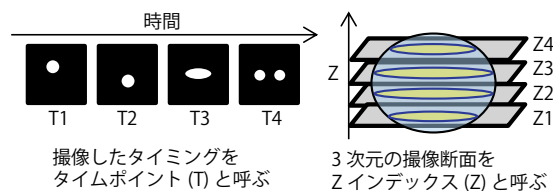
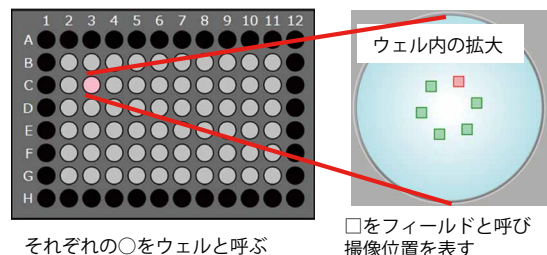


図3 Z 方向の実距離の計算



同一の位置（ウェル，フィールド，タイムポイント，Z）を異なる波長で撮像したものをチャンネル（Ch）と呼ぶ

図4 画像の意味付けについて

(2) 画像の対応付け

画像と正しい位置情報や波長などを関連付けることは、薬剤における変化を解析する上で非常に重要となる。具体的には、**図 4** に表すウェル（培養容器の番地）、フィールド（ウェル内の撮像位置）、タイムポイント、Z インデックス（Z 位置情報）、チャンネル（蛍光の色情報）をそれぞれの画像と対応付ける必要がある。また、容量の大きな画像ファイルは変換を行わないようにし、画像の対応付けと相対パスを記録する。

2.3 計測メタデータの実装

計測メタデータの実装には、C# 言語から利用可能な関係データベース管理システム（RDBMS: Relational Database Management System）である System. Data. SQLite<sup>(2)</sup> を使用した。1 度の計測において数万枚の画像（データ容量で数 GB～数 TB）の大量の画像データを撮像しても、RDBMS の効果により高速なアクセスが可能となる。RDBMS ではデータをテーブルという単位に分解し、それぞれ関連付けて管理する。具体的には、**表 1** に示す情報を保持している。

表 1 計測メタデータの保持する情報

保存する情報	DB 内のテーブル名
メタデータ作成時の基本情報	MetaDBInfo
計測データの基本情報	MeasurementInfo
計測したウェルプレート情報	SampleContainer, SourceContainer
計測条件 (位置、時間、波長、 対物レンズ、カメラなど)	Well, Timelaps, MeasurementBlock, MeasurementWell, Channel, ObjectiveLens, Camera, EmissionFilter, LightSource, LightSourceParameter
画像情報	Image, ImageSize, Projection
その他 (ログなど)	FocusValue, MeasurementLog, OtherFile

共通化した計測メタデータを解析ソフトウェアが読み込むことで、装置間の違いを意識することなく解析することができるようになる。計測メタデータの共通化により、様々な装置への対応が容易になり、横河の開発効率が向上した。このことにより、新機能の追加やユーザからの要望への対応へもスムーズにでき、ユーザの満足度の向上に繋がっている。

3. 4次元画像処理技術

3.1 4次元画像処理とは

生き物である細胞は常に何かしら変化しており、ライプセルの時系列観察では、タイムポイント毎に固定したサンプル観察からは得られない知見を与えてくれることが多い。また、細胞の動的な振る舞いを理解することは、発生物学のゴールの一つであると言われている。したがって、4次元画像処理の重要性はますます高まっている。

3次元画像処理とは、Z スタック画像（平面画像を Z

方向に積み重ねた画像群）から 3 次元（XYZ 成分）でオブジェクト（細胞）を認識し、種々の特徴量（形状情報、位置情報、輝度情報など）を算出することである。3次元画像処理の特徴として、2次元画像処理や MIP (Maximum Intensity Projection) 画像処理では得ることができない Z 成分を加味した細胞の位置情報や形状変化などを、数値化することができる。

4次元画像処理では、Z スタック画像に時系列の次元を加えた情報を入力とする。各タイムポイントにおいて、3次元（XYZ 成分）でオブジェクト（細胞）を認識し、各タイムポイントで認識したオブジェクト同士を関連付ける。そこで、3次元画像処理で得られる結果に加え、時系列に関する特徴量（移動速度、総移動距離など）を出力する。その概念を**図 5** に示す。

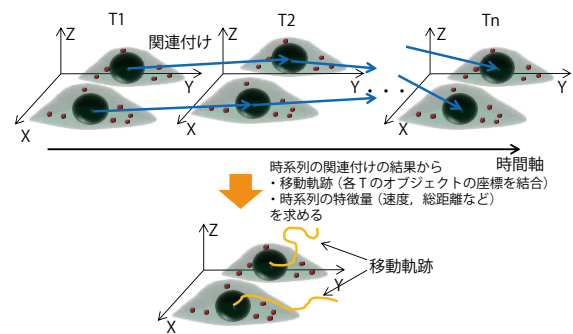


図 5 4次元画像処理

3.2 時系列の関連付け方法

4次元画像処理の性能及び精度は、3次元画像処理における認識精度と時系列の関連付け方法で決まる。3次元画像処理の認識では、5つの処理であるフィルタリング、2値化、2値化画像変形、ラベリング、ラベル画像変形を実行する<sup>(3)</sup>。

時系列の関連付け方法については、全タイムポイント (Ti, i=2, 3, ..., n) において、タイムポイント間 (Ti-1 と Ti) で認識した 3次元オブジェクトに対して、近さを評価し、最近傍のオブジェクトを関連付ける (**図 6**)。近さの評価関数には、3次元オブジェクトの距離、形状変化、分裂の有無、分裂前後のサイズ、動きの予測値の重み付き和を利用する。また、重みを可変にして、細胞追跡方法に柔軟性を持たせることにより、多種多様なライブセルイメージングに対応している。

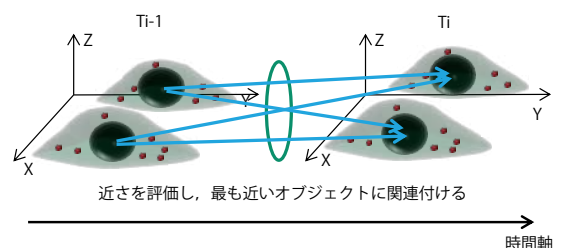


図 6 時系列の関連付け方法

#### 4. 機械学習技術

近年、急速に機械学習技術が発展し、画像認識の分野においては、人力を凌駕する認識結果が得られるようになってきた。我々は、この技術に着目し、バイオイメージングの分野においても、良質な結果を得ることができると考えて導入している。

機械学習とは、データ集合を学習し、未知のデータについて推定することであり、大きく分けて2つのステップがある。一方は、与えられたデータ集合から特徴（パターン）を見付け出すための学習ステップである。他方は、学習した結果を利用して、未知のデータを分類する推定ステップである。

学習ステップで用いるデータには、分類ラベル（答え）が付与されている場合と付与されていない場合がある。ここでの機械学習では、学習ステップで用いるデータに、分類ラベルが付与されている教師ありデータを利用して

##### 4.1 機械学習フィルタの概要

バイオイメージングにおいて、機械学習の適用方法はいくつか存在する。次節で説明する特徴を持たせるために、撮像画像から多値画像への画像変換処理に、機械学習を適用する。ここではそれを機械学習フィルタと呼ぶ。機械学習フィルタの出力である多値画像の各画素値は、分類ラベル（正の整数値：1, 2, 3, …）を表す。

図7に、機械学習フィルタの概念図を示す。機械学習フィルタの学習ステップでは、ユーザは学習用撮像画像において、各分類ラベルに対応させる場所をクリッピングし、それを学習データ（教師ありデータ）として与え、学習させる。図7では、分類ラベル1, 2, 3, に核, 細胞質, 背景(核と細胞質以外)をそれぞれ割り当てている。一方、機械学習フィルタの推定ステップでは、その学習結果を基に、推定対象画像の各画素に対して、分類ラベルを推定し、多値画像に変換する。

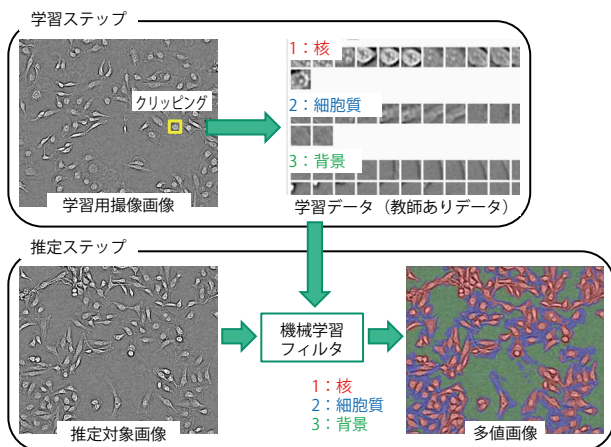


図7 機械学習フィルタの概念図

##### 4.2 機械学習フィルタの特徴

バイオイメージングにおいて、細胞認識の精度は重要なファクタである。細胞画像は極めて複雑で、多様な生体解析に対応するためには、機械学習の性能だけでは細胞認識の精度に限界がある。しかし、機械学習フィルタには、細胞認識の精度をさらに向上させるための機構が2つある(図8)。1つ目は、機械学習の推定精度を上げるためのプレ処理である。2つ目は、機械学習の推定ミスフォローするためのポスト処理である。

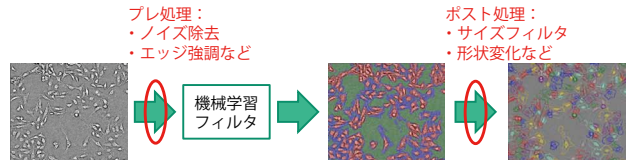


図8 機械学習フィルタのプレ・ポスト処理

プレ処理では、画像内のカメラノイズ除去や特定情報を強調するためのエッジ強調などの処理があり、撮像画像によって、ユーザは任意に処理を組み合わせることができる。プレ処理の実行により、学習データを洗練・強調し、学習ステップで学習データから特徴（パターン）を見付けやすくする。それにより、推定ステップでの推定精度を上げることができる。

ポスト処理では、細胞認識後のサイズフィルタや形状変化などの処理があり、プレ処理と同様に任意に組み合わせることができる。たとえ、機械学習の推定ミスにより細胞の誤認識が起きたとしても、ポスト処理の実行により、その誤認識を解消することができる。結果として、誤認識が少ない細胞認識が可能となる。

#### 5. おわりに

本稿では、顕微鏡画像をベースとした高度な画像処理技術を紹介した。当社は、これらの技術を解析ソフトウェアに組み込み、CellPathfinderとして製品化し、基礎研究から創薬応用まで幅広い分野の研究に貢献している。さらに、現行技術に留まらず、新しい技術の獲得及び洗練を図り、当社のハイコンテンツ画像処理技術の基盤を築き上げていきたい。

#### 参考文献

- (1) 小畑秀文, “モルフォロジー”, コロナ社, 1996
- (2) SQLite, System.Data.SQLite Home, <http://system.data.sqlite.org/index.html>
- (3) 坂下浩史, 大橋功治, 他, “共焦点定量イメージサイトメータとそのバイオアプリケーション”, 横河技報, Vol. 58, No. 1, 2015, p. 29-33

\* 文中の商品名及び名称は、各社の登録商標または商標です。