

# ビッグデータ時代の高速な知の発見 アルゴリズムの追究

## 古賀 研究室



古賀 久志  
Hisashi KOGA

「データ」を解析しています。これに対して、古賀久志准教授は、データマイニングの手法を「画像」や「動画」の解析に適用した、新しい研究に取り組んでいます。

日々増え続ける多様かつ膨大な情報は「ビッグデータ」と呼ばれ、その活用の重要性が年々高まっています。ビッグデータの解析方法の一つに、高度なアルゴリズム(コンピュータを使った計算手法)に基づいて大規模なデータから有用な知識を取り出す「データマイニング」技術があります。

しかし、データマイニングは通常、膨大な顧客データや多様なセンサデータなどを扱うことが多く、大抵は「数値データ」や「文字

近年、ユーチューブなどの動画共有サイトが普及し、自分で撮影した画像をインターネット上に投稿する例が増えていきます。こうした背景から、ネット上の膨大な画像データにどんなものが多いというニーズも高まっています。古賀准教授は、データマイニングの手法を使って画像処理を行い、このような要求に応える技術を開発しました。

その中核となる研究が、繰り返し出現するパターンを認識して抽

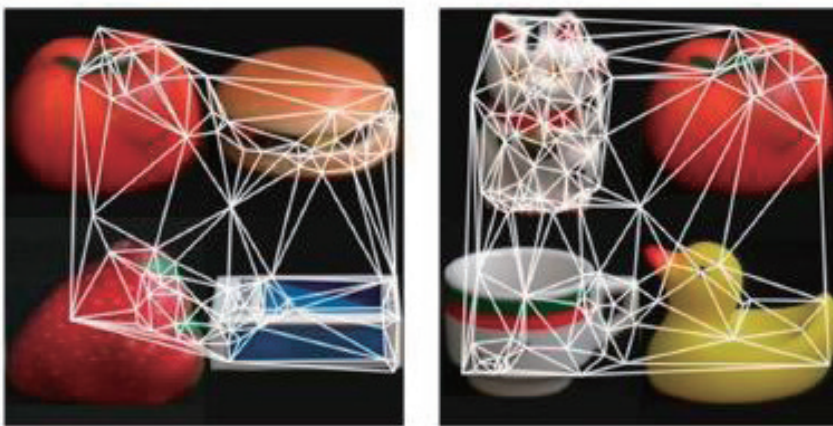


図1 画像のグラフ表現

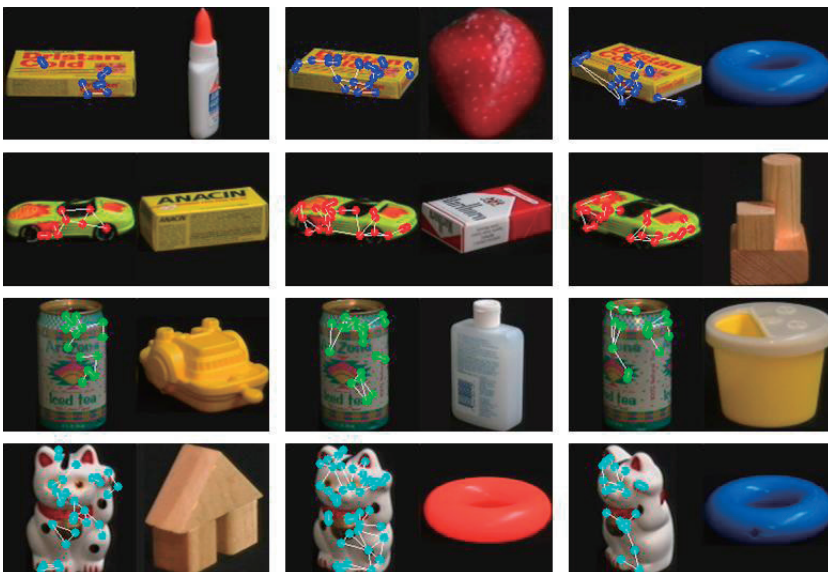


図2 オブジェクト抽出結果

### キーワード

適応知能情報システム、マルチメディア処理、ネットワーク情報システム、大規模データの高速処理、静止画像・動画像解析、画像マイニング、ウェブマイニング、インターネットの適応制御、確率アルゴリズム、学習アルゴリズム

|        |                              |
|--------|------------------------------|
| 所属     | 大学院情報理工学研究科<br>情報・ネットワーク工学専攻 |
| メンバー   | 古賀 久志 准教授                    |
| 所属学会   | 情報処理学会、<br>電子情報通信学会          |
| E-mail | koga@is.uec.ac.jp            |

出する「画像処理アルゴリズム」の研究です。例えば、一つの画像に複数の物体(オブジェクト)が写っている大量の画像データの中から、同じオブジェクトが写っている画像だけをすべて見つけるという処理が可能です。

実験では、約300枚の画像データから、トマトが写っている異なる画像を取り出すことができました(図1)。また、特定のオブジェクトが繰り返し現れる様子を、画像上に点で表示することもできます(図2)。応用としては、ネット上に公開されている大量の画像データから、特定の製品が写っている画像だけを選択的に取り出すことなどが想定できるでしょう。

そのほか、監視カメラが撮影した動画から「人の部分だけを抽出して解析することなどもできます」ところで、これらに類出するのは「背景」です。背景は動画の中で常に存在しているため、背景のパターンを認識し、それを全体の動画から取り除けば、人を抽出することができます。まず、動画を複数の画像として扱い、一枚の画像を色の違いに応じて複数の領域

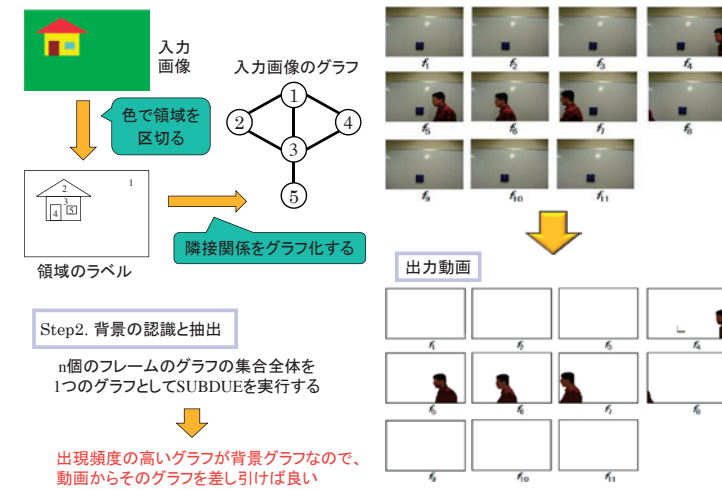


図3 背景除去手法

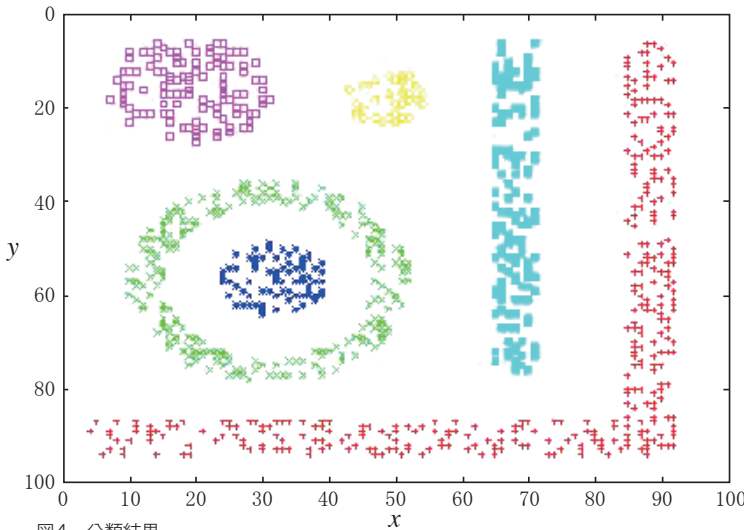


図4 分類結果

計算時間の比較

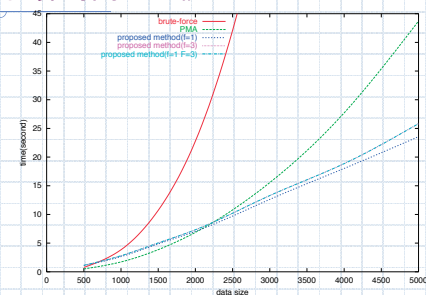


図5 データ数と処理時間

に区切ります。各領域に番号を振り、隣接する関係を考慮すると、一枚の画像がグラフとして示されます。この作業を複数枚の画像について行った後で、最も頻出するパターンを抽出すると「背景」が発見できます。実際のデータに適用し、人の動画を明瞭に取り出すことに成功しました。

もう一つの重要なテーマが、ビッグデータから似ているデータを一つのクラスター(集団)として分類する、パターン認識技術に応用

した新しいアルゴリズムの研究です。膨大なデータをいかに分類するかという類似検索技術の一種であり、「ビッグデータの高速な解析に役立つ」と古賀准教授は考えています。

類似する点を高速に見つける、既存の局所鋭敏ハッシュアルゴリズム(LSH)をデータの分類に初めて適用しました。従来はn個のデータがあった場合、すべてのデータ(点)間の距離を計算していたた

め、その計算量は $n^2$ と、nに応じて爆発的に増えていました。これに対して、LSHを効果的に応用し、似ている可能性があるデータ同士に対してのみ距離計算を行ってデータを分類することにより、計算量を「n」まで削減できることが分かりました。これによって、ビッグデータの処理を大幅に高速化することができます。

産業界にとってビッグデータはまさに「宝の山」です。しかし、有用なデータも、それを高速に解析

して活用しなければ「宝の持ち腐れ」になりかねません。企業は最新の研究成果を貪欲に取り入れていく必要があります。そのため、古賀准教授は「最新のアルゴリズムを駆使して、実世界のビッグデータ解析に応用できる技術の開発に力を入れている」のです。今後はLSH技術を拡張し、通常のデータだけでなく、より複雑な「グラフ」同士の類似度まで検索できるように、新たな研究のステージを目指しています。

【取材・文】藤本信穂