

研究テーマ

対象を限定しない一般画像・映像認識、Webからの視覚的概念の自動獲得、画像と言語概念との対応付け、位置情報と画像の意味の関心の分析およびそれを利用した画像認識、Web画像マイニング

キーワード

画像認識、画像理解、物体認識、機械学習、情報検索

所属	電気通信学部 情報工学科
メンバー	柳井 啓司 准教授
所属学会	情報処理学会、電子情報通信学会、人工知能学会、IEEE Computer Society、ACM
E-mail	yanai@cs.uec.ac.jp
研究設備	計算専用コンピュータ(140コア) 50台、データストレージ 約45TB



柳井 啓司  
Keiji YANAI

研究概要

言語に結びついた一般的な画像・映像の認識

インターネットが世界的に普及したことで、画像や映像を共有するサービスが登場し、多くのユーザーが利用している。

当研究室では、Webサイトにアップロードされている画像や映像を使って、対象を限定しない一般的な画像・映像をマイニングして、多種多様な用途に利用できるように、さまざまな研究を進めて

いる。

まず、画像認識に関して、言語と画像の関連付けを研究している。この技術は、言語に対応する画像、逆に画像に対応する言語を割り出すというものだ。もちろん、Googleでも画像検索は行えるが、これはあくまでも画像に付いているタグ(テキスト)情報を検索しているだけだ。当研究室が行っているのは、言語による画像の特性を理解し、画像を見てその特徴を判断するものだ。これを

実現するために、最初に名詞に対応する多くの画像をWebから自動収集し、その特徴の解析を行った。具体的には、画像検索エンジンから「ラーメン」という名詞を検索すると、多くの画像を集められ

る。その画像をマイニングする

と、どんな緑とチャーシューが共通の特徴であることが分かる。このように、多くの名詞に対して同様の作業を行い、名詞に対する画像特性を集めている。

次に、黄色い・黒い、堅い・柔らかいといった形容詞について研究している。これには当研究室で規定したエントロピーを使って、6000種類の名詞に対してランキングを実施し、それぞれの特性を集めた。

この2つの認識手法を使ったシステムとして、企業と共同で研究開発した「食べ物認識システム」がある。これは、食べ物の名前から集めた画像をマイニングして得られた特徴的なパターンである局所

パターンと、形容詞である色の認識を、独自の比率でミックスすることで実現した認識システムだ。

この手法を使えば、画像から50種類の食べ物を50%以上の認識率で判断できる。この認識率は、国内でもトップレベルと言える。この技術を使って、ユーザーは携帯で食べたものを撮り続けるだけで、1日の平均摂取カロリーや1週間に食べたものの傾向などを表示させるなど、さまざまな利用方法が考えられる。

最近始めたものとして、単語に位置情報を加えた画像認識がある。Flickrなどの画像投稿サイトでは、画像に加えて位置情報も得られる。Web画像自動収集の際、この情報も集め、比較を

行う。例えば「hoodie」と入れて検索すると、日本ではパーマなどに対して、ヨーロッパはバスタが多いという世界各地の傾向や差異が分かる。今後、いろいろな単語について検索を行い、国別の特性を割り出してみたいと考えている。

当研究室では、映像に関してもさまざまな研究を行っている。地上波デジタル放送の付加情報(テキスト)を使って必要な箇所だけを切り出したり、YouTubeなどの動画サイトから大量の画像を集めて、名詞による画像特性抽出のところで説明したような作業を映像でも行えるよう研究している。

映像解析の成果として、指定の

映像から目的の映像を自動的に切り出すアルゴリズムを競うTRECVID(国際映像処理ワークショップ)のコンテストで、国内第2位の成績を収めている。

アドバンテージ

より人間が考えていることに近い画像・映像認識を研究

当研究室で研究している画像認識は、あくまで人間の目を見た感覚にいかにか近づけるかに主眼を置いている。今、デジタルカメラに

多く搭載されている顔認識機能は、対象を顔に限定しているのだから、比較的簡単に実現できる。しかし、当研究室で行っているのは、写真に写っているものすべてが対象となる。例えば、通常のポートレートでは人の顔以外にも、背景にある風景や建物、テーブルの上の料理、足元にいる犬などさまざま

なものと同時に写っている。このように、対象を絞り込まない認識ができることは大きなアドバンテージだと言える。

今後の展開

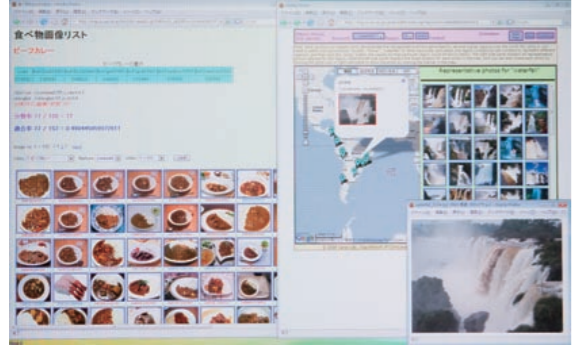
一般的な単語に対応する画像認識を深めたい

当研究室で現在研究中の言語と画像の認識について、もっとい

いたな単語について対応していきたい。言語は非常に幅広いので、一般的な単語に対応する画像認識ができれば利用範囲は非常に大きくなるかと考えている。また、食べ物の認識で使っている手法は画像認識の精度が高いので、そのアルゴリズムを使って効果的に応用していきたいと思っている。

また、単語と画像の関係に加えて、位置情報を合わせたものに関して、さらに多くのサンプルを集めて、より有用なものにしたい。

例えば「城」という単語一つとっても、日本人とアメリカ人では思い浮かべるイメージがずいぶん違う。このように、言葉によって認識される画像の違いを分析・理解することで、国や地域ごとの特徴を割り出すことができることから、これをうまく利用すればさまざまな分野に応用できると考えている。



食べ物認識システム(左)および位置情報付き画像にマイニングシステム(右)



140コアのクラスターシステム



大容量ストレージ