

特集「大規模画像データ処理」にあたって

谷口 倫一郎
(九州大学)

和田 俊和
(和歌山大学)

近年、大量の画像を対象とした検索や処理に関する研究が盛んに行われるようになってきている。この背景としては、1) SIFT (Scale-Invariant Feature Transform) などに代表される画像の安定な局所特徴記述の開発、2) この画像特徴をベクトル量子化したものをビジュアルワードとみなして、テキストマイニングで用いられる Bag of Words のフレームワークを画像に適用する Bag of Features (BOF) のアプローチが検索や認識で成功したこと、3) さらに LSH (Locality-Sensitive Hashing) をはじめとする近似最近傍探索技術が進んだこと、4) 著作権保護などを目的とした効率的な画像・映像検索に対する社会的ニーズが高まっていること、5) 一般物体認識にとどまらず、画像とそのアノテーションデータとの相互変換の可能性が見えてきたこと、などがあげられる。本特集では、これら大量画像を対象とした検索を支える基盤技術とその応用を紹介する解説として、以下の 5 編を掲載する。

(1) 「画像局所特徴量 SIFT と最近のアプローチ」:

藤吉 弘亘 (中部大学)

画像データそのものは極めて情報量が多いので、大量の画像データを効率的に処理しようとする、何らかの特徴抽出が必要不可欠である。当然のことながら、抽出された特徴の善し悪しが、画像処理の性能に大きく影響を及ぼすことになる。画像処理研究が始まって以来、数多くの特徴量が提案されてきたが、David Lowe が 1999 年に提案した局所的画像特徴である SIFT (Scale-Invariant Feature Transform) は、画像のスケールや回転の変化に頑健であるという優れた性質をもっているため、非常に良く利用されている。最近では、SIFT の考え方をもとに、画像特徴抽出をさらに高速化、高精度化する手法も数多く提案されている。本解説では、SIFT 特徴の基本的な考え方を説明するとともに、代表的な高速化、高精度化の手法について紹介する。

(2) 「高次元空間における近似最近傍探索技術の進歩とその展望」:

和田 俊和 (和歌山大学)

最近傍探索は、Bag of Features における特徴ベクトルの生成や、画像検索や認識などにおいて主要な役割を果たす基本的な技術である。特に画像のように高次元の特徴ベクトルを大量に扱う場合には、通常の線形探索を行うと長い処理時間を要してしまう。このため、高速な最近傍探索アルゴリズムが数多く考案されたが、高次元のデータ分布に対しては線形探索と同等の処理時間が必要となることがわかってきた。そのため、正確な最近傍探索ではなく、ある程度の誤差を許容する近似最近傍探索の手法が研究されている。これら近似最近傍探索手法のうち LSH は計算量と誤差の期待値の関係が明らかであり、Random Forest などほかの手法を包含する枠組みである。本解説では、これらの手法に関して概略を説明するとともに、著者らの考案した LSH の変種である PCH (Principal Component Hashing) をはじめとする最近の近似最近傍探索の手法について紹介する。

(3) 「局所特徴量を用いた画像照合による特定物体認識」:

黄瀬 浩一 (大阪府立大学)

特定物体認識とは、画像内にある特定の物体 (物体のインスタンス) を認識する処理である。換言すれば、物体の「見え」から物体 ID への変換を意味する。これが実現できれば、画像中の物体を Web のリンクアンカーとして用いることができ、Web と実世界をリンクすることが可能となる。Google Goggles に代表されるように、このような動きはすでに部分的ではあるが現実のものとなりつつある。我々の周囲には無数の物体インスタンスがあるため、特定物体認識の実現は、規模との戦いであるといえる。本解説では、先の二つの解説をもとに、局所特徴量とそれを用いた Bag of Features 表現を基礎とする特定物体認識の基本的アプローチを説明する。さらに、特定物体認識における最近の動向について紹介する。

(4) 「大量映像時代が拓く映像コピー検出技術の新展開」:

佐藤 真一 (国立情報学研究所)

映像コピー検出技術とは、Webなどの映像アーカイブや放送映像のような映像ストリーム中から、データベース中の映像と全く、もしくはほぼ同じ映像を検出する技術であり、コマーシャル映像検出・同定、違法コピー映像検出などを目的として、比較的古くから研究されてきている。最近では、マルチメディアやコンピュータビジョン系の国際会議で多くの発表が見られるようになってきている。これらには、大量の映像から頑健にコピー検出を行うため、SIFTを代表とする局所特徴量による頑健な照合、時系列情報の利用による照合精度の向上などの技術革新が見られる。また、映像アーカイブに対して映像コピー検出技術を適用し、検出パターンを解析することによって、映像ランキング、ドキュメンタリー映像自動生成、重要映像マイニングなどを行う新しい技術も開発されてきている。本解説では、これら映像コピー検出に関連する最近の研究動向について概観する。

(5) 「大規模画像データセットを用いた自動画像アノテーション」:

原田 達也 (東京大学)

自動画像アノテーションとは、ラベル(タグと呼ばれることも多い)の付与されていない画像群に計算機が自動的にラベルを付与する処理である。自動的なラベル付与を実現するためには画像とラベル間の対応関係の学習

が課題となるが、Webに存在する大規模なデータを活用することで、自動的に知識を獲得し、多様な画像に対して自動的にラベルが付与できると期待されている。本稿では、自動画像アノテーション手法の最近の動向と大規模データへの適用法について紹介する。画像アノテーションは、画像とそれに付随するラベルペアのデータセット構築、画像とラベルの特徴抽出、画像とラベルの関係性モデルの構築、モデルを用いた画像のラベル予測からなっている。つまり、データ、画像特徴、モデルが画像のラベル予測性能に貢献するため、これら三つのサブテーマを取り上げる。また、自動画像アノテーションを大規模画像データに適用するには、システムのスケラビリティが重要となるため、それについても言及する。

最近では、インターネット上や大容量記憶装置上に大量の画像データが蓄積されるようになってきており、これらの画像データをどのように有効活用するかというのはこれからの重要な課題である。携帯電話のほとんどにカメラが搭載され、簡単にインターネット上に画像(ビデオ映像を含む)が蓄積されるようになってきており、今後、インターネット上の画像の量は飛躍的に増大するものと思われる。一方で、画像は低レベルの情報の集合でもあるので、大量の画像データから得られる情報を構造化し、意味のある情報を抽出することは難しい問題である。本特集が、このような課題に取り組もうとする方々にとって、少しでも役に立てば幸いである。