

審査の結果の要旨

氏 名 張 拳石 (ちゃん・ちゅあんし)

極めて大量・多様な画像データがネット経由で利用可能になって久しく、画像データへのラベリングは大きな研究課題となっている。画像ファイル一つ一つについて、「山岳風景」「相撲」「野球」などのラベルを付与することを目的とした研究は多数存在する。しかし、画像には自動車、机、椅子、人物、動物などのさまざまなオブジェクトが含まれている。こうした個々のオブジェクトへのラベリングが実現すれば、自然風景の中に含まれているシカの画像を検索できるなど、一層精緻な画像検索が実現するばかりでなく、画像中のオブジェクトの認識や追跡などの画像解析の高精度化にも大きく貢献する。

しかし、ラベルを付与するためには対象物のカテゴリモデルが必要であり、そのモデルを獲得するために、従来の多くの研究では人がトレーニングデータを与えることが前提となっている。膨大で多様な画像データを対象に人力でトレーニングデータを作成することは現実的ではないが、トレーニングデータなしにラベル付けのためのカテゴリモデルを獲得する方法について研究が進んでいない。

本研究は多数の画像を対象に、その中に含まれるさまざまなオブジェクトを抽出し、カテゴリモデルを、トレーニングデータの付与を基本的に不要としたうえで自動的に取得する手法、すなわちオブジェクトレベルのビジュアルマイニング手法の開発を目的としている。

第1章はイントロダクションであり、上記のような研究の背景と目的を述べている。

第2章は、本研究で開発されたビジュアルマイニングの手法のうち、基礎的な考え方とアルゴリズムを記述している。具体的には画像の中に多数存在するローカルな特徴量とその相互関係からなるグラフ構造により、画像に含まれる情報を表現する。このグラフを *Attributed Relational Graphs (ARGs)* と名付け、カテゴリモデルを特定の属性や構造を持った ARG で表現する。同時に、同じカテゴリに属するオブジェクトでもさまざまな色のものが存在するので、それを表現するために、ARG の属性はある幅をもって変動すると仮定している。すると、同じカテゴリに属するオブジェクトを含んださまざまな画像は、ある共通の部分 ARG を含んでいるはずで、それがそのカテゴリのモデルに対応すると考えられる。その場合には特定の部分 ARG が共通に存在するという条件のみを利用して該当するカテゴリモデルを見出しているため、人が介入してトレーニングデータを与える必要がない。本章ではさらに、共通の最大グラフを発見するというアルゴリズムを定式化し、RGB 画像や RGB-D 画像から作成された ARG に適用してその性能を従来の方法に比較して評価している。その結果、従来のグラフマッチングの方法に比べ、概ね 10%以上精度が向上し、全体として 9 割以上のマッチング精度を達成した。

第3章は、ARG のマッチングに基づく教師なし学習手法について提案している。サブ ARG

で記述されるカテゴリーモデルを共通の最大グラフとして抽出することで教師なし学習を実現するというのが基本的な考え方であるが、マッチングにおいては、単なるマッチングパラメータの調整に留まらず、カテゴリーモデルの基礎となる共通のグラフを学習過程において適宜調整し学習精度の向上を図っている。また、学習対象となるカテゴリーモデルが含まれている「正」の学習データだけでなく、モデルの含まれていない「負」の学習データをも与えてモデルの調整を行う点にも特徴がある。この方法を用いて学習におけるグラフマッチングと、学習したモデルによる認識精度の比較実験を行っており、従来の方法に比べ同等から10~20%の精度向上が実現していることを確認した。

第4章は、ここまで開発された一連のビジュアルマイニング手法をRGB-D画像に適用した例を示している。奥行き(D)次元を持つRGB画像はRGB画像に比べ情報量が多く、空間的な構造などの把握が容易である。そのためRGB-D画像を利用してARGによるカテゴリーモデルを取得し、そのモデルをRGB画像に適用することでオブジェクトのカテゴリー認識を行う方法、あるいはRGB-D画像から得られたモデルを、RGB画像のみから得られたモデルの修正に利用し、修正後のモデルで認識を行う方法等である。これらについても実証実験を通じて、既存方法に対して精度が向上していることを確認した。

第5章は手法の応用(その1)であり、単一のRGB画像から得られる2次元情報から3次元構造を推定する方法を実現している。すなわちRGB-D画像でカテゴリーモデルを学習しておき、そのモデルをRGB画像に適用してカテゴリー認識をすることで、3次元構造を推定する方法である。

第6章も手法の応用(その2)であり、動画像中に撮影されている走る動物などの変形モデルの発見と学習手法を開発している。まずARGを構築するにあたり、SIFT(Scale Invariant Feature Transform)による特徴点を抽出してグラフノードとする。そして動画の最初のフレームにたいして最初のグラフテンプレートを構築することで、それ以降の各フレームに対してのマイニングを行い、変形モデルのマイニングを実現している。

第7章は、本研究で構築された約1200カテゴリーのオブジェクトに対するRGB-D画像アーカイブについて記述している。これはカテゴリーモデルの構築手法の比較研究等を支援する研究基盤となり得る。

第8章は結論と今後の課題を述べている。

以上まとめると本論文は、グラフ表現を利用した画像情報のモデルを利用して、画像中に点在するオブジェクトの分類(カテゴリー化)を教師なしに行う手法を開発し、その精度も実験的に確認したものである。カテゴリーモデルをグラフ表現し、取り得る属性にも幅を持たせているため柔軟な学習が可能である。またRGB-D画像から学習したモデルをRGB画像でのオブジェクト抽出や分類に適用できるなど幅広い応用が可能であることを示した。手法の新規性はきわめて高く、検証実験などもさまざまな画像を用いて丹念に行っており、空間情報工学やコンピュータビジョン分野に多大の貢献をしている。よって本論文は博士(工学)の学位請求論文として合格と認められる。