

令和元年6月11日現在

機関番号：82626

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K12455

研究課題名（和文）物体の分かりやすい説明表現のための絵描き歌自動生成に関する研究

研究課題名（英文）Research on automatic generation of drawing songs for easy-to-understand object descriptions

研究代表者

金崎 朝子（Kanezaki, Asako）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：00738073

交付決定額（研究期間全体）：（直接経費） 2,600,000円

研究成果の概要（和文）：写真に写っている物体を分かりやすい表現で説明することを目指し、物体認識技術を応用した絵描き歌自動生成システムの要素技術開発に取り組んだ。未知の物体を言語で表現するとき、人はよりイメージしやすい物体を用いて「 \times のような」「 \times を \times に乗せたような」といった比喩的表現を生成する。これを機械が実現するために、様々な角度から物体を撮影して一般的な名称を推論する物体認識技術の開発、および事前知識を使用せず、画像内の物体をパーツに分解する教師無し画像セグメンテーション技術の開発を行った。

研究成果の学術的意義や社会的意義

第一に、様々な角度から物体を撮影し、一般的な名称を推論する三次元物体認識技術を開発した。本研究は深層学習を用いており、物体の姿勢の教師信号を人間が与えることなく、自動的に獲得できる点が新しい。提案手法は三次元物体検索の国際的コンペティションSHREC'17にて、二部門で世界第一位の性能を記録した。第二に、画像内の物体をパーツに分解する教師無し画像セグメンテーション技術を開発した。深層学習を用いた画像セグメンテーションの研究例は多く存在するが、そのほとんどが教師あり学習である。提案手法は畳み込みニューラルネットワークを用いるが、事前学習等の学習を一切必要としない点が新しい。

研究成果の概要（英文）：With the aim of explaining unknown objects in images with easy-to-understand expressions, we tackled to develop elemental technology for the automatic generation of drawing songs based on general object recognition. In order to describe an unknown object, humans generate a metaphorical expression such as "Like xxx" or "Like placing xxx on yyy", using some common and imaginable objects. In order to develop such a system, we proposed a novel 3D object recognition method using multi-view images. We also proposed an unsupervised image segmentation method that decomposes unknown objects in images into separated regions without using any prior knowledge.

研究分野：画像処理

キーワード：物体認識 画像セグメンテーション 機械学習 マルチビュー画像 姿勢推定 教師なし学習

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

物体のイメージを伝達するとき、人は言語を使用する。ここで物体のイメージを正しく伝達するためには、物体を表現する言語は分かりやすいもの、すなわちイメージしやすいものである必要がある。例えばある人が「りんご」と言ったとき、その言語の発信者と受信者の脳内に浮かぶ物体のイメージはおそらくよく似た、赤くて丸い物体であろう。これに対して、たとえば「人形」と言ったとき、イメージされる物体の外観は多岐にわたる。このため、「のような人形」のように、よりイメージしやすい他の物体を用いて説明する必要が生まれる。あるいは、受信者にとって未知の物体のイメージを伝達する場合も同様に、他物体を用いた比喩的表現が必要となる。例として、シロチドリという鳥を挙げる。シロチドリの知名度は高くないが、「ウズラの卵を二つ重ねて枝を二本刺したようなもの」と説明されれば想像しやすい。このように、説明対象の物体をよりイメージしやすいパーツ部位に分解し、各パーツ部位を他の物体で比喩的に表現することは分かりやすい説明表現の生成に有用である。このようにして生成される物体説明表現は、実は「絵描き歌」の構造と同じである。絵描き歌は、隣接あるいは包含関係にあるパーツ物体を順に提示していくことで生成できる。

昨今の画像からの物体認識技術の発展は目覚ましく、このタスクの最高性能を競う大規模画像認識コンペティションでは、1000 カテゴリの物体の認識率 95%以上という人間に迫る性能が記録されている。2014 年度は、申請者のグループが[Kanezaki et al., 2014]を用いて 36 チーム中 4 位の成績をおさめた。しかしながら、本システムは学習データセットに存在しない物体の認識が不可能である。本申請課題では申請者がこれまで取り組んできた一般物体認識技術を応用し、未知物体をより分かりやすいパーツ物体に分割して比喩表現を生成する。これにより柔軟な物体認識を実現し、人と機械のインタラクションに真に役立つ物体認識技術の開発に取り組む。

2. 研究の目的

写真に写っている物体を分かりやすい表現で説明することを目指し、本研究では物体認識技術を応用した絵描き歌自動生成システムの要素技術開発に取り組む。未知の物体を言語で表現するとき、人はよりイメージしやすい物体を用いて「のような」「××をに乗せたような」といった比喩的表現を生成する。これを機械が実現するためには、物体の外観特徴を抽象的にとらえた上で他の物体との類似性を上位レベルで評価する必要があり、人工知能の本質的な課題といえる。このような物体の比喩的言語表現を代表するアプリケーションとして、「絵描き歌」がある。そこで本研究では、写真(画像)を入力とした絵描き歌の自動生成システムの開発を目指し、これを通して物体を認識・表現する知能の本質に取り組む。

3. 研究の方法

(1) 様々な角度から物体を撮影し、一般的な名称を推論する三次元物体認識技術の開発

絵描き歌に登場する物体は、誰もが名前を知っており、かつその概形を容易に想像できるような一般的な物体で構成される必要がある。このため、本研究では ImageNet のようにカテゴリが細分化され知名度の低い物体までもが含まれる画像データセットではなく、形状の違いに着目した 40~60 程度の一般物体カテゴリに所属する物体のみを集めた三次元物体データセットを用いて物体認識システムを学習する。具体的には、「コップ」や「椅子」等の 40 カテゴリからなる ModelNet40 と、(類似のカテゴリ分類からなる) 55 カテゴリの ShapeNet55 という三次元物体ベンチマークデータセットを扱う。これらは共に三次元物体認識および三次元形状検索分野で非常によく使われるデータセットであり、主に CAD で作られた三次元データで構成されるため、様々な角度から見た物体の画像を生成することができる。同じカテゴリの物体であっても、それを見る角度が変われば概形は大きく異なって見える。特に絵描き歌生成システムにおいては、絵描き歌を構成する物体パーツの概形の分かり易さ・伝わり易さが重要であり、様々な角度からの物体の見え方を知識として保有する必要がある。そこで、本研究では多視点から見た物体を観測して得るマルチビュー画像を学習し、これらを入力として物体の一般名称を推論する技術を新たに開発する。

(2) 事前知識を使用せず、画像内の物体をパーツに分解する教師無し画像セグメンテーション技術の開発

絵描き歌自動生成システムにおいて、説明対象の複雑な物体を、簡単な形状の物体パーツに分解する処理が必要である。このため、本研究では教師無し画像セグメンテーション技術を開発する。画像セグメンテーション技術とは、画像全体を複数個のまとまりを持った領域に分解する技術であり、色ヒストグラム等の特徴を手掛かりにして境界を決定する手法がよく知られている。昨今では、深層学習を用いてひとまとまりの物体の領域を推定するセマンティックセグメンテーションが盛んに研究されているが、そのほとんどが大量の画像と人手で正解領域をアノテーションされた大規模データセットを必要とする教師あり学習である。よって、見たことのない物体が写っている未知画像を領域分けすることができない。そこで、本研究では事前知識や事前学習を一切必要とせず、深層学習を用いて、未知の画像を教師無しでセグメンテーションする技術を新たに開発する。

4. 研究成果

(1) 様々な角度から物体を撮影し、一般的な名称を推論する三次元物体認識技術の開発

絵描き歌の素材として様々な物体を認識するためには、各物体カテゴリを代表するような一つの決まった姿勢のみを学習するのでは不十分であり、あらゆる回転姿勢の物体をさまざまな方向から観測して認識できるように学習する必要がある。そこで、回転台を用いた撮影およびインターネット上で収集した三次元物体の CAD モデルのレンダリングを用いて、物体を様々な方向から見たマルチビュー画像を学習し、物体のカテゴリと姿勢を同時に認識する手法を開発した。本研究は深層学習を用いており、物体の姿勢の教師信号を人間が与えることなく、自動的に獲得できる点が新しい。提案手法を用いた認識率のグラフを図 1 に示す。左から、ModelNet40, ModelNet10, 我々が構築した新しいデータセット MIRO を用いて評価した結果である。縦軸が物体カテゴリ識別率を、横軸が識別に用いた画像枚数を示している。赤い線が提案手法の結果を示しており、他のベースライン手法や、マルチビュー画像を用いた従来手法である MVCNN に比べて高い性能を示している。提案手法は、ベンチマークデータセット ModelNet10 と ModelNet40 において、現在世界トップの識別性能スコアを保持している。さらに提案手法は、三次元物体検索の国際的コンペティション SHREC17 にて、二部門で世界第一位の性能を記録した。なお、これらの知見をまとめた論文はコンピュータビジョン分野のトップ国際会議である IEEE CVPR 2018 に採択された。

また、実物体を認識・学習するための新たな画像データセットとして、様々な角度から撮影した 12 種 132 個 21,120 枚からなる日用品画像データセット (Multi-view Images of Rotated Objects; MIRO) を作成した。撮影した多視点画像の例を図 2 に示す。本データセットは Web 上で公開している。

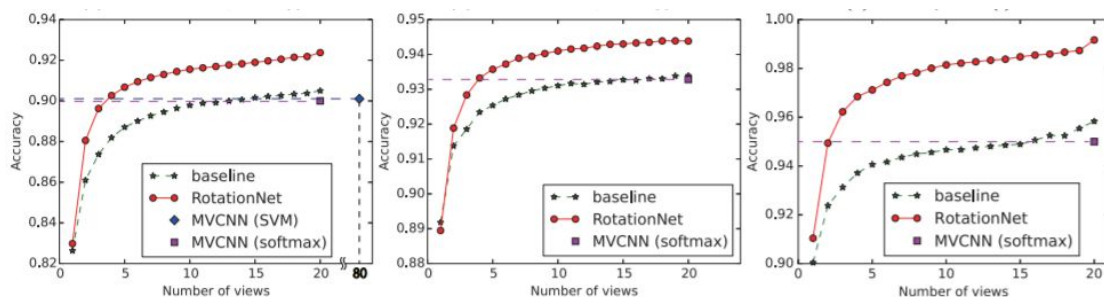


図 1 識別時の入力画像数に対する物体カテゴリ識別率の変遷



図 2 作成した日用品画像データセットの画像例

(2) 事前知識を使用せず、画像内の物体をパーツに分解する教師無し画像セグメンテーション技術の開発

物体を絵描き歌のような抽象化された表現で表すとき、あるまとまりを持ったパーツの集合体として物体を領域分割する必要がある。そこで、深層学習を用いて画像をセグメンテーションする手法を開発した。深層学習を用いた画像セグメンテーションの研究例は多く存在するが、そのほとんどが教師あり学習である。深層ニューラルネットワークの教師あり学習を行うためには膨大な学習データが必要であり、人手のアノテーションコストが大きい。また、学習データに存在するカテゴリの物体以外はセグメンテーションができないという問題がある。そこで、本研究では完全教師無しで画像のセグメンテーションを行うアルゴリズムを開発した。提案手法は畳み込みニューラルネットワークを用いるが、事前学習等の学習を一切必要としない。セグメンテーション対象のテスト画像が入力されると、画像ピクセルのグルーピングを行うタス

クを通して、畳み込みニューラルネットワークはランダムな重みからより良いパラメータへと自己学習する。画像セグメンテーションのベンチマークデータセット BSDS500 において、提案手法は従来手法を上回る性能を示した。提案手法を用いて未知の画像をセグメンテーションした結果の例を図 3 に示す。本研究成果は信号処理のトップ国際会議 IEEE ICASSP ' 2018 に論文採択された。なお、本手法は教師無し学習のみならず、一部の画像領域のラベルを指定するだけで画像全体をセグメンテーションするユーザインタラクティブな使用が可能であり、良く知られる従来手法グラフカットを適用した場合よりも性能が高いことを実験で確認した。また、本手法を用いて画像内の顕著なパターンを学習し、顕著性の高い物体の領域を検出するタスクに応用して知見をまとめ、信号処理のトップ国際会議 IEEE ICASSP ' 2019 に論文採択された。画像セグメンテーションはカラー画像に加え距離画像等の多種類のセンサ情報を統合することでより高い精度が得られる。このようなマルチモーダル画像を用いた深層学習アプローチについて知見をまとめ、Elsevier 出版の書籍 "Multimodal Scene Understanding 1st Edition" の第 2 章に寄稿した (2019 年 8 月刊行予定)。

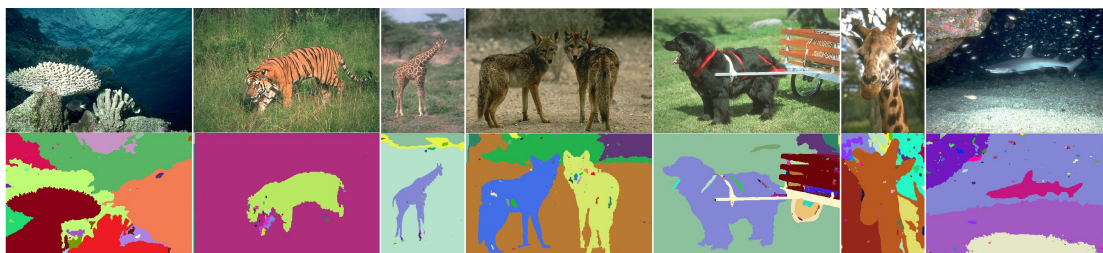


図 3 提案手法を用いた教師無し画像セグメンテーションの結果例

5 . 主な発表論文等

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 3 件)

Nevrez Imamoglu, Guanqun Ding, Yuming Fang, Asako Kanezaki, Toru Kouyama, and Ryosuke Nakamura. Salient object detection on hyperspectral images using features learned from unsupervised segmentation task. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), accepted, 2019.

Asako Kanezaki, Yasuyuki Matsushita, and Yoshifumi Nishida. RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.5010-5019, 2018.

Asako Kanezaki. Unsupervised Image Segmentation by Backpropagation. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1543-1547, 2018.

〔図書〕(計 1 件)

Asako Kanezaki, Ryohei Kuga, Yusuke Sugano, and Yasuyuki Matsushita (Chapter authors). Chapter 2: "Multimodal Scene Understanding: Algorithms, Applications and Deep Learning." Elsevier, August 2019. (525 pages)

〔産業財産権〕

出願状況 (計 0 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年：

国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

<https://kanezaki.github.io/pytorch-unsupervised-segmentation/>

6 . 研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。