

機関番号： 12401
 研究種目： 基盤研究（B）
 研究期間： 2007～2010
 課題番号： 19300055
 研究課題名（和文） 人間の物体・空間表現の分析に基づく対話を援用した統合物体認識
 研究課題名（英文） Object Recognition Integrating Autonomous and Interactive
 Methods Based on the Analysis of Human Expressions of Object
 Attributes and Spatial Relationships
 研究代表者
 久野 義徳（KUNO YOSHINORI）
 埼玉大学・大学院理工学研究科・教授
 研究者番号： 10252595

研究成果の概要（和文）：サービスロボットが作業を行うためには、対象となる物体を複雑な環境の中で認識できなければならない。物体認識の研究は進んできてはいるが、まだ自動で確実に認識することは難しい。そこで、自動で認識できないときには人間に対象物体がどういう物体かを聞く対話物体認識を検討した。このシステムでは物体の色や物体間の位置関係を人間に聞くことで認識を可能にする。実際にシステムを開発し、実験により有効性を確認した。

研究成果の概要（英文）：Service robots need to be able to recognize objects located in complex environments. Although there has been recent progress in this area, it remains difficult for autonomous vision systems to recognize objects in natural conditions. In this research, we have proposed an interactive object recognition system. In this system, the robot asks the user to verbally provide information about an object that it cannot detect. In particular, it asks the user questions regarding color and spatial relationship between objects depending on the situation. Experimental results confirm the usefulness and efficiency of our interaction system.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	3,200,000	960,000	4,160,000
2008年度	6,000,000	1,800,000	7,800,000
2009年度	2,200,000	660,000	2,860,000
2010年度	2,100,000	630,000	2,730,000
総計	13,500,000	4,050,000	17,550,000

研究分野：コンピュータビジョン、知能ロボット、ヒューマンインタフェース

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：コンピュータビジョン、物体認識、ヒューマンコンピュータインタラクション

1. 研究開始当初の背景

研究代表者はコンピュータビジョンを利用したヒューマンインタフェースについて研究を行っている。特に、今後の高齢化社会において人間を助けてくれる介護ロボットに関して検討を進めている。介護ロボットの作業の一つとして頼まれたものを取って来ることがある。この作業を実行するためには、頼まれたものを実世界の中で認識することが必要である。これはコンピュータビジョンにおける物体認識の問題になる。

物体認識には多くの研究があるが、工場の生産ラインの自動化のように環境を制限できる場合には実用的なものが実現できるが、介護ロボットが働く一般家庭のような変動の大きい環境で、多様な物体を確実に認識できるものはまだ実現していない。しかし、介護ロボットの場合には、そこに人間がいるので、全自動でなくても、その人間からの補助を考えることが可能である。そこで、自動で認識できない場合には、物体に関する情報を人間に対話を通じて教えてもらい認識を行

う対話物体認識を検討することにした。

2. 研究の目的

自動の物体認識は特定の定まった物体を認識する特定物体認識と、物体をカテゴリーのレベルで認識する一般物体認識に分けられる。特定物体認識に関しては、かなりの認識率が得られるようになってきている。しかし、一般物体認識については、統計的機械学習に基づく手法により性能は向上してきているが、まだ十分な認識率ではない。そこで対話物体認識の導入を検討する。しかし、常に対話が必要ではユーザに受け入れられない。したがって、まず、自動の物体認識を行い、それが失敗した場合に対話物体認識に入る統合物体認識システムを検討する。

対話物体認識においては、画像物体認識に関して知識のない一般のユーザが普通に使用する表現で対話を行える方法の実現を目指す。そのために、会話分析などを専門とする社会学の研究者と連携し、観察実験を通じて、人間同士の場合、指示した物体が分からないときにどのような表現を用いるかを明らかにする。そして、そのような表現に対応できる対話物体認識システムを実現する。対話においては、物体の属性だけでなく、物体間の位置関係についての表現についても検討する。

3. 研究の方法

(1) 人間の物体表現の調査

介護などのサービスロボットに取って代わることを頼むと思われる図1に示すような食品や日用品を集めた。そして、図2のような設定で人間が物体をどのように表現するか調べた。この実験は2人の参加者A、Bにより行う。2人の間には互いが見えないように仕切りがある。Bの前のテーブルには20個ほどの物体を置く。Aにその中の1つの物体を示し、物体名を言わないで、Bにその物体が何かを当てさせるために言語で表現するように指示した。これをBが正しい物体を認識するまで繰り返した。10組の参加者に対して、1組あたり10個の物体について実験を行った。



図1 実験対象物体の例

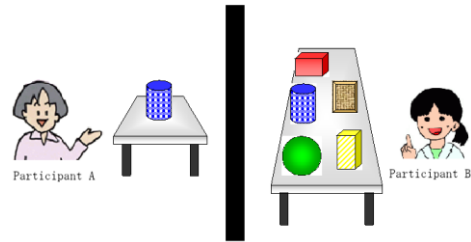


図2 人間の物体表現の調査実験

物体の空間位置関係についても同様の実験を試みたが、物体名を言わない条件では指示がしにくいため、十分なデータが得られなかった。そこで、人間の用いる表現については、言語学等の関連の文献を調査した。

(2) 対話物体認識システムの開発

前項の調査で分かった人間が用いる表現に対応する部分を画像から検出する方法を検討した。それを用いて、対話を通じて人間が指示した物体を認識するシステムを開発した。

(3) 自動物体認識法の検討

最近、研究が進んでいる統計的機械学習に基づく方法をもとに検討を行い、対話物体認識と組み合わせて使える自動物体認識を実現した。

(4) 統合システムの開発と評価実験

自動物体認識と対話物体認識を組み合わせ、自動認識が失敗したら対話認識に入る、統合物体認識システムを開発した。それを用いて、このシステムの有効性を実験により評価した。

4. 研究成果

(1) 人間の用いる物体表現

研究の方法のところで述べた調査実験を実施し、277個の物体の表現を得た。そのうちの67%は視覚で認識できる属性に関するものであった。視覚に関連する表現を属性別にまとめた結果を図3に示す。この図から分かるように、色について表現することが最も多かった。そこで、色に関しての表現をさらに詳しく分析した。図1に示したように多くの

物体は複数の色の部分から構成されている。しかし、色について表現した場合の81%は1つの色について述べただけであった。14%が2つの色で、3色以上を同時に表現したのは5%程度であった。そこで、複数色の中のどの部分の色を表現したかを調べると、背景色、あるいは面積が最大の領域の色の場合が、合わせると83%であった。

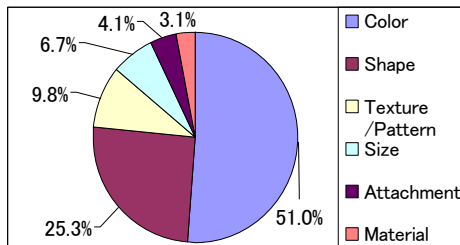


図3 表現された属性の内訳

色の次に表現されることの多かった形については、さらに同様の実験を形について表現するようにとの指示を加えて実験を行った。形については同じ表現であっても、それが全体の形状について言う場合、特定の方向から見た形を言う場合、部分や表面の模様などの形を言う場合などがあることが分かった。また、同じ表現でも違った形を表す場合がある(例えば「丸いもの」は円を表す場合もあるし、球を表す場合もある)ことも分かった。このように形の表現は多様で複雑であり、さらに検討する必要がある。

物体の空間位置関係については言語学等の文献調査から、人間の表現について検討した。Levinsonによれば参照物(*relatum*)に対する対象物(*referent*)の空間位置関係の記述には3種類の参照枠組(*frame of reference*): *absolute system*, *relative system*, *intrinsic system*が物体や状況に応じて用いられるということである。*absolute system*は東西南北のような絶対的な関係、*relative system*は、ある方向から見て(明示されていない場合は、話し手あるいは聞き手)、参照物に対して対象物がどういう位置にあるかというものである。*intrinsic system*は、参照物に固有の物体中心の座標系が設定され、それに対しての位置関係を述べるものである。例えば、テレビなら、それが話し手や聞き手に対してどこにあっても、「テレビの前」と言えば、表示画面の前方を指す。ロボットの視覚としては、このうちの*relative system*と*intrinsic system*に対応する必要があると考える。

(2) 対話物体認識システム

前項の結果、人間がある色の物体と言っても、その色の単色でない場合があることがわかった。そこで、ある色を指示されたとき、

その色が背景色か最大面積領域の色である物体を検出する画像処理法を開発した。形については、表現が非常に複雑なので、今回は丸や四角という表現に対して、全体形状がその形の物体であるものとして、それを検出する画像処理法を開発した。状況に応じて、指示された表現を解釈し、適切な部分の形状を検出することは今後の課題である。

空間位置関係に関しては、*reference system*と*intrinsic system*に対応できるものを開発した。両システムのどちらを使うかは物体により決まっているとして、人間が発話した物体名により決めている。

(3) 自動物体認識

自動物体認識に関しては、pLSA (*probabilistic Latent Semantic Analysis*)による仮説生成とその仮説のSVM(*Support Vector Machine*)による検証という方式により、10カテゴリー程度の小規模なものではあるが、物体認識研究でよく使われるデータセットに対して、他の先端研究と同等以上の認識性能を示すものを実現した。

(4) 統合物体認識システム

自動物体認識と対話物体認識を統合した物体認識システムを開発した。人間が物体を名称で指示した場合、システムはまず自動物体認識を行う。その物体が認識できない場合、あるいは認識が誤っていた場合に対話物体認識に入る。空間位置関係の対話の場合には、例えば、「テレビの前の本を取って」というように基準になる参照物について、人間は普通は物体名を使う。その参照物が自動認識できないと、それについて対話物体認識を行わなければならない。人間同士でも「○○の前の××取って」「○○ってどこ」というように参照物が分からない場合もあるが、現在の自動物体認識の能力では、このようなことが頻繁に起こる可能性が高い。そこで、今回の実装では、あらかじめシーンに対して自動物体認識を行い、認識できた物体の名称をロボットが人間に対して言うようにした。そうすれば、人間の方は、その物体を基準とした空間関係表現を行うことが期待される。ただし、一度に多くの物体名を言っても、人間は覚えきれないし時間もかかるので、一度には認識できたもののうち、空間的に離れた2つの物体名を言うようにしている。

この統合システムをロボットに実装した。図4にロボットの動作の様子を示す。ロボットは認識ができたと思ったら、指でその物体を指しながら「これですか」と人間に確認を求める。正しい物体が認識できるまで対話物体認識を繰り返す

テーブルの上に色のバリエーションの多い場合と少ない場合、それぞれ10個の物体

を置き、そのうちの1個、3個、6個が自動認識できたと想定する。すなわち、図5のように6つのケースを準備して、10人の実験参加者を用いて対話物体認識の実験を行った。それぞれのケースについて各人3例ずつ、総計180回の実験を行った。すべての場合に正解に達し、対話数はケースにより異なるが、平均してだいたい3回から4回であった。



図4 ロボットの動作

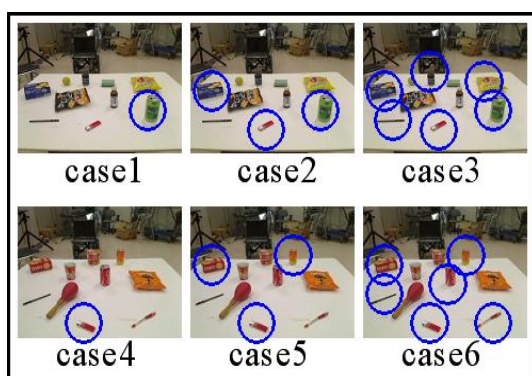


図5 実験に用いたシーンの例
上段は色のバリエーションが大きい場合、下段は小さい場合、青い丸は自動認識できたと想定した物体

(5) 研究成果の位置づけ・今後の展望

対話により物体を特定するというのは、これまでもロボットや人工知能の分野で研究されてきた。しかし、対象が単色で幾何形状も簡単な場合が多かった。それに対し、本研究では実世界の中の複雑な物体を対象にして対話物体認識を検討した。人間が通常用いる自然な表現で対話できるシステムを実現するために、まず、人間が実際に用いる表現を調査した。この調査で分かったことの一つは、人間は複雑な対象に対しても簡単な表現を用いるということである。例えば、複数の色が用いられている物体でも、1つの色だけと言う場合が多い。このように、人間の実際に用いる表現の意味に着目して対話物体認識を考えるというのは、人間に使いやすいシステムの実現のための重要なポイントである。自動物体認識の性能も向上しているが、まだ、必ず認識できるものの実現は困難であろう。また、人間同士でも物体の指示に名称だけでなく属性や位置関係を使う場合がある。したがって、対話物体認識はサービスロボットにとって重要な機能の一つである。本研究は、その実現に対して貢献するものである。

しかし、本研究の段階では、まだ実用化のためには課題が多い。今回の研究では人間の物体表現について調査を行い、それに基づき、対話物体認識を検討した。しかし、実装できたのはまだ一部である。調査を通じて、人間の物体表現の複雑さがさらに明らかになってきた。人間が用いる表現自体は簡単な場合が多いが、それをロボットが解釈するという点では非常に複雑である。例えば、形についての表現のところで述べたが、人間はいろいろな部分についての形について言及するが、必ずしもどの部分の形かを明示的に言わない場合が多い。また、同じ表現でも異なる幾何形状を指す場合もある。ロボットは画像処理によりある属性をもった物体を取り出したいわけだが、人間の発したこととそれが指す属性概念は状況によって変わってくる。したがって、対話物体認識において人間の表現を理解するためには、ことばと状況と概念の間の関係の知識と状況を認知する技術についてさらに検討が必要である。

また、対話物体認識では個々の物体がセグメンテーションされていることを前提としている。しかし、セグメンテーションも認識と同様に難しい問題である。完全なセグメンテーションは物体の認識ができていないと不可能な場合もある。最近、手軽に距離画像が得られるセンサが出てきているが、距離画像を利用すればセグメンテーションの問題は軽減できるので、その利用を検討していきたい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

論文公開

<http://sucra-rd.saitama-u.ac.jp/search/profile.do?lng=ja&id=gmkeumcN>

[雑誌論文] (計6件)

- ① D. Das, Y. Kobayashi, and Y. Kuno, Object detection in cluttered range images using edgel geometry, IEEJ Trans. EIS, Vol.130, No.9, pp.1572-1580, 2010 (査読有)
- ② D. Das, Y. Kobayashi, and Y. Kuno, Multiple object category detection and localization using generative and discriminative models, IEICE Trans. Information and Systems, Vol.E92-D, No.10, pp.2112-2121, 2009 (査読有)
- ③ A. Mansur and Y. Kuno, Specific and class object recognition for service robots through autonomous and interactive methods, IEICE Trans. Information and Systems, Vol.E91-D, No.6, pp.1793-1803, 2008 (査読有)

[学会発表] (計27件)

- ① L. Cao, Y. Kobayashi, and Y. Kuno, Spatial resolution for robot to detect objects, International Conference on Intelligent Robots and Systems (IROS2010), 2010年10月21日, Taipei (Taiwan)
- ② D. Das, Y. Kobayashi, and Y. Kuno, Sub-category optimization for multi-view multi-pose object detection, IAPR International Conference on Pattern Recognition (ICPR2010), 2010年8月24日, Istanbul (Turkey)
- ③ Y. Kuno, K. Sakata, Y. Kobayashi, Object recognition in service robots: Conducting verbal interaction on color and spatial relationship, IEEE 12th ICCV Workshops (Human-Computer Interaction), 2009年10月4日, 京都大学
- ④ Y. Kuno, K. Sakata, Y. Kobayashi, Robot vision for recognizing complex objects through simple interaction, Mecatronics 2008, 2008年5月23日, Le Grand-Bornand (France)

[図書] (計1件)

- ① D. Monekosso, P. Remagnino, and Y. Kuno (eds.), Intelligent Environments: Methods, Algorithms and Applications, Springer, pp.1-11, pp.191-210, 2009

[その他]

ホームページ等

<http://www.cv.ics.saitama-u.ac.jp/research/index.html>

6. 研究組織

(1) 研究代表者

久野 義徳 (KUNO YOSHINORI)
埼玉大学・大学院理工学研究科・教授
研究者番号：10252595

(2) 研究分担者

小林 貴訓 (KOBAYASHI YOSHIORI)
埼玉大学・大学院理工学研究科・助教
研究者番号：20466692

(3) 連携研究者

山崎 敬一 (YAMAZAKI KEIICHI)
埼玉大学・教養学部・教授
研究者番号：80191261