

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 15 日現在

機関番号：12401

研究種目：基盤研究(B)

研究期間：2011～2014

課題番号：23300065

研究課題名(和文) サービスロボットのためのオントロジーに基づく対話を援用した統合物体認識

研究課題名(英文) Integrated Object Recognition for Service Robots Based on Ontology with the Interactive Support from Humans

研究代表者

久野 義徳 (KUNO, Yoshinori)

埼玉大学・理工学研究科・教授

研究者番号：10252595

交付決定額(研究期間全体)：(直接経費) 12,300,000円

研究成果の概要(和文)：頼んだものを取ってきてくれるような介護ロボットの開発を目指している。その実現のためには頼まれた物体を認識する技術が必要だが、実世界のさまざまな状況で物体を確実に認識するのは、まだ困難な状況である。そこで、自動物体認識に失敗した場合には、ユーザにことばにより物体に関する情報(例えば色や形)を教えてもらい認識を行うという対話物体認識に入る統合物体認識を検討した。自然言語では、物体についての表現と、その示す物理的な属性が1対1に対応するとは限らず複雑である。この関係をオントロジーにまとめ、人間の言語表現を適切に解釈して認識できる枠組を実現した。

研究成果の概要(英文)：We have been developing a helper robot that is able to fetch objects requested by users. Such a robot must recognize the desired object(s) in order to carry out its tasks. However, it is difficult for a system to recognize objects autonomously without fail under various real-world conditions. To address this problem, we have proposed an integrated object recognition system. The system first attempts autonomous object recognition. If it fails, it switches to an interactive object recognition mode where the system asks the user to verbally provide information about the object that it is unable to detect autonomously. However, natural language descriptions from humans can be complex. Such descriptions may not even have a one-to-one correspondence to the physical attributes of objects. We have constructed an ontology for representing such complex relationships between human descriptions and physical attributes and developed a method to recognize objects based on this ontology.

研究分野：知能ロボット

キーワード：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 オントロジー 物体認識 対話システム

## 1. 研究開始当初の背景

研究代表者は、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の実現を目指して研究を進めている。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。物体認識はコンピュータビジョンの主要なテーマであり、その初期から活発に研究されてきたが、その進展にはしばらく停滞が見られた。しかし、近年、言語処理における bag-of-words のアナロジーの bag-of-visual-words など統計的機械学習に基づく手法により急速に発展してきている。しかし、まだカテゴリーレベルの認識では対象物体により異なるが、認識率が 70% というのは良い方という段階である。認識率は今後さらに高まるだろうが完全なものの実現は困難であろう。身体が不自由でロボットにものを取ってくるように頼んだのに認識できずにサービスができないというのでは困る。しかし、想定するユーザは、身体は不自由でも言語による指示はある程度できる場合が多いと考えられる。そこで、自動認識に失敗した場合は、ユーザに物体に関する情報をことばで教えてもらって認識を行う対話認識に入るといった統合物体認識を提案し、研究を進めていた。

対話を通じてシーンを理解するというのは Winograd の古典的な研究以来、多くの研究がある。しかし、それらの研究はコンピュータビジョンによる物体認識が研究の中心課題ではないので、記号として簡単に表せるような物体によるシーンを対象にしていた。例えば、単色の赤いボールや青い箱からなるようなシーンでは、ロボットが対象物がわからなかった場合、「赤いもの」と言えばロボットは理解できる。しかし、実際にはこのような単色の物体は少ない。例えば、スナック菓子のパッケージなどには多くの色が使われている。それをロボットが自動認識できなかった場合、ロボットに対してどう言えばよいであろうか。通常の間人間の会話では使われている色を全部言うようなことはしないであろう。そこで、人間同士でいろいろな物体について物体名を言わないという条件で、どういう表現で相手に物体をわからせるかを社会学の研究者と共同で調査した。その結果、人間は色について表現することが多く、多色の物体でも、ほとんどの場合、一つの色名しか言わないことがわかった。その色は、多くの場合、背景色か面積の一番大きい部分の色であった。そこで、色名を言われたとき、単色の物体ではなく、人間の使う意味での色の物体を検出する対話認識システムを開発した。

また、人間は既知の物体との位置関係で対象物を指示することも多い。これについても人間同士のインタラクションの分析と言語

学の文献調査で検討した。その結果、基準になる物体によって同様な位置関係にあっても異なる言い方をすることがわかった。そこで、それに対応できる対話認識システムを開発した。

## 2. 研究の目的

以上のように、色や位置関係という個別の属性に関して検討を行っていたが、さらに多くの属性に対応し、人間の多様な指示に対応できるようにするためには、人間の表現と物理的な属性、そしてそれを検出する画像処理法に関連する知識を統一的に扱える枠組が必要である。そこで、知識表現法として注目されているオントロジーを利用して、家庭内でのサービスロボットの対象物になると考えられる物体に対して、対話物体認識が可能なシステムの実現を目指して研究を行った。

### (1) ロボット視覚用オントロジーの構築

物体の指示には色や形などの属性と物体間の空間位置関係などの関係が用いられる。そこで、人間の指示に用いられるこれらの概念を体系化したオントロジーを構築する。しかし、指示に用いられることばを集めるだけでは対話物体認識用のオントロジーにはならない。オントロジーは本来、概念を体系化するものである。その概念はことばで表すが、概念とことばが 1 対 1 に対応するとは限らない。「丸いもの」が球を指す場合もあれば円盤を指す場合もある。また、周囲にある物体が角ばった物体ばかりの場合、少し角が丸まった物体を「丸いもの」というかもしれない。また、先に色の例を述べたが、ことばが物理や幾何学的に表す概念と人間が物体を指す場合の概念の関係も単純ではない。したがって、人間が物体の属性や関係についてもつ概念を体系化し、それを表すことばを整理した上で、ことばが状況に応じてどの概念を表すものになるかを関係付けたオントロジーを構築しなければならない。本研究では、このような関係を表現するオントロジーの枠組を考案し、サービスロボットの物体認識用オントロジーを構築する。

### (2) 統合対話物体認識システムの開発

作成したオントロジーを用いて、指示された物体を認識できるシステムを開発する。最初に自動認識を試み、認識できない場合は対話物体認識を行う。これをロボットに実装し、サービスロボットが対象にすると思われる一般家庭にある物体を認識できるシステムを実現する。

## 3. 研究の方法

### (1) 人間の用いる物体表現の調査

被験者を用いた実験で、人間が物体名を使わない時に、どのように物体を表現するか調べる。食べ物や文房具のように、サービスロボットに持ってきて欲しいと思う物体を集め、その中から無作為にいくつか選んでテーブルの上に置く。そのシーンの画像を撮影し、

データとする。被験者にその画像を提示し、そのうちの1つを他人に取ってもらうと、どう表現するかを聞く。その際、その物体の名前や機能についての説明はしないように指示する。ここで使われた表現を集めて、人間がどのように物体を表現するか分析する。

## (2) オントロジーの構築

前項の結果に基づき、人間の表現を理解するためのオントロジーを構築する。オントロジーの構築には大阪大学溝口研究室で開発されているシステム「法造」を利用する。

## (3) ロボットシステムの開発

カラー画像と距離画像の得られるマイクロソフトの Kinect をセンサとして用い、そのデータから認識に必要な属性（色、形、位置等）を求めるプログラムを開発する。小型人型ロボット (Aldebaran Robotics 社の Nao) を用いて、対話物体認識のできるロボットシステムを開発する。

## (4) 実験

ロボットシステムにより、さまざまな状況で対話を通じて物体が認識できるか確認する。また、定量的に有用性を評価するために(1)で採取したデータを用いて実験を行う。この場合は対話ではなく、人間が表現したデータをテキストとしてシステムに入力し、その表現をもとに物体が認識できるかを調べる。

## 4. 研究成果

### (1) 人間の用いる物体表現の分析

3(1)で述べた実験のために、20個の物体を集め、その中から無作為に5個の物体を選び、15のシーンの画像を作成した。各画像につき画像中の5つの物体をそれぞれ指示対象とする合計75パターンについて、大学生27人を対象にして人間の物体に対する表現を収集した。

このような設定のもと実験を行った結果、合計で物体に対する181の表現を収集した。得られた表現を幾つかのグループに分け、分類した結果を図1に示す。結果としては、物体の色を用いて物体を表現する方法が多く用いられ、その次に形状、位置関係、材質、模様、大きさの順となった。この結果の数字

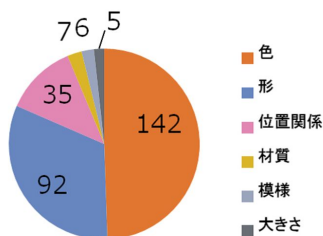


図1 物体の表現の分類結果

は重複を含んでカウントしてある。例えば「赤い球」なら色と形状の両方でカウントしている。

物体表現について物体認識用のオントロジーを構築するにあたり、同じ物体について様々な表現があることは当然だが、それ以外に以下の2点を検討する必要のあることが分かった。

### 表現の示すもの

同じ表現でも異なる物理的属性を示すことがある。例えば、「丸い」という表現は、ボールのような球を指すこともあるし、皿のような円盤のようなものを指すこともある。したがって、人間がある表現で目的の物体を指示した場合、その表現が示す可能性のあるすべての物理的属性をもつものが検討の対象になる。しかし、人間がその表現を用いたということは、そのシーンでは、目的の物体が最もその表現に合った物体ということであろう。すなわち、同じ表現で表される可能性のある物理属性の間に、その表現の当てはまる度合の優先順位があると考えられる。

そこで、特に形の表現でよく用いられた「丸い」という表現について、優先順位を調べた。9つの形を用意し、そのうちの2つをディスプレイに表示し、どちらが丸いと思うかを聞くという一対比較法による実験を行った（実験の詳細は5の雑誌論文を参照）。この実験により「丸さ」の尺度を求めた結果を図2に示す。これによると、「球」の方が「円盤」より丸く感じられるということになる。したがって、シーンの中にボールと皿がある場合に、「丸いもの」と言われたら、一般にはボールを指している可能性が高いと考えられる。

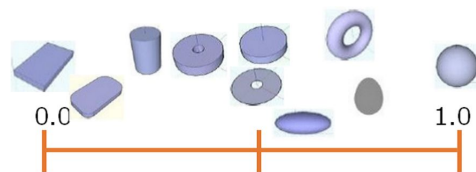


図2 丸さの尺度（右に行くほど丸いと感じる）

### 表現の示すところ

1で述べたように、色については、多くの色が使われている物体でも、背景色や最大面積部分の一色で指示されることが多い。色についてよく用いられる形については、全体の形についていう場合、ある方向から見た形についていう場合、あるいは、その物体の一部や付随したものの形についていう場合などがあった。その表現が示すところについて明示的という場合もあったが、多くの場合は、省略されていた。したがって、人間がある表現で指示した場合、その表現がどこを示すものかを検討する必要がある。

(2)対話物体認識用オントロジーの構築

人間の用いる物体表現の分析に基づき、人間の多様な物体表現に対応できる対話物体認識用のオントロジーを構築した。

基本的なオントロジーは、図3に示すように、対話物体認識の枠組で主に扱う、Object(実在物)、Human Description(人間の表現)、Object Attribute(属性)の3つのサブオントロジーで構成される。オントロジーの構築には法造を用いた。法造ではスロットは part-of 関係および attribute-of 関係を示す。図ではそれぞれ p/o、a/o と示されている。Object のサブオントロジーではロボットが取ってくる対象物体を定義し、Human Description では人が物体に用いる表現を定義する。画像処理により検出される物体の属性は、そのまま Object Attribute のサブオントロジーの概念として定義しておく。

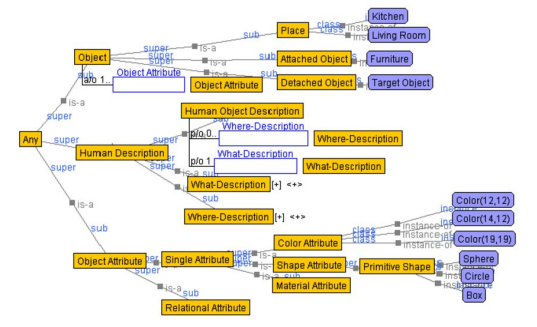


図3 オントロジーの概略図

以下では、このオントロジーの中心となる Human Description のサブオントロジーについて述べる。前節で検討が必要と述べた「表現の示すもの」と「表現の示すところ」に対応した構造になっている。人間の表現の概念は、その属性を表現する What-Description と、その表現がどの部分を示すものかという Where-Description を part-of 関係に持つ。「上から見ると丸い」という表現では、「上から見ると」が Where-Description として、「丸い」が What-Description として解釈される。図4は What-Description と Where-Description の構造を示している。

What-Description には人間の表現がどのような概念(物体の属性)を示すことがあるかという情報を記述する。この What-Description は part-of 関係に Object Attribute を持つ。この Object Attribute にはどのクラスに当てはまるかという制約が示されている。図4中の「アカイ」という What-Description では、Object Attribute のクラス制約として Color(12,12) | Color(14,12) という記述があり、「アカイ」という表現は属性 Color(12,12) と Color(14,12) を示すことが分かる((Color(m,n)はmが最大領域色、nが背景色を示す。それぞれの数値はYCbCr空間を20に分割して付けたもの)。このクラス制約にその表現が示す属性を記述しておくことで、

その表現が示す概念の範囲を決定する。また、複数候補がある場合には、どちらがよりその表現に合ったものかを調べなくてはならないが、それはクラス制約に優先順に属性を記述しておくことで示す。Color(12,12) | Color(14,12) という記述では Color(12,12) という属性の方が「アカイ」という表現がされた場合に優先順が高いことを示している。

Where-Description では、その表現が物体のどの部分を対象としているかを attribute-of 関係の「Part」に記述しておく。どの部分かを示す「Part」は、多くはその対象物全体を示すが、「上から見ると」といった表現ではその方向から3次元情報を射影したものが表現の対象部分となる。また、どの What-Description に対して行われる表現を「What-Description」というスロットに示しておく。「上から見ると」などと言った表現は形状表現にだけ用いられる表現と考えられるため、どの What-Description に合わせて使われる表現かを限定しておく。「上から見ると」や「下地」といった表現が実際に使われると、システムはその概念を探索し、それに対応した画像処理を用いて属性を検出し、認識を進める。

図5に実装した What-Description の一部の詳細を示す。What-Description には「赤い」「丸い」「木製の」などの140程度の表現が入力されている。これらは、実験で実際に用いられた表現や、一般的に用いられるだろう表現を日本語辞書から選択し、手動で入力したものである。最下層の概念は日本語で実際に用いられる表現に対応しているため、日本語で表記している。図5の例では、色を用いた表現のうち赤いものを示す表現について詳細を示してある。色に関する表現では、まず基本的な色で分類していき、最後にその色に対応する言葉(表現)を結び付けた。似た表現でまとめることで、ある表現が与えられたときに、似た表現が参照されるようになっている。

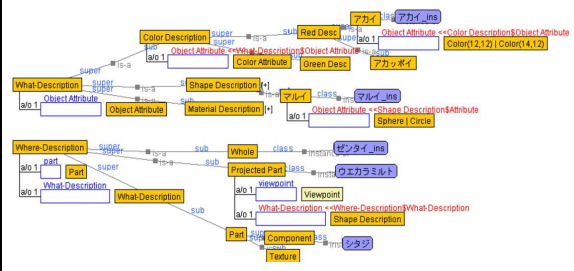


図4 人間の表現サブオントロジー

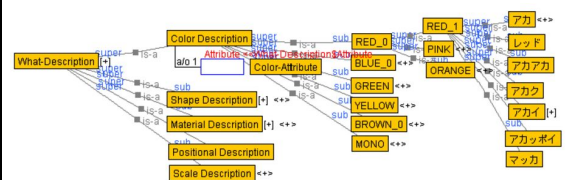


図5 What-Description の一部の詳細

### (3) ロボットシステムの開発

色、形、材質、位置等の属性を検出する画像処理法を開発し、図6のようなロボットシステムを実現した。音声認識関連は既存の技術を用いた。中心となる物体認識の部分は以下のように動作する。ユーザから何かを取るよう頼まれた場合、それが自動物体認識により分かっている既知物体にあった場合、それをユーザに提示し、正しければ認識のプロセスを終了する。認識に失敗した場合は対話を援用して物体を認識するよう試みる。

実際に表現が与えられると、例えば、「赤い」という表現が与えられた場合、その表現の概念を探索し、属性 (Object Attribute) のクラス制約を調べる。クラス制約にある属性を持つ物体を検出し、対象物の候補とする。クラス制約に記述されていない属性がその表現によって示されたということが最終的な結果として分かった場合は、その属性をクラス制約に新たに追加し、オントロジーを更新する。複数候補があった場合には、どちらがよりその表現に合ったものかをクラス制約を参照し、設定した優先度の高い物を候補として選択する。

表現の概念を探索し、属性のクラス制約から対象の物体を検出しようとしたとき、候補物体を検出できなかった場合は、まずその概念の下位概念があれば下位の概念を参照する。それでも検出できなかった場合は、その上位の概念を遡って物体の検出を試みるようになっている。

### (4) 実験による有用性の実証

開発したロボットシステムを用い、対話により指示した物体を認識できることを確認した。対話物体認識なので、対象物が他と分離して検出されていれば、次々に「これですか」と確認していけば、そのうちに認識できることになる。そこで、定量的な評価を行うために、4(1)で述べた実験で収集したデータを用いた実験を行った。この場合は、ロボットは使用せずに、収集した表現をロボットに与えたときに、画像から対象物が認識できるかを調べた。比較のためにオントロジーを用いずに属性と表現との対応関係だけでシステムを構築した場合についても実験した。同一表現に対する属性間の優先順位や、指示された表現に当てはまる物体がないときに関

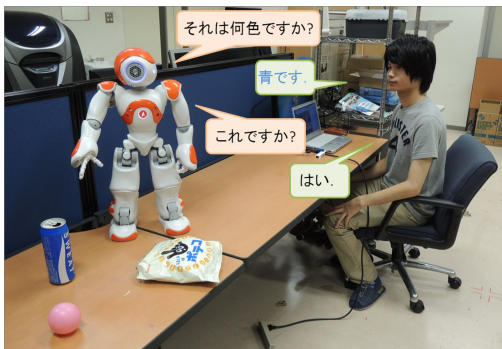


図6 対話を援用したロボットシステム

連の概念を探索できるなどにより、提案のオントロジーを用いる方が人間の表現にあてはまる物体を効率的に認識できることが示された。実験結果の詳細は5の学会発表を参照して欲しい。

### (5) 研究の位置づけと今後の展望

ロボットの知識表現の方法としてオントロジーは注目を集めてきている。IEEEのRobotics and Automation SocietyにもOntologies for Robotics and Automationというワーキンググループが設立された。日本でもオントロジーを用いたロボットが発表されている。しかし、これらの研究においては、ロボットの関連する世界の記述が中心で、本研究で考えているような人間とロボットのインタラクションのための人間の言語表現の多様性への対応は考えられていない。この点に本研究の学術的意義がある。一方、自動物体認識の技術は非常に進歩してきているが、実環境で必ず動作するものの実現は難しい。したがって、本研究で提案する対話物体認識は実用面での価値も大きい。

しかし、現在のシステムにはまだ多くの問題がある。対話物体認識の大きな問題は、対象物が他と分離して検出されている(セグメンテーションされている)と仮定していることである。この仮定により、分離された物体の中から指示された属性をもつものが対象物だと認識できる。距離画像データを用いることにより、この問題はかなり解決されているが、いつも完全に分離・検出できるとは限らない。セグメンテーションの技術を向上するとともに、誤りがあった場合にも対話を通じてシステムにそれを理解させるような方法の検討が必要である。

現在のオントロジーはほとんどを人手により作成している。今後は、システムを人間が使っているうちに、オントロジーをその人やその使用される環境に合ったものになるように学習させていく枠組の検討が期待される。

また、ここでは人間の言語行動しか考えていないが、人間のコミュニケーションでは非言語行動も重要な役割を果たしている。この非言語行動への対応もシステムに取り入れていきたい。

### <引用文献>

J. Ponce, M. Hebert, C. Schmid, and A. Zisserman (Eds.), *Toward Category-Level Object Recognition*, LNCS 4170, Springer, 2006

T. Winograd, *Understanding Natural Language*, Academic Press, 1972

A. Mansur, K. Sakata, Y. Kobayashi, and Y. Kuno, *Human robot interaction through simple expressions for object recognition*, RO-MAN2008, pp.647-652,

2008

L. Cao, Y. Kobayashi, and Y. Kuno, Spatial resolution for robot to detect objects, IROS2010, pp.4548-4553, 2010

Hozo: Ontology Editor,  
<http://www.hozo.jp/>

E. Prestes et al., Towards an upper ontology and methodology for robotics and automation, UbiComp'12, pp.873-892, 2012

## 5 . 主な発表論文等

[雑誌論文](計4件)

森 智史、福田 悠人、小林 貴訓、久野 義徳、加地 大介、物体形状オントロジーに基づく指示形状物体の認識、画像電子学会誌、vol.42、no.4、pp.477-485、2013、査読有

<http://www.iieej.org/gakkaishi.html>

H. Fukuda, S. Mori, K. Sakata, Y. Kobayashi, and Y. Kuno, Object recognition for service robots based on human description of object attributes, IEEJ Trans. Electronics, Information and Systems, vol.133-C, no.1, pp.18-27, 2013, 査読有

DOI: 10.1541/ieejieiss.133.18

D. Das, Y. Kobayashi, and Y. Kuno, Sub-category optimization through cluster performance analysis for multi-pose object detection, IEICE Trans. Information and Systems, vol.E94-D, no.7, pp.1467-1478, 2011, 査読有

DOI: 10.1587/transinf.E94.D.1467

[学会発表](計19件)

福田 悠人、森 智史、小林 貴訓、久野 義徳、加地 大介、サービスロボットのための物体表現オントロジーに基づいた物体認識、第20回画像センシングシンポジウム、2014年6月12日、パシフィコ横浜(神奈川県横浜市)

H. Fukuda, S. Mori, Y. Kobayashi, Y. Kuno, and D. Kachi, Object recognition for service robots through verbal interaction based on ontology, 9th International Symposium on Visual Computing (ISVC2013), 2013年7月30日, Rethymnon (ギリシア)

L. Cao, Y. Kobayashi, and Y. Kuno, Spatial-based feature for locating objects, 8th International Conference on Intelligent Computing (ICIC2012), 2012年7月27日、Huangshan (中国)

H. Fukuda, S. Mori, Y. Kobayashi, and Y. Kuno, Object recognition for service robots through verbal interaction about multiple attribute information, 8th International Symposium on Visual Computing (ISVC2012), 2012年7月17日, Rethymnon (ギリシア)

L. Cao, Y. Kobayashi, and Y. Kuno, A spatial-based approach for groups of objects, 8th International Symposium on Visual Computing (ISVC2012), 2012年7月17日, Rethymnon (ギリシア)

S. Mori, Y. Kobayashi, and Y. Kuno, Understanding the meaning of shape description for interactive object recognition, 7th International Conference on Intelligent Computing (ICIC2011), 2011年8月12日, Zhengzhou (中国)

M. Mannan, H. Fukuda, L. Cao, Y. Kobayashi, and Y. Kuno, 3D free-form object material identification by surface reflection analysis with a time-of-flight range sensor, 13th IAPR Conference on Machine Vision Applications (MVA2011), 2011年6月14日, なら100年会館(奈良県奈良市)

L. Cao, D. Das, Y. Kobayashi, and Y. Kuno, Object spatial recognition for service robots: where is the front?," IEEE International Conference on Mechatronics and Automation (ICMA2011), 2011年8月9日, 北京(中国)

[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, Yoshinori)  
埼玉大学・大学院理工学研究科・助教  
研究者番号：20466692

加地 大介 (KACHI, Daisuke)  
埼玉大学・教養学部・教授  
研究者番号：50251145