

人間が用いる物体表現の調査に基づいた対象物体の検出

Detection of Objects Based on Research of Human Expression for Objects

坂田 克俊 久野 義徳

Katsutoshi Sakata, Yoshinori Kuno

埼玉大学 〒338-8570 埼玉県さいたま市桜区下大久保 255

Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama, Saitama 338-8570, Japan

E-mail: {sakata, kuno}@cv.ics.saitama-u.ac.jp

Abstract

生活環境内で動作する福祉ロボットの一つとして、本研究室では『利用者に頼まれたものを取ってくるロボット』を想定し、その際に必要となる物体認識の手法として人間との対話を利用したシステムを提案してきた。そして、これまでの研究によってその有用性を示すことに成功した。しかし、それらの研究で扱われていた作業命令の形式はごく限られたものであり、システムの実用性についてはほとんど考慮されていない。そこで、本研究では人間が物体をどのように表現するのか調査し、その結果に基づいた画像処理を作成する、という人間の特性に沿った実用的な対象物体の検出法を提案する。

1 はじめに

近年の急速な高齢化、それに伴って生じる介護者の人員不足もあり、生活環境内で人間の手助けをしてくれるような福祉ロボットの研究が現在数多くされている。

本研究室では、そのようなロボットのの一つとして『利用者に頼まれたものを取ってくるロボット』を想定し、その際に必要となる物体認識の手法として人間との対話を利用したシステムを提案してきた[1]。利用者から「その本を取って」や「あれ取って」といった物体名や簡略化された発話[2]による依頼があったときに、現状の物体認識技術では必ずしも対象物体の検出に成功するとは限らない。そこで、ロボットが利用者に「どんな色ですか？」や「どんな形ですか？」といった質問をすることによって、対象物体の検出に必要な情報を与えてもらうという手法である(以後、この手法を対話物体認識と称す)。そして、これまでの研究によって対話物体認識の有用性を示すことに成功した[3]。しかし、その際に扱っていた物体は赤くて丸いものといった簡単な物体であったため、作業命令の形式はごく限られたものであった。実生活で見られる物体のほとんどは複数の色を持っていたり複雑な形状をしているため、従来の作業命令の形式では対象

の検出が非常に困難となる。福祉ロボットの活躍の場が我々人間の生活環境内であることを考慮すると、そういった物体においても対話物体認識が機能しなければ実用的なシステムとはいえない。そこで、本研究では生活環境内で見られる様々な物体に対しても対話物体認識によって対象の検出ができるようなシステムの開発を試みる。

2 人間が用いる物体表現の調査

目標とするシステムを実現するために、まず食べ物や飲み物や雑貨といった実生活で見られる物体を実際に多数、用意した(図1)。前項でも述べたように、ロボットは物体名を言われただけではそれを認識できない場合が多い。そこで、物体名を言うてはいけなかったときに、これらの物体を人間がどう表現するのか調べることにした。人間が表現した部分を検出できるような画像処理を作成すれば、対話を通じた物体認識が可能になると考えている。



図1: 用意した物体

2.1 個々の物体に対する表現の調査

個々の物体に対して人間が用いる表現を収集するために、以下のような実験を行った。

2.1.1 概要

2人1組の被験者を指示者と受け手に分け、指示者にはこちらが提示した物体を音声のみで受け手に伝えてもらい、その際に使用された表現を収集する。図2がその実験環境となる。受け手側のテーブルには図1の中の物

体を 20 個程度配置し, 相手側のテーブルは間に壁を挟むことによって見る事ができないようになっている。

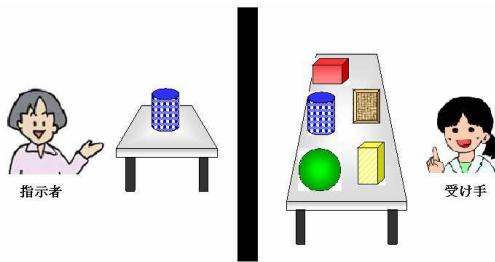


図 2: 実験環境 1

2.1.2 手順

実験の手順は以下のようになる。

受け手側のテーブルにある物体と同一の物体を 1 つだけ指示者側のテーブルに配置する。

指示者はそれがどんな物体なのかを音声のみで受け手に伝える。

受け手は指示者からの指示を聞き, それがどの物体であるのか予想して答える。

受け手の回答が正解なら, それ以外なら へ。

～ の手順を複数回繰り返す。

2.1.3 結果・考察

実験結果について述べる。10 人の被験者から 227 の発話を収集した。図 3 は, 指示者の一回の発話内において無知識情報が含まれていた確率と, 知識情報が含まれていた確率を表したものである。ここで, 無知識情報というのは「赤い」や「丸い」といった物体に関する知識がなくても表現できる情報を意味する。一方, 知識情報というのは「食べ物」や「ジュース」といった物体に関する知識がないと表現できない情報を意味する。本研究では, ロボットが対象物体についての知識を持たなくても, その認識を可能とすることなので, 主に前者の無知識情報からの認識を目指すことになる。そこで, 無知識情報にはこういったものがあるのかを調べた。それを表したのが図 4 である。物体の色, 形, 模様, 大きさ, 付属物, 素材の 6 種類による表現があった。今回は, その中でも一番多く利用されていた色に関する表現 (51.0%) の分析を行った。色による発話については, 対象の物体が複数の色を持っているにもかかわらず, その 80.8% が 1 色で表現されていた (図 5)。このことから, 人間は物体を 1 色で表現する傾向にあることがわかり, 自然な対話物体認識を行うためには, その色が物体のどの部分の色に相当するのかを把握することが重要となる。そこで, 1 色による表現の分析を行った。図 6 がその結果である。対象の 82.5% (36.3+37.5+8.7) が物体の下地の色または最も割合の多い色を表現していた。このことから, 人間の捉える傾向にある色は物体の下地の色や最も割合の多い色であることがわかり, そういった箇所の色を検出する画像処理が対話物体認識において有用であるといえる。

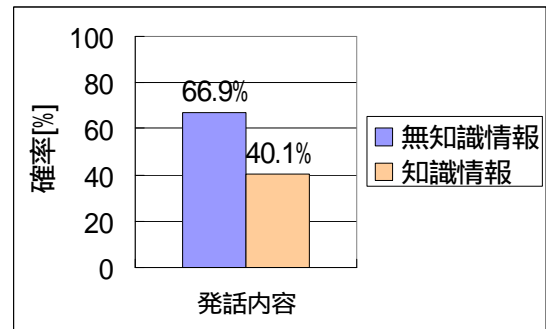


図 3: 発話内容の内訳

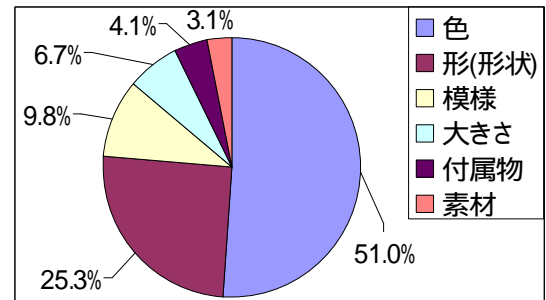


図 4: 無知識情報の内訳

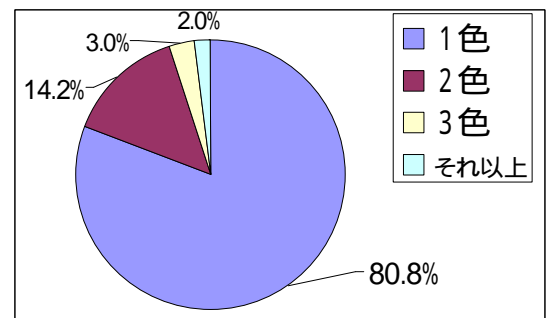


図 5: 表現された色の種類

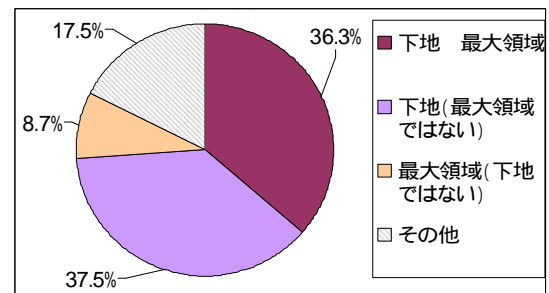


図 6: 1 色で表現された箇所

2.2 空間関係を利用した表現の調査

前項の調査では, 個々の物体に対する表現を調べたが, 人間が物体を指定するのに, その物体個々の表現に加え, 物体間の空間関係を使うことも多い。Levinson[4]は人間の reference system を intrinsic, relative, absolute の 3 つに分けている (intrinsic は「家の前」など物体名を言えばその前後関係を特定できるものを利用する方法, relative は「A から見て B の C にある」というようなもの, absolute は東西南北など絶対的に決まっている指標を利用するもの)。特に, 介護場面では人間またはロボッ

トを A とした relative system が中心になると考えられる。「(話者から見て)赤い本の右」というような表現である。relative system を使用する上で問題となるのは何を B(reference object)に選ぶかである。ロボットの場合は認識が正確に行えて、人間に言葉で伝えやすく、他の物体の位置を表しやすいものである必要があると考えられる。加えて、C の位置関係を表す言葉についても、人間がどういった表現をするのか把握する必要がある。

そこで、以下に述べる実験を行うことによって、どのような物体を reference object に選び、それに対してどういった位置関係を表現するのか調べた。

2.2.1 概要

前の調査と同様に、2 人 1 組の被験者を指示者と受け手に分け、指示者に提示した物体を受け手に伝えてもらうことによって表現の収集を行う。図 7 がその実験環境となる。テーブルには図 1 の物体を 10 数個配置している。また、指示者には受け手の後ろに立つてもらうことによって、受け手とほぼ同じ視線でテーブル上にある物体を見ることができ、加えて、指示者の視線や体の向き、ジェスチャ等は受け手に伝わることはなく、受け手が得られる情報は指示者の音声のみとなっている。

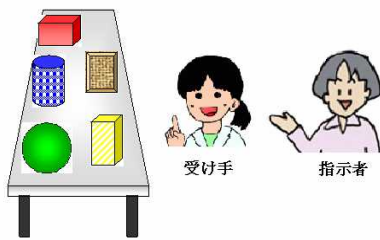


図 7: 実験環境 2

2.2.2 手順

実験の手順は以下になる。

正解の物体を指示者だけに伝える。

指示者はそれがどんな物体なのかを音声のみで受け手に伝える。

受け手は指示者からの指示を聞き、それがどの物体であるのか予想して答える。

受け手の回答が正解なら、それ以外ならへ。

～ の手順を複数回繰り返す。

2.2.3 結果・考察

途中結果として 6 人の被験者から 223 の発話を収集している。現時点で得られた表現の一部を以下に記す。

表現例 1 ... 「赤くて細長い」、「四角い」

表現例 2 ... 「右のほうにある」、「奥のほう」

表現例 3 ... 「結構でかい」、「小さめ」

表現例 4 ... 「その左」、「その手前」

表現例 1 のように、今回の調査でも前調査のような個々の物体を捉えた表現があった。場合によっては他の物体と違っている部分を指定した表現であるとも考えら

れるが、今回の調査ではその違いを明確にすることはできなかった。

表現例 2 は、「(テーブルの) 右のほうにある」や「(テーブルの) 奥のほう」といったテーブル全体を参照した表現であると考えられる。「テーブル」という言葉が省略されているのは、「テーブル」という言葉を言わなくても、受け手にはそれが伝わるという指示者の考えによるものであると推測できる。ゆえに、この省略された表現の理解もロボットにとって重要となる。

表現例 3 は、物体全体を比較・参照した表現であると考えられる。テーブルにある全ての物体を把握し、それらの物体と対象物体を比較したときの表現である。このことから、ロボットがこれらの表現に対応するには、周辺にある全ての物体の特徴を把握しておく必要がある。

表現例 4 は、受け手が答えた物体を参照した表現であると考えられる。この表現は、指示者の指示に対して受け手が答えた物体が間違いであったときに、そこからの訂正として使用されていた。ゆえに、ロボットがこれらの表現に対応するには、直前に認識した物体を記憶しておく機能を実装する必要がある。

以上のように、現時点では本来の目的である「(A から見て)B の C にある」といった表現を得ることはできなかった。表現例 4 が若干それに触れた表現ではあるが、B(reference object)は受け手が直前に答えた物体に限定されてしまっている。目的の表現が得られなかった原因については、テーブル上にあるどの物体も複雑な特徴を持っていたためであると考えている。B として複雑な物体を表現するより、単に対象の物体だけを表現したほうが簡単ということである。今後、この実験を何回か続けても目的の表現が得られない場合は、赤くて丸いものといった単純な物体を意図的に配置するなどした実験環境への変更を考えている。

3 対象物体の検出

ここでは 2.1 節で明らかになった最も割合の多い色または下地の色、という人間の捉える傾向にある色を画像内から検出する手法について述べる。

3.1 最も割合の多い色の検出

実験の結果から、人間は最も割合の多い色または下地の色を捉える傾向にあるということが明らかになった。しかし、そのどちらが使用されるのかについては明確にできなかった。そこでまず、人間は最も割合の多い色を表現すると仮定して、それに対応する色を抽出する手法を考えた。以下が提案するアルゴリズムである。

背景除去を行い、物体の領域を割り出す。

カラーセグメンテーションを用いて領域分割を行い、各色の占める面積を求める。

面積が最大となる色を最も割合の多い色とする。

図 8(a)を例とする．ここでは，主観的に赤色と黄色のどちらかが最も割合の多い色だと判断し，各色のカラーセグメンテーションを行った（実際にはアルゴリズムの，を行う）．その結果がそれぞれ図 8 の(b)と(c)になる．そして，これらの各色の占める面積を表したのが表 1 である（表内の数値は画像内の画素数に対応．画像の大きさは 320×240 ）赤色より黄色の面積のほうが大きいため，抽出結果は『黄色』となる．

実際に図 8(a)の物体に対して，調査で得られた色による表現は「黄色」のみであった．



図 8: カラーセグメンテーション 1

表 1: 分割された領域の面積 1

	segm_area
赤	4609
黄	14426

同様な処理を図 9(a)の物体に対しても行った(ここでも、主観的に赤色と黄色のどちらかが最も割合の多い色であると判断した). 各色のカラーセグメンテーション結果はそれぞれ図 9 の(b)と(c)であり、これらの各色の占める面積を表したのが表 2 である. 黄色より赤色の面積のほうが大きいため抽出結果は『赤色』となる. 調査の結果でも、1人だけが「赤くて黄色い」と表現しただけで、残りの表現は「赤」という1色での表現だった.



図9: カラーセグメンテーション2

表 2: 分割された領域の面積 2

	segm_area
赤	6712
黄	6661

以上より，このアルゴリズムを用いれば人間が捉える傾向にある色を抽出できると考えられる．しかし，ここで問題となるのが人間は最も割合の多い色というのをどこまで正確に認識できるかということである．表2では，赤色と黄色の面積の差はわずか 51[pixel]の差ではない．このわずかな差を人間が識別できるとは考え難い．

そのため、指示者によっては1色で「黄色い」と表現するケースも予想できるが、前述したように調査結果からそのような表現は得られなかった。ゆえに、最も割合の大きい色だけでは対象の認識は不十分となる。

3.2 下地の色の検出

次に、人間は下地の色を表現すると仮定する。下地の色ということは、もともと対象の物体は大部分がその色で覆われていたということである。そして、物体はその状態から文字や模様などといった色が上乗せされることになる。つまり、下地の色というのはもともとその物体の中で一番面積が大きかった色であるといえる。以上のことを踏まえ、下地の色を抽出する手法として以下のようなアルゴリズムを提案する。

背景除去を行い、物体の領域を割り出す。

カラーセグメンテーションで領域分割を行う。

で得られた各領域のエッジを抽出する。

で得られたエッジの特徴点を求める。

凸包(convex hull)を用いて の特徴点に対する
外周を得る.

で得られた外周内の面積が最大となる色を下地の色とする。

図 9(a)を例とする．カラーセグメンテーションを行った図 9 の(b)と(c)に対して、エッジを抽出し、その特徴点を取ったのがそれぞれ図 10(a)と図 11(a)である．それらの画像に対し、凸包を用いて外周を抽出したのが図 10(b)と図 11(b)であり、その外周内の領域が今回提案した下地領域の候補となる(図 10(c), 図 11(c))．そして、これらの領域の占める面積を表したのが表 3 である．黄色より赤色の面積のほうが大きいいため、下地の色の抽出結果は『赤色』となる．

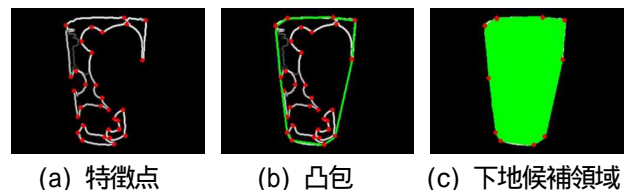


図 10: 『赤』に対する下地領域の検出過程

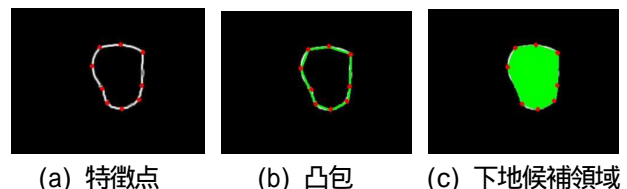


図 11: 『黄色』に対する下地領域の検出過程

表3: 下地候補領域の面積

	ground_area
赤	25574
黄	8621

しかし,この手法では図 12 のような物体があった場合うまくいかない.この物体を見たときに主観的な判断ではあるが,「赤」と表現する人がほとんどになると考えられる.だが,この画像に対して本アルゴリズムを適用すると青が抽出されることになる.下地の色としては正しい結果かもしれないが,このことから下地の色だけでは人間の用いる表現に基づいた対象の認識は困難となる.



図 12: サンプル画像

3.3 最も割合の多い色と下地の色の統合

最も割合の多い色と下地の色の両方を利用した手法として以下のようなアルゴリズムを提案する.

背景除去を行い,物体の領域を割り出す.

カラーセグメンテーションを用いて領域分割を行い,各色の占める面積を求める.

の面積が顕著に大きくなる色が1つだけなら,その色を記憶し. へ. 2つ以上あるなら. へ.

で得られた各色領域のエッジを抽出する.

で得られたエッジの特徴点を求める.

凸包(convex hull)を用いて. の特徴点に対する外周を得る.

で得られた外周内の面積が最大となる色を記憶する.

記憶した色を目的の色として処理を終了する.

つまり,まず始めに 3.1 で提案した最も割合の多い色を抽出するアルゴリズムを適用し,1つの色が抽出されればそれを検出結果とする.もし複数の色が抽出された場合には,3.2 で提案した下地の色を抽出するアルゴリズムを適用するという検出方法である.

例として,もし利用者から「黄色いの取って」という依頼があったときに,その付近に図 8(a)と図 9(a)の2つの物体があるなら,本アルゴリズムを利用することによって,ロボットは図 8(a)の物体を対象物体として認識する(表 4,表 5).

表 4: 図 8(a)の検出結果

	segm_area	ground_area	検出結果
赤	4609	-	×
黄	14426	-	

表 5: 図 9(a)の検出結果

	segm_area	ground_area	検出結果
赤	6712	25574	
黄	6661	8621	×

なお,下地の色の検出でも有意な差が見られない物体の場合には,ロボットは人間がその物体を2色以上で表現するものと判断する.例えば車の初心者マークだと,左半分が黄色で右半分が緑色であるため,最も割合の多い色の検出でも下地の色の検出でも有意な差は見られないと予測できる.この場合は,人間が「黄色と緑」という2色で表現するとロボットは判断する.人間から「黄色いの取って」という依頼があったとき,周辺に黄色に該当する物体があるなら,そちらの物体を優先して対象物体と認識し,周辺に黄色に該当する物体がないなら,初心者マークを対象物体と認識する.

3.4 課題

課題となるのが,各色の面積の差がどの程度大きくなったときに有意な差と判断するかであり,今後検討していく必要があると考えている.

また,人間には感知しやすい色とそうでない色があり,一般には赤や黄色などの暖色系の色は他の色より知覚し易く,青や黒などの寒色系の色は知覚されにくいとされている.日本の小学生のランドセルや帽子が黄色なのは,知覚し易い色を採用することで自動車等の事故を減らす狙いがあるものである.ゆえに,そういった指標が人間の表現する物体の色にも影響を与えているのか知る必要がある.

4 まとめ

本論文では,実用的な対話物体認識システムの実現を目指し,人間が用いる物体表現の調査を通して,利用者から色情報を含んだ依頼があった際に対象物体を検出する手法を提案した.今後は,提案した手法の有効性を検証していくのと同時に,今回の調査で得られたその他の表現にも着手していく予定である.

参考文献

- [1] 久野義徳: “サービスロボットのための視覚と対話の相互利用”, 情報処理学科論文誌, Vol. 47, No. SIG15(CVIM16), pp.22-34, (2006).
- [2] ザリヤナ=モハマド・ハナフィア, 山崎千寿, 中村明生, 久野義徳: “視覚によるサービスロボットのための簡略化発話の理解”, 電子情報通信学会論文誌, Vol. J88-D-11, No.3, pp.605-618, (2005).
- [3] 吉崎充敏, 中村明生, 久野義徳: “ユーザと環境に適応する指示物体認識のための視覚音声システム”, 日本ロボット学会誌, Vol.22, No.7, pp.901-910, (2004).
- [4] S.C. Levinson, P. Bloom, M. Peterson, L. Nadel, and M. Garrett, Eds.: “Frames of reference and Molyneux’s question: cross-linguistic evidence”, Language and Space, MIT Press, pp.109-169, (1996).