人間の物体と空間関係表現の調査に基づく対話物体認識

Object Recognition Based on Analysis of Human Description of Appearance and Spatial Relation

プロジェクト代表者：　久野義徳（大学院理工学研究科・教授）

Yoshinori Kuno

Professor, Graduate School of Science and Engineering

1　はじめに

頼んだものを取ってきてくれるような介護サービスロボットの実現のためには，頼まれたものを認識する機能が必要である．しかし，コンピュータによる物体認識は困難な問題で，どのような場合にも確実に動作する方法は実現されていない．そこで，頼まれたものを自動で認識しようとしてもできなかった場合に，人間にどんな物体かを聞いて，それを見つけることで物体認識を実現するという対話物体認識を検討している．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となる情報を与えてもらうという手法である．これまでに単純な色や形の物体を対象にしたシステムを開発し，その有効性を示した[2]-[5]．しかし，実際の生活環境に現れる物体は、複雑な色のパターンや形をしている。このような物体の検出に失敗したときに対話物体認識を使うためには，人間がこのような物体についてどのような表現をするか調べる必要がある．その表現にあたる部分を画像から検出する処理法を実現すれば，物体を検出できると考えられる．また，指示した物体を相手が認識できない場合に，他の物体との空間的な位置関係で物体を教える場合もある．これについても人間がどのような空間関係表現をするか調べ，それに対応する画像処理法を準備する必要がある．本プロジェクトでは，以上の考えに基づき，人間の物体や空間関係の表現法をロボットの視覚情報処理を開発するという観点から調査を行った．そして，人間の用いる表現に対応する部分を検出する画像処理について，基本的なものを開発した．

2　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)．前節で述べたように，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，物体名を言ってはいけないとしたときに，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られる．

図 1：対象物体

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，次のような実験を行った．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だけ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 に実験環境を示す．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ことができないようになっている．

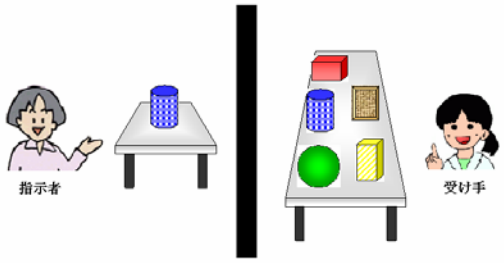


図 2：物体表現に関する実験

実験の手順は以下のとおりである．

① 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置く．
② 指示者はそれがどんな物体なのかを音声だけで受け手に伝える．
③ 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．
④ 受け手の回答が正解なら，この物体について

の実験は終了，それ以外なら②へ．
①～④の手順を複数回繰り返す．

以上のような実験を 10 人の被験者（指示者）に対して行い，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割合と，知識情報が含まれていた割合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，無知識情報にはど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る．したがって，自然な対話物体認識を行うためには，その１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．全体の 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，人間が表現に使う色は，物体の下地の色や最も割合の多い色であることがわかった．したがって，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では必要になる．

３ 空間関係表現の調査

前節の調査では，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[1]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．前の調査と同様に，２人１組の被験者を指示者と受

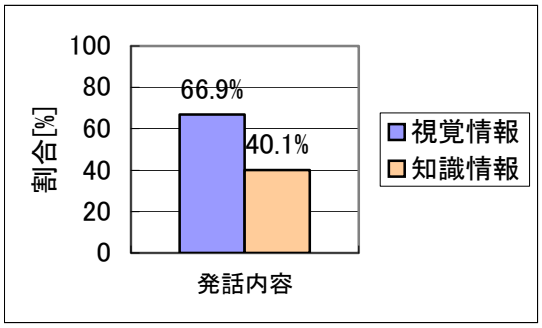


図 3：発話内容の内訳

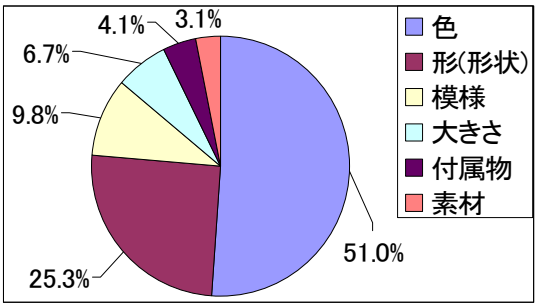


図 4：無知識情報の内訳

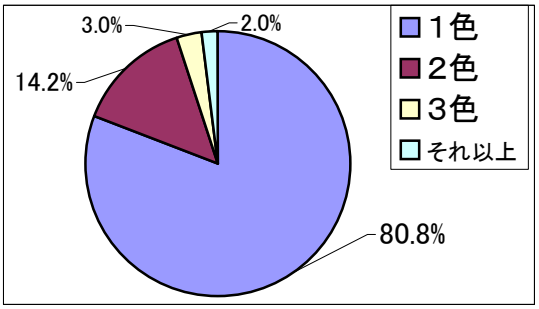


図 5：表現された色の種類

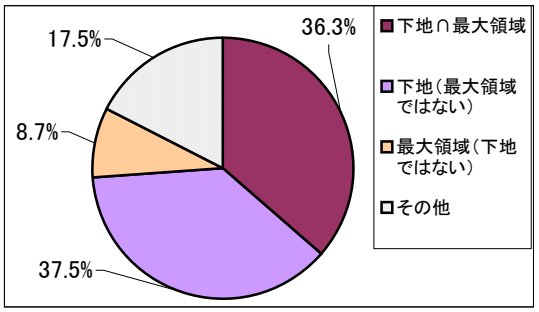


図 6：１色で表現された箇所

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った．図７に実験環境を示す．テーブルには図１の中の物体を 10 数個配置している．また，指示者には受け手の後ろに立ってもらうようにした．これにより，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，ジェスチャ等が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だけとなる．

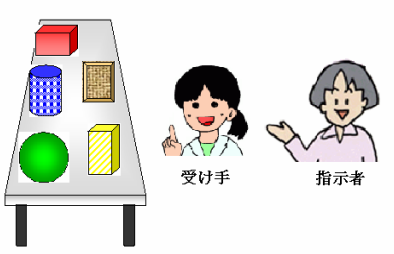


図 7：空間関係表現に関する実験

実験の手順は以下のとおりである．

① 正解の物体を指示者だけに伝える．
② 指示者はそれがどんな物体なのかを音声だけで受け手に伝える．
③ 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．
④ 受け手の回答が正解なら，この物体についての実験は終了，それ以外なら②へ．

①～④の手順を複数回繰り返す．

実験により 6 人の被験者（指示者）から 223 の発話を収集した．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 …「それの左」，「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．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．テーブルにあるすべての物体を把握し，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．したがって，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，現時点では本来の調査の目的である「(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がそれに関連した表現ではあるが，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．物体名を使うのを禁止したため，B となる可能性のある物体を簡単に表現できないため，対象物体を直接表現する方法をとったものと考えられる．空間関係表現については，さらに検討が必要である．

４　物体検出のための画像処理

物体表現の調査によれば，人間は多色の物体に対しても１色で指示することが多く，その１色は最も割合の多い色，または下地の色であるということがわかった．そこで，このような色の領域を検出する画像処理法を開発した．

最も割合の多い色と下地の色を検出するアルゴリズムを以下に示す．

① 背景除去を行い，物体の領域を割り出す．
② 色情報による領域分割（カラーセグメンテーション）を行い，各色の占める面積を求める．
③ ②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⑧へ．２つ以上あるなら④へ．
④ ③で得られた各色領域のエッジを抽出する．
⑤ ④で得られたエッジの特徴点を求める．
⑥ 凸包(convex hull)を用いて⑤の特徴点に対する外周を得る．
⑦ ⑥で得られた外周内の面積が最大となる色を記憶する．
⑧ 記憶した色を目的の色として処理を終了する．

このアルゴリズムでは，②，③で顕著に大きな面積を占める１つの色があれば，その色の領域を出力する．２つ以上の色が，ある程度大きな面積を占める場合には，④から⑦の処理により，その中から，

その色の領域を包む凸包を調べ，その囲む面積が最大のものを出力する．

図 8 に処理の 1 例を示す．この例では，『黄』領域が顕著に大きいので，それを出力する．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」だけであった．

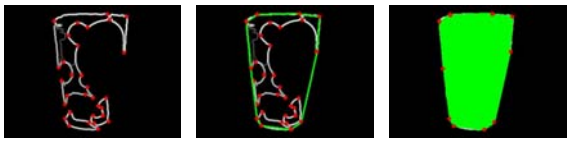
図 9 に別の例を示す．この例では『赤』領域の面積が 6712 画素，『黄』領域の面積が 6661 画素で，２つのほぼ等しい面積の色領域がある．そこで，凸包を調べる処理を行う．図 10 が『赤』領域，図 11 が『黄』領域に対する結果である．この場合，凸包で囲まれる部分(両図の(c))の面積が『赤』領域の方が大きいので，「赤色」が下地の色であると判定する．この物体については，人間の調査の結果でも，１人だけが「赤くて黄色い」と表現しただけで，他の 9 人は「赤」という１色での表現であった．

(a) 原画像　(b)『赤』領域　(c)『黄』領域

図 8：カラーセグメンテーション結果１

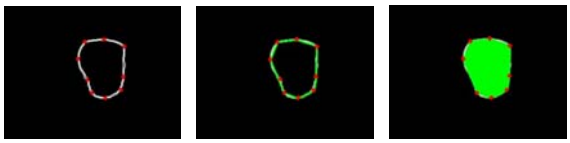
(a) 原画像　(b)『赤』領域　(c)『黄』領域

図 9：カラーセグメンテーション結果２



(a) 特徴点　(b) 凸包　(c) 下地候補領域

図 10：『赤』に対する下地領域の検出過程



(a) 特徴点　(b) 凸包　(c) 下地候補領域

図 11：『黄色』に対する下地領域の検出過程

なお，下地の色の検出でも面積に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．

５．まとめ

本研究では，実用的な対話物体認識システムの実現を目指して，人間が用いる物体表現と空間表現の調査を行った．物体表現については，多色の物体に対して人間は最も大きな割合を占める色，あるい下地の色の 1 色で表現することが多いことがわかった．そして，そのような色の領域を検出する画像処理法を開発した．空間関係については，人間の使用する表現の収集はできたが，対話物体認識に利用するためにはさらに検討が必要である．今後は空間関係表現について調査をさらに進めるとともに，さらに多様な物体表現に対応できる画像処理法を開発し，実用的な対話物体認識システムを実現していく．

参考文献・発表論文

[1] S.C. Levinson, “Frames of reference and Molyneux’s question: Cross-linguistic evidence,” P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett, Eds., Language and Space, MIT Press, pp.109-169, 1996.

[2] R. Kurnia, M. A. Hossain, A. Nakamura, and Y. Kuno, “Generation of efficient and user-friendly queries for helper robots to detect target objects,” Advanced Robotics, Vo.20, No.5, pp.499-517, 2006.

[3] M. A. Hossain, R. Kurnia, A. Nakamura, and Y. Kuno, “Interactive object recognition through hypothesis generation and confirmation,” IEICE Trans. Information and Systems, Vol.E89-D, No.7, pp.2197-2206, 2006.

[4] 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,” 情報処理学会論文誌：コンピュータビジョンとイメージメディア, Vol.47, No. SIG15(CVIM16), pp.22-34, 2006.

[5] R. Kurnia, M.A. Hossain, and Y. Kuno, “Use of spatial reference systems in interactive object recognition,” The Third Canadian Conference on Computer and Robot Vision, CD-ROM, 2006.

[6] H. Tsubota, H. Niwa, Y. Kuno, N. Akiya, and K. Yamazaki, “Recognition of objects indicated by deictic pronouns for helper robots,” SICE-ICASE International Joint Conference, pp.1437-1440, 2006.

[7] 坂田克俊, 久野義徳, “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” 第 13 回画像センシングシンポジウム講演論文集, CD-ROM, 2007.