

物体形状オントロジーに基づく 指示形状物体の認識

2014年3月

埼玉大学大学院理工学研究科(博士後期課程)

理工学専攻(主指導教員 久野 義徳)

森 智史

概要

頼まれたものを取ってきてくれるようなサービスロボットを実現するためには、物体認識の能力が必要である。物体認識の技術は進歩しているが、まだ複雑な環境で自動で確実に物体の認識を行うことは難しい。そこで、自動で認識できない場合には、ユーザから対象物の属性についての情報を、対話を通じて与えてもらうことで認識を行う対話物体認識を検討している。しかし、人間の表現は多様であり、同じものに対して様々な表現を用い、同じ表現でも状況により違った物体を指すことがある。そこで、そのような多様な表現を整理して、それに対応したシステムを実現するために、対話物体認識用のオントロジーを検討した。今回は、表現の多様性が特に顕著な形状属性について実際にオントロジーを構築し、対話による形状認識の実験システムを実現した。様々な日用品を用いた実験によりオントロジーに基づく認識の有効性を確認した。

第1章では、現状での世の中の技術の問題点を整理し、本研究の目的、背景、課題を示す。また、第2章以降には、問題のアプローチや実験方法、解決方法、実験評価等を示す。

第2章では、参加者を用いた会話実験で、どのような形状表現が使われるか調べた。人間同士の対話の観測実験(実験1)を行うために、サービスロボットに持ってきて欲しいと思う食べ物や文房具などの物体を集めた。テーブルに多数の生活用品をランダムに配置しておき、「指示者」と「回答者」を組とする。指示者は回答者に欲しい物の形状を伝える。回答者は依頼された物が何かを判断して、推定物を回答する。今回の実験1で、回答者には「形状を表現するように」と指示した。ただし、それ以上の細かい指示はしなかった。10組の参加者(全員、大学生)に対して伝達実験を行い、143の形状表現を収集した。

また、実験1を通じて、同じ表現でも異なる幾何学的形状を指す場合があることに気付いた。同じ言葉で表現される可能性のある物体が複数あった場合、その言葉ではどの物体が指示されるか調べることにした。例として「丸いもの」という表現について、Thurstoneの比較判断の法則に基づく一対比較法を用いて実験的に調べた(実験2)。57人の参加者(全員、大学生)の協力を得て、「丸い」に該当する形状の優先度を計測することができた。

実験1,2により収集した「人の形状表現の語句」と「ロボットが理解できる表現」を

結び付け、オントロジー（体系化）を作成した。オントロジーの構築には法造を用いた。この形状表現に関するオントロジーは、ロボットが人間の依頼内容を理解するための知識を支援するものとなる。

第3章では、収集した物体表現に基づいて作成した物体認識システムについて述べる。画像検出用のカメラは、Kinect（深度センサー付き赤外線カメラ、RGB カメラ）を用いる。Kinect の深度センサーにより、Kinect を用いて、物体の領域を検出し、対象物の領域を構成する画素の3次元座標を検出する。3次元座標データから、物体の幾何学形状を判定する。また、法線ベクトルを計算し、ベクトルの数と向きにより、物体を構成する面の数と向きを確認し、物体の凹凸部分を検出することを検討している。

また、第2章で作成した「人が表現に用いる形状特徴」を体系化したオントロジーの構成を元に、各々の形状特徴検出のプログラムの構成や関連性を整理することができた。

第4章では、第3章で作成した物体認識システムを用いた対話形状認識システムについて述べる。ロボットは、フランスの LDEBARAN Robotics 社が開発・生産している NAO を用いる。ロボット（もしくは PC）の視覚機能として、マイクロソフト社のカメラ Kinect を用いる。

実験にユーザに名前でも指示された物体に対して自動認識を行い、それが検出できない場合に対話物体認識に入るという統合物体認識システムの実現を目指して開発を進めている。そこでは形状だけでなく色や材質、それに位置関係など、その場の状況に応じた適当な属性について対話を行う方法を検討することを考えている。今回は、第2章で提案した「物体形状のオントロジー」の有効性を確認するために、「形状」に関する対話の部分に限った実験システムを開発した。「物体形状のオントロジー」を用いて、ロボットに人の命令を理解させ、人が求めている物体に該当する物をカメラ画像上から検出する。候補が複数ある場合には、第2章の実験2のように優先度をつける設定にする。

第5章では、対話物体認識システムの有用性を検証するために行った実験について述べる。2章の実験で用いた物体から3～5個程度を選び、テーブルの上に置き、形状に関する対話の実験を行った。以下に代表的な例を示す。ここでは、システム(S)は初めに「どんな形ですか」と聞く、以下では、それ以降のユーザの入力(U)とシステム(S)の対応を示す。実験で使用された対話回数と、認識の成功・失敗の割合を調べ、システムの性能を確認した。

最後に、第6章で上記の研究のまとめと今後の課題について述べる。

目次

第1章	序論.....	1
1.1	研究背景.....	1
1.2	関連研究とその問題点.....	6
1.2.1	統計的機械学習.....	6
1.2.1.1	1960年代の研究.....	6
1.2.1.2	1990年代以降の研究.....	7
1.2.1.3	既存の自動物体認識に対する問題.....	10
1.2.2	対話学習.....	13
1.2.3	オントロジーによる知識表現.....	14
1.2.3.1	システム開発におけるオントロジー.....	14
1.2.3.2	ロボットに導入するオントロジー.....	15
1.3	研究の目的.....	18
1.4	論文の構成オントロジーによる知識表現.....	19
第2章	人間の用いる表現に対応する物体形状オントロジー.....	22
2.1	人の用いる物体の表現の特徴（会話の特徴）.....	22
2.2	調査1：形状に関する人の言葉の特徴.....	26
2.2.1	概要.....	27
2.2.2	結果と考察.....	28
2.3	調査2：同一表現による異なる形状の表現.....	30
2.3.1	概要.....	30
2.3.2	結果と考察.....	31
2.4	対話物体認識用オントロジーの全体の構成.....	34
2.5	形状表現オントロジーの構成.....	37
第3章	形状認識の手法.....	42
3.1	3次元座標情報と視点方向による形状判定.....	43
3.2	法線ベクトルのグループ分けによる形状判定.....	47

3.3	形状特徴の検出.....	52
3.3.1	全体的な幾何学形状検出	52
3.3.2	全体的な非幾何学形状検出.....	55
3.3.2.1	凹凸面.....	55
3.3.2.2	文字形状検出.....	56
3.3.3	上面・側面形状.....	57
3.3.4	部分形状.....	58
3.3.4.1	角形状判定	58
3.3.4.2	穴の空いた物体	61
3.3.4.3	模様検出	62
3.3.5	物体の大きさの判定	63
第4章	オントロジーに基づいた対話認識認識システム構成.....	64
4.1	物体認識手法の構成	64
4.2	対話認識実験でのハードウェア構成	67
4.2.1	使用するロボット.....	67
4.2.2	使用するカメラ.....	69
4.2.3	使用するマイク.....	70
4.3	対話認識システムの構成	70
4.3.1	音声認識モジュール	71
4.3.2	言語理解モジュール	74
4.3.3	物体認識モジュール	75
4.3.4	音声発話モジュール	75
4.4	オントロジーによる言葉理解.....	76
4.5	画像形状検出.....	77
4.6	ロボットの動作	79
4.6.1	Nao の位置と関節角度	79
4.6.2	ワールド座標の取得.....	81

第5章	実験.....	86
5.1	基本動作確認実験	86
5.2	性能評価実験	89
5.3	対話認識失敗の例	94
5.4	対話認識失敗の対処方法.....	96
第6章	まとめ.....	98
謝辞	100
発表文献	101
参考文献	102

目次

図 1.1 : 現在活躍しているサービスロボット	4
図 1.2 : 人とのコミュニケーション時のロボット	4
図 1.3 : 人から頼まれた物を取ってくるロボットの様子	5
図 1.4 : 統計的機械学習(領域分割された各画像領域と単語の対応付け) ..	5
図 1.5 : SIFT による特徴検出	10
図 1.6 : 既存の物体認識の問題点	12
図 1.7 : オントロジーによる対象物の体系化の例	16
図 1.8 : 既存のオントロジーの対話認識の問題点	17
図 1.9 : 章の全体の構成	20
図 1.10 : 対話による物体の特徴検出プログラムの全体	20
図 1.11 : 対話物体認識の全体図	21
図 2.1 : 実験の環境	23
図 2.2 : 指示者・回答者間で行われた対話回数	23
図 2.3 : 指示者の 1 回の発話に含まれていた物体情報の数	24
図 2.4 : 視覚属性の内訳	25
図 2.5 : 実験に用いた物体	26
図 2.6 : 形状表現の収集実験	27
図 2.7 : 指示者の用いた形状表現の割合	29
図 2.8 : 指示者の用いた形状表現の例	29
図 2.9 : 丸みの度合いを計算するために使用した物体	30
図 2.10 : 図 2.9 に対応する実物体	31
図 2.11 : 丸みの度合い	34
図 2.12 : 対話物体認識用オントロジーの全体構成	35
図 2.13 : 形状の表現部分のオントロジー	39
図 2.14 : 対象表現のオントロジー	40
図 2.15 : 属性表現のオントロジー	41

図 3.1 : 物体と背景の分別	42
図 3.2 : 物体の 3 次元データの例	44
図 3.3 : 3 つの直交面に投影されたポイントデータ	45
図 3.4 : 丸いと思われる物体の順番+三方向から見たときの各々の形状	45
図 3.5 : 丸み度の計算式	46
図 3.6 : 形状判定	46
図 3.7 : 正 12 面体に対する法線ベクトル	48
図 3.8 : 物体の面検出と方向単位ヒストグラム	49
図 3.9 : 各物体の面分割と法線ヒストグラム	50
図 3.10 : 法線ベクトル方向による凹 (へこみ) の認識	51
図 3.11 : 3 次元幾何学形状検出の例 A	53
図 3.12 : 3 次元幾何学形状検出の例 B	54
図 3.13 : 法線ベクトル方向による凹凸の認識	55
図 3.14 : 文字形状の輪郭線検出	56
図 3.15 : 部分的な幾何学形状検出	57
図 3.16 : 角形状判定	58
図 3.17 : 角形状物体の検出	60
図 3.18 : セグメントとラベリングによる検出	61
図 3.19 : 法線ベクトルの向きによる検出	62
図 3.20 : 矩形模様検出	62
図 3.21 : 物体の高さ判定 (対象: 円柱)	63
図 3.22 : 細長い物を検出	63
図 4.1 : 提案する物体手法の構成	64
図 4.2 : 対話物体認識の全体像	65
図 4.3 : 形状検出手順のフローチャート	67
図 4.4 : ロボット NAO	68
図 4.5 : Kinect (Windows 用) 赤外線カメラ	69
図 4.6 : ECM-C10 の外観	70
図 4.7 : 対話認識システム構成	71

図 4.8 : 利用者の声を認識	72
図 4.9 : 音声対話用に作成した文法ファイルの一部	73
図 4.10 : 表 4.5 に対する形状の検出例.....	78
図 4.11 : Nao と Kinect の初期配置図.....	80
図 4.12 : 点 P のワールド座標上での位置.....	80
図 4.13 : 指さしの角度算出	81
図 4.14 : 点 P のワールド座標上での位置.....	82
図 4.15 : NAO の対話内容.....	83
図 4.16 : 水平方向で見たときの物体を指す様子	84
図 4.17 : 優先順位を考慮した対話形状認識	85
図 5.1 : 「上から見たら丸い」 の検出例	86
図 5.2 : 「丸いもの」 の検出例	87
図 5.3 : 「丸みのあるもの」 の検出例	88
図 5.4 : 実験 1 で用いた 3 シーン	90
図 5.5 : 対話の回数(1)	91
図 5.6 : 実験 2 で用いた 10 シーン	92
図 5.7 : 対話の回数(2)	93
図 5.8 : 失敗パターン 1	95
図 5.9 : 失敗パターン 2	95
図 5.10 : 失敗パターン 3	96
図 5.11 : PC (ロボット) の学習方法	97

表目次

表 2.1 : 度数行列 (参加者数 $N=57$)	32
表 2.2 : 比率行列 (p 行列)	33
表 2.3 : 尺度距離行列 (z 行列)	33
表 3.1 : 形状判別図	48
表 4.1 : NAO 仕様	68
表 4.2 : ECM-C10 の仕様	70
表 4.3 : 語彙データベースの一部	74
表 4.4 : 人の表現に対応した形状の表	76
表 4.5 : 対象物に対する特徴形状の情報のまとめ	77

1 章 序論

1.1 研究背景

21 世紀では、産業が成長するにつれて、ロボットが人間社会を支援するのに必要不可欠な存在となっている。手塚治虫氏が描いた科学マンガの「鉄腕アトム」のような電子頭脳を持つロボットの実現は、まだまだ遠い未来の話ではあるが、身近な世界において、一步一步、人の生活を支援するロボットの開発が進んでいる。はじめは、工場の生産ラインで使用される産業用ロボットが開発され、人間よりも迅速・正確に作業を行、日本の産業を大きく発展させた。このことで、ロボットの重要性を証明させ、次は、人が踏み込むことができない危険な場所（原発、宇宙、水中など）ロボットが開発された。宇宙を探索したり、危険な場所で作業して、被害にあっている人々を救助するのに役立っている。そして現在、人の生活を支援する身の回りのサービスロボット（会話、介護、医療、事務的な仕事するロボット等）に注目が寄せられている。サービスロボットの中でも、特に介護ロボットが必要とされている。

現在、日本は、高齢化社会に入り、高齢者が増加してゆくが、介護を行う人が増えるということは難しい。介護者には負担がかかり、介護を受ける人も生き活きた生活を行うことに遠慮がちとなることが予測される。世の中に役立つ福祉型ロボットがより必要とされている[1]。図 1.1 のようなロボット[107, 108, 109]は、友達や身寄りのない人に代わって、人の話し相手になってくれるロボットである。「21 世紀に向けて、夢のある製品を」というコンセプトを基に、人と共に暮らすために開発されたロボットである。主に、物体探索、人とのコミュニケーション、介護等といった機能が充実している。

介護用コミュニケーションロボットとしては、「PALRO (パルロ)」(富士ソフト)[107]が開発されている。PALROは、図1.2のように、人との会話の時の仕草、話す言葉、話しかけるタイミング、声が人間に近い。高齢者に人気が高い対話ロボットである。

「ASIMO (アシモ)」(本田)[108]は、顔と音声を同時に認識することによって、人間では難しい、複数人の発話を同時に聞き分けることができる対話認識機能が適用されている。図1.3のように、人が飲みたいものを言うと、ASIMOは人の言葉を理解し、依頼者が求めた飲み物を取りに行く。オーダーを受けたASIMOは、傍にいる人に

オーダーを伝え、トレイにドリンクをセットしてもらい受け取る。欠点として、ドリンクをセットするのに実際に人の手を借りているため、ドリンクの種類を分別して自分でセットするという物体認識がまだできていない。他にも会話ロボット「wakamaru (ワカマル)」(三菱重工)、「enon (エノン)」(富士通フロンテック株式会社)、お手伝いロボット「東芝買い物ロボット」(東芝) [109]、介護ロボット「RI-MAN (リーマン)」という人の生活を支援する優れたロボットがある。

我々もまた、ASIMO等のように、ユーザの指定する物を取ってくれるような対話物体認識のシステム開発を行っている。実現には、人との対話で「人の求めているもの」を正しく理解(対話認識)し、カメラ画像中から当てはまる物を検出する機能(物体認識)が必要である。

物体認識では、近年、画像からの物体認識に関しては、SIFT[2]のような局所特徴量を用いて対象物を画像中から検出する方法がある。また、図1.4のように、あらかじめ画像全体に対して数個のキーワードが付けられているCorel画像データベースを用いて、領域分割された画像への自動アノテーションを行う。そして、Normalized Cuts等で領域分割し、画像と単語の対応のみで領域と単語の対応付けがされていない学習データを用いて、領域分割された各画像領域と単語の対応付けを統計的に推定する手法が用いて、物体と背景を分類する統計的機械学習手法[4, 110]が進展しているが、画像情報だけでは、ロボットに物体認識を100%成功させることは困難である。

対象物の特徴点を指摘したパターン画像を何百種類も用意し、物体を形状認識させる手法もあるが、「あらかじめ与えられた特徴情報」でしか物体認識ができない。このために、誤認識をするようなケースや対象物体の特徴が学習されていないようなケースが生じる。例えば、福祉ロボットが働く介護施設や一般家庭では、家具等の大きな設置物の形状は固定的であるが、日用品のような物(例: テッシュ箱)は、製品の種類も多いため、サイズや色模様の種別があり、物の認識が複雑となるという問題点がある。また、置き方により、画像上での形状の見え方も変化する。

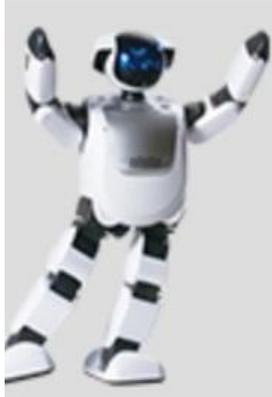
身体が不自由で、物を自由に取りに行くことが困難な人に対して、ロボットが代わりに物を取りに行くことを確実にすることが望まれる。ある状況によって、認識できたり、認識できなかったりということでは実用的とはいえない。そのため、物の理解ができない場合には、人がロボットに情報を与えることで、ロボットが正しく物の形状や特徴を理解し、物を認識できるように補助する必要がある。

対話認識では、人の言葉を正しく理解し、対話物体認識として「人から与えられた言葉」に当てはまる持物体を認識し、コミュニケーションとしてロボットが人に言葉を返さなければならない。ましてや、人の言葉表現は、多種多様で曖昧な表現がある。そのため、オントロジーという概念で、ロボットに学習・対話理解させる研究が行われている。

例として、Suhらは、オントロジーとルールによりロボットの知識フレームワーク OMRKF (Ontology-based Multi-layered Robot Knowledge Framework) [30]を構築し、その有用性を運搬タスクによって示している。小林らは、ロボットの実行可能動作を体系的にまとめ構築した動作オントロジーにより対話と動作の融合を実現した。また、日本語Wikipediaを利用した対話から、異機能ロボット連携による動作インストラクション、すなわち環境やタスクに適した身体的特徴を持つロボットが、それぞれ連携し、タスクを実行することを可能にした[31]。しかし、対話物体認識のためのオントロジーについては深く検討されたものはない。

また、対話により、画像による物体認識を補助することができる。例として、Maillotら[33]は視覚概念のオントロジーとして、画像分析の応用の専門家の知識とそれに対応する部分の画像認識処理の関係を体系化している。Dasiopoulouら[34]は対象シーンから物体を検出するための画像処理法をオントロジーとしてまとめている。

上記のことから、本研究では、「人から頼まれたものをオントロジーで認識し、画像中から検出する」研究を行っている。また、人の対話は曖昧で、該当する事柄がいくつも重なる場合があるが、今までの研究にはそれが触れられていない。そのため、本研究では、該当する事柄に優先順位を付けることも考慮する。



(a) PALRO

(文献[107]より引用)



(b) ASIMO

(文献[108]より引用)



(c) 東芝買い物ロボット

(文献[109]より引用)

図 1.1 : 現在活躍しているサービスロボット

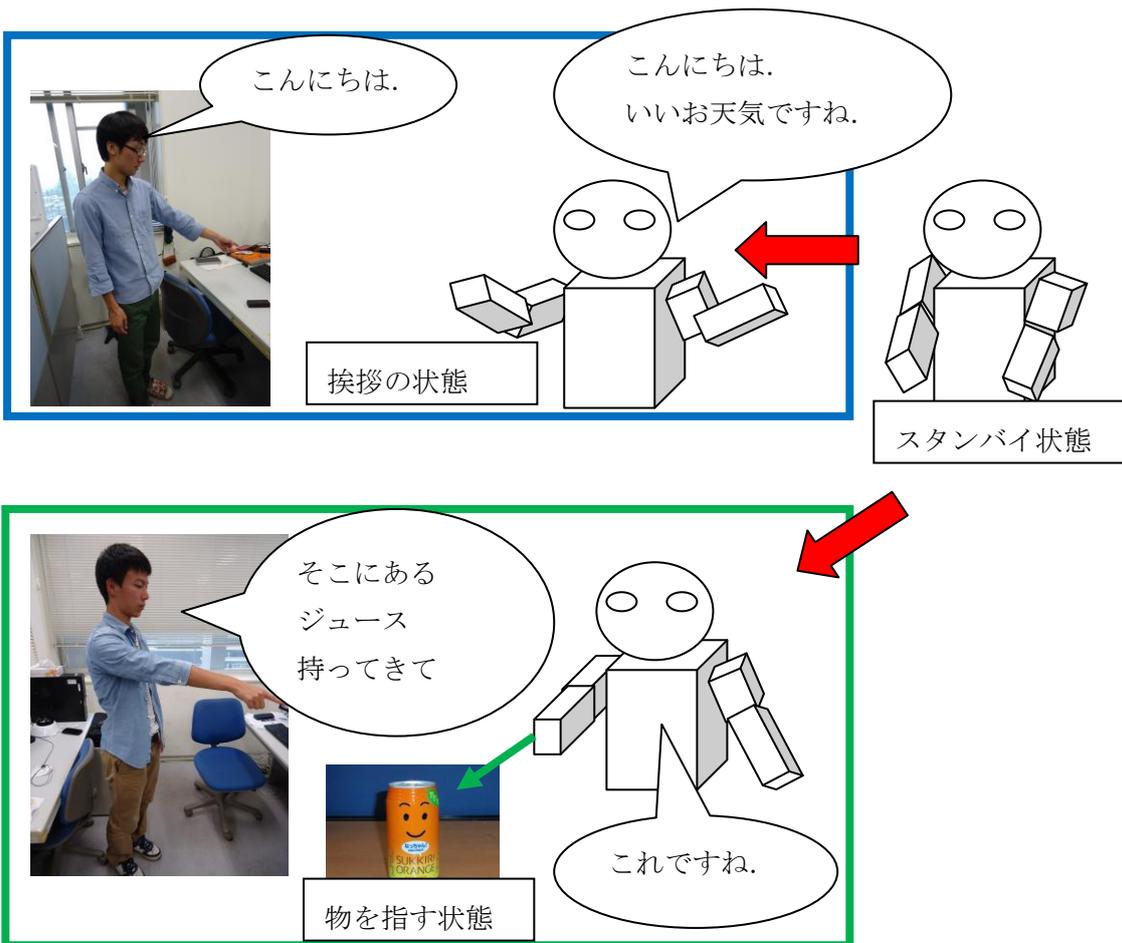


図 1.2 : 人とのコミュニケーション時のロボット

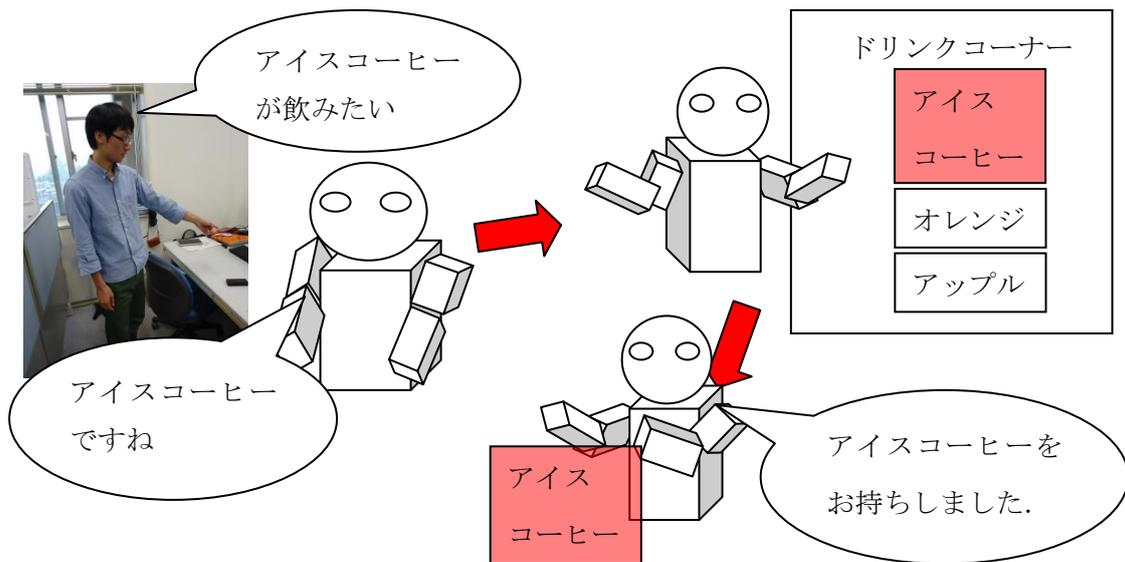


図 1.3 : 人から頼まれた物を取ってくるロボットの様子

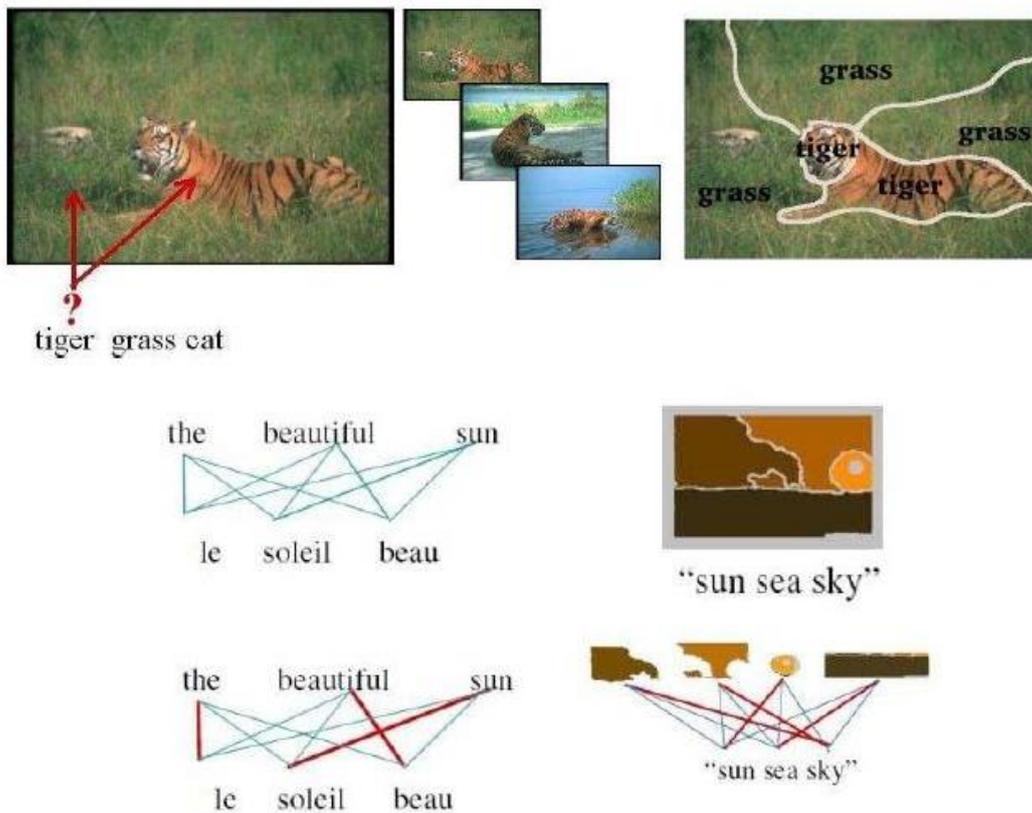


図 1.4 : 統計的機械学習
 (領域分割された各画像領域と単語の対応付け)
 (文献[110]より引用)

1.2 関連研究

関連研究として、自動物体認識、対話認識、オントロジーによる知識表現について記述する。

1.2.1 自動物体認識

1.2.1.1 1960年代の研究

物体認識技術はコンピュータビジョンの研究の主要なテーマの一つであり、一般物体認識もしくは一般画像認識は、画像認識の研究が始まった1960年代当初より研究が行なわれていた。しかし、当初より物体認識はとても困難な問題であることは認識されており、最初に成功を見た研究は、限定された世界『積木の世界』を対象としたものであった。その代表例の線画解釈[66]は多くの研究が行われたが、線画そのもの、もしくは容易に線画を得られる画像のみが対象となり、実世界の画像からいかに正しく線画を抽出するかに関しては問題が解決されることはなかった。

その後、一般に実世界を対象とする物体認識は、大きく分けて、**identification**(同定)と**classification**(分類)の2種類となった。**Identification**は、既知の物体を画像中から見つけ出す認識であり、入力画像とデータベース中のモデルの照合を行い、どのモデルに対応する物体が画像中に存在するかが出力となる。一方、**classification**は、画像中の未知の物体をそれが属するべき既知のクラスに分類する人閥である。画像中の物体とクラスとの対応付けの結果が出力となる。

実世界画像に対する研究として、2次元的な取扱いのできる画像、例えば、航空写真などの様な画像に対する理解システムが盛んに研究されるようになった。認識の方法は領域分割の延長線上にあり、同じ対象を表している領域を切り出して、その形状や色、模様、領域間の関係などを手掛かりにしてラベリングすることによって認識を実現していた。予め物体の完全な形状モデルが得られない場合の実世界シーンの認識は、古くはTenenbaumら[67]の領域分割した領域に対する緩和法によるラベリングによる認識があるが、こうした方法は非常に単純な方法であり、複雑な画像に対しては有効ではなかった。その後は、Draperら[68]などの画像中の物体毎に認識手法を用意する知識ベース型の画像理解システムが登場した。認識のためのモデルはルールとして表現されていたが、ルールは人手によって記述していたため、認識対象を増やすことが困難であると

いう。

当時の研究のほとんどが3次元画像を航空写真と同じ様に2次元的な画像として取り扱っており、領域分割を行なった後に、関係や構造の情報を利用してそれぞれの領域にラベリングを行い、認識を実現していた。このような方法では、初期の領域分割の結果が最後まで結果に影響してくることや、対象が3次元であるのにも拘らず、3次元的な取り扱いがなされていないという問題点があった。そのため、その後、3次元情報の復元が重視されるようになり、こうした領域分割+ラベリング規則の様な2次元的な物体認識の手法は下火となった[70]。その後、3次元の実世界を対象とする認識については、モデルベースによる物体認識の研究が盛んに行なわれた。モデルベース物体認識では、認識の対象とする物体の形状モデルを知識として予め用意しておいて、画像とモデルの照合を行うことにより、画像中にモデルの表す物体の存在を認識する方法である。モデルの表現の最も一般的な方法は、物体の3次元幾何形状をモデルとするものである。他にもパラメータによって形状モデルの形に幅を持たせることなども行なわれた[72]。

上記の認識の方法は、どの表現方法も物体の形状を直接認識に利用していた。そのため、認識する対象の形状が完全に既知でないと、正しい認識が不可能である。

identificationには向いているが、classificationに適用することことは困難である。例外的に、プロトタイプモデルによって、モデルベースアプローチでclassificationを目指した研究[73]があったものの、実際に実世界の画像を認識しようとする時、実世界に存在する物体の形状は無限ともいえる程あり、そのすべての形状が既知であることはあり得ず、また、海や道路などのように明確な形状を定義することできない物体も多く存在する等の問題を解決することは不可能であった。一方、違うアプローチからの手法として、物体の機能を推測して機能から物体を認識するfunction based recognition、物体の候補を複数出して物体間の関係によって最終的な結果の選択を行なう

Context-based recognition[74]、画像エキスパートシステム[75]などが提案されたが、結局ルールベースの認識手法には変わりなく、一般化することは出来なかった。

1.2.1.2 1990年代以降の研究

特に近年、SIFT[2]のような局所特徴量を用いて画像を記述し、統計的機械学習により分類する手法[4]の開発などにより、物体認識技術は急速に発展している(図1.5)。以下にいくつかの研究を紹介する。

相関テンプレートマッチングにおいては、対象物の回転やスケール変化などに対応するために、姿勢の異なる複数のテンプレート画像を何度も照合を行う。一方で、顔認証などでは、多数のテンプレート画像を主成分分析により圧縮して照合を行う固有顔法や、同様の考えに基づく固有空間法が知られている [5, 6, 7, 8].

一方で、SIFT などの特徴点検出法 [2, 9] においては、入力画像を異なるパラメータでぼやかして生成した多数枚のガウスぼけ画像（実際には、その差分である DoG 画像 [10, 11]）を生成してから特徴点を検出する (図 1.5)。他にも、L.Fei-Fei ら [16] は、局所パターンを SIFT 特徴量で表現し、13 クラスの学習画像 650 枚分すべての特徴量を k-means クラスタリングすることで 174 種類の **visual word** を作成し、確率的文書分類手法の LDA を用いることによって 13 種類のシーンを 64% の精度で分類することに成功している。Dipankar ら [94] は、対象物の輪郭線を検出し、検出した輪郭線所の局所特徴量から物体を検出している。他にも SIFT 特徴量を用いて交通標識を識別する研究 [100] もある。

他にも領域に基づく方法として有名なので Translation model がある。Duygulu, P (2002) の方法 [12] では、予め画像全体に対して数個のキーワードが付けられている Corel 画像データベースを用いて、領域分割された画像への自動アノテーションを実現している。

しかし、領域分割による方法では、オクルージョンがある場合や、形状が複雑で領域分割がうまくいかない場合への対処は難しい。そこで、M. C. Burl ら [13] は、局所領域の特徴とその位置関係を確率モデルで表現する **Constellation model** を提案している。そして、この方法をより多くの種類に対応可能として一般化させたのが、R. Fergus ら [14] の研究である。オペレータには特徴点周辺のパターンのスケール情報も出力される Kadir-Brady detector が用いられ、これによってある程度のスケール変化への対応も可能となった。また、**Constellation model** では局所領域の相対的位置の情報も確率モデル化しているが、局所領域の特徴量のみで認識を行う **Bag-of-keypoints** という方法も提案されている [15].

上記の手法は単一シーンを認識するのに用いられるが、実世界シーンの画像内には複数の物体が含まれ、それぞれが何らかの関係を持って存在しているのが普通である。そのため、最近では **context** を確率モデルによって表現し、学習によってモデルを構築する研究が行われるようになってきている。A. Torralba ら [17] は、確率モデルをグラフ

構造で表現するグラフィカルモデルを用いて、研究室 シーンの画像に対して机やキーボードなどの認識を行っている。また、D.Hoiem ら[18]は、消失点を用いた簡単な3次元復元を行い、ベイジアンネットワークによって視点位置、地面、空、垂直領域、歩行者、自動車の関係をモデル化し、街中のシーンの画像に対する歩行者や自動車を認識している。

距離濃淡画像と特徴量を用いてレジストレーションを実現している研究もある。Kurazums ら[19]は、距離濃淡画像中のエッジ点とカラー画像中のエッジ点間の誤差を、ロバストM推定を用いて最小化することにより位置合わせする手法を提案している。また Elstrom ら[20]は、まず距離濃淡画像とカラー画像中からそれぞれコーナー抽出し、コーナー点近傍の相関係数を計算することによりそれらの対応関係を決定する。そしてステレオ視の原理により対応点の奥行きを求め、それと距離画像から得られる奥行きが一致するように稜線さの相対位置姿勢を推定している。

Bohum ら[21]は、マッチングを行う特徴量の抽出に SIFT(Scale-Invariant Feature Transform)[22, 79]を利用している。Mikolajczyk ら[23]は、スケール・アフィン不変性を用いて、特徴点を検出している。猪股ら[24]は、SIFT は画像の拡大縮小、回転、オクルージョン、証明変化に対してロバストな特徴量の記述を行うため、高精度なマッチングを行うことができる。この手法では、SIFT を用いて距離濃淡画像とカラー画像の対応付け後、得られる対応関係から剛体変化を計算し、位置合わせを行っている。SIFT は、X線CTやMRIなどの装置で撮影される3次元画像の統合にも用いられており、SIFT 特徴量を3次元に拡張して用いる手法を提案されている[41]。

そして、現在知られている多くの手法の中で一番認識性能が優れた学習モデルの一つであるが、サポートベクターマシン(Support Vector Machine:SVM)である。SVMは高い性能を持ったクラス分類手法であるため、従来から様々な画像認識問題に応用されている[97]。K. Grauman らは'Pyramid Match Kernel[99]という2つのbag同士の部分マッチングに基づいて類似度を計算するカーネル関数を提案し、bag-of-keypoints approachにおいてSVMを用いた画像分類を行った。Lazebnikら[106] は、'Pyramid Match Kernel[99]に局所特徴の位置も考慮するように改良を加えたMatchingを提案しきわめて良い性能を示した。

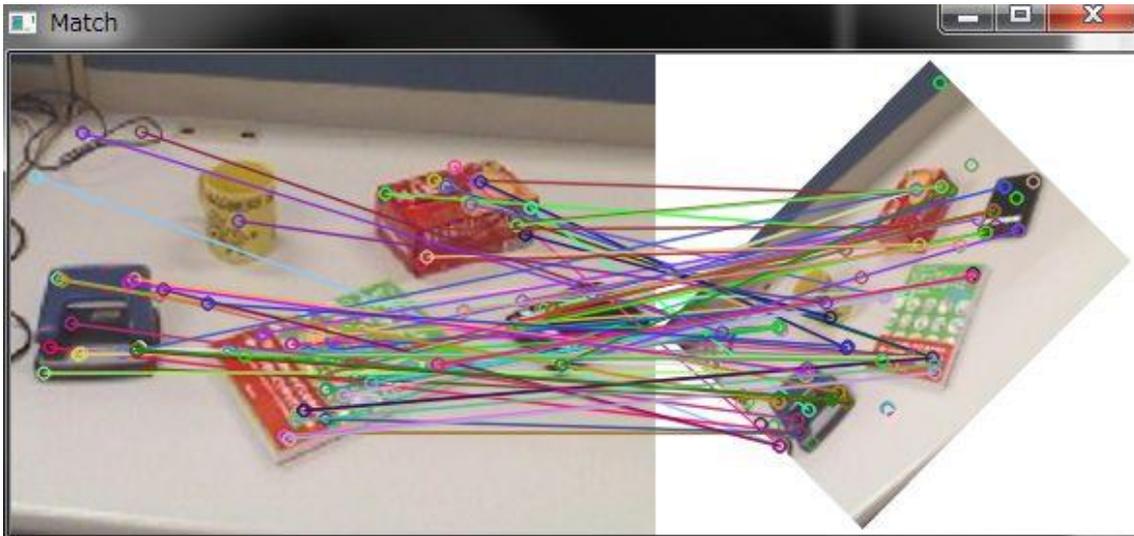


図 1.5 : SIFT による特徴検出

1.2.1.3 既存の自動物体認識に対する問題

物体認識に関する研究は、上で挙げたような手法が今後も進展していくものと予想される。しかし、現状ではそれらの手法だけで実世界にある全ての物体を正確に認識するのは非常に困難である。その一例を図 1.6 に示す。誤認識をするようなケースや対象物体の特徴が学習されていないような ケースが考えられる。また、ロボットが活動するような実環境においては、様々な条件の変化があり、福祉ロボットが働く介護施設や一般家庭のような変動の大きい環境では、その認識がさらに難しくなると考えられる。このように、画像情報だけで全ての認識を補完することはできないという問題が生じる。身体が不自由で、ものを取りに行けないような人にとっては、ある状況では正確に認識できるが別の状況では認識に失敗するといったシステムでは実用的とはいえない。そのため、こういった自動での認識に失敗した場合には、画像データ以外の何らかの情報を与えることによってシステムを補助する必要がある。

解決策として、久野研究室では、ロボットが人との対話により協調して対象物体を認識する手法を提案している[3]。これは、ロボットが自動的な物体認識に失敗した場合に、対話によりユーザから物体に関する情報を教えてもらい、物体を認識するという対話を援用して物体認識を行うものである。また、本研究で最も参考にするには、近年のコンピュータの処理能力やビデオ機器の発達 にもない、多視点カメラによって撮影された映像データをコンピュータ内部で融合し、その情報を用いて自由視点からの見え

方を提示する研究 [45, 50, 52, 53] である。多視点映像データの融合には、3次元形状推定やモデリングといったコンピュータビジョンの技術が用いられ、自由視点映像の生成には、イメージベースまたはモデルベースのレンダリングに基づくコンピュータグラフィックス技術が用いられている。R.T.Collinsら [54] と同様、3次元形状を複数の平面上の形状、つまり2次元形状の集合として定義する手法である。N. Dalalら [55] は、勾配方向ヒストグラム (HOG) を使用し、粗い形状を描写する研究を行っている。本研究の形状認識でも、[54,56] のように、3次元形状を2次元形状の集合として定義し、法線ベクトルによる形状を判定することを検討している。

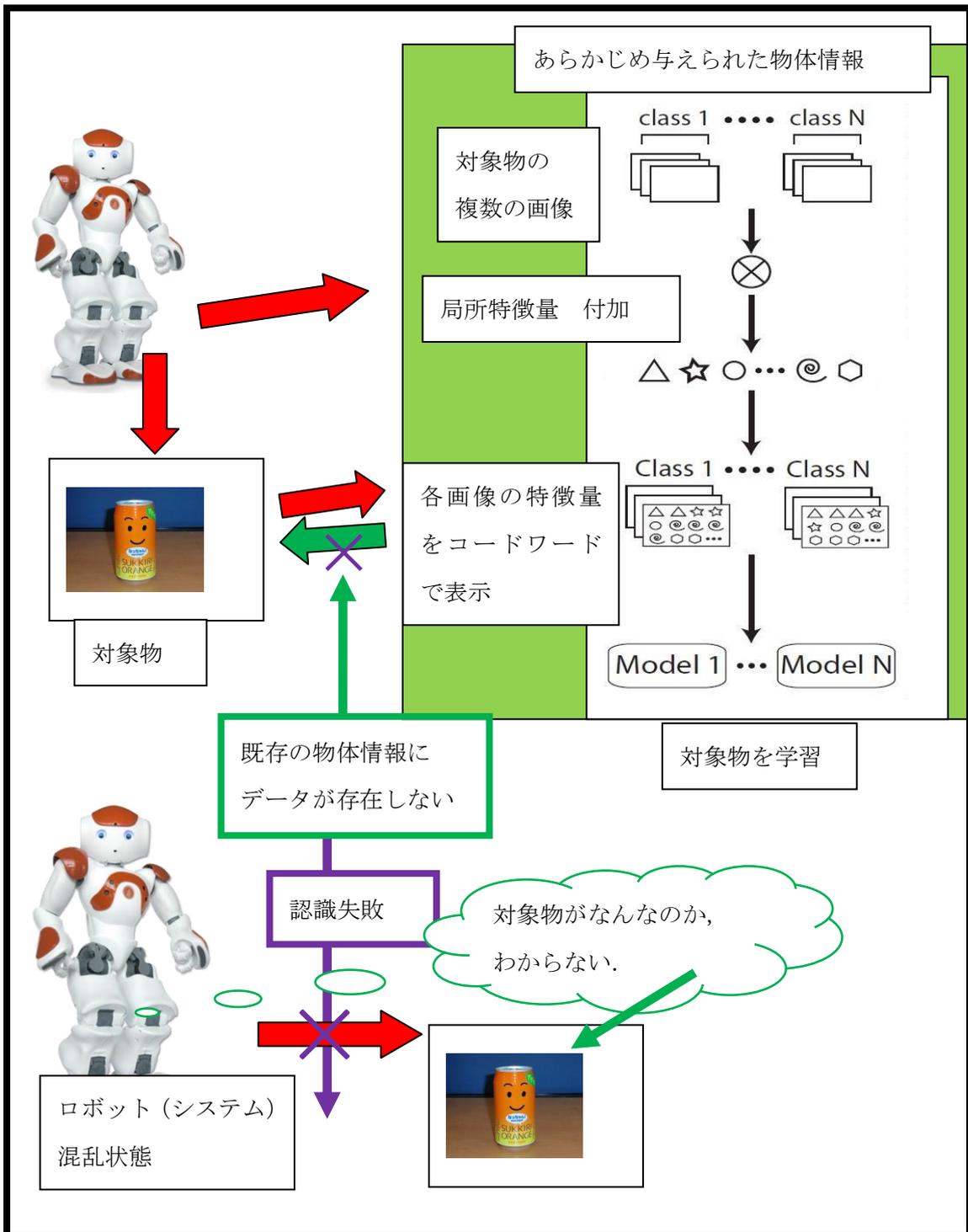


図 1.6 : 既存の物体認識の問題点

1.2.2 対話認識

Winograd ら[25]の研究以来、ユーザとの対話を用いてシーンやタスクの認識を行う研究は多く報告されている。しかしこれらは、コンピュータビジョンの物体認識が中心課題の研究ではなく、また簡単なシーンを想定している。ビジョンを用いた研究としては、見せられた物体の名前、カテゴリーを学習しようという研究が多い。Roy ら[26]は指さしをしながら「青い四角形」と発話することにより、「青」や「四角形」という概念を学習させるコンセプトを提案している。しかし、それらについても簡単なシーンであることや、物体が画像中で大きく捉えられているなどの制約がある。我々が考えているような物体認識に関しては、滝澤らの研究がある[27]。ここでは、病院や家庭で冷蔵庫から指定した飲み物（缶、瓶、ペットボトル）を取ってくるサービスロボットを開発している。そこでは、認識に必要な情報をユーザとの対話によって取得し、物体認識を可能にする対話システムを提案している。しかし、対話の内容は簡単なものに限定されている。

これらの研究に対し、Kuno ら[111]は、実際に人間が用いる物体の表現を調査し、それに対応できる対話物体認識の研究を進めている。久野研究室の「人間の用いる表現に基づく対話を援用した物体認識に関する研究」[3]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た際に必要となる物体認識の機能において、統計的学習による手法だけでは全ての物体を正確に認識できないという問題に着目し、その解決策として利用者との対話を援用した認識方法を提案した。目的のシステムを実現させるために、人間同士の会話実験を行うことで、人が物体を言葉でどのように表現しているのかを調べ、その分析結果に基づいた物体認識処理を作成した。その結果視覚属性として、色、形、大きさなどといった計6種類の表現が得られた。しかし、いきなりこれら6種類全ての表現に取り組むのは困難であるため、研究ではその中でも一番多く利用されていた色に関する表現に焦点を当てた。

例えば、菓子のパッケージには多くの色が使われているが、人間は背景色、あるいは使われている面積の大きい主要な色の一つ選択して表現することが多い。そこで、このような表現に対応できるシステムを提案している。その後、作成した色認識プログラムを実際のロボットに実装し、さらに音声認識やモーション再生などの機能を加えることによって、利用者との対話を行いながら目的の物体を見つけ出す、という一連の動作をこなすシステムを開発した。さらに形や材質などの属性についても検討し、初期的な統合

システムを開発している[28].

1.2.3 オントロジーによる知識表現

1.2.2で述べたような「対話認識に関する知識の表現」を検討する必要がある. 知識表現として, オントロジーが注目されている. 以下の図1.7に示すのは「台所にある物体」の属性を体系化したものである. 例として, コーヒーとお茶ならば, 熱い飲み物(hot drink)で, 飲むためのもの(drinkable)と定義できる.

1.2.3.1 システム開発におけるオントロジー

教育における情報技術活用が発展するため, 多様なシステム開発がおこなわれている. しかし, それぞれのシステムの多くは, 個別に開発されているのが現状である. システム間の連携が進まず, 生じる相互運用性(Interoperability)の問題を解決するために, 平田ら[49]は, オントロジーで, 体系化情報内容の体系化をして相互作用を行っている. 金来ら[51]は先行研究において電力系統事故復旧操作訓練システム構築用オーサリングツールSmart-Trainer/ATを開発した. その研究においては, 電力系統事故復旧操作の訓練システムを対象を特化したオントロジーアウェアな設計環境を開発した. 林ら[47]は, この開発で得られた知見を出発点として, より汎用的な枠組みとして学習コンテンツ設計支援環境*iDesigner*の開発を進めてきた.

溝口ら[64, 102, 103, 104]は, オントロジー工学的手法[56, 57, 61]を採用し, 学習内容と学習プロセスに関する概念をオントロジー化することによって解決を目指している. また, 情報基盤技術として設計に関わる活動をモデル化しオントロジーを策定している[48]. 稲葉ら[65]は, 協調学習における概念や概念間の関係を明示的・体系的に記述した協調学習オントロジーを作成している. 森田ら[58]は, 日本語専門文書を対象とした意味検索や計算機による意味解釈を行うために, 日本語概念を対象とした領域オントロジー構築支援環境である日本語DODDLEを提案した.

他にも, ソフトウェアシステムのユーザーと開発者との共通理解のために, ビジネスオントロジー[59, 60]や, ソフトウェアシステムの要求獲得を行うために, Webマイニングの技術を用いてドメインオントロジ(システムを適用する問題領域の知識)を取得する研究がある[69]. 他にも, タスクテンプレートとエージェントテンプレートを導入

してこれらの利用と具象化などをオントロジーを用いることにより、各レベルでの開発支援を統合的に行う手法[76]や、ソフトウェアエージェントやエキスパートシステムのような知識システムの開発を支援する手法[77]を提案されている。

1.2.3.2 ロボットに導入するオントロジー

上記のようにシステム開発の際に、システム間の相互作用をとるためにオントロジーを用いていたが、最近では、ロボットの知能処理にもオントロジーが使用されている。

Suhらは、オントロジーとルールによりロボットの知識フレームワークOMRKF (Ontology-based Multi-layered Robot Knowledge Framework) [30]を構築し、その有用性を運搬タスクによって示している。小林ら等は、ロボットの実行可能動作を体系的にまとめ構築した動作オントロジーにより対話と動作の融合を実現した。また、日本語Wikipediaを利用した対話から、異機能ロボット連携による動作インストラクション、すなわち環境やタスクに適した身体的特徴を持つロボットが、それぞれ連携し、タスクを実行することを可能にした[31, 83, 84, 87, 90]。

しかし、対話物体認識のためのオントロジーについては深く検討されたものはない。

対話物体認識の対象となる物体は物理的存在であり、その表現の対象になる属性の概念についてのオントロジーに関しては統一的なものが構築できるはずである。しかし、先に例を述べたように、人間の表現については、一つの表現が多くの属性の概念を指すのに用いられることがある。また、一つの属性の概念を指すのに多くの表現がある場合もある。

対話物体認識用オントロジーとしては、このように対象表現の体系化とともに、実際に用いられる多様な言葉や概念を整理する必要がある。ここでは、物体の形状表現について、この点を検討する。例えば、以下の図 1.8 のように、人間は球も円盤も「丸いもの」と言うことがある。また、球に対しては、その形状を典型として持つ「ボール」という表現を使うこともある。

上記のように、ロボットの扱う対象物をオントロジーにより用途別に体系化して、ロボットを支援する研究[32, 85]も行われている。Maillotら[33]は視覚概念のオントロジーとして、画像分析の応用の専門家の知識とそれに対応する部分の画像認識処理の関係を体系化している。Dasiopoulouら[34]は対象シーンから物体を検出するための画像処理法をオントロジーとしてまとめている。これらの研究では、対象世界に存在する物

体を物体の名称というシンボルレベルで体系化したり，物体を認識するために必要な画像処理に関する知識をオントロジーにより表現し[71, 91]，ロボットに知識を与えることができる[86, 88, 89, 92, 105]．

上記の研究に対し，本研究ではロボットが人間による自然言語表現の多様性に対応するための知識を表現するオントロジーを提案する．物体の特徴として形状に注目し，法造[42]を用いて，Koujiら[63]のように，オントロジーで定義した概念とそれらの関係を示すオントロジーエディタを作成する．形状の概念に関するオントロジーエディタを作成し，特徴検出プログラムも体系化することができ，ロボットに形状の構成を理解させることを検討している．

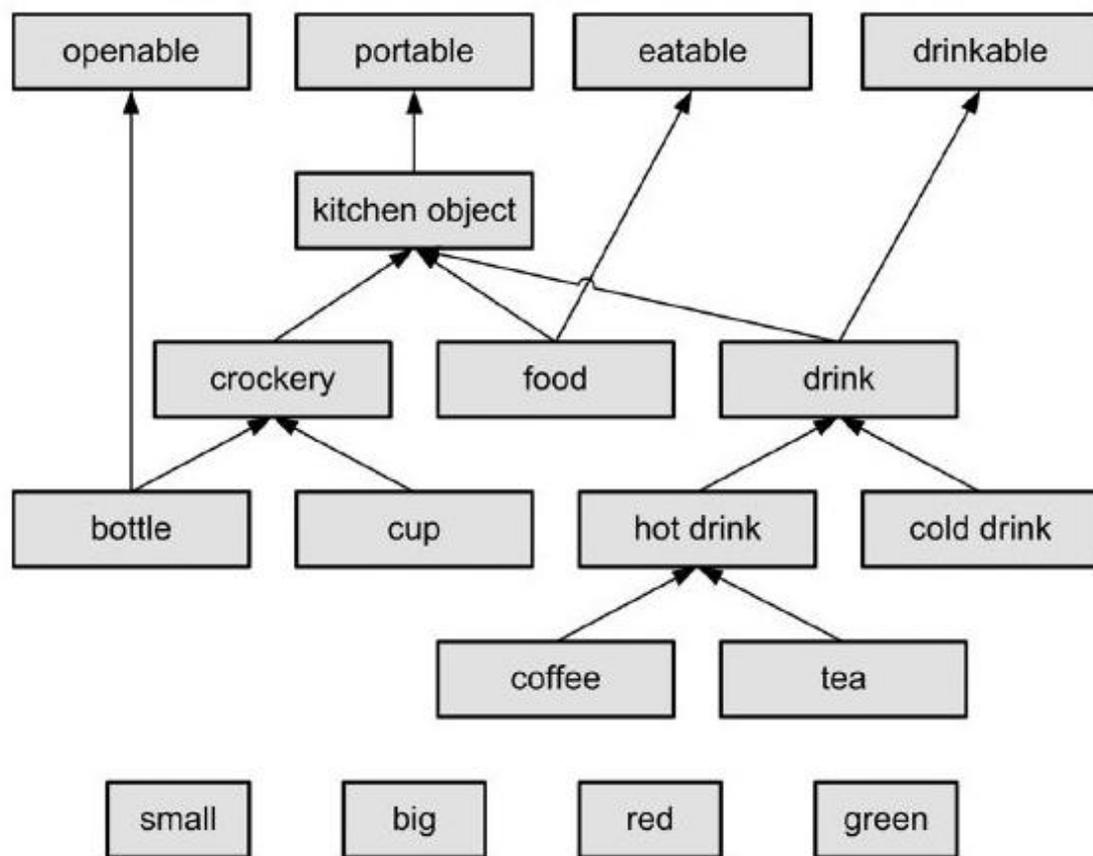


図 1.7 : オントロジーによる対象物の体系化の例
(文献[32]より引用)

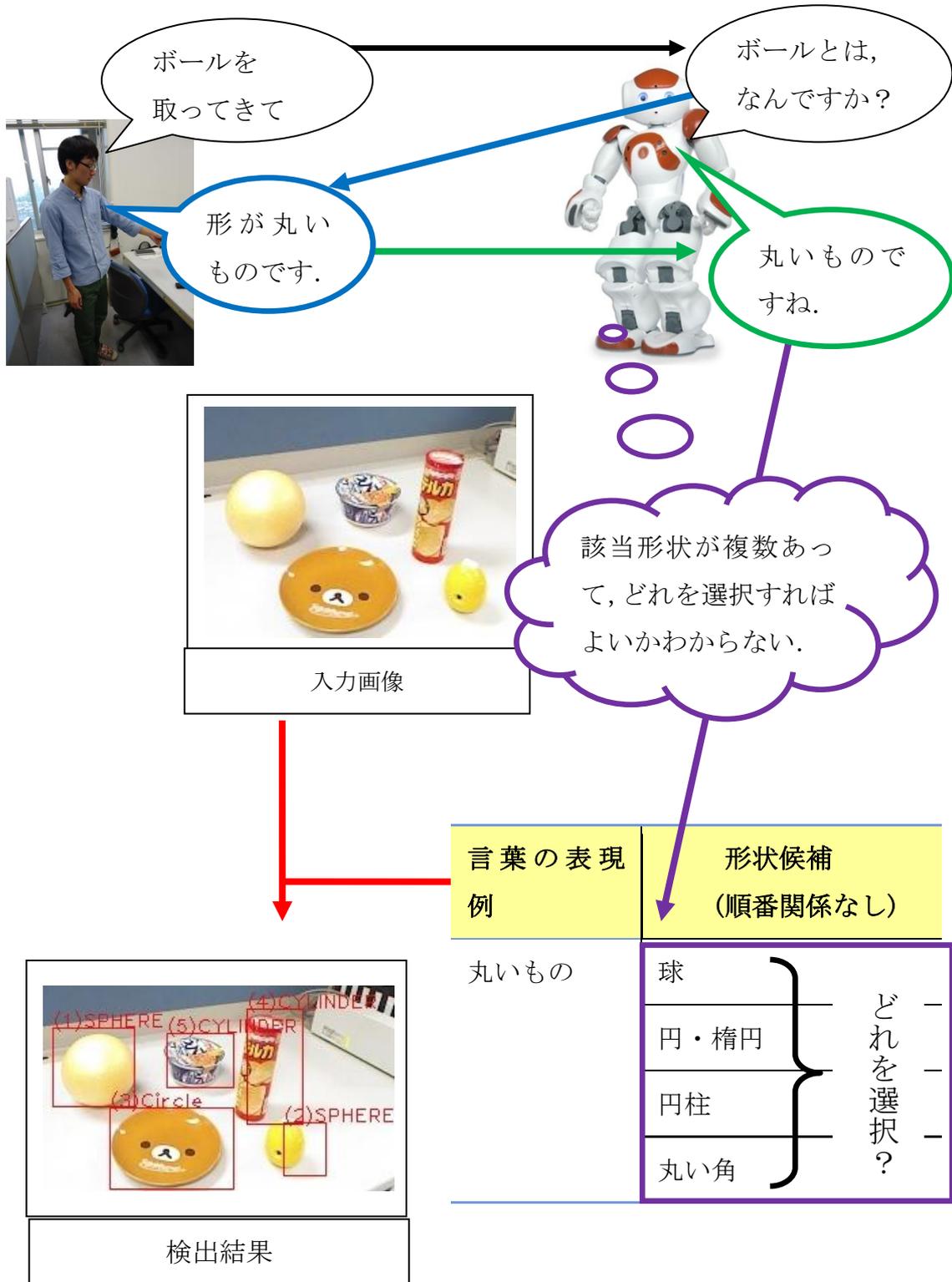


図 1.8 : 既存のオントロジーの対話認識の問題点

1.3 研究の目的

本研究は、ユーザの指定する物を取ってくれるようなサービスロボットの開発を目的としている。このようなロボットシステムの実現には対象の物体を正しく認識するための物体認識の技術が不可欠である。

物の特徴として形状に注目する。人の形状表現を収集し、オントロジーを用いて、ロボットが理解できる対話認識を行う。また、物体の形状の構成をオントロジーで体系化・整理し、ロボットに、ユーザの求めている物体の形状を理解させ、理解した形状に該当する形状をカメラ画像中から検出することを目指している。

[1] まず、人が日常生活で使用する形状表現を調査・収集する必要がある。参加者を用いた会話実験で、どのような形状表現が使われるか調べた。人間同士の対話の観測実験を行うために、サービスロボットに持ってきて欲しいと思う食べ物や文房具などの物体を集めた。テーブルに多数の生活用品をランダムに配置しておき、「指示者」と「回答者」を組とする。指示者は回答者に欲しい物の形状を伝える。回答者は依頼された物が何かを判断して、推定物を回答する。収集した「人の形状表現」に関する語句と、ロボットが理解できる表現を結びつけて体系化(オントロジー化)する。体系化した形状表現に関するオントロジーをデータにして整理し、ロボットの知識・辞書とする。

[2] 次に、ロボット(PCで構成)の視覚機能としてkinect(深度センサー付き赤外線カメラ, RGBカメラ)を用いる。Kinectの深度センサーにより、Kinectを用いて、物体の領域を検出し、対象物の領域を構成する画素の3次元座標から対象物の形状を求める。法線ベクトルも計算し、ベクトルの数と向きにより、対象物体の形状、物体を構成する面の数と向きを確認する。各々の「特定の形状特徴を検出するプログラム」を[1]のオントロジーと同様の構成で体系化・整理できる。

[3] 作成した「形状表現に関するオントロジーのデータ」と「物体形状認識システム」を組み込んだ対話物体認識のシステム開発を作成する。音声認識には、Julius[98]を用いる。ユーザに対象物の形状を指示され、該当する物体に対して画像認識を行う。該当する物体が複数あれば、物体の順位付けをする。

[4] 最後に、システムの評価として、処理速度、ユーザが求める形状の物体を画像中から検出した際のシステムの精度から評価を行う。

1.4 論文の構成

本論文の構成は以下のとおりである。以下の図 1.9 に 2～5 章の関連・構成を表示する。

第1章 序論

第2章 人の言葉の理解

第3章 提案する物体認識手法

第4章 対話認識実験でのハードウェア構成

第5章 実験

第6章 まとめ

第2章：人は物体をどのように表現するかの調査実験（人間同士の会話実験）について述べており，物体の形状の言葉表現を調査対象としている。

第3章：収集した物体表現に基づいて作成した物体認識システムについて述べる。本来なら図 1.10 のように様々な形状特徴を使用するが今回は形状情報のみ（前回の久野研の研究[3]は色情報のみ）を用いる。

第4章：使用したハードウェアと認識方法を述べ，図 1.11 のような NAO を用いた対話物体認識を行う。

第5章：対話物体認識システムの有用性を検証するために行った実験について述べる。

第6章：研究のまとめと今後の課題について述べる。

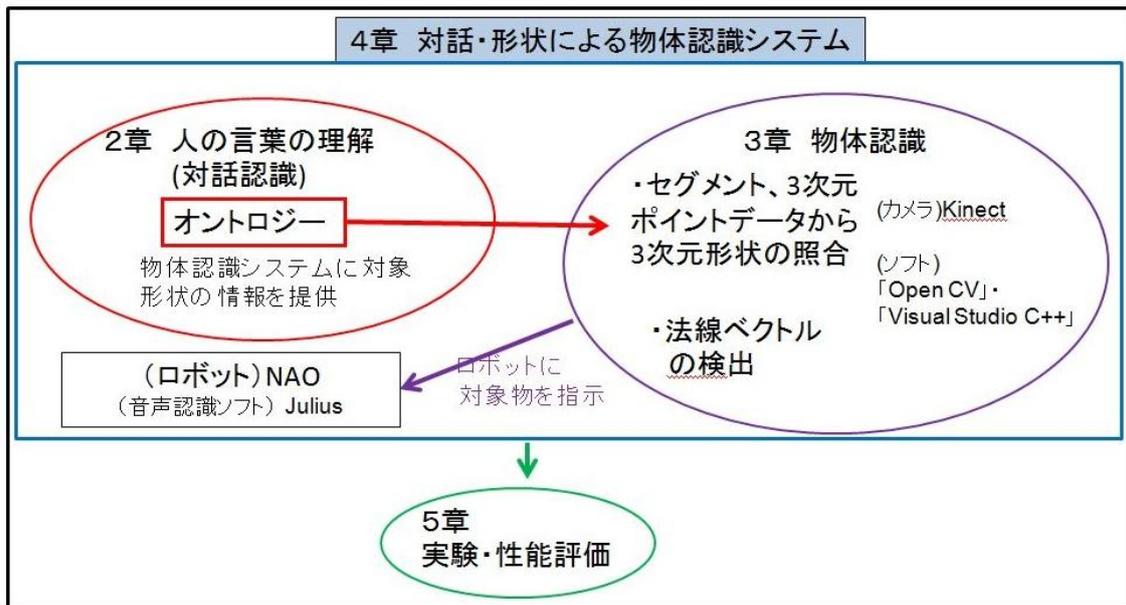


図 1.9 : 章の全体の構成

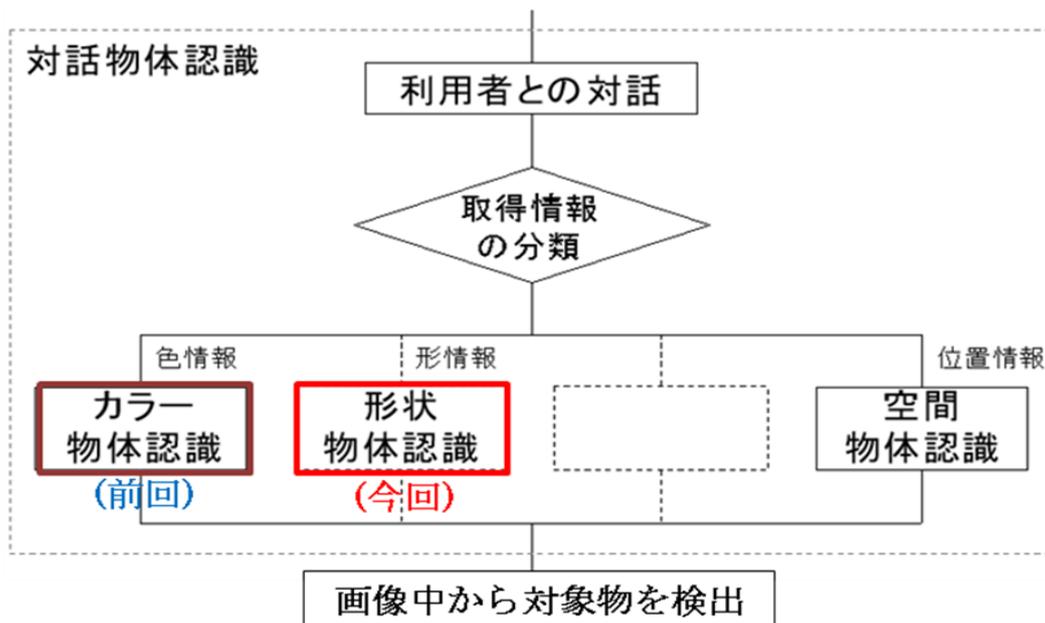


図 1.10 : 対話による物体の特徴検出プログラムの全体

ロボットが「人の形状表現」を理解して物体を認識

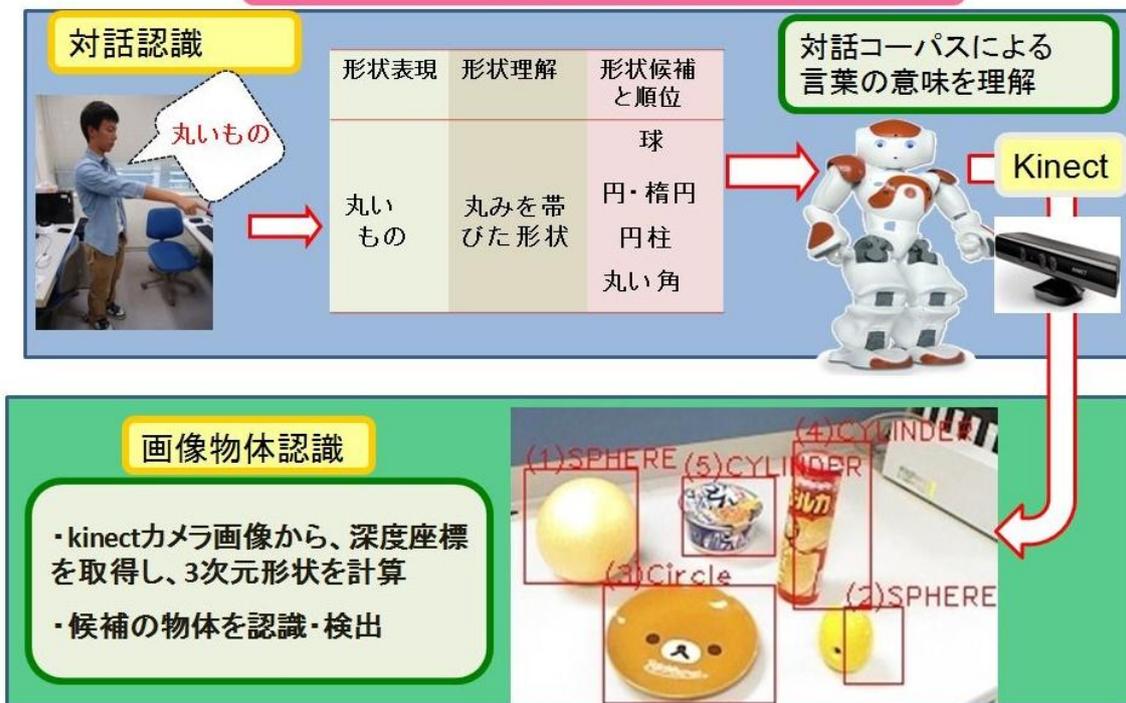


図 1.11 : 対話物体認識の全体図

2章 人間が用いる物体表現の調査

この章では、物体への表現を収集するために行った人間同士の会話実験について述べる。

1章でも述べたように、対話による物体認識を行わせるためには、まず人が物体を言葉でどう表現する傾向にあるのかを知る必要がある。そこで、食べ物や飲み物や雑貨などといった実生活で見られる物体を実際に数十個用意し、複数人の参加者にそれらの物体を表現してもらうことによってその傾向を調べた。このようにして得られるデータは、単なる言葉の集合として扱われるだけでなく、自然な対話を実現する上で重要となる会話の流れにも配慮した形で扱うことができると考えられる。

2.1で、坂田[3]は、特徴として最も多く用いられる「色」に焦点を当てて対話物体認識を行った。2.2以降で、本研究は、2番目に多く用いられている「形」(図2.4)に関する対話物体認識の研究についての説明をする。

2.1 人の用いる物体の表現の特徴（会話の特徴）

坂田[3]は、人同士の対話実験を行い、人の用いる「物体の属性に関する表現」を収集した。方法として、参加者を2人1組にして、指示者と回答者に分け、指示者には提示した物体を声のみで回答者に伝えてもらい、その際に発せられた言葉を収集する。

以下の図2.1に調査の環境を示す。指示者と回答者の前にそれぞれテーブルを置き、回答者側のテーブルの上には20個程度の物体を配置する。また、互いのテーブルの間に壁を挟むことによって、相手に伝えることのできる情報を音声のみとしている。



(a) 全体



(b) 回答者側のテーブル

図 2.1 : 実験の環境 (文献[3]より引用)

実験の結果、10人の参加者から227の発話を収集した。以下の図2.2は、回答者が対象の物体を見つけ出すまでに、指示者と回答者の間で、対話が何回行われていたのかを表したグラフである。大半が1回または2回の対話で対象の物体を見つけることができる。この結果から、できるだけ少ない対話数で、目的の物体を見つけ出せるような「対話物体認識システム」が必要となることがわかった。

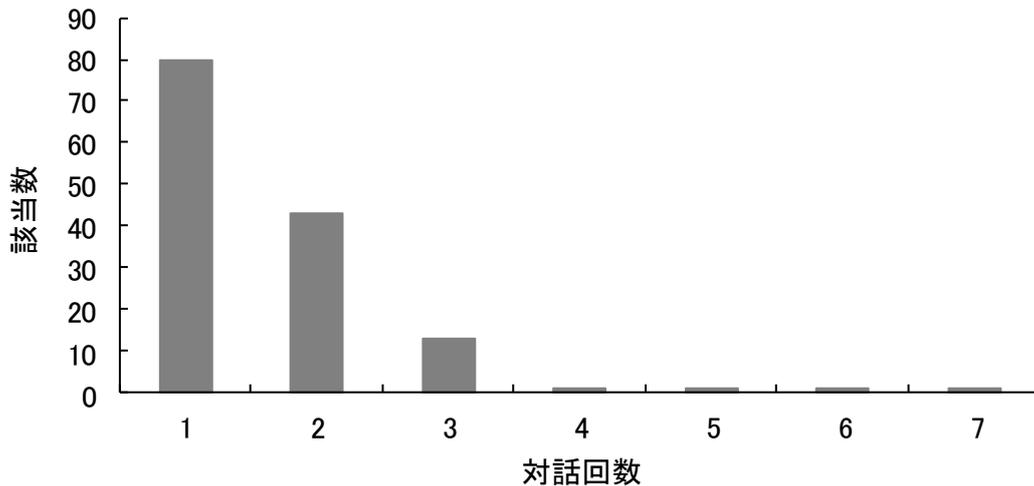


図 2.2 : 指示者・回答者間で行われた対話回数

(文献[3]より引用)

以下の図 2.3 は、指示者の 1 回の発話内に、対象物体に関する情報が何種類含まれていたのかを表したグラフである。グラフを見ると、物体情報の数が少ない発話ほど、最もよく使用されていることがわかった。目的の物体を正確に伝えるためには、できるだけ多くの情報を与えるほうがよいと考えられるが、一度に多くの情報を与えすぎると、返って相手には伝わりづらくなってしまふ。そのため、ロボットが対象物体の特徴を聞き出す時には、「何色ですか？」などと、できるだけ少ない情報を当てることにする。

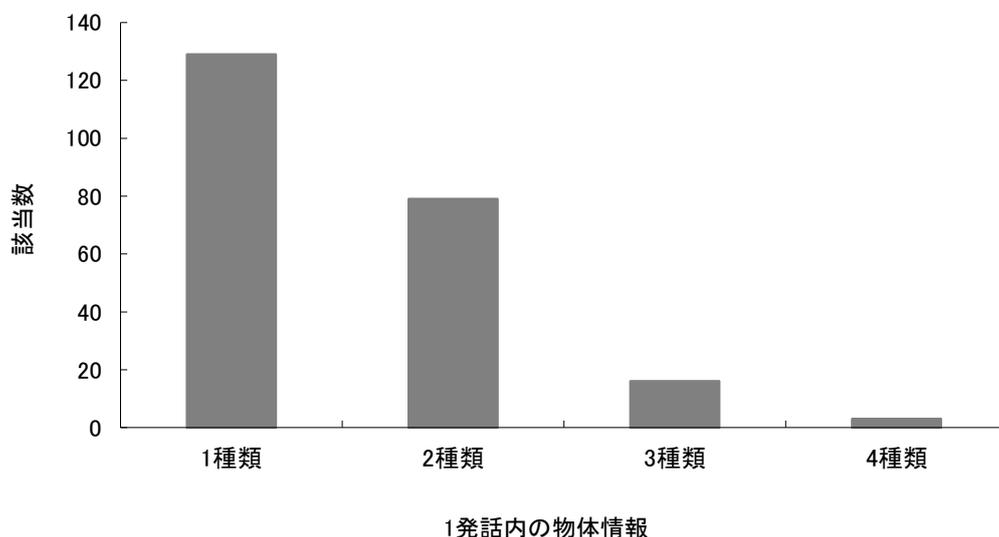


図 2.3 : 指示者の 1 回の発話に含まれていた物体情報の数

(文献[3]より引用)

図 2.4 は、対話実験で使用された「物体の視覚属性」の比率を表したものである。「色、形、模様、大きさ、付属物、素材」の 6 種類の表現を得た。しかし、全ての表現に取り組むのは非常に困難であるため、坂田[3] は、その中でも一番多く利用されていた色に関する表現に焦点を当て、「色情報を用いた対話による物体認識」を行った。

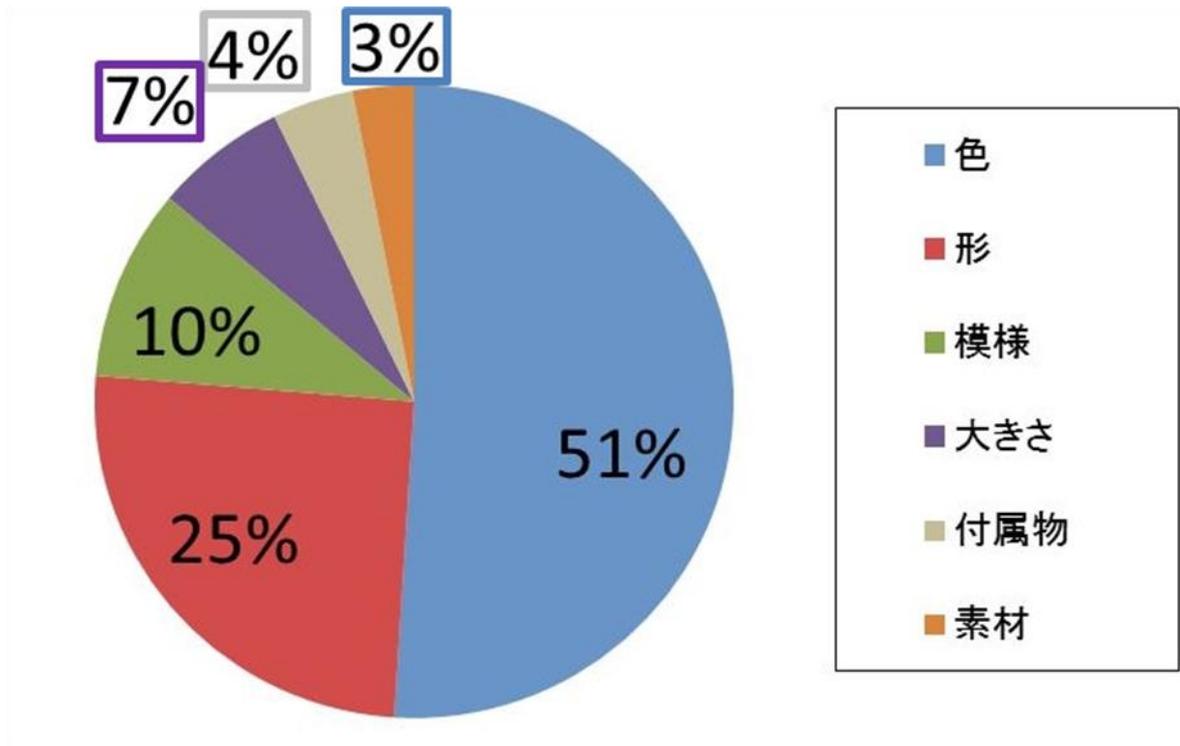


図 2.4 : 視覚属性の内訳 (文献[3]より引用)

2.2 調査 1 : 形状に関する人の言葉の特徴

視覚情報（図 2.4 参照）として、色情報の次に多く用いられている「形」に関する対話物体認識を行った。回答者には「形状を表現するように」と指示した。ただし、それ以上の細かい指示はしなかった。参加者を用いた実験で、どのような形状表現が使われるか調べた。人間同士の対話の観測実験を行うために、図 2.5 に示す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た。

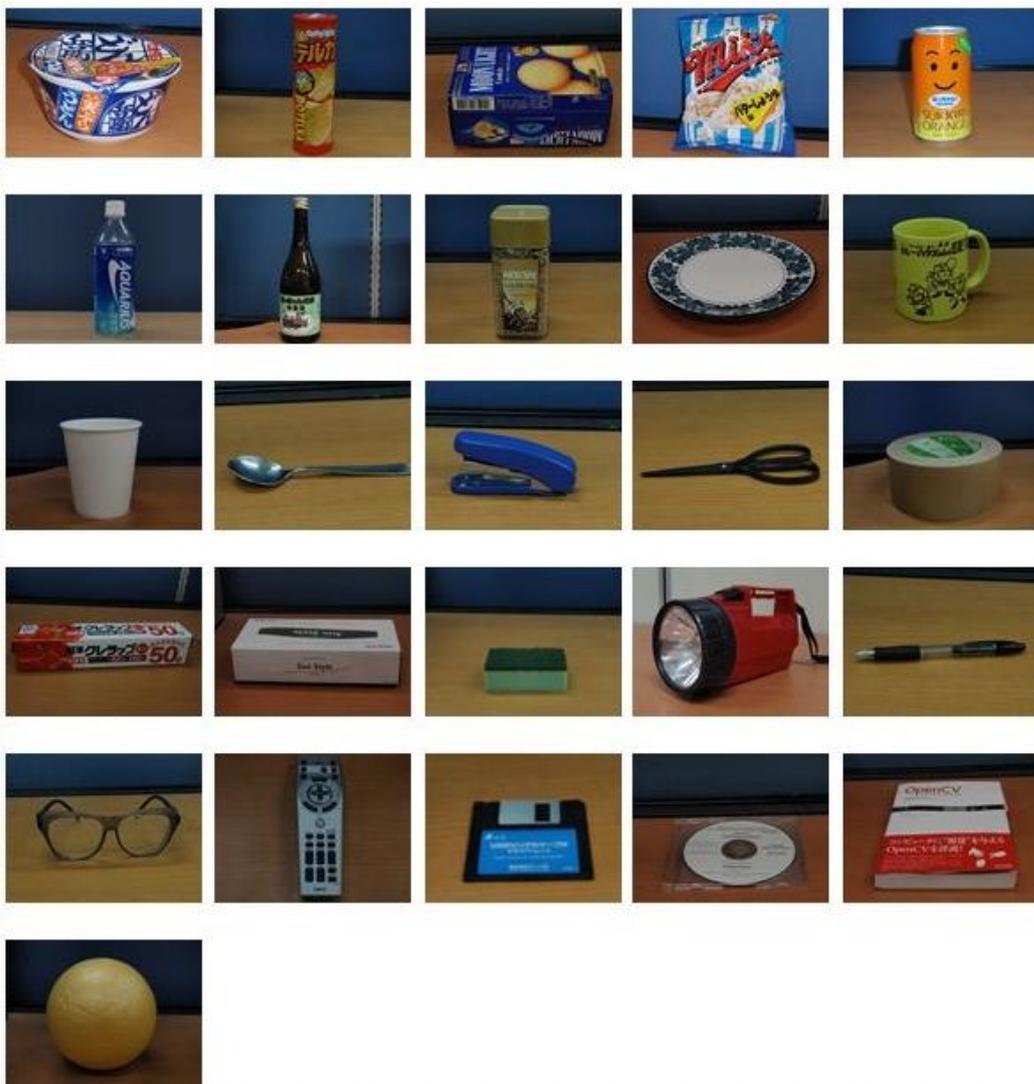


図 2.5 : 実験に用いた物体

2.2.1 調査 1 の概要

実験手順を図 2.6 に示す。テーブルに多数の生活用品をランダムに配置しておき、「指示者」と「回答者」を組とする。指示者は回答者に欲しい物の形状を伝える。回答者は依頼された物が何かを判断して、推定物を回答する。回答者には「形状を表現するように」と指示した。ただし、それ以上の細かい指示はしなかった。

本調査の手順を説明する。図 2.1 を参照に以下の手順で行う。

- ① 指示者はテーブルにある物体から 1 つを選ぶ
- ② 指示者は、選択した物体を形状情報で表現し、回答者に伝える
- ③ 回答者は指示者が表現したと思う物体を、テーブルにある物体の中から選んで答える
- ④ 回答者が対象物体を当てるまで①～③の手順を繰り返す



図 2.6 : 形状表現の収集実験

2.2.2 調査 1 の結果と考察

10 組の参加者（全員，大学生）に対して伝達実験を行い，143 の形状表現を収集した．収集した表現を分類した結果を図 2.7 に示す．図の上のグラフは「指示者の形状表現の割合」を示す．視点に依存しないと考えられる 3 次元形状全体を表現する場に加え，正面，上面，側面などの特定の方向から見た場合の 2 次元形状で表現する場合も多く見られた．今回の実験では「形状」表現と指示しただけなので，付着などした部分の形状や表面の模様の形状を表現した場合もあった．図 2.8 の下のグラフは全体形状の表現をさらに分類したものである．球，直方体，円柱，円，四角形といった幾何学的な表現だけでなく，V,U 字型，ドーナツ型など，その形状をもつ文字や物体名により形状を表現する場合も多かった．

形状を表現する語句については以下のことが分かった．

[1] 視点に依存しないと考えられる 3 次元形状全体を表現する場合

例) クッキー箱：直方体 ボール：球 コップ：円柱

[2] 正面，上面，側面などの特定の方向から見た場合の 2 次元形状で表現する場合

例) ボール：（正面，上，横）円

コップ：（正面）矩形，（上）楕円，（横）矩形

[3] 付着などした部分の形状や表面の模様の形状を表現した場合

例) リモコンのボタン：複数の矩形

[4] 物体全体の形状を，幾何学的な表現だけでなく，V, U 字型，ドーナツ型など，その形状をもつ文字や物体名により形状を表現する場合

例) ガムテープ：ドーナツ， ホッチキス：V 字， 眼鏡：U 字

[5] 同一表現で異なる形状を指示する場合がある．

例) 丸い→球，2 次元上での円，円柱

同じ形状表現で，該当する複数の形状がある場合に，該当形状に優先順位を付ける必要がある．再び，本大学の学生に協力してもらい，どの物体を優先するのかを聞き出し，各々の優先物体の消結果を収集・整理し，全体的な物体の優先度を計算して導き出すことを考慮する．

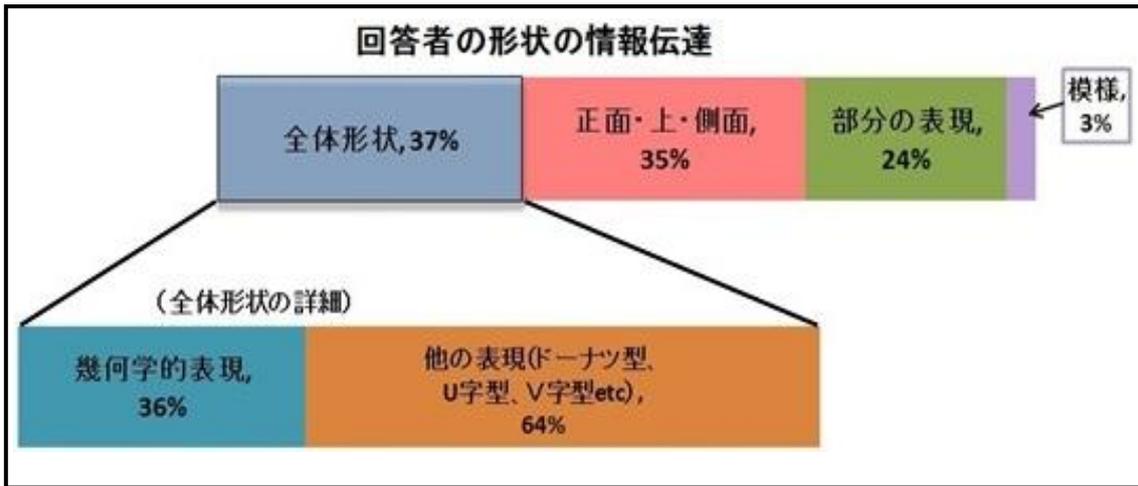


図 2.7 : 指示者の用いた形状表現の割合

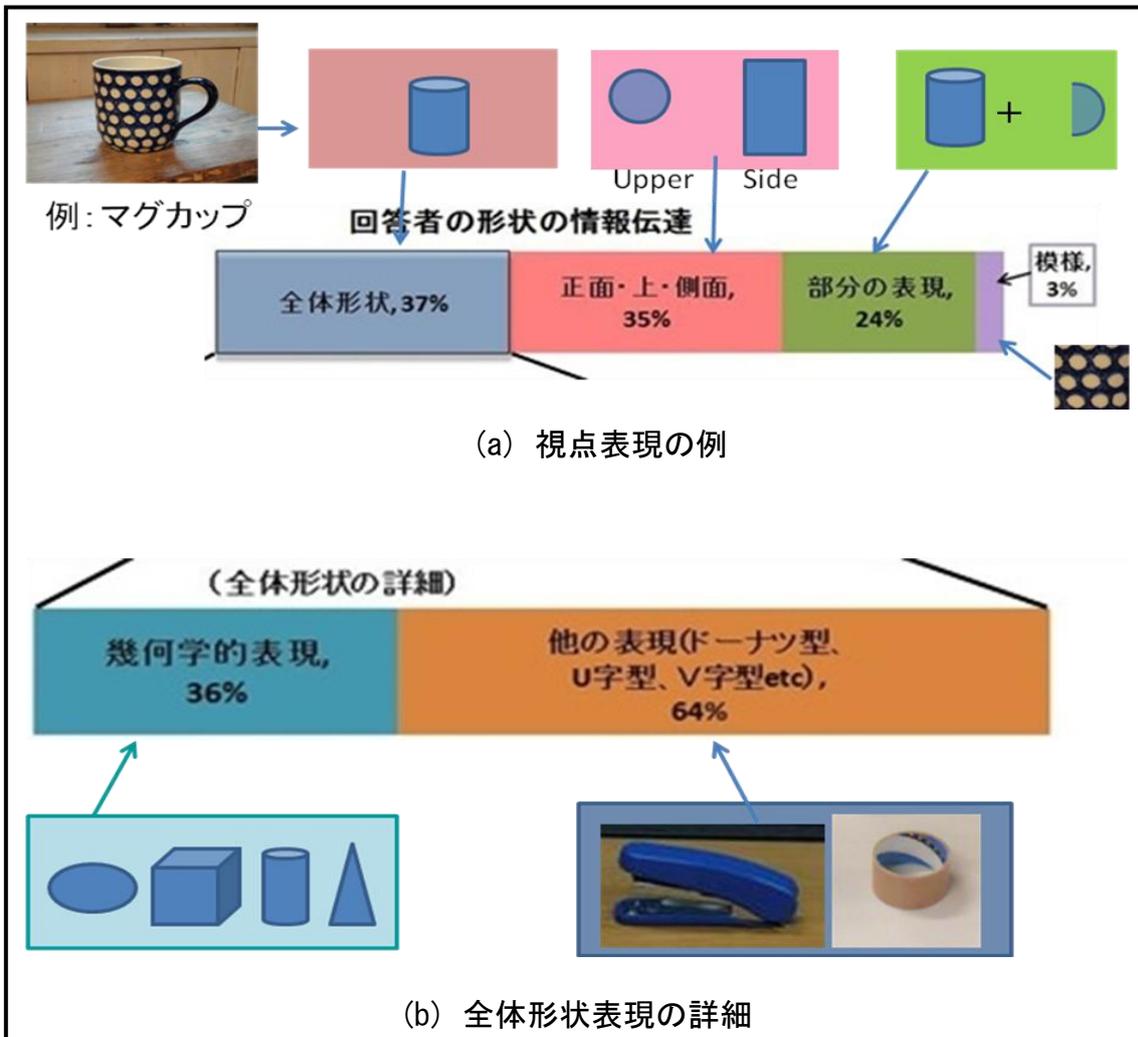


図 2.8 : 指示者の用いた形状表現の例

2.3 調査 2 : 同一表現による異なる形状の表現

2.1 で述べた形状表現の収集実験を通じて、同じ表現でも異なる幾何学的形状を指す場合があることに気付いた。よく用いられた表現として「丸い」という言葉があったが、これは 2 次元の円、3 次元の球、さらにはときには全体として丸みをおびた物体を表現するときにも用いられていた。そこで、同じ言葉で表現される可能性のある物体が複数あった場合、その言葉ではどの物体が指示されるか調べることにした。

2.3.1 調査 2 の概要

例として「丸いもの」という表現について、Thurstone の比較判断の法則に基づく一対比較法を用いて実験的に調べた[35, 36]。この手法は、各刺激を他の各々の刺激と比較し、上位にあると判断された回数とその比率を基にして各刺激間の距離を求め、特定尺度上に位置づけるものである。

図 2.9 のような CG で作成した 10 種類の物体について、1 つずつに刺激番号(S)を付け、2 つずつを組にした全組み合わせ 45 組を印刷したものを作り、各組について 2 個のうち「より丸いもの」を選ばせた。57 人の参加者（全員、大学生）について実験を行った。図 2.9 の各々の形状に該当する実在物を図 2.10 に示す。

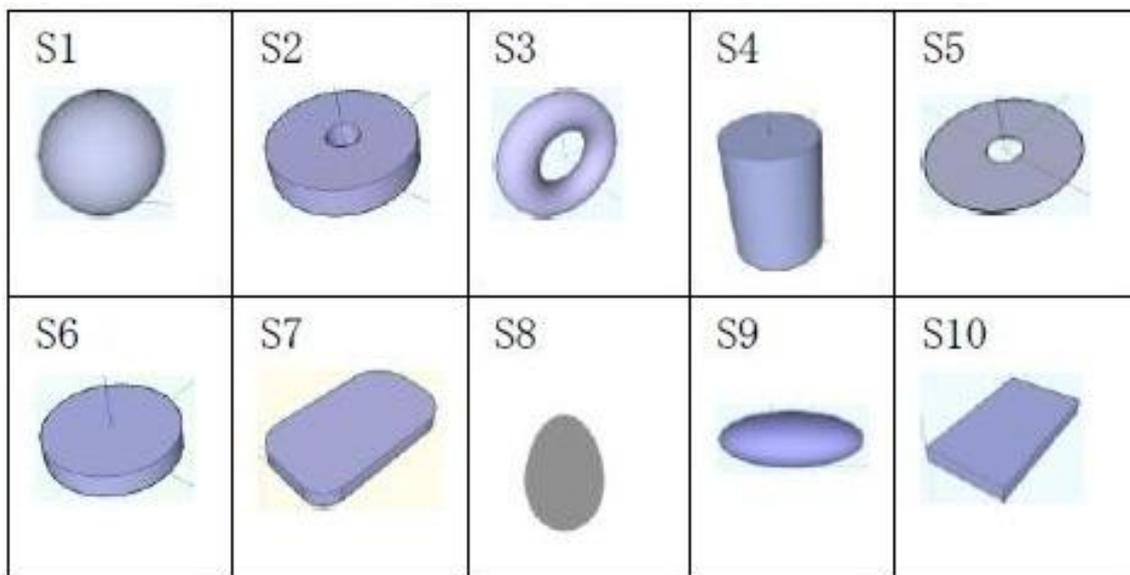


図 2.9 : 丸みの度合いを計算するために使用した物体

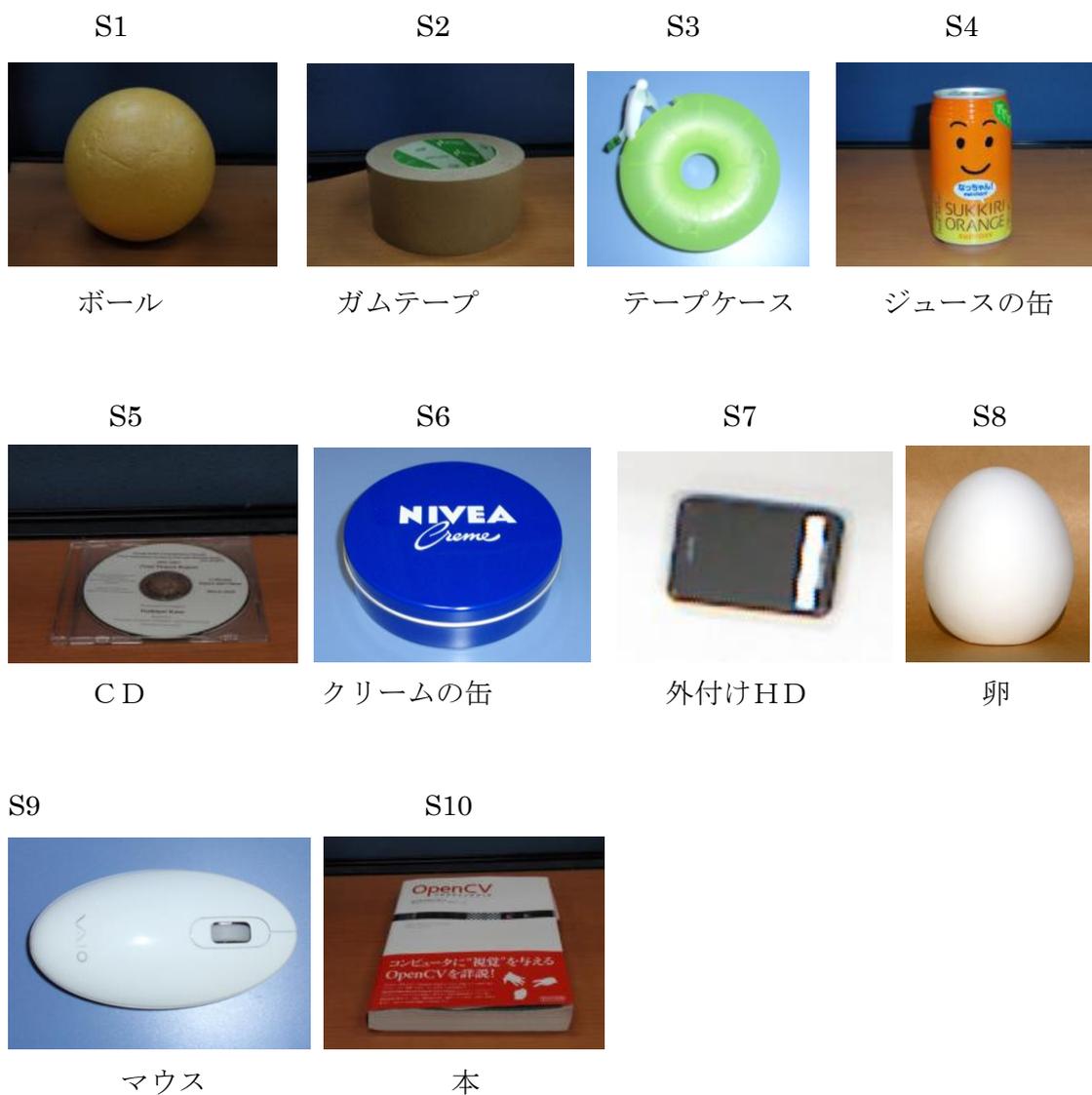


図 2.10 : 図 2.9 に対応する実物体

2.3.2 調査 2 の結果と考察

この結果から「丸い」ものの程度について尺度構成を行った結果を以下の図 2.11 に示す。「丸い」という表現は直接的には 3 次元の球と 2 次元の円の両者を指す場合があると考えられるが、この実験結果によると球の方が円より「丸い」ということになる。全般的に球、ドーナツ型、卵型のように、多くの方向から見た形（投影形状）が円形で、円形に見える部分の投影形状面積も大きいものがより「丸い」と判定されている。

る。

もし、複数の丸いものと呼ばれる可能性のあるものがあつた場合、人間が「丸いもの」と言ったならば、この実験結果で尺度値の大きい順に対象物候補とするような利用法が考えられる。ただし、今回の実験は人為的な設定であり、実際に複数の丸いと思われる物体があつたとき、人間は他の表現を用いるかもしれない。しかし、一般的な知識としては、ロボットの物体認識に有用なものであると考えられる。実際に、4隅が丸まった板が、角張った板と同時に存在する場合には、「角が丸まっている」というように、修飾語は付くが、「丸」という表現が用いられることは考えられる。さらに省略して「丸い方」というような表現が用いられる場合もあるかもしれない。このように同じ表現で異なる形状を指す場合については、さらに検討が必要だが、その手始めとして興味深い知見が得られたと考えている。

(1) 表 2. 1 は、列刺激が行刺激よりも「より丸い」と判断された度数を集計して度数行列である。横欄の i 番目の形状と縦欄の j 番目の形状とを比較し、どちらが丸いかを投票する。 i 番目の形状が丸いと 1 投票すると、 S_{ij} の欄に +1 を加算する。反対に j 番目の形状が丸いと 1 票投票すると、 S_{ji} の欄に +1 を加算する。

表 2. 1 : 度数行列 (参加者数 N=57)

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1		0	1	0	3	0	0	51	1	0
S2	57		53	10	31	35	2	40	37	0
S3	56	4		0	9	6	1	23	19	0
S4	57	47	57		42	43	5	47	42	0
S5	54	25	48	15		27	3	38	36	0
S6	57	22	51	13	29		2	40	38	0
S7	57	55	56	52	54	55		55	55	0
S8	51	16	33	10	19	17	2		19	0
S9	55	19	37	15	21	19	2	38		0
S10	57	56	56	55	57	56	56	57	57	
合計	501	244	392	170	265	258	73	389	304	0

(2) 表 2.1 から列ごとに度数の合計値を算出し、これの少ない刺激から順番に番号をつける。各セルの度数を $N(=10)$ で割って比率 p_{ij} を求め、刺激行列を番号順に並べ替えて表 2.2 のような比率行列 (p 行列) を作る。

表 2.2 : 比率行列 (p 行列)

順位		1	2	3	4	5	6	7	8	9	10
		S10	S7	S4	S2	S6	S5	S9	S8	S3	S1
1	S10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	S7	0.00	0.50	0.91	0.96	0.96	0.95	0.96	0.96	0.98	1.00
3	S4	0.00	0.09	0.50	0.82	0.77	0.74	0.74	0.82	1.00	1.00
4	S2	0.00	0.04	0.18	0.50	0.61	0.55	0.66	0.71	0.93	1.00
5	S6	0.00	0.04	0.23	0.39	0.50	0.52	0.67	0.70	0.89	0.89
6	S5	0.00	0.05	0.26	0.45	0.48	0.50	0.63	0.67	0.84	0.95
7	S9	0.00	0.04	0.26	0.34	0.33	0.37	0.50	0.67	0.66	0.98
8	S8	0.00	0.04	0.18	0.29	0.30	0.33	0.33	0.50	0.59	0.50
9	S3	0.00	0.02	0.00	0.07	0.11	0.16	0.34	0.41	0.50	0.98
10	S1	0.00	0.00	0.00	0.00	0.11	0.05	0.02	0.50	0.02	0.50
	合計	0.50	1.80	3.52	4.82	5.17	5.17	5.85	6.95	7.42	8.81

(3) p_{ij} から尺度距離 z_{ij} を求める。 z_{ij} は、単位正規分布で $-\infty$ の方から図った面積 p より求められる横座標である。 z_{ij} は、単位正規分布表を利用して求め、 $p > .5$ なら符号は正、 $p < .5$ なら符号は負とする。これをもとに、表 2.3 のような尺度距離行列を作成した。

表 2.3 : 尺度距離行列 (z 行列)

順位		1	2	3	4	5	6	7	8	9	10
		S10	S7	S4	S2	S6	S5	S9	S8	S3	S1
1	S10	0.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
2	S7	-0.50	0.00	0.41	0.46	0.46	0.45	0.46	0.46	0.48	0.50
3	S4	-0.50	-0.41	0.00	0.32	0.27	0.24	0.24	0.32	0.50	0.50
4	S2	-0.50	-0.46	-0.32	0.00	0.11	0.05	0.16	0.21	0.43	0.50
5	S6	-0.50	-0.46	-0.27	-0.11	0.00	0.02	0.17	0.20	0.39	0.39
6	S5	-0.50	-0.45	-0.24	-0.05	-0.02	0.00	0.13	0.17	0.34	0.45
7	S9	-0.50	-0.46	-0.24	-0.16	-0.17	-0.13	0.00	0.17	0.16	0.48
8	S8	-0.50	-0.46	-0.32	-0.21	-0.20	-0.17	-0.17	0.00	0.09	0.00
9	S3	-0.50	-0.48	-0.50	-0.43	-0.39	-0.34	-0.16	-0.09	0.00	0.48
10	S1	-0.50	-0.50	-0.50	-0.50	-0.39	-0.45	-0.48	0.00	-0.48	0.00
	合計	-4.50	-3.20	-1.48	-0.18	0.17	0.17	0.85	1.95	2.42	3.81
	平均値	-0.45	-0.32	-0.15	-0.02	0.02	0.02	0.09	0.19	0.24	0.38
	尺度値	0.00	0.13	0.30	0.43	0.47	0.47	0.54	0.64	0.69	0.83

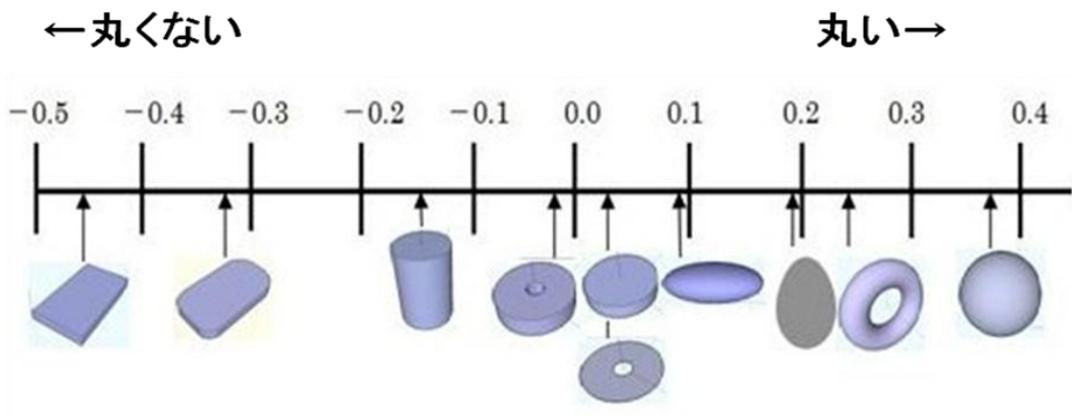


図 2.11 : 丸みの度合い

2.4 対話物体認識用オントロジーの全体の構成

本論文では形状表現について検討するが、対話物体認識用オントロジーの全体の構成としては図 2.12 に示すものを検討している。オントロジーの構築には法造[42]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。これ以降の図ではそれぞれ p/o, a/o と示されている。この図に示すように、最上位の概念として主に 4 つの概念を考える。4 つの概念としては、「実在物」、「属性」、「自然言語表現による属性表現」、「関係」である。ただし、今回は「関係」概念だけはシステム上で使用していない。

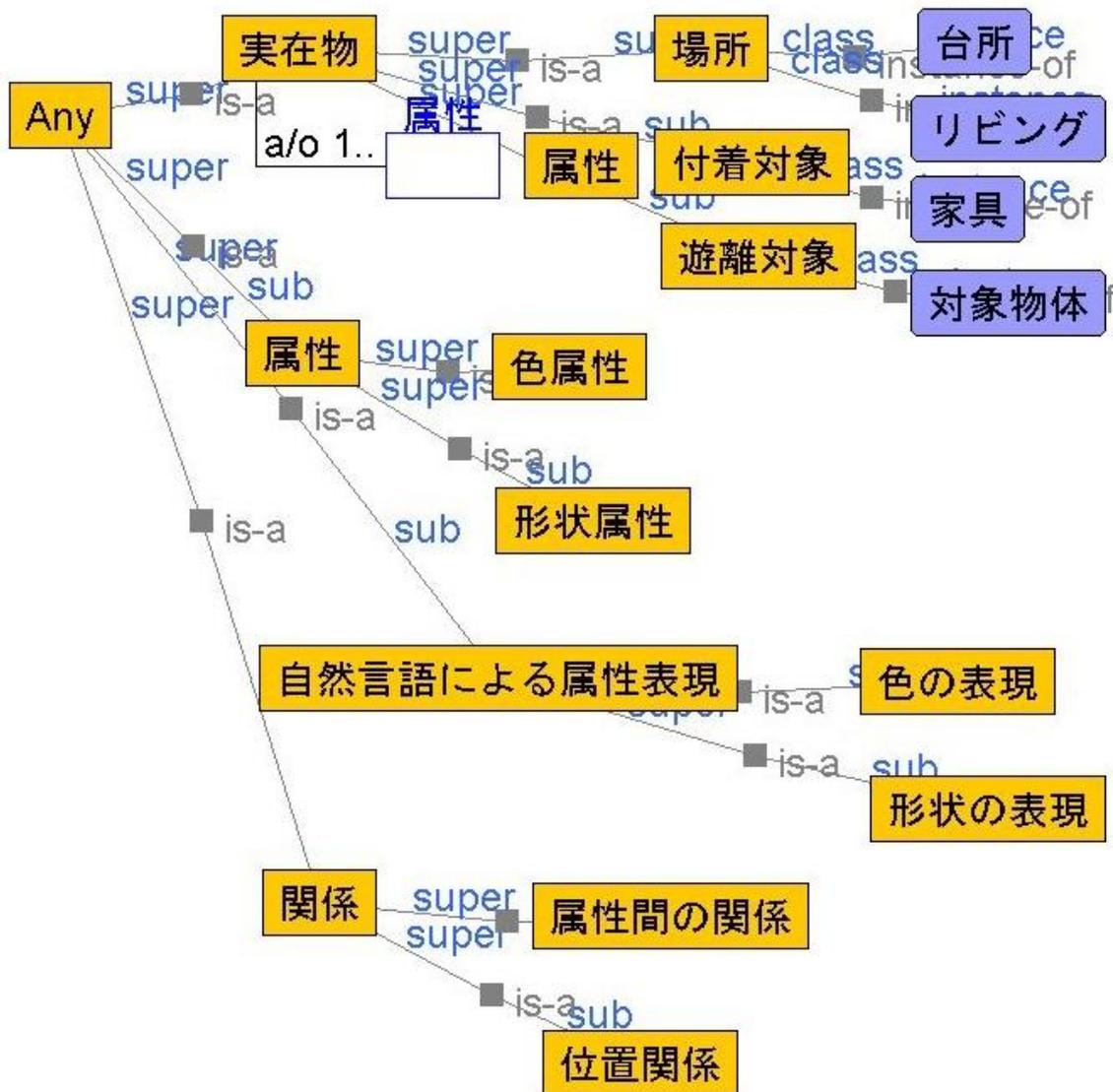


図 2.12 : 対話物体認識用オントロジーの全体構成

対話物体認識では、まず対象世界にある物体を物体名のレベルで体系化する。これが「実在物」である。「実在物」の概念では、実世界にある物体は、場所、付着対象、遊離対象に分類される。これらの用語はアフォーダンスの提唱者のギブソンが視覚で知覚されるものとして提案したものから採った[37]。本研究では、遊離対象をロボットが取ってくる対象物体とする。検討している対話物体認識システムは、まず、指示された対象物の自動認識を試みる。そして、それが失敗したときに対話物体認識に入る。実在物のクラスには、その属性が記述されるとともに、自動物体認識がそのクラスやインスタンスについて学習されていれば、その認識が起動されるようになってい

る.

「属性」の概念は色, 形, 材質など物理的な属性を体系化したものである. 形状属性について言えば, 形状を 1 次元, 2 次元, 3 次元の形状に分け, そしてそれぞれをさらに種々の形状に細分化していく. そして, それぞれの形状について, 例えば 4 角形なら頂点 (角) を 4 つもつなどの属性が記述されている. また, その形状の検出用の画像処理プログラムにもリンクされている.

対話物体認識用オントロジーの検討の中心になるのは, 「自然言語による属性表現」である. 「属性」は物理的に存在するもので, 一つに体系化できるが, 人の表現は多様である. この部分は「自然言語による属性表現」を整理し, その中のクラスと物理的な「属性」のクラスの可能な対応関係を示す.

「実在物」と「属性」に関しては, 具体的な内容は異なるが, 実世界に存在する物体を体系化し, その物体を認識するために, その物体の属性を検出する知識をまとめるといふ観点からは従来研究[32~34]で検討されていたものである. これらは対象世界と利用できる画像処理に関する知識として, 対話物体認識でも必要である. ここでは, それに加えて, 「自然言語表現による属性表現」を考えている.

今回の対話物体認識では, 一つの物体が存在するだけで決まる属性, 特に形状を考えている. しかし, 属性には, 複数の物体が存在することにより決まるものがある. それを表すのが「関係」である. これには属性の程度の比較 (例: A は B より丸い) や位置関係 (例: A は B の右にある) がある. 実は, この「関係」属性についても複雑な自然言語による表現があり, この部分についても詳細な検討が必要である. 特に, 位置については物体を指示するのによく使われが, Cao らの検討も別にあるので[38], ここでは, 単独物体の属性についての検討にしぼっている. 「関係」の部分については, 対話物体認識のためのオントロジーの要素として重要なので示しているが, 内容の検討は今後の課題である.

2.5 形状表現オントロジーの構成

図 2.13 に、2.3.2 での調査 2 の結果をもとに、人の属性表現の中の形状表現の部分のオントロジーを構築した結果を示す。形状の表現は、その表現がどこを示すものかという「対象表現」と、その属性を表現する「属性表現」からなる。図 2.13 でのそれぞれの拡大版のオントロジーを図 2.14 と図 2.15 に示す。

対象表現(図 2.14)は、2.1 で分かったように、全体の形状を表現する場合、ある方向から見た形を表す投影形状の場合、物体の部分を表す場合がある。物体の部分については、付着物を表す場合、表面の模様を表す場合、さらに「角が丸い」というときの「角」のように構成要素を表す場合がある。この図の末端が実際に用いられる言葉を示す。ただし、実際の人の発話の中では対象表現は明示されていない場合がある。「…から見る」というような投影の際の視点方向など、対象表現で対象が明示されていれば、システムはそれで示される対象の検出を試みる。このような明示的な対象表現がなければ、システムは属性表現がどの対象を示す場合もあり得るとして処理を進める。

属性表現(図 2.15)は、実際に形状の属性を表すのに用いられる言葉である。これは名詞的なものと形容詞的なものに分けられる。「的」としたのは、例えば前者が形容詞として修飾語として用いられることもあるからである。後者は、関係の概念とも連携させて検討する必要があるが、今回は前者についてだけ検討した。名詞的な表現は、幾何形状を表すものと、その形状をもつ典型的な物体名によるものがある。以上のような分類は Galton による自然言語による形の表現の分類[39]と同様である。さらに、幾何形状については、幾何学で使われる用語と、より一般的な表現がある。特に注目されるのは、「丸い」のような一般的な表現である。属性表現のクラスには物理的属性として表現される可能性のある「属性」概念の中の属性が「物理形状」として示される。複数ある場合は、それらが列挙される。2.2 で調べたように、その程度に順序があれば、その順序の順に対応する属性が記述される。これにより、一つの表現が多様な属性を表す可能性があるという問題に対応する。また、図 2.13 , 2.15 では煩雑になるので示していないが、例えばボールの物理形状属性には球が示されるようになって

ている。このようにして、人間が使った多用な表現が指す物理形状がオントロジーから分かるようになっている。今回は、人間が使った表現を理解するのが中心なので、以上により多様性の問題が解決できる。ロボットの方から物体を表現する必要がある場合には、多様な表現が可能な対象については、どれを用いるかという問題が起こる。自然な表現ということ考えると、表現の選択は必要であるが、基本的にはロボットの方からは、その物理属性にあてはまる最も詳細な表現を使うということによいと思われる。これについては、今後、検討したい。なお、ここで示された物理形状がどういふものかについては、「人の属性表現」ではなく、「属性」の部分に記述される。例えば、「四角形」には頂点（角）が4つあることなどは「属性」の部分に示される。

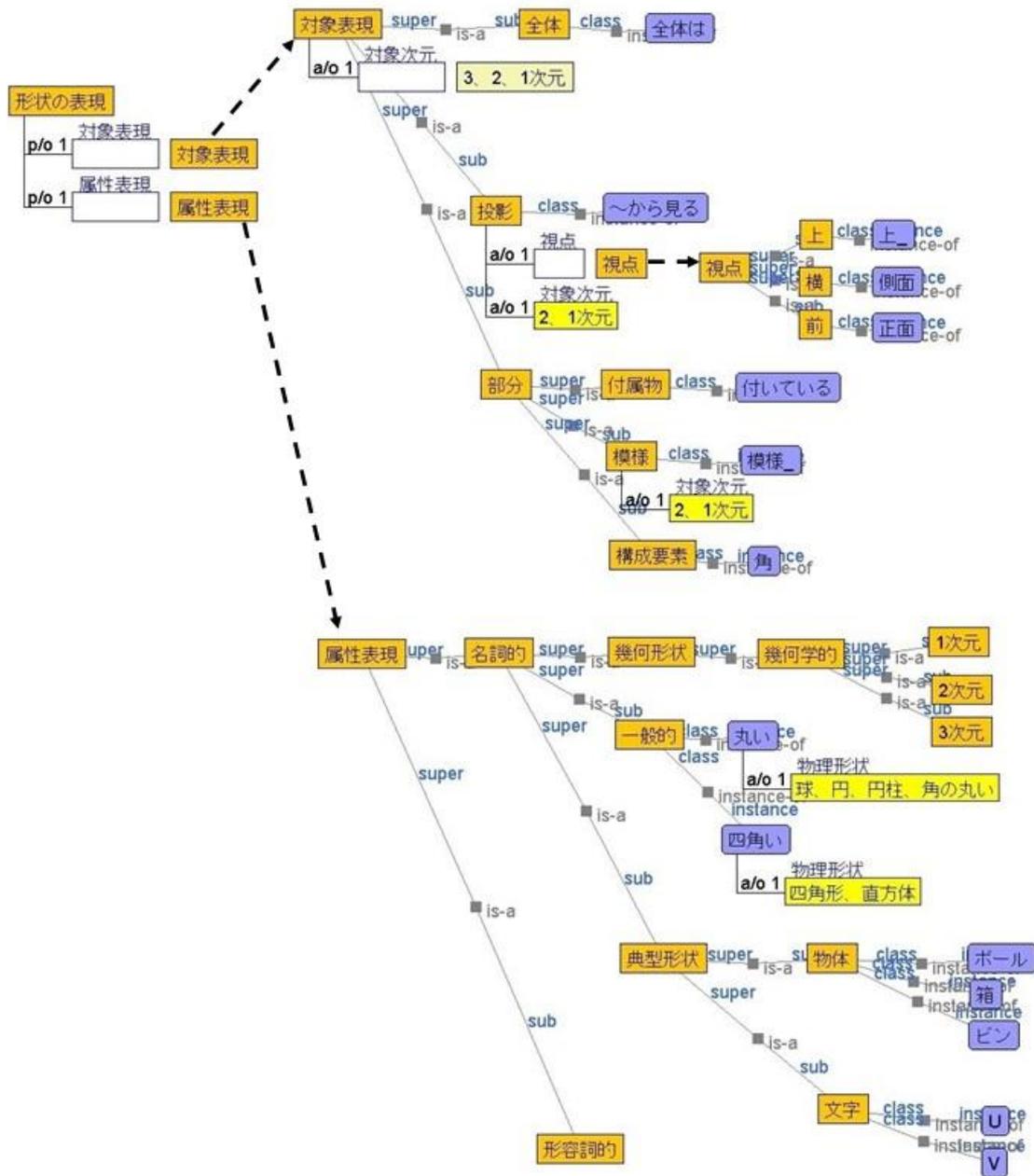


図 2.13 : 形状の表現部分のオントロジー

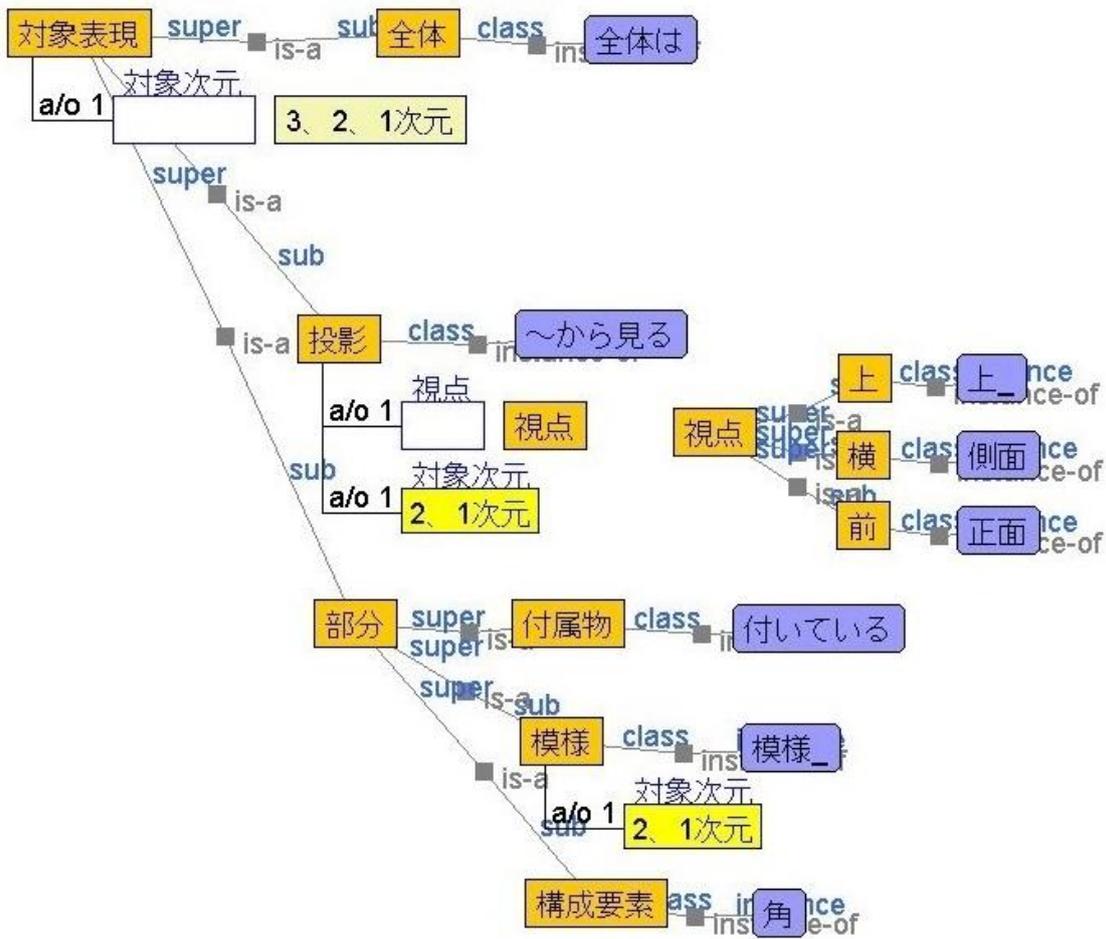


図 2.14 : 対象表現のオントロジー

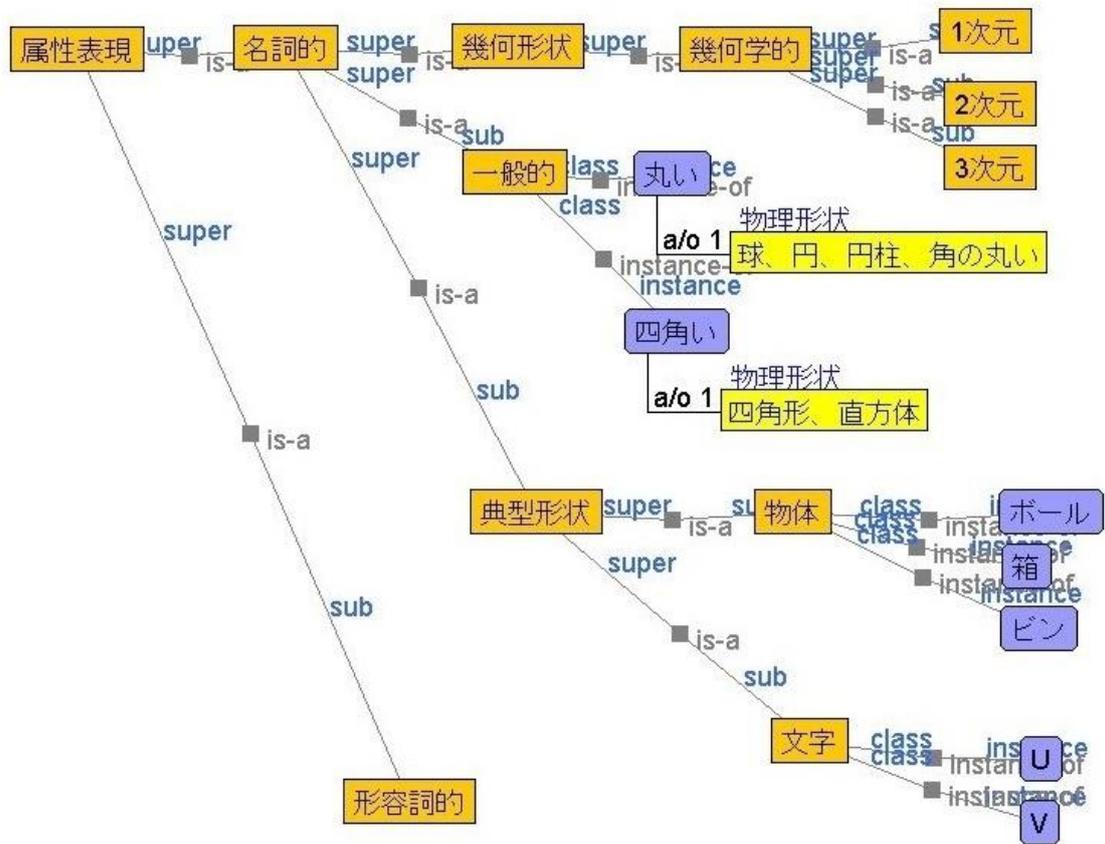


図 2.15 : 属性表現のオントロジー

3章 形状認識の手法

この章では、本研究で提案・作成した物体認識手法[93]について述べる。本研究の形状検出プログラムは、福田[28]らのプログラムを拡張して作成した。

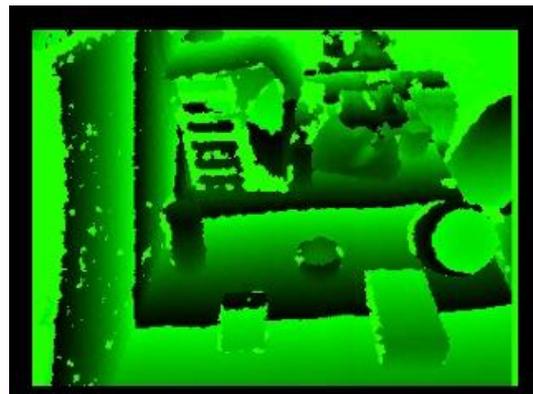
ロボット (PC で構成) の視覚機能としてマイクロソフト社の kinect[44] (深度センサー付き赤外線カメラ, RGB カメラ) を用いる。以下の図 3.1 では、kinect で深度画像を取得し、RGB 画像上の物体の 3 次元座標を取得する。最後に深度画像をマスクングして、物体と背景をセグメントする。

対象物の領域を構成する画素の 3 次元座標から法線ベクトルを計算する。法線ベクトルの数と向きにより、対象物体の形状、物体を構成する面の数と向きを確認する。

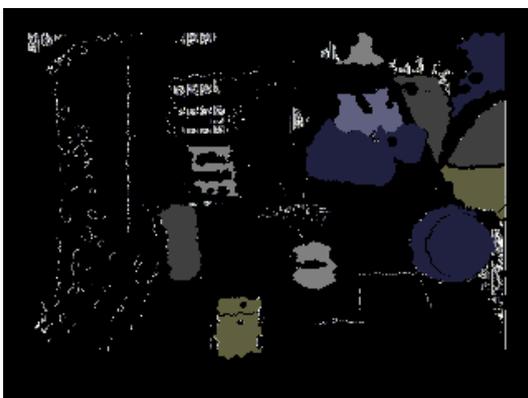
ソフトは、「visual studio C++2010」と「point cloud library(PCL)」を用いる。



(a) RGB 画像



(b) 深度画像



(c) マスク画像 (セグメンテーション)

図 3.1 : 物体と背景の分別

3.1 節では、福田ら[28]の提案した「3次元形状の判定方法」について述べる。この形状判定方法では、凹みのない3次元形状しか判定できないため、3.2 節以降の形状認識プログラムは3.1 節の形状判定方法に、法線ベクトル検出手法と分析手法を加えることで作成した。

3.2 節以降では、物体の凹凸形状の判定を含めた3次元形状認識手法について述べる。幾何学的な模様等をカメラ画像（平面画像）で見た場合には、模様表現なのか立体投影表現であるのかを区別するのは困難となる。このため、対象物体に対しての法線ベクトルを抽出し、法線ベクトル群の全体的な傾向、及び局所的な傾向を分析して、物体形状を判定する方法を実現した。

3.1 3次元座標情報と視点方向による形状判定

この節では、福田ら[28]の作成した「3次元形状判定」の手法について述べる。

形状判定方法の1つで、対象物の「全体形状」と「特定の方向からの形」を考慮する。第2章の対話実験で、「指示者の用いた形状表現の割合（2章の図2.7）」を求め、70%が「全体形状」と「特定の方向からの形」だったためである。

システムは、「球体」および「箱」といった物体の全体形状を認識するために、物体の3D ボリューム・データを分析する。それは、最初に、互いにセグメントされたボリューム・データ上で、3次元座標系を固定する。物体が横たわっている平面を検出し、Y軸の方向として平面の法線方向を検知する。その後、メソッドを3D座標データに適用した平面モデルを用いて、Y軸に対して最も平行な平面を見つける。この平面のX軸の方向に対する法線方向を検出でき、座標の重心として起源をとることができる。下記の図3.2にいくつかの例を示す。

システムは「3D物体」、「2D物体」、および「1D物体」に分類する。それらは、X、YおよびZ軸に沿った座標データの分配に依存する。それらのポイントデータが3つの軸に沿って分配される場合、体操物は3D物体として分類される。

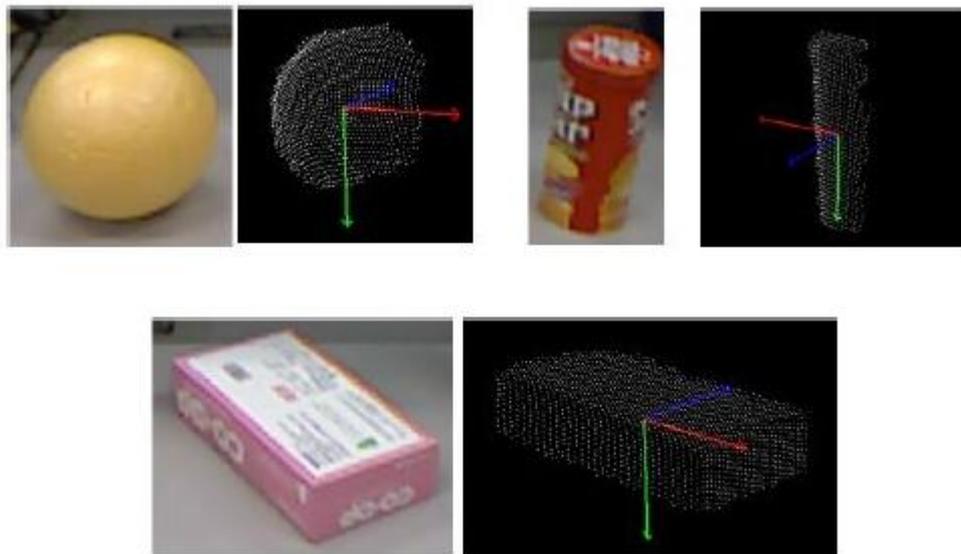


図 3.2 : 物体の 3 次元データの例 (文献[28]より引用)

2D と 1D の物体は同様の方法で決定される. 2D 物体は「平面物体」として処理され, 1D 物体, 「直線」と見なされる.

我々は, 球体, 円柱および箱のような基礎的な形状を検知するために RANSAC アルゴリズムに基づいたモデル取り付け方法の使用により, いくつかのプロセスを用意した. さらに, 我々は, 図 3.3 のように, 一方の XY 平面, YZ 平面あるいは ZX 平面に投影された座標データの検討に, 特別の方向から見られた形を認識するためにプロセスを準備した. R.T.Collins ら[54]と同様, 3次元形状を複数の平面上の形状, つまり 2次元形状の集合として定義する手法である.

人間は, 彼らがどの形記述方法を使用するか通常は言わない. 「丸い」は, 全体の物体形状が球体か, 円柱のような特別の方向から見られて, 物体が円形であることを意味するかもしれない. または, 全体の形として「円」と表現するかもしれない. しかしながら, 第 2 章で実施した実験結果は, 「丸い」の物体が球体(ボール)および円柱(缶)がある背景で球であるべきであることを示すかもしれない. 下記の図 3.4 でのように実験からの発見は概略的に示すことができる. 3 方向(X, Y および Z 軸からの)から見られた 2D 形状が全て, 円であることから, 球体は真球に近い物体を感じるかもしれない図 3.4 での左・中央の図を比較する場合, 上から見たとき円形部であるが, 残り 2 方向から見たとき非円形であることから, 中央のものはより円に近いと考えられる,

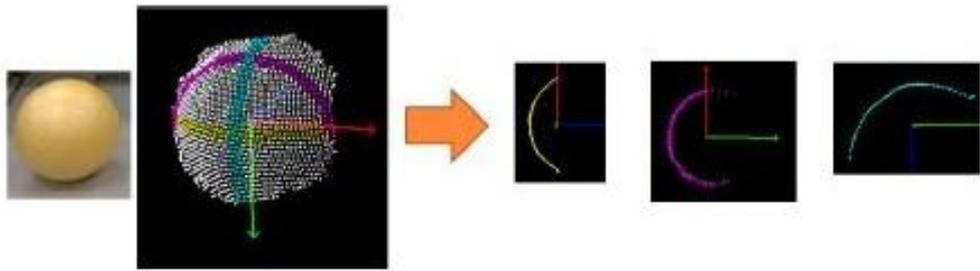


図 3.3 : 3 つの直交面に投影されたポイントデータ
(文献[28]より引用)

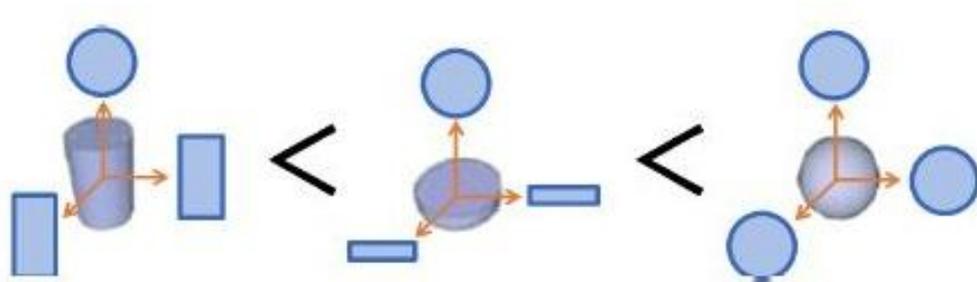


図 3.4 : 「丸いと思われる物体の順番」
& 「三方向から見たときの各々の形状」(文献[28]より引用)

丸さ度合いを計算する式を下記の図 3.5 (a,b,c)に示す.

(a) XY平面, YZ平面およびZX平面にポイントデータを投影し, 各投影データに対し,
丸さ度合い R_p を計算.

L_p : 領域最大の長さ S_p : 領域エリア

三方向データの合計により, 領域サイズの重みづけを計算する.

(b) w_p :領域サイズの重みづけを計算する

(c) 私たちは丸さの基準値を求める式

欠点として, 単一の観察からのみでは, 完全に3つの視点を得ることができないこの方法によって, 丸みの度合いを得ることができる.

$$R_p = \frac{\pi L_p^2}{4S_p} \quad (a)$$

$$w_p = \frac{S_p}{S_1 + S_2 + S_3} \quad (b)$$

$$\text{degree_of_roundness} = \sum_{p=1}^3 w_p R_p \quad (c)$$

図3.5：丸み度の計算式（文献[28]より引用）

下記の図3.6 は，図中の5つの物体を形状判定の結果である．この場面では，システムはプリミティブ検出によって球体，箱および2本の円柱を検知し，これらの物体の丸み度も計算している．

丸さは，ボール(球体)が最大で，箱は最低値である．ガムテープは，ポテトチップスの円筒構造のパッケージより丸み度があると考えられる．

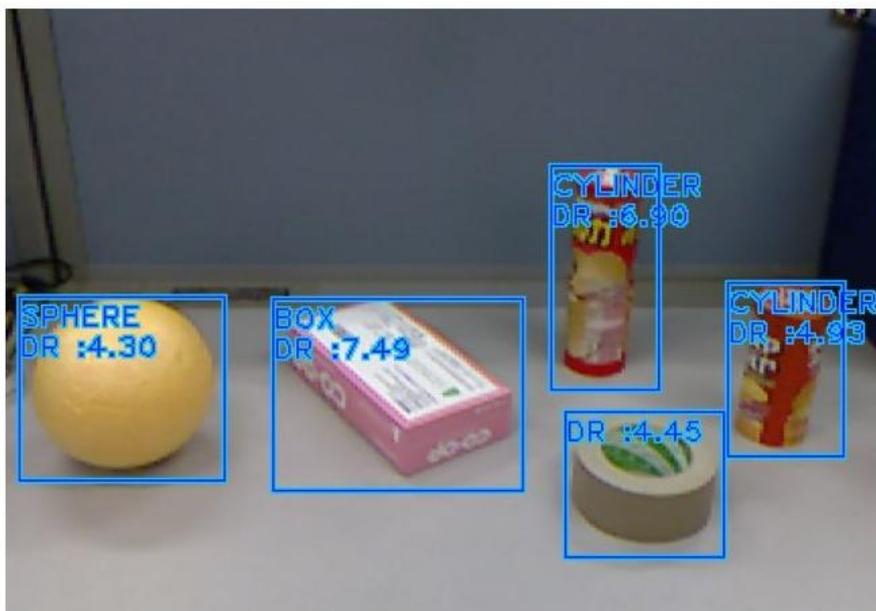


図3.6：形状判定（文献[28]より引用）

3.2 法線ベクトルのグループ分けによる形状判定

物体の面を分割し，面の向きにより形状を判別する．

N. Dalal ら[55]は，勾配方向ヒストグラム（HOG）を使用し，粗い形状を描写する研究を行っている．

本研究では，正 n 面体で n 個の平面を通る法線ベクトルにグループ化・ヒストグラムにし，各グループの法線数により，物体形状を検出する方法を検討した．分割面 i （ $1 \sim n$ ）と法線数の関係をヒストグラムで表現できる．ヒストグラムから形状を推定する（表 3.1）．

法線ベクトルを計算したが，画像上のノイズにより，平面図形においても，法線ベクトルの向きが一樣に揃わず，思わぬところで分散してしまうという欠点がある．

そこで，物体の法線ベクトルと 12 面体の 12 種類の方向を持つ法線ベクトル（図 3.7）との内積を計算し，結果が最も 0 に近い 12 面体法線ベクトルの方向に，対象物の法線ベクトルを投票する．これにより対象物の法線ベクトルを 12 種類の方向に集約できる．すなわち，物体の面の数，各面の方向性を検出でき，物体の形状概略を判別できる．さらに，カメラ映像では識別できない平面的模様と立体形状とを分別できる．

正多面体（面数を n ）には， $n = 4, 6, 8, 12, 20$ ，が考えられる．法線ベクトルの方向性をいくつかのグループに分けることを考え，6 グループ，12 グループ，20 グループ等をグループ数の候補と考えた．日用品を中心とした物体を検索対象としており，日用品には，直方体，円柱，球体，等の形状が主であるため，面方向のグループは 6 以上必要である．また，以下の理由により，グループ分けの演算処理負担が少ない， $n = 12$ を採用した．

6 面体：位置合わせ近似が困難

12 面体：位置合わせ近似ができ，面数が比較的少ない

20 面体：位置合わせ近似に優れるが，面数が多く演算大（処理時間の増加）

表 3.1 : 形状判別図

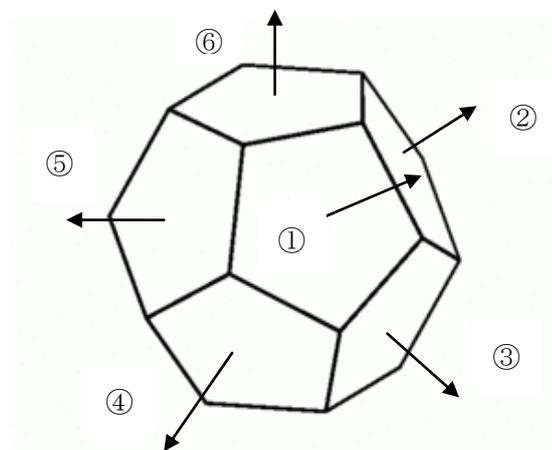
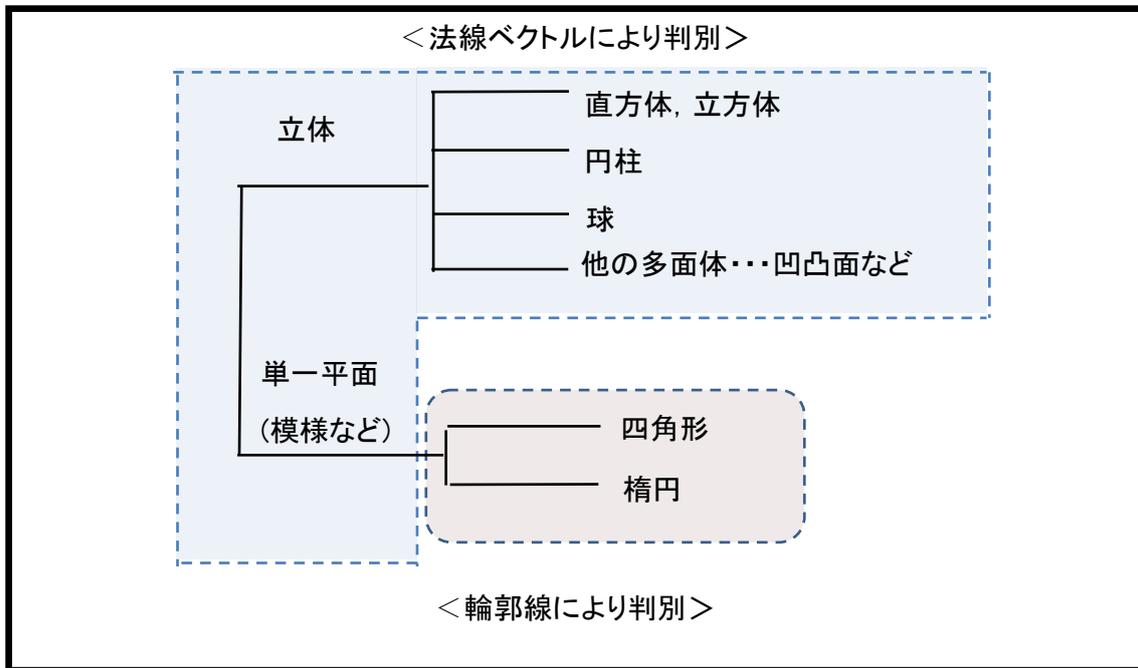


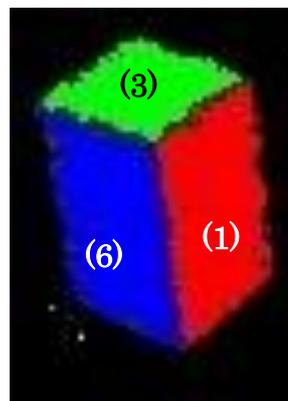
図 3.7 : 正 12 面体に対する法線ベクトル

下記の図 3.8 に示す対象物(形状 : 直方体)の法線ベクトルを, 上記の図 3.7 に示す 12 面体の法線ベクトルに投票して, 面の分別を行う. 図 3.8 は直方体を斜めから見て, 面を 3 面に分割した結果である. 次に, 分割面 i (1~12) と法線数の関係をヒストグラムで表現する.

カメラ角度 (向き) を対象物に合わせて, 対象物の形状や向きによって, 法線グルー

プのヒストグラムは変化するが、ヒストグラムの全体的な特徴や局所的な特徴を判定することで、形状も判定できる。

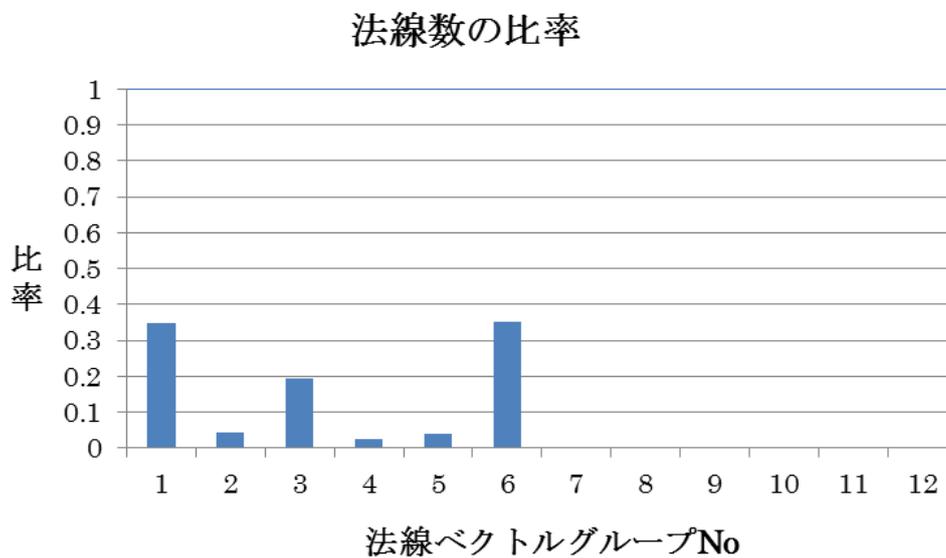
図 3.8 (c) のヒストグラムでは、高レベルの法線グループが 3 個あるので、直方体での頂点部分と見なしている。



() 内の番号は法線ベクトルグループ

(a) 対象物体

(b) 物体の面分割



(c) 方向単位のヒストグラム

図 3.8 : 物体の面検出と方向単位のヒストグラム

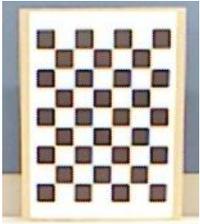
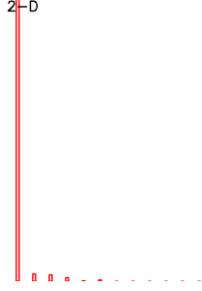
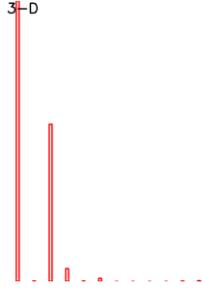
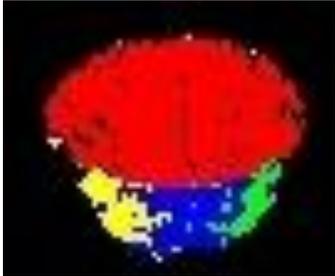
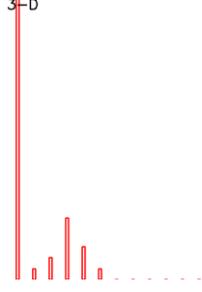
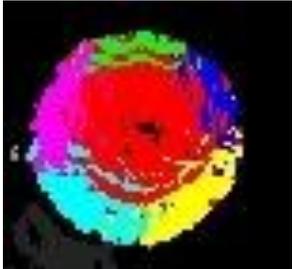
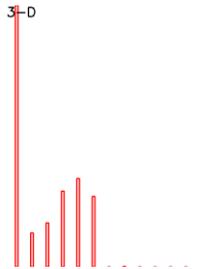
原画像	面検出	法線ベクトル
		
		
		
		

図 3.9 : 各物体の面分割と法線ヒストグラム

幾何学的な模様等をカメラ画像（平面画像）で見た場合には、模様表現なのか立体投影表現であるのかを区別するのは困難となる。このため、対象物体に対しての法線ベクトルを抽出し、法線ベクトル群の全体的な傾向、及び局所的な傾向を分析して、物体形状を判定する。

日用品の主要な形状である直方体、円柱、球に対して、法線ベクトルのヒストグラ

ムを上記の図 3.9 に示す。ヒストグラムのパターンマッチングにより形状を判定し、形状の区別ができないときは、カメラの位置をずらす。撮影位置により、面が単面、2面、3面での各法線ベクトルのヒストグラムには夫々の特徴がある。

直方体の場合、ヒストグラムでの高レベル値のグループ数が単面状態では1つ、2面状態では2つ、3面状態では3つと明確に出る。

円柱の場合、単面（曲面）の場合には高レベル値のグループは1個で中レベルのグループが分散し、2面状態の場合には高レベルのグループが分散している。

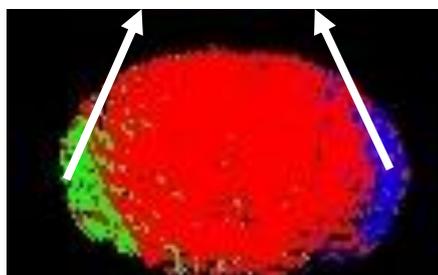
球（単面のみ）では、高レベル値のグループが1個と、高レベル値のグループのまわりに中レベル値のグループが集中している。

また、カメラの角度と位置を変えることで、対象物領域内においてヒストグラムの特徴変化がどのように生じるかで形状判定を高めることができる。

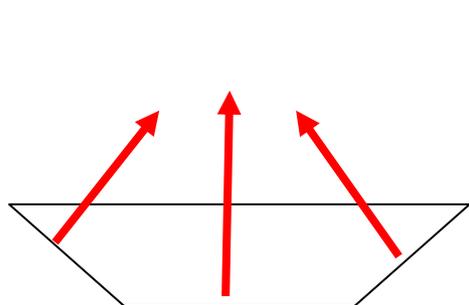
図 3.10 のように、分割した面の向き、各々の面の位置関係から、物体の形状の凹凸を知ることができる。図 3.10 はお皿を面分割したものであり、赤い面を中心に、青・緑の面の法線ベクトルが赤い面に向かって傾いていることから凹みのある形状と判別できる。



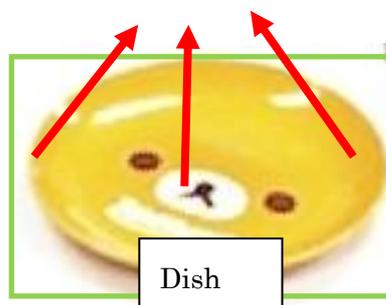
(a) 元画像



(b) 法線ベクトルによる面の分割



(c) 凹を示す法線ベクトルの向き



(d) 凹みのある物体

図 3.10 : 法線ベクトル方向による凹（へこみ）の認識

3.3 形状特徴の検出

第2章の「図 2.13：形状の表現部分のオントロジー」での形状特徴をオントロジーで整理したことから，形状の特徴プログラムもそれに合わせて整理することができた。

非幾何学的形状においては，まだ文字認識が成功していないので，今後の研究課題の1つとする。

検出対象とする形状の種類は以下の5つである。

[1] 全体的な幾何学形状検出

— 球，楕円，円柱，矩形，直方体

[2] 全体的な非幾何学形状検出

— 凹凸面

— 文字形状：V字

[3] 投影

— 上面・側面形状

[4] 部分形状

— 角形状判定：角張り，角が丸い（曲線）

— 穴の空いた物体

— 模様検出：格子

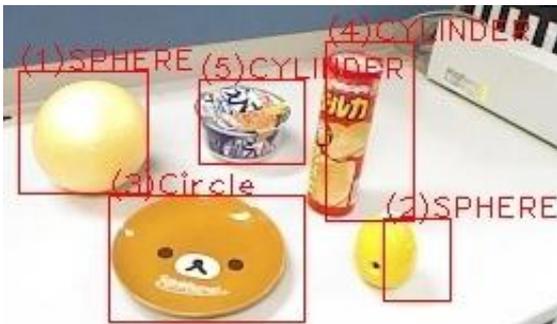
[5] 物体の大きさ

— 大小関係，縦・横の長さの比較

3.3.1 全体的な幾何学形状検出

全体的に幾何学的な形状を検出する場合，法線ベクトルの向きと数から3次元幾何学形状を判別する。以下の図 3.11, 3.12 では，左側の物体グループに対し，3.1 の手法を用いて形状を判別する。右側は，3.2 の手法を用いて，法線ベクトルの向きの違いを色分けし，物体の面を分割している。

(a) 形状判定



(b) 法線による面分割

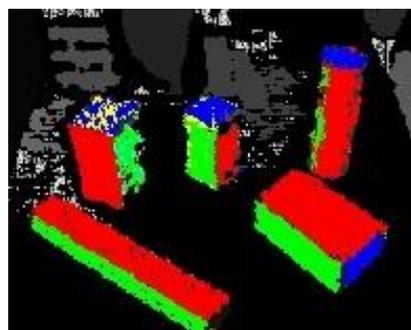
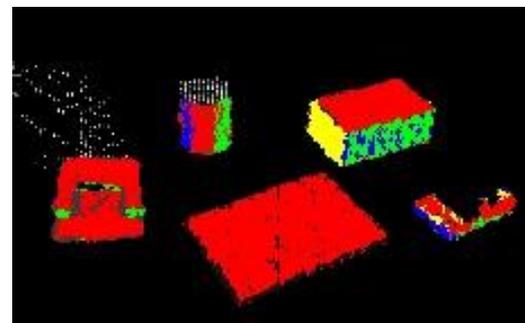
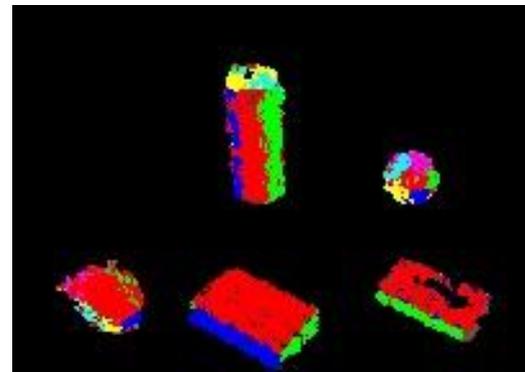
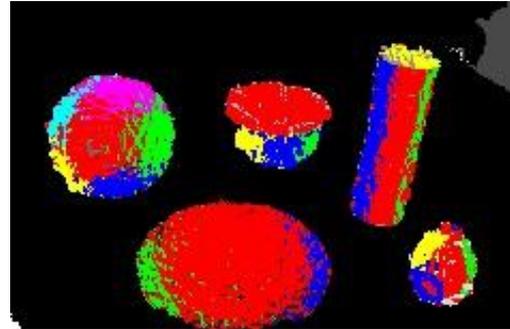


図3.11：幾何学形状認識の例A

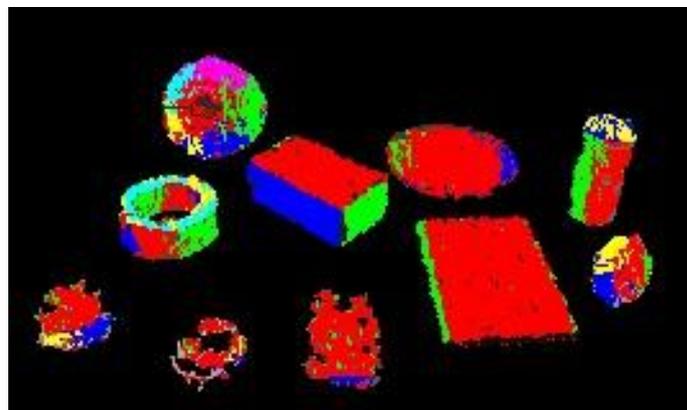
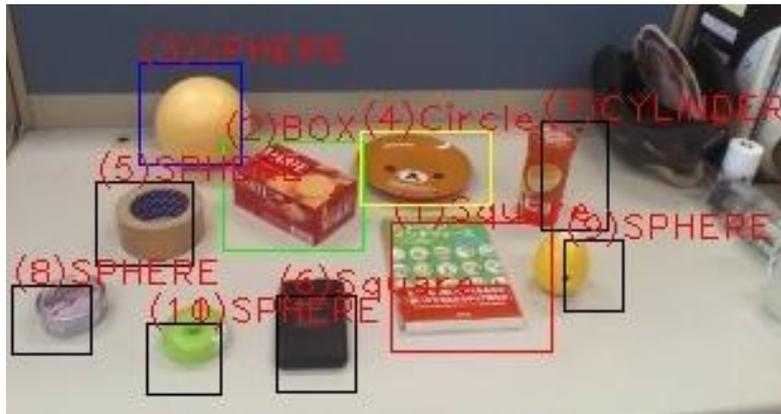


図3.12：幾何学形状認識の例B

考察

35 個の「幾何学的な形状の物体」に対し，3.1 の手法を用いて全体的な幾何学形状の判定実験を行ったところ，35 個の内，28 個の物体の形状を正しく判定できた．すなわち 35 個の物体に対する成功率は，80%であった．

失敗の原因は，1 方向からの 3 次元ポイントデータしか取得していないため，完全には，正確な形状を判定できないことがわかる．例として細長い直方体（サランラップ）を円柱と見なしてしまい，高さが低い円柱（ガムテープ）を球と見なしてしまうことがある．

また，物体をセグメントする際，1 方向での 3 次元ポイントデータを全て取得することはできず，形の重要な部分（角）をセグメントできない場合，ご認識をしてしまう．例として，矩形の角部分を検出できないと，円と誤認識してしまうことがある．

3.3.2 全体的な非幾何学形状検出

3.3.2.1 凹凸面

下記の図でお皿を検出したい場合、皿の特徴として、「丸い」という情報だけでは、他の候補として球も対象となってしまふ。他の特徴で、例として「凹んでいる」という情報をロボットに与える。

図 3.13 のように、分割した面の向き、各々の面の位置関係から、物体の形状の凹凸を知ることができる。図 3.13 はお皿を面分割したものであり、赤い面を中心に、青・緑の面の法線ベクトルが赤い面に向かって傾いていることから凹みのある形状と判別できる。



図 3.13 : 法線ベクトル方向による凹凸の認識

3.3.2.2 文字形状検出

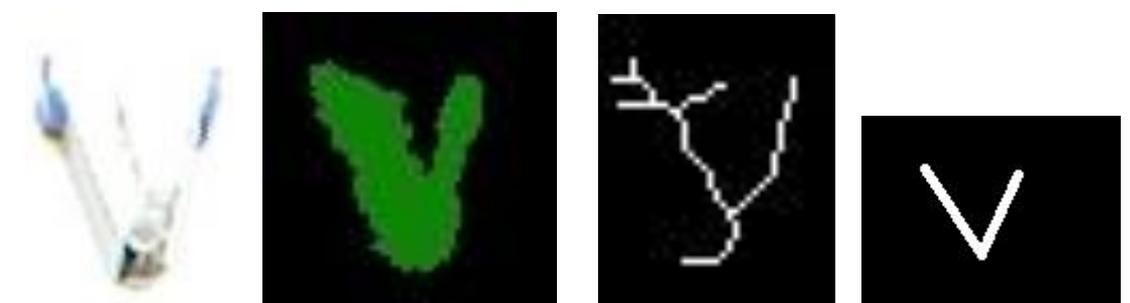
第2章で、対処物を幾何学的な形状で表現できない場合、『文字（アルファベットなど）』で表現する場合がある。そのため、非幾何学形状を文字認識で認識することを考えている。

ただし、この文字形状認識は、まだ認識が成功していないので、[78, 81, 82]を参考に今後の課題とする。

認識は、まだ成功していないが、検出の手順のみを以下の図 3.14 に記す。

例として、ホッチキスは、「V字」と表現されたので、ホッチキスの画像(図 3.14(1-a))から、kinect を用いて領域(図 3.14 (1-b))をセグメントする。今度は、検出した領域を細線化(図 3.14(1-c))し、目的の検出する文字の書かれた画像(図 3.14(1-d))とパターンマッチングし、目的の文字形状なのかを判定する。

同様に、メガネは「U字」で表現されているので、Uで文字認識する(図 3.14 (2-a ~2-d))。



(1-a) 元画像

(1-b) 領域検出

(1-c) 領域の細線化

(1-d) 文字画像

(1) ホッチキス（V字）の検出



(2-a) 元画像

(2-b) 領域検出

(2-c) 領域の細線化

(2-d) 文字画像

(2) メガネ（U字）の検出

図 3.14 : 文字形状の輪郭線検出

3.3.3 上面・側面形状

全体的な形状では表現できない場合、物体の一部分の形状、すなわち底面や角の状態から物体形状の候補を絞る。

下の例(図 3.15)では、円柱を検出するために、円柱の上面を構成する円を検出する。「円が付属している」と人が言うとき、法線ベクトルで分別した各々の面の円軽度を計算し、円形度が一定値なら円と見なす。



(a) 元画像(赤矩形：目的の物体形状)



(b) 領域画像



(c) 面分別



(d) 円検出

図 3.15 : 部分的な幾何学形状検出

3.3.4 部分形状

3.3.4.1 角形状判定

画像中の全物体の形状が矩形の場合、2章の対話実験で、依頼者は、物体の角の状態（角ばっているか、丸みがあるか）を用いていた。

形状情報として、「角が丸い」というと、画像中から物体のコーナー部分の輪郭線を検出する。コーナーの検出に成功すれば、角が丸いか、角ばっているかを判定する。角が丸いかどうかの判定は、楕円を当てはめ、一定値以上の円形度を持っていれば丸いとみなす。

下の図 3.16 に示すように丸い角には、青い円を当てはめ、緑の矩形で囲む。角張っているコーナーは赤い矩形で囲む。以下の図 3.17 でも複数の物体から、「角が丸い形状」の物体を検出した例を示す。

もし丸い角を持つ物体が複数検出した場合には、丸い角が多い順に候補番号をつけ、ユーザに順番に確かめさせる。

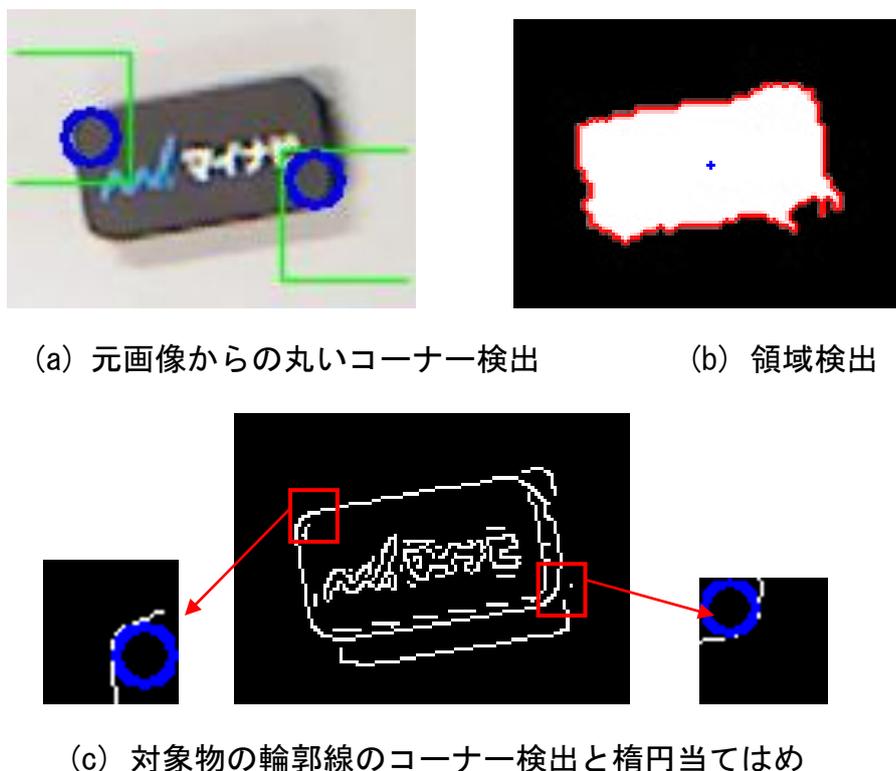


図 3.16 : 角形状判定

また、距離画像データと RGB 画像データのそれぞれに対して、エッジ検出を行い、

そこから円，楕円，多角形を検出する．角が丸められた物体についても，「丸い」という表現が用いられることがあるので，多角形の検出後，そのコーナーの丸み度を求める．具体的には，Harris[40]のコーナー検出により，コーナーを検出した後，その周辺の輪郭線に円を当てはめる．その当てはめの誤差を丸み度とし，それが一定の値以下の場合，丸いコーナーと見なす．

[1]直線・曲線エッジを検出

検出した輪郭点列からなるべく多くの直線・円弧を検出するために，点列全体に区間長と開始点を変化させながら，近似函数として直線・円弧をあてはめ，点列と近似函数の平均誤差 e_1 と最大誤差 e_2 を計算して，その値がある閾値以内ならその区間を直線・円弧として記憶しておくことを考える．しかし，輪郭線全てに直線と円弧を当てはめるのでは計算量が膨大になってしまうので，まず，輪郭点列から明らかな隅点をDavis法のデジタル曲率[43]によって抽出し，その隅点で区切ったセグメントに対して二分探索的手法を用いることで直線・円弧を検出することにする[95]．

[2]楕円検出手法

第1ステージとして，検出した曲線エッジ上から選んだ5つの画素から，楕円を抽出(楕円方程式との照合)する[96]．

ここで用いたくりこみ法はエッジ片が短くても楕円全体が精度良く当てはまる手法であり，この特性により上記の処理が有効に機能する．環状領域の計算は次のように行う．楕円を含む2次曲線は次の形の方程式をもつ．

$$Ax^2+2Bxy+Cy^2+2f(Dx+Ey)+f^2F=0 \quad (1)$$

f はデータと同じオーダの定数(例えば画像サイズ)である．上式が楕円を表す条件は

$$AC - BB > 0 \quad (2)$$

である．

また， $AC - BB \leq 0$ の場合，双曲線の式となる．

ベクトル x と行列 Q を

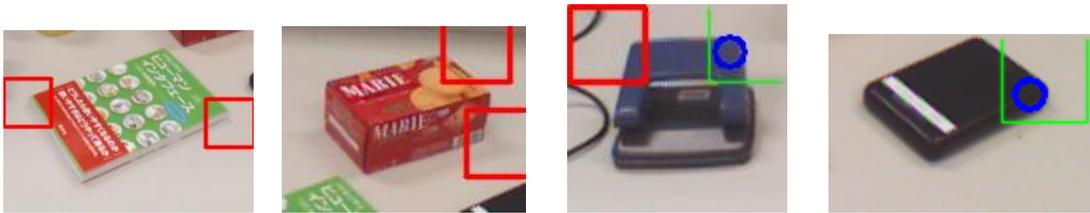
$$x = \begin{pmatrix} x/f \\ y/f \\ 1 \end{pmatrix} \quad Q = \begin{pmatrix} A & B & D \\ B & C & E \\ D & E & F \end{pmatrix} \quad (3)$$

と定義すると、式 (1) は次の式(4)と書ける.

$$(x \cdot Qx) = 0 \quad (4)$$

ただし $(a \cdot b)$ はベクトル a, b の内積を表す.

これから、上記に示したように、初期の段階で、曲線エッジは形状によって、楕円か双曲線のどちらかの一部に判定される. 第2段階以降で楕円行列式を再計算する場合、初期段階で楕円か双曲線とみなされたエッジでは、計算方法が若干異なる.



(a) 元画像からの丸いコーナー検出



(b) 輪郭線の検出



(c) 対象物の輪郭線のコーナー検出と楕円当てはめ

図 3.17 : 角形状物体の検出

3.3.4.2 穴の空いた物体

今度は図 3.18 に示すガムテープといった、穴の空いた立体の物体を斜めから見て認識した結果である。斜めから見たときのガムテープを面分割し、表側の面A（オレンジの矩形）に対し、面B（紫の矩形）は、面Aに向かい合わせの面の裏側であることを識別できる。図 3.19 (c) を見るように面Aの法線ベクトル群は発散していることから面Aは凹であることがわかる。それに比べ面Bの法線ベクトル群は外側に向かうにつれ収束していることから凹であることがわかる。対象物体上で、位置関係が向かい合わせであり、一方が凹であり、もう一方が凸であることから、対象物は穴の空いた立体物であることが推測できる。



(a) 穴の空いた物体検出

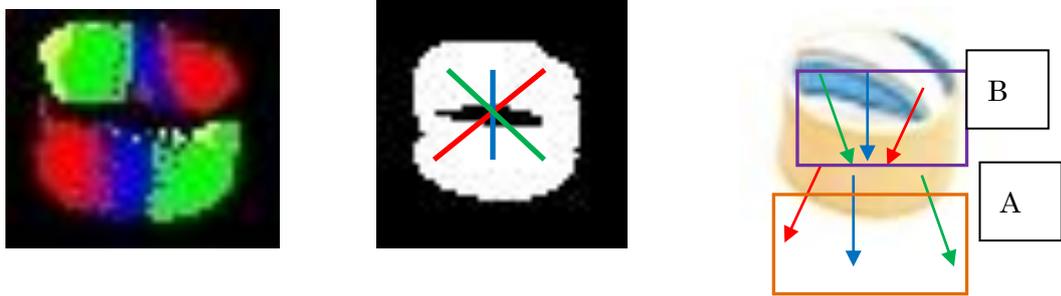


(b) セグメント



(c) ラベリング

図 3.18 : セグメントとラベリングによる検出

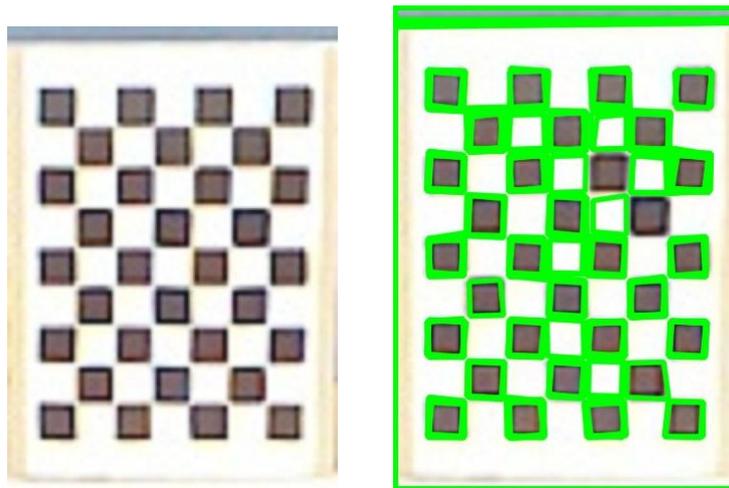


(a) 元画像 (b) 面分割 (c) 法線ベクトルによる凹凸の認識

図 3.19 : 法線ベクトルの向きによる検出

3.3.4.3 模様検出

下記の図 3.20 のような対象物の場合、立体的な凹凸部分であるのか、それとも平面的な白黒のチェック模様（描画）であることが考えられる。このため、対象物に対しての法線ベクトル（赤外線カメラで検出したによる 3 次元形状の各面に対しての法線ベクトル生成）の分布傾向を解析して、対象物が平面と判別された場合には、模様等の描画と見なす。



(a) 元画像 (b) 模様結果

図 3.20 : 矩形模様検出

3.3.5 物体の大きさの判定

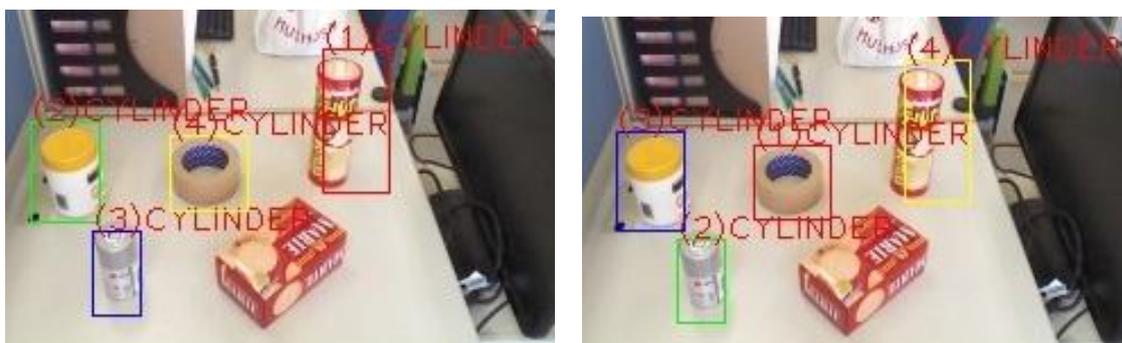
物体で同じ形状が複数ある場合、高さや大きさに形状を判定する。

下の例（図 3.21）では、対象形状で円柱に該当する物体が複数ある場合、人は、高さ（長い、短い）情報を用いる場合がある。

「長い（もしくは短い）円柱」というと、形状が円柱である物体のみで、長さ（長い、短い）順に並べ、候補順とする。

また、「細長いもの」と言った場合、

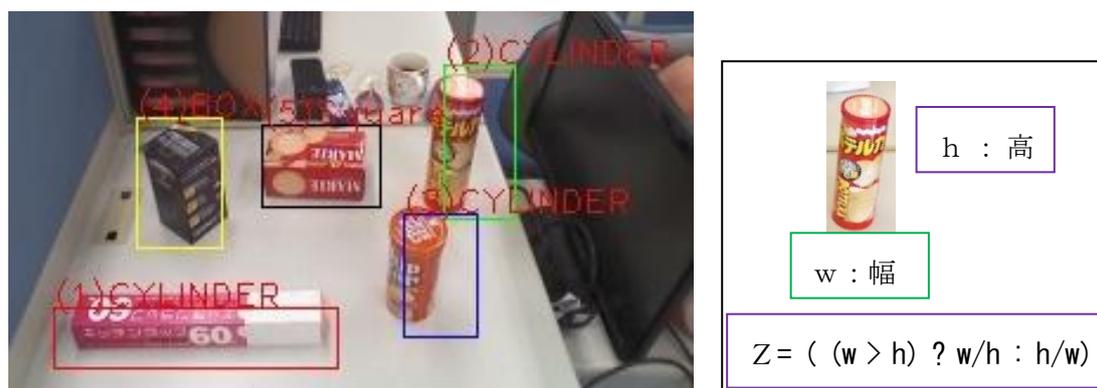
物体の縦横の比率 $Z = A$ （縦・横のどちらか長い方） / B （縦・横のどちらか短い方）を計算し、図 3.22 (a)のように比率の高い順に候補順に番号を付ける。



(a) 長い順

(b) 短い順

図 3.21 : 物体の高さ判定（対象：円柱）



(a) 細長い順

(b) 細長い比率 Z の四季

図 3.22 : 細長い物を検出

4章 オントロジーに基づいた

対話認識認識システム構成

4.1 物体認識手法の構成

本研究で提案する物体認識システムの構成を図 4.1 に示す。また、図 4.2 に対話物体認識の全体像を示す。自動での認識に失敗したら、まず利用者との対話を行うことによって、利用者から対象物体に関する情報を教えてもらう[62]。次に、取得した対話情報から物体の特徴情報を抽出し、該当する特徴をカメラ画像内から画像検出する。例として、物体の形状情報を取得したら、形状物体認識を行う。そして、候補となる物体を見つけたら、その物体が目的の物体であるのかを利用者へ確認する。もし、候補となる物体を見つけれなかった場合や、見つかった物体が目的の物体でない場合は、再び利用者と対話して、新しい物体情報を教えてもらう。

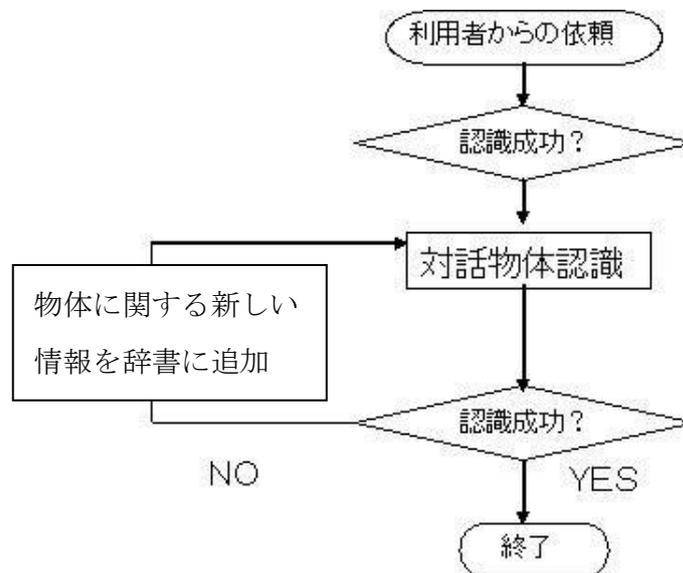


図 4.1 : 提案する物体手法の構成

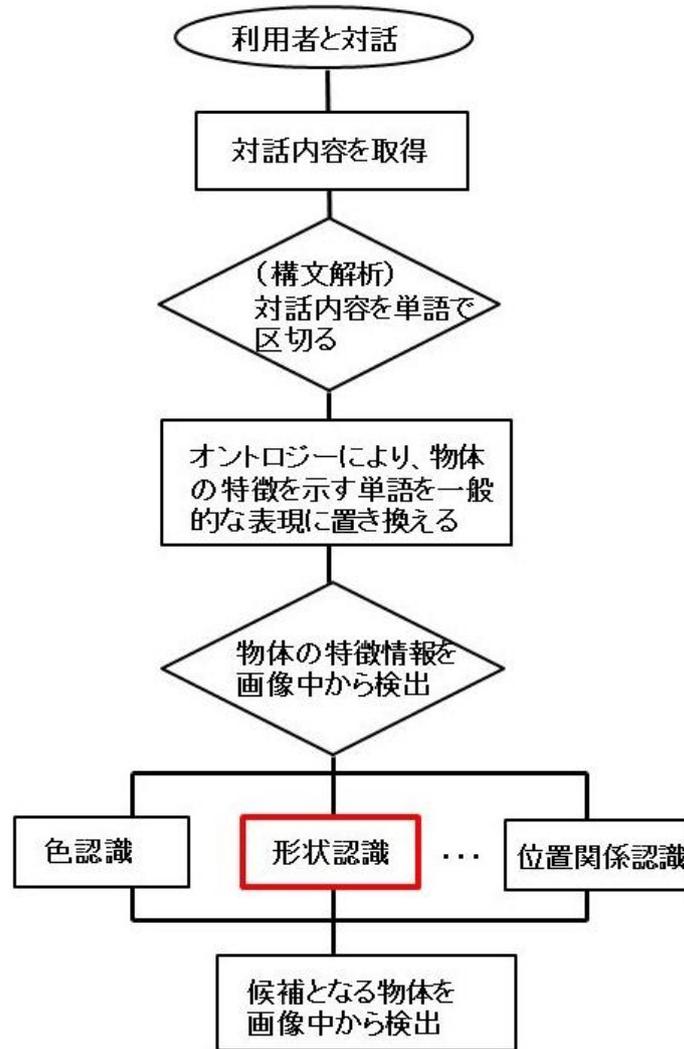


図 4.2 : 対話物体認識の全体像

人がロボットに依頼すると、ロボットは「人の言葉」を理解しようとする。このとき、人から聞いた物体の特徴が辞書（特徴言語を記載したデータ）にない場合は、新しい特徴言語をこの辞書に追加することで、ロボットは物体に関する知識を獲得する。2章で述べたように、ロボットに人の言葉を理解させるために、人が日常使う言葉を調査・収集し、ロボットが理解できる表現に置き換える必要がある。そこで、オントロジーという言葉同士の関係を結ぶ概念を用いり、「人の多種多様な表現」とそれに結びつく「一般的形状表現」をセットにし、辞書に登録する。そして、ロボット(コンピュータ)は、辞書から画像処理で探索可能な物体特徴を参照し、実際の物体を画像中

から探索する。

利用者に名前で指示された物体に対して自動認識を行い、それが検出できない場合に対話物体認識に入るという統合物体認識システムの実現を目指している。形状だけでなく、色や位置関係など、その場の状況に応じた適当な属性について対話を行う方法を検討している。

今回は、提案の物体認識用オントロジーの有効性を確認するために、人の表現が最も多種多様である形状情報に注目し、形状に関する対話の部分に限った実験システムを開発した。対話の部分についてもオントロジーの有効性の確認が主目的なので、音声認識の失敗などを除くため、音声ではなくキーボードからの文章入力を用いることにした。

以下に、対象物の判別方法と流れ（図 4.3）を示す。

- (1) 入力カメラ画像から物体群と背景とを領域分けする。
- (2) 深度センサーで物体の3次元座標値を取得し、3次元座標値から物体の法線ベクトルを計算する。
- (3) 依頼者が持ってきてほしい物の名称と形状情報(形状表現語)をコンピュータ(PC)に入力し、オントロジーに基づいた辞書(データ)により形状表現の理解をする。
- (4) DBに記憶しておいた対象物の視点对応の形状、法線ベクトルの特徴(法線グループのヒストグラム)等、に基づいて、依頼(目的)の物をカメラ画像から検出する。
- (5) 候補となる形状を発見できなかつたり、2個以上発見したりすると、PCは再度ユーザと対話を行い追加の形状情報を得る。
- (6) 形状情報に該当する形状を1つ検出できれば検出は成功と見なす。もし、目的の形状でなければ、またユーザから新たな形状情報を聞き出して画像中から形状検出をするという動作を繰り返す。

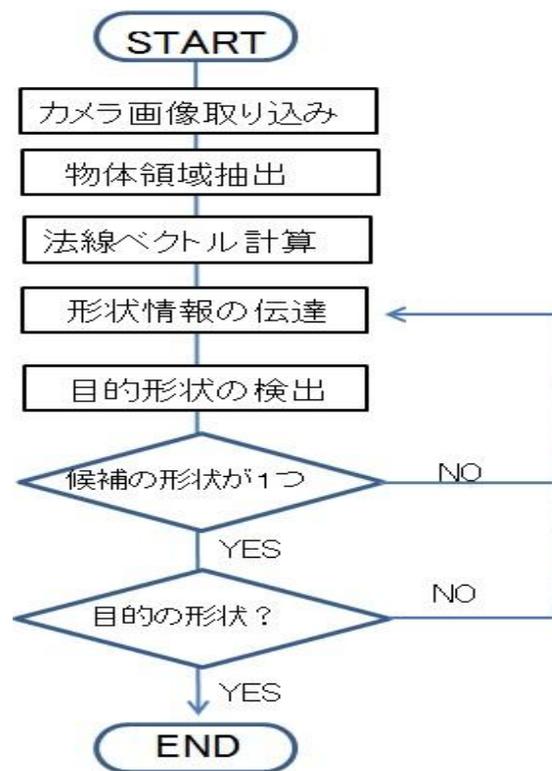


図 4.3 : 形状検出手順のフローチャート

4.2 対話認識実験でのハードウェア構成

4.2.1 使用するロボット

対話物体認識では、NAO[80](下記の図 4.4)というロボットを用いる。Nao は、自立歩行する小型ヒューマノイドロボットであり、フランスのパリに拠点を置く Aldebaran Robotics 社により開発された。自立的でありながらプログラム可能な NAO は初心者から上級者まで広範囲のユーザーの使用に適している。Aldebaran Robotics 社が開発したソフトウェア「CHAOREGRAPHE」によってユーザーはロボットの動作と対話シナリオを簡単に作成して編集することができる。

下記の表 4.1 に NAO の仕様を記載する。



図 4.4 : ロボット NAO

表 4.1 : NAO 仕様

<p style="text-align: center;">MOVE</p> <ul style="list-style-type: none"> • 25 自由度 • コアレスモーター • ソフトウェア制御 	<p style="text-align: center;">THINK</p> <ul style="list-style-type: none"> • CPU:Geode 500MHz • 256MB SDRAM • 2GB フラッシュメモリ • 各種ソフトウェア+SDK
<p style="text-align: center;">SENSE</p> <ul style="list-style-type: none"> • カメラ×2 • マイク×4 • FSR 圧力センサー(各足裏に 4×2) • バンパー×2 • ソナー×2 	<p style="text-align: center;">INTERACT</p> <ul style="list-style-type: none"> • スピーカー×2 • 複数の LED • 触覚センサー (頭部&両手) • 物をつかめる手 • 赤外線センサー • Wi-Fi (IEEE802.11b/g)

4.2.2 使用するカメラ

ロボット（PC で構成）の視覚機能として kinect（赤外線カメラ：図 4.5）を用いる。Kinect（キネクト，開発コードネーム：Project Natal）とは，既発売の Xbox360 全機種対応の専用周辺機器の一つである。図 4.5 の kinect は，Microsoft Windows 向けに使用する Kinect for Windows である。

Kinect センサー（動作中に青白く光る RGB カラー映像認識用カメラ 1 基，奥行き測定用赤外線カメラ 1 基，赤外線照射発光部（IR エミッタ）1 基の三つの光学系センサーと音響センサーとして，4 本のマルチアレイマイク）が前面に並び，左右に首を振る電動チルト機構が装備された横に細長い光沢仕上げの黒色（リキッドブラック）直方体の基本形状に前面はかまぼこ型に緩やかな丸みを帯びた本体（筐体）デザインである。

内部には 3 軸加速度センサー，64MB の DDR2 SDRAM メインメモリ，信号処理専用プロセッサが内蔵されている。



図 4.5 : Kinect（Windows 用）赤外線カメラ

4.2.3 使用するマイク

音声認識用のマイクとして、SONY の ECM-C10 (エレクトレットコンデンサーマイクロホン) (図 4.6, 表 4.2) を使用した. このマイクを用いて, 利用者からの命令を音声認識する.



(左図) マイク正面

(右図) マイク後方

図 4.6 : ECM-C10 の外観

表 4.2 : ECM-C10 の仕様

名称	ECM-C10 (エレクトレットコンデンサーマイクロホン)
型式	エレクトレットコンデンサー型
プラグ	金メッキL型ミニプラグ
外形寸法	約 $\phi 8 \times 19\text{mm}$ (最大径 \times 全長)
質量	約 10g (コード含む)
周波数特性	50~15,000Hz

4.3 対話認識システムの構成

はじめに, 図 4.7 のような対話認識システムの全体像を述べる. システムは, 人音声による依頼があると, その内容に応じた「物体の特徴」をカメラ画像内から検出するようにしている. 例として, 「丸いものを取って」という依頼があった場合には, カメラ画像中から候補となる物体を検出すると, それが利用者の求めている物体なのかをロボットのNAOを通じて, NAOに対象物に向けて指を指し, 依頼者に「これで

すか？」と問い合わせる。

対話認識システムの構成を以下の図 4.7 に示す。システムは以下の 4 つのモジュールから構成される。

- [1] 音声認識モジュール
- [2] 言語理解モジュール
- [3] 物体認識モジュール
- [4] 音声発話モジュール

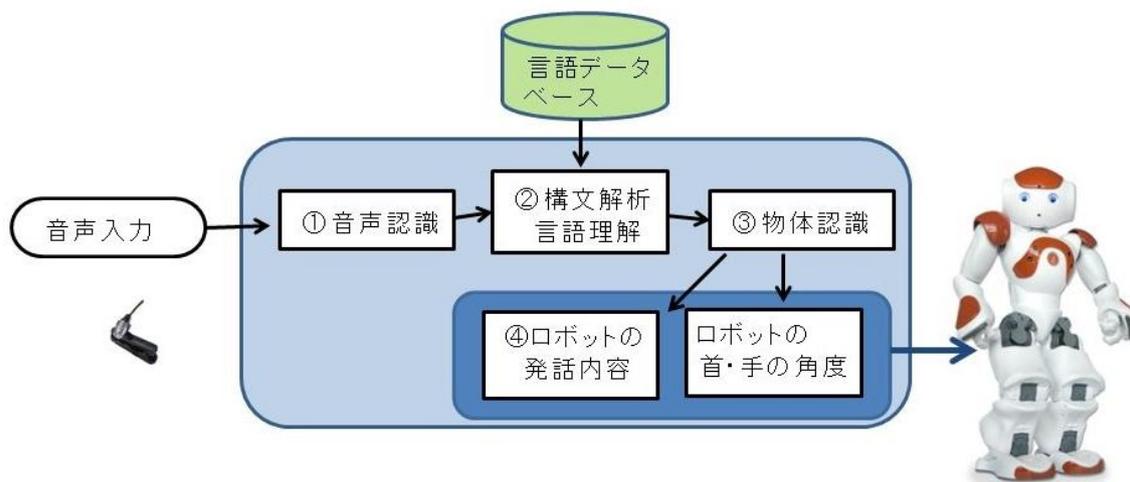


図 4.7 : 対話認識システム構成

4.3.1 音声認識モジュール

「ティッシュ取って」や「その四角くて、角の丸いもの」などといった利用者が発した音声の認識を行う。音声認識用ソフトとして、オープンソースの連続音声認識ソフトウェア Julius [98]を用いる。構文解析用のソフトでは、日本語係り受け解析器 CaboCha [112]と、オープンソース 形態素解析エンジン MeCab [113]を用いる。人がマイクに向かってしゃべると、音声が入力され、以下の図 4.8 のように、CaboCha によって、コマンド上で、「取得した音声の情報」が文字列に分割されて出力される。

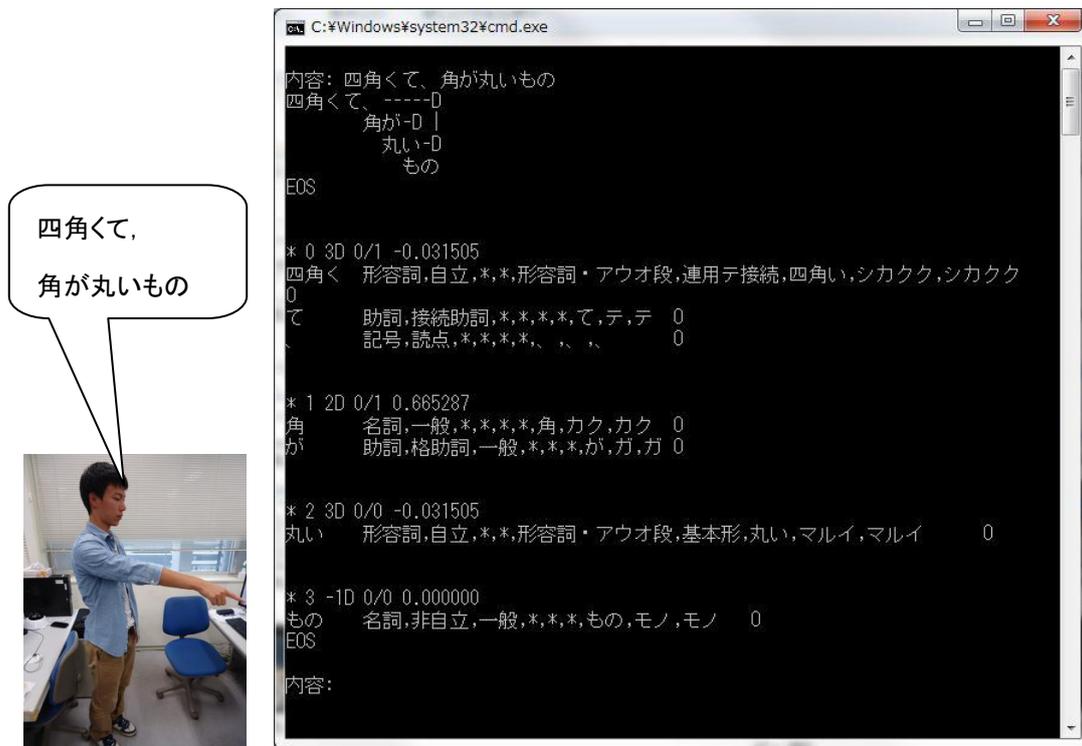


図 4.8 : 利用者の声を認識

CaboCha と MeCab を用いり，想定される「発話」と「文法法則」をあらかじめ記述することによって，「発話内容の物体の特徴」を物体認識プログラムに高い精度で認識させることができる．それを行うには，以下の図 4.9 に示す，(a)「人の形状表現と一般的形状語句を結びつける単語辞書」と(b)「想定される発話 case の文法規則」を使用する．まず，(b)の中から，人が発した形状表現に該当する「文法規則」を見つけ，検出した「文法法則」のとおり，発話内容を単語で区切り，その中から形状語句を検出する．そして，取得した「人の形状表現 ((a)の赤い項目)」に該当する「一般的形状 ((a)の青い項目)」を探索し，物体認識プログラムとロボットに，対象形状の情報を送信する．

人の表現		一般的形状語	
A		句 B	
1	球	球	
2	ボール	球	
3	筒	円柱	
4	直方体	直方体	
5	立方体	直方体	
6	箱	直方体	
7	円	円用語.csv	
8	楕円	楕円	
9	円柱	円柱	
10	まん丸	まん丸用語.csv	
11	丸	丸用語.csv	
12	丸い	丸用語.csv	
13	丸いもの	丸用語.csv	
14	丸く	丸用語.csv	
15	丸っぽい	丸用語.csv	
16	丸み	丸用語.csv	
17	ドーナツ	輪	
18	輪	輪	
19	わっか	輪	
20	空洞	輪	
21	穴	輪	
22	四角	四角用語.csv	
23	四角形	四角用語.csv	
24	四角い	四角用語.csv	
25	四角柱	直方体	
26	矩形	四角用語.csv	
27	台形	四角用語.csv	
28	長方形	四角用語.csv	
29	V字型	文字形状.csv	
30	角が丸い	角丸	
31	角丸	角丸	
32	角張り	角張り	
33	角張っている	角張り	
34	穴	穴	
35	縦長	縦長	
36	横長	横長	
37	長い	長い	
38	短い	短い	

(a) 「人の形状表現」に関する単語辞書 (b) 想定される発話 case の文法規則

図 4.9 : 音声対話用に作成した文法ファイルの一部

4.3.2 言語理解モジュール

音声認識モジュールから送られてきた認識文の解析を行う。音声認識モジュールで認識された文字列がどういった意味を持っているのかを、ロボットに理解させる。

表 4.3 : 語彙データベースの一部

ID	単語名	カテゴリ
1	皿	物体名
2	ボール	
3	ティッシュ	
...	...	
1	箱	物体の種類
2	本	
...	...	
1	球	形
2	円	
3	円柱	
4	直方体	
5	凹んでいるもの	
6	角が丸いもの	
7	角張っているもの	
...	...	
1	長い	長さ
2	短い	
...	...	
1	取ってきて	動作
...	...	

本モジュールでは、受け取った文字列に含まれている語彙を抜き出すという処理を行う。語彙の抜き出しには、あらかじめ作成しておいた語彙データベース（上記の表 4.3）を利用する。データベースには、各語彙の ID と名前、カテゴリを登録しておき、与えられた文字列とデータベース内の文字列を比較していくことによって認識文に含まれる語彙を抜き出す。

4.3.3 物体認識モジュール

Kinect カメラが取得した入力画像内から目的とする物体の認識を行う。言語理解モジュールから得た物体の特徴情報から対象物体を認識する。

例として、言語理解モジュールから“ボール”という物体情報が送られてきたなら、形状物体認識により、カメラ画像中から“球”の検出を行う。そして、そこで得られた認識結果を、次の音声発話モジュールへ送る。

4.3.4 音声発話モジュール

ロボット(NAO)に、依頼者に対してどのような発話をさせる場合、前もって、NAO に発話させたい音声ファイルを作成して、状況に合わせてそれらのファイルを読み込ませて NAO に発話を行わせる。例として、物体認識によって候補となる物体が見つかった場合には、その方向に向かって指を差すのと同時に、「これですか？」などという発話をさせる。もし、違っていたら「すみません、間違えました」と発話させ、合っていたら「わかりました、この物体を取ります」等と発話させる。

4.4 オントロジーによる言葉理解

第2章で、対話実験によって収集した「人の形状表現」と「形状語句に関するオントロジー表」を結びつけ、対象物に対する「人の表現」、「候補形状」、「検出方法」を、以下の表 4.4 のようにまとめて、ロボットに人の言語を理解させる。

例として

- ・丸い→全体形状が丸い、特定の方向から見たとき丸い、物体の一部分が丸い
- ・A (A : 真上, 横, 斜め) から見ると丸い→Aから見たときに丸いもの
- ・角が丸い→物体の角部分が丸い
- ・V字型である→物体の全体形状がV字である。
- ・格子模様である→対象物上に複数の矩形がある。

表 4.4 : 人の表現に対応した形状の表

対象物	人の言葉表現	注目部分	候補となる形状	優先度	検出方法
ボール	丸い	全体	球	1	3次元ポイントデータの追跡
			円・楕円	2	3次元ポイントデータの追跡
			円柱	3	3次元ポイントデータの追跡
		特定の方向	円・楕円	4	面検出 +楕円当てはめ
		一部分	曲線(角)	5	角検出 +楕円当てはめ
チップス	A (A : 真上, 横, 斜め) から見ると丸い	特定の方向 (真上, 横, 斜め)	円・楕円 円柱 (真上のみ)		3次元ポイントデータの追跡
ホッチキス	V字型である	全体形状	V字		SIFTによる 特徴点の一致
ボード	格子模様である	全体形状	複数の矩形		矩形検出

4.5 画像形状検出

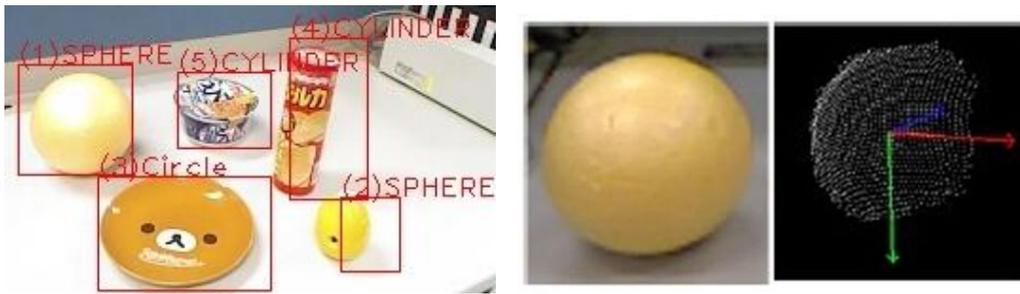
ユーザに名前で指示された物体に対して自動認識を行い、それが検出できない場合に対話物体認識に入るという統合物体認識システムの実現を目指して開発を進めている。ここでは、形状だけでなく、色や材質、それに位置関係など、その場の状況に応じた適当な属性について対話を行う方法を検討することを考えている。今回は、提案の物体認識用オントロジーの有効性を確認するために、オントロジー中の「属性」の「形状属性」および「自然言語による属性表現」の「形状の表現」に関する対話の部分に限った実験システムを開発した。対話の部分についてもオントロジーの有効性の確認が主目的なので、音声認識の失敗などを除くため、音声ではなくキーボードからの文章入力を用いることにした。

画像入力には Kinect を用いた。実験システムでは対象物体はテーブル上にあると仮定し、背景のテーブルとその上の物体を、距離画像を用いて分離し、個々の物体を切り出す。切り出された物体に対して法線ベクトルを計算し、その方向を正 12 面体の方向に粗く離散化しヒストグラムを求める。そして、そのパターンから球、円柱、直方体等の概略の 3 次元形状を判別する。

形状認識をする際に、以下の表 4.5 のように、「対象物の形状の特徴をまとめた表」を作成する。図 4.10 のように、画像中から表 4.5 に当てはまる特徴を持つ物体を検出する。

表 4.5 : 対象物に対する特徴形状の情報のまとめ

対象物名 :	ハードディスク	ボール	チップスの箱
形状 :	矩形	球	円柱
視点方向 :	真上	全体	全体
構成形状・数 :	角 × 4		
丸み度(最大 1) :	0.15 未満	0.85 以上	
長さ :			細長い
角形状 :	曲線		
模様 :			水玉

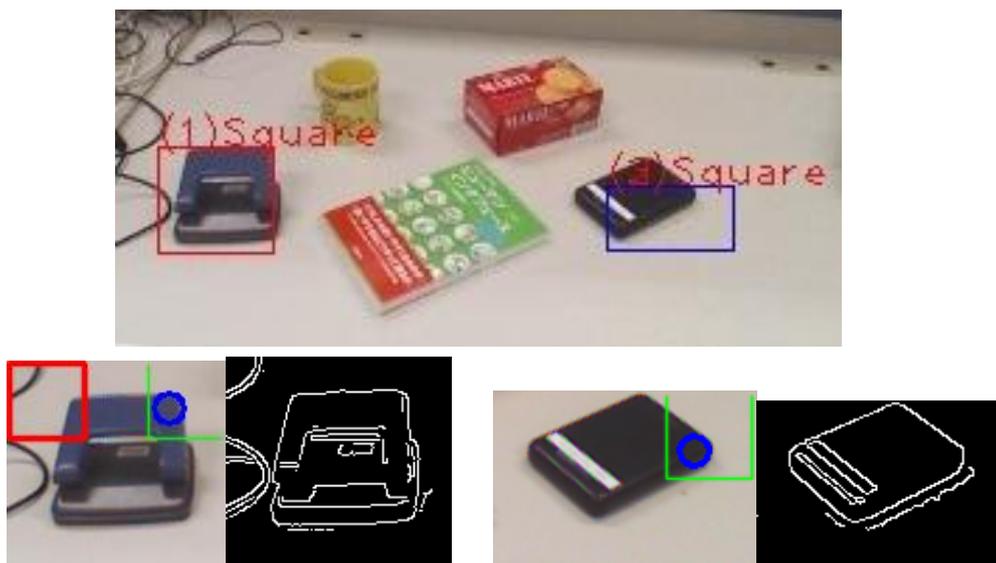


(a) 3次元ポイントデータによる形状検出
(候補はボールで、丸いもの順にソート)



長さの割合 $N=W(\text{幅}) / H(\text{高さ})$

(b) 対象物の長さの大小関係の判定



(c) 角の形状判定・物体検出

図 4.10 表 4.5 に対する形状の検出例

形状に関する対話は以下のように進められる。ユーザの入力に対象表現が明示されていれば、それに応じて表現の対象が限定される。例えば、「上から見たら丸い」と入力されれば、対象は投影された形状であり、視点は上方からと解釈される。そのような対象表現がない場合は、システムはすべての対象を考慮する。システムは、その対象について、属性表現で示された形状を持つ物体を画像中から検出することを試みる。属性表現が物理的に複数の形状を表す可能性がある場合は、その物理形状のところに示された優先順位で形状を調べる。指示された物体が検出できたら、ユーザに確認を求める。目的の物体でなかった場合は、次候補があれば、それを示すか、追加の形状情報の入力をユーザに求める。

4.6 ロボットの動作

ロボットNaoの動作として、目的の物体に指を指す動作を行う。

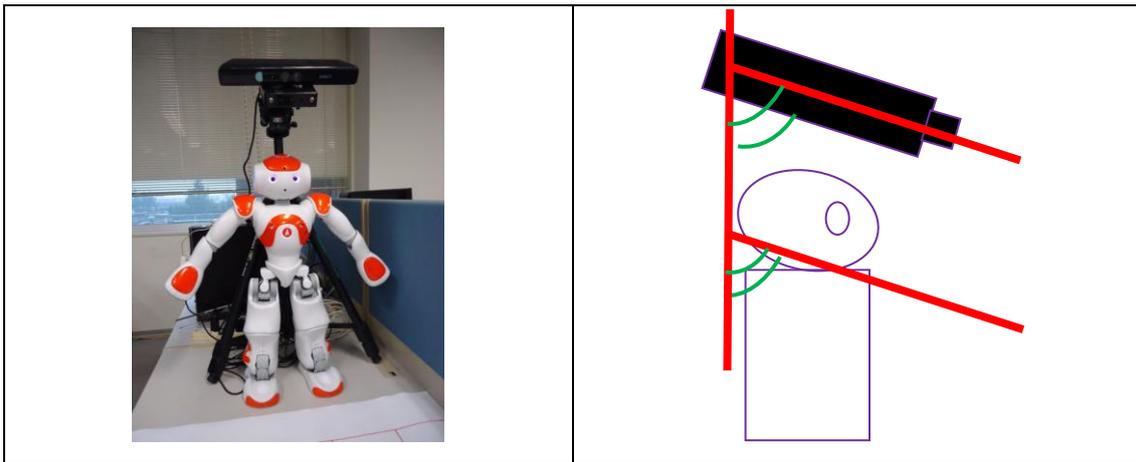
Naoに指を指す動作を行うには、対象物の現実上での位置座標を取得し、取得した座標からNaoの手の角度を計算しなければならない。

まず、kinectカメラから、距離画像を取得し、距離データから各画素の世界座標系（単位はメートル）を取得する必要がある。

4.6.1 Naoの位置と関節角度

取得した距離画像の世界座標系を用い、ロボットNaoの手と首の角度を計算する。

図4.11で示すように、Naoは、kinectの真下に配置し、kinectのカメラ角度と、Naoの首の角度を合わせる。使用するNaoの関節と角度を以下の図4.12、4.13に示す。



(a) 正面

(b) 真横

図 4.11 : Nao と Kinect の初期配置図

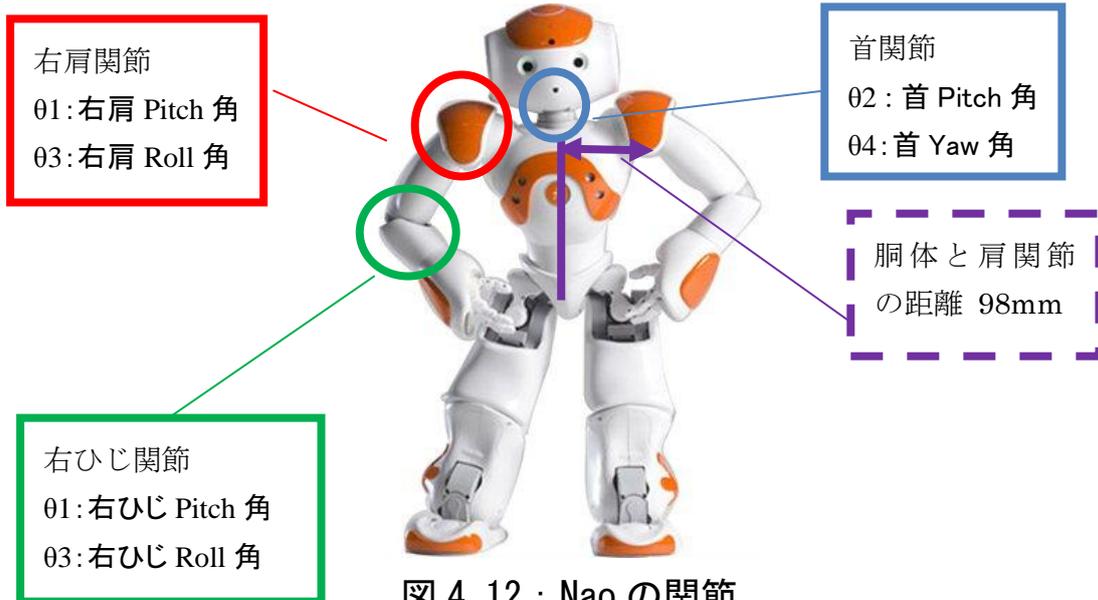


図 4.12 : Nao の関節

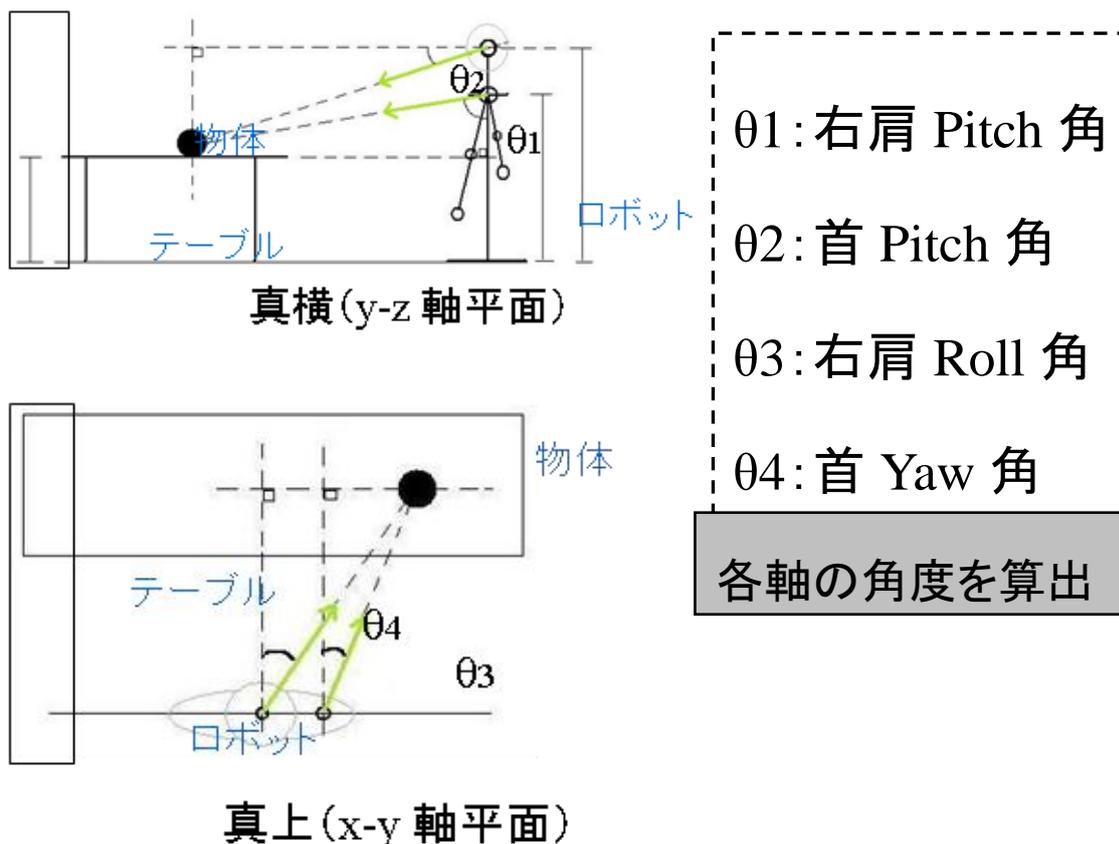


図 4.13 : 指さしの角度算出

4.6.2 ワールド座標の取得

Kinect の深度センサーの出力は画像データ，すなわちデプス画像である．デプス画像は 640×480 画素で，各画素 13 ビットの精度を持つ．画素の数値は深度センサーから被写体表面までの奥行きを M(メートル) 単位で表したものである．X-Z 軸上では，Nao (胴体) と kinect をワールド座標値の中心とし，右側の物体群には右手で差し，左側の物体群には左手を用いる．この場合，Nao の胴体と左右の肩の距離 (0.098 m) も考慮する．

画像データの各画素は被写体上の 1 点に対応している．したがって，深度センサーを中心とする 3 次元座標系において水平方向 (X 軸)，垂直方向 (Y 軸)，奥行き方向 (Z 軸) の座標値，すなわちワールド座標値を計算することができる．このなかで，画素の数値は Z 軸の値に相当する．X 軸，Y 軸の座標値は，画像座標と深度センサーの視野角

および奥行き (Z 値) から, 次のように求める.

Kinect の深度センサーから得られるのは物体からセンサーまでの奥行きである. センサーから物体までの水平垂直位置は画素単位で取得する. 画素単位の座標と Kinect の視野角, 距離から, 三角関数を用い水平垂直位置の実寸を求める.

被写体表面の点 P の, 実空間での 3 次元座標を (X, Y, Z) とする. 図 4.14(a) は水平方向で, X 軸に平行で点 P と投影中心 (C) を通る平面と, 近赤外カメラの投影空間が交わることで形成される三角形 AOC である. 図で, 角 AOC は, 近赤外カメラの水平視野角の半分であるから 28.5 度になる. w はデプス画像における中心から点 P までの X 軸方向の画素単位の距離である. 三角関数を用いると, 辺 AC の長さは $Z \cdot \tan 28.5$ となる. この長さはデプス画像の解像度を 640×480 画素とした場合, 320 画素分に相当する. 同様に, 図 4.14(b) の垂直方向での角度も計算する.

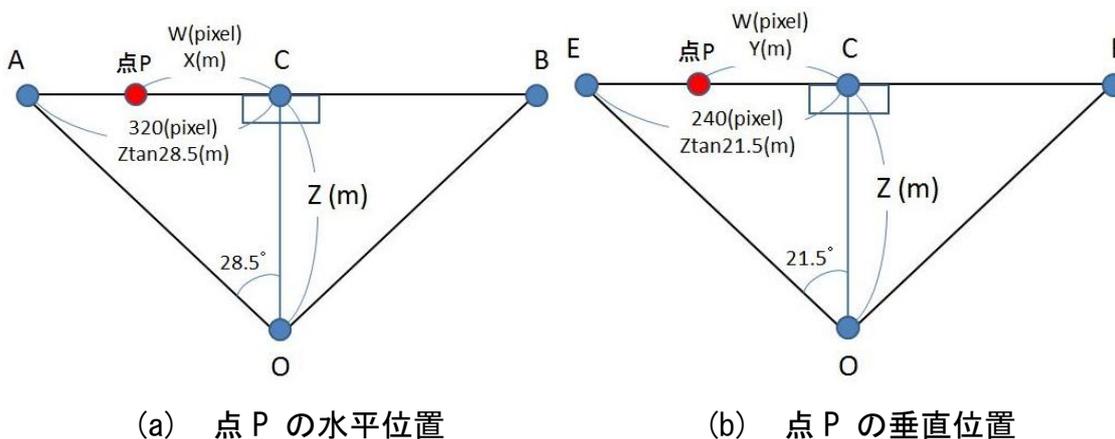


図 4.14 : 点 P のワールド座標上での位置

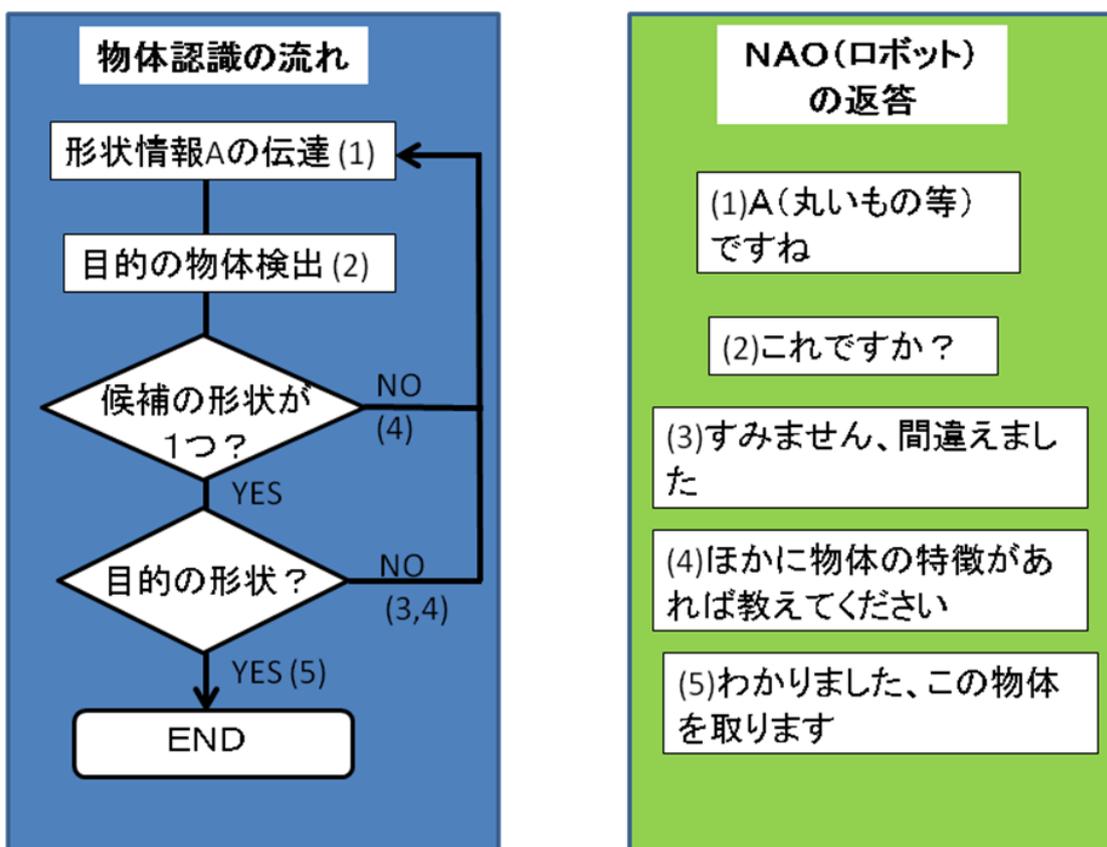


図 4.15 : NAO の対話内容

物体をセグメントした順に該当形状かどうか確かめ、該当物でもすぐにはユーザに確認しない。セグメントした物体の形状を全て確認し、優先順位を付けてソートしてから、上記の図 4.15 のような対話で、優先順にユーザに問いただしていく。ユーザの求めている形の順に聞くことで、余計な対話を減らすことができる。

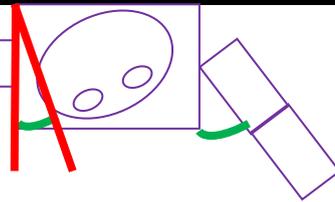
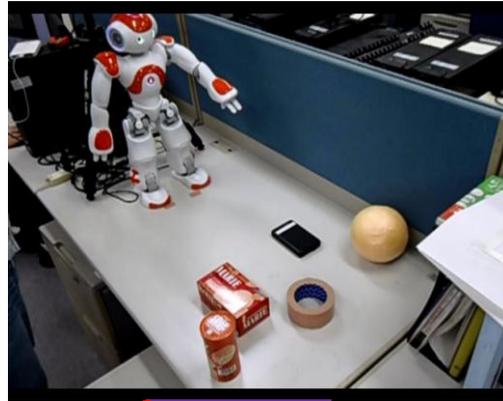
図 4.16 では、各物体を指すときのロボットの手の角度を水平方向から見たときの様子を示す。図 4.17 の「優先順位にソートされる物体群」で、上の物体群は、左から右にかけてセグメント検出し、なおかつ「丸いもの」に当てはまる形状物を並べたものである。下の物体群は、上の物体群を優先度順に並べたものである。もし、人が形状情報で「丸いもの」と答えると、形状に関する対話収集実験で、優先度が高かった順（球、円、楕円、円柱）にソートされる。この例だと、黄色い目覚まし時計を 5 番目から 2 番目の候補と移すことができ、対話回数を 3 回分減らすことができる。

★必ずしも、誰もが『丸いもの→1番：球』とは限らないが、今後の形状認識の成功・失敗で得られた形状情報の比率から、優先形状を更正することができる。

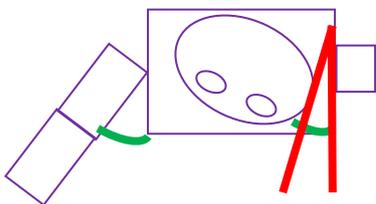


真上から見たときの
ロボットの水平方向での
手・首の角度

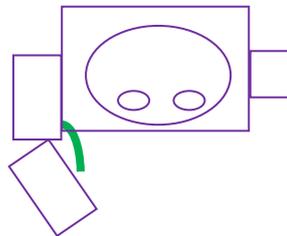
(a) 初期状態



(b) 水平距離 (x メートル) > 0.098
にある物体に対しての指差し

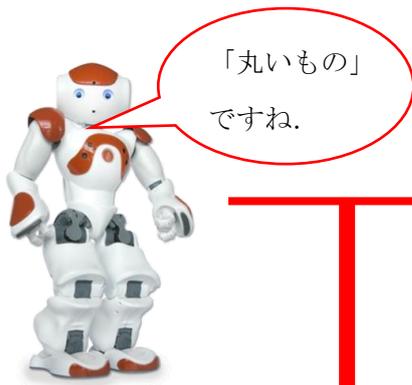


(c) $x < -0.098$ にある物
体に対しての指差し



(d) $(-0.098 \leq x \leq 0.098)$ にあ
る物体に対しての指差し

図 4.16 : 水平方向で見たときの物体を指す様子



形状表現	形状理解	形状候補と順位
丸いもの	丸みを帯びた形状	① 球 ② 円・楕円 ③ 円柱 ④ 丸い角



優先順位を考慮した形状オントロジー

優先順位にソートされる物体群



候補順にソート



図 4.17 : 優先順位を考慮した対話形状認識

5章 実験

5.1 基本動作確認実験

2.1 の実験で用いた物体から 3～5 個程度を選び、テーブルの上に置き、形状に関する対話の実験を行った。以下に代表的な例を示す。ここでは、システム(S)は初めに「どんな形ですか」と聞く、以下では、それ以降のユーザの入力(U)とシステムの対応を示す。

(1) 対象表現(投影)が明示された場合

図 5.1(c) のシーンで、カップ麺が対象物体の場合である。

U: 「上から見たら丸い」、S: 「これですか」、U: 「はい」

図 5.1(a) は上の対話を表示したシステムのディスプレイ画面である。「上から」という明示的な対象表現があるので、上から見て「丸い」物体を検出する。投影なので「丸い」の示す物理形状のうち、2次元か1次元のものとして「円」が対象物体の形状であるとシステムは理解する。ここでは詳細は省略するが、「円」が2次元形状であるというようなことは、「人間の属性表現」ではなく、2.4での図 2.12 に示される「形状属性」の中に記述されている。システムは輪郭線から円を検出した結果、図 5.1(b) に示す物体を候補として選択し、ユーザに確認を求め、正解であることを知る。

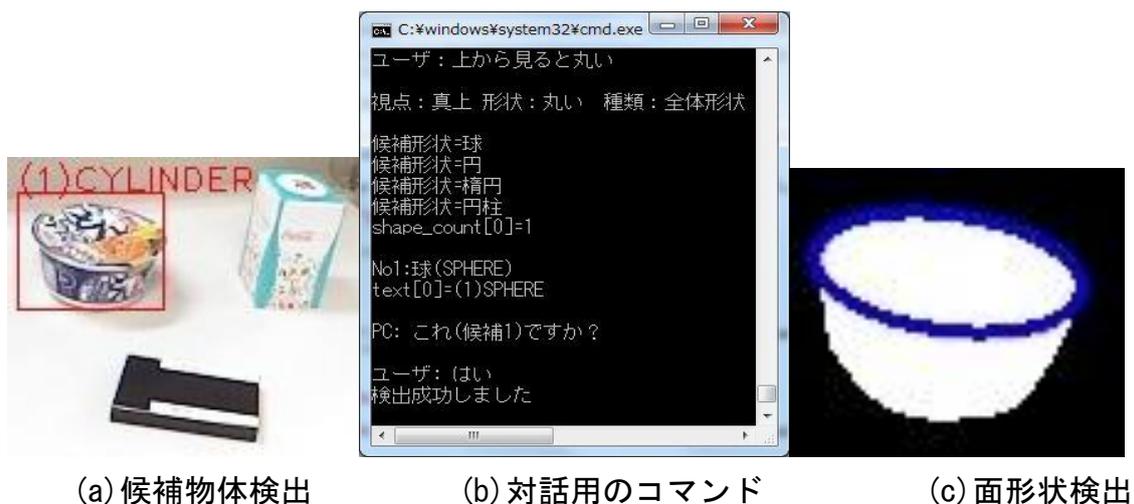


図 5.1 : 「上から見たら丸い」の検出例

(2) 対象表現が明示されていない場合

図 5.2 のシーンで左側のボールが対象物体の場合である。

U: 「丸いもの」, S: 「これですか」, U: 「はい」

対象表現が明示されていないので、丸いものと言われる可能性のある物体をすべて検討する必要がある(図 5.2 (a)). その中でシステムは、オントロジーの記述に従い、「球」(図中に(1)と示されたボール)を第一候補として、ユーザに確認を求める(図 5.2 (b)). この場合は、これが目的の物体だったので、ここで対話は終了する(図 5.2 (c)). もし、これが目的の物体でなかった場合は、このシーンにはもう一つ球形の物体(時計)があるので、それが指示された物体ではないかと確認を求める(図 5.2 (d)).



図 5.2 : 「丸いもの」の検出例

(3) 対象表現（構成要素）が追加された場合

図 5.3(a) のシーンで、左下のリモコンを対象物体としている。

U: 「四角いもの」

S: (候補 1 をディスプレイ上で示しながら) 「これですか」, U: 「いいえ」

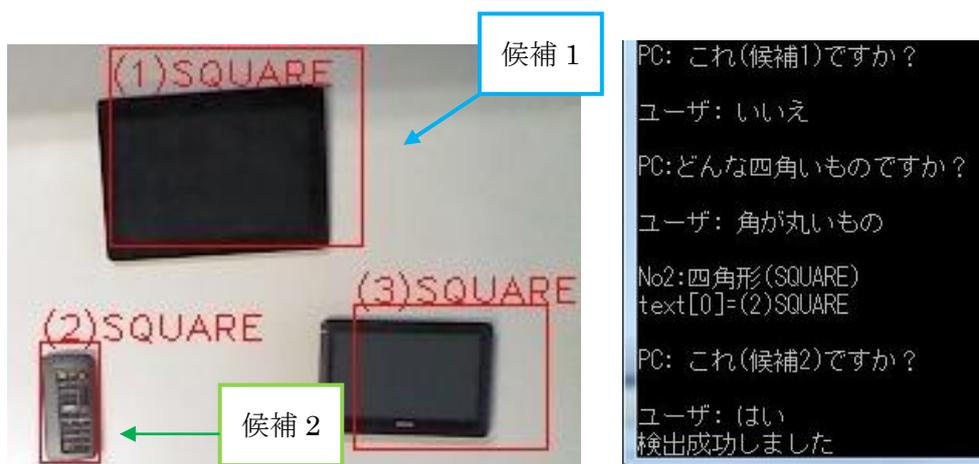
S: 「どんな四角いものですか」

U: 「角が丸いもの」

S: (候補 2 をディスプレイ上で示しながら (図 5.3 (d)) 「これですか」

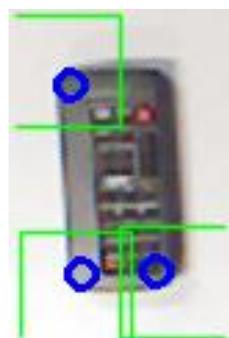
U: 「はい」

この例では、システムはまず候補 1 を四角いものと認識してユーザに確認を求める。しかし、それが違ったので、追加の情報を求める。「角が丸い」という表現が用いられたので、コーナーの丸み度を求め (図 5.3 (c)), 候補 2 ではないかとユーザに確認を求めて、認識に成功する。

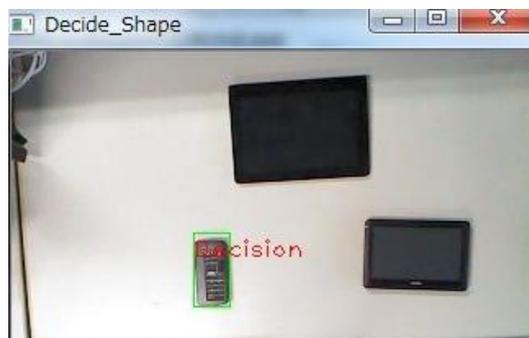


(a) 候補形状検出

(b) 対話用のコマンド



(c) 角形状の判定結果



(d) 候補物体検出

図 5.3 : 「丸みのあるもの」の検出例

5.2 性能評価実験

より定量的な評価を行うために 10 人の参加者（全員大学生）により，対話による形状認識の実験を行った．5 個の物体を置いたシーンをいくつか用意し，その中から対象物を形状認識により検出する．各実験参加者はシーン中の 5 個の物体のすべてについて，一つずつ順に対象物だと思ってもらい，それについて物体の形状情報を，対話で発してもらった．用いたシーンに対しては，システムは物体を正しくセグメンテーションできたので，参加者が「これですか」と確認して，違った場合はそれを除き，残りを対象にしていけば，必ず正解に達する．しかし，今回は参加者が思いついた形状表現でシステムがどれくらい認識できるかを調べるため，同じ表現を再度用いることは許さない条件で実験を行った．すなわち，参加者が思いつく「対象物の形状の表現」を全て発しても，対象物を検出できなかった場合は失敗とみなした．対話による形状認識の実験は，下記の 2 種類で行った．

(1) 参加者全員に同じシーンを用いた場合

5 個の物体を置いた 3 つのシーン(図 5.4)を用意した．各実験参加者は 3 つのシーンの合計 15 個の物体のすべてを対象物として形状情報による対話物体認識実験を行った．つまり全部で 150 回の実験を行った．実験結果を図 5.5 に示す．対話として，参加者が伝えた形状表現が 1 回，あるいはそれ以上で正しく伝えられた場合を「成功」とする．1 回で成功しない場合は，追加の対話を行う．追加の対話を行っても，それ以上の形状表現を伝えることができない場合を「失敗」とみなした．1 つの物体を認識するのに必要な対話の回数は平均 1.56 回であった．150 回試行で 133 回成功（追加の対話により成功したものも含む）し，形状に対する対話だけで成功率 89%であった．



(a) 5個全て丸いもの



(b) 四角いもの4個+丸い物1個



(c) 丸い物3個+四角いもの2個

図 5.4 : 実験 1 で用いた 3 シーン

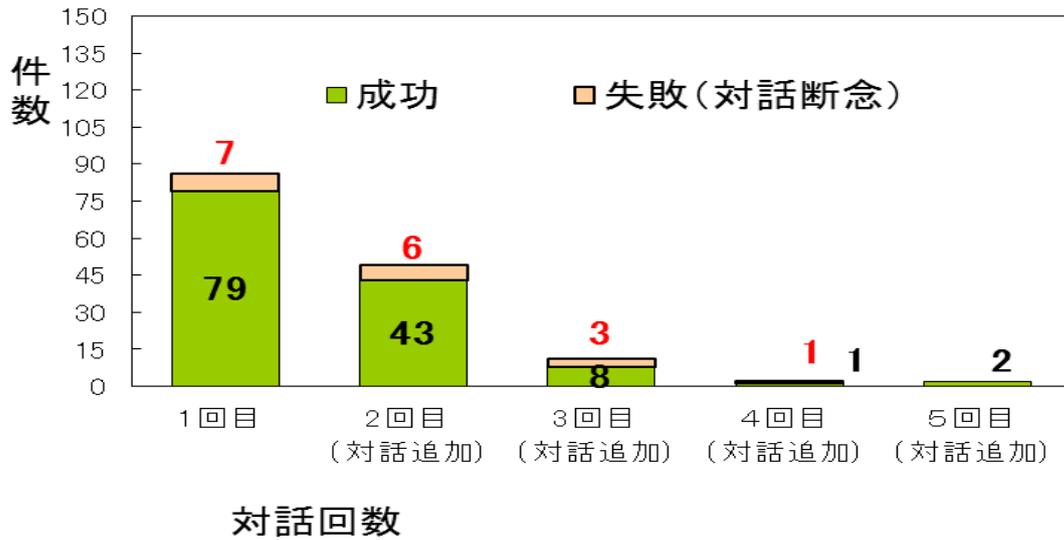


図 5.5 : 対話の回数(1)

(2) 参加者によって異なるシーンを用いた場合

20種類の物体からランダムに5個の物体を選択し、それをテーブルに並べた10パターンのシーン用意した。実験に用いたパターンを図5.6に示す。各実験参加者はそのうちの1つのシーンについて、(1)と同様に、シーン中のすべての物体を順に対象物として実験を行った。実験結果を図5.7に示す。したがって、全部で50回の実験を行った。1つの物体を認識する対話の回数は平均1.78回であった。50回試行して43回成功し、形状に対する対話だけで成功率86%であった。

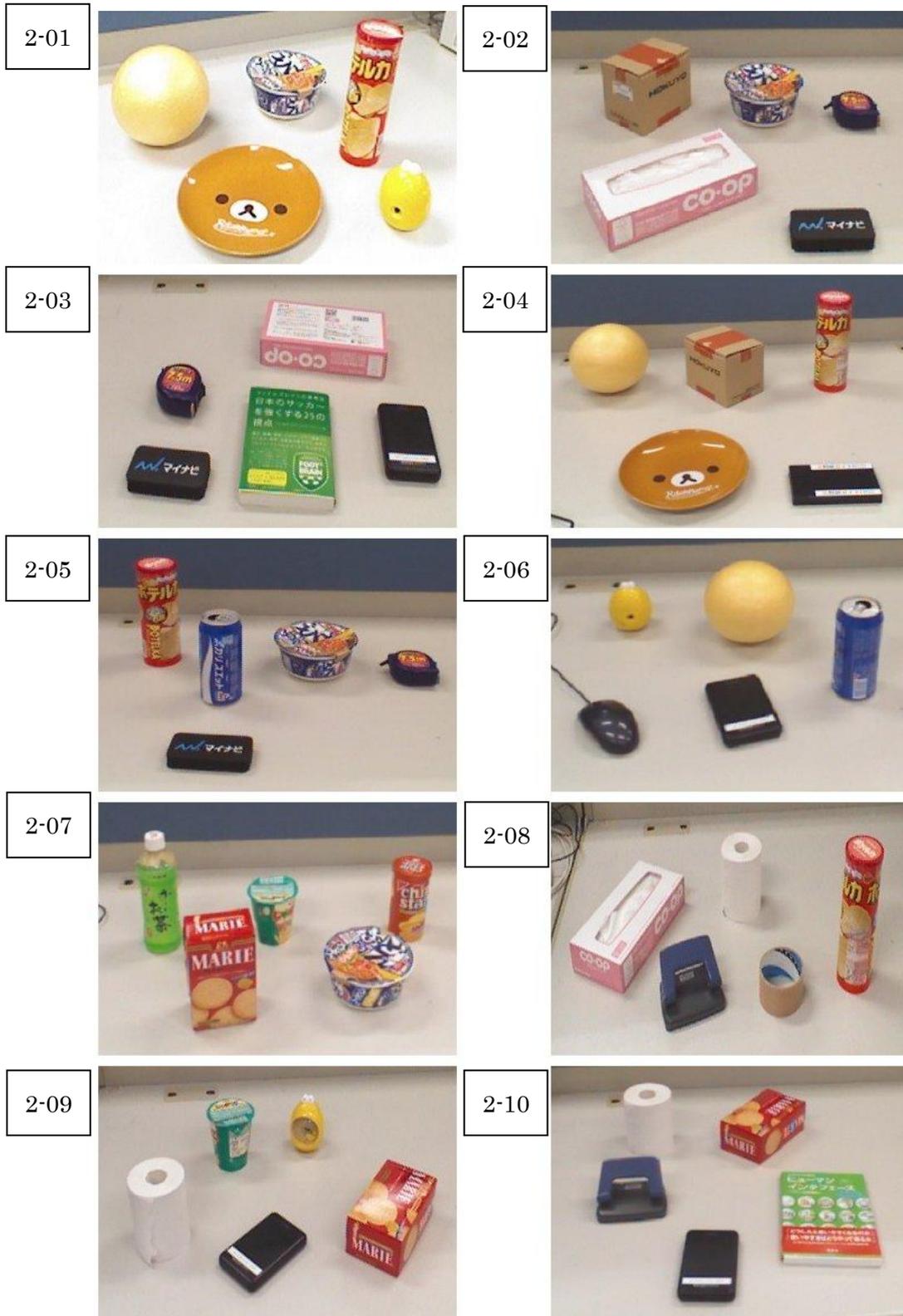


図 5.6 : 実験 2 で用いた 10 シーン

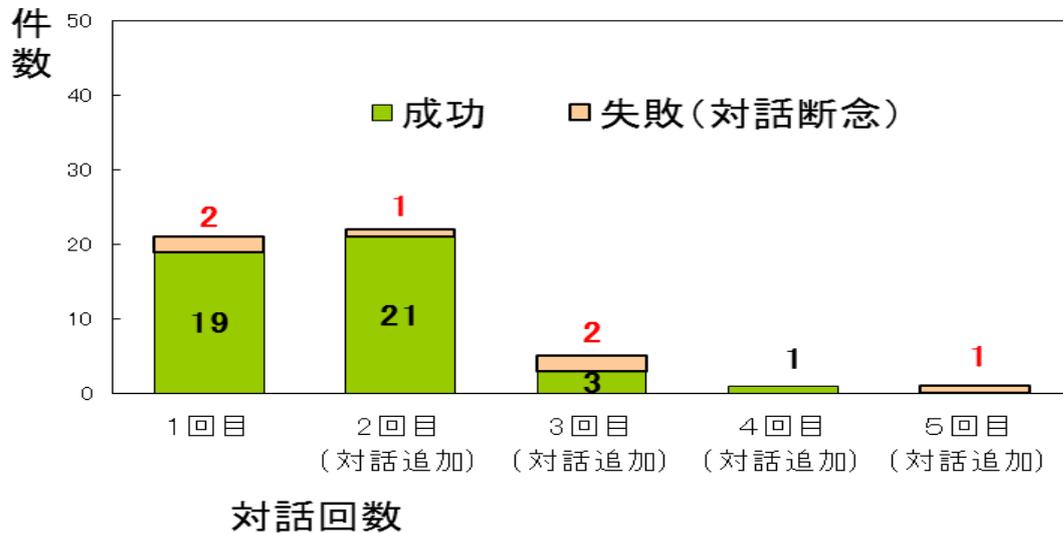


図 5.7 : 対話の回数(2)

まだ対象物体の数など限定された実験ではあるが、この2つの結果から、システムが人間の形状表現を適切にとらえ、形状表現だけでかなりの認識ができることは示されたと考えられる。

考察

実験(1)では、5個の物体を置いた3つのシーンを用意した。各実験参加者は3つのシーンの合計15個の物体のすべてを対象物として形状情報による対話物体認識実験を行った。したがって、全部で150回の実験を行った。1つの物体を認識するのに必要な対話の回数は平均1.56回であった。150回試行で133回成功（追加の対話により成功したものも含む）し、形状に対する対話だけで成功率89%であった。

実験(2)では20種類の物体からランダムに5個の物体を選択し、それをテーブルに並べた10パターンのシーンを用意した。各実験参加者はそのうちの1つのシーンについて、(1)と同様に、シーン中のすべての物体を順に対象物として実験を行った。したがって、全部で50回の実験を行った。1つの物体を認識する対話の回数は平均1.78回であった。50回試行して43回成功し、形状に対す

る対話だけで成功率 86%であった。

4.3 節の対話実験(1)(2) の対話回数で、(1) は 1 回が多く、(2) では 2 回が多かった。

実験(1)(2) で、P C (ロボット) は、1, 2 回の対話で、対象物を検出する。

対話物体認識が失敗した場合、ロボットが検出した物体を P C 画面で順番に見せ、依頼主に確認してもらう。

5.3 物体認識での失敗の例・考察

3 章でも述べたように、1 方向のみの 3 次元ポイントデータしか取得していないため、必ずしも正しい形状判定はできない。形状情報を全て発しても対象物を理解できない場合は、見つけた物体の順に、依頼人に問いただしていく。以下に失敗したときの例を示す。

例として、以下の図 5.8(a), 5.9(a), 5.10(a) の 3 パターンから、物体の形状の判定をした結果、正しく形状を判定できなかった物がある。

[1] 図 5.8 は、「角の丸い直方体」である「黒板消し」(図 5.8(b)) を誤って、円柱と見なして

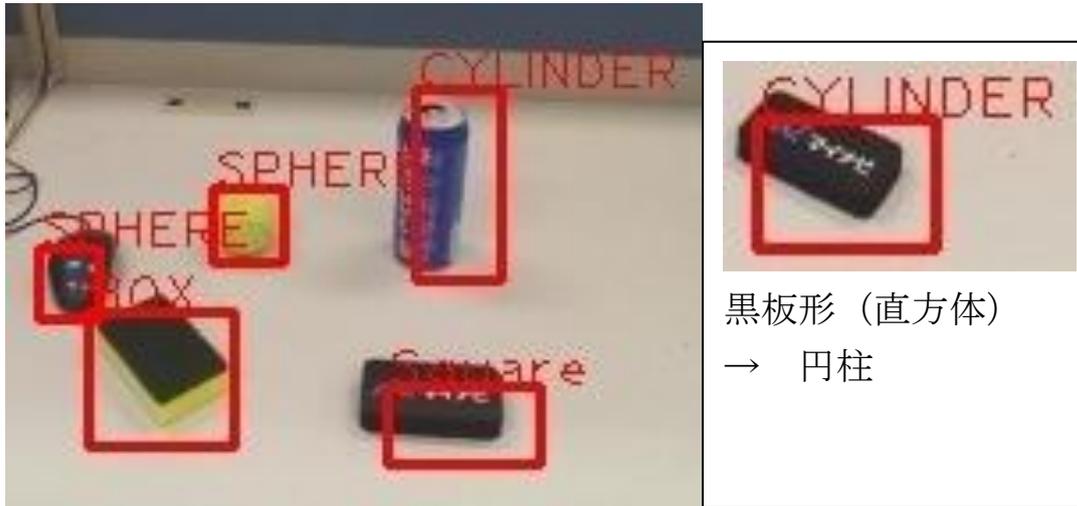
しまった。他にも「細長い直方体」である「サランラップ」も円柱と見なしてしまったり、3 章で作成した「3 次元ポイントデータによる 3 次元形状認識」は、あまり細長く角に丸みを帯びている直方体を円柱と誤認識してしまう。

[2] 図 5.9 では、「円柱」である「ガムテープ」と「缶詰」を「球」と誤認識している。

円柱の物体をいくつか判定したところ、あまりに低い円柱は、球と見なされてしまう。

[3] 図 5.10 では、「円柱」である「どんぶり」・「巻尺」の形状が判定できない場合がある。

今後の課題として、以上の問題を解決するために、kinect を 2 台用いて、別々の方角から物体をセグメントして、より正確な 3 次元データを取得し形状判定を行うつもりである。また、3 次元データを伸縮して、円柱・直方体を検出率を上げることを検討している。



(a) 形状判定結果

(b) 間違った形状判定

図 5.8 : 失敗パターン 1

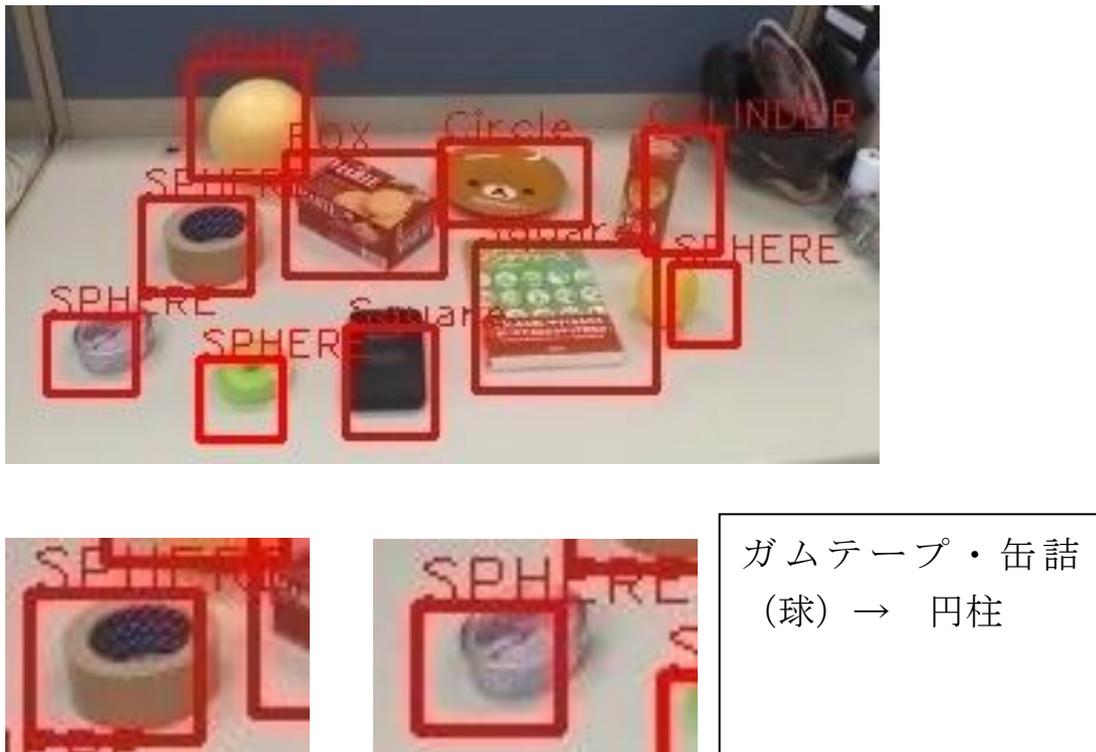


図 5.9 : 失敗パターン 2



図 5.10 : 失敗パターン 3

5.4 対話認識失敗の対処方法

対話として、参加者が伝えた形状表現が 1 回、あるいはそれ以上で正しく伝えられた場合を「成功」とする。1 回で成功しない場合は、追加の対話を行う。そして、追加の対話を行っても、それ以上の形状表現を伝えることができない場合を「失敗」とみなした。

対話物体認識が失敗した場合、ロボットが検出した物体を PC 画面で順番に見せ、依頼主に確認してもらう。

今後、認識が失敗したときの対処法として、以下の図 5.11 のように、PC (ロボット) は、依頼主から対象物の写真と、特徴情報を受け取り、セットにしてオントロジーに基づいた DB (データベース) に履歴として記録する。

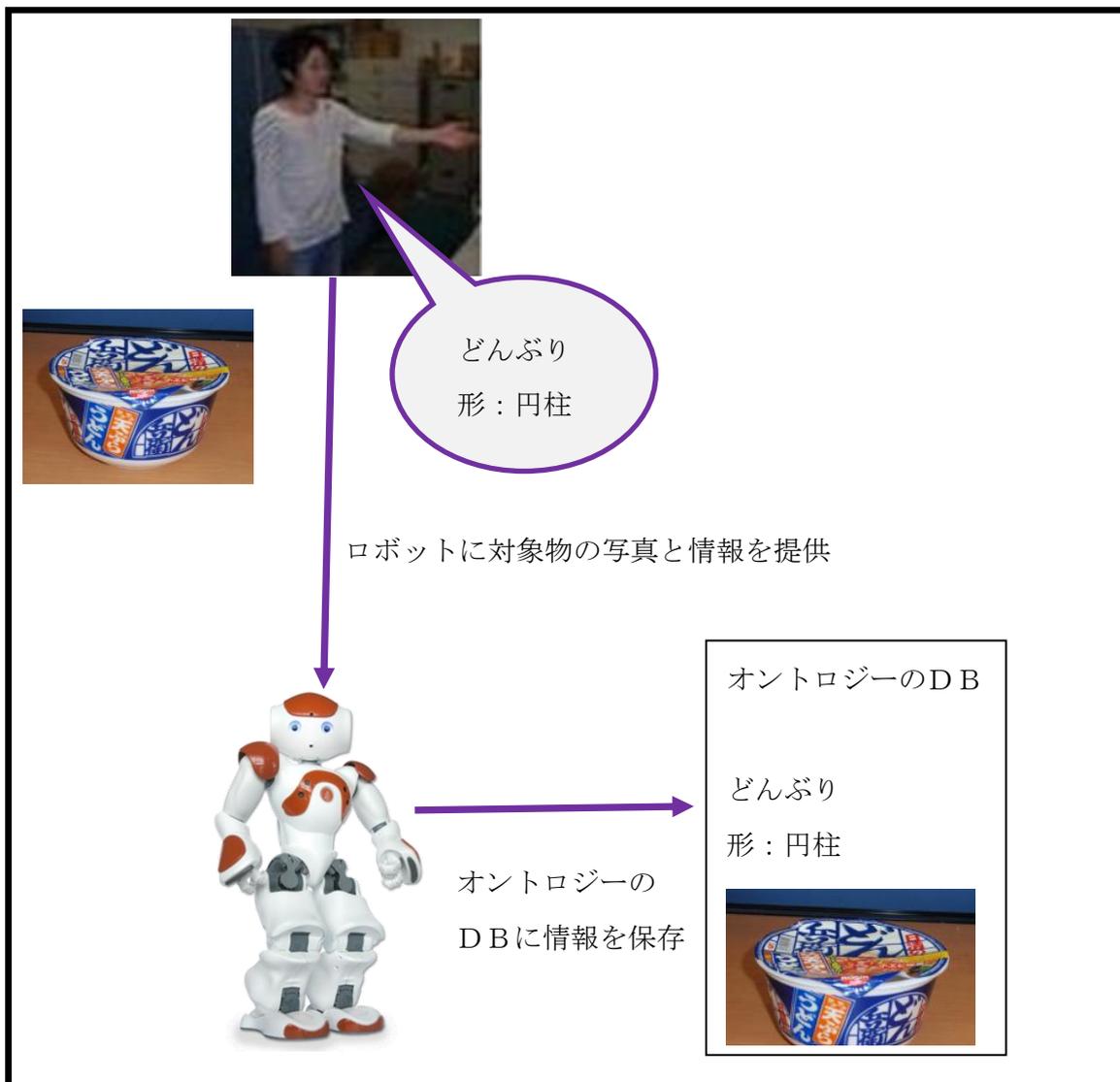


図 5.11 : PC (ロボット) の学習方法

6章 まとめ

頼まれたものを取ってきてくれるようなサービスロボットを実現するためには、物体認識の能力が必要である。しかし、確実に物体を認識するのは難しい。そこで、認識できない場合には、ユーザから対象物の属性についての情報を対話を通じて与えてもらい、物体の形状の認識を行う対話物体認識を検討している。これまでの検討で、人間の属性の表現が多様であることが分かった。多様な表現を整理して、それに対応したシステムを実現するために、対話物体認識用のオントロジーを構築することを提案した。今回は、表現の多様性が顕著だった形状属性について実際にオントロジーを構築して、対話による形状認識の実験システムを実現した。そして、実験によりオントロジーに基づく物体認識法の有効性を確認した。

しかし、対話により対象物が認識できたと言っても、このようなシステムでは、オントロジーに登録された表現がされる限り、動作するというのは、画像処理の失敗がなければ当然とも言える。

今後の課題として

- 3次元ポイントデータを取得する際、2つ以上の kinect を用いて別々の方向から撮影した対象物の3次元データを繋ぎ、より精密なデータから形状をプログラムに正しく判定させる。
- 物体をセグメントする際、なるべく形の重要な部分（角）を欠けることなく検出できるようにプログラムを改善する。
- 第3章に形状特徴の検出で、『文字（アルファベットなど）』による認識をできるように、対象物の領域の細線化方法を検討する。
- 今後、認識が失敗したときの対処法として、以下の図 5.11 のように、PC（ロボット）は、依頼主から対象物の写真と、特徴情報を受け取り、セットにしてオントロジーに基づいたデータを履歴として記録する。
- ロボットにオントロジーで人間の生活環境を理解せ、家庭内で人を支援するための知能処理[46]を検討する。

今回のシステムは事前に行った実験で使われた表現に対応できるように構成されたものだが、実際に人間が使用する表現はさらに多様であると思われる。それに対応するためにはオントロジーをさらに拡張する必要がある。さらに興味深いのは、使用しながら人間との対話の結果に基づきオントロジーを自動的に拡張していくことであるが、これについても検討を進めていきたい。そして、形状だけでなく他の属性についても検討を進め、対話物体認識を完成させていきたい。

謝辞

研究ならびに生活面においてご指導を賜りました久野義徳教授，小林貴訓助教，本当にありがとうございました。この場を借りてお礼申し上げます。そして，久野研究室の学生の皆様，物体認識プロジェクトとして共に開発を携わった皆様，並びに私を暖かく見守って頂いた両親をはじめとする周囲のすべての皆様に深く感謝致します。

発表文献

■査読付き国際会議

“Understanding the meaning of shape description for interactive object recognition”

著者：S. Mori, Y. Kobayashi, and Y. Kuno.

論文誌：Lecture Notes in Computer Science, vol.6838,
pp.350–356, 2012.

学会：ICIC2011, 2011-8-11

査読：有 本研究の一部は科学研究費補助金(23300065)による.

■ジャーナル論文

“物体形状オントロジーに基づく指示形状物体の認識”

著者：森智史, 福田悠人, 小林貴訓, 久野義徳, 加地大介

論文誌：ビジュアルコンピューティング論文特集号 2013,
画像電子学会誌第 42 巻第 4 号, pp 477–485, 2013-7-30.

査読：有 本研究の一部は科学研究費補助金(23300065)による.

■口頭発表

[1] “オントロジーに基づく対話物体認識”

著者：森 智史, 小林 貴訓, 久野 義徳

論文誌：電子情報通信学会総合大会講演論文集
2009 年_情報・システム(2), pp.183, 2009–03–04.

査読：無 本研究の一部は科学研究費補助金(19300055)による.

[2] “状況に応じて形状表現の意味を理解する対話物体認識システム”

著者：森 智史, 小林 貴訓, 久野 義徳

学会：HAI シンポジウム 2010, 2010-12-14.

査読：無 本研究の一部は科学研究費補助金(19300055)による.

[3] “オントロジーに基づく形状認識”

著者：森智史, 福田悠人, 小林貴訓, 久野義徳, 加地大介

学会：第 16 回画像の認識・理解シンポジウム MIRU2013, 2013-8-1.

査読：有 本研究の一部は科学研究費補助金(23300065)による.

参考文献

- [1] 土肥健純, “医療におけるロボティクスの現状と将来,” 日本ロボット学会誌, vol.18, no.1, pp. 29–32, 2000.
- [2] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman (Eds.) , “Toward category–level object recognition,” Lecture Notes in Computer Science, LNCS4170, Springer, 2006.
- [3] 坂田克俊, “人間の用いる表現に基づく対話を援用した物体認識に関する研究,” 平成21年度埼玉大学修士論文, 2009.
- [4] D.G. Lowe, “Object recognition from local scale–invariant features,” In Proc.ICCV'99, pp.1150–1157, 1999.
- [5] M.Kirby and L.Sirovich, “Application of the karhunen–loeve produre for the characterization of human faces,” IEEE Trans. Pattern Anal. Mach. Intell., vol.12, no.1, pp.103–108, Jan.1990.
- [6] M. Turk and A.Pentland, “Eigenfaces for recognition ,” J.Cognitive Neuroscience, vol.3, pp.71–86, Jan. 1991.
- [7] 村瀬 洋, “パラメトリック固有空間法による画像認識,” 電子情報通信学会論文誌, vol.85, no.4, pp.252–258, April 2002.
- [8] 出口光一郎, 岡谷貴之, “固有空間法はなぜうまく働くか,” 情報処理学会 CVIM 研報, vol.2001, no.66, pp.1–8, July 2001.
- [9] 高木雅成, 藤木弘旦, “SIFT 特徴量を用いた交通通路標識認識,” 電子情報通信学会論文誌 (C), vol.129, no.5, pp.824–831, 2009.

- [10] D. Marr and E.C. Hildreth, “Theory of edge detection,” Proc. Royal Society of London, vol.B-207, pp.187–217, 1980
- [11] L.Pap and J.J.Zou, “Sub-pixel edge detection for photogrammetry using laplace difference of Gaussian and 4th order ENO interpolation,” ICIP10, pp.2841–2844, 2010
- [12] Duygulu, P., Barnard, K., Freitas, N. d. and Forsyth, D., “Object Recognition as Machine Translation: Learning a Lexicons for a Fixed Image Vocabulary”, European Conference on Computer Vision, pp. 97–112, 2002.
- [13] Burl, M. and Perona, P., “Recognition of planar object classes”, IEEE Computer Vision and Pattern Recognition, pp.223–230, 1996.
- [14] Fergus, R., Perona, P. and Zisserman, A., “Object Class Recognition by Unsupervised Scale-Invariant Learning”, Computer Vision and Pattern Recognition, pp.264–271, 2003.
- [15] Csurka, G., Bray, C., Dance, C. and Fan, L., “Visual categorization with bags of keypoints”, ECCV Workshop on Statistical Learning in Computer Vision, pp.1–22, 2004.
- [16] Fei-Fei, L. and Perona, P., “A Bayesian Hierarchical Model for Learning Natural Scene Categories”, IEEE Computer Vision and Pattern Recognition, pp.524–531, 2005.
- [17] Torralba, A., Murphy, K. and Freeman, W., “Using the Forest to See the Trees: A Graphical Model Relating Features, Objects and Scenes, Advances in Neural Information Processing Systems”, 2003.

- [18] Hoiem, D., Efros, A.A. and Hebert, M., “Putting Objects in Perspective”, IEEE Computer Vision and Pattern Recognition, pp.2137–2144, 2006.
- [19] R.Kurazume, K.Nishino, Z.Zhang, and K.Ikeuchi,, “Simutaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute,” Proc. Fifth ACCV, pp99–106,2002.
- [20] M.D. Elstrom and P.W. Smith, “Stereo-based registration of muti-sensor imagery for enhanced visualization of remote environments,” Proc. 1999 Int. Conf. on Robotics Automation, pp.1948–1953, 1999.
- [21] J.Bohm and S.Becker, “Automatic maker-free registration of terrestrial laser scans using reflectance features,” 8th Conf. on Optical 3D Measurement Techniques, 2007.
- [22] D.G. Low, “Distinctive image geatures from scale-invariant keypoints,” Int. J. Comput. Vis., vol.60, no.2, pp.91–110, 2004.
- [23] K.Mikolajczyk and C.Schmid, “Scale & affine invariant interest point detectors,” Int.J.Comput. Vis., vol.60, pp.63–86, Oct. 2004.
- [24] 猪股 亮, 寺林 賢司, 梅田 和昇, ギ- ゴダン, “柔軟な誤対応除去と距離濃淡画像を用いた高精度な 2D/3D レジストレーション手法,” 画像・理解シンポジウム (MIRU2011), pp.1508–1515, 2011.
- [25] T.Winograd, “Understanding Natural Language, Academic Press,” 1972.
- [26] D. Roy, B. Scheile, and A. Pentland, “Learning audio–visual associations using mutual information,” In Proc. ICCV’99.Workshop on Integrating Speech and Image Understanding, pp.147–163, 1999.

- [27] 滝澤正夫, 榎原 靖, 白井良明, 島田伸敬, 三浦 純, “サービスロボットのための対話システム,” システム制御情報学会論文誌, vol.16, No.4, pp.174–182, 2003.
- [28] H. Fukuda, S. Mori, Y. Kobayashi and Y. Kuno, “Object Recognition for Service Robots through Verbal Interaction about Multiple Attribute Information,” 8th International Symposium, ISVC 2012, Advances in Visual Computing Lecture Notes in Computer Science , vol.7431, pp 620–631, 2012.
- [29] 溝口 理一郎, “オントロジ-工学(知の科学),” オ-ム社, 2005.
- [30] I.H.Suh, G.H.Lim, W.Hwang, H.Suh, J.H.Choi, “Ontology-based Multi-layered Robot Knowledge Framework (OMRKF) for Robot Intelligence,” Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.429–436 , 2007.
- [31] 小林 昭太郎, 樋川 暁, 山口 高平, “対話と動作の連携を目指したオントロジ-に基づく知能ロボット,” 人工知能学会全国大会 (第26回), 3K1-R-11-7, 2012.
- [32] H.Holzapfel, D.Neubig, A.Waibel, “A dialogue approach to learning object descriptions and semantic categories,” Robotics and Autonomous Systems vol.56 pp.1004–1013, 2008.
- [33] N.E.Maillot, M.Thonnat, “Ontology based complex object recognition,” Image and Vision Computing 26, pp.102–113, 2008.
- [34] S.Dasiopoulou, V.Mezaris, I.Kompatsiaris, V.-K.Papastathis, M.G.Strintzis, “Knowledge-assisted semantic video object detection,” IEEE Transactions on Circuits and Systems for Video Technology vol.15 no.10, pp.1210–1224, 2005.
- [35] 利島 保, 生和秀敏, “心理学のための実験マニュアル-入門から基礎・発展へ,” 北大路書房, 1993.

- [36] 井口 征士, 小林 重順, 長田 典子, 猪田 克美, 田辺 新一, 中村 敏枝,
“感性情報処理 (ヒューマンコミュニケーション工学シリーズ),” pp.13–21, 1994.
- [37] J. J. ギブソン, “生態学的視覚論–ヒトの知覚世界を探る,”サイエンス社, pp.254–257,
1985.
- [38] L. Cao, Y. Kobayashi, and Y. Kuno, “Object spatial recognition for service robots :
Where is the Fronts? ,” In Proc. International Conference on Mechatronics and
Automation 2011(ICMA 2011), pp.875–880, 2011.
- [39] Galton, “Qualitative Spatial Change,” Oxford University Press, pp.195–201,
2000.
- [40] C. Harris and M. Stephens, “A combined corner and edge detector,” Proc. Alvey
Vision Conference, pp.147–151, 1988.
- [41] G. Flitton and T. Breckon, “Object recognition usin 3D SIFT in complex CT
volumes,” Proc. British Machine Vision Conference, pp.11.1–12, 2010.
- [42] 溝口 理一郎, “オントロジ-構築入門,” オ-ム社, 2006.
- [43] Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada, “Object
Recognition Supported by User Interaction for Service Robots,” Proc. of the 5th
Asian Conf. on Computer Vision, vol. 2, pp. 719–724, Melbourne, Australia, Jan.
2002.
- [44] マイクロソフト社 Kinect for Windows <http://kinectforwindows.org/>.

- [45] K.Sato, I.Kitahara, Y.Ohta, “3D Image with Motion Parallax by Camera Matrix Stereo,” Proceedings of IEEE International Conference on Multimedia Computing and Systems (ICMCS’96), pp.349–357, 1996
- [46] 山口 亨, 福里 友介, 飯島 隆弘, 芝浦工業大学 水川 真, “生活空間における RT オントロジ-を用いたサービス支援システム,” 第 10 回(社)計測自動制御学会システムインテグレーション部門講演会(SICE SI2009), CD-ROM, pp.1205–1206, 2009.
- [47] 林 雄介, 山崎龍太郎, 池田 満, 溝口理一郎, “オントロジ-アウェアな学習コンテンツ設計環境,” 情報処理学会論文誌, vol. 44, No.1, pp. 195–208, 2003.
- [48] 平田謙次, 池田満, 溝口理一郎, “コンピテンシ-・オントロジ-に基づく人的資源設計支援,” 教育システム情報学会誌, vol. 18, no. 34, pp.340–351, 2001.
- [49] 平田謙次, 池田満, 溝口理一郎, “コンピテンシ-・オントロジ-に基づく人的資源設計支援,” 教育システム情報学会誌, vol. 18, no. 34, pp.340–351, 2001.
- [50] T.Kanade, P.W.Rander, P.J.Narayanan, “Virtualized Reality: Constructing Virtual Worlds from Real Scenes,” IEEE Multimedia, vol.4, no.1, 1997
- [51] 金来, 林雄介, 池田満, 溝口理一郎, 太田衛, 高岡良行, “訓練システムSmartTrainer 構築用オ-サリングツ-ル,” 教育システム情報学会誌, vol.16, no.3, pp.139–148, 1999.
- [52] H.Saito, T.Kanade, “Shape Reconstruction in Projective Grid Space from Large Number of Images,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’99), pp.49–54, 1999
- [53] S.Moezzi, A.Katkere, D.Y.Kuramura, R.Jain, “An Emerging Medium: Interactive Three-Dimensional Digital Video,” Proceedings of IEEE International Conference on Multimedia Computing and Systems (ICMCS’96), pp.358–361, 1996

- [54] R.T.Collins, "A Space-Sweep Approach to True Multi-Image Matching,"
Proceedings of IEEE Conference on Computer Vision and Pattern Recognition
(CVPR '96), pp.358-363, 1996
- [55] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection,"
In CVPR 2005, pages 886-893, 2005.
- [56] 溝口理一郎, "オントロジ-研究の基礎と応用,"人工知能学科論文誌, vol. 14,
pp.977-988, 1999.
- [57] Mizoguchi, R. and Bourdeau, J, "Using Ontological Engineering to Overcome AI-ED
Problems," International Journal of Artificial Intelligence in Education, vol.11, no.2,
pp.107-121, 2000.
- [58] 森田武史, 山口高平, "日本語概念を対象にした領域オントロジ-構築支援環境
DODDLE の機能拡張," 電子情報通信学会論文誌, 信学技報 (知能ソフトウェア工
学研究会), KBSE2006-1, 2006.
- [59] N.Izumi, T. Yamaguchi, "Integration of Heterogeneous Repositories Based on
Ontologies for EC Applications Development'," International Journal of
Electronic Commerce Research and Applications, vol.1, no.1, pp.77-91, 2002.
- [60] 和泉憲明・吉岡 健・山口高平, "ビジネス指向サ-ビス関連オントロジ-,"
人工知能学会誌 19(2), 151-158, 2004-03-01.
- [61] Ikeda, M., Seta, K., and Mizoguchi, R, "Task Ontology Makes It Easier To Use Authoring
Tools," Proc. of IJCAI'97, Nagoya, Japan, pp.342-347, 1997.

- [62] 榎原 靖 , 滝澤 正夫 , 白井 良明, 島田 伸敬, “対話を用いた物体認識のための照明変化への適応,” 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 vol.J87-D-II(2), 629-638, 2004-02-01.
- [63] Kouji, K., Kitamura, Y., Ikeda, M., and Mizoguchi, R, “Development of an Environment for Building Ontologies which is Based on a Fundamental Concideration of ”Relationship” and ”Role”, ” The Sixth Pacific Knowledge Acquisition Work Shop(PKAW2000), pp.205-221, Sydney, Australia, December 11-13, 2000.
- [64] 來村徳信, 笠井俊信, 吉川真理子, 高橋賢, 古崎晃司, 溝口理一郎, “機能オントロジに基づく機能的知識の体系的記述とその機能構造設計における利用,” 人工知能学会論文誌, vol.17(1), pp. 73-84, 2002
- [65] 稲葉晶子, ThepchaiSupnithi, 池田満, 溝口理一郎, 豊田順一, “学習理論に基づく協調学習グループの構成のための学習目的,”電子情報通信学会論文誌, vol.J83-D-I, no.6, pp.569-579, 2000
- [66] Clowes,M.B, “On Seeing things, Artifical Intelligence,” vol.2, no.1, pp.79-116, 1971.
- [67] Tenenbaum, “J.M.and Barrow, H.G.:Experiments in Interpretation Guided Segmentation,” Artifical Intelligence, vol.8, pp.241-274, 1997.
- [68] Draper, B.,Collins, R., Brolio, J.,Hanson,A.and Riseman, E, “The Schema System, International Journal of Computer Visiom,” vol.3, no.2, pp.209-250, 1989.

- [69] 海谷 治彦 , 清水 悠太郎 , 安井 浩貴 , 海尻 賢二 , 林 晋平 , 佐伯 元司, “要求獲得のためのオントロジ-を Web マイニングにより拡充する手法の提案と評価,” 情報処理学会論文誌 vol.53(2), 495–509, 2012–02–15.
- [70] Batlle,J.,Cansals, A.,Freixent, J. and Marti, J. , “A review on strategies for recognizing natural objects in color images of outdoor scenes,” Image and Vision Computing, vol.18, no.6–7, pp515–530, 2000.
- [71] 秋間 雄太 , 川久保 秀敏 , 柳井 啓司, “Folksonomy を用いた画像特徴とタグ共起に基づく画像オントロジ-の自動構築(物体認識, 文字・文書, 医用画像, <特集>画像の認識・理解論文), ” 電子情報通信学会論文誌. D, 情報・システム vol.J94–D(8), 1248–1259, 2011–08–01.
- [72] Brooks, R.A , “ Model–Based Three–Dimensional Interpretations of Two–Dimensional Image,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.5, no.2, pp.140–150, 1983.
- [73] Basri, R. , “Recognition by Prototypes,” International Journal of Computer Vision, vol.10, no.2, pp.147–167, 1996.
- [74] Strat, T.M. and Fischler, M.A. , “Context–Based Vision: Recongnizing Objects Using Information from Both 2–D and 3–D Imagery,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.13, no.10, pp.1050–1065, 1991.
- [75] 濱田 敏弘, 清水 昭伸, 長谷川 純一, 鳥脇 純一郎, “画像処理エキスパートシステム IMPRESS における少数の設計標本からの手順構成に関する検討,” 電子情報通信学会論文誌 vol.J84–D–II(9), pp.2029–2039, 2001–09–01 .

- [76] 白川真澄, 中山浩太郎, 荒牧英治, 原隆浩, 西尾章治郎, “Wikipedia と Web の情報を組み合わせたオントロジ-構築の試み,” 電子情報通信学会和文論文誌, vol.J94-D, no. 3, pp.525-539, 2011.
- [77] 和泉 憲明, 山口 高平, “オントロジ-に基づくソフトウェアエージェントのパターン指向開発,” 電子情報通信学会和文論文誌, vol.J84-D1, no.8, pp.1181-1190., 2001.
- [78] 孫寧, 田原透, 阿曾弘具, 木村正行, “方向線素特徴量を用いた高精度文字認識,” 電子情報通信学会論文誌, vol. J74-D- II, no.3, pp.330-339, 1991.
- [79] F. Mohd-Zaid, K.W. Baruer, and M.A. Friend, “Face recognition via ensemble SIFT matching of uncorrelated hyperspectral bands and spectral PCTs,” Int. J. Tomography & Statistics, vol.19, no.1, pp.1-13, 2012.
- [80] ロボット NAO <http://www.revast.co.jp/service/humanoid/type03.html>
- [81] 鶴岡信治, 栗田昌徳, 原田智夫, 木村文隆, 三宅康二, “加重方向指数ヒストグラム法による手書き漢字・ひらがな認識,” 電子情報通信学会和文論文誌, vol.J70-D, no.7, pp.1390-1397, 1987.
- [82] 森 稔, 倉掛 正治, 杉村 利明, 塩 昭夫, 鈴木 章, “背景・文字の形状特徴と動的修正識別関数を用いた映像中テロップ文字認識,” 電子情報通信学会和文論文誌, vol.J83-D2, no.7, pp.1658-1666, 2000.
- [83] 玉川奨, 桜井慎弥, 手島拓也, 森田武史, 和泉憲明, 山口高平, “日本語 Wikipedia からの大規模オントロジ-学習,” 人工知能学会論文誌, vol.25, no.5, pp.623-636, 2010.

- [84] Shotaro Kobayashi, Susumu Tamagawa, Takeshi Morita, Takahira Yamaguchi, “Intelligent Humanoid Robot with Japanese Wikipedia Ontology and Robot Action Ontology,” Proceedings of the 6th ACM/IEEE International Conference on Human–Robot Interaction, pp.417–424, 2010.
- [85] Benjamin Johnston, Fangkai Yang, Rogan Mendoza, Xiaoping Chen, Mary–Anne Williams, “Ontology Based Object Categorization for Robots,” Proceedings of the 7th International Conference, Practical Aspects of Knowledge Management, pp.219–231 , 2008.
- [86] Il Hong Suh, Gi Hyun Lim, Wonil Hwang, Hyongwon Suh, Jung–Hwa Choi, “Ontology–based Multi–layered Robot Knowledge Framework (OMRKF) for Robot Intelligence,” Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.429–436, 2007.
- [87] T.Yamaguchi and T.Morita, “Building up a Large Ontology from Wikipedia Japan with Infobox and Category Tree”, 3rd Interdisciplinary Ontology Conference (InterOntology10), pp.121–134, 2010.
- [88] Shotaro KOBAYASHI, Susumu TAMAGAWA, Takeshi MORITA, and Takahira YAMAGUCHI, “Human Robot Interaction Based on Wikipedia Ontology and Robot Action Ontology,” International Joint Conference on Knowledge–Based Software Engineering 2010, Kaunas University of Technology, pp.119–132, 2010.
- [89] Tatsuya ISHIKAWA, Keido KOBAYASHI, Masao OKABE, and Takahira YAMAGUCHI, “Support for Externalization of Intelligence Skill Using Ontology and Rule–based System,” International Joint Conference on Knowledge–Based Software Engineering 2010, Kaunas University of Technology, pp.145–159, 2010.

- [90] Susumu TAMAGAWA, Shinya SAKURAI, Takuya TEJIMA, Takeshi MORITA, Noriaki IZUMI, and Takahira YAMAGUCHI, “Learning a Large Scale of Ontology from Japanese Wikipedia,” 2010 IEEE/WIC/ACM International Conference on Web Intelligence, pp.279–286, 2010.
- [91] Chie IIJIMA, Makito, KIMURA, and Takahira YAMAGUCHI, “Implementing an Image Search System with Integrating Social Tags and DBpedia”, Knowledge-Based and Intelligent Information and Engineering Systems, PART3, KES2010, Springer, LNAI6278, pp.264–272 , 2010.
- [92] Hideki Mima, Susumu Ota, Koji Nagatsuna, “Ontology-based query processing for understanding intentions of indirect speech acts in natural-language question answering”, International Journal of Computer Applications in Technology 2009 – vol.35, no.2/3/4, pp.271–280, 2009.
- [93] 大崎紘一, 神代充, 宗澤良臣, 梶原康博共著, “画像認識システム学,” 共立出版, 2005.
- [94] Dipankar Das, Yoshinori Kobayashi, and Yoshinori Kuno, “Object Detection and Localization in Clutter Range Images Using Edge Features,” ISVC2009, pp.172–183 .
- [95] 河副文夫, “細線画像のフル-エンシ函数近似とその地図道路情報自動函数化への応用に関する研究,” 筑波大学大学院博士課程 システム情報工学研究科修士論文.
- [96] 岡部, “楕円成長法による円形物体の自動検出,” 電子情報通信学会和文論文誌, vol.J85–D–2, no.12, pp.1823–1831, 2002.
- [97] Dipankar Das, Al Mansur, Yoshinori Kobayashi, and Yoshinori Kuno, “An Integrated Method for Multiple Object Detection and Localization”, Proc. ISCV08, vol.2, pp.133–144, 2008.

- [98] 大語彙連続音声認識エンジン Julius <http://julius.sourceforge.jp/>
- [99] Grauman, K. and Darrell, T, “The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features”, Proc. Of IEEE International Conference on Computer Vision, pp.1458–1465, 2005.
- [100] 高木雅成, 藤木弘旦, “SIFT 特徴量を用いた交通通路標識認識,” 電子情報通信学会和文論文誌, vol.129, no.5, pp.824–831, 2009.
- [101] 山崎, 野沢, 植田, 楨, 森, 岡田, 松本, 稲葉, “日用品データベースを利用する家事支援ロボットによる 思い出し・片付け支援,” 第27回日本ロボット学会学術講演会, vol.2E2–05, 2009.
- [102] 笠井 俊信, 山口 晴久, 永野 和男, 溝口理一郎, “オントロジ-理論に基づく情報教育目標の体系的記述,” 電子情報通信学会 (D) vol.J88–D– I, no.1, pp.3–15, January 2005.
- [103] 瀬田和久, 林 雄介, 池田 満, 角所収, 溝口理一郎, “学習コンテンツのオントロジ-と設計パターン,” 電子情報通信学会技術研究報告, OFS99–17, AI99–29, pp.1–8.
- [104] 來村徳信, 笠井俊信, 吉川真理子, 高橋賢, 古崎晃司, 溝口理一郎, “オントロジ-に基づく機能的知識の体系的記述とその機能構造設計支援における利用,” 人工知能学会論文誌, vol.17, no.1, pp.73–84, 2002.
- [105] 結城 啓, 富樫 敦, “オントロジ-に基づく知的コミュニティシステム,” 情報処理学会研究報告. マルチメディア通信と分散処理研究会報告 vol.58, pp.47–50, 2005 .

- [106] Lazebnik, S., Schmid, C. and Ponce, J. , “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” Proc. Of IEEE Computer Vision and Pattern Recognition, pp.2169–2178, 2006.
- [107] PALRO (パルロ) <http://palro.jp/feature>
- [108] ASIMO (アシモ) <http://www.honda.co.jp/ASIMO/>
- [109] 東芝買い物ロボット http://www.toshiba.co.jp/rdc/rd/detail_j/0904_01.htm
- [110] Pinar Duygulu, Muhammet Baştan, David Forsyth ,“Translating Images to Words for Recognizing Objects in Large Image and Video Collections,” Toward Category-Level Object Recognition Lecture Notes in Computer Science Volume 4170, 2006, pp 258-276.
- [111] Y. Kuno, K. Sakata, and Y. Kobayashi, “Object recognition in service robot: Conducting verbal interaction on color and spatial relationship,” In Proc. IEEE 12th ICCV Workshops, pp.2025–2031, 2009.
- [112] 日本語係り受け解析器 CaboCha <http://code.google.com/p/cabocha/>
- [113] オープンソース 形態素解析エンジン MeCab
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>