

## RGB-D カメラを用いたサービスロボット のための対話物体認識

福田 悠人<sup>†1</sup> 小林 貴訓<sup>†1,†2</sup> 久野 義徳<sup>†1</sup>

サービスロボットは実環境などの複雑な環境でも物体を認識することが求められる。しかし、実環境において自動的な物体認識により間違えることなく物体を認識することは困難な問題である。このような背景のもと我々は対話を用いてユーザから対象物体の情報を獲得、利用することで、確実に物体を認識するシステムを提案している。本稿ではそのような物体認識システムにおいて、RGB-D カメラにより得られる物体の色、形、材質といった属性情報や位置関係を用いて認識を行う手法を提案している。複数の情報を用いることで、ロボットがより効率的に物体を認識することが期待できる。

### Interactive Object Recognition for Service Robot using an RGB-D Camera

HISATO FUKUDA,<sup>†1</sup> YOSHINORI KOBAYASHI<sup>†1,†2</sup>  
and YOSHINORI KUNO<sup>†1</sup>

Service robots need to be able to recognize objects located in complex environments. However, it is difficult to recognize objects autonomously without any mistakes in natural conditions. Thus, we have proposed an object recognition system using information about target objects acquired from the user through simple interaction. In this paper, we propose an interactive object recognition system using multiple attribute information such as color, shape and material, and positional information among objects by using an RGB-D camera. Experimental results show that the robot can recognize objects by using multiple information obtained through interaction with the user.

### 1. はじめに

高齢者や障害者の増加、それに伴う介護者不足という問題から、高齢者、障害者の支援を目的としたサービスロボットに注目が寄せられている。我々もまた、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行っている。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。

物体認識技術はコンピュータビジョンの研究の主なテーマの一つであり、その初期の頃から盛んに研究が行われている。特に近年になり、SIFT<sup>1)</sup>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<sup>2)</sup>の開発などにより、物体認識技術は急速に発展している。しかし、一般的に物体認識と呼ばれる画像からの物体の検出、カテゴリ認識を行った場合の認識率は70%程でも良いとされる段階である。今後の発展により、認識率はさらに向上するだろうが、完全な認識を実現することは困難であると考えられる。加えて実環境においては、様々なアウトライアにより完全な認識はさらに困難になると考えられる。誤った認識はサービスロボットがサービスを行うとき頼まれたものが分からず、何もできない状況を作ってしまう。そのため我々は、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<sup>3)</sup>。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、それらを手掛かりに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。

このように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 Winograd ら<sup>4)</sup>の研究以来、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。例えば「青い箱」や「赤いボール」などは「青いもの」、「赤いもの」として表すことを想定している。ビジョンを用いた研究としては、見せられた物体の名前、カテゴリを学習しようという研究は多い。Roy ら<sup>5)</sup>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。実際にサービスロボットの活動するような状況では、人間は「赤いもの」などのように単純な表現を用いるが、多くの物

<sup>†1</sup> 埼玉大学 理工学研究科

Graduate School of Science and Engineering, Saitama University

<sup>†2</sup> 科学技術振興機構 さきがけ

Japan Science and Technology Agency (JST), PRESTO

体は複雑な物体である場合がほとんどである。例えばお菓子のパッケージのデザインは複雑に構成されているが、人間はその構成を細かく伝えることはない。色に限って言えば、人間は主要な色を一つ選択し伝える。このように人間の表現は単純であるが物体は複雑であり、円滑なインタラクションを行うためには、人間の用いる表現の理解とそれらに基づいた画像処理技術が必要である。

これらの知見に基づき、これまでに我々は人の用いる物体の表現に関する調査を行い、色、位置関係の情報を用いて物体を認識する技術の開発を行ってきた<sup>3),6),7)</sup>。本稿では、物体の形に基づき物体を認識する手法、物体の材質に基づき物体を認識する手法を既存の物体認識システムに追加し、ロボットシステム上で統合を行った。物体の認識の手掛かりに複数の表現を用いることで、ロボットがより効率的に物体を認識することが期待できる。

## 2. 人間が用いる物体の表現

人間は物体の名称が分からない場合、どのような表現を用いて、取ってほしい物体を伝えるかについて検討した。これについては文献 3), 6)–9) で発表済みであるが、論理解のために調査結果をまとめておく。物体の表現には主に 2 つの方法があると考えられる。一つは色や形などの対象物体の属性について説明を行う方法であり、もう一つは対象物体と他の物体との空間関係について説明を行う方法である。インタラクティブなビジョンシステムでは、これらの人の用いる表現の理解が、円滑なインタラクションを実現するために重要であるといえる。

### 2.1 物体の属性に関する表現

物体の名称が分からない場合、人はどのように多くの物体の中から対象の物体を表現するか調査を行った<sup>3)</sup>。調査から人は物体を表現するために、2 つの表現方法を用いることが分かった。我々は表現の分析のために、これらを視覚的表現と知識的表現に分類した。前者は「赤い」や「丸い」などの視覚的な特徴を用いた表現であり、後者は「食べ物」や「飲み物」などの知識に基づく表現である。知識的表現に基づく物体の認識は本稿で注目するビジョンシステムでの解決が困難であるため、以降では視覚的表現に関する分析を行った。

視覚的表現に関する先の調査 3) の結果を図 1 に示す。視覚的表現では、人は物体の色、形、模様、大きさ、付属物、材質を用いることが分かり、物体の色についての表現を最も多く用いることが確認できた(図 1)。また、このとき人間は多色の物体に対しても 1 色で指示することが多く、その色は最も割合の多い色、または下地の色であることが分かった。ロボットのビジョンシステムにおいてもこれらの色を認識することが重要であると考えられる。

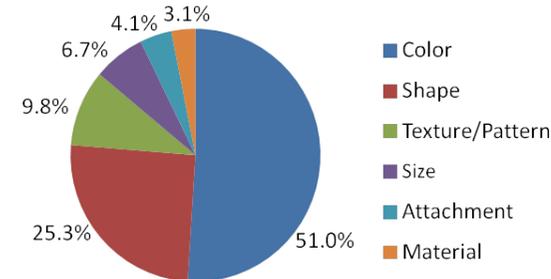


図 1 視覚的表現の割合  
Fig. 1 Classification of vision-based descriptions

2 番目に多く用いられたものは形についての表現であり、人間は主に 3 つの方法により形を記述することが分かった。1 つは物体の全体的な形を用いる表現であり、ボールなどの物体に対して「球形のもの」という表現を行うことが観られた。2 つ目は、ある視点からの外輪の形状に関する表現である。円柱の形をした物体は、側面からみた形状は四角い形状だが、上面からみた形状は円形であるため「丸い」という表現をされることがあった。3 つ目は物体の一部分に対しての形状を伝える方法であった。物体の形状については文献 8), 9) において、より詳細な調査、考察を行っている。

調査では材質を用いた表現は少なかったが、物体の認識を行うために材質を知ることは重要だと考えられる。なぜなら、材質は名称との関係性が深いからである。例えば、人間は知識により「紙コップ」は紙、「グラス」はガラスで構成されていると分かる。このように物体の名称から推定できる材質を、物体を認識する手掛かりとして利用できると考えられる。

### 2.2 物体の空間関係に関する表現

物体同士の空間の関係に対しても、調査、考察を行った<sup>3),6),7)</sup>。Levinson<sup>10)</sup> は人間が位置関係を示すのに用いる reference system を intrinsic, relative, absolute の 3 つに分類している。intrinsic は「彼の前の本」など名前を言えばその前後関係の特定が可能な物を参照する方法であり、relative は「彼からみて箱の左にある本」など、第 3 者や視点を基準として物体の位置関係を参照する方法である。absolute は東西南北など絶対的な指標により位置関係を表す方法である。ロボットに物を取ってもらうシーンでは、relative system が中心となると考えられる。つまり、「(ユーザからみて、もしくはロボットからみて) 赤い箱



図 2 自動物体認識の例  
Fig. 2 Autonomous object recognition result

の左」というような表現である。

### 3. 統合物体認識システム

本節では自動物体認識と対話物体認識を統合した物体認識システムについて述べる。システムは始めに自動物体認識により物体の認識を試みる。自動認識に失敗した場合、対話を援用した物体認識を開始する。システムは対話を用いて、対象物体の色、形、材質、位置関係をユーザから獲得し、ユーザの指定する物体を認識する。

#### 3.1 自動物体認識

自動物体認識は、Mansur<sup>11)</sup>らの手法を用いる。この手法は、対象物体の種類や問題設定に応じて、特定物体の認識(特定のメーカーのジュースの缶かの認識)、カテゴリーの認識(一般的なジュース缶としての認識)のどちらも行うことができる(図2)。

#### 3.2 物体の色による物体の同定

前節で述べた通り、人は色で物体を表現する場合、その物体で最も割合の多い色、下地の色に着目することが分かった。そのため、我々はこのような色領域を検出する画像処理法を開発している<sup>3)</sup>。以下に物体の色を検出するアルゴリズムについて説明を行う。検出アルゴリズムでは始めに1)物体領域の各画素の色情報に基づき領域分割を行い、各色の面積を求める。このとき、面積が顕著に大きくなる色が1つの場合、その色を目的の色としてその後の処理を行わない。2)1)において複数の色がある程度大きな面積を占める場合、1)で得られた各色領域の輪郭を求め、変曲点を特徴点として、凸包を計算し、3)各色領域の凸包内の面積が最大となる色を目的の色とする。この処理の流れを図3に示す。

#### 3.3 物体の形による物体の同定

前節での調査から、物体の認識に形を用いる手法では、全体的な形状の理解、特定の視点

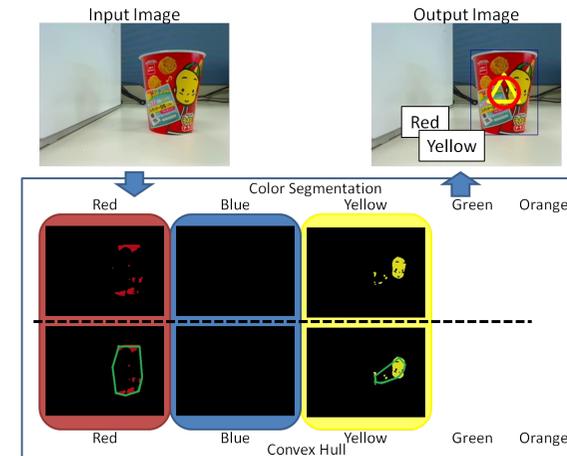


図 3 物体色の検出例  
Fig. 3 Identification of human description color

からの外輪形状の理解、物体の一部分に対する形状の理解が求められることが分かった。本稿では、使用される割合の高かった全体的な形状、外輪形状に着目し、これらを用いた手法を開発、実装した。

全体的な形状の理解のために、本手法では始めに物体の3次元のポリウムについて解析を行う。距離画像から獲得した物体の3次元点を用いて物体のX軸、Y軸、Z軸上での広がりから物体を、3次元の物体、2次元の物体、1次元の物体に分類する。物体の3次元点が全ての軸の方向に広がりを持つ場合、物体を3次元の物体に分類し、2方向にのみ広がりを持つ場合は2次元の物体、1方向にのみ広がりを持つ場合は1次元の物体と分類する。物体が3次元の物体の場合、球や円柱や箱型などのプリミティブな形状に対してモデル当てはめを行うことで、物体の全体的な形状を認識する。2次元の物体は「平たい」、1次元の物体は「細長い」と全体的な形状を分類する。

外輪形状を用いた表現では「丸い」や「四角い」といったような、他の物体との形状の比較を用いた表現が多かった。本システムでもこのような形状の度合いを評価する必要がある。本手法では各物体の座標系のXY平面、YZ平面、ZX平面において、各平面において近傍の点を平面に射影し、射影された点群の凸包領域の丸さを評価した。丸さの指標には一般的に用いられる(1)式を用いた。ここでLは領域内での点と点の最大長であり、Sは領

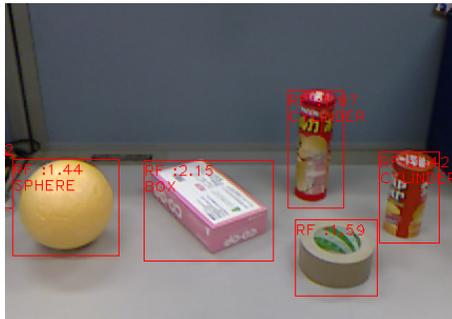


図 4 物体の形の認識  
Fig.4 Shape recognition result

表 1 認識に用いた Visual words  
Table 1 Visual words used in the recognition

Feature	Dim.	Feature num/image	Cluster num
Color	27	4315	150
Texture	128	1598	200
Shape	120	5504	100
Reflection	72	571	100

域の面積である．この評価値を，3 平面に対し計算し，各平面の面積で重みづけし，丸さの評価値とした．

図 4 に形状の認識結果の例を示す．

$$roundness\_factor = \frac{\pi L^2}{4S} \quad (1)$$

### 3.4 物体の材質による物体の同定

本システムでは，プラスチック，布，紙や金属などの物体の材質情報を対象物体の同定に用いる．材質の推定手法としては，校正された光源，カメラシステムを用いて反射光を観測し，そのパターンから材質を推定する方法<sup>12)</sup>なども考えられるが，物体は一つの材質で構成されているなどの制約や，追加のデバイスが必要となる．そのため，本稿ではカラー画像から物体の材質を推定する方法<sup>13)</sup>を実装した．以下に文献 13) で開発した手法について説明する．実装した手法では，材質の認識に有効に働くと考えられる物体の色，テクスチャ，形状，反射という情報を用いて材質を推定する．物体の画像から色，テクスチャ，形状，反射を表すような特徴を抽出し，言語処理における bag-of-words のアナロジーである bag-of-visual-words<sup>14)</sup>により画像表現を行う．そしてそれらの画像パターンを文書クラスタリングの手法である Latent Dirichlet Allocation<sup>15)</sup>により統計的な学習を行い，学習により得られるパターンから物体の材質らしさの推定を行う．この手法は，画像中の局所的な特徴に基づき推定を行うため，複数の材質により構成される物体に対しても有効に働くと考えられる．図 5 に材質らしさの推定の流れを示す．

以下に本手法において用いた特徴に対してまとめる．また，表 1 に抽出した各特徴の特徴

量の次元数，画像一枚あたりの特徴数，作成した Visual Words の数を示す．

色特徴 色特徴には， $3 \times 3$  の窓を用いてその窓の RGB 値を抽出し並べた，27 次元のベクトルを特徴として用いた．

テクスチャ特徴 テクスチャ特徴には一般物体認識においても多く用いられている Scale-Invariant Feature Transform(SIFT)<sup>1)</sup>を用いた．ただし SIFT による特徴点抽出は行わず，特徴点の抽出はグリッド状に密にサンプリングし，SIFT の特徴量記述のみを用いて，特徴量を抽出する．

形状特徴 形状には，Histograms of Oriented Gradients(HOG)<sup>16)</sup>を用いている．HOG は，SIFT 同様に局所領域における輝度の勾配方向をヒストグラム化した特徴量である．HOG は一定領域に対する特徴量を記述するため，大まかな物体の形状を表現することが可能であり，人検出などに用いられている<sup>16)</sup>．

反射特徴 反射を表すような特徴としては，HOG を特定領域に適用することで反射特徴とした．ハイライトなどがある場合，特定の方向に輝度変化が大きく観測されるため，画像をグリッド上に分割し，各グリッド内で輝度の分散が最大となる方向の短径領域に対して HOG を適用した．

このように画像表現には複数の特徴の組み合わせを用いている．しかし，画像表現に最適な特徴の組み合わせは材質毎に異なり，新たに特徴を追加しても，認識率は逆に下がることが考えられる．布という材質を例にとると，布という材質は様々な色であることがあり，色という特徴を用いても上手く識別できないということが想像できる．そのため本稿ではロボットシステムに組み込むにあたり，認識する材質毎に用いる特徴の組み合わせを選択し，材質らしさを推定することで，対象物体の認識に利用した．

### 3.5 物体の位置関係による物体の同定

物体の空間関係による同定では，空間関係の表現で最も多く用いられると考えられる rel-

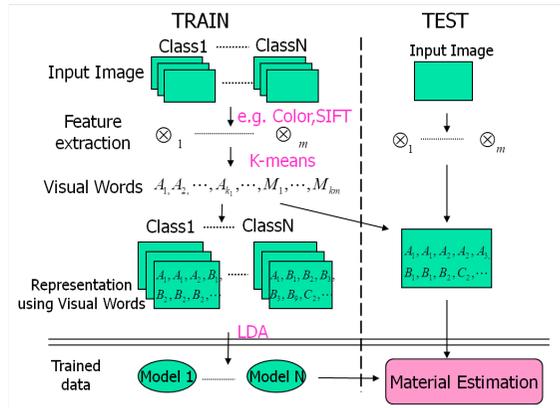


図 5 材質らしさの推定  
Fig. 5 Flow of the estimation of material likelihood

ative system を実装した。relative system は 3 つの要素：referent(対象物体), relatum(参照物体), origin(視点) から構成される。本システムでは最も用いられると考えられる聞き手(ロボット)を視点として仮定した。relative system では、参照物体を用いて物体の位置関係を表す。ユーザ主導の対話では、ユーザはロボットが参照物体を認識できているか分からないため、効率的に対話を進めることができないと考えられる。そのため本システムでは、ロボットが主導的に対話を進める。ロボットはユーザに認識できた物体の名前を告げ、その物体との位置関係を尋ねる。このとき、人間は 3 次元空間での位置関係で表現を行うが、本システムでは、距離画像センサーにより各物体の 3 次元位置を獲得している。そのため、物体の 3 次元位置をセンサーとロボットの位置関係を用いて、ロボットの視点を原点とした座標系に変換し、ロボットから見た各物体の位置関係を求める。ロボットはこれら物体の位置情報とユーザからの情報を用いて対象の物体を認識する。

### 3.6 統合システム

統合システムでは前述した 4 つの物体認識手法により物体を認識する。本システムへの入力には、カラー画像、距離画像を用いている。物体の検出には距離画像を用い、距離画像の各画素と周囲の画素の距離の差に基づいて領域を分割し、物体を検出する。本システムが想定する物体認識のシナリオは以下である。

ロボットはリビングルームのような特定の部屋に在りうる物体のリストを保持している。

ロボットは始めに自動物体認識によりリストにある物体の検出を行い、検出できた物体を既知物体リストに記憶する。ユーザから「コップ取って」のように特定物体を取ってくるように依頼を受けると、既知物体リストにその物体があるか検索を行う。ユーザの指定した物体が既知物体にある場合、ロボットは指さしをしながら「これですか?」のようにユーザに対して確認を行う。確認により誤認識だと分かった場合や、既知物体リストに依頼された物体がない場合は、ロボットは対話を用いた物体の認識を開始する。

対話を援用した物体の認識では、少ない対話数で速やかに物体を認識することが望ましい。そのため、各シーンにおいてどのような質問を行うかということが重要となる。そこで本システムがユーザに対してどう質問を行うかは、シーン中の物体の属性のバリエーション数、既知物体の数に基づき決定するようにした。例えば、赤い物体しかない状況(シーン中での色のバリエーション数が 1) では「それは何色ですか」と物体色に関する質問をしても意味がなく、このような状況では別の質問をしたほうが効率的である。本システムでは、物体の色、形、材質、位置関係に基づき物体の同定を行っており、物体の属性情報については、色の情報は、形、材質と比べて、信頼性の高い検出が可能であり、検出できる種類も多い。そのため、本システムはシーン中の物体の色のバリエーション数と、位置関係を表現するために用いる既知物体の数に基づいて対話を進める。

シーン中での物体の色のバリエーション数が一定数以上の場合(現在の実装では色のバリエーション数  $\geq 3$  の場合)、システムは物体の色についてユーザに尋ねる。シーン中での既知物体の数が一定数以上の場合(現在の実装では既知物体の数  $\geq 3$  の場合)、システムは物体の位置関係についての対話を行う。そのときシステムは「これとこれは在ることが分かります」などとユーザに伝える。多くの物体が既知物体である場合、全ての既知物体をユーザに伝えたとユーザが覚えきれなく、円滑に対話を進めることが困難なため、ユーザに伝えるのは既知物体の中で最も位置の離れた 2 つとしている。色のバリエーション数、既知物体の数がどちらも一定数以上の場合には色についての対話を優先して行う。物体の形、材質に関する質問は、色のバリエーション数が 1 の場合、既知物体の数が 0 の場合、または物体色、物体の位置関係への質問がすでに一度行われている場合に行う。形、材質についての質問では、形についての質問を優先して行う。質問により、選択肢を絞れない場合などは、まだ選択していない物体の中から、ランダムに選択し、逐次指さしと発話による確認を行うことで、ユーザの指定する物体の認識を行う。

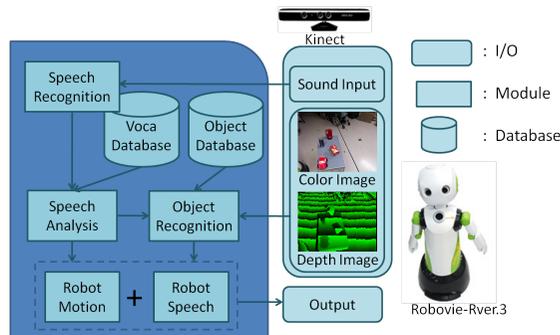
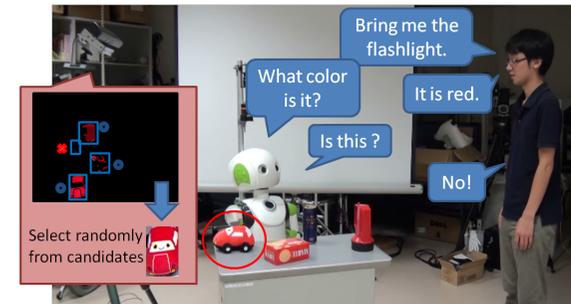


図 6 システムの構成  
Fig.6 System configuration

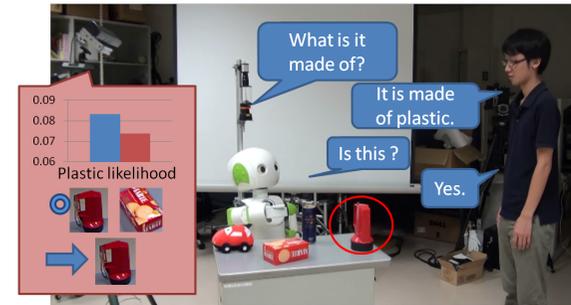
#### 4. ロボットシステム

開発した統合物体認識システムをロボットに搭載し、ロボットシステムを構築した。ロボットプラットフォームとしては Robovie-Rver.3<sup>17)</sup> を用い、カラー画像、距離画像の取得には、Kinect<sup>18)</sup> を用いた。音声認識エンジンには、Julius<sup>19)</sup> を用いた。ロボットシステムの構成について図 6 に示す。ロボットは指の曲げ伸ばしが可能になっており、指さしの動作を行うことができる。本ロボットシステムでは Kinect により物体の三次元位置を獲得し、ロボットと Kinect の位置関係を用いて物体の方向に指さしを行うことで、確認の動作を行う。

ロボットが対話によりユーザの指定する物体を同定する様子を図 7 に示す。図 7(a) のシーンではユーザはロボットに「懐中電灯をとって」と指示を行っている。すると、ロボットはユーザの指示した懐中電灯を認識できなかったため、ユーザに物体の色について尋ねると、ユーザは「赤です」と答えている。しかし、シーン中にロボットが赤の物体と認識したのは複数存在したため、ロボットはそれらの中からランダムに選択し、「これですか?」とシーンの左にある赤い車のおもちゃを指さしている。このロボットの振る舞いに対しユーザは「違います」と答え、ロボットの示した物体は対象の物体でないことをロボットに伝えている。図 7(b) のシーンでは、ロボットは対象物体の物体の材質について尋ねている。これに対し、ユーザは「プラスチックです」と答えている。この情報からロボットは候補として残っていた二つの物体のプラスチックらしさを推定、比較している。その後、ロボットはそれら二つからよりプラスチックらしいシーンの右の懐中電灯を指さし、再び確認を行うことで、ユー



(a) Object detection using color



(b) Object detection using material

図 7 対話を援用した物体の認識  
Fig. 7 Operation scene with robot

ザの指定する物体を同定している。

#### 5. ま と め

本稿では、自動物体認識技術と対話を援用した認識手法(対話物体認識)を組み合わせた統合物体認識システムを提案し、ロボットに実装した。ロボットは自動物体認識に失敗すると、対話により色、形、属性といった対象物体の属性や位置関係についてユーザに尋ねる。ロボットは得られた情報に基づき対象物体を認識する。このときロボットは物体の色のパリエーション数、既知物体の数により対話の内容を変化させ、効率的な物体の同定を試みる。今後、実際に人との対話を用いて本ロボットシステムの評価実験を行う予定である。また、

人間が指示に用いる言葉の概念をロボットが理解し、正確に対象物体を認識するシステムを目指していく。現在「赤いもの」というような指示を受け、物体の同定を行っているが「赤いもの」は状況により変化する。例えば赤い物体が少ない環境の中では、赤みがあった物体を「赤いもの」というかもしれないが、赤い物体が多い状況では同じ物体を「赤いもの」と言うとは限らない。このように人が言葉で指し示す概念は状況により異なる。そのため、ことば - 状況 - 概念の関係を体系化し、人間の言葉を正確に理解するロボットを目指す。

謝辞 本研究の一部は科学研究費補助金 (19300055,23300065) による。

### 参 考 文 献

- 1) D.G. Lowe. Object recognition from scale-invariant keypoints. In *ICCV*, pp. 1150–1157, 1999.
- 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) 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.
- 4) T. Winograd. *Understanding Natural Language*. Academic press, 1972.
- 5) D. Roy, B. Scheile, and A. Pentland. Learning audio-visual associations using mutual information. Proc. ICCV '99. Workshop on Integrating Speech and Image Understanding, pp. 147–163, 1999.
- 6) L. Cao, Y. Kobayashi, and Y. Kuno. Spatial resolution for robot to detect objects. In *Proc. International Conference on Intelligent Robots and Systems (IROS 2010)*, pp. 4548–4553, 2010.
- 7) 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.
- 8) 森智史, 小林貴訓, 久野義徳. 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. HAI シンポジウム, 2010.
- 9) S. Mori, Y. Kobayashi, and Y. Kuno. Understanding the meaning of shape description. In *Proc. International Conference on Intelligent Computing (ICIC2011)*, pp. 350–356, 2011.
- 10) S.C. Levinson. Frames of reference and molyneux's question: Crosslinguistic evidence. *Language and Space*, pp. 109–170, 1999.
- 11) A. Mansur and Y. Kuno. Specific and class object recognition for service robots through autonomous and interactive methods. *IEICE Trans. Information and Systems*, Vol. E91-D, No.5, pp. 1793–1803, 2008.
- 12) M. Mannan, H. Fukuda, L. Cao, Y. Kobayashi, and Y. Kuno. 3d free-form object material identification by surface reflection analysis with a time-of-flight range sensor. In *12th IAPR Conference on Machine Vision Applications (MVA2011)*, 2011.
- 13) 福田悠人, 小林貴訓, 久野義徳. 対話物体認識のための材質情報の獲得. 第 17 回 画像センシングシンポジウム (SSII2011), 2011.
- 14) G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *Proc. ECCV Workshops on Statistical Learning in Computer Vision*, pp. 1–22, 2004.
- 15) D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, Vol.3, pp. 993–1022, 2003.
- 16) N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pp. 886–893, 2005.
- 17) Intelligent robotics and communication laboratories.  
[http://www.vstone.co.jp/products/robovie\\_r3/index-en.html](http://www.vstone.co.jp/products/robovie_r3/index-en.html).
- 18) Kinect for windows.  
<http://kinectforwindows.org/>.
- 19) Open-source large vocabulary csr engine julius.  
<http://julius.sourceforge.jp/index-en.html>.