

サービスロボットのためのインタラクティブビジョン

島田 伸 敬[†] 三 浦 純^{††} 白 井 良 明[†]

ユーザの代わりに指示されたものを取ってくるロボット（サービスロボット）に必要なユーザとの対話機能、とくに対話的に画像認識を行う技術について報告する．たとえば冷蔵庫内の飲み物を取ってくるといったタスクの場合、庫内にある様々な物体が互いに隠蔽しあうように置かれていることが多く、また庫内外の照明条件の変化によって色の見え方が変化する．したがって自動認識型のアルゴリズムではうまく認識できずにタスクが遂行できないことがしばしば起こる．そこで、ロボットが取得した画像を加工してユーザに提示したり、音声による状況説明を行ったりすることで、ユーザからタスク遂行のための助言をもらいながら認識処理を進める．もし音声発話認識に一時的な間違いがあった場合でも画像認識の途中経過を考慮することで訂正したり、画像認識の失敗回復のために適切な助言をユーザから効率的に引き出したりすることができれば、実際の生活環境で使えるサービスロボットを実現できる．本稿ではサービスロボット実現のための対話的画像認識（インタラクティブビジョン）の実例を示してその現状と今後の課題を述べる．

Interactive Vision for Personal Service Robot

NOBUTAKA SHIMADA,[†] JUN MIURA^{††} and YOSHIAKI SHIRAI[†]

This paper reports user-interaction techniques, especially interactive image understanding, required of a personal service robot which brings user-specified objects. For example, consider a task of bringing some drinks from a refrigerator. Various objects often look overlapping together and their appearances of colors and textures widely change according to the illumination conditions inside and outside the refrigerator. In such situations full-automatic object recognition often fails and cannot recover its failures by itself. Our object recognition system explains the situation in some words and ask the user to make an advice via speech and help the system recognizing the object correctly. Speech recognition is also fragile but complementary visual and speech recognition can be done based on the user-interaction. Such an interactive interface makes the service robot useful in a real home environment. The paper shows an implementation of *interactive vision* and its effectiveness, and discuss the on-going and the future problems to study.

1. はじめに

現在、高齢化社会の到来により、サービスロボットの研究が数多く行われている^{1),3)}。サービスロボットの1つとして、ユーザの指定した物体を取ってくるロボットが考えられるが、このようなロボットは全自動では実現不可能である。ロボットが自動で行動できないときには何らかの外部情報を用いる必要がある。

指定された物体を取ってくる機能にとって必要な要素の1つとして、画像中から指定した物体を認識する

ことがあげられる。ユーザの指定するものとしては、飲み物、薬、本など様々であるが、これらは一般的にどこかに収納してあり、複雑背景下で物体認識を行う必要がある。画像を用いた物体認識については様々な研究が行われているが、複雑背景下では画像処理のみを用いた自動的な認識は困難な場合がある。このような場合には、画像情報以外の情報を用いて物体認識を行う必要がある。またサービスロボットには、認識結果を情報として与える能力と、認識に必要な情報を得る能力が必要であり、情報の流れが一方通行であるものはサービスロボットには適していない。

渡辺ら⁵⁾は、植物図鑑の花や果物を、図に添えてある説明文を用いて認識するシステムを提案している。これは情報が一方通行であり、サービスロボットには適していない。Ahlrichsら³⁾、Wachsmuthら²⁾はユーザの助言を認識に用いている。これらは指定さ

[†] 立命館大学情報理工学部知能情報学科

Department of Human and Computer Intelligence,
Ritsumeikan University

^{††} 大阪大学大学院工学研究科機械工学専攻

Department of Mechanical Engineering, Osaka University

れた物体の特徴をユーザの助言から得ることで、指定された物体の検出を可能にしている．しかし、これらの方法は認識が失敗したときに対処できない．また高橋ら⁴⁾は、りんごや本などの物体を人との対話やジャスチャを用いて認識する研究を行っている．しかし、対象とするシーンはそれほど複雑ではないので、物体の切り出しができるものとし、その中で物体を選択するために人との対話を用いている．

また、ロボットは物体認識をした結果をユーザに伝える必要がある．藤井ら¹⁴⁾、岩田ら¹⁵⁾は、認識結果に基づいて説明文を生成する研究を行っている．これらは情報の流れが一方通行であり、たとえば認識に失敗したときなどには対応していない．

稲邑ら⁶⁾は、ユーザとの言語対話によって移動ロボットの障害物回避のための行動決定モデルを自律的に獲得していくことを試みている．ロボットは周囲状況の認識結果に基づき実際に移動し、ユーザはロボットに搭載されたカメラ映像の変化を見てロボットの認識結果を理解して、必要であれば新たな教示を行う．しかし、物体を取ってくるといった、移動よりも複雑なタスクの場合には、認識結果をより詳細にユーザに説明することでユーザが教示をしやすく必要がある．

我々の研究グループでも画像認識に言語情報を利用することを考え、自動物体認識に失敗した場合には、ユーザとの音声による対話を通して認識の補助となるような情報を得て、それをもとに物体認識を試みることを考えた^{7),8),12),13)}．

たとえば冷蔵庫内の飲み物を取ってくるといったタスクの場合、庫内にある様々な物体が互いに隠蔽しあうように置かれていることが多く、また庫内外の照明条件の変化によって色の見え方が変化する．したがって自動認識型のアルゴリズムではうまく認識できずにタスクが遂行できないことがしばしば起こる．そこで、ロボットが取得した画像を加工してユーザに提示したり、音声による状況説明を行ったりすることで、ユーザからタスク遂行のための助言をもらいながら認識処理を進める．もし音声発話認識に一時的な間違いがあった場合でも画像認識の途中経過を考慮することで訂正したり、画像認識の失敗回復のために適切な助言をユーザから効率的に引き出したりすることができれば、実際の生活環境で使えるサービスロボットを実現できる．

本稿では、まず冷蔵庫から必要なものを取ってくるサービスロボットのためのシステムの概要について述べる．次に、冷蔵庫内の缶やペットボトルといった物体の見え方モデルと、認識時の照明条件に合わせて画

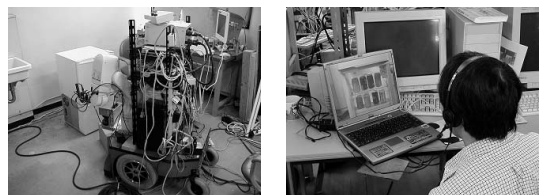
像照合する自動物体認識方法について述べる．次に、自動物体認識の失敗によってうまくタスクが遂行できなかった場合に、ユーザと音声対話を基づいて物体認識を行う方法について述べる．指定された物体が見つからなかった場合は、現在分かっていることを音声でユーザに説明し、物体のありかや認識するための手がかりについてアドバイスを求める．その際、音声認識に失敗することもしばしばあるため、ユーザの発話中に「誤認識された登録語」や「登録語と同義な未登録語」があるかどうかを、画像認識やタスクの進捗状況、発話の文脈を考慮して検出し、その誤認識後の意味を推定してユーザの発話を解釈する．さらに、ロボットが実際に冷蔵庫のドアを開閉する際の対話的な教示法についても述べる．ロボットの教示はスキルごとにカスタマイズされた知識が必要となるが、ロボットの側が何を教えてもらえばよいか、その要点をユーザに問い合わせるようにすれば、確実かつスムーズな教示が可能となる．

実際に構築したサービスロボットプラットフォームを使って実現されている対話的な画像認識（インタラクティブビジョン）の技術を中心に、対話的サービスロボットの実例を示すとともに、今後の課題を述べる．

2. 対話型サービスロボットシステムの概要

サービスロボットの概観を図 1 (a) に示す．ユーザは手元のディスプレイを見ながらロボットに音声で指示を出す（図 1 (b)）．ロボットはあらかじめ用意された地図情報とセンサ情報に基づいて目的地（冷蔵庫）の場所まで移動する．ロボットには冷蔵庫の扉を開ける作業スキルが教示によって与えられており、扉を開けて冷蔵庫の中をカメラで観察する．

ロボットは最初にあらかじめ登録された物体モデルを用いて指定された物体（缶・瓶・ペットボトルなど）を自動で認識し、結果をディスプレイと音声を用いてユーザに説明する．認識に失敗した場合やユーザが間違いを指摘した場合には、ユーザとの音声対話を通して得られる情報を用いて再度物体認識を試みる．この



(a) Our service robot

(b) User interface

図 1 サービスロボットシステム

Fig. 1 Our service robot system.

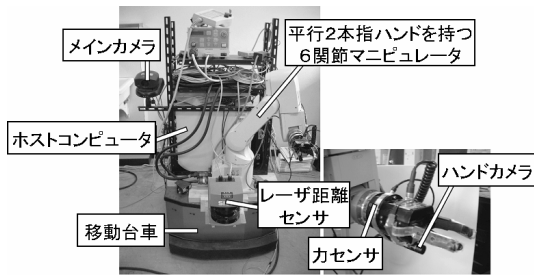


図 2 サービスロボットの実機
Fig. 2 Service robot.

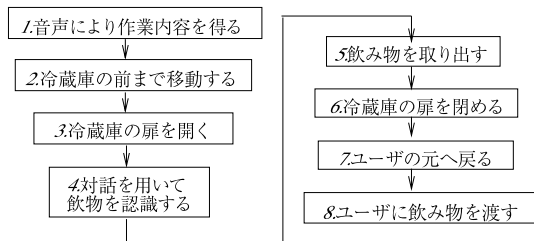


図 3 サービスロボットの作業手順
Fig. 3 Task flow of the service robot.

対話的物体認識は、照明変動を考慮した自動物体認識、物体認識のための対話生成と発話認識、対話結果に基づく物体認識からなっている。

図 2 に我々が試作したパーソナルサービスロボットを示す。ロボットの構成要素は以下のとおりである。(1) 移動台車 (2) 平行 2 本指ハンドを持つ 6 関節マニピュレータ (3) レーザ距離センサ (4) ハンドカメラ (5) カセンサ (6) ホストコンピュータ (7) メインカメラ。これらに加え、ユーザとの音声対話を行うためのマイク、スピーカ、ディスプレイがネットワーク越しに接続されている。

図 3 に全体の処理の概要を示す。このうち、1 で音声対話、2, 7 で移動、4 で音声対話と物体認識、3, 5, 6, 8 で冷蔵庫扉や物体のマニピュレーションが行われる。

3. 照明変動を考慮した自動物体認識

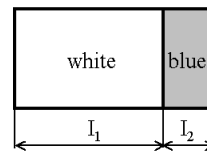
3.1 物体モデルの登録

ここでは、ビンや缶のような円柱状の物体を対象とする。物体がどの方向を向いても認識できるためには、多方向から見た画像を登録する必要がある。そこで効率化のために複数枚の画像ではなく、それらをモザイクした 1 枚の投影画像 (図 4) を作成する。その投影画像から物体の特徴として、物体の実際の大きさ、代表色 (最大の面積を持つ一様色領域の色)、お



(a) Original image (b) Piece image (c) Developed image
top: can bottom: square type PET bottle
図 4 投影画像の生成

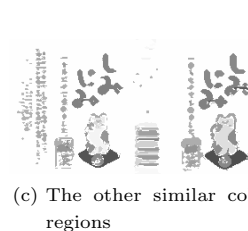
Fig. 4 Aquisition of developed image.



(a) Example of interval



(b) Segmented image



(c) The other similar color regions



(d) Features

In (d), top line: representative color, middle line: color of secondary feature 1, bottom line: color of secondary feature 2

図 5 モデル特徴の抽出

Fig. 5 Extraction of model features.



(a) 原画像 (b) 缶として認識された領域 (c) 認識結果 (緑領域: 2 次的特徴)

図 6 缶の認識

Fig. 6 Recognition of a can.

よび二次特徴 (代表色以外の一様色領域の色、位置、面積) をモデルに登録する (図 5)。本システムではビン、缶、ペットボトルなど 150 種類の物体モデルを作成した。図 6 は物体モデルとの照合例である。

また、大きさや色にばらつきを持つ楕円体近似可能な果物についても認識対象に加えた¹⁰⁾。最初に原画像 (図 7 (a)) から作成した正規化画像に対して、指定された果物の色にあてはまる候補領域 (図 7 (b)) を抽出する。次に、エッジ画像 (図 7 (c)) を用いて、ハフ

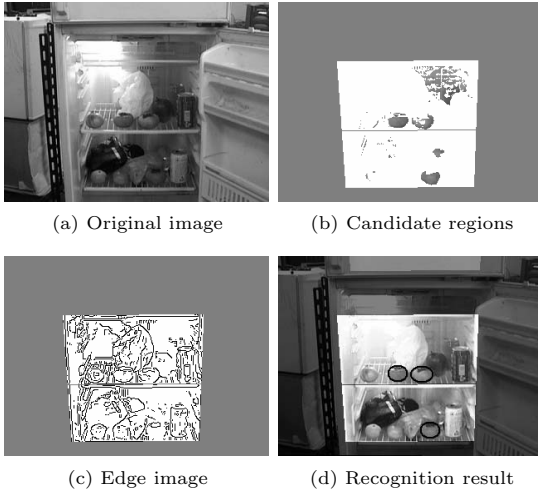


図 7 果物の認識

Fig. 7 Recognition of fruits.

変換により候補領域付近で楕円を抽出して、モデルの大きさにあてはまるものを物体候補 (図 7 (d) の黒枠) とする。

3.2 自動物体認識

まずシステムは以下の手順に従って自動物体認識を行う。

- (1) 照明条件に合わせて物体モデルを変換
- (2) 物体の候補領域を抽出
- (3) 物体の種類を照合して物体の領域を決定
- (4) 抽出された領域を物体モデルと照合

図 6 は自動物体認識の例である。この例は、ユーザが左上の黄色の缶 (グレープフルーツジュース) を指定した場合である。代表色を用いて缶の領域を抽出すると、図 6 (b) のように指定物体に加えて下の棚の黄色の缶 (レモンティー) も抽出される。しかし、次に指定物体の二次特徴を照合すると左上の缶だけから特徴が抽出される (図 6 (c) の緑の領域) ので、左上の缶を指定物体として認識することができる。

3.3 未知照明条件下でのモデル色変換の推定

観測される物体の色は照明条件によって大きく変化する。そこで、物体認識時の未知照明条件 L のもとで観測される物体 O の色 \vec{c}^{OL} を推定するために、参照物体 (ここでは冷蔵庫の扉や枠) の色を利用して物体の色がどのように変化するかを予測し物体を探索する¹¹⁾ (図 8)。

Lambertian 表面での光の反射の物理モデルより、照明条件 L のもとで観測される物体 O の色 \vec{c}^o は、有限次元線形モデルを用いると

$$\vec{c}^{OL} = A^L \vec{s}^O \quad (1)$$

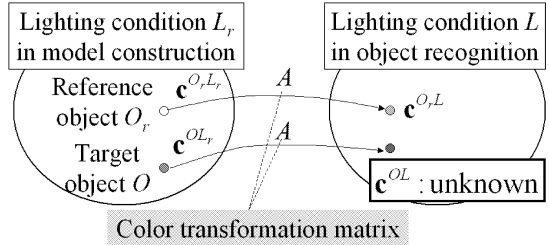


図 8 モデル登録時と異なる照明条件下における色推定問題
Fig. 8 Color estimation problem under different illumination.

と線形表現されることが分かっている。これから、異なる照明 L_1 と L_2 のもとで観測したときの物体 O の色 \vec{c}^{OL_1} 、 \vec{c}^{OL_2} は

$$\vec{c}^{OL_2} = A^{L_2 L_1} \vec{c}^{OL_1} \quad (2)$$

で線形変換される。未知照明下で観測された参照物体の色 \vec{c}^r をもとに、基準照明下での物体モデルの色 \vec{c}^m から未知照明下の物体の色 \vec{c}^o への変換行列 $A^{L_2 L_1} = A(\vec{c}^r; \vec{\alpha})$ を線形推定する。

$$\vec{c}^o = A(\vec{c}^r; \vec{\alpha}) \vec{c}^m \quad (3)$$

$$A_{ij}(\vec{c}^r; \vec{\alpha}) = \sum_k \alpha_{ijk} \vec{c}_k^r \quad (4)$$

変換行列 A の要素は $4 \times 3 = 12$ 個あり、これを参照物体の色 \vec{c}^r (3 成分) から線形推定するので、48 の推定パラメータ α_{ijk} を環境ごとに同定する必要がある。

本システムでは、最初は generic な推定パラメータを用いて色変換の推定を行い、物体検出を試みる。うまく見つからなかった場合には次節で述べるユーザとの対話に基づいて物体を発見する。一度見つければ、現在の照明条件下での指定の物体の色が分かるので、基準照明下でのモデル色と現在の照明下の物体色、現在の参照物体色の組合せを用いて線形変換の推定パラメータを修正する¹¹⁾。これにより、システムがユーザとのインタラクションを通じて次第に環境に適応していくことが期待できる。

3.4 認識結果の画面提示法

複数の物体候補が認識結果として得られた場合、ユーザが候補を選択する際に心理的負担の少ない方法で情報を画面提示する必要がある¹⁰⁾。本システムでは、物体候補領域のラベル付け同時表示や、順番に表示していく方法、候補選択を塊ごとに階層化して表示する方法について、候補選択肢の数とユーザが候補選択にかかる時間の関係を実験で確認し、選択時間が最も短くなる方法で提示する方法を切り替える。候補数が少ないときは 1 つずつ表示してそのつどユーザに確認



(a) Numbering for overlapping candidates

(b) Coloring for many candidates

図 9 認識結果の画面提示方法

Fig. 9 Display method of recognition results.

するものが良く、多くなるに従って、逐次表示して最後にどれかユーザに聞くものか、色と番号を付けて同時表示して選択させる方法が良いという結果となった(図9)。

4. 物体認識のための対話の生成と認識

システムはユーザとの対話を利用して物体認識を行うが、なるべく自然な対話を目指し、ユーザに煩わしさを感じさせないようにする必要がある。そのためには、以下のことに留意して対話を行う必要がある。

- なるべく自動で物体を絞り込む
たとえば、同じ物体を2つ検出したときに、「どちらにしますか?」と言ってユーザに選んでもらうよりは、1つ選んで「こちらにしますか?」と質問した方が良い。
- ユーザから必要な情報を得やすいような質問をする
たとえば物体を検出できなかったときは、ただ単に「見つかりませんでした」と答えるよりは、「どのへんにありますか?」と質問した方がユーザは情報を与えやすい。

システムはこのようなコンセプトに基づいて対話生成を行う。

4.1 音声認識処理

本研究のシステム構成を図10に示す。音声認識には当初 IBM ViaVoice を使って開発を行い、現在は連続音声認識コンソーシアム¹⁶⁾ が開発した大規模語彙連続音声認識ソフトウェア Julius/Julian¹⁸⁾ を用いている。本システムでは、文脈自由文法をサポートしたエンジンによって定型文を認識し、定型文にマッチしないときはディクテーションエンジンによって音声を仮名漢字文字列に変換する。システムは文字列からあらかじめ用意した発話文の確率モデルを用いて、未登

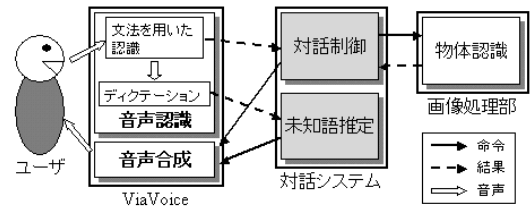


図 10 音声対話システム

Fig. 10 Our speech dialog system.

録の単語や誤認識された単語の推定を行いユーザの自由な発話ができるだけ許容するようになっている。

音声発話については当初 ViaVoice を利用して開発し、現在は Galatea プロジェクト¹⁹⁾ の開発した擬人化対話エージェントツールキット Galatea toolkit を用いている。

4.2 対話生成と制御

システムは現在のタスクの状態を保持しており、最初はユーザからの依頼待ち状態にある。この状態ではシステムはユーザの発話が発生した時点で発話内容を認識する。文脈自由文法を定義しているため、それにマッチする文、たとえば「コーラをとって」が正しく認識されれば、「コーラ」を対象として前述の自動画像認識を行う。

物体候補が検出された場合、システムは「これでいいですか?」とユーザにアドバイスを求める。確信度の高いものが見つからない場合、それに近い物体があれば、ユーザに注意深く見てもらうため「自信がないですが、これで正しいですか?」とユーザからのアドバイスを促す。ユーザが「はい」と答えれば決定するが、もし誤認識している場合には、ユーザは、(1) 指定物体の位置の情報を与えるか、(2) 必要に応じてシステムに認識が間違っていると指摘する、と仮定して、そのような内容のアドバイスがユーザの発話に含まれていないかどうか、ユーザ発話の認識結果を検証して情報を取得する。ユーザのアドバイスが誤認識の指摘のみにとどまる場合、システム側が積極的にユーザに「どこにありますか?」のように問合せを行う。これらのシステム発話の生成は、現在の認識状況とそれまでユーザ発話の履歴に基づいて行われる。その具体的な手順については文献12)に詳しい。

ユーザが、絶対的な位置情報(たとえば「上の棚の左の方にあります」)や、相対的な位置情報(たとえば「物体Aの右にあります」)を与えた場合には、システムは与えられた位置情報をもとにして指定物体の認識を試みる。ユーザが隠蔽情報(たとえば「物体Aの後ろです」)を与えた場合には、システムはAの周囲を詳細に検査する画像認識処理を起動する。具体的

平成15年度に活動を終了し、現在音声対話技術コンソーシアム¹⁷⁾が引き継いでいる。



(a) Original image (b) Recognition result

図 11 対話に基づく隠蔽物体の認識例

Fig. 11 Recognition of an occluded object based on speech dialog.

には、弱いエッジや色照合の閾値を緩める、部分的な隠れを考慮する処理を行う¹³⁾。

候補となる領域も指定物体も検出されなかった場合には、システムはユーザから物体の位置を問い合わせると同時に、冷蔵庫内の全物体の検出を試み、ユーザが指定しやすいよう、検出された物体を表示する。システムから「物体 A と物体 B が見つかりました」という情報を与えられると、ユーザは「その A の後ろの缶を取って」などと情報を与えやすいからである。

対話による物体認識の例を図 11 に示す。ここで行われた対話は以下のとおりである。

ユーザ：白い缶（ダカラ）とって

システム：見つかりませんでした。どのへんにありますか？

ユーザ：青い缶（アクエリアス）の後ろ

システム：1 個見つかりました。これを取りますか？(図 11 (b))

ユーザ：はい

システム：取りに行きますので、しばらくお待ちください

4.3 画像認識結果を利用した未知語の推定

現状の音声認識エンジンでは誤認識を避けることが難しい。そこで、画像認識処理の進捗状況や文法における文脈を考慮して、誤認識された発話を推定する¹³⁾。

- (1) ディクテーション文字列から未知語部分を検出する。
- (2) 状況、文脈、ディクテーション文字列を考慮して、各登録語の確率を計算する。
- (3) 最大の確率を持つ推定結果を採用する。

(2) では未知語が、a) 登録語の同義語、b) 誤認識された登録語、である場合に分ける。すべての確率が閾値以下の場合には雑音であると推定し、不明な部分をユーザに質問する。

本システムはシステムの状況として以下の 3 つを考慮している。

- 最初にユーザが発話する状況
- ユーザの発話に対して質問している状況
「何を取りますか?」「何色の物体ですか?」など
- 認識結果に対して質問している状況

表 1 確率的文法モデル

Table 1 Probabilistic grammar model.

モデル	意味
$P_{c-pre}(C_1 C_2)$	C_2 の前に C_1 が発話される確率
$P_{c-next}(C_1 C_2)$	C_2 の次に C_1 が発話される確率
$P_c(C S)$	S のもとで C が発話される確率
$P_{w-pre}(W_1 W_2)$	W_2 の前に W_1 が発話される確率
$P_{w-next}(W_1 W_2)$	W_2 の次に W_1 が発話される確率
$P_w(W S)$	S のもとで W が発話される確率

C : 単語カテゴリ W : 単語 S : 状況

表 2 認識成功例 1

Table 2 Successful recognition result 1.

(下線部は登録語を表す)

>

システム	見つかりませんでした。 どのへんにありますか？
ユーザ	のほほん茶の左
認識	五本茶の <u>左</u>
カテゴリ推定	物体名
単語推定	のほほんちゃ
システム	のほほん茶の左ですか？

「見つかりませんでした。どのへんにありますか?」「1 個見つかりました。これを取りますか?」など

また文脈は、ある単語カテゴリ（物体クラス、位置名称、品詞など）の前後に出現する単語カテゴリをさす。上記の状況ごとに、単語カテゴリや個々の単語の発生しやすさを条件付き確率モデルとして学習しておく（表 1）。

登録語の同義語であると仮定して推定する場合は、両者の発音は似ていないと考えられるので、状況 S 、文脈 $cont$ だけを考慮した登録語 W の事後確率を最大にするものを求める（式 (5)）。

$$\hat{W} = \arg \max_W P(W|S, cont) \quad (5)$$

誤認識された登録語と仮定して推定する場合は、ディクテーション文字列 R に発音が似た単語の可能性が高いので、 R と W の発音類似度も考慮する。

$$\hat{W} = \arg \max_W P(W|S, cont, R) \quad (6)$$

上記の確率を表 1 の確率モデルからベイズ推定によって求める。実際の対話における未知語推定例を表 2、表 3 に示す。

冷蔵庫画像 10 枚を研究室内の被験者に見せて発話してもらい、本システムの発話解釈の評価を行った結果、文法を用いた認識のみでは 54.5%しかユーザの発話を解釈できなかったのに対し、未知語推定の機能を備えた本システムでは 68.2%まで向上した。残りについてはユーザに質問することによって再度情報を得る

表 3 認識成功例 2

Table 3 Successful recognition result 2.
(下線部は登録語を表す)

システム	二つ見つかりました。 どちらにしますか？
ユーザ 認識	赤いキャップがついてる方
カテゴリ-推定	破壊 <u>キャップ</u> が <u>ついてる</u> 方
単語推定	あかい
システム	赤いキャップですか？



図 12 冷蔵庫の発見

Fig. 12 Detection of an refrigerator.



図 13 ユーザの発見

Fig. 13 Detection of an user.

ことで対処している。

5. ロボットの部屋内移動

冷蔵庫のおおよその位置は与えられるものとし、障害物を回避しながら冷蔵庫を発見できる場所への移動と、冷蔵庫の発見および作業のできる位置までとの移動との2つの移動を実現する。前者ではレーザー距離センサで移動量推定と障害物回避を行いながら、冷蔵庫が発見できる位置まで移動する。冷蔵庫を検出できる位置まで移動したら、次に視覚により冷蔵庫を発見し(図12)、さらにレーザー距離センサにより位置計測を行って作業可能位置まで移動する。飲み物を取り出した後は元の位置まで戻り、黒い髪と肌色を手がかりにユーザを発見し(図13)飲み物を渡す。

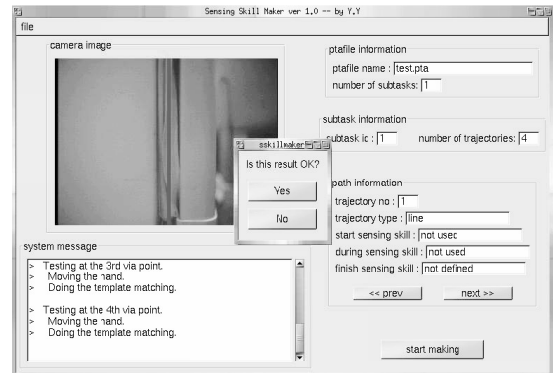


図 14 GUIによる対話的ロボットスキルの登録

Fig. 14 GUI-based interactive registration robot skills.

6. 物体の操作と対話的教示

冷蔵庫を開けたり物をつかんだりする操作は、対象の物体の種類や個別の形状などによって微妙に異なるが、基本的な動作の構造やセンサで検知しておかなければならないポイントは操作ごとにある程度決まっていることが多い。そこで本システムでは、基本的な物体操作スキル、たとえば冷蔵庫のドアを開けたり、缶やビンを取り出したりする作業の基本的な構造をあらかじめ与えておく。ただし、本体やドアの形状、大きさといった個別の冷蔵庫に依存する部分については変数にしておく。ロボットは初めて作業を行う際、作業の遂行に必要な部分でまだ入力されていない情報の提供をユーザに対して対話的に促す。とくにドアを開けるときにハンドを差し込めるところはどこか、どのような画像特徴を手がかりにすればよいか、というセンシングスキルについては、図14のようなGUIを介してユーザに必要なパラメータ、あるいはカメラによる注目点のテンプレート画像などの指定を促す。あるスキルを実行するのにどのような情報が必要かを順にロボット側からユーザに問い合わせていくことによって、不慣れなユーザでもロボットと対話的に必要な作業教示を行うことが可能となる⁹⁾。

図15に冷蔵庫を開けて缶を取り出すタスクの様子を示す。

7. まとめ

サービスロボットに必要とされるユーザとの対話機能、とくに対話的に画像認識を行う技術について報告した。自動の画像認識は環境の変動に弱くしばしば失敗が避けられないので、画像認識と音声認識の相互補完的な協調によってユーザから適切な助言を引き出し、認識および指定した物を取ってくるタスクを成功に導

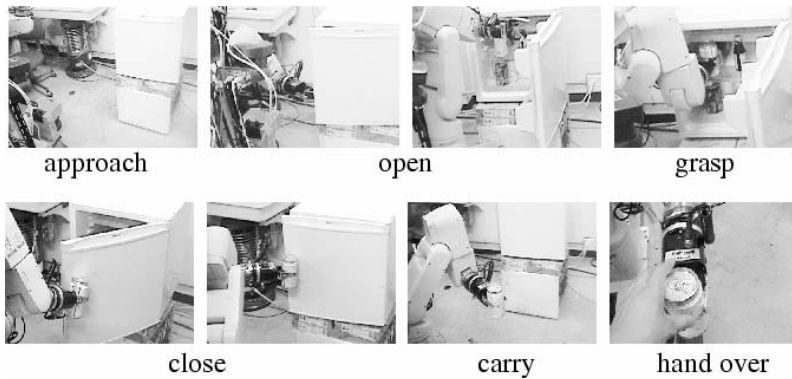


図 15 教示後のアーム動作の様子

Fig. 15 Robot arm operations based on the interactive teaching.

くインタラクティブビジョンの考え方について述べた。また、ロボットの作業スキルの教示方法にこの対話的な方策を適用し、ロボットの側が作業に必要な情報をユーザに問い合わせていくことで、不慣れなユーザでもロボットを教示できる手法についても紹介した。

実際にサービスロボットシステムを構築して実験を行っているが、認識できる物体の種類は限定的であり、対話できる内容もそれほど多くはない。ロボットが実行できるタスクも物を取ってくることに限定されている。さらに多様なタスクが実行できるためには、画像認識、音声認識といった個々の技術の性能向上とそれらの相補的な組合せをさらに推し進めることが課題となる。その際、音声、画像、その他の情報が不完全な断片としてしか得られていない状態で、どうやってユーザから有用な情報を聞き出すことができるか、という対話の本質的課題についてより深く研究していくことが必要であると考えられる。

参考文献

- 1) Takahashi, Y., Komeda, T., Uchida, T., Miyagi, M. and Koyama, H.: Development of the mobile robot system to aid the daily life for physically handicapped, *Proc. ICMA2000*, pp.549–554 (2000).
- 2) Wachsmuth, S. and Sagerer, G.: Integrated Analysis of Speech and Images as a Probabilistic Decoding Process, *Proc. 16th ICPR*, Vol.2, pp.588–592 (2002).
- 3) Ahlrichs, U., Fischer, J., Denzler, J., Drexler, C., Niemann, H., Noth, E. and Paulus, D.: Knowledge Based Image and Speech Analysis for Service Robots, *Proc. workshop on Integration of Speech and Image Understanding*, Kerkyra, Greece, pp.1–27 (1999).
- 4) 高橋拓弥, 中西周作, 久野義徳, 白井良明: 音声とジェスチャによる対話に基づくヒューマンロボットインターフェース, *インタラクション'98 論文集*, pp.161–168 (1998).
- 5) 渡辺靖彦, 長尾 真, 岡田至弘: 画像の内容を説明するテキストを利用した画像解析, 画像の認識・理解シンポジウム (MIRU'96), Vol.2, pp.271–276 (1996).
- 6) 稲邑哲也, 稲葉雅幸, 井上博允: ユーザとの対話に基づく段階的な行動決定モデルの獲得, *日本ロボット学会誌*, Vol.19, No.8, pp.983–990 (2001).
- 7) Makihara, Y., Takizawa, M., Ninokata, K., Shirai, Y., Miura, J. and Shimada, N.: A Service Robot Acting by Occasional Dialog — Object Recognition Using Dialog with User and Sensor-Based Manipulation, *J. Robotics and Mechatronics*, Vol.14, No.2, pp.124–132 (2002).
- 8) 矢野恵生, 榎原 靖, 滝澤正夫, 白井良明, 三浦 純, 島田伸敬: 福祉用サービスロボットの開発, 第2回福祉工学シンポジウム講演会講演論文集, pp.145–148 (2002).
- 9) 矢野恵生, 三浦 純, 白井良明, 島田伸敬: ロボットに対する認識を要する作業のインタラクティブ教示, 2004年ロボティクス・メカトロニクス講演会, 名古屋 (2004).
- 10) 榎原 靖, 白井良明, 島田伸敬: 対話を用いた物体認識のための候補選択戦略, 画像の認識・理解シンポジウム MIRU2005, CD-ROM (2005).
- 11) 榎原 靖, 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*, Vol.J87-D-II, No.2, pp.629–638 (2004).
- 12) Makihara, Y., Takizawa, M., Shirai, Y., Miura, J. and Shimada, N.: Object Recognition Supported by User Interaction for Service Robots, *Proc. 5th ACCV*, Vol.2, pp.719–724 (2002).
- 13) 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).

- 14) 藤井憲作, 杉山和弘: 歩行者ナビゲーション支援のための場所案内文生成手法, 電子情報通信学会論文誌, Vol.J82-D-II, No.11, pp.2026-2034 (1999).
- 15) 岩田 満, 鬼沢武久: 絵のつながりを考慮した絵情報の言語的表現, 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.2, pp.337-350 (2001).
- 16) 連続音声認識コンソーシアム.
<http://www.lang.astem.or.jp/CSRC/>
- 17) 音声対話技術コンソーシアム.
<http://www.astem.or.jp/istc/>
- 18) 連続音声認識ソフトウェア Julius/Julian.
<http://julius.sourceforge.jp/>
- 19) Galatea プロジェクト. <http://hil.t.u-tokyo.ac.jp/~galatea/index-jp.html>

(平成 18 年 1 月 10 日受付)

(平成 18 年 7 月 21 日採録)

(担当編集委員 佐藤 洋一)



島田 伸敬 (正会員)

平成 4 年大阪大学工学部電子制御機械工学科卒業, 平成 7 年同大学大学院博士後期課程修了. 博士(工学). 同年同専攻助手. 平成 13 年同研究科研究連携推進室情報ネットワーク部門講師, 同助教授を経て, 平成 16 年より立命館大学情報理工学部知能情報学科助教授, 現在に至る. コンピュータビジョン, ジェスチャ認識, ヒューマンインタフェース, インターネットソリューションの研究に従事. 電子情報通信学会, 人工知能学会, IEEE 各会員.



三浦 純 (正会員)

昭和 59 年東京大学工学部機械工学科卒業. 平成元年同大学大学院工学系研究科情報工学専攻博士課程修了, 工学博士. 同年大阪大学助手. 現在同大学大学院工学研究科機械工学専攻助教授. 知能ロボット, 人工知能, コンピュータビジョンの研究に従事. 平成 6~7 年米カーネギーメロン大学客員研究員. 平成 9 年ロボット学会論文賞受賞. 人工知能学会, 電子情報通信学会, 日本機械学会, IEEE, AAAI 各会員.



白井 良明 (正会員)

昭和 39 年名古屋大学工学部機械工学科卒業. 昭和 44 年東京大学大学院工学系研究科機械工学専攻博士課程修了. 同年電子技術総合研究所研究官. 昭和 60 年 5 月同制御部部長. 昭和 63 年 4 月大阪大学工学部 (後に大学院工学研究科) 教授. 平成 8 年 4 月~平成 11 年 3 月東京大学大学院工学研究科教授併任. 平成 14 年 8 月現在情報学研究所客員教授. 平成 17 年 4 月立命大学情報理工学部教授. 知能ロボット, 画像処理, ヒューマンインタフェースの研究に従事. IEEE, 電子情報通信学会, 人工知能学会, 日本ロボット学会等の会員.