

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[†]

[†]立命館大学情報理工学部知能情報学科 [‡]大阪大学大学院工学研究科機械工学専攻

〒 525-8577 滋賀県草津市野路東 1-1-1 〒 565-0871 大阪府吹田市山田丘 2-1

E-mail: shimada@ci.ritsumei.ac.jp

ユーザの代わりに指示されたものをとってくるロボット(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おこなう技術について報告する。たとえば冷蔵庫内の飲み物をとってくるといったタスクの場合、庫内にあるさまざまな物体が互いに隠蔽しあうようにお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なうことで、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すことができれば、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[†]

[†]Dept. of Human and Computer Intelligence, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, 525-8577, Japan

[‡]Dept. of Mechanical Engineering, Osaka University, 2-1 Yamadaoka, Suita, 565-0871, Japan

E-mail: shimada@ci.ritsumei.ac.jp

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. Because various objects are often overlapping together in the refrigerator and their colors and textures looks various dependently on the illumination conditions inside and outside the refrigerator, full-automatic object recognition often fails and cannot recover the failure 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. This report shows an implementation of *iterative vision* and its effectiveness.

1 はじめに

近年、高齢化社会の到来により人を助けるサービスロボットの必要性が高まっている。たとえば、画像を用いて認識した物体を持ってくるサービスロボットの研究が行われている [1]。しかし、サービスロボットの対象シーンは複雑で自動認識が困難な場合があ

る。このような場合に、視覚以外の情報を利用することが考えられる。その中には、画像情報と言語情報を組み合わせた研究 [2, 3, 4, 12, 13] がある。

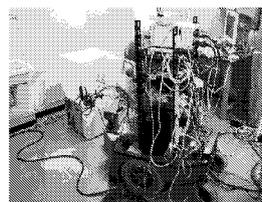
我々の研究グループでも画像認識に言語情報を利用することを考え、自動物体認識に失敗した場合には、ユーザとの音声による対話を通して認識の補助となるような情報を得て、それを元に物体認識を試

みることを考えた [5, 6, 10, 11].

たとえば冷蔵庫内の飲み物をとってくるといったタスクの場合、庫内にあるさまざまな物体が互いに隠蔽しあうようになっていることが多く、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なうことで、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すことができれば、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。

本報告では、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、現在わ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、音声認識が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。ロボットスキル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



(a) Our service robot



(b) User interface

図 1: サービスロボットシステム

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。認識が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通りである。(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

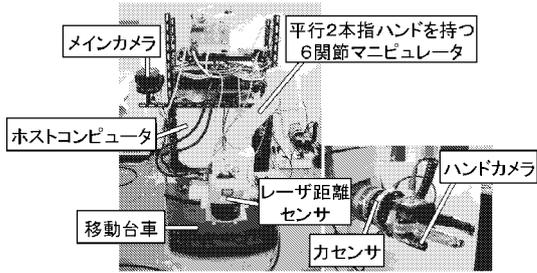


図 2: パーソナルサービスロボット

- | | |
|----------------------|---------------|
| 1. 音声により作業内容を得る | 5. 飲み物を取り出す |
| 2. 冷蔵庫の前まで移動する | 6. 冷蔵庫の扉を閉める |
| 3. 冷蔵庫の扉を開く | 7. ユーザの元へ戻る |
| 4. 対話を用いて
飲物を認識する | 8. ユーザに飲み物を渡す |

図 3: サービスロボットの作業手順

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)、及び二次特徴(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。

また、大きさや色にばらつきをもつ楕円体近似可能な果物についても認識対象に加えた[8]。最初に原画像(図6(a))から作成した正規化画像に対して、指

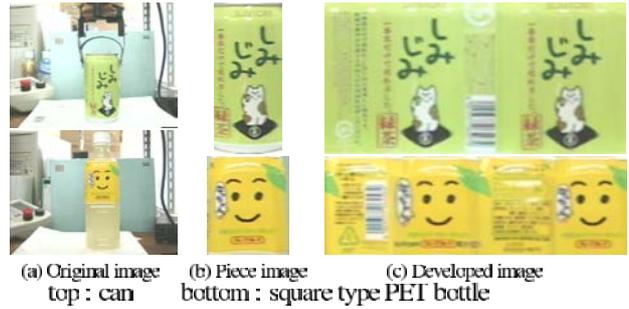


図 4: 投影画像の生成

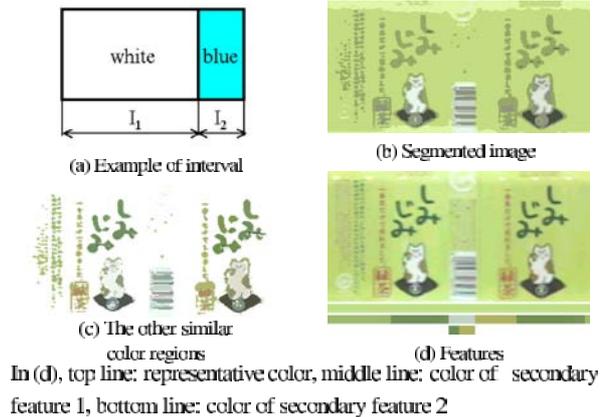


図 5: モデル特徴の抽出

定された果物の色に当てはまる候補領域(図6(b))を抽出する。次に、エッジ画像(図6(c))を用いて、ハフ変換により候補領域付近で楕円を抽出して、モデルの大きさに当てはまるものを物体候補(図6(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7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。代表色によって候補を絞り(図7(b))、二次特徴を探索して照合すると左上の缶を指定物体として認識することができる(図7(c))。

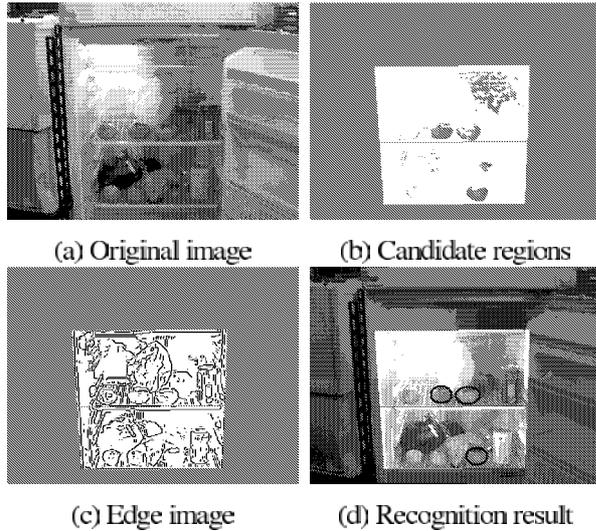
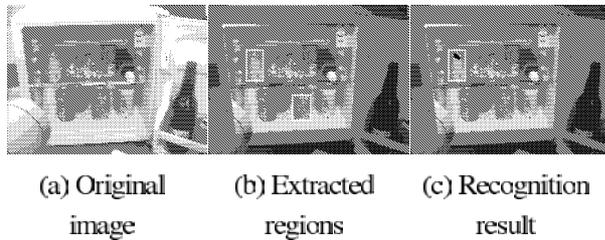


図 6: 果物の認識



In (c), black region is extracted secondary feature.

図 7: 物体モデルとの照合

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 \tilde{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 [9] (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 \tilde{c}^O は、有限次元線形モデルを用いると

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

と線形表現されることがわかっている。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 \tilde{c}^{OL_1} 、 \tilde{c}^{OL_2} は

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

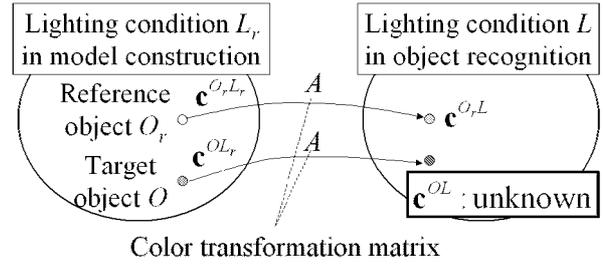


図 8: 色推定の問題設定

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 \tilde{c}^r をもとに、基準照明下での物体モデルの色 \tilde{c}^m から未知照明下の物体の色 \tilde{c}^o への変換行列 $A^{L_2 L_1} = A(\tilde{c}^r; \tilde{c}^m)$ を線形推定する。

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

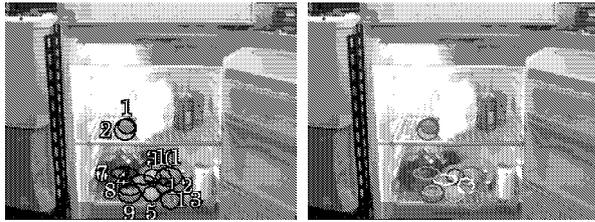
$$A_{ij}(\tilde{c}^r; \tilde{c}^m) = \sum_k \alpha_{ijk} \tilde{c}_k^r \quad (4)$$

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 \tilde{c}^r (3 成分) から線形推定するので、48 の推定パラメタ 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 な推定パラメタをもち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わかるので、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み合わせを用いて線形変換の推定パラメタを修正する [9]。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 [8]。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、もっとも選択時間の短くなる方法で提示する方法を切り



(a) Numbering for overlapping candidates (b) Coloring for many candidates

図 9: 認識結果の画面提示方法

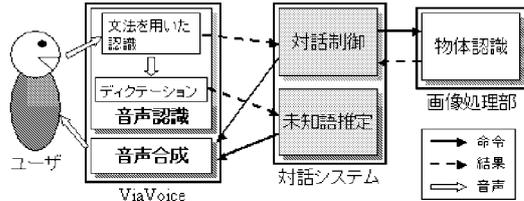


図 10: 音声対話システム

替える。候補数が少ないときはひとつずつ表示してそのつどユーザに確認するものがよく、多くなるにしたがって、逐次表示して最後にどれかユーザに聞くものか、色と番号をつけて同時表示して選択させる方法がよいという結果となった。

4 物体認識のための対話の生成と認識

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 [14]¹ が開発した大規模語彙連続音声認識ソフトウェア Julius/Julian [16] 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をできるだけ許容するようになっている。

¹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 [15] が引き継いでいる。

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 [17] 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。

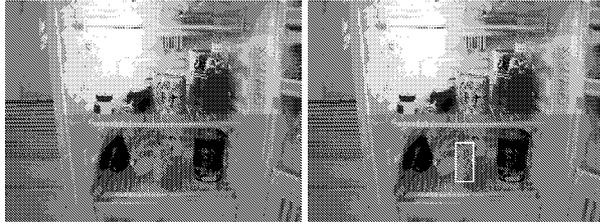
4.2 対話生成と制御

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を発生した時点で発話を認識する。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?」とユーザからの応答を促す。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、発話に対応する。

ユーザが、絶対的な位置情報 (例えば「上の棚の左の方にあります」) や、相対的な位置情報 (例えば「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 (例えば「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 [11]。

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。



(a) 原画像 (b) 認識結果

図 11: 対話に基づく隠蔽物体の認識例

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 [11]。

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、である場合にわけ。全ての確率が閾値以下の場合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
「見つかりませんでした。どのへんにありますか

表 1: 確率モデル

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: 成功例 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 <u>ち</u> ゃ
システム	のほほん茶の左ですか？

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、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

表 3: 成功例 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	赤いキャップがついてる方
認識	破壊 <u>キャップ</u> が <u>ついてる方</u>
カテゴリー推定	色
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

図 13: ユーザの発見

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 参照)、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 参照)飲み物を渡す。

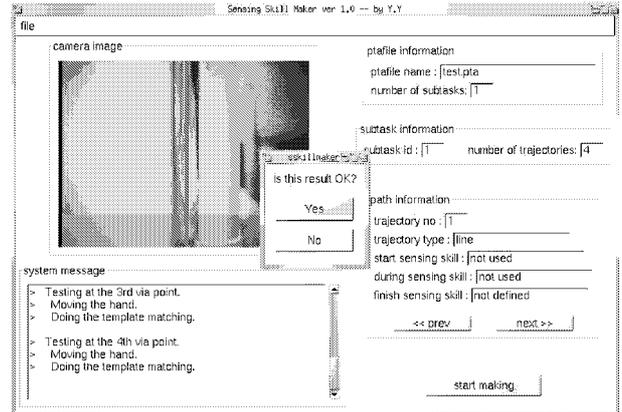


図 14: GUI による対話的ロボットスキルの登録と検証

6 物体の操作と対話的教示

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、缶やピンを取り出す作業の基本的な構造はあらかじめ与えておく構造をとっている。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空欄になっている。そこで、ロボットはユーザに対し、作業の遂行に必要な部分でまだ空欄になっている情報の提供を対話的に促す。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はロボットがすべてわかっているので、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 [7]。

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



図 15: 作業の様子

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とってくるタスクを成功に導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ることで、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と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み合わ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ます。

参考文献

- [1] Y. Takahashi, T. Kameda, T. Uchida, M. Miyagi, and H. Koyama, "Development of the mobile robot system to aid the daily life for physically handicapped", Proc. ICMA2000, pp.549-554, 2000.
- [2] S. Wachsmuth and G. Sagerer, "Integrated Analysis of Speech and Images as a Probabilistic Decoding Process", Proc. 16th ICPR, vol.2, pp.588-592, 2002.
- [3] U. Ahlrichs, J. Fischer, J. Denzler, C. Drexler, H. Niemann, E. Noth, and D. Paulus, "Knowledge Based Image and Speech Analysis for Service Robots", Proc. of workshop on Integration of Speech and Image Understanding, pp.1-27, Kerkyra, Greece, 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", インタラクシオン'98 論文集, pp.161-168, 1998.
- [5] Y. Makihara, M. Takizawa, K. Ninokata, Y. Shirai, J. Miura, and N. Shimada, "A Service Robot Acting by Occasional Dialog -Object Recognition Using Dialog with User and Sensor-Based Manipulation-", J. of Robotics and Mechatronics, vol.14, no.2, pp.124-132, 2002.
- [6] 矢野恵生, 横原靖, 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.
- [7] 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年6月.
- [8] 横原靖, 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.
- [9] 横原靖, 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.
- [10] Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada, "Object Recognition Supported by User Interaction for Service Robots", Proc. 5th ACCV, vol.2, pp.719-724, 2002.
- [11] 滝澤正夫, 横原靖, 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.
- [12] 渡辺, 長尾, 岡田: 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)
- [13] 岩田, 鬼沢: 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, Vol. J84-D-II, No. 2, pp. 337-350 (2001)
- [14] 連続音声認識コンソーシアム <http://www.lang.astem.or.jp/CSRC/>
- [15] 音声対話技術コンソーシアム <http://www.astem.or.jp/istc/>
- [16] 連続音声認識ソフトウェア Julius/Julian <http://julius.sourceforge.jp/>
- [17] Galatea プロジェクト <http://hil.t.u-tokyo.ac.jp/galatea/index-jp.html>