

インタラクティブビジョンにおいて ユーザから有用な助言を得るための対話戦略

A Dialog strategies to get helpful advice from a user in Interactive Vision

井本 浩靖*1 白井 良明*2 島田 伸敬*2 三浦 純*1
Hiroyasu Imoto Yoshiaki Shirai Nobutaka Shimada Jun Miura

*1大阪大学大学院工学研究科

*2立命館大学情報理工学部

Graduate School of Engineering, Osaka University

Information Science and Engineering, Ritsumeikan University College

This paper describes an interactive vision system for a Service robot which finds an object specified by a user and brings it to the user. The system first registers object models automatically. When the user specifies an object, the system tries to recognize it automatically based on the object model. But, we have cases that the system cannot recognize an object correctly in the situations when the object is occluded by other objects. In that case, the system shows object recognition results to the user, and ask the user a question to get helpful advice what kind of situation the object is or what kind of object is occluded by. Then the system recognize the object again based on the advice. And, when the system fails to interpret speech recognition results, we enable the system to interpret purpose of user's utterances by setting keywords depending on each situation. Experiments using real-world refrigerator scenes are shown.

1. はじめに

近年、高齢化社会の到来によりサービスロボットの必要性が高まっている。そのようなロボットの機能の中で、ユーザに指定されたものを持ってくるといった機能は重要である。本研究では、指定物体を画像中から認識することでその位置を求める。

画像情報と言語情報を組み合わせた研究として、Wachsmuthら [1] はユーザの発話内容に一致する特徴を画像から抽出することで物体を認識するシステムを提案したが、認識が失敗したときにそこから再度認識を試みてはいない。それに対して、榎原ら [2] は物体認識が失敗したときにユーザからの助言により再度認識を行うシステムについての研究を行った。しかし特定の状況のみしか扱っておらず、またどのような助言がロボットにとって適切であるかをユーザが知らず、ユーザから有用な助言を得られない場合がある。

そこで我々は榎原ら [2] を基に、どのような助言が有用であるかをユーザに伝えることによって、適切な助言を得るシステムについての研究を行った。システムはまずあらかじめ登録された物体モデルを用い可能な限り自動での物体認識を試みる。しかし、隠蔽や照明条件等により候補領域が適当な大きさに抽出できない場合などには、指定された物体を認識できない。そのような場合、ロボットが確実に作業を行うためにはどのような情報が必要であるかを認識結果から判断し、ユーザに尋ねる。そしてユーザから得られた助言に基づいて物体の再認識を試みる。また、ユーザの発話が完全に認識できない場合においても対話を円滑に進めるため、状況に応じたキーワードを設定しておくことによりユーザの発話意図をできるだけ解釈する。

本研究のサービスロボットは、マニピュレータを用いた物体の操作も行うが、その詳細については文献 [3] を参照されたい。

2. 自動物体認識

ユーザから対話によって取ってほしい物体 (以下、指定物体と呼ぶ) が伝えられると、システムは可能な限り自動での物体認識を試みる。

連絡先: 井本浩靖, 〒565-0026 大阪府吹田市山田丘 2-1 大阪大学大学院工学研究科機械工学専攻三浦研究室, TEL/FAX 06-6879-7333, imoto@cv.mech.eng.osaka-u.ac.jp

システムは最初にどの方向からでも認識できるように図 1(a) のような投影画像を用いて物体モデルを作成しておく [4]。そして、冷蔵庫の扉の色 (図 1(b) の矢印で指し示した白枠) を参照色として用いて、原画像 (図 1(b)) をモデル作成時の基準照明条件に変換した正規化画像を作成する (図 1(c))。その画像から指定された物体の代表色の範囲に入る候補領域を抽出する (図 1(d)) [5]。そして、図 2 のような物体の形状のモデルを用い、候補領域周辺からエッジやペットボトルのキャップなどそれぞれの形状に関する特徴 (以下、形状特徴と呼ぶ) を抽出して物体の形状を認識する (図 1(e))、赤色のグラフはエッジ強度、グラフ上の白線はそのしきい値を表し、形状モデルの範囲内でしきい値以上の部分を候補エッジとし、物体上の白線で表す。このようにして認識した候補物体をディスプレイを通してユーザに表示する (図 1(f)) なお、詳細な物体認識の手法については文献 [2], [6] を参照されたい。

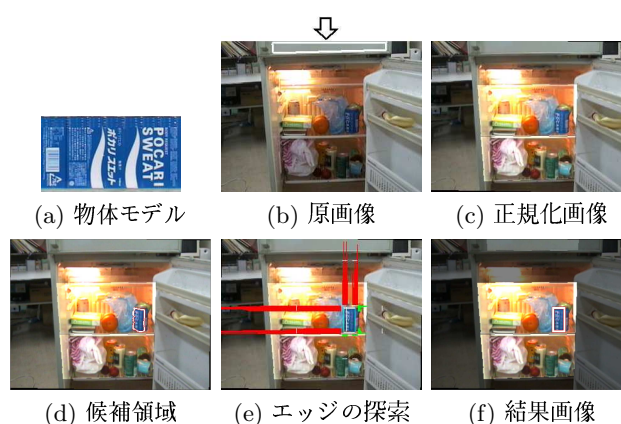


図 1: 物体認識の概要

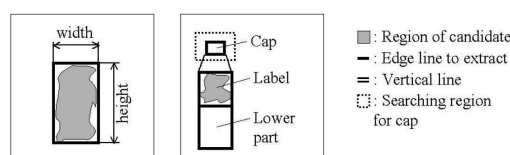


図 2: 物体の形状モデル

3. 物体認識結果と候補領域のクラス分け

実際の冷蔵庫環境において、指定物体には他の物体に隠蔽されているなどいくつかの状況が考えられる。そのような状況に対して2. で述べたような物体認識を行うと、物体モデルより大きく見つかる候補物体や形状特徴の一部が認識できない候補物体など、いくつかの認識結果が得られる。このとき、システムは指定物体である可能性の高い候補からユーザに表示するために、認識結果に基づいて候補領域のクラス分けを行う。

3.1 冷蔵庫環境での物体の状況

冷蔵庫環境内における指定物体の状況としては以下のような状況を考える。

1. 他の物体に隠蔽されていない
2. 違う色の物体にある程度隠蔽されている
(ある程度隠蔽されているとは、画像上である程度の大きさは見える程度隠蔽された状況である)
3. 違う色の物体に大きく隠蔽されている
(大きく隠蔽されているとは、画像上ではほとんど見えなほど大きく隠蔽された状況である)
4. 似た色の物体と重なっている

3.2 物体認識によって得られる結果の分類

物体認識によって得られる結果を以下のように分類する。

1. 幅、高さがともに物体モデルの許容範囲内
2. 幅、高さのどちらか一方が物体モデルの許容範囲内、他方が許容範囲より大きい
3. 幅、高さのどちらか一方が物体モデルの許容範囲内、他方が許容範囲より小さい
4. 幅、高さのどちらか一方が許容範囲より大きく、他方が許容範囲より大きいか小さい
5. 幅、高さがともに許容範囲より小さい
6. 候補領域が抽出できない

ただし、認識結果1から5については、形状特徴の認識結果よりさらに次のaとbの2つに分類する。

- a. 形状特徴がすべて認識される
- b. 形状特徴が一部、もしくは全く認識されない

また、候補領域の幅、高さを物体モデルの許容範囲と比較した場合、9つの認識結果が考えられるが、候補領域の幅が許容範囲内であり高さが許容範囲より大きい場合と、候補領域の高さが許容範囲内であり幅が許容範囲より大きい場合は、似た色の物体と重なっている状況のときに得られる認識であり、ユーザへの表示方法、物体の再認識方法が同じであることより、同一の認識結果2に分類した。認識結果3、認識結果4についても同様に複数の場合を同一の認識結果に分類した。

指定物体に対して2. で述べた物体認識を行うと、各状況においてそれぞれ表1の物体認識結果が得られる。

3.3 候補領域のクラス分け

3.2 で述べた物体認識結果に基づいて、候補領域が得られた物体について指定物体である可能性の高い候補から表示するために、3つのクラスに分類する。具体的には、候補領域の幅、高さが物体モデルの許容範囲内に入っているかどうか、形状特徴がどれくらい認識できたかで以下のように分類する。

クラス1：幅、高さが共に許容範囲内であり形状特徴がすべて認識された場合(認識結果1-a)

クラス2：幅、高さが共に許容範囲内であるが一部もしくは全ての形状特徴が認識されなかった場合、または幅、高さのどちらか一方が許容範囲内でない場合(認識

表1: 物体の状況と得られる認識結果の関係

	状況1	状況2	状況3	状況4
認識結果1-a	○			○
認識結果1-b	○			○
認識結果2-a				○
認識結果2-b				○
認識結果3-a	○	○		○
認識結果3-b	○	○		○
認識結果4-a				○
認識結果4-b				○
認識結果5-a	○	○		○
認識結果5-b	○	○		○
認識結果6	○	○	○	○

結果1-b, 2-a, 2-b, 3-a, 3-b)

クラス3：幅、高さが共に許容範囲内ではない場合
(認識結果4-a, 4-b, 5-a, 5-b)

4. 認識結果に基づいた対話

候補領域が得られた場合、候補が複数であれば対話により限定してもらおう。さらに、一つに限定されてもその候補の状況や正確な位置が特定できていない場合はユーザに伝え、得られた助言によって物体の再認識を行う。また、候補領域が抽出できない場合は指定物体の位置を知ることができないので、位置に関する対話を行うことにより候補領域の再抽出を行う。

4.1 候補領域が得られた場合

4.1.1 候補領域の表示と限定

候補領域を表示する際には、3.3 で述べたクラス分けによって分類された候補領域のうち上位クラスからそのクラスに含まれる候補をすべて表示し、候補が1つしかない場合には確認を行う。候補が複数ある場合は対話により候補を限定する。

候補の限定 ユーザが候補領域から指定物体の限定する際の発話としては、以下の場合がある。

- "缶" や "赤" など、物体の特徴で指定
- "上の棚" や "右" など、絶対位置で指定
- "赤い缶の左" など、他物体との相対位置で指定
- "どれでもいいです" のように、システムに選択を委譲

の4つの場合がある。前者の2つの場合、システムは画像上での候補の形状や色、および座標により限定することができる。また相対位置で指定された場合は、まず2. で述べた方法により相対された物体(以下、相対物体と呼ぶ)の認識を行う。そして、相対物体の認識により得られた候補を指定物体の認識結果と同様にクラスに分類する。さらに、相対物体候補のうち上位のクラスに属する候補がユーザの相対した物体である可能性が高いと考えられるので、指定物体候補の中から相対物体候補との画像上での距離が近く、かつその相対物体候補のクラスが上位であるものからユーザに表示し、確認する(図3)。またユーザがシステムに候補の選択を委譲した場合は、ユーザの対話による負担を少なくするため、状況や正確な位置を特定するための対話が少ない候補を優先的に選択する。状況や位置を特定するための対話については4.1.2で述べる。

クラス1に対する処理 クラス1を表示し候補が1つ限定されれば、それ以上助言を得る必要がないのでロボットにより物体を操作する。もし、候補に指定物体が存在しなければ次善クラスを表示する。次善クラスが存在しない場合は、候補領域が抽出できていないと判断できるので4.2で述べる対話に移る。



図 3: 候補領域の選択

クラス2に対する処理 クラス2を表示し候補が1つに限定されれば、指定物体の状況と正確な位置を知るための対話を行う。もし、候補の中に指定物体が存在しなければ次善クラスを表示し、次善クラスが存在しない場合は4.2の対話に移る。

クラス3に対する処理 クラス3を表示し候補が1つに限定されれば、クラス2と同様の処理を行う。もし、候補の中に指定物体が存在しなければ4.2の対話に移る。

4.1.2 指定物体の状況と位置の特定

指定物体の認識結果がクラス2もしくは3であった場合、3.2で述べた表1によって認識結果からあり得る状況を数え上げ、指定物体がどの状況であるか判断するためにユーザとの対話を行う。以下に、各クラス各認識結果においてどのような対話を行うかについて述べる。なお、ユーザの返答に基づいた物体の再認識の詳細な手法については、文献[6]を参照されたい。

●クラス2

○幅、高さとともに許容範囲内であるが一部の形状特徴が認識されない場合(認識結果1-b)

認識結果1-bが得られるのは、他の物体に隠蔽されていない状況(状況1)、似た色の物体と重なっている状況(状況4)の2つの状況である。そこで、認識されなかった特徴が存在するであろう位置を物体モデルより考えて候補物体を囲み、ユーザに正しいかどうかを確認する。もし正しいのであれば状況1であると判断し、候補物体が認識できる(図4)。正しくなければ状況4と判断する。この場合、候補物体の位置を知るためには、候補領域周辺で形状特徴を探索する方法も考えられるが、どのような物体と重なっているか判断できず、どの領域を探索すれば良いかわからない。そこで、状況4の場合は本来ならばもっと大きい候補領域が抽出されるはずであるので、代表色の範囲を広げることにより候補領域を拡張して物体の再認識を行う。

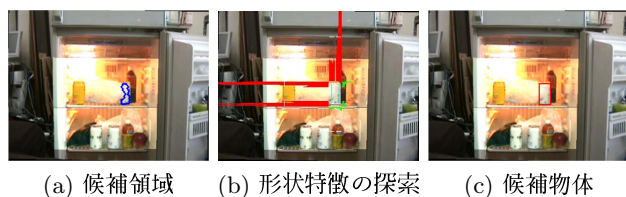


図 4: 認識1-bにおける候補領域の認識

○幅、高さのどちらか一方が許容範囲内、他方が許容範囲より大きい場合(認識結果2-a, 2-b)

認識結果2-a, 2-bが得られる物体の状況は、似た色の物体と重なっている状況(状況4)である。そこで物体の正しい位置を判断するため、認識された形状特徴を含み、かつ候補領域の中で最も代表色の割合が高くなるような領域を表示し、ユーザに正しいかどうかを確認する。もし正しいのであれば候補物体が認識できる(図5)。正しくない場合は順次割合の高い領域を表示し、ユーザに確認する。

○幅、高さのどちらか一方が許容範囲内、他方が許容範囲より小さい場合(認識結果3-a, 3-b)

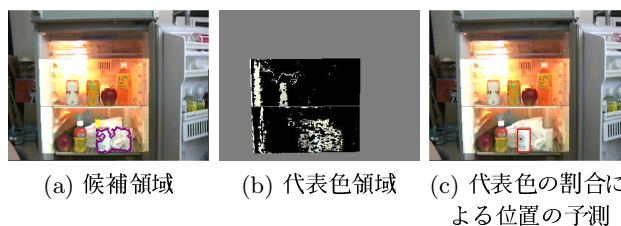


図 5: 認識2-aにおける候補領域の認識

認識結果3-a, 3-bが得られるのは、他の物体に隠蔽されていない状況(状況1)、違う色の物体にある程度隠蔽されている状況(状況2)、似た色の物体と重なっている状況(状況4)である。そこでまず、指定物体が他の物体に隠蔽されていないかをユーザに助言してもらい、ユーザの助言により指定物体が他の物体に隠蔽されているとわかれば状況2か4であると判断でき、さらにユーザにどの物体に隠蔽されているかを助言してもらい、その隠蔽している物体(以下、隠蔽物体と呼ぶ)の認識を試みる。認識できればマニピュレータで隠蔽物体を移動させ再度指定物体の認識を試みる。もし隠蔽物体の状況や位置が正確に認識できない場合は、指定物体の認識と同じようにユーザとの対話を行い助言を得る。また、隠蔽されていない場合は状況1であると判断でき、物体モデルの許す幅まで形状特徴の認識を試みる。形状特徴が認識できたら候補領域を変更し、再度ユーザに表示する(図6)。もし形状特徴がうまく認識できなかった場合には、代表色の範囲を広げることにより候補領域を拡張して再度認識を試みる。

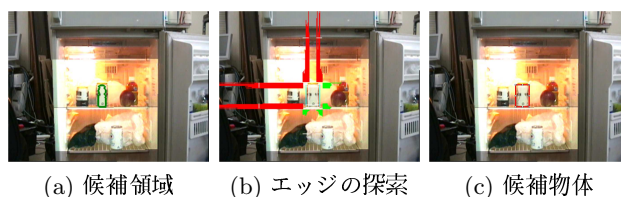


図 6: 認識3-aにおける候補領域の認識

●クラス3

○幅、高さのどちらか一方が許容範囲より大きく他方が許容範囲より大きい場合(認識結果4-a, 4-b)

認識結果4-a, 4-bが得られた場合は、クラス2における幅、高さのどちらか一方が許容範囲で、他方が許容範囲より大きい場合と同様の方法によって、どの状況であるかと物体の正しい位置を判断する。

○幅、高さともに許容範囲より小さい場合(認識結果5-a, 5-b)

認識結果5-a, 5-bが得られた場合は、クラス2における幅、高さのどちらか一方が許容範囲で、他方が許容範囲より小さい場合と同様の方法によって、どの状況であるかと物体の正しい位置を判断する。

4.2 候補領域が抽出できない場合

候補領域を抽出できない状況としては、違う色の物体に大きく隠蔽されている場合(状況3)と、全ての状況において画像上での色の範囲がずれている場合が考えられる。候補領域の抽出に失敗した場合、システムはまず、違う色の物体に大きく隠蔽されている状況(状況3)ではないかと考え、ユーザに他の物体に隠蔽されていないかを尋ねる。もし隠蔽されているなら4.1.2と同様に隠蔽物体の物体認識を試みる。また、隠蔽されていない場合は、システムが事前に登録しておいた色の範囲と正規化画像上での色の範囲が大きくずれていると考えられるので、代表色を範囲を大幅に広げて候補領域を拡張する。

5. キーワードによる発話意図の解釈

ユーザとの対話を行う場合、システムはあらかじめ各状況においてユーザが発話すると考えられる単語の並びを文法として登録しておく。そして、音声認識エンジンから得られた結果をその状況における文法とマッチングする。音声認識結果の全てが文法にマッチしない場合においても、その一部が文法とマッチしないかを調べ発話意図を解釈する。しかし、音声認識結果が不完全な場合にはその一部も文法にマッチせず、発話意図を解釈することができない。そのような場合においてもユーザとの対話を円滑に進めるため、各状況においてユーザが発話するであろう単語を状況に応じたキーワードとして設定しておく。認識結果にキーワードが存在するかを調べることによってユーザの発話意図をできるだけ解釈する。また、認識結果の一部が文法にマッチした場合にも、他の部分に更なる情報が存在するかをキーワードにより認識することを試みる。

5.1 候補を限定する場合のキーワード

複数の候補からユーザとの対話により指定物体を限定する状況の場合、システムは同一クラスに属する候補を区別できるようなキーワードを状況毎に設定し、発話解釈を行う。各状況において設定するキーワードは以下のものである。

- 候補の中に複数の形状が存在する状況
「缶」、「ペットボトル」等をキーワードに設定
- 候補の中に複数の色が存在する状況
「赤」、「赤い」、「赤色」等をキーワードに設定
- 候補が上下の棚の両方に存在する状況
「上」、「下」、「上の棚」等をキーワードに設定
- 一つの棚に複数の候補が存在する状況
「右」、「左」、「真ん中」等をキーワードに設定

ただし、「右」と「左」は絶対位置を表す単語のみではなく、「赤い缶の左」のように相対位置を表す単語として使われる場合がある。そのような場合に対処するため、キーワード「右」が認識された場合はその1つ手前の認識結果が「の」である場合は、キーワード「右」を相対位置を表す単語として解釈する。

5.2 状況と位置を特定する場合のキーワード

4.1.2 で述べた指定物体の状況と位置を特定するような対話と、4.2 で述べた候補領域が抽出できないときの対話を行う場合、システムは各クラス各認識結果に応じてシステムの問いかけに対するユーザの助言を解釈できるようなキーワードを設定する。各状況において設定するキーワードを以下に示す。

- 幅、高さがともに許容範囲内であるが一部の形状特徴が認識されない場合 (認識結果 1-b)
「はい」、「いいえ」、「そうです」等をキーワードに設定
- 幅、高さどちらか一方が許容範囲内もしくは小さく、他方が許容範囲より大きい場合 (認識結果 2-a, 2-b, 4-a, 4-b)
「はい」、「いいえ」、「そうです」等をキーワードに設定
- 幅、高さどちらか一方が許容範囲内もしくは小さく、他方が許容範囲より小さい場合 (認識結果 3-a, 3-b, 5-a, 5-b) と候補領域が抽出できない場合
 - 隠蔽されているかを尋ねている状況
「はい」、「隠されています」、物体名等をキーワードに設定
 - 隠蔽物体を尋ねている状況
物体名をキーワードに設定

5.3 キーワードを用いた対話例

音声認識結果が文法にマッチしなかった場合に、キーワードを用いてユーザの発話意図を解釈した対話例を以下に示す。な

お、認識結果とは音声認識エンジンから得た結果を示し、解釈結果とは認識結果を文法とマッチングした結果のことを示す。



(a) 候補領域 (b) 候補の限定 (c) 隠蔽物体の認識

図 7: キーワードも用いた対話例

システム「候補が3つ見つかりましたが、どれを取りましようか?」(図 7(a))

ユーザ「上のほうのやつ取って」

→ 認識結果: 上 が 小 の やつ 取って

→ 解釈結果: 取って

→ キーワードによる解釈結果: 上

システム「こちらですが、小さく見つかったのですが他の物体に隠蔽されていますか?」(図 7(b))

ユーザ「グレープフルーツの缶に隠されています」

→ 認識結果: グレープフルーツ の 欄 に 隠されています

→ 解釈結果: 隠されています

→ キーワードによる解釈結果: グレープフルーツに隠されています

システム「わかりました。それではまずグレープフルーツを探します。」(図 7(c))

6. おわりに

本論文では、対話を用いた物体認識システムにおいて、自動的な物体認識のみでは指定物体の状況や正しい位置が判断できないような場合に、ユーザから有用な助言を得るための対話戦略について述べた。最初に、冷蔵庫環境における指定物体の状況と物体認識結果の関係を与え、そして得られた結果に基づいて候補物体のクラス分けを行った。次に、候補領域が得られた物体について、その物体の状況と正しい位置を知るために助言を要求し、その助言に基づいた物体の再認識を行った。候補領域が得られない物体についても、その物体の位置に関する助言を得ることにより物体の再認識を行った。また、文法のマッチングが不完全な場合にも対話を円滑に進めるため、状況に応じたキーワードを設定しておくことによりユーザの発話意図をできる限り解釈した。今後の課題は、より複雑な指定物体の状況を考慮することとサービスロボットへの実装である。

参考文献

- [1] S. Wachsmuth et al.: Connecting Concepts from Vision and Speech Processing, Workshop on Integration of Speech and Image Understanding, 1999.
- [2] Y. Makihara et al.: "Object Recognition Supported by User Interaction for Service Robots", Proc. of 16th Int. Conf. on Pattern Recognition, Vol. 3, pp. 561-564, 2002.
- [3] 西村ら: "対話機能を備えたパーソナルサービスロボットの開発—多様なセンサ情報に基づく物体操作—", 第6回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 講演論文集, 2005.
- [4] 榎原ら: "ユーザとの対話を用いたサービスロボットのための物体登録", 第19回日本ロボット学会学術講演会 講演論文集, 2001.
- [5] 榎原ら: "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, Vol. J87-D-II, No. 2, pp. 629-638, 2004.
- [6] 井本ら: "インタラクティブビジョンにおいてユーザから有用な助言を得るための手法", 動的画像処理実用化ワークショップ 2006 講演論文集, 2006.