

対話を用いた物体認識のための候補選択戦略

楨原 靖[†] 白井 良明^{††} 島田 伸敬^{††}

[†] 大阪大学 産業科学研究所 複合知能メディア研究分野 〒 567-0047 大阪府茨木市美穂ヶ丘 8-1

^{††} 立命館大学 情報理工学部 知能情報学科 〒 525-8577 滋賀県草津市野路東 1-1-1

E-mail: [†]makihara@am.sanken.osaka-u.ac.jp, ^{††}{shirai,shimada}@ci.ritsumei.ac.jp

あらまし 本論文では、対話を用いた物体認識システムにおいて、複数の物体候補が認識結果として得られた場合の、ユーザにとって負担の少ない候補選択戦略について述べる。最初に候補を認識し易くするための候補表示方法と一回当たりの選択肢数を減らすための階層的選択法を導入して、各手法によってどの程度選択が容易になるかを主観テストにより検証する。次に、これらの手法を用いて選択にかかる時間を最小化する戦略を提案し、候補選択時間をユーザの表示認識時間と音声対話時間に分けて、各々の定式化を行う。最後に、候補選択時間に基づく戦略とユーザの主観による戦略の比較を行う。

キーワード 階層選択, 決定木, 物体認識, 音声対話

Strategy of Candidate Choice for Interactive Vision

Yasushi MAKIHARA[†], Yoshiaki SHIRAI^{††}, and Nobutaka SHIMADA^{††}

[†] Department of Intelligent Media, The Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, JAPAN

^{††} College of Information Science and Engineering, Ritsumeikan University
1-1-1, Noji Higashi, Kusats, 525-8577, JAPAN

E-mail: [†]makihara@am.sanken.osaka-u.ac.jp, ^{††}{shirai,shimada}@ci.ritsumei.ac.jp

Abstract This paper describes a strategy of object candidates choice to ease user's burdens for an interactive object recognition system when the system obtains multiple candidates as a recognition result. First, we propose several display methods so as to recognize candidates easily and hierarchical choices to reduce candidates per one choice, and verify their effectiveness by subjective tests. Next, we divide the time spent for the choice into time for recognizing displayed candidates and time for dialog via voice, and formulate each time to repute the methods quantitatively. Last, we compare the strategy based on choice time with user's subjective strategy.

Key words hierarchical choice, decision tree, object recognition, dialog

1. はじめに

近年の高齢化に伴い、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来るといった機能は重要なものである。その機能を実現するためには、指定された物体を画像を用いて認識する必要がある。

但し、サービスロボットが扱うような物体が収納されているシーンは複雑である場合が多く、誤認識を含む複数の物体候補が得られる場合がある。このような場合には、ディスプレイと音声によって認識結果をユーザに伝えて、ユーザの希望する物体を選択してもらうことが考えられる。しかし、非常に多数の候補が得られた場合には、選択時のユーザの音声対話の負担や

ディスプレイ上に表示された多数の候補を認識するための表示認識の負担が大きくなる。

ユーザの音声対話の負担を軽減する研究としては、音声対話ヘルプシステムにおいて対話回数を削減する研究[1]や、交通案内システムにおいて対話時間を短縮する研究[2]等がある。

これらに加えて、ユーザが候補を選択し易いようにディスプレイ上に認識結果を表示することで、表示認識の負担を軽減することも必要である。

また、他のユーザの負担の軽減方法としては、選択を階層化して一回当たりの選択肢を減らすという方法もある。選択候補の階層化の例としては、コンピュータのアプリケーションソフトウェアにおけるメニュー選択が挙げられる。この場合はメニュー内容が決まっているので、人間があらかじめ機能毎に分

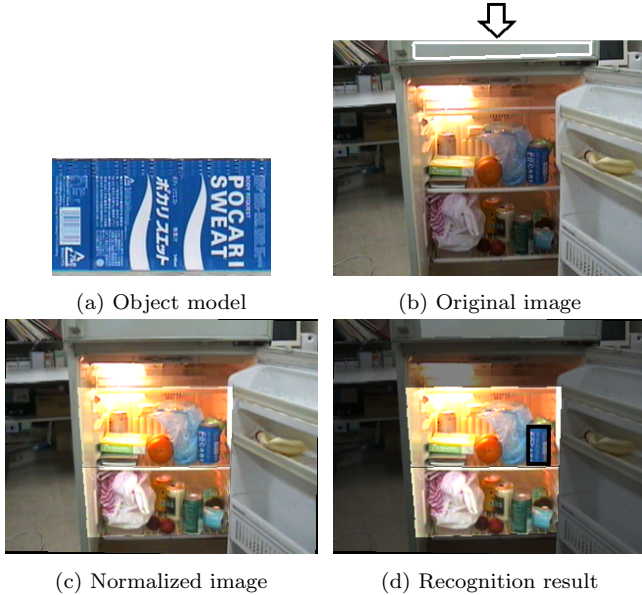


図 1 物体認識の概要

Fig. 1 Overview of object recognition

類して階層化することができる。

これに対して、物体認識システムの場合は、得られる物体候補が状況によって変化するので、認識結果に応じて適切な階層化を行う必要がある。

本論文では、冷蔵庫内のシーンにおける飲料や果物を認識対象として、最初にユーザにとって選び易い候補表示方法や階層化手法を提案する。また、それらの表示方法や階層化を利用して、候補から一つの正しい物体を選択するのにかかる時間を最小化するような候補選択の計画方法について述べる。

最後に、本論文の構成を以下に示す。2. では本研究で対象とする対話を用いた物体認識システムについて述べる。3. ではユーザにとって選び易い候補表示方法や階層化について述べ、4. ではその表示方法や階層化を利用した場合の選択時間の定式化を行う。最後に 5. で候補選択時間に基づく候補選択戦略とユーザの主観による戦略の比較を行い、6. で結論を述べる。

2. 対話を用いた物体認識

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、缶・瓶・ペットボトル等の物体モデルを作成しておく [3]。次に、認識時の照明条件の変化に対応するために、参照色（ここでは図 1(b) の矢印で指し示した白枠で囲まれた冷蔵庫の扉の色）を用いて原画像（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)） [4]。ユーザに音声で取って欲しい物体を指定してもらうと、対応する物体モデルを用いて正規化画像に対して、可能な限り自動認識を試みる [5]。認識結果として物体候補（図 1(d) の黒枠）が得られると、それをディスプレイと音声を通してユーザに伝える。ここで、候補が複数ある場合や認識に失敗した場合には、ユーザとの対話を通して認識の補助となるような情報を得て、その情報をもとに候補限定や再認識を試みる [6]。

また、これまで扱っていた缶・瓶・ペットボトルといった定形物体に加えて、大きさや色にばらつきをもつ楕円体近似可能

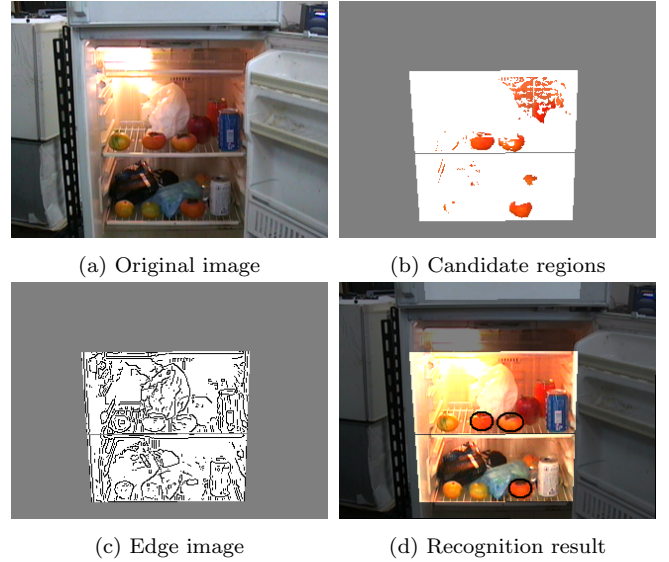


図 2 果物の認識

Fig. 2 Recognition of fruit

な果物についても認識対象とする。最初に原画像（図 2(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当てはまる候補領域（図 2(b)）を抽出する。次に、エッジ画像（図 2(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 2(d) の黒枠）とする。

一方、本研究で用いるサービスロボットは、音声対話や移動マニピュレータを用いた物体操作も行うが、その詳細についてはそれぞれ文献 [7] [8] を参照されたい。

3. 候補選択における表示方法と階層化

本節では、最初に 3.1 でユーザにとって見易い候補の表示方法について述べ、次に 3.2 で候補が多数存在するとき一回当たりの選択肢を減らすための階層的選択について述べる。

3.1 候補の表示方法

認識結果の候補を表示する際には、以下の点について考慮する必要がある。

- (1) 候補同士を区別するためのラベル付け
- (2) 候補の表示順序

3.1.1 候補同士を区別するためのラベル付け

候補同士を区別するために以下の四つの方法を導入する。

- L_{no}

これはラベル付けせずに候補を表示する方法である（図 3(a)）。ユーザは「下の棚の左から二番目のものを取って」や「柿の左上にある蜜柑を取って」というように場所の情報を与える必要がある。このような選択の仕方は、ユーザにとって負担が大きく、かつ音声対話システムにとっても文章が長くなることによる誤認識の恐れがあるため望ましくない。

- L_{ctr}

これは候補の輪郭を色分けして表示する方法である（図 3(b)）。この場合、ユーザに対して「何色で囲まれた候補ですか？」と尋ねることで、「青色」や「赤色」のような色名で選択してもらうことが可能になる。但し、一般的なユーザが区別して指定で

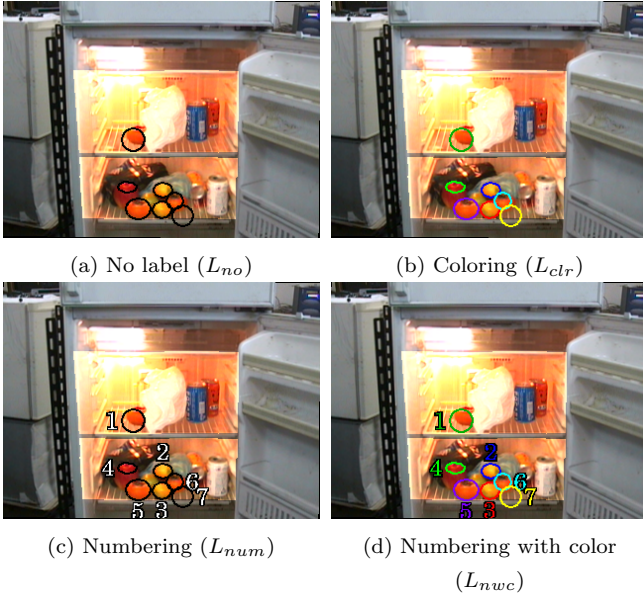


図 3 候補同士を区別するためのラベル付け

Fig. 3 Labeling to distinguish candidates

きる色の数には限りがあるため、候補数が多すぎる場合には色分け表示を使うことができない。

- L_{num}

これは候補の近くの空き領域に番号を表示する方法である (図 3(c))。ここで、空き領域の設定方法は以下の通りである。

- (1) 数字を表示するのに必要な大きさの矩形を決定して、それを候補の外接矩形の上下左右に隣接するように各々設定
- (2) 各矩形内に含まれる他の候補の面積割合を計算して、その割合が最も小さいものを空き領域として採用
- (3) 面積割合が等しい矩形が存在する場合は、上・左・右・下の順で優先順位を付けて空き領域を決定

このようなラベル付けをした場合、ユーザに対して「何番の候補ですか？」と尋ねることで、「2 番」や「3 番」のように番号で選択してもらうことが可能になる。番号付けの場合は、色分けとは異なり候補数に対する制限が無いという利点がある。しかし、表示候補数が増加して密集しているような場合には、番号同士が重なってしまうことや番号と候補との対応関係が一見して分かり辛くなる^(注1)という問題がある。

- L_{nwc}

これは、候補の輪郭を色分け表示した上で、その色と同じ色で番号を表示する方法である (図 3(d))。候補と番号の対応付けを明確にするという方法がある。しかし、番号同士が重なったときの見辛さの問題は残る。

よっていずれの方法にしても一長一短が存在することになる。

3.1.2 候補の表示順序

図 3 で挙げたような認識結果をユーザに対して静的に表示すると、原画像の一部が候補の輪郭や番号によって隠されたままになるため、システムが表示している候補の輪郭が本当に正しいかどうかをユーザが判断することが困難になる。そこで、原画像と認識結果画像を交互に表示して、ユーザの確認を補助す

(注1)：例えば、図 3(c) の下の棚の一番右の候補に対する番号が 6 番か 7 番かが一見して分かり辛い。

表 1 主観テストに用いる表示方法と評価値

Table 1 Display method and value for subjective test

記号	ラベル	表示順序	評価値	評価内容
			0	極めて選び難い
D_1	L_{no}	B_{sim}	1	かなり選び難い
D_2	L_{clr}	B_{sim}	2	選びにくい
D_3	L_{num}	B_{sim}	3	やや選び難い
D_4	L_{nwc}	B_{sim}	4	どちらかといえば選び難い
D_5	L_{no}	B_{turn}	5	普通
D_6	L_{clr}	B_{turn}	6	どちらかと言えば選び易い
D_7	L_{num}	B_{turn}	7	やや選び易い
D_8	L_{nwc}	B_{turn}	8	選び易い
D_9	L_{no}	B_{cnf}	9	非常に選び易い
			10	極めて選び易い

ることを考える。ここで、暗い照明条件下での原画像を表示するとユーザによる確認が困難になるため、2. で述べた基準照明へ変換した正規化画像を原画像の代わりに用いる。

ここでは、表示順序について以下の三つの方法を導入する。

- B_{sim} すべての候補を同時に表示した認識結果画像と正規化画像を交互に表示する方法
- B_{turn} 各候補を個別に順次表示する方法
- B_{cnf} 候補の一つを表示して、ユーザに対して逐次確認することを繰り返す方法

順次表示 B_{turn} を用いると、番号付け L_{num} の問題点で取り上げられた候補と番号の対応付けの問題点が無くなるという利点がある。但し、同時表示 B_{sim} の場合と比較して、全候補を表示し終わるまでの時間がかかるため、ユーザによっては煩わしく感じる可能性もある。この点を改善するために表示切り替え時間 T_{blk} を短くするという方法も考えられるが、あまりに T_{blk} が短いとユーザが一巡目の表示で候補を確認しきれずに、二巡目以降で確認しなければならない場合があり、結局時間がかかってしまうということも考えられる。特に、体の不自由な年配の方が本システムを利用することを考えると、あまりに速く移り変わる表示は当惑する原因ともなるため、 T_{blk} は一巡目の表示で十分に確認できる程度の長さしておく必要がある。

一方、逐次確認 B_{cnf} は、一つの候補に対してのみ注目して、それが正しいか正しくないかということのみを評価すれば良いため、一回当たりの選択にかかる労力は非常に小さい。また、この場合は「はい」または「いいえ」で返答するため、ラベル付けは不要 (L_{no}) である。但し、逐次確認表示をする場合は、同時表示や順次表示に比べて一般的に時間がかかるという点が問題となる。

3.1.3 表示方法に関する主観テスト

上記で述べた表示方法が、実際にユーザにとってどのように感じるかを確かめるために、図 3 の画像と候補を用いて、被験者 6 人に対して表 1 左側の 9 通りの表示方法に関する主観テストを行った。

また、各表示手法のテスト順序による依存性を軽減するために、被験者にはあらかじめ 9 通りすべての表示方法を見てもらい、どのような表示方法があるかを知った上で、もう一度最初

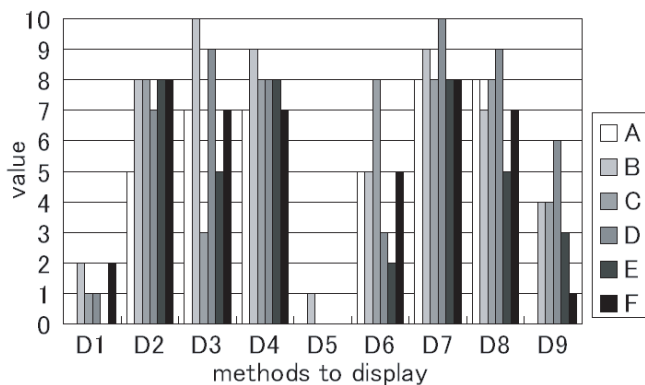
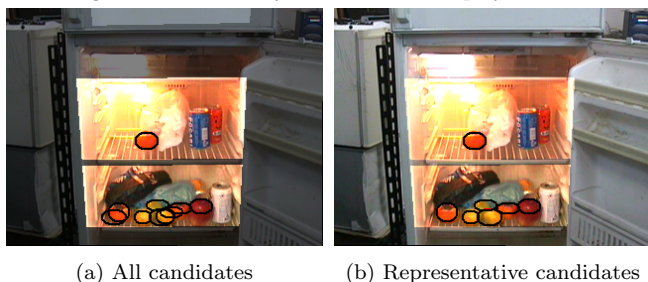


図 4 表示手法に関する主観テストの結果

Fig.4 Result of subjective tests for diaplay method



(a) All candidates (b) Representative candidates

図 5 代表候補による階層化

Fig.5 Hierarchy by representative candidates

から順に各表示方法の主観テストを行い、評価を受けた。評価には表 1 右側に示す 11 段階の絶対評価を用いた。

主観テストの結果を図 4 に示す。被験者によって多少の個人差が存在するが、表示方法としては D_4 (色分けと番号付けの両方をして同時表示) や D_7 (番号付けして順次表示) 等が高い評価を受けている。よって、システムとしてはこれらの高い評価を受けた表示方法をいくつか実装しておき、状況に適した表示方法を使うことが望ましい。

3.2 候補選択の階層化

3.1 では、表示すべき候補が決まっている場合の表示方法について扱ったが、表示すべき候補数が多い場合には、どのような表示手法を用いてもユーザーの選択の負担は大きいものとなる。特に大きさや色にばらつきのある果物を認識する場合には、他の物体も一緒に誤検出して図 5(a) のように候補が多数になることがある。このような場合には、最初は候補を大まかに絞るような選択をユーザーにしてもらい、その後、絞られた候補をユーザーに表示して選択してもらうことで一回当たりの選択肢数を減らす階層的選択を行うことが有効である。以下では、その選択の階層化について述べる。

3.2.1 棚による階層化

単純な階層化としては、上下の棚による階層化が挙げられる。図 5(a) のように片方の棚に候補が偏っている場合にはあまり有効ではないが、上下の棚共に多数の候補が存在する場合には有効であると考えられる。

3.2.2 大まかなグループによる階層化

物体が密集して置かれているような場合には、大きな一塊の候補領域が得られるが、それをいくつかの小さな領域の分割しておき、最初にその領域から選択してもらい、次に選択された



(a) Segmentation by position (b) Segmentation by color

図 6 候補領域の分割方法

Fig.6 Segmentation of candidate regions

表 2 階層的選択に関する主観テストの項目

Table 2 Items of subjective test for hierarchical choice

記号	変更前	変更後
H_1	直接選択	代表候補による階層選択
H_2	代表候補による階層選択	色で分割した大まかなグループ(領域レベル)と代表候補による階層選択
H_3	二分法で分割した大まかなグループによる階層選択	色で分割した大まかなグループによる階層選択

領域に含まれる候補を表示するという方法が考えられる。

このような大きな候補領域の分割方法としては、図 6(a) のような位置による二分法を用いた分割がある。しかし、様々な果物が集まってできた候補領域に対しては、図 6(b) のように色に基づく分割を行った方が物体の種類毎に分割される可能性が高いため望ましい。

また、このような領域レベルでのグルーピングの他に、物体候補レベルでのグルーピングも考えられる。特に色によるグルーピングを行う場合には、候補が隣接しているかどうかにかかわらず、物体の種類毎(例えば緑系の蜜柑、黄色系の蜜柑、誤検出した柿等)にグルーピングすることができ、後に述べるユーザによる選択の委譲を行い易くするという利点がある。

3.2.3 代表候補による階層化

図 5(a) の候補に着目すると、選択時の見辛さの主な原因となっているのは互いに重なっている候補であると考えられる。そこで、互いに重なった候補がある場合には、認識結果の信頼性が低い方の候補を、信頼性の高い方の候補の代替候補として扱い、代替候補を一時的に候補から外しておく。この操作を重ねている候補が無くなるまで行い、残った候補を代表候補(図 5(b))とする。そして、最初に代表候補を表示して「この中に正しい物体があれば選択して下さい。」と言って、選択されれば終了し、正しい物体が無いと言われれば、代替候補を表示して選択してもらう。この階層化により、ある程度の選択効率を保ったままで候補表示を見易くすることができる。

3.2.4 階層的選択に関する主観テスト

上記で述べた階層的選択の有効性を確かめるために、図 3 の画像を用いて、表 2 の階層選択に関する変更内容を適用した場合の改善度合いに関する主観テストを、被験者 6 人に対して行った。

主観テストの結果を図 7 に示す。評価尺度としては、5 を変えないとして、4 以下を改悪、6 以上を改善とする 11 段階の相対評価を用いた。

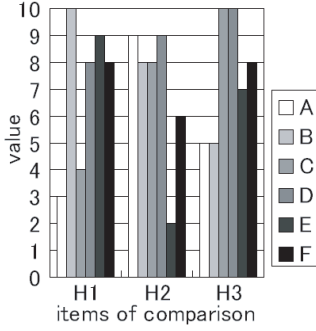


図 7 階層表示に関する主観テストの結果

Fig. 7 Result of subjective test for hierarchical display

結果より、色に基づく分割に関してはいずれの被験者に関しても「5. 変わらない」以上の評価を得ており、効果的であると考えられる。一方、代表候補による階層化に関しては、多くの被験者が改善されたと感じているものの、一部の被験者で悪化したという評価を受けた。これは、選択回数が増えることによって煩わしさが増加したことが主な原因である。また、大まかなグループによる階層化についても多くの被験者が改善されたと評価したものの、一名の被験者は悪化したという評価を下している。これは、物体を取って欲しいと頼んだにも拘わらず、領域を表示されて違和感を受けたためである。

4. 候補選択時間の定式化

本節では 3. で述べた表示方法や階層化を利用した場合の候補選択にかかる時間を定式化する。ここで候補選択時間は、最初に認識結果を表示してから最終的に候補を決定するまでに要した時間として定義する。但し、ユーザは認識結果が表示される前に、冷蔵庫内の状況を映したディスプレイを見てどの物体を選択するかを決めておくものとする。また、ユーザが選択する物体は認識結果の候補に必ず含まれているものとする。

今回は、表示方法の内、表示順序 B_{sim} , B_{turn} を用いる場合にはラベル付けとして L_{ctr} , L_{num} のいずれかを用い、 B_{cnf} を用いる場合は L_{no} を用いる。結局、用いる組み合わせとしては、表 1 左側で定義した、 $\{D_2, D_3, D_6, D_7, D_9\}$ の 5 通りとなる。また、階層としては、上下の棚の選択、大まかなグループ選択 (物体候補レベル)、代表候補選択、代替候補選択を考える。

4.1 決定木の作成

候補選択の階層を、選択をノードで表現した決定木として表現する。本研究で考慮する階層を全て入れた場合の決定木の例を図 8 に示す。最上位のノードは上下の棚の選択ノードであり、その下位のノードは棚毎の大まかなグループの選択ノードであり、候補の分割方法 (位置による等分法または色による分割法) によって下位の決定木の構造が変化する。更にその下位のノードは代表候補の選択ノードであり、その下位のノードは代替候補の選択ノードである。

実際には、これらの階層による選択ノードの有無を組み合わせることにより様々な決定木 tr_i を作成し、その中から候補選択時間の期待値が最小になる決定木 tr^* を採用する。

$$tr^* = \arg \min_i \{\bar{T}_{tr}(tr_i)\} \quad (1)$$

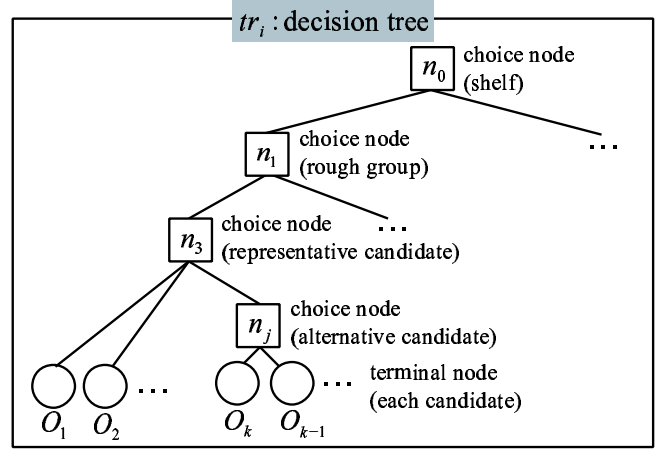


図 8 候補選択の階層に対する決定木

Fig. 8 Decision tree for hierarchy of candidate choice

ここで、 $\bar{T}_{tr}(tr_i)$ は決定木 tr_i によって候補を選択するのにかかる時間の期待値である。

4.2 選択時間の分解

候補選択時間の期待値 $\bar{T}_{tr}(tr_i)$ を直接定式化するのは困難であるので、 $\bar{T}_{tr}(tr_i)$ を個別要素に分解して考える。まず、物体候補数 N_O 、ユーザが各候補 O_k を選択する事前確率 $p(O_k)$ 、その物体を選択するのにかかる時間の期待値 $\bar{T}_{obj}(tr_i, O_k)$ を用いると、 $\bar{T}(tr_i)$ は以下ようになる。

$$\bar{T}_{tr}(tr_i) = \sum_{k=1}^{N_O} p(O_k) \bar{T}_{obj}(tr_i, O_k) \quad (2)$$

また、 $\bar{T}_{obj}(tr_i, O_k)$ は、決定木 tr_i において候補 O_k に辿り着くまでに通る N_{node} 個の選択ノード $\{n_1, \dots, n_{N_{node}}\}$ と、ノード n において選択にかかる時間の期待値 $\bar{T}_{node}(n)$ を用いると、以下のように表現できる。

$$\bar{T}_{obj}(tr_i, O_k) = \sum_{l=1}^{N_{node}} \bar{T}_{node}(n_l) \quad (3)$$

更に、各ノードでの選択時間 T_{node} は、以下の二つの時間の和で表現される。

- 表示認識時間 T_{rcg} : 表示された候補から正しい候補を認識して、その候補に対応するラベル (番号や色) を認識するのにかかる時間
- 音声対話時間 T_{dlg} : ユーザが返答を始めてからシステムとの音声対話を経て選択肢を決定するのにかかる時間

但し、逐次確認を行う場合には、一つのノードに対して選択肢毎に表示と対話を繰り返すので、表示認識時間 T_{rcg} と対話時間 T_{dlg} は選択肢毎にかかった時間を各々合計したものとして定義する。

以下ではこれらの時間をモデリングする。

4.3 表示認識時間と表示認識率のモデリング

表示認識時間 T_{rcg} は候補表示方法や候補数に依存する。また、表示された候補が複雑な場合にはユーザが表示を誤認識することもある。ここでは、表示認識成功率を p_{rcg_tp} 、誤認識率を $p_{rcg_fp} (= 1 - p_{rcg_tp})$ として、表示認識時間の期待値 $\bar{T}_{rcg}$ と共に表示方法毎にモデリングする。



(a) Numbering for overlapping candidates

(b) Coloring for many candidates

図 9 多数候補に対するラベル付け

Fig.9 Labeling for many candidates

・同時表示の場合 (D_2, D_3)

ユーザはあらかじめ選択すべき候補をディスプレイ上で確認しているので、多数の候補が表示されても表示認識はその一つの候補についてのみ行う。

但し、番号付けの場合 (D_3) には、図 9(a) のように候補が多数になり候補同士が重なる場合には、候補の輪郭と番号との対応が曖昧になることや、番号が重なって見辛くなることがあるため、表示認識時間が長くなり、表示認識率が低下する。ここでは、簡単のためこれらが候補数 m に依存するとし、その依存関係 $\bar{T}_{rcg} = T_{rcg-ol}(m), p_{rcg-tp} = p_{rcg-ol-tp}(m)$ を実験的に決定する。

一方、色分け (図 9(b)) の場合 (D_2) には、このような問題は無いものとして、表示認識時間の期待値 $\bar{T}_{rcg}$ は、候補数によらずユーザの認識能力のみに依存する基本的な表示認識時間 $T_{rcg-base}$ と等しいと仮定する。

・順次表示の場合 (D_6, D_7)

ユーザが選択すべき候補が i 番目の候補であったとすると、その候補が現れるまでにかかる時間は以下になる。

$$T_{rcg} = iT_{disp-intv} \quad (4)$$

但し、 $T_{disp-intv}$ は表示切替時間である。ここでは、ユーザが一度の表示で十分に候補を確認できるように、 $T_{disp-intv}$ を基本的な表示認識時間 $T_{rcg-base}$ に合わせておく。各候補が選択される確率が等しいと仮定すると、 m 個の候補に対する表示認識時間の期待値 $\bar{T}_{rcg}$ は以下になる。

$$\bar{T}_{rcg} = \frac{m+1}{2} T_{rcg-base} \quad (5)$$

また、候補を個別表示することから、表示を誤認識することは無く、 $p_{rcg-tp} = 1$ となる。

・逐次確認の場合 (D_9)

逐次確認時の表示認識時間は、選択肢毎にかかった時間を合計したものとなるので、順次表示の場合と同様になる。

最後に、表示方法毎の $\bar{T}_{rcg}, p_{rcg-tp}$ を表 3 にまとめる。

4.4 音声認識率を考慮した音声対話時間のモデリング

音声対話時間 T_{dlg} は以下の二つで構成されているものとする。

- $T_{dlg-base}$: 色や番号を答えるユーザの発話、音声認識処理時間、システムの返答にかかる時間の和
- $T_{dlg-cnfr}$: 選択が正しいかどうかの「はい」「いいえ」で答える確認の対話にかかる時間

表 3 表示認識時間の期待値 $\bar{T}_{rcg}$ と表示認識率 p_{rcg-tp}

Table 3 Time spent for recognizing displayed candidates and recognition ratio

表示方法	表示認識時間の期待値 $\bar{T}_{rcg}$	表示認識率 p_{rcg-tp}
D_2	$T_{rcg-base}$	1
D_3	$T_{rcg-ol}(m)$	$p_{rcg-ol-tp}(m)$
D_6	$\frac{m+1}{2} T_{rcg-base}$	1
D_7	$\frac{m+1}{2} T_{rcg-base}$	1
D_9	$\frac{m+1}{2} T_{rcg-base}$	1

ここでは、簡単のため $T_{dlg-base}, T_{dlg-cnfr}$ は候補数や表示方法によらず一定である仮定して、各々実験的に決定する。

また、音声認識結果の分類とその後の対応は以下になる (括弧内は各結果が得られる確率)。

- 認識成功 (p_{dlg-tp}): 選択されたものが末端ノードの場合は、それを表示した上で確認して、最終的に候補を決定する。それ以外の場合は次の選択ノードに移る。
- 誤認識 (p_{dlg-fp}): 認識成功時と同様にして確認もしくは次の選択を行う。その際にユーザから間違いを指摘されるので、その候補を除いた選択肢からユーザに選択してもらう。
- 認識失敗 (p_{dlg-n}): 再び同じ選択肢からユーザに選択してもらう。

これらの音声認識率は、候補数 m やユーザの発話内容に影響を及ぼすラベル付け方法 L に依存する。候補数 m が増加すると発話の種類も増加するため、他の単語との誤認識率が増加する。特にラベル付け方法として色分けを用いて、図 9(b) のように似た色が用いられる場合には、例えば「黄緑色」を「緑色」と誤認識するようなことがあるために認識成功率が大きく低下する。一方、確認時の対話では「はい」「いいえ」の 2 種類のみを認識すればよいので、常に成功するものと仮定する。

また、選択の対話の成功率を考える際には、表示認識率についても考慮する必要がある。ここでは、表示認識結果と音声認識結果の事象が互いに独立であると仮定して、二つを総合した場合の成功率 p_{tp} 、誤り率 p_{fp} 、失敗率 p_n について考える。成功となるのは、表示認識と音声認識の両方で成功した場合が主で、誤りとなるのは、表示認識に成功して音声認識で誤認識する場合と、表示認識で誤認識して音声認識で認識成功または誤認識する場合の二つの場合が主であるとみて、各々以下のように定義する。

$$p_{tp} = p_{rcg-tp} p_{dlg-tp} \quad (6)$$

$$p_{fp} = p_{rcg-tp} p_{dlg-fp} + p_{rcg-fp} (p_{dlg-tp} + p_{dlg-fp}) \quad (7)$$

$$p_n = 1 - (p_{tp} + p_{fp}) \quad (8)$$

以下では、上記の確率を考慮した音声対話時間の定式化を行う。

・同時表示または順次表示の場合 (D_2, D_3, D_6, D_7)

候補数 m の末端ノードからの選択ノードに対する音声対話時間の期待値 $\bar{T}_{dlg}(m)$ は、選択失敗を何度か繰り返した後に成功して最終確認をして終了する場合と、選択を誤って訂正された後に選択肢が一つ減った選択に移る場合に分けて考えると、以下になる。

$$\begin{aligned}
\bar{T}_{dlg}(m) &= \sum_{i=1}^{\infty} (p_n)^{i-1} p_{tp} (iT_{dlg_base} + T_{dlg_cnf}) \\
&+ \sum_{i=1}^{\infty} (p_n)^{i-1} p_{fp} (iT_{dlg_base} + T_{dlg_cnf} + \bar{T}_{dlg}(m-1)) \\
&= \frac{T_{dlg_base}}{p_{tp} + p_{fp}} + T_{dlg_cnf} + \frac{p_{fp}}{p_{tp} + p_{fp}} \bar{T}_{dlg}(m-1) \\
T_{dlg}(1) &= T_{cnf}
\end{aligned} \tag{9}$$

また、末端ノード以外からの選択ノードでは、認識成功時の確認はせずに、誤認識をした場合には次の選択ノードで訂正してもらうので、以下ようになる。

$$\bar{T}_{dlg}(m) = \frac{T_{rcg_base}}{p_{tp} + p_{fp}} + \frac{p_{fp}}{p_{tp} + p_{fp}} \{\bar{T}_{dlg}(m-1) + T_{dlg_cnf}\} \tag{10}$$

・逐次確認の場合 (D_9)

最大で m 回の確認をすることになるので、音声対話時間の期待値は以下ようになる。

$$\bar{T}_{dlg}(m) = \frac{m+1}{2} T_{dlg_cnf} \tag{11}$$

4.5 ユーザが選択を委譲する場合

これまでユーザが末端ノードに辿り着くまで逐一選択していくことを仮定していたが、選択したノードの先にある候補が全て同じ種類の物体である場合には、ユーザが途中で選択をシステムに委譲することがある。その場合には、そのノードの先での選択時間を省くことができるため、ユーザにとっては負担を軽減する有効な手段となる。

例えば、認識結果として得られた候補が全て同じ種類であった場合には、最初から選択を委譲することもできる。また、上下の棚によって選択する場合や、位置や色によって大まかなグループに分ける場合、代表候補から選択する場合にも、それが同じ種類で構成されている場合には同様に選択を委譲することができる。

一般にユーザが物体を収納する場合、同じ種類の物体は同じ場所に集めて置くことがあるため、位置による等分法で得られるグループが同じ種類の物体で構成される可能性がある。しかし、位置による等分法では、種類毎の量に偏りがある場合や整頓せずに様々な場所に散らばって収納している場合には同じ種類でグループ分けされないこともある。これに対して、色によるグループ分けを行う場合には、同じ種類のものが同じような色を共有している可能性が高いので、収納状況によらず同じ種類にグループ分けされる可能性が高くなる。よって、色によるグループ分けを行うと他のグループ分けに比べてユーザによる委譲が生じやすく、負担が軽減されるものと考えられる。

5. 候補選択時間に基づく戦略とユーザの主観による戦略の比較

本節では、候補選択時間に基づく戦略とユーザの主観による戦略が候補数に応じてどのように変化するかを比較する。今回は、戦略の種類として全候補から直接選択する単層選択と大まかなグループ分けによる二層選択の二種類を考える。定性的には候補数が増えるにしたがって単層選択から二層選択へ移行

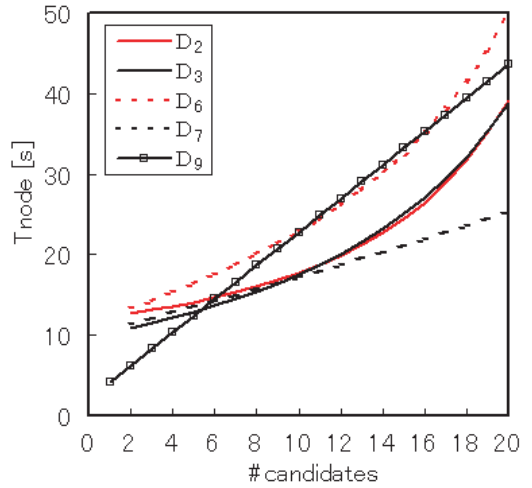


図 10 候補数に対する選択時間の期待値

Fig. 10 Relation between #candidates and expected choice time

するので、ここでは単層選択を採用する場合の最大の候補数 m_{max} について注目する。また、表示方法としては前節で取り扱った 5 通りの組み合わせを用いる。

5.1 候補選択時間の定式化に用いたパラメタの決定

まず、ユーザの基本的な表示認識時間 T_{rcg_base} と等しいと仮定した順次表示における表示切り替え間隔 T_{disp_intv} を求める。これは、実際に候補の順次表示を T_{disp_intv} を変化させながら見てもらい、十分に見落とさない長さを被験者毎に決定してもらった。また、同時表示の番号付けの場合の表示認識時間 $T_{rcg_ol}(m)$ については、いくつかシーンに対して、候補を表示してから発話を始めるまでにかかった時間を計測して、候補数 m との関係求めた (ここでは、簡単のため線形の関係とした)。加えて、表示認識率 p_{rcg_tp} についても、同じ実験においてユーザが候補を誤って指定した回数を用いて求めておく。最後に、音声認識率については音声認識ソフトウェア Julius [9] を用いた音声認識実験によって求めた。

5.2 候補選択時間のシミュレーション結果

このようにしてある被験者に対して求めたパラメタを用いた場合の単層選択の候補選択時間と候補数の関係を図 10 に示す。これより、候補数が少ないときには逐次確認が有利で、候補数が増えるに従って同時表示の番号付けや順次表示の番号付けが有利になっている。実際には、システムは候補数に応じて候補選択時間を最小にする表示方法を採用する。

次に、大まかなグループによる二層選択を行った場合の選択時間と単層選択の場合の選択時間の比較を図 11 に示す。この場合、候補数 $m_{max} = 7$ までは単層選択の方が選択時間が短く、それを超えると二層選択の方が選択時間が短くなっている。即ち、 m_{max} より候補数が多いか少ないかで、選択を階層化するかどうかが決まることになる。

5.3 ユーザの主観による戦略との比較

選択時間に基づく戦略と比較するために、ユーザに候補数がいくらまでであれば単層選択の方が負担が少ないかを尋ねる主観テストを行った。単層選択を採用する最大候補数 m_{max} の比較結果を図 13 に示す。これより、被験者 E を除いて、ユーザ

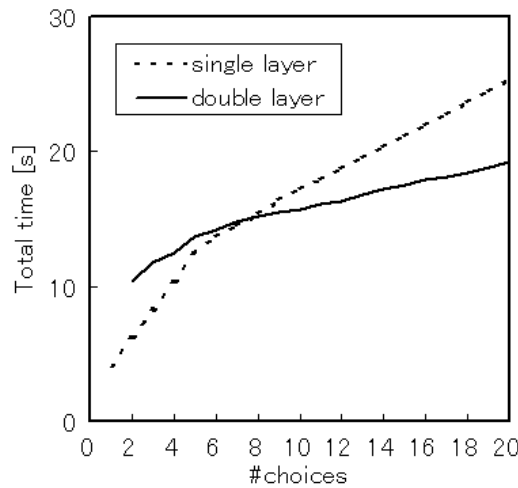


図 11 単層選択と二層選択の選択時間の期待値の比較

Fig. 11 Comparison of expected time between single-layer choice and two-layers choice

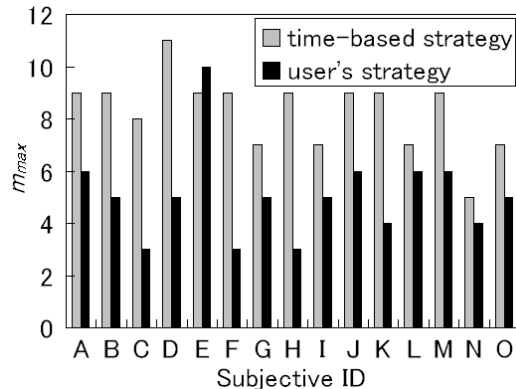


図 12 単層選択の最大候補数に対する主観テスト

Fig. 12 Subjective test for #max candidates for single-layer choice

がより少ない候補数で二層選択への切り替えを望んでいることが分かる。よって、ユーザの負担を軽減するためには、候補選択時間のみならず、選択の際に要する集中力といった精神的な労力についても考慮する必要があると考えられる。

5.4 表示切替時間と戦略の関係

ユーザが設定した表示切替時間 T_{disp_intv} と単層選択の最大候補数 m_{max} の関係を図 13 に示す。これより、幾人かの例外的な被験者がいるものの、全体としては表示切替時間が長いユーザ程、単層選択の最大候補数が少ないことが分かる。今回の主観テストは 20 代の健康者を被験者として行ったが、実際にサービスロボットを用いるような高齢者の方々は視覚機能の衰えなどがあるため、表示切替時間をより長く設定する必要がある、単層選択の最大候補数もより少なくなる、即ち二層選択の方をより好むようになることが予想される。

6. む す び

本論文では、対話を用いた物体認識システムにおいて、複数の物体候補が認識結果として得られた場合の、ユーザにとって負担の少ない候補選択計画について述べた。最初にラベル付けや表示順序を考慮した候補の表示方法や候補選択の階層化について述べ、各手法に対して主観テストを行い、その効果を確か

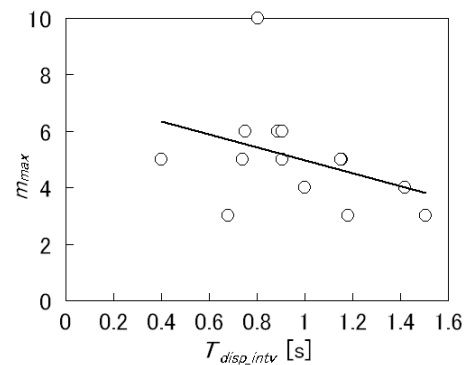


図 13 表示切替時間と単層選択の最大候補数の関係 (実線: 線形近似)

Fig. 13 Relation between display interval time and #max candidates for single-layer choice (solid line: linear approximation)

めた。次に、選択の階層を選択ノードによる決定木によって表現して、候補選択にかかる時間の定式化を行った。各選択ノードでかかる時間は表示認識時間と音声対話時間に分けて考え、表示認識時間については候補の重なりによる対応の曖昧さ等を考慮して定式化し、音声対話時間については音声認識率を考慮して定式化した。最後に、候補選択時間に基づく戦略とユーザの主観による戦略の比較実験を行い、ユーザの負担の尺度として時間に加えて選択に要する労力等を導入する必要があることを述べた。

今後の課題は以下の通りである。

- 一般ユーザを想定した主観テスト
- 選択時の集中力等を考慮した候補選択戦略の考案

文 献

- [1] 伊藤 亮介, 駒谷 和範, 河原 達也, "機器操作マニュアルの知識と構造を利用した音声対話ヘルプシステム", 情報処理学会論文誌, Vol. 43, No. 7, pp. 2147-2154, 2002.
- [2] 上野晋一, 駒谷和範, 河原達也, 奥乃博, "京都市バス運行情報案内システムの試験評価とユーザモデルの導入", 人工知能学会全国大会論文集, 2C2-03, 2003.
- [3] 横原靖, 滝澤正夫, 白井良明, 三浦純, 島田伸敬, "ユーザとの対話を用いたサービスロボットののための物体登録", 第 19 回日本ロボット学会学術講演会 講演論文集, CD-ROM, 2001.
- [4] 横原靖, 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, Vol. J87-D-II, No. 2, pp. 629-638, Feb. 2004.
- [5] Y. Makiyara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada, "Object Recognition Supported by User Interaction for Service Robots", Proc. of 16th Int. Conf. on Pattern Recognition, Vol. 3, pp. 561-564, Quebec, Canada, Aug. 2002.
- [6] Y. Makiyara, M. Takizawa, Y. Shirai, J. Miura, and N. Shimada, "Object Recognition Supported by User Interaction for Service Robots", Proc. of 5th Asian Conf. on Computer Vision, Vol. 2, pp. 719-724, Melbourne, Australia, Jan. 2002.
- [7] 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, 2002.
- [8] 矢野恵生, 三浦純, 白井良明, 島田伸敬, "移動マニピュレータのための直観的作業教示法", 第 20 回日本ロボット学会学術講演会 講演論文集, CD-ROM, 2002.
- [9] "大語彙連続音声認識システム Julius", <http://julius.sourceforge.jp/>