社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE
NLC2004-124, PRMU2004-206(2005-2)

対話を用いた物体認識のための候補選択戦略

槇原 靖† 白井 良明† 島田 伸敬††

† 大阪大学大学院工学研究科 〒1565-0871 大阪府吹田市山田丘 2-1
†† 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,shirai}@cv.mech.eng.osaka-u.ac.jp, ††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．ここでは，候補を見やすく表示することや一回あたりの選択肢数を減らすための階層的な選択を導入することで，選択にかかる時間を最小化する戦略を採る．最初にいくつかの候補表示方法と階層的選択法を提案して，次に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．最後に，これらを定量的に評価するために，候補選択時間をユーザの認識時間，返答作成時間，音声対話時間に分けて，各々の定式化を行う．

キーワード 階層選択，決定木，物体認識，対話

Strategy of Candidate Choice for Interactive Vision

Yasushi MAKIHARA†, Yoshiaki SHIRAI†, and Nobutaka SHIMADA††

† Faculty of Engineering, Osaka University Yamadaoka 2-1, Suita-shi, 565-0871 Japan
†† College of Information Science and Engineering, Ritsumeikan University Noji Higashi 1-1-1, Kusatsu-shi, 525-8577 Japan
E-mail: †{makihara,shirai}@cv.mech.eng.osaka-u.ac.jp, ††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. We take the strategy to minimize time spent for candidate choice by displaying the candidates visibly and by introducing hierarchical choice to reduce candidates per one choice. First, we propose several display methods and hierarchical choices. Next, we verify effectiveness of the methods by subjective tests. Last, we divide the time spent for the choice into user's recognition time, making answer time, and vocal dialog time, and formulate each time to repute the methods quantitatively.

Key words Hierarchical Choice, Decision Tree, Object Recognition, Dialog

1. まえがき

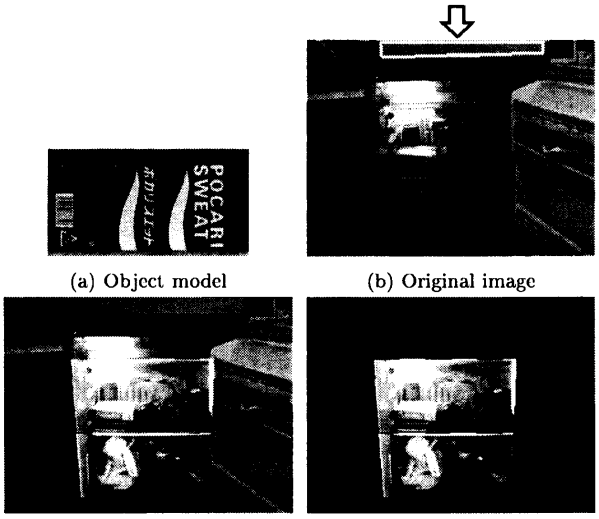
近年，高齢化社会の到来によりサービスロボットの必要性が高まっている．そのようなロボットの機能の中で，ユーザに指定されたものを持って来るという機能は重要なものである．その機能を実現するためには，指定された物体を認識してその位置を知ることが必要となるが，その方法としては画像を用いて物体を認識することが考えられる．

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，認識した物体候補が一つに定まらず，多数の候補になる場合がある．このような場合には，ユーザとの対話によって指定物体を決定する．その際には，候補の表示方法を工夫することや，選択を階層化して一回当たりの選択肢を減らすことで，ユーザの選択の負担を減らすことができる．そこで，本論文では冷蔵庫内のシーンを対象として，多数の候補が得られた場合に，様々な表示方法や階層化を利用して，候補から一つの指定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. では本研究のベースとなる対話を用いた物体認識システムの概要について述べる．3. では候補選択時に用いる表示方法や階層化の種類について述べ，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い，5. で結論を述べる．

2. 対話を用いた物体認識の概要

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 [1]．ユーザから音声対話によって取って欲しい物体を伝えられると，参



(a) Object model (b) Original image

(c) Normalized image (d) Recognition result

図 1 物体認識の概要

Fig. 1 Overview of object recognition

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して (図 1(c)) [2], その画像を基に可能な限り自動認識を試み [3], 図 1(d) の黒枠のように認識した物体候補をディスプレイを通してユーザに表示する. 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, その情報を基に候補限定や再認識を試みる [4]. また, 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の操作も行うが, その詳細についてはそれぞれ文献 [5] [6] 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やす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に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.

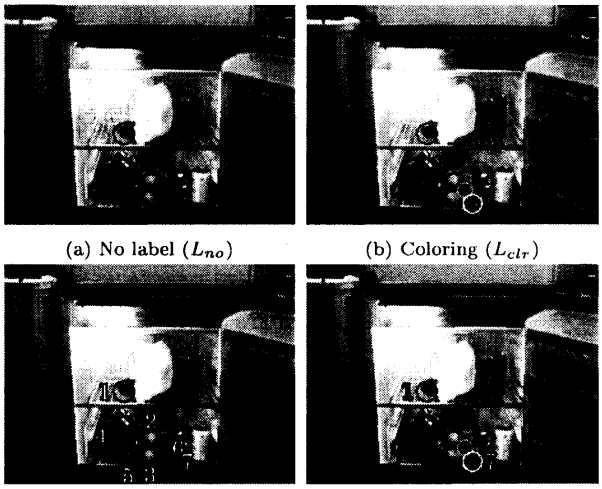
(1) 候補同士を区別するためのラベル付け

(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

システムが認識結果として表示している候補をユーザが区別して指定する際に, 図 2(a) のような候補のラベル付けをしない単純な結果表示をした場合 (以下 L_{no} とする), 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. そこで候補を区別するためのラベルを付け, 場所以外の情報でも候補を指定できるようにする.

ラベル付けの一つの方法としては, 図 2(b) にあるように候補を色分け表示すること (以下 L_{clr} とする) が考えられる. 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.



(a) No label (L_{no}) (b) Coloring (L_{clr})

(c) Numbering (L_{num}) (d) Numbering with color (L_{nwc})

図 2 候補同士を区別するためのラベル付け

Fig. 2 Labeling to distinguish candidates

別の方法としては, 図 2(c) にあるように候補の近くの空き領域に番号を表示すること (以下 L_{num} とする) が考えられる. この場合, 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.

これに対して, 図 2(d) のように候補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こと (以下 L_{nwc} とする) で, 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 2 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枠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範囲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ることを考える. ここで, 暗い照明条件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. で述べたような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.

このように画像を点滅させて表示させる場合, 図 2 に示したように認識結果の候補を同時に表示する方法 (以下 B_{sim} とする) に加えて, 各候補を個別に順次表示するという方法 (以下 B_{turn} とする) も考えられる. この場合は, 番号付け L_{num} の

(注1): 例えば, 図 2(c) の下の棚の一番左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 reputation value for subjective test

記号	ラベル	表示順序
D_1	L_{no}	B_{sim}
D_2	L_{clr}	B_{sim}
D_3	L_{num}	B_{sim}
D_4	L_{nwc}	B_{sim}
D_5	L_{no}	B_{turn}
D_6	L_{clr}	B_{turn}
D_7	L_{num}	B_{turn}
D_8	L_{nwc}	B_{turn}
D_9	L_{no}	B_{cnf}

評価値	評価内容
0	極めて選び難い
1	かなり選び難い
2	選びにくい
3	やや選び難い
4	どちらかといえば選び難い
5	普通
6	どちらかと言えば選び易い
7	やや選び易い
8	選び易い
9	非常に選び易い
10	極めて選び易い

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にしておく必要がある．

一方，候補の切り替え時間に追われずに，一つ一つ確実に確認する方法として，候補を一つ表示して，ユーザに対して「これですか？」というように逐次確認するという方法 (以下 B_{cnf} とする) がある．この方法であれば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 2 の画像と候補を用いて，被験者 6 人に対して表 1 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，被験者には予め 11 通り全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 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 3 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．

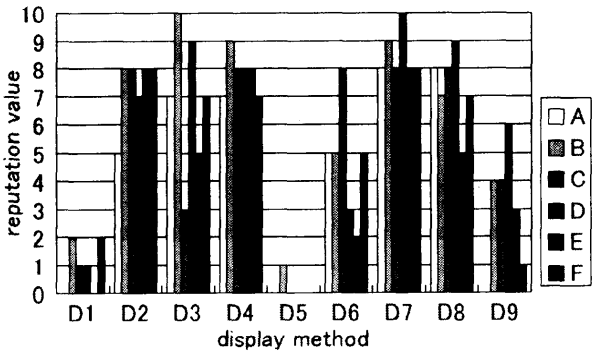


図 3　表示手法に関する主観テストの結果
Fig. 3　Result of subjective tests for diaplay method

(a) All candidates　(b) Representative candidates

図 4　代表候補による階層化
Fig. 4　Hierarchy by representative candidates

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あまりに多い場合にはどのような表示手法を用いてもユーザ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 4(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にしてもらい，その後，絞られた候補をユーザ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 4(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 5(a) のような二分法による分割がある．しかし，様々な色の果物が集まってできた候補領域に対しては，図 5(b) のようの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あるための望ましい．

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，語

(c) Segmentation by dichotomy (b) Segmentation by color

図 5 候補領域の分割方法

Fig. 5 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	変更内容
H_1	代表候補による階層化の導入
H_2	大まかなグループによる階層化 (領域レベル) を導入
H_3	大まかなグループへの分割を二分法から色に基づく分割に変更

検出した柿等) にグルーピングすることができ，次節で述べるグループ認識が可能になるという利点がある．

3.2.3 代表候補による階層化

図 4(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，認識結果の信頼性の低い方の候補を高い方の候補の代替候補として扱い，一時的に候補から外しておく．この操作を重なっ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 4(b)) とする．そして，最初に代表候補を表示して「この中に指定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，指定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 2 の画像を用いて，表 2 の階層選択に関する変更内容を行った場合に以前に比べてどの程度良くなったかという主観テストを，被験者 6 人に対して行った．

主観テストの結果を図 6 に示す．評価尺度としては，5 を変わらないとして，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．

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. 選択にかかる時間の定式化

本章では 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．表示方法の内，表示順序

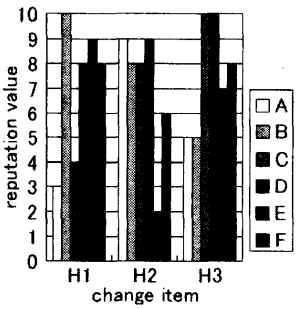


図 6 階層表示に関する主観テストの結果

Fig. 6 Result of subjective test for hierarchical display

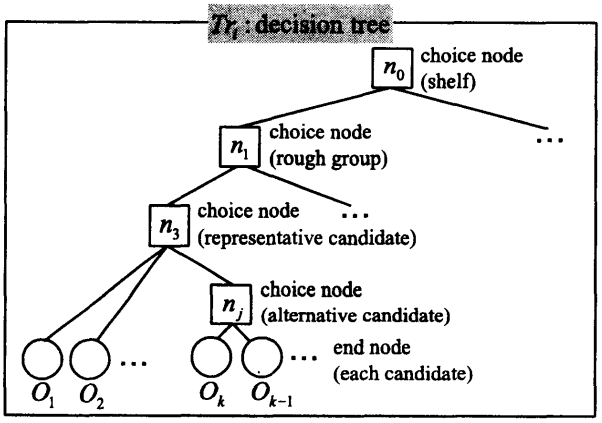


図 7 候補選択の階層に対する決定木

Fig. 7 Decision tree for hierarchy of candidate choice

B_{sim}, B_{turn} を用いる場合にはラベル付けとして L_{clr}, 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 $\{B_{sim}, L_{clr}\}$, $\{B_{turn}, L_{clr}\}$, $\{B_{sim}, L_{num}\}$, $\{B_{turn}, L_{num}\}$, $\{B_{cnf}, L_{no}\}$ 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

候補選択の過程を，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 7 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選択ノードであり，候補の分割方法 (等分法または色分割) によりそれ以下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表候補内に指定物体が無い場合の代替候補の選択ノードである．

システムは，これらの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にかかる時間の期待値を最小にする決定木 Tr^* を採用する．

$$Tr^* = \arg\min_i \{T(Tr_i)\} \tag{1}$$

ここで，$T(Tr_i)$ 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

選択時間 $T(Tr_i)$ を直接定式化するのは困難であるので，ここでは $T(Tr_i)$ 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 $T(Tr_i, O_k)$ を用いると，$T(Tr_i)$ は以下のようになる．

$$T(Tr_i) = \sum_{k=1}^{N_O} p(O_k) T(Tr_i, O_k) \tag{2}$$

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

$$T(Tr_i, O_k) = \sum_{l=1}^{N_c(Tr_i, O_k)} T_c(n_l(Tr_i, O_k)) \tag{3}$$

各ノードでの選択時間の期待値は，表示順序 B とラベル付け L を含む表示方法 $D = \{B, L\}$ に依存する．よって，ノード n で表示方法 D_s を用いた場合にかかる時間を $T_c(n, D_s)$ として，その中で最小値を与える表示方法を採用するものとすると，以下のようになる．

$$T_c(n) = \min_s T_c(n, D_s) \tag{4}$$

また，各ノードでの選択時間の期待値は，以下の三つの時間の期待値に分解できる．

- T_{rcg}: ユーザが表示された候補を見てそれらの内のどの候補が指定すべき候補であるかを認識するのにかかる時間の期待値
- T_{make_ans}: その候補を指定するための返答文(注2)を作成するのにかかる時間の期待値
- T_{dlg}: ユーザが返答してからシステムとの音声対話をして選択肢を指定するのにかかる時間の期待値

よって，ノードでの選択時間 T_c は以下のように表される．

$$T_c = T_{rcg} + T_{make\_ans} + T_{dlg} \tag{5}$$

以下ではこれらの時間について定式化していく．

4.3 候補認識にかかる時間

ユーザが候補から指定物体を認識する過程について考えてみると，認識結果として候補を同時表示した場合には，表示された候補を順に追っていき，指定物体が見つかった時点でユーザの認識が終了する．よって，同時表示の場合にはユーザのペースで認識を進めていくので，個人差が生じる．一方，順次表示の場合には一つ一つ表示される候補をユーザが目で追って行き，指定物体に辿り着いたところで認識が終了する．逐次確認表示の場合には一つ表示したものが指定物体であるかどうかを確認しながら次の候補に移っていき，指定物体に辿り着いた時点で認識が終了する．よって，順次表示や逐次確認の場合には認識にかかる時間の期待値が表示切り替え時間に依存する．

このような認識過程の下で，ユーザが m 個の候補から認識を行う場合にかかる時間を考える．まず，表示切り替え時間 T_{blk} の間にユーザが認識できる候補数 $N_{rcg/blk}$ について考える．順次表示及び逐次確認表示の場合には上記の認識過程を辿るため，$N_{rcg/blk} = 1$ となる．一方，同時表示の場合には，個人差はあるものの一度の表示で複数個の候補を認識することが可能であるので，$N_{rcg/blk} = N_{rcg/blk\_sim} (> 1)$ となる．これらを用いると順次表示及び逐次確認表示で i 番目に表示される候補，または同時表示でユーザが i 番目に見る候補を認識するのにかかる時間 $T_{rcg,i}$ は以下のようになる．

$$T_{rcg,i} = i \frac{T_{blk}}{N_{rcg/blk}} \tag{6}$$

各候補が選択される確率が等しいと仮定すると，m の候補から認識を行うのにかかる時間の期待値 T_{rcg} は，以下のようになる．

$$T_{rcg} = \frac{m+1}{2} \frac{T_{blk}}{N_{rcg/blk}} \tag{7}$$

上式の T_{rcg} と候補数 m の関係を模式的に表すと，図 8(a) のようになる．

但し，既に以前の選択ノードで指定物体の位置を認識していると，以降の選択ノードでは最初からその候補だけに注目していれば良いため，同時表示の場合には表示された時点で認識が終了し，認識時間は 0 となる．一方，順次表示や逐次確認表示の場合には，その候補が表示されるまでは指定すべき番号や色が分からないので，式 (7) で定義した時間と同じだけの時間がかかる．

4.3.1 個別認識とグループ認識

これまでは末端ノードの候補物体を直接選択する場合について考えてきたが，選択ノードが上下の棚や大まかなグループから選択するノードである場合には，そのノードでの選択肢数 m_{node} とユーザが指定物体の属するグループを認識するために見る必要のある候補数の最大値 m が一般には異なる．このことを扱うために，ユーザが指定物体の属するグループを認識する方法として以下の二つを考える．

- 個別認識

グループに関係無く個別候補 (末端ノード) から指定物体を認識して，それが属しているグループを認識する．これは，上下の棚や等分法によるグループを選択する場合に行われるものであり，候補数はそのノードの先にある末端ノードの数 $m = m_{end}$ になる．個別認識を行う場合に，T_{blk} の間に認識できる最大の候補数は N_{rcg/blk_sim} となるが，これは，同時表示以外でも，一つ一つのグループに属する候補が同時に表示されるため，グループ内では同時表示しているのと同じことになるためである．今回は簡単化のために，最大個数認識できるものと仮定して，$N_{rcg/blk} = N_{rcg/blk\_sim}$ とする．

- グループ認識

各グループの内に属する任意の一つの物体を見て，それがユーザが指定しようとしている物体と同じ種類 (例えば，蜜柑，柿，林檎等) であれば，そのグループを選択対象として認識する．これは，色分割によって蜜柑や柿といったような種類毎のグループに分かれたものから選択する場合に行われるものであり，候補数はそのノード自体の選択肢数 $m = m_{node} (< m_{end})$ に減る．また，グループ認識を行う場合の T_{blk} の間に認識できる候補数 $N_{rcg/blk}$ は，同時表示の場合には $N_{rcg/blk} = N_{rcg/blk\_sim}$，それ以外では $N_{rcg/blk} = 1$ となる．

4.4 返答文作成にかかる時間

逐次確認 B_{cnf} の場合には，「はい」または「いいえ」という

(注2): 例えば，番号付けで同時表示する場合には，指定する候補に対応する番号を認識して，その番号を答える文

返答を作成するだけであるので非常に短い時間となる．また，順次表示で番号付けをする場合 $\{B_{turn}, L_{num}\}$ にも番号を読み取るだけであるので短い時間となる．しかし，同時表示で番号付けをする場合 $\{B_{sim}, L_{num}\}$ には，ある程度以上に選択肢数が多くなり番号が密集すると，ユーザが指定物体と番号の対応を確かめるのに時間がかかる，一方，色付けを行う場合 L_{clr} には，表示された色を見てユーザがその色名を考えるために多少時間がかかる．また，ユーザが区別可能な色の数を m_{ltd} とすると，それ以上の選択肢に対しては返答を作成することができなくなる．これらの性質を模式的に表すと，図 8(b) のようになる．

4.5 音声対話にかかる時間

音声対話時間は選択肢数 m とユーザの指定方法に変化を及ぼすラベル付け L に主に依存すると考えられる．また，対話にかかる時間 T_{dlg} はユーザの発話，音声認識処理時間，システムの返答にかかる時間 T_{vr} とその選択が正しいかどうかの確認の対話にかかる時間 T_{cnf} に分けることができる．ここで，音声認識が失敗しない場合には，

$$T_{dlg}(m, L) = T_{vr} + T_{cnf} \tag{8}$$

となる．しかし，実際には音声認識の結果は以下のように分かれる (括弧内は各結果が得られる確率)．

- 認識成功 (p_{tp}): ユーザの発話を正しく認識して正しい候補を表示する．確認後，次の選択ノードに移る．
- 誤認識 (p_{fp}): ユーザの発話を誤って認識して他の候補を表示する．この場合，確認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_n): ユーザの発話を解釈しても選択肢に関する情報が得られないため，再び同じ選択肢からユーザに選択してもらう．

これより，音声対話時間 T_{dlg} は，認識失敗を何度か繰り返した後に認識成功して次のノードに移る場合と誤認識して選択肢が一つ減った選択に移る場合に分けて考えると，以下のようになる．

$$\begin{aligned} T_{dlg}(m, L) &= \sum_{i=1}^{\infty} p_n^{i-1} p_{tp}(iT_{vr} + T_{cnf}) \\ &+ \sum_{i=1}^{\infty} p_n^{i-1} p_{fp}(iT_{vr} + T_{cnf} + T_{dlg}(m-1, L)) \\ &= \frac{1}{p_{tp}+p_{fp}} T_{vr} + T_{cnf} + \frac{p_{fp}}{p_{tp}+p_{fp}} T_{dlg}(m-1, L) \end{aligned} \tag{9}$$

$$T_{dlg}(1, L) = T_{cnf} \tag{10}$$

ここで，ユーザの発話やシステムの応答内容は，厳密には選択肢数 m やラベル付け方法 L に依存するが，発話にかかる時間自体は 2〜3 秒程度であるので，ここでは T_{vr}, T_{cnf} は m, L によらず一定と仮定する．一方，音声認識率については発話内容によって大きく影響を受ける．選択肢数 m が増加すると発話内容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．また，選択肢数 m の増加による影響は，ラベル付け方法によって変化する．番号付けの場合は増加しても数字内容が変化するだけでそれほど認識率は低下しないものの，色分けの場合には例えば「黄緑色」を「緑色」と誤認識すること等があるために認識率が大きく低下する．

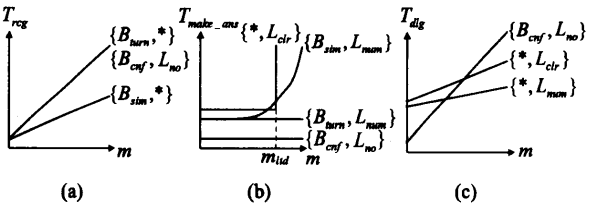


図 8 表示方法に対する選択時間の性質 (横軸:選択肢数，縦軸:時間)
Fig. 8 Property of choice time for each display method (horizontal axis: #choices, vertical axis: time)

一方，逐次確認の場合は最大で m 回の確認をすることになり，その場合の対話にかかる時間の期待値は以下のようになる．

$$T_{dlg}(m, L) = \frac{m+1}{2} T_{cnf} \tag{11}$$

最後に表示方法毎の音声対話時間の性質を模式的に表すと，図 8(c) のようになる．

5. 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を候補認識時間，返答文作成時間，音声対話時間に分けて考え，候補認識時間については色分割によるグループ認識等を考慮して定式化し，音声対話時間についてはラベル付け方法毎に変化する音声認識率を考慮して定式化した．

今後の課題は以下の通りである．

- 定式化に用いた各パラメタの実験による算出
- 候補選択にかかる時間の定式化の検証実験
- 一般ユーザを想定した主観テスト

文　献

[1] 槇原靖，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.
[2] 槇原靖，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.
[3] Y. Makihara, 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.
[4] Y. Makihara, 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.
[5] 滝澤正夫，槇原靖，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, 2002.
[6] 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.