サービスロボットのための視覚と対話の相互利用

久 野 義 徳†

頼まれたものを取ってきてくれるような，サービスロボットの実現を目指している．ここでは，このようなロボットにおける視覚と対話の統合的利用の研究について述べる．研究は大きく 2 つに分けられる．1 つは対話を通じて人間がロボットに依頼を行う場合，その依頼を理解するために必要な視覚に関するものである．人間の自然な依頼発話を理解するには，人間の非言語的行動や，周囲の状況を視覚で理解する必要がある．もう 1 つは，依頼の中に現れる物体の認識に関するものである．ロボットが物体を認識できない場合，対話により人間に支援してもらい，物体を認識する方法について述べる．

Reciprocal Functions of Vision and Speech for Service Robots

Yoshinori Kuno†

We are developing a service robot that can help humans by getting objects asked by a user's speech. This paper presents our research on the integrated use of vision and speech for such robots. The research is divided into two areas. One is about vision systems necessary for understanding spoken orders of humans. We show that the robot needs a vision system recognizing the situation and the nonverbal behaviors of humans to understand natural utterances. The other is about systems asking humans for help verbally when they cannot recognize objects along with the vision system.

1. はじめに

日常生活の中にロボットが入り込むことを考えると，人間とロボットのコミュニケーションが重要になってくる．コンピュータビジョンの研究者として，コンピュータビジョンがこの人間とロボットのコミュニケーションの中でどのように用いられるべきか検討している．視覚を持ったロボットやロボット用のビジョンに関しては多くの研究があるが，それらでおもに検討されているコンピュータビジョン単独での能力向上より，コンピュータビジョンの使用法の新しいコンセプトを提案し，その有効性を実証するということに主眼をおいて研究を進めている．その際に，以下の 2 つを基本ポリシに考えている．

(1) 自然なコミュニケーション

人間同士のコミュニケーションでは非言語的行動が重要であるといわれている．そこで，コンピュータビジョンのコミュニケーションの代表的な応用として，ジェスチャ認識が活発に研究されている．しかし，特定の身体の形や動きを意識的に行って，意図を伝えるというのは，人間の非言語的行動の中では中心的なものではない．そのように意図して起こした行動ではない，自然な行動がコミュニケーションに役立っていることが多い．そこで，このような自然なコミュニケーションの中で起こる行動の認識を利用して，自然なコミュニケーションができる方法を検討する．

(2) コミュニケーションはループ・双方向

これは 2 つの意味で考えている．1 つはロボットが人間の意図を理解する方向だけでなく，自分の状況を，たとえば，非言語的行動で人間に伝えることも考えなければならないということである．ただし，この部分はコンピュータビジョンとは直接はつながらない．もう 1 つは，コミュニケーションは参与者全員が 1 つのシステムになって，目標を達成するものだということである．ロボットは人間の要求に応え人間を助けるものだが，その要求の達成に必要なら，人間から助けてもらってもよい．全体システムの中で人間の負担が軽減されて目的が達成されるような枠組みになっていれば，ロボットの価値がある．

以上を基本ポリシとして，頼まれたものを取ってきてくれるような，福祉目的を中心としたサービスロボットの実現を目指して，そのための視覚技術につい

† 埼玉大学大学院理工学研究科数理電子情報部門
Division of Mathematics, Electronics and Informatics, Graduate School of Science and Engineering, Saitama University

て研究を進めている．

コンピュータビジョンのコミュニケーションへの応用というと，先にも述べたが，まず非言語的行動の認識があげられる．これに関しては，非言語的行動のヒューマンインタフェースへの利用ということで多くの研究がある．手のジェスチャや視線を認識し，計算機への入力手段にしようというものである．著者らのグループでもそのような研究を行ってきた[1]．そこにおいても，自然なコミュニケーションという基本ポリシに基づき，無意識的非意図的な行動のヒューマンインタフェースへの利用を検討した[2]．それらの研究の続きとして，サービスロボットに関しては，非言語的行動のうちでアイコンタクトについて研究している[3)~5)]．人間同士の場合，人に用があるとき，目を合わせるだけで意図が伝わることが多い．サービスロボットにもこのような能力を持たせようという研究である．人間同士のように自然なアイコンタクトがいかにしたら可能かを検討している．

次に検討しているのは，コンピュータビジョンを用いた発話の理解である．人間とロボットのコミュニケーションの手段としては，一般的には音声対話が最も適当なものと考えられる．ここで考えているロボットでも，音声でロボットに依頼をすることを想定している．ロボットはその発話を理解して必要な作業を実行する．この発話の理解の部分に必要な視覚について検討している．発話の理解のどこに視覚が必要かと思われるかもしれないが，まず，発話中に現れる物体などを実世界に対応付けるために視覚が必要になる．たとえば，「その本を取って」という発話をロボットが本当に理解したといえるためには，ロボットは「その本」を実世界の中で見つけなければならない．これは当然必要なものだが，それに加えて，人間の行動やその場の状況に関する情報を視覚で得ることが発話の理解に必要になる場合も多い．これらは広くいえば発話を実世界に対応付けて理解するシンボルグラウンディングの問題[6)]だが，ここではサービスロボットという応用の中で，コンピュータビジョンの技術に関連する部分に絞って問題を考えている．

研究の実際としては，視覚を用いた簡略化発話の理解を検討している．対面コミュニケーションの場合，人間同士ではあまり言葉で細かく言わなくてもコミュニケーションが成り立つ場合が多い．実際には様々な情報が利用されると考えられるが，まず，その時点での人間の行動に強く関わっている物体については，それについて名前を言うなどの詳細な言及が省略されるのではないかと考え，「あれ取って」というような簡略化された発話を理解するシステムを開発した[7)~9)]．

以上は自然なコミュニケーションというポリシに基づく研究である．ただし，以上の研究では，視覚情報が発話理解に必要であることを示すのが主眼で，発話で言及される物体などは簡単なものに限定していた．実際に役立つサービスロボットを実現するためには，発話中に明示的にせよ暗示的にせよ言及される物体を実世界の中で認識できるようにする必要がある．これは物体認識の問題である．物体認識はコンピュータビジョンの重要な課題であるが，サービスロボットで考えているような一般的な環境で多様な物体を認識できるようにすることは困難な問題である．

そこで，この問題の解決法としてコミュニケーションはループ・双方向という 2 番目のポリシに基づき，人間による支援を用いることを検討している．コンピュータビジョンのシステムでは，認識に失敗したらそれで終わりという場合が多い．一般に，失敗から回復するには，まず失敗していることを知り，それから成功に導くための指針を求めることが必要になる．これをコンピュータビジョンシステムが自動的に行うのは難しい．しかし，サービスロボットの応用なら，依頼をした人間がその場にいるので，その人に助けてもらうことが考えられる．成功しているか聞き，失敗ならどうしたらよいかを聞くというわけである．もちろん，ロボットのユーザが簡単に助けられるような質問をしなければならない．これが研究の課題になる．

このようなアプローチは，人間とコンピュータからなるシステムで，人間を結果の判断から修正へというフィードバックループに入れたものと考えられる．このアプローチについては，文献 10) で有望性を提案したが，それ以来，簡単な場合から次第に複雑な状況へ対応できるようにと研究を進めている[11)~15)]．ヨーロッパでは，CHIL（Computers in the Human Interaction Loop）というプロジェクトが進められている[16)]．字義どおりにとると，これは人間の世界のループの中にコンピュータが入るもので，著者らが考えているコンピュータのループに人間が入るものとは違うともいえるが，人間とコンピュータのつながりを考えている点では共通性がある．ただし，CHIL は大きなプロジェクトで，広範な問題を扱っている．人間の普通の営みの中にコンピュータが自然に入ってきて，人間を助けるというのがプロジェクトの目的である．著者らの研究は，人間が自然にロボットと対話していれば，それがロボットの視覚を助けるものになることを目指している．

以下，本稿では，人間の行動認識のコミュニケーショ

ンへの利用の研究について概略を述べ，それから対話物体認識の問題について，これまでの研究を紹介し，今後の課題を議論する．

2. 人間とロボットのアイコンタクト

人間同士のコミュニケーションでは頭部の動きや視線は重要な情報伝達手段になっている．ROBITA [17] や Robovie [18] などのロボットでは，話すときに頭部を聞き手に向け，また，聞くときは話者の方に向けることが実現されている．また，遠隔操縦のロボットで，操作者が見ている遠隔場面の映像の方向にロボットの頭部を動かすことが，ロボットの周囲の人に操作者の意図を予期させるのに有効であることが報告されている[19]．著者らのグループでは，これをさらに進め，頭部を動かすだけでなく，よりアクティブな行動，すなわち，目を合わせてアイコンタクトを行うロボットを開発した[3)~5)]．

アイコンタクトの成立には以下の 2 つの条件が必要だといわれている[20]．

(1) 互いに相手の目を見る．
(2) 両者とも相手に見られたことを知る．

1 番目の条件は，ロボットの目を人間の目（顔）に向ければ満足できる．2 番目の条件は，人間は特に意識せずに実現できるようだが，これをロボットに人間のように実現するのは難しい．そこで，ロボットを見ている人間を見つけたら，その人間の方を向き，それでも人間がロボットの方を見続けていたら，人間に見られていると判断し，見られていることに気づいたことを表情変化により伝えるようにした．

図 1 にロボットの外観を示す．移動ロボットの上にノート PC を搭載し，そのディスプレイに CG で顔を表示する．PC の上にパン・チルト・ズーム機構を持つカメラを取り付けてある．はじめにズームを広角側にしておき，その画像から人間の顔候補を求める．候補が見つかったらズームを望遠側にして，顔候補の中に目などの特徴があるか調べる．それがあれば顔と判断し，顔がロボットの方を向いているか求める．一定時間以上，ロボットの方を向いていれば，ロボットは身体をその人の方に向ける．その後，さらに一定時間以上，その人の顔がロボットの方を向いていたら，ロボットは人間に見られていると判断し，それに気づいたことを微笑むなどの表情変化により知らせる．

実験により，この方法により人間にアイコンタクトをしたと感じさせられることを確認した[3),4)]．そして，アイコンタクト後に手を少し動かせば，その人の方に来るロボットを実現した[3),4)]．さらに，手を動かさなくても目を合わせるだけで来るロボットを実現した[5)]．また，2005 年 12 月に東京の科学技術館で行われた磁性流体を用いたインタラクティブアートの展示会において，その 1 つの展示の説明を行うロボットを開発して展示と実験を行った．作品の近くにいる訪問者が近くにいるロボットの方を向きアイコンタクトをすれば，ロボットがやってきて，音声で作品の説明をしてくれるというものである．実験に協力してくれた 16 人の被験者に対して予期したとおりに動作することを確認した．

図 1 アイコンタクトロボット
Fig. 1 Eye contact robot.

3. 簡略化発話の理解

先に述べたように，人間同士の会話においては，それまでに会話で触れられていない物体に対しても「あれ取って」で意味が通じることがある．このような簡略化発話の理解を一般的に扱うのは困難なので，サービスロボットへの依頼に限って検討した[7)~9)]．ロボットへの依頼は「何を」（目的語，対象物体）「どうしてほしい」（動詞）という 2 つの要素からなると考えた．そして，それぞれについて，(1) 明確に言われている，(2) 指示語（目的語の場合），代動詞（動詞の場合）が使われている，(3) 省略されている，のどれかを判断するようにした．そして，対象物体に関して (2) か (3) の場合，人間が明確に言わなくても分かると判断して発話したのは，それが会話の当事者（人間とロボット）の行動に関連しているからだと仮定した．実際には，その物体を指差している，手で扱っている，その物体の近くにいる，そして，その物体を見ている（視線が

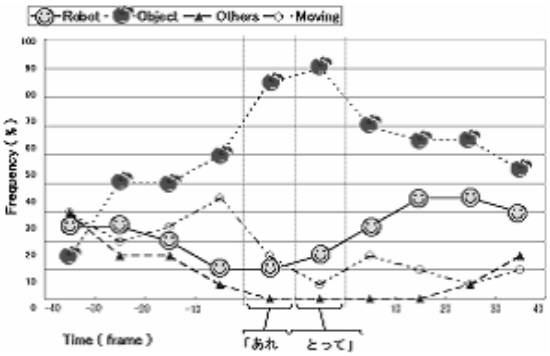


図 2 発話のタイミングと顔の向き
Fig. 2 Relation between utterance and face direction.

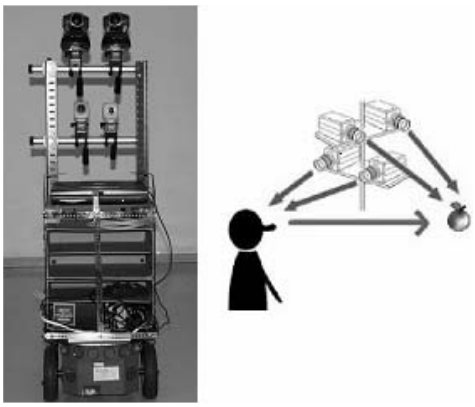


図 3 簡略発話理解ロボット
Fig. 3 Robot understanding simplified utterances.

向いている，実装の際は概略値として顔の向き）という 4 つの行動を考え，それに関わっている物体を視覚で検出し，検出された物体を対象物体であるとした．動詞が (2)，(3) の場合は，対象物によって，それに対して人間がしてほしい行為は決まっているとして，解釈するようにした．

4 つの行動のうち，視線以外はそれほど速く動かないので，発話された時点の行動から関連物体を検出すればよいと考えられる．しかし，視線は速く動くので，どの時点の視線上の物体を考えればよいのか検討する必要がある．そこで，以下のような実験を行った[9]．10 個程度の物体を周囲に置いた環境で，被験者にあるものを取ってもらいたいと思ってもらったうえで，「あれ取って」とロボットに対して言ってもらった．ロボットは後述のものだが，画像入力以外の動作は行わない．被験者 5 人に各 10 回の試行を行ってもらい，顔の向きを調べた．

図 2 に結果を示す．10 フレームごとに，その間に最も顔が向いた方向を求め，全試行の平均を表示している．総和が 100%を超える場合があるが，これはロボットと物体が同じ方向にある場合などは，双方を見たとカウントしたためである．この結果から，発話の始まる前から物体を見始め，「あれ取って」という発話の最中にはほとんどの場合，対象物体を見ていることが分かる．これは，まだ予備的な実験ではあるが，この結果に基づき，発話の少し前から発話終了までの間の顔の向きを求め，その間にロボット以外の方向で最も顔を向けた方向の物体を探すことにした．

図 3 に，これまでの検討の結果を実装したロボットの外観を示す．このロボットは 2 組のステレオカメラを持つ．下段のステレオカメラはつねに人間の方を向き，その視線や指差しの方向を求める，上段のステレオカメラは下段のステレオカメラで得られた視線あるいは指差しの方向上の物体を検出するのに用いる．視線あるいは指差しの方向の半直線上を 2 つのカメラの光軸が交わるようにカメラを動かしていき，zero-disparity filter[21] により，物体を検出する．

このロボットを用いて実験を行い，想定した簡略化発話の理解が視覚を用いて行えることを確認した．ただし，このシステムでは 4 つの行動に関連する物体を検出すればよいと仮定している．これは妥当ではあると思われるが，発話理解には他にも多くの情報を利用している可能性がある．そこで，現在，実験室に普通の家庭内のような環境を作り，車椅子使用者と介護者を想定した被験者に過ごしてもらい，そこに現れる依頼と，それがどのように理解されるかを調べている．また，老人介護福祉施設にビデオカメラを設置して，被介護者の意図を介護者がどのように理解しているかも調べている．この調査は社会学のエスノメソドロジ[22] の専門家と共同して行っている．エスノメソドロジは人間の行動の方法について調べる研究分野だが，そこで用いられる会話分析の手法で調査を行っている．会話分析は，ビデオデータから発話前後の行動と発話の関連を分析するものである．調査はまだ途中であるが，人間同士では発話にかなりの簡略化があるのが確認されている．特に，介護施設の場合はその傾向が大きい．仮定した 4 つの行動に関する簡略化が実際に出現することも確かめられている．さらに，ポリシ (2) に関するが，簡略化発話が理解されたかどうかを行動で互いに確認しながらコミュニケーションが進むことが分かってきた．現時点での調査結果については秋谷ら[23] に報告したが，さらに調査を進めていく．これにより，簡略化発話と関連行動の関係を明らかにし，その行動認識に基づき自然な依頼を理解できるロボットを実現していく計画である．

4. 対話物体認識

これまでに述べた研究はコンピュータビジョン技術の利用法に関する研究であり，その中で用いられるコンピュータビジョン技術に関しては，既存技術を利用している．また，3 章のシステムは簡単な環境で対象物体も顕著なものに対して動作を確認しただけである．実際に有用な福祉サービスロボットを実現するためには，これまで述べた応用システムに加えて，家庭内のような一般的環境で，依頼対象になる物体を認識できるコンピュータビジョンの技術が必要である．1 章で述べたように，これを人間との対話を通じて，人間からの支援を得て実現することについて研究を進めている．

コンピュータと対話しながら物体を特定したりシーンを理解したりすることについては Winograd の古典的研究[24] 以来，多くの研究がある．しかし，これらの研究では，シーン内の物体は形や色などで記述されており，記号として扱えるようになっている場合が多い．実際にロボットを扱う研究では，実世界が信号レベルから考慮されるようになってきている．たとえば Inamura らの研究[25] では獲得情報のあいまい性を考えてマルチモーダルな情報から人間の意図を理解している．しかし，視覚の部分に限れば対象物体は簡単なものに限定されている．それに対し，著者らのグループでは物体認識というコンピュータビジョンの問題を中心に検討を進めている．

4.1 物体認識システムの構成

物体認識システムの詳細に入る前に，著者らの物体認識の問題に対する考え方を述べる．物体認識の問題は多くの課題に分けられるが，ここでは以下のように分類して考える．

A. 対象に対する先験的知識がある場合

① 概念レベルの認識：物体に与えられた一般的な名前で認識できる．たとえば，どんな「本」でも「本」と認識できる．

② 特定物体レベルの認識：特定の物体を事前に見たことがあり，その特定の物体を認識できる．たとえば，特定の雑誌の特定の号を以前に見ており，今回，その雑誌を認識できる．

実際には，対象の一般性に関して ① と ② の間にさらに多くのレベルの設定が考えられる．

B. 対象に関する先験的知識がない場合

これは再認（recognition）という意味では認識ではないが，ある 1 つの物体を他の物体と切り分けて認知できる能力である．

人間の場合はほとんどの場合，A① のレベルで認識できる．知らない物体については，その名前を言われても認識はできないが，B のレベルの認識は通常は問題なくできる．ここで主に考えている「ものを取ってくる」というような依頼において，人間の場合でも頼まれた方が対象物が分からず聞き返すことはある．考えられる場合としては，3 章で扱ったような簡略化発話では対象が分からなかった場合がまずあげられる．「あれ取って」と言われても「あれ」が何か分からなかったような場合である．この場合は，人間の場合は「そこの本」というように，物体名を告げるなどの詳細情報を与えると考えられる．それにより A① の能力により物体を認識できる．ときには，物体名を与えられても分からない場合もある．これには，まれではあるがその物体名を知らなかったり，その物体名の中でのバリエーションを知らなかった場合，その物体名の物体が複数あり，そのどれか分からない場合，視野にあるのにたまたま目に入らず見つけられない場合がある．この場合は物体の属性や位置に関する情報を依頼者がさらに与えることで，普通は認識できる．物体名を知らない場合でも，B の場合に対する能力で切り分けて認知した物体の中で，依頼者の属性や位置情報に合うものを認識することができる．

以上のような人間の場合に対し，ロボット（コンピュータビジョン）では，A① のレベルの認識能力は現状ではきわめて限定的なものである．A② のレベルについてはかなり進歩してきているが，様々な条件変化に対して必ず動作が保障できるレベルではない．したがって，A①，② に対応しようとした視覚を持ったロボットでは認識の失敗を避けられない．そこで，そのような場合を救うために，対話物体認識を検討している．A①，② に対応した物体認識が失敗したということは，それに用いた先験的知識が有効でなかったということになる．対話により，知識を補えば認識できる場合もあるが，一般的にこの場合を救おうとするなら，B の先験的知識がない場合に対応する物体認識を準備しておく必要がある．しかし，ここでも人間とコンピュータの能力の大きな違いが問題になる．B の場合へ対応できるということは，コンピュータビジョンでいえばセグメンテーションができるということである．人間にとっては，ふつう，これは問題ない．したがって，先に述べたように物体名を知らない場合でも属性や位置の情報を与えてもらえば対象を認識できる．しかし，コンピュータビジョンでは完全なセグメンテーションは難しい．そこで，ここにも対話による人間からの支援を考える．

研究の流れとしては，きわめて限定された場合であるが A① のレベルの視覚の失敗を記号レベルの世界の中で補うシステムから対話物体認識にとりかかった．これを 4.2 節で述べる．それから，4.3 節で述べるセグメンテーションが完全な場合への対応に進んだ．そして，4.4 節で述べるセグメンテーションが不完全な場合に，対話でそれを補うシステムへと研究を進めている．

対話物体認識に関しては，さらに 2 つの考えを述べておきたい．1 つは以下で述べる対話物体認識だけでサービスロボットの視覚を構成するものではないということである．実際のロボットとしては，A①，② に対応するシステムを持ち，それらが失敗したときに対話物体認識を使用するという階層的なシステムを考えている．特に，先験的知識がなく，セグメンテーションもうまくいかないという場合に対応する 4.4 節の研究を進展させたものを対話システムの中でも最下層におき，手間はかかるかもしれないが，対象物が分からないことはないというシステムを目指している．

もう 1 つは，対話物体認識に関して，1 章で述べた 2 つのポリシの間の関係である．対話物体認識は 2 番目のポリシに基づくものであるが，人間でも，分からなければ相手に聞く．その意味では，対話物体認識も 1 番目のポリシに合う自然なものである．しかし，物体名のレベルで認識ができずに対話に入った場合は，その対話内容は，普通の人間が答えられるようなものを検討はしているが，人間同士ではあまり見られないものである．1 番目のポリシからすれば，これは避けたいことになる．そこで，先に述べたように階層的構成にして，その中で A①，② のレベルの視覚の能力を向上させ，対話物体認識の下位のレベルに来ることが少なくなるようにすることを考えている．しかし，自分では動けない人が何か取ってきてほしい場合には，多少の手間はかかっても，必ず成功するものが望ましい．それを実現するための基礎部分として，対話物体認識の研究をとらえている．

4.2 記号レベルのシステム

対話を通じた物体認識として，最初はまず記号レベルでの支援から検討を開始した[11]．人間同士では対話を通じてお互いの共通理解を目指していると考えられる．したがって，対話においては，何が分かっていて，何が分かっていないかを相手に伝えることがコミュニケーションを進める鍵になる．それが伝われば，相手は分かっていないことに関する情報を与えてくれることが期待できる．この考えに基づいて，サービスロボットの視覚システムを開発した．

このシステムでは，物体は色や形の属性で記述する．システムでは「○○を取って」という依頼が来ると，知識ベースの中からその物体の属性を調べてそれをゴールに設定する．一方，属性検出の画像処理を起動する．実装された属性は色，形，個数の 3 つだけである．検出された領域の属性とゴールを照合する．もし，すべての条件を満たす領域があれば，それが対象かどうか人間に確認する．一部しか照合するものがない場合は，照合する部分は肯定で，照合しない部分は否定にして，認識結果を言葉で説明する．それに対して人間が属性に対しての発話を行えば，ゴールの属性をそれに変更して判断処理を繰り返す．このシステムでは以下のような対話物体認識が可能である．たとえばリンゴは「色：赤；形：円」というように記述されているとする．ロボットは「リンゴを取って」という依頼が来たら，画像中から赤い丸い領域が検出できたら，人間に，それが目的物かどうか確認する．検出できない場合は，画像処理の結果を人間に伝える．たとえば，シーンには黄色いリンゴしかなく，人間が頼んだのはその黄色いリンゴだったとする．この場合，ロボットは黄色で丸い領域は検出できたが，赤い丸い領域は検出できない．そこで，「赤色でない丸いものを見つけました」と音声で人間に言う．それを聞くと人間はロボットが黄色いリンゴの存在を知らないということが分かり，「黄色だよ」というような色に関する情報をロボットに伝えてくれることが期待できる．このような発話があると，ロボットはゴールの色の部分を黄色にして，リンゴを検出できる．

実際のこの研究は従来の記号レベルの研究と同様なもので，コンピュータビジョンは完全に物体についての記号レベルの記述を与えてくれるものと仮定している．4.1 節で述べたが，A① のレベルの視覚が備わっていると仮定していることになる．しかし，実際には A① といっても，利用している先験的知識は赤い丸いものがリンゴ，四角いものが本というレベルのものであり，シーンの中にも物体が 2〜3 個しかないと仮定している．したがって，これだけでは限定した場合にしか動作しない，トイシステムにすぎない．しかし，4.1 節で述べたように，階層的なシステム構成を考え，失敗したら下位にいけばよいとできれば，限定した場面でしか動作しなくても意味があることになる．ただし，現時点のシステムでは限定が強すぎるので，A①，② レベルのさらに能力の高い物体認識と組み合わせることを検討する必要がある．以上のように，このシステム自体ではまだ技術的課題は多いが，「分かっていることと分からないことを伝えると，人間から分か

らないことについての情報が得られる」という考え方は，以降の研究につながっている．

4.3 物体の特定

4.2 節の研究では対象シーンが簡単ということで暗黙のうちに，画像中には対象物体の他には少数の物体しかないと仮定していた．しかし，実際のシーンでは画像をセグメンテーションして，その中から対象物体を検出しようという場合，セグメンテーション結果の中に多数の領域が含まれるのが普通である．そこで，多数の領域の中からどれが対象物体かを対話により特定する方法を検討した[12]．これも領域の属性で処理を考えており，まだ記号レベルでの扱いともいえるが，対話を生成する際には画像処理のことを考慮に入れている．

ここで扱う問題は，画像中の多数の物体の中から人間が決めた物体を人間に対象について質問をすることにより情報を得て特定することである．質問に対する答えから該当する候補を絞っていき，物体を特定するわけだが，結果を得るまでの質問の数ができるだけ少なく，また，それぞれの質問が人間にとって答えやすいものであることが望まれる．研究課題は，画像が与えられたとき，このような質問を生成する方法を検討することである．まず，画像に対して，色情報に基づくセグメンテーションを行う．ここでは，セグメンテーション結果の各領域が 1 つの物体に対応していると仮定する．したがって，どの領域が対象物に対応するかを決定することが課題になる．各領域について，色，形などの特徴を求め，その特徴に関して人間に質問を行う．質問の生成に際しては，特徴の性質を考慮して，人間に答えやすく効率的に対象を絞り込める質問を作成する．ここでは，以下の 4 つの性質を考える．

① 語彙の豊富さ
② 分布により影響を受けるか
③ 唯一性
④ 絶対的か相対的か

① はその特徴を表す語彙がたくさんあり，人間が言葉でその特徴を表すことが簡単にできるかどうかということである．色はこれにあてはまる．形も語彙は豊富だが，言葉で表しにくい形も多いので，これにあてはまらないと考える．② は周囲に他の物体が存在するときに影響を受けるかどうかということである．たとえば位置に関する表現は，物体の分布により，使えるものが変わってくる．それに対し，色ではそういうことはない．③ は同じ値を持つものが他にあるかということである．④ は大きさの大小などは他のものの存在により変わる可能性があるので相対的なのに対し，形はそういうことがなく絶対的であるというようなことである．今回のシステムでは，特徴としては，色，形，大きさ，位置の 4 つしか扱っていないが，それらの特徴についてここで述べた性質をまとめたものを**表 1** に示す．

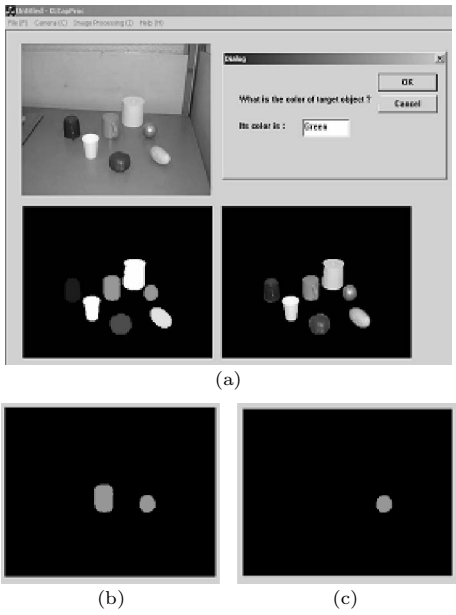
表 1 特徴の性質
Table 1 Features and their characteristics.

Characteristic	Color	Size	Position	Shape
Vocabulary	rich	-	rich	-
Distribution	-	-	dependent	-
Uniqueness	-	-	unique	-
Absolute/Relative	absolute	relative	relative	absolute

セグメンテーション結果の領域の特徴の分布と特徴の性質から，答えやすく効率的な質問を生成する．質問の形としては (1)「何？」という質問（たとえば「どんな色ですか」というような質問），(2)「はい・いいえで答えられる質問」，(3)「A，B，C のどれですか」というような質問が考えられる．対象物体の候補を少数に絞り込む効率という観点からいえば，(1) の質問が良いことになる．しかし，(1) の質問は人間には答えにくい場合がある．(2) の質問は (1) と反対の性質を持つ．したがって，どういう特徴に関して，どちらの形の質問をするのが良いかを考えることになる．なお，今回のシステムでは (3) は特別の場合にしか用いていない．

紙数の関係で質問文生成の方法の詳細は省略するが，概略の方針は以下のようなものである．まず，それぞれの特徴について，分布を調べる．たとえば色特徴は 7 つのクラスに分けているが，それぞれに何個の領域があるかを求める．そして，分布が一番分散している特徴を求める．その特徴が語彙が豊富なものであれば，その特徴に関して「何？」という質問を行う．たとえば，領域の色が多くの色のクラスに分かれていれば，色は語彙が豊富な特徴なので,「何色？」と聞く．こう聞かれても，語彙が豊富な特徴に対してなら人間は容易に答えられる．ある特徴についてある特定の値を持つ領域が多数ある場合には，それについては「はい・いいえで答えられる」質問を行う．たとえば，四角形の領域が多く，少数が円の場合は,「四角形ですか」と聞く．そして，対象の数が絞られてきたら，位置関係（右・左など）の質問を用いることも検討する．**図 4** に対話の例を示す．この例では，最初にロボットは「物体は何色か」と聞く．人間が「緑」と答えると図 4 (b) の物体が候補として残る．そこで，ロボットは「左の物体か」と聞く．人間が「違う」と答えたので，ロボットは図 4 (c) に示す物体を対象物と認識する．

4 つのシーンで 10 人の被験者に実験者が意図した



(a)

(b) (c)

図 4 対話物体認識の例 (a) 原画像（左上），カラーセグメンテーション結果（左下），切り出された物体（右下），(b) 最初の質問の後に残った物体，(c) 最後の質問の結果

Fig. 4 Example of interactive object recognition. (a) Original image (upper left), Color segmentation result (lower left), Extracted objects (lower right); (b) After the first question; (c) After the final question.

物体を質問をして当ててもらう実験を行い，提案システムの場合と比較した．その結果，提案システムが質問数で多くなる例はなかった．したがって，生成した質問は効率的であると判断できる．また，提案システムでは語彙が豊富な特徴でなければ「何？」という質問は行わないので，答えやすい質問が生成されていると考えられる．

4.4 対話によるセグメンテーション

前節の方法により，多数の物体の中から目的の物体を対話により特定できるようになった．しかし，この方法では，画像のセグメンテーションに誤りがなく，1つの領域が1つの物体に対応していると仮定している．画像特徴の性質を考慮して質問文を生成するという基本的な方法は，一般的に使えるものとして重要な成果と考えられるが，実際のシーンに対しては，この仮定が成り立つことは期待できない場合も多い．物体どうしに重なりがあったり，1つの物体が複数の色で構成されていたりすれば，この仮定が成り立たなくなる．さらに，そのような場合でなくても，セグメンテーションが完全でない場合もある．そこで，対話を通じてセグメンテーションを修正する方法を検討した[13),14)]．

最初に提案した方法[13)]の基本的な部分は以下のようなものである．領域の境界となる可能性のある部分の両側の詳細な特徴を調べる．その結果，境界の両側の部分が同一物体か，異なる物体か確かに判断できる場合以外は，人間にどちらであるか尋ねる．しかし，この方法には2つの問題点がある．1つは，境界候補ごとに人間に判断をあおぐので対話数が多くなってしまうことである．もう1つは，コンピュータの方で確かと判断した場合でも間違いがありうることである．これを防ぐにはすべての境界候補に対して人間に質問をすればよいが，これではさらに対話数が増えてしまう．そこで，以下のように改良した方法を考案した[14)]．まず，境界候補の詳細特徴を調べる．そして，特徴の値を利用して，可能性の高い順にセグメンテーション結果の候補を生成する．その際，必要ならばさらに別種の詳細特徴を調べる．それから，最も可能性の高いセグメンテーション結果から人間に説明して，正しいか確認する．この説明の際には，最初に解釈した物体の数を伝え，人間に確認を求める．この数が違えば，次の候補に進む．数が同じ場合は，どういう物体があるかセグメンテーション結果の詳細説明を行う．人間がそれに同意すれば解釈が正しかったことになる．違いを指摘された場合は他のセグメンテーション結果候補のうち，その数のものの検証に進む．なお，以上のプロセスの際に，特徴量からでは判定ができない境界候補があれば，それについては，最初に提案した方法[13)]と同様に，その境界についての判断を人間に聞く．現時点の実装では，その境界部分をディスプレイ上に示し，その両側が同一物体か別の物体か聞く（これに関しては 4.5.1 項参照）．

実装したシステムでは最初に調べる境界部分の詳細特徴に reflectance ratio を用いた[26)]．reflectance ratio は境界の両側が同一面上にあれば境界上で同じ値をとる．そこで，境界線をはさむ2点間のこの値を求め，その分散を計算する．分散が小さければ，境界の両側は同一物体の可能性が高い．逆に大きければ，2つの領域は画像上では隣接しているが空間的には離れた別の物体である可能性が高い．しかし，2つの物体が密着している場合は，当然，分散は小さくなる．他にも，いろいろな場合があり，この値による判断が絶対というわけではないが，分散が実験的に定めた上限しきい値以上なら2つの物体，下限しきい値以下なら同じ物体と判断して仮説生成に進む．この2つのしきい値の間の場合は，次に示す特徴を調べる．

reflectance ratio で判断できない場合には，境界線を通る線分上の輝度プロファイルを調べる．照明条件や物体の反射特性など，いろいろの条件が関わるが，shape-from-shading の研究で示されたように，輝度

表 2 輝度プロファイルのパターンによる判定
Table 2 Decision based on intensity profile patterns.

Intensity profile of region 1	Intensity profile of region 2	Decision
Line:	Line:	Same object
Line+:	Line-	Same object
Line-:	Line+:	Same object
Curve:	Curve -:	Same object
Curve +:	Curve -:	Same object
Curve -:	Curve +:	Same object
Other combinations		Different objects
Too small or big region (s)		Unknown

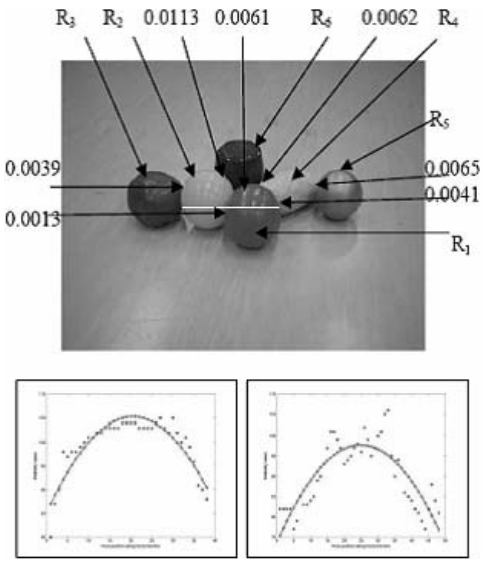


図 5 対話によるセグメンテーションの例．(a) 画像データ（上），(b) 黄色物体上の輝度プロファイル（左下），(c) 赤色物体上の輝度プロファイル（右下）
Fig. 5 Segmentation example through interaction. (a) Image data (upper), (b) Intensity profile on the yellow object (lower left), (c) Intensity profile on the red object (lower right).

の変化は 3 次元の形の変化に対応する場合がある．そこで，境界の両側の輝度プロファイルを調べ，それに直線あるいは 2 次曲線をあてはめる．そして，境界をはさんだ両側が 3 次元世界で接続している面上にありうるパターンかどうか判定する．判定法を**表 2** に示す．面が接続していても，たとえば色がそこで変わるなどで輝度の値が連続しているとは限らないので，この判定法では輝度プロファイルの形という定性的な指標で連結性を判定している．この判定も絶対に正しいというわけではないが，境界の両側が同一物体かどうかの可能性についての判断の指標として仮説生成に用いる．

図 5 に実験結果の例を示す．図中の数字は reflectance ratio の分散だが，中央付近の黄色と赤色のボールの境界の分散が 0.0013 と小さいため，最初は同一物体だと判定し，「物体数は 5 つですか」と人間に聞く．人間が「6」と答えると，図 5 (a) に白線で示す部分の輝度プロファイルを調べる．その結果が図 5 (b)，(c) である．このようなプロファイルの組合せは表 2 の中の同じ物体と判定される中にない．したがって，この 2 つの部分は別の物体であり，物体数は 6 であるという仮定を得る．物体数が 6 ということはすでに分かっているので，「2 つの赤い物体，2 つの黄色い物体，1 つの青い物体，1 つの緑の物体がありますか」と仮定の内容を人間に説明して確認を求める．この例のように仮説が違っている場合，輝度プロファイルを調べないで判定を下していた境界があれば，輝度プロファイルを調べる．物体数を多く考える必要がある場合には，reflectance ratio の分散が下限しきい値より小さくて，同一物体と判定した部分を対象に調べる．逆の場合は上限しきい値を超えた部分を調べる．複数ある場合は，しきい値に近い方から調べる．この例を含めて，17 種類の物体から 2〜6 個を選び，80 シーンを構成して実験を行った．実験では，全部で 335 個の境界があったが，そのうちの 81%は自動判断を正しいとして仮説を生成し処理を進め，確認以外の対話は必要なかった．残りの 19%については，対話による支援が必要であった．

この方法の基本的な考え方をまとめると以下のようになる．まず，対話システムといっても，できるだけ人間の負担を減らすためには，正しい解釈が自動的にできるのが望ましい．そこで，多くの詳細な特徴を調べるようにする．さらに，人間が簡単に確認できるように，境界候補ごとでなく，全体の解釈の仮説を立てる．一方で，間違いの可能性を考え，最終的には人間に解釈結果を示して確認してもらう．実際に，これをシーン全体に行うと実用的でないおそれがあるが，ここでは，サービスロボットの視覚として，人間の意図した物体を検出するのが目的であり，その目的物体周辺についてセグメンテーションができればよいので，適当な対話数で目的が達成できると考えられる．

4.5 対話物体認識の課題

物体の特定とセグメンテーションを対話を通じて行う研究について述べたが，これらはまだ研究の初期段階で課題が多く残されている．ここでは，そのうちの主なものについて議論する．

4.5.1 空間関係の表現

研究課題としては，まず，4.3 と 4.4 節で述べた物体の特定とセグメンテーションの両者を統合する必要がある．ここで考えている視覚はシーンを理解するこ

とでなく必要な物体を検出することである．したがって，先に述べたように，セグメンテーションの修正は，特定の物体の検出に必要な部分について行えばよい．対話を通じて，目的の物体のある部分を限定していき，物体の特定に必要な部分に対してだけ，対話によるセグメンテーションを行えばよい．しかし，ここに大きな問題がある．実は，4.3，4.4 節の個々のシステムの場合にも同じ問題が存在する．それは，人間とロボットがどの部分について対話しているのかを，どのように理解しあうかという問題である．4.3，4.4 節の現状のシステムではロボットが処理対象画像をディスプレイに表示して，この問題を回避している．すなわち，ロボットが発話する際に，発話の対象になっている，たとえば境界候補などを，ディスプレイの上で人間に分かるように表示している．これは実用的な方法ともいえるが，やはり，音声対話だけで問題を解決したい．

この解決の方法として reference system の利用を検討している[15]．reference system とは，空間表現の枠組みを考えて，それで物体の位置を指定しようというものである．たとえば，ある物体を基準にとって，それに対する位置関係で他の物体を示すようなことである．文献 15) では物体に重なりのある場合には，形や大きさなどの特徴が画像から得られないので，それを補うものとして，reference system の利用を提案したが，話をしている部分を会話の両者で共通理解するための方法として，重要な働きをするものとして研究を進めている．

reference system については，人間がどのように空間を認知して表現するかという観点から心理言語学などで多くの研究がある．Levinson[27] は人間の reference system を intrinsic，relative，absolute の 3 つに分けている（intrinsic はものの名を言えば，前後などが決まっているものを利用する方法，relative は A から見て B の C にあるというようなもの，absolute は東西など絶対的に決まっている指標を利用するもの）．文献 15) では，シーンの中から顕著な物体を選び，それを基準にして，ロボットあるいは人間から見て，その基準物体からどこにあるものという形で，relative system を利用することを提案した．また，複数の物体が固まっている場合には，固まり全体を 1 つの対象と考え，固まりの中のどこという形での空間表現も提案した．これは Tenbrink らが移動ロボットの行き先の指令の際に考えた group-based reference system[28] に相当する．対話の際に，人間とロボットで共通の認識ができた物体があれば，それを基にした reference system で位置関係を表現し，別の物体が共通認識できる．これを順次行えば，人間の意図した物体にたどり着けると考えられる．たとえば，シーン中に赤い物体が 1 つしかなければ，赤い物体と言えば，簡単に共通認識ができる．こういった物体を手始めにして対話を進めていけばよい．文献 15) では基準物体の選択法や利用などについて簡単なものを提案したが，さらに研究を進める必要がある．

4.5.2 対話と学習

ここでは対話を通じた物体認識について述べたが，最初のうち，対話を使わなければならないのはよいとしても，いつも同じように対話が必要では人間は使う気にならなくなる．しかも，会話は色や形や位置関係などの属性に関する語彙による会話である．人間にとっては，やはり物体の名称を使った会話が自然である．提案したような対話物体認識システムを使った場合でも，人間は対象物の名称をどこかで言うと考えられる．そこで，その物体を属性に関する語彙の対話で認識できたら，その物体と人間の使った名称を対応付けて，次回からは名称で指示されても認識できるようにすることが課題として考えられる．もちろん，次回に名称による指示で認識を試みて，できない場合には対話を用いればよい．これは行動を通じながら言葉の意味を理解させようという Roy らの研究[29] と通じるものであるが，実用的なロボットの実現のために重要な研究課題である．

この学習の問題については，実際の検討はまだあまり行っていない．関連するものとしては，成功した画像処理法を場所に結び付けて記憶しておくことを提案した程度である[30]．アフォーダンスの提唱者の Gibson は視覚認知の対象の分類の中で，付着対象と遊離対象というカテゴリをあげている[31]．付着対象は環境に固定されていて動かないもの，遊離対象はそうでないものである．ここで考えているサービスロボットに関していえば，人間が取ってきてもらいたいようなものは遊離対象である．そして，付着対象は，家具などであり，そこに遊離対象が置かれるなどしている．そこで，遊離対象の認識のときに，関連する付着対象を人間に言ってもらい，そこで成功した物体認識のための画像処理法やそのパラメータを付着対象に関連付けて記憶しておく．次に，その遊離対象が言及されたとき，関連する付着対象として記憶した付着対象が指示されれば，記憶されている情報を最初に使って認識を試みる．もちろん，それでうまくいかなければ，対話で支援を求める．これは，照明条件などが変わる可能性もあるが，付着対象により環境が規定されるという考えに基づいている．また，付着対象は動かないの

で，一度，どこにあるかが分かれば，それ以降は視覚で認識する必要はない（ロボットが自己の位置を知っている必要はあるが）．しかも，人間は付着対象をその名称で指示できる．不確実な視覚情報処理を避けて，環境情報を得られるという点でも，付着対象の利用は有効である．4.5.1 項で述べた空間関係の表現と組み合わせれば，さらに利用範囲が広がると考えられる．このように付着対象の活用は有効だと思われるが，学習に関しては多くの課題があり，今後，検討していきたい．

4.5.3 セグメンテーションと認識

ここでは，セグメンテーションをしてから認識という手順を考えている．より正確にいえば，セグメンテーションで個々の物体に対応する領域を得て，その領域の属性を調べて人間の意図した物体を検出している．先験的知識がまったくない場合には，このようにセグメンテーションが基本的には必要になる．しかし，先験的知識が利用できる場合にはセグメンテーションをしなくても，知識に合う部分の検出をすることが認識になるという方法も考えられる．このような物体認識としては，SIFT アルゴリズム[32] をもとにした方法などが提案され，複雑な背景での動作などの有効性が示されている．ここで提案の方法とこのような物体認識法との統合も今後考えていきたい．

5. ま と め

福祉を中心とした目的のサービスロボットの開発を目指している．ここでは，その実現に必要な技術として対話に関連する視覚について述べた．研究は大きく 2 つに分けられる．1 つは対話を理解するための視覚である．人間は対話において視覚情報を相互が得ていることを前提にして発話を考える．このような人間の自然な発話を理解するためには，視覚が必要である．もう 1 つは，ロボットの視覚が十分に機能しないときに，人間に対話により助けてもらおうという研究である．両者について別々に行っている研究の成果を述べたが，将来的には前者の中に後者が組み込まれるべきである．すなわち，視覚情報を得て対話を理解しているが，必要な視覚情報が得られないときは，自然に相手に必要なことを聞くというシステムの実現が望まれる．まだ，個々の部分についても研究すべき課題が多いが，それらの研究を進めて，統合したシステムの実現を目指していきたい．また，これらの研究は人間に深く関わっている．そこで，工学の技術面からだけの検討だけでなく，人間の性質を調べて研究を進めようということで，社会学の専門家と共同研究を進めている．分野融合的な研究を進めることで，人間に真に有用な技術を開発していきたい．

謝辞 本研究の一部は総務省戦略的情報通信研究開発推進制度および科学研究費補助金（14350127）による．

参 考 文 献

1) 久野義徳：ポインティングデバイスとしての身体動作，情報処理学会論文誌：コンピュータビジョンとイメージメディア，Vo.43, pp.43–53 (2002).
2) Kuno, Y., Ishiyama, T., Nakanishi, S. and Shirai, Y.: Combining Observations of Intentional and Unintentional Behaviors for Human-Computer Interaction, *CHI '99 Conference Proceedings*, pp.238–245 (1999).
3) Miyauchi, D., Sakurai, A., Nakamura, A. and Kuno, Y.: Active Eye Contact for Human-Robot Communication, *CHI2004 Extended Abstracts*, pp.1099–1102 (2004).
4) Miyauchi, D., Nakamura, A. and Kuno, Y.: Bidirectional Eye Contact for Human-Robot Communication, *IEICE Trans. Inf. & Syst.*, Vol.E88-D, No.11, pp.2509–2516 (2005).
5) Kuno, Y., Miyauchi, D. and Nakamura, A.: Beckoning Robots with the Eyes, *Proc. International Workshop on Intelligent Environments*, pp.260–266 (2005).
6) Deacon, T.W.: *The Symbolic Species*, W.W. Norton & Company (1997).
7) Hanafiah, Z.M., Yamazaki, C., Nakamura, A. and Kuno, Y.: Human-Robot Speech Interface Understanding Inexplicit Utterances Using Vision, *CHI2004 Extended Abstracts*, pp.1321–1324 (2004).
8) ザリヤナ・モハマド・ハナフィア，山崎千寿，中村明生，久野義徳：視覚によるサービスロボットのための簡略化発話の理解，電子情報通信学会論文誌，Vol.J88-D-II, No.3, pp.605–618 (2005).
9) 山崎千寿，久野義徳，中村明生：人間とロボットのコミュニケーションにおける顔の向き情報の利用，画像の認識・理解シンポジウム（MIRU2005），CD-ROM (2005).
10) 久野義徳：コンピュータはものを見ることができるようになるか，情報処理学会フロンティア領域ジョイント研究会 1998 全体パネル「人とコンピュータ」，情報処理学会研究報告，CVIM-111-12, pp.95–100 (1998).
11) 高橋拓弥，中西 和，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フェース，インタラクション'98 講演論文集，pp.161–168 (1998).
12) Kurnia, R., Hossain, M.A., Nakamura, A. and Kuno, Y.: Generation of Efficient and User-

friendly Queries for Helper Robots to Detect Target Objects, *Advanced Robotics*, Vol.20, No.5, pp.499–517 (2006).
13) Hossain, M.A., Kurnia, R., Nakamura, A. and Kuno, Y.: Interactive Object Recognition System for a Helper Robot Using Photometric Invariance, *IEICE Trans. Inf.& Syst.*, Vol.E88-D, No.11, pp.2500–2508 (2005).
14) Hossain, M.A., Kurnia, R., Nakamura, A. and Kuno, Y.: Interactive Object Recognition through Hypothesis Generation and Confirmation, *IEICE Trans. Inf. & Syst.*, Vol.E89-D, No.7, pp.2197–2206 (2006).
15) Kurnia, R., Hossain, M.A., Nakamura, A. and Kuno, Y.: Using Reference Objects to Specify Position in Interactive Object Recognition, *Proc. International Conference on Instrumentation, Communication and Information Technology*, pp.709–714 (2005).
16) CHIL: Computers in the Human Interaction Loop. http://chil.server.de
17) Matsusaka, Y., Kubota, S., Tojo, T., Furukawa, K. and Kobayashi, T.: Multi-person Conversation Robot Using Multi-modal Interface, *Proc.SCI/ISAS*, Vol.7, pp.450–455 (1999).
18) Kanda, T., Ishiguro, H., Ono, T., Imai, M. and Nakatsu, R.: Development and Evaluation of an Interactive Humanoid Robot "Robovie", *Proc. IEEE ICRA 2002*, pp.1848–1855 (2002).
19) Kuzuoka, H., Kosaka, J., Yamazaki, K., Yamazaki, A. and Suga, Y.: Dual Ecologies of Robot as Communication Media: Thoughts on Coordinating Orientations and Projectability, *CHI '99 Conference Proceedings*, pp.183–190 (2004).
20) Cranach, M.: The Role of Orienting Behavior in Human Interaction, *Behavior and Environment*, Esser, A.H. (Ed), pp.217–237, Plenum Press, (1971).
21) Coombs, D. and Brown, C.: Real-time Binocular Smooth Pursuit, *International Journal of Computer Vision*, Vol.11, No.2, pp.147–164 (1993).
22) 山崎敬一（編）：実践エスノメソドロジー入門，有斐閣 (2004).
23) 秋谷直矩，丹羽仁史，久野義徳，山崎敬一：福祉ロボット開発のための依頼のプロセスに関する基礎的考察，電子情報通信学会技術研究報告 福祉情報工学，Vol.105, No.684, pp.35–40 (2006).
24) Winograd, T.: *Understanding Natural Language*, Academic Press, New York (1972).
25) Inamura, T., Inaba, M. and Inoue, H.: Dialogue Control for Task Achievement based on Evaluation of Situational Vagueness and Stochastic Representation of Experiences, *Proc. International Conference on Intelligent Robots and Systems*, pp.2861–2866 (2004).
26) Nayar, S.K. and Bolle, R.M.: Reflectance based Object Recognition, *International Journal of Computer Vision*, Vol.17, No.3, pp.219–240 (1996).
27) Levinson, S.C.: Frames of Reference and Molyneux's Question: Crosslinguistic Evidence, *Language and Space*, Bloom, P., Peterson, M., Nadel, L. and Garrett, M., (Eds), pp.109–169, MIT Press, Cambridge, MA (1996).
28) Tenbrink, T. and Moratz, R.: Group-based Spatial Reference in Linguistic Human-Robot Interaction, *Proc. European Cognitive Science Conference* (2003).
29) Roy, D.K. and Pentland, A.P.: Learning Words from Sights and Sounds: A Computational Model, *Cognitive Science*, Vol.26, No.1, pp.113–146 (2002).
30) 吉崎充敏，中村明生，久野義徳：ユーザと環境に適応する指示物体認識のための視覚音声システム，日本ロボット学会誌，Vol.22, No.7, pp.901–910 (2004).
31) Gibson, J.J.: *The Ecological Approach to Visual Perception*, Houghton Mifflin (1979). 古崎 敬ほか（訳）：生態学的視覚論，サイエンス社 (1985).
32) Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91–110 (2004).

(平成 18 年 1 月 7 日受付)
(平成 18 年 7 月 21 日採録)

(担当編集委員 奥富 正敏)

久野　義徳（正会員）

1977年東京大学工学部電気工学科卒業．1982年同大学大学院工学系研究科博士課程修了．同年（株）東芝入社．1987〜1988年カーネギーメロン大学計算機科学科客員研究員．1993年大阪大学工学部電子制御機械工学科助教授．2000年より埼玉大学工学部情報システム工学科教授．工学博士．コンピュータビジョン，知能ロボット，ヒューマンインタフェースの研究に従事．電子情報通信学会，日本機械学会，日本ロボット学会，人工知能学会，計測自動制御学会，電気学会，IEEE，ACM各会員．