協調的物体認識のためのマン・マシンインタラクション設計*

近藤　一晃†　　西谷　英之††　　中村　裕一†

Designing Man-Machine Interactions for Collaborative Object Recognition*

Kazuaki KONDO†, Hideyuki NISHITANI††, and Yuichi NAKAMURA†

あらまし　人間とシステムが協調することで画像認識の適用範囲や精度を高める「協調的認識」の考えが提案されている．本研究では協調的物体認識を効果的に形成して認識を改善するインタラクションの設計について提案する．具体的には，認識結果やシーンの状況といった項目をシステムが評価し，認識を行う上で悪状況であった場合には，その説明とともに協力してほしい内容を人間に提示する．これにより，どうすれば認識が良い方向に向かうのかを簡単に知ることができるため，小さな負担で認識が改善されるとともに，人間にとって分かりやすい道具となることが期待される．協調的認識を行うための状況評価，提示インタフェースの設計・構築，及び実験を行い，認識困難な状況の検出が行えること，また人間の協力により認識が改善されることを確かめた．

キーワード　協調的認識，マン・マシンインタラクション，状況評価，認識困難な状況の検出

1. まえがき

現在，種々の画像認識手法が提案されており，また計算機処理の高速度化も相まって，自動認識の適用可能な範囲がますます広がっている．従来のような高度に統制された環境下だけでなく，より一般的な環境下での画像認識の利用とその高性能化が期待される．しかし，一般環境における高度な自動認識には，様々な条件，例えば，向き・照明・動き・隠れ・個体差などの多様性やそれらの経時的な変化への対応が必要である．特に観測対象や系に人間が含まれている場合，複雑な状況や想定外の事態が多く発生すると考えられるため，それらの対応はより強く求められる．なぜなら，人間の行動や振舞いは多岐にわたりかつ変化に富んだものであり，実利用の面から考えてそれらを強く拘束すべきではないからである．ただし，人間が系に含まれていることは，認識にとって悪い点だけではなく，実は良い点も持ち合わせている．例えば，人間の認識・判断能力をうまく利用する手法や，人間に手伝ってもらうことで認識問題を簡単にする，などが考えられる．

このような背景から，画像認識を行うシステムがある程度の性能をもち，加えて利用者がシステムに協力できる場合に，その適用範囲をより広くしたり精度を更に高める方法，また，その際に人間の負荷をできるだけ増やさないようにする枠組みとして，協調的認識の検討を行ってきた[2]．言い換えれば，人間が協力することで，画像認識の種々の環境への適応能力を高め，人間にとってもマニュアルレスで使えるようなシステムとする方法論を検討してきた．

協調的認識の枠組みには，利用者の協力が全く望めない場合，例えば監視カメラによる侵入者検知などには向かないという制限があるが，人間を支援するようなシステムでは利用者自身が受益者であることが多く，ある程度の協力が期待できる．具体的な導入例としては以下が考えられる．

調理支援：キッチンの状況や調理進行度などが認識できれば，調理者＝被支援者に対して状況に応じた支援や情報を選択して提供することができる．食材と調理器具が認識できればその食材の適切な調理法を教示することができ，調理進行度が認識できれば次に何をすればよいかを前もって知らせることができる．協調的認識の枠組みは調理者の協力を得ることで認識精度を向上させ，この支援シナリオを強化する．

† 京都大学学術情報メディアセンター，京都市
Academic Center for Computing and Media Studies, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto-shi, 606–8501 Japan
†† 京都大学大学院工学研究科，京都市
Graduate School of Engineering, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto-shi, 606–8501 Japan
* 本論文は第 13 回画像の認識・理解シンポジウム推薦論文である．

ジェスチャーインタフェース：利用者のジェスチャーを用いて家電機器等を制御する試みが実用化されつつあるが，一般的なシーンにおいて様々な人物のジェスチャーを精度良く認識することは難しい．協調的認識の枠組みでは，システムにとって認識しやすいように動作してもらう等の協力を利用者から得る．これにより，操作対象の家電が何であるのか，どのような操作をしたいのかを精度良く認識することができ，利用者にとって便利な入力インタフェースとなる．

本研究では，これら協調的認識を効果的に形成して認識を改善するインタラクションの設計を行う．その特長は，対象の認識と並行して状況の認識を行い，それにより認識にとって悪い状況と判断された場合には改善策を提示し，ユーザはそれを受けて状況を改善することである．特に状況の認識は，単なる失敗原因の推定とは異なるアプローチで，本研究独自の手法である．状況の認識では，悪状況，例えば，認識結果の信頼性が低いことや，物体の位置・照明や人間の操作などによって認識が困難になっていることなどを検出する．利用者への提示では，悪状況やその改善策を分かりやすく提案することで，できるだけ利用者に認知的負担や手伝う負担をかけないようにする．

2. 関連研究及び本研究の位置付け

悪状況が検出された場合にはそれを改善して正しい認識に導く，という考え方は，センシング結果によって次のセンシング方法が変わるアクティブビジョン [4], [5] とも似た概念である．しかし，協調的認識の枠組みには，(1) 人間には目的に合致するようにシーンを変更することが許される場合が多い，(2) 人間には知的な振舞いを期待でき，種々の状況改善を知的な判断のもとに行うことができる，(3) 認識の目的が利用者に内在している場合には，その目的自体を変えたり，タスク自体のスキップを指示できる，(4) 人間には答が分かっている場合がある，など，人間が系に含まれていることに起因する相違点が多い．

部分的に人間の協力を得て各種情報処理を高精度に行うという観点から見れば，半自動化 [3], [6]，半教師付学習 [1] などが関連する従来手法として挙げられる．これらの提案に共通するのは，計算機にとって困難な問題を部分的に人間に解いてもらう点である．一部の処理を人間が行う，データの一部分に人手で正解ラベルを与える，などが例として挙げられる．一方，協調的認識では，物理的に状況を変更することで認識を改善させるため，人間による協力内容の点で従来手法と大きく異なる．加えて，半自動化や半教師付学習は，いつも同じ協力内容が特定の場面で必ず必要とされるのに対し，人間の手伝いが必要である場合にのみ協力を依頼する，協力内容も状況に応じて変化する点も協調的認識がもつ相違点として挙げられる．

対話認識はより協調的認識に近い考えであるが，以下のような違いがある．SHRDLU 等に代表される初期の対話認識は，記号化された属性の集合で個々の物体を定義しており，与えられた属性だけでは一意に識別できない場合や計算機側の知識不足である場合，言い換えれば本質的に認識不可能な場面を問題として取り扱っていた．これは実世界の情報が正しくセンシング・記号化されるという前提に基づいているが，実際には必ずしも成り立つとは限らない．協調的認識では，その前提が満たされない場合，すなわち認識対象やそれを取り巻く環境がシステムの想定外である状況を問題としている．近年では，同様の問題を取り扱った対話認識手法 [7]〜[9] も提案されている．しかし，依然として対話を通して認識を改善させるため，対話数の増大や質問の設計法に関する問題が課題として残されている．効果的な質問についての提案 [10] もなされているが，現時点では有効な知見はまだ得られていない．協調的認識は対話を通じて情報を追加する代わりに，認識対象や環境を物理的に変更することで認識の改善を目指す．

自動認識と協調的認識の関係も興味深い．ここで強調したいのは，それらが互いに競合・排他するものではなく，両者の組合せが更に高精度な認識を可能とする点である．すなわち，対象の認識問題に対して，(1) 自動認識の観点からの高度化（観測方法の工夫・アルゴリズムの改善など），(2) 協調的認識の観点からの高度化（より簡単な協力内容・分かりやすい提示など），の 2 面からのアプローチが可能なのである．更に，自動認識が不得手とする状況を協調的認識が補助し，協調的認識における「人間が手伝う負担」を自動認識が軽減する，というように両者が相補的に機能して認識を良い方向へ導くことが期待される．

最後に協調的認識と以下二つの「適応」についての関連性を考えてみよう．

- システムに対するユーザの適応
- 認識対象や環境に対するシステムの適応

前者は，ユーザがシステムに順応し，徐々に簡単に使えるようになることを意味する．協調的認識では，悪

状況であった場合にそれを通知するため，ユーザは悪状況が実環境の何に対応するかやその改善策を学習することができる．そのため，悪状況とならないように作業を進める，たとえ悪状況に陥ってもすぐに復帰させることができる，などが期待される．一方，後者は，システムが徐々に賢くなり，認識に関する能力が向上することを意味する．協調的認識では，新たな悪状況・改善策をシステムが学習し，可能であればユーザの協力なしに正しく認識できるようになることに対応する．ユーザの協力により認識が改善された事実から，当初は想定していなかった悪状況・改善策を推定できると考えられるため，システムの利用を通じたオンライン学習が期待できる．これらの特性は本論文では議論しないが，協調的認識がもつ良い特徴として挙げられる．

3. 協調的認識の枠組み

3.1 人間の協力による認識改善

本論文では，図 1 に示すモデルに基づいてインタラクション設計を行う．ここでは人間（利用者）が系に包含されており，従来の自動認識である (A) コンピュータによる検出や認識，に加え，(B) 利用者への情報フィードバック，(C) 利用者によるシーンの改善や問題自体の変更，を含むループ系が構成される．このような枠組みがうまく働くためには，以下の前提が満たされる必要がある．

前提 1：シーンの状況が良い場合にはシステムは物体を正しく認識できる

前提 2：利用者はシーンを良い状態に変化させることができる

前提 3：利用者はシステムの認識結果の正しさを判断できる

利用者の協力なしにシステムが一方的に物体認識を行う場合は，前提 1 だけが満たされている状態である．この場合，シーンの状態が悪くなるにつれて性能が悪くなる．従来の自動認識手法の多くはこのケースに当てはまる．前提 2 では，利用者がその場にいて，良好な認識が得られるように協力することを意味する．既に述べたように，人間を支援するシステムであれば，この前提が満たされる場合が多く，また協力を得られるような問題設定とすることも可能である．そして前提 3 が満たされていれば，提示された情報によって利用者は状況を改善できる．

このような協調的認識の効果を利用者への負担の面から考えてみよう．多くの場合，利用者にかかる負担の主な要因は，「誤認識による負担」と「手伝う負担」であり，これら二つの「負担」の和を小さくすることが良いシステムの条件となる．利用者が手伝うことによって認識状況が改善されるならば，これら二つの「負担」は図 2 に示すようにトレードオフの関係にある．しかし，インタラクションの設計をうまく行うことによって，少ない「手伝う負担」でより多くの「誤認識による負担」を解消し，このトレードオフを軽減することができるはずである．本研究では，利用者への情報フィードバックをうまく設計し，小さな協力で認識精度の大きな改善を目指す．具体的にはシステム側から利用者へ以下のような情報提示を行う．

認識の状態：認識結果や認識の失敗など．認識状態を提示することで，利用者は正しく認識できていること，改善の必要があること，などを知ることができる．これは協調的認識におけるループのトリガとなる他，認識作業全体をスムーズに進める効果があると考える．

悪状況：認識が困難な状況であることや，認識失敗の原因など．利用者は認識を困難にしている原因，及び，それを改善すれば正しい認識を得られることに気づく．

状況改善策：悪状況を改善するための方策．利用者

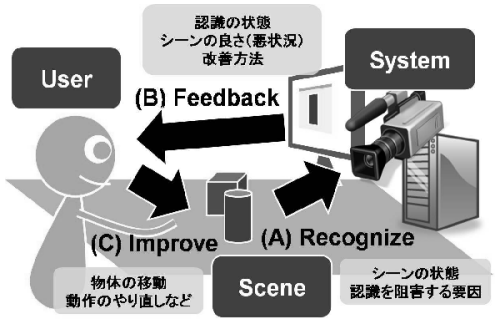

図 1 協調的認識のモデル
Fig. 1 Model of collaborative recognition.

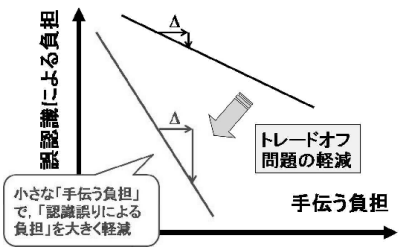

図 2 負担のトレードオフとその軽減
Fig. 2 Trade-off of burden given to a user.

はどのように協力すればシーンや観測の状態が良くなり，正しい認識に近づくことができるかを判断できる．

第 2 項，第 3 項の各々は，システムが状況の認識を行うこと，悪状況の改善策を推定・選択すること，で実現する．単純に考えれば，認識失敗の原因を推定する，その解決策を推定する，となりそうだが，そうではないところに本研究の特徴がある．詳細については次節にて述べる．

3.2 認識改善方法の概念

本提案におけるインタラクションを形式的な表現を用いて順に説明する．まず，認識状態の良さを R と表す．認識アルゴリズムを変えない場合，R は入力 $\mathbf{x}$ のみに依存する関数となる．これは，正しい認識結果が一意に得られていることを評価する関数ともいえる．

$$R = f(\mathbf{x}), \quad \mathbf{x} = [x_1, x_2, \ldots] \tag{1}$$

ここで，x_i は認識を左右する要因，例えば，認識対象物体の状態，人間の振舞い，周囲の環境等である．R が小さい場合，認識は失敗している可能性が高い．そのため R を最大にするような変位 $\Delta\mathbf{x}$ を求めて現状に適用すれば正しい認識が得られると考えられる（図 3）．すなわち，

$$S_d = \mathrm{argmax}\{f(\mathbf{x} + \Delta\mathbf{x})\} \tag{2}$$

となる改善策 S_d を提示すればよい．しかし，人間を含む一般環境下で最適解 $\Delta\mathbf{x}$ を解析的に求めることは困難である．これは，環境要因が多岐であること，常に変化し得ること，人間が系に含まれていること，などによる状況の複雑化によるところが大きい．人間を系に含むことや様々な環境への適応を考えた場合，要因 $\mathbf{x}$ が多岐にわたりかつ複雑であり，またそれらが互いに独立でないため，正しく $f(\mathbf{x})$ をモデル化することがほぼ不可能なのである．また，たとえ最適解が求められたとしても，複雑な改善策は利用者がそのとおりに行えるとは限らず，そもそも実施不可能なこともあり得る．

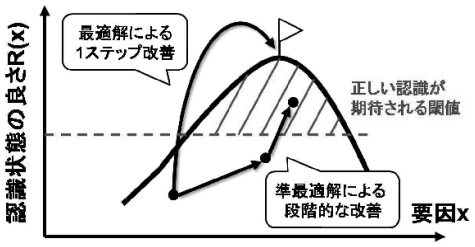

図 3 提案する認識改善方法のイメージ．最適解による 1 ステップ改善に限らず，準最適解による段階的な改善でも正しい認識を導くことができると考える．
Fig. 3 Illustration of a proposed method for improving recognition.

そこで本研究では，「最適解を求める」という厳密な制限を「準最適解を推定する」ことへ広げることで上記問題を緩和する．ここでいう準最適解とは，認識を失敗させている原因の一つは恐らくこれであり，このようにすればそれは改善される，といった多少曖昧なものでよい．つまり，失敗の原因そのものだけではなく，（失敗の原因を含んだ）状況を悪くさせている要因という大きな枠で捉えており，それが前節で述べた「悪状況」と「状況改善策」の考えに結び付いている．これを，状況の良さ R を用いて説明すると，状況改善策の集合 S は R を増加させるような $\mathbf{x}$ の変位とみなすことができる．できるだけ大きく改善されることが望ましいので

$$S = \{\Delta\mathbf{x}; \Delta R = f(\mathbf{x} + \Delta\mathbf{x}) - f(\mathbf{x}) >> 0\} \tag{3}$$

と表すことができる．この S を利用者がシーンに適用していくことにより段階的に正しい認識へと近づく．ここで $\Delta\mathbf{x}$ は準最適であればよいので，単一の要因 x_i に注目して推定することもでき，それは以下の二つの利点を生む．まず，多数の要因間の相互関係を考慮しなくてもよいので，簡便な手法で高精度に準最適解を求められる．また改善策を，例えば，物体を静止させる，位置を変える，といった単純かつ利用者への手伝う負担が少ない試行として提案できる．

上記方法では一度のインタラクションで正しい認識が得られる保証は必ずしもないが，たかだか数回のインタラクションで十分であると考えている（図 3）．これは，

- 最適でなくとも，それなりに良い認識状態であれば，システムは正しい認識ができる（図 3 におけるしきい値が高くない）
- 状況を悪くしている要因はたかだか数個であり，順に解決したとしても利用者に大きな負担をかけることはない

という前提，言い換えれば，**3.1** で挙げた前提 1 が高い水準にあることに基づいているが，認識技術が発展した昨今では決して厳しい条件ではない．ただし，ニュートン法や最急降下法のような繰返し演算により最適解に近づく数値解法とは異なる概念であることに注意したい．なぜなら本手法の目的は，式 (2) のよう

に R を最大化する状態に遷移することではなく，あくまで正しい認識を導く状態に遷移することだからである．言い換えれば「いかにして最良の状況にするか」ではなく，本来の目標である「いかにして正しい認識を導くか」に主眼を置いた発想といえる．

3.3 悪状況と改善策の具体的な導出法

以上は概念的な考えであり，やはり δR や S を解析的に求めることは簡単ではない．そのため，以下の流れで擬似的な $\{S\}$ を求めて利用者へ提示する．まず，悪状況であることを検出する．これは，(1) 認識失敗の状態から検出する，(2) 認識とは別の処理により認識困難な状況を検出する，の二つの観点から行う．続いて検出された認識困難な状況のそれぞれに対して改善策を推定する．具体的な手続きは以下のとおりである．

(1) では，認識に失敗している状態 E_i $(i = 1, 2, \ldots, m)$ を検出する．E_i には，認識の評価が悪い，候補が一つに絞れない等，種々のものが考えられる．次に，検出された E_i を回復させる方策の集合（S_{E_i}）をあらかじめ登録されたデータから選ぶ．この手続きは失敗原因の推定に基づいたアプローチといえるが，従来とは異なり改善策は最適でなくてよい．

$$S_{E_i} = \{S_{E_i 1}, S_{E_i 2}, S_{E_i 3}, \ldots\} \tag{4}$$

並行して，(2) の処理，すなわち認識を困難にしている要因 C_k $(k = 1, 2, \ldots, p)$ を物体の認識処理とは別の処理で検出する．要因 C_k には，正反射やモーションブラーが起きている等，種々のものがあり，それぞれの C_k に対し，それを改善するための方策が考えられる．こちらは認識を行う上での状況の良さを評価しているので，失敗の原因を直接取り除くものではない．状況の改善により結果的に正しい認識に進むことを期待したものである．

$$S_{C_k} = \{S_{C_k 1}, S_{C_k 2}, S_{C_k 3}, \ldots\} \tag{5}$$

C_k は要因 $\mathbf{x}$ と密接に関係するため，認識状態の良さ R を増加させるために，どの要因 $\mathbf{x}$ をどの方向に動かせばよいか（$\Delta\mathbf{x}$）は比較的容易に分かる．

以上のようにして得られた，$S = \bigcup_i S_{E_i}, \bigcup_k S_{C_k}$ から，有効な改善策を少数選んで S_p とし，利用者に提示する．ただし，信頼度と有効性の面から両者の性質は異なるため，どの改善策を提示すればよいのかには議論が残る．一般に認識の失敗はあくまでも結果であり，必ずしもその原因が明らかになるとは限らない．すなわち，推定された原因が果たして本当に失敗を引き起こした要因であるかどうかは疑わしい．しかし，それらが合致していれば，失敗の原因を直接取り除くことができるので認識改善の効果は大きいといえる．一方，認識を困難にしている状況は基礎的な画像処理により比較的高精度に検出できるため，その信頼性は高いが，認識の改善に対して間接的に働くので効果は多少低い．本論文では，両極端な方法よりも，確実に良い状況へ向かう方法が効率的であると考え，S_E と S_C の両方が得られている場合には，S_C を優先的に提示する．利用者の目的・習熟度・身体性などによっても提示の優先度が変化すると考えられるが，本論文では扱わず今後の課題としたい．

4. 協調的物体認識システムの設計と実装

4.1 問題設定

これまでに述べた協調的認識の枠組みを検証するために，実問題を設定し，協調的に認識を行うシステムの設計と実装を行った．想定した認識問題は調理や部品組立てなどの机上作業における物体認識である．机上作業を選択した理由は，**2.1** で挙げた前提 2，3 が満たされていること，また作業内容が十分に複雑で状況に即した支援を必要とする場面が多く，利用者がシステムに協力することで得られる利潤が高い（誤認識による負担を下げる意味が大きい）からである．協調的認識は上記のような条件がそろう問題設定であれば，**1.** で触れたような机上作業以外のアプリケーションでも効果的に働くと考える．

机上作業を行うシーンは以下のように定義した．

- 複数の物体がシーン中に出たり入ったりする
- 未知の物体は出現しない，または，認識する必要がない
- 物体の移動は利用者の手によってのみ起こる
- 照明環境，物体の性質，手と物体，物体同士の相互作用等により，種々の悪状況が発生する

本論文では個々の物体を認識することでシーン全体を認識する．また具体的な作業タスクの設定や利用者に対する支援などは行わず，物体認識の精度のみを問題とする．物体認識はカメラでシーンを撮影した画像に基づいて行い，利用者への情報提示はモニタにより行う（図 4）．できるだけ直感的に情報が伝達されるように，記号・図・簡単な説明文を入力画像に重畳して提示する（図 5）．

4.2 物体認識アルゴリズム

物体認識は画像から抽出した特徴量の比較により行

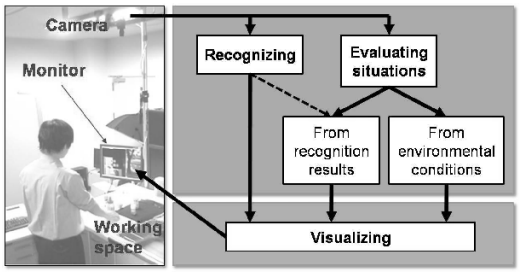

図 4　システム外観図と処理の流れ
Fig. 4　Overview of the prototype system and the processing flow.

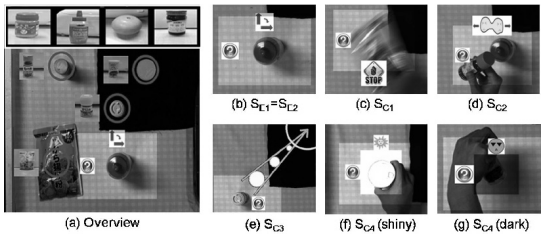

図 5　提示インタフェース例
Fig. 5　Examples of display for users.

う．全体的な特徴量として物体の大きさ，形状，色分布，局所的な特徴量として SIFT 特徴量を選択した．認識対象となり得る物体をその特徴量との対応で事前に登録しておき，作業中は以下の流れで認識を行う．(1) 背景差分処理により前景領域（物体領域）を切り出して特徴ベクトルを抽出する．(2) 抽出された特徴ベクトルと登録物体の特徴ベクトルをマハラノビス距離により比較する．ただし，まず全体的な特徴量で絞り込みを行い，一意の認識結果を得られない場合は更に局所特徴量を用いて類似度を計算する段階的な認識を行っている．一意の認識結果が得られており，かつ後述する悪状況が検出されない状態が一定時間以上続いた場合，認識結果が確定される．未確定状態では図 5 (a) 上部に示すように，複数の認識候補が次節で述べる情報提示法に従って提示される．

4.3　協調的認識のための情報提示

情報提示のための各種処理は図 4 に示すように行われる．以下にその詳細について説明する．

4.3.1　認識状態の提示

認識状態は認識の最終結果や途中状態から大きく三つに分類される．システムは以下の基準に基づいてどの認識状態であるかを判断し提示する．

単一の認識結果：認識結果の登録物体が他の上位の類似度に対して突出した類似度をもつとき，信頼度の高い単一の認識結果であるとみなし，その物体のみを提示する．

候補が複数：上位の類似度がほとんど同じであるとき，類似度最大の物体であっても認識結果の信頼性は低いとみなし，候補が複数あると判断する．また，それら候補を認識状態として提示する．

候補がない：登録物体について算出された類似度が全てしきい値以下であったとき，認識対象物体は登録されていないとみなし，その旨を提示する．

4.3.2　悪状況の検出と改善策の提案

悪状況，すなわち，認識の失敗や認識困難な状況が検出された場合には，システムによる認識結果が信頼性に欠けることを記号で示すとともに改善策を提示する．**2.2** で述べたように，改善策は，認識失敗の状態と認識困難な状況の各々から導き出される．本論文では表 1 に示すような悪状況とその改善策を設定した．これらは因果関係や事前知識に基づいて人手で構成したものであるが，実際の認識作業中に発生した悪状況や改善策から $\{S_{E_i}\}$ や $\{S_{C_k}\}$ を収集することも重要である．**3.3** で述べた手続きに従い，以下の二つの処理で悪状況の検出と改善策の提案を行う．

(1)　認識失敗の状態

候補が複数，または候補がない：候補が複数あるような認識状態であった場合，一意に識別するための情報が足りないために失敗したとみなす．認識候補がない場合も登録時とは異なる向きが見えているために失敗していると推察する．このときは追加の情報を得る，若しくは登録時と同様の向きとするために，物体の向きを変える改善策を提案する．一般に改善策は複数あると考えられるが，本論文では改善策とそれを実施した結果との因果関係を明らかにするため，原因に対応した一つの改善策のみを扱う．

候補が二つ：候補を二つにまで絞り込むことができたが，悪状況がこれ以上検出されない場合は，人間にしか答えが分からない状況だとみなし，ユーザが正解を簡単に教示できる二者択一のインタフェースを提供する．ただし利用者に判断を委ねてしまうのは認知的な負担の面から望ましくないため，提示の優先度は最も低く設定している．

(2)　認識困難な状況

表 1 に挙げた認識困難な状況の検出法を以下に示す．認識失敗の場合と同様，各々の認識困難な状況に対して単一の改善策を対応させている．

表 1　悪状況とその改善策
Table 1　Correspondences of bad situations and their improvements.

i	認識の失敗 E_i	改善策 S_{E_i}
1	認識結果として複数候補が残存している	物体の向きを変える
2	認識結果の候補がない	同上
3	認識結果の候補が二つで，かつ悪状況が検出されない	正解を教示できるインタフェースを提供する
k	認識困難な状況 C_k	改善策 S_{C_k}
1	物体の移動が速い	一度動きを止める
2	複数物体が近接している	近接している物体を離す
3	物体色が背景色と類似している	異なる色をもつ背景に移動させる
4	物体が鏡面反射を起こしている	良い照明環境となるように向きを変える/移動させる
5	物体が暗すぎる	同上

物体の移動が速い：未知物体及び手領域部が高速に移動していることを追跡により検出する．手領域の追跡には肌色の検出と腕の幾何学制約を用いている．この状況が検出されたときには，一度対象物体を静止させることを提案する．

複数物体が近接している：既に配置されている物体に未知の物体が近接してきたことで検出する．近接には物体同士の隠れ・接近運動の有無・前景領域が複数物体がもつ特徴量を併せ持っている等で判定している．この場合はそれらの物体を離す提案を行う．

物体色が背景色と類似している：背景差分により背景であると識別されたにもかかわらず，その近傍で影領域が検出された場合，背景に似た物体が存在していると判断し，異なる背景領域に移動させる改善策を提案する．

物体が鏡面反射を起こしている：入力画像上で輝度値が飽和している領域から判断し，対象物体の移動を依頼する．

物体が暗すぎる：同様に輝度値がしきい値以下である領域から判断する．

4.3.3　提示する改善策の選択

認識の失敗と認識困難な状況の検出処理は独立しており，かつ認識困難な状況が複数発生している場合も考えられるため，上記に挙げた改善策は同時に複数検出される可能性がある．このときは **2.2** における議論に従って優先度が最も高い改善策のみを提示する．本実装における提示の優先度は，表 1 に列挙した $S_{C_1} > S_{C_2} > S_{C_3} > S_{C_4} > S_{E_1} = S_{E_2} > S_{E_3}$ の順である．具体的な提示インタフェースとしては，記号を用いた直感的な提示（図 5）と簡単な文による説明の両者を実装している．提示した改善策に基づいて状況が改善されたにもかかわらず依然として悪状況が検出される場合は，再度，認識状態の提示と状況改善策の提案を行う．

5. 実　　験

協調的認識が効果的に達成されるには以下の項目が適切に機能する必要がある．ただしこれまでに挙げた各種前提は全て満たされているとする．

（1）　悪状況を検出し，改善策の集合 S を推定する
（2）　S から最も効果的な改善策 S_p を選択する
（3）　改善策 S_p が状況を改善し正しい認識を導く
（4）　改善策 S_p の提示により適切なインタラクションを誘導し認識を改善する

本来ならば各項目が適切に機能することを検証した上で，総合的なシステムの評価を行うべきであるが，項目 1，3 に関しては予備実験により確認されているため，紙面の関係上省略する．本論文では，実装したシステムが項目 1，2，3 を満たしているという前提のもとに，項目 4 を含めた協調的認識全体の機能を被験者実験により検証する．**3.3** で述べたように項目 2 についても議論・検証すべきであるが，本論文では扱わず今後の課題としたい．

5.1 実験内容

実験では **4.** で設計した協調的物体認識システムを用いて被験者が実際に認識作業を行い，認識精度及び提示インタフェースの有効性を検証する．認識精度は，認識率及び認識に要する時間を対象として定量的に評価する．また，提示インタフェースの有効性については各提示情報に対する被験者の振舞いを観察することで定性的に議論する．評価項目は，悪状況とその改善策を被験者が正確かつ即座に把握できたか，被験者がシステムに対して自然に協力できたか，などである．実験は以下に示すと二つの認識システムを用いて各被験者ごとに 5 回ずつ行い，その結果を比較した．なお，用いた物体認識アルゴリズムは共通であり，良い条件下では正しい識別結果を出力することは事前に確認済である．

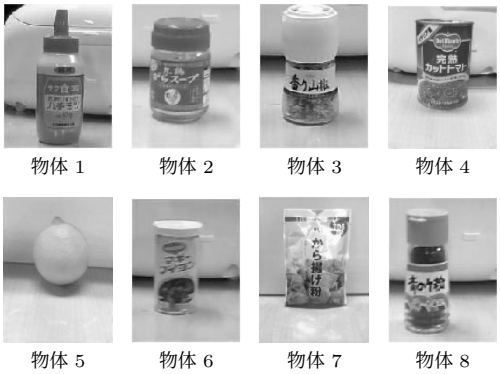

図 6　実験で用いた認識対象物体
Fig. 6　Recognition target objects used in the experiments.

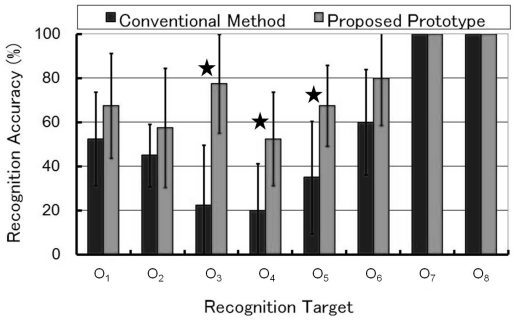

図 7　物体ごとに算出した認識率の平均値と標準偏差．星印は 2 手法の結果に有意差があることを示す（5%有意水準の t 検定）
Fig. 7　Results of recognition accuracy. Stars mean significant differences between results of two methods.

従来システム：対象物体に対して従来の自動認識を行い，最ゆうの認識結果を一つだけ提示する．本来ならば認識結果すら提示されないのが従来手法であるが，それでは被験者が認識失敗にすら気づくことができず改善が必要と分からない．つまり，状況改善行動を誘起するためにあえて認識結果の提示を行った．ただし，失敗の原因を推定するには判断材料が不足しており，認識を改善する方法も被験者自身で推測するしかない．

提案システム：**4.** で述べた協調的物体認識システムを用いる．従来システムと違って認識の状態や状況改善策などが提示されるため，被験者は認識失敗やその改善について多くの情報を得ることができる．

実験は両システムを利用したことのない初心者 8 名の被験者に対して以下の条件下で行った．

- 登録物体は調味料や食材 20 個．
- 認識対象物体は上記 20 個の登録物体から選択した 8 個（図 6）で，各々が特定の悪状況を引き起こしやすい性質をもつ．
- 全被験者で共通の物体を用い，それらを 1 個ずつ順に認識させる．順序は被験者ごとにランダムとする．
- 種々の認識困難な状況を再現するために，認識作業を終えた物体でも作業領域内に残す．
- 各被験者は，両システムを順に用いて実験を行う（被験者内計画）．ただし，全ての被験者に対して従来システムを先に用い，提案システムを経験することによる学習効果が現れないようにする．

5.2　結果と考察

認識対象物体ごとに認識率を算出し，全被験者で平均した結果を図 7 に示す．実利用における負担を考慮し，制限時間（10 秒）以内に正しく識別できれば認識に成功したとみなしている．認識対象物体は各々が特定の悪状況を引き起こしやすい性質をもつため，図 7 の横軸は各悪状況にほぼ対応している．

物体 7，8 は，両システムでともに認識率が 100%であった．これらの物体は，識別に適したテクスチャを有している，他に類似した物体が登録物体中に存在しない，などの理由から，状況が多少悪くても容易に認識できたと思われる．物体 3 は背景に非常に近い色をもつ物体，物体 4 は鏡面反射を起こしやすい物体である．これらの物体は多くの場合正しく特徴量を抽出することが困難であるため，従来システムでは低い認識率となった．一方，提案システムでは特徴量抽出を阻害する状況が改善されやすくなるので認識率が大きく向上したと考える．また，物体 1，2，及び物体 5，6 は互いに似た特徴をもつ物体であり，かつ他の登録物体にも似た物体が含まれている．そのため従来システムでは他の物体と誤認識されることが多かった．提案システムでは候補物体が提示されるとともに，見せ方を変えるという提案がなされるため，被験者の協力の下に認識率が向上したと考えられる．ただし，複数候補まで認識結果が絞り込めているので，従来システムでも正しい認識結果を出力する場合がある．結果として物体 3 や 4 に比べ，相対的に本手法の効果が薄かった（認識率の向上幅が小さい）と思われる．認識率の平均値に対して t 検定を行った結果，有意差の現れた物体を図 7 中の星印で示す．物体 7，8 を除いた全物体では明らかな有意差が見られた．物体ごとでは有意差が出るもの出ないものに分かれたが，総合的には有

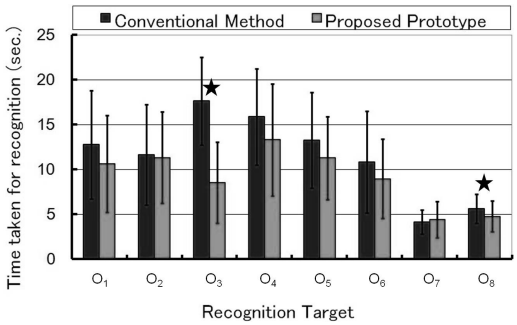

図 8　物体ごとに算出した認識に要した時間の平均値と標準偏差
Fig. 8　Results of time taken to correct recognition.

意差が現れていたことから，サンプル数が足りないことが原因と考えられる．

認識に要した時間についても評価する（図 8）．なぜなら認識率が向上しても認識に長時間要していては使いづらいシステムとなってしまうからである．被験者への情報伝達及び状況改善の実施が必要であるため，提案システムは長時間のインタラクションを要すると当初予想していた．しかし実際には，提案システムを用いることにより認識に要する時間が短縮されるという結果を得た．すなわち，情報伝達の曖昧性を含めたとしても提案手法は効果的であると推察できる．検定結果を見ると，物体ごとでは有意差がまちまちであるが，総合的には明らかな時間短縮がうかがえる．ただし，本実験でではできるだけ早く認識を行うというルールを被験者に課していなかった．そのため，上記考察はあくまで参考と考えてほしい．

実験中の被験者の様子から定性的な解析も行った．認識状態に関する種々の情報や改善策が提示された場合，被験者はそれに気づき積極的に状況を改善する様子が確認された．認識作業全体においては，多数の悪状況が認識を阻害している場面は少なく，たかだか 1，2 回のインタラクションで正しい認識が導かれていた．一方で，認識精度が低い場面や正しい認識に長時間要した場面では，単一の悪状況がなかなか改善されずにいることが多かった．すなわち，改善すべき要因は分かっているものの，指示どおりにしても改善されないという状態である．これは提示された改善策が準最適解ではなかったことを示唆している．単純な施行で改善されないときは具体的な改善策の提示を行う，因果関係や事前知識に基づいたものだけでなく実例から収集された改善策を導入する，などが将来的に考えられる．

陽に提示された改善策以外にも，一度シーン外に物体を取り出してから入れ直すなど自発的に試行錯誤する振舞いも見られた．同時に，被験者が混乱したり戸惑ったりする場合も観察された．例えば，認識候補の中に正解が含まれていない場合では，被験者はどうしたらよいか分からず，物体の見せ方を変え続ける，あるいはそのまま何もしないで提示モニタを見続けるといった場面が見られた．このように人間にしか正しい判断を下せない場合には別途対応法を考える必要がある．提示内容が理解できず，被験者の動きが止まる場面や反応が遅い場面も同様に観察されたが，これはシステムを使用するにつれて出現頻度が減少した．被験者の習熟度と認識精度の関係を調べた別実験では，システムを使用する回数が増えるにつれて，悪状況や改善策の把握が速くなる，改善策の実施が効果的になる，ことにより認識精度が向上する結果が得られている．このことからも，システムの利用法を理解していくにつれて，スムーズなインタラクションが形成されていくと期待できる．

6. むすび

本研究では，人間とシステムがインタラクションを行いながら画像認識の精度を高める協調的認識の枠組みに基づき，それを効果的に形成して認識を改善するインタラクション設計を提案した．提案手法では，システムから種々の情報を提示することで，利用者にかかる負担をできるだけ軽減する．提示情報は大きく分けて認識の状態と状況を改善する方法である．状況改善に関しては最適解ではなく準最適解を適用していくことでを正しい認識を導く手法を考案した．また，本提案の有効性を検証するために，机上作業を想定した上で協調的物体認識のシステムを設計・実装し，被験者実験を行ったところ，単一の認識結果のみを提示する従来手法に対して，認識率が大きく向上することが確認された．

今後の課題としては，協調的認識の枠組みをより高めること，例えば，事例ベースで悪状況の検出や改善策の提案を行う，複数の改善策から現状に最適なものを選択する，認識改善の進行度合や利用者のシステム習熟度に応じた提示を行う，などが挙げられる．将来的にはユーザ支援を行うシステムに協調的認識を組み込むことを考えているが，この場合，認識と支援は完全には独立しておらず，互いに関連して働くと考えら

れるため，種々の議論を必要とするであろう．

文　　献

[1] K. Nigam, A.K. McCallum, S. Thrun, and T. Mitchell, “Text classification from labeled and unlabeled documents using EM,” Mach. Learn., vol.39, pp.103–134, 2000.

[2] M. Ozeki, Y. Miyata, H. Aoyama, and Y. Nakamura, “Collaborative object recognition through interactions with an artificial agent,” Proc. Int. Workshop on Human-Centered Multimedia, pp.95–101, 2007.

[3] B. Suh and B.B. Bedersona, “Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition,” Interacting with Computers, vol.19, no.4, pp.524–544, 2007.

[4] N. Takemura and J. Miura, “View planning of multiple active cameras for wide area surveillance,” Proc. 2007 IEEE Int. Conf. on Robotics and Automation, pp.3173–3179, 2007.

[5] M. Shibata, Y. Yasuda, and M. Ito, “Moving object detection for active camera based on optical flow distortion,” Proc. 17th World Congress 2008, International Federation of Automatic Control, pp.14720–14725, 2008.

[6] M. Guttmann, L. Wolf, and D. Cohen-Or, “Semi-automatic stereo extraction from video,” Proc. Int. Conf. on Computer Vision, pp.417–424, 2009.

[7] 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認識,” 第 20 回日本ロボット学会学術講演会，2002.

[8] M.A. Hossain, R. Kurnia, A. Nakamura, and Y. Kuno, “Interactive object recognition system for a helper robot using photometric invariance,” IEICE Trans. Inf. & Syst., vol.E88-D, no.11, pp.2500–2508, Nov. 2005.

[9] M.A. Hossain, R. Kurnia, A. Nakamura, and Y. Kuno, “Interactive object recognition through hypothesis generation and confirmation,” IEICE Trans. Inf. & Syst., vol.E89-D, no.7, pp.2197–2206, July 2006.

[10] G. Sekhon, A. Kimura, Y. Minami, H. Sakano, and E. Maeda, “Action planning for interactive visual scene understanding based on knowledge confidence in latent spaces,” 信学技報，PRMU-187, 2010.

（平成 22 年 10 月 15 日受付，23 年 2 月 22 日再受付）

近藤　一晃　（正員）

2002 阪大･基礎工･システム科学卒．2004 同大大学院基礎工学研究科システム人間系専攻博士前期課程了．2007 同大学院情報科学研究科コンピュータサイエンス専攻博士後期課程了．同年同大学産業科学研究所特任研究員．2009 京都大学学術情報メディアセンター助教就任後現在に至る．反射屈折光学系，知能ロボット，マンマシンインタラクション，知的行動支援に関する研究に従事．博士（情報科学）．情報処理学会会員．

西谷　英之

2008 京大・工・電気電子卒．2010 同大大学院学工学研究科電気工学専攻博士前期課程了．知的行動支援に関する研究に従事．

中村　裕一　（正員）

1985 京大･工･電気工学第二学科卒．1990 同大大学院博士課程了．同年京都大学工学部助手．1993 筑波大学電子・情報工学系講師．1999 機能工学系助教授，2004 京都大学学術情報メディアセンター教授，現在に至る．博士（工学）．画像理解，映像処理，自然言語処理などの研究に従事．1996 カーネギーメロン大学ロボティクス研究所客員研究員．1998〜2001 科学技術振興事業団さきがけ 21 研究「情報と知」領域研究員（兼任）．情報処理学会，人工知能学会，ACM，IEEE 各会員．