

平成 21 年 6 月 26 日現在

研究種目：基盤研究（B）
 研究期間：2006～2008
 課題番号：18300065
 研究課題名（和文）人間・物体間インタラクションの理解に基づく動作と物体の自動認識
 研究課題名（英文）Automatic Recognition of Human Activities and Functional Tools
 Based on Understanding of Interaction Between User and Objects
 研究代表者
 島田 伸敬（SHIMADA NOBUTAKA）
 立命館大学・情報理工学部・准教授
 研究者番号：10294034

研究成果の概要：

影の影響を減らし、人の動きを検出し、人の持ち去りや持ち込み物体を検出する方法を研究し、さらに物体同士が重なって置かれたり、背後のものが持ち去られることによる複雑な画像上の変化を正しく解釈するため、物体シーンのレイヤ構造化表現を考え、持ち込みと持ち去りを適切に区別方法を確立した。また、道具を掴む手の形を学習させ、手と物体が重なって見える画像から、手の形を物体を認識する方法を提案した。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	4,900,000	0	4,900,000
2007年度	4,900,000	1,470,000	6,370,000
2008年度	3,500,000	1,050,000	4,550,000
総計	13,300,000	2,520,000	15,820,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：物体認識・インタラクション・イベント検地・ジェスチャ・人体姿勢計測・ヒューマンインターフェイス
対話的・領域分割

1. 研究開始当初の背景

人の行動をテレビカメラで撮影して、認識する研究は、米国メリーランド大学で行われたが、屋外における歩行や人との出会いのような大まかな行動の理解であった。また、荷物の持ち去りも研究されていたが、かばんのような大きなものの持ち去りに限定されていた。

部屋内に多くのセンサを配置して人間の行動を観察し、必要に応じてユーザを支援する試みとして、東大のロボティックルーム、産総研のコピキタスルームなどがあるが、事前登録した物体や動作を単独で認識するにとどまっている。

防衛大、慶応大、阪大などでは RFID タグやタッチセンサによって物体の操作を検知し

たが、すべての物体にセンサを取り付けることは現実的ではない。

米ジョージア工科大では人間動作と物体の関連付けに基づく物体認識が行われ、阪府大は物体の機能モデルと人間の動作を関連付けたシーン理解の研究が行われていた、いずれも物体と行動パターンは予め手動登録する必要がある。

また人の意図・動作を理解するインタフェースについてはさまざまな研究が報告されているが、全身の大きなジェスチャ（万歳をしたり、おいでおいで、バイバイといった動作）の認識といった、粒度の荒い認識を応用したものがほとんどであり、詳細な手指動作を理解することによる物体認識への応用はなされていなかった。

2. 研究の目的

本研究では、一般の部屋内のような複雑な画像シーンにおいて、画像中における物体と人間の詳細な動作の見え方を手がかりに物体を認識する方法を研究する。

このとき、棚、椅子、本、カップなどをあらかじめ登録してできる物体はその種類と見えを登録するが、すべての物体を登録しなくてもよいとする。人間の動作についても、掴む、握る、はさむ、といった手指動作から、口元や机の上などに移動させるなど代表的な動作について、3次元人体モデルの運動パターンとして事前登録し、各動作と動作の対象となりうる物体を関連付けておく。

はじめは登録されていない物体や動作の認識は不可能であるが、いずれかが既知であれば、他方を関連付けて新規登録することで認識する。たとえば、形や色は見たことがない物体でも、掴み方と動作がすでに知ったもの（飲む動作）であれば、飲み物を入れる物体（コップ類）としてカテゴリ化できる。また物体と動作の双方が未知の場合にも、物体の画像パターンと手指の3次元形状、動作の運動パターンを画像から計測して、それらのパターンの組み合わせとしてひとまず保存し、類似の組み合わせが一定数収集できた段階でユーザに物体名称や動作の意味を問い合わせ、新規物体・新規動作として登録する。このような前提となる部分知識と経験を融合して自動的に新たな物体を認識できるようにする、認識枠組みのブートストラップを実現することが本研究の目的である。そのようなブートストラップを実現するために、

(1) あらかじめ登録しておく必要のある物体および動作記述によってどのような物体が認識できるのか、

(2) 物体の見えの識別や動作認識自体のあいまいさをどのように解決すればよいかを研究期間内に明らかにする。

また、複数の場所におけるシーン変化の関係を記録しておいて、ある物体への動作が離れた場所のシーンの変化（イベント）を惹起した場合、（たとえば、はさみと紙をもって動作すると二つにわかれる、リモコンを操作するとテレビの画面が変化する、棚の前のカップがなくなると食卓の上に出現するなど）、新規の物体・動作の組み合わせがもつ意味をできるだけ自動獲得すれば、複数の物体や環境との関わりで物体の機能をブートストラップ的に記述することもできる。その際、

(3) どのようなイベントを検知するか、またどのイベント同士をどのように関連づければよいか、ならびにどのような物体機能の記述を獲得できるかを明らかにすることを最終的には、人間が部屋内のさまざまな場所で物体を操作したり持ち運ぶ様子を観察しながら、新規の物体や動作を次々と登録して

いき、部屋内の物体に関するイベントを自動記録し、紛失したもののありかや最後に触った人物をユーザが調査できるシステムを構築する。

3. 研究の方法

1台のテレビカメラとキャプチャボード、PC、映像保存用大規模ファイルサーバを接続した環境で人体の動作画像を長時間にわたって記録した。一般の部屋内のテーブルや棚などを撮影対象とし、背景の照明変化に影響されないように拡張した背景差分法により、物品を操作する人物領域を検出し、その姿勢を推定した。

画像から、人間の顔、腕、手、胴を抽出して、物の持ち込み、持ち去りを検出した。さらに、複数人物による同時操作に対応できるようにした。これに基づいて既知動作のデータベース構築をした。

本や皿、コップなど典型的な物体について物体の登録方法を検討する。照明条件や向きにより見え方がことなることを前提とした方法を検討した。

人物の影がテーブル上や壁に投影される場合に、背景差分としてできるだけ検出されないようにする手法を研究した。

姿勢表現の時系列パターンをモデルとして、歩く、立ち止まる、右手（あるいは左手）の伸ばす、腕をスイングする、ものを指し示す、などの動作を取得し、複数の動作を識別した。このとき、体型の個人差や人体の体や腕の向き、動作速度のばらつきがあることを前提とした手法を検討した。

手元をズームアップする映像取得機構を実装し、手指形状の3次元推定システムを組み込み、物体を持つことによって手の一部が隠蔽されるシーンに対する手指形状の推定法を研究した。

影領域の誤検出があった場合に生じる、物体のレイヤ記述の誤りをジェスチャインターフェースおよび音声を用いて対話的に訂正することで、時間をさかのぼって正しい解釈をやり直す手法を検討した。

既知物体と未知把持パターンの組み合わせ、および未知物体と既知把持パターンの組み合わせを観察することで未知物体、未知動作の意味づけを対話的に行うための手法を研究した。

登録された物体の見えと人間の動作の関連付けに基づく相補的認識の基本枠組み実装と認識実験を行い、ある特定のカップの見えとそれを持つ手形状および飲む動作を登録しておき、見えのちがう別のカップをカップと認識できるか、また同じカップをもった別の人物が単にカップを持ったのか、飲み物を飲んでいるのかを区別できるか確認した。

4. 研究成果

水場横や会議卓の上の映像に対して、ロバス

ト統計処理に基づく背景画像推定手法を用いて、シーン変化イベントの検知を行った。身体構造の知識を用いた認識手法を用いることにより、人間のそばで起こった変化をイベントとしてとらえることができた。

物体同士が重なって置かれたり、背後のものが持ち去られることによる複雑な画像上の変化を正しく解釈するため、物体シーンのレイヤ構造化表現を考えた。シーン変化が検知された際、それがあつた物体レイヤを持ち去ったときに予想されるシーン変化と整合するかを検証することで、持込みと持ち去りを適切に区別できた。これにより、背後に持ち込まれたあるいは背後から持ち去られた物体の解釈が隠れのために曖昧になることを適切に考慮し、手前のものがなくなって背後が観測できた時点で解釈を構築しなおすことができた。

物体持込・持去イベントの時系列をデータベースに蓄積し、簡単な音声・ジェスチャーインターフェースによって、現場で持込・持去を問い合わせると、物体の見えと手指による操作の観察による物体認識では、物体領域、手領域、背景領域の分割結果と手領域のテクスチャ情報を連関させて学習させる方法を提案した。

手形状の詳細な推定において、2通りの手法を考案した。

1) 3次元手形状モデルの当てはめに基づき、背景中からの手領域の切り出しと姿勢推定を同時に行う。この手法では、あらかじめ用意した一般的な手指形状モデルから、どのような輪郭線や内部のテクスチャ(皺や隠蔽による境界線)が生じるかを解析しておき、取得された実画像に対する尤度が最大になる候補を探索することによって実現した。あらかじめどのような照合誤りがありうるかを想定して照合度を評価することにより、従来手法よりロバストに背景から手領域を切り出すことができ、かつ正確に手指の姿勢(位置、スケール、向き、および各関節角度)を推定することができた。

2) 道具を持つことによる背景および隠蔽を含んだ場合に人間の手が物体を把持するパターンを分類し、それぞれを部分空間に繰り返し射影を行うことによる前景画像の修復をおこなう手法を開発した。背景が比較的単純で照明を制御できる環境下では6種類ほどの把持パターンを識別して、コップやペン、カッターなど、手のみあるいは物体のみの映像では判別の難しいものを、手と物体の相互作用を学習することで、不完全な前景を適切に復元しつつ物体識別することができることを示した。

把持物体と種々の把持パターンを認識するため、それぞれの画像を部分空間に基づいて抽出しておく(学習)。認識時

には、可能な把持パターンを仮定して、把持物体の識別を試み、識別結果から復元した推定画像と実画像の誤差が小さいものを把持パターンの認識結果とすればよいことがわかった。

把持パターンと対象物体のいずれかが未知の場合は、部分空間から復元した画像から、復元ができていない部分を抽出し、その部分が何であるかをユーザに尋ねることにより、未知な把持パターンあるいは物体を学習する方法を検討した。

把持パターン識別のため、新しく画像の勾配方向のヒストグラム(HOG特徴)を用いる方法を研究し、認識対象の位置決め精度が十分あれば、多種類の手形状の認識を行えることがわかった。

室内で動いている人物を抽出する場合、影領域を検出する方法を得た。影領域としての条件は、背景画像より暗い、背景画像とテクスチャの正規化相関が高い、および色が類似しているとし、すべての条件を満たすものを影として検出した。この方法により、従来誤っていた領域の大部分を修正することができた。

種々の原因で、物体が誤検出される場合、ユーザとの対話によって訂正する方法を研究した。この場合、物体の変化の履歴をレイヤーとして蓄えておき、これに基づいて誤りの可能性を絞ることによって、ユーザの負担を減少させる方法を確立した。

持ち込、持ち去りの検知と検索を行うプロトタイプシステムにより、数日間にわたるシーンの取得と識別およびユーザによる検索と対話的修正の実験を行い、極端な照明条件や多数の人物の同時出現などを除けば、ほぼ正しく動作することが確認された。

本研究の最終的な目的である、観察に基づく物体の見えと人間の行動の結びつけに関しては十分検討できていない。プロトタイプシステムの実装では、シーン変化が起こったとき、持ち込む、持ち去るといった事前に定義された簡単な動作と、見えが未知の物体の関係について記述するようにした。そこで、本研究の成果である手形状・動作検出をプロトタイプシステムに用い、シーン変化時に検出された物体の見えと手形状・動作の組み合わせを多量に自動収集・分類することによって、動作の意味すなわち物体の機能のラベリングが可能かどうか確かめることが今後の課題である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 16 件)

(1) 今井章博, 島田伸敬, 白井良明, 照合

の誤り予測と見え変化の学習に基づく複雑背景下での手指形状推定,情報処理学会 コンピュータとイメージメディア研究会,2006-CVIM-155-20 2006年 pp.151-158

(2) K. Kawahigashi, Y. Shirai, J. Miura, and N. Shimada, Automatic Synthesis of Training Data for Sign Language Recognition Using HMM, Proc. 10th International Conference on Computers Helping People with Special Needs 2006 pp.623-626

(3) Atsushi Matsumono, Yoshiaki Shirai, Nobutaka Shimada, Takuro Sakiyama, and Jun Miura, Face, Gesture, and Action Recognition Robust Face Recognition under Various Illumination Conditions, IEICE Trans. on Info. and Sys Vol.E89-D, No.7, 2006 pp.2157-2163

(4) 島田伸敬, 三浦純, 白井良明, サービスロボットのためのインタラクティブビジョン,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,Vol. 47, No. SIG15(CVIM 16),2006 pp.1-9

(5) 島田伸敬, 上條真継, 松下浩一郎, 松谷剛, 田中弘美, ハプティックビジョンに基づく関節物体のインタラクティブモデリング 電子情報通信学会論文誌 D-II Vol.J89-D No.9 2006 pp.2013-2024

(6) 片山憲昭, 島田伸敬, 白井良明, 部屋内シーン変遷の自動管理システム ViEW2006 ビジョン技術の実利用ワークショップ講演論文集 2006 pp.72-77

(7) 原田哲也, 伴重尚, 李周浩, 映像投影装置を持つ移動ロボットによる避難誘導システム 第7回 計測自動制御学会(SICE)システムインテグレーション部門講演会 211-1, 2006 CDR0M

(8) 島田伸敬, 片山憲昭, 牧和宏, 白井良明 対話とジェスチャーインターフェースに基づく部屋内シーン変遷問い合わせシステム 第7回 計測自動制御学会(SICE)システムインテグレーション部門講演会 212-1 2006 CDR0M

(9) Hironori Kasahara and Nobutaka Shimada Object Recognition by Observing Grasping Scene Proceedings the 13th Japan-Korea Joint Workshop on Frontiers of Computer Vision (FCV06) 2007 pp.375-378

(10) 浜田康志, 島田伸敬, 白井良明, 手話認識のための複雑背景下で高速に運動する手指の形状推定 電子情報通信学会論文誌 D Vol.J90-D No.3 2007 pp.617-627

(11) 望月智則, 植田勝彦, 島田伸敬, 白井良明 ゴルフスイング診断のための特徴抽出 動的画像処理実用化ワークショップ(DIA) 講演論文集 2007 pp.11-16

(12) 島田伸敬, 白井良明, (招待講演) 人体ジェスチャの画像計測手法とその応用 情報処理学会研究報告 CVIM-158 2007

(13) 大田博義, 木村朝子, 島田伸敬, 田中弘美, Analysis by Reality-Based Simulationに基づく関節物体の力学的機能推定 電子情報通信学会論文誌 D 2007

(14) 大田博義, 木村朝子, 島田伸敬, 田中弘美, Analysis by Reality-Based Simulationに基づく関節物体の力学的機能推定 電子情報通信学会論文誌 D Vol.J90-D No.7 2007 pp.1799-1811、査読有

(15) 今井章博, 島田伸敬, 白井良明 複雑背景下におけるモデルの照合誤りを考慮した手指形状推定 電子情報通信学会論文誌 D Vol.J91-D, No.3 2008 pp.784-792、査読有

(16) K.Morioka, J.-H.Lee, Y.Kuroda and H.Hashimoto Hybrid tracking based on color histogram for intelligent space Artificial Life and Robotics Vol.11 No.2 2007 pp.204-210、査読有

【学会発表】(計 9 件)

(1) 松尾直志, 白井良明, 島田伸敬 手話認識のための状態遷移構造推定 画像の認識・理解シンポジウム(MIRU)予稿集 2007年7月30日-8月1日 広島、日本

(2) 今井章博, 島田伸敬, 白井良明 複雑背景下におけるモデルの照合誤りを考慮した手指形状推定 画像の認識・理解シンポジウム(MIRU)予稿集 2007年7月30日-8月1日 広島

(3) 片山憲昭, 牧和宏, 島田伸敬, 白井良明, 動画に基づく部屋内シーン変遷の自動検出と対話的イベント検索システム 画像の認識・理解シンポジウム(MIRU)予稿集 2007年7月30日-8月1日 広島、日本

(4) 新庄 智子, 山田 寛, 松尾 直志, 白井良明, 島田 伸敬, HMMを用いた手話認識のための学習データの自動合成 ヒューマンインタフェースシンポジウム 2007年9月3日-6日 東京、日本

(5) A. Imai, N. Shimada and Y. Shirai, Hand Posture Estimation in Complex Backgrounds by Considering Asian Conf. on Computer Vision (ACCV) 2007年10月18日-22日 東京

(6) M.Ueda, Y.Shirai, M.Oonuki, Automatic Diagnosis System of Golf Swing 3rd Congress on Sports Technology 2007 Singapore

(7) Kazuhiro MAKI, Noriaki KATAYAMA, Nobutaka SHIMADA, Yoshiaki SHIRAI Image-Based Automatic Detection of Indoor Scene Events and Interactive Inquiry, FCV2008 2008年1月25日 大分、日本

(8) 宮下智至, 李周浩, 環境非固定型パンチルトプロジェクトの姿勢変化に起因する

歪みの補正手法 計測自動制御学会システム
ムインテグレーション部門講演会, 3L2-5
2007

(9) 李周浩, 移動投影装置を用いた強化現
実システムの開発 日本機械学会ロボティ
クス・メカトロニクス講演会 2007年5月10
日-12日秋田, 日本

6. 研究組織

(1) 研究代表者

島田 伸敬 (SHIMADA NOBUTAKA)
立命館大学・情報理工学部・准教授
研究者番号: 10294034

(2) 研究分担者

白井 良明 (SHIRAI YOSHIAKI)
立命館大学・情報理工学部・教授
研究者番号: 50206273

田中 弘美 (TANAKA HIROMI)
立命館大学・情報理工学部・教授
研究者番号: 10268154

李 周浩 (LEE JOOHO)
立命館大学・情報理工学部・准教授
研究者番号: 80366434

(3) 連携研究者