

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 16 日現在

機関番号：34315

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500224

研究課題名(和文) 指使いの観察によるモノの機能分類能力の創発

研究課題名(英文) Emergence of Discriminating Object Function by Fingering Observation

研究代表者

島田 伸敬 (SHIMADA, NOBUTAKA)

立命館大学・情報理工学部・教授

研究者番号：10294034

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：指使いと物体の見えの関連性を事例から学習する枠組みを提案した。初見の物体の見えを与えると、手指がその物体をどのように把持するかを想起できることを実験的に示した。

1) 手指Joint局所特徴をHough Forest法による手首位置を推定し被把持物体領域を推定2) 物体把持のインタラクション状態を表現する把持状態空間をSparse Stacked Convolutional Autoencoderによるクラスタリングを応用して獲得3) 把持状態表現と物体の見えの関連をConvolutional Neural Netを用いて学習して初見の物体の見えからその把持状態を想起

研究成果の概要(英文)：This research finally proposed a framework of learning relationship between 'fingering' and object appearances from samples which enables to infer the functional class of objects and remind the fingering and grasping way for a given object appearance, and experimentally showed its availability:

1) The method of estimating wrist position based on Hough Forest voting of Joint hand image features,2) The method of segmentation of the grasped object based on suitable grasping pattern bound to a certain object function class,3) The method of constructing 'grasping pattern feature space' which describes interaction states in object grasping by hand by using Sparse Stacked Convolutional Autoencoder,4) The method of reminding the grasping approach by hand from stimulus of a object appearance by using Convolutional Neural Network .

研究分野：画像認識

キーワード：指使い 画像認識 把持状態空間 機械学習 想起 CNN プロセスモデリング

1. 研究開始当初の背景

人間の生活環境で、体の不自由な人の代わりにモノを取ってきてくれたり、探している人やモノを見つけてくれるロボットが望まれている。その実現のために画像に基づく特定物体認識、一般物体認識の手法が数多く提案されているが、その多くは特定の物体のカテゴリに属する学習サンプルを手動で収集しており、新しいモノが増えたときに対応できなかった。また対象も人体全体の大まかな動作に留まり、細かな指使いは扱えていなかった。一方で、人間が手で物体を操作する時の指使いとその物体の機能は深く関連していることが指摘されており、人間の指使いとモノの形を突き合わせてみることで初めてモノの機能や意味を特定できる可能性があり、人間が物体を操作する際の手指の姿勢や動きを詳細に観察することができれば、物体認識の手がかりとなりうる。そのような人間の動作と物体の機能の関連を創発的に獲得する枠組みの研究として、本課題は適切なフィールドとなりうると思われた。

2. 研究の目的

本研究では、機能をもった道具の見えおよび形と、それを操作する人間の指使いを画像認識の技術をつかって解析することによって、機能の不明なモノ(物体)を、指使いの種類によって自動的に機能をカテゴリ化する。さらに、モノの色・模様・形のようなさまざまな画像特徴のうち、なにが道具としての機能と関連しているのかを抽出し、同じ機能をもつ初見の物体を検出する創発的な画像認識枠組みを提案することを主な目的とした。具体的には、卓上における生活シーンを想定した上で、1)モノの見えと形、2)モノを操作する指の詳細な姿勢と動き、を手がかりに物体を機能に基づき識別する枠組みについて研究する。「コップで飲む」「本をめくる」「鋏で切る」というような物体の機能に固有な一連の指使いを過去の視覚経験から獲得し、モノの見え、形、および指使いの共起性を一体的に記述したモデルを生成し、このモデル照合することによって物体の機能を分類する。

3. 研究の方法

まず研究課題を2つの軸にそって分解した。一つ目の軸は「静的な道具の把持」→「動的な動作による道具使用」の観点で、手指の姿勢は非常に複雑なパターンを呈するので、静的な把持パターンを単一の道具把持画像の集合から分類するための特徴をいかに取得するか、という研究と、道具を使用する際の姿勢の時間的変遷＝プロセスをマクロに捉えてそれをどのようにモデリングし照合するか、という研究にわけ並列に検討することとした。また二つ目の分解軸として、手指そのものの姿勢を記述する際の粒度を段階的に表現す

ることとした。すなわち、「手指の「見え」を特徴とみなして記述」→「手指の三次元関節モデルの内部パラメータを特徴として記述」という記述の詳細度の観点である。以上の観点で研究の具体的着手課題を次の4つのサブ課題に分けた。

- (1) Convolutional Autoencoder-decoderを用いた手・物体インタラクションの学習と想起
- (2) 階層的イベント検知構造に基づく物体登録検知システムとジェスチャキーによる類似機能物体の検索
- (3) 糸結び手技の自動訓練のためのプロセスモデリング
- (4) 指運動リハビリテーションのための深度センサによる指先可動範囲の自動計測と可視化

(1)は全体課題の中心的課題であり、静的な画像+見え特徴を入力とし、物体把持の様相＝物体自体の見えと手指形状の見え、を記述するための特徴空間の自動抽出と、それを用いた新規物体の見えから物体把持の様相を想起する枠組みを検討した。(2)と(3)は道具を扱う動的なプロセスを記述して照合することに焦点を置き、前者はテーブル上に持ち込まれた物体を検知しつつ、その物体を使用したときの全身動作をモデリングし照合することで、物体機能を分類した。後者では手術シーンにおける糸結び手技を題材に、深度センサを用いて繊細な手指の動作をモデリングし照合する枠組みについて検討した。

(4)では(3)で用いた三次元計測技術のスピンアウトとして、指先の詳細な可動範囲を可視化する技術に応用した。

4. 研究成果

(1) Convolutional Autoencoder-decoderを用いた手・物体インタラクションの学習と想起

物体を把持した状態では手と物体が相互に隠蔽し合うため、手の全体像において、検出や姿勢推定、領域分割を正しく行う事が一般的に困難である。そこで、手の部分特徴と手首位置の位置関係を学習する事で、隠れがあっても把持状態を推定できるシステムを提案する。道具ごとに存在する典型的な持ち方を学習して、把持パターンに対応する物体領域を予測する。予測した物体領域から検出された手首の位置と物体の重心位置を利用して手首-物体座標系を設定し、物体領域から手首位置を推定する。手首-物体座標系を基準に正規化した道具の把持画像を利用して、道具のみを写した画像を持ち方ごとに分類し、持ち方を表すパラメータと道具のみの画像を組み合わせて対応する把持パターンを想起する。

具体的には、画像中の手に現れる2点のSURF特徴と位置関係を組み合わせたJoint特徴と手首位置の相対位置関係をRandomized Treesを用いて学習し、手首位置

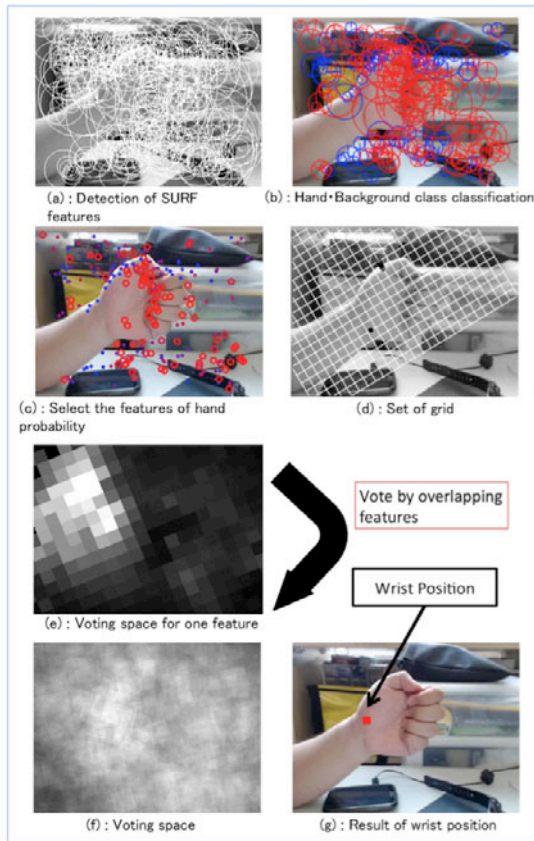


図 1 手首位置の検出

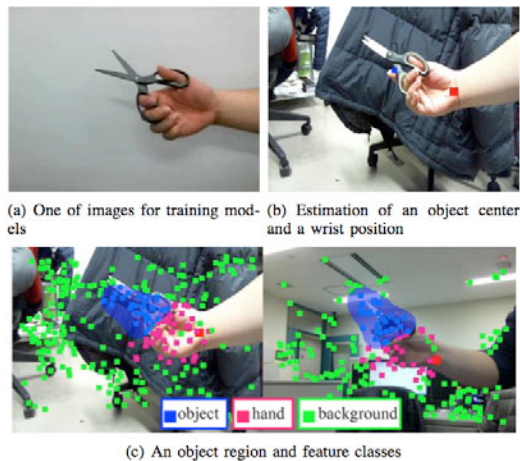


図 2 持ち方に基づく道具領域の自動切り出し
 確率分布の合算から手首位置の検出を行う (図 1)。手首位置の検出と同様に、手の場所に現れる SURF 特徴と物体の重心位置との相対位置関係を Randomized Trees を用いて学習することで、物体位置確率分布の合算から物体位置検出を行い、手首-物体座標系を設定する。手首-物体座標系を基準に、正規化した道具の把持画像 (図 2) から、道具ごとの持ち方の様相を表現するパラメータ (持ち方パラメータ) を抽出するように Stacked Convolutional Auto-Encoders を用いて学習を行う (図 3)。抽出した持ち方パラメータを教師信号として、Convolutional Neural Network を用いて道具のみ画像から持ち方パラメータを予測する学習を行う。予想した持ち方パラメータを Stacked Convolutional Auto Encoders と対になる decoder に入力す

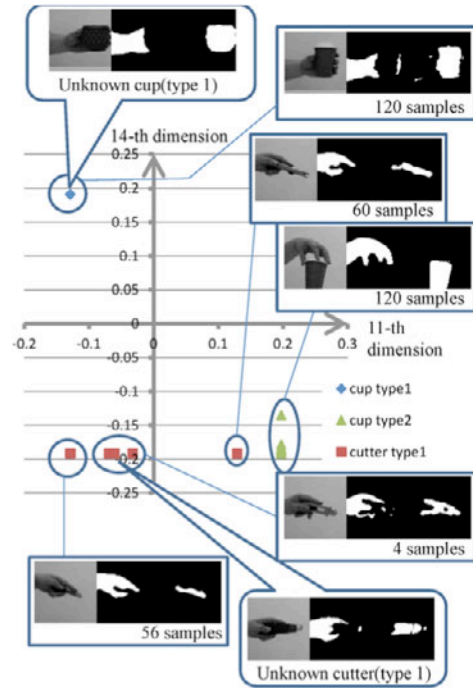


図 3 持ち方パラメータ空間の自動獲得

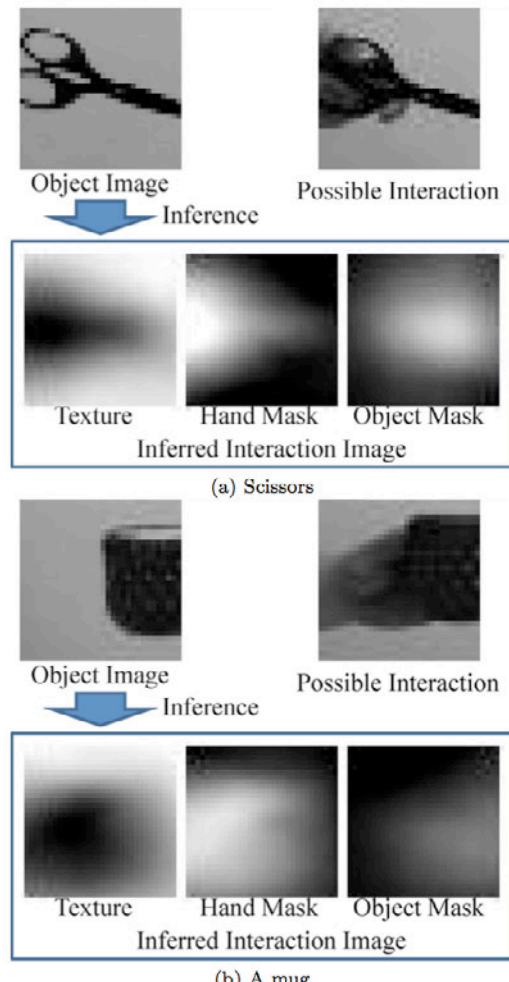


図 4 道具の見えからの持ち方の自動想起の例
 ると、道具に対応する把持パターンが想起される (図 4)。手首位置と物体位置の検出精度実験と把持パターンの想起実験を行い、提案手法の有効性を示した。

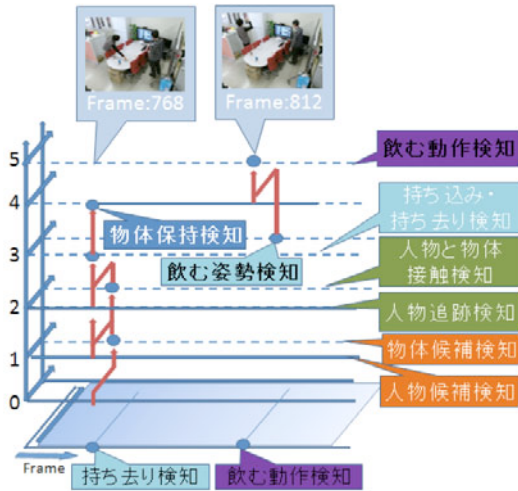


図 5 階層的イベント検知:「飲む」イベントの構成例

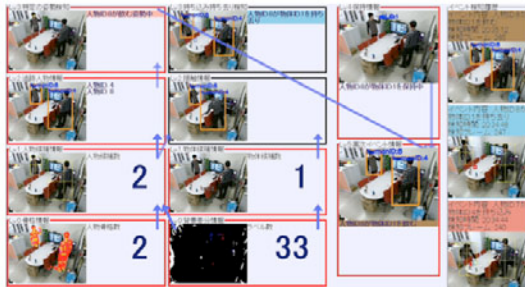


図 6 「飲む」イベントの検知例

(2)階層的イベント検知構造に基づく物体登録検知システムとジェスチャキーによる類似機能物体の検索

人物の行動と物体の関わりを検知することができれば、人物の行動をもとに物体の機能に基づく分類ができる。例えば飲み物を飲む動作や、シーン内にある物体を服やかばん等にしまう動作をした時、その物体がどういった時に使う物かや誰が持っていた物なのかを物体の検索時に利用できる。このような人物の特定の動作の検知についての研究は多くなされているがほとんどは事前に当該人物動作の特徴を学習させ、それと一致するかどうかで検知する必要がある。

このように複雑な動作の検知に加え、様々な動作をシステムで検知させるには膨大な学習用のモデルを集める必要がある。本研究で開発した物体操作システムは画像センサにより得られる背景差分や動きなどの低次のイベント情報を元に階層的にそれらを積み上げることによって、持ち込み・持ち去り・移動、接触、把持などの中間的イベント、さらに飲み物を飲む、かばんに物を隠す、といったより高次のイベントを検知する手法を提案した(図5)。これにより複雑なイベントをそのイベントより下の階層の情報の組み合わせにより検知することで、学習や作りこんでおくモジュールを共有化し、事前の準備を最小限にとどめながら、システムをより発展させる余地をもった検知システム構築法を提案する。

高次動作のモデリングでは、kinect 深度センサより得られる人体スケルトンの時系列情報を、Dirichlet Process Model によるキー姿勢抽出と階層化 Pitman-Yor 言語モデルを用いた動作の分節化を行うことでパターンの記号化を行うことでモデル記述を行う。シーンに持ち込まれる物体と動作を関連付け記録するシステムを試作し、そこにこの高次動作モデルとの照合モジュールを追加した。このとき、物体持ち込みや接触・把持といった低次イベントを検知の必要条件に設定することで検知のロバスト性を向上させた。これにより飲む動作をキーにしてペットボトルの持ち込みシーンを検索するインターフェースを実現できた(図6)。

(3)糸結び手技の自動訓練のためのプロセスモデリング

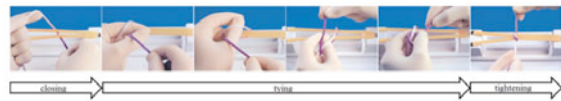


図 7 糸むすび手技

外科手術における手術手技の訓練システムという具体的な課題場面を設定して、傷縫合時の糸結び手技(図7)を比較同定し、教師の模範動作に対して異なる箇所を指摘する課題を設定した。糸結びは対象の糸を両手で繰り返しながら一方の端で輪を作ってからもう一方を通し、その後端同士をひっばる、といった手指自体と対象物=糸の状態変化を伴う一連のプロセスとして定義できる。手指の複雑な隠れに対応するために kinect 深度センサを用いて三次元点群情報を得、各糸結び時系列データをモデリングする。

kinect センサにより時系列で三次元点群を取得し、その三次元形状特徴(ここでは 3D SHOT 特徴を採用した)を用いて k-means クラスタリングにより生成した特徴単語辞書に基づく bag-of-words 特徴によって手の形状分類を行った。これに基づいて典型的な糸結び手技の各プロセス状態を分節記述しておき、入力テストシーケンスとの照合を行って手技の進捗状況を識別することを試みた。その結果単純な SHOT 特徴の利用では分節化や照合が不安定であり、手技プロセスの識別には記述性能が不足することがわかった。そこで、教師の模範動作の深度センサ情報の時系列フレーム自体を直接モデルとし、訓練者動作の入力フレームをモデルフレームと照合する事例ベースのモデリング手法を採用した。

訓練動作に対して模範動作との差異を指摘する、という応用に鑑みれば入力時系列の途中に本来の動作とは異なる揺れ動作が挿入されることを想定しなくてはならず、これが本研究の課題である。そこで、入力時系列の各フレームに対しモデル動作には含まれない未知のフレームへの照合パスを導入した動的計画法による照合方法を提案する。これにより数種類の糸結び動作を別人物間で

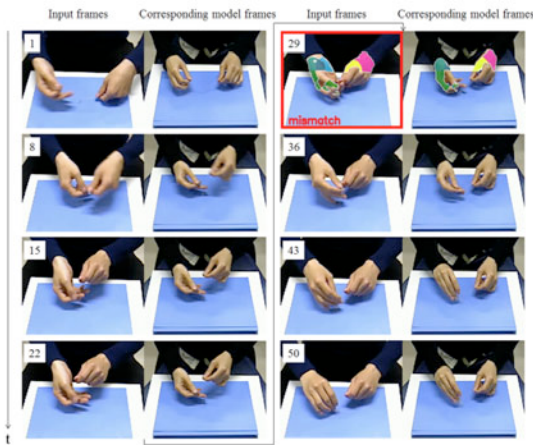


図 8 結び動作タイプの同定とモデルとの差異動作の検知

照合し時空間的に照合できない部分を実際に指摘できることを示した(図 8)。これによる糸結び訓練システムの試作に向けてモデルと入力系列の照合のロバストネスを向上させることが課題として残った。

(4) 指運動リハビリテーションのための深度センサによる指先可動範囲の自動計測と可視化

手首や指の骨折時には手術の影響により一時的に手指の運動を制御する神経や筋肉の動きに障害が発生することがある。とくに骨折の治療過程においてギブス等の装具によって関節を固定することで関節が固まってしまうことがあり、その防止や回復にリハビリテーションによる運動が欠かせない。その際関節の運動可能範囲がどの程度なのかを実測するために、リハビリの現場ではものさしや分度器を直接手指にあてがうことによって計測している実態がある。これを非接触の画像計測によって実現できれば、より計

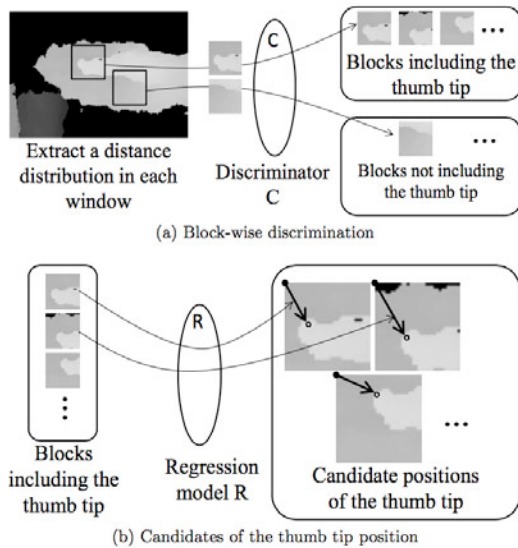


図 9 深度センサ情報の機械学習に基づく指先検出

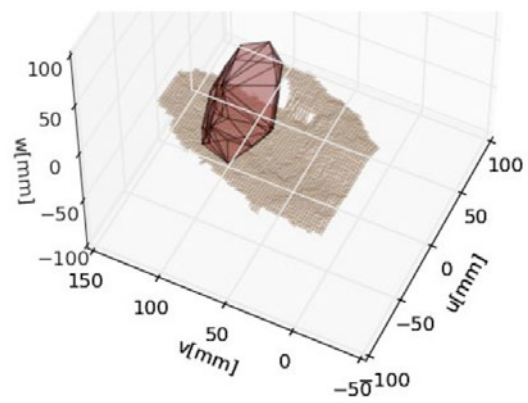


図 10 親指先の三次元可動範囲の可視化

測精度を高めかつデジタルデータをグラフィックスによって可視化することで回復の度合いを時系列的に負うことができ、治療者、患者の双方にとって治療効果を直観的に共有することができるようになる。

そこで本研究では前述までの手指画像特徴の検出技術をスピノフさせ、とくに kinect 深度センサによる距離情報を用いることで、指先の三次元位置を計測する手法を開発した。その設計仕様として、動きの自由度が大きく手指による動作の中で重要な役割を果たす親指先を対象とし、測定時に手のひらが多少動いてしまうことを許容するために手のひらの向きを同時計測して、手のひらに対する相対的な親指先位置の軌跡を求める手法を開発した。指先の深度センサーパターンをテンプレートとして持つだけだとその照合に失敗することがあるため、正解位置の周辺のパターンに対して正解位置へのオフセットを回帰する randomized trees を学習させ、投票によって親指位置を求める手法とした(図 9)。手のひらの位置と向きを同定するために、人差し指、小指の付け根部分と手首位置の3箇所の深度パターンについて、親指先と同様の回帰モデル学習を行った。これにより手のひら相対の親指位置を計測し、得られた点群の凸包を計算して可視化することができた(図 10)。

(5) 今後の課題

物体の見えと把持パターンを統合した持ち方パラメータ空間を機械学習により自動的に獲得することができたが、2-D の見え画像そのものを用いているため、各指や手のひらがどのように用いられているかなどの物理的な現象を正しく反映したモデルにはまだなっていないのが現状である。そこで、切り分けたサブ課題で用いた深度センサ情報を組み合わせモデル獲得を試みるのが今後の課題である。また動作の時系列プロセスを含めた動的な持ち方パラメータ空間をどう構成するかも課題として残っている。これらを組み合わせる動的な状態変化を含む動作プロセスを認識しロボットなどが再現することまでは今回の課題期間中にはたどり着かなかったため、これについて次期の研究

課題で実施したいと考えている。

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 3 件)

1. Shinya Morioka, Tadashi Matsuo, Yasuhiro Hiramoto, Nobutaka Shimada, Yoshiaki Shirai, “Automatic Image Collection of Objects with Similar Function by Learning Human Grasping Forms”, Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, Lecture Notes in Computer Science, 査読有, Vol. 8869, Springer International Publishing, 2015, pp. 3-14,
2. 松尾直志, 山田寛, 白井良明, 島田伸敬, “HMM を利用した画像処理による手話単語の認識のための特徴抽出および状態分割”, ヒューマンインタフェース学会論文誌, 査読有, Vol. 15, No. 1, 2013, pp. 85-94
3. 神山泰典, 松本俊昭, 白井良明, 島田伸敬, 植田勝彦, “ゴルフスイング診断のための画素値の時間ヒストグラムとグラフカットに基づく人体シルエット抽出”, 電気学会論文誌 C(電子・情報・システム部門誌), 査読有, Vol. 132, No. 11, 2012, pp. 1840-1846, DOI:10.1541/ieejieiss.132.1840

[学会発表] (計 24 件)

1. SIQI JIA, 松尾直志, 島田伸敬, “3-D 手指姿勢推定のための掌の表裏識別に基づく左手右手の自動判別” 電子情報通信学会 2015 総合大会, 立命館大学 (滋賀) 3. 10. 2015
2. 川本祥悟, 池上貴之, 川北真也, 島田伸敬, “階層型イベント検知に基づく人と物の関わり履歴の自動記録システム”, 電子情報通信学会 2015 総合大会, 立命館大学 (滋賀) 3. 10. 2015
3. 松尾直志, 島田伸敬, “手指リハビリテーションのための RGB-D センサーを用いた親指の 3D 稼働領域計測”, 第 2 回看護理工学会学術集会, 大阪大学 (大阪), 2014 年 10 月 4 日
4. Shinya Morioka, Tadashi Matsuo, Yasuhiro Hiramoto, Nobutaka Shimada, Yoshiaki Shirai, “Automatic Image Collection of Objects with Similar Function by Learning Human Grasping Forms”, The 3rd International Workshop on Multimodal pattern recognition of social signals in human computer interaction

(MPRSS2014) (ICPR 2014 Satellite Workshop), Stockholm Waterfront Congress Centre, Stockholm, Sweden, 8. 24, 2014.

5. 森岡慎也, 島田伸敬, 松尾直志, 白井良明, “Joint 特徴に基づく手と物体の関係性記述を用いた把持状態の検出”, 第 17 回画像の認識・理解シンポジウム (MIRU2014), 岡山コンベンションセンター (岡山), 2014. 7. 29
6. 森岡慎也, 平本康裕, 島田伸敬, 松尾直志, 白井良明, “局所特徴の位置関係の学習に基づく複雑背景下での把持状態の検出”, 計測自動制御学会 システム・情報部門 学術講演会 SSI2013, ピアザ淡海 (滋賀県大津市), 2013. 11. 20
7. Takafumi Marutani, Kazuyoshi Tagawa, Nobutaka Shimada, Hiromi T. Tanaka, Yoshimasa Kurumi, Masaru Komori and Sigehiro Morikawa, “Trainee’s Action Archiving System for Supporting Laparoscopic Surgery Training”, Proceedings of the 3DSA2013 International Conference on 3D Systems and Applications, poster, Grand Front Osaka, Osaka, June 28, 2013
8. Tadashi Matsuo, Yoshiaki Shirai, Nobutaka Shimada, “Construction of General HMMs from a Few Hand Motions for Sign Language Word Recognition”(poster), IAPR INTERNATIONAL CONFERENCE ON MACHINE VISION APPLICATION (MVA2013), Suzaku Campus, Ritsumeikan University, Kyoto, 2013. 5. 20
9. Sho MIYAMOTO, Tadashi MATSUO, Nobutaka SHIMADA, Yoshiaki SHIRAI, “Real-Time and Precise 3-D Hand Posture Estimation Based on Classification Tree Trained with Variations of Appearances”, 21st International Conference on Pattern Recognition (ICPR 2012), Tsukuba International Congress Center, (Tsukuba Ibaraki), November 12, 2012.

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・教授
研究者番号: 10294034