

平成 30 年 6 月 6 日現在

機関番号：34315

研究種目：基盤研究(B) (一般)

研究期間：2015～2017

課題番号：15H02764

研究課題名(和文) 指使いの自動プロセスモデル解析に基づく緻密手作業のロボットハンド模倣

研究課題名(英文) Automatic Process-Modeling of Precise Fingering and Its Application to Robotic Hand Operation

研究代表者

島田 伸敬 (Nobutaka, Shimada)

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) 11,600,000円

研究成果の概要(和文)：物体操作の指使いプロセスをモデリングする手法の研究を行った。課題を(1)手指姿勢や指先位置の精密な画像計測、(2)物体の形状に基づく把持パタンの想起、(3)操作による物体および手の状態遷移の記述モデル、(4)物体状態に基づく次の取るべき行動の想起、(5)ロボットハンドの制御、の各部分問題に還元し、それぞれ個別に検討した。物体と手指の間関係性を、人の行動観察から静的ならびに動的時系列として獲得する枠組みとして、深層学習に基づく学習・想起の道具立てを開発することで上記の部分問題に共通して適用し、制約はあるもののプロセスモデリング手法の端緒を提案することができた。

研究成果の概要(英文)：We studied modeling method of fingering process of object manipulation. We broke down whole problem into 5 sub-problems:(1) precise measurement of fingers postures and fingertip positions from image,(2) recall of human's grasping patterns based on 3-D shape of objects by DNN,(3) describing fingering process in object manipulation as state transition model,(4) recall of next action to be taken which causes required scene/object state change,(5) controlling the robot hand based on recalled grasping pattern.We applied DNN frameworks including CNN and RNN commonly to the above sub-problems and then acquired relationships between human action and object's state as static frame and dynamic sequence from human behavior observation. We proposed the first step methods to build process model of human's operation of object, under several limitations.

研究分野：コンピュータビジョン、知能ロボティクス

キーワード：深層学習 CNN 持ち方パラメータ 物体機能 想起 プロセスモデリング

1. 研究開始当初の背景

モノの種類や機能を自動認識するロボットビジョンでは、近年非線形パターン識別器や深層機械学習を応用した成果が目立っているが、主に静的な物体の形状や見え方に着目した認識能力の開発に主眼がおかれてきた。しかし鎌倉が指摘したように人が手でモノを操作する指使いとそのモノの機能は関連しており、動きを伴う指使いとモノの形を突き合わせて初めてモノの機能を発現させる動的過程が理解できる。さらに、ロボットが人と同じ環境で道具類を人と共有し使用するには、人がモノを操作する指使いをまねることが一番の近道である。

近年では、ロボットの構造が既知の場合に視覚・力覚フィードバックを用いて固形物や紐などの対象を把持し操る制御手段の研究も進められてきた。これらの数理的な制御モデリング手法は特定の状況下での把持など、基本的な「状態維持」動作の実現に留まり、例えば紙箱の蓋をこじあけて中身を取り出すなど、対象の状態を順番に変化させていく操作は扱えていなかった。

手指の操作形態と道具の見えの関係性に着目して物体認識に活かそうという試みは近年精力的に研究されてきた。これらはいずれも大まかな手の配置と道具および環境の大雑把な位置関係を利用するに留まり、指使いを詳細に観測して物体機能を認識する研究はまだほとんどなかった。

2. 研究の目的

本研究は、人の指使いを詳細に計測する画像認識技術を応用して、緻密で複雑な指使いを自動的にロボットが獲得するための計算論的枠組みを明らかにすることを目的とする。とくに複数の指使いの連鎖による対象物体の状態変化の遷移をプロセスモデルとしてとらえ、おのおのの状態遷移を引き起こす指使いに着目してモデルを構成することにより、ロボットが人と同様の物体操作を実現する足がかりとすることを目的とした。

- 対象の形状が複数の状態変化を伴う緻密な物体操作のプロセスモデルの表現方法とその取得方法を明らかにする。
- 操作対象の状態変化を計測/記述する手法、それを手の動作と結びつけたモデリング手法を検討する。
- 物体操作のプロセスモデルは、人とは必ずしも一致しない(指の本数が足りない、大きい)ロボットの身体性に転写することを念頭に記述される必要があり、その方法を明らかにする。
- 紐結び、紙箱の蓋あけなどの例についてプロセスモデリングとロボットへの動作転写を試み、評価や問題点の分析を行う。

3. 研究の方法

物体操作のプロセスモデリング手法を開発するに当たり、全体の問題を(1)手指姿勢や指先位置の精密な画像計測、(2)物体の形

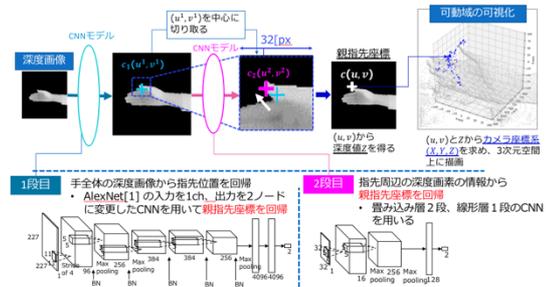


図1 Coarse-to-Fineによる他段階CNN回帰に基づく深度画像からの親指先位置の推定

状に基づく把持パタンの想起、(3)操作による物体および手の状態遷移の記述モデル、(4)物体状態に基づく次の取るべき行動の想起、(5)ロボットハンドの制御、という各部分問題に分解し、それぞれについて個別に検討する方策で臨んだ。ただし物体(シーン)と手指(人体姿勢)の間の関係性を、人の行動観察から静的ならびに動的時系列として獲得する枠組みとして、深層学習に基づく学習・想起の道具立てを開発することで、上記の部分問題に共通して適用することができた。

4. 研究成果

- (1) 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

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し、計測精度を高めかつデジタルデータをグラフィックス

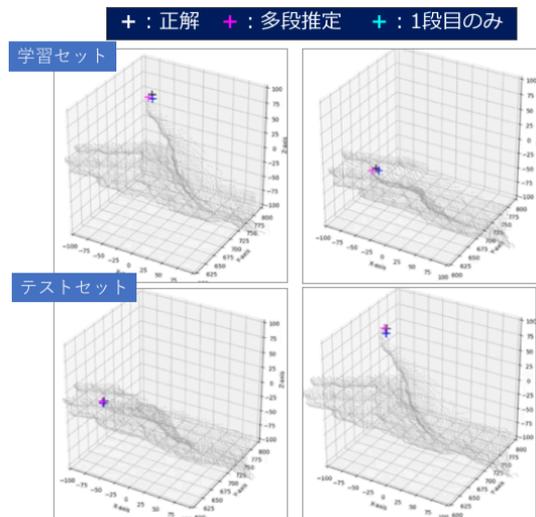


図2 推定された親指位置の3-D空間上への投影

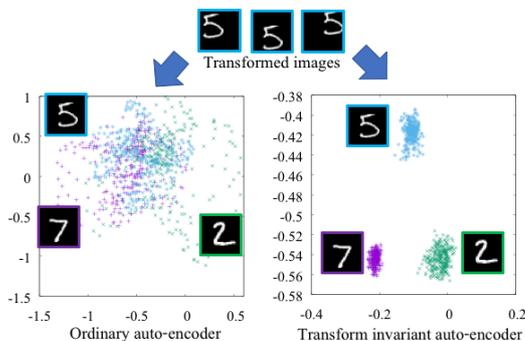


図 3 Transform-Invariant Auto-Encoder によるクラスタリング

によって可視化することで回復の度合いを時系列的に追い、治療者、患者の双方が治療効果を直観的に共有することができる。

本研究では kinect 深度センサによる距離情報を用い、親指先の三次元位置を計測する手法を検討した。先行研究において指先周辺の深度パターンと指先位置を randomized tree をつかって回帰したが、その検出推定精度が不十分であった。そこで本研究では Convolutional Neural Network (CNN) を用いて指先位置を深度パターンから回帰することを試みた。また最初は治具に固定した手全体の深度画像から親指周辺を粗く検出し、次に検出された領域について親指先の画像内位置へのオフセット  $(u, v)$  を回帰する Coarse-to-Fine 手法 (図 1) を用いて  $\pm 3$  画素程度に精度を向上することができた。推定された 2D 座標の深度画像の値を用いて 3次元座標を計算し可視化し結果を図 2 に示す。2D 座標  $(u, v)$  を経由するのではなく直接 3次元の  $(X, Y, Z)$  座標を回帰する手法も同時に検討したが、こちらは奥行きに関するサンプルの少なさから 2D 推定を上回る精度を達成できなかったため、より深い層構造の深層学習モデルを用いるなど引き続き手法の検討を続ける予定である。

(2) Transform-Invariant Auto-Encoder の開発とそれに基づく物体形状／把持パタンの関連性学習と想起

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、物体の見えと把持パタンの対応関係を事前に学習しておくことで、物体から把持状態を推定する枠組み提案した。先行研究では 2D の物体濃淡画像から把持パターンを回帰したが、本研究ではこれを 3D の物体深度画像に拡張して、物体形状と物体ごとに存在する典型的な持ち方の対応を学習させた。これを深層学習のいち手法である Auto-Encoder/Decoder を用いて実装したが、カメラと物体の位置関係や物体の姿勢にバリエーションが存在するため、Encoder による符号化がそれらに強く依存してしまい、物体形状や把持パターンを適切にクラスタリングすることができなかった。

そこで本研究では、撮影時に生じるバリ

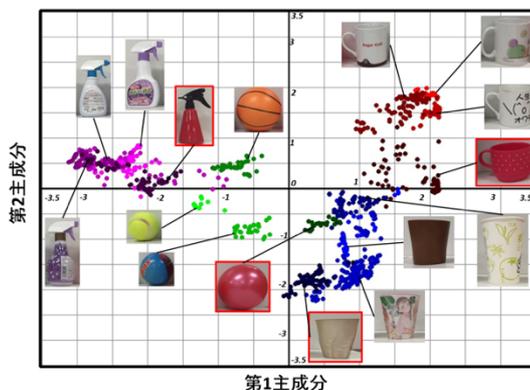


図 4 物体ごとの把持パターンを符号化した特徴量空間におけるクラスタリングの様子

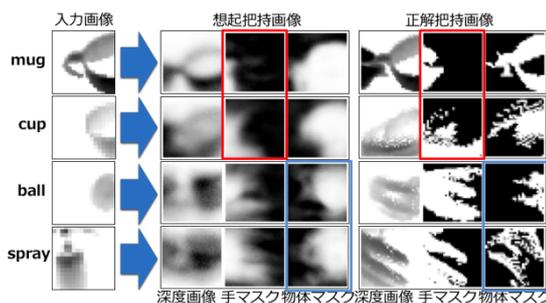


図 5 部分物体画像からの把持パターン画像の想起結果

エーションを画像変換クラスと捉え、変換クラスに不変な画像特徴を自動的に抽出する Transform-Invariant Auto-Encoder を開発した (図 3)。この Auto-Encoder は予め想定したクラスの変換を施した画像を訓練画像から生成し、符号がそれらの変換に不変な場合に損失が小さくなるような損失関数を設計することで実現される。

この新たな Encoder によって物体把持をしている深度画像、手領域マスク、物体領域マスクを各チャンネルに格納した多チャンネル画像を符号化し持ち方クラスタを得る (図 4)。符号化された把持特徴量をさらに物体だけの深度画像と対応付けした CNN を別途訓練することで、物体深度画像から持ち方を想起させた。

このとき物体の把持様式は物体のどのパーツを持つかによって複数の可能性があるため、物体全体をつかっただけでクラスタ化・想起を行うとそれらが干渉してしまうことがわかった。そこで、物体の各パーツにフォーカスした部分画像を切り出し、それらに対して部分把持パターンを対応付けて同様の訓練をおこなったところ、各パーツの 3D 形状に応じた持ち方をクラスタリングし、想起することができた (図 5)。また、共通の持ち方符号を想起する画像領域をマージしていくことで、物体全体のうちどの部分がどの持ち方をサポートしているかを表示することもできた (図 6)。

(3) 正常／異常状態対に基づく状態遷移モデリングおよび結紮手技中の異常動作検出

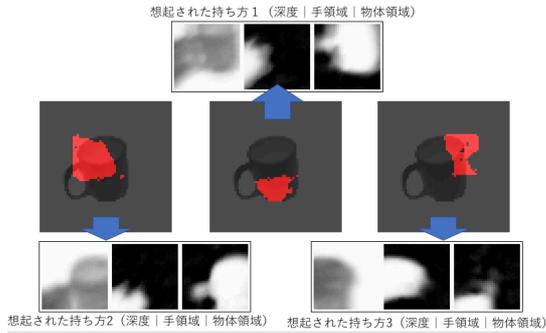


図 6 物体画像全体に対する複数の把持パタンの想起結果とその位置分布



図 7 手術手技における結紮動作 (糸結び)

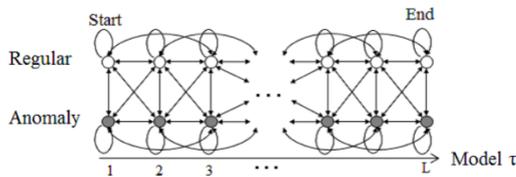


図 8 正常異常状態対に基づく確率的状態遷移モデル

手指による緻密作業モデリングの対象として、外科手術における手術手技の訓練システムの開発という具体的な課題場面を設定し、傷縫合時の糸結び手技(図7)を取り上げた。教師の模範動作に対して被訓練者の動作が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が、当然模範動作とは手の形状や動きのタイミングなどが異なるため、何を等価とみなし何を異常とみなすかが課題となる。時系列全体としてみれば正常な動作であっても、個別の画像フレームを単独で比較すると完全に一致することは少ない。その上手戻りや糸の取り落としなどのモデルにはない動作が交じる場合にそもそもモデルにない動作状態を確率モデルとしてどのように取扱い、どのように入力とモデルを照合するか、は明らかではない。

本研究では先行研究と同様に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した。本研究での新たな提案として、少ない学習サンプルから構築できる正常と異常の状態対から構成される確率的状態遷移モデル(図8)を考案した。正常状態は模範動作データからサンプリン

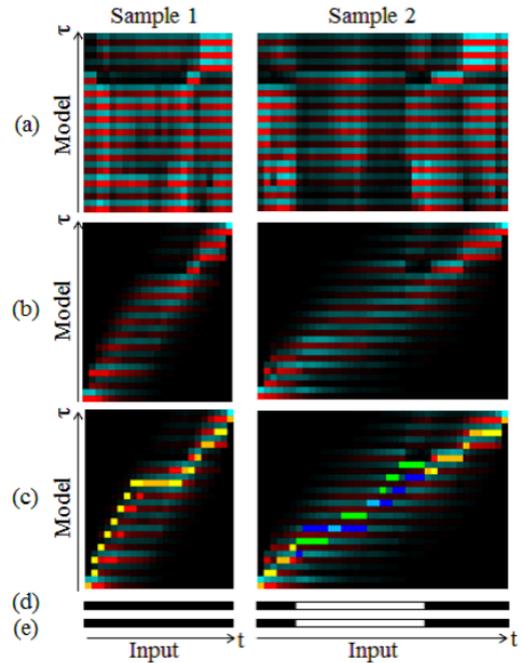


図 9 入力/モデル間フレーム対応付けに基づく異常フレーム区間の検出

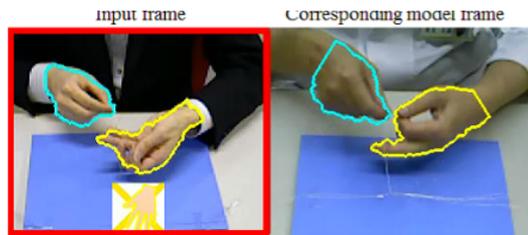


図 10 異常動作フレーム検知結果の提示

グした遷移確率及び画像特徴の尤度分布を用い、異常状態のそれらは分散の大きなガウス関数を仮定する。入力画像特徴の時系列と結紮動作モデルを frame-to-frame に対応付ける際、ICP(Iterative Closest Point)アルゴリズムによる三次元点群の位置合わせを行って位置ずれや手形状の差を緩和したのち、マルコフモデルに基づくベイズフィルタリングを用いて確率的に最適なフレーム間対応付けを決定する(図9)。このとき正常異常状態対のうち異常状態に対応付けられた入力フレームを異常状態区間とし、被訓練者に対して提示を行った(図10)。

13の学習用模範結紮シーケンス(合計421フレーム)、72の評価用入力シーケンス(合計3038フレーム)を用いて評価実験を行い、異常フレーム指摘の正確度91%、再現率94%、F値0.85の性能を確認した。

(4) Recurrent Neural Networkによる物体の使用手順のプロセスモデリングと想起

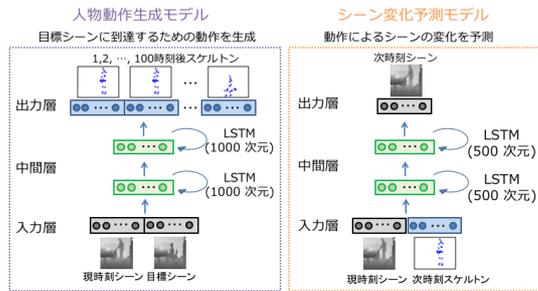


図 11 物体使用手順のプロセスモデル

(3) で取り組んだ緻密指使いのプロセスモデリングでは操作対象である結紮の糸を画像センサで計測することが困難であり、主に指使いの手順そのものをモデリングして入力と照合することを主題においた。しかし人が道具を使う場合には道具自体の位置や姿勢、他の物体や人との関係など、その状態の変化に応じて次に取り得る操作が決まることが多い。たとえば椅子に座る場合には、椅子がテーブルに引き込まれているならばまず椅子を手前に引いて、座るスペースが十分確保されたことを確認の上、その座面に腰掛ける動作を行う。もしすでに椅子が手前に引かれているならば、改めて椅子を引く動作は不要であり、すぐに腰掛ける動作に移ればよい。この例は物体を使用するにあたって、人の行動が対象物体の状態を変化させ、変化した状態によって次に行動が発現するという一連のプロセスが道具を使用することの中心に位置することを示している。

そこで本研究では椅子に腰掛けテーブルにある物体を持ち上げる、という事例を対象に、行動と対象シーンの変化の連鎖をいかにモデリングし、その結果として目前のシーンの状態観測に基づいて次に取るべき行動を生成する方策について研究を行った。

本研究では時系列の入出力を取り扱うため、内部状態の再帰的入力を持つ Recurrent Neural Network (RNN) を用いたモデル構造を提案した。とくに長い時系列を記憶／想起できるとされる Long Short-Term Memory (LSTM) を構造の中核に用いた。図 11 に椅子を引いて座る行動とシーン変化のプロセスを学習する際に用いたモデル構造を示す。図 11 左側は、現時刻と動作完了後の目標となるシーンの深度画像を入力とし、現時点から目標シーンに至るために必要な人物動作をスケルトンの時系列として 100 フレーム分出力する、動作生成ネットワークモデルである。図 11 右側は現時刻シーンの深度画像と次時刻での人の動作姿勢を表すスケルトン情報を入力とし、次時刻のシーン状態を深度画像として出力するシーン変化想起ネットワークモデルである。この 2 つのモデルを連鎖的に組み合わせ、実際に椅子を使用する人物の画像シーケンスを訓練データとして入力して、同時に学習を行った。

学習後にテストデータとして椅子が収納

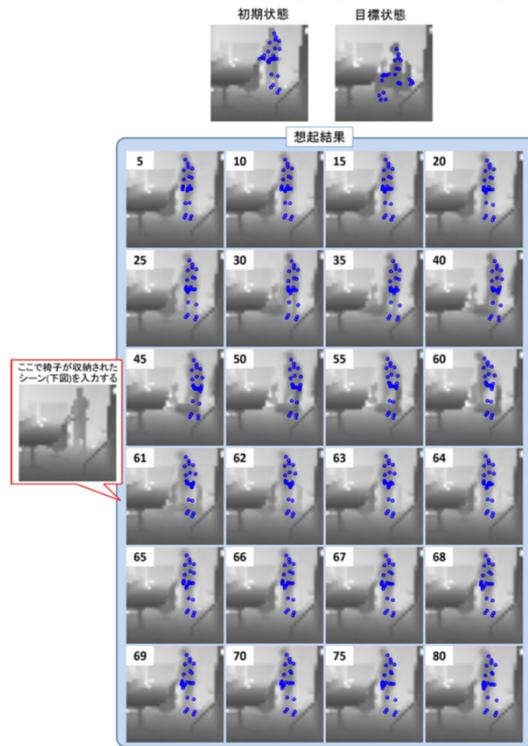


図 12 椅子を引いて腰掛ける動作におけるやり直し動作の発現例

された現時刻シーン画像と、椅子を引いて座っている目標シーン画像を動作生成モデルに入力すると、まず椅子を引いてそれから腰掛ける動作が生成された。このとき、途中で椅子を引くのに失敗したことをシミュレートするために、現時刻シーン画像として急に椅子がしまい込まれた画像に差し替えたところ、途中まで生成されていた椅子を引く動作を巻き戻し、最初から椅子を引き始める動作を生成できた (図 12)。椅子を引き終わって腰掛ける動作の途中で椅子を収納する画像に差し替えると、やはり腰掛ける動作を中断して、再度椅子を引く動作を生成し始めることも確認できた。

本研究では全身のマクロな動作を対象としたにとどまり、微細な指使いによる物体操作手順をプロセス・モデリングにはまだ至っていないため、今後これを指使いに適用していくことを検討する予定である。

(5) Transform-Invariant Auto-Encoder と半教師付き学習に基づく多指ハンドによる物体把持姿勢制御

(2) で取り組んだ物体形状からの把持パタンの想起によって、物体を見たときにどの部分をどのような指使いでつかめばよいかを得ることが可能となった。しかし指使いはあくまでも深度画像およびマスク画像で得られるため、そのままではロボットハンドの各関節モータをどう駆動してよいかはわからない。

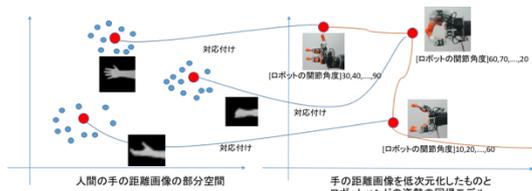


図 13 半教師付き学習による手指画像からロボットハンドのモーターコマンド生成



図 14 手指形状によるロボットハンドの姿勢制御例 (SIAE が提案手法による結果)

そこで、想起された手領域の深度画像を入力としてロボットハンドのモーターコマンドを生成する回帰モデルの学習を試みた。基本的には CNN を用いて想起された手指深度画像を入力とし、モーターコマンドを出力するネットワークである。しかし、人の手指とロボットハンドの構造は必ずしも一致せず、今回用いたロボットハンドは 3 指 9 関節のものでその見えも一致しない。

そこで、人の手指の見え一つ一つに対して、どのようなロボットハンドの姿勢を対応させるかを教示することとした。しかしこの教示もすべての手指姿勢について細かく与えることは手間がかかり現実的ではない。そこで、手指の深度画像を (2) で開発した Transform-Invariant Auto-Encoder で符号化した特徴量空間に写像し、写像後の限られた種類の手指姿勢についてのみロボットハンドのモーターコマンドに対する教師信号を与えて訓練を行う半教師付き学習を行った (図 13)。すると単純な PCA 等の空間で学習するよりも直観的に正しいと感じるロボットハンドの姿勢が出力されるようになった (図 14)。このロボットハンド制御プログラムは実時間で動作するため、実時間の把持パターン想起と組み合わせると、いわゆる pre-shaping と呼ばれる物体形状に応じた形状をつくってから力学的な物体把持制御に進むことを模倣できる可能性が出てきた。一方で数値的な客観評価が不十分であり、引き続き検討を重ねていく予定である。

## 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 6 件)

1. 小川陽子, 松尾直志, 島田伸敬, 白井良明, 来見良誠, 小森優, “手技訓練補

助のための正常異常状態対で構成される状態遷移モデルによる異常動作の検出” (DOI: 10.14923/transinfj.2017PDP0033), IEICE Trans. on Info. and Sys., Vol. J101-D, No. 3, pp. 549–559, 2018 (査読あり)

2. T. Matsuo, N. Shimada, “Construction of Latent Descriptor Space and Inference Model of Hand-Object Interactions” (DOI:10.1587/transinf.2016EDP7410), IEICE Trans. on Info. and Sys., Vol. E100-D, No. 6, pp. 1350–1359, 2017 (査読あり)
3. Dinh T. T., R. Sakurai, H. Yamazoe, and J.H. Lee, “Phase Segmentation Methods for an Automatic Surgical Workflow Analysis” (DOI:10.1155/2017/1985796), Int’l J. of Biomedical Imaging, vol. 2017, Article ID 1985796, 2017 (査読あり) [学会発表] (計 69 件)

1. T. Kawakami, T. Matsuo, Y. Ogawa, N. Shimada, “Grasping Pattern Estimation Based on Co-occurrence of Object and Hand Shape”, 24th Int’l WS on Frontiers of Computer Vision, 2018 (Best Paper Award).
2. 浅野奈生, 小川陽子, 松尾直志, 島田伸敬, “手指リハビリテーションのための拇指先端可動域の非接触自動計測システム”, コンピュータビジョンとイメージメディア研究会 2018 年 3 月研究会 (CVIM211), 2018
3. T. Matsuo, H. Fukuhara, N. Shimada, “Transform Invariant Auto-encoder”, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2017
4. M. Yano, T. Matsuo, N. Shimada, “Description and recall of the object using processes with scene change using LSTM”, 2nd WS on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics, 2017

## 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

### (2) 研究分担者

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 (LI J00-H0)  
立命館大学・情報理工学部・教授  
研究者番号：80366434