力覚情報のメル周波数ケプストラム係数に基づく接触動作の認識

辻 俊明^{*1} 佐藤 航陽^{*1} 境野 翔^{*2*3}

Recognition of Contact Event Based on MFCC of Force Signals

Toshiaki Tsuji^{*1}, Koyo Sato^{*1} and Sho Sakaino^{*2*3}

Recognition of contact events is often required for contact-based tasks such as assembly and grinding. This paper proposes a recognition method based on the Mel Frequency Cepstrum Coefficient (MFCC) of force signals, which can be derived from an ordinary robot configuration with a force sensor and a manipulator. We demonstrated that the combination of MFCC and time delayed neural network is effective in the case of the contact event recognition. It was confirmed that the above method has the ability to recognize instantaneous responses that do not generate repetitive waveforms.

Key Words: MFCC, Force Sensing, Neural Network, Contact Recognition, Imitation Learning

1. 序 論

組立ての自動化はロボット工学の中でも古くから研究されてきたテーマであり、かつ今日においても学界・産業界双方で重要な課題として議論されている。すでに多くの組立て工程が実用レベルで自動化されているが、作業が高次になるに従って教示のコストが膨大になるのが課題である。接触を伴う作業のため、接触モデルの非線形性と複雑さによってプログラミングによる教示が困難なケースがしばしば生じる。プログラミングコストを削減するために人の作業データを用いて教示する模倣学習 [1] [2] が研究されている。

模倣学習の問題は動作検出、動作認識、動作生成に分類できる [3]。最も広く研究されているのは動作そのものを模倣する動作生成技術である [4] [5]。高次の動作を自律的に生成するためには複雑な動作を簡易な動作の組み合わせとして構造化し、解析する必要がある [6]。そのためにタスクの分節化を論じた研究も活発に進められており [7]、動作認識はその基盤となる技術である。サブタスクの成否を判定することができればより確実に複合的なタスクを遂行できるようになることから、成否判定に動作認識を利用することも想定される [8]。

動作認識は位置情報に基づくものが多く、主に画像や内界センサが利用される。動作の内容は接触状態に強く依存することから、上記のセンサ類に基づいて接触状態を検知し、動作認識する研究もなされている [9]。しかし、特に精密組み立てのように対象物どうしの接触を伴い、かつ位置のクリアランスが小さ

いタスクにおいてはオクルージョンや画像の分解能の影響を受けずに直接接触状態を検知し、動作認識する必要がある。

接触状態を分類する研究は触覚アレイセンサを用いた例が多く、その分類結果を利用して作業の質を高める試みが進められている [10]。また、接触状態から把持の安定性を評価し、不安定と判断された場合に再度把持をする手法も提案されている [11] [12]。

ほかのセンサを使用した例として、対象物表面をなぞったときの加速度情報からその材質を分類する深層学習手法が挙げられる [13]。力覚センサを用いたものとしては力の大きさの情報から接触状態の遷移を検出する手法が古くから利用されている [14]。また、Kikuuwe らは対象物のインピーダンスから表面の特徴を検出する手法を提案している [15]。素材の分類性能を向上するために素材をなぞった力をケプストラム解析する手法も提案されている [16]。これらの手法はアレイ状のセンサよりもハードウェアの制約が少ない点で優れているが、振動パターンが連続することを前提とした手法であり、衝突やクリック応答のような瞬時の力応答に対応するものではない。タスクの分節化や成否判定に動作認識を利用する場合には瞬時の力応答からどのようなイベントが発生したかを判定する必要がある。

そこで本研究では、組み立て作業の動作認識を目的として接触の分類を力覚センサのみで行う。力覚応答値のメル周波数ケプストラム係数 (MFCC) に基づく分類手法を提案する。本論文が主に明らかにするのは以下の2点である。

- (1) 力覚センサを用いた場合には MFCC とニューラルネットワークを組み合わせた分類手法の有効性が高いことを示す。
- (2) 上記の手法がなぞり動作のように繰り返し波形を発生しない、瞬時的な動作を認識する能力を持つことを示す。

一般的には動作認識という呼称は操作者の主体的動作を対象とするものである。しかし、本研究のように操作者と対象物が接触によって強度の力学的インタラクションを持つ系では操作者の主体的動作と対象物からの応答を分離することはできない。そのため、本論文では対象物からの応答値に着目し、特に状態

原稿受付 2020 年 8 月 18 日

^{*1}埼玉大学

^{*2}筑波大学

^{*3}JST さきがけ

^{*1}Saitama University

^{*2}University of Tsukuba

^{*3}JST PRESTO

■ 本論文は有用性 (システム設計・構築分野) で評価されました。

Fig. 1 System composition

遷移の検出に有効性の高いクリック応答の認識を試みる。

2. 実験機の構成

Fig. 1 に本研究で用いるシステムの構成を示す。一般的な産業用ロボットと同じ 6 自由度のマニピュレータの手先に 6 軸力覚センサを搭載し、その先にグリッパを装着している。組み立ては基本的に剛体の部品どうしの接触状態を遷移させる作業とみなせるので、把持している部品に生じる合力と部品の位置・姿勢が分かれば、人の組立て動作と等価の情報が得られることになる。換言すれば力覚センサとロボットという一般的なシステムは、原理上組み立て技能解析に十分な構成である。グリッパの手先にペンの蓋を持たせ、1 [ms] のサンプリングタイムで制御を行いながら蓋を閉める動作を学習させる。ペンの種類によって閉める作業を完遂するために必要な力や位置は異なるため、ペンの種類を変更した場合でも接触動作の認識が可能かを調べる。実際の組立作業では位置と力の応答からオンラインで接触動作を認識する必要がある。すでに実用化されている一般的な産業用ロボットのセンサ・アクチュエータと計算機のみで接触動作の認識が可能になるかを検討する。

3. 力覚センサの応答値に基づく特徴量抽出

3.1 MFCC の構成

MFCC は音声処理分野でしばしば利用される特徴量であり、信号の周波数を対数ケプストラムに変換したのちに周波数知覚特性を考慮した重み付けをして得られる [18]。組み立ての接触認識でははめ合い動作の大きな押し付け力のなだらかな変動とクリック動作のような微小な力の変動の特徴量を両方同時に抽出するスケーラブルな処理が必要である。対数スペクトル領域で表現することにより、これらを同時抽出・分離することが容易になるため TDNN の前処理に実装する。

力覚センサの応答値 $f(t)$ をフーリエ変換したものを $F(e^{j\omega})$ とすると、

$$c(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |F(e^{j\omega})| e^{j\omega m} d\omega \quad (1)$$

よりメルケプストラム $c(m)$ が求められる。ただし、 ω はメル周波数目盛とし、本研究では離散コサイン変換によりメルケプストラムの低次成分五次元を取り出したものを MFCC とする。音声処理の MFCC の先行研究では、メルケプストラムの 12 次元程度を取り出すものが多いが、接触動作の記録のサンプリングレートが高くないことや、クリック応答の特徴量が比較的低次元でも抽出できることから、本研究の扱う対象では五次元で十分な性能が得られる。

Fig. 2 MFCC and TDNN

3.2 TDNN の構成

Fig. 2 に提案手法の概要を示す。力センサの応答から上述のアルゴリズムで得た五次元の MFCC を時系列順に並べたもの TDNN への入力としてロボットの接触状態を認識する。TDNN は時系列順に並んだ情報の 1D-CNN と等価であり、1 層の畳み込み層と 2 層の全結合層から構成される。いずれの層も活性化関数は Leaky-Relu としている。畳み込み層は 5x1 の重みフィルタを時間方向にスライドさせる演算となり、クリックや衝突などの短時間の応答、および接触時の振動パターンなどを認識する。それに対して全結合層はより長い時間レンジで下位層で得られた特徴量を統合した特徴マップを得る。畳み込み層と全結合層がそれぞれの時間レンジを扱うかは畳み込み層のフィルタ長によって調整される。本研究では 100 [ms] ごとに抽出された時系列の力応答値を 5x1 のフィルタに通していることから、0.5 秒以下の時間レンジの特徴量を畳み込み層で抽出することになる。最終層はタスクの成否をそれぞれ出力するものであり、交差エントロピー誤差により損失関数を求める。パラメータの更新は Adam を用い、学習率は初期値を 0.001 とする。

4. 実験

4.1 データの取得

本研究では MFCC と TDNN に基づく接触状態の認識手法を提案する。そのためには学習用の運動データとその接触状態を表すラベルが必要となる。多くの先行研究では別のセンサを付加することでラベルを自動的に取得するアプローチがとられている [17]。蓋のはめ合いを検出するほかのセンサが存在しないことや、組立ての現場を想定した際にセンサを追加できるケースが少ないという理由により、本研究では人手によるラベリングを行った。

ロボットによる軌道生成ではその再現性の高さのため、同様の条件で実施した場合に成功もしくは失敗のデータのみが蓄積されることになりやすい。また、収集できるデータの数が画像や音声の情報と比較して少ないため、不均衡データになりやすい。その課題を回避するために本研究では異なる 3 種類の失敗軌道を 25 例ずつ生成した。まず、アドミッタンス制御されたロボットに正弦波関数を用いた簡易な軌道生成アルゴリズムで蓋

Fig. 3 Pens used for experimental verification

Table 1 Comparison of feature amounts for each pen model

	Max. force[N]	Amplitude of click[N]
Pen 1	33.5 ± 1.0	9.0 ± 3.1
Pen 2	37.5 ± 1.6	5.8 ± 1.2
Pen 3	35.8 ± 4.4	7.6 ± 1.9
Pen 4	42.1 ± 1.0	4.8 ± 1.6

のはめ合いの位置指令を与えた。試行ごとに固定位置が変化する状況を再現するため、指令位置を左右方向、上下方向に少しずつ平行移動して試行を繰り返し、成功動作のデータが25セット得られるまで続けた。次に上記の過程で得られた失敗軌道に加えて左右方向へのずれ、上下方向のずれによる失敗動作を記録し、これらの合計がそれぞれ25回になるようデータをさらに蓄積した。そして最後にペンがないためにはめ合いが実施されなかった動作を25回生成し、上記と合わせて100回の動作を記録した。種類の異なるペンに対する汎化性能を評価するため、Fig. 3に示すような4種類のものを使用した。訓練データを左の2種で収集したのちにテストデータを全4種から集めた。

試行ごとのばらつきとペンの種類に依存する応答の差を把握するため、成功時の力の応答についてその最大値とクリック応答の振幅およびそれらの標準偏差をあらかじめ評価した。その結果を、Table 1に示す。本研究で利用した4種類のペンの応答を比較すると、最大値、クリック振幅ともに標準偏差を上回る差異が見られることが確認された。Fig. 4に成功時と失敗時の応答を比較した一例を示す。(a)、(b)はクリック応答がある場合とない場合の力応答のMFCC、(c)は両方の加工前の値を示している。成功時に発生するクリック応答ははめ合いの力と比較して振幅が小さいが、それに対して(a)の10秒前後ではクリック応答と同じタイミングでMFCCの高い周波数帯域の値が増大していることが確認された。本結果は力計測において、MFCCが小さい特徴量を大きな力の変動に埋もれさせずにスケラブルに検出することを示唆するものである。

4.2 性能評価

今回は瞬時の特徴量を検出できるかどうかを検証するためにクリック動作を含むペンの蓋閉め動作を対象とした。ペンの種類によって蓋閉めに必要な力の大きさは異なるが、動作の成功の可否はクリック応答が得られたか否かで判定可能である。比較対象の手法としてFNNおよびLSTMを採用し、その精度と演算時間を比較した。また、一般的なロボットの制御系での実時間処理が可能かを判断するため、演算時間はロボットの制御用PC(Corei7-8700K, 3.70 [GHz], 32 [GB] RAM)のCPUのみで400回テストを実施した時間の平均値を導出した。

FNNは全結合層を5層重ねたネットワーク構成とし、隠れ層

(a) MFCC of force response (with click)

(b) MFCC of force response (without click)

(c) Raw data of force response

Fig. 4 Comparison of response values of acquired data

Table 2 Performance comparison of each model

Model	Acc(%)	Computation time (ms)
FNN	78.0	1.0
LSTM	96.2	8.8
TDNN	87.5	0.9
MFCC+LSTM	97.5	14.8
MFCC+TDNN	98.0	1.8

と出力層のノード数をそれぞれ50, 2とした。Mean Squared Errorを損失関数とした。LSTMではLSTM層に入力し、その出力をノード数50の全結合層2層に入れた。各層のユニット数は入力を一次元×データ長、LSTM層を100、全結合層を50、出力を2とした。ドロップアウト率0.3でLSTM層と全結合層にドロップアウトを設定した。損失関数は交差エントロピー誤差により求めた。いずれのモデルも活性化関数はすべての層でReLU、バッチサイズは20で設定している。

他手法との性能比較をTable 2に示す。LSTM, TDNN いずれもFNNに比べて高い判定精度を示したが、特にTDNNではMFCCを前処理に加えることで精度が飛躍的に向上することが確認された。LSTMは高い判定精度を示すことが確認されたが、オンラインでの判定を想定した場合には演算時間の短いMFCC+TDNNの組み合わせの有効性が高いと判断される。

次に訓練データに使用するペンの種類が性能に及ぼす影響について検証する。訓練データの取得コストを下げるためにはできるだけ少ない種類のペンで学習したものがその他の種類にも適用できることが望ましい。しかし先にTable 1に示したように、本研究で利用したペンの応答には種類に依存した差異が見られることから、一種類のペンのみの学習では十分な汎化性能が得られないと予想される。そこで訓練データのペンの種類を1種類から3種類まで変化させたときの認識率を評価した。た

Fig. 5 Comparison by pen models used for training

だし、訓練データの数を揃えるために1種類での訓練データは200回、2種類と3種類での訓練データはそれぞれ100回ずつと66回ずつとした。また、成功軌道の比率は前節の記述と同様になるよう調整した。テストデータには、いずれも前節の条件と同じ4種類のペンの結果を利用した。その比較結果をFig. 5に示す。1種類のペンで学習させた場合には認識率が低下していることから、種類の異なるペンへの汎化性能は低いことが確認された。一方、訓練データの種類を増やすと認識率が向上する結果が得られた。特に訓練データを2種類に増やしたときの改善の幅が大きいことから、種類が少なくても複数の訓練データを得ることの効果は高いと推察される。

5. 結 論

本稿では力覚センサの応答から導出したメル周波数ケプストラム係数(MFCC)に基づく接触分類手法を提案した。力覚センサを用いた場合にはMFCCとTDNNを組み合わせた分類手法が有効であることが示された。そして上記の手法がなぞり動作のように繰り返し波形を発生しない、瞬時的な動作を認識する能力を持つことが確認された。組み立てや研削の作業においては接触状態を認識し、その状態に基づき次の動作を決定する必要があることから上記の技術を実装する利点大きい。そのため実作業の能率がどの程度改善されるかを今後検討する必要がある。また、本稿で示した結果はクリック応答の認識に限定されたものであることから、一般的な作業の分節化を行うためには、一点接触および多点接触の認識や締結状態の認識等に問題を拡張する必要がある。

謝 辞 本研究は埼玉県AIを活用した機器等開発・実証補助金の助成の元を実施したものであることをここに記し、関係各位に謝意を表する。

参 考 文 献

- [1] S. Schaal: "Is imitation learning the route to humanoid robots?" Trends in cognitive sciences, vol.3, no.6, pp.233–242, 1999.
- [2] B.D. Argall, S. Chernova, M. Veloso and B. Browning: "A survey of robot learning from demonstration," Robotics and

- autonomous systems, vol.57, no.5, pp.469–483, 2009.
- [3] 杉浦孔明: "模倣学習における確率ロボティクスの新展開", システム／制御／情報, vol.60, no.12, pp.521–527, 2016.
- [4] S. Scherzinger, A. Roennau and R. Dillmann: "Contact Skill Imitation Learning for Robot-Independent Assembly Programming," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2019), pp.4309–4316, 2019.
- [5] D. Furuta, K. Kutsuzawa, S. Sakaino and T. Tsuji: "Motion Planning With Success Judgement Model Based on Learning From Demonstration," IEEE Access, vol.8, pp.73142–73150, 2020.
- [6] K. Kase, R. Nakajo, H. Mori and T. Ogata: "Learning Multiple Sensorimotor Units to Complete Compound Tasks using an RNN with Multiple Attractors," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp.4244–4249, 2019.
- [7] T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, H. Asoh and M. Kaneko: "Segmenting continuous motions with hidden semi-markov models and gaussian processes," Frontiers in neurobotics, vol.11, no.67, 2017.
- [8] K. Sato, M. Oikawa, K. Kutsuzawa, S. Sakaino and T. Tsuji: "Success/Failure Identification of Skill Movement by Neural Network Using Force Information," Proc. 45th Annual Conference of the IEEE Industrial Electronics Society (IECON2019), pp.3641–3646, 2019.
- [9] M. Wachter and T. Asfour: "Hierarchical segmentation of manipulation actions based on object relations and motion characteristics," Proc. IEEE 2015 Int. Conf. on Advanced Robotics (ICAR), pp.549–556, 2015.
- [10] Y. Zhang, W. Yuan, Z. Kan and M. Y. Wang: "Towards Learning to Detect and Predict Contact Events on Vision-based Tactile Sensors," Conf. on Robot Learning (CoRL), pp.1395–1404, 2019.
- [11] R. Calandra, J. Lin, A. Owens, J. Malik, U.C. Berkeley, D. Jayaraman and E.H. Adelson: "More Than a Feeling: Learning to Grasp and Regrasp using Vision and Touch," IEEE Robotics and Automation Letters, vol.3, no.4, pp.3300–3307, 2018.
- [12] B. Wu, I. Akinola, J. Varley and P. Allen: "MAT: Multi-fingered adaptive tactile grasping via deep reinforcement learning," Conf. on Robot Learning (CoRL), pp.142–161, 2019.
- [13] H. Zheng, L. Fang, M. Ji, M. Strese, Y. Ozer and E. Steinbach: "Deep learning for surface material classification using haptic and visual information," IEEE Trans. on Multimedia, vol.18, no.12, pp.2407–2416, 2016.
- [14] M. Skubic and R. A. Volz: "Acquiring robust, force-based assembly skills from human demonstration," IEEE Trans. on Robotics and Automation, vol.16, no.6, pp.772–781, 2000.
- [15] R. Kikuuwe and T. Yoshikawa: "Recognizing surface properties using impedance perception," IEEE Int. Conf. on Robotics and Automation (ICRA), pp.1539–1544, 2003.
- [16] T. Miyagi and S. Katsura: "Analysis and Modeling Based on Cepstrum for Haptic Presentation Considering Frequency Features of Vibration in Rubbing Motion," IEEEJ Journal of Industry Applications, vol.4, no.2, pp.74–82, 2015.
- [17] J. Li, S. Dong and E. Adelson: "Slip detection with combined tactile and visual information," IEEE Int. Conf. on Robotics and Automation (ICRA), pp.7772–7777, 2018.
- [18] L. Muda, M. Begam and I. Elamvazuthi: "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," Journal of Computing, vol.2, no.3, pp.138–143, 2010.