

LSTMによる力覚データを用いた把持物体の識別

山崎雅幸† 山下隆義† 山内悠嗣† 藤吉弘亘† 堂前幸康‡ 白土浩司‡

† 中部大学 ‡ 三菱電機(株)

E-mail: hf@cs.chubu.ac.jp

Abstract

本研究では、力覚センサより取得したデータを用いてロボットが把持した物体の識別を目的とする。力覚センサはロボットに接触を用いて精密な作業を行わせるために使用されており、ハンドにかかる微細な力を計測することができる。力覚センサの出力は時系列データであり、一般的なニューラルネットワークでは連続的な変化を捉えることができない。そこで、本研究では系列データから有用な特徴を捉えることが可能な Long Short Term Memory(LSTM)を用いて、把持物体の識別を実現する。入力には力覚データとともに電流フィードバック(FB)を使用する。電流FBを用いることでロボット関節にかかった負荷を同時に得ることができ、より多くの特徴を得られる。評価実験より、力覚センサと電流FBを同時に用いることでより高い精度の識別を実現した。また、剛体の物体に対する識別率が高く、非剛体を1クラスにまとめることでより高い精度の結果を得られた。

1 はじめに

従来の物体識別は、Deep Convolutional Neural Network[3]等の機械学習を用い、ビジョンセンサにより取得した画像より、識別することが一般的である。しかし、画像からでは箱の内包物の種類や量などの取得できない情報がある。そこで、本研究は、画像を用いずに、ロボットハンドに装着された力覚センサ[2]からの把持物体識別を対象とする。力覚センサは主にロボットのハンドに取り付けられ、ロボットが接触を用いて精密な動作をするために活用されている。力覚センサをハンドにつけることで接触以外にも、物体を把持した時にハンドにかかる力を測ることができる。本研究では物体を把持した時の物体ごとに異なるモーメントの変化に注目する。また、力覚センサはハンドにかかる力のみを測定するため、ロボットの姿勢を加味することができない。そこでロボット関節の電流フィードバックを用いることでロボットの関節にかかった負荷を見ることができ、力覚データを測定するシーンは

アームを動作させ、停止した後のデータを用いる。停止後を用いることで慣性による影響や、非剛体の形状変化が生じるため、重量を測るよりも有効な特徴を得ることができる。

Recurrent Neural Network(RNN)[4]の一種である Long Short Term Memory(LSTM)[1]は動画、音声などの系列データを扱うときに用いられる。RNNは1時刻前の中間層の出力をサンプルとともに入力する。学習時のパラメータ更新は、Back Propagation Through Time(BPTT)[6]と呼ばれる方法を用い中間層から時系列をさかのぼりながら行うため連鎖的に過去の情報を得ることができる。LSTMは通常のRNNで起こる勾配消失問題を解決したもので、長時系列を扱うことができる。

本研究では力覚センサと、電流フィードバック(FB)よりLSTMを用いて把持物体の識別方法を検討する。また、電流FBを利用する有用性、データ取得について検討を行う。

2 Long Short Term Memory

LSTMはRNNの1種であり、系列データを扱うニューラルネットワークとして用いられる。RNNは時刻 t の時、中間層へ入力サンプルと同時に時刻 $t-1$ の中間層の応答値を入力することで系列を加味することができる。RNNでの学習にはBPTTが用いられ、過去のデータを遡るように更新していくが、長期のデータになるにつれ、誤差の勾配が消失、または爆発的に大きくなってしまふ。LSTMはRNNの問題点である長期記憶における勾配消失問題を解決した手法である。LSTMはRNNの中間層をLSTMユニットに置き換えている。LSTMユニットは記憶メモリと入力、出力、忘却ゲートで構成されている。図1にLSTMの構造を示す。

LSTMは通常のRNNと同様に入力と時刻 $t-1$ の中間層の応答値を結合する。同時に入力ゲートの計算を行う。入力ゲートには $t-1$ の時の中間層 h^{t-1} 、記憶メモリ U_0^{t-1} と入力 x^t から U_{in}^t を求める。 U_1^t を活性化関数 \tanh に与え、 U_{in}^t をシグモイド関数に与え U_2^t を求める。 U_{in}^t はシグモイド関数により、0から1の値をとる。これにより $t-1$ の入力の値による影響を調整し

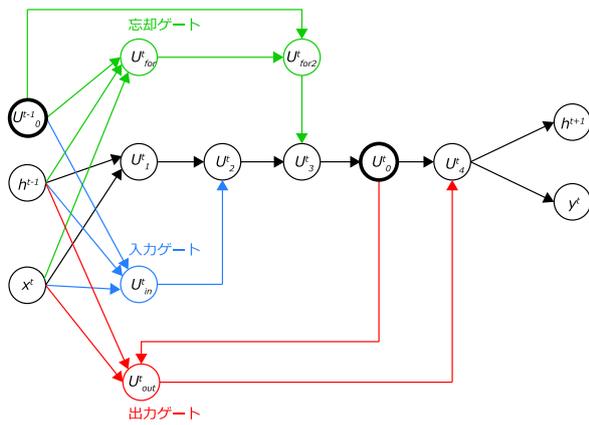


図1 LSTMの構造

ている。次に忘却ゲートに計算を行う。入力ゲートと同様に $t-1$ の時の中間層 h^{t-1} 、記憶メモリ U_0^{t-1} と入力 x^t を足して U_{for}^t を求める。 U_{for}^t をシグモイド関数に与え $t-1$ の記憶 U_0^{t-1} の値と乗算して U_{for2}^t を求める。忘却ゲートでは記憶メモリの内容を忘却させられるかを調整している。記憶メモリを更新した後、出力ゲートの計算を行う。 $t-1$ の時の中間層 h^{t-1} 、記憶メモリ U_0^{t-1} と入力 x^t から U_{out}^t を求める。 U_{out}^t を活性化関数 \tanh に与え、 U_{out}^t をシグモイド関数に与えて U_4^t を計算する。出力ゲートでは、記憶メモリの出力を次の計算に伝達するかを調整する。

入力ゲート、出力ゲートにより不要な誤差計算を行わないため、勾配消失を防いでいる。また、忘却ゲートにより不要な記憶を忘れることでより効率的に誤差伝播を行える。そのため、長期的な記憶も正しく出力に反映できるため、高精度な識別が行える。

3 力覚センサと電流フィードバック

本章では識別に使用される力覚センサ、電流FBについて述べる。力覚センサはロボットに触覚を付与させる目的で、様々なロボットに装備されている。特に産業ロボットの微細な部品はめ込み等に利用されている。力覚センサを用いることで接触を利用した動作を可能とさせている。力覚センサは図2に示すようにセンサ上部と下部との力の変位を計測し、静電容量の変化から変位を計測し力とモーメントを求める。取得できるデータは力ベクトル (F_x, F_y, F_z) と回転ベクトル (M_x, M_y, M_z) を合わせた6軸のデータである。

電流FBは、アームが動作する際に関節のモータの電流を測定した値である。そのため、電流FBは動作や姿勢維持するための値なので、物体ごとに影響を受けたロボットの状態を取得できる。取得されるデータは6軸関節に流れる電流データ ($J_1, J_3, J_3, J_4, J_5, J_6$) である。

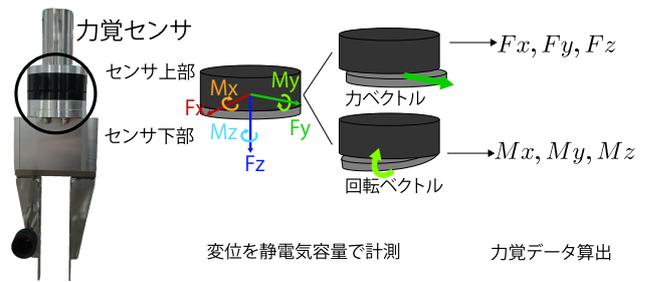


図2 力覚データの測定

3.1 把持物体ごとの力覚センサ、電流FBの変化

図3は2つのアイテム (book, mailer) を把持し、ロボットアームを左右に動作させ停止した後の力覚値 F_y の遷移である。途中、力覚センサデータの波形のピークの位置が異なっている。これにより、物体形状が似ていてもセンサ値の変動が異なることが分かる。

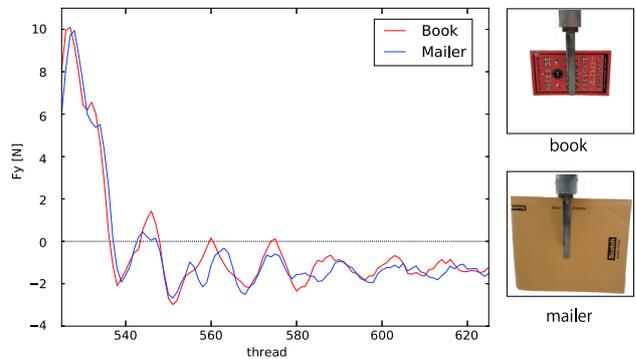


図3 アイテムごとの力覚データの比較

図4は2つのアイテム (scissors, curtain) を把持し、ロボットアームを左右に動作させ停止した後の電流FBの遷移である。停止動作を実現するために、電流FBの収束が把持物により変化する。把持する物体により関節にかかる電流が変化することが分かる。

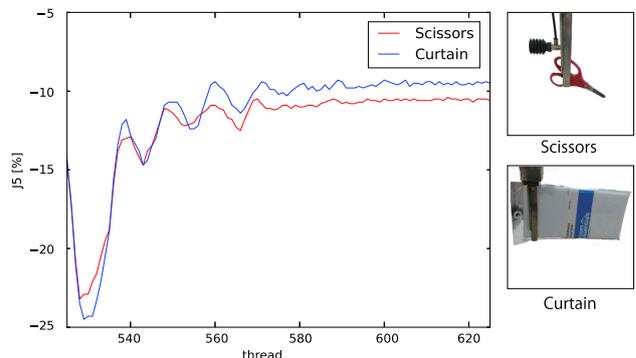


図4 アイテムごとの電流FBの比較

3.2 把持位置による力覚データの変化

図5は1つのアイテム (book) を2種類の把持位置で把持し、ロボットアームを左右に動作させ停止した後の力覚値 F_y の遷移である。同一アイテムであるが把

持位置に波形のピークの位置が異なっている。これにより同一アイテムでも把持位置によりセンサ値の値が変化するため、様々な把持パターンを学習する必要がある。

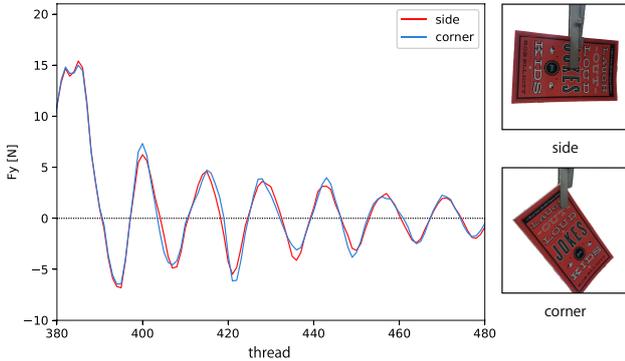


図5 把持点による力覚データの比較

4 提案手法

本章では、提案手法の流れを述べる。本研究では、産業用ロボットに装着された力覚センサと電流フィードバック (FB) より得られたデータを用いて LSTM により学習する。取得するシーンはアイテムを把持した後の手を動作させた時のデータを用いる。取得されたデータを正規化したのち、LSTM に入力する。フレームごとに出される LSTM の出力を一つの結果に統合し、識別を行う。

4.1 取得シーン

本手法では図6のようにロボット座標系で x , y , z 方向にそれぞれ1往復させた時データを取得する。学習時には3方向を混在させたものを使用する。

取得した力覚データには移動時のデータ停止、制動までのデータが含まれており、1回往復した長い系列データを扱うことになる。この中でも本実験では動作停止後100フレームを学習データとして使用する。図7に1往復データから抽出する学習箇所を示す。

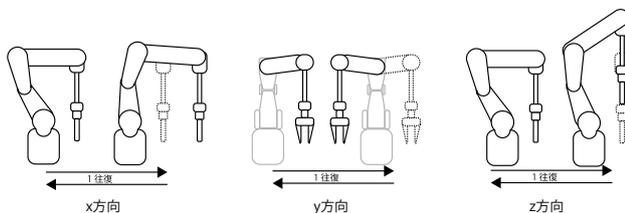


図6 取得するロボットの動作

4.2 入力データの正規化

入力データは、力覚センサから得られる3次元の力ベクトル (F_x, F_y, F_z) と回転ベクトル (M_x, M_y, M_z) を合わせた6軸のデータ、電流FBから得られる6軸

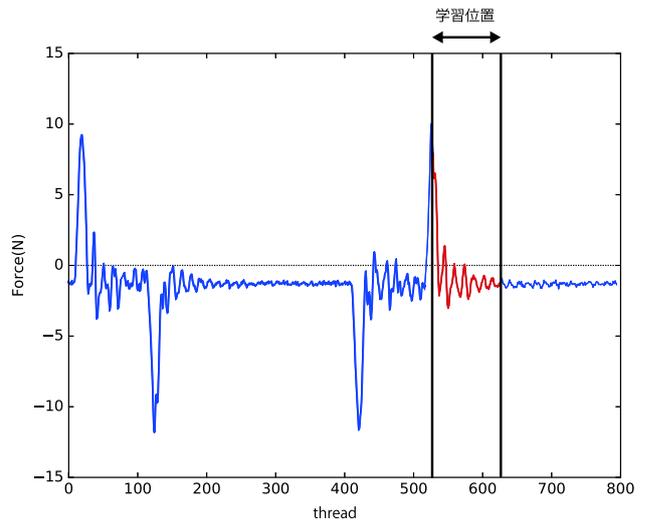


図7 1往復より抽出する F_y 軸データ範囲

関節に流れる電流データ ($J_1, J_3, J_3, J_4, J_5, J_6$) である。入力データはベクトルごとの関係を持たせるためベクトルごとに正規化を行う。正規化は力ベクトル、回転ベクトル、電流FBごとの最大値をもとに求める。力ベクトルの最大値を F_{\max} 、回転ベクトルの最大値を M_{\max} 、電流FBの最大値を J_{\max} としたとき、 i フレーム目の入力データ \mathbf{x}_i を求める式 (1) を示す。

$$\mathbf{x}_i = \left[\frac{F_{x_i}}{F_{\max}}, \frac{F_{y_i}}{F_{\max}}, \frac{F_{z_i}}{F_{\max}}, \frac{M_{x_i}}{M_{\max}}, \frac{M_{y_i}}{M_{\max}}, \frac{M_{z_i}}{M_{\max}}, \frac{J_{1i}}{J_{\max}}, \frac{J_{2i}}{J_{\max}}, \frac{J_{3i}}{J_{\max}}, \frac{J_{4i}}{J_{\max}}, \frac{J_{5i}}{J_{\max}}, \frac{J_{6i}}{J_{\max}} \right] \quad (1)$$

4.3 ネットワーク構成

力覚データは系列データであるため、LSTM を用いたネットワークを使用する。本研究で使用するネットワーク構造を図8に示す。学習係数は0.01、入力層には力覚データ、電流FBのを12ユニット与え、出力層は25クラスに対応した25ユニット、中間層は250ユニットのLSTMを2層を使用する。

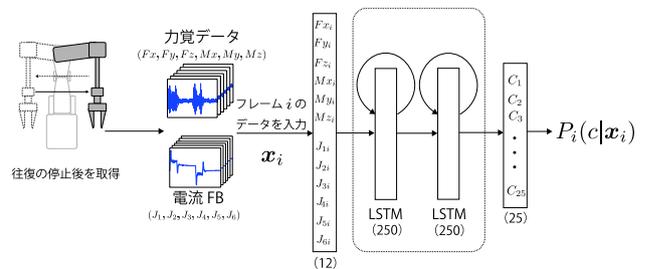


図8 ネットワークの構成

4.4 識別結果の統合

RNN は i フレーム毎にクラス c の確率 $P_i(c|\mathbf{x}_i)$ を出力するため、1つの系列データの結果を統合する。式

(2)により100フレームの確率を統合し、最終的な識別結果 \hat{C} を式(3)より求める。

$$P(c) = \frac{1}{100} \left(\sum_{i=1}^{100} P_i(c|\mathbf{x}_i) \right) \quad (2)$$

$$\hat{C} = \arg \max_c P(c) \quad (3)$$

5 評価実験

力覚センサを用いた把持物体の識別の有用性を示すために評価実験を行う。本実験では、Amazon Picking Challenge 2016[7] データセットの25アイテムを対象とする。使用するデータセットを図9に示す。

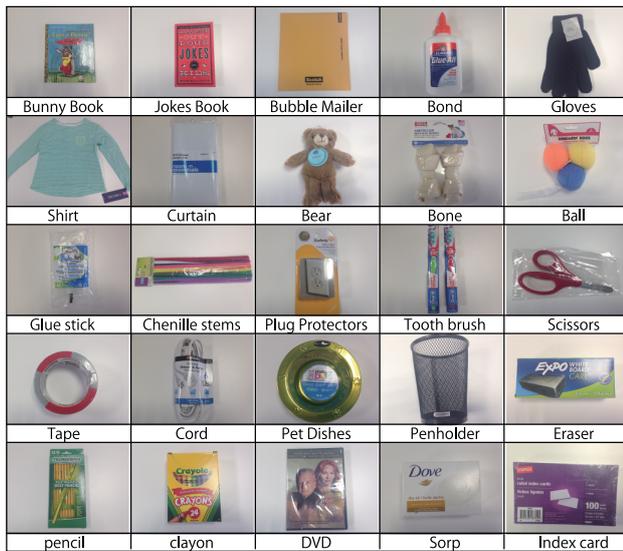


図9 識別アイテムリスト

取得する動作シーンはロボット座標系でX, Y, Zの3軸方向へ往復させたデータを、学習用データ3,326セット、評価用データは375セットを使用する。データの内約を表1に示す。

表1 使用データ内約

	X方向	Y方向	Z方向	合計
学習用データ	1124	1124	1078	3326
評価用データ	125	125	125	375

本実験では、電流FBの導入、非剛体クラスの導入、動作方向の入力を用いた識別を行う。

5.1 動作方向ごとの識別

本実験では3方向への動作を学習している。3方向それぞれで学習した場合の識別率を表2に示す。また、アイテムごとの識別率を図11に、Confusion matrixを図10示す。

表2より3方向の中でZ方向に動作させた時が識別率が高い結果となった。図11から、X方向の識別率結

表2 動作ごとの識別率 [%]

動作方向	X方向	Y方向	Z方向	全方向
識別率	14.4	62.5	63.2	79.7

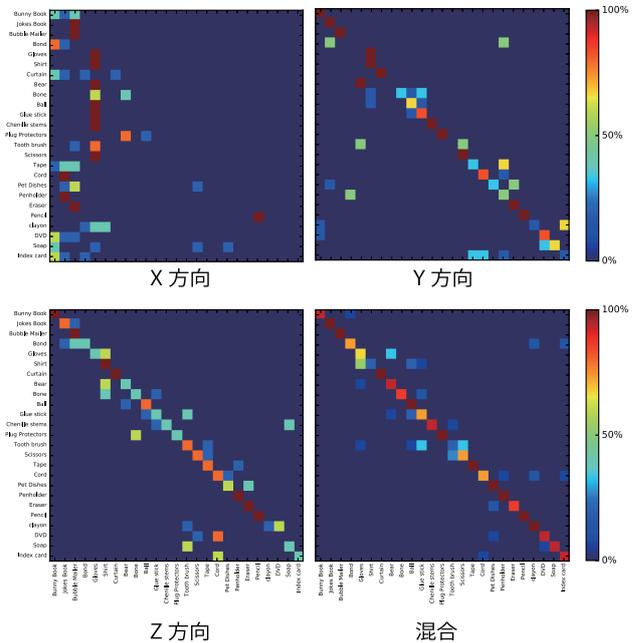


図10 動作方向ごとの Confusion matrix

果が偏っていることが分かる。原因はハンドの把持していない方向に振動してしまうため、力が逃げてしまい特徴が捉えられなかったと考えられる。しかし、図11より、それぞれの軸のみで識別ができるアイテムがあるため、統合することで汎用性を上げられたと考えられる。

5.2 電流FBの導入

提案手法による識別率を表3に示す。表3より、力覚センサのみでは識別率が79.7%であった。一方、電流FBのみでは4.0%と識別率が大幅に低下した。しかし、力覚センサとともに電流FBを入力することで、識別率を85.3%に向上させることができた。図12より電流FBを加えることで全体的に識別率が向上した。電流FBのみでは特徴を捉えることができないが、力覚センサと同時に入力することで力覚データのみとは違

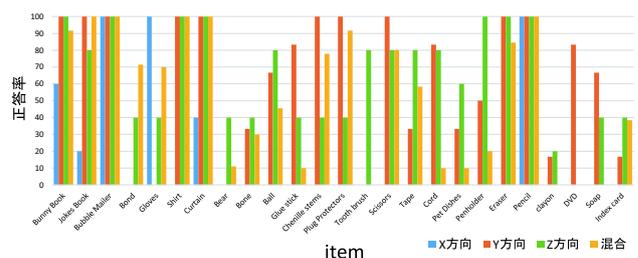


図11 動作方向ごとのアイテム識別率

う特徴を捉えることができる。

表 3 識別率 [%]

入力	力覚センサ	電流 FB	力覚センサ+電流 FB
識別率	79.7	4.0	85.3

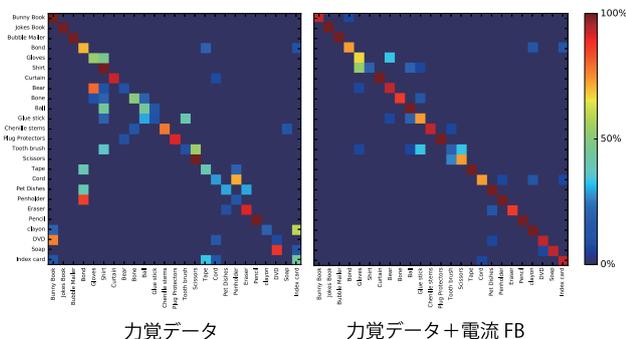


図 12 電流 FB を追加した Confusion matrix

5.3 動作方向の導入

本実験ではロボットに複数の動作を利用していることから、図 13 に示すように入力データに対して力覚データと同時に動作方向を入力する。入力データは力覚データ、電流 FB に加え X, Y, Z 方向の動作パターンの合計 15 ユニットとなる。力覚データと移動方向を導入した場合の識別結果を表 4 に示す。

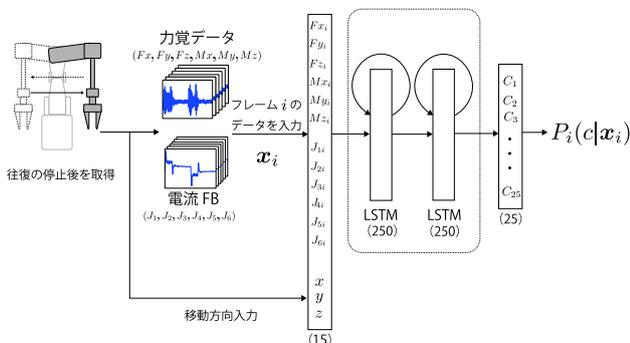


図 13 移動方向を導入したネットワーク構成

表 4 動作方向を導入した場合、電流 FB の導入の有無にかかわらず識別率の向上は見られなかった。力覚センサデータや、電流 FB データは動作方向ごとに大きく異なるため、動作方向を導入しない状態でも動作情報を加味できていると考えられる。そのため、動作情報なしでの識別が行えるが動作方向の情報から有効な特徴が捉えられなかったと考えられる。

5.4 非剛体クラスの導入

形状においてアイテムを剛体のアイテム、非剛体のアイテムを分けて識別を行う。図 14 に示すように 25 アイテム中 5 アイテムを非剛体アイテムとしてクラスを分ける。非剛体クラスをまとめた Confusion matrix を

表 4 識別率 [%]

入力	力覚センサ+移動方向	力覚センサ+電流 FB+移動方向
識別率	76.2	75.3

図 15 に示す。図 15 において赤枠で囲まれている箇所は非剛体のアイテム同士の混合行列である。剛体は平均 88.1% と高い識別率であるが、非剛体は平均 22.1% と低いことがわかる。これは、非剛体のアイテムは不規則に振動するため、学習誤差が収束せず、その結果誤識別の要因となっている。非剛体を 1 クラスとした全 21 クラスを LSTM で学習すると、平均識別率は 90.1% を得た。

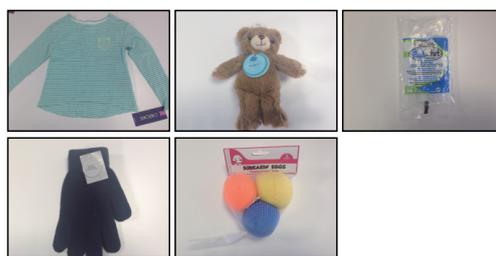


図 14 非剛体アイテムの例

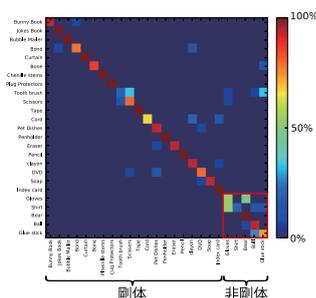


図 15 非剛体クラスの Confusion matrix

6 おわりに

本研究では、画像を用いないアプローチとして、力覚センサによる把持物体の識別法を提案し、その有効性を示した。力覚センサと電流 FB を用いることで、25 アイテムに対して約 85% の識別性能を実現した。さらに、剛体の識別率が高いため非剛体クラスを作成することで、約 90.1% の識別性能を得た。

参考文献

- [1] S.Hochreiter, “Long Short-Term Memory”, Neural Computation, 9(8): 1735-1780, 1997.
- [2] 緒方 浩二郎, 柏木 邦雄, 小野 耕三, “力覚センサ”, 日本ロボット学会誌 6 (9): 759-765, 1991.

- [3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, pp. 2278-2324, 1998.
- [4] J. Elman, "Finding Structure in Time", Cognitive Science 14 (2): 179-211, 1990.
- [5] T. Mikolov, Martin Karafiat, Lukas Burget, Jan "Honza" Cernocky, and Sanjeev Khudanpur, "Recurrent neural network based language model", INTERSPEECH, 2010
- [6] P. Werbos, "Generalization of backpropagation with application to a recurrent gas market model", Neural Networks 1 (4): 339-356, 1988.
- [7] Amazon Robotics, "https://www.amazonrobotics.com" (2017/4/17)