1R-03

持ち上げ動作からの所持物体の重量推定

氏　名　仲田 仁†　　氏　名　田村 仁‡

日本工業大学工学研究科機械システム工学専攻†　日本工業大学工学部創造システム工学科‡

1　はじめに

物体を持ち上げる場合，その物体が持ち手に与える荷重の大きさに応じて人体の挙動に差が表れ，動作に変化が生じるものと考えられる．

本研究では持ち上げた物体の重量を，物体を持つ手の動きの変移から推定することを目的とする．

同様，近しい内容の先行研究としては機械学習を用いたジェスチャの認識などが行われており，物体を持ち上げる動作の判別も同様に可能なのではないかと考えられる．そこで，本研究では重量判別の手法として，対象物を持ち上げる動作を撮影した動画を元に，機械学習を用いて手に所持している物体の重量を推定する．

2　研究概要

実験では撮影した動画から対象物を持ち上げる動作を行うところを取り出す．

機械学習では CNN(畳み込みニューラルネットワーク)を利用する．CNN は通常行う機械学習に対して畳み込み層とプーリング層を追加することにより物体の認識，判別に用いられている．

畳み込み層では小領域を一つの特徴として圧縮するフィルタを 1 ピクセルごとにスライドしながら画像全体に処理を施す．

プーリング層では画素領域を分割し，その中からそれぞれ画素を取り出して画像を作る収縮を行う層であり，今回は画素領域から最大値を取り出すマックスプーリングを使用する．

本研究は動画を用いての機械学習となっているが，CNN を利用するにあたり動画をベクトルデータにする必要がある．動画を元にした機械学習には映像を連続画像として整理し縦，横，時系列データを三次元ベクトルとして入力して学習を行う手法があり，畳み込み層のフィルタを三次元化して学習を行う．

本研究では入力データの時系列部分に着目し，持ち上げ動作を行う開始フレームを取得できれば連続画像を一枚の画像として並べることで同一時系列のフレームが画像内の同一箇所に来るため，時系列ごとの特長量の差を抽出，判別することができるのではないかと考えた．この仮説を元に，本研究では連続画像を一枚の単純な画像として扱っての CNN による持ち上げ動作からの所持物体の重量推定を試みる．

3　実験

今回行った実験では 500ml のボトルを持ち上げる動作を撮影し，持ち上げる動作の変移の差を獲得する．撮影対象となる物体の重量はボトル内に水を 100ml ごとに 100ml～500ml の間隔で注ぎ込んで撮影を行い，重量推定用の教師データ，および試験データの作成を行う．

被写体となるボトルとカメラの距離は持ち上げ動作が獲得できるようにボトルの全体像が映る 30cm 以上離して撮影を行う．

撮影した動画からボトルに手が接触したフレームから持ち上げ動作を行う分の 81 フレームを獲得し，それを 9×9 の範囲に並べて一枚の画像としてフレームの連結を行い機械学習のデータとして縦 480，横 640 に大きさを変換する．作成したデータの例を図 1 および図 2 に示す．

今回作成したデータ数は 100ml ごとに 20 枚ずつの合計 100 枚程度となっており，半分の 50 枚を教師データに，残りの半分を試験データとして学習を行う．

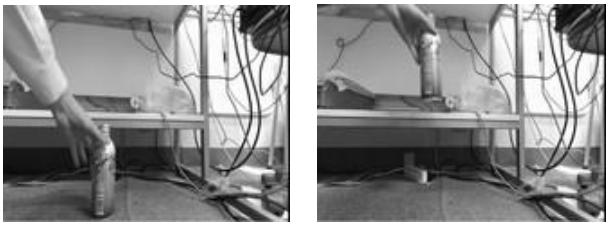

図 1. 撮影フレーム例

Estimation weghit of possessed object from lifting motion
†Nakada Hitoshi　†Nippon Institute of Technology
‡Tamura Hitoshi　‡Nippon Institute of Technology

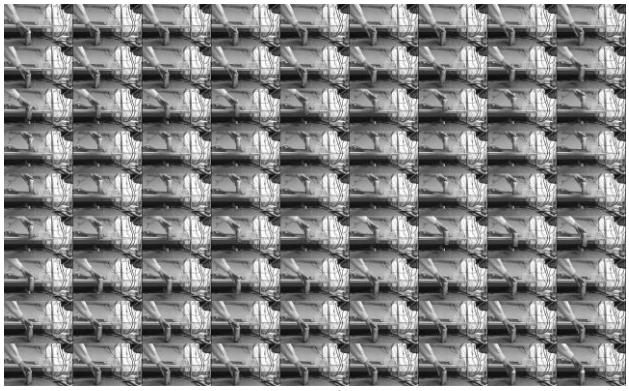
図 2. 作成データ例

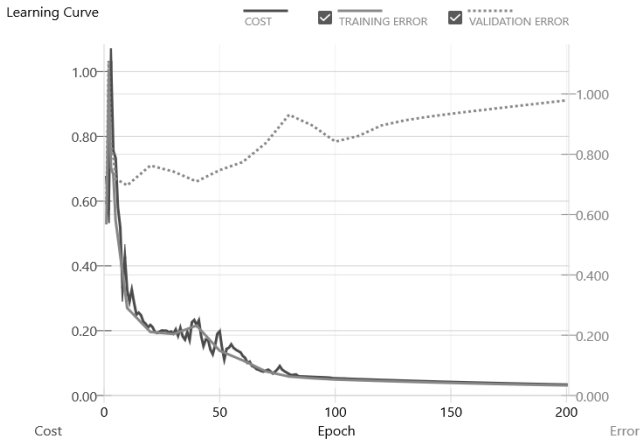

図 3. 学習結果

4　結果と考察

CNN による学習結果を図 3 に示す.

今回の実験では 300ml 以下の水を

学習結果から, 複数枚の画像を一枚の画像に並べた教師データの学習はできていると考えられるが, 試験データ側の認識率が低くなってしまっている. これに関して, 今回使用したデータ数が少なかったために教師データに対する過学習が生じてしまい認識率が低下してしまったのではないかと考えられる.

また, 連続画像を一枚絵にしたことでプーリング層での処理の際に局所的な部分が潰れてしまったため, 特長量の取得がうまくいかなかったという可能性も考えられる.

5　今後の研究, 改善案

過学習回避のために教師データ及び試験データの枚数を増やす必要性があると考えられる. 画像加工によるデータの増量を行う場合は, 通常の一枚絵を学習させる場合のように画像に回転処理を施して枚数を獲得することができないため, 各フレームに対して左右反転などの画像処理を行った後に 9×9 の一枚絵に並べる方法が有効だと考えられる.

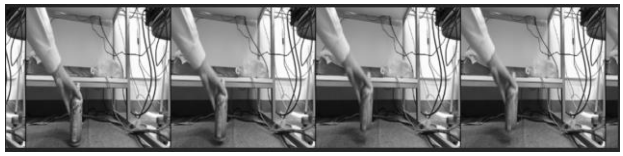
図 3. 横並び作成データ例

今回機械学習に使用した画像は 81 フレームを 9×9 に並べることで一枚の画像として扱ったが, 使用するフレーム数を増やす場合は学習用データにする際に行う縮小処理で 1 フレーム当たりのサイズが小さくなってしまうため, 最低限の特長量が取れるよう増加に伴ったサイズ調整を行う必要があると考えられる. また, 作成するデータを 9×9 ではなく横一列に繋げて一枚の画像とした場合の実験も行う. こちらの画像はフレームの折り返し部分が存在しないため動作の変移判別の学習精度が上がるのではないかと考えられる. 画像の例を図 4 に示す.

特長量を取得する精度の向上を図るため, 背景をブルーバックのような単一色の状態で物体の持ち上げ動作の撮影を行う. これにより手の変移のみを取得することができるため, 収縮処理による特長量の減少に対応が取れるのではないかと考えられる. また, 背景が単一色である場合には背景画像を複数枚用意することで撮影画像と背景画像の差し替えにより乗算的なデータ数の増加が見込める.

今回の実験では連続画像の学習を一枚絵で行ったが, 実験結果の比較のためにも入力されるデータを三次元ベクトルデータとしての学習実験も行っていく.

参考文献

[1]機械学習を用いたジェスチャー認識精度向上方法の研究
情報処理学会 ゲームプログラミングワークショップ 2012 論文書 6 号 167-170 頁

[2]深層学習を用いた会話中の人物頭部ジェスチャ認識
人工知能学会全国大会(31 回)　2H3-OS-35a-4in1

[3]立体フィルタを用いた Convolution Neural Network による三次元物体認識
情報処理学会 第 78 回全国大会講演論文集　2016(1), 37-381 頁