1R-06

ジェスチャー分類に対するオプティカルフローを前処理とした機械学習手法の検討

多和田 真悟† 遠藤 聡志‡ 山田 孝治‡ 當間 愛晃‡ 赤嶺 有平‡

琉球大学 工学部 情報工学科† 琉球大学 工学部 工学科 知能情報コース‡

1. はじめに

ジェスチャー分類とは動画像を入力とした分類問題の一つで、本稿では人間の手の動作認識を扱う。ロバストな学習モデルを作成するためのデータセットとして THE 20BN-JESTER DATASET[1]が公開されており、これを使用することでハンドジェスチャーの分類を行うことができる。ジェスチャー分類のモデルを構築するために、Guangming Zhu らは時系列情報を畳み込んだニューラルネットワーク[3]を使用したアプローチで高精度な分類を可能としている。しかし、彼らのアプローチでは、ネットワークそのものが複雑になり多くの計算量を伴う点や、ハイパーパラメータの調整が困難となるなどの問題が残る。ハンドジェスチャー分類では着目すべき対象が手の領域と決まっているため、手の動きの情報だけで分類ができると考えられる。

本稿では SSD[2]とオプティカルフローによる前処理と XGBoost[4]及び CNN を組み合わせた手法を構成し実験により検討する。SSD とオプティカルフローによる前処理を行うことで入力に対して次元を圧縮することができる。このことによって XGBoost のような分類器による学習が可能となり、ニューラルネットワークのハイパーパラメータの調整に関する問題の改善が見込める。

2. 検討手法

2.1 SSD 及びオプティカルフローによって抽出されるベクトルについて

本稿で検討する手法では初めに手の範囲を検出するために Single Shot MultiBox Detector(SSD)[2]を使用し手の範囲矩形を判別する。次に取得した手の範囲矩形からオプティカルフローの追跡に必要な特徴点を探索し以降、Lucas-Kanade 法[5]を用いてこれらの点を繰り返し追跡していき、各点の時間ごとの座標を抽出した上で、次元圧縮したデータを時系列を加味した座標ベクトルとして扱う。

XGBoost[4]の学習では行を各動画番号、列を座標ベクトルとした行列をテーブルデータとしてまとめ入力に使用し、CNN の学習では座標ベクトルを、行を座標、列を時間とする行列に変換し、これをグレースケール画像として入力に使用する(図 1)。

Consideration of optical flow as a preprocess in gesture classification.

† University of the Ryukyus, Depertment of Information Engineering

‡ University of the Ryukyus, Computer Science and Intelligent Systems

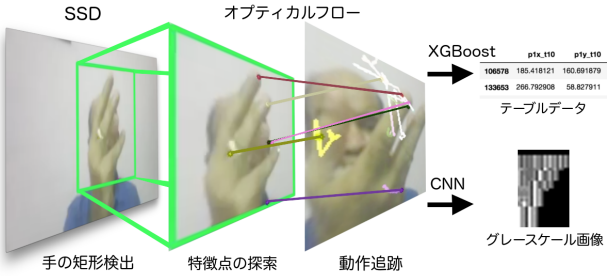

図 1 オプティカルフローによる抽出の流れ

2.2 eXtream Gradient Boosting (XGBoost)[4]

XGBoost とは、Gradient Boosting と Random Forests を組み合わせたアンサンブル学習のアルゴリズムである。XGBoost の特徴には、高い柔軟性をもつ木構造を構築すること、要素の欠損や、0 が多いような偏ったデータについて予め分岐の方向を決めるアルゴリズムを導入していることが挙げられる。前処理で抽出した座標ベクトルには欠損しているものや、座標をうまく追跡できなかったものがあるためそういったノイズを含んだテーブルデータに対し XGBoost が有効であると考えた。XGBoost に基づく以下の方法(図 2)を構成し、これを構成手法(1)とする。

構成手法(1)

1. SSD とオプティカルフローによりベクトルデータを抽出する
2. 特徴点の座標ベクトルをジェスチャーの正解ラベル毎に 1 つのテーブルにまとめる
3. ラベル付データセットに対し XGBoost で学習させ、グリッドサーチを行いパラメータチューニングを行う

2.3 畳み込みニューラルネットワーク(CNN)

深層学習の一つである CNN は畳み込み層とプーリング層と呼ばれる部分的に結合された層と 1 層以上の全結合層で構成される順伝播型ネットワークである。畳み込み層とプーリング層では入力値に対して特徴を抽出しその特徴量を計算し、全結合層で分類を行いその確率を出力するモデルである。座標ベクトルには欠損値などのノイズが含まれており、畳み込みによって有益な特徴を抽出できると考えた。CNN に基づく以下の方法(図 2)を構成し、これを構成手法(2)とする。

構成手法(2)

1. SSD とオプティカルフローによるベクトルデータを抽出する
2. 抽出したベクトルをグレースケール画像と

してみなす

3. 画像をラベル情報を元に、CNN で学習させる。

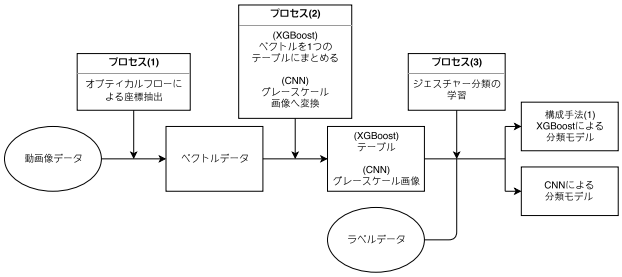

図 2 構成手法の流れ

3. 実験

3.1 実験概要

実験に使用する THE 20BN-JESTER DATASET[1]は 27 種類のハンドジェスチャーラベルがあり、各ラベルには約 5000 のデータが用意されている。また、予め各ジェスチャー動画を 12 フレームごとに静止画としてサンプリングされている。

初めに、このデータセットに構成手法の前処理を行う。本実験では 27 種類の中から「Swiping Up」「Swiping Down」「Turning Hand Clockwise」「Turning Hand Counterclockwise」「Stop Sign」の 5 つのハンドジェスチャー(図 4)を選択し簡単化した問題を扱うこととした。前処理を含む 2 つの構成手法で分類モデルを学習実験しモデル性能評価を行う。

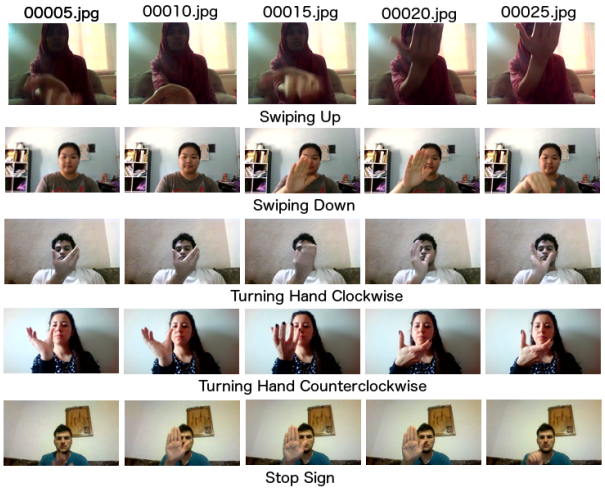

図 3 5 つのハンドジェスチャー

3.2 実験結果

表 1 に構成手法と分類精度を示す。次元数について入力画像を 32×32 にリサイズすると仮定すると、ベクトルでは元データに対して 1%の次元数、グレースケール画像なら 2%の次元数で学習を行った事となる。結果より手法(1)では 98.48%、手法(2)では 39.48%であった。XGBoost を用いた手法(1)では、十分な分類精度を得ることができた。一方、手法(2)では(1)に比べて大幅に精度が低い。その原因としては、CNN で吸収できると想定していた座標ベクトルの欠損値などのノイズ(すなわち手が含まれない動画像など)が悪影響を及ぼしたと考えられる。

Guangming Zhu ら[3]は本稿とは異なる SKIG Data Set[6] に対してジェスチャー分類実験を行っており 98.89%の精度を示した。実験では、データセットに対して RGB 画像とセンサーにより奥行きをグレースケール化した画像の二つを入力として用いていた。このことから動作対象を抽出し入力とすることは精度の向上につながり、また動作対象から抽出する情報は追跡座標以外にも必要であると考えられる。

	データ形式	次元数	分類精度
前処理 + XGBoost	特徴点の座標ベクトル	1.2×10^7	98.48%
前処理 + CNN	グレースケール画像	1.9×10^7	39.48%

表 1 性能評価

4. まとめ

本稿では、ジェスチャー分類に対して SSD とオプティカルフローを前処理とした機械学習手法を構成しジェスチャー分類精度を検討した。構成手法は前処理を通して情報量を削減していることから学習速度の向上が期待できる。

今後は、手の追跡情報だけでなく背景差分法等の画像処理による動体検出や、手だけでなく指等のより細かい特徴の前提情報を組み込んだ抽出手法を考察することで分類精度の向上を試みる。

参考文献

[1] twentybn."THE 20BN-JESTER DATASET V1". twentybn.2017-08-22. https://www.twentybn.com/datasets/jester. 参照(2018-1-11)

[2] Wei Liu, Dragomir Anguelov, Dumitru Erhan,Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg.: SSD: Single Shot MultiBox Detector. arXiv:1512.02325 v5.(2016)

[3] Guangming Zhu and Liang Zhang and Peiyi Shen and Juan Song: Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM. IEEE Access pages 4517-4524 (2017)

[4] Tianqi Chen, Carlos Guestrin: XGBoost: A Scalable Tree Boosting System. arXiv:1603.02754 v3. (2016)

[5] BD Lucas, T Kanade.: An iterative image registration technique with an application to stereo vision.In the 7th International Joint Conference on Artifical Intelligence page.674-679 (IJCAI 1981)

[6] The University Of Sheffield." Sheffield KInect Gesture (SKIG) Dataset". THe University Of Shelffield.2018-1-11. http://riemenschneider.hayko.at/vision/dataset/task.php?did=170. 参照(2018-1-11)