

RGB-D を用いた CNN による行動認識

馬場 樹[†] 尾関 智子^{††}

[†] 東海大学大学院工学研究科電気電子工学専攻

^{††} 東海大学情報理工学部情報科学科

1. はじめに

近年、監視カメラや高齢者の見守り、自動ブレーキなどにおいて人の行動認識が必要とされている。特に自動ブレーキにおいては自動車の自動運転の実用化に向け、衝突回避のために歩行者の行動の認識・予測が必要である。また人の検出は古くから行われており、Navneetらの研究^[1]では HOG 特徴量を用いて検出を行っている。

畳み込みニューラルネット(convolutional neural networks, CNN)を用いた画像認識の研究が従来手法より高い認識精度を出したことで、CNN を用いた研究に関心が高まっている。Jiらの研究^[2]では、画像数フレームを CNN の入力とすることで、動画から人の行動認識において高い精度を出すことに成功している。また、RGB 画像と深度画像(RGB-D)を同時に用いることで RGB 画像のみの認識よりも高い精度を出している研究もある^{[3][4]}。

本研究では、RGB-D 画像数フレームを CNN への入力として人の行動を認識する。深度情報を用いない場合^[1]と用いた場合で精度を比較する。DHA データセット^[5]で CNN を学習させ、未知データにより行動分類の認識精度を測定する。

2. Convolutional Neural Network

CNN とは、多層パーセプトロンの 1 つで中間層において、畳み込み処理とプーリング処理を複数回繰り返す、特徴抽出をするニューラルネットワークである。CNN は教師あり学習を前提としており、誤差逆伝播法を用いた勾配降下最適化法で学習を行う。

3. 提案手法

RGB-D 画像数フレームを CNN の入力データとして、特徴抽出、行動の識別を行う。本研究では日常、公共の道路における歩行者の行動を対象とし、「立っている」「電話をしている」「走っている」などの行動を認識する。

4. まとめ

本研究では RGB-D 画像を入力とした CNN による行動分類手法を提案する。今後様々なデータベースを用いて他手法との比較や CNN 構造の違いによる認識精度の違いの検証を行う予定である。

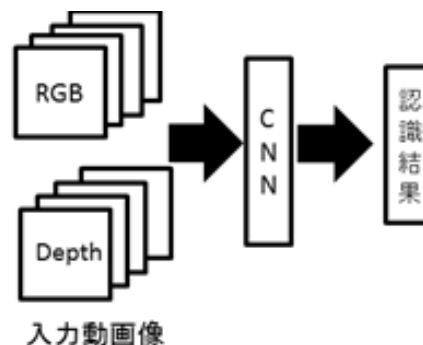


図1. RGB-D データを用いた認識システム

参考文献

- [1] D. Navneet, T. Bill, "Histograms of Oriented Gradients for Human Detection", *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, 2005, pp. 886-893
- [2] S. Ji, W. Xu, M. Yang, K. Yu, "3D Convolutional Neural Networks for Human Action Recognition", *IEEE Transactions on pattern analysis and machine intelligence*, vol.35, no.1, 2013, pp.221-231
- [3] J. Zhang, W. Li, P. O. Ogunbona, P. Wang, C. Tang, "RGB-D-based action recognition datasets: A survey", *Pattern Recognition*, vol.60, 2016, pp.86-105
- [4] K. Lai, L. Bo, X. Ren, D. Fox, "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset", *ICRA*, 2011, pp.1817-1824.
- [5] Y.-C. Lin, M.-C. Hu, W.-H. Cheng, Y.-H. Hsieh, and H.-M. Chen, "Human Action Recognition and Retrieval Using Sole Depth Information", *Proc. 20th ACM Int'l Conf. Multimedia(MM'12)*, 2012, Nara, Japan.