

シミュレーションの訓練データを用いた 深層学習による油圧ショベルの動作認識

Action Recognition of Excavator by Deep Learning Using Simulated Training Data

○ 沈鎮赫 (東大) ルイ笠原純ユネス (東大) 筑紫彰太 (東大)
山川博司 (東大) 正 田村雄介 (東北大) 正 永谷圭司 (東大)
千葉拓史 (フジタ) 山本新吾 (フジタ) 茶山和博 (フジタ)
正 山下 淳 (東大) 正 浅間 一 (東大)

Jinhyeok SIM, The University of Tokyo, sim@robot.t.u-tokyo.ac.jp
Jun Younes LOUHI KASAHARA, The University of Tokyo
Shota CHIKUSHI, The University of Tokyo
Hiroshi YAMAKAWA, The University of Tokyo
Yusuke TAMURA, Tohoku University
Keiji NAGATANI, The University of Tokyo
Takumi CHIBA, Fujita Co., Ltd.
Shingo YAMAMOTO, Fujita Co., Ltd.
Kazuhiro CHAYAMA, Fujita Co., Ltd.
Atsushi YAMASHITA, The University of Tokyo
Hajime ASAMA, The University of Tokyo

Measuring and recording an action performed by construction machinery is a very effective task for improving productivity of construction sites. However, measuring and recording the action by construction machinery is time consuming and expensive because the construction site managers have to observe and record data manually. Therefore, it is important to automatically recognize the action of construction machinery. Action recognition is achieved with high performance on humans using deep learning techniques but those approaches require large amounts of training data. There is no data set for action recognition of construction machinery. Therefore, the proposed method uses training data generated from a simulator. In this study, action recognition is performed for excavator that is most commonly used at construction sites. Experiments were conducted with a remote control excavator in laboratory conditions.

Key Words: Action recognition, Deep learning, excavator

1. 序論

建設業界において、施工現場の生産性向上は重要な課題であり、建設機械が行う動作を測定・記録することは作業を完了するのに必要な時間や費用が推定できるため生産性向上に非常に効果的なタスクである[1].

しかし、建設機械が行う動作の測定・記録は、現場の管理者が手動で観察および記録する必要があるため、時間や費用がかかる問題がある[2]. したがって、このような問題を解決するには、建設機械が行う動作を自動的に認識することが重要である.

建設機械が行う動作を自動的に認識する先行研究として、Pradhananga らと Akhavian らは全地球測位システム (GPS) や慣性測定ユニット (IMU) などセンサから情報を分析することにより自動的に建設機械の動作認識を行った[3][4]. しかし、これらの先行研究はセンサを建設機械に取り付ける必要があり、レンタル建設機械や古い建設機械には適さない問題が発生する[5]. そこで、本研究ではセンサの取り付けが不要で高い認識精度を示す深層学習を用いて、施工現場で最も用いられる油圧ショベルの動作認識を目指す.

深層学習を用いて油圧ショベルの動作認識を行うためには、膨大な訓練データが必要だが、実際の現場から油圧ショベルの訓練データを生成することは困難である. 本研究では、シミュレーションを用いて訓練データを生成し、油圧ショベルの動作認識を行う.



Fig. 1 Excavator in a simulation environment

2. 油圧ショベルの動作認識

2.1 深層学習を用いた油圧ショベルの動作認識

動作認識の代表的な手法として、CNN と Long Short Term Memory (LSTM) を用いた手法[6]や 3D CNN を用いた手法[7] など深層学習を用いた手法がある. 深層学習を用いた動作認識は、他の手法より高い認識精度を示している. そこで、本研究でも深層学習を用いた油圧ショベルの動作認識を行うこととした. ただし、深層学習を用いた動作認識を行うためには、膨大な訓練データが必要となる. 従来研究の殆どが人間の行動認識を中心に研究されてきたため、油圧ショベルの動

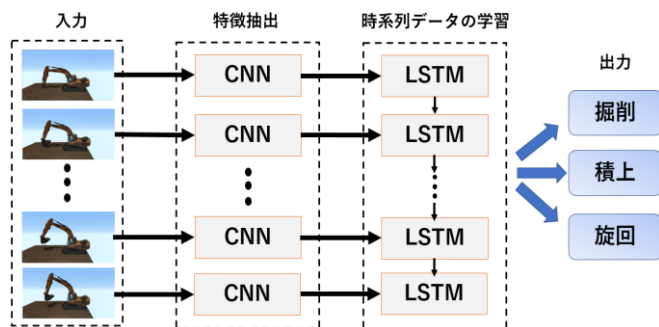


Fig. 2 Overview for action recognition of excavator

作認識用のデータセットは存在しない。そこで、本研究では、油圧ショベルの動作認識用のデータセットを自ら用意し、深層学習を用いて油圧ショベルの動作認識を行う。

2.2 動作認識用の訓練データ

油圧ショベルの動作認識の訓練データとして最も理想的な訓練データは、実際の施工現場で作業を行う油圧ショベルの映像だが、実際の作業映像を大量に集めることは現実的に困難である。そこで、本研究では機械システム操作のリアルタイムシミュレータである Vortex Studio[8]を用いて訓練データを生成する。図 1 は、Vortex Studio からの油圧ショベルの様子である。今回深層学習の入力データとしては、RGB データを用いることにした。しかし、RGB データは背景やカメラ視点に影響されやすい問題がある。そこで、本研究では図 1 のように油圧ショベル以外の背景に影響されないよう油圧ショベル 1 台とその周辺に土だけを置いた背景を用いることとした。また、カメラ視点による影響を減らすために、複数のカメラ視点(多角度)から訓練データを生成した。

訓練データの作り方としては、まず Vortex Studio 上でのカメラ視点を固定させた後、コントローラを用いて自ら油圧ショベルを操作する。その時、認識したい複数の動作を均等かつ継続的に操作しその様子の動画を収録する。複数の視点からも同じく油圧ショベルの様子の動画を収録する。その後、複数の視点から得られた全ての動画から 1 つの動作ごとに 1 つの動画を切り取り、訓練データを生成する。つまり、1 つの訓練データには、1 つの動作(動画)とその動作のラベルが存在する。

2.3 学習モデル

本研究では、学習モデルとして CNN と LSTM を用いた。CNN は中間層が畳み込み層とプーリング層で構成され、空間情報を含んだ特徴マップを抽出するネットワークで、LSTM は時系列データに適した長期的な依存関係が学習可能なネットワークである。そのため、CNN と LSTM を用いれば、空間的情報と時間的情報を同時に考慮した油圧ショベルの動作認識が可能となる。

図 2 にモデルの概要を示す。まず、フレームレートが 30fps である訓練データ(動画)をフレームごとに分割し RGB データ(画像)を抽出する。次に抽出された RGB データを $298 \times 298 \times 3$ のサイズに縮小する。 $298 \times 298 \times 3$ のサイズに縮小した RGB データを CNN に入力し特徴を抽出する。今回用いた CNN ネットワークは、100 万枚以上の画像から学習した Inception V3[9] という学習済みモデルを用いた。その後、Inception V3 より特徴抽出された結果を LSTM に入力する。LSTM は 3 つのレイヤで構成され、最後の softmax レイヤで動作ラベルを分類する。

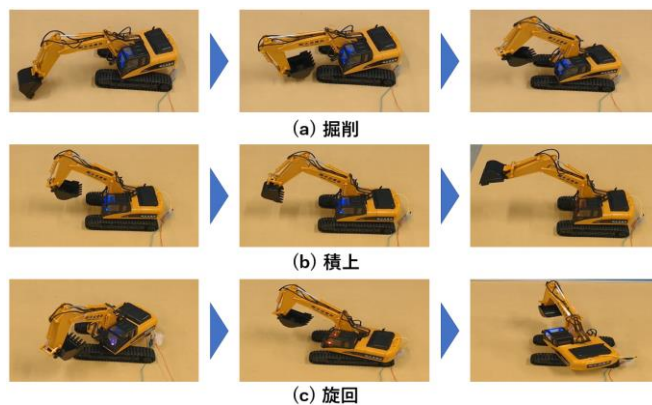


Fig. 3 Examples of working of excavator:
(a) Digging; (b) Loading; (c) Swinging



Fig. 4 Remote control excavator

3. 動作認識実験

3.1 訓練データ

今回は、掘削・積上・旋回の 3 つの動作の認識を行うことを目標とし、2.2 節の手順に従い 4 つの視点から訓練データを生成した。その結果、動作ごとに約 55 個の訓練データ(動画)が生成された。動画の平均長さは 7 秒で、最短長さは 4 秒、最長長さは 13 秒である。生成した訓練データより 2.3 節モデルの学習を行った。図 3 に掘削・積上・旋回の様子を示す。

3.2 テストデータ

学習させたモデルを用いて油圧ショベルの動作認識の実験を行った。油圧ショベルの動作認識用のテストデータを、訓練データより学習させたモデルに入力すると、油圧ショベルの動作ラベル(掘削・積上・旋回)が出力される。

本研究では、シミュレーションの訓練データから学習させたモデルが実際の油圧ショベルの動作認識にどのくらい精度が出るか確認・考察するため、2 つのテストデータを用意した。1 つ目は、3.1 節と同じ環境で生成したテストデータで、2 つ目は、ラジコンの油圧ショベルで生成したテストデータである。今回実際に施工現場で動いている油圧ショベルの作業映像でテストデータを生成することが困難だったため、ラジコンの油圧ショベルを用いることとした。それぞれのテストデータより油圧ショベルの動作認識実験を行った。

3.3 シミュレーション環境での動作認識実験

テストデータは動作ごとに約 20 個の動画を生成した。シミュレーション環境での油圧ショベルの動作認識を行った結果、85.5%の精度が得られた。

Table 1 Recognition accuracy

	シミュレーションより生成したテストデータ	ラジコンの油圧ショベルより生成したテストデータ
認識精度	85.5%	36.7%

3.4 ラジコンの油圧ショベルを用いた動作認識実験

まず、ラジコンの油圧ショベル 1 台を用意し、シミュレーションの背景とできるだけ同じになるようにした。その後、コントローラを用いて掘削・積上・旋回の 3 つの動作を行い、その時のラジコンの油圧ショベルの様子を 4 つの視点から三脚に固定されたビデオカメラで撮影した。図 4 にラジコンの油圧ショベルを用いた実験の様子を示す。テストデータは動作ごとに約 20 個の動画を生成した。ラジコンの油圧ショベルの動作認識を行った結果、36.7%の精度が得られた。2 つのテストデータから得られた認識精度の結果を表 1 に示す。

3.5 考察

訓練データと同じ環境（シミュレーション環境）で生成したテストデータは高い認識精度が出たものの、ラジコンの油圧ショベルで生成したテストデータの認識精度は非常に低いものとなった。これは、訓練データの数や視点が少なかったことに原因があると考えられる。一般に、深層学習を用いた動作認識で高い精度を出すには、膨大な訓練データが必要であり、今回用いた訓練データの数や視点の数をさらに増やす必要がある。また、今回入力データとして RGB データを用いたが、RGB データは背景や照明、視点に影響されやすく、これがシミュレーション環境と実環境でのギャップを縮めなかった原因でもあったと考えられる。

4. 結論

本研究では、深層学習を用いて油圧ショベルの動作認識を行った。油圧ショベルの動作認識用のデータセットが存在しなかったため、Vortex studio を用いて訓練データを生成した。その後、訓練データより CNN と LSTM のモデルを学習させ、学習させたモデルからテストデータの動作認識を行った。その結果、訓練データと同じシミュレーション環境で生成したテストデータでは 85.5%、ラジコンの油圧ショベルで生成したテストデータでは 36.7%の精度が得られた。今後の予定として、訓練データの数と視点を増やすことと背景や照明、視点に影響されにくい新しい入力データを用いることを検討する。

参考文献

- [1] Jinwoo Kim, Seokho Chi, and Jongwon Seo: "Interaction Analysis for Vision-Based Activity Identification of Earthmoving Excavators and Dump Trucks," *Automation in Construction*, Vol. 87, pp. 297-308, 2018.
- [2] Hyunsoo Kim, Changbum R.Ahn, David Engelhaupt, and Sanghyun Lee: "Application of Dynamic Time Warping to the Recognition of Mixed Equipment Activities in Cycle Time Measurement," *Automation in Construction*, Vol. 87, pp. 225-234, 2018.
- [3] Nipesh Pradhananga, and JochenTeizer: "Automatic Spatio-Temporal Analysis of Construction Site Equipment Operations Using GPS Data," *Automation in Construction*, Vol. 29, pp. 107-122, 2013.
- [4] Reza Akhavian, and Amir H.Behzadan: "Construction Equipment Activity Recognition for Simulation Input Modeling Using Mobile Sensors and Machine Learning Classifiers," *Advanced Engineering Informatics*, Vol.29, pp. 867-877, 2015.
- [5] Ehsan Rezazadeh Azar, Sven Dickinson, and Brenda McCabe: "Server Customer Interaction Tracker: Computer Vision-Based System to Estimate Dirt-Loading Cycles," *Journal of Construction Engineering*

and Management, Vol. 139, No. 7, pp. 785-794, 2013.

- [6] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell: "Long-term Recurrent Convolutional Networks for Visual Recognition and Description," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, pp. 2625-2634, 2015.
- [7] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Tor-resani, and Manohar Paluri: "Learning Spatiotemporal Features with 3D Convolutional Networks," In *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2015)*, pp. 4489-4497, 2015.
- [8] Vortex Studio. <https://www.cm-labs.com/vortex-studio/> (accessed 2020.02.19)
- [9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna: "Rethinking the Inception Architecture for Computer Vision," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 2818-2826, 2016.