

OpenPoseおよび複数識別器を用いた 自転車の回避行動の検出に関する検討

八木 雅大¹・高橋 翔²・萩原 亨³

¹学生会員 北海道大学 大学院工学院 (〒060-8628 北海道札幌市北区北13条西8丁目)
E-mail:masahiro@eis.hokudai.ac.jp

²正会員 北海道大学准教授 大学院工学研究院 (〒060-8628 北海道札幌市北区北13条西8丁目)
E-mail:stakahashi@eng.hokudai.ac.jp

³正会員 北海道大学教授 大学院工学研究院 (〒060-8628 北海道札幌市北区北13条西8丁目)
E-mail:hagiwara@eng.hokudai.ac.jp

近年、カメラを搭載した車の数が増加しており、車から前方を撮像した映像の解析が可能となってきた。一方、自転車と車の接触事故が起こりやすい状況の一つとして自転車による回避行動がある。これを検出可能とすることで、より安全な交通を実現するための分析手法の構築につながる。また、自転車は荷重移動によって駆動するものであり、自転車の走行とライダーの姿勢には関連があると考えられる。そこで、本文では、動画に撮像された自転車の回避行動をライダーの姿勢に基づいて検出する手法を提案する。提案手法ではまず、人間の骨格データを抽出可能なOpenPoseを用いて、ライダーの姿勢に関する特徴量を求める。最後に、得られた特徴量を入力とする識別器を複数構築し、それらの識別結果に対する確信度を用いて、自転車の回避行動を検出する。

Key Words : *Bicycle, Avoidance behavior, OpenPose, Machine learning, Deep learning*

1. はじめに

近年、カメラを搭載した車の数が増加しており、車から前方を撮像した映像の集積が可能になりつつある。このため、それらの映像から道路状況を把握可能とすることによって、道路交通のさらなる高度化が期待できる。中でも特に、交通事故に関する場面の解析は、より安全な道路交通を実現するうえで効果的であり、種々の研究が行われている^{1,2)}。具体的にそれらの研究は、交通事故およびヒヤリハットが発生した場面の検出や交通事故の発生の予測を主な目的としている。しかし、道路状況は複雑であり、時間や場所によって常に変化するものである。したがって、上記の研究のように交通事故が発生するか否かといった点に注目するだけでなく、車や自転車、歩行者などがどのような行動を行っているかなどを把握することが、道路交通の安全性を向上させる上で、有用であると考えられる。

交通事故に関わる場面の一つに自転車による回避行動がある。図-1の例が示すように、自転車による回避行動は、通常時より車との距離が短くなるため、接触事故が発生しやすい行動の一つである。したがって、回避行動

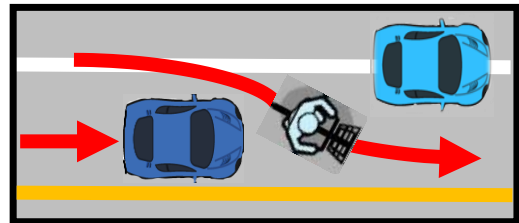


図-1 自転車の回避行動の例

が発生している場面を検出し、分析することで、接触事故を減らすための対策の構築につながると考えられる。そこで我々は以前、回避行動が発生している場面の検出につながる手法として、映像内に撮像された自転車が、回避行動を行っているか否かを識別する手法を提案した³⁾。文献³⁾では、自転車が、ライダーの荷重移動により車輪を回転させることで駆動する車両であることに着目した。具体的には、自転車の動きとライダーの姿勢には高い関連があると考え、ライダーの姿勢に関する特徴量を入力とするSupport Vector Machine (SVM)⁴⁾を構築することで、回避行動の検出を実現した。また、機械学習においては、識別器を複数構築し、それらの識別結果を統合することで、精度の向上が期待できることが知られて

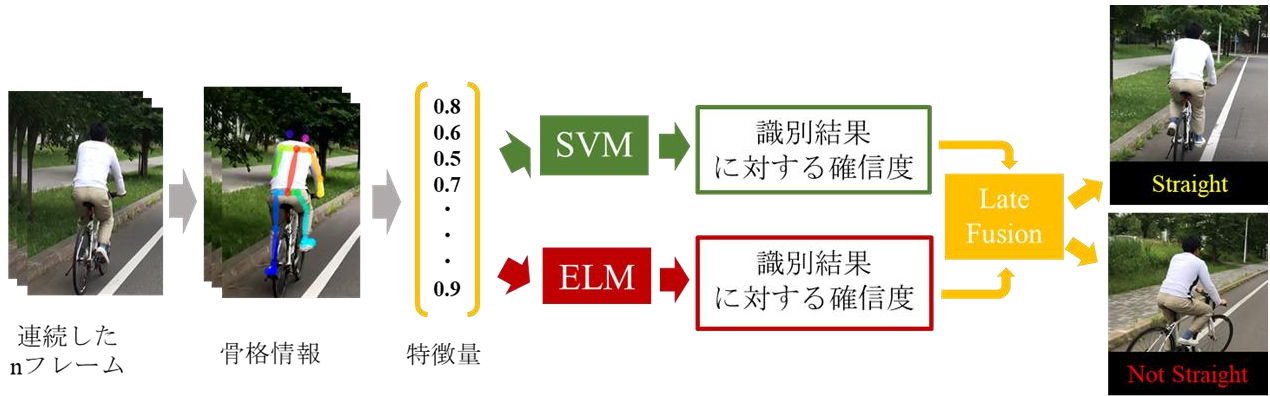


図-2 提案手法の概要

表-1 ライダーの姿勢に関する特徴量

| ライダーの姿勢に関する特徴の種類 | 次元数 |
|------------------|-----------|
| 相対座標の平均値 | 50 |
| 隣接フレーム間での相対座標の差分 | $50(n-1)$ |
| 関節角度の平均値 | 10 |
| 計 (d_n) | $50n+10$ |

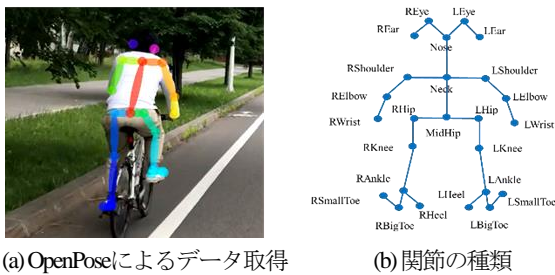


図-3 特徴量の算出

いる⁵⁾。また、識別結果の統合方法の一つとして、識別結果に対する確信度を用いた手法がある⁶⁾。

そこで本文では、ライダーの姿勢に関する特徴量を入力とする識別器を複数構築し、それらの識別結果に対する確信度を用いることで、回避行動を検出する手法を提案する。本手法では、回避行動が発生する状況を、自転車がまっすぐに走行できない状況と考え、映像に撮像された自転車の走行が、まっすぐか否かを識別することで、回避行動を検出する。具体的に、提案手法ではまず、ライダーの骨格データをOpenPose⁷⁾を用いて取得する。OpenPoseは、畳み込みニューラルネットワークを用いて、画像に撮像された人間の関節の位置座標を取得可能である。次に、得られたデータから、ライダーの姿勢に関する特徴量を求める。最後に、求めた特徴量を入力とする識別器として、SVMおよび3層のニューラルネットワークからなるExtreme Learning Machine (ELM)⁸⁾を構築し、それらの識別結果に対する確信度を用いて、回避行動を検出する。

2. 提案手法

本章では、提案手法について説明する。提案手法の概要を図-2に示す。提案手法では、連続した n ($n \geq 2$) フレーム(以降、ビデオクリップ)から回避行動を検出する。具体的に、まず初めに、各フレームに対して、OpenPoseを

適用することで、ライダーの骨格データを求める。次に、求めた骨格データをもとに、ライダーの姿勢に関する特徴量を算出する。続いて、求めた特徴量を入力とする識別器として、SVMおよびELMを構築する。最後に、構築した識別器を用いて、まっすぐな走行によるものか否かをビデオクリップごとに分類し、識別結果に対する確信度を用いて、回避行動を検出する。

(1) ライダーの姿勢に関する特徴量の算出

本節では、ライダーの姿勢に関する特徴量の算出方法について説明する。提案手法では、まず、各フレームから、図-3に示すライダーの関節の位置座標を、OpenPoseを用いて求める。次に、求めた関節の位置座標から、腰の座標を起点とした関節それぞれの相対座標を得る。最後に、取得した相対座標をもとに、表-1に示す特徴量を算出する。

(2) 識別器の構築および確信度の算出

本節では、前節で求めた特徴量を入力とするSVMおよびELMを構築し、識別結果に対する確信度を算出する。

a) SVMの構築及び確信度の算出

SVMは画像認識等を目指した研究で広く用いられている識別器であり、特徴ベクトルを高次元特徴空間に写像し、識別超平面を構築することで分類を行う。具体的に、提案手法では、前節までに求めたライダーの姿勢に関する d_n 次元の特徴ベクトル v を、以下に示す識別関数 f_{svm} に入力することで、ラベル z ($z \in \{st, n_st\}$)を決定

する。ただし、 st は“まっすぐな走行”， n_st は“まっすぐでない走行”を意味する。

$$f_{svm}(v) = w^T v + b, \quad (1)$$

$$\begin{cases} st & f_{svm}(v) > 0 \\ n_st & otherwise \end{cases} \quad (2)$$

ただし、 w, b は学習によって、求められるパラメータである。SVMでは、識別超平面と識別超平面に最も近い特徴ベクトルとの距離を最大化することでパラメータを求める。そこで提案手法では、特徴ベクトル $v_i (i \in \{1, \dots, M\})$ およびそのラベル z_i からなる M 個の学習データを用いて w, b を求める。具体的には、制約条件

$$\min |w^T v_i + b| = 1 \quad (3)$$

の下で、

$$\min \frac{1}{2} \|w\|^2 \quad (4)$$

を解くことで、 w, b を求める。また識別結果の確信度 P_{SVM}^z は、識別超平面からの距離を用いて算出する。また、 $0 \leq P_{SVM}^z \leq 1$ である。

b) ELMの構築および確信度の算出

ELMは3層のニューラルネットワークからなる識別器である。提案手法では、学習データとして、 d_n 次元の特徴ベクトル $v_i = [v_{i1}, v_{i2}, \dots, v_{id_n}]^T$ およびそのラベル $z_i = [z_{st}, z_{n_st}]^T$ (ただし、 v_i の属するクラスが st の場合、 $z_{st} = 1$ となり、 $z_{n_st} = 0$ となる。)を用いる。また、提案手法では、最終層の重み β を以下によって求める。具体的には、シグモイド関数 G を用いて、特徴変換を行うことで、

$$h(v_i) = [G(a_1, b_1, v_i), \dots, G(a_L, b_L, v_i)]^T \quad (5)$$

を算出する。ただし、 $a_l (l = 1, 2, \dots, L)$ 、および $b_l (l = 1, 2, \dots, L)$ はシグモイド関数 G のパラメータであり、 L は隠れ層のノード数を表す。次に、

$$\beta = (H^T H)^{-1} H^T T \quad (6)$$

により最終層の重み β を求める。ただし、 $T = [z_1, z_2, \dots, z_M]^T$ 、 $H = [h(v_1), h(v_2), \dots, h(v_M)]^T$ である。最後に、テストにおいて、特徴ベクトル v をELMに入力したとき、出力値は $f = h(v)^T \beta$ であり、クラスラベルは f の内、最も大きな値を出力したノードに対応するクラスラベルとなる。また、出力値 f にソフトマックス関数を適用することにより、識別結果の確信度 $P_{ELM}^z (z \in \{st, n_st\})$ を算出する。

$$P_{ELM}^z = \frac{e(f_z)}{e(f_{st}) + e(f_{n_st})} \quad (7)$$

ただし、 f_z は出力値 f において、ラベル z に対応するノードの値である。また、 $0 \leq P_{ELM}^z \leq 1$ である。

(3) 識別結果に対する確信度を用いた検出

本節では、前節までに構築したSVMおよびELMの、識別結果に対する確信度を用いて、回避行動を検出する。

表-2 実験で用いるデータセットの内訳

| | ラベル | 映像数 | ビデオクリップ数 |
|------|---------|-----|----------|
| 学習用 | st | 47 | 22,320 |
| | n_st | 52 | 22,320 |
| テスト用 | st | 20 | 9,998 |
| | n_st | 20 | 10,965 |

表-3 各手法における再現率, 適合率, F値

| | st | | | n_st | | |
|-----|--------------|-------|-------|--------------|--------------|-------|
| | PM | CM1 | CM2 | PM | CM1 | CM2 |
| 再現率 | 0.891 | 0.881 | 0.816 | 0.919 | 0.919 | 0.783 |
| 適合率 | 0.909 | 0.908 | 0.774 | 0.902 | 0.894 | 0.824 |
| F値 | 0.900 | 0.894 | 0.795 | 0.910 | 0.906 | 0.803 |

具体的に、まず、構築した識別器を用いて、まっすぐな走行か否かをビデオクリップごとに分類する。この時、SVMによる識別の確信度 P_{SVM}^z およびELMによる識別の確信度 P_{ELM}^z を算出する。最後に、次式により、“まっすぐな走行”か否かを決定する。

$$\begin{cases} st & (P_{SVM}^{st} - P_{ELM}^{n_st}) > 0 \\ n_st & (otherwise) \end{cases} \quad (8)$$

以上により、提案手法では、自転車の回避行動を検出する。

3. 実験

本章では、提案手法の有効性を確認するための実験を行う。本実験では、表-2の学習データを用いて識別器を構築する。これらの映像は、自転車が後方から撮像されている映像である。また、走行の種類は、“まっすぐな走行”および“まっすぐでない走行”の2種類である。本実験では、“まっすぐでない走行”を回避行動と仮定し、各ビデオクリップ ($n=3$) の走行種類を分類する。本実験で用いるデータセットの内訳を表-2に示す。また本実験では、次の二つの比較手法を用いる。比較手法1は識別器がSVMのみの手法であり、比較手法2は識別器がELMのみの手法である。また、評価指標として、以下に示す、ラベル z における再現率 $Recall_z$ 、適合率 $Precision_z$ およびF値 $F - measure_z$ を用いる。

$$Recall_z = \frac{\text{ラベル}z\text{に正しく分類されたビデオクリップの総数}}{\text{ラベル}z\text{であるビデオクリップの総数}}, \quad (9)$$

$$Precision_z = \frac{\text{ラベル}z\text{に正しく分類されたビデオクリップの総数}}{\text{ラベル}z\text{に分類されたビデオクリップの総数}}, \quad (10)$$

$$F - measure_z = \frac{2 \times Recall_z \times Precision_z}{Recall_z + Precision_z} \quad (11)$$

また、SVMではガウシアンカーネルを用い、パラメータはグリッドサーチ⁹⁾により決定する。また、ELMのノード数Lは実験的に800と設定した。実験結果を表-3に示す。ここで、PMおよびCMはそれぞれ、提案手法、比較手法を表す。表-3より、全ての評価指標において、提案手法の精度が、比較手法の精度と同等、あるいはそれを上回ることが確認できる。以上より、自転車の回避行動の検出において、ライダーの姿勢に関する特徴量を入力とする複数の識別器を構築し、それらの識別結果に対する確信度を用いることの有効性が確認された。

4. 結論

本文では、OpenPoseおよび複数識別器を用いた、自転車の回避行動の検出手法を提案した。また、実験により、その有効性が確認された。

謝辞：本研究の一部は、JSPS 科研費 JP17K00148, JP19H02254および公益財団法人戸田育英財団の助成を受けて行われた。

参考文献

- 1) L. Taccari, F. Sambo, L. Bravi, S. Salti, L. Sarti, M. Simoncini and A. Lori : Classification of Crash and Near-Crash Events from Dashcam Videos and Telematics, Proc. In the IEEE International Conference on Intelligent Transportation Systems, pp. 2460-2465, 2018.
- 2) F.-H. Chan, Y.-T. Chen, Y. Xiang and M. Sun : Anticipating Accidents in Dashcam Videos, Proc. In the Asian Conference on Computer Vision 2016, pp. 136-153, 2016.
- 3) M. Yagi, S. Takahashi and T. Hagiwara : An Evaluation Method of Obstacle Avoidance Behavior on Bicycle Trip Using Rider's Gesture, Proc. In the IEEE 8th Global Conference on Consumer Electronics, (October 2019, Accepted for presentation).
- 4) C. Cortes and V. Vapnik : Support-vector networks, Machine learning, Vol. 20, no. 3, pp. 273-297, 1995.
- 5) F. Sebastiani : Machine learning in automated text categorization, In ACM computing surveys, Vol. 34, no. 1, pp. 1-47, 2002.
- 6) 齊藤直輝, 小川貴弘, 長谷山 美紀 : 画像共有サービスに投稿される画像の観光名所に関するカテゴリ分類, 電子情報通信学会論文誌 D, Vol. 99, no. 9, pp. 848-860, 2016.
- 7) Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh : Realtime multi-person 2d pose estimation using part affinity fields, Proc. In the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291-7299, 2017.
- 8) G.-B. Huang, Q.-Y. Zhu and C.-K. Siew : Extreme Learning Machine: a new learning scheme of feedforward neural networks, Proc. In the IEEE International Joint Conference on Neural Networks, Vol. 2, pp. 985-990, 2004.
- 9) C.-W. Hsu, C.-C. Chang and C.-J. Lin : A practical guide to support vector classification, Technical Report, Department of Computer Science, pp. 1-16, 2003.

(?)

A NOTE ON DETECTION OF AVOIDANCE BEHAVIOR ON BICYCLE TRIP USING OPENPOSE AND MULTIPLE-CLASSIFIERS

Masahiro YAGI, Sho TAKAHASHI and Toru HAGIWARA

Recently, cameras are equipped on many vehicles, which makes it easy to acquire the video and analyze it. Also, the avoidance behavior on bicycle trip is a cause for the bicycle-car accident. Thus, to grasp the scene where the avoidance behavior occurs is thought to be effective to improve the road traffic safety. Then, since a bicycle is a vehicle which is ridden by rotating a wheel with load transfer of a rider, there is a high correlation between bicycle behavior and rider's gesture. Therefore, we propose a method to detect the avoidance behavior on bicycle trip in videos based on rider's gesture. In the proposed method, features about rider's gesture are obtained based on OpenPose which can detect human's 2D skeleton data. In addition, multiple-classifiers are constructed based on the features. Finally, the avoidance behavior is detected by fusing the probability of estimation obtained with the classifiers. The experimental results show the effectiveness of our method.