情報処理学会論文誌　データベース　Vol.10　No.4　26–30 (Dec. 2017)

テクニカルノート

映像とセンサ信号を用いたドライブレコーダデータからのヒヤリハット検出手法

山本 修平[1,a)]　遠藤 結城[1,†1]　戸田 浩之[1,b)]

受付日 2017年6月9日, 採録日 2017年8月2日

概要：本論文では，ドライブレコーダデータからの，映像とセンサ情報を組み合わせたヒヤリハット検出手法を提案する．提案手法では，各時刻のセンサ情報と画像を全結合ニューラルネット，および畳み込みニューラルネットによってそれぞれ効果的な特徴表現へとエンコードする．センサと画像の特徴表現を結合し，再帰型ニューラルネットワークによってデータの時間的な状態遷移を考慮する．実際のドライブレコーダデータを用いたヒヤリハットの検出実験を行った結果，センサか映像いずれかの情報を欠損させた手法や，時系列を考慮しない手法に比べ，提案手法が高い検出性能を示すことを明らかにした．

キーワード：深層学習，ヒヤリハット検出，ドライブレコーダ，マルチモーダルデータ

Near-miss Scene Detection from Event Recorder Data Using Video and Sensor

Shuhei Yamamoto[1,a)]　Yuki Endo[1,†1]　Hiroyuki Toda[1,b)]

Received: June 9, 2017, Accepted: August 2, 2017

Abstract: In this paper, we propose a method to automatically detect near-miss scenes using multi-modal data including video and sensor data from event recorder data collected by vehicles. Our method encodes image and sensor data at each time-step into efficient feature representation via convolutional neural networks and fully-connected networks. Also, it can handle sequential information of the concatenated features of them by using recurrent neural networks. In our experimental evaluation of near-miss scene detection from real event recorder data, we demonstrate that our proposed method shows higher detection performance than three methods of without using sensor, video, or sequential information.

Keywords: deep neural network, near-miss scene detection, event data recorder, multi-modal data

1. はじめに

ドライブレコーダとは，車内に設置され，前方映像，加速度，ウィンカ操作，ブレーキ操作などの車両運行状況を記録する装置である．全時刻の情報を記録することは，記憶容量などの問題もあるため，実際には加速度に一定の閾値を設け，車両に何らかの衝撃が見られたと考えられるタイミングを検知し，その前後十数秒を記録することが多い．このような映像やセンサ情報は，交通事故の証拠だけでなく，交通違反をした映像や交通事故に近いヒヤリハット映像をドライバに実際に視聴してもらうことによって，安全運転を促進することにも活用されている[*1]．GfK ジャパン社の調査によると，ドライブレコーダの国内販売台数は 2016 年に前年比 29%増の 79 万台となっており，今後もさらなるヒヤリハット映像の収集が見込まれる[*2]．

一方，ドライブレコーダで記録されたすべてのデータが

[1] 日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, NTT Corporation, Yokosuka, Kanagawa 239–0847, Japan
[†1] 現在，筑波大学
Presently with University of Tsukuba
[a)] yamamoto.shuhei@lab.ntt.co.jp
[b)] toda.hiroyuki@lab.ntt.co.jp

[*1] https://www.npa.go.jp/bureau/traffic/anzen/drive_recorder.html
[*2] http://www.gfk.com/fileadmin/user_upload/dyna_content/JP/20170330_2016drivingrecorder2.pdf

© 2017 Information Processing Society of Japan

ヒヤリハットに関するものではない．たとえば，道路の激しい起伏が原因で車両に大きな衝撃が加わり記録されたものや，狭い道路での方向転換のために小刻みな加速を必要とされ記録されたものなど，ヒヤリハットでない映像も多く含まれる．収集された膨大なドライブレコーダデータの中から，安全運転教育に資するヒヤリハットを人手で抽出することは，その判定者に多くの労力や注意力を要するため，これらの蓄積されたデータの中から，自動的にヒヤリハットを含むデータを抽出することが望まれている．

これまでも，ヒヤリハットを自動的に検出することに取り組んでいる研究がいくつか存在する．ドライブレコーダに含まれるセンサ信号に対して，経験的にルールを設定しヒヤリハットの抽出を試みている研究 [1] や，センサデータを適切な特徴量に変換して右折時に限定したヒヤリハット発生形態を分析する研究 [2] などあるが，これらは前方映像には触れていない．一方で，センサ情報からは非ヒヤリハットと推定されるが，前方映像からはヒヤリハットと推定できるデータも十分に考えられる．たとえば，見通しの悪い信号のない交差点への侵入を試みる自動車が，一時停止後にアクセルを踏んで発進した直後に自転車が急に横切るケースでは，センサに大きな変化はないもののヒヤリハットと認められる．実際，このような出会い頭の事故は，警察庁が公開する平成 28 年の交通事故の発生状況によれば，類型化された交通事故の中で 2 番目に多いパターンである[*3]．このような理由から，より網羅的なヒヤリハット検出のためには，映像も考慮すべきであると考えられる．

以上の考察をふまえ，本論文ではセンサと映像の両情報源と時系列の遷移を考慮した，深層学習を活用したヒヤリハット検出手法を提案する．センサデータは全結合ニューラルネット，映像データはフレーム画像に分割し畳み込みニューラルネットによってそれぞれ適切な特徴表現へエンコードする．それらの特徴表現を各時刻で結合した後，再帰型ニューラルネットによって時間的な時系列モデリングし，最終的な分類結果（ヒヤリハットか否か）を得る．

2. 関連研究

本章では，ドライブレコーダに含まれるセンサを対象とした関連研究と，前方映像を対象とした関連研究を述べ，本研究の位置付けを整理する．

センサデータに着目した研究． ドライブレコーダに含まれるセンサを対象とした研究として，ドライバの運転操作からヒヤリハットの起こりやすい潜在リスクを持つ道路の推定に関する研究 [3], [4]，ヒヤリハットを起こしやすい危険運転ドライバの分類に関する研究 [5] などがある．特に，本論文で対象とするヒヤリハットデータの検出と同じタスクに取り組んでいる研究 [1], [2] もある．これらはスピードや加速度などを詳細に分析し，センサのみからヒヤリハットデータを検出することを目指している．しかし，道路の起伏が原因で加速度は異常と見なされるが，実際にはヒヤリハットの発生していないデータを考慮するためには，センサだけでなく映像も組み合わせた検出手法が望ましい．本論文では，センサと映像を用いたヒヤリハット検出手法の実現を目指す点で，これらの研究とは異なる．

映像データを対象とした研究． 車両の前方映像は，近年，自動車の自動運転に関する研究で幅広く利用されている．自動運転タスクでは，前方映像に基づいてハンドルの操舵角を予測する研究 [6]，車の直進，左折，右折，車線変更，停止などの運転操作を予測する研究 [7]，同様のタスクに車内映像も用いた研究 [8]，前方で発生する交通事故を予測する研究 [9] などがある．これらはいずれも深層学習に基づく手法を提案しており，画像解析は Convolutional Neural Netowork（CNN），時系列モデリングは Recurrent Neural Network（RNN）において有効性の知られている Long Short-Term Memory（LSTM）を用いている．本論文でも CNN や LSTM を用いたモデル化に取り組むものの，上記のタスクでは運転操作やハンドル操舵角，交通事故の起こりやすさを予測するが，ヒヤリハット検出タスクでは入力の最終時刻でヒヤリハットか否かの判定する点で，これらの研究と提案するモデルやタスクが異なる．

3. 深層学習に基づくヒヤリハット検出手法

本論文で提案する，深層学習に基づくヒヤリハット検出のネットワーク構造を**図 1** に示す．入力データは，センサデータと画像から構成される固定長の系列を想定する．以下，それぞれのデータの処理過程を，提案ネットワークを構成する要素を用いながら説明する．

センサデータ処理ユニット． 入力されるセンサデータは複数の次元（加速度やスピードなど）からなるベクトルであり，入力された系列データに対してそれぞれの次元で平均 0，標準偏差 1 になるよう正規化される（**Norm**）．その後，センサデータは Fully-connected layer によって，256 次元の特徴ベクトルへエンコードする（**FC-S**）．

画像データ処理ユニット． 画像データは 3 層の CNN によってエンコードする（**Conv1**，**Conv2**，**Conv3**）．それぞれの層では，画像を CNN で畳み込んだ後，ニューラルネットワークの学習を高速化する技術として使用される Batch Normalization [10] をし，対象とする特徴量の出現位置のわずかな変化に対しても頑健に特徴を抽出できる Max Pooling [11] を行う．畳み込みは，いずれの層でも Stride は 1 とし，フィルタサイズは 1 層目で 7×7，2 層目で 5×5，3 層目で 3×3 としている．Max Pooling は，いずれの層でも Stride を 2，フィルタサイズを 2×2 としている．ここで，すべての Batch Normalization と Max Pooling の間で用いる活性化関数には ReLu を用いている．

[*3] http://www.e-stat.go.jp/SG1/estat/Pdfdl.do?sinfid=000031559551

© 2017 Information Processing Society of Japan

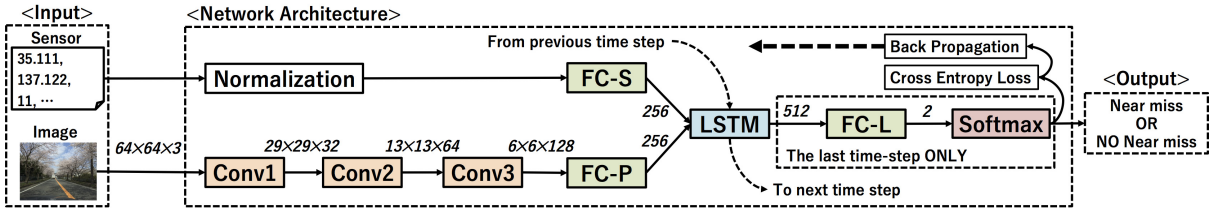


図 1　ヒヤリハット検出のネットワーク構造
Fig. 1　Network architecture of near-miss scene detection.

たとえば，64 × 64 × 3 の解像度の画像が与えられた場合には，図 1 に示しているように，29 × 29 × 32，13 × 13 × 64，6 × 6 × 128 と特徴表現が変換される．このようにしてエンコードされた画像データを行列からベクトルへ変換した後，Fully-connected layer を用いてセンサデータと同様に 256 次元の特徴ベクトルにさらにエンコードする（**FC-P**）．

時系列データのモデリングユニット．それぞれ適切な特徴表現へと変換されたセンサデータと画像データを結合し，LSTM ユニットを持つ RNN へ入力する（**LSTM**）．LSTM は系列データの長期的な依存関係を学習することができることから，ヒヤリハット検出に必要な時間的な状態遷移のモデル化を期待できる．

ヒヤリハット分類ユニット．LSTM によって獲得した時間的な状態遷移も考慮した特徴表現を，Fully-connected layer によって 2 次元の特徴量へエンコードする（**FC-L**）．この 2 次元が，ヒヤリハットと非ヒヤリハットの 2 ラベルに対応し，SoftMax 関数を通じて最大の値を持つラベルを推定結果として出力する（**Softmax**）．

LSTM は各時刻の入力に対しても過去の状態を考慮して適切な推定が可能であるが，本論文で対象とするヒヤリハット検出では，1 つの系列データに対し 1 つの正解を想定している．このため，系列データの最終時刻以外では，FC-L 以降へ順伝播はしない．また，下記で説明するネットワークの学習も最終時刻の結果に基づいて実行する．

ネットワークの学習．ネットワークの学習では，誤差逆伝播法で求められた交差エントロピー誤差の勾配に基いて，Adam によって最適化する[*4]．このとき，mini-batch のサイズは 10 に固定している．

4.　評価実験

実験データと設定．本章では，実際のドライブレコーダデータを用いた実験を行い，提案手法の検出性能を評価する．実験には，東京農工大学スマートモビリティ研究拠点の提供する「ヒヤリハットデータベース」を利用した[*5]．ヒヤリハットデータベースでは，各データは衝撃が加わった時刻を起点に前 10 秒，後 5 秒の約 15 秒からなり，各ドライブレコーダデータに訓練された人間によるヒヤリハットレベル（高，中，低，反応）と，非ヒヤリハットのラベルが付与されている．本論文では，ヒヤリハットレベルが低以上のデータを「ヒヤリハット」とし，ヒヤリハットを 100 件，非ヒヤリハットを 100 件，ヒヤリハットデータベースからそれぞれをランダムに抽出した．実験では 1 秒ずつフレーム画像とその時刻のセンサデータをサンプリングし，1 データにつき 15 個の画像とセンサデータを統一して抽出した．また，各フレーム画像は RGB 形式で 64 × 64 の解像度に線形変換した[*6]．センサデータは，GPS，ウィンカ操作，ブレーキ操作などある中から，本論文では，先行研究 [1], [2] や予備実験を参考に，X，Y，Z 方向の加速度とスピードの 4 種類を使用した．用意した 200 件の正解データに対し，70%の 140 件を訓練データ，30%の 60 件をテストデータに用いた．提案手法の有効性を確認するため，**表 1** に示す 5 つの手法を用意した．ここで，Sensor はセンサデータのみを提案ネットワーク構造に入力する手法，Video は画像フレームのみ提案ネットワーク構造に入力する手法，SVM はセンサデータと画像フレームを特徴変換せず，すべての画像フレームとセンサデータを同一の SVM 分類器（LIBSVM[*7]）に入力し分類する手法[*8]，none RNN は提案ネットワーク構造において LSTM を Fully-connected layer に置き換え，センサデータは FC-S，画像フレームは Conv1，Conv2，Conv3，FC-P で変換した特徴量をすべての入力フレームから得た後，それらを並列に Fully-connected layer に入力し，得られた特

表 1　比較手法の一覧とその条件
Table 1　List of comparison methods and each method condition.

ラベル	センサ	映像	時系列の考慮
Proposal	✓	✓	✓
Sensor	✓		✓
Video		✓	✓
SVM	✓	✓	
none RNN	✓	✓	

[*4] 最適化のハイパーパラメータは，$\alpha = 0.001$，$\beta_1 = 0.9$，$\beta_2 = 0.999$，$\epsilon = 10^{-8}$ とした．
[*5] http://web.tuat.ac.jp/~smrc/drcenter.html
[*6] 前方映像のカメラの設定位置によっては，ボンネットが大きく映り込むものや，映像が傾いているものもあるが，画像に対する編集操作は線形変換を除いていっさい行っていない．
[*7] https://www.csie.ntu.edu.tw/~cjlin/libsvm/
[*8] カーネルは最も精度の高かった線形カーネルを選択した．

© 2017 Information Processing Society of Japan

徴ベクトル（2 次元）に対して Softmax 関数を通じて分類結果を得る手法である．ニューラルネットの誤差逆伝播回数は，予備実験で交差エントロピー誤差が十分に収束していた 200 回としパラメータを学習した．このときのネットワークを用いてテストデータに対する Accuracy（正解率）[*9]を評価した．いずれの手法も Python で実装し，特にネットワークの構造と学習は Chainer[*10]によって実装した．

実験結果．上記の設定で全手法で評価した正解率を**図 2** に示す．横軸は入力に使用した前からのフレーム数，縦軸が正解率である．全フレームを入力に用いた条件では，Proposal が比較手法の中で最も高い正解率となり，このときの Proposal はすべての比較手法に対して t 検定（有意水準 1%，片側検定）で，正解率に有意な向上が認められた．Sensor は入力フレーム数の増加にともない正解率が大きく向上し，Video は入力フレーム数が少ないときでも比較手法の中では高い正解率となった．このことから，ヒヤリハット検出においてセンサは長期の時系列を考慮する必要があり，画像は短期の時系列でも有効な特徴をとらえられていると考えられる．また，SVM に比べて none RNN が高い正解率であることから，画像とセンサの異なる情報源をニューラルネットで効果的な特徴表現に変換できたと考えられる．前方映像を用いたことで正解率が向上した要因としては，CNN でヒヤリハット発生要因となりやすい画像中の領域（たとえば歩行者や自転車が急に現れた際に映る，画像左端や右端の領域）を学習できたことが考えられる．

入力フレームに関しては，今回は等間隔に 15 フレームをサンプリングし，先頭から t 秒間の画像フレームとセンサデータを入力している．ここで，ドライブレコーダデータが記録されたトリガ時刻を中心に前後 t 秒間を抽出することで，ヒヤリハット発生時刻付近の画像フレームとセンサ情報を集中して学習できることも考えられるため，今後の評価・検証を検討している．

最後に，手法 Video で検出できなかったヒヤリハット例を**図 3** に，手法 Sensor で誤検出した非ヒヤリハット例を**図 4** に示す．ここでは，評価実験で使用した系列データのスピードと，最後から 3 フレームの画像を抽出している．図 3 のケースでは，一時停止すべき道路で停止せず，スピードを緩めて交差点に侵入したところに，左からの通行車を確認し急ブレーキを踏んでいる．夜間で画像全体が暗く，Video はヒヤリハットを引き起こす要因を画像からとらえられなかったと考えられる．一方，センサには大きな変化があったことから，Proposal や Sensor ではこのケースは正しく検出できていた．図 4 のケースでは，センサに大きな変化を確認したことからヒヤリハットと分類したが，実際にはヒヤリハットは起きていない．Proposal や Video では，前方映像から適切に特徴をとらえることができたため，このケースは非ヒヤリハットと分類している．

[*9] テストデータのうち，正しく識別できたデータの割合
[*10] https://chainer.org

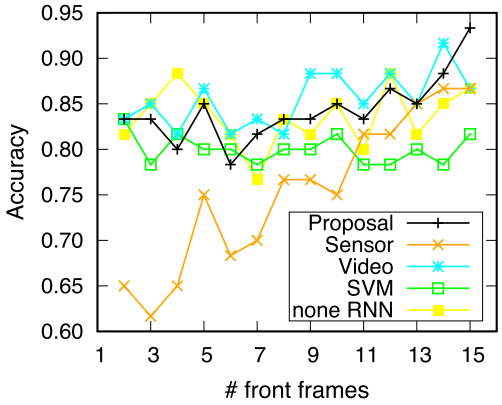
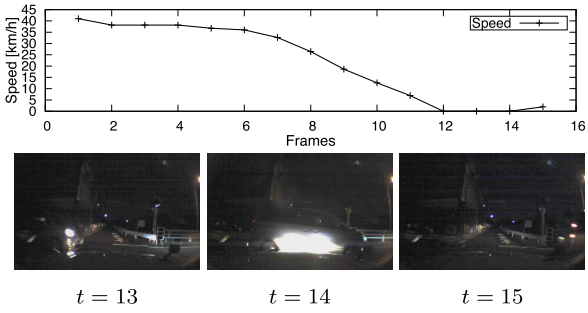
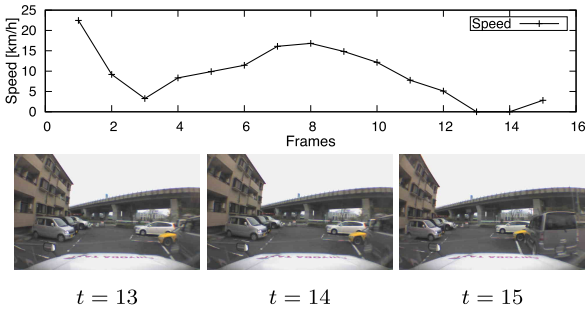


図 2　各手法の入力フレーム数に対する Accuracy
Fig. 2　Accuracy for number of inputting frames in each method.



$t = 13$　　$t = 14$　　$t = 15$

図 3　手法 Video で検出できなかったヒヤリハット例
Fig. 3　An example of near-miss data that Video method could not detect.



$t = 13$　　$t = 14$　　$t = 15$

図 4　手法 Sensor で誤検出した非ヒヤリハット例
Fig. 4　An example of no near-miss data that Sensor method wrongly detected.

5.　結論

本論文では，映像とセンサのマルチモーダルデータを用いた，深層学習に基づくヒヤリハット検出手法を提案した．提案手法では，映像とセンサをそれぞれ適切なニューラルネットワークで特徴変換し，RNN で時間的な遷移をモデル化する．実際のドライブレコーダデータを用いた評価実験の結果，提案手法が比較手法に比べて高い正解率でヒヤ

© 2017 Information Processing Society of Japan

リハット検出できることを明らかにした．

今後の課題は，より少ない訓練データで高い検出性能を目指した手法の実現と，各ヒヤリハットケースをより細かなクラスに自動的に分類することである．

謝辞　本研究は，国立大学法人東京農工大学スマートモビリティ研究拠点の提供する「ヒヤリハットデータベース」を利用した．ここに記して謝意を示す．

参考文献

[1] 菊池理人，日景由華，御室哲志：ドライブレコーダデータからの自動分別の試み，計測自動制御学会東北支部第290回研究集会 (2014).
[2] 森村哲郎，谷澤悠輔，山崎慎也，井手　剛：統計的機械学習を用いたプローブカーデータからのヒヤリハット発生形態の推定，自動車技術会秋季学術講演会 (2011).
[3] 豊田正史，横山大作，伊藤正彦：運転状況を考慮したドライブレコーダデータからの潜在リスク交差点検知手法，*DEIM Forum* (2017).
[4] Lam, H.T.: A Concise Summary of Spatial Anomalies and Its Application in Efficient Real-time Driving Behaviour Monitoring, *SIGSPATIAL*, pp.30:1–30:9 (2016).
[5] Yokoyama, D., Toyoda, M. and Kitsuregawa, M.: Understanding Drivers' Safety by Fusing Large Scale Vehicle Recorder Dataset and Heterogeneous Circumstantial Data, *PAKDD*, pp.734–746 (2017).
[6] Kim, J. and Canny, J.F.: Interpretable Learning for Self-Driving Cars by Visualizing Causal Attention, *CoRR*, Vol.abs/1703.10631 (online), available from ⟨http://arxiv.org/abs/1703.10631⟩ (2017).
[7] Xu, H., Gao, Y., Yu, F. and Darrell, T.: End-to-end Learning of Driving Models from Large-scale Video Datasets, *CVPR*, (online), available from ⟨http://arxiv.org/abs/1612.01079⟩ (2017).
[8] Jain, A., Koppula, H.S., Raghavan, B., Soh, S. and Saxena, A.: Recurrent Neural Networks for Driver Activity Anticipation via Sensory-Fusion Architecture, *ICRA*, pp.3118–3125 (2016).
[9] Chan, F.-H., Chen, Y.-T., Xiang, Y. and Sun, M.: Anticipating Accidents in Dashcam Videos, *ACCV*, pp.136–153 (2017).
[10] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *ICML*, pp.448–456 (2015).
[11] Scherer, D., Müller, A. and Behnke, S.: Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition, *ICANN*, pp.92–101 (2010).

山本 修平　（正会員）

2012年筑波大学情報学群知識情報・図書館学類卒業．2014年同大学大学院図書館情報メディア研究科博士前期課程修了．2016年同大学院同研究科博士後期課程修了．博士（情報学）．同年日本電信電話株式会社に入社．データマイニング，時空間データ分析に関する研究開発に従事．日本データベース学会会員．

遠藤 結城　（正会員）

2017年3月筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻博士後期課程修了．博士（工学）．2012年4月から2016年7月まで日本電信電話株式会社に勤務し，データマイニングの研究開発に従事．2016年8月より筑波大学に勤務し，現職は同大学システム情報系・助教．現在はコンピュータグラフィクス，特に画像処理や機械学習の応用に興味を持つ．電子情報通信学会会員．

戸田 浩之　（正会員）

1997年名古屋大学工学部材料プロセス工学科卒業．1999年同大学大学院工学研究科プロセス工学専攻博士前期課程修了．同年日本電信電話株式会社に入社．以来，情報検索，データマイニングの研究開発に従事．2007年筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻博士後期課程修了．博士（工学）．ACM，電子情報通信学会，日本データベース学会各会員．

（担当編集委員　井手 一郎）

© 2017 Information Processing Society of Japan