

## 車載カメラとGPSを組み合わせた 走行環境適応型歩行者検出器に関する検討

鈴木 大地<sup>†</sup> 吉田 英史<sup>†</sup> 出口 大輔<sup>††</sup> 井手 一郎<sup>†</sup> 村瀬 洋<sup>†</sup>  
町田 貴史<sup>†††</sup> 小島 祥子<sup>†††</sup>

<sup>†</sup> 名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

<sup>††</sup> 名古屋大学 情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

<sup>†††</sup> (株) 豊田中央研究所 〒480-1192 愛知県長久手市横道 41-1

E-mail: †{suzuod,yoshidah}@murase.m.is.nagoya-u.ac.jp, †{ddeguchi,ide,murase}@is.nagoya-u.ac.jp

あらまし 近年、車載カメラ画像から歩行者を検出する研究が注目されている。しかし、時刻や天候、季節、走行場所などの環境要因によって歩行者の見えは大きく変動する。そのため、様々な環境で歩行者を高精度に検出することは難しい。画像中の歩行者の見えを特徴とした従来の検出手法では、事前に大量の学習用歩行者画像および非歩行者画像を手で収集することで、このような変動に対応していた。しかし、あらゆる見えの変動を想定して学習用画像を用意することは難しい。そこで本報告では、車載カメラ映像から学習用画像を自動収集して歩行者検出器を走行環境に適応させる枠組みを提案し、この問題の解決を図る。具体的には、走行シーンを見えに基づいて分類し、GPSから得られる走行場所と関連付けて学習用の非歩行者画像を自動収集する。これにより、走行環境に適応させた歩行者検出器を構築する。同一経路を複数回走行して撮影した映像を用いて、環境適応させた歩行者検出器の精度を検証した。その結果、従来手法と比較して大幅な精度向上が見られ、提案手法の有効性を確認した。

キーワード 歩行者検出, 適応的検出器, ITS

## A study on a pedestrian detector adaptive to driving environment using in-vehicle camera and GPS

Daichi SUZUO<sup>†</sup>, Hidefumi YOSHIDA<sup>†</sup>, Daisuke DEGUCHI<sup>††</sup>, Ichiro IDE<sup>†</sup>, Hiroshi MURASE<sup>†</sup>,  
Takashi MACHIDA<sup>†††</sup>, and Yoshiko KOJIMA<sup>†††</sup>

<sup>†</sup> Graduate School of Information Science, Nagoya University

<sup>††</sup> Information and Communications Headquarters, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

<sup>†††</sup> Toyota Central Research and Development Laboratories, Inc.

Yokomichi 41-1, Nagakute-shi, Aichi, 480-1192 Japan

E-mail: †{suzuod,yoshidah}@murase.m.is.nagoya-u.ac.jp, †{ddeguchi,ide,murase}@is.nagoya-u.ac.jp

**Abstract** Recently, pedestrian detection from in-vehicle camera images is being focused. However, it is difficult to detect pedestrians accurately due to the variety of their appearances caused by various environmental factors. To deal with changes of driving environment, most existing appearance-based methods require the manual preparation of many pedestrian and non-pedestrian images. However, it is difficult to prepare training images considering all possible environmental factors. To tackle this problem, we propose a method to construct a pedestrian detector that is adaptive to driving environment by collecting non-pedestrian images automatically from in-vehicle camera video. The proposed method classifies driving scenes and then collects pedestrian images for each scene accompanied with GPS positions. We confirmed the effectiveness of the method compared to existing methods.

**Key words** Pedestrian detection, Adaptive detector, ITS

## 1. まえがき

平成 22 年の交通事故死者数 4,863 人のうち、歩行者はその 35% (1,714 人) と大きな割合を占めている [1]。そのため、運転者へ歩行者の存在を知らせるなどの安全運転支援技術が注目されている。その中でも高精度な歩行者検出技術は重要であり、車載カメラ画像を対象とした検出手法が広く研究されている。

Dalal らは、Histograms of Oriented Gradient (HOG) 特徴量と Support Vector Machine (SVM) を用いた歩行者検出手法を提案し [2]、現在はこれを発展させた手法が広く提案されている。一方、このような学習に基づく歩行者検出手法では、高精度な検出を実現するために大量の学習用画像が必要となる。そのため、多様な学習用画像を準備できない場合は、検出精度が低下する可能性がある。これは、学習時の走行環境以外の環境下では歩行者検出の精度が低下することを意味する。

この対策として、歩行者を含む映像中から学習用画像を自動収集するというアプローチがある。Wöhler らは、歩行者検出と歩行者追跡の組み合わせにより、検出された歩行者を学習用画像として自動収集する枠組みを提案した [3]。このアプローチでは、誤検出結果の収集を避けるための何らかの手法が必要である。そのため、Wang らは画像情報のみを用いる方法を提案し [4]、服部は時系列情報を用いる方法を提案している [5]。これらは、学習用画像の自動収集により単独の識別器をあらゆる環境 (もしくは特定の環境) で使用可能とすることを目指している。しかし、極めて多様な走行環境の変化に対して一つの識別器のみで対応することは難しい。

これに対して、それぞれの走行環境で学習用画像の自動収集と識別器の学習を行い、検出時は走行環境に合わせて適切な識別器を選択することで、高精度な歩行者検出が可能となる。このアプローチは、歩行者検出器の走行環境適応と呼ばれる。我々のグループでは走行場所の変化を環境要因としてとらえ、走行場所に適応する枠組みを検討し (図 1)、大幅な精度向上が達成できることを確認した [6]。そこで本報告ではこの検討結果を発展させ、走行環境の自動分類および学習用画像の自動収集を行い、歩行者検出器を自動的に走行場所へ適応させる手法を提案する。学習用画像は歩行者画像と非歩行者画像からなるが、走行場所の変化においては非歩行者画像の見えの変動が特に大きくなる。そのため、提案手法では非歩行者画像の自動収集に焦点を当てる。

以降、2 節で歩行者検出の精度に影響を与えると考えられる環境要因について整理し、3 節で歩行者検出器を走行場所に自動的に適応させる枠組みを提案する。4 節で提案手法による歩行者検出器の環境適応の有効性を検証する実験方法について述べ、5 節で実験結果および考察を述べる。最後に、6 節で検討課題を整理し、本報告をむすぶ。

## 2. 走行環境の変化と歩行者検出精度の関係

歩行者検出器の精度に影響を与えると考えられる走行環境の変化として、次の 3 つが挙げられる。

- 走行場所の変化による背景の見えの変化

- 時刻の変化による照明の変化
- 天候・季節の変化による歩行者の服装の変化

このうち時刻や天候・季節による歩行者の見えの変動は限定的であり、対処は比較的容易である。一方、走行場所の変化による変動は極めて大きく、事前に全ての変動を想定した学習用画像の収集は困難である。このため、走行場所に適応可能な検出器を構築することが最も重要である。

## 3. 走行場所へ適応した歩行者検出器の構築

提案手法は、学習段階と検出段階からなる。学習段階は、歩行者検出器を走行場所に適応させる処理である。ここでは、アピアランスに基づく走行シーンの自動分類を行い、分類した走行シーンに対応する地点の映像から学習用画像を自動収集し、識別器を学習する。その後、各識別器と GPS 情報を関連付けた地点-識別器テーブルを構築する。検出段階では、GPS 情報を用いて地点-識別器テーブルから現在位置に対応する識別器を選択し、選択された識別器を用いて車載カメラ画像から歩行者検出を行う。

提案手法の概要を図 2 に示す。以降、各処理について詳しく説明する。

### 3.1 アピアランスに基づく走行シーンの自動分類

提案手法では、過去に同一地点を走行した際に撮影した映像を入力とし、アピアランスに基づいた走行シーン分類を行う。

この処理では、映像中のフレームごとに SURF 局所特徴量 [8] に基づいた Bag of Visual Words (BoVW) [7] を求め、該当フレームの走行シーン特徴量とする。そして、求めた特徴量に対して  $k$ -means クラスタリングを適用し、走行シーンの分類を行う。ここで、 $k$  は走行シーンの種類の数を表す。

なお、本報告では BoVW のコードブックサイズは 100 とし、走行シーン特徴量を 100 次元で表現する。また、走行シーン分類は映像の 30 フレーム毎に行う。

### 3.2 学習用の非歩行者画像の自動収集と識別器の学習

走行環境への適応のために、分類された走行シーンごとに映像から学習用の非歩行者画像を自動収集する。非歩行者画像の収集は、同一地点で異なる時刻に撮影された 2 画像から前景と背景を分離する手法 [9] により行う。

この手法では、まず画像間の射影変換により微小な位置ずれを補正する。続いて、位置合わせされた画像から局所特徴量に基づく画像間の特徴点の対応を求める。そして対応付けられた特徴点同士の距離が近い場合は、その特徴点周辺に変化はないと仮定し、背景領域とする。また、対応がない特徴点や互いに遠方に対応付いた特徴点の周辺画素は何らかの変化が含まれると仮定し、前景領域とする。これにより画素単位で前景領域の尤度マップが得られる。この結果に対してしきい値処理することで前景領域を抽出する。

図 3 は、同一地点で異なる時刻に撮影された 2 枚の画像を入力した場合の前景領域の抽出結果を示している。自動車や歩行者により画像間に差が生じており、それらの領域を正しく抽出できていることがわかる。

学習用画像の収集は、まず検出された前景領域を除外した後、

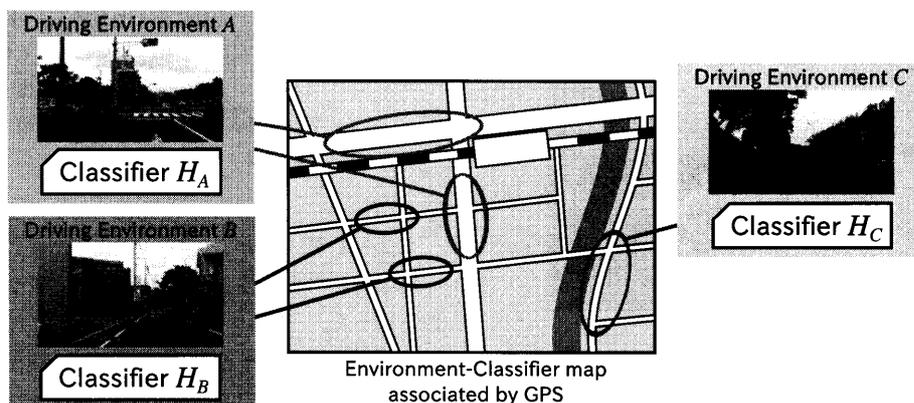


図 1 走行場所へ適応した歩行者検出器の概念. 走行場所の変化による車載カメラ画像の見えの変化に着目し, それぞれに適応した歩行者検出器を構築する.

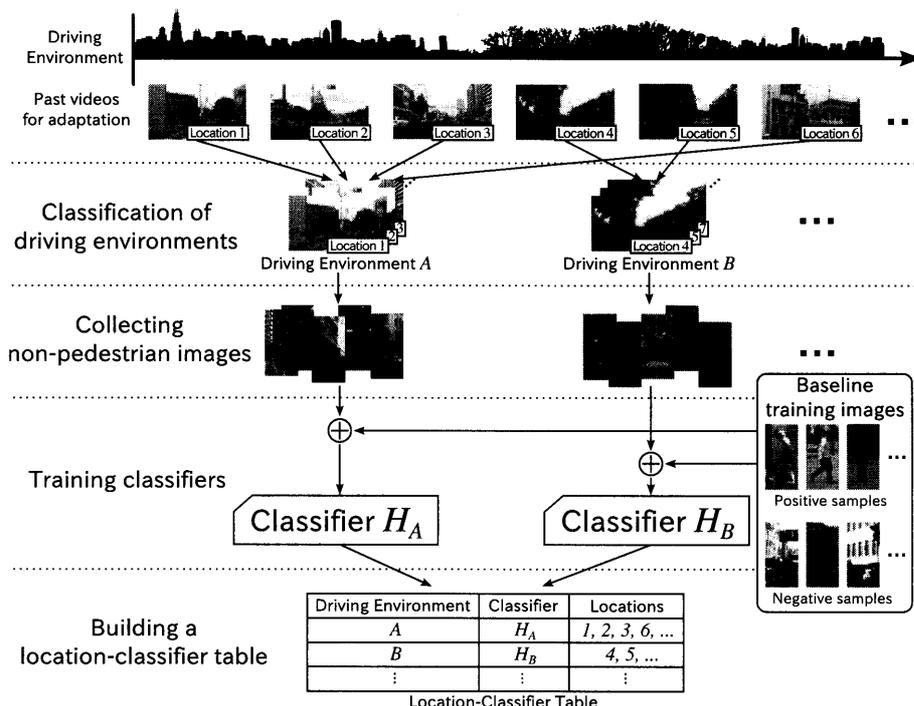


図 2 過去に同地点を走行した際の映像をもとにした歩行者検出器の走行環境適応の枠組み.

ランダムな位置および大きさの矩形を切り出すことで行う. この処理を走行シーンごとに行い, 事前に収集した学習用画像と併せて識別器を学習する. これにより, 走行シーンに適応した歩行者検出器を構築する.

### 3.3 走行場所と検出器の対応表の構築

3.2 節の処理により走行シーンに適応した識別器が得られる. また, 走行シーンの分類結果から, GPS により得られる走行位置と走行シーンの対応が得られる. よってこれらにより, 地点-識別器テーブルを構築する. このテーブルを用いることにより, 自転車位置の GPS 情報から各走行位置において適切な識別器の選択が可能になる. これにより, 検出時には走行シーン認識を行う必要がなくなり, 検出の高速化が可能になる.

### 3.4 検出段階

検出段階では, 自転車位置を表す GPS 情報をキーとして地点-識別器テーブルから識別器を選択する. ここで, 地点-識別器テーブルに記録された地点は離散的であるため,  $k$  近傍探索



図 3 前景領域抽出処理の入力画像と抽出結果.

(本報告では  $k = 5$  とする) により  $k$  個の識別器を選ぶ. そして,  $k$  個の中で最も多く現れたシーンに対応する識別器により



学習用 1 の画像例。



学習用 2 の画像例。



検出用の画像例。

図 4 実験に用いた車載カメラ画像の例。異なる時刻・天候下で撮影された同一地点の画像を示している。

歩行者検出を行う。

#### 4. 走行環境適応の有効性評価実験

本節では、提案手法の有効性を確認するための実験について述べる。まず実験に用いたデータについて説明する。実験では、提案手法の有効性を確認するため、走行環境に適応した歩行者検出器の精度評価を行う。次に、走行シーンの分類数  $k$  と検出精度の関係を調査する。

##### 4.1 実験データ

実験データとして、各フレームに GPS 情報を付加した車載カメラ映像を用いる。本実験では、解像度  $1920 \times 1080$  のプログレッシブモード (24fps) にて撮影したものをを用いた。

映像は、幹線道路や住宅地、郊外などの走行シーンを含む経路を走行したものである。表 1 に示す条件下でこの走行経路を 3 回撮影した。このうち 2 本を走行場所への適応に用い、残る 1 本を検出精度の評価に用いた。映像の例を図 4 に示す。

また環境適応前の歩行者検出器を構築するための学習用画像として、前述した映像とは異なる地域で撮影された映像から事前に人手で収集した歩行者画像および非歩行者画像を用いた。

表 1 実験に用いた映像データの概要

映像名	経路長	映像長	天候	時間帯
検出用	約 7km	約 24 分	曇天	日中
学習用 1		約 35 分	晴天	日没
学習用 2		約 28 分	雨天	日没

#### 4.2 比較手法

提案手法の有効性を検証するため、表 2 に示す各手法との比較を行った。

ベースラインの手法は環境適応前の検出器であり、人手で事前に準備した学習用画像のみで構築した。比較手法 1 は事前に準備した学習用画像の枚数を増やした検出器、比較手法 2 は走行シーンを分類せずに映像全体から学習用画像を収集して構築した検出器である。これらの検出器の構築に用いる学習用画像は、提案手法と同数である。

#### 4.3 実験方法

今回の検出対象となる歩行者は、映像中の高さが 192 画素以上であり、大きなオクルージョンのない歩行者である。検出用映像には、このような歩行者は述べ 372 人存在した。また、歩行者検出器の特徴量には HOG を用い、識別器としてソフトマージン型の線形 SVM を用いた。ここで、liblinear [10] を識別器の構築に利用した。

### 5. 実験結果および考察

#### 5.1 走行シーンへ適応した歩行者検出器の精度評価

精度評価には f-ROC カーブを用いた。これは横軸に画像 1 枚当たりの平均過検出数、縦軸に検出率 (再現率) をとるグラフであり、識別器のパラメータを変化させながら検出処理を行った場合の精度をプロットしたものである。従って、グラフが左上に行くほど、少ない過検出で多くの歩行者を検出できることを意味する。走行環境へ適応した歩行者検出器を実映像に用いた場合の精度評価の結果を図 5 に示す。

図 5 より、提案手法が各比較手法と比べて最も高精度であることがわかる。比較手法 1 と比較手法 2 の結果から、事前に準備する学習用画像を単に増やすより、実際に検出を行う走行環境の映像から学習用画像を自動収集することが検出精度の向上に効果的であると言える。さらに、提案手法と比較手法 2 の結果から、走行環境の変化に応じて学習した検出器を適切に切り替えることが有効であると言える。特に、 $FPPF = 0.16$  における提案手法の検出率は、ベースラインの手法に対して約 23% 向上している。比較手法 1 および 2 との比較においても、提案手法は高精度であることがわかる。

図 6,7 は、ベースラインの手法と提案手法での実際の検出結果の例である。この図から分かるように、提案手法により検出器の環境適応を行うことで過検出を抑制できており、高精度な歩行者検出器を構築することができた。

一方、提案手法では図 8 に示すような自動車の領域を過検出する例が見られた。自動車領域上の画像は非歩行者画像として収集されることが望ましいが、今回提案した収集方法では自動車のような移動物体が除外されてしまう。今後、このような問

表 2 精度比較に用いる歩行者検出器の構築手法一覧

手法	人手で準備した学習用画像数		自動収集した学習用画像数		学習用画像の総数		シーン分類に基づく収集
	歩行者	非歩行者	歩行者	非歩行者	歩行者	非歩行者	
ベースライン	4,104	5,000	-	-	4,104	5,000	-
比較手法 1	4,104	10,000	-	-	4,104	10,000	-
比較手法 2	4,104	5,000	-	5,000	4,104	10,000	×
提案手法	4,104	5,000	-	5,000	4,104	10,000	○

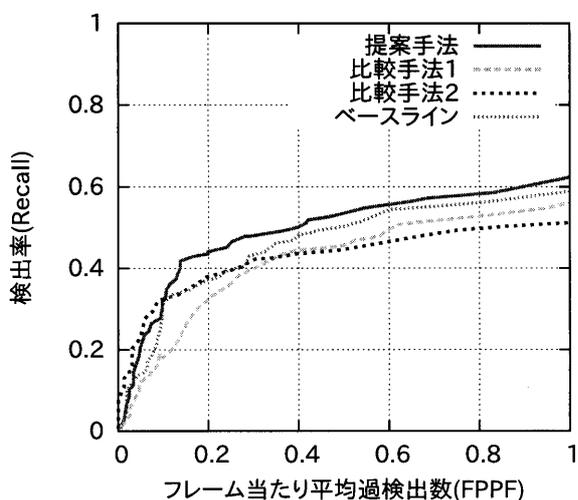


図 5 走行環境へ適応した歩行者検出器の精度評価 (走行シーン分類数  $k$  を 10 とした場合)。

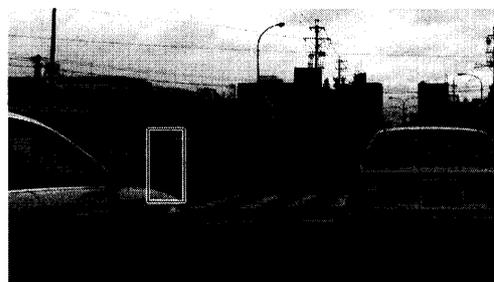
題への対処が必要である。

### 5.2 走行シーン分類数と検出精度の関係の検証

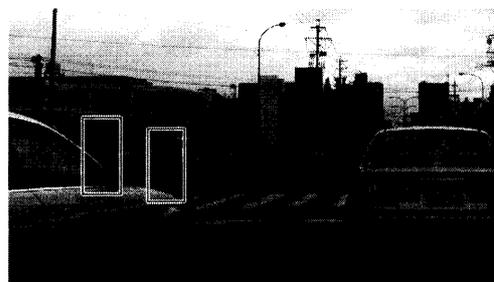
今回の提案手法では、走行シーンの分類数に固定の値を用いた。しかし、走行シーンをいくつに分類するのが適切であるかは実際に使用する環境に依存する。この走行シーン分類数を増やすことで、環境適応した歩行者検出器の検出精度は漸近的に向上すると期待できる。ただし、これは保持すべき識別器の数の増加を意味しており、走行シーン数と計算コストがトレードオフの関係となる。そのため、走行シーン分類数には適切な値を用いることが必要である。図 9 では、走行シーン分類数  $k$  をさまざまに変えながら環境適応した歩行者検出器の精度を示している。

自動収集された学習用非歩行者画像の例を図 10 に示す。歩行者の多い走行シーンにおいても歩行者を含まない画像を収集できており、これにより歩行者検出器を効果的に学習することで精度向上につながったと考えられる。

また図 11 は、走行シーン分類数を  $k = 5$  としたときの分類結果を示したものである。ほぼ同一の地点であっても異なる走行シーンに分類される例があった。これはフレームごとに独立に求めた画像特徴量により分類を行なったためである。例えばカメラの方向の急激な変化などにより画像に大きなブラーが生じる場合、画像の大部分が前方車両に覆われた場合、などが該当すると考えられる。これらに対しては、時系列情報を用いた外れ値除去などの導入により対応可能であると考えられる。



提案手法により走行場所へ適応した検出器。



ベースラインの検出器。

図 6 ベースラインの検出器と提案手法により走行場所へ適応した検出器の検出結果の例 1。



提案手法により走行場所へ適応した検出器。



ベースラインの検出器。

図 7 ベースラインの検出器と提案手法により走行場所へ適応した検出器の検出結果の例 2。

## 6. ま と め

本報告では、車載カメラ画像を用いた歩行者検出器を走行場

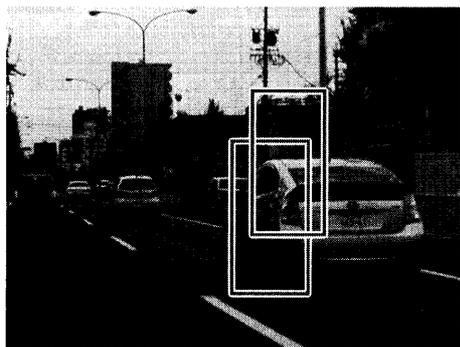


図 8 提案手法で抑制できなかった過検出の例.

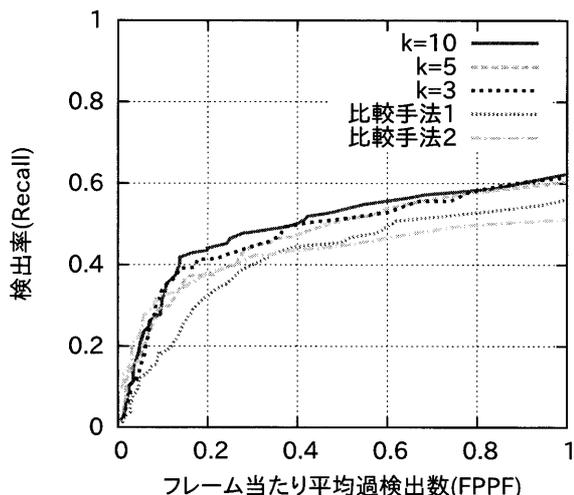


図 9 走行シーン分類数  $k$  と検出精度の関係. なお比較手法 2 は走行シーン分類数を 1 とした提案手法と等価である.



図 10 自動収集された学習用非歩行者画像の例.

所へ適応することで識別器を高精度化する手法を提案した. 具体的には, 走行シーンごとに画像から非歩行者画像を収集して歩行者検出器を学習し, それらを GPS と対応付けて用いる走行環境適応型歩行者検出手法を提案した. 実験の結果, 提案手法の有効性を確認した. 今後の課題として, 時刻や季節等の走行場所以外の環境要因への適応を可能とするための手法の拡張, 走行シーンの分類手法についてのさらなる検討, 学習用の非歩行者画像の自動収集手法の改善, および歩行者画像の自動収集手法などが挙げられる.

謝辞 日頃より御指導および御討論をいただく名古屋大学村瀬研諸氏に深く感謝する. 本研究の一部は JST 戦略的創造研究推進事業 CREST および科学研究費補助金による.

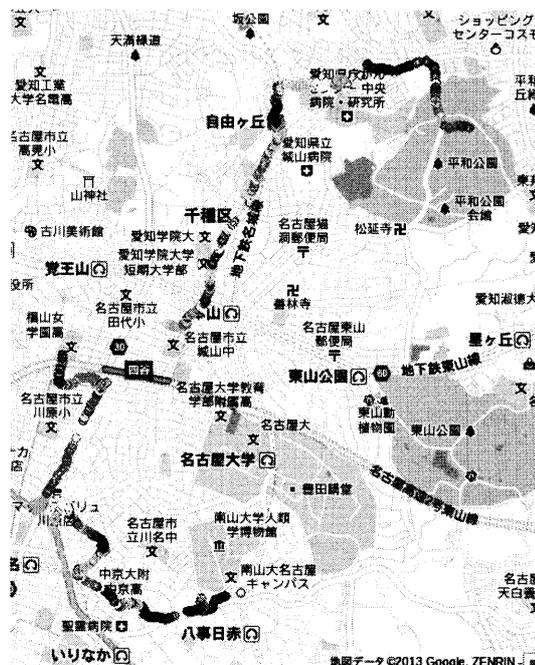


図 11 走行シーン分類結果の色分けして表示した地図.

## 文 献

- [1] 警察庁交通局, “平成 22 年中の交通死亡事故の特徴及び道路交通法違反取締り状況について,” January 2011.
- [2] N. Dalal, and B. Triggs, “Histograms of oriented gradients for human detection,” Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.886–893, June 2005.
- [3] C. Wöhler, “Autonomous in situ training of classification modeules in real-time vision systems and its application to pedestrian recognition,” Pattern Recognition Letters, vol. 23, no. 11, pp.1263–1270, September 2002.
- [4] M. Wang, and X. Wang, “Automatic adaptation of a generic pedestrian detector to a specific traffic scene,” Proceedings of IEEE Computer Society Conference on Computer Vision and Patter Recognition, pp.3401–3408, June 2011.
- [5] 服部 博憲, “固定カメラ映像を対象とした HOG 人物検出器のシーン適応手法,” 京都大学大学院情報学研究所知能情報学専攻修士論文, February 2010.
- [6] 鈴尾 大地, 吉田 英史, 出口 大輔, 井手 一郎, 村瀬 洋, 町田 貴史, 小島 祥子, “車載カメラを用いた走行環境適応型歩行者検出器に関する予備的検討,” 電子情報通信学会 パターン認識・メディア研究会 (PRMU) 技術報告, Vol. 112, No. 197, pp. 99–104, 2012.
- [7] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual Categorization with Bags of Keypoints,” In Workshop on Statistical Learning in Computer Vision, ECCV, May 2004.
- [8] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features,” Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, September 2008.
- [9] 小山 俊隆, 鈴木 崇啓, 別所 史浩, 原田 達也, 國吉 康夫, “移動体の自己位置情報を利用した画像からの異常領域検出,” 情報処理学会第 72 回全国大会, pp. 2.715–1.716, March 2010.
- [10] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, “LIBLINEAR: A Library for Large Linear Classification,” Journal of Machine Learning Research 9, pp. 1871–1874, August 2008.