

人物検出器の高精度化に向けた走行映像群からの ネガティブ学習サンプルの自動抽出に基づく人物検出器の追加学習

本谷 真志† 久徳 遙矢† 出口 大輔† 川西 康友† 井手 一郎† 村瀬 洋†
†名古屋大学

E-mail: hontanim@murase.m.is.nagoya-u.ac.jp

Abstract

近年、運転支援を目的とした車載カメラ映像からの人物検出技術が広く研究されている。一般に、検出器が誤って人物と判定した背景画像（ネガティブ学習サンプル）を追加学習することで、検出器の精度を改善できることが知られている。そこで本発表では、人物検出器の追加学習を行なうために、検出器が誤って人物と判定した背景画像をネガティブ学習サンプルとして自動抽出する手法を提案する。提案手法では、同一地点を長時間観測しても背景領域は大きく変化しないという性質を利用し、同一経路を走行した映像群を比較することでネガティブ学習サンプルを抽出する。そして、抽出したネガティブ学習サンプルを追加学習することにより、検出器の精度向上を図る。提案手法の有効性を確認するため、同一経路を4回走行して撮影した映像を用いて提案手法の評価を行なった。その結果、誤って人物を抽出することなくネガティブ学習サンプルの自動収集が可能であり、また得られたサンプルを用いた追加学習により精度が高い検出器を再構築できることを確認した。

1 はじめに

近年、交通事故発生件数は減少傾向にある。しかし、平成28年には499,232件もの事故が発生しており、依然として解決すべき重要な社会問題の1つとなっている[1]。そのため、さらなる交通事故の削減に向けた有効な対策が求められている。その一例として、自動車が運転者を支援する技術が近年実用化されつつあり、大きな注目を集めている。

上述のようなシステムを実現するための要素技術として、自車の前方に存在する人物や障害物を自動的に検出する技術が必要である。そのために、ミリ波RADARやLIDAR、カメラなど、自動車に様々な外界センサが搭載されるようになった[2]。特に、単眼カメラは安価かつ小型であり、映像記録を目的としたドライブレコーダが広く普及していることから、他のセンサに比べて

普及面の優位性が高い。以上の背景から、本研究では、単眼カメラを用いて撮影された走行映像を対象とした人物検出器の精度改善を目指す。

検出器を構築する際には、人物の見えの多様な変化を識別器に学習させることが重要であり、一般に数千から数万枚という大量の学習サンプルが必要である。しかし、大量の学習サンプルを手で収集するには、多大なコストを要するという問題がある。その対応策として、まずベースラインとして用いる初期人物検出器を構築し、アノテーションデータから初期人物検出器による検出結果の抽出を行ない、検出結果をさらなる学習データとして用いることで、限られたアノテーション付き学習サンプルからより高精度な検出器を再構築する手法が提案されている[3, 4]。

白砂らは、誤検出特性を利用した非標識画像の自動抽出による高精度な標識検出器の構築手法を提案している[3]。走行映像中の標識は、連続したフレームで安定して検出でき、また標識の周りで検出枠の位置をずらしても標識と判定可能であるという2つの特性がある。これらを利用し、初期人物検出器が標識と判定した候補に対して、その前後のフレームにおける検出の有無、候補が存在するフレーム内で他の検出枠と重なる部分の多寡に基づいて、検出候補が誤検出か否か判定する。この判定処理により誤検出と判断されたものを抽出し、それらをネガティブ学習サンプルとして学習することで、より高精度な検出器を再構築している。また服部らは、固定カメラ映像の撮影場所に適応した人物検出器の構築方法を提案している[4]。この手法は、移動体である人物と背景では検出のされ方に差異がある点を利用し、人物が検出されたフレームの前後のフレームにおける、同一物体の検出の有無を調べる。同一物体が検出されたフレーム数が少ない場合は移動体である可能性が高いため正検出と判定し、それ以外は背景または人物の一部に対する誤検出と判定する。そして最終的に正検出、誤検出サンプルを抽出して追加学習することにより、撮影場所に頻出しやすい正検出、誤検出について十分に学習された検出器を再構築する。

以上のように、初期人物検出器で検出したサンプル



図 1 異なる時刻に同じ場所を撮影し、それぞれ人物検出を行なった 3 枚の画像（実線：人物の位置）。

からネガティブ学習サンプルまたはポジティブ学習サンプルを抽出し、追加学習することが高精度な検出器の構築に有用であることが示されている。しかし、白砂らの手法は移動しない物体の特性を利用したものであるため、移動体である人物検出器には適用できない。また、服部らの手法は固定カメラ映像を想定したものであるため、移動カメラにより撮影された走行映像には適用できない。そのため、移動カメラで撮影された走行映像に対して移動体である人物を検出した結果から、自動で追加学習を行なうためのサンプルを抽出する手法の実現が望まれる。

以上の背景から、本発表では走行映像に対する初期人物検出器の検出結果から自動で抽出したネガティブ学習サンプルを追加学習することにより、より精度が高い人物検出器を再構築する手法を提案する。

ここで、実際の走行映像からサンプルを抽出して学習に利用する手法を考える。自動で学習用サンプルを抽出するためには、得られたサンプルがポジティブ学習サンプルかネガティブ学習サンプルかを人手によるアノテーション作業なしに判定する必要がある。この問題に対して提案手法では、人物と背景の検出のされ方に差異が表れる点に注目する。一般に同じ場所を長時間観測した場合、人物は移動するため、常に同じ位置に存在することはないと仮定できる。一方、背景は長時間観測しても同じ位置に一貫して存在する。この性質を利用し、同一経路を複数回走行した映像を用いてネガティブ学習サンプルを抽出する。ポジティブ学習サンプルについては正確なアノテーションを要し、抽出には人手による介入が不可欠と考えられる。そこで、本稿では抽出対象をネガティブ学習サンプルに絞り、学習サンプルの自動抽出手法を提案する。

以降、2 節では、同一経路を複数回走行した映像を用いたネガティブ学習サンプル抽出、およびそれを用いた人物検出器の追加学習手法を提案する。また、3 節では提案手法を用いたネガティブ学習サンプルの抽出実験、4 節では抽出したネガティブ学習サンプルを用いた人物検出器の再構築実験について述べる。5 節ではそれらの実験を考察し、6 節で本発表のまとめと今後の課題

について述べる。

2 提案手法

まず、人物検出器による誤検出には、以下の 2 種類のものが考えられる [4]。

1. 人物と似た特徴量をもつ背景を検出
2. 人物の一部分のみを検出

このうち、誤検出 1 は、同一地点を走行する度に発生する可能性が高く、その走行経路における誤検出率が上昇する要因になり得る。一方、本研究で想定する人物検出器は、自転車前方の人物に対する警告や自動ブレーキのような車載システムに用いるものであるため、誤検出 2 は誤検出 1 と比べて大きな問題にならないと考える。そのため、提案手法では誤検出 1 に着目してネガティブ学習サンプルを抽出する。

同一地点を長時間観測して撮影した 3 枚の画像の例を図 1 に示す。この画像群から分かるように、人物は時間とともに移動するため、撮影時刻が異なれば同じ位置に同一人物は写らない。一方、標識や木などの背景は静止物であるため、撮影時刻が異なっても同じ位置に写る。従って、撮影時刻が異なる画像間で常に同じ位置に写る物体は背景であると仮定できる。提案手法ではこの仮定に基づいて、人物候補の誤検出判定を行なう。

提案手法では、まず同一経路を走行して撮影した映像を複数本用意する。このうち、人物候補を検出する映像を基準映像、その人物候補が誤検出か否か判定するために参照するその他の映像を参照映像と呼ぶ。図 2 に提案手法の処理手順を示す。

提案手法は、基準映像と参照映像の対応付けを行なう対応付け段階と、人物検出候補が誤検出か否かを判定する判定段階（図 2(a)）、ネガティブ学習サンプルの追加学習による検出器の再構築段階（図 2(b)）の 3 段階から構成される。

2.1 対応付け段階

対応付け段階では、基準映像と参照映像を比較するために、各映像を時間的・空間的に対応付ける。

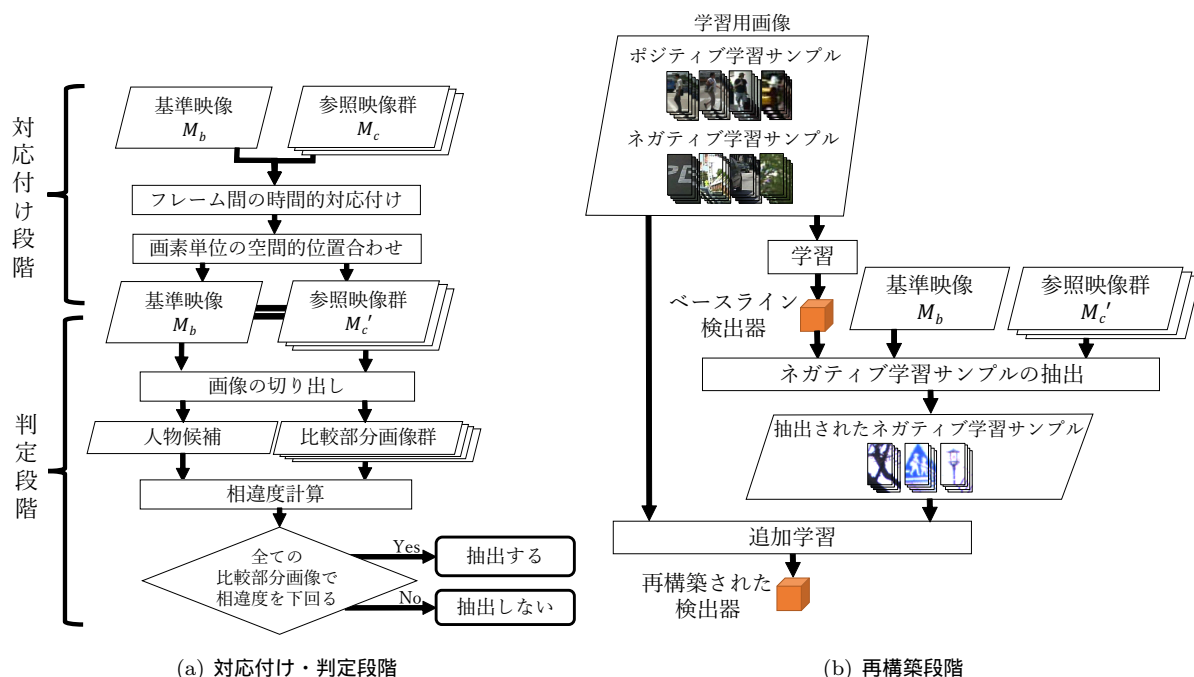


図2 提案手法の構成と各段階における処理手順。

まず、時間的な対応付けを行なう。基準映像と参照映像を比較するためには、基準映像中の各フレームを撮影した地点と同一地点で撮影された参照映像中のフレームを求める必要がある。しかし、同一経路上で撮影された走行映像であっても、走行軌跡や障害物による遮蔽による見えの違いが存在する。そのため、拡大・縮小・回転などを用いた見かけ上の単純な画像照合では精度良くフレームを対応付けることが困難である。

そこで、久徳らが提案したカメラ幾何に基づくフレーム間距離尺度によるフレームの対応付け手法 [5] を用いて、基準映像に対して参照映像から撮影位置が可能な限り近いフレーム同士を対応付ける。

また、対応付けられたフレーム間には視点位置や向きの違いが存在し、これらは画像座標系の位置ずれとして現れる。そのため、対応付けられたフレーム間で特徴量比較を行なう場合、同じ画像座標系上の位置で直接比較を行なうことができない。そこで、DeepFlow [6] で求めた密な Optical Flow を用いて、基準映像中のフレームに適合するように参照映像中のフレームを画素単位で変換する。これにより、参照映像を基準映像に空間的に対応付ける。図 3 に画素単位の空間的対応付けによる画像変換の例を示す。図 3(a) の上部の画像は基準映像中のあるフレームであり、左下の画像はそれに対応する参照映像中のフレームである。図 3(b) 左にこれらのフレームの差分画像を示す。図から、これらのフレームには大きなずれが含まれることがわかる。一方、図 3(a) の右下に基準映像のフレームに参照映像のフレームを画素単位で対応付けた画像を、図 3(b) 右にこの画像と基準映像のフレームとの差分画像を示す。画



基準映像中のフレーム



対応する参照映像中のフレーム

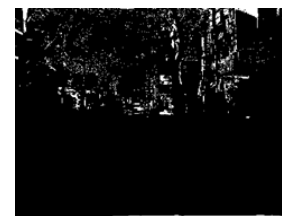


Optical Flow を用いて画素単位で対応付けた参照映像中のフレーム

(a) 基準映像中のフレームに対する参照映像中のフレームの画素単位の対応付けの例



参照映像中のフレーム



Optical Flow を用いて画素単位で対応付けた参照映像中のフレーム

(b) 参照映像中のフレームに対する差分画像

図3 Optical Flow を用いた画素単位の対応付けの例。

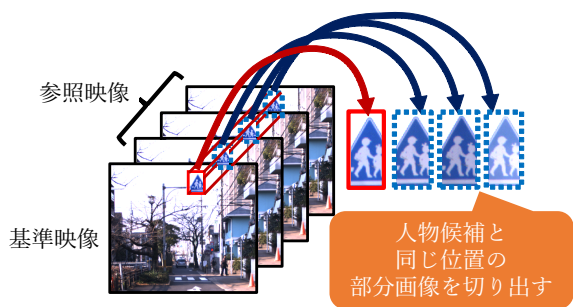


図 4 部分画像切り出し処理の概念図。

素単位での位置合わせにより、基準映像のフレームとそれに対応する参照映像のフレーム間でずれが小さくなったことが分かる。これにより、2つのフレーム間において同じ画像座標系上の位置で正しく比較を行なうことができるようになる。

以上の手順により、基準映像に対して時間的・空間的に対応付いた参照映像を得る。

2.2 判定段階

基準映像中で初期人物検出器が検出した人物候補に対して、基準映像と参照映像群を比較することで誤検出か否か判定し、ネガティブ学習サンプルを抽出する手法について説明する。

まず、既存の人物検出器を初期検出器として人物候補を検出する。この人物検出器は、抽出したネガティブ学習サンプルを追加学習するための初期人物検出器となる。本稿では、この初期人物検出器として予め別のデータセットを用いて学習した Aggregated Channel Features (ACF) に基づく検出器 [8] (以降、ACF 検出器) を用いる。ACF 検出器は、初めに画素単位で特徴量を計算し、局所領域ごとに画素単位の特徴量を統合することで特徴量ピラミッドを高速に求めることが可能な検出器である。

次に、基準映像から初期人物検出器が出力した人物候補をネガティブ学習サンプルであるか否か判定するために、参照映像から人物候補に対応する部分画像を切り出す。図 4 に切り出し処理の概念図を示す。基準映像とそれに時間的・空間的に対応づいた参照映像を重ね合わせ、基準映像上の人物候補の位置から部分画像を切り出す。これにより、入力画像と参照映像群から同一位置を映した部分画像群を得る。そして、基準映像と参照映像から切り出した部分画像それぞれから HOG 特徴量を抽出し、それらの SSD (Sum of Squared Difference) を相違度として出力する。相違度は、部分画像間で映っている物体の見えの差が小さいほど低い値となり、大きいほど高い値となる。ここで出力される相違度は、人物候補 1 つに対して参照映像の数だけ出力される。

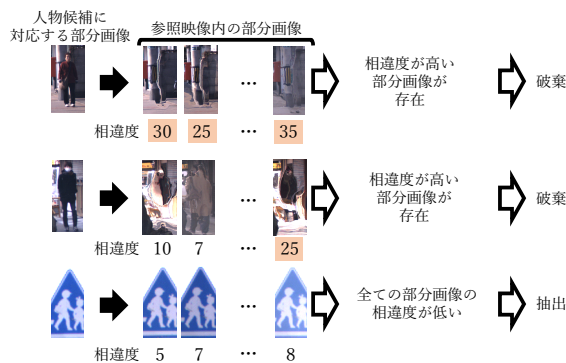


図 5 相違度を用いたネガティブ学習サンプル判定の概念図。

最後に、相違度を用いて人物候補が誤検出か否か判定する。基準映像から切り出した部分画像と参照映像から切り出した部分画像との相違度が低い場合、2つの画像には同じ物体が写っていると考えられるため、検出結果は背景、すなわち誤検出である可能性が高い。ここで、提案手法が人物画像を誤って誤検出と判定した場合、追加学習により性能を低下させてしまうおそれがある。そのため、確実に誤検出であるもののみを抽出したい。そこで、人物候補に対して相違度が全てしきい値を下回る場合のみ、その人物候補は誤検出であると判定し、ネガティブ学習サンプルとして抽出する。図 5 に、相違度を用いたネガティブ学習サンプル判定の概念図を示す。

2.3 追加学習による検出器の再構築

初期人物検出器を構築する際に用いたポジティブ学習サンプル(人物画像)とネガティブ学習サンプル(背景画像)に対し、2.2 項の処理で抽出したネガティブ学習サンプルを加えた学習サンプル群を用いて検出器を再構築し、最終的な人物検出器とする。

3 抽出実験

3.1 ネガティブ学習サンプル抽出実験

提案手法である複数の車載カメラ映像を用いたネガティブ学習サンプル抽出手法の抽出精度を確認するため、複数の車載カメラ映像を用いた評価実験を行なった。

3.1.1 実験用データセット

本実験では、車載カメラを用いて日中に市街地の同一経路を同一方向に 4 回走行して撮影した走行映像(1 映像当たり 2,000 ~ 2,600 フレーム)を用いた。撮影には、Point Gray Research 社のカメラ Grasshopper3 GS3-U3-28S4C-C とスペース社の望遠 8 mm レンズ VP-JHF8M-3MP を用いた。各走行映像には、人手により正解人物枠(1 映像当たり 3,000 ~ 7,000 枠)を付与した。次に、走行映像群に対して初期人物検出器である

表 1 提案手法の抽出精度評価に関する記号 .

		提案手法による判定	
		人物	背景
真値	人物 (TP)	T-TP	F-TP
	背景 (FP)	F-FP	T-FP

ACF 検出器を適用し、人物候補を検出した。本実験で用いる ACF 検出器は、車載カメラ映像に正解人物枠が付与された Caltech Pedestrian Detection Benchmark Dataset [9, 10] (以降、Caltech データセットと呼ぶ) を用いて予め構築した。

そして、ACF 検出器で求めた人物候補全てを本実験のデータセットとした。このうち、人物を正しく検出した人物候補 (TP : True-Positive) 数は 3,925、背景を誤って検出した人物候補 (FP : False-Positive) 数は 80,884 であった。

3.1.2 実験方法

走行映像 4 本のうち、ある 1 本を基準映像、残りの 3 本を参照映像として、基準映像から検出された人物候補に対して提案手法を適用し、人物候補が誤検出か否かが判定した。この処理を 4 本の走行映像各々を基準映像として行ない、誤検出と判定された全ての人物候補をネガティブ学習サンプルとして抽出した。

提案手法では、HOG 特徴量を用いて相違度を計算することで誤検出か否かが判定する。実験に用いた HOG 特徴量のパラメータは以下のように設定した。

- 画像サイズ: 128 × 96 画素
- ブロック: 3 × 3 画素
- セル: 5 × 5 ブロック

3.1.3 評価指標

本研究の目的は、初期人物検出器が誤検出した人物候補 (FP) を抽出することである。よって、提案手法がネガティブ学習サンプル (背景) と判定した人物候補のうち、真に背景であるものを T-FP (True-FP)、実際は人物であるものを F-TP (False-TP) とする。同様に、提案手法がポジティブ学習サンプル (人物) と判定した人物候補のうち、真に人物であるものを T-TP (True-TP)、実際は背景であるものを F-FP (False-FP) と表す。表 1 にこれらの記号をまとめる。

以上の定義をふまえ、本実験で用いる評価指標は、縦軸に T-FP 抽出数、横軸に F-TP 抽出数をとって描画した曲線を用いる。T-FP 抽出数が高いほど、より多くのネガティブ学習サンプルを集めることができる。逆に F-TP 抽出数が多いほど、人物画像をネガティブ学習サンプルとして誤抽出することになるため、追加学習で再構築した検出器の検出精度の低下につながる。この

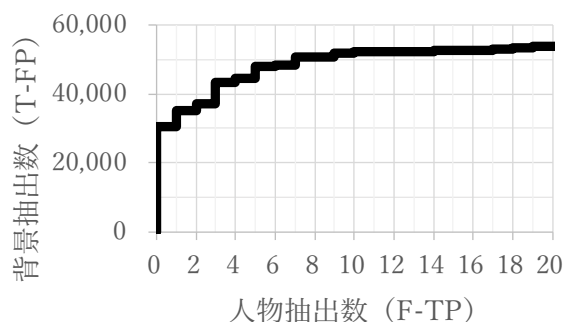


図 6 提案手法のネガティブ学習サンプル抽出性能。

ことから、本曲線は描画領域の左上に位置するほど性能が高いことを表す。

3.1.4 実験結果

図 6 に F-TP 抽出数及び T-FP 抽出数を描画したものを示す。F-TP 抽出数が 0 枚の時、T-FP 抽出数は 23,778 枚 (FP 全体の 37%)、F-TP 抽出数が 10 枚の時、T-FP 検出率は 52,170 枚 (FP 全体の 63%) であった。

3.2 再構築実験

提案手法で抽出したネガティブ学習サンプルの学習面での有用性を確認するため、3.1 項の抽出実験で抽出したネガティブ学習サンプルを用いて人物検出器を再構築する実験を行なった。

3.2.1 実験用データセット

本実験で用いる学習用データは、3.1 項で抽出したネガティブ学習サンプルに加えて、3.1.1 で初期人物検出器の構築に用いた Caltech データセット中の学習用データを用いた。

3.1 項で誤検出と判定した人物候補について、検出枠の大きさが 20 × 50 画素になるように拡張し、特徴量計算のためのマージンを確保するために、検出枠の位置を中心に 96 × 128 画素の大きさに切り出した画像をネガティブ学習サンプルとした。

評価用データセットは、Caltech データセットの評価用として配布されている 4,024 フレームの車載カメラ画像を用いた。この評価用データセットは、学習用データとして用いた Caltech データセットとは別のものである。

3.2.2 実験方法

Caltech データセットのみを用いて学習した ACF 検出器と、提案手法で抽出したネガティブ学習サンプルを加えて追加学習した ACF 検出器をそれぞれ構築した。各手法においてネガティブ学習サンプルは以下のように構成した。なおポジティブ学習サンプルについては、いずれの手法も Caltech データセットから人物の正解

表 2 各手法で検出器構築に用いるネガティブ学習サンプル枚数

	学習枚数	
	無作為に抽出	抽出実験で抽出
提案手法 A	76,222 枚	
	50,000 枚	23,778 枚
提案手法 B	50,000 枚	
	26,222 枚	23,778 枚
比較手法	50,000 枚	
	50,000 枚	0 枚

人物枠が付与された画像 24,498 枚を用いた。

提案手法 A: Caltech データセットから正解人物枠が付与されていない部分画像を無作為に抽出 (50,000 枚) するとともに, 3.1 項で述べた F-TP 抽出数が 0 枚かつ T-FP 抽出率が最大になるようにしきい値を設定して抽出したネガティブ学習サンプル (23,778 枚, 人物画像は含まれていない) を加えた 76,222 枚を利用。

提案手法 B: Caltech データセットから正解人物枠が付与されていない部分画像を無作為に抽出 (26,222 枚) するとともに, 提案手法 A と同様に抽出したネガティブ学習サンプル (23,778 枚) を加えた 50,000 枚を利用。

比較手法: Caltech データセットから正解人物枠が付与されていない部分画像を無作為に抽出した 50,000 枚を利用。

提案手法 A で用いる Caltech データセットの背景画像 50,000 枚は, 比較手法で用いる Caltech データセットの背景画像 50,000 枚と同一のものである。また, 提案手法 B で用いる Caltech データセットの背景画像 26,222 枚は, 提案手法 A および比較手法で用いる Caltech データセットの背景画像 50,000 枚の部分集合である。表 2 に各手法で学習するネガティブ学習サンプルの枚数を示す。

3.2.3 評価方法

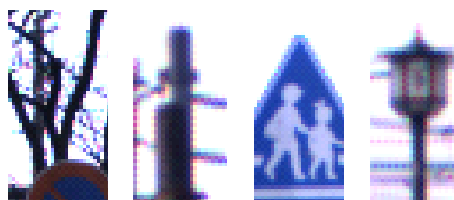
評価用データセットに対して, 各手法で構築した検出器を用いて人物検出を行ない, 検出結果と用意した正解人物枠を比較することで各検出器の性能を評価した。評価指標には, Log-average Miss Rate [7] を用いた。

3.2.4 実験結果

表 3 に Miss Rate を用いて算出した Log-average Miss Rate を示す。比較手法と提案手法 A との比較より, 抽出したネガティブ学習サンプルを加えて構築する提案手法 A の方が, 比較手法よりも Log-average Miss Rate が約 2% 低くなることを確認した。また, 比較手法と提

表 3 各手法で構築した検出器の Log-average Miss Rate .

	学習枚数	Log-average Miss Rate
提案手法 A	76,222 枚	26.18%
提案手法 B	50,000 枚	26.70%
比較手法	50,000 枚	28.19%



(a) 木 (b) 電柱 (c) 標識 (d) 電灯

図 7 抽出されたネガティブ学習サンプルの種類。

案手法 B との比較より, ネガティブ学習サンプルを同じ枚数学習させた場合でも抽出したネガティブ学習サンプルを加えて構築する提案手法 B の方が, 比較手法よりも Log-average Miss Rate が約 1.5% 低くなることを確認した。

4 考察

3.1 項のネガティブ学習サンプル抽出実験では, 真値が人物である人物候補 (F-TP) を全く抽出することなく, 真値が背景である人物候補 (T-FP) 80,884 枚のうち 23,778 枚 (37%) を正しく背景と判定して抽出できた。また, F-TP をごく少量だけ許容する場合, T-FP 全体の約 6 割を抽出することが可能であった。よって, 提案手法によりネガティブ学習サンプルを高精度に抽出できることを確認した。3.2 項の人物検出器の学習実験では, 提案手法で抽出したネガティブ学習サンプルを用いて検出器を学習することにより誤検出率が低下することを確認した。これにより, 提案手法で抽出したネガティブ学習サンプルの学習面での有用性を確認した。

ここで, 3.1 項のネガティブ学習サンプル抽出実験の結果について考察する。

4.1 抽出されたネガティブ学習サンプル

F-TP 検出数が 0 枚のときに抽出されたネガティブ学習サンプルの内容を確認したところ, 代表的なものは木 (約 7 割), 電柱 (約 3 割), 標識 (若干), 電灯 (若干) の 4 種類の背景であった。それぞれの例を図 7 に示す。このことから, 本実験で用いたデータセットの走行環境の偏りにより, 抽出されたネガティブ学習サン



(a) 基準映像中のフレーム



(b) 対応する参照映像中のフレーム



(c) Optical Flow を用いて対応付いた参照映像中のフレーム

図 8 フレームの対応付けに失敗した画像の対応付け例。



(a) 基準映像中のフレーム



(b) 対応する参照映像中のフレーム



(c) Optical Flow を用いて対応付いた参照映像中のフレーム

図 9 影を含む画像の対応付け例。

ブルの種類が少なかった可能性がある。そのため、今後多様な走行環境において、多種類のネガティブ学習サンプルを抽出できることを確認する必要がある。

4.2 画素単位の位置合わせ

図 8~9 に示すように、一部の画像間で Optical Flow を用いた画素単位の空間的位置合わせに失敗した。図 8 は画像全体の見えの違いにより画素単位の位置合わせに失敗した例である。これは、フレーム間の時間的対応付けの失敗により、そもそも異なる撮影地点のフレーム同士を対応付けてしまった例である。そのため、2枚の画像間で同じ物体までの距離が大きく異なり、画像中の位置を補正しきれなかったものと考えられる。このような誤った対応付けの発生を避けるため、フレーム間の時間的対応付けの精度向上が必要である。

また図 9 は、影の影響により対応付けに失敗した例である。Optical Flow を用いた画像変換では、2枚の画像が大きな見えの変化を含む場合に失敗しやすい。このような失敗は F-TP の発生を招くため、画素単位の空間的位置合わせの精度向上が必要である。

5 むすび

本発表では、走行映像を用いた人物検出器の追加学習を目的とし、同一経路を走行して撮影した複数の走行映像からネガティブ学習サンプルを自動抽出する手法を提案した。具体的には、同じ場所を長時間観測した際に背景のみが一貫して同一位置に存在するという性質を利用し、同一経路を複数回走行した映像群から自動でネガティブ学習サンプルを抽出して追加学習することを考えた。

市街地の同一経路を複数回走行した映像 4 本に提案手法を適用してネガティブ学習サンプルの抽出実験を行なった。その結果、人物画像を 1 枚も誤ってネガティブ学習サンプルとして抽出することなく 37% (23,778 枚) の背景画像をネガティブ学習サンプルとして抽出できることを確認した。また、抽出したネガティブ学習サンプルを用いて人物検出器の学習実験を行なった結果、検出器の Log-average Miss Rate を改善できることを確認した。以上から、ネガティブ学習サンプル抽出における提案手法の有効性、そして抽出したネガティブ学習サンプルの学習上の有用性を確認できた。

本研究に関する今後の課題として、フレーム間の時間的対応付け及び画素間の空間的対応付けの精度向上、相違度計算に用いる特徴量検討などが挙げられる。

謝辞 本研究の一部は科学研究費補助金による。

参考文献

- [1] 警視庁, “平成 28 年中の交通事故死者数について”. <http://www.e-stat.go.jp/SG1/estat/Pdfdl.do?sinfid=000031518193>, Jan. 2017. (2017/04/06 参照).
- [2] 二宮芳樹, “車載周辺のセンシング・認識技術”, システム制御情報学会誌, vol.55, no.1, pp.8-13, Jan. 2011.
- [3] 白砂光教, 出口大輔, 井手一郎, 村瀬 洋, “誤検出特性を利用した非標識画像の自動収集による標識検出器の構築”, 電子情報通信学会総合大会講演論文集, Vol. 情報・システム(2), p.174, Mar. 2011.
- [4] 服部博憲, 道上育久, 椋木雅之, 美濃導彦, “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方法”, 電子情報通信学会技術研究報告, HIP2009-261, Mar. 2009.

- [5] 久徳遙矢, 出口大輔, 高橋友和, 目加田慶人, 井手一郎, 村瀬 洋, “自車位置推定のための車載カメラ映像と市街地映像データベースの位置ずれや遮へいに頑健なフレーム対応付け,” 電子情報通信学会論文誌 (D), vol.J95-D, no.11, pp.1973–1982, Nov. 2012 .
- [6] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, “DeepMatching: Hierarchical deformable dense matching,” *Int. J. of Computer Vision*, vol.120, pp.300–323, May 2015.
- [7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp.886–893, June 2005.
- [8] P. Dollár, R. Appel, S. Belongie, and P. Perona, “Fast feature pyramids for object detection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.36, no.8, pp.1532–1545, Aug. 2014.
- [9] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of the art,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, no.4, pp.743–761, Apr. 2012.
- [10] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” *Proc. 2009 IEEE Computer Society conf. on Computer Vision and Pattern Recognition*, pp.304–311, June 2009.