同一経路走行映像群からの ネガティブサンプル自動抽出による人物検出器の高精度化

本谷 真志[1,a)]　久徳 遙矢[1]　出口 大輔[1]　川西 康友[1]　井手 一郎[1]　村瀬 洋[1]

概要

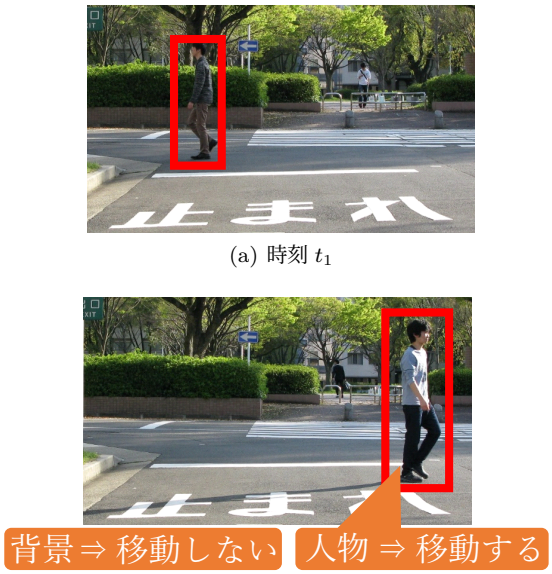
近年，ドライバーに対する高度運転支援システムへの期待から，高精度な人物検出器への需要が高まっている．高精度な人物検出器の構築のためには，学習サンプルを大量に収集することが重要である．本発表では，特に人物に間違えやすいような背景画像をネガティブサンプルとして自動抽出する手法を提案する．同一経路の走行映像群を用いて提案手法の評価を行なった結果，誤って人物を抽出することなくネガティブサンプルの自動収集が可能であり，また得られたサンプルを用いることで高精度な検出器を構築できることを確認した．

1. はじめに

近年，交通事故対策の 1 つとして前方にいる人物を検知して警報を鳴らすなどの、ドライバーに対する自動車の高度運転支援システムへの期待がある．そして，そのための要素技術の 1 つとして，車載カメラを用いた人物検出器への需要が高まっている．

人物検出器を構築する際には，人物や背景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人手で収集するには，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，少数のアノテーション付き学習サンプルと多量のアノテーション無しデータを組み合わせることで高精度な検出器を再構築する手法が提案されている [1], [2]．これらの手法は，まずベースラインとして用いる初期検出器を構築する．そして，アノテーションデータに対して初期検出器を用いて物体検出を行ない，その結果が正検出であるか誤検出であるかを自動で判定する．そして，それぞれをポジティブサンプルおよびネガティブサンプルとして追加学習し，高精度な検出器を構築する．しかし，これらの手法は固定カメラ映像や移動しない検出対象の性質を利用したものである．そのため，移動カメラから得られる走行映像中の移動体である人物を検出対象とした手法の実現が望まれる．



(a) 時刻 t_1

(b) 時刻 t_2

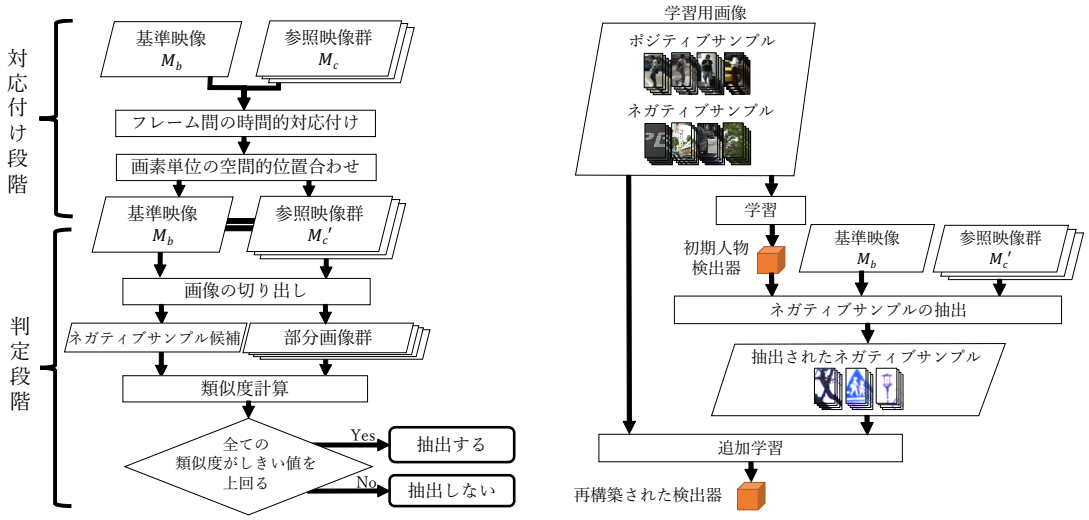
図 1　異なる時刻に同じ場所を撮影し，それぞれ人物検出を行なった 2 枚の画像（実線：人物の位置）．

以上の背景から，我々の研究では走行映像に対する初期人物検出器の検出結果から学習データを自動で抽出することを考える．ここで一般的には，学習データの中でも誤検出しやすいようなネガティブサンプルの学習が，誤検出の少ない検出器の構築に有効である．そこで我々は，これまでにネガティブサンプルの抽出を自動で行なうことにより，高精度な人物検出器を再構築する手法を提案してきた [3]．本発表では，最新検出技術を用いた検出器の再構築実験を実施することで，最新の検出器にも自動抽出したネガティブサンプルが有効であることを示す．

2. 提案手法

人物検出器による検出結果は，人物に対する正検出と，背景に対する誤検出の 2 つに分類できる．このうち，誤検出結果のみをネガティブサンプルとして抽出したい．そのため，我々の提案手法では人物検出器による検出結果をネガティブサンプル候補とし，そのうち誤検出だと判定され

[1] 名古屋大学
[a)] hontanim@murase.is.i.nagoya-u.ac.jp



(a) 対応付け・判定段階　　(b) 再構築段階

図 2　提案手法の構成と各段階における処理手順.

たネガティブサンプル候補をネガティブサンプルとして抽出する.

ここで，同一地点を長時間観測して撮影した 2 枚の画像の例を図 1 に示す．これらの画像から分かるように，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映らない．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映る．したがって，撮影時刻が異なる画像間で常に同じ位置に映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，ネガティブサンプル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，初期人物検出器によりネガティブサンプル候補を抽出する映像を基準映像，そのネガティブサンプル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，ネガティブサンプル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．以降，それぞれについて詳細に説明する.

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走行映像を時間的・空間的に対応付ける [3].

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．そこで，久徳らが提案した走行軌跡や障害物の有無に伴う見えの違いに頑健な，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 [4] 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するため，同じ画像座標系上の位置で直接特徴量を比較することができない．そこで，DeepFlow [5] 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，参照映像と基準映像を空間的に対応付ける.

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

判定段階では，基準映像から初期人物検出器が検出したネガティブサンプル候補に対して，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，ネガティブサンプルとして抽出する.

まず，既存の人物検出器を初期検出器として人物を検出し，この検出結果をネガティブサンプル候補とする．次に，基準映像から初期人物検出器が出力したネガティブサンプル候補が誤検出であるか否か判定するために，参照映像からネガティブサンプル候補に対応する部分画像を切り出す．基準映像と時間的・空間的に対応づいた参照映像を重ね合わせ，基準映像上のネガティブサンプル候補と同じ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 [3].

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との類似度を計算する．本発表では，類似度の計算に Ahmed らが提唱した CNN モデル [8] を用いる．この CNN モデルは 2 枚の画像を入力とし，画像間の類似度を

表す尤度を出力する．提案手法では，ネガティブサンプル候補 1 つにつき参照映像の数だけ類似度が出力される．

最後に，類似度を用いてネガティブサンプル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類似度が高い場合，2 つの画像には同じ物体が映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，参照映像群中の同一部分画像全てに対して類似度がしきい値以上の場合のみ，そのネガティブサンプル候補は誤検出であると判定し，ネガティブサンプルとして抽出する．

3. 抽出実験

最新の人物検出技術に対する提案するネガティブサンプル抽出手法による検出器構築における有効性を確認するために，複数の車載カメラ映像を用いたネガティブサンプルの抽出実験および抽出したネガティブサンプルを用いた検出器の再構築実験を実施した．以降，それぞれについて詳細に述べる．

3.1 ネガティブサンプル抽出実験

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．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，初期人物検出器として，Aggregated Channel Features (ACF) に基づく検出器 [6]（以降，ACF 検出器）を構築した．構築には，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）を用いた．

そして，ACF 検出器を用いて人物検出を行なうことでネガティブサンプル候補を出力し，このネガティブサンプル候補全てを本実験のデータセットとした．

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ネガティブサンプル候補に対して提案手法を適用し，ネガティブサンプル候補が誤検出か否か判定した．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ネガティブサンプル候補をネガティブサンプルとして抽出した．

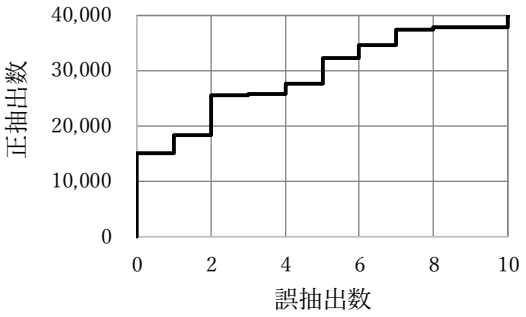


図 3 提案手法のネガティブサンプル抽出性能．

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ネガティブサンプル候補を抽出することである．そのため，以下の 2 つを評価指標とする．

- **正抽出数:** 実際に背景である画像をネガティブサンプルとして抽出した数
- **誤抽出数:** 実際には人物である画像を誤ってネガティブサンプルとして抽出した数

3.1.4 実験結果および考察

ネガティブサンプル候補を誤検出だと判定するための類似度のしきい値を変化させた時の正抽出数および誤抽出数を描画したグラフを図 3 に示す．人物画像を全く抽出することなく，真値が背景であるネガティブサンプル候補 80,884 枚のうち 15,031 枚（18%）を正しく背景と判定して抽出できた．

3.2 再構築実験

提案手法で抽出したネガティブサンプルの有用性を確認するため，3.1 の抽出実験で抽出したネガティブ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．なお，ここで構築する検出器は最新の検出技術である Faster R-CNN に基づく検出器 [6]（以降，Faster R-CNN 検出器）である．

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として，Caltech データセットと The KITTI Vision Benchmark Suite [11]（以降，KITTI データセットと呼ぶ）の学習用データ，そして 3.1 で抽出したネガティブサンプルを用いた．Caltech データセットは，3.1.1 で初期人物検出器の構築に用いたものと同一である．また，3.1 で抽出したネガティブサンプルは，誤抽出数が 0 枚かつ正抽出数が最大になったしきい値を用いて抽出されたものである．

評価用データセットは，3.1.1 で用いたデータセットと同じ車載カメラを用いて日中に市街地を走行して撮影した走行映像（15,445 フレーム）を用いた．この評価用データセットは 3.1.1 で用いたデータセットとは別のものである．

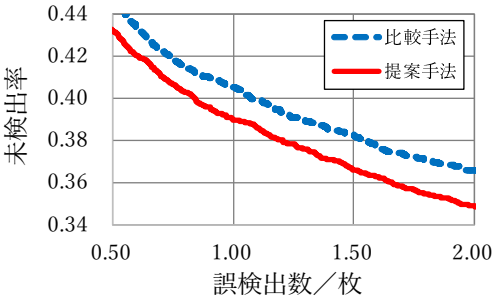


図 4 各手法で構築した検出器の DET 曲線.

3.2.2 実験方法

以下のデータセットを用いて，Faster R-CNN 検出器を構築した.

提案手法: Caltech データセットと KITTI データセットの学習用データに加え，3.1 で抽出したネガティブサンプルを利用.

比較手法: Caltech データセットと KITTI データセットの学習用データを利用.

そして，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，1 枚あたりの誤検出数に対する未検出率を示す DET（Detection error tradeoff）曲線を用いた.

3.2.3 実験結果および考察

図 4 に提案手法と比較手法の DET 曲線を示す．提案手法を用いて学習した検出器の未検出率が，既存データセットのみで学習した検出器の未検出率を下回った．これにより，提案手法で抽出したネガティブサンプルを用いた検出器の再構築が最新の検出技術においても有用であることを確認した.

4. 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サンプルとして抽出することなく 15,031 枚（全体の 18%）の背景画像をネガティブ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サンプルを用いて最新技術による人物検出器の学習実験を行なった結果，検出器の精度を改善できることを確認した．以上から，抽出ネガティブサンプルを用いた検出器の再構築が最新の検出技術においても有用であることを確認した.

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および画素間の空間的対応付けの精度向上，類似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] Y. Yuan, Z. Xiong, and Q. Wang, “An incremental framework for video-based traffic sign detection, tracking, and recognition.” IEEE Trans. on Intelligent Transportation Systems, vol. 18, no. 7, pp. 1918–1929, July 2017.

[2] I. Mitsugami, H. Hattori, M. Minoh, “Improving Human Detection by Long-Term Observation” IEEE Trans. on The 4th Asian Conference on Pattern Recognition, pp.662–666, Nov. 2013.

[3] 本谷 真志, 久徳 遙矢, 出口 大輔, 川西 康友, 井手 一郎, 村瀬 洋, “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,” 第 23 回画像センシングシンポジウム, IS3-30, May 2017.

[4] 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.

[5] 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.

[6] 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.

[7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.39, no.6, pp.1137–1149, June 2017.

[8] E. Ahmed, M. Jones and T. K. Marks, “An improved deep learning architecture for person re-identification.” IEEE Conf. on Computer Vision and Pattern Recognition, , pp.3908–3916, June 2015.

[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,” IEEE Conf. on Computer Vision and Pattern Recognition, pp.304–311, June 2009.

[11] A.Geiger, P.Lenz, and R.Urtasun, “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite,” IEEE Conf. on Computer Vision and Pattern Recognition, pp. 3354–3361, June 2012.