

# 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抽出手法

本谷 真志 (指導教員: 村瀬 洋, 井手 一郎, 出口 大輔, 川西 康友)  
名古屋大学 工学部

## 1. はじめに

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. このことから, 検出器が判定を誤るネガティブ学習サンプルを自動で大量に抽出することができれば, 検出性能の改善が期待される. そこで本発表では, 同じ経路を走行した複数の車載カメラ映像を用いることで, 人物検出器の学習に必要なネガティブ学習サンプルを自動抽出する手法を提案する.

## 2. 複数の同一経路走行映像を用いたネガティブ学習サンプルの抽出手法

一般に人物は移動するため, 図1に示すように, 同じ地点の画像であっても撮影時刻が異なれば全く同じ場所に人物が写る可能性は低い. 一方, 木や標識などの静止物は, 撮影時刻が異なっても同じ場所に写る. よって, 撮影時刻が異なる画像で常に同じ場所に写る対象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き, 同一経路を異なる時刻に走行して撮影した複数の映像を用いてネガティブ学習サンプルを抽出する. 具体的には, ベースライン検出器を用いて人物候補を検出する入力映像と, 同一経路を異なる時刻に撮影した参照映像を比較し, 参照映像でも同じ場所に写るような人物候補をネガティブ学習サンプルとして抽出する.

入力映像と参照映像には, 走行軌跡や障害物による遮蔽から見えの違いが含まれる. 提案手法はまず, フレーム間距離尺度によるフレームの対応付け手法 [1] を用いて, 入力映像と参照映像で同じ場所において撮影されたフレームを時間的に対応付ける. 更に, 入力映像と参照映像では同一の物体についても画像座標がずれて観測されるため, DeepFlow [2] で求めた Optical Flow により, 参照映像内のフレームを入力映像内のフレームに適合するように, 空間的に変換する. 以上の手順により, 入力映像に対して時間的・空間的に対応づいた参照映像を生成する. 次に, 入力映像からベースライン検出器を用いて人物候補を検出し, 対応づいた参照映像から人物候補と同じ位置の部分画像を切り出す. 全ての参照映像から切り出した部分画像において人物候補間の相違度がしきい値を下回る場合, その人物候補は常に同じ場所に写る背景, すなわちネガティブ学習サンプルであるとして抽出する.

## 3. 実験

提案手法の性能を確認するため, 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, ネガティブ学習サンプルを抽出した. まず1本を入力映像, 残り

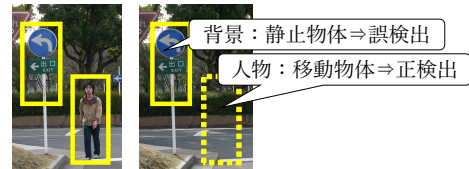


図1 ネガティブ学習サンプルの抽出コンセプト

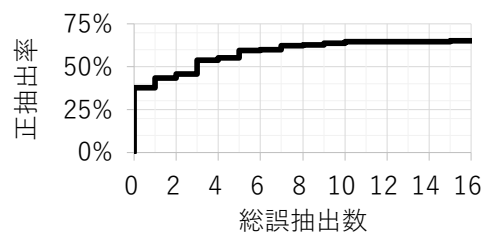


図2 総誤抽出数-正抽出率曲線

3本を参照映像として, 提案手法によりネガティブ学習サンプルの抽出を行なった続いて, 4本それぞれに対して同様に処理し, 全ての映像からネガティブ学習サンプルを抽出した. 評価指標には以下の2つを用いた.

- 正抽出率: 真値が背景である人物候補 (False-Positive) のうち, 提案手法が抽出できた割合
- 誤抽出数: 真値が人物である人物候補 (True-Positive) のうち, 提案手法が抽出した数

図2に, 4本の走行映像に対する総誤抽出数-正抽出率曲線を示す. 誤抽出がない条件において, 検出器が歩行者と判定した背景画像のうち, 37%を正しく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.

## 4. むすび

本発表では, 同一経路を走行した複数の映像を用いてネガティブ学習サンプルを抽出する手法について検討した. 実験により, ベースライン検出器の検出結果から高精度にネガティブ学習サンプルを抽出できることを確認した. 今後の課題として, 時間的・空間的対応付けの精度向上を狙うとともに, 相違度計算の更なる改善を行なう予定である.

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] 久徳遙矢ら, “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.
- [2] J. Revaud et al., “DeepMatching: Hierarchical deformable dense matching,” Int. J. of Computer Vision, vol.120, pp.300-323, May 2015.