

HOG 特徴量を用いた時系列画像からの人物検出

荻 荘 裕太^{†1} 阿部 亨^{†1,†2} 菅沼 拓夫^{†1,†2}^{†1} 東北大学大学院情報科学研究科 ^{†2} 東北大学サイバーサイエンスセンター

1 はじめに

映像中での人物領域の検出は、監視、制御、解析など幅広い分野のアプリケーションで必須の処理となっている。人物領域検出のために様々なアプローチが提案されているが、その中でも、Histograms of Oriented Gradients (HOG) 特徴量 [1] を用いたアプローチが多くの手法で用いられている。フレーム単位で HOG 特徴量を求め人物領域を検出する従来手法では、各フレームでの抽出結果が映像の状態により大きく変化する。この問題に対処するため、本稿では、対象フレームの HOG 特徴量に加え、前後のフレームの情報も用いることにより、安定した人物領域検出の実現を図る手法について述べる。

2 関連研究

HOG 特徴量は、図 1 に示すように、画像中の局所領域における輝度勾配方向の分布をヒストグラムで表現した特徴量であり、照明の変動や多少の位置ずれ、回転に対して頑健なため、人物領域検出にも広く利用されている [2]。HOG 特徴量を用いた人物領域検出手法では、図 2 に示すように、ウィンドウのサイズを変化させながら画像内を走査し、各々で求めた HOG 特徴量が人物領域に対応するか否かの判定を、サンプル（人物領域および他の領域の HOG 特徴量）を SVM で学習して作成した検出器を用いている。

時系列画像中の人物領域を検出する場合は、HOG 特徴量により人物領域の判定をフレーム毎に行う手法の他に、フレーム間の情報や時間方向の情報も併用することで人物領域検出精度の向上を図る様々な手法が提案されている。例えば、Dalal らは、連続した 2 フレームの画像から得られるオプティカルフローの勾配の分布をヒストグラムで表現した Motion Boundary Histograms (MBH) 特徴量を用い、フレーム間の情報を運動情報として導入した手法を提案している [3]。また、Klaser らは、HOG 特徴量を時間方向へ拡張した 3DHOG 特徴量を人物領域検出へ適用する手法を提案している [4]。しかしながら、これらの手法では、SVM の学習を行う際に、多様な動作を行う人物のサンプルが必要となり、適切な検出器を作成するための条件が厳しくなると考えられる。

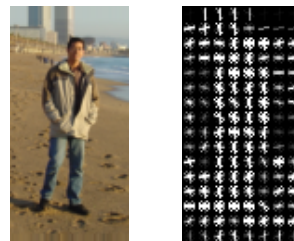
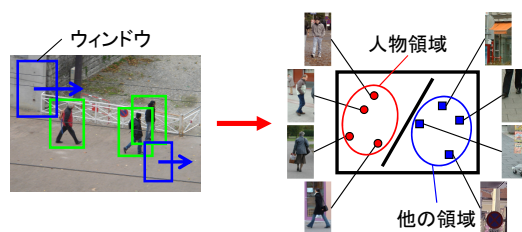


図 1 入力画像（左）から求めた HOG 特徴量（右）



(a) 画像内の走査

(b) HOG特徴量による検出器

図 2 HOG 特徴量を用いた人物領域検出

3 提案手法

提案手法では、時系列画像から人物領域を検出する際、HOG 特徴量による人物領域の検出をフレーム毎に行い、その結果をフレーム間で統合することで、より安定した人物領域の検出を図る。提案手法の概要を図 3 に示す。

提案手法では、まず、各フレームに対し、ウィンドウのサイズを変化させながら画像内を走査し、各箇所でも求めた HOG 特徴量が人物領域に対応するか否かを検出器により判定し、人物領域に対応すると判定された箇所を人物領域候補として記録する。注目フレーム t で人物領域の検出を行う際は、前後の N フレーム $t-N, \dots, t, t+N$ で記録されている人物領域候補を重ね合せ、重なり大きい箇所をフレーム t の人物領域として出力する。具体的には、まず、フレーム中の各画素 (x, y) での重なりを記録する配列 $O_t(x, y)$ を用意し、初期値を $O_t(x, y) = 0$ に設定する。注目フレーム t から n 離れたフレーム $t+n$ ($n = -N, -N+1, \dots, 0, \dots, N-1, N$) において画素 (x, y) が人物領域候補の矩形内にあるならば $O_t(x, y)$ に $\left(1 - \frac{2}{N+2}\right)^{|n|}$ を加える。このように、注目フレーム t から離れたフレームほど影響が小さくなるように重み付け評価した人物領域候補の重なり $O_t(x, y)$ に対し閾値処理を行い、閾値 T_t 以上の $O_t(x, y)$ を持つ画素 (x, y) を注目フレーム t での人物領域として出力する。なお、閾値 T_t は、注

Human Detection in Image Sequence Using HOG Features
Yuta OGISHO^{†1}, Toru ABE^{†1,†2}, and Takuo SUGANUMA^{†1,†2}
^{†1}Graduate School of Information Sciences, Tohoku University
^{†2}Cyberscience Center, Tohoku University

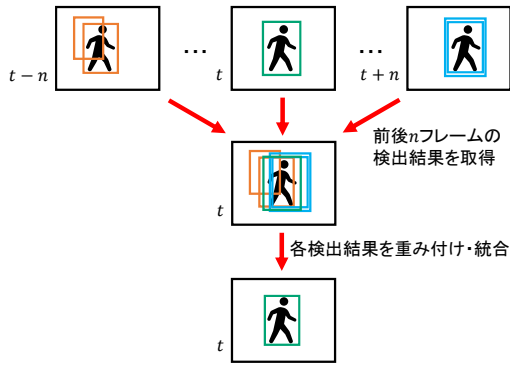


図3 提案手法概要

目フレーム t での重なるの最大値 $\max_{x,y} O_t(x,y)$ に基づき次式で決定する。

$$T_t = \max_{x,y} O_t(x,y) \times C$$

ここで、 C は、事前に与える定数であり、全フレームを通じ同じ値である。

4 実験

4.1 実験内容

前後 N フレームの結果を統合し人物領域の検出を行う提案手法の有効性を検証するために、フレーム単位で HOG 特徴量を求め人物領域を検出する単純手法との比較実験を行った。実験には、KITTI Vision Benchmark Suite [5] 中の時系列画像 (1392 × 512 画素, 74 フレーム, 全フレームが 1 箇所の人物領域を含む) を用いた。画像中の人物領域に対し、正解となる領域を矩形として目視で設定し、出力領域のうち設定した矩形の内側に対応する画素を正解、それ以外の画素を不正解として判定した。

4.2 実験結果

単純手法と提案手法 ($N = 4, C = 0.6$) による検出結果の例を図4に示す。また、単純手法と N を変化させた提案手法による抽出結果の適合率、再現率、F 値を表1に示す。なお、単純手法では、人物領域として検出された矩形同士の重なるの割合が一定以上の場合にそれらを統合している。提案手法を用いることで、単純手法と比較して誤検出が抑制できることが確認できた。一方、 n の値により精度に差が生じているが、 n が小さい場合は統合に利用できる領域の数が少ないため、誤検出された領域も人物領域として統合されてしまうことが原因として考えられる。また、 n が大きい場合は統合に利用する検出結果の位置にばらつきがあることが検出精度に影響を与えていると考えられる。

5 おわりに

本稿では、フレーム単位で HOG 特徴量を求め人物領域を検出する従来手法に対し、対象フレームの HOG 特徴量に加え、前後のフレームの人物領域の検出結果を用いることにより、安定した人物領域検出の実現を図る手法について提案した。今後は人

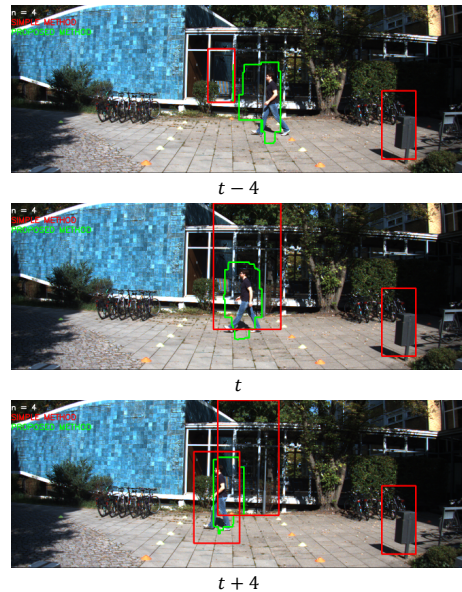


図4 検出結果の例 (赤：単純手法, 緑：提案手法)

表1 人物領域検出精度

	Precision	Recall	F-measure
単純手法	0.201	0.666	0.308
$N = 1$	0.469	0.537	0.501
$N = 2$	0.518	0.549	0.533
$N = 3$	0.556	0.557	0.556
$N = 4$	0.584	0.567	0.575
$N = 5$	0.568	0.580	0.574
$N = 6$	0.536	0.579	0.557
$N = 7$	0.475	0.571	0.519
$N = 8$	0.447	0.559	0.497
$N = 9$	0.412	0.537	0.466
$N = 10$	0.391	0.518	0.446

物領域検出の精度向上を図るため、人物領域の統合方法についてさらに検討を進める予定である。

参考文献

- [1] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *IEEE Conf. Comput. Vision Pattern Recognit.*, Vol. 1, pp. 886–893 (2005).
- [2] 山内悠嗣, 山下隆義, 藤吉弘巨: 画像からの統計的学習手法に基づく人検出, *電子情報通信学会論文誌*, Vol. 96, No. 9, pp. 2017–2040 (2013).
- [3] Dalal, N., Triggs, B. and Schmid, C.: Human detection using oriented histograms of flow and appearance, *European Conf. Comput. Vision*, pp. 428–441 (2006).
- [4] Klaser, A., Marszałek, M. and Schmid, C.: A Spatio-Temporal Descriptor Based on 3D-Gradients, *British Machine Vision Conf.*, pp. 995–1004 (2008).
- [5] Geiger, A., Lenz, P., Stiller, C. and Urtasun, R.: Vision meets Robotics: The KITTI Dataset, *Int. J. Rob. Res.*, Vol. 23, No. 11, pp. 1231–1237 (2013).