

HOG と SVM による上半身検出器の特徴の抽出位置に関する考察

A Study on the Position of the Feature Extraction
for Upper Body Detector with HOG and SVM藤井 龍也† 中島 克人† 野口 祥宏‡ 西田 健次‡
Tatsuya Fujii Katsuto Nakajima Yoshihiro Noguchi Kenji Nishida

1. はじめに

俯瞰カメラ等から人物検出を行うことで、人物の位置情報の取得や人数カウント等を行うことが可能になり、セキュリティや人流解析の分野に応用することができる。

Dalalらは、局所領域内の勾配方向ごとの強度に着目した特徴量であるHOGを使用して、人物検出を行った[1]。

Kobayashiらは、HOGとSVMを用いた人物検出において、DOGによって選択された各位置のHOGを使ってPCAのスコアを計算し、SFSやSBSを用いて特徴の選択を行い、特徴量の次元数を減らしつつ識別器の精度を向上させた[2]。

我々は、人通りが多く下半身のオクルージョンが多い状況で人物検出を行う場合や教室等で着座している人の人数カウントには、人影の全身ではなく上半身のみを検出するのが望ましいと考え、人物検出のために用いられる事が多い[3]HOGと、識別性能に優れた学習モデルとして知られているSVMを組み合わせた上半身検出器を作成した。

検出器の実時間性を重視する応用では、識別率だけでなく識別器の処理速度も重要である。そこで我々は、HOGにおける特徴選択が識別率の向上だけでなく、特徴の抽出速度やSVMにおける識別速度の向上にも効果があるのではないかと考え、特徴量の抽出位置の限定による処理速度の向上と検出精度への影響に関する評価を行った。

2. 提案手法

2.1 概要

前述のように、我々の上半身検出器では特徴量としてHOG、識別手法としてSVMを使用する。そして、学習画像中でHOGが大きく現れる領域だけを残し、他の領域を識別に用いない場合の識別速度と識別率を比較する。

2.2 HOG(Histograms of Oriented Gradients)

2.2.1 領域の定義

本研究では64×64画素の画像に対して、HOGを計算する。画像内で16×16画素の領域をBlockと定義する。また、Blockの4隅から12画素の幅を持つ正方領域をCellと定義する(図1)。

2.2.2 勾配の計算

64×64画素の画像の各画素に対して、垂直・水平方向の輝度値の差を計算する。計算した垂直・水平方向の輝度値の差から、勾配方向と勾配強度を計算する。勾配方向は

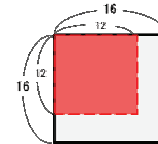


図1. Block領域(16×16)と左上端のCell領域(12×12)

-90度から90度までの範囲を0度から180度の範囲に修正し、それを20度ごとに区切った9方向で定義する。即ち、180度反対方向の勾配は区別せず、同じ方向として扱う。

2.2.3 Block領域での計算

各画素の勾配方向をbinとし、周辺部分の影響を少なくするためにガウス重みをかけた勾配強度を度数とした9binのヒストグラムを作成する。ヒストグラムはBlock領域内の4つのCell領域ごとに作成する。Cellごとに作成された4つのヒストグラムを結合した36binのヒストグラムに対して、各binの勾配強度の2乗和の平方根でヒストグラムを正規化する。

2.2.4 ヒストグラムの結合

64×64画素の画像内でBlockを右上から左下まで8画素ずつ動かし、Block領域の各位置に対して上記の36binのヒストグラムを得る。作成されるヒストグラムの個数は、 $(64/8 - 1) \times (64/8 - 1) = 7 \times 7$ で49個となる。これら49個のヒストグラムを結合したものがHOGとなる。HOGの次元数は36bin×49個で1764次元となる。

2.3 SVM(サポートベクターマシン)

SVMは、教師あり学習を用いる識別手法の1つで、線形入力素子を利用して2クラスのパターン識別器を構成する。SVMは、学習データを2クラスに分離する超平面の中でマージンが最大となる超平面を選択するのが特徴で、未知のデータに対する識別率が高さに定評がある。線形SVMのフリーライブラリとしてLinらがLIBLINEAR[4]を開発・公開している。本研究では、SVMによる精度評価、および速度評価に対して、このLIBLINEARを使用する。

2.4 領域選択

2.4.1 事前実験

HOGは勾配の大きさと方向を特徴として捉えるため、勾配の大きな部分が識別に大きな影響を与えていると考えられる。そこで、学習に使用する正例画像8,000枚において、各画素の勾配強度を積算し、ある閾値によって2値化を行ったところ図2のような結果になった。このことから人物の頭から肩にかけて強い勾配が出ていることがわかる。

† 東京電機大学 未来科学研究科 情報メディア学専攻
Graduate School of Science and Technology for Future Life,
Tokyo Denki University‡ 産業技術総合研究所 The National Institute of Advanced
Industrial Science and Technology



図2. 勾配強度の高い領域

2.4.2 マスク作成

勾配の少ない部分は識別に対して影響が少ないと考え、正例画像群に対して、積算勾配強度が小さい領域を識別対象から除外する。各 Block 領域内の勾配強度の平均を求め、それが画像全体の平均よりも小さい Block を除外するマスクを生成する。また、その全体平均の 0.8 倍よりも小さい Block を除外するマスクと、画像の上半分と下半分では勾配強度の違いがあると考え、それぞれの平均を閾値とした領域除外マスクを作成する。また、上記の方法を Cell 単位で行ったマスクも作成する。

3. 実験

3.1 学習データ

学習に用いる上半身画像の正例と負例は、Webで公開されている人物画像データベース[4,5,6,7,8]から切り出したものを使用する。

本研究では正例となる上半身画像を「人物の頭頂部が画像の上辺から縦幅の1/8の位置にあり、かつ、頭部が画像の縦幅の約3/5、横幅の約1/2の領域に収まっている画像」と定義する。人物画像データベースの画像から、正例として定義した条件を満たすように画像を切り出し、最終的に64×64画素の画像にリサイズする。

また、正例を切り出したデータベースの画像から、上半身でない部分をランダムに切り出し、64×64画素の画像にリサイズしたものを負例として使用する。本研究では、正例画像8,000枚と負例画像22,000枚の合計30,000枚の画像を使用し実験を行う。

3.2 評価実験

2.4.2で作成したマスクによって抽出部分を限定したHOGを使用し、速度評価と精度評価を行う。

表1にBlock単位で領域限定を行った場合の、次元数、HOGの生成時間、SVMによる識別時間、学習データ自身に対する識別率、5-fold Cross-validation(CV)の識別率の結果を示し、同様に、表2にCell領域単位で領域限定を行った結果を示す。実験環境は表3に示す。

3.3 考察

HOGを抽出する領域を限定することにより、HOGの生成時間・識別時間の削減が見られる。Block領域単位でもCell単位でも領域限定により識別率の低下が見られるが、小幅に留まっているため、速度向上には提案手法は有効と考える。

識別率の低下は、識別に有効な特徴まで除外してしまったためであろう。勾配が弱い部分であっても勾配の方向に関して強い特徴が存在する可能性もあることから、除外領域を勾配方向の偏り度合い等によって決定することも有効と考えられる。

表1. Block 単位での領域限定の結果

マスク方法	なし	平均	平均×0.8	上下別平均
マスク画像	なし			
次元数	1764	1008	1188	972
HOG生成時間[μs]	126.1	77.0	84.9	64.3
識別時間[μs]	5.6	3.2	4.1	5.7
自身の識別率[%]	97.88	97.26	97.2667	97.66
CVの識別率[%]	97.11	96.68	96.78	96.84

表2. Cell 単位での領域限定の結果

マスク方法	なし	平均	平均×0.8	上下別平均
マスク画像	なし			
次元数	1764	1008	1503	882
HOG生成時間[μs]	126.1	63.5	100.8	58.7
識別時間[μs]	5.6	3.4	5.0	3.0
自身の識別率[%]	97.88	97.42	97.64	97.50
CVの識別率[%]	97.11	96.83	97.05	96.83

表3. 実験環境

OS	Windows7 (32bit)
CPU	Intel® Core™ i7 CPU M 620 @2.67GHz
メモリ	8.00GB (2.98GB 使用可能)

参考文献

- [1]N.Dalal and B.Triggs, "Histograms of Oriented Gradients for Human Detection", Proc.CVPR, vol.1, pp.886-893, 2005.
- [2]T. Kobayashi, et.al, Selection of Histograms of Oriented Gradients Features for Pedestrian Detection, ICONIP, 2007
- [3]山下隆義, 藤吉弘亘, "特定物体認識に有効な特徴量", 情報処理学会 研究報告 CVIM 165, pp. 221-236,2008.
- [4]LIBLINEAR -- A Library for Large Linear Classification, <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>
- [5]INRIA Person Dataset, <http://pascal.nrialpes.fr/data/human/>
- [6]People Playing Musical Instrument, <http://ai.stanford.edu/~bangpeng/ppmi.html>
- [7]The Center for Biological & Computational Learning, <http://cbcl.mit.edu/index.html>
- [8]ETHZ PASCAL Stickmen V 1. 0, http://www.vision.ee.ethz.ch/~calvin/ethz_pascal_stickmen/index.html
- [9]Buffy Stickmen V2. 1, <http://www.robots.ox.ac.uk/~vgg/data/stickmen/index.html>