H-042

二段階式上半身検出器における前段検出器の再現率向上
Recall Improvement for the First-stage Detector in a Two-staged Upper Body Detector

遠藤 怜† 中島 克人†
Rei Endo Katsuto Nakajima

1 はじめに

カメラを用いたコンピュータによるリアルタイム人物位置検出は，安全運転支援システムにおける車載カメラによる歩行者検出や監視カメラによる人の挙動解析等への応用で需要が高まっている．

人検出器としては顔検出器や全身検出器が挙げられるが，前者では横顔や後ろ姿の検出が困難である．また，後者はオクルージョンに弱く，混雑した状況での検出が困難となる．もし，上半身だけを手掛かりに人検出が出来ればこれらの問題は解決する．

本研究では，近年画像認識分野で高い精度が報告されている Deep Convolutional Neural Network(DCNN)[1]を上半身検出に適用する．DCNN は高い識別性能が期待できるが，シーン走査するには低速である．そのため，全身検出で実績が有り実時間動作が可能だが偽陽性が高い，Histogram of Oriented Gradient(HOG) 特徴 [2] と Linear Support Vector Machine(LSVM) による上半身検出器(HOG/LSVM)により，上半身候補を抽出し，その候補領域のみに DCNN による識別を行う二段階式上半身検出器を提案する（図 1 参照）．二段階式検出器における前段検出器は，上半身の候補を漏らさずに検出する必要があるため，再現率が重視される．しかし，同じ学習データを用いた浅利らの実装評価[3]では，上半身検出における HOG/LSVM は再現率が比較的低いことが分かった．

そこで，本研究では上記の二段階式検出器における前段の再現率向上を目的とし，前段の学習データの改善と LSVM のパラメータ最適化を試みた．

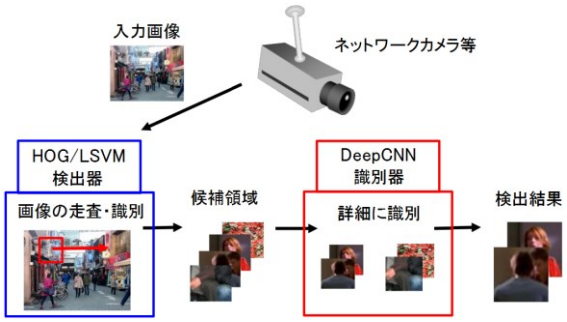


図1 二段階式上半身検出器

2 二段階式上半身検出器

2.1 HOG 特徴による上半身検出器（前段）

HOG 特徴量とは，局所領域単位で輝度勾配方向と勾配強度をヒストグラム化した特徴量である．HOG 特徴により，上半身の形状の関係を捉えている．

識別には LSVM を使用する．これは，正例・負例を分類する超平面を両データ間のマージンが最大になるように決定する識別器である．HOG/LSVM による検出は実時間処理が可能で，再現率は比較的高いが誤検出が多い．

2.2 DCNN による上半身識別器（後段）

DCNN は中間層を 2 層以上持たせた Convolutional Neural Network(CNN)を指す．CNN は多層パーセプトロンの中間層に畳み込み・Pooling 層を取り入れた識別器である．DCNN はシーン画像に対する実時間での検出処理は困難であるが，高い識別性能が期待できる．

2.3 学習データ

検出対象の上半身画像を「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 の領域に収まるもの」と定義する．正例は Web 画像検索エンジンを用いて収集した集合写真，歩行者画像等から切り出し，負例はそれらの中から非上半身となる部分をランダムに切り出すことで収集されている．このデータセットは正例データが 25,000 枚，負例データが 75,000 枚の計 100,000 枚の画像から成る．画像は全て 64×64 画素のカラー画像である．

3 学習データの改善

上記のように，上半身の正例データは負例データの 3 分の 1 しか無く，不均衡である．

LSVM で正例数と負例数が不均衡なデータを学習する場合，少数派データの方に超平面が寄ってしまう問題がある．その対処法には多数派データを減少させる手法（アンダーサンプリング，以降 UnderSmpl），少数派データを増加させる手法（オーバーサンプリング，以降 OverSmpl），多数派データを減少させつつ少数派データを増加させる手法（ハイブリッドサンプリング，以降 HybridSmpl）の 3 種類がある．

本研究ではこれらの手法を前段の学習データに適用した．なお，OverSmpl と HybridSmpl のためには，少数派の正例データに平滑化，コントラスト変更，ガンマ変換，ノイズ付加等の画像処理による Data Augmentation を行い，これらを元の正例データに加える事とする．

4 評価

4.1 HOG/LSVM による学習データの評価

本研究で行った前段の学習データに対するサンプリングは以下の計 29 通りである．これらの中で最良の識別性能を示すものと元データによるそれを比較する．

- OverSmpl：元の正例データと，正例加工データの 2 種を選んで組み合わせたものを 21 種作成する．
- UnderSmpl：元の負例データから 3 分の 1 をランダムに選んだものを 1 種作成する．
- HybridSmpl：元の負例データから 3 分の 2 をランダムに選び，元の正例データと正例加工データの 1 種を選んで組み合わせたものを 7 種作成する．

† 東京電機大学大学院未来科学研究科
Graduate School of Science and Technology for Future Life, Tokyo Denki University

なお，正例加工データ作成の画像処理の具体的な内容は，3×3 のカーネルサイズの平滑化，コントラストを 0.6 と 1.4 の傾きで変換，ガンマを 0.6 と 1.4 で変換，ガウシアンノイズをシグマ 20 での付加，ゴマ塩ノイズを画像の 2.4%への付加，とした．そのため，正例加工データは 7 種となる．

学習データと共に HOG/LSVM の識別器側のパラメータ（以降，LSVM パラメータ）も性能に影響を及ぼす要因のため，各学習データでこれのグリッドサーチを行い HOG/LSVM に最適なデータと LSVM パラメータの組み合わせを探索する．評価に用いる LSVM のパラメータはコストパラメータ C，データ内で許容される残差を指定する回帰用の正則化パラメータ ε の 2 つとした．グリッドサーチでは C を 0.01 から 0.08 までの範囲で 1.5 倍ずつ変化させ，ε を 0.1 から 0.8 までの範囲で 1.5 倍ずつ変化させた．グリッドサーチにおけるデータセットの評価には 10-分割交差検証を用いる．

なお，元データを含む全 30 通りの学習データに対してグリッドサーチを行うことは時間的に困難であるため，事前検証によりグリッドサーチを行うデータの選定を行った．この事前検証ではパラメータを C=0.01，ε =0.1 に固定した LSVM で学習し，学習データセットとは別の正例画像 1,000 枚，負例画像 1,000 枚の計 2,000 枚のテストデータの分類を行う．この結果，元データよりも再現率，F 値が共に高いデータ 9 種と元データのみにグリッドサーチを行うこととした．

グリッドサーチの結果，UnderSmpl における C=0.01，ε =0.1 の組み合わせの F 値が 0.972 で最も高かった．また，再現率も 0.970 と高いことから，これを元データによるものとの比較対象とする．なお，元データにおいては C=0.01，ε =0.1 の組み合わせの F 値が 0.922 で最も高かった．

4.2 映像による性能評価

元データを C=0.01，ε =0.1 で学習した HOG/LSVM（以降，元の HOG/LSVM）と，4.1 の評価実験で求めた最も高い F 値を示す UnderSmpl で学習した HOG/LSVM（以降，最適化 HOG/LSVM）を映像により評価し，比較する．

評価用映像は画像サイズが 640×480 画素，フレーム数が 10 枚である 4 つの映像である．これらの映像に現れる上半身の位置には予め正解となる検出位置として，矩形のアノテーションを付与しておく．

これらの映像に対して，上半身検出器の検出結果の矩形とアノテーション矩形との和集合に対して，両者の重複面積が 50%以上のものを検出成功と見なし，それ以外の検出結果は誤検出と見なす．

元の HOG/LSVM と最適化 HOG/LSVM の評価結果の平均再現率を表 1 に示す．結果から，元の HOG/LSVM に対して最適化 HOG/LSVM では再現率が大幅に向上出来ており，前段検出器に適していることが分かった．

次に最適化 HOG/LSVM の検出結果を DCNN で識別する二段階式上半身検出器の評価を行った．本研究の DCNN では，畳み込み層と Pooling 層が交互に 2 層ずつ存在し，その後，全結合層が 2 層，出力層の順で構成される．

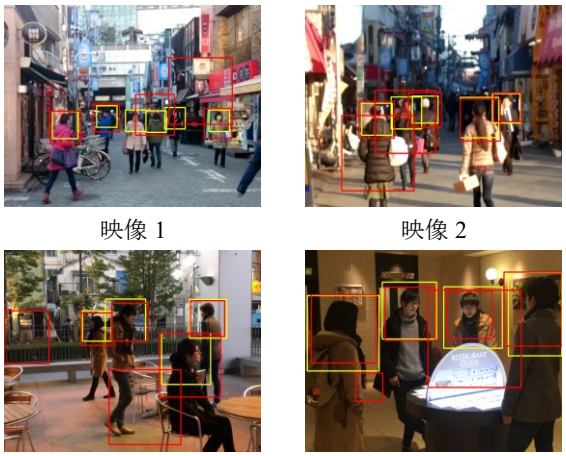
最適化 HOG/LSVM + DCNN の評価結果を表 2 に，その検出結果例を図 2 に示す．図中の黄枠は上半身のアノテーションを示し，赤枠は二段階式上半身検出器の検出結果を示す．

表1　平均再現率の比較

映像番号	元のHOG/LSVM	最適化HOG/LSVM
1	0.347	0.595
2	0.759	1.000
3	0.875	1.000
4	0.750	1.000

表2　最適化HOG/LSVM + DCNN 検出性能

映像番号	平均再現率	平均適合率	平均F値
1	0.423	0.546	0.472
2	0.740	0.696	0.714
3	1.000	0.906	0.946
4	0.875	0.735	0.792



映像 1　映像 2　映像 3　映像 4

図 2　最適化 HOG/LSVM + DCNN による検出結果例

4.3 考察

表 1 より，全映像で元の HOG/LSVM よりも最適化 HOG/LSVM の再現率が高いことを確認できる．しかし，映像 1 の様な背景が暗いシーンではまだ再現率が不十分である．これはこのようなシーンの学習データの拡充により解決できるものと考える．

5 まとめ

学習データのアンダーサンプリングにより，前段検出器の再現率向上について一定の成果を上げることができた．しかし，乱数によるアンダーサンプリングのため，サンプリングごとの検出器の性能変動は確認を要する．最終的には実時間で上半身検出を行うことを目標としているため，速度の評価と，必要ならば高速化の検討も必要だろう．

参考文献

[1] A.Krizhevsky, et al., “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS 2012, 2012.

[2] N.Dalal, et al., “Histograms of Oriented Gradients for Human Detection”, Proc.CVPR, vol.1, pp.886-893, 2005.

[3] 浅利広織，中島克人，“上半身検出の手法とその評価”,情報処理学会第 77 回全国大会，1D-03，2015.