

歩行者検知における Visual Word 選択手法

An Approach to Simplify the Visual Word for the Pedestrian Detection

張 興国[†] 猿田 和樹[‡] 寺田 裕樹[‡] 陳 国躍[‡]
Xingguo Zhang Kazuki Saruta Yuki Terata Guoyue Chen

1. まえがき

画像認識分野において、G.Csurka らによって提案された Bag-of-Features(以下、BoF)手法が近年注目され一般物体認識分野などで多くの研究者が改良を重ねている [1-3]. BoF モデルは画像を局所特徴量の集合と見なし、クラスタリング手法より作成した Visual Word を辞書として、その出現頻度をカウントしたヒストグラムより画像の認識を行う。しかし、辞書中のすべての Visual Word が分類に有効であるとは限らず、また Visual Word の数は学習や認識の時間にも影響する。そこで、筆者らは BoF 手法を改良し、検知に有効な Visual Word を選択する手法を提案した[4]. 本稿では、提案手法において選択される Visual Word に対応する特徴点分布を解析し、検知結果への影響を検討する。

2. Visual Word 選択手法

BoF モデルは、学習画像から抽出された局所特徴量をまとめて特徴空間内でクラスタリングを行い、各クラスターの重心点を Visual Word として作成する。次に、各学習用画像から特徴量のベクトル量子化を行い、Visual Word の出現頻度をカウントしたヒストグラムを作成し(図 1 における h_m と h_n)、このヒストグラムを特徴量とする。しかし、辞書の中には学習や判定に影響の少ない Visual Word も含む可能性がある。提案する Visual Word 選択手法の流れを図 1 に示す。まず、BoF 手法で Visual Word の出現頻度ヒストグラムを positive と negative に分かれ、Visual Word によるヒストグラムの和 H_{pos} , H_{neg} を次式より求めて、ヒストグラムを作成する(図 1 (a)(b)).

$$\begin{cases} H_{pos}(x) = \sum_{m=1}^M h_m(x) \\ H_{neg}(x) = \sum_{n=1}^N h_n(x) \end{cases} \quad x \in [1, X] \quad (1)$$

ここで、 M は学習用の positive サンプルの数、 N は negative サンプルの数、 X は Visual Word の数。図 1 のヒストグラム H_{pos} と H_{neg} の横軸は Visual Word の番号、縦軸はすべての画像から抽出した局所特徴量が各 Visual Word に属する総数を表す。

次に、作成したヒストグラムの正規化を式(2)により行う。ここで、 ε は分母が 0 の場合に計算不能になることを防ぐ係数である。これにより、学習画像の枚数に対して影響されない平均ヒストグラムが得られる。続いて、positive と negative の平均ヒストグラムの差分を求め、差分ベクトル

が得られる(図 1 (c)). 差分ベクトルの絶対値の大きさはこの Visual Word の分類性能を表す。差がプラスである時、この Visual Word は positive サンプルに対して分類効果が高く、マイナスである時 negative サンプルに有効であることを表す。さらに、差分ベクトルの絶対値を図 1 (d) のように大きい順にソーティングする。このとき、適切な閾値 u を設定して、上位 u 個以外の Visual Word は取り除き(図 1 (e)), 上位 u 個の Visual Word は分類性能が高いとして、新たな辞書を登録する。最後に、上位 u 個以外の Visual Word と対応する頻度ヒストグラム h_m と h_n の次元を取り除く(図 1 (f)). そして、削減したヒストグラムを学習サンプルとして SVM に入力することにより、SVM の学習を行う。識別サンプルも図 1(g) のように、BoF 手法で作成した頻度ヒストグラムを new- Visual Word により削減する。

$$\begin{cases} H_{pos}(x) = \frac{H_{pos}(x)}{\sum_{i=1}^X (H_{pos}(i))^2 + \varepsilon} \\ H_{neg}(x) = \frac{H_{neg}(x)}{\sum_{i=1}^X (H_{neg}(i))^2 + \varepsilon} \end{cases} \quad (\varepsilon = 0.0001) \quad (2)$$

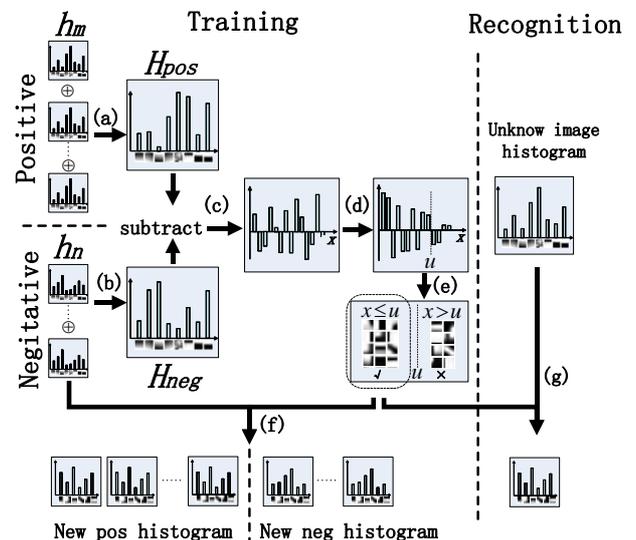


図 1 提案手法の流れ

文献[4]の評価実験より、選択 Visual Word 数 u の減少に伴って、SVM で学習と識別に要する時間時間は大きく減少できること、さらに $u > 150$ の場合、検知精度はほとんど変化しないことが示されている。選択 Visual Word 数 u を 250 から 150 に減少する場合、検知精度がほぼ同程度でありながら、処理時間は 40% を削減できることが明らかとなっている。

[†] 秋田県立大学大学院 Graduate School of Systems Science and Technology, Akita Prefectural University

[‡] 秋田県立大学 Akita Prefectural University

4. 特徴点分析

本研究では, Dense-SIFT 特徴を画像の局所特徴量として用いる. dense-SIFT における特徴量は, 画像全体から抽出できるため, dense-SIFT 特徴は比較的高い検知性能が得られることが分かっている. その一方で, 検知には不要な情報を特徴量に含む可能性がある. そこで, 提案手法における特徴点分析を行い, 検知に有効な特徴点分布を明らかにする. 提案手法における特徴点分析の概要を図 2 に示す. 歩行者画像全体から得られる特徴点に対応する Visual Word を求め, 提案手法で選択される Visual Word に対応する特徴点以外を除去して表示およびカウントすることで, 歩行者検知に有効となる特徴点の分布を明らかにする. 図 2 においては, 左の画像で全体に分布している特徴点に対応する Visual Word を求め, 選択された Visual Word である場合のみ右の画像に特徴点として表示している.

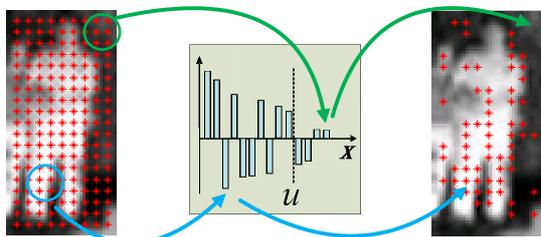


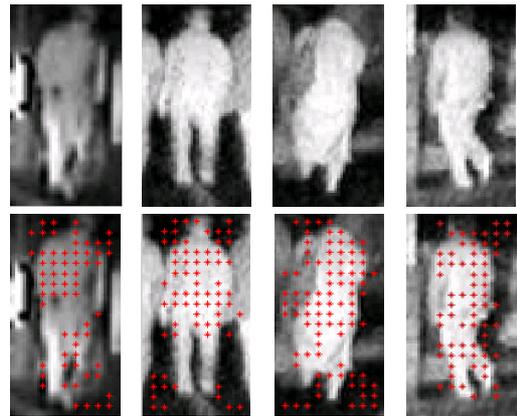
図 2 検知に有効な特徴点の選択

歩行者画像および非歩行者画像に対して選択した特徴点の分布を図 3 に示す. ここでは提案手法における初期 Visual Word 数を 250 に設置し, 有効な Visual Word 数 u を 150 に設定し, 各画像中で上位 150 個の Visual Word に属する特徴点だけを表示することで, 検知に有効な特徴点の分布を示している. 図 3 に示すように, 歩行者においては画像中央部のとくに上半身に特徴点が多く, 非歩行者画像では画像周辺や背景部分に特徴点が存在する傾向が見られた.

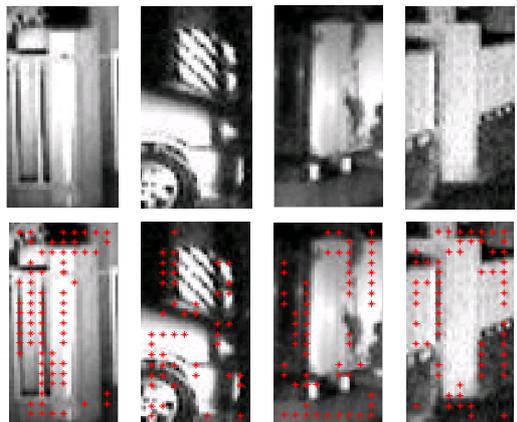
続いて, 学習済みの SVM を用いて歩行者検知実験を行い, TP(true positive), TN(true negative), FP(false positive), FN(false negative) の 4 種類に画像を分類し, 分類結果ごとの特徴点分布を解析する. TP 画像 6000 枚, TN 画像 12000 枚, 誤検知画像 FP1200 枚, 未検知画像 FN1600 枚を使用し, 図 3 と同様に提案手法有効とされた Visual Word に対応する特徴点のみをカウントする. ここでは分類結果別の違いを明確に可視化するために, 特徴点の位置と数をもとに分布画像を作成する. 実験結果を図 4 に示す. 図 4 の画像において白い部分ほど平均特徴点の数が多いことを示している. 図 4 から, 判定結果が正解かどうかに関わらず, 歩行者と判定された場合は有効な特徴点は主に身体の周辺, とくに肩付近に多く, 非歩行者と判定された場合は, 特徴点が全体的に広く分布していることが分かる. これは検知精度に影響する原因の一つであると考えられる.

4. まとめ

本稿では, Bag of Features を基に提案した歩行者検知における Visual Word 選択手法における特徴点分布について解析し, 特徴点分布の面から検知精度への影響の原因を検討した. 今後, さらに詳細な分析をし, 検知手法の改善を図る予定である.

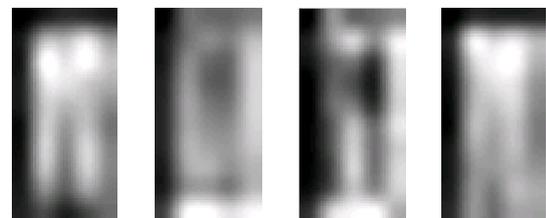


(a) 歩行者画像の有効な特徴分布



(b) 非歩行者画像の有効な特徴分布

図 3 有効な特徴点の分布の例



(a)TP (b)TN (c)FN (d)FP

図 4 分類に有効な dense-SIFT 特徴点の分布

参考文献

- [1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray, "Visual categorization with bags of keypoints," ECCV International Workshop on Statistical Learning in Computer Vision, pp. 1–22, 2004.
- [2] J. Uijlings, A. Smeulders and R. Scha, "Real-time bag-of-words, approximately," in ACM International Conference on Image and Video Retrieval, 2009.
- [3] B. Besbes, B. Labbe, A. Bensrhair and A. Rogozan, "SVM-based fast pedestrian recognition using a hierarchical codebook of local features," IEEE International Workshop on machine learning for signal processing (MLSP'10), pp. 226–231, 2010.
- [4] 張, 猿田, 寺田, 王, 陳, "歩行者検知における Bag-of-Features の codebooks 選択手法", 電子情報通信学会総合大会講演論文集, AS-5-7, pp.s-65-66 (2012)