

SIFT 特徴を用いた歩行者検知に関する検討 Research on The Pedestrian Detection Using SIFT Feature

張 興国[†] 築瀬 千草[‡] 猿田 和樹[†] 寺田 裕樹[†] 陳 国躍[†]
Xingguo Zhang Chigusa yanase Kazuki Saruta Yuki Terata Guoyue Chen

1. はじめに

近年、車載カメラ映像から歩行者を検知し、運転手の注意を喚起するシステムが実用段階となっている[1]。歩行者検知手法では、EOH, HOG, Edgelet などのエッジに着目した特徴量を用いた研究がなされているが、歩行者の姿勢や方向、自転車などパターンが様々であり、特に歩行者かどうかを判定する矩形領域の画像内における歩行者の位置ずれは、検知精度に大きく影響する。

Lowe は Schmid らの局所領域の特徴量記述という考えを拡張し、スケールスペースを用いることで、画像のスケール変化や回転に不変な特徴量を記述する Scale-Invariant Feature Transform(SIFT)を提案した[2]。SIFT 特徴は位置ずれ、スケール変化、照明変化等に頑健である。PCA (Principal Component Analysis)を用いて勾配情報を部分空間へ射影しマッチング精度を向上させる PCA-SIFT[3]や、SIFT の特徴量記述時における背景の影響を軽減する BSIFT [4]など多くの研究者が改良手法を提案している。

本研究では、画像認識による歩行者検知システムにおいて SIFT 特徴を用い、スケール変化、位置ずれに強いことを確認する。また、SIFT 特徴を用いた場合の学習条件の変化に対する検知精度の違いを明らかにする。さらに、SIFT 特徴と HOG 特徴を用い、双方の検知結果の違いを明らかにする。

2. 歩行者検知システムの概要

本研究では、SIFT 特徴を用いて夜間歩行者検出を目指す。この手法の流れを図1に示す。処理プロセスは学習フェーズと識別フェーズに分類される。学習フェーズでは、学習画像から SIFT 特徴点を抽出し、特徴点周辺パターンを SIFT 特徴ベクトル空間内でクラスタリングを行い、各クラスタの重心点を学習サンプルとして SVM で学習、識別器を構築する。

識別フェーズでは、同様に、各認識画像から抽出されたSIFT特徴点をSVMに識別サンプルとして入力し、歩行者画像かどうかを判断する。

SIFT特徴の算出方法は、特徴点の検出と特徴点の記述の2段階からなり、最終的に1つの特徴点に対し128次元の特徴量を抽出する。特徴点の数は画像によって異なるため、画像によって次元数が異なり、このままではSVMに用いることができない。そのためクラスタリング手法(K-means法)を用いて特徴点の数を一定にする。

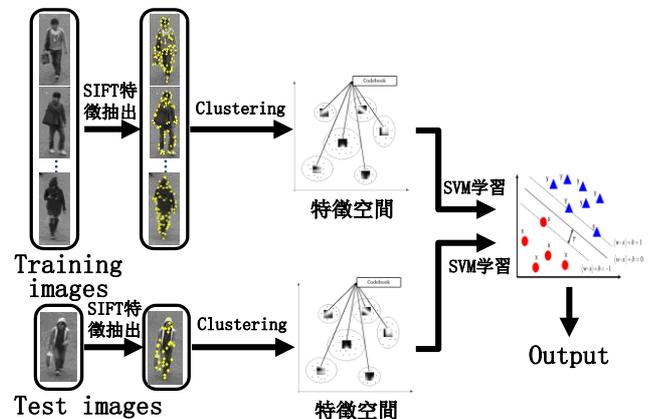


図1 処理の流れ

3. 位置ずれ画像に対する検知実験

3.1 実験概要

ここでは SIFT 特徴がスケール変化、位置ずれに強いことを確認する。なお、結果を検知(歩行者を歩行者と判断)、未検知(歩行者を歩行者以外と判断)、誤検知(歩行者以外を歩行者と判断)、棄却(歩行者以外を歩行者以外と判断)の4種類に分類して検証する。

実験用データとしては、フレーム画像から位置をずらした画像を作成する。歩行者画像を40枚、非歩行者画像を40枚切り出し、各画像20枚、計40枚を学習用データとして、残りの40枚は評価用データとする。フレーム画像及び切り出した画像の例を図2に示す。



図2 フレーム画像及び切り出した画像の例

3.2 実験結果

表1に検知結果を示す。評価用データに対して正解率が100%となり、SIFT 特徴がスケール変化、位置ずれに強いことを確認することができた。

表1.位置ずれ画像の検知結果

	検知	未検知	誤検知	棄却
評価データ	20	0	0	20

[†] 秋田県立大学大学院システム科学技術研究科
Graduate School of Systems Science and Technology of
Akita Prefectural University

[‡] アキタ電子システムズ Akita Electronics Systems

4. 学習条件の変化に対する検知実験

4.1 実験概要

ここでは学習条件を変化させた場合の検知精度を明らかにする。実験用データは SIFT 特徴点が 8 個以上得られたグレースケール画像で、学習用データとして 16×32 ピクセルの歩行者画像 2500 枚、非歩行者画像 7500 枚を使用する。また評価データとして 16×32 ピクセルから 57×114 ピクセルの歩行者画像 8490 枚、非歩行者画像 25004 枚を使用する。学習条件として、クラスタ数を 3 から 8 に変化させる。また学習データを 3 つの条件に設定して学習を行う。学習条件としては歩行者画像 2500 枚、非歩行者画像 2500 枚 (条件 1)、5000 枚 (条件 2)、7500 枚 (条件 3) とする。

4.2 実験結果

学習条件の変化に対する検知結果を図 3 に示す。平均正解率はクラスタ数 4 以降、条件 1 で高くなっており、その中でもクラスタ数 7 の場合が最も高いことがわかる。最も平均正解率が高かった条件 1、クラスタ数 7 の検知結果を表 2 に示す。学習用データに対しては正解率が 100% となっているが、評価用データに対しては 71.5% と正解率が低くなっており、中でも未検知率が高いことがわかる。

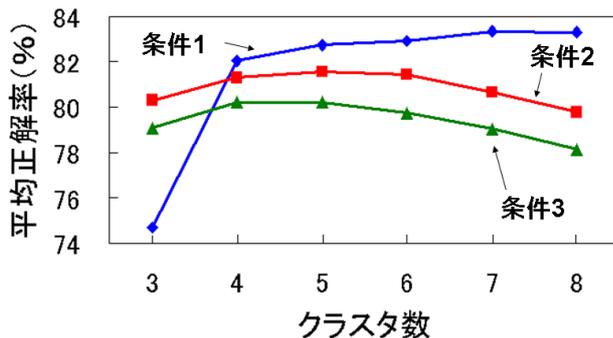


図 3 学習条件の変化に対する検知結果

表 2. 条件 1, クラスタ 7 の検知結果

	学習データ	評価用データ
検知数	2500	4843
未検知数	0	3647
誤検知数	0	5921
棄却数	2500	19083
正解率 (%)	100.0	71.5

5. 考察

最初に SIFT 特徴と HOG 特徴での検知率を比較する。SIFT 特徴の検知率としては表 2 の結果を用いる。また HOG 特徴の検知率を求める際に使用する学習データは条件 1 と同じである。

SIFT 特徴と HOG 特徴の検知率を表 3 に示す。棄却率に大きな違いはみられないが、検知率は HOG 特徴の方が高いことがわかる。

次に、図 4 に示すように、SIFT 特徴と HOG 特徴の未検

知画像を比較する。SIFT 特徴では白黒がはっきりしない画像が未検知となりやすく、HOG 特徴では位置ずれ画像が未検知となりやすい傾向が確認できた。

ここで、HOG 特徴で未検知になった位置ずれ画像 488 枚に対し、SIFT 特徴を用いて検知を行った。結果を表 3 に示す。SIFT 特徴を用いることで約半数の 227 枚が検知できることがわかる。このことから、位置ずれ画像に対しては HOG 特徴よりも SIFT 特徴が有効であるといえる。

表 3. HOG 特徴量と SIFT 特徴量の検知率

	SIFT 特徴	HOG 特徴量
検知率 (%)	57.0	78.8
棄却率 (%)	76.3	78.6

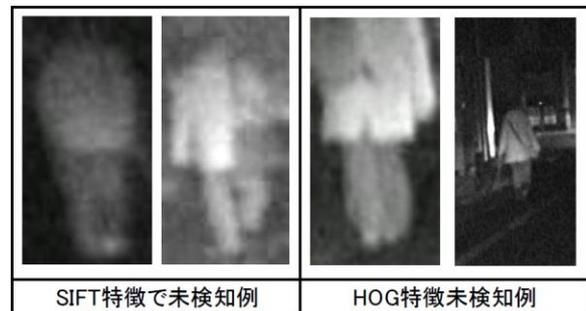


図 4 SIFT 特徴と HOG 特徴未検知例比較

表 4. 位置ずれ画像の検知結果 (SIFT 特徴)

	検知	未検知
位置ずれ画像	227	261

6. まとめ

本研究により、SIFT 特徴がスケール変化、位置ずれに強いことを確認できた。また学習環境の変化による検知精度の違いを明らかにすることができた。今後は、SIFT 特徴とそれぞれの特徴量を組み合わせさせた特徴量を用いて識別器を構築することにより、検知精度向上の効果について検討を行う予定がある。

謝辞

本研究の一部は科学研究費補助金基盤研究(C)(22500161)の補助による。

参考文献

- [1] 豊福邦彦他, “ナイトビュー (歩行者検知機能付) の開発”, TOYOTA Technical Review Vol.56 No.1(2008)
- [2] Lowe, D. Distinctive image features from scale-invariant key-points. International Journal of Computer Vision, 60(2), 91-110 (2004).
- [3] Y. Ke, R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors”, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511-517, (2004).
- [4] A. Stein, M. Herbert, “Incorporating background invariance into feature-based object recognition”, Proc. of IEEE Workshop on Applications of Computer Vision (WACV), pp. 37-44, January, (2005).