

ドライブレコーダーからの歩行者認識技術の研究

A Study on Recognition of Pedestrian from Drive Recorder

稲垣 宏樹†, 江口 一彦††

Hiroki INAGAKI, Kazuhiko EGUCHI

Abstract: The total number of accident itself and injured people are still numerous, although the number of casualty has been decreasing. Many studies and researches have been reported to identify/recognize pedestrians by using onboard video camera to prevent accidents. However current drive assist systems which identify pedestrians automatically and make warning are expensive in general. They are limited to install on luxury cars as optional equipment.

This paper proposes to recognize pedestrians from the movie taken by an inexpensive drive recorder. It is difficult to accurately recognize pedestrians from drive recorder images since the resolution and quality of them is low. We apply support vector machine identifier to reduce false positives and provide accurate recognition. This identifier uses HOG(Histograms of Oriented Gradients) features which is effective to human recognition.

1. はじめに

平成 21 年の交通事故死者数は 4914 人で、9 年連続の減少となっているが、事故発生数・負傷者数は依然として高い値を示している^[1]。

また、これまで高齢者が関係する交通事故では、歩行中に被害者になるケースが大半であったが、最近の傾向として自動車等の運転中に加害者になるケースが増えている^[2]。

運転支援技術の研究として、車載カメラからの動画像を利用した映像認識技術の研究が盛んに行われている。認識の対象としては歩行者や障害物、車線や標識、信号などが上げられ^{[4][5][6][7][8]}、中でも見落とした場合、人命に関わる歩行者を認識する研究が数多く行われている。

歩行者事故の原因は脇見運転や安全不確認などが多く、高齢者ドライバーの事故原因と重なっている。そのため、増加している高齢者ドライバーの事故を未然に防ぎ、安全な運転を可能にする運転支援システムは必要不可欠であるが、既に実用化されている運転支援システム^{[7][8]}は高級車に限られた高価なオプションであり、普及は

進んでいない。

そこで、最近低コスト化され普及が進んでいるドライブレコーダーに着目した^[3]。

本研究では、ドライブレコーダーから取得した動画像を用いて歩行者の認識を行う。しかし、ドライブレコーダーの映像は品質が低いため誤検出が多く正確な検出が難しい。そこで人認識において有効な特徴量である Histograms of Oriented Gradients(HOG)特徴量^[11]を用いてサポートベクターマシンに^[12]より識別器を構築し、誤認識を減らすことを試みた。

2. 歩行者検出の流れ

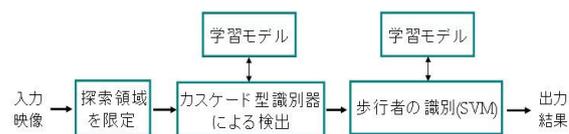


Fig.2.1 フローチャート

本研究の検出手法のフローチャートを Fig.2.1 に示し、それぞれの処理について簡潔に説明する。

・入力映像

ドライブレコーダーにより撮影された動画像から 1 フレーム分の画像を取得する。

† 愛知工業大学大学院 工学研究科
電気電子工学専攻 (豊田市)

†† 愛知工業大学 工学部 電気学科 (豊田市)

・探索領域の限定

探索領域を歩行者が存在しない、上部の上空部分、下部のボンネット部分以外に限定する。

・カスケード型識別器による検出

探索領域から Haar-like 特徴量を算出する。この特徴量を AdaBoost^[10]に基づくカスケード型の高速度なパターン識別器^[9]を用いて歩行者の検出を行う。

・歩行者の識別

検出された領域には誤検出が含まれており、誤検出を減らす必要がある。検出領域から HOG(Histograms of Oriented Gradients)特徴量を算出し、SVM(Support Vector Machine)識別器により識別を行う。

・出力結果

歩行者として識別された領域を画像中に描画する。

3. Haar-like 特徴を用いたカスケード型識別器

Haar-like 特徴を弱識別器とし、それを複数連結して高精度の検出を行なう手法であり、Viola と Jones らによって提案され顔検出などに多く用いられている。

3.1 Haar-like 特徴量

Haar-like 特徴量は Fig.3.1 のように白い矩形と黒い矩形の 2 つの矩形の輝度平均値の差で表される。2 つの矩形のパターンは多数あり、大きさや矩形の位置を変化させることで縦方向や横方向の明暗差を求めることができる。

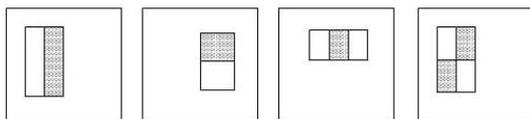


Fig.3.1 Haar-like 特徴の例

3.2 カスケード型識別器

得られた特徴量はパターン識別のための学習方式の 1 つである AdaBoost を用いて識別器を算出する。AdaBoost による学習では、識別したいクラスに属するパターンの例題 (学習サンプル) を与えると、属するクラスが未知の入力パターン (未学習サンプル) を識別するための関数が得られる。AdaBoost は、逐次的に学習サンプルの重みを変化させながら異なる識別器を作り、これら複数の識別器の重み付き多数決によって最終的な識別関数を与える。個々の識別器は弱識別器と呼ばれ、それらを組み合わせたものを強識別器と呼ばれる。Haar-like 特徴に基づく弱識別器 $h_t(x)$ は式(3.1)のように設計される。

$$h_t(x) = \begin{cases} +1 & \text{if } p \cdot z(x) > p \cdot \theta \\ -1 & \text{otherwise} \end{cases} \quad (3.1)$$

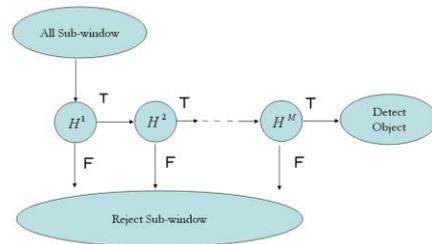


Fig.3.2 カスケード型識別器

ここで、 z は 1 つの Haar-like 特徴から算出された特徴量、 θ は閾値である。また、 p は特徴量と閾値を比較する不等号の向きを決定する変数である。

カスケード型識別器は Fig.3.2 に示す H^1 から H^M の M 個の AdaBoost 識別器を直列に接続したものである。画像をスキャンする走査ウィンドウを入力とし、そのウィンドウ内に検出対象が含まれるか否かを識別する。それぞれの AdaBoost 識別器 H^m は式(3.2)のように定義される。

$$H^m(x) = \text{sign}\left(\sum_{t=1}^m \alpha_t^m h_t^m(x) - \theta^m\right) \quad (3.2)$$

ここで、 T は用いられる弱識別器の数、 θ^m はしきい値、 α_t は信頼度で学習サンプルに対する誤り率 ϵ_t から式(3.3)のように表わされる。

$$\alpha_t = \frac{1}{2} \log\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \quad (3.3)$$

カスケード前段の識別器 (H^1 や H^2) は例えば壁や空といった単純な背景領域を棄却する役割を持つ。後段の識別器は、前段で棄却されなかった複雑な背景領域と対象を区別できるように、より多くの弱識別器を用いる。最終段の識別器 (H^M) を通過したウィンドウのみが対象であると判断される。

4.HOG 特徴量

Dalal 等により勾配ベースの特徴量として Histograms of Oriented Gradients(HOG)が提案されている。局所領域における輝度の勾配方向をヒストグラム化した特徴量である。近接画素の勾配を局所領域によってヒストグラム化するため、照明の影響を受けにくく、局所的な幾何学変化に頑健という特徴がある。以下に HOG 特徴量の算出方法について述べる

4.1 輝度勾配算出

各ピクセルの輝度 L から勾配強度 m と勾配方向 θ を式(4.1)(4.2)(4.3)より求める。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (4.1)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (4.2)$$

$$\begin{cases} f_x(x, y) = L(x + 1, y) - L(x - 1, y) \\ f_y(x, y) = L(x, y + 1) - L(x, y - 1) \end{cases} \quad (4.3)$$

4.2 セル毎に勾配方向と強度をヒストグラム化

求めた輝度の勾配強度 m と、勾配方向 θ より 6×6 ピクセルを1セルとして、セルごとに輝度の勾配方向をヒストグラム化する。輝度の勾配方向ヒストグラムは $0^\circ \sim 180^\circ$ を 20° ずつに分割するため、9方向の勾配方向ヒストグラムになる。

4.3 ブロック毎に正規化

各セルで作成した輝度の勾配方向ヒストグラムを 3×3 セルを1ブロックとして正規化を行う。 i 行 j 列のセル (i, j) の特徴量(9次元)を $F_{i,j} = [f_1, f_2, \dots, f_9]$ とすると、 k 番目のブロックの特徴量(81次元)は $V_k = [F_{i,j}, F_{i+1,j}, F_{i+2,j}, F_{i,j+1}, F_{i+1,j+1}, F_{i+2,j+1}, F_{i,j+2}, F_{i+1,j+2}, F_{i+2,j+2}]$ と表すことができる。正規化後の特徴量を v としたとき、式(4.4)より正規化する。

$$v = \frac{f}{\sqrt{\|V_k\|_2^2 + \epsilon^2}} \quad (\epsilon = 1) \quad (4.4)$$

正規化はブロックを1セルずつ移動させることによって行う。そのため、特徴量 f は異なるブロックの領域によって何度も正規化される。入力画像を 30×60 ピクセルとした場合、横方向3ブロック、縦方向8ブロック、合計24ブロックに対して正規化を行う。ブロックごとに正規化されたHOG特徴量は24ブロック \times 81次元=1,944次元となる。

5. サポートベクターマシン

サポートベクターマシン (Support Vector Machine, SVM) は1990年代にVapnikらによって考案された識別手法で、非線形の識別手法へと拡張されたことで手書き文字認識、画像認識、タンパク質配列分析など幅広い問題へと適用されている。

SVMはニューロンを単純化したモデルで、入力データ $(x_1, y_1), \dots, (x_n, y_n)$ に対し $(x_i \in R^m, y_i \in \{-1, 1\})$ 、識別関数は式(5.1)のように表わされる。

$$f(x) = \mathbf{w}^T x_i + b \quad (5.1)$$

ここで、 \mathbf{w} はシナプス荷重に対応するパラメータであり、 b はしきい値である。このパラメータ \mathbf{w}, b は冗長性を持っており、一意に決まらない。そのため、式(5.2)のような制約を入力データと超平面の最小距離を表す式(5.3)に加える。

$$\min_{i=1, \dots, l} |\mathbf{w}^T x_i + b| = 1 \quad (5.2)$$

$$\min_{i=1, \dots, l} \frac{|\mathbf{w}^T x_i + b|}{\|\mathbf{w}\|} \quad (5.3)$$

このような制約を加えることで、最小距離は $1/\|\mathbf{w}\|$ とな

る。 \mathbf{w}, b は、最小距離を最大にするように決めるので、次のように定式化される。

目的関数：

$$\frac{1}{2} \|\mathbf{w}\|^2 \rightarrow \text{最小化} \quad (5.4)$$

制約条件：

$$y_i(\mathbf{w}^T x_i + b) \geq 1 (i = 1, \dots, l) \quad (5.5)$$

これは凸2次計画問題を解くことと等価になる。この最適化問題をラグランジュ乗数法を用いると、より簡単な問題に帰着する。ラグランジュ乗数 $\alpha_i (\geq 0)$ を導入すると、

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T x_i + b) - 1) \quad (5.6)$$

を得る。 \mathbf{w}, b に関する停留点から式(5.7)が得られる。

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i x_i \quad (5.7)$$

これを式(5.1)に代入すると

$$f(x) = \sum_{i=1}^n \alpha_i y_i x_i \cdot x + b \quad (5.8)$$

という、最適な識別関数が得られる。 α_i が0でない入力データ x_i のことをサポートベクターと呼ぶ。

5.1 ソフトマージン

上述のSVMは、訓練サンプルは超平面によって線形分離可能という仮定であった。しかし、パターン認識のような実問題で線形分離可能な場合は稀である。その場合、制約条件(式(5.5))を満たす \mathbf{w}, b が存在しないため、最適化を行うことができない。このような場合に対処するために「ソフトマージン法」考案されている。最適化を行うには、制約条件を緩めることが必要となるので、スラック変数 $\xi \geq 0, i = 1, \dots, l$ を導入し、制約条件を次のように変更する。

$$y_i(\mathbf{w}^T x_i + b) \geq 1 - \xi \quad (5.9)$$

そして、最適化においては、次の目的関数を最小化する。

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (5.10)$$

ここで、 C が、どこまで制約条件を緩めるかを指定するパラメータであり、設定は実験的に行うことになる。

5.2 カーネルトリック

ソフトマージン法を用いたとしても、本質的に非線形で複雑な識別課題に対しては、必ずしも良い性能の識別器を構成できるとは限らない。そこで、前処理として、非線形な写像を用いて、より高次元の空間に写像を行い、線形分離性を高める「カーネルトリック」と呼ばれる方法が知られている。

SVMは識別関数が入力パターンの内積のみに依存しているため、内積が計算できれば最適な識別関数を構成

することができる。入力データ x を非線形に写像した空間での二つの要素 $\phi(x_1)$ と $\phi(x_2)$ の内積が、

$$\phi(x_1)^T \phi(x_2) = K(x_1, x_2) \quad (5.11)$$

のように、入力特徴 x_1 と x_2 のみから計算できるなら、非線形写像によって変換された特徴空間での特徴 $\phi(x_1)$ と $\phi(x_2)$ の代わりに、 $K(x_1, x_2)$ から最適な非線形写像を構成できる。ここで、このような K のことをカーネルと呼んでいる。

6. 実験方法

実験には株式会社安心マネジメント製「あんしん mini(DRA-01)」より撮影した動画を用いた。データは市街地や郊外など実環境を走行して収集した 1953 フレームである。動画中には画面端で体の一部が隠れている人物や数フレームしか出現しない人物が含まれているが、より実用的な評価のために一般環境のデータをそのまま用いている。撮影時刻は朝方～夕方までの昼間とした。

画像のサイズは 640×480 ピクセルで、不要と思われる上空やボンネット部分を除いた 640×240 ピクセルを処理の対象とした。

認識の対象は歩行者、自転車に搭乗した人とし、学習用サンプルとして対象画像 4170 枚、非対象画像 7000 枚を用いてカスケード構造の歩行者検出器を構築した。

SVM には台湾国立大学の Lin らによって作られた SVM ライブラリの LibSVM^[19]を用いた。識別には学習用サンプルとして対象画像 2000 枚、非対象画像 2150 枚、性能評価用サンプルとして対象画像 400 枚、非対象画像 1000 枚を用意した。Fig.6.1 に対象画像と非対象画像の一部を示す。



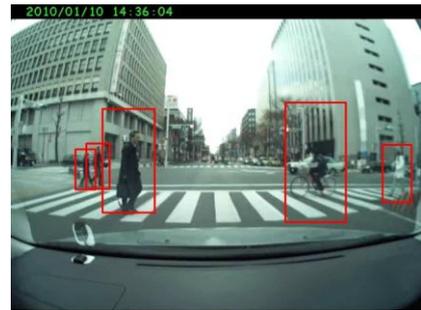
(a) 対象画像 (b) 非対象画像

Fig.6.1 学習用サンプルの一部

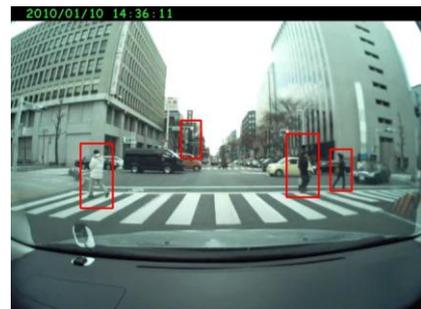
SVM の事前設定パラメータは実験的に行い最適なものを選択した。なお開発環境は Microsoft Visual Studio 2005 で、使用した PC は CPU:Core™2 Duo E4500 @ 2.20GHz、メモリ:2GByte で実験を行っている。

7. 実験結果

検出結果の例を市街地、郊外、その他に分けて Fig.7.1 ~ Fig.7.3 に検出成功例と失敗例で示す。



(a) 検出成功例



(b) 誤検出例

Fig.7.1 市街地における検出結果



(c) 検出成功例



(d) 誤検出例

Fig.7.2 郊外における検出結果



(e) 検出成功例



(f) 誤検出例

Fig.7.3 その他における検出結果

検出率と誤検出率を以下のように定義する。

$$\text{検出率} = \text{対象検出数} / \text{対象数}$$

$$\text{誤検出率} = \text{誤検出数} / \text{フレーム数}$$

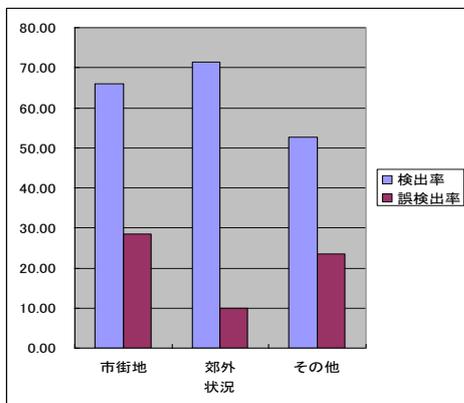


Fig.7.4 各状況の検出率と誤認識率

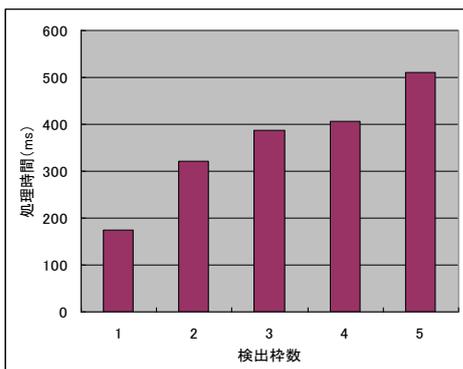


Fig.7.5 1フレーム当たりの処理時間と検出枠数

8. 評価

市街地

市街地は建物などで背景が複雑になるため、最も誤検出が多く出る結果になった。また、人の往来によってオクルージョンの発生が多くなることも検出率に影響を与えている。検出対象や誤検出が増えることで、SVMによる識別の回数が増し処理に時間がかかってしまっている

郊外

市街地に比べて複雑な背景がないため、誤検出は最も少なかった。また複数人が一度に映り込むことが少なかったため、処理時間の増加は少なかった。

その他

その他として屋内駐車場の例を示した。天井からの照明により映像右側が白くぼやけてしまっている。使用しているドライブレコーダーが明るさの変化に対応するまでに時間がかかるためだと考えられる。Fig.8.1のように右半分が光により劣化し、歩行者をほとんど判別することができない例もあった。



Fig.8.1 光による画像の劣化

検出が困難と思われる状況

Fig.8.2は高架の影から出た瞬間で逆光も受けている。暗い場所から明るい場所へと移った場合、数秒の間画面が真っ白な状態になる。逆に明るい場所から暗い場所へ移る場合は画面が数秒真っ暗になる。Fig.8.3は立体駐車場に入る瞬間である。この二つの状況では歩行者がいなかったため、検出は行っていないが、恐らく正常な検出はできないと考えられる。



Fig.8.2 暗い場所から明るい場所への移動



Fig.8.3 明るい場所から暗い場所への移動

9. 結果

ドライブレコーダー映像からの歩行者検出を行った。複雑な背景の中であっても HOG 特徴を入力ベクトルとした SVM による識別で歩行者を検出することができた。今回は映像を記録するドライブレコーダーに手を加えず市販されている状態で使用したが、逆光時の対処などソフトウェア以外にもハードウェア側の工夫が必要なのことがわかった。

10. 考察・今後の課題

今回はオクルージョン（部分隠れ）に対して特別な工夫を行わなかったが、局所特徴量を用いたことで、低度のオクルージョンであれば、検出は可能であった。しかし、Fig.10.1 のように検出ができなかった場合もあった。そのため、体の部分的な特徴（上半身や左半身、右半身）だけでも識別できるように改良することが考えられる。



Fig.10.1 検出ができなかった例

処理速度の改善については、SVM より高速な識別手法（カスケード構造にした AdaBoost など）を用いた方法が考えられる。

精度の改善については、時系列データの特徴量として用いることで、誤認識の削減と検出精度の向上が可能と考えられる。

また、ソフトウェアの改良以外にも、外部の明暗による画質の劣化を抑えるハードウェア的な工夫も必要だと考えられる。

実用機器とするため、マイコンなど組み込み機器への実装も必要である。

参考文献

- [1] 警察庁「平成 21 年中の交通事故死者数について」
URL: http://www.npa.go.jp/toukei/kouki/0102_H21dead.pdf
- [2] 総務省「平成 21 年度版高齢社会白書」
URL: http://www8.cao.go.jp/kourei/whitepaper/w-2009/zenbun/21pdf_index.html
- [3] 国土交通省「平成 20 年度映像記録型ドライブレコーダー活用モデル事業報告書」
URL: <http://www.mlit.go.jp/jidosha/anken/03driverec/resource/data/dora-houkoku20.pdf>
- [4] 高木雅成, 藤吉弘亘: “SIFT 特徴量を用いた交通道路標識認識”, 第 13 回画像センシングシンポジウム SSII07, LD2-06, 2007.
- [5] 中田裕士, 子安大士, 前川仁: “車載カメラを用いた動画からの交通信号認識”, 電子情報通信学会技術研究報告. ITS, Vol.107(486) pp.121-125, 2008.
- [6] 澤野弘明, 岡田稔: “Snake を用いた道路両端白線の両エッジ追跡”, 電子情報通信学会技術研究報告. PRMU, Vol.105(674) pp.265-270, 2006.
- [7] 藤田浩一ほか: “プリクラッシュセーフティの開発”, トヨタテクニカルレビュー, Vol.53(219), pp.70-75, 2004.
- [8] 勝野歳泰, 石崎 覚: “夜間の運転視界支援技術”, トヨタテクニカルレビュー, Vol.56(2), pp.58-63, 2009.
- [9] P. Viola and M. Jones.: “Robust real-time face detection”, IJCV, Vol. 57(2), pp. 137-154, 2004.
- [10] Y. Freud and R.E. Schapire: ”A Short Introduction to Boosting”, J. Japan. Soc. for Artif. Intel. Vol.14(5), pp. 771-780, 1999.
- [11] N. Dalal and B. Triggs: “Histograms of Oriented Gradients for Human Detection”, CVPR, pp. 886-893, 2005.
- [12] V.N.Vapnik: “Statistical Learning Theory”, John Wiley & Sons, 1998.
- [13] 八木康史, 斎藤英雄: “CVIM チュートリアルシリーズ コンピュータビジョン 最先端ガイド 1, アドコム・メディア, 2008.
- [14] Cristianini, N. & Shawe-Taylor J: “サポートベクターマシン入門”, 共立出版, 2005.
- [15] 津田宏治: “サポートベクターマシンとは何か”, 電子情報通信学会誌, Vol.83(6), pp.460-466, 2000.
- [16] 栗田多喜夫 :サポートベクターマシン入門
URL: <http://www.neurosci.aist.go.jp/~kurita/lecture/svm/svm.html>
- [17] C. C. Chang, C. J. Lin:

URL:<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

[18] 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, November, 2008.

[19] 藤吉弘亘: “一般物体認識のための局所特徴量(SIFTとHOG)”, PCSJ/IMPS2008 ナイトセッション.

[20] 鷺見和彦, 関真規人, 波部斉: “物体検出- 背景と検出対象のモデリング-“, CVIM, pp. 79-98, 2005.

[21] 竹内大輔, 丸谷健介: “安全運転支援システムにおける歩行者認識技術の開発”, Panasonic Technical Journal, Vol.54, No.4, Jan. 2009.

(受理 平成 22 年 3 月 19 日)