

ステレオ視と HOG を用いた歩行者認識手法

外谷 文人[†] 吉田 武史[†] 胡 振程[†] 内村 圭一[†]
[†]熊本大学大学院自然科学研究科

概要 歩行中の交通事故死亡者数を低減するため、車載カメラを利用した歩行者認識システムの要求が高まっている。本論文では、車両に設置した左右2台のカメラから得られる視差情報と HOG フィルタを用いた歩行者認識手法について提案する。まず、左右の画像から視差画像を算出し、視差情報をもとに領域分割を行い、背景領域を除去する。その後、Sobel フィルタによりエッジ情報を抽出し、HOG フィルタを適用後、NaiveBayesClassifier を用いて歩行者の認識を行う。これらの手法を道路撮影画像に適用し、有効性の確認と検証を行う。

Pedestrian Recognition using StereoVision and Histogram of Oriented Gradients(HOG)

Ayato Toya[†] Takehumi Yoshida[†] Zhencheng Hu[†] Keiichi Uchimura[†]
[†] Graduate School of Science and Technology, Kumamoto University

Abstract In late years, traffic fatalities in the pedestrian percentage is growing. Therefore, pedestrian aimed at the prevention of traffic accidents is to build the system demands. In this paper, as a part of the prevention of traffic accident system for the pedestrian, I propose pedestrian recognition using stereo vision and histogram of oriented gradient(HOG). First, we abstract pedestrian area from disparity image which is obtained from stereo vision. Next, we created edge image from the original image with Sobel Filter. And abstract only edge included in pedestrian area. We perform HOG process for the edge image, and get the HOG feature. We perform pedestrian recognition with Naive Bayes Classifier using HOG feature. As inspection of this technique, I performed inspection by the ROC curve and inspection with a database for inspection. We showed that this approach was effective as pedestrian recognition method.

1. はじめに

ITS における自動車側からのアプローチとして予防安全技術がある。交通事故の発生原因には、認識ミス、判断ミス、操作ミスなどがある。この交通事故発生原因の中で、1番の原因となっているのは認識ミスである。ドライバーに対して周辺環境の障害物情報を提供することが出来れば交通事故の低減が可能であると考えられる。また、交通事故死亡者の中で約30%を歩行者が占めており、歩行者の交通事故の致死率は全体の約4.4倍になっている。歩行者を対象とした交通事故予防システムを構築できれば交通事故における死亡者数が低減できると考えられる。

そこで、我々はステレオ視とレーザーレーダのセンサフュージョンによる歩行者認識システムを提案した[1]。システムは大きく分けて2つの部分に分けられる。1つは歩行者検出の部分である。歩行者検出部分では、歩行者候補領域の検

出を行う。もう一つは、歩行者認識部分である。検出部分において検出された候補領域が歩行者認識にかけられ、歩行者であるかの判断を行う。

従来、歩行者認識では SVM やニューラルネットワークといった識別器を用いていた[2][3]。これらの識別器は高精度の認識精度を持っている。しかし、これらの識別器は認識速度が遅く、リアルタイム性に欠ける部分があった。

そこで本論文では、ステレオ視と HOG フィルタを用いた NaiveBayesClassifier による歩行者認識を行う。NaiveBayesClassifier は SVM やニューラルネットワークよりも認識速度が速い特徴を持っている。しかし、その反面単純な認識手法であるがゆえに SVM などの識別器と比べると認識率が若干劣ってしまふ。この問題を解決するために、ステレオ視を用いることで3次元情報を考慮し、HOG フィルタを用いることで周辺のエッジの影響を考慮した歩行者認識を行う。

2. 原理

2.1. ステレオ視

左右2つのカメラを用いて画像を撮影して、対象物までの距離情報を検出する方法をステレオ視という。図1に一般的なステレオ視の幾何学モデルを示す。左右2つのカメラ画像のずれを視差という。ステレオ視とはこの視差から幾何学的関係により奥行き情報を求める手法である。同一平面上にある左右2つのカメラを用いて撮影したステレオ画像の視差に関する基本的な式は式(1)で表される。

$$d = x_l - x_r = f \frac{b}{z} \quad (1)$$

ここで、 d は視差、 x_l 、 x_r は左画像、右画像の x 座標、 f は焦点距離、 b は基線距離、 z は奥行きである。

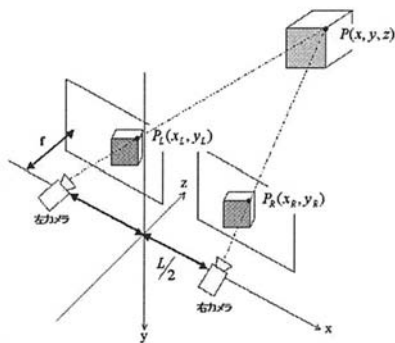


図1. ステレオ視基本モデル

2.2. レーザレーダ

レーザ光を放射し、反射光を受光する。このとき、照射した光と受光した光の間に生じる時間差や周波数差などによって距離情報や相対速度などを取得している。あらゆる方向に対してレーザ光を照射し、距離情報を取得する。これにより、レーザレーダは3次元情報を取得している。図2にレーダ系幾何学モデルを示す。この図のように基本的には二次元平面での探索となる。

2.3. センサフュージョン

カメラとレーザレーダは、それぞれ長所短所が存在する。車載カメラの視野角はレーダよりも広いものとした際、両センサの視野角の関係を図3に示す。レーダは距離・相対速度計測にお

いて高い精度を持つものの、物体認知を行うことは不可能である。一方、カメラは視野角が広く環境認識に優れているものの、ロングレンジや悪天候での精度に問題がある。そこで、両センサを合成することでレーザだけでは困難な対象物認知を可能とし、また、レーダ範囲外の領域に対しては、カメラを使用することで、道路状況認識を可能とする。

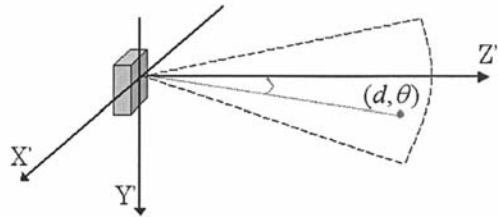


図2. レーダ系幾何学モデル

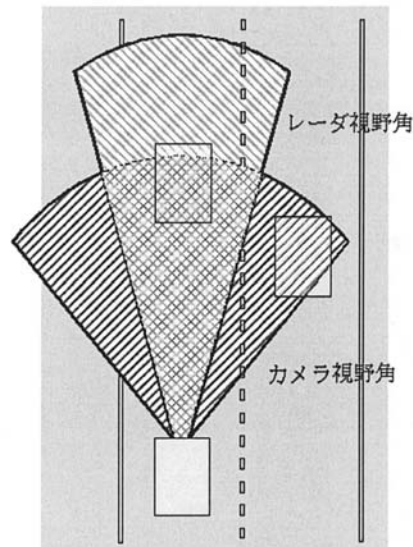


図3. カメラ視野角とレーダ視野角

3. 歩行者検出システム

本システムは、ステレオカメラ、レーザレーダの両センサで取得した情報を入力とする[2]。まず、ステレオカメラから視差画像を取得する。その際、視差情報をOFMに投影する。また、レーザレーダで取得したターゲットデータをステレオカメラ座標系に変換すると同時にOFMに投影する。その後、OFM上の両センサデータの関連付けを行い、それらの情報を用いることで候補領域ROIを作成する。検出したROIを歩行者認識にかけることにより、歩行者の判定を行う。

ここで歩行者として判定されなかったROIは削除する。歩行者認識により歩行者と認識されたROIは、次のフレームで追跡のモデルとして用いる。追跡にはパターンマッチングを用い、マッチング結果を”高”，”中”，”低”の3つで分類し，”高”の場合には歩行者として認識し，”中”の場合には候補領域ROIとして認識にかけられる。 ”低”の場合にはROIを削除する。

4. 歩行者認識手法

本歩行者認識手法は，ステレオ視によって得られる視差画像と原画像を入力データとして認識を行う。また，画像のサイズは64×128としており，入力として与えられる前に正規化してあることを前提として行っている。

4.1. ヒストグラムの平滑化

画像の輝度分布がある範囲に集中している場合，色が類似しているためエッジ特徴の抽出が難しい。そこで，ヒストグラムの平滑化を行い，画像の輝度分布を均一に調整することによりエッジ特徴の抽出を容易にする。

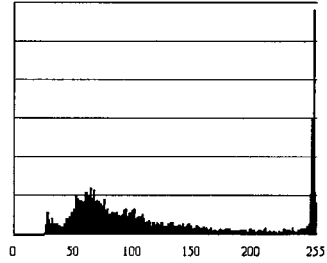
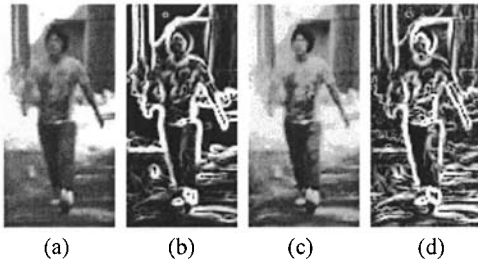
4.2. エッジ抽出

エッジ抽出はSobel Filterを用いて行う。Sobel Filterは式(2)のようなフィルタを適用することにより，縦エッジ，横エッジの抽出を行う。その後，式(3)，(4)を適用することで縦エッジ，横エッジからエッジ角とエッジの大きさを取得する。

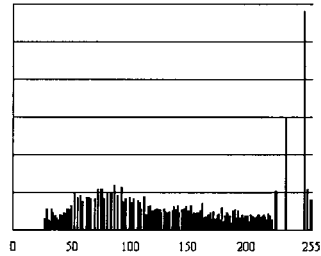
$$S_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad S_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (2)$$

$$|e_{i,j}| = \sqrt{e_{x_{i,j}}^2 + e_{y_{i,j}}^2} \quad (3)$$

$$\angle e_{i,j} = \arctan\left(\frac{e_{y_{i,j}}}{e_{x_{i,j}}}\right) \quad (4)$$



(e)画像値のヒストグラム



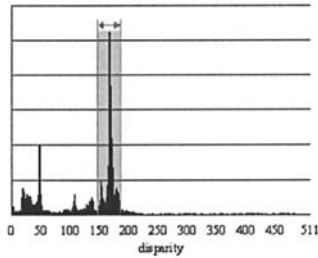
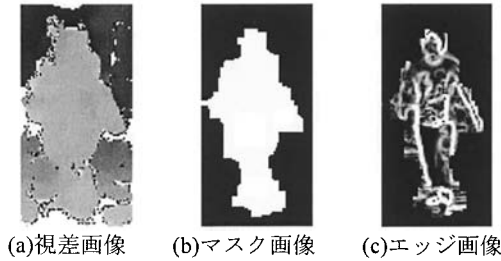
(f)画像値のヒストグラム(平滑化後)

図4. ヒストグラムの平滑化
(a)原画像 (b)エッジ画像
(c)平滑化後の原画像 (d)平滑化後のエッジ画像

4.3. 背景エッジの除去

対象となるエッジ情報のみを取得するために背景エッジなどの情報の除去を行う必要がある。そこで，視差画像からマスク画像を作成し，適用することで背景エッジの除去を行う。

マスク画像の作成には，視差画像から得られる視差情報を用いる。視差値は0～511の値をとり，各ピクセルに視差値がある。対象物の視差値を示すピクセルは画像中に多く存在していると仮定し，最も多い視差値付近だけを抽出する。このときに抽出する範囲は視差値の大きさによって決まる。抽出された部分のエッジ情報以外を除去するようなマスクを作成し，膨張・収縮処理等のノイズ除去を行う。マスク画像に対してノイズ除去を行った結果，マスクの粒子が一定の大きさ以上の場合には，そのマスク画像を用いて背景エッジの除去を行う。粒子が小さい場合には，背景エッジの除去は行わない。



(d)視差値によるヒストグラム
図5. 背景除去

4.4. HOG(Histogram of Oriented Gradient)

取得したエッジ情報に周辺エッジとの関連性を持たせるためにHOGを用いる。HOGは、一定領域内での量子化したエッジ角によるヒストグラムを用いた特徴抽出手法である[3].

4.4.1. セルによるヒストグラムの作成

ヒストグラムはセルと呼ばれる一定領域内で作成される。セルの領域は、4×4pixelや8×8pixelなどの正方形の領域が割り当てられる。セル内に含まれるエッジはエッジ角で20°毎に分類され、そのエッジの大きさによる重みを蓄積することによりヒストグラムを作成する。

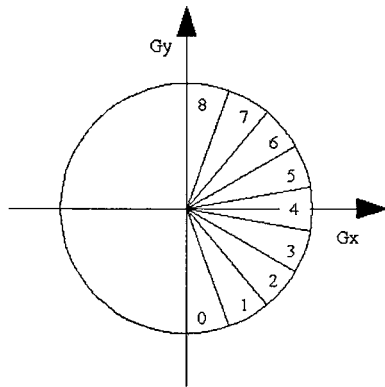


図6. エッジ角の量子化

4.4.2. ヒストグラムの正規化

作成されたヒストグラムはブロックと呼ばれる一定領域内で正規化する。ブロックは、3×3セルの正方形の領域が割り当てられる。ブロックを1セルずつ走査し、画像全体に対して正規化を行う。正規化は式(5)を用いて行う。

$$\frac{V}{\sqrt{\|V\|_2^2 + \epsilon^2}} \quad (5)$$

このとき、 V はブロック要素、 ϵ は正規化整数であり、 ϵ は結果に影響を及ぼさない。

4.4.3. 特徴量の量子化

抽出した特徴 v の値を量子化する。量子化はNaiveBayseClassisierで使用する。量子化は以下の表のように行う。

表1. 値の量子化

量子化値	値
0	~0.005
1	0.005~0.01
2	0.01~0.05
3	0.05~0.1
4	0.5~

4.5. 歩行者認識

量子化したHOG特徴をNaiveBayseClassifierの特徴量とすることで歩行者認識を行う。NaiveBayseClassifierは対象とする特徴が歩行者である確率と非歩行者である確率を調べることにより、歩行者の認識を行う。認識時に使用するそれぞれの確率は、それぞれのトレーニングデータでの出現確率を調べることで取得する。

4.5.1. LUTの作成

LUT(出現確率の結果)は、歩行者、非歩行者のデータベースから作成する。データベース内の画像データから、特徴を抽出し、データがあるか無いかをカウントしていく。データベース内のすべての画像に対してカウントを行い、それらをデータ数で割ることで、出現確率の算出を行う。このとき、出現確率は量子化した値ごとに算出を行う。

4.5.2. NaiveBayseClassifier による認識

NaiveBayseClassifier による認識は、式(9)を用いることにより行う。対象となる画像から抽出した特徴 \mathbf{v} が歩行者である確率 $P(\mathbf{v}|C_p)$ と非歩行者である確率 $P(\mathbf{v}|C_n)$ を求め、両者の確率を比較することにより認識する。また、それぞれの確率を求めるには、式(10)(11)を用いる。作成した LUT 内に記載してあるそれぞれの特徴の出現確率を利用し、抽出した特徴それぞれの出現確率を調べ掛け合わせることで全体の確率としている。

$$\frac{P(\mathbf{v}|C_p)}{P(\mathbf{v}|C_n)} > \lambda \quad (9)$$

$$P(\mathbf{v}|C_p) = \prod_i P(v_i|C_p) \quad (10)$$

$$P(\mathbf{v}|C_n) = \prod_i P(v_i|C_n) \quad (11)$$

また、これらの確率に対し対数をとることで認識時に乗算を使わずに加算と減算で認識を行うことができる。これにより、認識速度の向上が考えられる。

4.6. Bootstrapping による LUT 更新

作成したばかりの LUT では、高い性能を発揮することはできない。そこで、トレーニング用データベースと更新用データベースを用いて LUT の更新を行い、認識率の向上を行う。

作成した LUT を用いてトレーニング用、更新用データベースの画像を認識する。LUT の更新は、誤認識、未認識した画像の特徴を用いて行う。認識ミスをした画像をトレーニング画像として追加し、新たに LUT を作成する。更新した LUT に対して認識、更新を行う。認識率、誤認識率が設定した条件に達するまでこの処理を繰り返す。

5. 検証

歩行者認識手法の検証は、ROC 曲線による検証とテストデータベースに対する認識検証の 2 種類の検証を行う。

5.1. ROC 曲線による検証

5.1.1. 検証条件

ROC 曲線の測定には LUT 更新で使ったデータベースを使用し、データベースは歩行者 2940 枚、非歩行者 5035 枚の計 7975 枚の画像で構成さ

れている。画像サイズは 64×128 pixel で正規化されている。ROC 曲線は、トレーニングデータにおける認識率の限界を示したものであり、縦軸は認識率、横軸は誤認識率を示している。この曲線が左上に近ければ近いほど精度のいい認識を行えていることになる。

5.1.2. 検証結果

本手法における ROC 曲線は図 7 のようになった。曲線は左上に近く良い結果がでていることがわかる。認識率が約 75% の時に誤認識率は 0% となり、認識率が 100% の時には、誤認識率は約 10% となっている。しかし、認識率 80% 付近では ROC 曲線はほぼ平行になっている。そのため、80% 付近での認識の性能が良くないことがわかる。

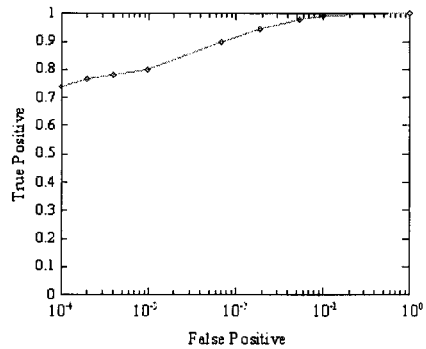


図 7. ROC 曲線

5.2. 認識検証

5.2.1. 検証条件

検証に使用するデータベースは歩行者 1260 枚、非歩行者 2157 枚である。また、認識に使用する LUT はそれぞれ学習を行い、認識率 100%、誤認識率 0% まで向上させたものを使用する。画像サイズは 64×128 pixel であり、すべての画像が車載カメラによって撮影されたものである。また、LUT はトレーニングにより認識率、誤認識率を 95%、2% まで向上させた LUT を使用する。

5.2.2. 検証結果

HOG 特徴を用いた NaiveBayesClassifier による認識結果を表 2 に示し、認識成功例、誤認識例、未認識例を図 8,9,10 に示す。検証用データベースに対して認識を行った結果、認識率、誤認識率はそれぞれ 96.19%、6.49% という結果になった。

歩行者認識の成功結果を見ると、視差がきれいに取れており、マスク画像も正常に作成されていることがわかる。画質が良く、視差もきれいに取れている場合には正確に認識できている。

しかし、図9の未認識例を見ると画像のコントラストが悪く、視差も正確に取得できていないことがわかる。また、画像中の歩行者の位置が端に寄っていることも未認識の原因の1つになっていると考えられる。この未認識に関する対策としては、トレーニングに用いるデータベース内にこのような画像データを取り入れることで未認識は改善されると考えられる。しかし、このような画像データを加えることで、非歩行者との区別が付きづらくなり、誤認識が増加する可能性がある。

また、図10に誤認識例を示す。誤認識例を見ると、中央付近に縦長の物体があり、視差画像も中央付近にきれいに視差が出ている。誤認識をしてしまった物体は縦方向のエッジが多くあり、歩行者の特徴と酷似している部分が多くある。しかし、これらの物体は縦方向のエッジが歩行者と比べ、直線性が強いいため、縦方向に強く出ている直線のエッジに対してフィルタをかけることで、誤認識が防げると考えられる。

表 2. 認識結果

	歩行者	非歩行者	計
検出数	1212	140	1352
未検出数	48	2017	2065
計	1260	2157	3417

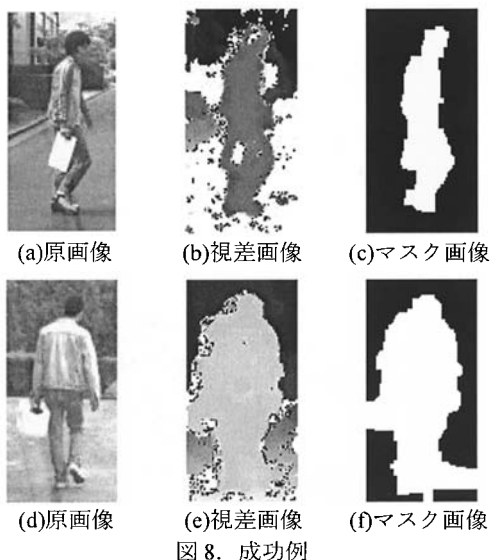


図 8. 成功例

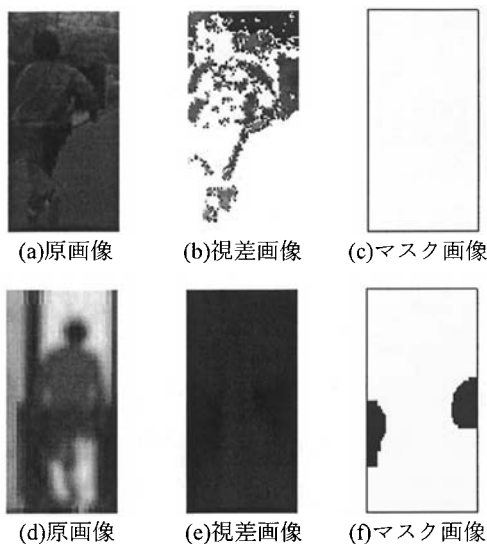


図 9. 未認識例

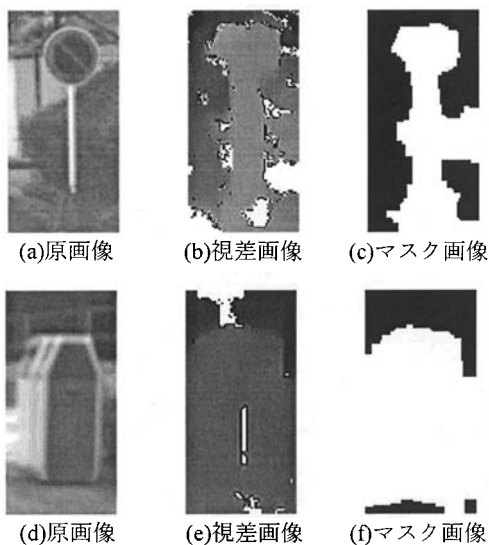


図 10. 誤認識例

6. まとめ

本稿では、ステレオ視と HOG を用いた Naive BayesClassifier による歩行者認識手法を提案した。本提案手法は、原画像から SobelFilter によりエッジ画像を作成し、視差画像から視差値のヒストグラムにより作成したマスク画像を用いて背景エッジの除去を行う。その後、背景エッジを除去したエッジ画像から HOG を用いて HOG 特徴の抽出を行う。抽出した HOG 特徴を特徴量とする NaiveBayesClassifier により歩行者認識を行う

た。その結果、高い認識率を得ることができた。

今後の課題としては、HOGのセルサイズ、ブロックサイズ等のパラメータの変更による認識率の変化の検証、認識速度に対する検証を行う必要がある。また、未認識、誤認識の対策として、トレーニングデータによる認識率の変化の検証や縦方向の直線エッジを除去した場合の誤認識率の調査を行う必要がある。

参考文献

- [1]木嶋啓貴,“センサフュージョンによる歩行者検知に関する研究”,2006年度 修士論文, February 9, 2007.
- [2]松島宏典,“ステレオ視による三次元歩行者認識”,第5回ITSシンポジウム2006,December 6-7, 2006,P1-14.
- [3]F.Suard, A.Rakotomamonjy,A.Bensrhair,A.Broggi,“Pedestrian Detection using Infrared image and Histograms of Oriented Gradients”, Intelligent Vehicle Symposium 2006, June 13-15, 2006, Tokyo, Japan, pp.206-212.