技術論文

# 人を検出するステレオ画像認識システム

## A Stereo Vision System for Human Detection

吉田 睦　山崎 章弘

要旨

本稿では、小型車両への画像認識技術応用を目指し試作した、ステレオビジョンをベースとするシステムについて報告する。本システムは、ステレオ法により距離情報を取得することで障害物候補を抽出し、続いて学習により構築した識別器を用いて、それが人かどうかの識別を行う。条件分岐が比較的少なく並列化しやすいステレオマッチングまでの処理を Field-Programmable Gate Array(FPGA）に、またそれ以降の処理を CPU に、それぞれ配することで、秒間 15 フレームのスループットを実現した。

Abstract

This paper describes a prototype stereo vision-based system for human detection developed with the objective of applying image recognition technology to small vehicles. This system utilizes distance information obtained by the stereo method to detect possible obstacles and then employs a machine-learned classifier to determine if the obstacle candidate is a human being or not. By allocating processing up to the stereo matching stage, with its parallel computational efficiency due to a relatively low degree of conditional branching, to a Field-Programmable Gate Array (FPGA) and the subsequent processing to the CPU, the prototype system yields a throughput of 15 frames per second.

## 1 はじめに

車両のロボット化において、前方障害物あるいは走行可能な領域など、周囲環境の認識は重要な技術であり、さまざまな取り組みが行われている。本稿では、電磁誘導ゴルフカーやバギー車ベースUnmanned Ground Vehicle(UGV)などを含む小型車両へ画像認識技術を応用することを目指し開発したシステムについて述べる。ここではその例のひとつとして、電磁誘導ゴルフカーのための障害物検出を挙げる。

(a) 電磁誘導ゴルフカー　　(b) UGV

図 1　小型車両の例

電磁誘導ゴルフカーは、地中に埋設された誘導線に沿って操舵制御することで自動走行を行う。このため、通常静止物である建物や立ち木などが誘導走行上の障害となることはない。一方で、プレイヤをはじめとする“人”は、自ら移動し誘導コース上に立ち入る可能性がある。これを検出する際に、単に前方の物体を検知するのみのシステムを用いると、カーブ外側の本来支障のない静止物や、勾配変化の大きい走路における前方の地面を検出してしまう場合があるため、画像による人検出が要望される。

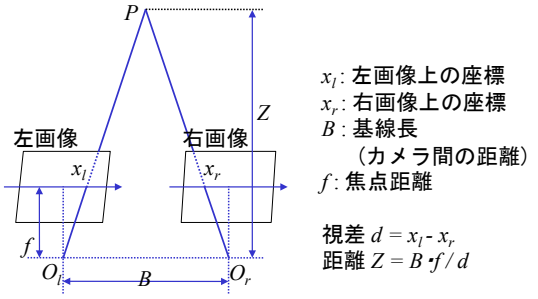



図 2　ステレオビジョンによる距離算出

ステレオビジョンは、通常左右2台のカメラ画像から同一の対象を撮像し、その投影位置の違いから図2に示す三角測量の原理により、対象までの距離を測定するものである。測距手段としては他にもレーダなど能動型のセンサがあるが、画像認識処理により、上述した人検出をはじめとする対象物の認識が可能という利点がある。

## 2 システムハードウェア

本システムでは、ステレオ法により距離情報を取得することで障害物候補の抽出を行い、続いてそれが人かどうかの識別を行う。試作したステレオ画像認識システムのハードウェアを図3に示す。

ステレオカメラは撮像素子にモノクロCMOSを用い、グローバルシャッターにより左右同期キャプチャが可能である。また

(a) ステレオカメラ　　(b) 画像処理ボード
図 3　ステレオ画像認識システムのハードウェア

有効画素数はVGA(640×480)である。出力インタフェースにはLVDSとUSBとを備え、後段装置の構成に柔軟に対応する。

画像処理ボードは左右2枚の画像を入力とし、FPGA(Xilinx社製Spartan-6)にて分岐の少ない前処理、具体的にはレクティフィケーション、エッジ検出、ステレオマッチングなどを、またその後段ではCPU(ルネサス社製SH4A)にて低信頼視差の除去、障害物候補の抽出、人かどうかの識別などを行う。処理の内容詳細については後述する。FPGAとCPUは並列に動作させることで、スループットを向上させることができる。

## 3 処理内容

本システムにおける処理の概要を、図4に示す。

### 3-1. ステレオマッチング

ステレオカメラから取り込まれる左右カメラ画像は、各々レクティフィケーションとエッジ検出とを経て、ステレオマッチングに入力される。

レクティフィケーションとは、左右カメラ画像を図2に示したような理想的な状態に近づけるよう変換するものである。レクティフィケーションに先立ってはキャリブレーションが必要となる。キャリブレーションにおいては、カメラの内部パラメータとして焦点距離のばらつき、レンズ歪み、投影面の傾き、また外部パラメータとして取り付け位置、姿勢のばらつきなどを推定し、変換マップを生成する。オンラインの処理ではこのマップを用い、左右カメラを平行等位とみなせるよう、両画像のエピポーラ線の高さを水平に揃えている。なおエピポーラ線とは、一方のカメラと注目点とがなす直線の、もう一方のカメラへの投影である。これにより、ステレオマッチングにおける探索を水平方向のみに限定し、高速に処理することができる。

ステレオマッチングはエッジ画像に対するブロックマッチングを用いる。ブロックマッチングは左右の対応点を求めるために、その周辺を含むブロック同士で類似度合いを評価するものである。また視差画像の生成は、左画像基準、右画像基準、それぞれについて行う。両者に差異がある画素、あるいはエッジ強度が低い画素については、視差の信頼度は低いものとして除去する。

### 3-2. 障害物候補の抽出

障害物候補の抽出においては、まず車両前方に検出対象空間を設定し、かかる空間内にてある程度の数を有する点群クラスタを抽出する。ステレオマッチングにおいては画角内のあらゆる撮像対象が2.5Dデータに変換されるが、その中には地面や無限遠点など、およそ検出対象とはならないものも多いため、これらについてはあらかじめ除外しておく。

### 3-3. 人の識別

抽出された障害物候補が人かどうかを識別するには、レクティフィケーション後の可視光画像に対する人検出処理を用いる。可視光画像の見えによる(appearance-based)人検出という課題に対しては、これまでさまざまな手法が提案されている[1][2]。

特に低レベルな特徴の抽出においては、ここではその代表例のひとつであるHistograms of Oriented Gradients(HOG)

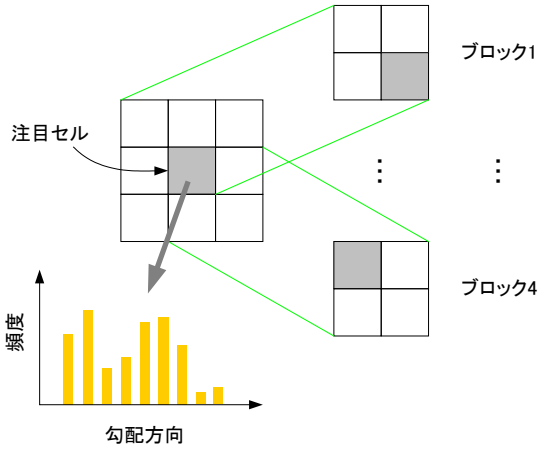



図 5　HOG におけるセルとブロック

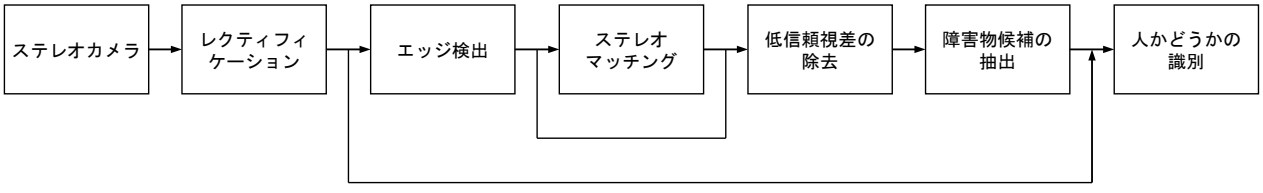



図 4　処理の概要

アルゴリズム 1　SODA-Boosting

入力:

学習サンプル $(x_1, y_1),...,(x_n, y_n)$

ここで $y_i = -1,1$　はそれぞれ陰性と陽性をあらわすラベル

1. 初期化：

   $y_i = -1,1$　に対するそれぞれのウェイト

   $$w_{1,i} = \frac{1}{2n^-}, \frac{1}{2n^+}$$

   ここで　$n^-, n^+$　はそれぞれ陰性サンプルと陽性サンプルの総数

2. For $t = 1,...,T$：

   (a) ウェイトを正規化: $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$

   (b) FLD 特徴による弱識別器 $f_{FLD}(x)$ 及び
   そのエラーレイト $\varepsilon_{\mathrm{FLD}}$ を計算

   (c) MRC+特徴による弱識別器 $f_{\mathrm{MRC}+}(x)$ 及び
   そのエラーレイト $\varepsilon_{\mathrm{MRC}+}$ を計算

   (d) MRC-特徴による弱識別器 $f_{\mathrm{MRC}-}(x)$ 及び
   そのエラーレイト $\varepsilon_{\mathrm{MRC}-}$ を計算

   (e) 最小エラーレイト $\varepsilon_t$ の弱識別器
   $f_t(x) \in \{f_{FLD}, f_{MRC+}, f_{MRC-}\}$ を選択

   (f) ウェイトを更新:

   $$w_{t+1,i} \leftarrow w_{t,i} \exp[-\alpha_t y_i f_t(x_i)]$$

   ただし $\alpha_t = \frac{1}{2} \ln \frac{1-\varepsilon_t}{\varepsilon_t}$

出力:

強識別器 $F(x) = sign\left(\sum_{t=1}^{T} \alpha_t f_t(x)\right)$

を用いる。HOGは所定の領域(セル)内輝度の勾配方向ヒストグラムを複数の異なる局所領域(ブロック)内にて正規化することで、照明環境や対象の形状、姿勢、向きなどの変動に対しても、比較的頑健に対象形状の特徴を捉えることができる。図5にブロックサイズ2×2セル、勾配方向の分割数9としたときのセルとブロックの正規化概念図を示す。

また、上記特徴を効果的に用いて対象を識別するには、学習による識別器の構築が有効である。ブースティングによる学習では、比較的単純な特徴を識別する弱識別器を複数個それぞれの重要度にあわせて重みをつけて組み合わせることにより、より強い識別器を構築する。ここではSODA-Boosting[3]を用いる。SODA-Boostingにおいては、フィッシャーの線形判別(FLD)あるいはMaximal Rejection Classifier（MRC）による射影変換を弱識別器に用いる。FLDは、クラス内変動に対するクラス間変動の比を最大にするような変換を求めるものであり、両クラスは均等に扱われる。一方MRCは、図6に示すように、対象とする“Target”クラスがその他の広範な“Clutter”クラスに取り囲まれているような分布を想定したものである。SODA-Boostingにおいては、2クラスのうちひとつを“Target”とするMRC-positive特徴と、逆に“Clutter”とするMRC-negative特徴をそれぞれ用いる。今回のように対象を検出する問題においても、実際、MRC-negative はMRC-positiveと遜色ない割合で選択される。

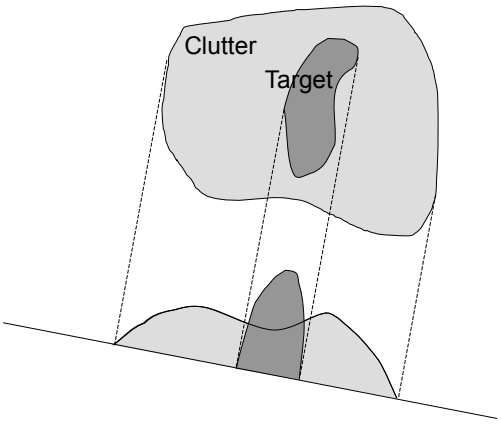



図 6　ターゲットイメージと MRC

なお、カメラから上記障害物候補までの距離を利用して識別器の処理ウィンドウサイズを限定することで、処理量の削減が可能である。HOGを用いた対象物体の検出において、画像内のどの位置にどのような大きさで対象が写るかを限定できない場合にはIntegral Histogramが処理の高速化に有効であるが、ここではステレオマッチングにより得られた距離から処理ウィンドウサイズを限定するため、特に用いない。

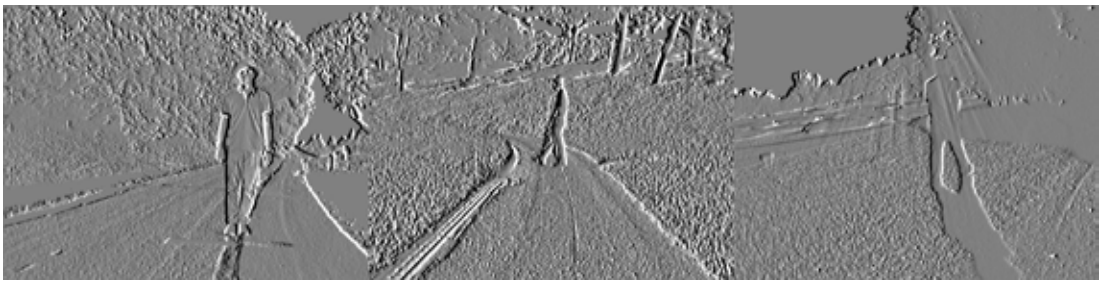
## 4 実験結果

### 4-1. 人検出単体でのプリテスト

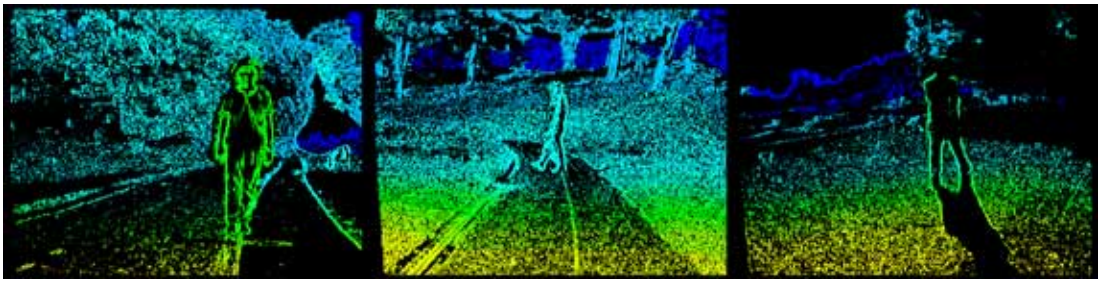
先述したハードウェアへの実装に先立ち、人検出器としてのプリテストを行った。用いた画像データは実際のゴルフコースにおいて、カメラを車両に搭載し走行しながら撮影したもので、学習サンプル、評価サンプルには重複のないそれぞれ1万枚以上の画像を抽出して用いた。HOGの構成は、最小セルサイズを16×16ピクセル、ブロックサイズを2×2セル、勾配方向の分割数を9、検出器の処理ウィンドウサイズを4×8セルとした。これら設定については、検出性能と処理時間とのバランスの良いところを事前実験から選択した。検出性能の評価には誤検出の少なさ(Precision)と未検出の少なさ(Recall)を考慮する必要があり、ここではその調和平均であるF値を用いた。結果、F値は約0.96であり、VGA画像 1 枚の全探索(約7,500ウィンドウ)に要する処理時間は、Intel社製Core2 Quad 3.0GHzを搭載したPCによる実行で、約90msであった。

ここで用いられた検出器の処理ウィンドウサイズは、4×8セルと先行事例[1]のパラメータと比して少ないが、セル1つあたりのカバー範囲を相対的に大きくすることで、ラスタスキャ

(a)　レクティファイ画像



(b)　エッジ画像



(c) 視差画像

(d)　人検出結果表示

図 7　画像処理例

ン時の位置及びスケールのステップを大きくすることが可能となり、誤検出の抑制と処理速度の向上に寄与することができる。

### 4-2. 統合テスト

図4の処理をシステムとして統合したテストを行った。図7に本システムによる各処理結果について示す。順にレクティファイ画像、エッジ画像、視差画像、人検出結果表示であり、いずれも左カメラ画像を元にしている。視差画像は暖色が近方、寒色が遠方を表している。黒い領域は視差信頼度が低く除去した部分である。人検出結果表示においては、視認しやすいよう人検出結果枠内をハイライト表示している。検出結果枠下の数字は人までの距離を表している。

図8には、ステレオ法により障害物候補として抽出された対象における人検出のPrecision-Recallを示す。F値は約0.95であった。

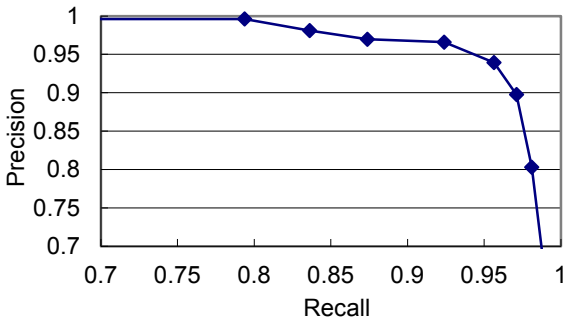



図 8　人かどうかの識別についての Precision-Recall

このように自然な環境中に人物が写っている状況でシステムを動作させ処理速度を計測した結果、スループットは秒間15フレームとなった。各処理の所要時間を図9に示す。相対的にCPU側の処理に時間がかかっており、FPGA側では小さくないウェイトが生じている。CPU側の処理量をさらに削減することで両者を均衡させることが、当面の今後の課題である。

## 5　まとめ

小型車両への画像認識技術応用を目指して試作したステレオ画像認識システムについて紹介した。本システムは、条件分岐の少ない前処理、具体的にはレクティフィケーション、エッジ検出、ステレオマッチングなどをFPGAにて行い、それ以降の処理、つまり低信頼視差の除去、障害物候補の抽出、人かどうかの識別などをCPUにて行う構成とした。人かどうかの識別にはHOGとSODA-Boostingとを組み合わせ、検出性能と処理時間とのバランスを図った。ステレオ法により障害物候補として抽出された対象における人検出で、約0.95のF値を得た。FPGAとCPUとは並列に動作させることで、秒間15フレームのスループットを実現した。今後はCPU側の処理量をさらに削減するとともに、距離画像からも人の識別のための特徴量を抽出していくことを検討する。

## 6　おわりに

本研究を進めるにあたり有益な助言をいただいた中部大学の藤吉弘亘教授に、感謝の意を表する。

また実験データの取得に協力いただいた葛城ゴルフ倶楽部に、感謝の意を表する。

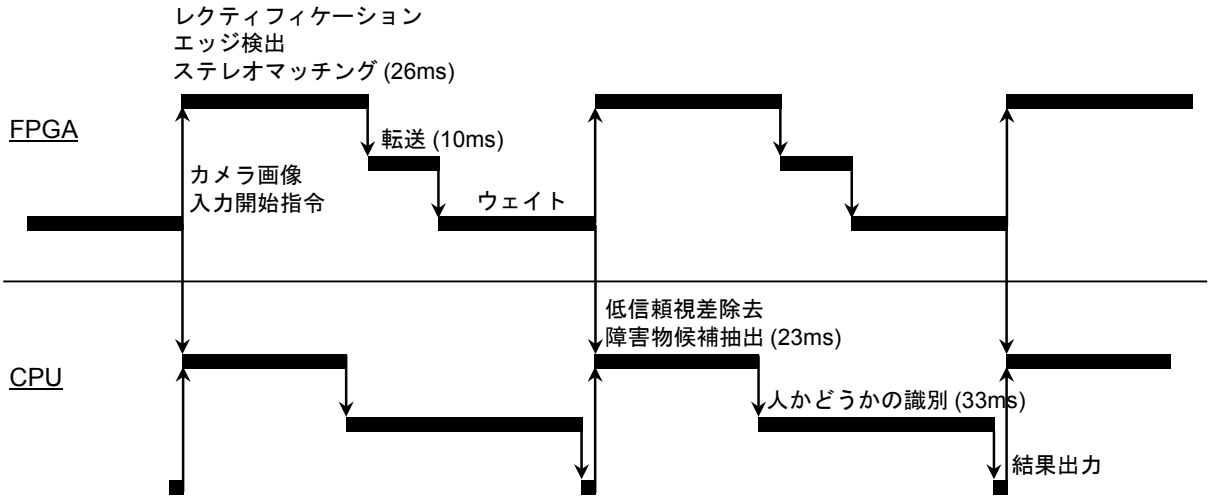



図 9　各処理の所用時間

## 7 参考文献

[1] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. IEEE Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.
[2] 松島千佳, 山内悠嗣, 山下隆義, 藤吉弘亘. Relational Binarized HOG特徴量とReal AdaBoostによるバイナリ選択を用いた物体検出. 第13回画像の認識・理解シンポジウム, 2010.
[3] X. Xu and T. S. Huang. SODA-Boosting and Its Application to Gender Recognition. IEEE Workshop on Analysis and Modeling of Faces and Gestures, 2007.

**■著者**

**吉田 睦**
Makoto Yoshida
技術本部
研究開発統括部
イノベーション研究部

**山﨑 章弘**
Akihiro Yamazaki
技術本部
研究開発統括部
イノベーション研究部