

基線長が短い車載ステレオカメラのための障害物検出手法

和泉 圭祐^{†a)} 三浦 衛[†] 伊藤 康一^{†b)} 青木 孝文[†]

Obstacle Detection Method for In-Vehicle Narrow-Baseline Stereo Camera

Keisuke IZUMI^{†a)}, Mamoru MIURA[†], Koichi ITO^{†b)}, and Takafumi AOKI[†]

あらまし 本論文では、(i) 位相限定相関法 (Phase-Only Correlation: POC) に基づく高精度なステレオ画像の対応付けを用いた視差画像の生成と、(ii) 微分処理に基づく等視差領域のクラスタリングを組み合わせた障害物検出手法を提案する。POC に基づくステレオ画像の対応付けを用いることで、基線長が短いステレオカメラで撮影した画像に対しても、密で正確な視差画像を生成することができる。カメラに正対する障害物が等しい視差をもつ領域で表現されることを利用して、微分処理に基づく簡便なクラスタリングにより障害物を検出する。基線長が短いステレオカメラで撮影した動画像を用いた性能評価実験を通して、提案手法が従来手法と比べて障害物を過不足なく検出できることを示す。

キーワード 運転支援システム, 障害物検出, ステレオビジョン, 位相限定相関法, 視差クラスタリング

1. ま え が き

近年、自動車の安全性や利便性を向上させるために、運転支援システムに関する研究開発が盛んに行われている [1]。運転支援システムは、センサで走行環境に関する情報を取得し、それらの情報に基づいて運転を制御したり、危険を通知したりすることで、運転を補助する。運転支援システムはさまざまな機能で構成されているが、その中でも、障害物検出は、最も重要な基本機能である [2]。

障害物検出を行うためには、センサで走行環境の情報を取得する必要がある。レーザやレーダなどのような能動的なセンサがよく使われているが、高精度に測距できる一方で、障害物の種類を判別できなかつたり、検出の範囲が限られていたりする問題がある。本論文では、障害物の検出や測距だけでなく、検出した障害物の種類の判別や、白線、路面、標識、信号などの走行環境の認識も行うことが可能であるステレオカメラに着目する。

ステレオカメラを用いた障害物検出は、(i) ステレ

オ画像の対応付けによる視差画像の生成と、(ii) 視差画像のクラスタリングで構成される。障害物の領域を過不足なく正確に検出することができれば、障害物の認識や追従などの処理を高精度に行うことが可能となる [2]。正確な障害物検出のためには、ステレオ画像の対応付け及び視差画像のクラスタリングの両方で高精度な手法が必要である。

ステレオカメラを用いて障害物検出を行うために、基線長が長いステレオカメラが用いられることが多い。基線長を長くすることで、ステレオ画像の対応付けにおける誤差が三次元計測の精度に与える影響を小さくすることができる。そのため、画像対応付けの精度ではなく、処理時間を優先した手法を用いることができるので、SAD (Sum of Absolute Differences) や SSD (Sum of Squared Differences) のような手法が用いられている。しかし、基線長を長くすると、ステレオカメラの設置位置が車両の両端に限定されてしまつたり、車両に後付けができなかつたりする問題がある。一方で、基線長が短いステレオカメラは、設置位置が限定されないため、利便性が高い。ただし、高精度な障害物検出手法、特にサブピクセル分解能で高精度に視差画像を生成することができる手法が必須である。現在までに報告されている多くの手法は、基線長が長いステレオカメラを前提としており、基線長が短いステレオカメラを用いることが考慮されていない。

[†] 東北大学大学院情報科学研究科, 仙台市
Graduate School of Information Sciences, Tohoku University, 6-6-05 Aramaki Aza Aoba, Aoba-ku, Sendai-shi, 980-8579 Japan

a) E-mail: izumi@aoki.ecei.tohoku.ac.jp

b) E-mail: ito@aoki.ecei.tohoku.ac.jp

もし基線長が短いステレオカメラを用いて高精度な視差画像を生成することができれば、小型のステレオカメラユニットを利用することができる。小型ユニットであれば、取り付けの自由度が上がるだけでなく、設置や調整などにかかるメンテナンスコストを低減することができる。そのため、基線長が短いステレオカメラに対する需要は高い。

具体的な例を用いて、ステレオカメラの基線長の長さ、測距精度、対応付け精度の関係について説明する。図1に、100m先にある物体をステレオカメラで計測するときの、基線長の長さ、測距の誤差、対応付けの誤差の関係を示す。この関係は、ステレオビジョンの原理から容易に導出することができる[3]。基線長が短くなるほど、対応付けの精度が距離の精度に与える影響が大きくなる。例えば、基線長の長さが500mmのステレオカメラを用いると、0.5画素の対応付けの誤差が7mの測距の誤差になる。一方で、100mmのステレオカメラを用いると、0.2画素の対応付け誤差が12mの測距の誤差になってしまう。ステレオ画像の高精度かつ密な対応付け手法として、Semi-Global block Matching (SGM) [4] や、Efficient LArge-Scale stereo matching (ELAS) [5] が提案されている。図2に、The KITTI Vision Benchmark Suite [6] のステレオ画像に対して、基線長が長い場合(540mm)と短い場合(60mm)にSGM及びELASを用いて生成した視差画像を示す。なお、基線長が長い場合は車両の屋根(ルーフ)の中央と端にそれぞれカメラを設置した場合であり、基線長が短い場合はカメラを隣接して設置した場合である。基線長が長い場合に比べて、基線長が短い場合は、SGMとELASのいずれも誤対応点の数が多く、正確に視差が求められていない。これらのことより、基線長が短いステレオカメラで撮影したステレオ画像からでも、障害物検出ができるような視差画像を生成することができる対応付け手法が必要であることが分かる。

上記の問題が解決され、高精度な視差画像を生成することができれば、視差画像上でカメラに正対する障害物が等視差領域として表現されるので、視差画像をクラスタリングすることで障害物を検出することができる。これまでに、簡便で高速なV-disparity画像に基づく手法が提案されている[7]~[10]。V-disparity画像に基づく手法は、視差画像から垂直及び水平方向のそれぞれに対して視差のヒストグラムを生成し、ヒストグラム上で頻度の高い領域を障害物として検出す

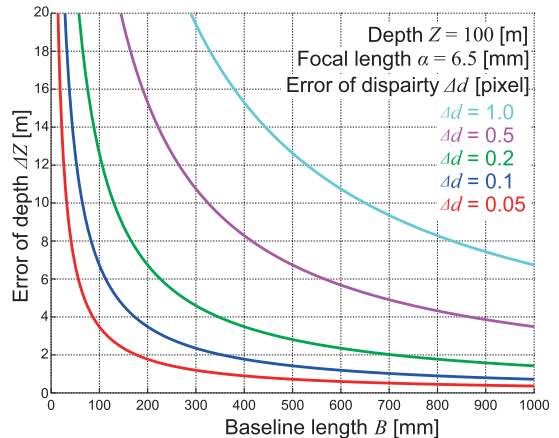


図1 基線長の長さ、距離計測の誤差、対応付けの誤差の関係

Fig. 1 Relation among length of stereo camera baseline, error of distance measurement and error of correspondence matching.

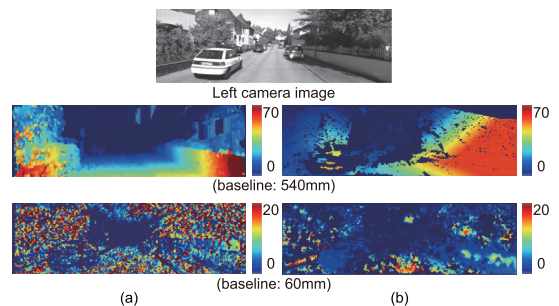


図2 (a) SGM 及び (b) ELAS によって生成された視差画像 (上段: 基線長が長い場合, 下段: 基線長が短い場合)

Fig. 2 Disparity maps generated by (a) SGM and (b) ELAS from the stereo image (upper: wide-baseline setting, lower: narrow-baseline settings).

る。基線長が短くなると、画素単位の視差を基本としているので、障害物領域とそれ以外の領域が同じビンに投票されてしまう。ヒストグラムのビンの幅をサブピクセル単位にすることで、障害物領域とそれ以外の領域を区別することができるが、煩雑なパラメータの設定及び最適化が必要になる。また、V-disparity画像に基づく手法は、ハードウェア実装を目的として、疎な視差画像を対象としている。そのため、近年に提案されている対応付け手法で生成された密な視差画像から必ずしも正確に障害物を検出できるとは限らない。

本論文では、(i) 位相限定相関法 (Phase-Only Correlation: POC) に基づく高精度なステレオ画像の対

応付けを用いた視差画像の生成と、(ii) 微分処理に基づく等視差領域のクラスタリングを組み合わせた障害物検出手法を提案する。POC に基づくステレオ画像の対応付けは、誤対応が少なく、サブピクセル分解能で高精度に視差を推定することが可能である [11]。高精度で密な視差画像を得ることができれば、微分処理に基づく簡便なクラスタリングで高精度な障害物検出が可能となる。提案手法では、視差画像を変換して得られる深度画像に対して、垂直及び水平方向のそれぞれについて微分処理を施し、それらの結果を統合することで等視差領域を検出する。クラスタリングのパラメータは、実空間上の距離に基づいて容易に決定することができるため、環境によらず設定することができる。性能評価実験では、基線長が短いステレオカメラで撮影した動画像を用いて、提案手法が従来手法と比べて障害物を過不足なく検出できることを示す。なお、本論文の実験では、定量的な評価のために、自車両の前方を走行している車両を障害物の対象とするが、その他の障害物も同様に検出することが可能である。また、提案手法の処理時間の評価を通して、リアルタイム処理の可能性について考察する。

2. 位相限定相関法に基づく密な視差画像の生成

提案手法で用いる位相限定相関法 (Phase-Only Correlation: POC) [12]~[14] に基づく視差画像生成について述べる。POC に基づくステレオ画像の対応付けは、サブピクセル分解能で高精度に視差を推定することができるため、基線長が短いステレオカメラにおいても利用することができる [11]。また、基準点の配置を任意に設定することができるため、用途に応じて視差推定の領域を容易に限定することも可能である。これらの特長は、車載向け小型のステレオカメラユニットにとって有効である。以下では、POC の基本原理及び対応付け手法について説明する。なお、画像の対応付けは、二次元の探索問題であるが、ステレオ画像の場合は、エッジ線に沿った一次元の探索問題として扱うことができる。そこで、以下の説明では、ステレオ平行化を行い、水平方向のマッチングを行うこととする。

2.1 位相限定相関法

N 点の二つの一次元画像信号 $f(n)$ 及び $g(n)$ が与えられたとき、これらの信号間の位置ずれ (平行移動量) を求めることを考える。ただし、 $n = -M, \dots, M$

($M > 0$) であり、 $N = 2M + 1$ である。 $f(n)$ の一次元離散フーリエ変換 (Discrete Fourier Transform: DFT) $F(k)$ を次式で計算する。

$$F(k) = \sum_{n=-M}^M f(n)W_N^{kn} = A_F(k)e^{j\theta_F(k)} \quad (1)$$

ここで、 $k = -M, \dots, M$ であり、 $W_N = e^{-j\frac{2\pi}{N}}$ は回転因子、 $A_F(k)$ は振幅、 $\theta_F(k)$ は位相である。 $g(n)$ の一次元 DFT $G(k)$ も同様に定義できる。このとき、正規化パワースペクトル $R(k)$ を次式で計算する。

$$R(k) = \frac{F(k)\overline{G(k)}}{|F(k)G(k)|} = e^{j(\theta_F(k) - \theta_G(k))} \quad (2)$$

ここで、 $\overline{G(k)}$ は $G(k)$ の複素共役、 $\theta_F(k) - \theta_G(k)$ は位相差である。位相限定相関関数 (POC 関数) $r(n)$ は、 $R(k)$ の一次元逆離散フーリエ変換 (Inverse Discrete Fourier Transform: IDFT) であり、次式で定義される。

$$r(n) = \frac{1}{N} \sum_{k=-M}^M R(k)W_N^{-kn} \quad (3)$$

二つの画像信号が類似しているとき、POC 関数は鋭い相関ピークを有する。一方で、二つの画像信号が類似していないとき、その相関ピークは著しく低下する。相関ピークの高さは、二つの画像信号の類似度に、相関ピークの位置座標は、二つの画像信号の平行移動量に相当する。POC を用いて高精度に平行移動量の推定を行うために、(i) 画像端での信号の不連続性を解消するための窓関数の適用、(ii) エイリアシングやノイズ、ぼけ、ひずみの影響を抑制するためのスペクトル重み付け、(iii) 複数の一次元信号の利用、(iv) 相関ピークモデルのフィッティングを用いる。なお、(iv) を用いることで、実際に計算された POC 関数の離散数値データに対して、解析的に導出された相関ピークモデルをフィッティングすることにより、二つの画像信号間の平行移動量をサブピクセル精度で求めることが可能である。(i)~(iv) の高精度化手法の詳細については、文献 [11], [15] を参考にされたい。

2.2 位相限定相関法に基づくステレオ画像の対応付け

位相限定相関法による画像信号のマッチングを局所的なブロックマッチングに適用し、画像ピラミッドによる階層探索と組み合わせることで、高精度なステレ

オ画像の対応付けを行うことができる。左カメラ画像 $I(u, v)$ 上の基準点の座標 $\mathbf{m} = (m_u, m_v) (\in \mathbf{Z}^2)$ を入力とする。ここで、 \mathbf{Z} は整数の集合である。このときに、右カメラ画像 $J(u, v)$ 上の対応点の座標 $\mathbf{q}(\mathbf{m}) = (q_u, q_v) (\in \mathbf{R}^2)$ をサブピクセル精度で求めることを考える。ここで、 \mathbf{R} は実数の集合である。なお、平行ステレオでは、視差が水平方向の平行移動に限られるので、 $m_v = q_v$ である。以下に、ステレオ画像の対応付けの詳細な手順を示す。

Step 1: 画像 $I(u, v)$ 及び $J(u, v)$ をそれぞれ 2^{-l} 倍だけ縮小し、階層画像 $I^l(u, v)$ 及び $J^l(u, v)$ を次式で生成する。

$$I^l(u, v) = \frac{1}{4} \sum_{i_u=0}^1 \sum_{i_v=0}^1 I^{l-1}(2u + i_u, 2v + i_v)$$

$$J^l(u, v) = \frac{1}{4} \sum_{i_u=0}^1 \sum_{i_v=0}^1 J^{l-1}(2u + i_u, 2v + i_v)$$

ただし、 $l = 1, 2, \dots, l_{\max} - 1$, $I^0(u, v) = I(u, v)$, $J^0(u, v) = J(u, v)$ である。処理時間を短縮するために、最上層 l_{\max} では、対応点候補座標を決めるだけで、平行移動量を推定しないので、階層画像を生成しない。

Step 2: 最上層 l_{\max} において、基準点の座標 $\mathbf{m} = (m_u, m_v)$ は、 $(\lfloor 2^{-l_{\max}} m_u \rfloor, \lfloor 2^{-l_{\max}} m_v \rfloor)$ に対応する。ここで、 $\lfloor z \rfloor$ は、実数値 z を負の無限大方向の最も近い整数に丸めることを示す。最上層 l_{\max} における対応点候補は、基準点と同じ座標をもつと仮定し、次式で与えられる。

$$\mathbf{q}^{l_{\max}} = (\lfloor 2^{-l_{\max}} m_u \rfloor, \lfloor 2^{-l_{\max}} m_v \rfloor) \quad (4)$$

ただし、上記仮定を満たすように1次元信号の長さ N と階層数 l_{\max} を適切に設定する必要がある。POCを用いて高精度に平行移動量を推定するために、ハニング窓を信号に適用する。そのため、推定できる平行移動量の範囲は、ハニング窓の半値幅を考慮すると、 N 画素の1次元信号に対して $\pm \lfloor \frac{N}{4} \rfloor$ 画素となる。最上層（第 l_{\max} 層）では対応点候補座標を決定するだけなので、第 $l_{\max} - 1$ 層において N 画素の信号間の平行移動量を推定すると、原画像（第0層）上で $\pm \lfloor \frac{N}{4} \rfloor 2^{l_{\max}-1}$ 画素の範囲の平行移動量を推定することに相当する。以上より、想定され得る画像間の最大視差を d_{\max} とすると次式を満たす N と l_{\max} を用いる必要がある。

$$\left\lfloor \frac{N}{4} \right\rfloor 2^{l_{\max}-1} \geq d_{\max} \quad (5)$$

$l \leftarrow l_{\max} - 1$ とし、次のステップに移る。

Step 3: 第 l 階層において、基準点の座標は、 $(\lfloor 2^{-l} m_u \rfloor, \lfloor 2^{-l} m_v \rfloor)$ に対応する。このとき、第 l 階層の右カメラ画像 J^l 上での対応点候補は、上位階層の対応点候補 \mathbf{q}^{l+1} を用いて $2\mathbf{q}^{l+1}$ で与えられる。左右の画像 I^l 及び J^l からそれぞれ $(\lfloor 2^{-l} m_u \rfloor, \lfloor 2^{-l} m_v \rfloor)$ 及び $2\mathbf{q}^{l+1}$ を中心として、水平方向に N 点の一次元信号を垂直方向に L ラインだけ切り出して画像ブロックを構成する。POCを用いて、切り出した画像ブロック間のピクセル精度の平行移動量 δ^l を推定する。ピクセル精度での平行移動量推定では、2.1で述べた高精度化手法のうち、(i), (ii), (iii)を用いる。なお、サブピクセル精度で推定することもできるが、縮小された画像であるため、そもそもサブピクセル精度で推定する必要がなく、処理時間を減らすためにもピクセル精度で推定する。平行移動量 $\delta^l = (\delta_1^l, 0)$ を考慮して、第 l 階層の対応点 \mathbf{q}^l を次式で求める。

$$\mathbf{q}^l = 2\mathbf{q}^{l+1} + \delta^l \quad (6)$$

Step 4: $l \leftarrow l - 1$ として、 $l \geq 0$ である限り、Step 3からStep 4を繰り返す。

Step 5: 以上の処理により得られるピクセルレベルの対応点を $\mathbf{q}(\mathbf{m}) = \mathbf{q}^0$ とする。左右の画像 I^0 及び J^0 から、それぞれ \mathbf{m} 及び $\mathbf{q}(\mathbf{m})$ を中心として画像ブロックを切り出す。POCを用いて、切り出した画像ブロック間のサブピクセル精度の平行移動量 δ を推定する。サブピクセル精度での平行移動量推定では、2.1で述べた高精度化手法のうち、(i), (ii), (iii), (iv)を用いる。推定した平行移動量 $\delta = (\delta_1, 0)$ を考慮して、対応点 $\mathbf{q}(\mathbf{m})$ を次式で更新する。

$$\mathbf{q}(\mathbf{m}) = \mathbf{q}(\mathbf{m}) + \delta \quad (7)$$

全ての基準点に対して以上の処理を行うことで、ステレオ画像間をサブピクセル精度で密に対応付けることができる。

2.3 視差画像の生成

ステレオ画像から密な視差画像を生成するために、左カメラ画像 $I(u, v)$ の全画素を基準点 \mathbf{m} とし、2.2の対応付け手法を用いて、右カメラ画像 $J(u, v)$ 上の対応点 \mathbf{q} を求める。基準点の座標 \mathbf{m} 及び対応点の座標 \mathbf{q} から視差画像 $d(\mathbf{m})$ を次式で求める。

$$d(\mathbf{m}) = (\mathbf{m} - u_0) - (\mathbf{q}(\mathbf{m}) - u'_0) \quad (8)$$

ここで、 u_0 及び u'_0 は、それぞれ左カメラ及び右カメラにおける画像中心の水平座標を示す。

3. 微分処理に基づく等視差領域のクラスタリング

視差画像から障害物を検出するためのクラスタリングについて述べる。提案するクラスタリング手法は、U-disparity 画像を用いたマスク画像の生成及び等視差領域のクラスタリングで構成される。まず、U-disparity 画像を用いて障害物がない領域を検出し、誤検出を防ぐためのマスクを作成する。次に、視差画像から深度画像に変換し、縦方向及び横方向に微分処理に基づいてクラスタリングを行うことで障害物領域を検出する。

3.1 U-disparity 画像を用いたマスク画像の生成

U-disparity 画像は、縦軸を視差 d 、横軸を水平方向の座標 u とする座標平面上に、視差画像の垂直方向のヒストグラムをマッピングすることで得られる [9]。視差画像において、障害物領域は垂直方向に視差が変化していないが、その他の領域は垂直方向に視差が変化している。これを U-disparity 画像に置き換えると、障害物領域のように垂直方向に視差が変化していない領域は値が大きくなるが、その他の領域のように垂直方向に視差が変化している領域は値が小さくなる。これより、U-disparity 画像にしきい値処理を施すことで、障害物がない領域を検出することが可能である。

図 3 に、U-disparity 画像を用いたマスク画像の生成の例を示す。視差画像 $d(u, v)$ (図 3 (b)) に対し

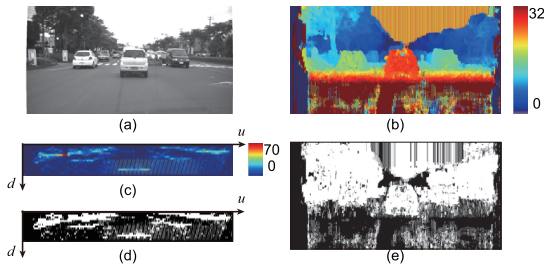


図 3 U-disparity 画像を用いたマスク画像の生成：(a) 左カメラ画像、(b) 視差画像、(c) U-disparity 画像、(d) 2 値化した U-disparity 画像、(e) マスク画像

Fig. 3 Mask generation using U-disparity image: (a) left camera image, (b) disparity image, (c) U-disparity image, (d) binarized U-disparity image and (e) generated mask.

て、垂直方向のヒストグラムを求め、U-disparity 画像 $U_o(u, d)$ を生成する (図 3 (c))。次式に基づいて、U-disparity 画像 $U_o(u, d)$ を二値化し、 $U_b(u, d)$ を求める (図 3 (d))。

$$U_b(u, d) = \begin{cases} 1 & U_o(u, d) \geq th(d) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

しきい値 $th(d)$ は、U-disparity 画像 $U_o(u, d)$ 上の各行について設定する。本論文では、次式を用いてしきい値を決定する。

$$th(d) = \frac{1}{15} \max_u \{U_o(u, d)\} \quad (10)$$

最後に、次式に基づいて、障害物のない領域を示すマスク M_r を生成する (図 3 (e))。

$$M_r(u, v) = \begin{cases} 1 & U_b(u, d(u, v)) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

3.2 等視差領域のクラスタリング

等視差領域のクラスタリングは、視差画像 $d(u, v)$ ではなく、深度画像 $Z(u, v)$ 上で行う。校正済みのステレオカメラであれば、視差画像 $d(u, v)$ を次式により深度画像 $Z(u, v)$ に変換することができる。

$$Z(u, v) = \frac{\alpha B}{d(u, v)} \quad (12)$$

ここで、 α はカメラの焦点距離、 B は基線長である。深度画像上でクラスタリングを行うことで、クラスタリングのパラメータを実空間上の距離に基づいて決定することができる。等視差領域のクラスタリングは、(i) 垂直方向のクラスタリング、(ii) 水平方向のクラスタリング、(iii) マスク画像の統合と閉領域の検出で構成される。図 4 に等視差領域のクラスタリングの例を示す。なお、図 4 (a) の深度画像は、3.1 で生成したマスク画像を適用し、障害物がない領域の深度を 0 としている。

(i) 垂直方向のクラスタリング

深度画像 $Z(u, v)$ の垂直方向の一次元微分を次式で計算する。

$$\frac{\partial}{\partial v} Z(u, v) = Z(u, v+1) - Z(u, v) \quad (13)$$

垂直方向に同じ深度をもつ連続する画素を次式により検出し、マスク画像 $M_v(u, v)$ を生成する (図 4 (b))。

$$M_v(u, v) = \begin{cases} 1 & \left| \frac{\partial}{\partial v} Z(u, v) \right| \leq \sigma_v \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

式 (14) は、垂直方向に隣接する画素間の深度が σ_v mm 未満の連続する画素を一つのクラスタとすることを意味する。

(ii) 水平方向のクラスタリング

(i) と同様の処理を水平方向にも施す。深度画像 $Z(u, v)$ の水平方向の一次元微分を次式で計算する。

$$\frac{\partial}{\partial u} Z(u, v) = Z(u + 1, v) - Z(u, v) \quad (15)$$

水平方向に同じ深度をもつ連続する画素を次式により検出し、マスク画像 $M_u(u, v)$ を生成する (図 4(c))。

$$M_u(u, v) = \begin{cases} 1 & |\frac{\partial}{\partial u} Z(u, v)| \leq \sigma_u \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

式 (16) は、水平方向に隣接する画素間の深度が σ_u mm 未満の連続する画素を一つのクラスタとすることを意味する。

(iii) マスク画像の統合と平領域の検出

3.1 で求めた障害物のない領域を示すマスク $M_r(u, v)$ と (i) 及び (ii) で求めたマスク画像 $M_v(u, v)$ 及び $M_u(u, v)$ から障害物検出のためのマスク画像

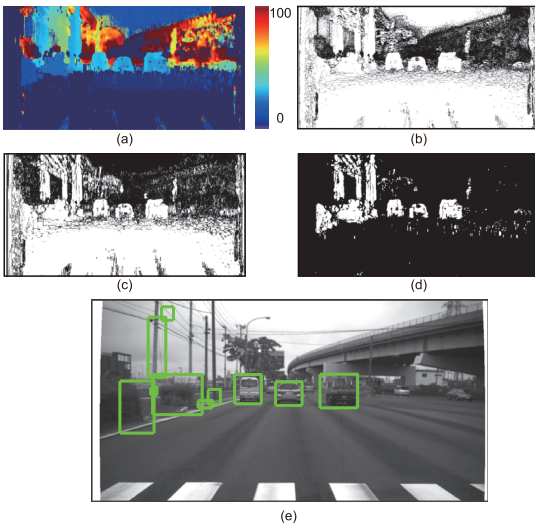


図 4 等視差領域のクラスタリングに基づく障害物検出：(a) 深度画像 $Z(u, v)$, (b) 垂直方向のマスク画像 $M_v(u, v)$, (c) 水平方向のマスク画像 $M_u(u, v)$, (d) 統合したマスク画像 $M(u, v)$, (e) 障害物検出の結果

Fig. 4 Obstacle detection using equi-disparity clustering: (a) depth image $Z(u, v)$, (b) mask image in the vertical direction $M_v(u, v)$, (c) mask image in the horizontal direction $M_u(u, v)$, (d) integrated mask image $M(u, v)$ and (e) result of obstacle detection.

$M(u, v)$ を次式により求める。

$$M(u, v) = M_r \cap M_u \cap M_v \quad (17)$$

マスク画像 $M(u, v)$ 上の閉領域のうち、一定の大きさの領域を障害物として検出する (図 4 (e))。本論文では、縦及び横ともに 500mm 以上 2,000mm 未満の閉領域を障害物として検出する。障害物までの距離は、深度画像 $Z(u, v)$ 上の閉領域内の深度の平均値とする。

4. 実験と考察

車両に取り付けたステレオカメラで撮影された動画像を用いて、提案手法の性能評価を行う。以下では、実験環境、評価方法、検出精度の評価、処理時間の評価について述べる。

4.1 実験環境

本論文では、車への搭載を検討するために、組込み機器に近い計算機環境に提案手法を実装して性能評価を行う。図 5 に示す計算機及びステレオカメラを用いる。計算機環境を表 1 に、ステレオカメラの仕様を表 2 に示す。図 5 に示すステレオカメラユニットを自動車のフロントガラスに取り付けて、自車両の前方を撮影した 50 フレームからなる 6 シーンの動画像を用いる。なお、本論文の実験では、定量的な評価のために、自車両の前方を走行している車両を計測対象 (障害物) とする。障害物として検出する車両は、図 6 に示す矩形で囲まれた 13 台の車両とする。これらの車両を近距離 (10~40m)、中距離 (40~70m)、遠距離

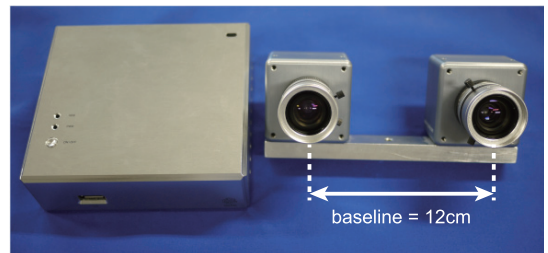


図 5 実験で用いる計算機 (左) とステレオカメラ (右)
Fig. 5 Small computer (left) and stereo camera unit (right) used in the experiment.

表 1 計算機環境

Table 1 Computer environment.

CPU	Intel Core i3-3217U (1.80GHz)
GPU	Intel HD 4000 graphics
Memory	DDR3 SO-DIMM 4GB
OS	Windows 7 (32bit)
Compiler	Visual C++ + 10.0 Compiler

表 2 ステレオカメラの仕様
Table 2 Specification of stereo camera.

Camera	PointGrey Scorpion
Focal length	6.5mm
Image size	1,280×960 pixels
Color mode	Grayscale
Frame rate	15fps
Baseline	120mm



図 6 入力画像 (左カメラ) と検出対象の障害物：近距離 (10~40m: 緑), 中距離 (40~70m: 青), 遠距離 (70~110m: 赤)

Fig. 6 Left camera images and target obstacles: short distance (10~40m, green), middle distance (40~70m, blue, long distance (70~110m, red).

(70~110m) に分けて評価する。本論文では、撮影したステレオ画像をあらかじめ平行化してから、入力画像とする。

4.2 F 値を用いた検出精度の評価

障害物の検出精度を定量的に評価するために、本論文では、F 値 (F-measure) を用いる。F 値は、図 7 に示すように、車両の真値領域と、検出アルゴリズムによって得られる検出領域との面積比で定義される。真値領域のうち検出されなかった領域を FalseNegative、検出領域のうち誤検出した領域を FalsePositive、真値領域と検出領域の共通領域を TruePositive とする。これらの領域の面積比を用いて、次式で定義される再現率 (recall) と適合率 (precision) を計算する。

$$\text{recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \quad (18)$$

$$\text{precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (19)$$

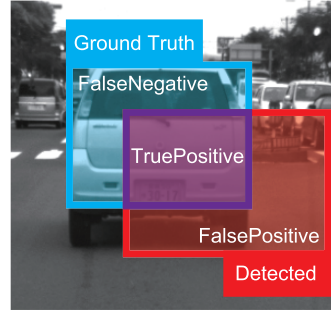


図 7 F 値の計算に用いられる領域
Fig. 7 Definition of FalseNegative, FalsePositive, TruePositive regions for calculating F-measure.

再現率は、真値領域をどれくらい不足なく検出できたかを表し、適合率は、真値領域をどれくらい超過なく検出できたかを表す。F 値は、これらの調和平均として次式で定義される。

$$\text{F-measure} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}} \quad (20)$$

F 値が 1 に近ければ、障害物を過不足なく正確に検出できていることを示す。

これまででは、動画の中で、障害物に検出枠が重なっているフレームの割合を検出率として評価することが多かった [16]。しかし、このような検出率のみでは、検出枠の位置や大きさを定量的に評価していない。そのため、障害物検出手法の優位性を正確に評価することができない。F 値を評価指標として用いることで、障害物検出が過不足なく正確に行えているかを評価することができる。

4.3 検出精度の評価

実験では、ステレオ画像対応付けとクラスタリングのさまざまな組合せにおいて障害物検出の精度を評価し、提案手法の有効性を示す。

ステレオ画像の対応付けとして、Semi-Global block Matching (SGM) [4], Efficient Large-Scale stereo matching (ELAS) [5], Normalized Cross-Correlation (NCC) に基づく手法と提案手法 (POC) を用いる。SGM 及び ELAS は大域的な最適化に基づく手法に、NCC 及び POC (Proposed) はブロックマッチングに基づく手法に分類される。SGM は、OpenCV のライブラリ [17] を用いる。ELAS は Geiger らが公開しているライブラリ [5] を用いる。NCC は、2.2 で述べた対応付けの処理で、POC を NCC に置き換えた手法である。NCC を用いたサブピクセル分解能の

平行移動量推定には、パラボラフィッティングを用いる [18]. NCC と POC は、C 言語で実装する. どちらの手法においても、画像ピラミッドの階層数を 3 とする. 探索ウィンドウのサイズは、NCC に基づく手法では 16 画素 \times 15 ラインとし、POC に基づく手法では 32 画素 \times 15 ラインとする. POC に基づく画像マッチングでは、ハニング窓を用いることでフーリエ変換の際に生じる信号端の不連続性の影響を抑えている. ハニング窓では、信号長の半分が半値幅になることより、POC に基づく手法において 32 \times 15 画素のウィンドウに含まれる情報量は、他の対応付け手法で用いる 16 \times 15 画素のウィンドウに含まれる情報量に相当する. POC に基づく対応付けにおいて、上記のパラメータを用いると、-64 ~ 64 画素の視差を探索することができる. これは、表 2 に示す条件下において、実空間上で 5m ~ 無限遠に相当する. 視差画像のクラスタリングは、V-disparity 画像に基づく手法 [7] と提案手法 (等視差領域のクラスタリング) を用いる. V-disparity 画像に基づく手法は、F 値が最大になるようにパラメータを設定する.

表 3 に、対応付けとクラスタリングの全ての組合せに関する F 値の平均を示す. それぞれのフレームで対象とする障害物に対して F 値を算出し、全 300 フレーム (50 フレーム \times 6 シーン) の平均を比較する. なお、障害物を検出できなかったフレームの F 値は 0 としている.

まず、クラスタリング手法に着目して結果を考察する. 障害物までの距離やステレオ画像の対応付け手法によらず、提案手法は、V-disparity 画像に基づくクラスタリング手法に比べて F 値の平均が大きい. 特に、距離が遠くなるにつれて、F 値の平均の差が大きくなっている. このことから、提案手法の方が障害物を正確に検出できていることが分かる. また、図 8 に、POC で生成した視差画像に対して、それぞれのクラスタリング手法で障害物を検出した結果の例を示す. V-disparity 画像を用いた手法は、基線長が短いために、障害物とそれ以外の領域との切り分けが難しく、障害物を大きな領域として検出してしまっている. 一方で、提案手法は、障害物を過不足なく正確に検出できており、F 値が大きい.

次に、画像対応付けに着目して結果を考察する. 以下では、提案手法である等視差領域のクラスタリングを用いることにする. 表 3 から、距離によらず、POC で生成した視差画像を用いたときに最も F 値が大き

表 3 F 値の平均 (上段 (V) : V-disparity 画像に基づく手法, 下段 (P) : 提案手法)

Table 3 Average of F-measure for each distance (V: V-disparity, P: proposed).

Distance	Clustering	SGM	ELAS	NCC	Proposed
Short 10~40m	V	0.7085	0.7659	0.5357	0.7207
	P	0.7385	0.6791	0.5916	0.8145
Middle 40~70m	V	0.3448	0.3379	0.3920	0.3201
	P	0.6848	0.6788	0.4743	0.8105
Long 70~110m	V	0.2032	0.1674	0.1364	0.1620
	P	0.4189	0.4390	0.2143	0.5493

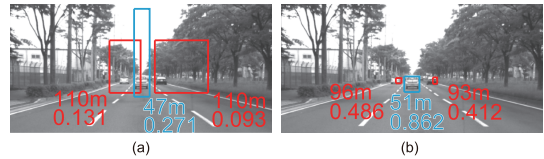


図 8 障害物検出の例: (a) V-disparity 画像に基づくクラスタリング, (b) 提案手法 (検出枠の下の数値は、障害物までの距離と F 値を示す)

Fig. 8 Example of obstacle detection result: (a) V-disparity-based clustering and (b) proposed clustering (values under the detected region are distance between camera and obstacle and F-measure).

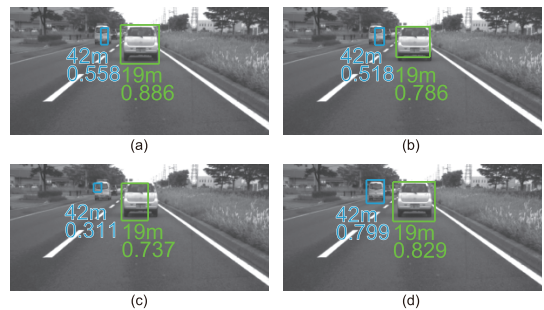


図 9 障害物検出の例: (a) SGM, (b) ELAS, (c) NCC, (d) Proposed (検出枠の下の数値は、障害物までの距離と F 値を示す)

Fig. 9 Example of obstacle detection result: (a) SGM, (b) ELAS, (c) NCC and (d) Proposed (values under the detected region are distance between camera and obstacle and F-measure).

い. 図 9 に、それぞれのステレオ画像対応付け手法で障害物を検出した例を示す. 近距離の障害物の検出は大差がないが、中距離の障害物の検出は POC を用いたときに F 値が最も大きい. 図 10 に、図 9 の障害物検出に用いた視差画像を示す. POC を用いて生成した視差画像は、誤対応が少なく、40m 程度の中距離でも障害物領域の視差をなめらかに推定できている.

図 11 に提案手法を用いたときの検出結果の例を示

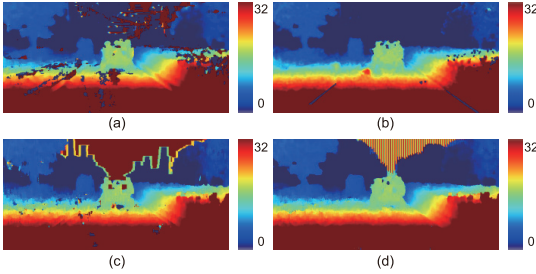


図 10 図 9 の検出に用いた視差画像：(a) SGM, (b) ELAS, (c) NCC, (d) Proposed

Fig. 10 Disparity image used to detect obstacle in Fig. 9: (a) SGM, (b) ELAS, (c) NCC and (d) Proposed.

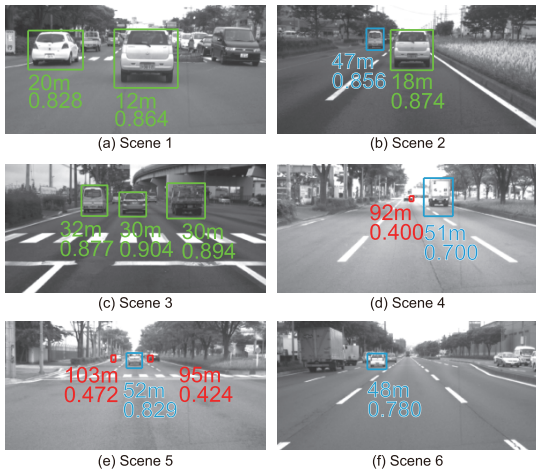


図 11 提案手法を用いた障害物検出の例（検出枠の下の数値は、障害物までの距離と F 値を示す）

Fig. 11 Example of obstacle detection result using the proposed method (values under the detected region are distance between camera and obstacle and F-measure).

す。以上から、POC に基づく画像対応付けと等視差領域のクラスタリング手法を組み合わせた提案手法が最も正確に障害物を検出できていることが分かる。

4.4 処理時間の評価

提案手法の処理時間の評価を通して、リアルタイム処理の可能性について考察する。提案手法は、対応付けを行う基準点の数に処理時間が依存する。例えば、配置する基準点の間隔を疎にすることで、生成する視差画像を小さくし、処理時間を大幅に短縮することが可能である。一方で、視差画像が小さくなることで、障害物の検出精度が低下する可能性がある。表 4 に、視差画像のサイズ（基準点の数）を変えたときの提案

表 4 視差画像のサイズを変えたときの処理時間
Table 4 Processing time of the proposed method with changing the size of disparity image.

Size of disparity image [pixel]	Total [ms]	Correspondence matching [ms]	Clustering [ms]
1,280×960	36643.1	31122.9	5520.2
640×480	8579.1	8076.4	502.7
427×320	3610.9	3496.6	114.3
320×240	2050.9	2001.6	49.3
256×192	1339.9	1314.1	25.8
213×160	963.5	945.5	18.0
183×137	733.9	720.5	13.4
160×120	542.5	530.6	11.9
142×107	429.6	421.1	8.5
128×96	360.0	352.3	7.7

表 5 視差画像のサイズを変えたときの F 値の変化
Table 5 F-measure of the proposed method with changing the size of disparity image.

Size of disparity image [pixel]	F-measure		
	Short 10~40m	Middle 40~70m	Long 70~110m
1,280×960	0.8145	0.8105	0.5493
640×480	0.7943	0.8215	0.5583
427×320	0.7843	0.7892	0.4893
320×240	0.7816	0.7462	0.4722
256×192	0.7491	0.7388	0.3566
213×160	0.7448	0.7118	0.2716
183×137	0.7079	0.6559	0.1624
160×120	0.6616	0.6280	0.0759
142×107	0.6253	0.5785	0.0310
128×96	0.6286	0.5667	0.0272

手法の処理時間を示す。処理時間の大半がステレオ画像の対応付けであり、対応付けの高速化が不可欠であることが分かる。ステレオ画像の対応付けは、基準点ごとに処理が独立していて並列処理が可能であるので、Graphics Processing Unit (GPU) による並列実装が効果的である [19]。表 5 に、視差画像のサイズを変えたときの提案手法の F 値を示す。視差画像が小さくなるほど全体的に F 値が小さくなるが、Short と Middle に比べて、Long の低下が著しい。これは、遠い車両を対象としたときに、視差画像上での対象車両の面積（画素数）が小さくなってしまい、等視差領域のクラスタリングを用いても正確に車両を検出できないためである。表 3 に示した V-disparity 画像に基づくクラスタリング手法と比較すると、視差画像のサイズが 213×160 画素より大きいときは、距離によらず提案手法の方が F 値が高い。これは、V-disparity 画像に基づくクラスタリング手法が、1,280×960 画素の視差画像を用いているのに対して、提案手法は、213×160 画素の視差画像で同等以上の性能を有することを示す。つまり、従来手法と比較して、1/6 のサイズの視差画

像から同等以上の精度で提案手法が車両を検出できることを示している。提案手法において、V-disparity画像に基づくクラスタリング手法よりもF値が高く、かつ、最も視差画像のサイズが小さいのは、 213×160 画素の視差画像を生成した場合である。このときの処理時間は、表4より、963.5 msである。本論文では、表1に示したように、低消費電力プロセッサを用いることで、組み込み機器に近い計算機環境のもとで評価を行っている。実験結果より、70m先の障害物を検出することはできるが、リアルタイムで運転手に提示するためには、特に対応付けの処理速度を改善する必要がある。GPUを利用した並列処理による高速化だけではなく、アルゴリズムの改善を含めた処理の最適化を行ってリアルタイム処理を実現することが今後の課題である。

5. む す び

本論文では、基線長が短い車載ステレオカメラに適した障害物検出手法を提案した。提案手法は、(i) 位相限定相関法 (Phase-Only Correlation: POC) に基づく高精度なステレオ画像の対応付けを用いた視差画像の生成と、(ii) 微分処理に基づく等視差領域のクラスタリングを組み合わせることで、障害物領域を過不足なく正確に検出することが可能である。性能評価実験を通して、基線長の短いステレオカメラを用いたときに、100m先の障害物を正確に検出できることを示した。本論文では、基線長が短いステレオカメラに焦点を当てて議論をした。基線長が長いステレオカメラで撮影した画像に対してSGMやELASを用いて作成した視差マップは、POCで作成した視差マップと同程度の精度である。等視差領域のクラスタリングの検出精度は、視差の推定精度に大きく依存するため、正確な視差マップが得られれば、正確に車両を検出することができる。基線長が長い場合は、等視差領域のクラスタリングで車両検出を行うと、SGM、ELAS、POCのいずれの場合でも検出精度が高いことを実験的に確認している。今後は、車載ステレオカメラ向けの対応付け手法を検討するとともに、大規模なステレオ画像データベースを作成して網羅的な性能評価実験を行う予定である。

文 献

- [1] H. Cheng, *Autonomous Intelligent Vehicles*, Springer, 2011.
- [2] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.28, pp.694–711, 2006.
- [3] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.
- [4] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol.2, pp.807–814, 2005.
- [5] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," *Proc. Asian Conf. Computer Vision*, pp.25–38, 2010.
- [6] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp.3354–3361, June 2012.
- [7] R. Labayrade, D. Aubert, and J.P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "V-disparity" representation," *Proc. IEEE Intelligent Vehicles Symposium*, vol.2, pp.646–651, 2002.
- [8] Y. Gao, X. Ai, Y. Wang, J. Rarity, and N. Dahnoun, "U-V-disparity based obstacle detection with 3D camera and steerable filter," *Proc. IEEE Intelligent Vehicles Symposium*, vol.4, pp.957–962, 2011.
- [9] Z. Hu and K. Uchimura, "U-V-disparity: An efficient algorithm for stereovision based scene analysis," *Proc. IEEE Intelligent Vehicles Symposium*, pp.48–54, 2005.
- [10] K. Matsushima, H. Matsuura, Y. Kijima, Z. Hu, and K. Uchimura, "Road scene analysis using trinocular stereo vision," *IEEJ Trans. Electronics, Information and Systems*, vol.126, no.12, pp.1551–1558, 2006.
- [11] T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "A sub-pixel stereo correspondence technique based on 1D phase-only correlation," *Proc. Int'l Conf. Image Processing*, pp.V–221–V–224, 2007.
- [12] C.D. Kuglin and D.C. Hines, "The phase correlation image alignment method," *Proc. Int'l Conf. Cybernetics and Society*, pp.163–165, 1975.
- [13] H. Foroosh, J.B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol.11, no.3, pp.188–200, March 2002.
- [14] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. Fundamentals*, vol.E86-A, no.8, pp.1925–1934, Aug. 2003.
- [15] 青木孝文, 伊藤康一, 柴原琢磨, 長嶋 聖, "位相限定相関法に基づく高精度マシビジョン—ピクセル分解能の壁を越える画像センシング技術を目指して," *IEICE Fundamentals Review*, vol.1, no.1, pp.30–40, Jan. 2007.
- [16] A. Seki and M. Okutomi, "Robust obstacle detection

in general road environment based on road extraction and pose estimation,” *Electronics and Communications in Japan (Part II:Electronics)*, vol.90, no.12, pp.12–22, 2007.

- [17] “Open Computer Vision Library,” <http://sourceforge.net/projects/opencvlib/>
- [18] M. Shimizu and M. Okutomi, “Sub-pixel estimation error cancellation on area-based matching,” *Int. J. Comput. Vis.*, vol.63, no.3, pp.207–224, 2005.
- [19] M. Miura, K. Fudano, K. Ito, T. Aoki, H. Takizawa, and H. Kobayashi, “GPU implementation of phase-based stereo correspondence and its application,” *Proc. Int’l Conf. Image Processing*, pp.1697–1700, 2012.

(平成 26 年 1 月 9 日受付, 7 月 22 日再受付)



青木 孝文 (正員)

1988 東北大・工・電子卒. 1992 同大学院工学研究科博士課程了. 同年同大学・工・助手, 1994 同大学院情報科学研究科助手, 1996 同助教授, 2002 同教授. 2012 より副学長を兼任. 現在に至る. 超高速デジタル計算の理論, 画像センシング, 映像信号処理, バイオメトリクス, VLSI 設計技術, 分子コンピューティングに関する研究に従事. 英国電気学会フレミング賞及びマウントパッテン賞ほかを受賞. IEEE, 計測自動制御学会, 情報処理学会各会員. 博士 (工学).



和泉 圭祐

2012 東北大・工・情報・卒. 2014 同大学院情報科学研究科修士課程修了. 画像処理に関する研究に従事.



三浦 衛 (学生員)

2010 東北大・工・情報・卒. 2012 同大学院情報科学研究科修士課程了. 現在, 同大学院情報科学研究科博士課程在学中. 画像処理に関する研究に従事.



伊藤 康一 (正員)

2000 東北大・工・電子卒. 2005 同大学院情報科学研究科博士課程了. 同年同大学院情報科学研究科助手, 2007 同助教. 2004~2005 日本学術振興会特別研究員. 現在に至る. 画像処理及びバイオメトリクスに関する研究に従事. IEEE, 情報処理学会各会員. 博士 (情報科学).