

差分ステレオ運動領域に注目したステレオ視の提案*

梅田和昇 ** 寺林賢司 ** 橋本優希 *** 中西達也 *** 入江耕太 ***

Subtraction Stereo - A Stereo Vision Method That Focuses On Moving Regions -

Kazunori UMEDA, Kenji TERABAYASHI, Tatsuya NAKANISHI, Yuki HASHIMOTO and Kota IRIE

In this paper, a practical method of stereo vision, “subtraction stereo” is proposed. A huge number of studies have been carried out for stereo vision until now, and several practical stereo vision systems have been reported. However, what is called the correspondence problem that stereo matching becomes difficult and not robust for weak textures or recurrent patterns is inevitable for stereo vision. Subtraction stereo realizes robust measurement of range images by detecting moving regions with each camera first and then applying stereo matching for the detected moving regions. Detection of moving regions is carried out with a subtraction process. Concept and fundamental algorithm of subtraction stereo are introduced. Then measurement of three-dimensional position, height and width of a target object using the subtraction stereo is discussed. The basic algorithm is implemented on a commercially available stereo camera and the effectiveness of the subtraction stereo is verified by several experiments using the stereo camera. Although objects are restricted to moving ones, subtraction stereo gives sufficient information robustly for many applications such as surveillance.

Key words: stereo camera, motion, detection of moving regions, subtraction, surveillance camera

1. 緒 言

ステレオ視に関しては、これまでに数多く研究開発が行われてきた^{1)~5)}。最近では、CPUやGPU (Graphics Processing Unit) の高速化に伴い、比較的容易にPCを用いた実時間でのステレオ画像取得が可能となっている⁶⁾⁷⁾。PCの代わりにFPGA (Field Programmable Gate Array) を用いた高速なステレオ画像取得も提案されている⁸⁾。また、PCに接続して用いる製品も市販されている⁹⁾¹⁰⁾。さらに、車載用などに実用化されているステレオカメラもある¹¹⁾。

このように、実用レベルにあると言えるステレオ視であるが、テクスチャが弱い場合や繰り返しパターンがある場合に左右の画像での対応づけを求めるのが困難になる、いわゆる対応点問題が本質的に不可避である。この問題に対し、多数台のカメラを用いることでロバスト化を行うマルチベースラインステレオ¹²⁾が知られている。また、実用的なアプローチとして、ランダムドットなどのパターンを照射し、ステレオマッチングをロバスト化する手法¹³⁾¹⁴⁾があるが、これらの手法は対象が近い場合に限定される。

本研究では、不審者検出や人の流れの計測を行う監視カメラなどの用途を想定し、実用的なステレオカメラの開発を行うことを目標とする。単眼のカメラで歩行者の検出などを行うシステムは色々提案されているが¹⁵⁾¹⁶⁾、ステレオカメラは、寸法情報を得ることができ、スケラブルな利用が可能となるため、これらの用途に有意義であると考えられる。すなわち、例えば、検出された運動領域が猫なのか人なのか、大人なのか子供なのか、何名程度の人がいるのか、などの情報を得ることが容易になると期待できる。

これらの用途では、計測したい対象物は一般に動きを伴うことが多い。本研究では、この性質を利用した新しいステレ

オ視の一手法：“差分ステレオ (Subtraction Stereo)”を提案することを目的とする。本手法は、運動情報に陽に注目することによって、ステレオカメラのロバスト化を実現しつつ、監視カメラなどの用途に適した情報を出力する。さらに、差分ステレオの基本的な特性の評価を行う。まず、各種の応用に用いるための基本として、対象とする領域の3次元特徴計測を行う手法の流れを示す。そして、市販のステレオカメラに提案手法の基本アルゴリズムを実装して実験を行うことで、差分ステレオが実装可能なことを示すと共に、一般的なステレオ視との違い、基礎的な計測誤差の評価、応用の基礎実験としての歩行人物の3次元計測結果を示す。

以下、2章で差分ステレオの基本アルゴリズムを提案し、3章で差分ステレオを用いた物体の3次元計測を論じる。基本アルゴリズムの市販ステレオカメラへの実装を4章に示し、このシステムを用いたいくつかの実験結果を5章に示す。6章で考察を行い、7章で結論を述べる。

2. 基本アルゴリズム

本研究で提案するステレオ視の基本アルゴリズムを図1に示す。通常のステレオ視では、左右カメラの画像をマッチングすることで視差画像を得る。これに対し、本研究で提案する差分ステレオでは、まず左右カメラそれぞれの入力画像から運動領域を抽出し、その後に抽出された運動領域に対して左右画像でのステレオマッチングを適用する。運動領域の抽出は、何らかの差分処理—最も単純には背景差分—によって実現する。

本手法は、視差が得られるのが運動している領域に限定される代わりに、以下の理由からマッチングをロバスト化することができると考えている。

- 対応点探索を左右カメラの運動領域に大きく限定することができる。
- マッチングのための情報に、元画像だけでなく運動情報

* 原稿受付 平成21年3月27日

** 正会員 中央大学/JST CREST (東京都文京区春日 1-13-27)

*** 中央大学大学院/JST CREST

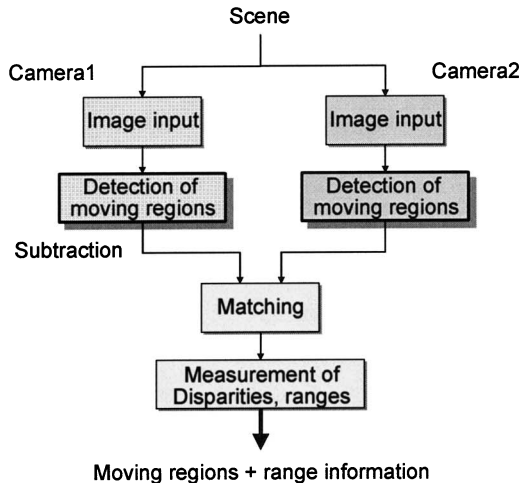


Fig. 1 Flow of subtraction stereo

も利用可能である。

これらを式で表現すると以下ようになる。画像中の点 $\mathbf{u}$ において視差 k を算出する処理は

$$k(\mathbf{u}) = \arg \min_k d(\mathbf{u}, k) \quad (1)$$

と表すことができる。ここで、 $d(\mathbf{u}, k)$ は左右の画像でのマッチングを表す何らかの指標で、SAD (Sum of Absolute Distance) が典型的である。また、 $\arg \min$ は、関数の最小値を与える引数の値を示す。これに対し、差分ステレオでの視差の算出処理は、

$$k(\mathbf{u}) = \arg \min_{k \in K_m} d(\mathbf{u}, k) \quad (2)$$

と表される。すなわち、視差 k を探索する範囲を運動領域 K_m に限定することができる。また、 $\mathbf{u}$ に関しても探索領域が運動領域に限定されるので、探索範囲は2重に削減可能と言える。さらに、左右の画像での運動情報（単純には差分値そのもの）のマッチングを表す指標 $d_m(\mathbf{u}, k)$ を加え、視差の算出を

$$k(\mathbf{u}) = \arg \min_{k \in K_m} f(d(\mathbf{u}, k), d_m(\mathbf{u}, k)) \quad (3)$$

とすることができる。

なお、対象は必ずしも文字通り運動している必要はなく、差分処理で検出可能であれば、静止していても差分ステレオは適用可能である。

3. 物体の3次元特徴計測

本節では、差分ステレオによる物体の3次元特徴の計測に関して論じる。3次元位置、高さ、幅などが計測可能である。

3.1 ラベリング

差分ステレオでは、得られる視差画像が元々運動領域に限定されているため、視差画像から移動する物体を抽出するのは容易であり、標準的なラベリング手法で行うことが可能である。その際、面積に関するしきい値処理により、ノイズの除去や抽出する対象物の大きさを限定することが可能である。

3.2 3次元位置の計測

ある画素において、視差が与えられれば、光軸方向の距離を求めることができる。以後、距離という用語を、光軸方向の距離の意味で用いることとする。2台のカメラが平行に配置され

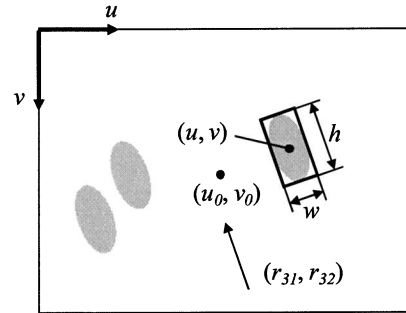


Fig. 2 Height and width of an object in the image coordinate system

ていると仮定し、視差を k 画素、距離を z とおくと、よく知られているように、距離 z は以下の式で求めることができる。

$$z = \frac{\alpha}{k}, \quad \alpha = \frac{b \cdot f}{p} \quad (4)$$

ここで b は基線長（レンズ中心間の距離）、 f はレンズの焦点距離、 p は1画素あたりの横幅である。

さらに、距離 z と画像座標 (u, v) (図2参照) から、3次元位置 $\mathbf{x}$ が得られる。スキューがなく、各画素のアスペクト比が1である場合には、 $\mathbf{x}$ は

$$\mathbf{x} = z \cdot \begin{bmatrix} \frac{p}{f}(u - u_0) & \frac{p}{f}(v - v_0) & 1 \end{bmatrix}^T \quad (5)$$

で与えられる。ここで、 (u_0, v_0) は画像中心の画像座標である。

3.3 座標変換

ワールド座標系におけるカメラの位置・姿勢が既知であれば、対象点のワールド座標系における位置 $\mathbf{x}_w$ が次式で計算される。

$$\mathbf{x}_w = R\mathbf{x} + \mathbf{t}. \quad (6)$$

ここで、 $\mathbf{t}$ はカメラの位置を表す3次元ベクトル、 R はカメラの姿勢を表す 3×3 回転行列である。

3.4 高さ・幅の計測

対象が近くて距離の計測誤差が小さい場合には、領域中の各点の3次元位置から、高さ・幅を求めることが可能である。一方、対象が遠いと、距離の計測誤差が大きくなり、各点の3次元位置から高さ・幅を求めることは適切ではない。そこで、対象領域の画像中での高さ・幅、ならびに対象領域を代表する距離を用いて、3次元空間中での高さ・幅を求める。まず対象領域を代表する距離としては、例えば領域の中心位置での距離、あるいは領域の各点の距離の平均値を用いれば良い。

次に、対象の画像中での高さ・幅と3次元空間中での高さ・幅との関係を求める。ワールド座標系の z 軸を鉛直方向とし、対象の鉛直方向の分布を高さ、それと垂直な方向の分布を幅とする。この時、3次元空間での高さ方向は、

$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$$

と表される。このベクトルをカメラ座標系から見ると、

$$\begin{bmatrix} r_{31} & r_{32} & r_{33} \end{bmatrix}^T$$

と表される。ただし、 r_{ij} は式(6)の R の第 (i, j) 成分である。このベクトルが画像平面に射影されたものが画像中での高さ方向である。カメラ座標系の z 軸方向を画像平面と垂直な方向とし、弱中心射影¹⁷⁾を仮定すると、高さ方向は、 (r_{31}, r_{32}) となる。さらにこの時高さが $\sqrt{r_{31}^2 + r_{32}^2} = \sqrt{1 - r_{33}^2}$ 倍に縮

小さくて観測されることになる。以上より、図 2 に示すように、領域の画像中での (r_{31}, r_{32}) 方向の幅を h 、それと垂直方向の幅を w とすると、領域の 3 次元空間中での高さ、幅は、それぞれ以下の式で求められる。

$$height = z_c \cdot \frac{p}{f} \cdot \frac{h}{\sqrt{1-r_{33}^2}} \quad (7)$$

$$width = z_c \cdot \frac{p}{f} \cdot w \quad (8)$$

ここで、 z_c は対象領域を代表する距離である。

3.5 距離計測の誤差解析

視差の不確かさを σ_k (標準偏差) とする。誤差の伝播則を式 (4) に適用することにより、計測される距離の不確かさ σ_z が次式で与えられる。

$$\sigma_z = \left| \frac{\partial z}{\partial k} \right| \sigma_k = \frac{p}{b \cdot f} z^2 \sigma_k \quad (9)$$

この式は、良く知られているように、計測される距離の不確かさが距離の 2 乗に比例して大きくなることを示している。さらに、不確かさが各画素の幅に比例し、基線長および焦点距離に反比例することも示している。例として、次節のステレオカメラのパラメータに対応した値を用いて、 b を 120mm, f を 3.8mm, p を $14.8\mu\text{m}$ とすると、

$$\sigma_z = 3.25 \times 10^{-2} z^2 \sigma_k \quad (10)$$

となる。例えば、 σ_k が 0.1 画素の時、 $z=1, 10, 20, 50\text{m}$ に対し、 $\sigma_z=0.003, 0.3, 1.3, 8\text{m}$ と計算できる。

4. 市販ステレオカメラへの実装

提案した差分ステレオの基本アルゴリズムを、市販のステレオカメラに実装した。カメラには、Point Grey Research 製の Bumblebee2 (カラー、 $f=3.8\text{mm}$) を用いた。画像サイズは 320×240 画素とした。

現段階では、運動領域の検出には単純な背景差分を用いている。またステレオマッチングには Bumblebee2 に付属のライブラリを利用している。背景差分で得られる左右の差分画像に対してステレオマッチングを行うことで、視差画像が得られる。処理速度は、Core2Duo T9300 (2.50GHz) の PC で、ラベリングなどを行わず基本アルゴリズムのみを実行して約 37fps である。

5. 検証実験

本節では、差分ステレオの有効性を検証するための実験結果を示す。前節で示したシステムを用いて実験を行った。

5.1 基本アルゴリズムの検証実験

まず、基本アルゴリズムの有効性を検証するため、屋内・屋外の典型的なシーンに対する差分ステレオの適用結果を示す。

5.1.1 屋内シーン

図 3(a) に屋内の対象シーンのカラー画像を示す。なお、このカラー画像はステレオカメラの右画像である。以後の実験でも同様にカラー画像はステレオカメラの右画像を示す。事前に登録しておいた背景画像を用いて背景差分を行い、得られた差分画像が図 3(b) である。左右カメラで得られた差分画像を用いて差分ステレオによって得られた視差画像を図 4(a) に、通常のステレオ法で得られた視差画像を図 4(b) に示す。これらの視差画像で色が視差を表し、より青い色がより大きな視



(a) Color image of the experimental scene



(b) Subtraction image
Fig. 3 Experimental scene: indoor



(a) Subtraction stereo



(b) Standard stereo

Fig. 4 An example of subtraction stereo: indoor

差、すなわち近距離を、より赤い色がより小さな視差、すなわち遠距離を表す。

図 4(a) より、差分ステレオによって視差画像が移動物体領域のみに対して適切に得られていることが分かる。ステレオマッチングを適用するのが移動物体領域に限定されているた

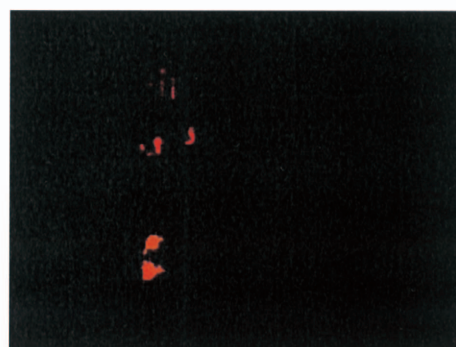


(a) Color image of the experimental scene

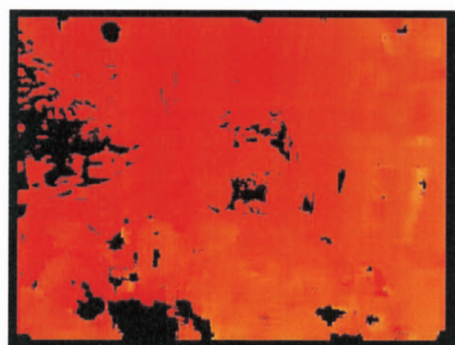


(b) Subtraction image

Fig. 5 Experimental scene: outdoor



(a) Subtraction stereo



(b) Standard stereo

Fig. 6 An example of subtraction stereo: outdoor

め、ステレオマッチングにおける誤対応を減少させ、また計算コストも削減することが可能であると示唆される。実際に、図 4(b) で見られる誤対応（例えば右側の赤い、すなわち視差が小さいと見なされた箇所）は、図 4(a) では生じていない。

5.1.2 屋外シーン

図 5(a) に屋外の対象シーンのカラー画像を示す。屋内シーンにおける実験と同様に、図 5(b) が差分画像、図 6(a), (b) がそれぞれ差分ステレオならびに通常のステレオ法で得られた視差画像である。この例では、対象となる人物は遠くて画像中で小さく、通常のステレオ法では困難なシーンであり、実際に、図 6(b) から人物を抽出するのは困難である。一方、図 6(a) では、人物領域が視差情報とともに適切に抽出されている。

以上の屋内・屋外シーンでの適用例より、差分ステレオにより運動領域の視差画像が実際に得られること、また監視用途のように移動物体の検出が意味のある対象においては、必要十分かつより有効な情報が得られることが示されている。

5.2 計測誤差の評価実験

距離の計測誤差を評価するための基本実験を行った。図 7(a) に示すように、屋内の廊下で、テクスチャを持つ平面物体をステレオカメラに正対するように設置した。平面物体の寸法は 440mm×1400mm である。距離 1m から 10m まで 1m おきに、また 10m から 35m まで 5m おきに計測を行った。各距離で、抽出された領域の中心における距離を 100 回計測し、平均とばらつき（標準偏差）を求めた。図 7(b) に結果を示す。距離が小さい時には、計測誤差は小さい。一方、距離が 10m を超えると、ばらつきが大きくなっていくと同時に、計測値に偏りが生じている。この偏りの原因は不明であるが、他の個体で異なる特性を示すことから、Bumblebee2 の個体ごとの特性のようである。いずれにせよ、偏りを考慮しても、距離 35m

においても計測誤差は距離の 10-20% 程度であり、監視用途などで利用可能な誤差の範囲内であると考えられる。また、特性には再現性が見られるので、事前にキャリブレーションを行うことで偏りを低減することも可能であると考えている。

5.3 光軸方向に歩行する人物の距離計測

図 8(a) に示すように、カメラの光軸方向に、カメラから遠ざかる方向に歩行する人物の距離計測を行った。前項の実験が静止した平面物体を対象としたのに対し、歩行している人物を対象としている点で、より現実的な状況での実験となっている。前項の実験と同様に、抽出された人物領域の中心における距離を計測した。得られた距離データを図 8(b) に示す。横軸はフレーム番号を示す。図中の直線は、距離データに当てはめた近似直線、上下の曲線は、この近似直線に式 (10) で σ_k を 0.1 画素として計算される標準偏差を足し引きしたものである。

まず、対象が歩行している人物であっても安定かつ妥当に距離が計測できていると言える。また、計測距離のばらつきが距離に応じて大きくなっていることが観測される。ばらつきの大きさは、図 7 の静止物体の場合と同程度である。さらに、上下の曲線との比較から、ばらつきの増加が理論にほぼ沿っていること、また式 (10) で仮定している 0.1 画素より σ_k が小さく、視差の計測が高精度であることが言える。なお、図中の最後のフレームにおける実際の距離は 37.5m であり、前項の実験と同様に偏りが見られる。

5.4 一般シーンでの歩行人物の 3 次元計測

最後に、一般的な屋外シーンで歩行する人物の計測を行った実験結果を示す。計測したのは、抽出された人物領域の中心における 3 次元位置、高さ・幅である。カメラは、図 9 に示すように、高さ 8.3m の位置に、下向きに 50° だけティルト



(a) Experimental scene

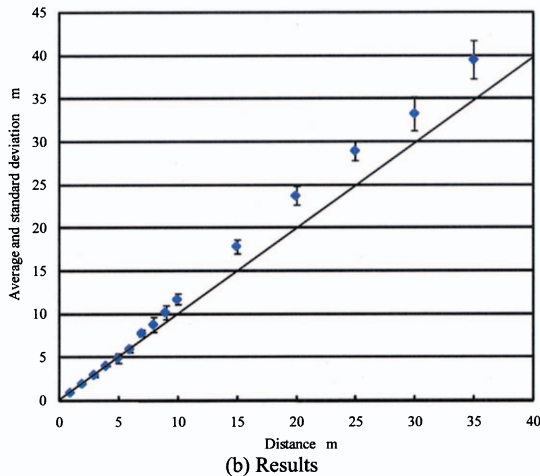


Fig. 7 Experiment to evaluate measurement error: still object



(a) Experimental scene

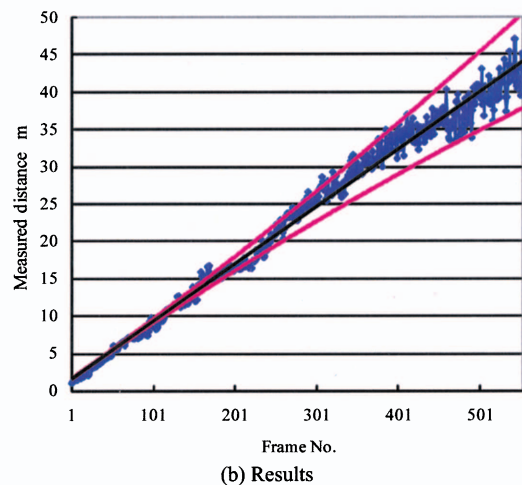


Fig. 8 Experiment to measure a person walking straight away

させて設置した。図 10(a) に対象シーンのカラー画像を示す。対象人物は身長 173cm で、右から左に歩行した。

図 10(b), (c) に実験結果を示す。図 10(b) が 3 次元位置、図 10(c) が高さ・幅を示している。なお、3 次元位置は、図 9 に示すワールド座標系での値に変換してある。また、カメラのディリットに対応して $r_{33} = \sin 50^\circ$ となるため、式 (7) の $\sqrt{1 - r_{33}^2}$ は $\cos 50^\circ$ となる。実験結果より、歩行人物の運動が適切に計測されており、また得られた高さ・幅も妥当な値と言える。幅が振動しているのは、歩行時の足の振れに対応したものである。

6. 考 察

実験結果より、差分ステレオの有効性、特に監視などの用途への適性が示されたと考えている。また、差分ステレオの基本アルゴリズムは、非常にシンプルであり、実装が容易である。そのため、差分ステレオの機能をハードウェア的に持つステレオカメラを構築することも可能であると考えている。

ステレオカメラの構築を含め、今後進めていく必要があると考えられることを以下に示す。

- 基本アルゴリズムの改良。

現段階では、単純な背景差分を用いているため、対象物を含まない背景画像が事前に得られ、かつ背景に変更がない場合しか適用できない。そこで、計算コストとのバランスを取りつつ統計モデルに基づいたロバストな背景画像推定手法¹⁸⁾¹⁹⁾などを導入することで、照明の変動や風などにより背景が変動する場合にも運動領域の抽出を行えるようにすることが必要であると考えている。

- 複数人物のトラッキング、群衆の計測。
- 単純な人物のトラッキングに留まらず、シーンの高次情報の抽出。
- スタンドアロンで稼働する小型ステレオカメラシステムの構築。
- 屋内、ビル内、屋外など多様なスケールの対象への応用の模索。

7. 結 言

本論文では、実用的なステレオ視の一手法として、差分ステレオを提案した。本手法は、左右カメラそれぞれでまず運動領域を抽出し、得られた運動領域に注目してステレオマッチングを行うことによって、ロバストなステレオ画像計測を実現するというものである。また、差分ステレオによる移動物体の 3 次元位置、高さ・幅の計測手法を示した。差分ステレオの基本アルゴリズムを市販のステレオカメラに実装し、本システムを用いた実験により、差分ステレオの有効性を示した。

謝 辞

本研究の実装・実験などに尽力してくれた元中央大学大学院生の松木優介氏、和田正樹氏に深謝します。

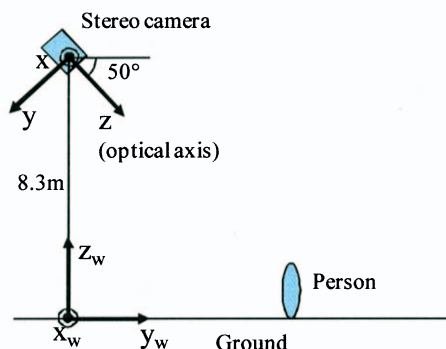
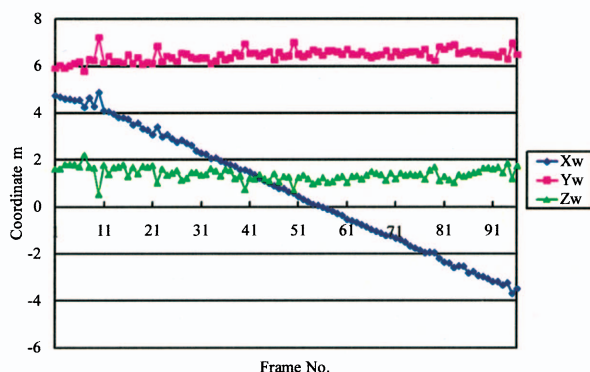


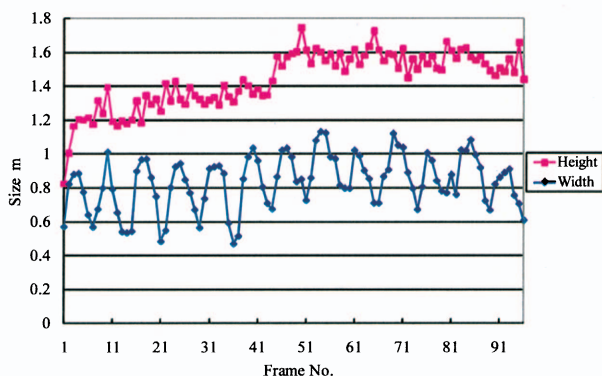
Fig. 9 Experimental setup of the outdoor experiments



(a) Color image of the experimental scene



(b) 3D Position



(c) Height and width

Fig. 10 Experimental result of measuring a walking person in a general outdoor scene

参考文献

- 1) M. Okutomi: Difficulties in Stereo Vision, J. Robotics Society of Japan, **16**, 6, (1998) 773 (in Japanese).
- 2) R. Hartley and A. Zisserman: Multiple View Geometry in Computer

- Vision, Cambridge Univ. Press, (2000).
- 3) M. Hebert: Active and passive range sensing for robotics, Proc. of ICRA'00, 1, (2000) 102.
- 4) M.Z. Brown, D. Burschka and G.D. Hager: Advances in computational stereo, IEEE Trans. PAMI, **25**, 8, (2003) 993.
- 5) S.M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, Proc. of CVPR'06, 1, (2006) 519.
- 6) K. Okada, S. Kagami, M. Inaba and H. Inoue: Onbody Realtime Disparity Image and Flow Generation System based on Highspeed Correspondance Matching using PC, J. Robotics Society of Japan, **18**, 6, (2000) 896 (in Japanese).
- 7) T. Ueshiba: An Efficient Implementation Technique of Bidirectional Disparity Search for Real-time Trinocular Stereo Vision, Meeting on Image Recognition and Understanding (MIRU2006), (2006) IS3-60 (in Japanese).
- 8) M. Hariyama, Y. Kobayashi, H. Sasaki and M. Kameyama: FPGA Implementation of a Stereo Matching Processor Based on Window-Parallel-and-Pixel-Parallel Architecture, IEICE Trans. Fundamentals, **E88-A**, 12, (2005) 3516.
- 9) Point Grey Research: <http://www.ptgrey.com/>
- 10) Videre Design: <http://www.videre.com/>
- 11) N. Sogawa and K. Hanawa: Forward Direction Recognition System with Stereo Vision for Automobiles, J. Society of Automotive Engineers of Japan, **56**, 4 (2002) 34 (in Japanese).
- 12) M. Okutomi and T. Kanade: A Multiple-Baseline Stereo, IEEE Trans. PAMI, **15**, 4 (1993) 353.
- 13) Y. Sato, M. Matsuyama and K. Sakaue: Investigation of projecting near-infrared pattern for electric wheelchair with a stereo vision, Dynamic Image Processing for Real Application Workshop (DIA2007), (2007) O6-1 (in Japanese).
- 14) Y. Matsuki and K. Umeda: Improvement of the range image sensor that integrates active and passive methods using multiple spotlights and a stereo camera, JSME Robomec2008, (2008) 2P1-G12 (in Japanese).
- 15) K. Terada: Observation of Cars and Pedestrians, J. Robotics Society of Japan, **22**, 7, (2004) 864.
- 16) T. Haga, K. Sumi and Y. Yagi: Walker Detection in Outdoor Scene Using Spatial-Temporal Motion Analysis, Trans. IEICE, **J87-D-II**, 5 (2004) 1104 (in Japanese).
- 17) G. Xu and S. Tsuji: 3D Vision, Kyoritsu Shuppan, (1998) (in Japanese).
- 18) H. Shimai, T. Kurita, S. Umeyama, M. Tanaka and T. Mishima: Adaptive Background Estimation by Robust Statistics, Trans. IEICE, **J86-D-II**, 6 (2003) 796 (in Japanese).
- 19) A. Shimada, D. Arita and R. Taniguchi: Increment and Decrement of Gaussians in Adaptive Mixture-of-Gaussian Background Models, Meeting on Image Recognition and Understanding (MIRU2006), (2006) 746-751 (in Japanese).