

差分ステレオを用いた歩行者計測における カメラの外部パラメータ推定

橋本優希 星川佑磨 寺林賢司 梅田和昇 (中央大, JST CREST)

Estimation of Camera's Extrinsic Parameters in Measurement of Pedestrians Using Subtraction Stereo

*Yuki HASHIMOTO, Yuma HOSHIKAWA, Kenji TERABAYASHI, Kazunori UMEDA
(Chuo Univ., CREST, JST)

Abstract— In this paper, estimation of camera's extrinsic parameters in measurement of pedestrians using "subtraction stereo" is discussed. Subtraction stereo is a stereo vision method that focuses on the movement of objects to make a stereo camera robust and produces range images for moving regions. Features of pedestrians such as 3D position, height and width are measured from range images obtained by subtraction stereo. Then a method to estimate extrinsic parameters of a camera is proposed. The basic algorithm of the subtraction stereo is implemented on a commercially available stereo camera, and the effectiveness of the method to estimate extrinsic parameters is verified by experiments using the stereo camera.

Key Words: Stereo Camera, Extrinsic Parameter, Self-Calibration, Pedestrians

1. 序論

ステレオカメラでは、テクスチャが弱い場合などに画像間の対応付けが難しくなる、対応点問題が不可避である。この問題に対し、我々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、運動情報に陽に注目した“差分ステレオ”を提案している [1]。本手法は、マッチングにおける探索範囲を運動領域へ制限することによりロバスト化を図るものであり、これまでに計測精度、歩行者の検出などの検討を行ってきた [2][3]。

本論文では、差分ステレオを用いたカメラの外部パラメータ推定に関して論じる。単眼カメラを用いたカメラの外部パラメータ推定に関しては、様々な手法が提案されている [4][5]。外部パラメータが求められれば、人物が平面上にいると仮定することで、単眼カメラでも人物の寸法情報や実環境での位置を取得することができる。しかしながら、計測システムを色々な場所へ移動させる場合やカメラが水平面に平行に近い場合などを考えると、距離やカメラ座標系における寸法情報を直接的に得ることのできるステレオカメラの利用 [6] が有効であると言える。

本論文では、差分ステレオによって得られる歩行者の領域の視差を用いてカメラの外部パラメータを推定する手法を提案する。平面当てはめを3次元空間ではなく視差空間に対して行うことで、より妥当な計測を実現する。屋内・屋外の異なる計測環境において実験を行い、手法の有効性を検討する。

2. 差分ステレオによる歩行者計測の概要

2.1 差分ステレオの概要

差分ステレオは、通常のステレオ手法と異なり、あらかじめ差分処理によって抽出した領域のみにマッチング処理を行う (図1)。これにより探索範囲を制限す

ることができ、対応点探索のロバスト化及び高速化を図ることができる。

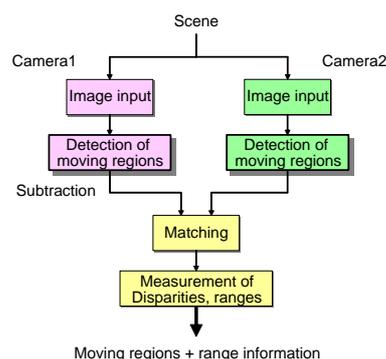


Fig.1 Flow of subtraction stereo

2.2 人物領域の検出と3次元特徴量計測

差分ステレオでは、視差画像が運動領域に限定されるため、基本的なラベリングによって人物領域を抽出することができる。ノイズに対しては、大きさ情報の閾値処理により人物領域を判定する。

人物領域の3次元位置は、一般の平行ステレオに基づき、以下のように計算される。カメラから光軸方向の距離 z_c は、

$$z_c = \frac{bf}{dp} \quad (1)$$

で与えられる。ここで、 d は視差、 b は基線長、 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) から、3次元位置 x_c 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、 x_c は

$$x_c = z_c \begin{bmatrix} \frac{p}{f}(u - u_0) & \frac{p}{f}(v - v_0) & 1 \end{bmatrix}^T \quad (2)$$

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．

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 x_w が次式で計算される．

$$\mathbf{x}_w = R\mathbf{x}_c + \mathbf{t} \quad (3)$$

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. カメラの外部パラメータ推定

カメラに映る人物の寸法情報や実環境での位置を計測するためには，まずカメラの外部パラメータである R と \mathbf{t} を求めなければならない．本研究ではカメラに映る歩行者の領域から重心の位置を抽出し，これらから RANSAC[7] により地面を推定し，カメラの外部パラメータを求める．ここで，位置を抽出する時に視差を利用し，視差空間で地面を推定する．これは，視差を距離に変換する際に距離の 2 乗に比例して誤差がのるため，より誤差の影響を受けにくい視差空間において推定の方が妥当であると考えられるからである [8]．手法を以下に示す．

(I) 人が歩いているところを平面と仮定し，歩行者の重心座標データを取得し蓄える．

(II) 蓄えた重心座標からランダムに 3 点を選ぶ．

(III) 3 点を通る平面を求め，この平面上にある重心座標の数 N をカウントする．点 $A(u_1, v_1, d_1)$ ， $B(u_2, v_2, d_2)$ ， $C(u_3, v_3, d_3)$ を通る平面の式は

$$a(u - u_1) + b(v - v_1) + c(d - d_1) = 0, \quad (4)$$

$$a = (v_2 - v_1)(d_3 - d_1) - (d_2 - d_1)(v_3 - v_1)$$

$$b = (d_2 - d_1)(u_3 - u_1) - (u_2 - u_1)(d_3 - d_1)$$

$$c = (u_2 - u_1)(v_3 - v_1) - (v_2 - v_1)(u_3 - u_1)$$

で与えられる．ただし，計測点の画像座標 (u, v) での視差を d とする．

(IV) (II)，(III) を適度に繰り返し， N が最も大きいものを最適モデル平面とする．

(V) 最適モデル平面より外部パラメータを求める．ここで，(IV) で求めた最適モデル平面をカメラ座標系に変換する． (u, v, d) をカメラ座標系の (x_c, y_c, z_c) に変換したのに対して， a, b, c それぞれを再度計算し直した値を a', b', c' とおく．カメラの地面からの高さ t_3 ，つまりカメラと地面の最短距離は

$$t_3 = \frac{|a'x_1 + b'y_1 + c'z_1|}{\sqrt{a'^2 + b'^2 + c'^2}} \quad (5)$$

より求まる．ただし， t_3 は式 (3) の \mathbf{t} の第 3 成分である．図 2, 3 のようにカメラ座標系における各軸 X_c, Y_c, Z_c と最適モデル平面の交点をそれぞれ A, B, C，カメラの光学中心を O とおくと，仰角 θ_t とロール角 θ_r はそれぞれ \overline{OC} と \overline{BC} ， \overline{OA} と \overline{BA} の関係を使って以下のように求めることができる．

$$\theta_t = \cos^{-1} \frac{z_0}{\sqrt{y_0^2 + z_0^2}} \quad (6)$$

$$\theta_r = \cos^{-1} \frac{x_0}{\sqrt{x_0^2 + y_0^2}} \quad (7)$$

ただし，

$$\begin{cases} x_0 = \frac{1}{a}(by_1 + cz_1) + x_1 \\ y_0 = \frac{1}{b}(ax_1 + cz_1) + y_1 \\ z_0 = \frac{1}{c}(ax_1 + by_1) + z_1 \end{cases}$$

である．

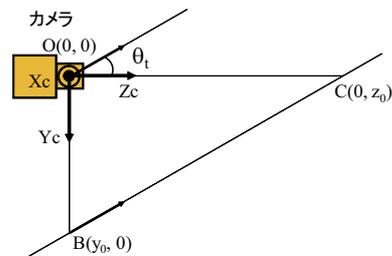


Fig.2 $y_c - z_c$ plane ($x_c = 0$)

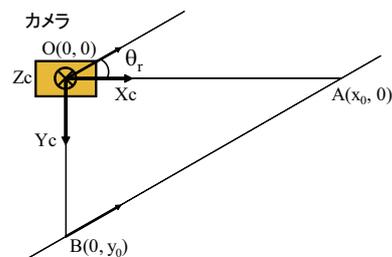


Fig.3 $x_c - y_c$ plane ($z_c = 0$)

4. 外部パラメータ推定実験

本節では，2 つの実験から，差分ステレオを用いたカメラの外部パラメータ推定手法の有効性を検証する．大学構内と研究室の 2 つの計測環境において，歩行者を対象として外部パラメータ推定を行った結果を示す．

ステレオカメラは，Point Grey Research 製の Bumblebee2 (color, VGA, 48[fps], $b = 120[\text{mm}]$, $f = 3.8[\text{mm}]$, $p = 7.4[\mu\text{m}]$) を用い，実験での画像サイズは $320 \times 240[\text{pixel}]$ とした．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．

4.1 歩行者重心と地面の取得データによる推定精度の比較実験

本節では図 4 に示すように，ステレオカメラを高さ $t_3 = 8.3[\text{m}]$ ，仰角 $\theta_t = 50[^\circ]$ ，ロール角 $\theta_r = 0[^\circ]$ とし設置した．

差分ステレオによる歩行者重心と地面の取得データによる推定精度の比較実験を行った．歩行者重心について，撮影範囲の周辺部では人の出入り時における身体の欠損により重心座標が不安定になるため，周辺部

Table 1 Estimated result of extrinsic parameters using the information of the ground

	Height[m]	Tilt[°]	Roll[°]
Ave.	8.71	49.01	2.43
S.D.	0.36	3.19	4.15
True value	8.30	50.00	0

Table 2 Estimated result of extrinsic parameters using the information of pedestrians

	Height[m]	Tilt[°]	Roll[°]
Ave.	7.64	46.06	-8.40
S.D.	0.34	2.21	4.00
True value	8.30 (7.45)	50.00	0

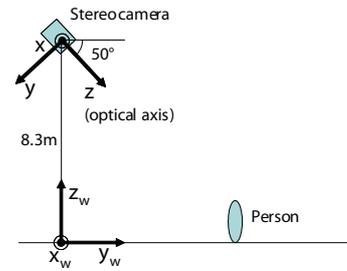
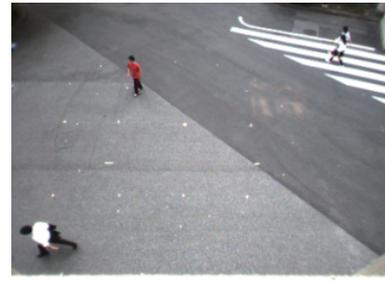
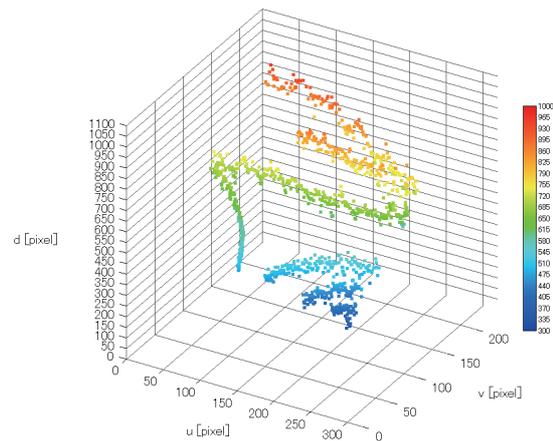
のデータをカットした。地面については、実験のシーンにおいて画像の周辺を除くほとんどが地面であることから、周辺部を除く 251×191 [pixel] の範囲内における視差の値をデータとして用い、3. で説明した手法を使って最適モデル平面を求め、外部パラメータの推定精度を求めた。歩行者重心と地面のそれぞれで 1008 個のデータを取得し、10 回ずつ外部パラメータ推定を行い、平均値と標準偏差を計算して精度を比較した。歩行者重心については、データ取得中に約 13 人がそれぞれ異なる経路でカメラの前を通過した。このときのシーンを図 5 に示す。

それぞれの推定の結果を表 1, 2 に示す。また、取得した歩行者の重心座標の点群と、これより精度検証のために繰り返し求めたモデル平面の 1 つを図 6, 7 に示す。Roll の欄にある符号は、正のとき時計回りであることを示している。表 2 の値などは重心の情報に基づいて求めている。そのため、高さは真値と異なる。カッコ内の数値は、身長を 170 [cm] として、その半分である 85 [cm] を引いたものである。これらより、地面での推定結果は直接平面上の点群から推定しているため、歩行者での推定結果に比べて誤差が小さいと考えられるが、歩行者の結果も真値に近い推定結果を出していることが分かる。より精度を求める場合、地面の座標を用いてパラメータを推定する方が有効であると考えられる。しかし、都市環境では建物、街路樹、車、人などにより、カメラに平面のみを映すことが難しい。これに対し、人は平面を歩くことがほとんどのため、取得する座標の領域を限定する必要がない。また、上述の結果より、精度にそれほど差がないことから、都市環境において提案手法は地面の座標から推定する手法に比べ優位性があると考えられる。

4.2 床面の特定が困難なシーンでの推定

本節では、ステレオカメラを高さ $t_3 = 2.54$ [m]、仰角 $\theta_t = 35^\circ$ 、ロール角 $\theta_r = 0^\circ$ として設置した。

都市環境での使用を想定し、研究室において地面が

**Fig.4** Experimental setup of the experiments**Fig.5** Experimental scene**Fig.6** Positions of the centroid of walking humans

あまり見えないシーンを人為的に作り、差分ステレオによる歩行者重心の取得データによる推定の基礎検討を行った。1 人の歩行者が障害物をよけながら研究室

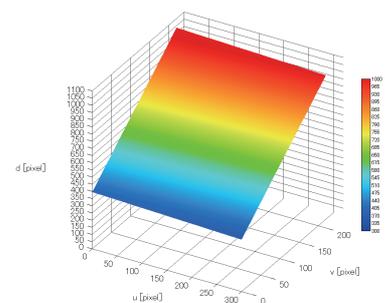
**Fig.7** Model plane obtained from Fig.6

Table 3 Estimated result of extrinsic parameters using the information of the pedestrian in the laboratory

	Height[m]	Tilt[°]	Roll[°]
Ave.	1.75	33.12	9.15
S.D.	0.79	6.39	27.03
True value	2.54 (1.73)	35.00	0

をカメラに映る範囲で1周するシーンを対象とし、精度を検討した。重心のデータは前節と同様に1008個取得し、10回ずつ外部パラメータ推定を行い、平均値と標準偏差を計算した。このときのシーンを図8に示す。

推定精度の結果を表3に示す。また、取得した歩行者の重心座標の点群と、これより精度検証のために繰り返し求めたモデル平面の1つを図9, 10に示す。Rollの欄にある符号は前節と同様に、時計回りが正である。重心の情報を基にパラメータなどを求めているため、高さは真値と異なる。身長162[cm]の人物で実験したため、その半分である81[cm]を重心の高さとする。この結果から、平均値は真値に近い値を示している。標準偏差は前節の結果と比べて値が大きくなっていることが分かる。これは、障害物によってオクルージョンが生じたり、歩行者の近くの障害物が影の影響を受けたことにより、歩行者の重心のぶれが前節の実験に比べて大きくなったためと考えられる。標準偏差は大きくなったものの、基礎検討としては今後につながる成果が得られたと考えられる。また、本手法を用いることで地面とそうでない部分の手動での切り分けが不要なため、簡単で実用的な外部パラメータ推定手法であると考えられる。今後、影の除去を行い歩行者の重心以外のノイズを極力減らすことが、更なる精度向上のための課題であると考えられる。



Fig.8 Experimental scene

5. 結論

本論文では、差分ステレオを用いたカメラの外部パラメータ推定手法を提案した。そして、市販のステレオカメラを利用して比較実験及び基礎検討を行い、提案手法が地面の切り分けに手間がかかる都市環境において外部パラメータを推定するのに有効であることを示した。

今後の課題は、影の除去の実装による外部パラメー

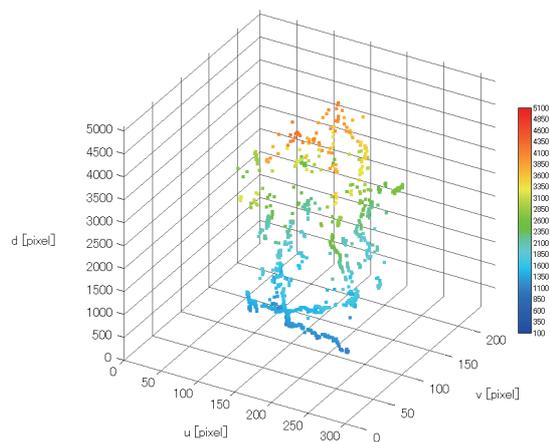


Fig.9 Positions of the centroid of the walking human

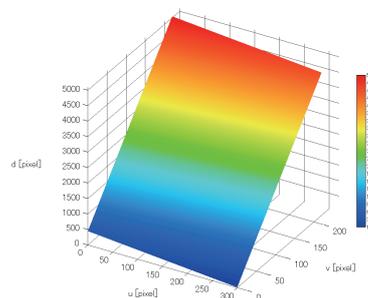


Fig.10 Model plane obtained from Fig.9

タの精度向上である。

- [1] 梅田和昇, 和田正樹, 松木優介: “運動領域注視型ステレオカメラシステムの提案 第1報: 基本コンセプトの提案”, 第25回日本ロボット学会学術講演会予稿集, 1N31, 2007.
- [2] Y. Hashimoto, Y. Matsuki, T. Nakanishi, K. Umeda, K. Suzuki and K. Takashio: “Detection of pedestrians using subtraction stereo,” SWDMNSS 2008, pp.165-168, July 2008.
- [3] 橋本優希, 中西達也, 寺林賢司, 梅田和昇, 鈴木慧, 高汐一紀: “差分ステレオを用いた歩行者計測”, 第26回日本ロボット学会学術講演会予稿集, 1L1-03, 2008.
- [4] Z. Zhang: “A flexible new technique for camera calibration,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.11, pp.1330-1334, Nov. 2000.
- [5] F. Lv, T. Zhao, R. Nevatia: “Camera calibration from video of a walking human,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.28, no.9, pp.1513-1518, September 2006.
- [6] 関晃仁, 奥富正敏: “ステレオ動画像を利用した平面領域抽出による障害物検出”, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, vol.45, no.SIG13(CVIM10), pp.1-10, 2004.
- [7] M. A. Fischler and R. C. Bolles: “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM, vol.24, no.6, pp.381-395, Jun. 1981.
- [8] 梅田和昇: “平面領域検出における相対視差画像の利用の検討”, 第20回日本ロボット学会学術講演会予稿集, 2A16, 2002.