人検出結果に基づくカメラの自己キャリブレーションと3次元位置推定

安藤寛哲，藤吉弘亘（中部大学）

A Method of Measuring 3D Position and Camera Self-Calibration Based on Results of Human Detection

Hiroaki Ando, Hironobu Fujiyoshi (Chubu University)

Abstract

A camera self-calibration method based on the results of human detection is proposed. This method extracts the positions and heights of people in the target scene from the results of human detection and human-area segmentation, and estimates camera parameters such as a location of camera in the world coordinate and a vanishing line in the image coordinate. Calibrating a camera generally requires intensive work, but the proposed technique can perform self-calibration using parameters that are automatically extracted from the target image. As a result, our method can estimate the three-dimensional position of an object even with a camera that had not been previously calibrated. Experimental results show that the accuracy of estimated camera parameters can be improved by using the results of human-area segmentation.

キーワード：自己カメラキャリブレーション，人検出，セグメンテーション，3 次元位置推定
(Camera self-calibration，Pedestrian detection，Segmentation，3D position estimation)

1. はじめに

近年，公共施設だけではなくオフィスや一般家庭への監視カメラの普及とともに，物体検出や追跡などの動画像処理技術による監視システムの需要が高まっている．特に，映像中から人を検出 [1] [2] し，人の 3 次元位置情報や身長を知ることは，場所に応じた人数カウントや人流測定において重要である．小川らは，人の身長と位置情報の推定に，3 次元空間の光線情報に基づく手法を提案した [3]．この手法では，Tsai モデル [4] に基づきカメラキャリブレーションを行い，カメラの内部パラメータと外部パラメータを用いて人の位置と身長の推定を行う．キャリブレーションにより求めたカメラパラメータを利用することで，画像中の 2 次元座標と世界座標における 3 次元座標の関係を求めることが可能となる．しかし，一般的にカメラキャリブレーションは，画像中の座標に対応する世界座標を与える必要があり，広い範囲でのキャリブレーションは難しく，正確に行うには大変な労力が必要となる．

一方，手間を必要としない自己カメラキャリブレーションの手法も多く提案されている．Lv らは，歩行者を背景モデルとの差分から算出し，歩行者の頭と足下の位置検出し，画像中の人の高さ世界座標と画像座標の関係を用いて，消失点と水平線を算出することにより，カメラの自己キャリブレーション手法を提案している [5]．しかし，Lv らの手法では，ノイズ等の影響により画像中から人の大きさを安定して取得できないという問題がある．また，Hoiem らは，入力画像から物体の検出を行い，検出結果から得られた複数の物体高さの情報と，入力画像における三次元の面構造と消失点を推定して，自己カメラキャリブレーションする手法を提案している [6]．これらの物体検出に基づくカメラの自己キャリブレーション手法では，参照する物体の高さ (例えば人の高さ) を精度良く求める必要がある．

そこで，本稿では人領域のセグメンテーション結果に基づくカメラの自己カメラキャリブレーション手法を提案する．提案手法は，推定するシーンに対し人検出と人領域のセグメンテーションを行う画像から抽出した複数の人の高さと位置情報と透視投影モデルにおける世界座標と画像座標の関係を用いてカメラの自己キャリブレーションを行う．また，自己キャリブレーションの結果を用いることにより，未校正のカメラにおいても 3 次元位置の情報を推定することができる．

2. 人領域のセグメンテーション

提案手法では，単眼のカメラで撮影された映像から，画像中の物体の高さと位置，透視投影モデルにおける世界座標と画像座標の関係を用いて，カメラの自己キャリブレーションを実現する．本手法では，画像中の物体の高さ情報を用いてカメラパラメータを高精度に推定するため，人領域を正確にセグメンテーションする必要がある．

本手法では，人領域のセグメンテーション手法に，村井らの提案した弱識別器の応答による類似シルエットの選択を用いたセグメンテーション手法 [7] を用いる．この手法は，人検出器の学習と学習サンプルに対応するシルエット画像のスコアからハッシュテーブルを求めるオフライン処理と，入力画像からラスタスキャンによる人検出を行い Chamfer Matching を用いてセグメンテーションを行うオンライン処理から構成される．処理の流れを図 1 に示し，以下に詳細を述べる．

〈2･1〉 人検出器の構築　人検出法として HOG（Histograms of Oriented Gradients）特徴量 [1] を用いた Real AdaBoost [8] による人検出器を用いる．HOG 特徴量は 1 つの局所領域内におけるエッジ方向ごとのエッジ強度に着目

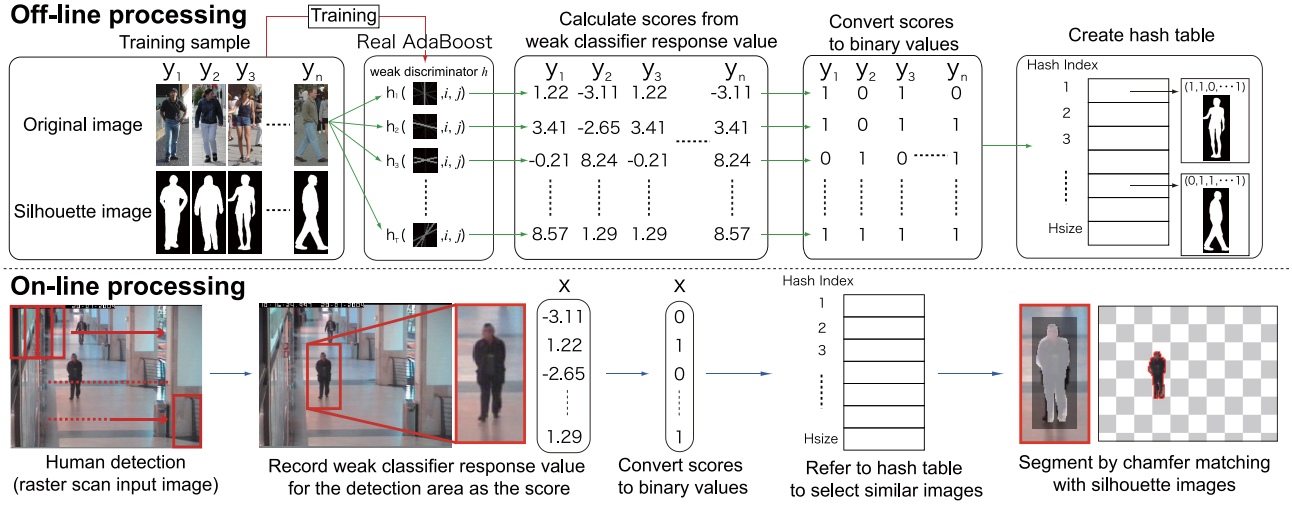

図 1　人領域のセグメンテーション処理の流れ

Fig. 1.　Processing flow in the human-area segmentation

した特徴量であり，照明変動と，局所的な幾何学的変化に頑健な特徴量である．人検出では，この HOG 特徴量を予め用意した学習サンプルから算出し，Real AdaBoost により人検出器の構築を行う．

〈2・2〉　シルエット画像のスコア算出　　入力画像から人検出を行い，シルエット画像のマッチングを行う候補領域を推定する．構築した検出器を用いてシルエット画像のスコアを算出する．シルエットのスコアとは，人検出器の各弱識別器の応答値を多次元のベクトルとして表現したものである．まず，学習サンプルからシルエット画像を作成する．次に，作成したシルエット画像に対応する学習サンプル (ポジティブのみ) を構築した人検出器に入力する．入力画像に対して各弱識別器は人か人以外を識別した結果として返す．T 個の弱識別器に対応した T 次元の特徴ベクトルを対応するシルエット画像のスコアとする．

〈2・3〉　ハッシュテーブルの作成　　算出したシルエット画像のスコアは，Chamfer Matching を行う対象領域に類似したシルエットを選択するために使用する．シルエット画像のスコアを記録したハッシュテーブルを作成することで高速な類似シルエット画像検索を実現する．ある一つのシルエット画像 y から算出された弱識別器の応答に基づくスコアを，T 次元の特徴ベクトル $\mathbf{y} = \{h_1(y), h_2(y), \cdots, h_T(y)\}$ とする．まず，次式により各次元の特徴量を 2 値化し，ビットベクトル $\mathbf{u} = (u_1, u_2, \cdots, u_T)$ を作成する．

$$u_j = \begin{cases} 1 & if \quad h_j(y) \geq 0 \\ 0 & otherwise \end{cases} \quad \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

次に，以下の式を用いて，ハッシュテーブルのインデックス H_{index} を求め，ハッシュテーブルに特徴ベクトルとシルエット画像の ID を登録する．

$$H_{index} = \left(\sum_{i=1}^{T} u_i 2^i \right) \mod H_{size} \cdots\cdots\cdots\cdots\cdots (2)$$

ここで，H_{size} はハッシュテーブルのサイズである．以上の処理をすべてのシルエット画像に対して行い，ハッシュテーブルを作成する．

〈2・4〉　類似シルエットの選択　　オンライン処理として実際の入力画像から人検出を行い，MeanShift クラスタリングによりウインドウを統合する．検出されたウインドウ内の領域が検出器に入力されたときの各弱識別器の応答値の T 次元の特徴ベクトルとシルエット画像のスコアとして記録した特徴ベクトルを比較することで類似シルエット画像の選択を行う．その方法として，作成したハッシュテーブルを用いて近似最近傍探索を行う．人検出により検出された領域から得られた弱識別器の応答に基づく特徴ベクトルを $\mathbf{x} = (h_1(x), h_2(x), \cdots, h_T(x))$ とする．この特徴ベクトルに対しても式 (1)，(2) を用いてハッシュのインデックスを求め，同じハッシュのインデックスを持つ特徴ベクトルをハッシュテーブルから探索することで類似する特徴ベクトルを持つシルエット画像を選択する．

〈2・5〉　Chamfer Matching によるセグメンテーション　人検出により検出された領域と，弱識別器の応答値から選択された類似シルエット画像を Chamfer Matching [9] によりマッチングすることで人領域のセグメンテーションを行う．入力画像として人検出により検出された領域，テンプレート画像としてシルエット画像を用いて相違度に基づいてマッチングを行う．このとき，人検出により検出された領域は必ずしもきれいに人領域を検出しているとは限らないため，Chamfer Matching を行う領域としてマージンを持たせて少し大きめに切り出した領域を使用する．また，Chamfer Matching はスケール変化に敏感であるため，シルエット画像のスケールを変化させながらマッチングを行う．最終的に，最も Chamfer Matching の相違度が小さくなる位置とスケールとシルエット画像を決定し，シルエット画像が重なっている部分を切り出すことで人領域のセグ

メンテーションを行う．セグメンテーション結果から，画像中の人の高さ h_i と位置 v_i を精度よく抽出することができる．

3. カメラの自己キャリブレーション

人領域のセグメンテーション結果から得られた画像中の人の高さを用いてカメラの自己キャリブレーションを行う．まず，文献 [6] を参考に，透視投影モデルにおけるカメラ位置と人の高さと人の位置の関係を定式化する．本章では，画像座標と世界座標とカメラパラメータの関係と，その関係からカメラパラメータの導出方法について述べる．

〈3･1〉 カメラ位置と人の高さの関係　図 2 に，カメラ位置と人の高さの関係を示す．

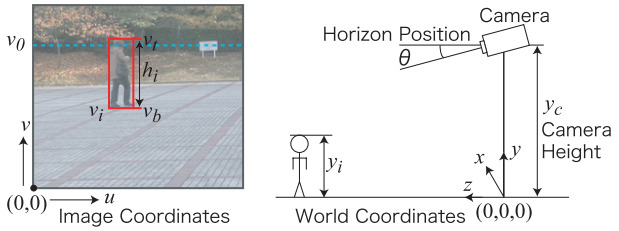

図 2　カメラと人の高さの関係
Fig. 2.　Relationship between camera and human height

画像の左下の座標を $(0,0)$ とし，画像の縦と横を画像の縦幅のサイズで正規化した画像の座標系を (u,v) で与える．また，世界座標系を (x,y,z) とする．ただし，y を高さ，z を奥行きとする．次に，カメラのチルト角を θ，焦点距離を f，カメラ中心 (u_c,v_c)，カメラの高さを y_c とする．世界座標系は，カメラ位置を基準として，$z_c=0, x_c=0$ とし，接地平面を $y=0$ と定義する．また，カメラ中心 (u_c,v_c) は画像中心とし，焦点距離 f は 1.4 とする．カメラにロールは発生せず，画像座標系の地面の消失ラインとして水平線を v_0 と定義する．このとき，カメラのチルト角（ラジアン）は次式で与えられる．

$$\theta = 2\arctan\frac{v_c - v_0}{2f} \cdots\cdots (3)$$

カメラモデルとして，歪み無しの単位アスペクト比である透視投影モデル（Perspective projection model）を使用すると，世界座標系から画像座標系への変換は以下の式で与えられる．

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 & u_c \\ 0 & f & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & y_c \\ 0 & \sin\theta & \cos\theta & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4)$$

〈3･2〉 カメラパラメータの導出　世界座標系と画像座標系とカメラパラメータの関係から，カメラパラメータの導出方法について説明する．式 (4) を物体の高さ y について解くと次式となる．

$$y = \frac{z(f\sin\theta - (v_c - v)\cos\theta) - fy_c}{(v_c - v)\sin\theta + f\cos\theta} \cdots\cdots (5)$$

ここで，画像中の人（直立と仮定）領域の上底と下底の位置をそれぞれ v_t と v_b として与える（図 2）．下底位置 v_b で地面と接する．つまり v_b のとき $y=0$ なので，物体の奥行き z は次式のように求められる．

$$z = fy_c / (f\sin\theta - (v_c - v_b)\cos\theta) \cdots\cdots (6)$$

式 (5) と式 (6) から，物体の高さ y は次式のようになる．

$$y = \frac{fy_c(f\sin\theta - (v_c - v_t)\cos\theta)/(f\sin\theta - (v_c - v_b)\cos\theta) - fy_c}{(v_c - v_t)\sin\theta + f\cos\theta} \quad (7)$$

ここで，カメラのチルト角が小さい（推定された水平線の位置が画像内に存在する）場合，$\cos\theta \approx 1$，$\sin\theta \approx \theta$，$\theta \approx \frac{v_c - v_0}{f}$ と近似できる．従って，式 (7) は以下のように表現できる．

$$y \approx y_c \frac{v_t - v_b}{v_0 - v_b} / \left(1 + (v_c - v_0)(v_c - v_t)/f^2\right) \cdots\cdots (8)$$

分母の $(v_c - v_0)(v_c - v_t)/f^2$ は，チルト角が小さいとき $v_c - v_0 \approx 0$ となるため，$(v_c - v_0)(v_c - v_t)/f^2 \approx 0$ と近似できる．従って，式 (8) は以下のようになる．

$$y \approx y_c \frac{v_t - v_b}{v_0 - v_b} \cdots\cdots (9)$$

ここで，画像中に存在する i 番目の人領域を考えたとき，$v_t - v_b$ は画像中の i 番目の人領域の高さ h_i とする．また，人領域の基準座標を矩形領域の左下に設定したとき，$v_b = v_i$ となる．従って，求めたい式は以下のようになる．

$$y_i \approx y_c \frac{h_i}{v_0 - v_i} \cdots\cdots (10)$$

式 (10) を変形し，n 人のサンプルとの関係を行列で表現すると次式のようになる．

$$\begin{bmatrix} v_0 \\ y_c \end{bmatrix} = \begin{bmatrix} y_i & -h_i \\ \vdots & \vdots \\ y_n & -h_n \end{bmatrix}^{-1} \begin{bmatrix} y_i v_i \\ \vdots \\ y_n v_n \end{bmatrix} \cdots\cdots (11)$$

式 (11) からカメラの高さ y_c と水平線の位置 v_0 を求める．

〈3･3〉 RANSAC によるカメラパラメータ推定　式 (11) において，既知のパラメータは，〈2･5〉で述べたように，人領域のセグメンテーション結果より画像中から抽出した人の高さ h_i と人領域の位置 v_i である．世界座標におけるカメラの高さ y_c と画像座標上の水平線の位置 v_0 を求めるためには，世界座標系での人の高さ y_i が必要である．しかし，与えられている映像からそれぞれの世界座標系の人の高さの真値を知ることは不可能である．そこで，提案手法では人の身長として平均値 170.0[cm]，標準偏差 8.0 の正規乱数により人の高さを与え，RANSAC(RANdam SAmple Consensus) [10] を用いてカメラの高さとチルト角の推定を行う．RANSAC の流れは以下の通りである．

Step1　ランダムに 2 点のサンプルを選択
Step2　最小二乗法によりカメラの高さと水平線位置を推定
Step3　推定値から全サンプルとの誤差を算出
Step4　誤差の中間値を選択
Step5　**Step1** ∼ **Step4** を繰り返し，誤差の中間値が最小となるパラメータを求め，最終的なカメラの高さと水平線位置とする

〈3･4〉 人の 3 次元位置推定　推定したカメラパラメータと画像中における物体の位置 (u,v) から，世界座標における物体の 3 次元位置 (x,y,z) を推定する．物体の奥行き z は，世界座標系と画像座標系の変換式から，$y=0$ のとき求めたカメラパラメータのカメラの高さと，画像座標における位置関係から次式となる．

$$z = f y_c / \left(f \sin\theta - (v_c - v_b)\cos\theta\right) \cdots\cdots\cdots\cdots\cdots (12)$$

次に，式 (12) から求めた奥行き z と画像中の位置から，x と y を求めることが出来る．世界座標における高さ y は，奥行き z と，カメラの高さ y_c と画像中での高さの関係から次式となる．

$$y = \frac{z(f\sin\theta - (v_c - v)\cos\theta) - f y_c}{(v_c - v)\sin\theta + f\cos\theta} \cdots\cdots\cdots\cdots (13)$$

世界座標における x の推定には，世界座標系から画像座標系に変換する式 (4) を画像座標 u について解くと次式となる．

$$u = \frac{fx + u_c(y\sin\theta + z\cos\theta)}{z} \cdots\cdots\cdots\cdots\cdots (14)$$

式 (14) から，x は次式のようになる．

$$x = \frac{z(u + u_c\cos\theta) - u_c y\sin\theta}{f} \cdots\cdots\cdots\cdots\cdots (15)$$

式 (15) から求めた世界座標における x 座標と，式 (13) と式 (12) から求まる y，z 座標の値を，自己カメラキャリブレーションで求めたカメラパラメータを用いて算出することで，対象物体の 3 次元位置を推定することができる．

4. 実　　験

〈4･1〉 カメラパラメータの推定

〈4･1･1〉 実験概要　本実験では，撮影した映像から人領域のセグメンテーションを行い，その結果から撮影したカメラの自己キャリブレーションを行う．求めるパラメータはカメラの高さ y_c と水平線位置 v_0 で，水平線位置 v_0 と式 (3) からカメラの角度 θ である．また，推定したパラメータを用いて現実世界での 3 次元位置の推定を行う．撮影した映像は，640×360 ピクセル，推定に用いたセグメンテーション結果のサンプル数は 600 である．

〈4･1･2〉 実験結果　図 3(a)，(b) にカメラパラメータの推定に用いた 600 サンプルの人検出結果の分布を示す．表 1 にパラメータ推定結果を示す．表 1 より，セグメンテーションありの場合にカメラ高さが真値との誤差が約 0.3 [cm]，カメラの角度が真値との誤差は約 3.3 [degree] の精度で推定できた．図 3(c)，図 4 に自己キャリブレーションにより推定した水平線位置 v_0，y_c と式 (10) から推定した検出した人の身長，θ と式 (12) から求めた奥行き z の値を示す．提案するカメラの自己キャリブレーション法は，カメラの内部パラメータを推定することはできないが，推定したカメラの高さ y_c と水平線位置 v_0 を用いることで人検出の誤検出抑制やトラッキング精度の向上が期待できる．

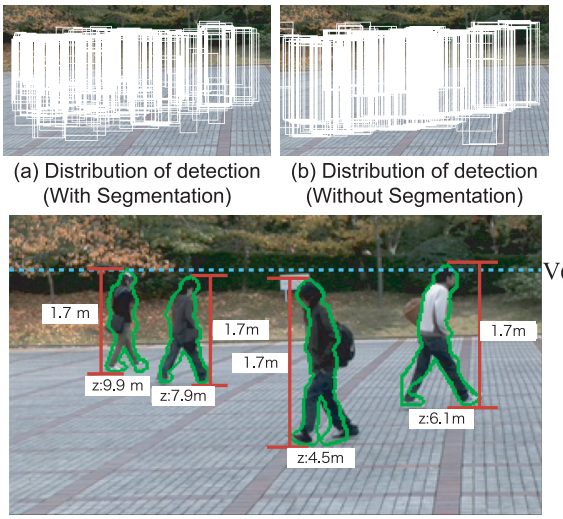

図 3　パラメータ推定結果 1(オリジナルシーケンス)
$y_c = 190.93$[cm], $v_0 = 0.85, \theta = 14.4$[deg]
Fig. 3.　Estimation result 1(Original Sequence)

表 1　カメラパラメータ推定結果
Table 1.　Estimation results of camera parameter

Camera parameter	True value	Segmentation	
		Proposed	Conventional
Camera height yc[cm]	184.6	184.3	214.1
Camera angle θ[degree]	10.0	13.3	18.1

〈4･2〉 人の 3 次元位置の推定　推定したカメラパラメータを用いて，検出した人の世界座標における 3 次元位置を推定をする．推定精度を世界座標の真値と比較することで，推定精度の評価を行う．図 5 に示す 15 点のランドマーク点の画像座標を用いて 3 次元位置の精度評価を行う．

〈4･2･1〉 評価実験　図 6 に 3 次元位置の推定結果を示す．また，表 2 に 3 次元位置推定誤差の x 軸，z 軸の平均と分散を示す．セグメンテーションした結果により推定したカメラパラメータを用いて 3 次元位置を推定することで，推定誤差を小さくすることができた．また，図 6 の結果から提案手法で 3 次元位置をおおまかに推定できていることから，数 m の領域に人がどのように存在しているかなどの位置関係を推定できることが分かる．

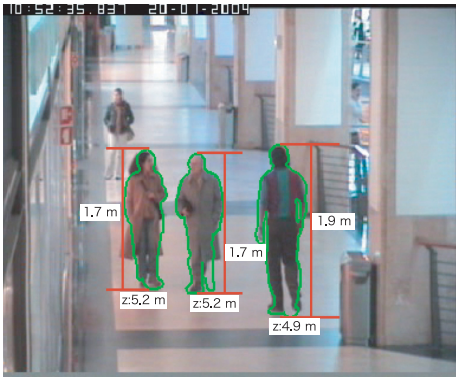

図 4　パラメータ推定結果 2(CAVIAR dataset 2)
$y_c = 350.05$[cm], $v_0 = 1.12, \theta = 29.2$[deg]
Fig. 4.　Estimation result 2(CAVIAR dataset 2)

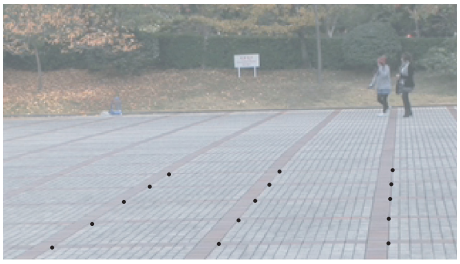

図 5　推定する 3 次元位置 (オリジナルシーケンス)
Fig. 5.　Estimation for 3D positions (Original Sequence)

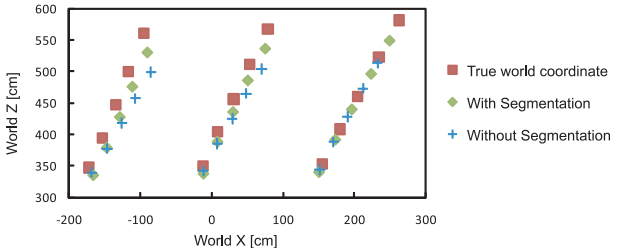

図 6　3 次元位置の推定誤差 (オリジナルシーケンス)
Fig. 6.　Estimation error of 3D positions(Original Sequence)

図 7，図 8 に提案手法を用いた 3 次元位置の推定結果を示す．未校正のカメラから自己キャリブレーションを行い，その結果を用いることで検出した人の 3 次元の位置関係や身長を推定することが可能となる．

表 2　3 次元位置の誤差の平均と分散 [cm]
Table 2.　Average and variance of error in 3D positions [cm]

	With Segmentation		Without Segmentation	
	Average	Variance	Average	Variance
x-axis	5.6	14.1	9.1	76.1
z-axis	20.6	43.9	33.8	573.0

5.　おわりに

本稿では，人検出とセグメンテーション結果に基づく，カメラの自己キャリブレーション手法を提案した．セグメンテーションの結果から画像中の人の高さと，位置を自動的に抽出することで自己キャリブレーションを行い，カメラパラメータの推定を行った．これにより，推定シーンにおける物体の 3 次元位置推定が可能となり，評価実験によりその有効性を確認した．

今後は推定した物体の 3 次元位置情報を用いた高精度な人の追跡を行う予定である．

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） Bo Wu and Ram Nevatia, “Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors”, International Journal of Computer Vision, 75(2), pp.247–226, 2007.

（3） 小川雄三, 藤吉弘亘, “実空間に対応した Master-Slaving による追尾カメラシステム”, 第 9 回画像センシングシンポジウム, June, 2003.

（4） R. Y. Tsai: “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses”, IEEE journal of Robotics and Automation, Vol. RA-3, No.4, pp. 323–344, 1999.

（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） D. Hoiem, A. A. Efros and M. Hebert, “Putting Objects in Perspective”, International Journal of Computer Vision, vol. 80, No. 1, 2008.

（7） 村井陽介，藤吉弘亘, “弱識別器の応答に基づく類似シルエット画像選択による Chamfer Matching を用いた人領域のセグメンテーション”, 第 12 回画像の認識・理解シンポジウム, 2009.

（8） R. E. Schapire and Y. Singer, “Improved Boosting Algorithms Using Confidence-rated Predictions”, Machine Learning, No. 37, pp. 297–336, 1999.

（9） D.M. Gavrila, “Multi-feature Hierarchical Template Matching Using Distance Transforms”, IEEE International Conference on Pattern Recognition, pp. 439–444, 1998.

(10) M. Fischer, and R. Bolles, “Randam sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”, communications of the ACM 24, pp. 381-385, 1981.

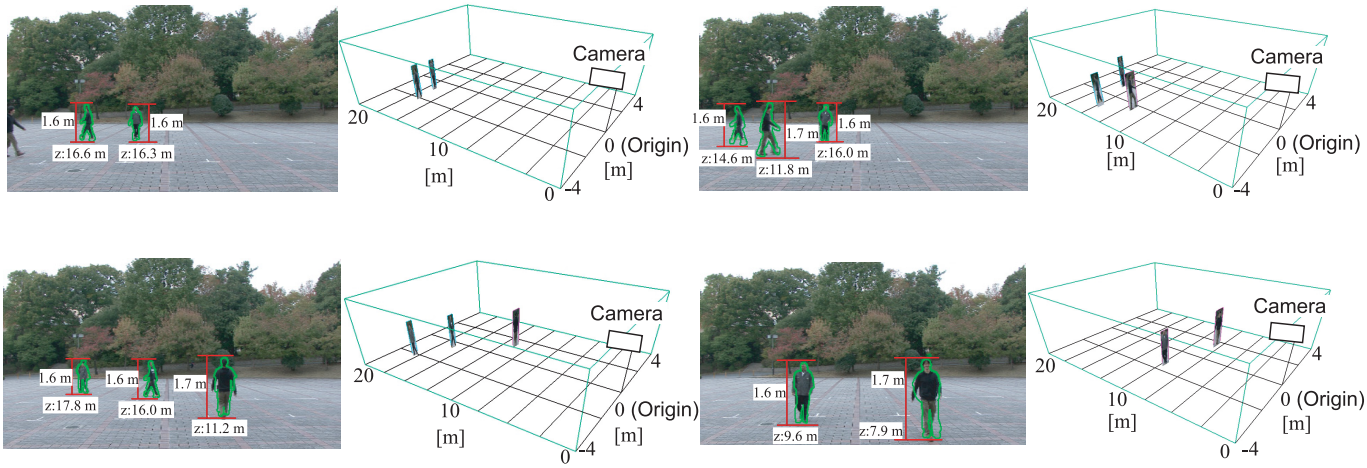

図 7　3 次元位置の推定結果 1(オリジナルシーケンス)
$y_c = 180.5[\mathrm{cm}], v_0 = 0.5, \theta = 0.0[\mathrm{deg}]$
Fig. 7.　Estimated results of 3D positions 1 (Original Sequence)

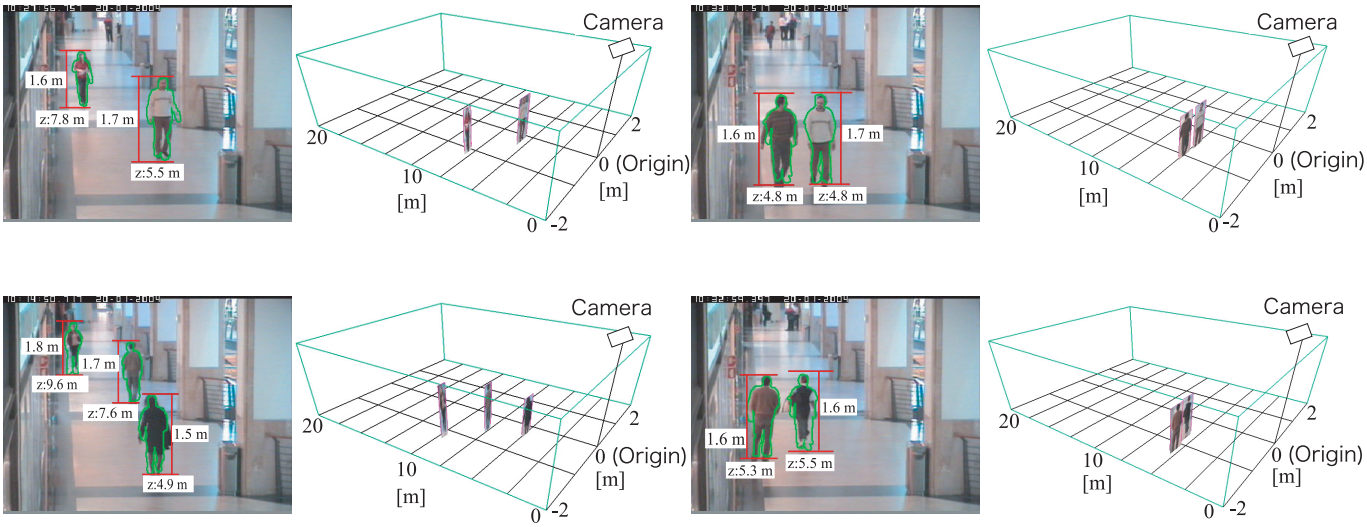

図 8　3 次元位置の推定結果 2(CAVIAR dataset 2)
$y_c = 350.05[\mathrm{cm}], v_0 = 1.12, \theta = 29.2[\mathrm{deg}]$
Fig. 8.　Estimated results of 3D positions 2(CAVIAR dataset 2)