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

村井　陽介†　　藤吉　弘亘†　　伊藤　誠也††　　数井　誠人††

† 中部大学大学院工学研究科 〒 487-8501 愛知県春日井市松本町 1200
†† 株式会社 日立製作所 日立研究所 〒 319-1292 茨城県日立市大みか町 7-1-1
E-mail: †{yoo,hf}@vision.cs.chubu.ac.jp, ††{masaya.itoh.pp,masato.kazui.bq}@hitachi.com

あらまし　人領域のセグメンテーションは，映像サーベイランスにおいて重要な問題である．従来法の多くは，背景差分等により得られた前景領域から個々の人領域を推定するが，カメラの動きの影響で背景画像の取得が困難な場合がある．そこで我々は，Real AdaBoost を用いた人検出結果とシルエット画像を Chamfer Matching によりマッチングすることで背景画像を必要としない人領域のセグメンテーションを実現する．Chamfer Matching ではテンプレート数が増加するほどマッチング精度が低下するという問題があるが，提案手法では Real AdaBoost の弱識別器の応答値に基づいて予めマッチング対象に類似したシルエット画像を選択することでセグメンテーション精度を向上させることが可能となる．また，提案手法の応用として，セグメンテーション結果を用いたカメラ位置の自己キャリブレーション手法について述べる．

キーワード　セグメンテーション，チャンファーマッチング，人検出，自己キャリブレーション

Human-area Segmentation Using Chamfer Matching by Selecting Similar Silhouette Images Based on Weak-classifier Response

Yosuke MURAI†, Hironobu FUJIYOSHI†, Masaya ITOH††, and Masato KAZUI††

† Dept. of Computer Science, Chubu Univ. 1200 Matsumoto, Kasugai, Aichi, 487-8501 Japan
†† Hitachi Research Laboratory, Hitachi, Ltd. Ohmika 7-1-1, Hitachi, Ibaraki, 319-1292 Japan
E-mail: †{yoo,hf}@vision.cs.chubu.ac.jp, ††{masaya.itoh.pp,masato.kazui.bq}@hitachi.com

Abstract　Human-area segmentation is a major issue in video surveillance. Many existing methods estimate individual human areas from the foreground area obtained by background subtraction, but the effects of camera movement can make it difficult to obtain a background image. We have achieved human-area segmentation requiring no background image by using chamfer matching to match the results of human detection using Real AdaBoost with silhouette images. Although accuracy in chamfer matching drops as the number of templates increases, the proposed method enables segmentation accuracy to be improved by selecting silhouette images similar to the matching target beforehand based on response values from weak classifiers in Real AdaBoost. A camera-position self-calibration technique using segmentation results is also presented as an application of the proposed technique.

Key words　segmentation, Chamfer Matching, pedestrian detection, self calibration

1. はじめに

近年，公共施設だけでなくオフィスや一般家庭への監視カメラの普及に伴い，映像サーベイランス技術への需要が高まっている．特に，映像中から人を検出する技術 [1] [2] は，どこに人がいるかを知ることができるため重要である．人領域を正確に知ることは，単眼カメラによるカメラ位置の自己キャリブレーション [12] や身長推定 [3] に利用できるため，人領域の正確なセグメンテーションは重要である．

従来人領域のセグメンテーションは，背景差分に基づく手法が主流であった．Zhao 等は背景差分により得られた前景領域に対し，人の形状モデルやカメラモデル，シーンモデルを導入することで高精度な人検出と追跡を実現している [4] [5]．しかし，このような背景差分ベースの手法では，必ず背景画像が取得できることが前提となるが，実世界においてはカメラの動きの影響等により，背景画像の取得が困難な場合がある．

Leibe 等は物体の局所的な特徴量を用いた part-based approach により人検出を行い，検出結果に対して人の

シルエット画像と Chamfer Matching を行うことで人領域のセグメンテーションを実現している [7]．この手法では背景画像を必要としないため，カメラの動きが発生する映像に対して適用が可能である．Leibe 等の手法では予め用意した約 400 枚のシルエット画像すべてと Chamfer Matching を行う必要がある．しかし，Chamfer Matching はマッチング対象のテンプレート（シルエット画像）数が増加するほどマッチング精度が不安定になるという問題があるため，全てのシルエット画像と Chamfer Matching を行うことはセグメンテーション精度の低下を招く危険性がある．

そこで我々は，予めマッチング対象に類似したシルエット画像を選択し，それらの候補から Chamfer Matching を行う手法を提案する．提案手法では，局所特徴量である HOG と統計的学習手法である Real AdaBoost を用いて人検出を行う．次に，検出された領域とシルエット画像それぞれの Real AdaBoost の応答値を比較することで，検出された領域内の人と類似したシルエット画像を選び出す．最後に，選び出したシルエット画像と人検出結果の領域を Chamfer Matching することで，人領域のセグメンテーションを行う．類似したシルエット画像のみを用いて Chamfer Matching を行うことで，人の形状のバリエーションによるミスマッチングを抑制することができる．これにより，マッチング精度の向上が期待できる．

本稿では，2. 章にて提案手法で用いる Chamfer Matching とその問題点について述べ，3. 章にて提案手法の詳細について述べる．そして，4. 章にて評価実験を行い，提案手法の有効性を示す．また，5. 章にて，提案手法の結果を用いた応用として，カメラ位置の自己キャリブレーション手法について述べる．

2. Chamfer Matching

提案手法は，映像中の人が存在する領域を局所的特徴量である HOG [1] と統計的学習手法である Real AdaBoost [8] により検出し，検出された領域と人のシルエット画像を Chamfer Matching することで人領域のセグメンテーションを実現する．本章では，Chamfer Matching の概要と問題点について述べる．

2.1 Chamfer Matching の概要

Chamfer Matching はテンプレートマッチングの一種であり，入力画像とテンプレート画像間のエッジの相違度に基づいてマッチングを行う手法である [6]．一般的なテンプレートマッチングが画像の色や輝度情報を用いるのに対し，Chamfer Matching ではエッジのような形状情報に基づいたテンプレートマッチングが可能である．

Chamfer Matching の処理の流れを図 1 に示す．まず，入力画像からエッジ特徴を抽出したエッジ画像を生成する．次に，生成したエッジ画像に距離変換処理を行い距離変換画像を得る．距離変換処理とは，エッジ画像において各画素から最も近いエッジ点までの距離をその画

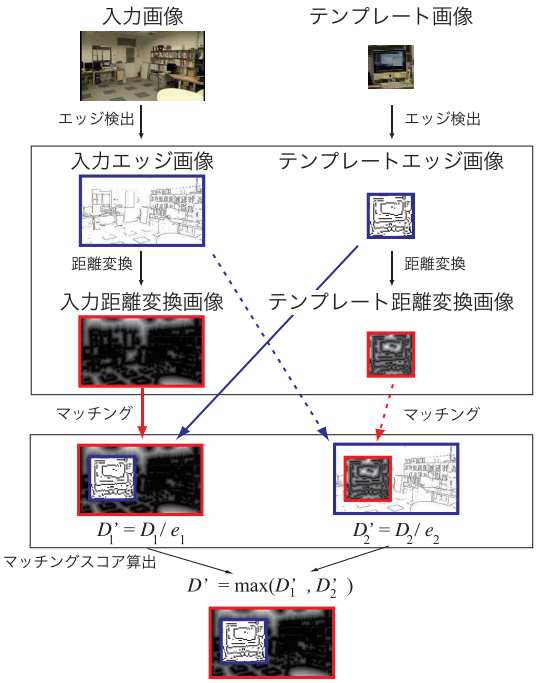


図 1　Chamfer Matching の処理の流れ

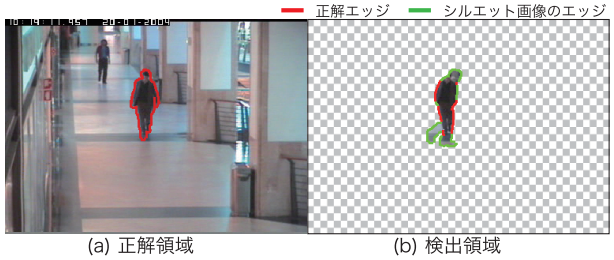
素の輝度値として与える処理である．図 1 では，周囲にエッジが存在する領域ほど黒くなり，エッジが存在しない領域ほど白く表現されている．このようにして生成された入力画像の距離変換画像と，テンプレート画像から生成したエッジ画像を相違度に基づいてマッチングを行う．ここで，入力画像から得られた距離変換画像を $I(x,y)$，テンプレート画像から得られたエッジ画像を $T_{w,h}(i,j) = \{(i,j)|0 \leqq i \leqq w, 0 \leqq j \leqq h\}$ とする．そして，入力画像中の点 (u,v) における入力距離変換画像 $I(x,y)$ とテンプレートエッジ画像 $T_{w,h}(i,j)$ との相違度 $D(u,v)$ を次式により求める．

$$D(u,v) = \sum_{i=0}^{w} \sum_{j=0}^{h} [T_{w,h}(i,j) I(u+i, v+j)] \qquad (1)$$

以上のように求められる相違度 D が最も小さくなる位置 $D(u,v)$ をテンプレートの位置として決定する．しかし，Chamfer Matchiing はエッジ画像と距離変換画像という異なる特徴間の距離に基づいた相違度を用いるため，入力距離変換画像とテンプレートエッジ画像のマッチング結果と，入力エッジ画像とテンプレート距離変換画像のマッチング結果が必ずしも一致するとは限らない．また，テンプレートに含まれるエッジ点数によっても相違度が大きく変化する．そこでこれらの問題を解決するために，入力距離変換画像とテンプレートエッジ画像間の相違度 D_1 と入力エッジ画像とテンプレート距離変換画像間の相違度 D_2 をそれぞれ求め，さらにそのときのエッジ画像に含まれるエッジ点数 e_1，e_2 で相違度を正規化する．正規化された相違度 D_1'，D_2' は次式のように定義する．

$$D_1' = D_1/e_1, \quad D_2' = D_2/e_2$$

以上のように求めた相違度から，次式により最終的なマッチングスコア D' を算出する．



(a) 正解領域　(b) 検出領域

図 2　Chamfer Matching の失敗例

$$D' = \max(D'_1, D'_2) \tag{2}$$

2.2　Chamfer Matching の問題点

Chamfer Matching はエッジ画像と距離変換画像という異なる画像間でエッジの相違度，つまり形状の一致具合からマッチングを行う．そのため人のように形状が様々に変化する物体に対して Chamfer Matching した場合，胴体が完全にテンプレート（本研究ではシルエット画像）と一致すると，手や足などの他の部分が多少一致していなくても最適なマッチングとして誤った結果を出力する危険性がある．特に，シルエット画像を用いた人領域セグメンテーションでは，予め様々な形状のシルエット画像を多数用意する必要があり，シルエット画像のバリエーションが増えるほどミスマッチングする危険性も増加するという問題がある．図 2 は，Chamfer Matching により人領域の側面が誤ったシルエット画像と完全に一致したために失敗した例である．

3.　提 案 手 法

提案手法の流れを図 3 に示す．提案手法は，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．本章ではこれらについて詳しく述べる．

3.1　オフライン処理

3.1.1　人検出器の構築

入力画像全体とシルエット画像を Chamfer Matching によりマッチングすることで人領域の検出を行うことは可能であるが，背景領域の影響によるミスマッチングや処理コストの問題があるため現実的ではない．そこで提案手法では，始めに入力画像に対して人検出を行いマッチング対象の候補領域を検出する．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した特徴量であり，照明変動による影響が少なく，局所的な幾何学的変化に頑健な特徴量である．局所領域はブロックとして表現され，ブロックは複数のセルと呼ばれるさらに小さな局所領域から構成される．特徴量は，セルの勾配方向ごとの勾配強度ヒストグラムを結合した多次元の特徴ベクトルとして表現される．入力画像を 48×120 ピクセルとした場合，HOG

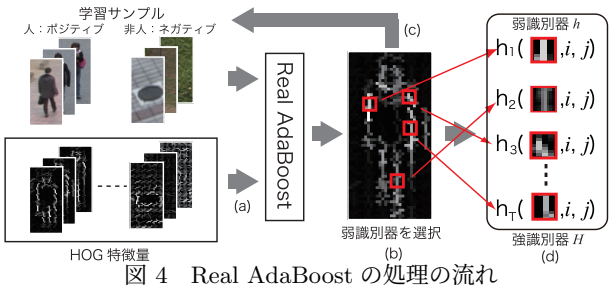


図 4　Real AdaBoost の処理の流れ

特徴量として 8748 次元の特徴ベクトルを得る．

次に，予め用意した学習サンプルから HOG 特徴量を算出し，Real AdaBoost により人検出器の構築を行う．Real AdaBoost とは，boosting 手法の一種であり，学習サンプルから算出した特徴ベクトルから最も識別に有効な特徴ベクトルを弱識別器として学習する手法である．図 4 に Real AdaBoost の処理の流れを示す．予め学習サンプルとしてポジティブサンプル（人領域）とネガティブサンプル（非人領域）を多数用意し，各サンプルに対して同等の重みを設定する．また，各学習サンプルから HOG 特徴量を算出する．次に，算出した HOG 特徴量を Real AdaBoost に入力する (図 4(a))．入力した HOG の各特徴ベクトルから学習サンプルを最も正しく識別することができる特徴ベクトルを弱識別器 $h_t(x)$ として選択する．選択された弱識別器には，一つの HOG 特徴ベクトルとその特徴ベクトルが算出された画像上の位置情報が含まれる (図 4(b))．選択された弱識別器により，学習サンプルの重みを更新する (図 4(c))．このとき，正しく識別できたサンプルの重みは小さくし，誤って識別したサンプルの重みを大きくする．これにより，次の選択時に正しく識別できなかったサンプルについて正しく識別することが可能な弱識別器が選択される．これらの処理を T 回繰り返すことで，複数の弱識別器が選択され最終的に強識別器が構築される (図 4(d))．強識別器 $H(x)$ は次式のように複数の弱識別器の線形結合で表現される．

$$H(x) = \mathrm{sign}\left(\sum_{t=1}^{T} h_t(x)\right) \tag{3}$$

3.1.2　シルエット画像のスコア算出

次に，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．シルエット画像のスコアとは，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/非人を識別した結果を返す．このとき，入力画像を人と正しく判別した場合は弱識別器の応答値として正の実数を返す．一方，入力画像を非人と誤って判別した場合は弱識別器の応答値として負の実数を返す．各弱識別器は，一つの HOG 特徴ベクトルとその画像上での位置情報を保持している．従って，弱識別器の応答値が正の実数のときは，入力画像のある位置において求められた一つの HOG 特

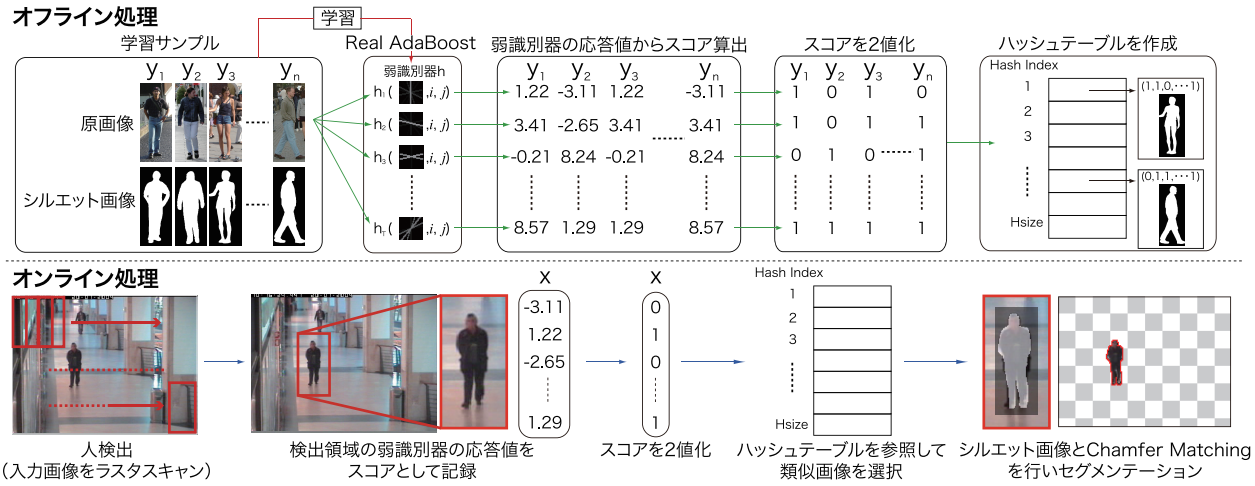


図 3　提案手法の流れ

徴ベクトルが人の勾配情報である確率が高いことを意味している．逆に，弱識別器の応答値が負の実数のときは，その位置における HOG 特徴ベクトルが背景の勾配情報である確率が高いことを意味する．最終的に，T 個の弱識別器に対応した T 次元の特徴ベクトルを得る．ここで得られた 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．

3.1.3　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画像を選択するために使用する．提案手法では，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．ここでは，野口等が提案するハッシュテーブルを用いた近似最近傍探索手法 [10] を用いる．そのために，オフライン処理としてシルエット画像のスコアを特徴ベクトルとしたハッシュテーブルを作成する．

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．

$$u_j = \begin{cases} 1 & if \quad h_j(y) \geq 0 \\ 0 & otherwise \end{cases} \tag{4}$$

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．

$$H_{index} = \left(\sum_{i=1}^{T} u_i 2^i \right) \bmod H_{size} \tag{5}$$

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．以上がオフライン処理である．

3.2　オンライン処理

3.2.1　人　検　出

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，検出領域に対してシルエット画像のマッチングを行う．まず，入力画像に対し検出ウィンドウを左上からスケールを変化させ複数回ラスタスキャンする．このとき，検出ウィンドウ内の領域を人検出対象領域として検出器に入力し，それぞれの領域ごとに人か非人かの識別を行う．人として識別されたウィンドウは，MeanShift クラスタリングによるウィンドウの統合処理を行う [9]．最後に，検出されたウィ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を T 次元の特徴ベクトルとして記録する．このとき，人と判別した弱識別器は正の実数を返し，非人と判別した弱識別器は負の実数を返す．検出された領域の弱識別器の応答に基づく特徴ベクトルと，オフラインで算出した各シルエット画像の弱識別器の応答に基づく特徴ベクトルを比較することで，類似するシルエット画像を選択する．

3.2.2　類似シルエット画像の選択

3.1.2 でシルエット画像のスコアとして記録した特徴ベクトルと 3.2.1 で算出した検出されたウィンドウ領域の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，3.1.3 で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 (4)，式 (5) 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シルエット画像をハッシュテーブルから探索する．探索された N 枚のシルエット画像の特徴ベクトルと，検出領域の特徴ベクトル間の類似度 s_i を次式により求める．

$$s_i = \frac{\mathbf{x} \cdot \mathbf{y}_i}{||\mathbf{x}||||\mathbf{y}_i||} \qquad i = 1, \cdots, N \tag{6}$$

比較する 2 つの特徴ベクトルの各次元の値が共に正である場合対応する領域が人領域である可能性が高く，また，共に負である場合は対応する領域が背景領域である可能性が高い．さらに，比較する 2 つの特徴ベクトルの各次元の値が近いほど，対応する領域の形状が類似している

可能性が高くなる．以上のように，検出領域と類似したシルエット画像を選択する．

図 5 に人検出結果領域の特徴ベクトルから類似したシルエット画像を自動的に選択した例を示す．図 5 を見ると，検出領域の形状と類似したシルエット画像が選択されていることがわかる．このように，予め検出対象領域に類似したシルエット画像を選択することで，Chamfer Matching のマッチング精度の向上が可能となる．

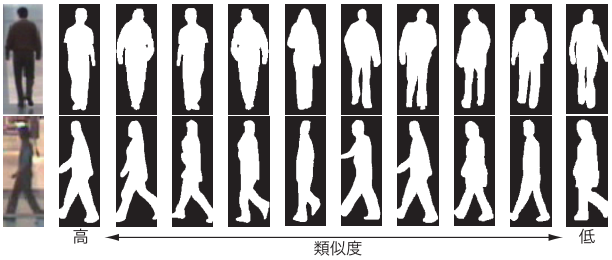


図 5　選択された類似シルエット画像例

3.2.3　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 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．2.1 で述べたように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 D' 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．提案手法では，人検出により検出された領域サイズの 0.8~1.2 倍まで 0.01 刻みでスケールを変化させ Chamfer Matching を行った．最終的に，最も Chamfer Matching の相違度 D' 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．

4.　評 価 実 験

4.1　実 験 概 要

提案手法の有効性を示すためにセグメンテーション精度の評価実験を行う．提案手法である弱識別器の応答値に基づいて予めマッチング対象に類似するシルエット画像を選択することの有効性を評価するために，以下の手法と比較を行う．

比較手法　作成したすべてのシルエット画像とマッチングを行う手法

提案手法　弱識別器の応答値に基づいて予め選択された類似シルエット画像とマッチングを行う手法

提案手法では，類似シルエット画像を 10 枚選択し，Chamfer Matching を行った．評価を行うシーケンスは CAVIAR Test Case Scenarios [11] のデータセット 2 から，正面方向への人の動きがあるシーン 300 フレーム (図 6(a)) と横方向への人の動きがあるシーン 300 フレーム (図 6(b)) を使用する．正面方向の動きがあるシーンでは人のスケール変化が大きく発生し，横方向の動きがあるシーンでは人の形状の変化が大きく発生する．評価方法として各フレームごとに再現率（Recall），適合率（Precision）および F 値（F-measure）を算出しシーケンス全体での平均値により評価を行う．セグメンテーション結果から，検出結果で正解した人領域と背景領域の面積をそれぞれ True Positive(TP)，True Negative(TN) とする．また，背景領域を人領域として誤検出した面積を False Positive(FP)，人領域を背景領域と誤検出した面積を False Negative(FN) とする．このとき Recall，Precision，F-measure をそれぞれ以下のように定義する．

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\mathrm{F-measure} = \frac{2 \times \mathrm{Recall} \times \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{9}$$

評価シーケンス全フレームの正解領域は手動により作成した．また，シルエット画像として人検出用学習サンプルから代表的な形状をしたサンプルを 800 枚選び出し作成した (図 7)．実験に使用する PC は，Intel Core 2 Duo 3.06GHz，メモリ 4GB である．

図 6　使用するシーケンス

図 7　使用するシルエット画像例

4.2　実 験 結 果

表 1，表 2 にシーケンスごとのセグメンテーション精度評価結果，図 8 にセグメンテーション結果例を示す．提案手法（シルエット選択あり）はシルエット選択数 10 フレームの結果である．また，比較手法（シルエット選択なし）は全シルエット数 800 枚とのマッチング結果である．

結果を見ると，正面方向のシーケンスに対して Recall で 0.02，Precision で 0.08 の精度の向上が確認できた．

表 1　セグメンテーション精度の比較 (正面)

	シルエット選択なし	シルエット選択あり
Recall	0.79	0.81
Precision	0.82	0.90
F-measure	0.80	0.85

表 2　セグメンテーション精度の比較 (横)

	シルエット選択なし	シルエット選択あり
Recall	0.72	0.73
Precision	0.71	0.76
F-measure	0.71	0.75

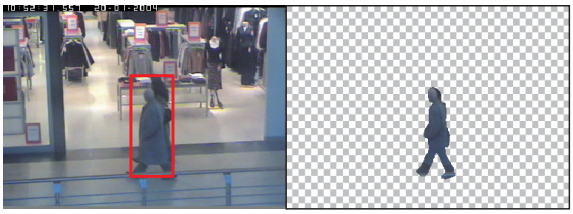
また，F 値で 0.05 の精度の向上が確認できた．一方，横方向のシーケンスに対して，Recall で 0.01，Precision で 0.05 の精度の向上が確認できた．また，F 値で 0.04 の精度の向上が確認できた．これらの結果から，シルエット画像を選択することで，False Positive を抑制することが可能であると言える．さらに，シルエット選択なしの場合，1 人のセグメンテーションにつき平均約 352.1s の処理時間に対し，選択ありの場合約 2.1s の処理時間であり，約 167 倍の高速化を実現した．次に，セグメンテーション結果例を見ると，大きなスケール変化が発生する正面方向のシーケンスにおいて比較的精度良くセグメンテーションできていることがわかる．また，人の形状変化が正面方向よりも発生する横方向のシーケンスにおいても精度良くセグメンテーションできていることがわかる．以上の結果から，提案手法の有効性が確認できる．

次に，セグメンテーション失敗例から提案手法の問題点について考察する．提案手法は，人検出器により検出された領域に対して Chamfer Matching を行うことで人領域のセグメンテーションを行うため，人検出結果がセグメンテーション結果に影響する．CAVIAR Test Case Scenarios から切り出した positive サンプル 1003 枚，negative サンプル 1134 枚に対して，我々が構築した人検出器を用いて検出実験を行った結果，誤検出率 10% のとき検出率約 87% であった．このとき，人検出器で検出に失敗した場合，その領域のセグメンテーションを行うことはできない (図 9(a))．また，人同士の重なりが大きい場合，その領域が 1 人の人物領域として検出されてしまうため，個々のシルエットを検出できず誤ったセグメンテーション結果になってしまう問題がある (図 9(b))．今後の課題として，人検出器の精度向上や人検出段階での失敗を補完するアルゴリズムの開発が必要である．

5.　カメラ位置の自己キャリブレーション

物体検出や追跡など映像サーベイランスにおいて，カメラの位置情報は精度向上のために有用な情報である．しかし，現実世界において撮影しているカメラの位置情報は必ずしも既知であるとは限らない．そこで我々は，提案手法の応用としてセグメンテーション結果を用いたカメラ位置の自己キャリブレーションについて述べる．

(a) 人領域の未検出によるセグメンテーション失敗例 (緑楕円：未検出領域)



(b) 人領域の重なりによるセグメンテーション失敗例

図 9　セグメンテーション失敗例

5.1　カメラ位置と人の高さの関係

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位置の自己キャリブレーションを行う．まず，文献 [12] を参考に，カメラ位置と画像中の人の高さの関係を定式化する．図 10 に，カメラ位置と人の高さの関係を示す．

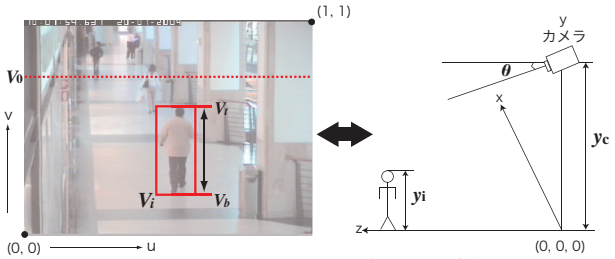


図 10　カメラと人の高さの関係

画像の左下の座標を $(0,0)$，右上の座標を $(1,1)$ と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，カメラ中心を (u_c,v_c)，カメラの高さを y_c とする．世界座標系は，カメラ位置を中心として，$z_c=0, x_c=0$ とし，接地平面を $y=0$ と定義する．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．

$$\theta = 2\arctan\frac{v_c - v_0}{2f} \tag{10}$$

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．このとき，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．

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z}\begin{bmatrix} f & 0 & u_c \\ 0 & f & v_c \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & y_c \\ 0 & \sin\theta & \cos\theta & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{11}$$

式 (11) を物体の高さ y について解くと次式となる．

$$y = \frac{z(f\sin\theta - (v_c - v)\cos\theta) - fy_c}{(v_c - v)\sin\theta + f\cos\theta} \tag{12}$$

(a) 原画像　(b) 比較手法（シルエット全探索）　(c) 提案手法（シルエット選択）　(d) 原画像　(e) 比較手法（シルエット全探索）　(f) 提案手法（シルエット選択）

図 8　セグメンテーション結果例 (a),(b),(c)：正面方向　(d),(e),(f)：横方向

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 10）．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．

$$z = f y_c / \left(f \sin\theta - (v_c - v_b)\cos\theta \right) \tag{13}$$

式 (12) と式 (13) から，物体の高さ y は次式のようになる．

$$y = \frac{f y_c (f\sin\theta - (v_c - v_t)\cos\theta) / (f\sin\theta - (v_c - v_b)\cos\theta) - f y_c}{(v_c - v_t)\sin\theta + f\cos\theta} \tag{14}$$

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 (14) は以下のように表現できる．

$$y \approx y_c \frac{v_t - v_b}{v_0 - v_b} / \left(1 + (v_c - v_0)(v_c - v_t)/f^2 \right) \tag{15}$$

分母の $(v_c - v_0)(v_c - v_t)/f^2$ は，チルト角が小さいとき $v_c - v_0 \approx 0$ なので，$(v_c - v_0)(v_c - v_t)/f^2 \approx 0$ と近似できる．従って，式 (15) は以下のようになる．

$$y \approx y_c \frac{v_t - v_b}{v_0 - v_b} \tag{16}$$

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なので 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．

$$y_i \approx y_c \frac{h_i}{v_0 - v_i} \tag{17}$$

式 (17) を変形し，同時座標で表現すると次式のようになる．

$$\begin{bmatrix} v_0 \\ \\ y_c \end{bmatrix} = \begin{bmatrix} y_i & -h_i \\ \vdots & \vdots \\ y_n & -h_n \end{bmatrix}^{-1} \begin{bmatrix} y_i v_i \\ \vdots \\ y_n v_n \end{bmatrix} \tag{18}$$

式 (18) からカメラの高さ y_c と水平線の位置 v_0 を求める．

5.2　パラメータ推定法

式 (18) において，既知のパラメータは画像座標系での人の高さ h_i と人領域の設置位置 v_i のみである．カメラの高さ y_c と水平線の位置 v_0 を求めるためには，世界座標系でのそれぞれの人の高さ y_i が必要である．しかし，今回与えられている映像からそれぞれの人の高さの真値を知ることは不可能である．そこで，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た．

カメラの高さとチルト角の推定法として RANSAC を用いる．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

ここで，乱数で与えた身長を各サンプルの真値と仮定し，推定値から算出した各サンプルの身長と真値の差を誤差としてパラメータ推定を行う．

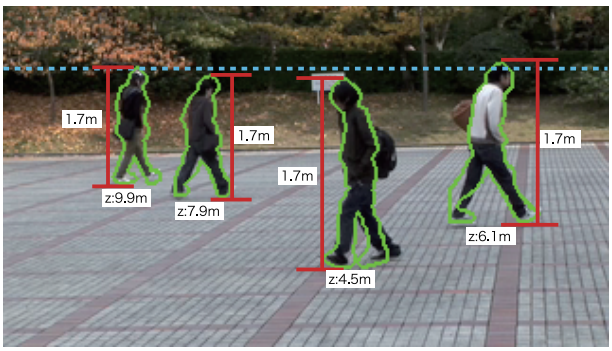


図 11　パラメータ推定結果 1(オリジナルシーケンス)
$y_c = 190.93[\mathrm{cm}], v_0 = 0.85, \theta = 14.4[\mathrm{deg}]$

5.3　パラメータ推定実験

5.3.1　実 験 概 要

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位置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 である．さらに，水平線位置 v_0 と式 (10) からカメラの角度 θ を求める．また，推定したパラメータを用いて現実世界での人の身長推定を行う．撮影した映像は，640×360 ピクセル，約 16,000 フレーム （約 70 人）である．

5.3.2　実 験 結 果

表 3 にパラメータ推定結果を示す．表 3 より，おおよそ真値に近いカメラの高さと角度が推定できていることがわかる．提案するカメラ位置の自己キャリブレーション法は，内部パラメータまで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 11，図 12 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 (17) から推定した身長，θ と式 (13) から求めた奥行き z の値を示す．図 12 はチルト角が鋭角な場合の推定例である．推定結果を見ると，水平線位置は画面外に存在している ($v_0 = 1.12$) が，お互いの身長の高低や位置関係はおおよそ求められていることがわかる．式 (15) はカメラのチルト角が小さい場合を仮定した式でありチルト角が鋭角な場合では高精度な推定は困難であるが，大まかな推定は可能であるため実用上の影響はない．以上のように，自己キャリブレーションによって求めたカメラ位置を用いることで，おおよその身長を推定することが可能である．

表 3　パラメータ推定結果

	真値	推定結果	誤差
カメラの高さ yc [cm]	184.6	190.9	6.3
カメラの角度 θ[degree]	10	14.3	4.3

6.　ま　と　め

本稿では，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手法を提案した．提案手法は，予めマッチング対象に類似したシルエット画像を選択することで，Chamfer Matching のミスマッチングを抑制し，マッチング精度の向上を実現した．また，セグメンテーション結果を用いたカメラ位置の自己キャリブレーションを行い，良好な結果を得た．今後は，シルエット画像の形状に合わせたセグメンテーション結果から，より正確な検出対象の領域セグメンテーションを行う予定である．

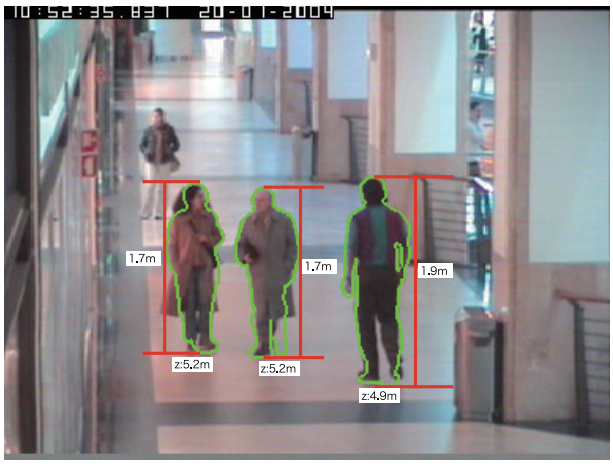


図 12　パラメータ推定結果 2(CAVIAR dataset 2)
$y_c = 350.05[\mathrm{cm}], v_0 = 1.12, \theta = 29.2[\mathrm{deg}]$

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] T. Zhao, R. Nevatia and F. Lv, “Segmentation and Tracking of Multiple Humans in Complex Situations”, IEEE Conference on Computer Vision and Pattern Recognition, 2001.
[5] T. Zhao and R. Nevatia, “Bayesian Human Segmentation in Crowded Situations”, IEEE Conference on Computer Vision and Pattern Recognition, 2003.
[6] D.M. Gavrila, “Multi-feature Hierarchical Template Matching Using Distance Transforms”, IEEE International Conference on Pattern Recognition, pp. 439–444, 1998.
[7] B. Leibe, E. Seemann and B. Schiele, “Pedestrian Detection in Crowded Scenes”, Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. 878–885, 2005.
[8] R. E. Schapire and Y. Singer, “Improved Boosting Algorithms Using Confidence-rated Predictions”, Machine Learning, No. 37, pp. 297–336, 1999.
[9] D. Comaniciu and P. Meer, “Mean Shift Analysis and Applications”, IEEE International Conference on Computer Vision, pp. 1197–1203, 1999.
[10] 野口和人, 中居友弘, 黄瀬浩一, 岩村雅一, “特徴ベクトルの近傍探索と物体認識の効率に関する実験的検討”, 情報処理学会研究報告 (CVIM2006), No. 93, pp.57–64, 2006.
[11] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
[12] D. Hoiem, A. A. Efros and M. Hebert, “Putting Objects in Perspective”, International Journal of Computer Vision, vol. 80, No. 1, 2008.