論　文

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[†]　安藤　寛哲[†]　藤吉　弘亘[†]　伊藤　誠也[††]
数井　誠人[†††]

Human-Area Segmentation Using Chamfer Matching by Selecting Similar Silhouette Images Based on Weak-Classifier Response

Yosuke MURAI[†], Hiroaki ANDO[†], Hironobu FUJIYOSHI[†], Masaya ITOH[††], and Masato KAZUI[†††]

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あらかじ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異なるカメラ間での人物同定について述べる．

キーワード　セグメンテーション，Chamfer Matching，人検出，人物同定

1. まえがき

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]〜[3], [11] 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．また，人領域を正確に知ることは，身長推定 [4] や異なるカメラ間での人物同定 [5], [6] に利用できるため，人領域の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 ら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 [7], [8]．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．

一方，背景差分を必要としない人検出ベースのセグメンテーション手法も複数提案されている [2], [3], [9]〜[11]．これらの手法では，背景差分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．Gavrila らは，様々な姿勢のエッジ画像のテンプレートから階層的な木を作成し，画像全体に対して Chamfer Matching を行うことで人検出を実現している [9], [10]．しかし，このような大域的なマッチングによる人検出では，背景の影響を強く受けるという問題がある．一方，Wu らは局所的な特徴量である Edgelet と boosting により，人検出とセグメンテーションを同時に実現している [3]．Wu らの手法では，弱識別器ごとに人検出に有効な Edgelet 特徴と，その Edgelet の形状から推定される人領域の部分的なシルエット情報を保持しているため，人検出とセグメンテーション

† 中部大学大学院工学研究科，春日井市
Dept. of Computer Science, Chubu Univ., 1200 Matsumoto, Kasugai-shi, 487–8501 Japan
†† (株) 日立製作所日立研究所，日立市
Hitachi Research Laboratory, Hitachi, Ltd., 7–1–1 Ohmika, Hitachi-shi, 319–1292 Japan
††† (株) サムスン横浜研究所，横浜市
Samsung Yokohama Research Institute Co., Ltd., Sugasawa-cho, Tsurumi-ku, Yokohama-shi, 230–0027 Japan

 © (社) 電子情報通信学会 2011

を同時に行うことが可能である．しかし，弱識別器がもつ部分的なシルエット情報のみから人領域の推定を行うため，側面方向のシーンのような人物の形状に大きな変化が発生しやすい場合では，手や足の領域の正確なセグメンテーションが困難である．人検出とセグメンテーションを同時に行う手法に対し，Leibe ら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 [11]．Leibe らの手法では，シルエット画像と Chamfer Matching を用いてセグメンテーションを行っているため人の形状の変化に対して比較的高精度なセグメンテーションが可能であるが，あらかじめ用意した 400 枚のシルエット画像すべてと Chamfer Matching を行う必要がある．Chamfer Matching は形状の一致具合からマッチングスコアを算出するため，人物のように形状が様々に変化する非剛体をマッチング対象とした場合，姿勢によっては本来異なるシルエットであっても対象の一部分が完全にテンプレートと一致してしまい，他の領域が多少一致していなくても最適なマッチングであるという誤った結果を出力することがある．したがって，シルエット画像が増加するほど類似する形状が増加するためミスマッチングの可能性が高くなる．このように，シルエット画像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り，すべ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，あらかじ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 800 枚のシルエット画像それぞれの Real AdaBoost で構築した各弱識別器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．

本論文では，**2.** 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.** にて提案手法の詳細について述べる．そして，**4.** にて評価実験を行い，提案手法の有効性を示す．また，**5.** にて，提案手法の結果を用いた応用として，異なるカメラ間での人物同定について述べる．

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 [12] 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 [9]．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

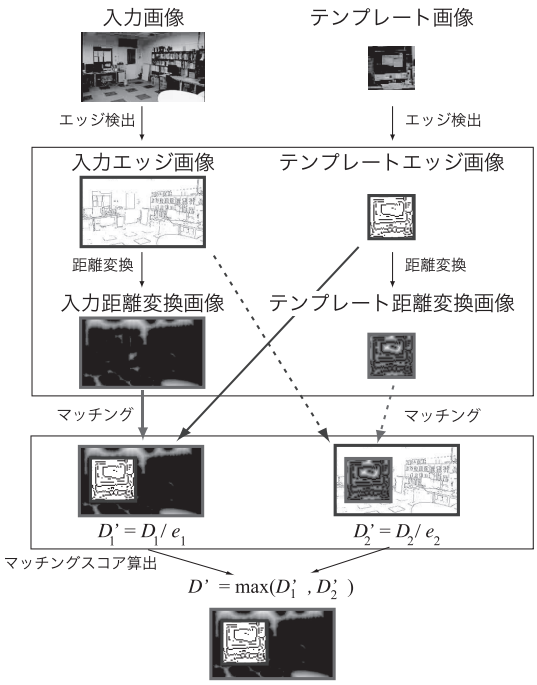


図 1　Chamfer Matching の処理の流れ
Fig. 1　Overview of Chamfer Matching.

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 $T_{w,h}(i,j) = \{(i,j)|0 \leq i \leq w, 0 \leq j \leq h\}$ 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．

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

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，更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．

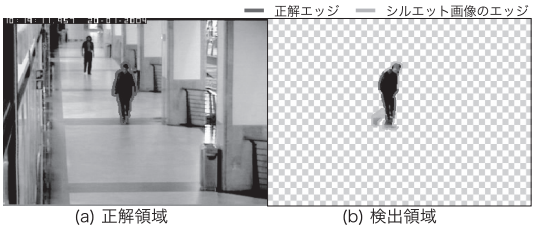
$$D'_1 = D_1/e_1, \quad D'_2 = D_2/e_2$$

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．

$$D' = \max(D'_1, D'_2) \quad (2)$$

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，あらかじ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．



(a) 正解領域　(b) 検出領域

図 2　Chamfer Matching の失敗例
Fig. 2　Exsample of failure of Chamfer Matching.

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，はじ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 [12] による人検出器を用いる．HOG 特徴量は一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

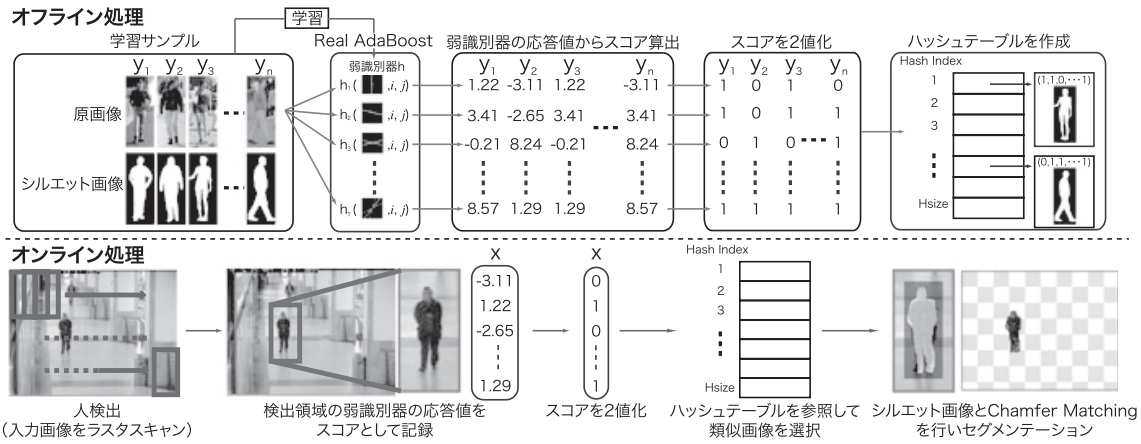


図 3　提案手法の流れ
Fig. 3　Overview of proposed method.

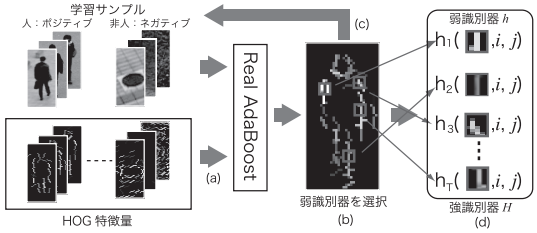


図 4　Real AdaBoost の処理の流れ
Fig. 4　Overview of Real AdaBoost.

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更に小さな局所領域から構成される．特徴量は，セルのこう配方向ごとのこう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．次に，あらかじ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．あらかじ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 w_t を設定する．次に，ポジティブクラスとネガティブクラスの HOG 特徴量の重み付き頻度ヒストグラム W_+, W_- を作成する．HOG 特徴量は，ブロック領域ごとにそのブロック内のこう配強度ヒストグラムのノルムにより正規化をするため，特徴量の値は [0, 1] の範囲で算出される．これを J 分割し，ヒストグラムの対応する量子化区間番号 j に学習サンプルの重み $w_t(i)$ を加算することで作成する．

$$W_+^j = \sum_{i:j \in J \wedge y_i = +1} w_t(i) \tag{3}$$

$$W_-^j = \sum_{i:j \in J \wedge y_i = -1} w_t(i) \tag{4}$$

ここで，t は学習回数，i は学習サンプルの番号，y_i はクラスラベル $y \in \{+1, -1\}$ である．本研究では，実験により量子化数 J を 32 とした．HOG 特徴量の重み付き頻度ヒストグラムは，HOG 特徴の各次元ごとに作成する．作成した HOG 特徴量の重み付き頻度ヒストグラムの評価値 Z を算出する．

$$Z_t = 2 \sum_j \sqrt{W_+^j W_-^j} \tag{5}$$

Z はヒストグラム間の Bhattacharyya 距離であり，Z の値が小さいほどポジティブクラスとネガティブクラスの分布が分離していることを意味する．すべての弱識別器候補 $\mathcal{H}$ の中より最も評価値 Z_t が小さい弱識別器候補を弱識別器 h_t として次式により選択する．

$$h_t = \arg\min_{h \in \mathcal{H}} Z_t \tag{6}$$

上記の処理は，HOG 特徴量の各次元ごとに作成した重み付き頻度ヒストグラムを弱識別器候補 $\mathcal{H}$ とし，弱識別器候補のプールから識別に最適な重み付き頻度ヒストグラム W_+，W_- を選択することを意味する．し

たがって，選択された弱識別器には識別に最適な次元の 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)）．RealAdaBoost における弱識別器の応答値は次式のように定義される．

$$h_t(x) = \frac{1}{2}\log\frac{W_+^j + \epsilon}{W_-^j + \epsilon} \tag{7}$$

ここで，ϵ は分母が 0 の場合に計算不能になることを防ぐための係数であり，学習サンプルの枚数に応じて決定する．j は，[0,1] の範囲に正規化された HOG 特徴量 x の値を J 分割した際の量子化区間番号である．弱識別器の出力は，学習サンプルより算出した HOG 特徴量 x の値から重み付き頻度ヒストグラムの量子化区間番号 j を算出し，それに対応した重み付き頻度ヒストグラム W_+^j，W_-^j から式 (7) により算出する．強識別器 $H(x)$ は，式 (7) で定義した複数の弱識別器の線形結合として次式のように表現される．

$$H(x) = \mathrm{sign}\left(\sum_{t=1}^{T} h_t(x)\right) \tag{8}$$

最終的に，強識別器の出力値がしきい値以上の場合は人，しきい値未満の場合には非人として識別する．本研究では，強識別器の出力値が正の実数の場合は人，負の実数の場合は非人として識別する．

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．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．各弱識別器の保持する位置に対応する入力画像上の位置から得られた HOG 特徴ベクトルを用いて識別をした結果，人と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，非人と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．したがって，弱識別器の応答値が正の実数のときは，各弱識別器が保持する位置情報と対応する入力画像上の位置において求められた一つの HOG 特徴ベクトルが人のこう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こう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，野口ら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 [14] 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．

$$u_j = \begin{cases} 1 & \text{if} \quad h_j(y) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

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．

$$H_{index} = \left(\sum_{i=1}^{T} u_i 2^i\right) \bmod H_{size} \tag{10}$$

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 [13]．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 (9)，式 (10) 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も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．

$$s_i = \frac{\mathbf{x} \cdot \mathbf{y}_i}{||\mathbf{x}||||\mathbf{y}_i||} \quad i = 1, \cdots, N \tag{11}$$

比較する二つの特徴ベクトルの各次元の値がともに正である場合，対応する領域が人領域である可能性が高く，また，とも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．更に，比較する二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分かる．図 6 に，選択されたシルエット画像のスコアの例を示す．図 6 を見ると，類似

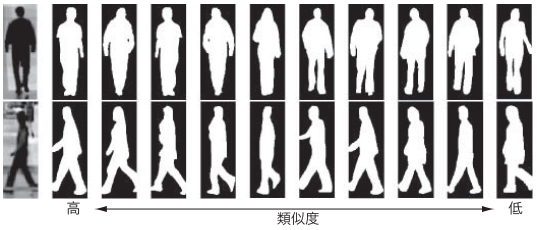


図 5 選択された類似シルエット画像例
Fig. 5 Example of selected similar silhouette image.

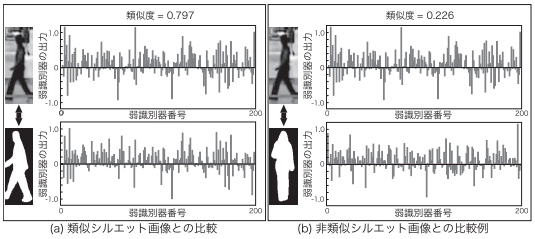


図 6 入力画像とシルエット画像のスコアの比較
Fig. 6 Comparison of scores of input image and silhouette image.

シルエット画像として選択された画像のスコアは入力画像のスコアと類似していることが分かる．一方，形状の類似していないシルエット画像のスコアは入力画像のスコアと類似していないことが分かる．このように，弱識別器の出力値から求めたスコアを用いて，あらかじ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．

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も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．実験 1 として，提案手法である弱識別器の応答値に基づいてあらかじ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あらかじ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，800 枚のシルエット画像から類似シルエット画像を 10 枚選択し，Chamfer Matching を行った．比較手法では，800 枚のシルエット画像すべてと Chamfer Matching を行った．評価を行うシーケンスは CAVIAR Test Case Scenarios [15] 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 7 (a)）と横方向への人の動きがあるシーン 300 フレーム（図 7 (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．次に，実験 2 として，背景差分ベースの手法では困難なシーンに対するセグメンテーションを行う．評価を行うシーケンスは，PETS2001 [16] のデータセットから照明変化のあるシーン 300 フレーム（図 7 (c)）と，独自に撮影したカメラのパンニングが発生するシーン 300 フレーム（図 7 (d)）を使用する．評価方法として各フレームごとに再現率（Recall），適合率（Precision）及び 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．

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{14}$$

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 8）．実験に使用する PC は，Intel Core 2 Duo 3.06 GHz，メモリ 4 GByte である．

4.2 実験結果 1

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 9 にセグメンテーション結果例を示

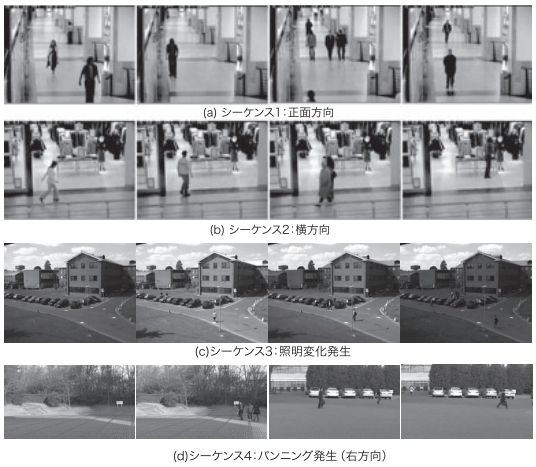
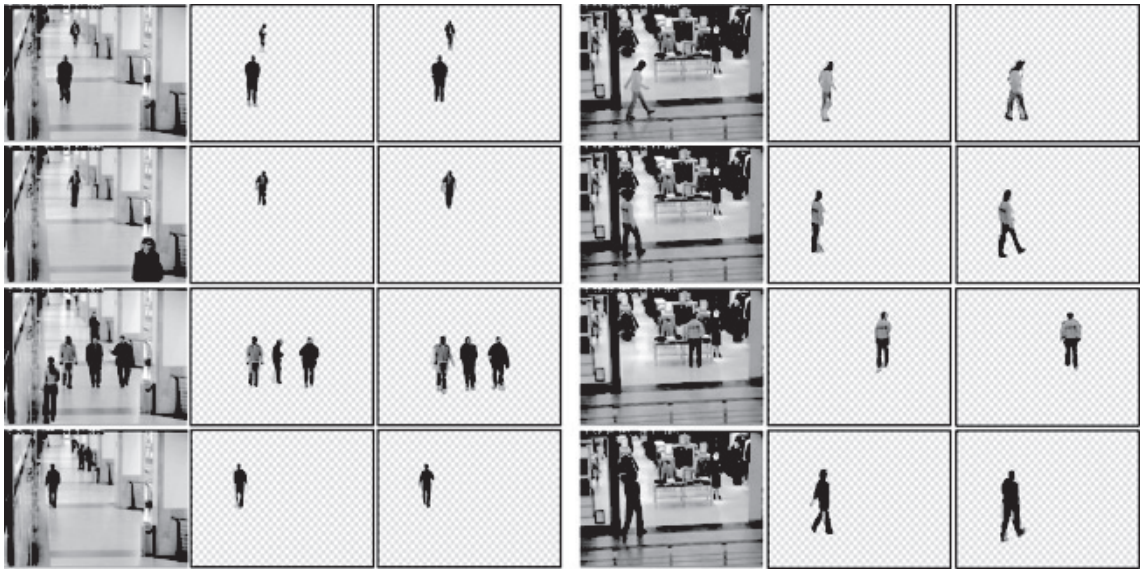


図 7 使用するシーケンス
Fig. 7 Example of sequences to use.

図 8 使用するシルエット画像例
Fig. 8 Example of silhouette images to use.

表 1 セグメンテーション精度の比較（正面）
Table 1 Comparison of segmentation accuracy. (frontal view)

	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



(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 9　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
Fig. 9　Segmentation results (a), (b), (c): frontal view (d), (e), (f): side view.

表 2　セグメンテーション精度の比較（横）
Table 2　Comparison of segmentation accuracy. (side view)

	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

表 3　照明変化・カメラパニング発生シーンにおけるセグメンテーション評価結果
Table 3　Evaluation result of segmentation in illumination change and camera panning scene.

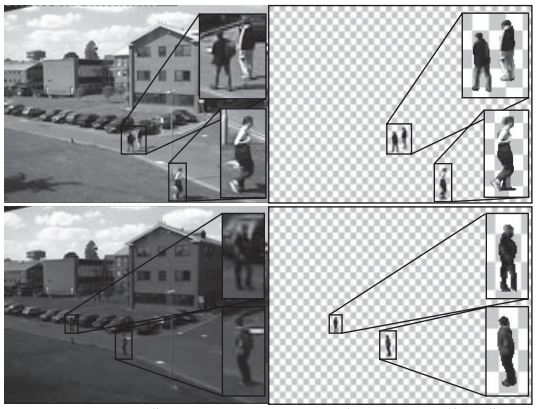
	照明変化発生シーン	パンニング発生シーン
Recall	0.72	0.79
Precision	0.76	0.85
F-measure	0.73	0.81

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．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いえる．更に，人検出後の 1 人の人物に対し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たときのセグメンテーションの処理時間は，シルエット選択なしの場合平均約 65 s の処理時間に対し，選択ありの場合平均約 0.9 s の処理時間であり，約 72 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分かる．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分かる．以上の結果から，提案手法の有効性が確認できる．

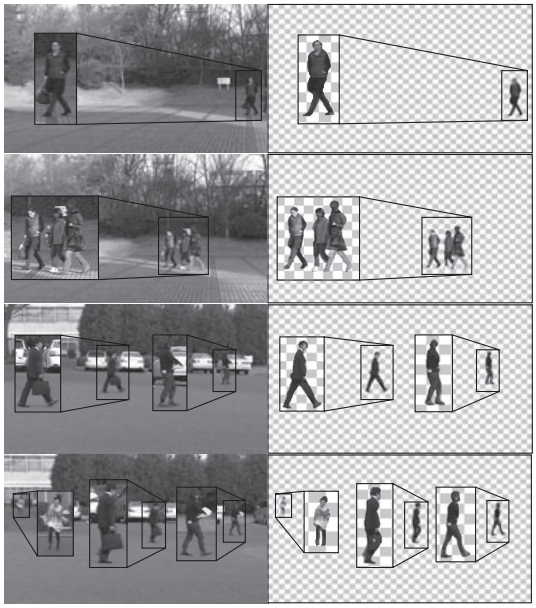
4.3　実験結果 2

表 3 に照明変化，パンニング発生シーンにおけるセグメンテーション精度評価結果，図 10，図 11 にセグメンテーション結果例を示す．結果を見ると，提案手法は実験 1 におけるセグメンテーション精度とほぼ同等の結果が得られていることが分かる．提案手法では，照明変化やカメラの動きに頑健なエッジ特徴に基づく人検出とセグメンテーションを行っているため，図 7 (c)，(d) のような背景差分が困難なシーンにおい



(a) 原画像　(b) セグメンテーション結果画像

図 10　照明変化発生シーンにおけるセグメンテーション結果例
Fig. 10　Example of segmentation result in illumination change scene.



(a) 原画像　(b) セグメンテーション結果画像

図 11　パンニング発生シーンにおけるセグメンテーション結果例
Fig. 11　Example of segmentation result in camera panning scene.

てもセグメンテーションが可能である．

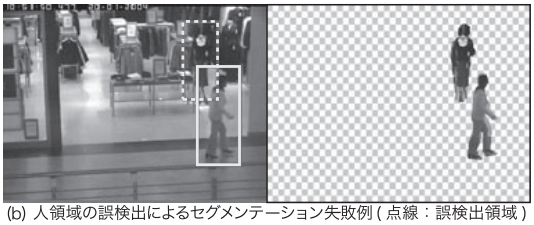
4.4 考　　察

提案手法では，ハッシュテーブルにより類似シルエット画像選択の候補を絞ることで高速化を実現している．しかし，ハッシュテーブルによる候補の選択は近似最適解であるため，類似シルエット画像選択の精度が低下する可能性がある．そこで，最近傍探索による類似シルエット画像選択とハッシュテーブルによる類似シルエット画像選択の精度の比較を行う．比較は以下の手順で行う．まず，手動で切り出した 300 枚の各人物画像に対し，最近傍探索により最も類似度の高いシルエット画像を 1 枚選択し正解画像とする．次に，ハッシュテーブルにより 300 枚の各人物画像の類似シルエット画像を 10 枚選択する．このとき，10 枚の類似シルエット画像内に，最近傍探索により選択した正解画像が含まれる率をハッシュテーブルの精度として評価する．表 4 に評価結果を示す．評価結果を見ると，ハッシュテーブルによる選択法では選択失敗をするケースが発生したが，全体として 98% という高精度で類似シルエット画像の探索ができた．選択失敗をした 6 枚についても，ハッシュテーブルにより上位 15 枚を選択した場合では正解することができた．また，処理時間についても，最近傍探索では 10 枚選択に約 20 ms 必要であったのに対し，ハッシュテーブルでは約 9 ms で選択可能であり，処理時間を約 1/2 に抑えることが可能である．ハッシュテーブルの特長として，探索対象数が増加しても処理時間の増加を抑えることが可能という点があり，シルエット画像枚数を現在の 800 枚から増やしたとしても処理時間を抑えることが期待できる．以上より，ハッシュテーブルによる類似画像選択が有効であると考えられる．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，人検出で未検出や誤検出した場合，その領域のセグメンテーションを正しく行うことはできない（図 12 (a) は未検出による失敗例，図 12 (b) は誤検出による失敗例）．また，人同士の重なりが大きい場合，その領域

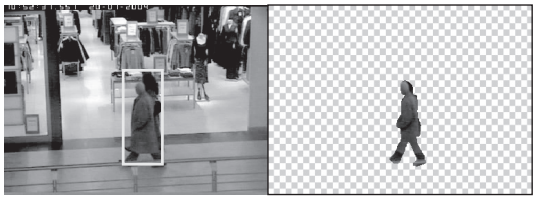
表 4　類似シルエット画像選択精度評価結果
Table 4　Evaluation result of similar silhouette image selection accuracy.

正解画像選択率 [%]	選択失敗画像枚数 [枚]
98.0	6

(a) 人領域の未検出によるセグメンテーション失敗例 (楕円：未検出領域)



(b) 人領域の誤検出によるセグメンテーション失敗例 (点線：誤検出領域)



(c) 人領域の重なりによるセグメンテーション失敗例

図 12　セグメンテーション失敗例
Fig. 12　Example of failure of segmentation.

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 12 (c)）．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.　セグメンテーション結果を用いた異なるカメラ間での人物同定

異なるカメラ間での人物同定は，監視カメラによる不審者の特定や追跡等，セキュリティ分野への応用が期待されている．人物同定では，映像から人領域のみをセグメンテーションすることが重要である．従来，人物同定法では人領域のセグメンテーションに背景差分ベースを用いる手法 [5], [6] が主流であるが，実世界においてはカメラの動きや照明の影響等により背景画像の取得が困難な場合がある．そこで我々は，セグメンテーション結果を用いた異なるカメラ間での人物同定により，提案手法の有効性を示す．

5.1　人物同定法

5.1.1　特徴量抽出

人物同定の特徴量には，セグメンテーションした領域を HSV 色空間に変換し，色相 H と彩度 S から作成

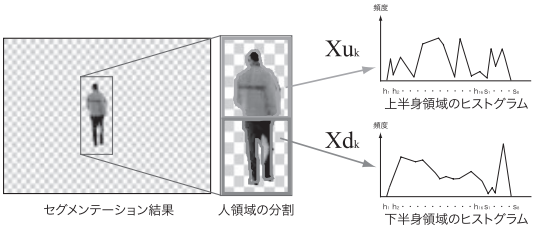


図 13　特徴量抽出の流れ
Fig. 13　Overview of feature extraction.

した HS ヒストグラムを用いる．人の服装は上半身と下半身で色が異なることが多いため，人領域全体から HS ヒストグラムを作成すると識別能力が低下する可能性がある．そこで，人領域を上半身領域と下半身領域に分割して，HS ヒストグラムを作成する（図 13）．まず，セグメンテーションにより検出された人領域 X_k を鉛直方向に 2 等分し，上半身領域 Xu_i と下半身領域 Xd_i とする．ここで，i $(i = 1, \cdots, n)$ は，画像から検出された人領域数である．次に，上下半身各領域から HS ヒストグラム $H_i(h, s)$ $(h = 1, \cdots, 16, s = 1, \cdots, 8)$ を算出する．最後に，求めたヒストグラムを $\sum_{s=1}^{8}\sum_{h=1}^{16} H_i(h, s) = 1$ となるように正規化する．以上のように求めたヒストグラムを特徴量として人物同定を行う．

5.1.2　人物同定

異なる画像から検出された人領域の HS ヒストグラムをマッチングすることで人物同定を行う．次式に示すように，二つのヒストグラム Hu, Hd 間の Battacharyya 距離を計算し，各人領域ごとのマッチングスコア ms を算出する．

$$ms(k,l) = \sum_{s=1}^{8}\sum_{h=1}^{16}\sqrt{Hu_k(h,s)Hu_l(h,s)} + \sum_{s=1}^{8}\sum_{h=1}^{16}\sqrt{Hd_k(h,s)Hd_l(h,s)} \quad (15)$$

ここで，k はテンプレート画像からセグメンテーションした各人領域，l はマッチング対象の画像からセグメンテーションした各人領域である．比較対象であるすべての人領域とのマッチングスコアを算出し，マッチングスコア $ms(k, l)$ が最大かつしきい値 th が一定値以上の人領域を同一人物領域として検出する．

5.2　人物同定実験

5.2.1　実験概要

異なるカメラで撮影された 2 種類のシーケンスを用

(a) シーケンス 1 の人物同定結果例

(b) シーケンス 2 の人物同定結果例

図 14　人物同定結果例
Fig. 14　Example of human identification.

表 5　人物同定結果
Table 5　Experimental result of human identification.

	シルエット選択なし	シルエット選択あり
シーケンス 1	35.0%	63.3%
シーケンス 2	34.0%	95.0%

いて人物同定を行う．実験では，類似シルエット画像選択ありの場合となしの場合の人領域セグメンテーション結果を用いて人物同定を行い，提案手法の有効性を示す．評価に用いるシーケンスは，PETS2006 [17] のデータセットから，正面方向への人の動きがあるシーン 150 フレーム（シーケンス 1）と横方向への人の動きがあるシーン 150 フレーム（シーケンス 2）を使用する．二つのシーケンスは，同一時刻に異なる位置から同一場面を撮影したものである．また，探索対象のテンプレートは，異なるカメラ映像からセグメンテーションした人領域を用いる．

5. 2. 2　実 験 結 果

表 5 に人物同定の結果を示す．また，図 14 に人物同定結果の例を示す．図中の白色方形は異なる画像間で同一人物として判定した結果であり，青色方形は他人と判定した結果である．表 5 を見ると，シルエット画像選択をしていないセグメンテーションを用いた人物同定精度に対して，提案手法であるシルエット画像選択をした場合の人物同定精度はシーケンス 1 で 28.3%，シーケンス 2 で 61% と大幅に向上していることが分かる．図 14 より，人検出段階の失敗により背景領域を人領域として誤検出した領域も，人物同定のマッチング候補に含まれているが，異なる画像間で正しく同一人物を同定していることが分かる．類似シルエット画像の選択を行い，人物領域のみのカラー情報を用いるため，人物同定の精度を向上させることができた．

6. む す び

本論文では，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，あらかじ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異なるカメラ間での人物同定実験を行い，提案手法の有効性を示した．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．

文　　献

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” IEEE Comput. Vis. Pattern Recognit., vol.1, pp.886–893, 2005.

[2] B. Wu and R. Nevatia, “Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors,” Int. J. Comput. Vis., vol.75, no.2, pp.247–226, 2007.

[3] B. Wu and R. Nevatia, “Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses,” Int. J. Comput. Vis., vol.82, no.2, pp.185–204, 2009.

[4] 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.

[5] 上村和広，池亀幸久，下山　功，玉木　徹，山本正信，“ネットワーク上の複数カメラを用いた実時間人物照合システ

ム,” 信学技報，PRMU2003-242, 2004.
[6] 高橋弘樹，中村克行，趙　卉菁，柴崎亮介，“レーザスキャナと画像センサの融合による人物同定,” 第 13 回画像センシングシンポジウム，June 2007.
[7] T. Zhao, R. Nevatia, and F. Lv, “Segmentation and tracking of multiple humans in complex situations,” IEEE Conf. Comput. Vis. Pattern Recognit., pp.194–201, 2001.
[8] T. Zhao and R. Nevatia, “Bayesian human segmentation in crowded situations,” IEEE Conf. Comput. Vis. Pattern Recognit., pp.459–466, 2003.
[9] D.M. Gavrila, “Multi-feature hierarchical template matching using distance transforms,” IEEE Int. Conf. Pattern Recognit., pp.439–444, 1998.
[10] D.M. Gavrila and V. Philomin, “Real-time object detection for “smart” vehicles,” IEEE Int. Conf. Comput. Vis., pp.87–93, 1999.
[11] B. Leibe, E. Seemann, and B. Schiele, “Pedestrian detection in crowded scenes,” Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.878–885, 2005.
[12] R.E. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” Mach. Learn., no.37, pp.297–336, 1999.
[13] D. Comaniciu and P. Meer, “Mean shift analysis and applications,” IEEE Int. Conf. Comput. Vis., pp.1197–1203, 1999.
[14] 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,” 情処学研報 (CVIM2006)，no.93, pp.57–64, 2006.
[15] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
[16] http://www.cvg.rdg.ac.uk/PETS2001/
[17] http://www.pets2006.net/

（平成 22 年 9 月 7 日受付，11 月 29 日再受付）

村井　陽介

2007 中部大・工・情報工学卒．2009 同大大学院工学研究科修士課程了．同年よりセコム（株）IS 研究所入所．現在，同社先端研究ディビジョンバイオメトリクスグループ研究員．画像情報処理・コンピュータビジョンに関する研究に従事．情報処理学会会員．

安藤　寛哲

2009 中部大・工・情報工学卒．同年より同大大学院工学研究科情報工学専攻に在籍．画像処理，パターン認識に関する研究に従事．

藤吉　弘亘　（正員）

1997 中部大学大学院博士後期課程了．1997〜2000 米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow．2000 中部大学講師を経て 2004 より同大准教授．2006 米カーネギーメロン大学ロボット工学研究所客員研究員．工博．計算機視覚，動画像処理，パターン認識・理解の研究に従事．2005 年度ロボカップ研究賞．2009 年度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優秀論文賞，2009 年度山下記念研究賞．情報処理学会，電気学会，IEEE 各会員．

伊藤　誠也

2002 北大・工・システム工学卒．2004 同大大学院工学研究科システム情報工学専攻修士課程了．同年（株）日立製作所入社．現在，同社日立研究所研究員．画像認識・理解，パターン認識の研究に従事．

数井　誠人　（正員）

1994 北大・工卒．1996 同大大学院工学研究科修士課程了．2001 同大学院博士課程修了．同年より（株）日立製作所入所．2005 米カーネギーメロン大学ロボット工学研究所客員研究員．現在，サムスン横浜研究所専任研究員．工博．画像処理・認識技術の研究に従事．情報処理学会会員．