論　文

準同型暗号を用いた人物対応付けのための画像選択による通信量の削減

福田　尚悟[†]　　森田　一成[†]　　西山　正志[†,††a)]　　岩井　儀雄[†,††]

Reduction of Communication Traffic by Selecting Pedestrian Images for Person Re-identification using Homomorphic Encryption

Shogo FUKUDA[†], Kazunari MORITA[†], Masashi NISHIYAMA[†,††a)], and Yoshio IWAI[†,††]

あらまし　本論文では，準同型暗号を用いた人物対応付けシステムにおける通信量を削減するために，足姿勢を用いて適切な人物画像を選択し，特徴量を抽出する手法を提案する．人物を対応付ける際に用いる特徴量は個人情報に該当するため，特徴量を暗号化し保護する必要がある．準同型暗号は復号化することなく特徴量間の距離を計算できるため，安全なシステム運用が可能となる．ただし，暗号化された特徴量のサイズは大きくなるため，特徴量をシステム内で送受信する際，通信量が増大する課題がある．そこで提案手法では，特徴量を抽出する前段階である人物画像の集合から，カメラ間で共通に表れる足姿勢を用いて画像枚数そのものを減らすことを狙う．準同型暗号を組み込んだシステムにおいて，人物対応付けの精度を同等以上に保った上で，特徴量を送受信する際の通信量を 73.2% ほど削減できることを実験で確認した．

キーワード　画像選択, 準同型暗号, 特徴量, 人物対応付け

1. はじめに

安心安全な社会の実現のために，複数カメラを用いた見守りシステムが必要とされている．本論文では，人物が歩行する動線を仮定できる通路での見守りシステムについて考える．通路は屋内の地点間をつなぐ役割として存在する．見通しが良く，人物が移動する際に必ず通る場所であるため，このシステムは犯人や迷子の捜索時に有効である．本論文では，通路に設置されている各カメラの高さと角度は同じであると仮定し議論を進める．

歩行する人物をカメラで撮影する場合，その人物の姿勢は時々刻々と変化する．同一人物の人物画像をカメラ間で高精度に対応付けるため，人物画像からその人物の個人性を表す特徴量を抽出し，特徴量同士の距離を計算することが一般的である．特徴量として例えば，色や勾配の分布を画像統計として抽出する手法 [1]〜[4] や，性別や年代などの人物属性を特徴として抽出する手法 [5]〜[8] が提案されている．

人物対応付けで使用される特徴量は，個人を特定できるものであるため個人情報に該当する(*1)(*2)．よって対応付けシステムでは，個人情報である特徴量を暗号化し，第三者からはアクセスできない形で取り扱う必要がある．以下では特徴量を保護するために，準同型暗号 [9] に着目する．準同型暗号は，対象を復号化することなく加算と乗算の演算が可能である．準同型暗号を用いたパターン認識の手法として，主成分分析への適用例 [10] や，深層学習への適用例 [11] が報告されている．本論文では，人物対応付けの特徴量に準同型暗号を適用し，暗号化されたままで距離計算を行うことを考える．

特徴量の距離計算において，スケーラビリティもまた考慮する必要がある．混雑で人数が多くなった場合，また，見守る範囲が広がりカメラの台数が多くなった

† 鳥取大学大学院持続性社会創生科学研究科，鳥取市
Graduate School of Sustainability Science, Tottori University, Tottori 680-8550, Japan
†† 鳥取大学大学院工学研究科，鳥取市
Graduate School of Engineering, Tottori University
a) E-mail: nishiyama@tottori-u.ac.jp

(*1)：個人情報保護に関する法律 (2017/5/30)
(*2)：General Data Protection Regulation (2018/5/25)

場合，多数の特徴量が獲得され距離計算のコストが大幅に増加する．システムにスケーラビリティを持たせるため，一般的にはクラウド上のサーバを利用することが考えられる．ただし，公開されている通信路やハードウェアを用いるため，悪意のある攻撃をサーバが受けた場合，特徴量が漏えいする危険性がある．準同型暗号で保護された特徴量をサーバへ送ることで，特徴量をサーバ上で復号化する必要がなくなり，安全な人物対応付けのシステム運用が可能になると考えられる．

準同型暗号を適用する際，各カメラクライアントとサーバとの間の通信量が課題となる．準同型暗号が施された特徴量はデータサイズが大きくなるため，特徴量 1 つあたりの通信量が増加する．さらに，特徴量に変換する前の段階である人物画像の枚数にもデータサイズは比例するため，通信量がさらに増加する．これらの課題を解決するために通信量を減らす工夫が必要となる．考えられる工夫として，特徴量の次元数を削減すること，送受信する特徴量を選択することが挙げられる．本論文では後者の工夫に着目し，特徴抽出の前段階である人物画像の枚数そのものを少なくし通信量を減らすことを考える．人物画像は，カメラ視野内に人物が存在している間に多数獲得されていく．具体的には，カメラクライアントの視野内に存在する人数，撮影される時間，カメラのフレームレートの乗算で人物画像の枚数は増加していく．さらにサーバ側ではカメラ台数も加わるため，人物画像の枚数は膨大となり，獲得された人物画像を単純に全て用いた場合，通信量が爆発的に増加すると考えられる．

そこで本論文では，人物画像の枚数を削減し通信量を抑えるために，各カメラで獲得された大量の人物画像の中から，対応付けの特徴量として適切な人物画像を選択することを考える．提案手法では，人物画像の選択の手掛かりとして，歩行時に繰り返して表れる足の姿勢に着目する．足姿勢を手掛かりとして，カメラ間で共通に表れる人物の見え方から特徴量を抽出することを狙う．本論文では足姿勢を用いた画像選択を，準同型暗号により保護された人物対応付けシステムへ組み込む．独自収集したデータセットおよび公開データセットにおいて，提案手法を用いて人物画像を選択することで，人物対応付けの精度を維持したままで，画像枚数の大幅な削減が可能であることを確認した．以下，2. では関連研究について述べる．次に 3. で人物画像の選択手法について述べ，4. で提案手法を組み込んだ対応付けシステムについて述べる．そして 5. で実験結果について述べ，最後に 6. でまとめる．

2. 関 連 研 究

人物対応付けシステムにおいて，カメラクライアントとサーバとの間で通信量を抑える工夫の 1 つとして，映像中の人物に追跡処理 [12]〜[14] を適用し，追跡で繋がった時系列の複数画像から 1 つの特徴量を抽出することが考えられる．文献 [15] では，時系列の複数画像において，見え方変化の少ない身体部位に注目し，特徴量を抽出する手法を提案している．この既存手法では，人物の身体部位ごとに特徴量を GMM (Gaussian Mixuture Model) で表している．GMM は平均ベクトルと共分散行列を保持する必要があるため，1 名あたりの特徴量サイズが大きくなる課題がある．このため，本論文の目的であるデータサイズ削減に向いていないと考えられる．文献 [16] では，歩容に着目し，時系列の複数画像を 1 枚の画像に統合する手法を提案している．この既存手法では，周期的に変化する手足の動きの特徴量を，統合された画像で表している．歩行の 1 サイクルを欠けなく取得できることを前提としているため，追跡処理が途切れた場合に特徴量が抽出できない課題がある．本論文では，追跡処理を用いずとも運用できることを目指し，提案手法の設計を行う．

通信量を削減する別の工夫として，特徴量の次元削減が挙げられる．例えば，ハッシュ関数を用いて特徴量を符号化する手法 [17]〜[19] の適用や，ベクトル量子化やスパースコーディングを用いて特徴量を軽量化する手法 [20]〜[22] の適用が考えられる．これらの手法は特徴量の次元数を小さくすることができるため，通信量削減への寄与が期待される．特徴量 1 つあたりのデータサイズを小さくするために，符号化や軽量化を特徴量へ適用する手法 [17]〜[22] が有効と考えられる．ただし，符号化や軽量化を適用したとしても，人物画像の枚数が増加した場合の課題は残る．本論文では人物画像の枚数増加の課題に焦点をあて手法の開発を行っていく．

人物対応付けではないが顔認識を目的として，文献 [23]〜[25] では，映像から検出された顔画像の集合から，正面向きに近い顔画像を，姿勢推定を用いて選択する手法が提案されている．これらの既存手法の狙いは提案手法と同じであると言えるが，顔を対象としているため，人物の全身を用いる対応付けシステムにそのまま適用することは難しい．顔認識における画像選択の手法 [23]〜[25] は，人物がカメラに対して後ろ

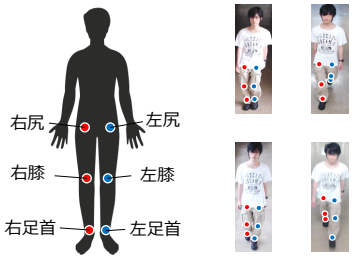


図 1　足姿勢の空間手掛りを表す間接部位.
Fig. 1　Body joints for representing spatial cues of leg pose.

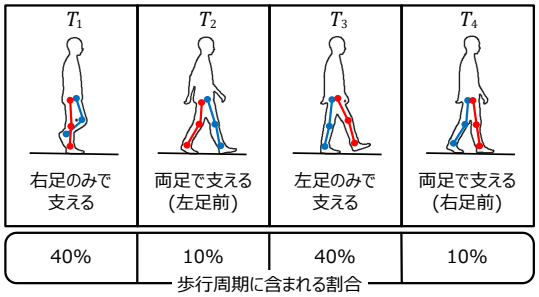


図 2　足姿勢の時間手掛りを歩行周期に基づく区分で表現.
Fig. 2　Representation of temporal cues of leg pose by using segments based-on walking cycle.

を向いている場合に，顔が取得できない課題がある．よって本論文では，カメラで獲得された人物画像の集合から，対応付けに適した人物画像を選択する手法の開発を行う．

3. 対応付けに用いる人物画像の選択

3.1 提案手法の考え方

各カメラで獲得された多数の人物画像の中から，対応付けに適切な人物画像を少数枚選択するために，本論文では足姿勢に着目する．通路を歩行する人物の足姿勢は，周期的な動きを繰り返している．このため各カメラにおいて，特定の同じ足姿勢を含む人物画像を獲得できる機会は多いと仮定できる．同じ足姿勢の人物画像を少数枚獲得できれば，精度低下を招くことなく人物を対応付けられる可能性は高いと期待できる．なお，歩行中の腕の動きは，荷物の所持や腕組みの影響で自由度が高いため，同じ腕姿勢を繰り返し獲得できる機会は足姿勢と比べて少ないと考えられる．頭姿勢や胴姿勢は足姿勢と比べて歩行中の変化は少ないが，そもそも同じ姿勢ばかりが得られる可能性が高く，画像枚数の削減に効果が少ないと考えられる．対応付けの精度を維持したまま人物画像の枚数を削減するため，本論文では足姿勢を採用する．

歩行する人物の足姿勢は，空間手掛りと時間手掛りの二つで表現される．空間手掛りは，足の関節部位の位置を指す．具体的には図 1 に示す間接部位の空間配置が手掛りとなる．時間手掛りは，時々刻々と変化する関節部位の周期的な動きを指す．画像選択の基準となる足姿勢を考えるため，次節では足姿勢の歩行周期について詳しく述べる．

3.2 歩 行 周 期

文献 [26] では，歩行時に足姿勢が変化していく事象を歩行周期と呼んでおり，人物が歩き始めてから終わるまで同じ動きを繰り返すと述べられている．歩行周期は，支える足の状態の違いにより四つの区分 $T_c(c = 1, \ldots, 4)$ に分けられている．それらの区分を図 2 に示す．文献 [26] では歩行周期の開始時刻が明確には定められていないが，本論文では工学的に取り扱う必要があるため，左足が地面から離れた瞬間を開始時刻とし，区分を時間方向に T_1 から T_4 とする．歩行周期の 1 サイクルの時間長は約 1 秒であり，T_1 と T_3 が時間長の 40% ずつ，T_2 と T_4 の時間長が 10% ずつ含まれている．

3.3 基準姿勢の設定手法

提案手法では，足姿勢の時間手掛りを歩行周期の各区分 T_c で表す．また各区分において空間手掛りを，関節部位の平均位置で表す．これにより，人物画像を選択する際の手掛かりとなる基準姿勢 $P_c(c = 1, \ldots, 4)$ を設定する．区分内で関節部位の位置を平均化する際，各人物における区分内の位置の時間変化，および，人物間の位置の空間変化を考慮する．本論文では，様々な人物から観測された歩行周期の時系列データを用いて，基準姿勢を設定していく．

基準姿勢の具体的な設定手法を以下で述べる．H 人の実験協力者について，歩行周期 1 サイクルに対して，関節部位の 3 次元位置がモーションキャプチャで観察されているとする．各カメラクライアントにおいて，カメラの高さと角度は一定であると本論文は仮定しているため，予め設定した透視投影行列を用いて画像平面上に関節部位を射影する．なお実験協力者の体型の違いを抑えるために，首から腰を結ぶ胴の長さが画像平面上で一定となるように，スケールと傾きを相似変換で正規化する．胴が正規化された関節部位の表現例

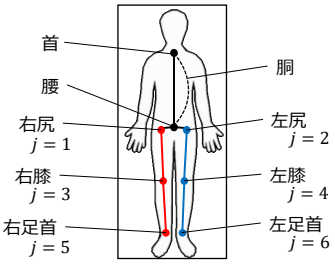


図 3　胴が正規化された関節部位の表現.
Fig. 3　Representation of body joints after normalizing torso.

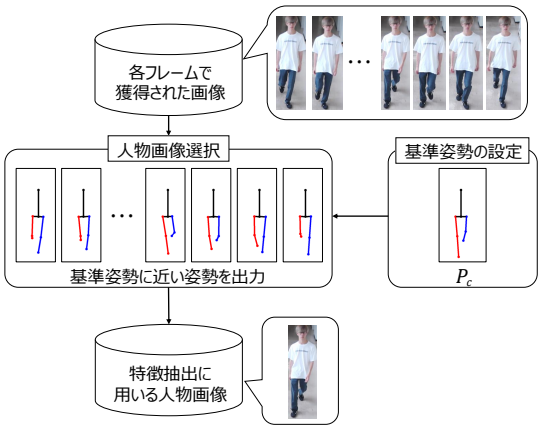


図 4　人物画像の選択手法の流れ.
Fig. 4　Overview of selecting pedestrian images.

を図 3 に示す．次に，3.2 で述べた歩行周期の定義に従い，各時刻における各協力者の関節部位 $j = 1, \ldots, 6$ の 2 次元位置を各区分 T_c に分類する．分類された関節部位の 2 次元位置の集合から平均値を求めることで，区分 T_c に対する基準姿勢 P_c を設定する．なお，四つの基準姿勢 P_c の中で，どれを用いるべきかについては 5. の実験にて検証する．

3.4　人物画像の選択手法

各カメラクライアントにおいて検出された人物画像の集合の中から，3.3 で設定した基準姿勢 P_c を用いて，対応付けに適した画像の選択を行う．この処理の結果として，基準姿勢に近い足姿勢をもつ人物画像が出力され，人物対応付けの特徴抽出に処理が進む．人物画像の選択処理の流れを図 4 に示す．

画像選択の具体的な手法について以下で述べる．人物画像が与えられると，関節部位の位置を 2 次元の姿勢推定を用いて算出する．人物間の体型の差異を吸収するため，胴の長さが基準姿勢と同じになるよう 3.3 で述べた相似変換で，関節部位の位置を正規化する．基準姿勢の関節部位の位置と，正規化された姿勢の関節部位の位置とが，どの程度離れているかを表すコストを L2 ノルムで計算する．このコストがしきい値 R 以下の場合，対応付けに適切な人物画像であると出力する．なお R は，各カメラクライアントにおいて，各人物がカメラの視野に入ってから出るまでの間，各人物の画像が 1 枚以上選択されるよう経験的に定める．

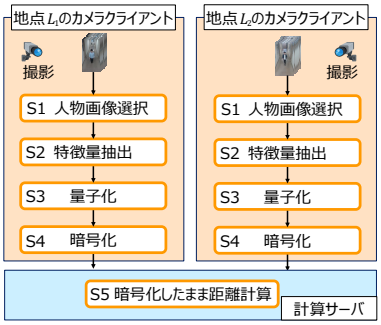


図 5　特徴量が保護された人物対応付けシステムの処理の流れ.
Fig. 5　Overview of your system for privacy-protected person Re-identification.

4.　準同型暗号により特徴量が保護された人物対応付けシステム

4.1　システムの概要

特徴量が保護された人物対応付けシステムの処理の流れを図 5 に示し以下で説明する．処理 S1 では，地点 L_1 や地点 L_2 のカメラクライアントにおいて，対応付けに用いる人物画像を，3.4 で述べた手法を用いて選択する．処理 S2 では，選択された人物画像からそれぞれ特徴量の $\boldsymbol{f}_q$ や $\boldsymbol{f}_t$ を抽出する．処理 S3 では，準同型暗号を適用するための前処理として，特徴量に線形量子化を適用する．処理 S4 では，量子化された特徴量に対し準同型暗号を適用し，各カメラクライアントからサーバへ送信する．処理 S5 では，受け取った特徴量に対して，サーバ内で復号化することなく距離を計算する．本論文では，人物対応付けの性能を高めるために，計量学習の 1 つである Large Margin Nearest Neighbor (LMNN) [27] を適用する．計量行列を $\mathbf{M} = \mathbf{L}^{\mathrm{T}}\mathbf{L}(||\mathbf{M}||_f \leq 1)$ で表すと，特徴量間の距離 d^2 は式 (1) で計算される．

$$d^2 = (\boldsymbol{f}_q - \boldsymbol{f}_t)^{\mathrm{T}}\mathbf{M}(\boldsymbol{f}_q - \boldsymbol{f}_t) \quad (1)$$

なお，準同型暗号を用いて行列演算を行うと計算量が非常に大きくなる．本論文では行列演算を暗号化前に行うこととし，距離を式 (2) で求める．

$$d^2 = (\boldsymbol{f}'_q - \boldsymbol{f}'_t)^{\mathrm{T}}(\boldsymbol{f}'_q - \boldsymbol{f}'_t) \quad (2)$$

ここで $\boldsymbol{f}'_q = \mathbf{L}\boldsymbol{f}_q$，$\boldsymbol{f}'_t = \mathbf{L}\boldsymbol{f}_t$ とする．算出された距離が小さいほど，それらの特徴量が同じ人物から抽出された可能性が高くなる．以下では，準同型暗号の詳細を 4.2 で，量子化の詳細を 4.3 で述べる．

4.2 準同型暗号

準同型暗号は有限体上での加算と乗算の演算処理を復号化することなく行うことが出来る．本論文の対応付けシステムでは，準同型暗号を利用することで，特徴量 $\boldsymbol{f}'_q, \boldsymbol{f}'_t$ を暗号化したままで距離 d^2 を計算する．距離計算だけであれば 1 回乗算が可能な Somewhat Homomorphic Encryption (SwHE) を利用することも考えられるが，Gentry らが公開している完全準同型暗号（Fully Homomorphic Encryption, FHE）ライブラリ HElib(*3)を本論文では利用する．N 次元の特徴量を暗号化するために，平文空間を表す整数環 $\mathbb{Z}_p$ を決定するパラメータ p，および，スロット数 $n(n \geq N)$ を決定するパラメータ m を制御する．ここで p は素数とする．パラメータ n は式 (3) で決定される．

$$n = \frac{\phi(m)}{\mathrm{order}(m, p)} \quad (3)$$

ここで，ϕ はオイラー関数，order は m を法としたときの p の位数である．n は m と p に支配されるため，自由に設定することは難しい．p と m に関して，特徴量の次元数 N 以上となるスロット数 n が得られるまで Brute-force 探索を行う．

4.3 量 子 化

準同型暗号ライブラリ HElib は整数環を使用するため，計算効率の観点から特徴量を整数で表すことが望ましい．本論文の人物対応付けシステムでは，画像から抽出された特徴量に対し量子化を適用することで，実数表現から整数表現へ変換する．式 (2) の距離計算では，計量行列を分解した $\mathbf{L}$ を用いるため，特徴量間の相対スケールを保持しつつ量子化を適用する必要がある．そのため本論文では線形量子化を単純に適用する．特徴量 $\boldsymbol{f}'_q$ と $\boldsymbol{f}'_t$ の各要素 f'_q, f'_t が取り得る範囲は $-1 \leq f'_q \leq 1$，$-1 \leq f'_t \leq 1$ として，ビット数 g の線形量子化を適用する．

暗号化された特徴量間の距離 d^2 を計算する際，オーバーフローに注意しなければならない．オーバーフローすることなく計算するためには，素数 p が式 (4) を満たす必要がある．

$$2^{2g}N < p \quad (4)$$

ここで g は量子化ビット数である．p と m を探索するとき，この制約条件も同時に用いる．参考として，特徴量の次元数 N を 95 に設定した時，素数 p とパラメータ m と量子化ビット数 g の組み合わせ例を表 1 に示す．

表 1 素数 p とパラメータ m と量子化ビット数 g の組み合わせ例 (N=95 としたとき)．
Table 1 Examples of prime number p, parameter m, and quantization bit g when $N = 95$.

素数 p	m	ビット数 g	素数 p	m	ビット数 g
509	10083	1	131071	14173	5
2053	12017	2	524341	14023	6
8191	12007	3	2097143	13943	7
32771	14023	4	8388617	7957	8

5. 実　　験

5.1 実 験 条 件

準同型暗号を組み込んだ人物対応付けシステムにおいて，基準姿勢を用いて人物画像を選択した場合に通信量が削減されるかどうかについて評価した．評価指標として，対応付け精度および削減率を用いた．対応付け精度では，あるカメラクライアントの人物画像をターゲットとし，別のカメラクライアントの人物画像をクエリとして，正しい人物同士がヒットするかどうかを第一位正解率で評価した．このとき P_1 から P_4 それぞれを用いて選択し，各地点間で対応付けを行った際の第一位正解率を平均して使用した．削減率では，画像選択を行わない場合の人物画像の枚数から，どれだけの画像枚数を減らすことができたかどうかの比率で評価した．このとき P_1 から P_4 それぞれを用いて選択した際の削減率を平均して使用した．

本実験では，我々が独自に収集した TUP (Tottori University Pedestrian) データセット，および，公開されている MARS (Motion Analysis and Re-identification Set) [28] データセットを用いた．各デー

(*3)：https://github.com/shaih/HElib

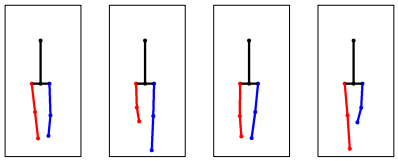


図 6　TUP データセットの透視投影行列を用いて生成した基準姿勢.
Fig. 6　Base poses generated using the projection matrix of the TUP dataset.

タセットの詳細は 5.2 と 5.3 でそれぞれ述べるが，以下ではデータセット間で共通する実験条件について先に述べる．

5.1.1　設　　定

まず基準姿勢を設定するためのパラメータについて述べる．本論文では AIST Gait Database 2015 [29] を用いた．このデータセットには，日本人健常者 214 名を対象とし，歩行中にモーションキャプチャで観測された間接部位の 3 次元位置が含まれている．ここではランダムに選択された $H = 100$ 名のデータを用いた．各人物の歩行周期 1 サイクルにおいて，間接部位の位置を 3.3 の手法で平均化した．さらに人物間で平均化することで，各区分 T_c の基準姿勢 P_c を設定した．なお体の表面に取り付けられたモーションキャプチャのセンサでは，図 3 の間接部位 $j = 3, ..., 6$ を直接観測することができないため，本論文では周辺の複数部位からの平均位置で代用した．TUP データセットと MARS データセットではカメラ配置が異なるため，透視投影行列はそれぞれ別のものを用いた．図 6 に，TUP データセットに対する透視投影行列を用いて設定した基準姿勢 $P_c (c = 1, \ldots, 4)$ を示す．

次に画像選択のパラメータについて述べる．本実験では，図 3 に示す間接部位の 2 次元位置を人物画像から推定をするために OpenPose [30] を使用した．また本実験では，データセット毎に画像選択のしきい値 R を設定した．3.4 でも述べたが，人物対応付け時に見落としを避けるため，各人物につき少なくとも 1 枚以上の画像が選択される R を探索した．その結果，TUP データセットで $R = 100$，MARS データセットで $R = 375$ が得られた．

5.1.2　人物対応付けのパラメータ

人物対応付けに用いる特徴量 $\boldsymbol{f}_q, \boldsymbol{f}_t$ として，文献 [31] の共起属性を用いた．共起属性は，人物の外見的な手掛かりと身体的な手掛かりの組み合わせで表現される．特徴量の次元数は $N = 95$ とした．特徴量 $\boldsymbol{f}_q, \boldsymbol{f}_t$ の各次元の要素 f_q, f_t の取り得る範囲は $-1 \leq f_q \leq 1$, $-1 \leq f_t \leq 1$ とした．線形量子化のパラメータとして，ビット数 $g = 8$ を用いた．準同型暗号のパラメータとして，表 1 に記載した素数 $p = 8388617$ と $m = 7957$ を用いた．計量行列 $\mathbf{M}$ を学習するために，TUP データセットの実験では SARC3D データセット [32] に含まれる人物画像 200 枚を用いた．また MARS データセットの実験では，そのデータセットに含まれている訓練用の画像からランダムに選んだ 1000 枚を用いた．

5.2　TUP データセットでの評価

5.2.1　撮 影 条 件

屋内の通路を歩行する人物を，複数地点に設置されたカメラから撮影した．図 7 に，TUP データセットのカメラクライアント (地点 L_1 から L_6) で撮影された映像の例を示す．床からの高さ 2.4 メートルにカメラを設置し，カメラ画像の中心が手前 3 メートルに合うように角度を設定した．撮影には logicool C920（1920 × 1080 画素）を使用した．歩行中の人物を可能な限り観測するために，カメラを水平の状態から反時計回りに 90 度回転させ，解像度が 1080 × 1920 画素の映像を取得した．フレームレートを 30 fps とした．撮影人数は 31 名とした．同時に歩行する人数は 1 名とした．歩行の開始位置は L_1，L_2，L_5，L_6 ではカメラの前方 7 メートルとした．L_3 と L_4 では動線上にドアが存在するためカメラの前方 3.5 メートルとした．全ての地点で，通路の中央を通るように指示した．図 8 に，TUP データセットに含まれる人物画像の例を示す．各地点における人物画像の枚数は，L_1 で 5117 枚，L_2 で 5060 枚，L_3 で 2895 枚，L_4 で 2418 枚，L_5 で 4841 枚，L_6 で 5082 枚であった．

5.2.2　評 価 結 果

TUP データセットにおける人物対応付けの精度を図 9 に示す．画像選択を行わず全ての人物画像を用いたベースライン手法の精度と比較して，基準姿勢 $P_c (c = 1, \ldots, 4)$ のいずれかを用いた提案手法の精度は同等以上であった．表 2 に，各基準姿勢を用いて人物の画像選択を行った場合の削減率を示す．実験結果より，P_1 と P_3 の削減率に比べて，P_2 と P_4 の削減率が高くなることが分かった．TUP データセットの撮影環境では，削減率と対応付け精度がともに高かった基準姿勢 P_2 と P_4 が適していると考えられる．

選択された人物画像の例を図 10 に示す．ここでは，地点 L_4 のカメラクライアントで基準姿勢 P_2 を用い

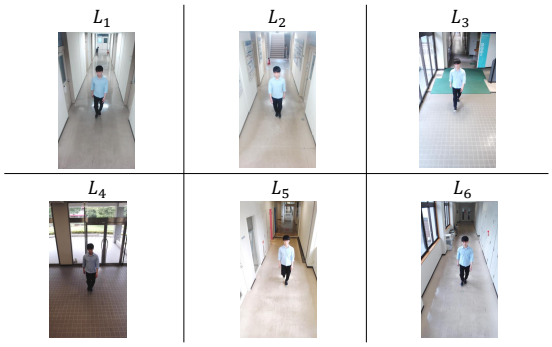


図 7　TUP データセットのカメラクライアントで撮影された映像の例.
Fig. 7　Examples of images acquired from camera clients in the TUP dataset.

図 8　TUP データセットに含まれる人物画像の例.
Fig. 8　Examples of pedestrian images contained in the TUP dataset.

表 2　TUP データセットにおける人物画像の削減率 (%).
Table 2　Reduction rate of pedestrian images in the TUP dataset.

P_1	P_2	P_3	P_4
58.0	**82.4**	53.4	**76.1**

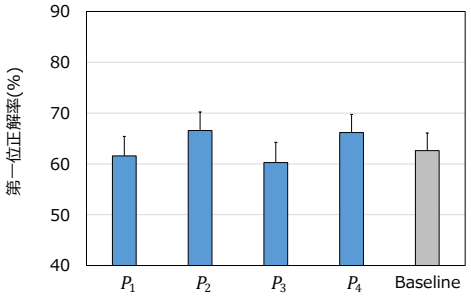


図 9　TUP データセットにおける人物対応付けの精度.
Fig. 9　Accuracy of person re-identification in the TUP dataset.

た場合とした．図中に，ある人物の歩行周期 1 サイクルから間引いた人物画像とその足姿勢のペアを列挙した．この例では 16 枚の人物画像のうち，3 枚の人物画像が提案手法で選択されていた．図 6 の基準姿勢 P_2 に近い足姿勢に対応する人物画像が残っていることが分かる．

5. 2. 3　カメラクライアントとサーバとの間の通信量

ここでは，L_1 と L_2 のカメラクライアントから，計算サーバへ特徴量を送信する場合について通信量を考察する．準同型暗号のパラメータは 5. 1. 2 で示したものとした．特徴量 1 つ当りの通信量は，暗号化の前後で 760 バイトから 524.4 キロバイトに増加した．各カメラクライアントから検出された人物画像の枚数は L_1 で 5117 枚，L_2 で 5060 枚であった．全ての人物画像から抽出された暗号化特徴量を用いるベースライン手法の通信量は L_1 で 2.68 ギガバイト，L_2 で 2.65 ギガバイトとなった．一方，基準姿勢 P_2 を用いて人物画像を選択した場合，人物画像の枚数が L_1 で 903 枚，L_2 で 807 枚に削減された．提案手法の通信量は L_1 で 0.47 ギガバイト，L_2 で 0.42 ギガバイトとなった．通信量と画像枚数について，L_1 では 82.4%，L_2 では 84.1% の削減となった．以上の結果より，準同型暗号を用いた人物対応付けシステムにおいて，提案手法により大幅に通信量が削減できたことを確認した．

5. 2. 4　歩行周期の区分を考慮しない場合の評価

人物画像を選択する際に，歩行周期の区分を考慮すべきかどうかを確認する実験を行った．提案手法では，歩行周期 1 サイクルを四つの区分に割りそれぞれ基準姿勢を設定しているが，ここでの比較手法では，区分を設けずに歩行周期 1 サイクルから基準姿勢を 1 つ設定した．基準姿勢を除いた全ての実験条件は提案手法と比較手法で 5. 2. 1 と同じとした．実験の結果，基準姿勢 P_2 を用いた提案手法の第一位正解率が 66.6% で画像枚数の削減率が 82.4% であったのに対し，比較手法の第一位正解率が 59.7% で画像枚数の削減率が 49.1% であった．よって人物画像を選択する際，歩行周期の区分を利用することは有効であると考えられる．

5. 2. 5　ランダムに選択した場合の評価

人物画像をランダムに選択した場合について評価した．ここでは 5. 2. 2 の実験で性能が高かった基準姿勢 P_2 と P_4 を用いた．提案手法とランダム手法の間で公平性を保つため，人物画像の選択枚数を同じとした．

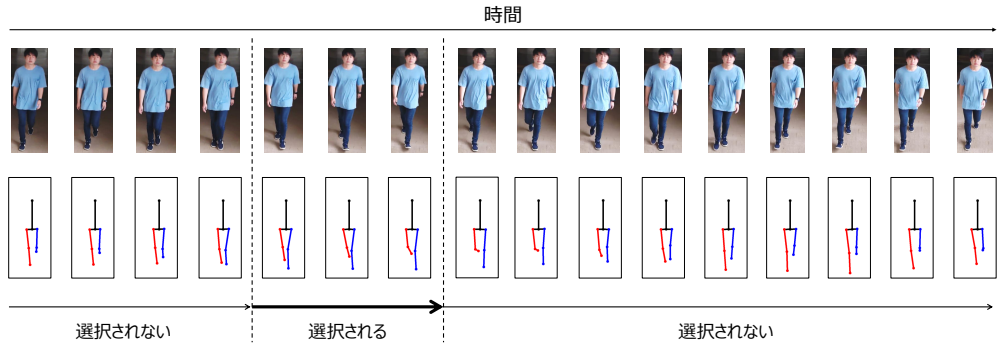


図 10　基準姿勢 P_2 を使用した場合に地点 L_4 のカメラクライアントで選択された人物画像の例.
Fig. 10　Examples of pedestrian images selected using P_2 in L_4.

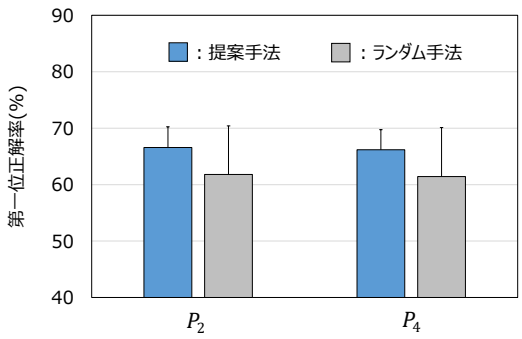


図 11　ランダムに選択した場合の人物対応付け精度の比較.
Fig. 11　Accuracy of person re-identification in the TUP dataset by random selection.

P_2 で 4385 枚，P_4 で 6093 枚を選択した．ランダム選択を 10 回行い，ランダム手法の対応付け精度の平均を求めた．上記以外の条件は，5.1 で述べた条件と同じとした．

提案手法とランダム手法の人物対応付け精度を図 11 に示す．提案手法は，ランダム手法と比べて，P_2 と P_4 ともに対応付け精度が高かった．この結果より，基準姿勢を用いて人物画像を選択する提案手法は，ランダムに選択する手法と比べて有効であると言える．

5.2.6　人物画像の検出漏れの評価

本論文の実験では，OpenPose [30] から出力された身体部位に基づき人物画像を検出している．ここでは，カメラ映像の各フレームにおける人物画像の検出漏れを評価した．TUP データセットにおいて，人物の身体が一部でも含まれるフレームの総数は 35094 枚であった．一方，OpenPose により検出された人物画像の総数は 25413 枚であった．各フレームには 1 名のみの人物が存在するため，人物画像の検出漏れフレーム数は 9681 枚であった．検出漏れとなったフレームをランダムに 100 枚選んで目視確認を行ったところ，カメラの近くを歩行する人物が全てに含まれていた．これらのフレームは，足がフレーム内で観測されていないため，人物対応付けに適していないと考えられる．よって，これらのフレームからの人物画像の検出漏れの影響は，対応付け精度に関して少ないと考えられる．なお，TUP データセットに含まれる全ての人物について，歩行周期 1 サイクル以上の人物画像が検出されていることも目視で確認した．

5.3　MARS データセットでの評価

5.3.1　撮影条件

公開されている MARS データセットを用いて実験を行った．図 12 に MARS データセットに含まれる人物画像の例を示す．屋外の広場を歩行する人物が撮影対象であるものの，動線は一定であると概ね仮定できるセットが含まれている．本実験では，それらのセットであるカメラ C1 と C2 で評価を行った．このデータセットでは，人物追跡が適用された後の画像列のみが公開されている．追跡処理が細かく途切れ足姿勢の変化が十分に含まれていない画像列が存在するため，1 つの画像列に 20 枚以上の人物画像が含まれているもののみを実験に用いた．以下では，C1 で撮影された 41 名に対する 4540 枚の人物画像，および，C2 で撮影された 40 名に対する 3948 枚の人物画像で評価を行った．本実験の目的として，カメラ C1 または C2 の視野内において，同じ方向に歩行する人物を，異な

図 12　MARS データセットに含まれる人物画像の例.
Fig. 12　Examples of pedestrian images contained in the MARS dataset.

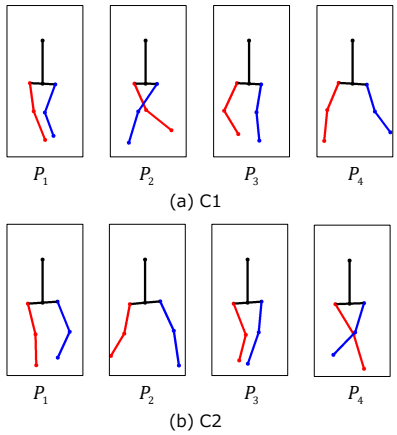


図 13　MARS データセットの透視投影行列を用いて生成した基準姿勢.
Fig. 13　Base poses generated using the projection matrix of the MARS dataset.

る時刻の間で対応付けることを想定した.

5.3.2　評価結果

MARS データセットの透視投影行列を用いて生成した基準姿勢を図 13 に示す．カメラ C1 と C2 の間で高さと角度が異なるため，それぞれの透視投影行列を適用した．カメラ C1 の評価では図中 (a) の基準姿勢を，C2 の評価では (b) の基準姿勢を用いた．MARS データセットのカメラ C1 と C2 における人物対応付けの精度を図 14 に示す．画像選択を行わないベースライン手法の精度と比較して，基準姿勢で画像選択を行った提案手法の精度は同等以上である結果が得られた．表 3 に，人物画像の選択を行った場合の削減率を示す．実験結果より，P_1 と P_3 の削減率に比べて，P_2 と P_4 の削減率が高くなることが分かった．MARS データセットの撮影環境では，削減率と対応付け精度の両方が良かった基準姿勢 P_2 および P_4 が適していると考えられる．以上より，公開されている MARS データセットにおいても，提案手法は対応付け精度を維持したまま画像枚数を削減できることを確認した.

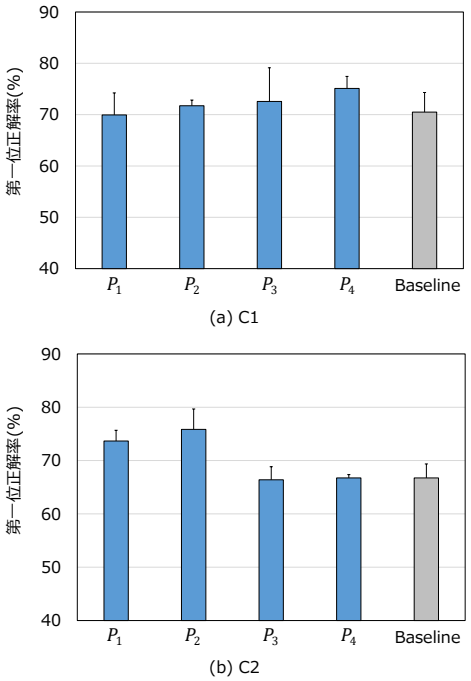


図 14　MARS データセットにおける人物対応付けの精度.
Fig. 14　Accuracy of person re-identification in the MARS dataset.

表 3　MARS データセットにおける人物画像の削減率 (%).
Table 3　Reduction rate of pedestrian images in the MARS dataset.

(a) C1				(b) C2			
P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4
8.1	**55.4**	15.2	**44.3**	9.5	**45.7**	2.5	**33.3**

5.3.3　足以外の身体部位を基準姿勢に用いた場合の評価

提案手法において足姿勢を基準に採用した理由を考察するために，足以外の身体部位を基準とした場合の性能を評価した．提案手法では，胴を中心とし，その胴に連結する足の関節部位に注目した．本実験では，足だけではなく，腕や頭の部位を用いた以下の手法について比較を行った.

- 提案手法 (足姿勢)： 図 3 の右尻，左尻，右膝，左膝，右足首，左足首
- 比較手法 1(腕姿勢)： 図 15(a) の右肩，左肩，右肘，左肘，右手首，左手首
- 比較手法 2(頭姿勢)： 図 15(b) の鼻

全ての手法において，胴の長さと位置を一定にするた

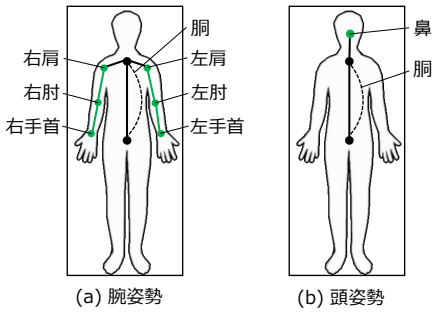


図 15　腕姿勢と頭姿勢を表すための身体部位.
Fig. 15　Body parts for representing arm pose and head pose.

めに，3.3 で述べた相似変換を用いて各部位の位置を正規化した．時間方向の区分の分け方は，提案手法と比較手法の間で同じとした．提案手法では，5.3.2 で性能が高かった基準姿勢 P_2 と P_4 を用いた．比較手法 1 では，提案手法と同じ時間区分に対応する腕の基準姿勢 P_2^a と P_4^a を用いた．比較手法 2 では，同様に P_2^h と P_4^h を用いた．各手法において，各人物につき少なくとも 1 枚以上の画像が選択されるしきい値 R を探索した．その結果，足姿勢で $R = 375$，腕姿勢で $R = 650$，頭姿勢で $R = 100$ が得られた．MARS データセットのカメラ C1 と C2 で評価し，平均削減率と平均対応付け精度を求めた．上記以外の実験条件は 5.1 と同じにした．

各手法を用いて人物画像の選択を行った場合の平均削減率を表 4 に示す．比較手法の腕姿勢および頭姿勢は，提案手法の足姿勢と比べて，人物画像の枚数を削減できていないことが分かった．この理由を考察するために，選択された画像を目視で確認したところ，歩行中に腕を組む人物や手を振る人物が含まれていた．これらの人物画像を選択するため，腕姿勢のしきい値 R の設定が緩くなり，多くの人物画像が削減されずに残っていた．同様に，歩行中に頭が下を向いている人物が含まれており，頭姿勢のしきい値 R の設定が緩くなっていた．腕姿勢や頭姿勢のしきい値 R の設定を厳しくしたところ，これらの人物画像が全く選択されない状況が発生した．このため，腕姿勢や足姿勢では，しきい値の設定を緩くせざるを得ない状況であった．一方，足姿勢は全ての人物から安定に観測されるため，しきい値を適切に設定できていた．このことより，提案手法は冗長な人物画像の多くを削減することができたと考えられる．

表 4　MARS データセットにおいて各手法を用いて選択された人物画像の平均削減率 (%).
Table 4　Reduction rate of pedestrian images selected by each method in the MARS dataset.

提案手法 (足姿勢)	P_2	**50.5**	P_4	**38.8**
比較手法 1(腕姿勢)	P_2^a	4.9	P_4^a	4.1
比較手法 2(頭姿勢)	P_2^h	1.4	P_4^h	1.4

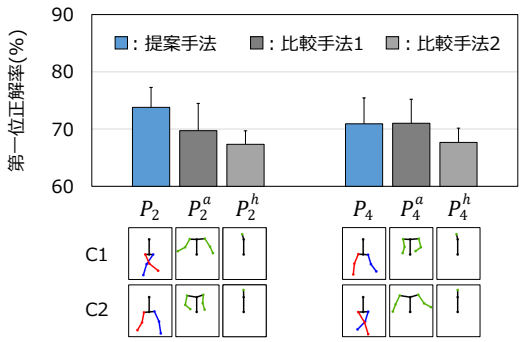


図 16　基準姿勢を設定して選択した際の MARS データセットにおける人物対応付けの平均精度.
Fig. 16　Accuracy of person re-identification using each pose in the MARS dataset.

各手法で選択された画像を用いて，人物を対応付けた時の平均精度を図 16 に示す．提案手法の足姿勢は，比較手法の腕姿勢と頭姿勢に比べて，同等以上の精度を得ることができた．以上より，足を基準姿勢に採用することは，足以外の身体部位を用いる場合に比べて有効であると言える．

5.3.4　人物対応付けで用いる特徴量の考察

本論文では，人物の個人性を表す特徴量として共起属性 [31], [33] を用いた．共起属性は，身体的な手がかりと外見的な手がかりの組合せで表現される．例えば，性別と年齢の共起，トップスとボトムスの共起，性別とボトムスの共起がある．それぞれの共起属性に該当する尤度を，人物画像から抽出された ELF(Ensemble of Localized Features) [34] 特徴量と線形サポートベクタマシンを用いて推定している．人物対応付けのための特徴量を，各共起属性の尤度を要素とするベクトルで表現している．

各共起属性の推定精度は，人物画像中の被写体の姿勢に依存することを予備実験で確認した．姿勢がほぼ同じであれば，共起属性の尤度は近い値をとる割合が多かった．提案手法を用いることで，足姿勢が近い人物画像が選択されることが 5.3.2 の実験で分かっている．提案手法の効果により，人物対応付けにおける特

徴量の要素である共起属性の尤度が類似する可能性が高まったため，対応付け精度が同等以上になったと考えられる．

6. ま と め

本論文では，人物対応付けシステムに準同型暗号を適用した場合の通信量の問題に対し，人物画像の集合の中から適切なものを選択する手法について述べた．提案手法では，画像選択の手掛りとして足姿勢の歩行周期に着目した．時間方向に分割された歩行周期の各区分において，空間方向に分布する関節部位の配置から，足の基準姿勢を設定した．画像選択の手法を，準同型暗号を用いた人物対応付けシステムに組み込み，実験でその有効性を評価した．独自データセットと公開データセットにおいて，足の基準姿勢を用いて人物画像を選択することで，人物対応付けの精度を同等以上としたままで，特徴量の通信量を大幅に削減できることを明らかにした．

今後の課題として，人物の動線が一定と仮定できない場合における基準姿勢の設計手法の開発が挙げられる．カメラの高さと角度が一定でない場合の基準姿勢の設計手法も開発が必要である．カメラ台数や人数が増加した場合の有効性の検証も挙げられる．符号化や軽量化を特徴量へ適用し，提案手法を組み合わせた場合の評価も挙げられる．IoT やエッジコンピューティングの分野で議論されている時系列データ量の削減手法 [35], [36] の考え方を取り入れることも考えられる．

謝　　辞

本研究の一部は，総務省 SCOPE(No. 172308003) および JSPS 科研費 18H04114 の助成を受けたものである．

文　　献

[1] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio. Full-body person recognition system. *Journal of Elsevier Pattern Recognition*, Vol. 36, No. 9, pp. 1997–2006, 2003.

[2] N. D. Bird, O. Masoud, N. P. Papanikolopoulos, and A. Isaacs. Detection of loitering individuals in public transportation areas. *Journal of IEEE Transactions on Intelligent Transportation Systems*, Vol. 6, No. 2, pp. 167–177, 2005.

[3] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2360–2367, 2010.

[4] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *Proceedings of XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 322–329, 2009.

[5] L. Bourdev, S. Maji, and J. Malik. Describing people: A poselet-based approach to attribute classification. In *Proceedings of International Conference on Computer Vision*, pp. 1543–1550, 2011.

[6] R. Layne, T. M. Hospedales, and Gong. Person re-identification by attributes. In *Proceedings of the British Machine Vision Conference*, No. 24, pp. 1–11, 2012.

[7] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev. Panda: Pose aligned networks for deep attribute modeling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1637–1644, 2014.

[8] T. Matsukawa and E. Suzuki. Person re-identification using cnn features learned from combination of attributes. In *Proceedings of International Conference on Pattern Recognition*, pp. 2428–2433, 2016.

[9] Z. Brakerski, C. Gentry, and V. Vaikuntanathan. Fully homomorphic encryption without bootstrapping. In *Proceedings of Innovations in Theoretical Computer Science Conference*, pp. 309–325, 2012.

[10] W. Lu, S. Kawasaki, and J. Sakuma. Using fully homomorphic encryption for statistical analysis of categorical, ordinal and numerical data. In *Proceedings of Network and Distributed System Security Symposium*, pp. 201–210, 2016.

[11] N. Dowlin, R. Gilad-Bachrach, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing. Cryptonets: applying neural networks to encrypted data with high throughput and accuracy. In *Proceedings of International Conference on Machine Learning*, Vol. 48, pp. 201–210, 2016.

[12] G. Shu, A. Dehghan, O. Oreifej, E. Hand, and M. Shah. Part-based multiple-person tracking with partial occlusion handling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1815–1821, 2012.

[13] Y. Kawanishi, D. Deguchi, I. Ide, and H. Murase. Trajectory ensemble: Multiple persons consensus tracking across non-overlapping multiple cameras over randomly dropped camera networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1471–1477, 2017.

[14] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel. A survey of appearance models in visual object tracking. *Journal of ACM Transactions on Intelligent Systems and Technology*, Vol. 4, No. 4, pp. 58:1–58:48, 2013.

[15] F. M. Khan and F. Brémond. Multi-shot person re-

identification using part appearance mixture. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 605–614, 2017.

[16] Ju. Han and B. Bhanu. Individual recognition using gait energy image. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322, 2006.

[17] K. Weinberger, A. Dasgupta, J. Attenberg, J. Langford, and A. Smola. Feature hashing for large scale multitask learning. In *Proceedings of Annual International Conference on Machine Learning*, pp. 1113–1120, 2009.

[18] Q. Shi, J. Petterson, G. Dror, J. Langford, A. Smola, A. Strehl, and S. V. N. Vishwanathan. Hash kernels. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, Vol. 5, pp. 496–503, 2009.

[19] W. Chen, J. Wilson, S. Tyree, K. Weinberger, and Y. Chen. Compressing neural networks with the hashing trick. In *Proceedings of International Conference on Machine Learning*, pp. 2285–2294, 2015.

[20] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 12, pp. 2916–2929, 2013.

[21] M. Ambai, T. Kimura, and C. Sakai. Fast and accurate object detection based on binary co-occurrence features. *Journal of IPSJ Transactions on Computer Vision and Applications*, Vol. 7, pp. 55–58, 2015.

[22] G. Irie, H. Arai, and Y. Tanigushi. Multimodal learning of geometry-preserving binary codes for semantic image retrieval. *Journal of IEICE Transactions on Information and Systems*, Vol. E100.D, No. 4, pp. 600–609, 2017.

[23] L. Chang, I. Rodés, H. Méndez, and E. del Toro. Best-shot selection for video face recognition using fpga. In *Proceedings of Iberoamerican Congress on Pattern Recognition*, pp. 543–550, 2008.

[24] Zhiguang Yang, Haizhou Ai, Bo Wu, Shihong Lao, and Lianhong Cai. Face pose estimation and its application in video shot selection. In *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 1, pp. 322–325, 2004.

[25] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 74–81, 2011.

[26] J. M.Burnfield J. Perry. *Gait Analysis: Normal and Pathological Function*. 2010.

[27] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, Vol. 10, pp. 207–244, 2009.

[28] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian. Mars: A video benchmark for large-scale person re-identification. In *Proceedings of Computer Vision and Pattern Recognition European Conference on Computer Vision*, pp. 868–884, 2016.

[29] Y. Kobayashi, H. Hobara, and H. Mochimaru. 2015: Aist gait database 2015. https://www.airc.aist.go.jp/dhrt/gait2015/index.html.

[30] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1302–1310, 2017.

[31] M. Nishiyama, S. Nakano, T. Yotsumoto, H. Yoshimura, Y. Iwai, and K. Sugahara. Person re-identification using co-occurrence attributes of physical and adhered human characteristics. In *Proceedings of International Conference on Pattern Recognition*, pp. 2085–2090, 2016.

[32] D. Baltieri, R. Vezzani, and R. Cucchiara. 3d body model construction and matching for real time people re-identification. In *Proceedings of Eurographics Italian Chapter Conference*, 2010.

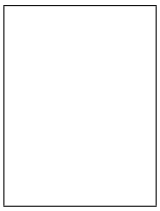
[33] 中野, 四元, 吉村, 西山, 岩井, 菅原. 身体と外見の共起属性を用いた人物対応付け. 電子情報通信学会 D, 第 J100-D 巻, pp. 104–114, 2017.

[34] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of European Conference on Computer Vision,*, pp. 262–275, 2008.

[35] R. Shinkuma, T. Nishio, J. Katto, and T. Tsuda. Data assessment and prioritization in mobile networks for real-time prediction of spatial information with machine learning. In *Proceedings of 2019 IEEE First International Workshop on Network Meets Intelligent Computations*, pp. 1–6, 2019.

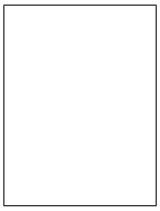
[36] K. Kanai, K. Ogawa, M. Takeuchi, J. Katto, and T. Tsuda. Intelligent video surveillance system based on event detection and rate adaptation by using multiple sensors. *Journal of IEICE Transactions on Communications*, Vol. E101.B, No. 3, pp. 688–697, 2018.

（平成 xx 年 xx 月 xx 日受付）



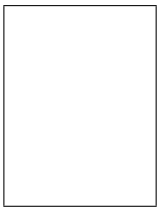
福田　尚悟

2019 年 鳥取大学工学部電気情報系科卒業．2019 年 鳥取大学大学院持続性社会創生科学研究科工学専攻博士前期課に在学中．



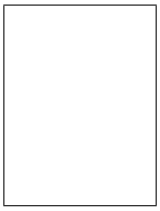
森田　一成

2018 年 鳥取大学工学部知能情報工学科卒業．2019 年 鳥取大学大学院持続性社会創生科学研究科工学専攻博士前期課に在学中．



西山　正志　（正員：シニア会員）

2000 年 岡山大学工学部情報工学科卒業．2002 年 同大大学院博士前期課程了．同年株式会社東芝入社．同社研究開発センターを経て，現在鳥取大学大学院工学研究科准教授．2011 年 東京大学大学院学際情報学府にて博士（学際情報学）を取得．カメラを用いた人物認識を始めとするパターン認識およびインタラクションの研究に従事．山下記念研究賞や画像センシングシンポジウム優秀学術賞など受賞．電子情報通信学会，情報処理学会各会員．



岩井　儀雄　（正員）

1992 年 (平成 4 年) 大阪大学基礎工学部情報工学科卒業．1997 年 (平成 9 年) 大阪大学大学院基礎工学研究科博士課程後期修了．同年同大学院助手．2003 年 (平成 15 年) 同大学院助教授．2004 年 (平成 16 年)5 月〜2005 年 (平成 17 年)3 月英国ケンブリッジ大学客員研究員．2007 年 (平成 19 年) 同大学院准教授．2011 年 (平成 23 年）鳥取大学大学院工学研究科教授．コンピュータビジョン，パターン認識の研究に従事．博士（工学）