

# 天井カメラを用いて観測された身体動揺における自己遮蔽の影響を考慮した人物対応付け

非会員 神谷 卓也\* 非会員 山口 優太\*  
非会員 西山 正志\* 非会員 岩井 儀雄\*

## Identification of People by Measuring Body Sway with Self-Occlusion using Top-View Camera

Takuya Kamitani\*, Non-member, Yuta Yamaguchi\*, Non-member,  
Masashi Nishiyama\*, Non-member, Yoshio Iwai\*, Non-member

We propose a method for identifying people using body sway measured from head regions under a condition that a top-view camera observes bodies disturbed by self-occlusion. In order to correctly represent identities of body sway, it is necessary that the appearances of people are accurately acquired from the camera. However, the deficits of the appearances are contained when occurring self-occlusion. The deficits cause a problem that the identification performance of an existing method is wrongly degraded because the method uses whole-body regions to represent identities. To overcome the problem of self-occlusion, we represent identities of body sway by using head regions acquired from a top-view camera. Our method computes the silhouette images of the head regions by applying a segmentation technique. To represent identities of body sway, we temporally measure the movements of the head regions by spatially dividing them into local regions. The experimental results show that the identification performance improved from 17.3 % using the whole-body regions of the existing method to 57.9 % using the head regions of our method.

キーワード：身体動揺, 自己遮蔽, 人物対応付け

**Keywords:** Body sway, Self-occlusion, Identification

### 1. はじめに

安心安全な社会を実現するために、様々な場所に設置されている防犯カメラを活用した人物対応付けシステム<sup>(1)(2)</sup>が求められている。高精度な対応付けを実現するために、人物の動きから個人性を捉える行動手掛かり<sup>(3)~(6)</sup>が注目を集めている。行動手掛かりの代表として、歩行中の手足の周期的な動きを表す歩容<sup>(7)(8)</sup>が広く利用されており、その対応付けの精度は高いことが知られている。人物は常に歩行している訳ではなく、セキュリティゲートやエレベータの前で立ち止まり、扉の開閉を待つことがある。立ち止まった人物からは、手足の周期的な動きが発生しないため、歩容を手掛かりとして用いることは難しい。このため近年では、身体動揺<sup>(9)</sup>が新たな行動手掛かりとして期待されつつある。身体動揺とは、人物が意識的に静止しても発生する

身体の僅かな動きであり、立ち状態の人物の個人性を捉える上で有効な手掛かりとなる。例えば、工場内のセキュリティゲートや自動ドアの前で立ち止まる人物を対応付けるために、身体動揺の活用が期待される。工場で働く人物は、同じ種類の制服を着用するため、それぞれの人物の見た目は非常に似ている。身体動揺から個人性を適切に捉えることで、見た目の違いが少ない服装においても、正しく人物を対応付けることを狙う。これにより、工場の物理セキュリティを強化する応用が期待される。

人物を正しく対応付けるためには、身体動揺に含まれる個人性を、空間方向と時間方向とで適切に捉える必要がある。空間方向で捉える個人性とは身体の形状であり、時間方向で捉える個人性とは身体の揺れである。以下では、これらの個人性を、一般的に利用されている防犯カメラで捉えることを考える。立ち状態の人物をカメラを通して見た場合、空間方向の身体形状は、人物の外見そのものとして観測される。また、時間方向の身体揺れは、人物の外見が連続的に変化する様子として観測される。このため身体動揺を用いた人物対応付けでは、人物の外見をカメラを通して

\* 鳥取大学大学院 工学研究科  
〒 680-8552 鳥取市湖山町南 4 丁目 101 番地  
Graduate School of Engineering, Tottori University  
4-101 Koyama Minami, Tottori Japan 680-8552

精度よく観測することが重要となる。ただしカメラから観測される人物の外見は、遮蔽によりその形に欠損が生じることが多い。欠損による外見の変化は、人物とカメラとの位置関係に強く依存する。身体動揺を用いた人物対応付けにおいて、遮蔽の問題は精度の低下を引き起こすため、早急な解決が求められている。

身体動揺における遮蔽の発生原因について考察する。遮蔽が発生する原因は大きく分けて二つ存在する。第一の原因として、ある人物を別の人物が隠すことで、カメラから観測される外見の一部が欠損することが挙げられる。これは相互遮蔽と呼ばれ、複数の人物を同時に観測する場合に発生する。カメラを天井に設置することで、相互遮蔽の発生を抑えることが可能である。第二の原因として、ある身体部位が、その人物の別の部位を隠すことで、外見の一部が欠損することが挙げられる。これは自己遮蔽と呼ばれ、天井カメラで観測する場合でも発生する。身体動揺を用いた人物対応付けでは、特に自己遮蔽の影響をどのように抑制するかを考慮しなければならない。

カメラから観測される人物の外見において、自己遮蔽の影響が小さい領域を考える。天井カメラを用いる場合、頭領域が自己遮蔽の影響を最も受けにくいと想定される。既存手法において頭領域は、人物対応付けを目的としていないが、人数カウント<sup>(9)~(11)</sup>や人物追跡<sup>(12)~(13)</sup>の目的で活用されている。一方、身体動揺を用いる人物対応付けの既存手法<sup>(6)</sup>では、カメラから観測される全身領域そのものを用いて個人性を捉えていた。既存手法は頭領域の活用を考慮していなかったため、自己遮蔽の影響を強く受ける問題があった。

そこで本論文では、身体動揺を用いた人物対応付けにおいて、自己遮蔽の影響を抑制するために、天井カメラから観測された頭領域を用いて個人性を捉える手法について述べる。提案手法では、天井カメラから観測された画像列から、空間方向のセグメンテーションを用いて頭領域を推定する。人物対応付けの特徴量を抽出するために、頭領域が時間方向で連続的に変化する様子を振動量として算出する。このように頭領域を用いることで、身体動揺に含まれる個人性を空間方向と時間方向とで頑健に捉えることを狙う。提案手法の有効性を確認するために、自己遮蔽の影響を含む場合について、実験協力者 50 名から身体動揺の画像列を収集した。実験結果より提案手法は、全身領域を用いるベースライン手法と比べて、人物対応付けの精度を 40.6 ポイント改善することを確認した。以下では、2 章で関連研究について述べ、3 章で自己遮蔽の影響について説明する。4 章で提案手法について述べ、5 章で提案手法の性能を評価し、最後に 6 章でまとめる。

## 2. 関連研究

**〈2・1〉 行動手掛かりを表す時空間特徴量** 行動手掛かりの一種である歩容を用いた人物対応付けにおいて、時空間の変化から個人性を捉える特徴量が提案されている。代表的な既存手法として、GEI (Gait Energy Image)<sup>(7)</sup> と STHOG

(Spatio-Temporal Histograms of Oriented Gradient)<sup>(14)</sup> が挙げられる。歩行中の人物の手足は、不規則な動きをする訳ではなく、同じ動きを繰り返す。GEI では、カメラから観測される外見をシルエット画像列として算出する。空間方向の個人性を捉えるシルエット画像列を、時間方向の動きについて一周分だけ平均化することで時空間特徴量を抽出している。一方、STHOG では、空間方向の勾配変化を表す HOG (Histograms of Oriented Gradients)<sup>(15)</sup> を、時間方向に拡張することで特徴量を抽出している。画像列に対し時空間のブロックを設け、その領域内の勾配からヒストグラムを生成している。これらの既存手法は、歩容の周期的な動きを表すために設計された時空間特徴量であり、本論文で対象とする身体動揺を表す特徴量として十分とは言えなかった。

行動認識の分野で Dynamic image<sup>(16)</sup> と呼ばれる時空間特徴量が提案されている。この特徴量では、画像列の時間方向の順序関係を維持しつつ、空間方向の外見変化を表す重みベクトルを Rank SVM<sup>(17)</sup> で算出している。また文献<sup>(18)</sup>では重みベクトルを近似式で算出する手法が提案されている。Dynamic image は、サッカーのリフティング、水泳のクロール、卓球のショットのように、手足など身体の姿勢が大きく変化する動きを表すために設計された時空間特徴量であり、身体動揺の微細な動きを表す特徴量として十分とは言えなかった。

文献<sup>(6)</sup>では、身体動揺における時空間特徴量を抽出するため、シルエット画像列を空間方向のブロックに分割し、そのブロック中で時間方向の振動量を算出している。ただし、人物の全身領域を用いることを前提としていたため、本論文で取り扱う自己遮蔽を考慮しておらず、人物対応付けの精度が低下する問題があった。

**〈2・2〉 頭領域の推定** 人数カウントや人物追跡を目的とし、天井カメラから観測された画像列から頭領域を推定する既存手法<sup>(9)~(13)</sup>が提案されている。文献<sup>(9)~(11)</sup>では、天井に深度カメラを設置し、人物までの距離が近い領域を頭としている。深度カメラは普及しつつあるものの、防犯カメラとしては一般的ではないため、本論文では利用しないこととする。文献<sup>(10)~(13)</sup>では、人物の頭の形状は楕円であると仮定し、頭領域のパラメータを推定している。これらの既存手法では、楕円として大まかに近似するため、個人間の頭の外見の違いを詳細に捉えることが困難である。そのため、人物の位置を検出する目的に適しているが、本論文で目的とする人物対応付けの手掛かりとして適しているとは言えない。文献<sup>(12)</sup>では、人数カウントを目的とし、画像セグメンテーションを用いて頭領域を推定している。この既存手法では、ピクセル単位で人物の頭領域を推定するため、個人性による外見の違いを詳細に捉えることが期待される。ただし、推定された頭領域が人物対応付けに有効であるかどうかは検証されていないため、本論文では頭領域を用いた人物対応付けの精度を新たに検証する。

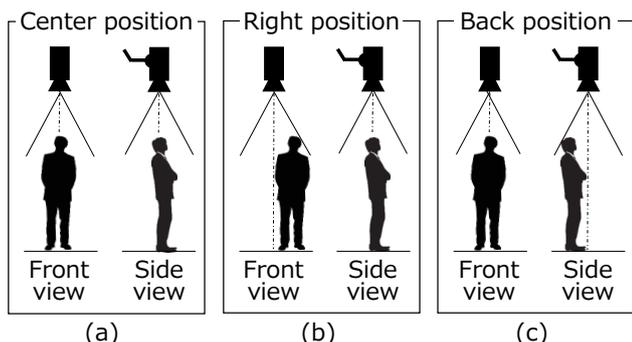


Fig. 1. The standing positions of individuals to investigate the influence of self-occlusion.

### 3. 自己遮蔽の影響

カメラから観測される人物の外見は、自己遮蔽により立ち位置毎で大きく変化する。予備調査を行うため、異なる立ち位置の人物をカメラで観測した時の外見を比較した。この調査では、天井カメラの位置を固定し、Fig. 1 に示す3つの位置に立つ人物を観測した。カメラ光軸が床と交わる点を中央位置とした。図中 (a) は、中央位置に立つ人物の観測条件を表す。また (b), (c) は、中央位置からそれぞれ右と後に移動した位置に立つ人物の観測条件を表す。

Fig. 2 に、人物の立ち位置が変化した時の外見の例を示す。ここでは二名の人物を観測の対象とし、上段が人物1、下段が人物2の外見を表す。図中 (a), (b), (c) を比較すると、カメラから観測される外見である全身領域は、どちらの人物においても立ち位置毎に大きく異なることが分かる。次に本論文で取り扱う頭領域について述べる。Fig. 3 に、頭領域を緑色で強調した例を示す。なおカメラの観測条件はFig. 2 と同じとした。Fig. 3 (a), (b), (c) を比較すると中央、右、後の全ての立ち位置において、頭領域は類似した形で観測されることが分かる。一方、頭以外の身体部位として観測される肩領域は、立ち位置毎に形が大きく異なり、自己遮蔽の影響を強く受けていることが分かる。図中 (a) の中央位置では、左肩領域が右肩領域と対称であるが、(b) の右位置では、中央では観測されていた右肩領域が隠れており、観測されていなかった左肩領域が新たに出現している。また (c) でも同様の傾向が見られる。以上より頭領域は、他の身体部位の領域と比べて、自己遮蔽の影響が小さく、それぞれの立ち位置において安定した観測ができると言える。自己遮蔽が生じた場合でも身体動揺に含まれる個性を捉えるために、本論文では頭領域を用いることとする。

### 4. 提案手法

〈4・1〉 概要 身体動揺に含まれている個性を捉えるために、頭領域を用いて時空間特徴量を抽出する手法について述べる。その流れを Fig. 4 に示す。提案手法では、立ち状態の人物の画像列を、天井カメラを用いて観測する。なお人物の立ち位置は自由とするが、直立姿勢を維

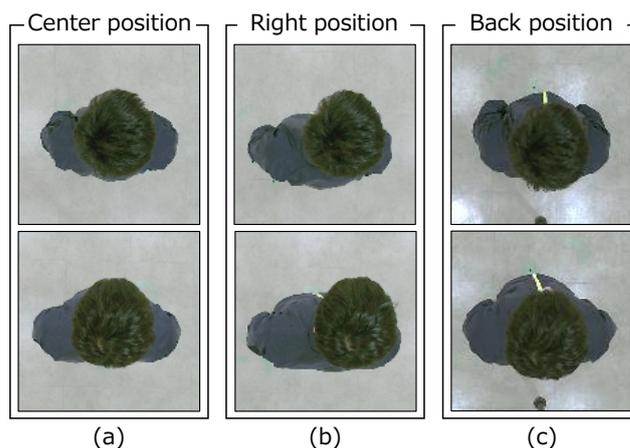


Fig. 2. Examples of the whole body region acquired from two individuals in different standing positions.

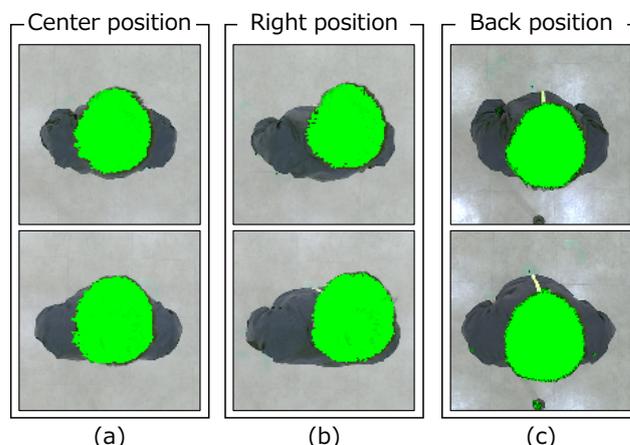


Fig. 3. Examples of the head region acquired from two individuals in different standing positions. The green pixels represent the head regions in the images.

持すると仮定する。自己遮蔽の影響を抑制するために、観測された画像列から、頭領域を表すシルエット画像列をセグメンテーション手法を用いて推定する。推定されたシルエット画像列を空間方向のブロックへ分割し、各ブロックにおける時間方向の振動量の特徴量として抽出する。以下、提案手法の詳細について述べる。

〈4・2〉 人物の頭領域の推定 自己遮蔽の影響を抑えるために、天井カメラから観測された画像列の頭領域を推定する手法について述べる。頭領域は、その他の身体部位や背景の領域と比べると、テクスチャの変化が細かいと想定される。大量の訓練サンプルから頭の外見の変動を統計的に学習することができるため、肩や背景と色合いが近くなる場合でも、頭領域を正しく推定することが期待される。本論文では、空間方向のセグメンテーション手法<sup>(19)~(24)</sup>を用いることで、頭領域を頑健に推定することを狙う。セグメンテーション手法において様々な統計アプローチが採用されているが、高い推定精度を実現できる深層学習<sup>(25)~(27)</sup>が近年注目を集めている。深層学習を適用するために、頭領域を示す教師ラベルと、全身や背景の様々な変動を含む

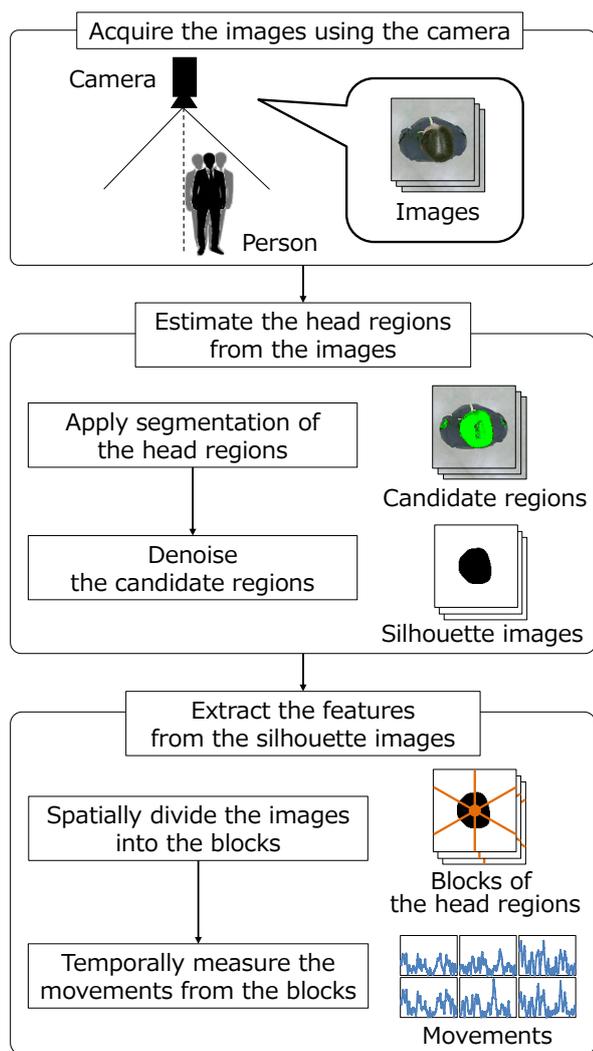


Fig. 4. Overview of our method.

大量のカラー画像を準備しておく. Fig. 5 (a) に教師ラベルの例, (b) にカラー画像の例を示す. 入力としてカラー画像を与え, 推定されたシルエット画像が教師ラベルに近づくようにネットワークモデルの学習を行う.

深層学習を用いたセグメンテーション手法では, 推定されるシルエット画像にノイズが含まれることが多い. ノイズとして, 頭以外の画素を頭とする誤推定と, 頭の画素を頭としない誤推定が存在する. Fig. 6 (a) にノイズの例を示す. 本論文ではノイズ除去を行うために, シルエット画像内に存在する複数の候補領域から面積最大のものを選択し, その領域内に存在する誤推定の画素を塞ぐ処理を行う. さらにメディアンフィルタを適用し, 頭領域と背景領域の境界付近のノイズを軽減する. 図中 (b) にノイズ除去後のシルエット画像の例を示す.

〈4・3〉 頭領域を用いた時空間特徴量の抽出 人物対応付けで用いる時空間特徴量を, 頭領域のシルエット画像列から抽出する手法について述べる. 本論文では時空間特徴量を抽出するために, 我々が提案した既存手法<sup>(6)</sup>を頭領域へ適用できるように拡張する. 立ち状態の人物から観測さ

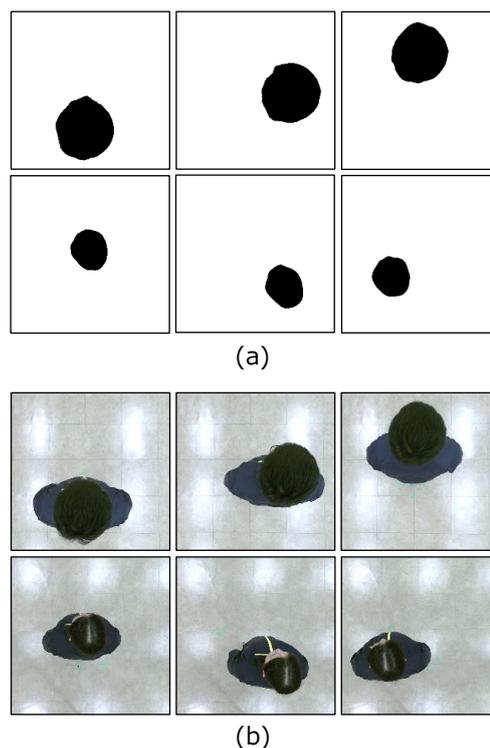


Fig. 5. Examples of the images and annotation labels to train the network model for head segmentation.

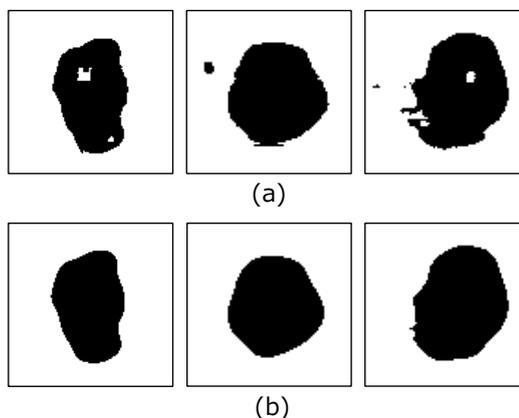


Fig. 6. Examples of the silhouette images containing noises and the shihouette images after denoising.

れる頭領域は, 前後左右に揺れながら微小に変化していく. 揺れの動きは, その中心となる位置に対応する参照時刻が基準となる. まず参照時刻を決定するために, 頭領域のシルエット画像列の中で, 互いに最も類似する画像を一枚選択する. この画像を参照時刻とし, その画像の頭領域の重心を求めることで中心位置を設定する. 空間方向の変化を捉えるために, 中心位置を基準に放射線状のブロックヘシルエット画像列を等分割する. さらに時間方向の変化を捉えるため, 各ブロックにおいて時間微分を行うことで振動量を算出する. 最後に, 頭領域の各ブロックの振動量に対し, パワースペクトル密度 (PSD: Power Spectral Density)<sup>(28)</sup>を適用することで特徴量を抽出する.



Fig. 7. Postures and clothes of the participants.

## 5. 実験

### 〈5・1〉 評価データセット

提案手法の有効性を確認するために、異なる位置に立つ人物の身体動揺について、天井カメラから観測された画像列を収集した。実験協力者の人数は50名(平均年齢  $22.7 \pm 3.0$  歳, 男性42名, 女性8名)とした。画像列の観測中は、手足を揃えた直立姿勢を維持するよう実験協力者へ指示した。工場の制服を想定し、紺色ナイロン製ジャケットの着用を指示した。Fig. 7に実験協力者の姿勢と服装の例を示す。カメラの取り付け位置を床からの高さ2.5メートルとし、カメラ光軸が床面と垂直となるようキャリブレーションを行った。カメラの解像度を  $1920 \times 1080$  画素とし、サンプリング周波数を30ヘルツとした。

実験協力者は、カメラ光軸が床と交わる点である中央位置、そこから前後左右にそれぞれ0.15メートル離れた位置に立つこととした。Fig. 8(a)に人物の立ち位置を上から見た場合を示す。実験協力者へ立ち位置を指示するため、床に円マーカを設置した。その際、図中(b)のように、両足の中心が円マーカと重なるよう実験協力者へ指示した。人物が前位置に立つ場合の観測条件を図中(c)に示す。実験協力者の正面に設置されたターゲット点を見ることで、観測中の実験協力者の顔向き変化を抑制した。なお立ち位置毎で人物の正面となるようターゲット点を設置した。それぞれの立ち位置での観測時間は60秒とした。実験協力者は1回の観測が終了する度に、別の場所に着席し休憩をとった。実験協力者1名に対し、5箇所の立ち位置で2回の観測を行った。なお立ち位置の観測順序はランダムとした。以下の実験では、観測された映像の両端から420画素ほどの領域を削除し、 $256 \times 256$ 画素にリサイズした画像列を用いた。

### 〈5・2〉 頭領域の推定精度

天井カメラから観測された画像列を用いて頭領域の推定精度を評価した。頭領域のセグメンテーション手法として文献<sup>(12)</sup>で利用されているU-Net<sup>(26)</sup>を用いた。ネットワーク構造として、ダウンサンプリング層を8個、アップサンプリング層を8個用意した。実験協力者50名の中から25名をランダムに選択し、U-Net

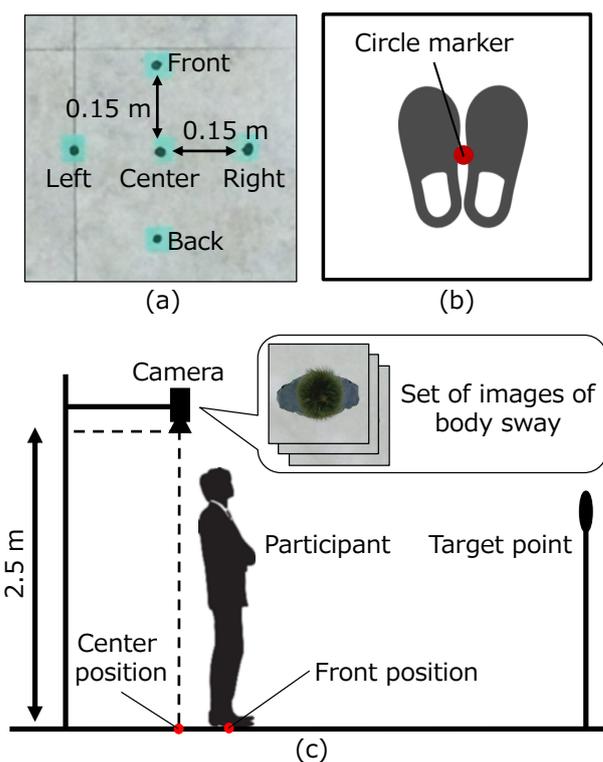


Fig. 8. The relationship between the participant and the standing position. In (a), the standing positions on the floor are represented. In (b), the marker to align the feet position of the participant is represented. In (c), the setup for acquiring the set of images of body sway in the case that the participants stand on the front position is represented.

の訓練サンプルとした。また、残りの25名をテストサンプルとした。全ての実験協力者の全フレームの合計である450,000枚から、それぞれランダムに選択された45,000枚を訓練サンプルとテストサンプルの枚数とした。また、訓練とテストで、サンプルの被りはないようにした。画像と教師ラベルのサイズは  $256 \times 256$  画素とした。訓練時のエポック数は200とした。頭領域の推定精度の指標として、適合率と再現率の調和平均であるF値を用いた。F値は1に近づくにつれ推定精度が高くなる。実験協力者のランダム選択を5回試行しF値の平均を求めた。

頭領域の推定精度を評価したところ平均F値が  $0.96 \pm 0.03$  となった。Fig. 9に、U-netから出力された頭領域の例を示す。この図では実験協力者4名、立ち位置5か所の結果を挙げる。図中(a), (b), (c)では、どの実験協力者についても、頭領域が精度よく推定されていることが分かる。一方(d)では、頭領域の形が髪留めの影響で大きく変わっていることが分かる。以上の結果より、誤推定が一部見られるものの平均F値が1に近いと、単純なU-netでも本実験環境では高い推定精度を得ることができると言える。

〈5・3〉 人物対応付けの基本性能 自己遮蔽が発生する場合において、頭領域を用いる提案手法が人物対応付けに有効であるかどうかを調査した。ここでは頭以外の身体

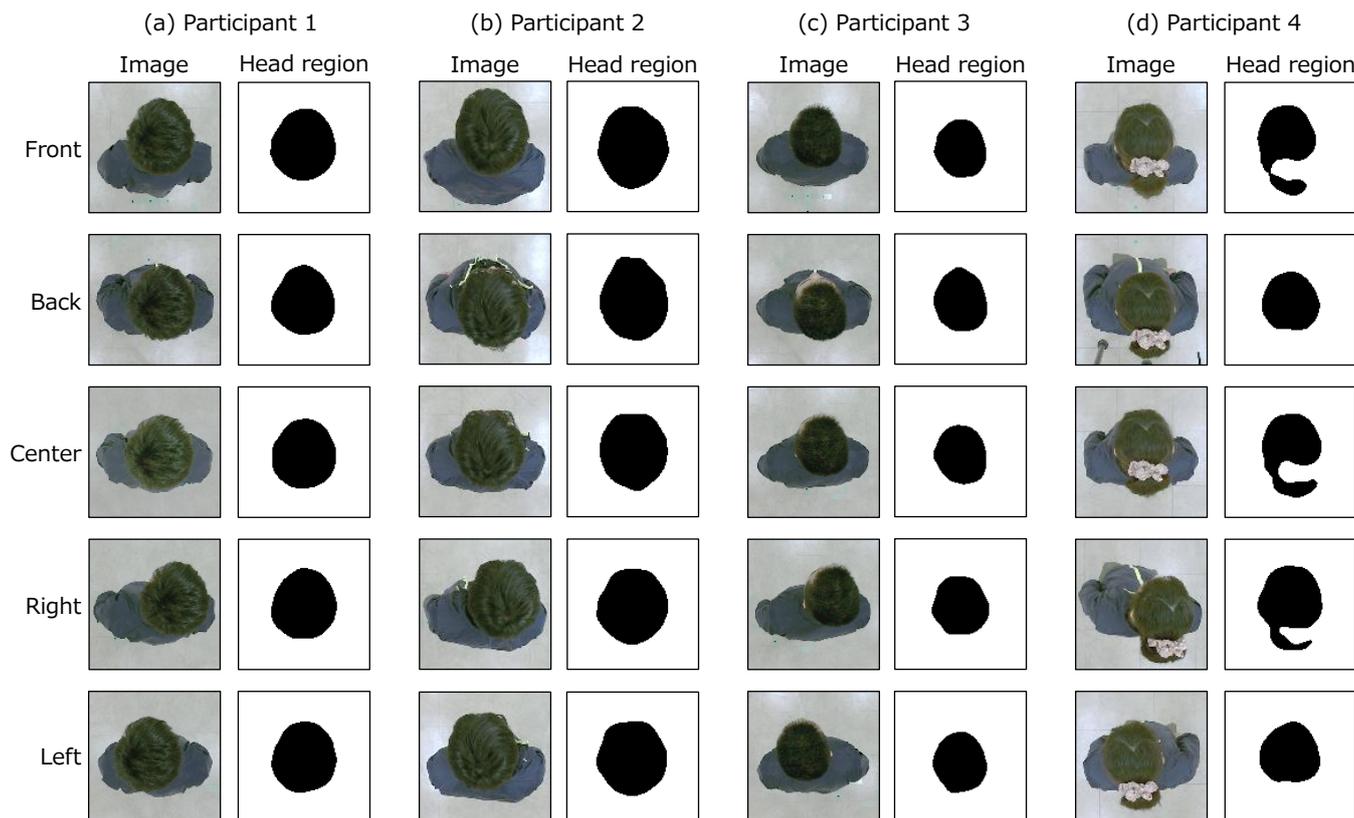


Fig. 9. Examples of the head regions estimated from the image of each standing position.

部位を含む領域を比較の対象とした。以下に実験条件を挙げる。

**Head:** 提案手法を用いて推定された頭領域を用いた。Fig. 10(a)にその例を示す。

**Whole body:** 文献<sup>(6)</sup>の全身領域を用いた。図中 (b)にその例を示す。

**Shoulder:** 全身から頭を除いた肩領域を用いた。図中 (c)にその例を示す。

なお全身領域および肩領域を推定するために、教師ラベルをそれぞれの領域とし〈4・2〉で述べた手法を適用した。人物対応付けに用いる特徴量を、各領域のシルエット画像列から〈4・3〉で述べた手法を用いて抽出した。空間方向のブロック分割数は、各領域とも25とした。人物対応付けのクエリとして前、左、右、後のいずれかの位置から観測されたシルエット画像列を用いた。また、ターゲットとして中央位置から観測されたシルエット画像列を用いた。クエリとターゲットを対応付けるために最近傍法を適用した。対応付け精度の指標は一位正解率とした。なお計量学習としてLMNN (Large Margin Nearest Neighbor)<sup>(29)</sup>を適用した。実験協力者50名のうち、クエリとターゲットに含まれない25名をランダムに選択し計量学習に用いた。ランダム選択の試行回数は5回とし、残りの実験協力者25名をクエリとターゲットとし対応付けに用いた。一位正解率の平均を算出し、対応付けの性能を評価した。

身体部位の領域ごとの人物対応付けの平均一位正解率を

Table 1. Comparison of the identification performance achieved with the regions of the body parts.

Region	First matching rate (%)
<b>Head</b>	<b>57.9 ± 11.1</b>
Whole body	17.3 ± 6.8
Shoulder	9.8 ± 5.3

Table 1に示す。提案手法の頭領域は、全身領域や肩領域と比べて、対応付け精度が高いことが分かる。一方、肩領域は、頭領域や全身領域と比べて、対応付け精度の大きな低下が見られた。以上の結果より、天井カメラより観測される頭領域は、全身領域や肩領域に比べて自己遮蔽の影響を大きく抑えることができ、人物対応付けの精度向上に有効であることが確認された。

**〈5・4〉 空間特徴量と時間特徴量に対する精度比較** 頭領域から抽出された時空間特徴量が有効であることを確認するために、空間方向または時間方向の個人性を捉える特徴量を用いて人物対応付けの精度を評価した。以下に比較条件を挙げる。

**Spatio-temporal:** 提案手法を用いて抽出した。

**Spatial:** 頭領域から空間方向のみの個人性を捉えるために、天井カメラで観測されたシルエット画像列から、一枚の画像を選択した。ここでは〈4・3〉で述べた参照時刻に対するシルエット画像を用いた。

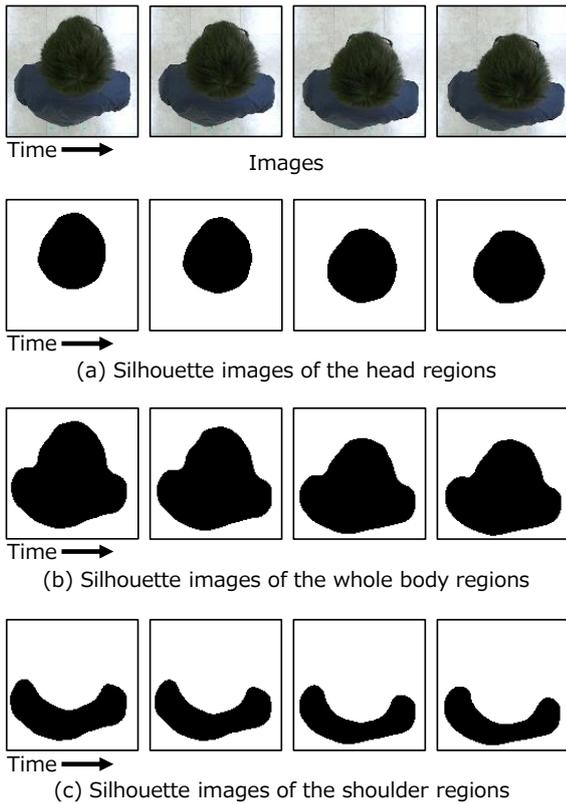


Fig. 10. Examples of the silhouette images of the body parts.

Table 2. Comparison of the identification performance using the temporal feature or spatial feature.

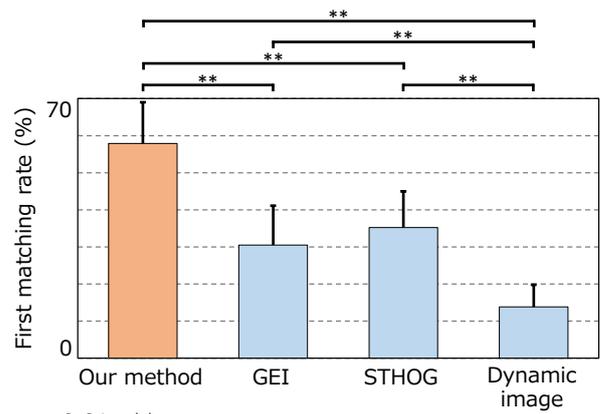
Feature	First matching rate (%)
<b>Spatio-temporal</b>	<b>57.9 ± 11.1</b>
Spatial	33.8 ± 10.7
Temporal	40.2 ± 9.9

**Temporal:** 頭領域から時間方向のみの個人性を捉えるために、各時刻のシルエット画像において重心位置を計測し、その時間方向の変化を記録した。重心位置は2次元ベクトルであり、そのままではPSDを適用できないため、直交する2本の基底へ重心位置を射影し、それぞれの基底上で振動量を算出した。

特徴量以外の実験条件は〈5・3〉と同じとした。

頭領域から抽出された空間特徴量または時間特徴量を用いた場合の人物対応付けの精度を Table 2 に示す。時間特徴量は、空間特徴量に比べ、精度が高いことが分かった。提案手法である時空特徴量は、空間特徴量や時間特徴量と比べ、対応付けの精度が高いことが分かった。以上の結果より、頭領域から空間方向と時間方向の個人性を同時に捉える提案手法は有効であると言える。

〈5・5〉 既存手法の時空間特徴量の評価 既存手法の時空間特徴量を頭領域から抽出した場合について、人物対応付け精度を評価した。ここでは、歩容認証の分野で広く利用されている GEI<sup>(7)</sup> と STHOG<sup>(14)</sup> を対象とした。また、行



$p < 0.01 : **$

Fig. 11. Comparison of the identification performance between our method and the existing methods.

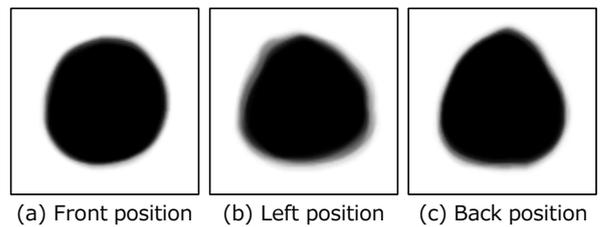


Fig. 12. Examples of features of the GEI.

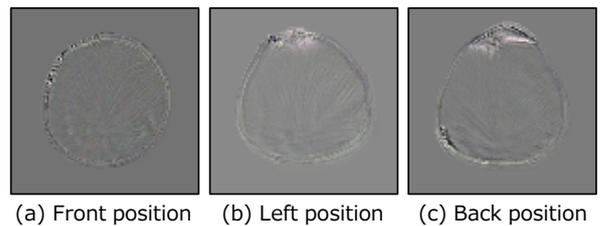


Fig. 13. Examples of features of the dynamic image.

動認識で利用されている Dynamic image<sup>(16)</sup> も対象とした。GEI を抽出するために、60 秒間で観測されたシルエット画像列から平均を求めた。STHOG を抽出するために、頭領域の画像列に対する時空間ブロックの数を  $6 \times 6 \times 6 = 216$  個とし勾配を算出した。Dynamic image を抽出するために、60 秒間で観測された頭領域の画像列に対して Rank SVM を適用した。時空間特徴量以外の実験条件は〈5・3〉と同じとした。頭領域から既存手法の時空間特徴量を抽出した場合の人物対応付けの精度を Fig. 11 に示す。提案手法は、既存手法の GEI や STHOG, Dynamic image と比べて、精度が高いことが確認された。ウィルコクソンの符号順位検定およびボンフェローニ補正を適用した。その結果、提案手法と全ての既存手法との間で有意差が見られた。

同じ人物の画像列から抽出された GEI を可視化した例を Fig. 12 に示す。図中 (a), (b), (c) は、異なる立ち位置で観測された同一人物の GEI である。同じ人物の画像列から抽出されているにもかかわらず、その頭領域の輪郭は人物の

立ち位置ごとに異なった形状であった。立ち位置変動による自己遮蔽は、頭領域であっても発生し、空間特徴である頭領域の輪郭形状を変化させたと考えられる。既存手法の GEI は空間特徴を主に捉え、そこに時間特徴を加える設計思想であると想定される。GEI の性能低下の原因は、頭領域に残る自己遮蔽の影響であると推測される。

Dynamic image を可視化した例を Fig. 13 に示す。同じ人物から生成された Dynamic image 内の頭領域の輪郭でも、GEI の結果と同様に、立ち位置ごとに異なった形状が見られた。Dynamic image は空間特徴を主に捉える設計思想であると想定されるため、頭領域に残る自己遮蔽の影響を受けて認識性能が低下したと推測される。

次に、STHOG について述べる。画像列から抽出された STHOG は、2つの勾配ヒストグラムで構成される。各勾配ヒストグラムは空間方向と時間方向のそれぞれから算出される。各ヒストグラムのビン数を同じ値に設定することで、空間特徴と時間特徴を同じ次元数で表現している。STHOG 中の空間特徴は、GEI と同様に、頭領域に残る自己遮蔽の影響を受けていると考えられる。この影響により、STHOG を使用した際の対応付け性能が低下したと推測される。

提案手法は、時間特徴を主に捉え、そこに空間特徴を加える設計思想である。提案手法は既存手法に比べて、頭領域に残る自己遮蔽の影響を受けにくく、身体動揺に含まれる時間方向の個人性を頑健に捉えることができたと考えられる。

〈5・6〉 同じ制服を着用した人物の対応付けにおける服装変化の影響調査 提案手法を実場面に適用するためには、人物の立ち位置以外の条件が変化した場合についても評価する必要がある。ここでは、Fig. 7 以外の服装を着用した人物に対して、提案手法を適用することを考える。本実験では、新たに2種類の服装を用意し、それぞれの服装において人物対応付けの性能を評価した。荷物の配達員を想定し、濃青色の帽子と薄青色の上着を服装 1 として準備した。工事現場の作業員を想定し、白色のヘルメットと濃緑色のジャケットを服装 2 として準備した。服装 1 と服装 2 を着用した例を Fig. 14 (a), (b) にそれぞれ示す。なお、人物の立ち位置を〈5・1〉の中央位置のみとした。

それぞれの服装を着用した人物の身体動揺を計測するため、天井カメラを用いて画像列を収集した。実験協力者の人数は 34 名 (平均年齢  $22.5 \pm 1.6$  歳, 男性 29 名, 女性 5 名) とした。すべての実験協力者に対し、服装 1 と服装 2 を着用した場合で、それぞれ 2 回ずつ撮影を行った。なお、各実験協力者において、服装の着用順序はランダムとした。実験協力者の人数および立ち位置以外の撮影条件は〈5・1〉と同じとした。

頭領域から抽出された時空間特徴量を用いて、服装 1 と服装 2 のそれぞれにおける人物対応付けの性能を評価した。実験協力者の人数および立ち位置以外の実験条件は〈5・3〉と同じとした。各服装における人物対応付けの平均一位正解率を Table 3 に示す。どちらの服装においても、その性能は



Fig. 14. Variation of the clothes of the participants.

Table 3. Comparison of the identification performance under different clothing conditions.

	First matching rate (%)
Clothes 1	$56.5 \pm 9.9$
Clothes 2	$52.4 \pm 9.7$

チャンスレート 5.9% より高いことが確認できた。ただし、Fig. 7 の服装を着用した人物を同条件で対応付けた際の性能である  $81.2 \pm 8.0\%$  (注: 実験協力者の人数は異なる) と比べると、Fig. 14 のどちらの服装を着用した場合でも人物対応付けの性能は低下していると考えられる。服装 1 や服装 2 を着用すると、帽子やヘルメットによって頭領域が隠され、それぞれの人物の見え方が近くなる。これにより、空間方向に存在する外見の個人性が表れにくくなるため、対応付けの性能が低下したと考えられる。この問題を解決するために今後の課題として、身体動揺に含まれる時間方向の個人性をより顕著に表す特徴量の設計が求められる。

## 6. まとめ

本論文では、身体動揺を用いた人物対応付けにおいて自己遮蔽の影響を抑制するため、天井カメラから観測された頭領域を用いて時空間方向と空間方向の個人性を捉える手法を提案した。天井カメラから観測された画像列の頭領域を推定するために、深層学習を用いたセグメンテーションを適用しノイズ除去を施した。身体動揺に含まれる個人性を捉えるために、シルエット画像列を空間方向のブロックに分割し、それぞれのブロックにおいて時間方向の微分を行うことで特徴量を算出した。提案手法の有効性を確認するために、自己遮蔽による頭領域の欠損を含む身体動揺の画像列を収集した。実験結果より、頭領域を用いる提案手法は、全身領域を用いる既存手法と比べて、精度が大幅に改善することを確認した。また、既存の時空間特徴量と比較して、頭領域を用いる提案手法は身体動揺を用いた人物対応付けに有効であることを確認した。

今後の課題として、帽子やヘルメットなど頭領域の外見が大きく変わる場合において、身体動揺の個人性を安定に捉える手法の開発が挙げられる。また、身体動揺の観測時間を短くした場合でも高精度に人物を対応付ける手法の開

発が必要になると考えられる。

## 謝 辞

本研究の一部は，総務省 SCOPE(No. 172308003) および JSPS 科研費 JP17K00238 の助成を受けたものである。

## 文 献

- (1) W. Xiaogang. Intelligent multi-camera video surveillance: A review. *Pattern Recognition Letters*, Vol. 34, No. 1, pp. 3–19, 2013.
- (2) R. Min, J. Choi, G. Medioni, and J.L. Dugelay. Real-time 3d face identification from a depth camera. *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 1739–1742, 2012.
- (3) S.P. Banerjee and D.L. Woodard. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, Vol. 7, No. 1, pp. 116–139, 2012.
- (4) K. Delac and M. Grgic. A survey of biometric recognition methods. *Proceedings of 46th International Symposium on Electronics in Marine*, pp. 184–193, 2004.
- (5) M.O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user authentication on mobile phones using biometric gait recognition. *Proceedings of sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 306–311, 2010.
- (6) T. Kamitani, H. Yoshimura, M. Nishiyama, and Y. Iwai. Temporal and spatial analysis of local body sway movements for the identification of people. *IEICE Transactions on Information and Systems*, Vol. 102, No. 1, pp. 165–174, 2019.
- (7) J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322, 2006.
- (8) Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. *Proceedings of 9th European Conference on Computer Vision*, pp. 151–163, 2006.
- (9) P. Vera, S. Monjaraz, and J. Salas. Counting pedestrians with a zenithal arrangement of depth cameras. *Machine Vision and Applications*, Vol. 27, No. 2, pp. 303–315, 2016.
- (10) S. Munir, R.S. Arora, C. Hesling, J. Li, J. Francis, C. Shelton, C. Martin, A. Rowe, and M. Berges. Real-time fine grained occupancy estimation using depth sensors on arm embedded platforms. *Proceedings of 2017 IEEE Real-Time and Embedded Technology and Applications Symposium*, pp. 295–306, 2017.
- (11) B.A.Y. Agusta, P. Mitrapriyanuruk, and P. Kaewtrakulpong. Field seeding algorithm for people counting using kinect depth image. *Indian Journal of Science and Technology*, Vol. 9, p. 48, 2016.
- (12) D. Liciotti, M. Paolanti, R. Pietrini, E. Frontoni, and P. Zingaretti. Convolutional networks for semantic heads segmentation using top-view depth data in crowded environment. *Proceedings of 24th International Conference on Pattern Recognition*, pp. 1384–1389, 2018.
- (13) S. Mukherjee, B. Saha, I. Jamal, R. Leclerc, and N. Ray. Anovel framework for automatic passenger counting. *Proceedings of 18th IEEE International Conference on Image Processing*, pp. 2969–2972, 2011.
- (14) C. Hua, Y. Makihara, and Y. Yagi. Pedestrian detection by using a spatio-temporal histogram of oriented gradients. *IEICE Transactions on Information and Systems*, Vol. 96, No. 6, pp. 1376–1386, 2013.
- (15) N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886–893, 2005.
- (16) B. Fernando, E. Gavves, J. Oramas, A. Ghodrati, and T. Tuytelaars. Rank pooling for action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 4, pp. 773–787, 2016.
- (17) A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, Vol. 14, No. 3, pp. 199–222, 2004.
- (18) H. Bilen, B. Fernando, E. Gavves, and A. Vedaldi. Action recognition with dynamic image networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, pp. 2799–2813, 2017.
- (19) J. Tighe, M. Niethammer, and S. Lazebnik. Scene parsing with object instances and occlusion ordering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
- (20) T. Brox, L. Bourdev, S. Maji, and J. Malik. Object segmentation by alignment of poselet activations to image contours. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2225–2232, 2011.
- (21) P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image seg-

mentation. *International Journal of Computer Vision*, Vol. 59, No. 2, pp. 167–181, 2004.

- (22) Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, pp. 122–1239, 2001.
- (23) J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, Vol. 81, No. 1, pp. 2–23, 2009.
- (24) P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in Neural Information Processing Systems*, pp. 109–117, 2011.
- (25) J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
- (26) O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.
- (27) V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp. 2481–2495, 2017.
- (28) P. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, Vol. 15, No. 2, pp. 70–73, 1967.
- (29) K.Q. Weinberger and L.K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, Vol. 10, No. Feb, pp. 207–244, 2009.

神谷卓也（非会員）2018年鳥取大学大学院博士前期課程修了。2019年同大学院博士後期課程在学中。ソフトウェア計測の研究に従事。2017画像の認識・理解シンポジウム学生奨励賞を受賞。

山口優太（非会員）2019年鳥取大学大学院博士前期課程在学中。

西山正志（非会員）2002年岡山大学大学院博士前期課程了。同年株式会社東芝研究開発センターに勤務。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。2015年より現在鳥取大学大学院工学研究科准教授。画像認識，インタラクシオンの研究に従事。2019画像センシングシンポジウム優秀学術賞などを受賞。電子情報通信学会，情報処理学会の会員。

**岩井 儀雄**（非会員） 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）