身体動揺を用いた人物対応付けにおける被り物変化に頑健な時間特徴の抽出 *

神谷卓也 * 中山晴貴 * 西山正志 *

Temporal Feature Extraction Robust to Headwear Variations for Person Identification using Body Sway

Takuya KAMITANI, Haruki NAKAYAMA and Masashi NISHIYAMA

We propose a method for extracting temporal features robust to headwear variations for person identification using the video sequences of body sway. When people put on headwear such as caps and helmets, their head shapes, observed from an overhead camera, change dramatically depending on the type of headwear. The existing method cannot obtain high accuracy of person identification in situations where the head shapes change because their features are directly affected by the headwear variations. We perform a learning-based low-pass filter for the time-series signal of head center positions representing body sway to extract our temporal features robust to the headwear variations. Experimental results show that our temporal features significantly improved the accuracy of person identification when the headwear variations occur, compared to the existing features.

Key words: person identification, headwear variations, body sway, temporal feature extraction, learning-based low-pass filter

1. はじめに

映像中の人物がカメラ間で同じであるかどうかを紐づける人物対応付けの技術 [1, 2, 3] が必要とされている．人物対応付けは，例えば工場やイベント会場など，見守りたい区間の入退出を把握する応用が期待される．精度よく人物を対応付けるためには，人物の個人性を表現する手掛りの選択が重要となる．人物を対応付けるための新たな手掛りとして身体動揺が注目されている．身体動揺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る．

これまでに，身体動揺を用いた人物対応付けの既存手法 [4, 5] が提案されてきた．これらの既存手法では，エレベータ待ちや信号待ちの場面を想定し，立ち止まる人物から身体動揺の映像を，頭上カメラを用いて観測している．観測された身体動揺の映像から，人物の頭部の形状と，その動きを利用して時空間特徴を抽出し，同一人物であるかどうかを確認している．これらの既存手法では，帽子などの被り物を，頭部に着用していないことを制約条件としており，図 **1** のように，人物が頭部に被り物を着用していることを想定していなかった．ただし，人物は場面に応じて適切な被り物を選択し，着用することがある．工場で働く人物であれば，生産ライン作業で帽子を着用することがあり，組み立て作業でヘルメットを着用することがある．また，仮装イベントに参加する人物であれば，イベント会場で仮装をするためにロングヘアウィッグやアフロウィッグを着用することがある．

着用する被り物が変化することによって，同じ人物であっても頭部形状が変化する．その例を図 **2** に示す．図中では，被り物を何も着用しない場合と，それぞれの被り物を着用する場合

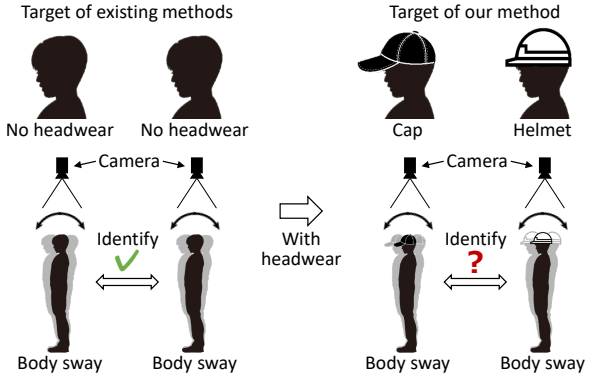

Fig. 1 The target of our person identification using body sway. This paper investigates whether our method can identify people with headwear.

であり，全て同じ人物の頭部である．被り物を着用しない場合と比べて，帽子を着用する場合は，帽子のツバによって頭部形状が楕円に近くなっている．また，ヘルメットを着用する場合は，頭部形状が円に近くなっており，ロングヘアウィッグとアフロウィッグを着用する場合は，頭部形状にウィッグの形そのものが表れている．このように被り物に応じて頭部形状は大きく異なっていく．

身体動揺を用いた人物対応付けにおいて，頭部形状が変化することによって生じる影響について考える．先に述べた既存手法 [4, 5] で抽出される時空間特徴は，頭部形状に依存している．空間的には人物の頭部形状そのものを用いて個人性を捉えており，時間的には人物の頭部形状の時間方向の微小移動を用いて個人性を捉えている．頭部形状が変化する場合は，既存手法の時空間特徴もその影響を受けて変化する．このため，既存の特徴をそのまま用いると，人物対応付けの精度が低下する．頭部形状を表す空間特徴を排除し，頭部形状の変化による影響を抑

* 原稿受付 令和 5 年 4 月 30 日
* 掲載決定 令和 5 年 8 月 31 日
* 鳥取大学大学院工学研究科 (鳥取市湖山町南 4 丁目 101)

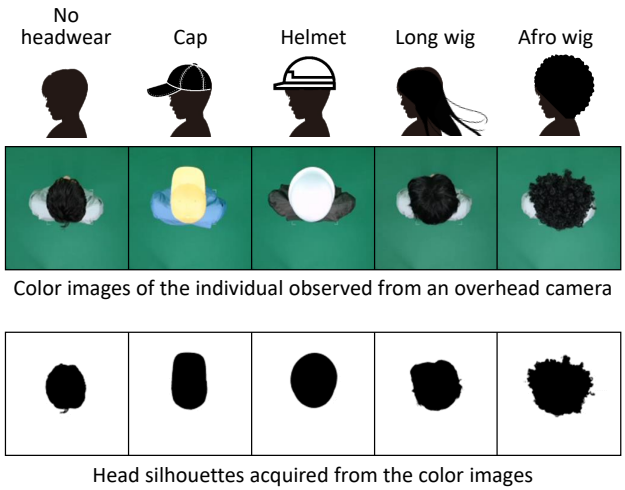

Fig. 2 Examples of the variation in the headwear types. These color images are acquired from the same individual with different types of headwear using an overhead camera. We see that the shapes of the head silhouettes are dynamically changed for each headwear.

える新たな時間特徴の設計が必要となる．

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，被り物変化に頑健な時間特徴を抽出する手法について述べる．提案手法では，カメラで撮影された身体動揺の映像から，頭部の中心位置のみの時系列信号を計測する．被り物変化の影響を受けにくい要素を取り出すため，学習型低域通過フィルタを設計し，頭部中心位置の時系列信号の低周波成分のみを残した時間特徴を抽出する．実験結果から，被り物が変化する場合，提案手法で抽出された時間特徴は，既存手法で抽出された時空間特徴と比べて，人物対応付けの精度を大幅に改善することを確認した．以下では，2. で身体動揺の時間特徴における被り物の変化が与える影響について述べ，人物対応付けの精度向上に向けた課題を明確にする．3. で課題を解決するための提案手法について述べ，その有効性を実験結果で示す．4. で被り物変化に取り組む価値について議論し，最後に，5. でまとめる．

2. 身体動揺の映像から抽出された時間特徴における被り物の影響

2.1 時間特徴の抽出手法の設計

被り物変化の影響を軽減した特徴抽出の設計について議論を進める．まず，身体動揺を用いた人物対応付けの既存手法 [4,5] の特徴抽出について述べる．次に，既存手法が持つ課題について述べる．最後に，その課題解決を目指し改善された既存手法について述べる．

既存手法 [4,5] では，身体動揺を用いた人物対応付けを実現するため，頭上カメラで獲得された映像から時空間特徴を抽出している．その既存手法の流れを図 **3** に示す．個人性を捉えるため，空間特徴の抽出と時間特徴の抽出とが導入されている．空間特徴の抽出では，頭上カメラで獲得された映像から人物の頭部シルエット映像を生成し，その頭部シルエット映像を複数の局所領域に分割することで，人物間における頭部形状の空間的な差異を個人性として表現している．頭上カメラの映像には，頭，肩，背中，腹などの身体部位が含まれているが，その中でも頭部形状だけを獲得することで，自己遮蔽による見え方の変

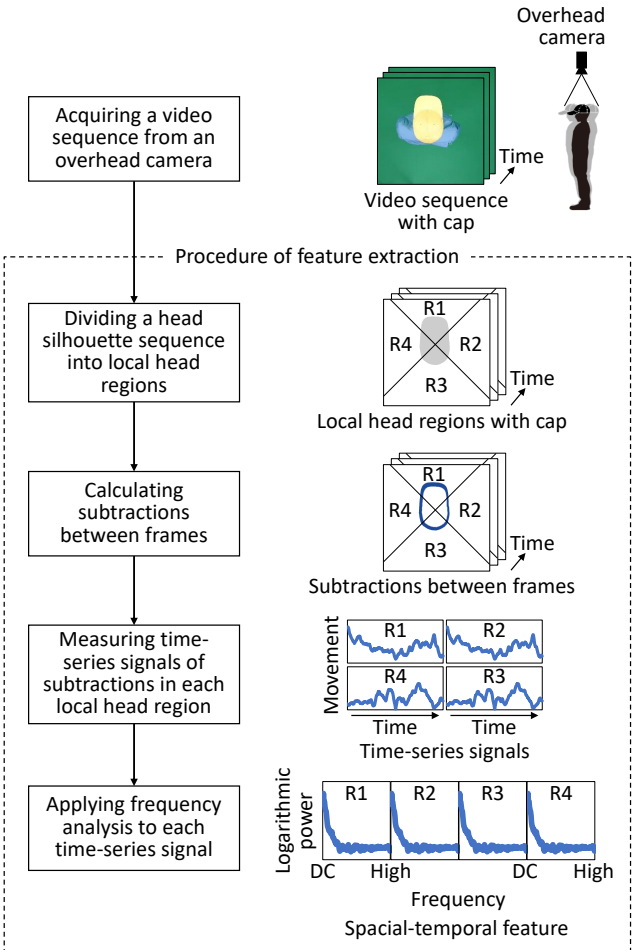

Fig. 3 The procedure of the existing methods [4,5] for extracting a spatio-temporal feature from the video sequence of body sway.

動を抑えていると報告されている [5]．なお，頭部シルエット映像の局所領域は放射状に分割されている．時間特徴の抽出では，分割された頭部シルエット映像の各局所領域において，時間方向のフレーム差分を算出し，頭部形状の微小移動の時系列信号を求めることで，人物間における動きの時間的な差異を個人性として表現している．最後に，各局所領域における微小移動の時系列信号に対して周波数解析を適用することで，人物対応付けに用いる時空間特徴を抽出している．

既存手法 [4,5] の時空間特徴の課題として，1. でも述べたように，頭部形状が変化する場合，特に空間特徴が影響を受けやすいことが挙げられる．このため空間特徴では個人性を捉えることが困難であり，時間特徴のみで個人性を捉えることを考える．ただし，既存手法の考え方で時間特徴のみを抽出したとしても，頭部形状の変化による影響を受けてしまう課題が残る．この既存手法の課題を改善する時間特徴のアイデアとして以下を挙げる．頭部形状の変化による影響を抑えるため，時間方向の微小移動の時系列信号を計測する際，既存手法のように頭部形状そのものを利用するのではなく，頭部の中心位置のみを利用することを考える．

既存手法 [4,5] の課題を改善した時間特徴を抽出する手法の流れを図 **4** に示す．まず，身体動揺を撮影したカメラ映像から，人物の頭部形状を表すシルエット映像を生成する．頭部シル

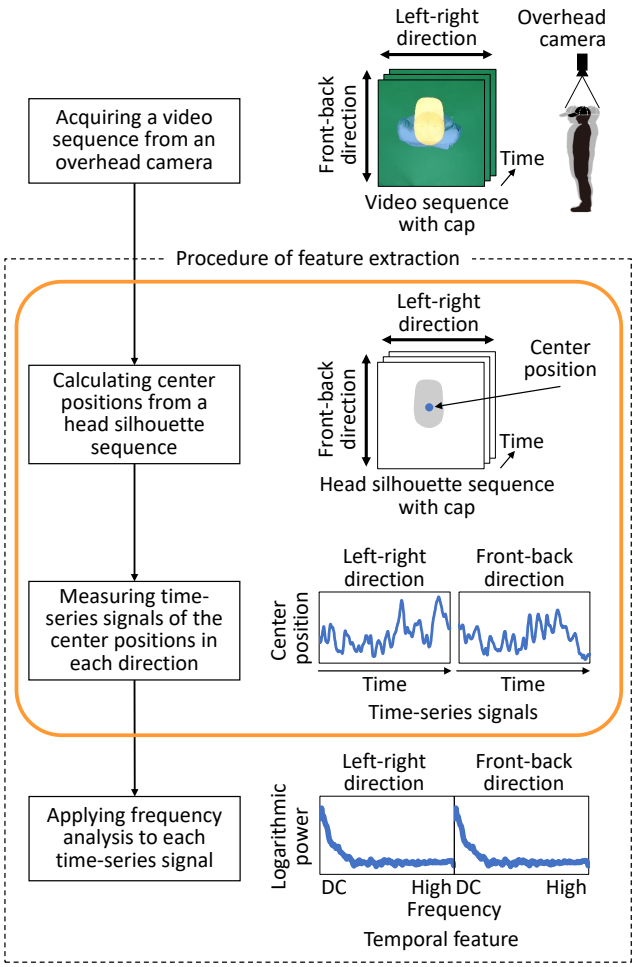

Fig. 4 The procedure of the improved method of the existing methods [4, 5] for extracting a temporal feature from the video sequence of body sway.

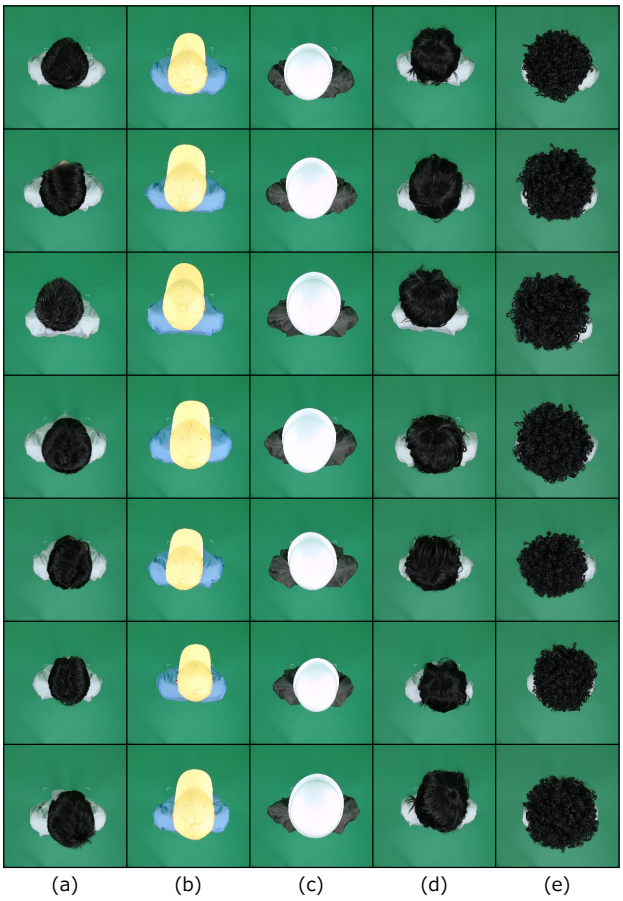

Fig. 5 Examples of the color images of participants when wearing different types of headwear. The participants are observed from an overhead camera. The row direction means the same individuals and the column direction means the same types of headwear. Columns (a), (b), (c), (d), and (e) show the color images with no headwear, a cap, a helmet, a long wig, and an afro wig, respectively.

エット映像の生成には，図 **3** の既存手法と同様に，Deep Image Matting [6] を用いる．本論文では先にも述べたように，人物の頭部の中心位置のみをシルエット映像から抽出することを考える．具体的には，頭部形状を表すシルエット映像の各時刻において，頭部形状を空間方向に見た時の 2 次元重心ベクトルを中心位置として算出する．各時刻において中心位置を求め，時間方向に並べることで頭部の中心位置の時系列信号を求める．以下では，中心位置の時系列信号を，左右方向成分と前後方向成分とでそれぞれ 1 次元の時系列信号として扱う．なお，原点を中心位置の時間方向の平均とする．各方向成分の時系列信号を周波数毎に見るため，既存手法と同様に周波数解析一種であるパワースペクトル密度 (Power spectral density: PSD) [7] を用いて対数パワーを算出する．左右方向成分に関する各周波数の対数パワーと前後方向成分に関する各周波数の対数パワーとを要素としてもつベクトルに結合することで時間特徴を抽出する．以降では，この特徴抽出手法を既存手法改と呼ぶ．ただし，この既存手法改を用いたとしても，被り物が変化した時の人物対応付けの精度を僅かにしか改善できない．以下では，この原因と解決手法について実験的に検証していく．

2.2 評価データセット

被り物が変化する場面において，人物対応付けの精度を評価するため，独自のデータセットを構築した．データセットの構築では，何も被り物を着用していない場合と，4 種類の被り物をそれぞれ着用した場合で身体動揺のカメラ映像を収集した．この評価データセットでは被り物の種類を以下とした．

被り物なし：通勤時など普段の人物を想定し，何も被り物を着用しなかった．各人物の髪型については制限を特に設けなかった．被り物を着用していない人物を撮影した例を図 **5** (a) に示す．また，被り物を着用していない人物を横から見た例を図 **6** (a) に示す．

帽子：小型民生品のライン作業に従事する人物を想定し，ツバ付きの黄色の帽子を選択した．ツバの向きが顔の向きと同じになるように帽子を各人物は着用した．帽子を着用した人物を撮影した例を図 **5** (b) に示す．また，帽子を着用した人物を横から見た例を図 **6** (b) に示す．

ヘルメット：プラント大型装置の組み立て作業に従事する人物を想定し，短いツバ付きの白色のヘルメットを選択した．短いツバの向きが顔の向きと同じになるようにヘルメットを各人物は着用した．ヘルメットを着用した人物を撮影した例を図 **5** (c) に示す．また，ヘルメットを着用した人物を横から見た例を図 **6** (c) に示す．

ロングヘアウィッグ：芝居の舞台に出演する人物を想定し，髪が長いウィッグを選択した．髪が長い部分が頭の後ろにくるよ

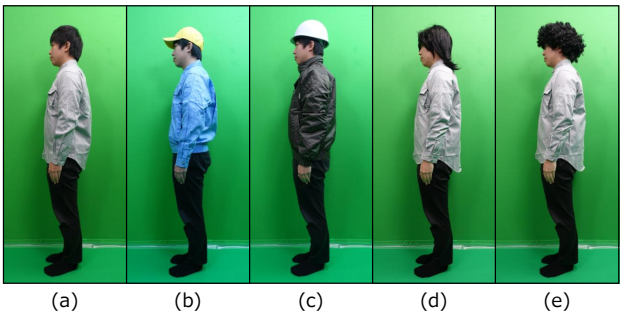

Fig. 6 Examples of the color images of a participant observed from a side-view camera. The participant is the same individual in the first row in Fig. **5**.

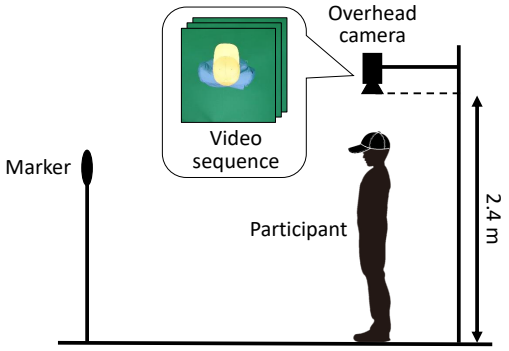

Fig. 7 The experimental setting of acquiring a video sequence of body sway using an overhead camera.

うにウィッグを各人物は着用した．ロングヘアウィッグを着用した人物を撮影した例を図 **5** (d) に示す．また，ロングヘアウィッグを着用した人物を横から見た例を図 **6** (d) に示す．

アフロウィッグ：仮装イベントに参加する人物を想定し，アフロ型のウィッグを選択した．ウィッグの仕様書に前後左右の指定がなかったため，任意の向きでアフロウィッグを各人物は着用した．アフロウィッグを着用した人物を撮影した例を図 **5** (e) に示す．また，アフロウィッグを着用した人物を横から見た例を図 **6** (e) に示す．

実験協力者は 50 名 (平均年齢: 22±1.9 歳) とした．身体動揺の撮影環境を図 **7** に示す．撮影中は，直立姿勢を維持するようにすべての実験協力者に指示した．撮影中に顔の向きが変動しないように，マーカを見るようにすべての実験協力者に指示した．カメラは床からの高さが 2.4 メートルに位置するように設置した．カメラの真下の床面に印したマーカーに，土踏まずを合わせるように実験協力者に立ち位置を指示した．カメラの解像度を 1920 × 1080 画素とし，人物を含む領域を 1000 × 1000 画素でクロッピングした．カメラの対角画角を 78 度，サンプリング周波数を 30 ヘルツとした．各被り物について 2 回ずつ撮影を行い，実験協力者 1 名あたりの撮影回数を計 10 回とした．撮影 1 回あたりの時間を 120 秒とした．各被り物を着用する順番をランダムとした．

2.3 既存手法改で抽出された時間特徴を用いた人物対応付けの精度

被り物が変化した場合において，既存手法改で抽出された時間特徴が，人物対応付けの精度向上に有効であるかを評価した．本論文での人物対応付けとは，入力された人物の動画が，

Table 1 Comparison of the accuracy of person identification using the existing methods [4, 5] and the improved existing method.

Method	n=1	n=5	n=10	n=15	nAUC
Existing method	3.7	14.2	26.4	36.2	56.2
Improved method	5.2	18.8	31.3	43.3	59.7

あらかじめ登録された辞書の中に存在する人物と同じであるか否かを識別することとする．辞書の被り物と入力の被り物について，そのペアの種類を ${}_5\mathrm{P}_2 = 20$ 通りとした．辞書と入力の間で，被り物の種類は異なることとした．個人の識別には時間特徴によるユークリッド距離を用いた最近傍法を適用した．この試行を 5 回繰り返した．人物対応付けに用いた人数を 50 人とした．人物対応付けの精度の評価指標として，n 位正解率，および，Cumulative Match Characteristics (CMC) 曲線の normalised area (nAUC) を用いた．n 位正解率は，入力の人物と距離が近い順に辞書の人物を並べた時，上位 n 番目までに入力と同一の人物が含まれている割合を示している．n 位正解率は，0 から 100 までの値をとり，100 に近いほど精度が良いことを示す．nAUC は，横軸を n 位，縦軸を n 位正解率とした時に描かれるグラフの下部の面積を求めたものであり，0 から 100 までの値をとる．100 に近いほど精度がよいことを示す．

既存手法改と既存手法における n 位正解率と nAUC を表 **1** に示す．実験結果から，微小な改善傾向はみられるものの，既存手法改では人物対応付け精度を大きくは改善できないと言える．次節では，その改善できない理由について考察し，精度を高める可能性があるかどうかについて検討する．

2.4 被り物の変化が時間特徴に与える影響の考察

既存手法改でも人物対応付け精度を改善できない理由を考察する．その理由として，被り物が変化した時，人物対応付けに有効でない周波数の対数パワーが，時間特徴の中に存在するためと考えられる．この対数パワーが存在すると，辞書の時間特徴と入力の時間特徴とが本人同士の場合でも，それら特徴間の距離が大きくなり，本人であるのに本人ではないと誤る試行が増加すると考えられる．また，辞書の時間特徴と入力の時間特徴とが本人と他人である場合，それら特徴間の距離が小さくなり，本人を他人と誤る試行が増加すると考えられる．

考察を進めるため，辞書の時間特徴と入力の時間特徴とについて，本人同士の間で差分を算出した．また同様に本人と他人との間で差分を算出した．具体的に差分を求めるため，時間特徴の要素である周波数毎の対数パワーについて，辞書の値から入力の値を引き，その値を 2 乗した．以下では，辞書と入力とが本人同士の場合を本人差分と呼び，本人と他人の場合を他人差分と呼ぶ．各周波数において，本人差分と他人差分とが類似する場合，対応付けの誤りの原因となる対数パワーが存在する可能性が高いと考えられる．

本人差分と他人差分を算出した結果を図 **8** に示す．図中のオレンジ線は本人差分を表し，紫線は他人差分を表す．中心位置の時系列信号について，左右方向に関する各周波数の対数パワーの差分をとった結果を (a) に示し，前後方向に関する各周波数の対数パワーの差分をとった結果を (b) に示す．図中より，(a) の左右方向と (b) の前後方向とのどちらでも，4.0 ヘルツに近づくに連れて，本人差分と他人差分が急激に類似していくことが分かった．4.0 ヘルツから 15.0 ヘルツまでの高周波

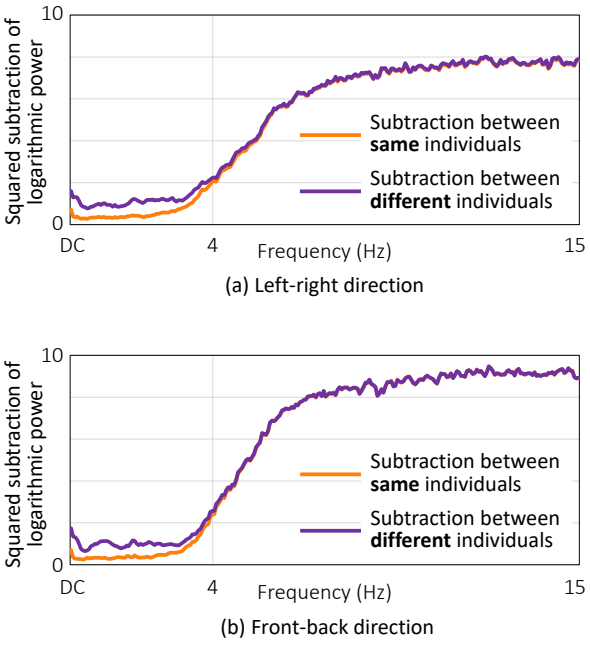

Fig. 8 We calculated the squared subtraction of the logarithmic powers for each frequency. The orange line is the subtraction between the same individuals, and the purple line is the subtraction between the different individuals. The upper graph (a) shows the subtractions in the left-right direction, and the lower one (b) shows those in the front-back direction.

数帯では，左右方向と前後方向のどちらでも，本人差分と他人差分が，非常に類似していることが分かった．一方，直流成分 DC から 4.0 ヘルツ手前までの低周波数帯では，左右方向と前後方向のどちらでも，本人差分は他人差分よりも小さいことが分かった．以上の結果より，既存手法改で精度を改善できない理由として，人物対応付けに有効でない対数パワーが高周波数帯で多く存在することの影響が大きいと考えられる．さらなる考察として低周波数帯についてみていく．本人差分が他人差分よりも小さい低周波数帯では，辞書と入力との間で本人同士の距離が小さくなり，また，辞書と入力との間で本人と他人との距離が大きくなることが考えられる．よって，低周波数帯の対数パワーのみを時間特徴として利用すると，人物対応付け精度を改善できる可能性があると考えられる．ただし，低周波数帯を 4.0 ヘルツ以下に単純に設定してよいかどうかは一概には言えない．実際に 4.0 ヘルツに近づくにつれ，本人差分と他人差分が近づいてきており，辞書と入力との間で，本人同士の距離と，本人と他人との距離との境界が曖昧になるため，精度改善の幅が小さくなる場合があると考えられる．このため，人物対応付けに有効な低周波数帯を，手作業で適切に決めることは難しいと考えられる．人物対応付けの精度を改善する時間特徴を抽出するため，低周波数帯を自動で設定する必要があると言える．

3. 提案手法の学習型低域通過フィルタにより抽出された時間特徴

3.1 学習型低域通過フィルタの設計

人物対応付けの精度を改善するため，有効な時間特徴の候補である低周波数帯の対数パワーを抽出する学習型低域通過フィルタを設計する．このフィルタにおいて，以下で説明する閾値をどのように決めるかが重要になる．閾値とは，直流成分 DC からどの周波数までの対数パワーを，前後方向と左右方向のそれぞれにおいて，低周波数帯として通過させるかを制御するパラメータである．人物を精度良く対応付けることができる閾値を決めるため，本人同士の時間特徴を近づけ，本人と他人との時間特徴を遠ざけることを考える．その際，2.4 で述べたように，4.0 ヘルツ付近では，本人同士の特徴間の距離と，本人と他人との特徴間の距離の境界が曖昧である．本論文では，適切な閾値を自動で決定するため，分離度 [8] の考え方を利用する．分離度の値が大きいほど，本人同士の特徴は近く，本人と他人との特徴が離れていることを意味する．

被り物の変化に頑健な分離度を計算するため，複数種類の被り物をそれぞれ着用している人物の時間特徴を予め準備する必要がある．ただし，辞書として登録されている人物から，複数種類の被り物の時間特徴を獲得できるとは限らない．本論文では，複数種類の被り物を着用している人物の時間特徴で構成される訓練セットを用意する．この訓練セットにおいて，分離度を最も高める低域通過フィルタの閾値を全探索で決定する．分離度が最大の時，本人同士の時間特徴が最も近く，本人と他人との時間特徴が最も離れている状態となる．全探索を行う時の閾値の範囲として，左右方向と前後方向とのそれぞれで，直流成分 DC から，観測できる最大周波数までの間を考える．以下で分離度の具体的な算出方法を述べる．分離度は，クラス間分散を分子とし，クラス内分散を分母とすることで計算される．クラス間分散とは，本人の時間特徴と他人の時間特徴との間での距離分散であり，クラス内分散とは，本人同士の時間特徴の間での距離分散である．なお，これらの距離は2.4 で述べた他人差分と本人差分そのものから算出される．本論文では，分離度が最大となる閾値で帯域が指定されるものを，学習型低域通過フィルタと呼ぶことにする．提案手法では，学習型低域通過フィルタで時間特徴を抽出することで，被り物変化に頑健な人物対応付けを行うことを狙う．

学習型低域通過フィルタで時間特徴を抽出する提案手法の流れを図 **9** に示す．図中の，映像の取得から，周波数解析までの流れは図 **4** と同じである．提案手法では，周波数解析の一種である PSD によって算出された全帯域の対数パワーに対して，学習型低域通過フィルタを適用し閾値以下の帯域からなる対数パワーのみを選択することで，低周波数帯のみの時間特徴を抽出する．

3.2 提案手法で抽出された時間特徴を用いた人物対応付けの精度評価

3.2.1 基本性能

被り物が変化した場合において，提案手法である学習型低域通過フィルタを用いて抽出された時間特徴が，人物対応付けの精度改善に有効であるかを評価した．提案手法では，フィルタ閾値を探索する必要があるため，2.2 で述べた実験協力者 50 人の内，ランダムに選択された 10 人を訓練セットに，残りの 40 人を人物対応付けの辞書と入力とに用いた．2.2 で述べた 5 種類の被り物をそれぞれ着用した人物の時間特徴を訓練セットに含むこととした．学習型低域フィルタにおいて，閾値を探索する際の候補を選ぶ時の刻み幅を 0.05 ヘルツとした．辞書と入力とのペアにおいて，どの被り物の組み合わせであっても閾値を同じとした．訓練セットとして，ランダムに 10 人を選択する試行を 5 回行い，それぞれの試行毎に人物対応付け精度を算

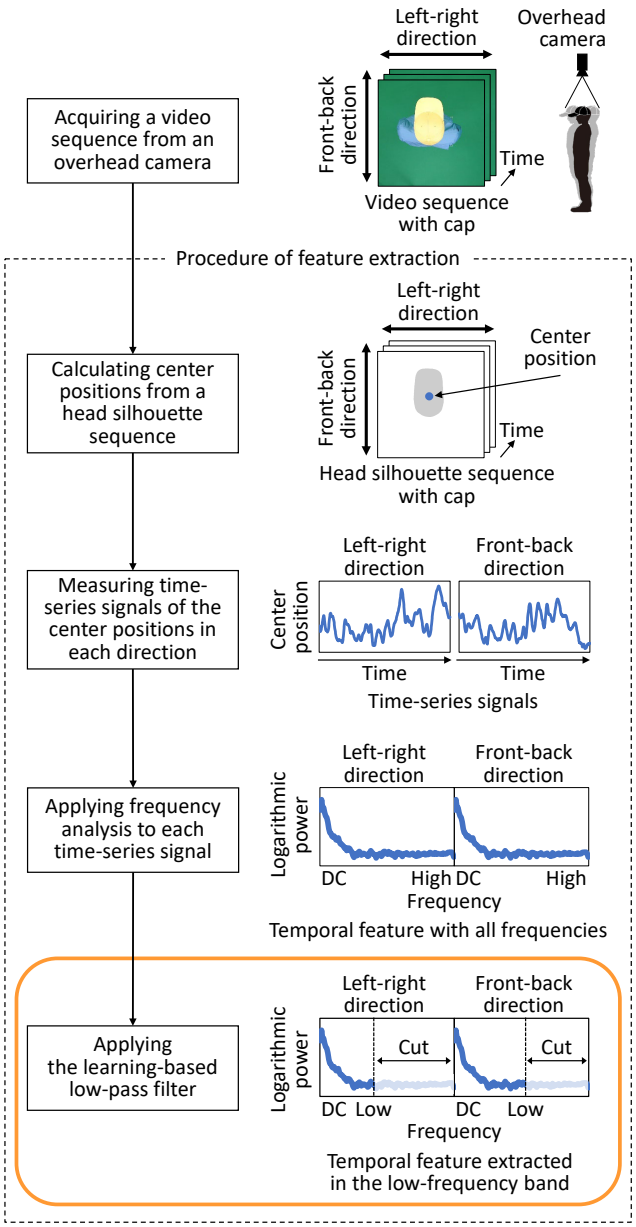

Fig. 9 The procedure of our method for extracting a temporal feature in a low-frequency band using the learning-based low-pass filter from the video sequence of body sway.

出した．その他の人物対応付けの実験条件を2.3 と同じとした．人物対応付けの精度の評価指標として n 位正解率と nAUC とを用いた．

提案手法，既存手法改，既存手法により抽出された時間特徴を用いて人物対応付けを行った時の n 位正解率と nAUC とを表 **2** に示す．なお，提案手法とは3.1 で述べたものを指し，既存手法改および既存手法とは2.1 で述べたものを指す．実験結果から，提案手法で抽出された時間特徴は，既存手法改または既存手法で抽出された時間特徴に比べて，人物対応付けの精度を高めることを確認した．このため，提案手法の学習型低域通過フィルタを適用して時間特徴を抽出することは，被り物が変化した場合の身体動揺を用いた人物対応付けに有効であると言える．

Table 2 Comparison of the accuracy of person identification using our method, the improved existing method and the existing methods [4, 5] when the headwear variations occur.

Method	n=1	n=5	n=10	n=15	nAUC
Our method	**44.2**	**78.6**	**89.2**	**94.2**	**92.1**
Improved method	6.3	21.9	38.1	51.0	60.1
Existing method	4.2	17.2	31.9	44.6	56.5

Table 3 Comparison of the accuracy of person identification using our method, the improved existing method and the existing methods [4, 5] when not putting on headwear.

Method	n=1	n=5	n=10	n=15	nAUC
Our method	62.5	92.5	98.0	98.5	97.1
Improved method	77.0	90.5	93.5	96.0	95.7
Existing method	98.5	100.0	100.0	100.0	100.0

3.2.2 被り物なしの場合の精度評価

身体動揺を用いた人物対応付けにおいて，被り物を頭部に着用しない場合の精度を評価した．ここでは評価データセットとして2.2 で述べた“被り物なし”のみを用いた．その他の人物対応付けの実験条件を3.2.1 と同じとした．

被り物なしの場合において，提案手法，既存手法改，既存手法を用いて人物対応付けを行った時の n 位正解率と nAUC とを表 **3** に示す．実験結果から，既存手法の精度は，被り物なしの場合において優れていることを確認した．一方で，既存手法の精度は，被り物変化がある場合において，3.2.1 の表 **2** で示した通り，大幅に低下していた．これらの結果より，既存手法は，被り物なしの場合に最適化されていることが示唆される．提案手法の精度は，既存手法の精度や既存手法改の精度と比べて，表 **3** の被り物なしの場合において低いものの，表 **2** の被り物変化がある場合において高くなっていた．これらの結果より，提案手法は，被り物変化がある場合のみに有効であることが示唆される．提案手法を運用する際，被り物がある場合と被り物がない場合とで，既存手法と適切に切り替えていく工夫を組み込む必要がある．今後の課題として，映像中の人物が被り物を着用しているかどうかを検出する機能を前処理として加えることが考えられる．

3.2.3 低域通過フィルタの閾値を手動決定した場合との比較

提案手法では，人物対応付けに有効な時間特徴を抽出するため，学習型低域通過フィルタの閾値を分離度に基づいて自動で決定した．ここでは，提案手法を用いて自動設定された閾値が，人物対応付けの精度にどの程度有効であるかを評価した．具体的には，一定の刻み幅で閾値を手動で設定し，人物対応付けの精度をそれぞれ算出した．手動設定の閾値の変更範囲を1.0 ヘルツから 5.0 ヘルツまでとし，刻み幅を 1.0 ヘルツとした．その他の人物対応付けの実験条件を3.2.1 と同じとした．

閾値を手動設定した場合と提案手法で自動設定した場合とにおいて，身体動揺を用いて人物対応付けを行った時の 1 位正解率を図 **10** に示す．実験結果から，3.0 ヘルツに手動設定した場合が提案手法の自動設定より精度が 4.0 ポイントほど高くなることが分かった．提案手法の精度は最高とは言えないが，それに近い値をとっていると考える．人物対応付けの精度を評価す

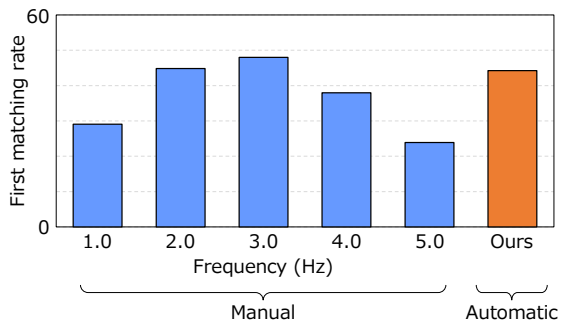

Fig. 10 The first matching rate (%) when the threshold is set manually or automatically.

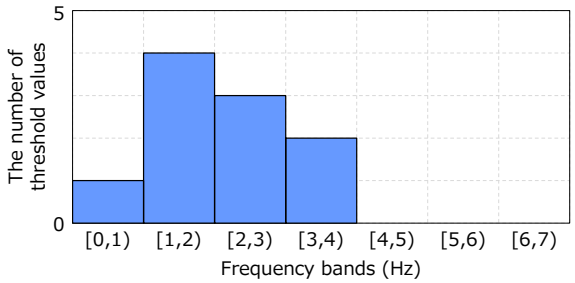

Fig. 11 The histogram of threshold values of the low-pass filter with our automatic setting. The threshold values are voted for the bins of the appropriate frequency bands.

るデータセットが事前に既知であれば，手動設定で適切な閾値を決めた上で人物対応付けを行うことができると言える．データセットが既知でなく人物対応付けの精度を事前に評価できない場合において，提案手法の自動設定は，一定の精度が得られる閾値を決めることができる可能性があると考える．

提案手法によって自動決定された閾値の分布を確認するため，図 **11** のヒストグラムを求めた．横軸は 1.0 ヘルツの間隔で区切られた周波数帯のビンを表し，縦軸は該当する周波数帯のビンに投票された閾値の数を表す．閾値の自動決定で用いる訓練セットにランダム性があるため，人物対応付けを 5 回試行し，左右方向と前後方向とで決定された計 10 個 (5 試行 $\times$ 2 方向) の閾値を用いた．図中から，提案手法で自動決定された閾値は，直流成分 DC から 4.0 ヘルツまでの低周波数帯に含まれる傾向にあることが分かった．特に，$[1, 2)$ ヘルツのビンへ投票された閾値の数が多かった．

3.2.4 高周波数帯から抽出された時間特徴との比較

前節までの実験では，被り物が変化した場合において，学習型低域通過フィルタを用いて低周波数帯から対数パワーを時間特徴として取り出すことで，人物対応付けの精度を改善できることを示した．ここでは低周波数帯ではなく，逆に高周波数帯から対数パワーを取り出した時について，人物対応付けの精度を評価した．比較する周波数帯として，低周波数帯だけの場合，高周波数帯だけの場合，および，全ての周波数帯の場合を準備した．その他の人物対応付けの実験条件を3.2.1 と同じとした．

低周波数帯，高周波数帯，および，全周波数帯から抽出された時間特徴を用いて，人物対応付けを行った時の n 位正解率と nAUC とを表 **4** に示す．実験結果から，時間特徴を抽出する際，低周波数帯だけを用いた場合は，他の周波数帯を用いた場合に比べて，人物対応付けの精度を高めることを確認した．

Table 4 Comparison of the accuracy of person identification using the temporal feature extracted from the low-frequency band, only high-frequency band, and all-frequency band.

Frequency band	n=1	n=5	n=10	n=15	nAUC
Low (ours)	**44.2**	**78.6**	**89.2**	**94.2**	**92.1**
High	4.5	17.1	31.5	44.6	55.6
All (low+high)	6.3	21.9	38.1	51.0	60.1

このため，時間特徴を抽出する周波数帯として，低周波数帯だけを用いることは有効であると言える．

3.2.5 体の動きを対象とする既存手法との比較

提案手法では，人物の体の動きの一種である身体動揺を対象とし，人物対応付けに有効な時間特徴を抽出した．体の動きを対象とした特徴抽出の手法は，提案手法以外にも多く提案されている．その例として，歩容認識の既存手法 [9] や行動認識の既存手法 [10, 11, 12] が提案されており，本論文の問題設定にも適用できると考えられる．ここでは身体動揺の映像に対して，それらの既存手法を適用し特徴を抽出することで，人物対応付けの精度を比較した．特徴抽出以外の実験条件を 3.2.1 と同じにした．前節までの実験と同様に，各既存手法を用いて抽出された特徴に対して最近傍法を適用することで人物対応付けを行った．既存手法で抽出された特徴の詳細を以下に述べる．

Gate energy image (GEI) [9]: 2.1 で述べた 120 秒間の頭部のシルエット映像から 8.5 秒の短区間をランダムに切り取ることで，GEI を映像 1 個あたり 15 個生成した．それぞれの GEI を生成するため，切り取られた短区間の中で時間方向に画素値を平均した．ここでは人物対応付けの精度を高めるため，ResNet101 [13] に GEI を入力し中間層の特徴マップを取り出すことで，人物対応付けのための特徴を抽出した．なお，ImageNet-1K で事前学習された ResNet101 の重みに，3.2.1 で述べた訓練セットから生成された計 1500 枚の GEI を用いてファインチューニングを行った．人物対応付けでは，映像 1 個あたり 15 個生成された特徴の平均を用いて距離が最も近い人物を出力した．

Dynamic image (DI) [10]: 上記の GEI と同様に，ランダムに短区間を切り取り DI を生成した．それぞれの DI を生成するため，短区間に属するカラー映像の中で Rank SVM を適用した．さらに上記の GEI と同様に，ResNet101 に DI を入力し中間層の特徴マップを取り出した．訓練セットから生成された計 1500 枚の DI を用いて ResNet101 のファインチューニングを行った．

C3D [11]: 上記の GEI と同様に，ランダムに短区間を切り取りネットワークに入力することで，C3D の中間層から特徴を抽出した．メモリ使用量を抑えるため，短区間に属するカラー映像のフレームの中から 16 枚を等間隔で選択した．C3D への入力の配列サイズを $16 \times 3 \times 256 \times 256$ とした．ここでは，3.2.1 で述べた訓練セットから生成された計 1500 個の短区間のカラー映像で C3D のネットワークを学習した．ネットワーク構造として，3 次元畳み込み層 4 個，3 次元プーリング層 4 個とした．

TimeSformer [12]: 上記の GEI と同様に，ランダムに短区間を切り取りネットワークに入力し，TimeSformer の中間層から特徴を抽出した．メモリ使用量を抑えるため，短区間のフレームの中から 8 枚を等間隔で選択した．TimeSformer への入力の配列サイズを $8 \times 3 \times 224 \times 224$ とした．ここでは Divided

Table 5 Comparison of the accuracy of person identification using our method, the existing method of gait recognition [9], and the existing methods of action recognition [10, 11, 12].

Method	n=1	n=5	n=10	n=15	nAUC
Ours	**44.2**	**78.6**	**89.2**	**94.2**	**92.1**
GEI [9]	8.0	30.2	53.3	69.9	72.4
DI [10]	15.8	45.9	67.5	81.4	80.4
C3D [11]	16.2	47.8	68.1	79.7	79.8
TimeSformer [12]	10.1	33.4	52.1	64.7	70.1

Table 6 Comparison of the accuracy of person identification using our method, the existing methods [13, 14] using color image feature extractions.

Method	n=1	n=5	n=10	n=15	nAUC
Ours	**44.2**	**78.6**	**89.2**	**94.2**	**92.1**
ResNet101 (e)	4.1	17.0	31.1	44.1	55.7
ResNet101 (c)	4.0	17.1	31.2	43.9	55.7
EfficientNet-B7 (e)	3.5	15.5	30.7	42.6	54.7
EfficientNet-B7 (c)	3.5	15.3	30.1	42.8	54.7

Space-Time Attention を用いた．なお，Kinetics-400 で事前学習されたネットワークの重みに，3.2.1 で述べた訓練セットから生成された計 1500 個の短区間のカラー映像を用いてファインチューニングを行った．

提案手法および各既存手法で抽出された特徴を用いて人物対応付けを行った時の n 位正解率と nAUC とを表 **5** に示す．実験結果から，提案手法で抽出された特徴は，既存手法で抽出された特徴に比べて，人物対応付けの精度をより高めることを確認した．この実験においても，被り物が変化する場合の人物対応付けにおいて，提案手法は有効であることが分かった．

3.2.6 静止画から特徴を抽出する手法との比較

本論文で対象とする身体動揺は非常に微細な動きである．このため，3.2.5 で述べた既存手法 [9, 10, 11, 12] では，人物の大きな動きを捉えるよう設計されているため，身体動揺を用いた人物対応付けに有効な特徴を抽出することが難しい可能性がある．そこで，映像ではなく静止画から特徴を抽出する ResNet101 [13] と EfficientNet-B7 [14] とを用いて，人物対応付けの精度を評価した．映像の各フレームを静止画として ResNet101 に入力し，それぞれの静止画について中間層の特徴マップを取り出すことで，人物対応付けのための特徴を抽出した．EfficientNet-B7 についても ResNet101 と同様の手順で特徴を抽出した．なお，ResNet101 と EfficientNet-B7 では，ImageNet-1K で事前学習済みのモデルを採用した．また，特徴がどれだけ似ているかを測るため，ユークリッド距離とコサイン類似度とを用いた．これら以外の実験条件を3.2.5 と同じにした．

提案手法，ResNet101，および，EfficientNet-B7 を用いて人物対応付けを行った時の n 位正解率と nAUC とを表 **6** に示す．表中の記号 e はユークリッド距離を用いた場合，c はコサイン類似度を用いた場合を指す．実験結果から，提案手法は，ResNet101 および EfficientNet-B7 を用いた場合に比べて，人物対応付けの精度を向上させることを確認した．提案手法は，立ち止まった人物の微細な動きから個人性を捉え，人物対応付けに有効な時間特徴を抽出していると考えられる．

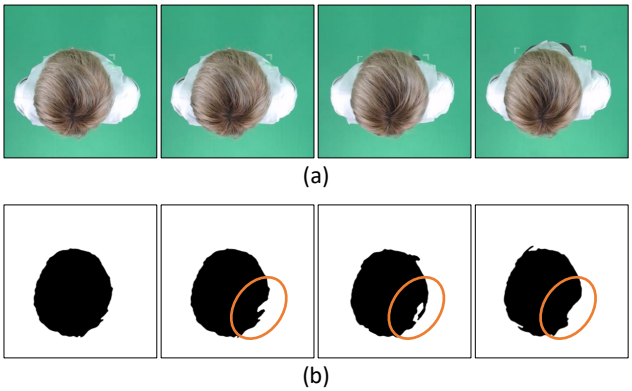

Fig. 12 Examples of the head silhouettes generated from a person with blond hair.

3.2.7 頭部シルエットの生成精度に関する考察

本論文では，頭部シルエットは高い精度で生成されていると仮定し，人物対応付けの精度を評価した．具体的には2.2 の評価データセットで述べたように，頭部シルエットを安定に生成できるよう全人物は黒髪である統制条件を設けた．ただし，実際には人物のヘアカラーは多様である．本論文で用いた評価データセットには含まれていない人物であるが，例えば図 **12**(a) のように金髪の人物も存在した．この人物から頭部シルエットを2.1 で述べた手法 [6] で生成した場合，図 **12**(b) で示す生成誤差が発生した．図中において，最も左の画像では，正常に頭部シルエットを生成できていたが，残り 3 つの画像では，頭部シルエットの一部に欠けが発生し正常に生成できていなかった．同じ欠けが常に繰り返し発生する訳ではなく，様々な欠けが時間方向に一時的に発生する傾向が見られた．このような頭部の欠けは，提案手法の時間特徴における全周波数帯の成分に影響を及ぼすと考えられる．高周波成分への影響は，提案手法の低域通過フィルタによって軽減されると期待できるが，低周波成分への影響は，そのまま強く残ると考えられる．これらの影響は再現性を保証できないため，人物対応付けの精度を大きく低下させる可能性があると考えられる．提案手法を運用する場合，ヘアカラーや髪質など頭部の見え方の変動要因を考慮しつつ，頭部シルエットの生成精度にも注意を払う必要がある．

4. 被り物変化に取り組む価値

身体動揺を用いた人物対応付けの技術を発展させるため，本論文で取り組んだ被り物変化の他にも多様な課題に取り組む必要がある．ここでは応用場面を想定した上で，それらの課題を列挙していく．本論文では1. で述べたように，エレベータ待ちや信号待ちなど，人物が立ち止まる場面で，それらの人物を撮影し対応付けることを想定している．人物対応付けの精度を低下させる課題として，大きく分けて以下の二つがあると考えられる．

撮影環境の課題: 人物対応付けの応用場面では，必ずしも同じ環境で人物を毎回撮影できるとは限らない．例えば，エレベータホールの天井や信号機の支柱に取り付けられた防犯カメラを活用するとき，それらのカメラは多様な角度や高さで設置されていると想定される．また，エレベータ待ちや信号待ちの人物が周囲に存在すれば，自身のパーソナルスペースを保つため，立ち位置が変化していくと想定

される．さらに，信号や階数や周囲を確認するため，顔を様々な方へ向けていくことも考えられる．カメラの向きや人物の立ち位置や顔の向きによる頭部の見え方の変化など，撮影環境の影響が人物対応付け誤りの原因となると想定される．

人物状態の課題: 履物や髪型や所持物など人物の状態が身体動揺に影響を与える課題について考慮する必要がある．本論文では髪型が頭部シルエットの形状に与える影響について議論したが，履物により身体動揺そのものが変化する可能性についても考慮する必要がある．例えば，ある人物が通勤時にサンダルを着用した状態で撮影され，その人物が勤務時に作業靴を着用した状態で撮影されたとする．この場合は履物の違いにより，同じ人物であったとしても身体動揺が変化し，人物対応付け誤りが発生することが考えられる．さらに，鞄やリュックを所持する場合においても，同様の課題が発生することが考えられる．履物や髪型や所持物などの人物状態の影響が身体動揺に表れることが，人物対応付け誤りの原因となると想定される．

身体動揺を用いた人物対応付けにおいて，上記のいくつかの課題で発生する影響を我々の先行研究[15, 16, 17]で調査してきた．先行研究[15]では，カメラ俯角が変化することで，観測される頭部シルエットの時系列信号も変化し，人物対応付けの精度が低下することを報告した．また，先行研究[16]では，靴の踵が高くなることで身体動揺そのものが変化し，人物対応付けの精度が低下することを報告した．人物対応付けを目的としていないが，先行研究[17]では所持物の重さを身体動揺から識別する手法を提案した．所持物の重さにより身体動揺に変化が表れることから，人物対応付けにおいて所持物は精度を低下させる課題になると考えられる．これらの課題は全体の一部であり，撮影環境の課題と人物状態の課題とについて，さらに調査することで，取り組むべき課題を明確化していく必要がある．

本論文では，身体動揺を用いた人物対応付けにおいて，人物状態の課題の一つである被り物変化に取り組み，その対処方法を示した．特に被り物が変化した場合，一般的な深層学習で用いられることが多いカメラ映像中の見た目に基づく特徴のみでは，3.2.5 と3.2.6 とで示した通り，人物対応付けを行うことは困難である．本論文では，頭部の見え方変化の影響を抑えつつ，身体動揺の時間特徴のみで，人物対応付けの精度がどの程度改善されるかを明らかにしたことに価値があると考える．

5. まとめ

本論文では，身体動揺の映像を用いた人物対応付けにおいて，人物の被り物が変化した場合に頑健な時間特徴を抽出する手法について述べた．映像中の人物頭部の形状は，被り物が異なると同一人物であっても大きく変化するため，その頭部形状の中心位置を用いることにした．被り物の変化から影響を受けにくい時間特徴を実験的に探ることで，頭部中心の時系列信号の低周波数帯に属する対数パワーが有効であることを明らかにした．この低周波数帯の対数パワーを頭部中心位置の時系列信号から適切に捉えるため，学習型低域通過フィルタを用いた時間特徴の抽出手法を設計した．実験結果から，提案手法の学習型低域通過フィルタは，被り物が変化した場合の人物対応付けに有効な低周波数帯の対数パワーを自動で決定でき，身体動揺を用いた人物対応付けの精度改善に有効であることを確認した．また，提案手法は行動認識の既存手法と比べて，被り物が変化した場合の人物対応付けにおいて精度を大幅に改善することを確認した．

今後の課題について以下で述べる．被り物が変化した場合でも身体動揺の特徴は本来であれば変化しないと考えられるため，身体動揺の本質を捉えることで人物対応付けの精度を更に向上させる手法の設計が挙げられる．また，被り物の種類にさらに多様性をもたせた実験が挙げられる．

謝辞

本研究を進めるにあたり，貴重なご助言やご意見をくださいました鳥取大学工学部教授の岩井儀雄先生に感謝の意を表する．

参 考 文 献

1) A. Bedagkar-Gala and S. K. Shah. A survey of approaches and trends in person re-identification. *Image and vision computing*, Vol. 32, No. 4, pp. 270–286, 2014.
2) L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 1116–1124, 2015.
3) L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian. Person re-identification in the wild. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1367–1376, 2017.
4) 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.
5) T. Kamitani, Y. Yamaguchi, M. Nishiyama, and Y. Iwai. Identifying people using body sway in case of self-occlusion. *In Proceedings of International Workshop on Frontiers of Computer Vision*, pp. 1–13, 2020.
6) N. Xu, B. Price, S. Cohen, and T. Huang. Deep image matting. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2970–2979, 2017.
7) 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.
8) N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, Vol. 9, No. 1, pp. 62–66, 1979.
9) 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.
10) 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.
11) D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 4489–4497, 2015.
12) G. Bertasius, H. Wang, and L. Torresani. Is space-time attention all you need for video understanding? *In Proceedings of the International Conference on Machine Learning*, 2021.
13) K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
14) M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *In Proceedings of the International Conference on Machine Learning*, pp. 6105–6114, 2019.
15) 中山晴貴, 西山正志, 岩井儀雄. 身体動揺を用いた人物対応付けにおけるカメラ俯角が認識精度に与える影響の評価. 画像の認識・理解シンポジウム, I32-21, 2021.
16) H. Nakayama, T. Kamitani, M. Nishiyama, and Y. Iwai. Effect of shoe insoles on body sway in video-based person identification. *In Proceedings of IEEE 4th Global Conference on Life Sciences and Technologies*, pp. 293–296, 2022.
17) 山口優太, 中山晴貴, 神谷卓也, 西山正志, 岩井儀雄, 櫛田大輔. 身体動揺の時系列深度画像における時間手掛りを用いた手荷物の軽重認識. 精密工学会誌, Vol. 88, No. 1, pp. 91–101, 2022.