

## 混雑環境下における人物追跡のための歩容特徴に基づく動線の対応付け

杉村 大輔<sup>†</sup> 木谷 クリス真実<sup>††</sup> 岡部 孝弘<sup>†</sup> 佐藤 洋一<sup>†</sup> 杉本 晃宏<sup>†††</sup>

<sup>†</sup> 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

<sup>††</sup> 電気通信大学大学院情報システム学研究所 〒182-8585 東京都調布市調布ヶ丘 1-5-1

<sup>†††</sup> 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: <sup>†</sup>{sugimura,takahiro,ysato}@iis.u-tokyo.ac.jp, <sup>††</sup>kitani@is.uec.ac.jp, <sup>†††</sup>sugimoto@nii.ac.jp

あらまし 本稿では、混雑環境下における安定した人物追跡を実現するための分断化される動線の対応付け手法を提案する。遮蔽に伴い、人物の動線は複数の動線片に分断されるため、一貫した人物追跡を実現するためには、分断化された動線片を対応付ける枠組みが必要となる。ここでは、歩容特徴を用いることにより、動線の対応付けを実現する。周波数領域における歩容特徴は、人物固有の性質であり、個人認証において利用されている指標である。人物に属する特徴点の動きの軌跡群から歩容特徴を抽出し、データ集合間の類似度を測る pyramid match kernel を用いることにより、対応付けるべき動線片間の類似度を計算する。計算される動線片間の類似度を用い、最尤推定を行うことにより、最適な人物動線の対応付けが実現される。実環境における実験により本手法の有効性を確認した。

キーワード 人物追跡, 動線の対応付け, 歩容特徴, 特徴点軌跡, Pyramid Match Kernel, 混雑環境

## Associating Trajectories using Gait Features for Tracking People in Crowds

Daisuke SUGIMURA<sup>†</sup>, Kris M. KITANI<sup>††</sup>, Takahiro OKABE<sup>†</sup>, Yoichi SATO<sup>†</sup>, and Akihiro SUGIMOTO<sup>†††</sup>

<sup>†</sup> Institute of Industrial Science, The University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan

<sup>††</sup> Graduate School of Information Systems, The University of Electro-Communications 1-5-1 Choufugaoka, Chofu-shi, Tokyo, 182-8505, Japan

<sup>†††</sup> National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan

E-mail: <sup>†</sup>{sugimura,takahiro,ysato}@iis.u-tokyo.ac.jp, <sup>††</sup>kitani@is.uec.ac.jp, <sup>†††</sup>sugimoto@nii.ac.jp

**Abstract** We propose a method for associating fragmented trajectories of a moving person in crowds. Due to the frequent partial occlusions, it is inevitable that the entire trajectory of a person are fragmented into multiple short tracks (tracklets). In order to achieve tracking individuals consistently over a scene, we exploit gait features, which is a person's unique spatio-temporal trait, for computing the affinity between tracklets. Gait features in the frequency domain have been shown to be an effective cue for identifying individuals. Acquiring the gait features from a set of motion trajectories of feature points that belong to the same individual, we measure the similarity between the tracklets via the pyramid match kernel algorithm. Under a maximum likelihood estimation, optimal tracklet associations can be searched even in crowds. Our experiments show that the use of gait features effectively work for associating tracklets.

**Key words** tracking people, tracklet association, gait features, feature trajectories, Pyramid Match Kernel, crowded scene

### 1. はじめに

近年、混雑環境下における人物追跡について様々な研究が行われている。朝夕のラッシュ時の駅構内やイベント会場などの混雑状況において、事故や事件の未然防止、人数の自動計測など様々な分野への応用が期待される。

混雑環境において頑健な人物追跡を実現するためには、様々な問題に対処する必要がある。なかでも障害物や他の人物により引き起こされる追跡対象の遮蔽は、人

物追跡において避けることができない問題である。遮蔽に対する追跡の頑健性を高めるために、人物の形状をモデル化する手法 [1], [2] や、局所的な見えを利用する手法 [3], [4] などが提案されている。また近年では特徴点軌跡のクラスタリングによる追跡手法 [5] ~ [8] や、公共スペースにおける人間の行動特性を利用した手法 [9] などが提案されており、その発展は目覚ましい。

しかしながら、遮蔽による追跡失敗を完全に防ぐことはできないため、視野内を移動する人物の動線は、いく

つかの動線片に分断されることが考えられる．そのため、視野内において一貫した人物追跡を実現するためには、分断された人物の動線片を対応付ける枠組みを確立する必要がある．

近年、検出器による検出結果、もしくは追跡器により得られる動線片を適切に対応付けることにより、観測視野内において一貫した人物追跡を実現する手法が報告されている [10] ~ [16]．これらの手法では、追跡対象の見え、位置、動きなどの情報から動線片間の類似度を計算し、最小コストフロー問題やハンガリアン法などの最適化アルゴリズムに基づき最適な動線の対応付けが実現されている．

このような最適化アルゴリズムが効果的に働くことが報告されている一方で、動線片間の類似度計算は、見えや動きに基づく比較的単純な指標を用いる場合が多い．混雑環境下において頻繁に発生する遮蔽により追跡対象の観測が不安定となるため、このような場合、見え、動きに基づいた単純な指標だけでは対処が難しいことが考えられる．このような混雑環境下において頑健な人物動線の対応付けを実現するためには、個々の人物が持つ固有の性質を活用することが有効であると考えられる．

これに対し本手法では、歩容特徴を用いた人物動線の対応付けを実現する．周波数領域における歩容特徴は、歩幅や歩行周期により特徴づけられる人物固有の性質であり、個人認証のための手掛かりとして生体認証分野で広く利用されている [20]．混雑環境下において互いに近接した複数の人物が同じ方向に同じ速度で歩いている場合においても、各人物の歩幅や歩行周期は異なるため、その周波数特徴には各人固有の性質が現われる．このことから、混雑環境下における個々の人物の動線の対応付けにおいて、周波数領域における歩容特徴は効果的な指標になることが期待される．

移動する人物から歩容特徴を獲得するために、本手法では特徴点軌跡のクラスタリングに基づく人物追跡手法の枠組みを利用する [5] ~ [8]．これらの手法は、同じ人物に属する特徴点は類似した動きをするという仮定に基づき、特徴点軌跡群を動きの類似性や位置関係などを手掛かりにクラスタリングすることにより、人物追跡を実現する方法である．このような追跡手法を利用することで、得られる人物の動線片は属する特徴点軌跡群により表現される．これにより、歩容特徴は、属する特徴点軌跡群の動きを周波数解析することにより獲得することができる．

周波数領域における歩容特徴は、属する特徴点軌跡それぞれについて計算される．これは、人物の動線片は歩容特徴のデータ集合を持つことを意味する．このような歩容特徴のデータ集合を利用することにより、動線片間の類似度計算は、多次元データ集合間の類似度を測る問題と見なすことができる．このような問題に対し、本手法では特徴量集合間の類似度を測る pyramid match

kernel (PMK) [17] を適用することでその類似度を計算する．PMK は、ヒストグラム交差に基づきそのデータ集合間のマッチングをとるため、データサイズが異なる集合間においても類似度を測ることができる．そのため、PMK を動線片間類似度の計算に応用することは、個数が不定である特徴点軌跡群を用いる本手法に適している．

このようにして計算される動線片間類似度を用い、最尤推定の枠組みのもとハンガリアン法を適用することで最適な対応付けを実現する．これにより、混雑環境下において一貫した人物追跡を実現することが可能となる．

本論文の構成は以下の通りである．続く第 2. 節において本手法に関連する動線の対応付けに基づく既存手法について概観し、本手法の位置づけおよび新規性について整理する．続く第 3. 節では、最適な人物動線の対応付けはマルコフ連鎖モデルを適用することにより最尤推定として捉えることができることを示し、第 4. 節において動線片間の類似度計算の詳細について述べる．第 6. 節では実映像を用いた実験結果について報告し、最後に第 7. 節においてまとめる．

## 2. 関連研究

検出器による人物検出結果、もしくは追跡器により得られる人物動線片を対応付けることにより人物追跡を実現する手法が報告されている．その多くは、最適な動線の対応付け探索を MAP 推定として定式化し、既存の最適化アルゴリズムを活用することにより有効な結果を得ている．

Kaucic らは、追跡対象の位置、速度、見えに基づき計算される動線間の類似度に加え、シーン認識により得られる遮蔽マップを事前知識として利用することで、MAP 推定により観測視野における複数人物追跡を実現している [11]．具体的には、MAP 推定をハンガリアン法により計算することで、最適な組み合わせの探索を実現している．また Kaucic らの枠組みを拡張し、複数の追跡対象が相互作用するような状況を取り扱うことで、頑健な対象の追跡を実現する手法が報告されている [14]．これ以外にも、対応付けるべき動線片の生成を高度化した手法もまた提案されている [15]．

ハンガリアン法のみならず、様々な最適化手法が動線片の対応付けについて応用されている．Leibe らは、個々の人物検出と人物の動線の推定（検出結果の対応付け）を quadratic boolean optimization の問題として捉えることにより、観測シーン全体の人物追跡を実現している [12]．彼らは、最適化の計算コストの問題から、適切な仮定に基づき EM アルゴリズムのような繰り返し演算によりその問題を解いている．これ以外にも、最小コストフロー問題として対応付けの問題を捉えた動線の対応付けに基づく追跡手法 [16] などが提案されている．

実環境においてより頑健な動線の対応付けを実現するために、Huang らはハンガリアン法を用いた階層的な動

線対応付け手法を提案している [10] . 具体的には, 隣接フレーム間における人物検出結果の対応付け (Low level), low level において得られた短い動線片集合を用いたハンガリアン法による Greedy な対応付け (Mid. level), シーン認識に基づく知識を組み入れた mid. level における対応付け結果の補正 (High level), の三段階の処理を行うことで実現している. このような Huang らの枠組みのもと, Li らは識別器を動線の対応付けに活用することで, 混雑環境下における人物追跡を実現している [13] . 具体的には, 動線の対応付けを, 対応付けるべき動線候補の序列付けと識別の二つの問題として考慮し, 学習した Rankboost 識別器と Adaboost 識別器を組み合わせることで動線の対応付けを実現している .

このように, 検出結果, 動線片の対応付けに基づく追跡手法の多くは, ハンガリアン法などの最適化アルゴリズムを有効に組み入れるアプローチに基づいている. その一方で, 動線片間の類似度計算は, 見えや動きに基づく比較的単純な指標を用いる場合が多く, 観測が不安定となる混雑環境下において有効に働かないことが懸念される. これに対し本手法では, 動線片間の類似度計算に人物固有の性質である歩容特徴を活用する. これにより, 混雑環境下においても安定な動線の対応付けが実現できる .

歩容特徴を利用した先行研究として, 歩容特徴を特徴点軌跡のクラスタリングの考え方に適用した人物追跡手法 [8] が報告されている. [8] の手法は, 歩容特徴を個々の人物を追跡するための手掛かりとして利用しているが, 本提案手法では, 分断化された人物動線片を対応付けるための指標として歩容特徴を利用しており, この点が大きな違いである .

### 3. 最尤推定に基づく人物動線の対応付け

追跡器により得られる人物の動線片の集合から, 観測領域における全ての人物の動きを表現するような最適な動線の組み合わせを探索する. 具体的には, 動線片の組み合わせをマルコフ連鎖で表現し, その対応付けの尤度を最大にする組み合わせを最適解とする. 大域最適解の探索には莫大な計算コストを要するため, 本手法では二部グラフの最大重みマッチング手法の一つであるハンガリアン法に基づいた Greedy な方法 [10] により対応付けを実現する .

得られる動線片の集合を  $\mathcal{T} = \{T_i\}$  とする. この動線片集合について最適な対応付けを行うことにより, 視野内を移動する全ての人物の一貫した動きを表現するような動線の集合  $\mathcal{S} = \{S_k\}$  が得られるものとする. ここで  $S_k = \{T_{k_0}, T_{k_1}, \dots, T_{k_{l_k}}\}$  である. これは, 対応付けられた  $k$  番目の動線  $S_k$  は,  $l_k$  個の動線片により構成されていることを意味する. 対応付けられた最適な動線の集合  $\mathcal{S}^*$  は, 動線集合  $\mathcal{S}$  の尤度  $P(\mathcal{T}|\mathcal{S})$  を最大化する組み合わせと見なすことができる. 一つの動線片  $T_i$  はいず

れか一つの人物動線  $S_k$  に属すること, また個々の人物の動きは独立に観測できることから, 最適な動線片の対応付けの探索問題は次式のような最尤推定の問題に帰着される .

$$\begin{aligned} \mathcal{S}^* &= \operatorname{argmax}_{\mathcal{S}} P(\mathcal{T}|\mathcal{S}) \\ &= \operatorname{argmax}_{\mathcal{S}} \prod_{S_k \in \mathcal{S}} P(S_k = \{T_{k_0}, T_{k_1}, \dots, T_{k_{l_k}}\}) \end{aligned} \quad (1)$$

更に, 尤度  $P(S_k)$  をマルコフ連鎖によりモデル化する .

$$\begin{aligned} P(S_k) &= P_{\text{init}}(T_{k_0}) P_{\text{link}}(T_{k_1}|T_{k_0}) \\ &\quad \dots P_{\text{link}}(T_{k_{l_k}}|T_{k_{l_k-1}}) P_{\text{term}}(T_{k_{l_k}}). \end{aligned} \quad (2)$$

ここで  $P_{\text{init}}(T_{k_0})$  と  $P_{\text{term}}(T_{k_{l_k}})$  は, 動線片  $T_{k_0}, T_{k_{l_k}}$  が属する動線  $S_k$  においてそれぞれ最初と最後の動線片となる尤度を意味し,  $P_{\text{link}}(T_{k_{n+1}}|T_{k_n})$  は, 動線片  $T_{k_n}$  の終端が  $T_{k_{n+1}}$  の先端へ連結される尤度を意味する .

(2) 式を (1) 式に代入し, 対数をとることで次式を得る .

$$\begin{aligned} \mathcal{S}^* &= \operatorname{argmax}_{\mathcal{S}} \sum_{S_k \in \mathcal{S}} [\ln P_{\text{init}}(T_{k_0}) + \ln P_{\text{link}}(T_{k_1}|T_{k_0}) + \\ &\quad \dots + \ln P_{\text{link}}(T_{k_{l_k}}|T_{k_{l_k-1}}) + \ln P_{\text{term}}(T_{k_{l_k}})]. \end{aligned} \quad (3)$$

このようにして得られた (3) 式に対してハンガリアン法を適用することで [10], 最適な対応付けを実現する .

安定した動線片の対応付けの実現のためには, 二つの動線片間の連結尤度  $P_{\text{link}}(T_{k_{n+1}}|T_{k_n})$  の計算が重要な役割を果たす. 本手法では, この二つの動線片間の連結尤度の計算に人物の歩容特徴を利用する .

## 4. 人物動線片間の連結尤度の計算

動線片間の連結尤度  $P_{\text{link}}(T_j|T_i)$  は, 人物に属する特徴点の動きの軌跡群により得られる性質に基づき, 次式のように計算される .

$$\begin{aligned} P_{\text{link}}(T_j|T_i) &= P_g(T_j|T_i) \cdot P_a(T_j|T_i) \cdot \\ &\quad P_m(T_j|T_i) \cdot P_c(T_j|T_i). \end{aligned} \quad (4)$$

ここで (4) 式右辺は各指標に基づき計算される類似度を表し, 歩容特徴 ( $P_g(T_j|T_i)$ ), 局所的見え ( $P_a(T_j|T_i)$ ), 動き ( $P_m(T_j|T_i)$ ), 時間関係 ( $P_c(T_j|T_i)$ ) を用いて計算される. 具体的には, pyramid match kernel [18] によりその類似度を計算する .

### 4.1 Pyramid Match Kernel

Pyramid match kernel (PMK) は, 多次元データ集合間のマッチング (i.e., 類似度) を高速に求める手法であり, 主に画像間のマッチング, 一般物体認識などの研究分野において頻繁に利用されている. 具体的には, 特徴量の集合を, 複数の分解能を持つ (ピン幅が異なる) ヒストグラムピラミッドへ写像し, それぞれのレベルにおいて

ヒストグラム交差を計算する．データ集合間の類似度  $\mathcal{C}(\cdot)$  は，各レベルにおけるヒストグラム交差の重み付き和により次式により与えられる．

$$\mathcal{C}(\Psi(\mathbf{X}), \Psi(\mathbf{Y})) = \sum_{i=0}^L w_i [\mathcal{I}(H_i(\mathbf{X}), H_i(\mathbf{Y})) - \mathcal{I}(H_{i-1}(\mathbf{X}), H_{i-1}(\mathbf{Y}))] \cdot (5)$$

ここで， $\mathbf{X}, \mathbf{Y}$  は比較すべきデータ集合を， $\Psi(\mathbf{X})$  はデータ集合  $\mathbf{X}$  のヒストグラムピラミッドを表し， $H_i(\mathbf{X})$  ( $i \geq 0$ ) はレベル  $i$  におけるデータ集合  $\mathbf{X}$  のヒストグラムを表す．また  $H_{-1}(\mathbf{X})$  は，それぞれの特徴量ベクトルが唯一いずれかのピンに射影されるような，十分小さいピン幅を持つヒストグラムを意味する． $w_i$  はレベル  $i$  における重みを表し， $w_i = 2^{-i}$  で与えられる． $\mathcal{I}(H_i(\mathbf{X}), H_i(\mathbf{Y}))$  は，データ集合  $\mathbf{X}, \mathbf{Y}$  間のレベル  $i$  におけるヒストグラム交差の値である．

PMK は，高速にデータ集合間の類似度を計算できることだけでなく，類似度を測る二つの集合の大きさが異なる場合においても利用することができる．これは，属する特徴点軌跡の数が不定である本手法に適した類似度計測手法であると言える．

しかしながら，PMK の利用は，高次元のデータ集合間の類似度計算には不適當であることが報告されている [18], [19]．本手法において取り扱うデータサンプルは，最大でおよそ百次元程度と高次元になるため，正しい類似度計算が行われないことが懸念される．このような問題に対処するために，本手法では PMK の拡張である Vocabulary-guided PMK (VG-PMK) を利用する [18]．VG-PMK は，データ集合に対し階層的  $k$ -means クラスタリングを行うため，データ分布に応じた適応的なヒストグラムピラミッドを設計することが可能となる．これにより，より効率的かつ高精度なデータ集合間の類似度計算が実現できる．

## 4.2 歩容特徴に基づく類似度

人物の歩容特徴は，歩行に伴う周期的な動きに表れる [8] と仮定し，特徴点の動きの軌跡からその特徴を獲得する．本手法では，人物に属する特徴点軌跡群それぞれに対して主成分分析を適用することにより，歩容に相当する周期成分を求める．

動線片  $T_i$  を持つ人物に属する  $m$  番目の特徴点の動きの軌跡  $\text{tr}_i^{(m)}$  を，位置，時間の三次元データ  $(x, y, t)$  の集合と見なし，主成分分析を適用する．本手法では，第一主成分の主成分ベクトル  $\text{tr}_i^{(m),P1}$  は，変位の大きい人物の並進運動にあたり，第二主成分が持つ主成分ベクトル  $\text{tr}_i^{(m),P2}$  に歩行に伴う周期成分が表れると考え，以後の類似度計算において後者を利用する．図 1 に例を示す．

得られる歩容特徴に相当する特徴点の動きの周期成分 (第二主成分ベクトル)  $\text{tr}_i^{(m),P2}$  に対し，高速フーリエ変換 (FFT) を行うことにより周波数特性を獲得する．FFT

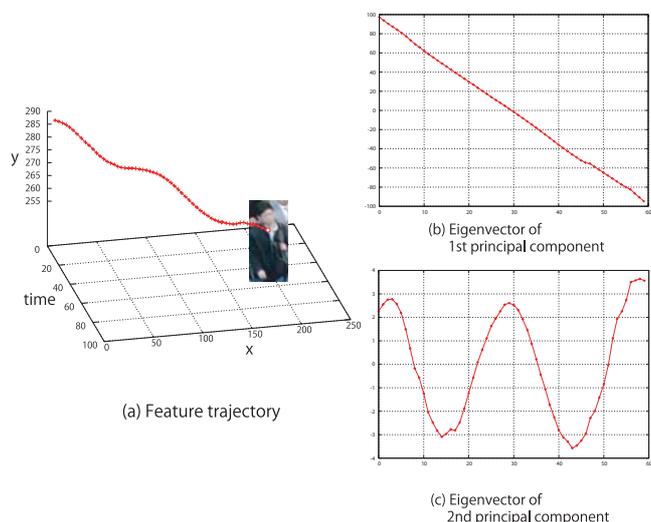


図 1 主成分分析による周期成分の抽出．(a) 特徴点軌跡．(b) (a) の特徴点軌跡について主成分分析を行うことにより得られた，第一主成分ベクトル  $\text{tr}_i^{(m),P1}$ ．(c) 第二主成分ベクトル  $\text{tr}_i^{(m),P2}$ ．この例からわかるように，第一主成分ベクトル  $\text{tr}_i^{(m),P1}$  は (a) の特徴点軌跡の並進成分に相当しており，第二主成分ベクトル  $\text{tr}_i^{(m),P2}$  は周期成分に相当していることが見て取れる．

に用いる窓サイズ  $w$  は，人間の歩行周期はおおよそ 1 秒程度である [20] ことに基づき，映像のフレームレートに応じて決定される．FFT により得られる周波数特性は，人物に属する特徴点軌跡群それぞれについて計算される．

動線片  $T_i$  に属する  $m$  番目の特徴点軌跡の周期成分  $\text{tr}_i^{(m),P2}$  について FFT を適用することにより得られる，周波数領域における振幅特性を  $\mathbf{A}_i^{(m)} = \{a_i^{(m),0}, \dots, a_i^{(m),w}\}$  ( $a_i^{(m),n}$  は  $n$  番目の周波数帯における振幅成分) とし，全特徴点軌跡に関する集合を  $\mathbf{A}_i = \{\mathbf{A}_i^{(m)}\}$  とする．これらを用い，歩容特徴に基づく類似度  $P_g(T_j|T_i)$  は VG-PMK により次式のように計算される．

$$P_g(T_j|T_i) = \mathcal{C}(\Psi_{\text{amp}}(\mathbf{A}_j), \Psi_{\text{amp}}(\mathbf{A}_i)) \cdot (6)$$

ここで  $\mathcal{C}(\mathbf{P}, \mathbf{Q})$  は，VG-PMK により計算されたデータ集合間の類似度を表す．

## 4.3 動きに基づく類似度

二つの動線片  $T_i, T_j$  それぞれの位置，速度を手掛かりに動線片間の類似度を測る．動線片  $T_i$  の終端位置を  $\mathbf{x}_i^{\text{tail}}$ ，終端位置における速度を  $\mathbf{v}_i^{\text{tail}}$  とし，対応付ける候補であるもう一つの動線片  $T_j$  の始端位置を  $\mathbf{x}_j^{\text{head}}$ ，速度を  $\mathbf{v}_j^{\text{head}}$  とする．また  $T_i$  の終端位置における時刻を  $t_i^{\text{tail}}$ ， $T_j$  の始端位置における時刻を  $t_j^{\text{head}}$  とし，その時間差を  $\Delta t$  とする．人物は等速直線運動すると仮定することにより，それぞれの時刻  $t_i^{\text{tail}}, t_j^{\text{head}}$  において，動線片  $T_i, T_j$  が互いにどの位置にいる (いた) のかを推定することができる (図 2)．推定される動線片の位置と実際の位置のペア，すなわち， $\tilde{\mathbf{x}}_i \equiv \mathbf{x}_i^{\text{tail}} + \mathbf{v}_i^{\text{tail}} \Delta t$  と  $\mathbf{x}_j^{\text{head}}$ ，

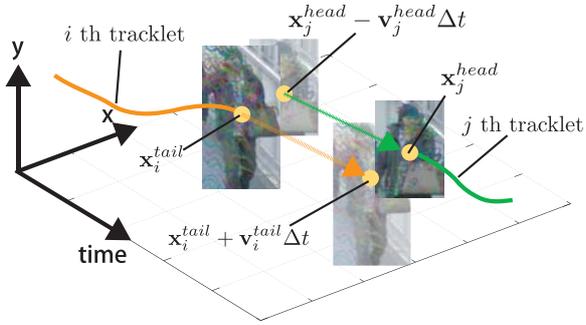


図2 動きに基づく類似度計算．推定される動線片と実際の動線片の位置のペア  $\tilde{x}_i \equiv x_i^{\text{tail}} + v_i^{\text{tail}} \Delta t$  と  $x_j^{\text{head}}$ ,  $\tilde{x}_j \equiv x_j^{\text{head}} - v_j^{\text{head}} \Delta t$  と  $x_i^{\text{tail}}$  それぞれの位置において属する特徴点軌跡群を用いて類似度を測る．

$\tilde{x}_j \equiv x_j^{\text{head}} - v_j^{\text{head}} \Delta t$  と  $x_i^{\text{tail}}$  それぞれが互いに近接した関係にあるならば，動線片間の類似度は高いと見なすことができる．特徴点軌跡群の位置を用い，VG-PMKにより類似度は次式のように計算される．

$$P_m(T_j|T_i) = C(\Psi_m(\tilde{\text{tr}}_j), \Psi_m(\text{tr}_i)) \cdot C(\Psi_m(\text{tr}_j), \Psi_m(\tilde{\text{tr}}_i)). \quad (7)$$

ここで  $\text{tr}_i$ ,  $\text{tr}_j$  はそれぞれ，時刻  $t_i^{\text{tail}}$ ,  $t_j^{\text{head}}$  における動線片  $T_i$ ,  $T_j$  がそれぞれ持つ特徴点軌跡群の位置の集合を表す．また  $\tilde{\text{tr}}_j = \{(\text{tr}_j^{(0)} - v_j^{\text{head}} \Delta t), \dots, (\text{tr}_j^{(n_j)} - v_j^{\text{head}} \Delta t)\}$  ( $n_j$  は  $j$  番目の動線片に属する特徴点軌跡の個数) は，動線片  $T_j$  の推定された始端位置  $\tilde{x}_j$  における特徴点軌跡群の位置の集合を表す．同様に， $\tilde{\text{tr}}_i = \{(\text{tr}_i^{(0)} + v_i^{\text{tail}} \Delta t), \dots, (\text{tr}_i^{(n_i)} + v_i^{\text{tail}} \Delta t)\}$  は，動線片  $T_i$  の推定された終端位置  $\tilde{x}_i$  における特徴点軌跡群の位置の集合を意味する．

#### 4.4 局所的見えに基づく類似度

動線片  $T_i$  を持つ人物に属する，時刻  $t = t'$  における  $m$  番目の特徴点軌跡の位置を中心とした矩形局所パッチを考える．この局所パッチにおいて HS (Hue-Saturation) 色ヒストグラムを計算する．ヒストグラムのビン数はそれぞれ，16, 8 とした．すなわち 128 次元のデータを意味する．

人物に属する全ての特徴点軌跡周りについて局所パッチ内色ヒストグラムは計算されるため，時刻  $t'$  における動線片  $T_i$  を持つ人物の見えは，局所パッチ内色ヒストグラム集合  $\mathbf{h}_{i,t'} = \{\mathbf{h}_{i,t'}^{(0)}, \dots, \mathbf{h}_{i,t'}^{(n_i)}\}$  により表現される．ここで  $\mathbf{h}_{i,t'}^{(m)}$  は， $i$  番目の動線片に属する， $m$  番目の特徴点軌跡周りの局所矩形パッチ内における，時刻  $t'$  での色ヒストグラムを意味する．

得られる色ヒストグラム集合を用い，VG-PMKにより類似度  $P_{\text{app}}(T_j|T_i)$  を計算する．動線片  $T_j$  の始端位置における時刻を  $t_j^{\text{head}}$ ,  $T_i$  の終端位置における時刻を  $t_i^{\text{tail}}$  とし，次式により与えられる．

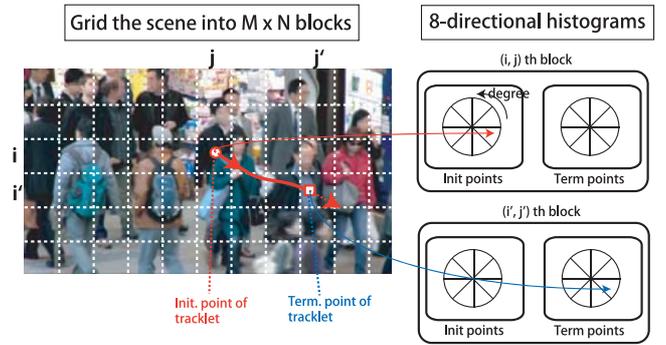


図3 方向ヒストグラムに基づく始端・終端確率分布の学習． $M \times N$  個のブロックに分割した観測視野において， $(i, j)$  のブロックにおいて動線の始端位置が， $(i', j')$  のブロックにおいて終端位置が存在する．このとき，各位置における動線の方向から，各ブロックが持つ始端，終端に関する8方向ヒストグラムが更新される．

$$P_a(T_j|T_i) = C\left(\Phi_{\text{app}}\left(\mathbf{h}_{j,t_j^{\text{head}}}\right), \Phi_{\text{app}}\left(\mathbf{h}_{i,t_i^{\text{tail}}}\right)\right). \quad (8)$$

#### 4.5 時間関係に基づく類似度

動線片  $T_i$  の終端時刻は，対応付けるべき動線  $T_j$  の始端の時刻よりも前でなければならない．また，二つの動線間の時間差が大きくなればなるほど連結しづらいと考えられる．このような対応付けるべき動線片間の始端・終端の時刻の関係に基づき，先行研究 [10], [13] と同様に時間関係に基づく類似度  $P_c(T_j|T_i)$  を計算する．

#### 5. 動線片の始端・終端尤度の計算

マルコフ連鎖によりモデル化された連結される動線の尤度の計算 ((1) 式) において，4. 節において述べた動線片間の連結尤度の計算に加え，始端，終端尤度  $P_{\text{init}}(T_i)$ ,  $P_{\text{term}}(T_i)$  を計算する必要がある．動線片が始端・終端となる確率はそれぞれ，全動線片の始端，終端となった個数に対する，人物の動線を連結した後に実際に始端・終端となる動線片の本数の割合として求められる．本手法では，画像平面上における観測領域を  $M \times N$  のブロックに分割し，各ブロックにおいて得られる動線片集合から確率分布を学習する．

観測視野において複数の出現，消失領域が存在する場合，人物の移動方向に依存して動線の始端，終端確率は変動することが考えられる．このような問題に対し，本手法では方向ヒストグラムを導入する．得られる動線片の始端，終端における方向を考慮することにより，始端，終端位置の方向ヒストグラムが得られる．これにより，分割した各ブロックは，方向ヒストグラムそれぞれのビンにおいて始端，終端に関する確率  $\{P_{\text{init}}^{(0)}, \dots, P_{\text{init}}^{(7)}\}$ ,  $\{P_{\text{term}}^{(0)}, \dots, P_{\text{term}}^{(7)}\}$  を持つことができる．本処理の概要を図3に示す．このように学習された学習された確率分布を用いることで，動線片の方向を考慮した始端・終端尤度  $P_{\text{init}}(T_i)$ ,  $P_{\text{term}}(T_i)$  が計算される．

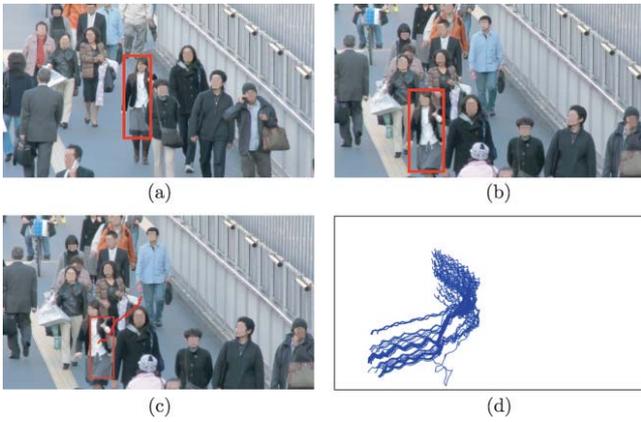


図4 動線片の生成．映像から人物領域を切り出すことにより，人工的に動線片を生成する．動線片の始端，終端それぞれにおいて人物領域を切り出し (a), (b))，両方の領域内に存在する特徴点の追跡結果に基づき，動線片を生成する (c)．図 (d) は追跡対象に属する特徴点の動きの軌跡群の抽出例を表す．

## 6. 実験

本手法の有用性を確認するために，実際の映像を用いて動線の対応付けの実験を行った．実験に用いた環境は，OS: Windows XP, CPU: Intel Core2 Quad 2.66GHz, RAM: 3.00GB である．実験に用いた映像は全部で2つであり，解像度は  $640 \times 360$ ，フレームレートは 60fps のものを用いた．動線の対応付け処理の性能について注目するために，事前に映像から人物領域を切り出し，半自動的に動線片を生成した (図4)．各 VG-PMK のパラメータ (クラスタ数  $k$ ，ピラミッドのレベル数  $L$ ) は次のようにそれぞれ設定した．歩容特徴に基づくものを  $k = 3$ ， $L = 5$ ，局所的見えに基づくものを  $k = 4$ ， $L = 3$ ，そして動き特徴に基づくものを  $k = 2$ ， $L = 2$  とした．これらのパラメータは全ての実験で同じものを用いた．また始端，終端尤度計算に利用する方向ヒストグラムのピン数は4とした．また歩容特徴の利用の性能への寄与を確認するために，歩容特徴に基づく類似度計算を除いたものを基準システムとして用いた．これはすなわち，動線片間の連結尤度を， $P_{\text{link}}(T_j|T_i) = P_a(T_j|T_i) \cdot P_m(T_j|T_i) \cdot P_c(T_j|T_i)$  というように計算することを意味する．

最初に，比較的人物の動きの観測が良好なシーケンスについて実験を行った．図5に用いた映像の例を示す．本実験において生成した動線片の数は11つであり，実際は6人の人物の動線である．

表1に対応付けの結果を真値と併せて示す．これらの結果を比較するとわかるように，歩容特徴を用いた場合には正しく対応付けが行われていることに対し，歩容特徴を用いなかった場合には誤った対応付けを行ってしまっている．これは，図6に示すように，見えが類似した二人の人物が非常に近接しており，見えや動きだけの指標では区別がつかなかったためと考えられる．



図5 シーケンス (1) の一部．図6 歩容特徴を用いない場合の対応付け失敗時の様子．赤丸で囲った2名の人物は互いに近接しており，正しい対応付けが難しい．

表1 シーケンス (1) における動線片対応付け結果．表中の数字は動線片の番号を表しており，矢印の順番で対応づいたことを意味する．

	提案手法	歩容特徴なし	真値
Track 0	0 → 9	0 → 9	0 → 9
Track 1	1 → 2 → 3	1 → 2 → 5	1 → 2 → 3
Track 2	4 → 5	4	4 → 5
Track 3	8 → 6	8 → 6 → 3	8 → 6
Track 4	10	10	10

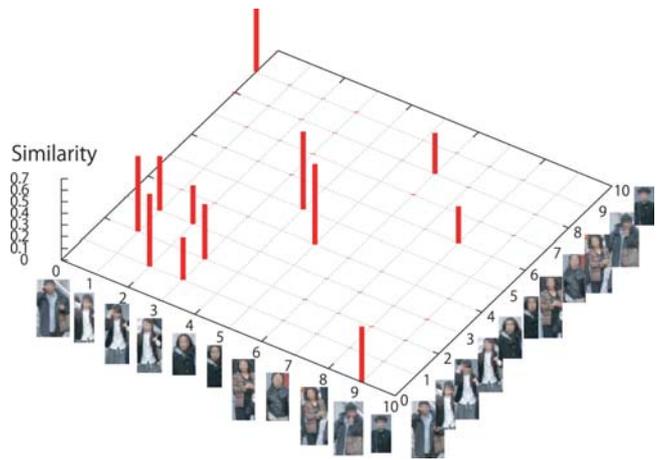


図7 歩容特徴に基づく動線片間の類似度．縦横方向に記される0から10までの数字は動線片番号を意味する．また，本手法では自分自身への対応は許可していないため，自分自身の間の類似度は計算されない．

歩容特徴の性能向上への寄与を確認するために，本実験において得られた歩容特徴に基づく動線片間の類似度  $P_g(T_j|T_i)$  を図7に示す．この結果からわかるように，対応付けるべき動線片間 (表1参照) での類似度は，他の対応付けるべきではない動線片との類似度に比べ明らかに大きな値を示していることが見て取れる．この結果から，歩容特徴は動線片の対応付けにおいて有効に働いているといえる．

次に，動線片の対応付けにおいて困難な状況の一つである，観測視野内に多数の人物が存在する混雑したシーケンスについて実験を行った．図8に用いた映像の一部を示す．本実験において生成した動線片の数は25であり，実際は15人の人物の動線である．



図8 シーケンス(2)の一部。赤丸で示す人物は、はじめ周りの群集と同様の動きをしている(左図(a))が、しばらくすると進路を変更して異なる方向へ移動した(右図(b))。これは、動線の対応付けにおいて難しいケースであると考えられる。

表2 シーケンス(2)における動線片対応付け結果。

	提案手法	歩容特徴なし	真値
Track 0	0 → 1	0 → 1 → 3	0 → 1
Track 1	24 → 17	22 → 23	24 → 17
Track 2	7 → 8	2 → 8	7 → 8
Track 3	20 → 16	5 → 18 → 1	20 → 16
Track 4	10 → 11	12 → 13 → 19	10 → 11
Track 5	18 → 19	4	18 → 19
Track 6	2 → 3	6	2 → 3
Track 7	22 → 23	7 → 21	22 → 23
Track 8	4	24 → 17 → 14 → 15	4
Track 9	5	10 → 20 → 16	5
Track 10	6	9	6
Track 11	9		9
Track 12	12 → 13		12 → 13
Track 13	14 → 15		14 → 15
Track 14	21		21

表2に対応付け結果を真値と併せて示す。この結果からわかるように、混雑状況であるにも拘らず、全ての人物の対応付けが行われていることが見て取れる。

以上の結果から、歩容特徴を用いることにより、混雑環境下においても頑健な動線の対応付けを実現することができると言える。

## 7. おわりに

混雑環境下における人物追跡安定化を目的とした、人物の歩容特徴に基づく動線の対応付け手法を提案した。人物に属する特徴点の動きの軌跡群から歩容特徴を抽出し、特徴量集合間の類似度を測る pyramid match kernel を利用することで、最尤推定のもと最適な動線の対応付けを実現した。実際の群集の映像を用いた実験を通じて、混雑環境において歩容特徴が動線の対応付けに有効に働くことを確認した。

本稿において報告した実験では、動線片の対応付けの性能評価のみを行っている。本対応付け手法を、報告されている人物追跡手法(たとえば[8])と組み合わせることにより、混雑環境下において一貫した人物追跡を実現することができると思われる。

謝辞 本研究の一部は、日本学術振興会特別研究員制度、文部科学省科学研究費補助金(課題番号:20300061)、および科学技術振興機構CRESTの助成により行われた。

- [1] B. Wu and R. Nevatia: "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors", International Journal on Computer Vision (IJCV), **75**, 2, pp. 247–266 (2007).
- [2] T. Zhao and R. Nevatia: "Bayesian human segmentation in crowded situations", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 459–466 (2003).
- [3] X. Song, J. Cui, H. Zha and H. Zhao: "Vision-based multiple interacting targets tracking via on-line supervised learning", Proc. of European Conference on Computer Vision (ECCV), pp. 642–655 (2008).
- [4] P. Tu, T. Sebastian, G. Doretto, N. Krahnstoeber, J. Rittscher and T. Yu: "Unified crowd segmentation", Proc. of European Conference on Computer Vision (ECCV) (2008).
- [5] G. Brostow and R. Cipolla: "Unsupervised bayesian detection of independent motion in crowds", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 594–601 (2006).
- [6] Y. Li and H. Ai: "Fast detection of independent motion in crowds guided by supervised learning", Proc. of IEEE International Conference on Image Processing (ICIP), Vol. 3, pp. 341–344 (2007).
- [7] V. Rabaud and S. Belongie: "Counting crowded moving objects", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 705–711 (2006).
- [8] D. Sugimura, K. M. Kitani, T. Okabe, Y. Sato and A. Sugimoto: "Using individuality to track individuals: Clustering individual trajectories in crowds using local appearance and frequency trait", Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 1467–1474 (2009).
- [9] S. Pellegrini, A. Ess, K. Schindler and L. V. Gool: "You'll never walk alone: Modeling social behavior for multi-target tracking", Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 261–268 (2009).
- [10] C. Huang, B. Wu and R. Nevatia: "Robust object tracking by hierarchical association of detection responses", Proc. of European Conference on Computer Vision (ECCV), Vol. 2, pp. 788–801 (2008).
- [11] R. Kaucic, A. G. A. Perera, G. Brooksby, J. Kaufhold and A. Hoogs: "A unified framework for tracking through occlusions and across sensor gaps", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 990–997 (2005).
- [12] B. Leibe, K. Schindler and L. V. Gool: "Coupled detection and trajectory estimation for multi-object tracking", Proc. of IEEE International Conference on Computer Vision (ICCV), Vol. 1, pp. 1–8 (2007).
- [13] Y. Li, C. Huang and R. Nevatia: "Learning to associate: Hybridboosted multi-target tracker for crowded scene", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 2953–2960 (2009).
- [14] A. G. A. Perera, C. Srinivas, A. Hoogs, G. Brooksby and W. Hu: "Multi-object tracking through simultaneous long occlusions and split-merge conditions", Proc. of CS Conference on Computer Vision and Pattern Recognition (CVPR), pp. 666–673 (2006).
- [15] J. Xing, H. Ai and S. Lao: "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses", Proc. of IEEE CS conference on Computer Vision and Pattern Recognition (CVPR), pp. 1200–1207 (2009).
- [16] L. Zhang, Y. Li and R. Nevatia: "Global data association for multi-object tracking using network flows", Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition

- tion (CVPR), pp. 1–8 (2008).
- [17] K. Grauman and T. Darrell: “The pyramid match kernel: Discriminative classification with sets of image features”, Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 1458–1465 (2005).
  - [18] K. Grauman and T. Darrell: “Approximate correspondences in high dimensions”, Proc. of Advances in Neural Information Processing Systems (NIPS), pp. 505–512 (2006).
  - [19] J. J. Lee: “Libpmk: A pyramid match toolkit”, Technical Report MIT-CSAIL-TR-2008-17, MIT Computer Science and Artificial Intelligence Laboratory (2008).
  - [20] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo and Y. Yagi: “Gait recognition using a view transformation model in the frequency domain”, Proc. of European Conference on Computer Vision (ECCV), Vol. 3, pp. 151–163 (2006).