

混雑シーンからの人領域抽出のための 関節動きに不変な距離を用いた領域分割手法

甲本亜矢子[†] 岩崎正宏[†] 登一生[†]

本稿では、混雑シーンを撮影した動画像から、密な人物領域の抽出が可能な領域分割手法を提案する。従来の動き領域分割法は、混雑シーンにおける人物ごとの領域分割と、密な人物領域の抽出とを、同時に実現することが困難であった。提案手法は、関節体の関係にある2つの小領域間で測地距離が不変となることに着目し、この距離に基づいて領域分割を行う。これにより2つの小領域が、同一人物上の領域か、異なる人物上の領域かを判定できるため、提案手法は、混雑シーンにおける人物ごとの領域分割と、密な人物領域の抽出とを同時に実現可能である。さらに、一定間隔で得た複数の領域分割結果に対して、確率的ラベル伝播法を適用することで、提案手法では長い時間にわたって一貫性を保った領域分割を実現する。実験では、提案手法が、混雑シーンにおいて、人物ごとの領域分割と、密な人物領域の抽出とを同時に実現できることと、遮蔽を多く含むシーンにおいても、時間的に一貫性のある領域分割が出来ることを示す。

An Articulated Motion Invariant Distance for Motion Segmentation in Crowded Scenes

Ayako Komoto[†] Masahiro Iwasaki[†] and Kunio Nobori[†]

This paper introduces an algorithm for dense motion segmentation using an “articulated motion invariant distance” in crowded video sequences. This algorithm solves two conventional problems: one is dense motion segmentation of pedestrians with complex articulated motion; the other is severe occlusions especially in a crowded scene. We focus on the temporal invariance of geodesic distance (similarity) between articulated segments, and utilize it as a criterion for motion segmentation. It can segment the whole appearance of each articulated object, while articulated object contains a wide variety of motions even on the same object. Also, label propagation method is applied to obtain temporally coherent but dense segments from short-term trajectories. The experimental results show that the algorithm can successfully obtain dense and temporally coherent segments under severe occlusion conditions.



(a)入力動画像

(b)領域分割結果

図1:混雑シーンからの人領域抽出の例

1. Introduction

動画像からの人領域抽出は、動画像処理における最も重要な課題の一つであり、さまざまな手法が提案されている[2,3,4,8,11,12,15,17]。これらを大別すると主に2つの手法が挙げられる。一つは人モデルに基づいて人領域を抽出する手法であり、もう一つは動きに基づいて人領域を抽出する手法である。

人モデルに基づいて人領域を抽出する手法[4,11,12]では、人モデルとして、人の見えや構造のモデルを事前に取得しておき、このモデルと動画像を比較して類似度が高い領域を人領域として抽出する。しかし、人の姿勢変化や遮蔽によって見えが大きく変化すると、人モデルとの類似度が低くなり、人領域の抽出が困難になる。特に混雑シーンの動画像では、人同士の重なりによって遮蔽が高い頻度で発生し、見えが大きく変化するため、人モデルに基づく手法では、人領域を抽出することが困難である。

これに対し、動きに基づいて人領域を抽出・追跡する手法が、近年数多く提案されている[2,3,8,15,17]。Brostowら[2]は、Johanssonらの提案した「生物学的運動(biological motion)」[7]の概念を利用し、動画像の動きに基づいて算出した多数の移動軌跡に対して、その位置や動きを基準にペイズの枠組みによってクラスタリングすることで、人領域を抽出・追跡する手法を提案した。この手法は、近接する2人の分離が困難であるという課題があり、これに対し、杉村ら[17]は歩容特徴を用いて、近接する2人の分離を容易にする手法を提案している。これらの、動きに基づいて人領域を抽出・追跡する手法は、人の輪郭や形状、色などの見えに強く依存しないため、混雑シーンにおいても人領域の抽出・追跡が可能であるという利点がある。そこで本研究では、動

[†] パナソニック株式会社
Panasonic Corporation

きに基づいて人領域を抽出する手法により、混雑シーンの動画像から人領域を抽出することを旨とする。

一方、前述の手法は、いずれも追跡時間分の移動軌跡が必要であり、一度も遮蔽されなかった一部の領域しか抽出できないため、密な人物領域の抽出が困難という課題があった。これに対し、遮蔽の影響を低減して領域抽出の密度を向上するには、短時間の移動軌跡を用いるというアプローチがある。しかし、短時間の移動軌跡には、長時間の移動軌跡には含まれなかった、手や足などの部位の移動軌跡が含まれる。これらの手や足などの多関節動きは、同一人物であっても部位によって動きが大きく異なる一方、短時間の移動軌跡では異なる人物間の動きの差異が小さく、両者の区別が困難になるという別の課題があった。

以上のように、動きに基づいて人領域を抽出する従来手法は、人物ごとの領域分割と、密な人物領域の抽出の両立が難しいという課題がある。

その理由としては、以下の2点を挙げられる。

1) 関節動き: 人物は関節動きを伴って移動するため、同一人物であっても部位(手・足・胴体など)ごとに動きが大きく異なる。そのため、同一人物の異なる部位の動きと、異なる人物の動きとを、判別することが困難である。

2) 遮蔽: 領域抽出の密度、領域抽出の精度の間にジレンマの関係がある。すなわち、移動軌跡長(時間)を短くすると、遮蔽の影響が減り、領域抽出の密度は高くなる半面、同一人物か異なる人物かの判定が難しくなる。逆に移動軌跡の時間を長くすると、同一人物か異なる人物かの判定は容易になる反面、領域抽出の密度が低くなる。

上記課題を解決するため、本稿では、以下の2点を特徴とする、動画像から人領域を抽出する動き領域分割法の新たな手法を提案する。

1) 関節動きに対しては、異なる人物の動きと、同一人物上の異なる部位の動きを判別可能な距離を提案する。具体的には、移動軌跡間の類似度の基準として、ユークリッド距離ではなく、測地距離[16,13]に基づく「関節動きに不変な距離」を用いる。「関節動き」は、最も単純には、1つの節(joint)と、2つの剛体がなす動きとしてモデル化できる。節を介する2剛体上の移動軌跡間では、関節動きに伴い、ユークリッド距離は変動するが、測地距離は変化しない。すなわち、「関節動きに不変な距離」が、人物領域の抽出において有効であることを以降で示す。

2) 遮蔽の影響を低減し、密な領域抽出を実現するために、提案手法では、まず、短時間の移動軌跡を用いて、画像を複数の小領域に分割する。そして、確率的ラベル伝播法を導入して、それらの小領域を時間方向に結合することで、長時間にわたり一貫性を保った領域分割を実現する。

以上の手順により、人物ごとの領域分割と、密な人物領域の抽出を両立できる。

本稿は以下の構成より成る。第2章で関連手法について、第3章と第4章で提案手法の概要を述べ、第5章で実験結果を示し、最後に第6章で結論を述べる。

2. 関連研究

混雑シーンに対する動き領域分割手法は、近年数多く提案されている[2,15,17]。

Brostow ら[2]は、ペイズの枠組みによる動き領域分割手法を提案しており、その中で、移動軌跡間類似度に基づくペアワイズなクラスタリング手法を採用している。提案手法でも、移動軌跡を初期統合する過分割(オーバークラスタリング)の部分で、ペアワイズなクラスタリングの枠組みを採用する。

また、杉村ら[17]は、人物ごとの分離に有効な歩容特徴を提案した。歩容特徴の取得には、歩行者の周期的な動きを表せるだけの長時間の観測(例: 30fpsの画像 40-64フレーム)が必要である。長時間の移動軌跡であっても、上半身が遮蔽されないようにカメラが比較的高い位置に設置され、かつ、複雑な動きをする手足は必ず遮蔽されるような混雑シーンの動画像に限れば、人物ごとに分離が可能である。一方、混雑度が低く、手足に長時間の移動軌跡が残る場合や、図1の低いアングルで撮影された画像のように、上半身までも遮蔽される場合は、領域分割が困難となる。

そこで提案法では、動画像から人領域を抽出するため、2つのアプローチをとる。第一に、「関節動きに不変な距離」を採用した動き領域分割を行なうこと、第二に、短時間の移動軌跡に対する領域抽出結果を、確率的ラベル伝播法を導入し、時間方向に結合することで、変形と遮蔽のどちらにも影響されにくく、人物ごとの領域分割と、密な人物領域の抽出を両立可能な動き領域分割の実現を試みる。

3. 関節動きに不変な距離を用いた動き領域分割

提案手法は、動画像中の人領域を分類・抽出するため、まず、移動軌跡全体を小領域に過分割(オーバークラスタリング)したのち、統合する。その際、「関節動きに不変な距離」を、小領域を統合する基準とする。

ここで、 $T+1$ フレームよりなるフレームセット τ を定義する。連続する2フレームセット間では、 T フレームごとに、同じ画像が共有される。(例: τ_f と τ_{f+1})。

提案手法は、図2に示す4ステップの処理で構成される。

まず、密な領域分割([3])を得るため、オプティカルフローに基づく動き検出により密な移動軌跡を生成する(図2(a), 3.1節)。次に、ペアワイズなクラスタリングの枠組みに基づき、移動軌跡を過分割(オーバークラスタリング)し小領域を生成する(図2(b), 3.2節)。次に、「関節動きに不変な距離」に基づく各小領域間の動きの類似度に従って小領域を統合する(図2(c), 3.3節)。最後に、前述の小領域に対し、時間軸上で確率的ラベル伝播法を適用し、各小領域に、全フレームセットを通してそれぞれ一貫

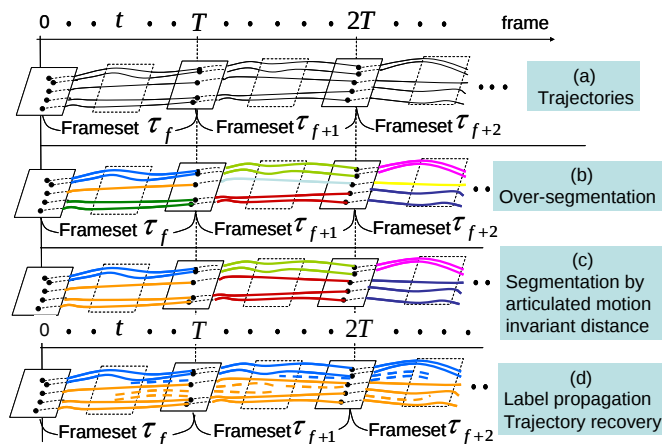


図 2：提案手法の概要（異なる領域を別の色で示す）
(a) 移動軌跡 (b) 過分割 (c) 小領域の統合 (d) ラベル割り当て

したラベルを付与する．なお、破線は、前後のフレームセットとの対応関係に基づいて復元した移動軌跡を示す(図 2(d), 4 章)．

3.1 移動軌跡

密な人物領域を抽出するには、空間的に密な移動軌跡が望ましいため、関連手法 [2, 15, 17] で用いられる KLT トラッカー[20]ではなく、グラフカットベースの動き検出手法[9]を用いて移動軌跡を生成した．移動軌跡は、画素座標を時系列に並べたベクトルとして表され、時刻 t での i 番目の軌跡は N を軌跡数として以下のように定義される．

$$\mathbf{x}_i(t) \in \mathbf{R}^2 \quad (t = 0, \dots, T, \quad i = 1, \dots, N)$$

なお、図 2(a) に示すように、各フレームセット τ から軌跡 $\mathbf{x}$ が生成される．

動き検出の結果、遮蔽等により、規定フレームセット ($T+1$ フレーム) に満たない移動軌跡は除去する．遮蔽の影響を低減し、移動軌跡をより多く残存させるには、 T (例： $T=8$) を、従来手法より小さく設定する必要がある[2, 17]．

3.2 過分割（オーバークラスターリング）

続いて、移動軌跡の過分割（オーバークラスターリング）を行う．各フレームセット

τ に属する移動軌跡を、移動軌跡間の類似度に基づいて分割した例を図 2(b) に示す．移動軌跡間の類似度は、以下に示すグラフ構造で表せる．

$$S_{\tau_f}(\mathbf{x}_i, \mathbf{x}_j) = w \cdot ptn_{ij} + mtn_{ij}, \quad (1)$$

ここで、 ptn_{ij} は移動軌跡間の平均ユークリッド距離、 mtn_{ij} は移動軌跡間の動き成分の

差の平均値、 w は式(1)上で ptn_{ij} と mtn_{ij} の値のバランスを調整する重み係数であり、

$w = 0.1$ とする．続いて、グラフを用いて移動軌跡を過分割する．デンドログラム[5] に似た手法ともいえるが、 $S_{\tau_f}(\mathbf{x}_i, \mathbf{x}_j)$ 値から移動軌跡間のギャップを検出し、領域を分

割する点で異なる．また、類似度の高い移動軌跡を逐次的に統合するデンドログラムとは異なり、本手法は、経験的に定めた領域数（例：100）に到達するまで、ギャップ検出パラメータの値を下げていきながら領域を分割する．密な移動軌跡を利用するため、処理すべき $S_{\tau_f}(\mathbf{x}_i, \mathbf{x}_j)$ の数は巨大になるが、領域数があまり多くない場合には、デンドログラムより高速、かつ同等の結果が得られる．本節の領域分割により得られた領域は、3.3 節での領域分割結果と区別するため、「局所小領域」と呼ぶ．なお、少数の移動軌跡（例えば 5 点以下）から成る局所小領域はアウトライアとして除去する．

3.3 関節動きに不変な距離

まず、関節動きの関係にある局所小領域 C_a と C_b を示した図 3 を参照しながら「関節動きに不変な距離」について述べる．関節動きに対し、移動軌跡 $\mathbf{x}_i \in C_a$ 及び $\mathbf{x}_k \in C_b$ 間のユークリッド距離（青線）が変動する一方、測地距離（緑線）は時間不変である．すなわち、局所小領域を統合する基準となる、局所小領域間の類似度を求めるには、「関節動きに不変な距離」である移動軌跡間測地距離の、フレームセット τ にわたる時間変動を求めればよい．

はじめに、移動軌跡間ユークリッド距離について、閾値 R より大きいもののみを無限大に変更する変換を行う．

$$F'_i(\mathbf{x}_i(t), \mathbf{x}_j(t)) = \begin{cases} F_i(\mathbf{x}_i(t), \mathbf{x}_j(t)) & \text{if } F_i(\mathbf{x}_i(t), \mathbf{x}_j(t)) < R \\ \infty & \text{otherwise} \end{cases}, \quad (2)$$

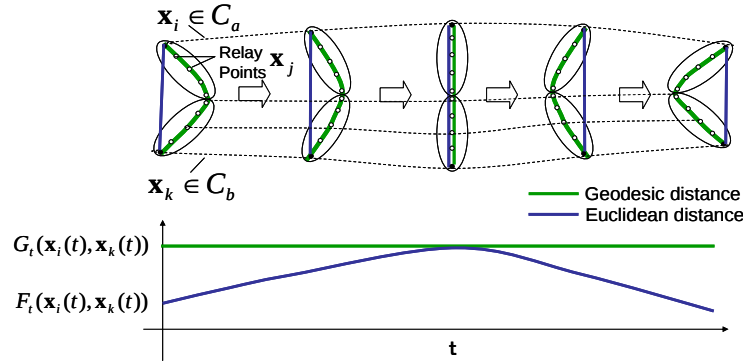


図3：関節動き時の、測地距離とユークリッド距離の違いを示す図

$F_t(\mathbf{x}_i(t), \mathbf{x}_j(t))$ は時刻 $t \in \text{frameset } \tau$ における移動軌跡間のユークリッド距離を表す。

次に、移動軌跡間測地距離 $G_t(\mathbf{x}_i(t), \mathbf{x}_k(t))$ を求める。

$$G_t(\mathbf{x}_i(t), \mathbf{x}_k(t)) = \min(F'_t(\mathbf{x}_i(t), \mathbf{x}_k(t)), F'_t(\mathbf{x}_i(t), \mathbf{x}_j(t)) + F'_t(\mathbf{x}_j(t), \mathbf{x}_k(t))) \quad (3)$$

ここで、 $\mathbf{x}_i \in C_a, \mathbf{x}_k \in C_b$ 、 C_a 及び C_b は空間的に隣接した局所小領域であり、 $\mathbf{x}_j(t)$ は、 $\mathbf{x}_i(t)$ から $\mathbf{x}_k(t)$ まで辿る経路の中継点である (図3)。測地距離は $\mathbf{x}_i(t)$ から $\mathbf{x}_k(t)$ まで、データ点 $\mathbf{x}_j(t)$ を辿って算出される最短経路の距離を示す。ユークリッド距離と異なり、最短経路の距離は、理想的には、体のある部分から別の部分まで、関節を通して辿った距離として表すことができる。 $\mathbf{x}_j(t)$ は Dijkstra 法 [18]により選択する。

さらに C_a 及び C_b を隣接する局所小領域とすると、領域間測地距離 $G_t(C_a, C_b)$ は、 C_a 及び C_b にそれぞれ属する移動軌跡間測地距離の平均として算出する ((4)式)。A 及び B は、それぞれ C_a 及び C_b に属する移動軌跡の軌跡数である。

$$G_t(C_a, C_b) = \frac{1}{AB} \sum_{i \in C_a} \sum_{k \in C_b} G_t(\mathbf{x}_i(t), \mathbf{x}_k(t)) \quad (4)$$

最後に、 C_a と C_b の統合・分離を判定する基準として、各フレームセット τ での $T+1$ フレームにわたる時間変動 $(1/T) \sum_{t=0}^{T-1} |G_{t+1}(C_a, C_b) - G_t(C_a, C_b)|$ を用いる。

剛体及び関節動きのケースでは、理想的には $G_t(C_a, C_b)$ は時間不変である。しかし実際には、検出された動きに含まれるノイズなどを考慮する必要があるため、局所小領域間を統合するかどうかを決める閾値については、実験的に定義する。

3.3.1 関節動きに不変な距離の近似距離

測地距離の時間不変性は、同一関節物体上の局所小領域を統合する基準として有効だが、式(3)の計算コストが高いことがネックである[16]。そこで我々は、測地距離の「関節動きに不変な」特徴を保ったまま計算コストを下げる別の距離を提案する。

局所小領域が十分小さいと仮定し、各小領域の動きを剛体動きとして近似すると、式 4(b)のように、同一の局所小領域に属する移動軌跡間では、測地距離をユークリッド距離で近似できる。

$$G_t(C_a) \approx F_t(C_a) = (1/A) \sum_{i,j \in C_a} F_t(\mathbf{x}_i(t), \mathbf{x}_j(t)) \quad (4b)$$

さらに、局所小領域のサイズが大きく変化しない場合、隣接した局所小領域の境界で隣接する移動軌跡間ユークリッド距離の時間変化は、測地距離と同様な性質を示すものと近似できる。以上より、隣接する局所小領域間の近似測地距離 $\hat{G}_t(C_a, C_b)$ の時間変動は以下(5)式であらわせる。

$$\frac{dG_t(C_a, C_b)}{dt} \approx \frac{d\{F_t(C_a) + F_t^{\min}(C_a, C_b) + F_t(C_b)\}}{dt} = \frac{d\hat{G}_t(C_a, C_b)}{dt} \quad (5)$$

図4に示すように、 $F_t^{\min}(C_a, C_b)$ は隣接する領域 C_a 及び C_b にそれぞれ属する移動軌跡間の最小ユークリッド距離である。図3に示すとおり、関節動きに対し G の時間変動が不変であることから、提案法では、隣接する小領域が同じ関節物体に属するかどうかを判定する基準として、領域間近似測地距離の時間不変性を利用する。

実際には、関節物体または剛体が画像平面に平行に動く場合のみ、 $d\hat{G}_t(C_a, C_b)/dt$ が、微小な値を示す。そこで、それ以外の場合に生じる、画像平面上での小領域のスケール変動に対応するため、(6)式による正規化を適用する。

$$\tilde{G}_{t+1}(C_a, C_b) = \frac{F_{t+1}(C_a) + F_{t+1}(C_b)}{F_t(C_a) + F_t(C_b)} \hat{G}_t(C_a, C_b), \quad (6)$$

ここで、隣接する局所小領域を統合するかどうかは、近似測地距離の時間変動((5)式)の平均値に基づき判定する((7)式)。閾値 $TH = 0.6$ とする。

$$\frac{1}{T} \sum_{t=0, t \in \tau} \left| \hat{G}_{t+1}(C_a, C_b) - \tilde{G}_{t+1}(C_a, C_b) \right| < TH, \quad (7)$$

統合・分割の判定は、全ての隣接する局所小領域間で行う。以降、フレームセット τ_f における領域統合によって得られた p 番目の領域を $D_{\tau_f}^p$ とする。

3.3.2 予備実験

提案法の距離が実際に関節動きに対し不変であることを確認するため、剛体と節で構成される単純な関節が動くCGの動画像を作成し、実験を行なった(図4)。

CG評価画像から求められた領域間距離を図4下部のグラフに示す。同一剛体上の小領域間では、ユークリッド距離、測地距離ともに時間不変である一方、関節物体の場合は、測地距離のみが時間不変であることがわかった。加えて、提案法の近似測地距離も測地距離に類似した時間不変性を示すことを確認できた。

4. 確率的ラベル伝播法に基づく小領域間の対応付け

混雑シーンの動画像において、短時間の移動軌跡を用いると、領域抽出の密度は高くなる半面、同一人物が異なる人物かの判定が難しくなる。一方で長時間の移動軌跡を用いると、同一人物が異なる人物かの判定は容易になる反面、遮蔽の影響により、移動軌跡が減り、領域抽出の密度が低くなる。上記のジレンマを解決するため、3章で統合した小領域に対し、確率的ラベル伝播法を時間方向に適用し、隣り合うフレーム間で領域が共有する移動軌跡の割合に基づき、全フレームセットの各領域に、一貫したラベルを割り当てる。以下で具体的な方法を説明する。

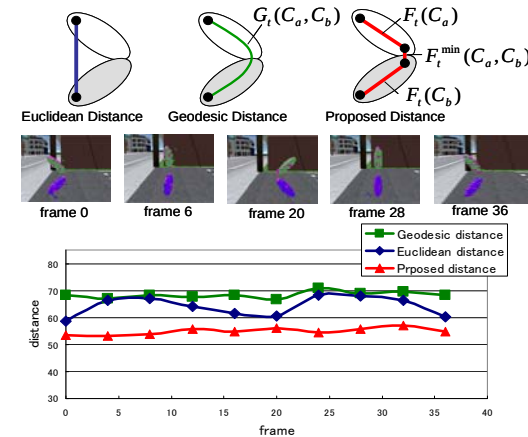


図4：測地距離及び提案法の距離の、時間的不変性を示す図

4.1 確率的ラベル伝播法

長時間にわたり一貫性のあるラベルを各領域に割り当てるには、各フレームセット τ の短時間移動軌跡に対し、それぞれ独立に求めた領域ラベルを、異なるフレームセット間で対応づけて、割り当てなおす処理が必要である。

一度も遮蔽されない移動軌跡は多くないため、長時間一貫したラベルが自明に得られるケースは稀であるため、時間的に隣接したフレームセット間の、異なる領域分割結果に対し、一貫したラベルをどのように割り当てるかを推定することになる。提案法では、局所接続性に基づく全体最適化を目的とし、確率的ラベル伝播法を採用する。

確率的ラベル伝播法[21, 22]は、本来半教師つき学習の手法として提案された手法であり、線形近傍の局所情報に基づいてラベルを伝播する。

まず、フレームセットに関係なく各領域は独立と仮定する。全フレームセット $\tau_f, \tau_{f+1}, \dots$ の小領域 $D_{\tau_f}^p$ に、異なるラベル γ を割り当てる。 Γ は全フレームセットでの領域数を、 D^γ は各領域を表す。このとき、確率的ラベル伝播法の式は以下のように定義できる。

$$\mathbf{Z}^{f+1} = \alpha \mathbf{W} \mathbf{Z}^f + (1 - \alpha) \mathbf{Y}, \quad (8)$$

$\mathbf{Z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_\gamma^T, \dots, \mathbf{z}_\Gamma^T] \cdot \mathbf{z}_\gamma^L = [z_1, z_2, \dots, z_\Gamma]$ は、逐次計算 l における各領域 D^γ の予測ラベルベクトルである。 $\mathbf{Z}^0 = \mathbf{Y}$ であり、 $\mathbf{Y}$ は各領域の信頼度とみなせる対角行列である。 $\mathbf{Y}_{\gamma\gamma}$ は各領域のサイズ、つまり、領域 D^γ に属する移動軌跡数である。 $\mathbf{W}$ は、 $\Gamma \times \Gamma$ の重み行列であり、各領域間の類似度を示す。領域間の類似度は、図 2 (d) に示すように、フレーム T^h で隣接したフレームセット(例: τ_f と τ_{f+1})での領域ペアが共有する移動軌跡の割合で示される。もし領域ペアが時間的に隣接していれば、

$$w_{pq} = \frac{N_{pq}}{N_p}, \quad w_{qp} = \frac{N_{pq}}{N_q},$$

或いは,

$$w_{pq} = w_{qp} = 0,$$

(9)

が成り立つ。 N_p 及び N_q は、時間的に隣接した領域 $D_{\tau_f}^p$ と $D_{\tau_{f+1}}^q$ にそれぞれ属する移動軌跡の数である。 N_{pq} は、領域 $D_{\tau_f}^p$ と $D_{\tau_{f+1}}^q$ が各 T^h フレームで共有する移動軌跡の数であり、 T^h フレームで隣接したフレームセットには、 T^h フレームの画像が共通して含まれる。また、 $\sum_q w_{pq} = 1$, $\alpha = 0.01$ である。式(8)は、ラベルが収束、つまり逐次計算で予測ラベルの値が変わらなくなるまで更新される([21]参照)。最終的なラベルは $\text{to } \mathbf{z}_p^{\text{Final}} = \text{argmax}_q \mathbf{Z}_{pq}$ にしたがって決まる。

T^h フレーム前後での、時間的に隣接した局所領域間の接続関係のみが含まれるため、 $\mathbf{W}$ は疎行列になるが、式(8)による逐次更新処理が行われるため、ランダムウォークタイプのクラスタリング[1]のように、フレームセット間で一貫性のあるラベル伝播が実現可能である。

4.2 移動軌跡の回復（リカバリ）

あるフレームセット中いずれかのフレームで遮蔽される移動軌跡は、3.1 節で述べ

た通り除去される。しかし、領域の密度を上げるためには、除去された移動軌跡を復元（リカバリ）することが必要である。割り当てられたラベルを用いて、遮蔽により除去された移動軌跡を復元した結果を図 2(d)に示す。復元とは、除去されラベルの割り当てられなかった移動軌跡にラベルを割り当てることである。除去された移動軌跡の多くは、フレームセットの境界である T^h フレームで、ラベルが付与された移動軌跡に接続している。これを利用することで、除去された移動軌跡にラベルを割り当てる。

例えば、フレームセット τ_f で除去された移動軌跡には、その移動軌跡に対応する移動軌跡の、フレームセット τ_{f+1} でのラベルを割り当てる。隣接したフレームセットの領域間には同一ラベルが割り振られることが多いためである。同様な移動軌跡の復元を、時間軸上で逆方向の場合についても行う。

5. 実験結果

提案法の有効性を検証するため、3つのシーケンスを用いて実験を行なった。1つは、シンプルな歩行シーケンス[10]であり、手足の大きな関節動きと、自己遮蔽への有効性を検証する。残りは混雑シーンのシーケンスである[6, 15]。一方は全身を、相互遮蔽に対する有効性を検証する。もう一方は従来法 [2, 17]との比較のため採用したUCSD シーケンスである。従来法は、[2]と類似した設定とした。提案法はPC (Core2 Quad 2.66GHz CPU / 3GB of RAM)上の処理で（動き検出を除き）約 60-70 秒/frame かかる。計算量削減のため、4 x 4 のグリッド上の移動軌跡のみを用いた。

歩行人物シーケンス:

40 フレームにわたる歩行者のシーケンスに対して検証を行なった(図 5(a)).

従来法は手足の追跡に失敗しており、[2, 17]同様に体の剛体部分のみが抽出された(図 5(b)). 個人を識別可能な[17]の手法は歩容計測のため40~60 フレームが必要だが、人物のような関節物体に対しそのような長いフレームセットを用いると、遮蔽の影響で多くの移動軌跡が除去される。

一方、短いフレームセットでは、密ではあるが時間的に一貫性がない領域分割となるが(図 5(c)), その後時間方向のラベル割り当てを行なうため、全身の領域を抽出できた(図 5(d)).

表 1 に、移動軌跡数に関する評価結果を示す。移動軌跡数はフレーム数で平均し、密度は全移動軌跡中、残存した移動軌跡の割合として算出した。提案法は従来法に比

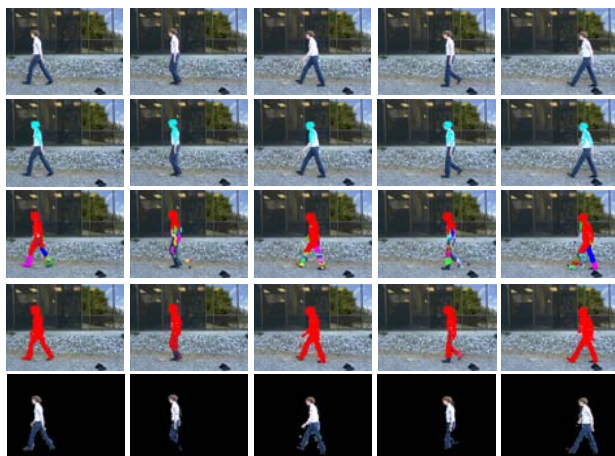


図 5 : 歩行者の画像に対する領域抽出結果

(a) : 入力 (40frames) (b) : 2行目) 領域分割結果 (従来法[1]) (c) : 3行目) 過分割
(d) : 領域分割結果 (提案法) (e) : 人物領域抽出結果

べ約 5 倍の 80%の移動軌跡が残存しており、遮蔽の影響を受けず、人領域を密に抽出できたといえる。

混雑シーン(ITE):

40 フレームにわたる混雑シーンについて検証した(図 6(a)).

ほとんどの歩行者が、40 フレームのうち 1 度は遮蔽されるため、従来手法は手前の被写体しか追跡できていない(図 6(b)).

短いフレームセット(例: $T=8$)を用いた領域分割では、全身を細かく分割される歩行者と、そうでない歩行者が混在する(図 6(c)). しかし、関節動きに不変な距離に基づいて小領域を統合し、時間方向のラベル割り当てを行なうと、各人物領域を、それぞれ密な領域として抽出できた(図 6(d)).

表 1 に示すように、提案法では従来法の約 3.5 倍の 83.7%の移動軌跡が残存しており、人物の手足の大きな動きや、すれ違いによる遮蔽を多く含む画像に対しても、提案法は有効であるといえる。また、図 6(e)のように特定の歩行者を切り出すことも可能である。ここで、手足の一部で抽出に失敗した原因としては、3.2 節で述べた過分割処理において、手足の領域にあたる小領域が、動き検出誤差を多く含むためにうまく統合されず、アウトライアとして除去された可能性がある。空間的に隣接した小領域でなければ統合されないため(式(7))、他の小領域に隣接しない局所小領域は孤立する。結果、手足に相当する領域の一部がうまく統合されなかった可能性がある。

experimental conditions	sequences	Walking	ITE	UCSD
	# of frames	40	40	80
	image size	320 × 240	720 × 480	720 × 480
	frame rate	15fps	60fps	30fps
	grid size	2 × 2	4 × 4	4 × 4
experimental results	T (See Fig.2)	8	8	16
	#of frameset	5	5	5
	number of trajectories (density) / frame			
	PROPOSED	713 (80.0%)	6330 (83.7%)	4900 (70.4%)
	PROPOSED without trajectory recovery	553 (62.0%)	5416 (71.6%)	3705 (53.2%)
experimental results	BASELINE	150 (16.8%)	1755 (23.2%)	3059 (44.0%)

表 1 : 実験結果

(PROPOSED:提案手法 (復元有/無) /BASELINE: 従来手法)

Crowded video sequence (UCSD) [15]: 80 フレームにわたる混雑シーンについて検証した。本シーケンスでは、下半身は遮蔽されるが、歩行者の上半身が余り遮蔽されないため、従来法は歩行者の一部を長時間追跡できている。それでも、表 1 に示すように、提案法は従来法よりも、25.6%多い、70.4%の移動軌跡が残存した。ITE のシーケンス同様に、所望の歩行者を切り出すことも可能であるが、下半身がほとんど全てのフレームで遮蔽されているため抽出できなかった。何人かの歩行者が同じ領域に統合される場合があり、動きが類似したため分離できなかった可能性がある。このように、動きが類似した歩行者を分離・識別するには、“個人指標” ([15]参照)の適用が有効であると考えている。

6. 結論

「関節動きに不変な距離」を用いた、混雑シーンからの人領域抽出のためのアルゴリズムを提案した。「関節動きに不変な距離」を採用した動き領域分割を行ない、短時間の移動軌跡に対する領域抽出結果を、確率的ラベル伝播法の枠組みで時間方向に一貫性のあるラベルを割り当てることで、人物ごとの領域分割と、密な人物領域の抽出を両立可能な動き領域分割を実現した。また、提案手法の距離が、関節動きに不変な性質を示すことを、CG を用いた実験により示し、最後に、実写の動画像を用いた実験により、大きな遮蔽や、手足の複雑な動きが含まれるような、従来手法では人物領

域の抽出が困難なケースに対しても、提案手法が有効であることを示した。提案手法は、人物ごとの領域分割と、密な人物領域の抽出を両立できるという点で優れており、応用として、映像編集における被写体の切り出しや、監視システムにも適用可能であると考えている。

参考文献

- 1) ArAmik and Z.Ghahramani. Spectral methods for automatic multiscale data clustering. In *CVPR*, 2006
- 2) G. J. Brostow and R. Cipolla. Unsupervised Bayesian detection of independent motion in crowds. In *CVPR*, 2006
- 3) T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In *ECCV*, 2010
- 4) P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection : A benchmark. In *CVPR*, 2009.
- 5) R.O. Duda, P.E. Hart and D.G. Stork. Pattern classification second edition. *Wiley Interscience*, page 581, 2001
- 6) ITE sequence, <http://www.ite.or.jp/eng/index.html>. *The institute of image information and television engineers*.
- 7) G.Johansson. Visual motion perception. *Scientific American* 14, pages 76-88, 1975
- 8) L. Kratz and K. Nishino. Tracking with local spatio-temporal motion patterns in extremely crowded scenes. In *CVPR*, 2010
- 9) V. Kolmogorov and R. Zabih. "Computing visual correspondence with occlusions using graph cuts". In *ICCV*, 2001
- 10) M.P. Kumar, P.H.S. Torr, and A. Zisserman. Learning layered motion segmentation of video. In *ICCV*, 2005
- 11) B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool. Coupled object detection and tracking from static cameras and moving vehicles. *IEEE TPAMI*, Vol.30, No.10, pages 1683-1698, 2008
- 12) J. C. Niebles, B. Han, and L. Fei-Fei. Efficient extraction of human motion volumes by tracking. In *CVPR*, 2010
- 13) Robert Pless. Using isomap to explore video sequences. In *ICCV*, pages 1433-1440, 2003
- 14) F. Perbet, A. Maki, and B. Stenger. Correlated probabilistic trajectories for pedestrian motion detection. In *ICCV*, 2009
- 15) V. Rabaud and S. Belongie. Counting crowded moving objects. In *CVPR*, 2006
- 16) V.de Silva and J.B. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. In *NIPS*, 2002
- 17) D. Sugimura, K. Kitani, T. Okabe, Y. Sato, and A. Sugimoto. Using individuality to track individuals: clustering individual trajectories in crowds using local appearance and frequency trait. In *ICCV*, 2009
- 18) A.Sundarasan and R.Chellappa. Model driven segmentation of articulating humans in Laplacian eigenspace. *IEEE TPAMI*, Vol.30, No.10, pages 1771-1785, 2008
- 19) J.B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500): pages 2319-2323, 22 December 2000
- 20) C.Tomasi and J. Shi. Good features to track. In *CVPR*, pages.593-600, 1994
- 21) F. Wang and C. Zhang. Label propagation through linear neighborhoods. In *ICML*, 2006
- 22) J. Wang, F. Wang, C. Zhang, H. C. Shen, and L.Quan. Linear neighborhood propagation and its applications. *IEEE TPAMI*, Vol. 31, No. 9, pages 1600-1615, 2009.
- 23) J. Yan and M. Pollefeys. A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *ECCV*, 2006



図 6 : 混雑シーン(ITE)に対する結果
(a) : 入力 (40frames) (b) : 領域分割結果 (従来法[1]) (c) : 過分割
(d) : 領域分割結果(提案法) (e) : 人物領域抽出結果