H-005

時系列の画像・深度情報を用いた人物と物体の領域抽出
Extraction of Human and Object Regions Using Image and Depth Sequences

菅原 勝也 † 阿部 亨 †,‡ 菅沼 拓夫 †,‡
Katsuya Sugawara Toru Abe Takuo Suganuma

1. はじめに

現在，監視や防犯，見守り支援などでの応用を目的として，映像に基づく人物動作認識手法の開発が進められている．しかし，映像から人物の詳細な行動を獲得するためには，人物の動作だけでなく周囲（特に，人物が接触した周囲の物体等）の状況も認識する必要があり，そのためには，先ず，人物と周囲の物体の領域を映像中で正確に分割・抽出する必要がある．

そこで本稿では，時系列の画像情報と深度情報を用いて，人物と人物が接触した物体の領域を正確に抽出する手法を提案する．

2. 関連研究

画像（映像中の各フレーム）から人物や物体の対象領域を抽出手法は多数提案されている [1, 2]．それらの手法では，各画素の輝度，色，運動，深度など様々な特徴（あるいは，その組み合わせ）を用い，画像中で特徴が類似した箇所を同一の対象領域（部分領域）として抽出している．また，時系列画像を対象とする場合，時系列全体で対象領域の抽出を行うために，例えば，各フレームで抽出された部分領域をフレーム間で対応付ける手法が提案されている [3, 4, 5]．

フレーム間で部分領域を対応付けるためには，どの部分領域が同じ対象に属するかを決定する必要があり，その手掛りには，各フレームで部分領域を抽出する場合と同様様々な特徴を利用できる．その際，輝度や色はフレーム間での変動が小さいため，それらの特徴の類似度をそのまま部分領域の対応付けに利用できる．一方，対象が動く場合，運動や深度はフレーム間で大きく変動するため，それらの類似度をそのまま対応付けに用いることはできない．そこで，フレーム間での部分領域の対応付けに深度を直接用いるのではなく，各フレーム内での部分領域間の深度の差を用いることにより，対象が動く場合の影響を抑える手法 [5] 等が提案されている．しかし，対象の姿勢や複数の対象の位置関係が変化する場合は，この手法でも対応が難しい．

3. 提案手法

本稿では，時系列の画像情報と深度情報を用い，人物と人物が接触した物体の領域を時系列全体から正確に抽出する手法を提案する．提案手法は，図 1 に示すように，各フレームを対象とした処理と時系列全体を対象とした処理で構成される．各フレームを対象とした処理では，画像情報（RGB の輝度値）と深度情報（対象までの距離）を用いて各フレームで部分領域を抽出する．時系列全体を対象とした処理では，各フレームで抽出された部分領域に対し，画像情報と深度情報を用いてフレーム間での対応付けを行う．その際，各部分領域の動き（シーンフロー）を推定し深度情報を補正することで，対象が動く場合でも，フレーム間での対応付けに深度情報を利用できるようにする．最後に，時系列全体で対応付けられた部分領域を分割・統合し人物と物体の領域を抽出する．

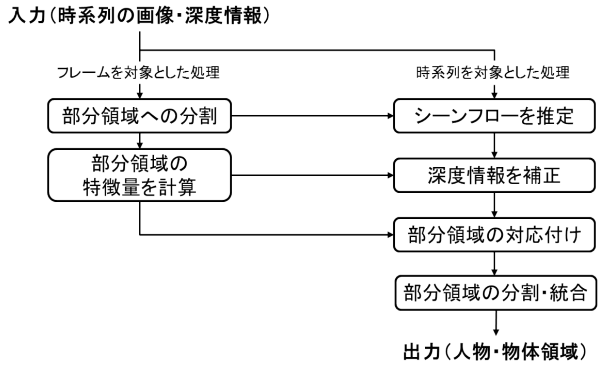

図 1: 提案手法の流れ

3.1. フレームを対象とした処理

各フレームで部分領域を抽出するために，提案手法ではグラフベース [1] の手法を用いる．この手法は，各画素をノード $v \in V$，画素の隣接関係をエッジ $e \in E$ として，フレームを無向グラフ $G = (V, E)$ で表現する．グラフ G で，隣接する部分領域（ノードの集合）C_1, C_2 が式 (1) の条件を満たすとき両者を統合し，条件を満たす部分領域が無くなるまでこれを繰り返す．

$$B(C_1, C_2) > \min(I(C_1) + \tau(C_1), I(C_2) + \tau(C_2)) \quad (1)$$

ここで，$B(C_1, C_2)$ は，部分領域 C_1, C_2 間のエッジの重み $w(e)$ の最小値を表し，$I(C)$ は，C 内の $w(e)$ の最大値を表す．また，$\tau(C)$ は閾値関数を表している．提案手法では，画像情報と深度情報を統合した 4 次元ベクトルを各画素の特徴量とし，特徴量同士のユークリッド距離をエッジの重み $w(e)$ に用いる．

3.2. 時系列を対象とした処理

時系列を対象とした処理では，まず，各フレームでシーンフローの推定を行う．シーンフローは，対象の動きを 3 次元ベクトル場で表現したものであり，これを推定するために，提案手法は，フレーム F_t と F_{t+1} の画像情報を用い，F_t におけるオプティカルフローを求める [6]．得られたオプティカルフローの始点と終点に対して，対応する画素の深度情報を反映すれば，その箇所のシーンフローを決定することができる．F_t で得

†東北大学 大学院情報科学研究科, Graduate School of Information Sciences, Tohoku University
‡東北大学 サイバーサイエンスセンター, Cyberscience Center, Tohoku University

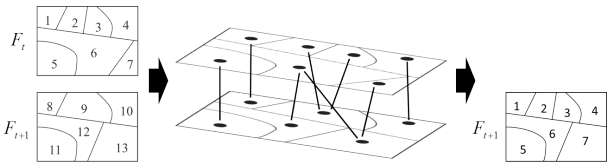

図 2: フレーム間でのグラフ表現

られたシーンフローの奥行き方向の成分を F_t の深度情報に加えることで，対象の動きにより深度情報が F_{t+1} で変動しても，その影響を補正することができる．

フレーム間で部分領域を対応付けるためには，Couprie らの手法 [3] を用いる．この手法では，フレーム F_t と F_{t+1} で抽出された部分領域を各々ノード $r, s \in V$，フレーム間で部分領域が重なる箇所をエッジ $e \in E$ とした図 2 に示すような無向グラフ $G = (V, E)$ に対し，エッジの重み $w(e)$ を式 (2) で計算する．

$$w(e) = \frac{(|r| + |s|)\, d(g_r, g_s)}{|r \cap s|} + a_{rs} \qquad (2)$$

ここで，$|r|, |s|, |r \cap s|$ は r, s および r, s 間で重なる箇所の画素数を各々表し，$d(g_r, g_s)$ は r, s の重心 g_r, g_s 間の距離を表す．a_{rs} には，r, s の特徴量同士のユークリッド距離を用い，特徴量には，RGB の輝度値，補正した深度情報を各部分領域で平均したものを用いる．$w(e)$ の昇順に，F_t のノードのラベル（部分領域の識別番号）を F_{t+1} の全ノードへ伝播することで対応付けを行う．

最後に，フレーム間での部分領域の対応付け結果に対し，シーンフローなどの推定結果をもとに人物と物体に属する部分領域の判定を行う．その後，各フレームでの部分領域の隣接関係から部分領域の分割統合を行い人物と物体の領域を抽出する．

4. 実験

提案手法の一部を実装し，各フレームで部分領域を抽出する実験を行った．実験では，Kinect で撮影した画像情報と深度情報（640 × 480 画素）を対象に，画像情報のみを特徴量として用い部分領域を抽出した場合と，画像情報と深度情報を統合した特徴量を用い抽出した場合を比較した．

実験結果の一部を図 3 に示す．この結果から分かるように，画像情報のみを特徴量に用い部分領域を抽出した場合，人物の腕の一部が背景と正しく分割できていないのに対し，深度情報を統合した特徴量を用いた場合は，両者の分割が正しく行われている．また，人物の右手先端では，異なるフレームで深度情報が大きく変化しており，フレーム間での部分領域の対応付けに深度情報を特徴量として用いるためには，対象の動きを考慮した補正が必要であることが確認できる．

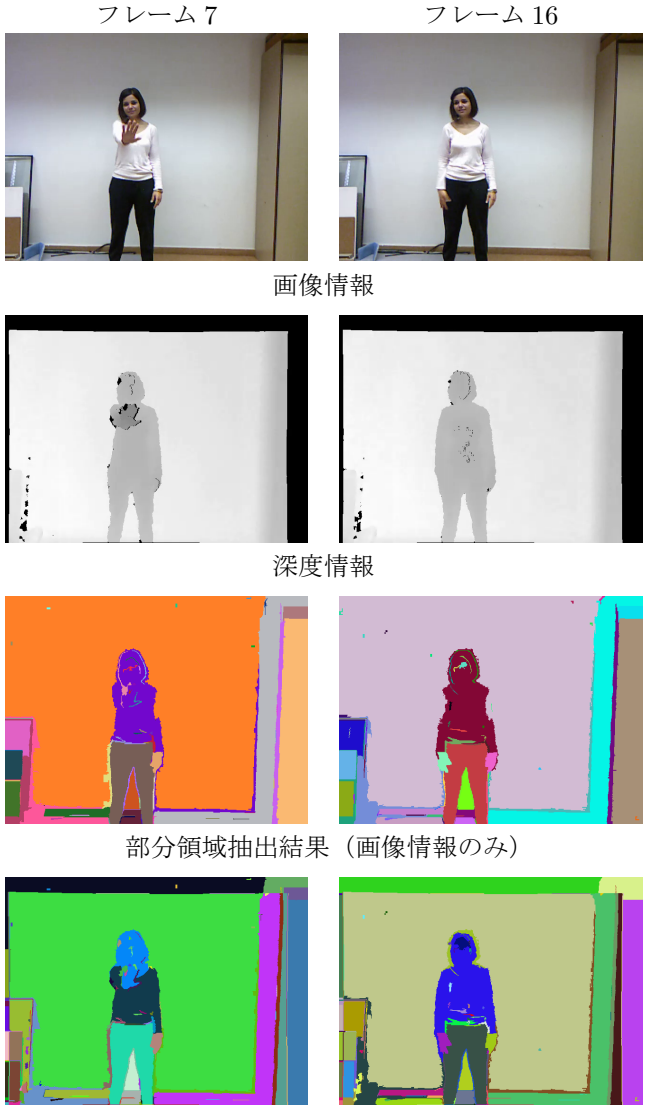

図 3: 部分領域抽出結果の例

5. おわりに

本稿では，時系列の画像情報と深度情報を用い，人物と人物が接触した物体の領域を抽出する手法を提案した．今後，フレーム間で部分領域を対応付ける処理の実装と，対応付けられた部分領域を分割・統合し人物と物体の領域を抽出する処理の設計実装を進める．さらに，種々の状況で獲得された画像情報と深度情報を対象に，提案手法の有効性を検証する予定である．

参考文献

[1] P. Felzenswalb and D. Huttenlocher, “Efficient graph-based image segmentation,” Int. J. Comput. Vision, Vol.59, No.2, pp.167–181 (2004).

[2] C. Cigla and A.A. Alatan, “Object segmentation in multi-view video via color, depth and motion cues,” ICIP, pp.2724–2727 (2008).

[3] C. Couprie, et al., “Causal graph-based video segmentation,” ICIP, pp.15-18 (2013).

[4] R. Trichet and R. Nevatia, “Video segmentation with spatio-temporal tubes,” AVSS, pp.330–335 (2013).

[5] A. Abramov, et al., “Depth-supported real-time video segmentation with the Kinect,” WACV, pp.457–464 (2012).

[6] G. Farneback, “Two-frame motion estimation based on polynomial expansion,” LNCS, Vol.2749, pp.363–370 (2003).