

2つの人物検出の組み合わせと複数特徴量の利用による人物追跡

川下 雄大† 増山 岳人‡ 梅田 和昇‡

† 中央大学大学院 ‡ 中央大学

E-mail: kawashita@sensor.mech.chuo-u.ac.jp

Abstract

本稿では、ステレオカメラを用いた人物検出・追跡手法を提案する。人物検出では、2つの人物検出手法より得た結果を組み合わせ、さらに前景領域と視差前景領域を用いて検出結果の妥当性を検証することで、検出性能を向上させる。人物追跡では、対応付け処理の際に位置・進行方向・色の評価値を組み合わせることで、ロバスト性を向上させる。さらに、色情報の評価値算出の際には人物領域を分割し、各ブロックの識別能力と関係性を考慮することで、ロバストな追跡を実現する。提案手法の有効性検証のため、実環境において人物追跡実験を行った。結果より、位置・進行方向・色（ブロック分割した場合）の特徴量を用いることで、その他の特徴量の組み合わせと比べて追跡成功率が上昇することが確認できた。

1 序論

画像処理において人物検出・追跡技術は、非常に重要な技術のひとつである。例えば、マーケティングや防犯、安全管理等の様々な分野への応用が期待されている。さらに、近年では自動車の運転支援システム技術にも応用されている。しかし、カメラ映像内の人物を検出・追跡する際、障害物や人同士によるオクルージョンという問題がある。この問題に対して、2次元情報しか得られない単眼カメラではなく、3次元情報を取得できるステレオカメラやレーザレンジファインダ等を用いることで、オクルージョンに対するロバスト性を高めている研究が多く報告されている。また、近年では低価格で距離精度が良い Microsoft Kinect を用いた研究 [1] が多く報告されているが、計測距離の短さ・屋外環境での使用の難しさが問題となる。その他に、同一の環境を複数カメラによって計測する、真上から俯瞰するようにカメラを設置するなどの方法により、オクルージョンを減らし、ロバスト性を高める方法も報告されている [2]。しかし、これらの方法では、設置コストの増加・設置環境の限定といった問題があり、実用化は難しい。



(a) Background subtraction (b) Shadow detection

図 1 前景検出

本論文では、画像情報と距離情報を同時に取得でき、かつ屋外でも比較的ロバストなステレオカメラを用いた人物追跡システムを提案する。人物検出では、2つの人物検出結果の統合・検証を行うことで、正確な検出を行う。人物追跡では、人物とトラックの対応付け処理の際に、距離情報・進行方向・色情報を用いる [3]。さらに、人物領域をブロック毎に分割し、各ブロックにおける識別能力と色情報の関係性を考慮することでロバストな追跡を実現する。

2 人物検出手法

2.1 差分ステレオ

画像中の前景領域と距離情報を取得する手法として差分ステレオ [4] を用いる。差分ステレオは、マッチングを行う前に左右カメラで背景差分を行うことで、処理する領域を限定し、ステレオマッチングの誤対応や処理時間の削減を行う。背景差分によって検出された前景領域を図 1(a) に示す。図中の青の領域が前景領域を示す。しかし、図 1(a) の状態では、影の領域も含まれている。そのため人物を正確に検出するためには、影領域を除去する必要がある。影領域は、前景と背景の近傍画素値の相違度と色相の相違度より判定を行う。今回用いた影検出手法の詳細は文献 [5] にて述べられている。影検出の結果を図 1(b) に示す。また、図中の緑の領域が影領域を表す。最終的に検出された前景領域をラベリングすることで、人物の検出を行う。

2.2 Joint HOG による人物検出

Joint HOG 特徴 [7] とは、複数の HOG 特徴の共起を表現して組合わせた特徴量である。Real AdaBoost [8] により識別に有効な組合わせを選択し、Joint HOG 特



(a) Case 1.

(b) Case 2.

図2 誤検出例

徴を生成する．2段階目の Real AdaBoost より識別に有効な Joint HOG 特徴を選択し，識別器を構築する．実際には，この識別器の出力値により人物検出を行う．2.1, 2.2 項の詳細は，文献 [9] を参照されたい．

3 人物検出結果の統合・検証

提案手法では，2章で示した2つの手法による人物検出結果を統合することで未検出を低減し，前景領域と視差前景領域を用いて結果の妥当性を検証することで誤検出を低減する．

3.1 検出結果の統合による未検出の補完

各人物検出結果を統合する際，同一人物に対する検出結果を統合する必要がある．そこで，図2のような人物領域の検出結果を表す矩形の中心座標間の距離値が，しきい値以下の場合同一人物に対する検出であると判定する．この処理によって各手法で生じた未検出を削減する．

3.2 検出結果の検証による誤検出の削減

各検出手法の誤検出の発生状況は，1. 同一人物に対して複数の検出，2. 人物がいない場所での検出の2点に分けられる．実際の誤検出発生状況を図2に示す．図2(a)の橙色の楕円に1.の誤検出例，同図(b)の赤い矢印で指す検出矩形に2.の誤検出例を示す．

1.の誤検出は，同一人物に対して検出が行われているため，検出矩形の大きさ・位置が共に非常に近いという特徴を持つ．そこで，画像座標上の検出矩形中心座標間の距離がしきい値未満である場合，同一人物に対する検出と判断し，誤検出の削減を行う．

2.の誤検出は，検出矩形内に人物がいないため，矩形内に前景領域や視差前景領域が少ないという特徴を持つ．そこで，検出矩形内に存在する前景領域と視差前景領域，及び各領域のマッチ度より結果の検証を行う．また，正しい検出の場合，矩形中心に人物がいるため，前景領域と視差前景領域は矩形中心に集中している．そこで，矩形中心に各領域が存在するほど高いスコアが出力されるように重み W_u を加える．検出矩形の横軸，縦軸座標をそれぞれ u, v とすると，重み W_u は次式で定義される．

$$W_u = k - \left(u - \frac{w}{2}\right)^2 \quad (1)$$

k は重みの最大値を設定する定数， w は矩形の幅である．そして，スコア S がしきい値以上の場合，検出結果は妥当だと判断し，しきい値未満の場合は誤検出と判断し検出結果を削除する．スコア S を以下の式で定義する．

$$S = \frac{\sum_{u=0}^{u=w-1} \sum_{v=0}^{v=h-1} W_u \times M_{uv}}{\sum_{u=0}^{u=w-1} \sum_{v=0}^{v=h-1} W_u (G_{uv} + D_{uv})} \quad (2)$$

M_{uv} は注目画素に前景領域と視差前景領域， G_{uv} は前景領域， D_{uv} は視差前景領域があれば，それぞれ1を返す変数である．それ以外の場合は，それぞれ0を返す． h は検出矩形の高さである．以上の処理を終えて，残った結果を最終的な人物検出結果とする．また，検出矩形の中心座標を人物の重心位置とする．

4 人物追跡手法

人物の追跡は，Particle Filter (PF) [10] を用いて行う．PF とは，非線形なモデルや観測値のノイズなどに対してロバストな時系列フィルタの一種である．複数人物追跡用の PF の処理手順を以下に示す．

i. Initialization

検出された人物に対してトラックを散布する．この処理は，初期検出時にのみ行う．

ii. Prediction

等速直線運動を運動モデルとして，追跡対象の位置を予測する．

iii. Weighting

各パーティクルとターゲットの距離より，各ターゲットに対する各パーティクルの重みを計算する．各パーティクルの重み W_i は，式 (3) より定義する．

$$W_i = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d_i^2}{2\sigma^2}\right) \quad (3)$$

d は人物と各パーティクルの世界座標系の X, Y 平面におけるユークリッド距離， σ は標準偏差， i はパーティクルの番号を表している．その後，重みが付いた状態のパーティクル群の重心位置をトラックの現在の位置とする．

iv. Data association

トラックの追跡対象を探索し，人物と対応付けを行う．対応付けは，トラックの保持している位置情報・進行方向・色情報と人物から得られる3つの特徴量を比較することで行う．詳細は，4.1 項にて述べる．

v. Resampling

各パーティクルの尤度に従い，追跡に有効なパーティクルの選択を行う．尤度の大きいパーティクルが存在する位置にはパーティクルを尤度に従い再配置し，尤度が小さいパーティクルは，その位置周辺に人物の重心がある確率が低いとみなし消滅させる．これにより，

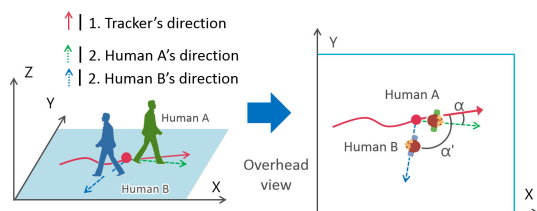


図3 進行方向算出例

追跡が続くにつれ, パーティクルは人物の重心位置近傍に収束し, 安定した人物の追跡が可能となる.

4.1 Data association

a. 位置情報による評価値の算出

位置情報による評価値は, 世界座標系の X, Y 平面におけるトラックの現在位置と人物の現在位置のユークリッド距離より求める. 評価値 D_d は, 次式で求められる.

$$D_d = \sqrt{(X_p - X_h)^2 + (Y_p - Y_h)^2} \quad (4)$$

(X_p, Y_p) はトラックの位置, (X_h, Y_h) は各フレームで検出された人物の位置を表している.

b. 進行方向による評価値の算出

進行方向による評価値は, 以下の値の角度差より求める.

1. トラックの位置情報より求めた進行方向

2. 各フレームで検出された人物の進行方向

1. の進行方向は, トラックの過去の位置情報より求められる (図3: 赤色の矢印). しかし, 2. に関しては, この処理を行う時点で人物の過去の位置が不明であるため, 直接的に算出することはできない. そこで, 各フレームの人物と一定距離内にあるトラックの過去の位置情報から進行方向を算出する. 図3の緑色の矢印は, Human A と赤トラックが対応付いていたと想定して求めた進行方向を表している. 青色の矢印は, Human B と赤トラックが対応付いていた想定して求めた進行方向を表している. α と α' は, それぞれ 1. と 2. の角度差を表している. 進行方向による評価値 D_a を, 以下の式より求める.

$$D_a = \begin{cases} k^\alpha - 1 & (\alpha < \alpha_{thr}) \\ k & (otherwise) \end{cases} \quad (5)$$

c. 色情報による評価値の算出

提案手法では, 図4上図のように人物領域全体から色ヒストグラムを取得するのではなく, 下図のように人物領域を分割し, ブロック毎に色ヒストグラムを取得し, モデルデータを作成する. さらに, 前処理として, 過去の追跡結果を用いて追跡対象の判別に有効なブロックに重みを与える. それに加え, 重みを考慮して評価値を求めることで, ロバストな対応付けを実現する.

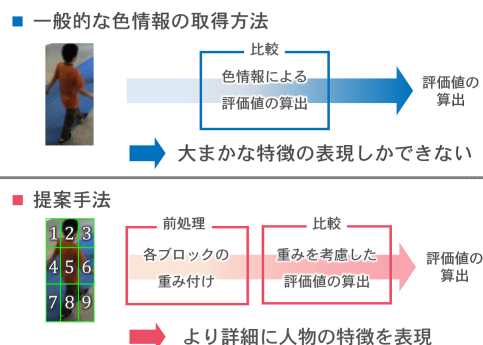


図4 色の評価値算出の流れ

図5 W_s 算出例

まず, 前処理の追跡対象の判別に有効なブロックの重み付け方法について述べる. 図5のように過去の追跡結果をブロック毎に分割し, 同一ブロック間で色相ヒストグラムの類似度の計算を行う. ヒストグラム間の類似度の計算には, Bhattacharyya 係数を用いた. 例えばブロック間で類似度が高い場合, その色情報は追跡対象に定常的に現れる特徴であり, 判別に有用だと考えられる. その逆であれば, 背景や安定しない色を含んでいるため, 追跡対象の判別に有用でないと考えられる. そこで, 各ブロックの類似度を全ブロックの類似度の平均で割ることで, 重み W_{si} を求める. ここで, i は, ブロックの番号である. 図5に算出例を示す.

次に, 比較を行う際の評価値の重み付けについて述べる. 例えば図5の人物を探索する際, モデルデータの定常性の高いブロック (2, 3, 5, 7) と類似度の高いブロックがマッチする数が多いほど, 図5の人物と似た特徴を持つと考えられる. そこで, モデルデータと人物で類似度を計算したとき, 類似度の高いブロックと定常性の高いブロックとがマッチした数に応じて重みを加える. 一致したブロックの数を c とすると, ブロック間の関係性による重み W_c は, 次式より求める.

$$W_c = c^2 \quad (6)$$

各ブロックでの色の類似度を L_i , ブロックの総数を n とすると, 色情報による評価値 D_c を以下の式より求める.

める.

$$D_c = \sum_{i=1}^n (L_i \cdot W_{si}) \times W_c \quad (7)$$

d. 最終的な評価値の算出

対応付けに用いる評価値 D は, 以下の式より求める.

$$D = \alpha(D_d + D_a) + (1 - \alpha) D_c \quad (8)$$

α は通常時 0.5 に設定されている変数で, トラックの追跡失敗数に応じて減少させる. これは, 追跡の失敗数が増加するにつれてトラックの位置と追跡対象の位置の誤差が大きくなり, 位置情報の信頼度が下がるため導入した.

5 実 験

提案手法の有効性を検証するために, 実環境において人物検出・人物追跡実験を行った. 検証には, Bumblebee 2 (Point Grey Research) を用いて撮影した距離動画像 (Frame rate: 20fps, resolution: 320 × 240) を用いた. また, 使用した PC の CPU は Intel Core i-7-4770 (3.4GHz), メモリは 12GB (RAM) である. 実験シーンでは, 歩行者同士の部分的なオクルージョン, 完全なオクルージョンが頻繁に発生する. また, 急に立ち止まる, 方向を変える人物, 走る人物なども動画像中に含まれる. また, 追跡には各人物毎 500 個のパーティクルを使用し, 追跡が 10 フレーム続いた場合に識別番号 (ID) を与えた.

5.1 人物検出実験

人物検出の評価には, オクルージョンが頻繁に発生する 1500 フレームを用いた. 実験結果を表 1, 従来手法との比較を Fig.6 に示す. 実験結果の $T.P. Rate$ は正検出率 (True Positive Rate), $F.P. Rate$ は誤検出率 (False Positive Rate), $F.N. Rate$ は未検出率 (False Negative Rate) を表す. 結果より提案手法では, 従来手法で発生した未検出・誤検出を低減出来ている. また, 図 6 では, 図中矢印で示す場所で差分ステレオによる誤検出, Joint HOG による未検出が補完できていることが確認できる. これは, 提案手法の検出結果の統合・結果の検証処理により未検出・誤検出を削減できたためだと考えられる.

5.2 人物追跡実験

次に, 人物追跡精度を検証した. 5.1 項と同様の 1500 フレームを用いて評価を行った. 評価は, 人物に ID が与えられてから画面外に出るまで ID の変化がなければ追跡成功とし, 評価を行った. 実験結果を表 2, 実験の様子を Fig.7 に示す. 表中の $D_{c:Full}$ は, ブロック分割を用いずに色の評価値を算出した場合の結果を表す. 図中の番号は, 人物毎に振られた識別番号を表している. 矩形は, 人物検出結果であり, 色は視認性向上のため

人物毎に与えられた識別番号に応じて色分けを行った結果である. 複数特徴量を用いることで, 複数人の重なりが生じた場合でもロバストな追跡が実現でき, 処理時間も 19fps とほぼリアルタイムの計測が可能である. しかし, $Frame = 230$ から $Frame = 263$ の図中の赤い楕円で囲まれた人物の追跡は失敗した. これは, 未検出の期間が長いため, トラックと追跡対象の位置が大きく離れてしまい, 再追跡が不可能だったためである.

6 結論と今後の展望

本論文では, 差分ステレオと Joint HOG による人物検出結果を統合・検証を行う人物検出手法及び, 対応付け処理に複数特徴量を用いた人物追跡システムを提案した. 人物検出では 2 つの手法による結果を統合, 検証することで検出率を向上させた. 人物追跡では, 対応付け処理の際に位置・進行方向・色情報を用いることで, 混雑状況下においても正確な追跡を実現した.

今後は, 異なる実験環境においての実験を行い, 提案手法の有用性の検証を行う. また, 異なる人物検出手法の組み合わせを検討し, さらに混雑した環境での追跡の実現を目指す.

参考文献

- [1] S. Gao, *et al.*, "Depth Structure Association for RGB-D Multi-Target Tracking," In Proc. IEEE International Conference on Pattern Recognition (ICPR), pp.4154-4157, 2014.
- [2] T. Tseng, *et al.*, "Real-Time People Detection and Tracking for Indoor Surveillance Using Multiple Top-View Depth Cameras," In Proc. IEEE International Conference on intelligent Robots and Systems (IROS), USA, pp.4077- 4082, 2013.
- [3] T. Kawashita, *et al.*, "Tracking of Multiple Humans Using Subtraction Stereo and Particle Filter", IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO), 2014.
- [4] 梅田 和昇 他, "差分ステレオ運動領域に注目したステレオ視の提案," 精密工学会誌, Vol.76, No.1, pp.123-128, 2010.
- [5] A. Moro, *et al.*, "Auto-adaptive threshold and shadow detection approaches for pedestrians detection," In Proc. AWSVCI, pp. 9-12, 2009.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CA, USA, pp. 886-893, 2005.

表 1 人物検出実験結果

Method	T.P. Rate[%]	F.N. Rate[%]	F.P. Rate[%]
Subtraction	78.0	22.0	9.8
Joint HOG	76.4	23.6	1.9
Proposed	83.2	16.8	0.9

表 2 人物追跡実験結果

Features	人数 [人]	追跡成功数 [人]	追跡成功率 [%]
D_d	72	43	59.7
$D_d + D_a$		48	66.7
$D_d + D_{c:Full}$		56	77.8
$D_d + D_a + D_{c:Full}$ [3]		58	80.6
$D_d + D_a + D_c$		67	93.1



図 6 1201 frame における人物検出結果

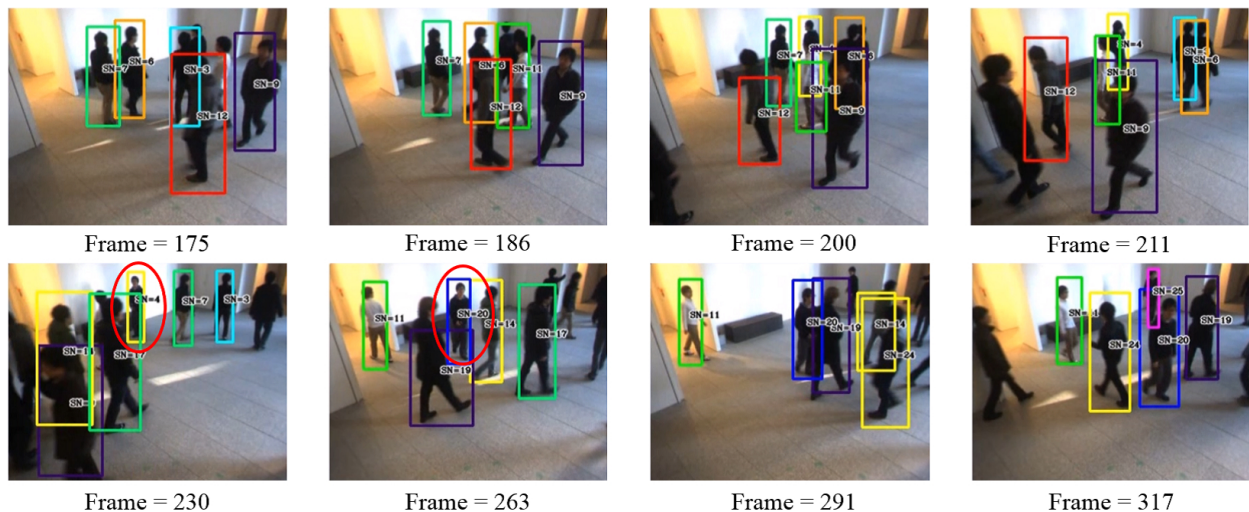


図 7 複数人が交差しすれちがう状況における人物追跡結果

- [7] T. Mita, *et al.*, “Discriminative Feature Co-occurrence Selection for Object Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 30, No. 7, pp. 1257-1269, 2008.
- [8] R. E. Schapire and Y. Singer, “Improved Boosting Algorithm Using Confidence-rated Predictions,” Machine Learning, No. 37, pp. 297-336, 1999.
- [9] 柴田 雅聡 他, “距離画像セグメンテーションに基づくリアルタイム人物検出,” 日本ロボット学会誌, Vol. 32, No. 6, pp. 558-565, 2014.
- [10] N. Gordon, *et al.*, “Bayesian state estimation for tracking and guidance using the bootstrap filter,” Journal of Guidance, Control, and Dynamics, Vol.18, No.6, pp.1434-1443, 1995.