

カメラ映像を用いた混雑環境に対応した人物追跡技術

磯部 翔 増田 誠
青木 義満

近年、IPカメラ（ネットワークカメラ）の普及に伴う防犯カメラの低価格化や防犯意識の高まりを背景に、公共施設やオフィスビル、商業施設への防犯カメラの導入が増加している。その結果、監視センターの一人のオペレーターに対する負担が増大している。そのため、カメラ映像から自動的に人物の行動や状況を認識し、オペレーター業務を支援するシステムが望まれている。例えば、膨大な映像の中から特定の人物を検出するブラックリスト照合や禁止エリアへの立ち入りを検知するシステムなどが挙げられる。

また、商業施設等では監視や防犯目的以外にも、カメラ映像から利用者の数や年齢・性別を推定し、マーケティングなどに活用したいというニーズがある。特に、POSシステムでは捉えられない、商品に興味を覚えたが購入までには至らなかった利用者の行動分析に注目が集まっている。

OKIでは、防犯用途やマーケティング用途に向けた行動認識技術の研究開発を進め、これまでに、映像中の人物の年齢・性別を推定して人数を計測する客層分析システム「RESCAT^{*1}」を製品化している。

現在、行動認識技術の用途拡大に向け、慶應義塾大学の青木義満教授と広域を対象とした人物追跡技術の高精度化に取り組み、駅や空港での混雑により隠蔽が発生する状況にも対応した人物検出技術¹⁾及び追跡技術²⁾について研究成果をあげている。本稿では、OKIが開発しているカメラ映像を用いた混雑環境に対応した人物追跡技術について紹介する。

人物追跡技術とこれまでの課題

映像中の人物を認識し、追跡する技術は、検出器ベースと特徴点軌跡ベースの大きく2つのアプローチに分けられる。

(1) 検出器ベース

検出器ベースの人物追跡技術は、予め人物の特徴を統計的に学習した検出器により、映像内から人物を検出し、フレーム間の検出結果を対応付けて人物を追跡する。映像内から一人の人物を正確に検出するため、写真1(a)に

*1) RESCATは、沖電気工業株式会社の登録商標です。

示すような狭域映像に対して比較的高い追跡精度を実現できる。しかし、写真1(b)に示すような混雑環境の広域映像において、追跡対象の人物が他の人物に隠蔽される事象が頻発する状況では、人物検出が困難となり追跡が途切れる課題がある。

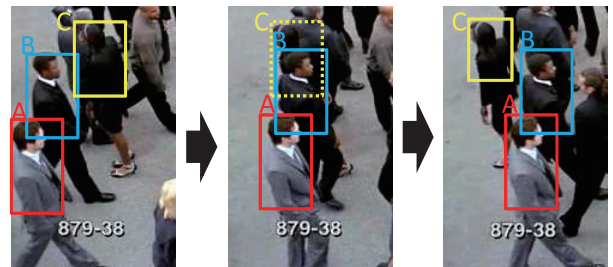
例えば、図1に示すようなすれ違いシーンにおいて、図1(a)と図1(c)では隠蔽が発生していないため3人(人物A、B、C)とも検出できるが、図1(b)では奥の人物Cの上半身が隠蔽するため検出できない。このため図1(a)の人物Cと図1(c)の人物Cは別の人物として追跡される。



(a) 狭域映像

(b) 広域映像

写真1 映像例



(a)

(b)

(c)

図1 検出器ベースの追跡イメージ

(2) 特徴点軌跡ベース

特徴点軌跡ベースの人物追跡技術は、一人の人物に属する特徴点軌跡の画像上での距離の近さや動きの類似性を利用して人物を追跡する³⁾。最初に追跡したい人物領域を指定する。次に指定した領域に対してフレーム間に対応する特徴的な点(特徴点)を求め、複数フレームに渡って特徴点を繋げた軌跡を作成する。最後に画像を色

と形状で小領域に分割して、小領域毎に追跡することで部分的な隠蔽が発生している状況でも、見えている部分の特徴点軌跡を用いて追跡を継続する。しかし、近くにいる複数の人物が同一方向へ移動する場合、特徴点軌跡を人物毎に分離することができず、一人の人物として誤って追跡する課題がある。

例えば、図2のようなすれ違いシーンでは、図2(a)に示すように3人の人物領域(A、B、C)を指定し追跡を追跡する。図2(b)に示すような人物Cが人物Bによって隠蔽されるシーンでは、人物Cの隠蔽されていない領域を追跡することで追跡が継続する。しかし、手前の人物Aと人物Bは、移動方向が同じで距離も近いため一人の人物としてまとめて追跡してしまう。

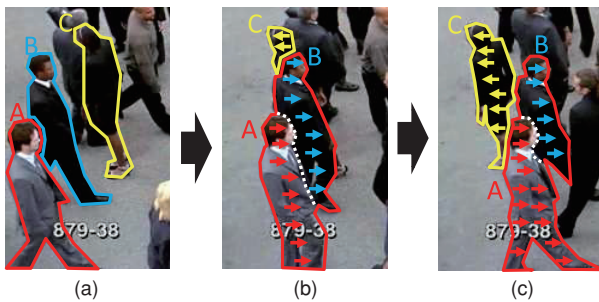


図2 特徴点軌跡ベースの追跡イメージ

混雑環境に対応した人物追跡技術

写真1(b)に示すような混雑環境の広域映像を対象とした人物追跡技術の高精度化に向けて、従来の特徴点軌跡ベースの追跡技術の課題を解決した技術を開発した。

開発した人物追跡技術の処理フローを 図3を用いて説明する。

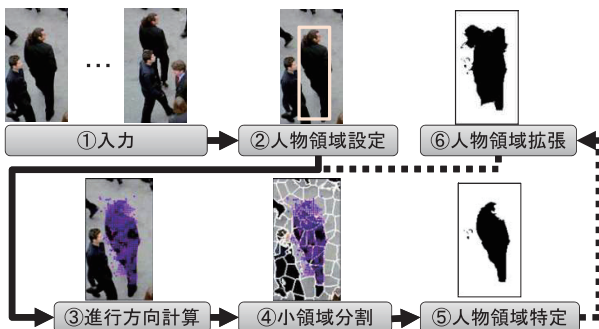


図3 追跡技術の処理フロー

①入力

時間的に連続した画像を入力する。

②人物領域設定

追跡したい人物の初期領域を矩形で指定する。特徴点軌跡ベースの人物追跡技術では、図2(a)のように人物の形状に沿って正確に領域を指定するのが一般的だが、検出結果を矩形で出力する人物検出技術と連携するため、図3②のように人物領域を矩形で与える。

③進行方向計算

人物領域の進行方向を推定するために、オプティカルフローを利用する。オプティカルフローとは、2枚の画像間での画素毎の対応付けを行い、ベクトル（移動速度を大きさ、移動方向を向き）として表現した技術である。

写真1(b)のような広域映像に対してオプティカルフローを計算すると、隠蔽領域や背景領域にノイズが多く含まれるため、2種類のフローチェック処理によりノイズを除去する。1つ目は、2枚の画像間でオプティカルフローを双方向に計算し、お互いに対応する点が一致しないものを除去する処理である。例えば、図4に示すような人物Cの隠蔽前後の画像で双方向にオプティカルフローを計算すると、人物Cの隠蔽されない領域（図4(a)(b)の実線領域）では対応する点が一致するが、隠蔽領域（図4(a)(b)の点線領域）では対応する点は一致しない。対応する点が一致しないものを除去することで隠蔽領域のノイズを除去する。2つ目は、一人の人物が同時に複数の方向へ移動しないという制約の基に、過去の進行方向と異なるフローを除去する。

ここで、正しい人物領域でも周辺と同じ色の領域や境界領域のオプティカルフローは安定せず、フローチェック処理で一定量削除されてしまうため、人物領域を復元する処理が必要になる。

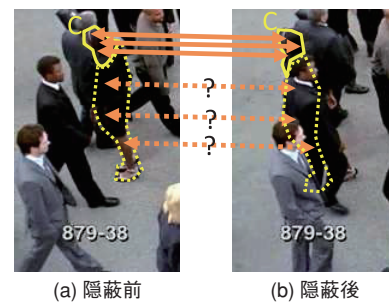


図4 追跡技術の処理フロー

④小領域分割

人物領域を復元するために画像全体を色や形状を考慮した小領域に分割する。分割する手法としてSimple Linear Iterative Clustering (SLIC)⁴⁾を利用した。SLICの計算結果

を 図5 に示す。 k は分割数を示し、 m は色距離およびユークリッド距離の重みを表している。 k を大きくすると画像を細かく分割し、 m を大きくすると矩形に近く分割する。

我々は、SLICを時間方向へ拡張することで追跡に適した小領域に分割する手法を開発した。画像の分割数を k 、時間方向の長さを $frameset$ 、時系列画像の枚数を T 、画像の総画素数を N とすると、時系列画像を分割する直方体数（ラベル数） K 、直方体の空間方向での1辺の長さ s および時間方向の長さ s' は以下の式で求められる。

$$K = k \times \frac{T}{frameset} \quad S = \sqrt{\frac{N}{k}} \quad S' = \frac{T}{frameset}$$

時系列画像内のある画素を注目画素としたとき、注目画素を中心とした $2S \times 2S \times 2S'$ の領域内に含まれる注目画素と各直方体の中心座標との距離 D は以下の式で求められる。

$$D = \sqrt{d_c^2 + m \times d_m^2} \quad \begin{aligned} d_c^2 &= (l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2 \\ d_m^2 &= (x_j - x_i)^2 + (y_j - y_i)^2 + (t_j - t_i)^2 \end{aligned}$$

ここで i, j は画素のインデックスを表す。 d_c は色距離を表し、 l, a, b は CIELAB 表色系における画素値を表す。 d_m はユークリッド距離を表し、 x, y, t は空間座標および時間方向の座標を表す。注目画素に対して距離 D が最小となる画素を含む直方体インデックス（ラベル）を注目画素のラベルとする。

図6に時間方向へ拡張したSLICの適用例を示す。ラベルを色で表現しているため、同一の直方体に属する領域は同じ色をもつ。人物の頭部（図6(a)の丸領域）に着目すると、人物の移動に合わせて小領域が追従している。

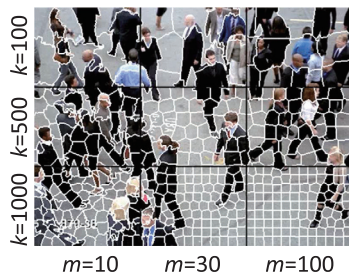


図5 SLIC の計算結果



図6 時間方向へ拡張した SLIC の適用例

⑤人物領域特定

フローチェック処理によって完全にノイズが除去できない場合に人物領域を復元すると、図7(a)に示すように背景を含んで復元する。そのため3つの補正処理によって図7(b)に示すように背景を除去する。1つ目は、人物サイズが縦長であることを制約条件として、小領域を結合する処理である。人物領域が横に広がることを防ぐ効果がある。2つ目は、人物の色情報を利用して背景や他人の領域を除去する処理である。人物の色情報は、開始フレームで設定した人物領域からk-means法によって代表的な色を2色抽出する。3つ目は、人物の移動軌跡と異なる動きをする小領域を除去する処理である。これにより追跡対象と似た色情報を持つ人物や背景を除去できる。



図7 人物領域特定処理の結果例

⑥人物領域拡張

追跡対象の人物の一部が隠蔽されるなどしてオブティカルフローの精度が悪くなると、フローチェック処理や補正処理によって正しい人物領域が削除されることがある。その結果、追跡領域が縮小されたまま追跡が継続する事になり追跡精度の低下を招く。そこで、過去に人物領域として対応付いたラベル情報を基に人物領域を拡張する。

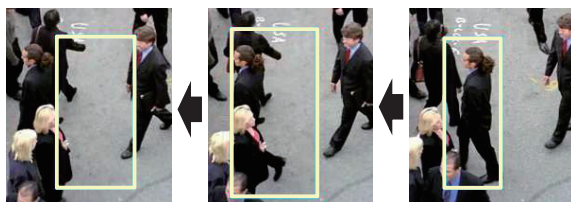
提案方式による人物追跡結果

開発した特徴点軌跡ベースの人物追跡技術の効果を確認するために、UCF Crowd Dataset⁵⁾の中から人物が多くすれ違うデータセット(フレーム数300、追跡人数68人)に対して人物追跡を実施した。評価指標は正解矩形と検出矩形の和集合に対する共通集合の割合が10%以上を占めるものを追跡成功として、総正解フレーム数に対する総追跡成功フレーム数で算出した。従来の特徴点軌跡ベースの追跡技術³⁾と比較した結果を表1に示す。比較手法と比べて本手法は、17.21%精度が向上した。追跡例を図8に示す。比較手法は初期領域として人物形状に沿った領域を想定しているため、本評価のように背景を含む矩形を初期領域

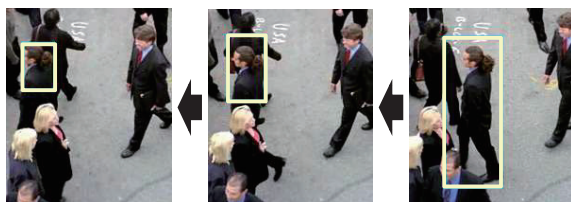
とすると背景領域を除去できず追跡が失敗している。本手法は初期領域を矩形で与えたとしても背景領域や他人の領域を除去できているため、見えている領域で追跡が継続している。

表 1 精度比較結果

手法	追跡精度
比較手法 ³⁾	64.21%
本手法	81.42%



(a) 比較手法



(b) 本手法



(c) 本手法による軌跡

図 8 追跡例

おわりに

本稿では、カメラ映像を用いた混雑環境に対応した特徴点軌跡ベースの人物追跡技術を紹介した。今後、OKI 固有の人物検出技術との融合による高精度な人物追跡技術の開発を計画している。本技術を用いることで防犯ソリューションの高度化による豊かな社会の実現に貢献していく所存である。◆◆

参考文献

1) 鈴木哲平、他：動き特徴を考慮した領域分割による動物体領域抽出、精密工学会誌、Vol.83 No.2 p.151-157、2017

2) Shota Takayama, et al: Tracking People in Dense Crowds Using Supervoxels, HTBA2016.

3) Clement Creusot: Local segmentation for pedestrian tracking in dense crowds, In Multimedia Modeling, pp.266-277, 2014.

4) Radhakrishna Achanta, et al: Slic superpixels compared to state-of-the-art superpixel methods, Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 34, No. 11, pp. 2274-2282, 2012.

5) Saad Ali and Mubrak shah, A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, CVPR2007.

● 筆者紹介

磯部翔: Sho Isobe. 情報・技術本部 研究開発センター センシング技術研究開発部

増田誠: Makoto Masuda. 情報・技術本部 研究開発センター センシング技術研究開発部

青木義満: Yoshimitsu Aoki. 慶應義塾大学 理工学部

TiPO 【基本用語解説】

フレーム

映像を静止画の連続データとして表現した場合に静止画1枚の事をフレームと呼ぶ。

SLIC

画像中の領域を色の類似度で分割する手法の一つ。最初に画像を格子状に分割してから色の類似度でグループ化することで高速に画像を分割する手法。

k-means法

データの集まりをデータ間の類似度によっていくつかのグループに分類する手法の一つ。予め設定したグループの数に従って、グループの代表値を更新しながらグループ化する手法。