

Camera Dropout と Trajectory Ensemble による多カメラ間複数人物追跡

川西 康友^{†a)} 出口 大輔^{††b)} 井手 一郎^{†c)} 村瀬 洋^{†d)}

People Tracking across Non-overlapping Camera Views via Camera Dropout and Trajectory Ensemble

Yasutomo KAWANISHI^{†a)}, Daisuke DEGUCHI^{††b)}, Ichiro IDE^{†c)},
and Hiroshi MURASE^{†d)}

あらまし 複数カメラにわたる多人数追跡は、通常、対応付けの候補を減らすために複数のカメラの隣接関係をカメラネットワークというグラフ構造で表現し、カメラネットワーク上の隣接カメラ間での人物対応付けにより実現される。しかし、カメラ間の照明環境の違いや撮影した人物の向きの違いによって対応付けを誤ることがあり、このような場合に追跡が途切れることがある。この問題に対し、隣接カメラだけでなく、カメラネットワーク上で離れたカメラ間での人物対応付けをすることが考えられるが、必要以上に対応付け候補が増加してしまい、誤対応の原因となる。本論文では、本来のカメラネットワークからカメラを無作為に削除したカメラネットワークを複数作る (Camera Dropout) ことで、一部のカメラについては離れたカメラまで考慮した対応付けを行い、更に各カメラネットワーク上での追跡結果をアンサンブルクラスタリングの考え方で統合し、最終的な追跡結果を得る Trajectory Ensemble という手法を提案する。実験では、公開データセットから仮想的に生成した複数のデータセット上で評価を行い、既存の追跡手法に加えて Trajectory Ensemble を導入することにより、精度良く追跡ができることを確認した。

キーワード Trajectory Ensemble, Camera Dropout, 人物追跡, カメラネットワーク, アンサンブルクラスタリング

1. ま え が き

近年、街のいたるところに防犯カメラなどの固定カメラが設置され、人々の様々な行動の観測に利用されている。その際、視野が重複しない複数の固定カメラ間で、多人数を同時に追跡し、移動履歴を獲得する技術が求められている。ここで移動履歴とは、複数のカメラ視野間を人物が移動した順序を表す。多人数の移動履歴が得られれば、ショッピングモール等における不審な人物の発見や、移動履歴に基づく商品や店舗の

配置の改善を目的とした分析など、防犯や市場調査の分野で活用できると考えられる。そこで、自動的に多人数を多数のカメラ視野間にわたって追跡する技術に期待が集まっている。特に、実時間での追跡が要求されないような場合、一定時間蓄積したデータを用いるオフライン追跡は、高精度な追跡結果が得られることから、注目されている。

多くの従来研究は、カメラで観測した人物の見えの特徴と人物の移動に関する時空間制約に基づいて、隣接するカメラ視野間で各カメラにおいて検出した人物が同一人物であるか否かを判定することにより、多人数追跡を実現している [1]~[5]。画像の照合だけで同一人物か否かを判定することは困難なため、多カメラ間の人物追跡では、ネットワークを用いてカメラの隣接関係を表現し、そのネットワークの情報に基づいて、観測された時刻が早いものから順に対応付けていく方法が一般的である。本論文では、このような、カメラネットワークを用いて候補を絞り込みながら多人数追跡することを、“カメラネットワーク上での多人数

[†] 名古屋大学大学院情報学研究科, 名古屋市

Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

^{††} 名古屋大学情報連携統括本部情報戦略室, 名古屋市

Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

a) E-mail: kawanishi@i.nagoya-u.ac.jp

b) E-mail: ddeguchi@nagoya-u.jp

c) E-mail: ide@i.nagoya-u.ac.jp

d) E-mail: murase@i.nagoya-u.ac.jp

DOI:10.14923/transinfj.2017IUP0010

追跡”と呼ぶ。Song ら [2] は、時間順に動的計画法により対応付けを行い、対応付いた系列の長期間の特徴量分布をもとに対応付けを修正する手法を提案している。Chen ら [4] は、対応付けをした系列のなかで誤対応を評価する Error Path Detection Function を定義し、誤対応があれば修正する手法を提案している。Alahi ら [6] らは、カメラ内人物追跡で用いられる Data Association [7] の考え方を応用し、グループなど人物間の関係性も用いた人物追跡手法を提案している。

カメラネットワーク上での多人数追跡をする際、同一人物か否かの判定をする候補に視野内の照明環境や設置向きが異なるカメラがあると、それらのカメラで観測された画像は見えの特徴が大きく異なる。その結果、同一人物であっても類似度が低くなり、同一人物であると判定されない場合や、別人物が同一人物であると誤判定される場合がある。前者の場合は、その時点で追跡が途切れ、そこから先の追跡系列には別人として追跡されることになる。後者の場合は、別人が同一人物であると判定されることを防ぐには、対応付けの際のしきい値を高くする必要があるが、それによって逆に同一人物の追跡が途切れてしまうおそれがある。

同一人物か否かの判定にカメラネットワークに基づく時空間制約を用いず、隣接する 2 カメラ間のみに限定しなければ、一時的な見えの変化による追跡の途切れに対応できる可能性はあるが、候補の増加により誤判定の可能性も高まる。カメラの隣接関係のある程度考慮しつつ、隣接する 2 カメラ間における対応付けの制限のみを緩和するには、時空間制約を緩和して、カメラネットワークの数ホップ先までのカメラとの間で同一人物か否かを判定すればよい。しかし、全てのカメラでそのような判定をすると、候補の増加により誤判定の可能性が増す。

そこで、カメラネットワークから幾つかのカメラを削除することで、基本的には時空間制約を用いて隣接した 2 カメラ間での同一人物判定をしつつ、部分的にはカメラの時空間制約を緩和して数ホップ先のカメラとの間で同一人物判定をすることを考える。しかし、どのカメラを削除すればよいかは事前にはわからないため、本論文では、無作為にカメラを削除 (Camera Dropout) して複数回の追跡を行い、その結果を教師なしアンサンブル学習の考え方で統合する Trajectory Ensemble を提案し、一時的な見えの変化による追跡の途切れに対処する。

本論文での貢献は以下のとおりである。なお、本論文は文献 [8] を発展させて問題の新たな定式化を与えるとともに、まとめたものである。

- Camera Dropout の提案：カメラを無作為に削除したネットワークを生成することで、カメラネットワークを利用した人物追跡における時空間制約の選択的な緩和を実現する。
- Trajectory Ensemble の提案：Camera Dropout で得られる複数の異なる追跡結果に対するアンサンブルクラスタリングにより、一時的な見えの変化に頑健な多カメラ間人物追跡を実現する。

本論文の構成は以下のとおりである。まず、2. で従来の複数カメラ間での人物追跡についてまとめ、従来手法の一般的な定式化について述べる。また、関連する研究についてまとめる。次に、3. では、提案手法である Trajectory Ensemble について詳述する。4. では評価実験により、提案手法の有効性を確認し、5. でまとめと今後の展望について述べる。

2. カメラネットワークに基づく多人数追跡

2.1 問題設定と一般的な定式化

ショッピングモール内の通路などの環境中に、視野が重複しないようにカメラが複数設置されており、その環境中を複数の人物が移動している状況を考える。人物がカメラの視野内に入ると、その人物が検出され、その人物がカメラの視野外に出るまで追跡される。そして、一つの人物画像系列 (トラックレット) $r_i = \{m_{ij}\}_{j=1}^{M_i}$ が得られるとする。ただし、検出もれや人物以外の誤検出はないものとする。ここで、 m_{ij} は追跡している人物を切り出した人物画像であり、 M_i は得られた人物画像の枚数である。本論文で対象とする多人数追跡は、カメラの集合 $C = \{c_1, \dots, c_N\}$ から得た全てのトラックレット $R = \{r_1, \dots, r_K\}$ について、同一人物だと判定したものに対して一意のラベルを割り当てる問題である。このとき、カメラの隣接関係及び、カメラ間の移動の所要時間などの情報が、カメラネットワーク $G = (C, E)$ として与えられる。ただし、 $C = \{c_1, c_2, \dots, c_N\}$ は頂点の集合、 $E = \{(c_j, c_k) | c_j, c_k \in C\}$ は辺の集合である。カメラネットワークの各頂点はカメラの視野に対応し、辺で結ばれたカメラは隣接していることを表す。2 台のカメラが隣接しているとは、他のカメラの視野内に入ることなくその 2 台のカメラ視野間を移動できることを言う。カメラネットワークの各辺には、その辺が結ぶ

二つの頂点間を人物が移動する際の所要時間の分布が与与される。所要時間の分布は、例えばガンマ分布で表現され、カメラ間の距離に応じた移動時間パラメータが与えられる [9].

この問題設定のもとで、カメラネットワーク上での多人数追跡の一般的な定式化をすると以下のようになる。

$$\hat{L} = \underset{L}{\operatorname{argmax}} P(L|R, G) \quad (1)$$

ただし、 P はトラックレットの集合 R とカメラネットワーク G が与えられたときの、ラベル割当てに関する事後確率であり、これを最大にするラベル割り当て \hat{L} を見つけることが、カメラネットワーク上での多人数追跡であると言える。これは離散最適化問題であり、一般に最適解を得ることは困難である。そのため、貪欲法や、動的計画法などのアルゴリズムを用いて近似的に解かれることが多い。最も基本的な多人数追跡のアプローチは、トラックレットを二つずつ比較し、そのトラックレット対が類似していれば同一人物と判断し、同一のラベルを割り当てるものである。また、一度対応付けた後に、全体の整合性をみて適宜対応付けが修正される [2], [4].

2.2 関連研究

カメラネットワーク上での多人数追跡の最も基礎的な技術は、視野の異なるカメラ間で人物の同一判定を行うことであり、人物再同定 (Person Re-identification) と呼ばれ、多くの研究がなされている [10], [11]. 人物再同定では、基本的に 2 台のカメラ間で人物の同一判定を行うが、その判定には、人物の見えに関する特徴と、人物の移動時間の尤もらしさからなる類似度を用いられることが多い。

人物の見えに関する特徴としては、色やテクスチャの特徴量が用いられることが多い [12], [13]. 画像 1 枚ずつではなく、各人物について複数枚の画像を得られる場合は、複数枚の画像から得られる統計的な情報が特徴量として使える [14]. また、多数の画像やそれらの特徴量から、有効なものを選択的に重み付けする距離指標学習も、この分野の重要な研究課題の一つである [15], [16]. 更に近年、他の画像認識分野と同様、深層学習に基づく特徴量も注目されている [17]~[20].

一方、移動時間の尤もらしさとしては、2 台のカメラ間の移動に要する時間に範囲を設けて候補を制限したり、その 2 台のカメラ間の一般的な移動時間との一致度を評価することが多い [21]~[25].

2 台のカメラ間ではなく、1 台のカメラ視野内における人物追跡問題についても、カメラ視野内における追跡対象の見失いからの復帰や追跡対象の入れ替わりを考慮して、各フレームでの人物検出結果やトラックレットを統合する手法の研究がなされている。特に、各フレームにおける人物検出結果に対する最適なラベル割当てを目的としたものは、Global Data Association と呼ばれ、様々な手法が提案されている。Zhang ら [7] は、最小費用流問題 (Minimum-cost Flow Problem) [26] として、Berclaz ら [27] は k -最短路問題 [28] として、Zamir ら [29] は、最小クリーク被覆問題 [30] として定式化し、各々解法を提案している。これらの手法は、各フレームにおける検出結果を、カメラ視野内における追跡結果に置き換えることで、本研究が想定する多カメラ間の人物追跡にも応用が可能であると考えられる。

本論文の内容は、これらの知見をもとに、カメラネットワーク上での多人数追跡を行うものである。

3. Trajectory Ensemble

3.1 概要

カメラネットワーク上での多人数追跡では、カメラの設置環境や向きの違いなどによる人物の一時的な見えの変化によって、複数カメラ間の追跡が途切れてしまうことが多々ある。本論文では、教師なしアンサンブル学習の一つであるクラスタアンサンブル [31] の考え方をカメラネットワーク上での多人数追跡に応用し、条件を変えた複数の追跡結果を統合することにより、最終的な追跡結果を得る Trajectory Ensemble を提案する。

クラスタアンサンブルの基本的な考え方は、クラスタリング基準などの条件を変えて複数回のクラスタリングを行った“弱クラスタリング”結果を得たのち、多数の弱クラスタリング結果において同一クラスタに割り当てられたデータは同一クラスタに属すると判断して統合することで、最終的なクラスタリング結果を得るものである。本論文では、この考え方をカメラネットワーク上での多人数追跡に応用する。提案手法は、条件を変えて複数回行った“弱追跡”結果を統合することで、最終的な追跡結果を得る。

様々な追跡結果を得る方法は幾つか考えられるが、本論文では、元のカメラネットワーク G からカメラを無作為に削除したカメラネットワーク (サブネットワーク) G_s を多数 ($s = 1, \dots, S$) 生成して多人数追

跡を行い、その結果を統合することを考える (図 1). カメラを無作為に削除することにより異なるサブネットワークが得られ、同一人物か否かの判定をするトラックレットの候補が変化する. その結果, 生成した異なるサブネットワークの数だけ異なる追跡結果が得られる (式 (2)).

$$\hat{L}_s = \operatorname{argmax}_{L_s} P(L_s | R_s, G_s) \quad (s = 1, \dots, S) \quad (2)$$

ただし, G_s はカメラネットワーク G のカメラ集合 C の部分集合 $C_s \subset C$ を頂点とするグラフであり, $R_s \subset R$ はトラックレットの集合 R のうち, カメラ集合 C_s に含まれるカメラにより観測されたトラックレットの集合である. このとき, 多くのサブネットワーク上での追跡において, 同一人物のトラックレット対にはそれぞれ同一のラベルが割り当てられる可能性が高いと考えられる.

この R_s と G_s を用いてカメラネットワーク上での多人数追跡を実行した結果のラベル割当てが \hat{L}_s であり, それらをアンサンブル学習の考え方をを用いて, 関数 \mathcal{M} により統合した結果, 最終的なラベル割当て \hat{L}^* を得る (式 (3)).

$$\hat{L}^* = \mathcal{M}(\hat{L}_1, \dots, \hat{L}_S) \quad (3)$$

その結果, 多くのカメラネットワーク上での追跡結果

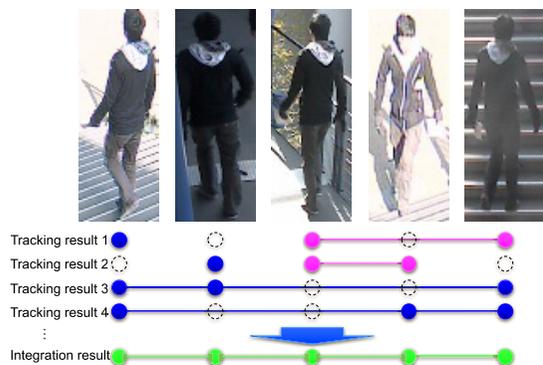


図 1 Trajectory Ensemble の概念図. 各画像は異なるカメラからの観測結果である. 破線の丸印は Camera Dropout によりカメラが削除されたために “弱追跡” に利用されなかったことを表す.

Fig. 1 Concept of Trajectory Ensemble. Each image is observed from a different camera. Dashed circles indicate that the corresponding tracklets were not used for “weak” tracking since the corresponding cameras were dropped in the sub-networks.

において, 同じラベルが割り当てられたトラックレット対には, 最終的に同じラベルが割り当てられることになる.

複数の追跡候補を統合するという点での類似研究は幾つか存在する [32]~[34] が, サブネットワークでの複数の追跡結果の統合を行う既存研究はない.

3.2 Camera Dropout によるサブネットワークの生成

サブネットワーク G_s の生成は, カメラネットワーク $G = (C, E)$ をもとにして, 削除するカメラ台数と同じ回数以下の手順を繰り返す.

- (1) 無作為に 1 台のカメラ $c_j \in C$ を選択する.
- (2) カメラ c_j に隣接するカメラの集合を $\mathcal{N}(c_j)$ とする.
- (3) $c_k, c_l \in \mathcal{N}(c_j), k \neq l$ の全ての組み合わせに対し, カメラネットワークに辺 (c_k, c_l) を追加し, その移動時間パラメータを, 辺 (c_k, c_j) と辺 (c_j, c_l) に付与されている移動時間パラメータの和とする. 辺 (c_k, c_l) に既に辺が存在する場合, 異なる移動時間パラメータをもつ辺 (c_k, c_l) が複数存在することになる.
- (4) カメラ c_j 及び, c_j に接続する全ての辺を削除する.

元のカメラネットワークの例と, そのカメラネットワークから生成されたサブネットワークの例を図 2 に示す. 図 2 (i) 矢印左側のカメラネットワークから, カメラ 2 を削除した例 (図 2 (i) 矢印右側の中央) のカ

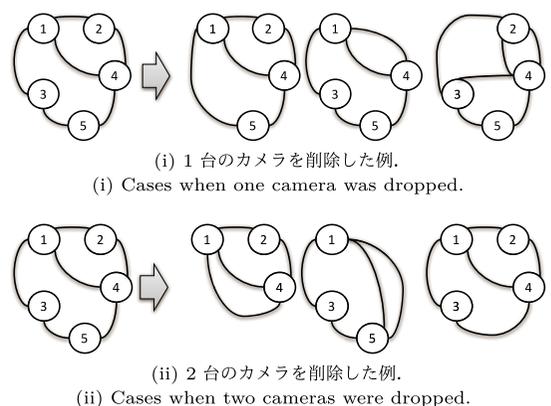


図 2 無作為なカメラの削除により得られたサブネットワーク. 幾つかのカメラの組は, 元の移動経路を維持するために複数の辺をもつ.

Fig. 2 Sub-networks generated by Camera Dropout. Some camera pairs have multiple edges to keep the original routes.

メラネットワークでは、カメラ1からカメラ4への辺が2本存在する。これは、元のカメラネットワークで存在したカメラ1からカメラ4への移動経路に、カメラ1からカメラ2を経由してカメラ4へ移動する経路が追加されたことによる。それぞれ経路が異なるため、異なる移動時間パラメータをもつ。

3.3 各サブネットワーク上での多人数追跡

カメラネットワーク G から生成したサブネットワーク G_s ($s = 1, \dots, S$) 上で多人数追跡を行い、 $P(L_s | R_s, G_s)$ を最大とするラベル割当て \hat{L}_i を得る。追跡手法は何を用いても良いが、ここでは、1台のカメラ視野内におけるオフライン追跡手法の中で、精度が良いとされている Global Data Association [7] による追跡を、カメラ間追跡へと応用する。この Zhang ら [7] の手法における各フレームで検出された人物画像を、本論文では各カメラで検出されたトラックレットと置き換える。具体的には、検出された各トラックレットを二つの頂点で表現し、ソースとシンクの二つの頂点を加えた有向グラフで表現する。各辺には、人物画像の類似度の代わりにトラックレット同士の類似度を用いたコストを付与し、ソースからシンクへの最小費用流問題 (Minimum-cost Flow Problem) [26] として解く。ただし、各カメラの撮影時刻は同期しているとは限らないため、隣接フレーム間で検出された画像同士の対応付けではなく、隣接カメラで直前に検出されたトラックレット同士の対応付けと考える。

各サブネットワーク G_s では、追跡により、同一人物と判断したトラックレット同士が対応付けられるため、同一人物のもと判断したトラックレットの集合に一意のラベルを割り当てる。ただし、追跡に利用しなかったトラックレットに対するラベルは欠損値 (NA) とする。その結果、全てのトラックレット $r_i \in R$ に対する S 回の追跡結果のラベル割当て \hat{L}_s ($s = 1, \dots, S$) が得られる。

3.4 追跡結果の統合

前節の追跡処理は独立に行われるため、同一トラックレットであっても、追跡結果 \hat{L}_a, \hat{L}_b ($a \neq b$) では同一ラベルが割り当てられているとは限らない。また、各追跡結果には、削除したカメラで観測されたトラックレットのラベルは欠損値となっている。そこで、各追跡結果 \hat{L}_s を関数 \mathcal{M} を用いて統合し、全てのトラックレットに対して、全体として一貫したラベル割当てを行い、最終的な追跡結果 \hat{L}^* を出力する。

ここでは、 \mathcal{M} として教師なしアンサンブル学習の

一つである、クラスタアンサンブル [31] を用い、各サブネットワーク G_s 上での追跡結果を統合する。以下、具体的な処理について述べる。

トラックレット r_i は、各追跡結果のラベル割当て \hat{L}_s ($s = 1, \dots, S$) において、それぞれ同一人物と判断したトラックレットに対して一意のラベルが割り当てられている。あるトラックレット r_i について、各ラベル割当て \hat{L}_s で割り当てられたラベルを並べたベクトルを $\ell_i = (\ell_{i1}, \dots, \ell_{iS})$ とする。

二つのトラックレット r_{i_1} と r_{i_2} に対し、多くのサブネットワーク上での追跡において同じラベルが割り当てられているならば、そのトラックレット対は実際に同一人物である可能性が高い。しかし、サブネットワーク間で、同一のラベルが割り当てられているとは限らない。また、ラベルが欠損値となっているトラックレットも存在する。そこで、 r_{i_1} と r_{i_2} に対応する ℓ_{i_1} と ℓ_{i_2} の距離を用いて再度クラスタリングをすることにより、最終的なラベル割当てを得る。

このとき、ベクトル ℓ_i の各要素の値は、ラベルであり、値の大小関係に意味がない。また、欠損値 NA を含む。そこで、 L_0 ノルムを改良し、かつ欠損値を考慮した以下の距離指標 d により、ベクトル ℓ_{i_1}, ℓ_{i_2} 間の距離を定義する。

$$d(\ell_{i_1}, \ell_{i_2}) = \sum_{s=1}^S I(\ell_{i_1s}, \ell_{i_2s}) \quad (4)$$

$$I(a, b) = \begin{cases} 1 & \text{if } a \neq b, a \neq \text{NA}, b \neq \text{NA} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

この距離指標を用いて、割り当てられたラベルを並べたベクトル ℓ_i をクラスタリングすることで、全てのトラックレット $r_i \in R$ に対する最終的なラベル割当て \hat{L}^* を得る。このとき、クラスタリング手法としては、階層型クラスタリングを用いた。ただし、 k の値は、各追跡によって得られた追跡軌跡の数の平均とした。

4. 実験

4.1 データセット

複数カメラを対象とした既存のデータセットは幾つかある [17], [35]~[39] が、各データセットは、あるカメラ配置の環境中を何人かの人物が移動するという、一つまたは少数のシナリオだけからなる。そこで本研究では、様々なカメラ配置及び移動のシナリオに対する評価を行うため、Kokura ら [9] と同様に、公開デー

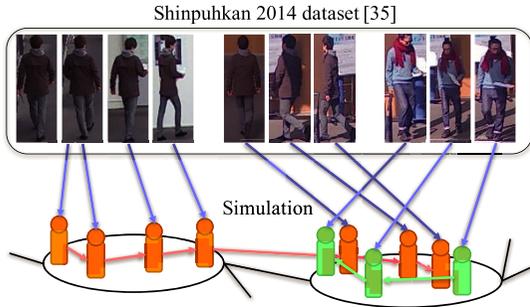


図3 仮想的なデータセットの生成. シミュレーションで人物がカメラ視野に入ると, 対応する人物の画像が Shinpuhkan 2014 dataset [39] から選択される.

Fig. 3 Virtual dataset generation. Once a simulated person is observed from a camera, an image corresponding to the person is sampled from the Shinpuhkan 2014 dataset [39].

タセットである Shinpuhkan 2014 dataset [39] を基に, 仮想的なデータセットを生成して評価を行った. 仮想的なデータセット生成は, 以下の手順で行った.

(1) 2次元平面上にカメラを配置し, 無作為に辺を追加してカメラネットワークを生成する. 各カメラを Shinpuhkan 2014 dataset のカメラに対応付ける.

(2) シミュレーションにより, カメラネットワーク上での歩行者の移動を生成する. 各人物は Shinpuhkan 2014 dataset 内の人物に対応付ける.

(3) 歩行者の移動の生成過程において人物がカメラの視野内に入った際には, Shinpuhkan 2014 dataset から対応する人物の画像を選択する (図3).

カメラネットワーク上を移動する人数は, Shinpuhkan 2014 dataset に含まれる人数と同じ 24 人とし, カメラ台数も同様に 16 台とした. 生成したカメラネットワークの例を図4に示す. 生成されたカメラネットワークには, カメラ間の距離に応じた移動時間パラメータをガンマ分布のパラメータとして設定した. カメラ間の移動に要する時間は, カメラ間の距離に応じて変化し, 最長約5分であった. 一方, カメラ内の移動時間については, 実際に Shinpuhkan 2014 dataset で対応する人物がカメラ内を移動するのに要した時間を用いた. 人物は, 少なくとも7台のカメラ視野内を移動するように, 各人物の移動軌跡を無作為に設定した. また, カメラネットワーク内には最短20秒程度から最大5分程度の間隔で人物が入って来て, 少なくとも7台のカメラ視野内を移動するものとした. この設定では, 人物は約15分間程度カメラネッ

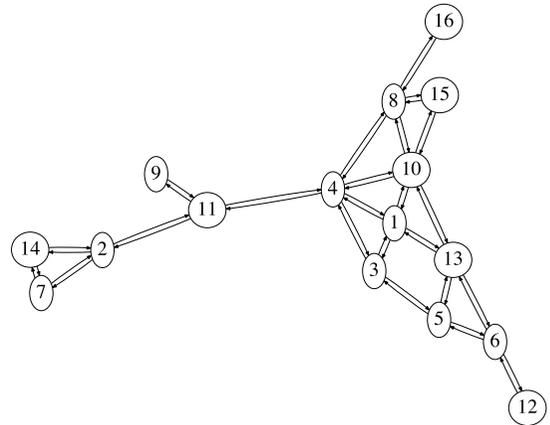


図4 生成したカメラネットワークの例
Fig. 4 Example of a generated camera network.

トワーク上を移動することになる.

4.2 類似度の比較に用いた特徴量

本論文の主眼は画像特徴量の提案ではないため, 本実験では単純に画像ごとの HSV 色空間ヒストグラムを用いた. 人物領域以外の背景の影響を抑えるため, 画像の中央部分 (縦横 1/2) のみを利用し, 照明変化への頑健性を高めるため, HSV 色空間への変換の前に, Adaptive Histogram Equalization [40] を適用した.

二つのトラックレット r_{i_1} と r_{i_2} の類似度 $f(r_{i_1}, r_{i_2})$ は, 画像の類似度 f_{app} と, カメラ間の移動時間のゆう度 f_{temp} の積として以下の式で求めた.

$$f(r_{i_1}, r_{i_2}) = f_{\text{app}}(r_{i_1}, r_{i_2}) f_{\text{temp}}(r_{i_1}, r_{i_2} | \mathbf{e}) \quad (6)$$

ただし, \mathbf{e} はトラックレット r_{i_1} と r_{i_2} が観測された二つのカメラを結ぶカメラネットワークの辺の集合である. また, 二つのトラックレットに含まれる画像の全ての組み合わせに対して HSV のヒストグラム交差で類似度を計算し, その最大値をトラックレット間の類似度とした.

$$f_{\text{app}}(r_{i_1}, r_{i_2}) = \max_{m_1 \in r_{i_1}, m_2 \in r_{i_2}} f_{\text{hist}}(m_1, m_2) \quad (7)$$

一方, カメラ間の移動時間のゆう度は, 各カメラ間の移動時間の分布をガンマ分布で表現し, その確率密度関数の値を移動時間のゆう度とした.

$$f_{\text{temp}}(r_{i_1}, r_{i_2} | \mathbf{e}) = \max_{e_j \in \mathbf{e}} f_{\text{pdf}}(r_{i_1}, r_{i_2} | e_j) \quad (8)$$

ただし, f_{pdf} は, 辺集合 \mathbf{e} の各要素 e_j がもつ移動時

間パラメータで表現されるガンマ分布の確率密度関数であり、形状母数 k_j と尺度母数 θ_j を用い、式 (9) のように表せる。

$$f_{\text{pdf}}(t) = t^{k_j-1} \frac{\exp(-t/\theta_j)}{\Gamma(k_j)\theta_j^{k_j}} \quad (9)$$

ただし、 t は r_{i_1} の最後の画像が観測されてから r_{i_2} の最初の画像が観測されるまでの時間であり、 Γ は階乗の概念を複素数に拡張したガンマ関数である。最終的に画像の類似度と移動時間のゆう度の積を、二つのトラックレット間の類似度とした。

4.3 評価方法

実験では、カメラネットワークの構造や人物の出現時刻が異なる仮想的なデータセットを5個生成して5回の追跡実験を行い、その結果の平均により評価した。

評価指標として修正 Rand 指標 (Adjusted Rand Index; ARI) [41] を用いた。修正 Rand 指標は、二つのラベル割当て結果の間の類似度を評価する指標であり、ラベル割当て結果が一致したときに最大値1となり、ラベル割当てがチャンスレート以下の際には負の値にもなりうる。この指標を用いて、全てのトラックレットに対し、追跡結果に対するラベルの割当てと、正解のラベルの割当ての一致を評価し、多人数追跡の精度とした。

クラスタリング結果のラベル割当てを $X = \{X_1, X_2, \dots, X_m\}$ と $Y = \{Y_1, Y_2, \dots, Y_n\}$ 、 n_{ij} を X_i と Y_j が割り当てられた要素の個数、 n_i と n_j を、それぞれ X_i と Y_j が割り当てられた要素の個数とすると、修正 Rand 指標は、

$$\text{ARI}(X, Y) = \frac{\sum_i \sum_j \binom{n_{ij}}{2} - \left[\sum_i \binom{n_i}{2} \sum_j \binom{n_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2} \right] - \left[\sum_i \binom{n_i}{2} \sum_j \binom{n_j}{2} \right] / \binom{n}{2}} \quad (10)$$

と定義される。ただし、 $\binom{a}{b} = {}_a C_b$ であり、 a 個の中から b 個選択する組合せ数を表す。この指標を用い、追跡結果のラベル割当てと、正解のラベル割当てとの類似度を測り、追跡の精度とした。

提案手法では、無作為にカメラを削除したサブネットワークを生成するため、削除されるカメラによって結果が変わる。そこで、実験1回につき、Trajectory

Ensemble を10回実施し、その修正 Rand 指標の平均を最終的な精度とした。

比較手法として、Trajectory Ensemble を実施せずに、時間順に貪欲法で追跡する手法、3.3 で述べた最小費用流問題として追跡する手法をそのまま元のカメラネットワークに対して適用する手法の二つを採用した。どの手法も、同一の特徴量及び類似度計算方法を用いて評価した。

4.4 結果

五つのデータセットに対し、各手法で追跡を行った結果の平均修正 Rand 指標の平均と分散を表1に示す。提案手法の修正 Rand 指標が最も高くなったことが分かる。また、分散が低くなっていることから、アンサンブル学習の効果である汎化性能の向上が確認できる。Trajectory Ensemble を実施しない最小費用流に基づく追跡結果と比較することで、本論文で提案する Trajectory Ensemble の効果を確認できた。

また、Trajectory Ensemble の安定性を確認するため、5個のデータセットでの追跡結果の各10回ずつの平均と分散を表2に示す。各試行の分散が小さく、安定した追跡を行えていることがわかる。

図5に比較手法の最小費用流に基づく追跡結果(図5(i))と提案手法の Trajectory Ensemble による追跡結果(図5(ii))の例を載せる。比較手法では途中で別の人物へ追跡が切り替わり、追跡に失敗しているのに対し、提案手法では対象人物を継続的に追跡できていることがわかる。

4.5 考察

Trajectory Ensemble の効果を分析するため、提案手法で生成するサブネットワークの個数 S 及び、削除

表1 5回の追跡結果の平均修正 Rand 指標
Table 1 Averaged ARI over five datasets.

手法	平均修正 Rand 指標 (分散)
貪欲法による追跡	0.169 (0.0072)
最小費用流による追跡	0.472 (0.0227)
提案手法による追跡	0.581 (0.0112)

表2 Trajectory Ensemble 各10回の修正 Rand 指標の平均と分散

Table 2 Average and variance of Adjusted Rand Index of 10 trials for each dataset.

	平均	分散
データセット1	0.580	0.00159
データセット2	0.509	0.00026
データセット3	0.672	0.00086
データセット4	0.446	0.00073
データセット5	0.696	0.00024



図 5 追跡結果の例

Fig. 5 Examples of tracking results.

表 3 パラメータに応じた平均修正 Rand 指標の変化
Table 3 Averaged ARIs with different parameters.

		サブネットワーク数 S		
		50	100	150
削除するカメラ台数 C_d	5/16	0.548	0.556	0.556
	7/16	0.573	0.572	0.576
	9/16	0.575	0.581	0.576
	11/16	0.457	0.526	0.578
	13/16	0.314	0.455	0.495

するカメラ台数 C_d を変化させて評価を行った。表 3 より、 $C_d = 9$ 、 $S = 100$ のとき、最も高い修正 Rand 指標が得られた。

C_d を増やしすぎると、追跡に用いられないトラックレットが増加する。その結果、各トラックレットに対して追跡結果のラベルを並べたベクトル ℓ_i の要素に、欠損値が増える。欠損値が多いベクトル同士を式 (4) により比較しても類似度がうまく求まらないため、追跡精度が低下したと考えられる。

一方、生成するサブネットワークの個数 S を増やすと、多様なサブネットワークを生成することができ、アンサンブル学習の効果が現れて精度が向上すると考えられる。一方、 S を増やすと、その分だけ多人数を多カメラ間で追跡処理することになるので、計算量が増加する。そのため、精度と計算量のトレードオフも考慮する必要がある。今後、精度と計算量のトレードオフを考慮した、サブネットワーク数 S の自動的な決定法の開発が望まれる。

5. む す び

カメラネットワーク上での多人数追跡は、カメラの設置環境や向きの違いなどによる人物の一時的な見

えの変化によって、複数カメラにわたる追跡に失敗しうる。この問題に対し、カメラネットワークから無作為にカメラを削除 (Camera Dropout) して生成した複数のサブネットワーク上で多人数追跡を行い、各追跡結果を教師なしアンサンブル学習の考え方で統合することにより追跡の途切れに対処する、Trajectory Ensemble を提案した。実験により、半数以上のカメラを削除したサブネットワーク上での追跡を統合することで、高精度に追跡ができることを確認した。

今後の展望として、提案手法のオンラインアルゴリズム化や、他のマッチング問題への応用が考えられる。

謝辞 本研究の一部は、科学研究費補助金による。

文 献

- [1] D. Makris, T. Ellis, and J. Black, "Bridging the gaps between cameras," Proc. 2004 IEEE Conf. Comput. Vision and Pattern Recognit., pp.205–210, June 2004.
- [2] B. Song and A.K. Roy-Chowdhury, "Robust tracking in a camera network: A multi-objective optimization framework," J. Sel. Top. Signal Process., vol.2, no.4, pp.582–596, Sept. 2008.
- [3] A. Alahi, P. Vanderghenst, M. Bierlaire, and M. Kunt, "Cascade of descriptors to detect and track objects across any network of cameras," J. Comput. Vis. Image Understand., vol.114, no.6, pp.624–640, June 2010.
- [4] K. Chen, C. Huang, S. Hsu, and I. Chang, "Multiple objects tracking across multiple non-overlapped views," Proc. 5th Pacific-Rim Symp. Image and Video Tech., pp.128–140, Nov. 2011.
- [5] G. Lian, J. Lai, and W.-S. Zheng, "Spatial-temporal consistent labeling of tracked pedestrians across non-overlapping camera views," Pattern Recognit., vol.44, no.5, pp.1121–1136, May 2011.
- [6] A. Alahi, V. Ramanathan, and L. Fei-Fei, "Socially-aware large-scale crowd forecasting," Proc. 2014 IEEE Conf. Comput. Vision and Pattern Recognit., pp.2211–2218, June 2014.
- [7] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," Proc. 2008 IEEE Conf. Comput. Vision and Pattern Recognit., pp.1–8, June 2008.
- [8] Y. Kawanishi, D. Deguchi, I. Ide, and H. Murase, "Trajectory ensemble: Multiple persons consensus tracking across non-overlapping multiple cameras over randomly dropped camera networks," Proc. 2017 IEEE Conf. Comput. Vision and Pattern Recognit. Workshops, pp.56–62, July 2017.
- [9] T. Kokura, Y. Kawanishi, M. Mukunoki, and M. Minoh, "Tracking pedestrians across multiple cameras via partial relaxation of spatio-temporal constraint and utilization of route cue," Proc. 12th Asian

- Conf. Comput. Vision Workshops, pp.587–601, Nov. 2014.
- [10] 井尻善久, 川西康友, 美濃導彦, 村瀬 洋, “視野を共有しない複数カメラ間での人物照合,” 信学技報, PRMU2011-119, Nov. 2011.
- [11] 川西康友, 伍 洋, 椋木雅之, 美濃導彦, 勞 世竈, “サーベイランスカメラ映像を対象とした人物照合の高精度化への取り組み,” 信学技報, PRMU2013-85, Dec. 2013.
- [12] D. Gray, S. Brennan, and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” Proc. 2007 Intl. Workshop on Performance Evaluation for Tracking and Surveillance, vol.3, pp.41–47, Oct. 2007.
- [13] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person re-identification by symmetry-driven accumulation of local features,” Proc. 2010 IEEE Conf. Comput. Vision and Pattern Recognit., pp.2360–2367, June 2010.
- [14] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, “Boosted human re-identification using Riemannian manifolds,” Image and Vision Comput., vol.30, no.6–7, pp.443–452, June 2012.
- [15] Y. Wu, M. Mukunoki, T. Funatomi, M. Minoh, and S. Lao, “Optimizing mean reciprocal rank for person re-identification,” Proc. 8th IEEE Intl. Conf. Adv. Video and Signal Based Surveillance, pp.408–413, Aug. 2011.
- [16] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao, “Set based discriminative ranking for recognition,” Proc. 12th European Conf. Comput. Vision, pp.497–510, Oct. 2012.
- [17] W. Li, R. Zhao, T. Xiao, and X. Wang, “Deep-ReID: Deep filter pairing neural network for person re-identification,” Proc. 2014 IEEE Conf. Comput. Vision and Pattern Recognit., pp.152–159, June 2014.
- [18] R. Zhao, W. Ouyang, and X. Wang, “Learning mid-level filters for person re-identification,” Proc. 2014 IEEE Conf. Comput. Vision and Pattern Recognit., pp.144–151, June 2014.
- [19] E. Ahmed, M. Jones, and T.K. Marks, “An improved deep learning architecture for person re-identification,” Proc. 2015 IEEE Conf. Comput. Vision and Pattern Recognit., pp.3908–3916, June 2015.
- [20] S. Bak and P. Carr, “Deep spatial pyramid for person re-identification,” Proc. 14th IEEE Intl. Conf. Adv. Video and Signal Based Surveillance, no.46, pp.1–6, Aug. 2017.
- [21] V. Kettner and R. Zabih, “Bayesian multi-camera surveillance,” Proc. 1999 IEEE Conf. Comput. Vision and Pattern Recognit., pp.253–259, June 1999.
- [22] F. Porikli and A. Divakaran, “Multi-camera calibration, object tracking and query generation,” Proc. 2003 IEEE Intl. Conf. Multimedia and Expo, vol.1, pp.653–656, July 2003.
- [23] O. Javed, K. Shafique, and M. Shah, “Appearance modeling for tracking in multiple non-overlapping cameras,” Proc. 2005 IEEE Conf. Comput. Vision and Pattern Recognit., pp.26–33, June 2005.
- [24] K. Tieu, G. Dalley, and W.L. Grimson, “Inference of non-overlapping camera network topology by measuring statistical dependence,” Proc. 10th IEEE Intl. Conf. Comput. Vision, pp.1842–1849, Oct. 2005.
- [25] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, “Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views,” J. Comput. Vis. Image Understand., vol.109, no.2, pp.146–162, Feb. 2008.
- [26] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, Network flows: Theory, algorithms, and applications, Prentice Hall, 1993.
- [27] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, “Multiple object tracking using k-shortest paths optimization,” IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.9, pp.1806–1819, 2011.
- [28] J.K. Wolf, A.M. Viterbi, and G.S. Dixon, “Finding the best set of k paths through a trellis with application to multitarget tracking,” IEEE Trans. Aerosp. Electron. Syst., vol.25, no.2, pp.287–296, March 1989.
- [29] A.R. Zamir, A. Dehghan, and M. Shah, “GMCP-tracker: Global multi-object tracking using generalized minimum clique graphs,” Proc. 12th European Conf. Comput. Vision, pp.343–356, Oct. 2012.
- [30] R.M. Karp, “Reducibility among combinatorial problems,” Proc. 1972 Symposium on the Complexity of Computer Computations, pp.85–103, March 1972.
- [31] A. Strehl and J. Ghosh, “Cluster ensembles—A knowledge reuse framework for combining multiple partitions,” J. Machine Learning Res., vol.3, no.35, pp.583–617, March 2003.
- [32] R. Tokola, W. Choi, and S. Savarese, “Breaking the chain: Liberation from the temporal Markov assumption for tracking human poses,” Proc. 14th IEEE Intl. Conf. Comput. Vision, pp.2424–2431, Dec. 2013.
- [33] Y. Xu, L. Qin, and Q. Huang, “Coupling multiple alignments and re-ranking for low-latency online multi-target tracking,” Computer Vision—ACCV 2014: 12th Asian Conf. Comput. Vision, Revised Selected Papers, part V, pp.534–549, Nov. 2015.
- [34] R. Kumar and D. Batra, “Pose tracking by efficiently exploiting global features,” Proc. 2016 IEEE Winter Conf. Applicat. Comput. Vision, pp.1–9, March 2016.
- [35] D. Baltieri, R. Vezzani, and R. Cucchiara, “3DPes: 3D people dataset for surveillance and forensics,” Proc. 2011 Intl. ACM Workshop on Multimedia Access to 3D Human Objects, pp.59–64, Nov. 2011.
- [36] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey, “A database for person re-identification

- in multi-camera surveillance networks,” Proc. 2012 Intl. Conf. Digital Image Computing: Techniques and Applications, pp.1–8, Dec. 2012.
- [37] J. Per, V.S. Kenk, R. Mandeljc, M. Kristan, and S. Kovacic, “Dana36: A multi-camera image dataset for object identification in surveillance scenarios,” Proc. 9th IEEE Intl. Conf. Adv. Video and Signal-Based Surveillance, pp.64–69, Sept. 2012.
- [38] W. Li and X. Wang, “Locally aligned feature transforms across views,” Proc. 2013 IEEE Conf. Comput. Vision and Pattern Recognit., pp.3594–3601, June 2013.
- [39] Y. Kawanishi, Y. Wu, M. Mukunoki, and M. Minoh, “Shinpuhkan2014: A multi-camera pedestrian dataset for tracking people across multiple cameras,” Proc. 20th Korea-Japan Joint Workshop on Frontiers of Comput. Vision, pp.1–5, Feb. 2014.
- [40] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.T.H. Romeny, and J.B. Zimmerman, “Adaptive histogram equalization and its variations,” Comput. Vision Graph. Image Process., vol.39, no.3, pp.355–368, Sept. 1987.
- [41] J.M. Santos and M. Embrechts, “On the use of the adjusted Rand index as a metric for evaluating supervised classification,” Proc. 19th Intl. Conf. Artificial Neural Netw., pp.175–184, Sept. 2009.

(平成 29 年 10 月 8 日受付, 30 年 2 月 1 日再受付,
5 月 2 日早期公開)



川西 康友 (正員)

平 18 京大・工・情報卒。平 24 同大学院情報学研究科博士後期課程修了。博士(情報学)。平 24 同大学術情報メディアセンター特定研究員。平 26 名大未来社会創造機構特任助教。平 27 同大情報科学研究科助教。平 29 同大情報科学研究科助教。現在に至る。防犯カメラ・車載カメラ映像を対象とした、人物検出・追跡・検索を含む人物画像処理に関する研究に従事。平 23 年度 PRMU 研究奨励賞受賞。IEEE ITS Society Nagoya Chapter Young Researcher Award 受賞。IEEE 会員。



出口 大輔 (正員)

平 13 名大・工・情報卒。平 18 同大学院情報科学研究科博士後期課程修了。博士(情報科学)。平 16~18 まで日本学術振興会特別研究員。平 18 名大大学院情報科学研究科研究員, 平 18 名大大学院工学研究科研究員, 平 20~24 まで同大学院情報科学研究科助教, 平 24 より同大情報連携統括本部情報戦略室准教授。現在に至る。主に画像処理・パターン認識技術の開発とその ITS 及び医用応用に関する研究に従事。CARS2004 Poster Award, CADM2004 大会賞, 平 18 日本医用画像学会奨励賞, 平 18 日本コンピュータ外科学会講演論文賞。IEEE 会員。



井手 一郎 (正員:シニア会員)

平 6 東大・工・電子卒。平 8 同大学院工学系研究科情報工学専攻修士課程了。平 12 同研究科電気工学専攻博士課程了。博士(工学)。同年国立情報学研究所助手。平 16 名古屋大学大学院情報科学研究科助教, 平 19 より准教授。平 29 同大学院情報科学研究科准教授, この間, 平 14~16 総合研究大学院大学数物科学研究科助手併任, 平 16~22 情報・システム研究機構国立情報学研究所客員助教・准教授兼任, 平 17, 18, 19 フランス情報学・統計システム研究所(IRISA) 招聘教授。平 22~23 オランダアムステルダム大学情報学研究所上級訪問研究員。パターン認識技術の実応用や映像メディア処理全般に興味をもっている。情報処理学会シニア会員, 映像情報メディア学会, 人工知能学会, IEEE, ACM 各会員。



村瀬 洋 (正員:フェロー)

昭 53 名大・工・電気卒。昭 55 同大学院修士課程了。同年日本電信電話公社(現 NTT) 入社。平 4 から 1 年間米国コロンビア大客員研究員。平 15 から名古屋大学大学院情報科学研究科教授。平 29 から同大学院情報科学研究科教授。現在に至る。文字・図形認識, コンピュータビジョン, マルチメディア認識の研究に従事。工博。昭 60 本会学術奨励賞, 平 6 IEEE-CVPR 最優秀論文賞, 平 7 情報処理学会山下記念研究賞, 平 8 IEEE-ICRA 最優秀ビデオ賞, 平 13 高柳記念奨励賞, 平 13 本会ソサエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, 平 22 前島密賞, 平成 24 紫綬褒章, 他受賞。IEEE フェロー, 情報処理学会会員。