

# 特別研究報告書

## 時空間制約の選択的緩和と 経路制約の導入による 多人数多カメラ間対応付け

指導教員 美濃 導彦 教授

京都大学工学部情報学科

小倉 暢

平成 26 年 1 月 31 日

## 時空間制約の選択的緩和と経路制約の導入による 多人数多カメラ間対応付け

小倉 暢

### 内容梗概

人物ごとの訪れた地点と時刻の記録である移動履歴の獲得を、多数の人物に対して行うことができれば、商業施設では店舗や商品の配置の改善や類似スポットの発見などに有用な情報が得られる。この移動履歴の獲得は、人手で行うには大変な時間と労力と費用が求められるため、自動化できることが望ましい。

そこで本研究では、一般に商業施設に設置されるような、既存の固定カメラで撮影された画像を用いて自動で移動履歴を獲得することを考える。これらの既存のカメラは、複数個が互いに視野の重複の無いように置かれているものとする。そのため、これらのカメラを用いて移動履歴を推定するには、視野同士の間にも他カメラの視野を経由せずに移動を行えるような経路が存在する隣接カメラ間で、撮影された人物の対応付けを行う必要がある。対応付けが全て正しく行われれば、各人物に対してその撮影時刻と撮影場所を時刻の昇順に参照することにより、人物ごとの移動履歴を推定することができる。

従来の人物対応付けでは、隣接する2カメラ間で、人物のカメラ視野間の移動時間と画像の特徴量という2つの情報をもとに類似度の計算を行い、類似度の高いものを対応付けする、という処理が行われている。この従来手法では、隣接するカメラ視野間を移動する時間は人物によって大きく変化せず、またカメラ視野間で人物の見えは大きく変化しない、という仮定が置かれている。さらに、あるカメラの視野から隣接するカメラの視野への移動は必ずある時間以上かつある時間以内の時間で行われる、という時空間制約を導入し、対応付け相手の候補を絞り、対応付け精度を高めているものもある。しかし、この従来の人物対応付けには、次の2つの問題が存在する。

第1に、観測される人物がカメラ視野間での立ち寄りや立ち止まり、即ち滞留を行うと、移動時間が大きく変化しうる。この滞留は、例えばカメラ視野間に小売店、便所、休憩所、看板、展示品などの、人物が立ち寄るべき箇所が存在した場合に発生しうる。その結果、時空間制約を用いて対応付け相手の候補を絞ることが難しくなる、という問題が発生する。

第2に、ごく単純な構造の通路を除いては、必ず人物を一定の方向から撮影

できるようにカメラを配置することはできないため、カメラごとに人物の撮影方向は変わりうる。加えて、屋内外での日照の有無、天候の変化、電燈の光の明るさや色相の違い、といった要因により、カメラごとに照明条件が変わりうる。これらの原因により、カメラ間で人物の見えが変化する問題が生じる。その結果、同一人物であっても外見の特徴を用いた対応付けが難しくなる。

以上の2つの問題に対し、提案手法では次の2つの方法で対処する。第1に、移動速度が変化することによって移動時間を用いた対応付けが難しくなる問題には、時空間制約を選択的に緩和し、画像の特徴量を重視した人物対応付けを行うことで対処する。これによって、当該人物が滞留を行ったか否かに係らず正しく対応付けができる。第2に、見えが変化することによって画像の特徴量を用いた対応付けが難しくなる問題には、経路制約を導入した対応付けを行い、他のカメラの情報を用いて対応付け相手の候補を限定して対応付け精度を上げる。ここに、経路制約とは、ある人物が異なる任意の2つのカメラで観測されているときに、その2つのカメラ視野を結ぶ経路上に存在する別のカメラでも必ずその人物が観測されているはずである、という制約である。この制約は、人物対応付けの際に参照する隣接する2カメラとは別の、対象の人物が撮影されているカメラの情報も参照することで、従来の時空間制約をさらに強力にしたものである。これによって、対応付け相手の存在する区間を限定して、外見の似た別人を対応付け相手候補から除外し、隣接2カメラ間で人物の見えが変化する問題に対処する。

固定カメラ映像を用いた人物対応付けにおける提案手法の有効性を確認するために、商業施設に設置された固定カメラの映像に対して移動履歴の推定を行い、人物対応付け精度の従来手法との差異を評価した。人物の滞留の発生率と見えの変化の発生率の2条件を変化させた多数のデータセットに対して提案手法を適用した結果、人物対応付けの精度が最大で13.1%上昇するという結果を得た。また、その他の全ての条件の組み合わせにおいて、提案手法を用いた人物対応付けの精度が従来手法の精度を上回り、提案手法の有効性が確認できた。

今後の課題としては、実環境での人物対応付けを行って精度を評価すること、時空間制約の緩和の程度の設定基準を検討すること、などが挙げられる。

# Multiple People Tracking across Multiple Views via Partial Relaxation of Spatio-Temporal Cue and Utilization of Route Cue

Toru KOKURA

## Abstract

In shopping malls, it has been beneficial to find similar spots and reallocate the positions of stores for sales growth. For that purpose, “movement histories” of many pedestrians can be valuable information. Movement history is a sequence of locations and their times for a person. Acquiring movement histories manually requires enormous time, labor and cost. So, estimating them automatically is desired.

We present a method for estimating the movement histories by utilizing videos taken by existing fixed cameras. We suppose that these cameras are often installed with their views non-overlapping. Therefore, in order to estimate the moving histories, we need to re-identify pedestrians across adjacent views by matching tracklets which are obtained by tracking pedestrians in each view. Here, adjacent views are the camera views which are connected by the direct path which is not crossing by other camera views. When re-identifications between all adjacent views are correctly achieved, moving history for each pedestrian can be estimated by traversing the camera views in which the person was observed in time order.

We use two kinds of information: traveling time between two adjacent views and appearance features of a tracklet has been used. Two tracklets are matched when the similarity calculated between the tracklets with the two information is bigger than a given threshold. This approach is based on the hypothesis that the traveling time between a pair of adjacent views varies little among the pedestrians, and that every pedestrian’s appearance varies little among the cameras. Additionally, some existing methods introduce the spatio-temporal cue, which reflects that traveling times between a pair of adjacent views can be bounded by a given time span. This cue reduces matching candidates and increases matching accuracy. However, these methods has following two drawbacks.

Firstly, when an observed pedestrian significantly delayed between a pair of

adjacent views, the traveling time is out of the given time span. A delay often happens when there are visiting places between the adjacent views, for example, stores, toilets, rest spaces, signboards, and exhibitions. This makes it difficult to reduce matching candidates by the spatio-temporal cue.

Secondly, except very simple environment such as a straight road, it is sometimes difficult to install cameras to observe pedestrians from the same direction. Thus, each pedestrian’s appearance varies depending on each camera. Additionally, lighting condition varies depending on observation time because of weather, light’s intensity and the existence of sunlight. These factors cause appearance change in camera views. This makes it difficult to match tracklets by appearance features.

The proposed method deals with these problems via following two ideas. Firstly, we tackle the traveling time variation by selectively relaxing the spatio-temporal cue in matching tracklets. This makes it possible to match people correctly regardless the occurrence of the delay. Secondly, we deal with the appearance variation by introducing a route cue. The route cue is a constraint that when a pedestrian is observed by different two cameras, the pedestrian should be observed by other cameras which exists on a route between the two cameras. This constraint is an enhanced version of the spatio-temporal cue. This constraint contributes to reducing matching candidates by excluding different pedestrians with similar appearance from matching candidates.

To show the effectiveness of the proposed method, we evaluate the method on simulated datasets generated from a public data set collected by 16 cameras in a shopping mall. The dataset includes several settings of camera adjacency and the number of pedestrians who delay. Under each setting, we conducted an experiment to estimate movement histories. The result shows that matching accuracy of the proposed method was higher than that of the existing method under any considered settings. It also shows the accuracy of the proposed method surpassed that of the existing method by 13.1 % at the maximum. These experiments prove that the proposed method can match tracklets in higher accuracy.

# 時空間制約の選択的緩和と経路制約の導入による 多人数多カメラ間対応付け

## 目次

第 1 章	はじめに	1
第 2 章	多人数多カメラ間人物対応付け	3
2.1	カメラ間人物対応付けの関連研究	3
2.2	多人数多カメラ間人物対応付け問題の定式化	4
2.3	多人数多カメラ間対応付けの従来手法	6
2.4	従来手法の問題点	8
第 3 章	時空間制約の選択的緩和と経路制約を用いた人物対応付け	11
3.1	時空間制約の選択的緩和と経路制約の導入	11
3.2	対応付け手順	13
3.2.1	手順 1: 強い時空間制約と外見制約を用いたラベル割り当て	15
3.2.2	手順 2: 時空間制約を緩和したラベル割り当て	16
3.2.3	手順 3: 経路制約を用いたラベル割り当て	16
3.2.4	手順 4: 局所的最適化によるラベル割り当て	17
第 4 章	実験・結果	17
4.1	利用データ	17
4.2	実験用データセットの生成	18
4.2.1	カメラネットワーク $G$ の生成法	18
4.2.2	トラックレット集合 $R$ の生成法	18
4.3	実験方法	19
4.3.1	実験 A: 滞留発生率の変化に対する対応付け精度の測定	20
4.3.2	実験 B: 日向のカメラの割合の変化に対する対応付け精度の測定	21
4.4	評価方法	21
4.5	実験結果	22
4.5.1	実験 A: 滞留発生率の変化に対する対応付け精度の測定	22

4.5.2	実験 B：日向のカメラの割合の変化に対する対応付け 精度の測定.....	23
<b>第 5 章</b>	<b>結論</b>	<b>24</b>
	謝辞	25
	参考文献	25

## 第1章 はじめに

近年，商業施設では店舗や商品の配置の改善や類似スポットの発見などが売上向上のために注目されている．これらを行う際に，人物ごとの訪れた地点と時刻の記録である**移動履歴**の獲得を，多数の人物に対して行うことができれば，有用な情報となる．この移動履歴を獲得するには目視や聞き取りを行う必要があり，大変な時間と労力と費用が求められるため，自動化できることが望ましい．

自動で人物の移動履歴を推定するには，人物に ID タグを配布して各地点で ID 情報を読み取る，人物の持つ携帯電話の発する電波を観測して位置情報を取得する，といった方法が考えられる．しかしこれらの方法は，いずれも特別な装置を必要としたり，多数の顧客への働きかけを要したりする．

一方で，既存のセンサを他の用途にも活用するセンシングウェブ [1] という概念がある．また，既存の防犯カメラから得られる画像を用いて人物の検出や追跡を行う研究が，コンピュータービジョンの分野で盛んに行われ，例としてカメラ視野内に進入した不審人物を検知する研究 (武田ら [2])，人物の頭部を追跡する研究 (小林ら [3])，道路上での複数カメラによる人物追跡と監視を行う研究 (西沢ら [4])，人物を検索する研究 (井関ら)，などがなされてきた．これらの研究と同様に既存のカメラから得られた人物画像を用い，人物の**対応付け**を行えば，人物の移動履歴を推定することができる．ここに，対応付けとは，複数のカメラから得られた人物画像が同一人物のものであると判定することである．対応付けが全て正しく行われれば，各人物に対してその撮影時刻と撮影場所を時刻の昇順に参照することにより，人物ごとの移動履歴を推定することができる．

本研究では，複数のカメラで撮影された各人物の対応付けをする**多人数多カメラ間対応付け**を行うことを考える．用いるカメラは，各々が視野重複の無いように配置されているものとする．通常，カメラの視野同士の間には，視野から視野への移動を行えるような経路が存在し，この経路はカメラで観測することができない．この経路のうち，カメラでの観測ができない部分を，本稿では**ブラインドエリア**と呼ぶ．また，本稿では，ある2つのカメラの視野を別のカメラの視野を経由せずに連結するような経路が存在する場合，その2つのカメラは**隣接**している，と表現する．

従来の人物対応付けでは，隣接する2カメラ間で，人物の画像の特徴量とカメラ視野間の移動時間という2つの情報をもとに類似度の計算を行い，類似度



の高いものを同一人物とする，という処理が行われている．この従来手法では，ブラインドエリア内では人物の移動速度と見えは大きく変化しないという仮定が置かれている．さらに，あるカメラの視野から隣接するカメラの視野への移動は必ずある時間以上かつある時間以内の時間で行われる，という**時空間制約**を導入し，対応付け相手の候補を絞り，対応付け精度を高めているものもある．しかし，この従来の人物対応付けには，次の2つの問題が存在する．

第1に，観測される人物がカメラ視野間での立ち寄りや立ち止まり，即ち滞留を行うと，移動時間が大きく変化しうる．この滞留は，例えばカメラ視野間に小売店，便所，休憩所，看板，展示品などの，人物が立ち寄るべき箇所が存在した場合に発生しうる．その結果，人物のカメラ視野間の移動時間が時空間制約の定める最長の移動時間を上回り，時空間制約を用いて対応付け相手の候補を絞ることが難しくなる，という問題が発生する．

第2に，ごく単純な構造の通路を除いては，必ず人物を一定の方向から撮影できるようにカメラを配置することはできないため，カメラごとに人物の撮影方向は変わりうる．加えて，屋内と屋外の差，天候の変化，電燈の光の明るさや色相の違い，といった要因により，カメラごとに照明条件が変わりうる．これらの原因により，カメラ間で人物の見えの変化が生じ，同一人物であっても外見の特徴を用いた対応付けが難しくなる，という問題が発生する．

以上の2つの問題に対し，提案手法では次の2つの方法で対処する．第1に，移動速度が変化する問題には，移動速度の変化が発生した人物に対して選択的に時空間制約を緩和した人物対応付けを行うことで対処する．これによって，当該人物が滞留を行ったか否かに係らず正しく対応付けができる．

第2に，見えが変化する問題には，**経路制約**を導入した対応付けを行い，他のカメラの情報を用いて対応付け相手の候補を限定して対応付け精度を上げる．ここに，経路制約とは，ある人物が異なる任意の2つのカメラで観測されているときに，その2つのカメラ視野を結ぶ経路上に存在する別のカメラでも必ずその人物が観測されているはずである，という制約である．この制約は，人物対応付けの際に参照する隣接する2カメラとは別の，対象の人物が撮影されているカメラの情報も参照することで，時空間制約をさらに強力にしたものである．これによって，対応付け相手の存在する区間を限定して，外見の似た別人を対応付け相手候補から除外し，隣接2カメラ間で人物の見えが変化する問題に対処する．

本稿の構成は以下のとおりである。まず2章では、多人数多カメラ間対応付けの既存手法を整理し、問題設定を明確化したうえで既存手法の問題点を述べる。次に3章では、本研究で提案する、時空間制約の選択的緩和と経路制約を用いた人物対応付けの内容について述べる。そして4章では、時空間制約の選択的緩和と経路制約を用いた人物対応付けにより、従来手法よりも対応付け精度が向上することを実験で示す。最後に5章では、本研究の結論と今後の課題を述べる。

## 第2章 多人数多カメラ間人物対応付け

### 2.1 カメラ間人物対応付けの関連研究

多人数多カメラ間対応付けを行うには、ある人物画像の組がともに同一人物のものであるか否かを判定する、人物の照合 [5] を、任意の人物画像の組に対して実行できればよい。人物の照合は、同一人物の外見は類似しているとの仮定のもと、人物の服の色や模様の特徴である外見特徴を用いて行われる。中でも Farenzena ら [6] は、服装の色と模様の情報を統合した人物照合に有効な特徴量を提案している。

これらの人物照合の手法を用いてカメラ間人物対応付けを行う研究として、Javed ら [7] は、時空間制約と時空間特徴を組み合わせた2カメラ間人物対応付けの手法を提案している。ここに、時空間制約とは、人物の遷移は必ず隣接するカメラ視野に対して最小遷移時間以上最大遷移時間以下の時間で行われるという制約であり、時空間特徴とはカメラ視野間遷移時間の尤度である。この手法では、外見特徴の類似度と、遷移時間の尤もらしさの積が最大となる人物同士を同一人物として対応付けることにより、照合精度を向上させている。ブラインドエリアを徒歩で移動する際の所要時間の尤度は、各ブラインドエリアに対してテストデータを用いてそれぞれ予め学習させている。

一方で、カメラに写る複数人物を同時に照合することにより照合精度を向上させる多人数2カメラ間対応付けの研究がおこなわれてきた。Liana ら [8] は、2カメラ間での人物対応付け問題を組み合わせ最適化問題として解く手法を提案している。上記の手法では、2カメラ間での人物対応付けを行う際に、カメラに映る複数の人物の対応付けを同時に最適化することにより、人物対応付けの精度を向上させている。

多人数多カメラ間人物対応付けを行うには、2カメラ間人物対応付けを、隣接する全てのカメラの組み合わせに対して行えば実現できる。しかし、対応付けの際に特定の2カメラの情報しか用いないよりも、それ以外のカメラの情報も用いた方が、対応付けの際に使用できる情報が増加し、対応付け精度を向上させることができる。複数のカメラの情報をを用いてさらに対応付け精度を向上させる、多人数多カメラ間対応付けの研究として、Songら[9]は複数のカメラ視野にわたる人物の移動履歴を獲得する手法を提案している。Songらの手法では、人物の移動履歴を獲得する問題は、各撮影記録に対して人物ラベルを付与する、ラベル割り当て問題であるとし、その解から各人の遷移履歴を獲得している。

## 2.2 多人数多カメラ間人物対応付け問題の定式化

本節では、2.1節で述べた関連研究を踏まえ、問題を定式化する。まず、カメラネットワークの定義と人物の移動のモデル化について述べ、次に対応付けの入力と出力について詳述する。

### カメラネットワークの定義

本研究では視野重複の無い複数のカメラ  $c_1, c_2, \dots, c_{|V|}$  を協調させたシステムを用いて人物を観測するという状況を仮定する。この複数のカメラの視野を頂点とし、ブラインドエリアを辺とするグラフを、カメラネットワーク  $G = (V, E)$  として表す。ここに、 $V$  は頂点集合であり、各頂点  $v_a$  がカメラ  $c_b$  の視野に対応している。また、 $E$  は辺集合であり、カメラ  $c_a$  と  $c_b$  ( $a \neq b$  かつ  $a, b \in [1, |V|]$ ) の視野間に、他のカメラ視野を通過せずに移動できる経路がある場合には  $(v_a, v_b) \in E$  であり、そうでない場合には  $(v_a, v_b) \notin E$  となっている。

### 人物の移動モデル

本稿では、各人物の移動を、 $G$  内の頂点から頂点への辺を通っての移動としてモデル化する。この際、人物の移動時間は辺ごとに決められた確率分布に従うものとする。即ち、ある人物が、ある頂点  $v_a$  から  $v_b$  へ辺  $(v_a, v_b)$  を通って滞留を行わずに移動する際の移動時間  $t$  の分布はある関数  $\zeta_{c_a c_b}(t) : \mathbb{R} \rightarrow [0, 1]$  で表される。この関数は確率密度関数であるので、 $\int_{-\infty}^{\infty} \zeta_{c_a c_b}(t) dt = 1$  が成り立つ。また、移動時間が負の値になることは有り得ないため、 $t < 0 \Rightarrow \zeta_{c_a c_b}(t) = 0$ 、という性質を持つものとする。

カメラ視野間の移動時には、人物はある確率  $\beta$  で滞留を行い、その際の移動

時間はある滞留時最短移動時間  $\tau_{min}$  以上かつある滞留時最長移動時間  $\tau_{max}$  以内となるものとする. この滞留発生時の移動時間を  $t$  とすると,  $t$  の分布を表す関数を  $\eta_{c_a c_b}(t) : \mathbb{R} \rightarrow [0, 1]$  とする. ここに, 滞留発生時の移動時間は, 発生時間の予測が困難な事象が発生する時間を表現するのに用いられることの多い,  $[\tau_{min}, \tau_{max}]$  の区間で定義される一様分布で決定される値になると定義する. 従って,  $\eta_{c_a c_b}(t)$  は以下の関数で与えられる.

$$\eta_{c_a c_b}(t) = \begin{cases} \frac{1}{\tau_{max} - \tau_{min}} & (\tau_{min} \leq t \leq \tau_{max}) \\ 0 & (0 < t < \tau_{min}, \tau_{max} < t) \end{cases} \quad (1)$$

以上の2つの関数のうち, 確率  $\beta$  で  $\eta_{c_a c_b}(t)$  が, 確率  $1 - \beta$  で  $\zeta_{c_a c_b}(t)$  が, 人物がカメラネットワークの頂点間を移動するときに選択され, 移動時間の分布は選ばれた関数に従う.

#### 対応付け問題に対する入力

人物がカメラの視野に進入すると, カメラで撮影された映像の各フレームの画像から, 人物検出とカメラ内人物追跡により, ある人物があるカメラに一度写り始めてから写り終わるまでの人物画像列を得ることができる. この人物画像列は, 当該人物の映っている領域のみが外接矩形で切り出されたものであるとする. ここに, 観測開始から  $i$  番目に得られた人物画像列を  $I_i$  とし,  $I_i$  に含まれる  $1, 2, \dots, n_i$  番目の画像を  $I_i^1, I_i^2, \dots, I_i^{n_i}$  ( $n_i$  は  $I_i$  に含まれる画像枚数) とする. そして, これを各カメラ映像に対して行うことで, 追跡系列の人物画像列の集合  $I = \{I_i\}_{i=1}^N$  が得られる. ここに,  $N$  は得られた人物画像列の総数である.

$\{I_i\}_{i=1}^N$  の中から, 任意の  $I_i$  と  $I_j$  ( $i \neq j$ ) の組に対して互いに同一人物の画像列であるかどうかを判定できれば, 全人物の移動履歴を推定できる. この対応付けを行うために, 本稿ではトラックレット  $r_i = (\mathbf{f}_i, s_i, e_i, c_i, v_i)$  を定義する. ここに,  $\mathbf{f}_i$  は人物画像の特徴量列,  $s_i$  は  $I_i$  の観測開始時刻,  $e_i$  は観測終了時刻,  $c_i \in \mathcal{C} = \{c_j\}_{j=1}^{|V|}$  は撮影カメラ,  $v_i \in \{v_j\}_{j=1}^{|V|}$  は  $c_i$  に対応するカメラネットワークの頂点を表す. また,  $\mathbf{f}_i = \{\mathbf{f}_i^m\}_{m=1}^{n_i}$  は特徴量  $\mathbf{f}_i^m$  の列として表され, 各特徴量  $\mathbf{f}_i^m$  は人物画像  $I_i^m$  に対して特徴抽出を行うことで得られる.

以上より, 全てのカメラから得られたトラックレットの集合  $R = \{r_i\}_{i=1}^N$  を問題を解く際の入力データとする. 但し,  $R$  内の各トラックレットは観測開始時刻について昇順に整理されているものとする.

## 対応付け問題に対する出力

集合  $R$  内から，同一人物のものであるあらゆるトラックレットの組  $(r_i, r_j)$  を対応付けしたとき，その結果は行列  $M = (m_{ij})$  の形で表される．ここに， $M$  は  $N$  行  $N$  列行列であり，各成分はトラックレットの組  $(r_i, r_j)$  が対応付けられているなら  $m_{ij} = 1$ ，対応付けられていないなら  $m_{ij} = 0$  となっている．

この対応付け結果から，同一人物のトラックレットには同一のラベルが割り振られ，別人物のトラックレットには別のラベルが割り振られる，というようなラベル割り当て  $\mathbf{h}$  が得られれば，トラックレットをラベルの種類ごとに集計することで，人物の移動履歴が得られる．このラベル割り当て  $\mathbf{h}$  は，ベクトル  $\mathbf{h} = (h^{(1)}, h^{(2)}, \dots, h^{(N)}) \in \mathbb{N}^N$  の形で表すことができる．また， $\mathbf{h}$  に使用されているラベルの種類が  $k (= |\{q | \exists p, h^{(p)} = q\}|)$  種類であったとき， $\forall p, h^{(p)} \in [1, k]$  であるとする．この  $\mathbf{h}$  は，以下の性質を満たすように  $M$  から作成することができる．

$$\forall i, j, m_{ij} = 1 \Rightarrow h^{(i)} = h^{(j)} \quad (2)$$

$$\forall i, j, m_{ij} = 0 \Rightarrow h^{(i)} \neq h^{(j)} \quad (3)$$

$\mathbf{h}$  から人物ごとの移動履歴を得るには，同じラベルが割り当てられてるトラックレットの添字の集合  $G_q = \{i | h^{(i)} = q\}$  を  $q = 1, 2, \dots, \max_i h^{(i)}$  について作成すればよい．これにより，ラベル  $q$  に対応する人物が  $n$  回目に観測されたのはカメラ  $c_{qn}$  であり，時刻  $s_{qn}$  から時刻  $e_{qn}$  までそのカメラ視野に存在した，という情報をあらゆる  $q, n$  について求めた移動履歴  $\{(c_{qn}, s_{qn}, e_{qn})\}_{n=1}^{\max_i h^{(i)}}$  を得ることができる．

以上より，人物の対応付け結果  $M$  が得られれば人物ごとの移動履歴は容易に得られる．また，対応付け結果  $M$  を得るのとラベル割り当て  $\mathbf{h}$  を得ることは同値である．よって，解としてはラベル割り当て  $\mathbf{h}$  が得られれば良い．

## まとめ

以上より，多人数多カメラ間人物対応付け問題は，トラックレット集合  $R$  を引数に取り，ラベル割り当て  $\mathbf{h}$  を解として返す問題であると定式化できる．

## 2.3 多人数多カメラ間対応付けの従来手法

前節で述べたカメラ間人物対応付け問題は，最適な出力  $\mathbf{h}$  を求めるラベル割り当て問題であると言える．従来では，正解のラベル割り当て  $\hat{\mathbf{h}}$  を引数に取っ

たときに最小値をとるコスト関数  $F(\mathbf{h})$  が導入され、このコスト関数  $F$  の最小化問題を解くことで、ラベル割り当て問題が解かれた。この  $F$  は、例えば以下の形式で表される。

$$F(\mathbf{h}) = \sum_{i=1}^N \sum_{j=i+1}^N \{a_{ij}P(r_j, r_i) + (1 - a_{ij})(1 - P(r_j, r_i))\} \quad (4)$$

$$\text{for } a_{ij} = \begin{cases} 0 & (h^{(i)} \neq h^{(j)}) \\ 1 & (h^{(i)} = h^{(j)}) \end{cases} \quad (5)$$

ここに、 $P(r_j, r_i)$  はトラックレット  $r_j$  と  $r_i$  の類似度である。この最小化問題を解くには、全てのラベルの組み合わせに対して  $F(\mathbf{h})$  の値を計算する必要があるが、この計算の時間計算量は  $O(N^N)$  となり、最小化問題は NP 困難であることがわかる。

これに対し Chen ら [10] は、ブラインドエリア内では人物の移動速度も見えも大きく変化しないものという仮定を置き、計算時間を短縮する手法を提案している。この手法では、対応付け相手の限定のため、あるカメラ  $c_j$  の視野から隣接するカメラ  $c_i$  の視野への遷移は必ずある時間以上かつある時間以内の時間で行われる、という時空間制約を用いている。その時間の最小値である最短遷移時間  $t_{min}(c_j, c_i)$  と、最大値である最長遷移時間  $t_{max}(c_j, c_i)$  がカメラ  $c_j$  と  $c_i$  の各組に対して予め統計的に値が求められている。この制約を用いることで、遷移時間が制約の範囲内に入らないトラックレットは対応相手の候補から外すことが可能となり、対応相手の探索範囲を限定することが可能となる。この時空間制約を用い、時空間的に狭い範囲のみを参照して局所的に最適化を行うことで対応付けの相手を特定することで、ラベリング問題を  $O(N^2)$  の時間計算量で解くことができ、計算時間の短縮が可能となる。Chen らの手法の詳細な処理手順を、以下で述べる。

この手法では、まず、 $r_i \in R$  に着目して  $prev(r_i)$  を探す。ここで、ある人物があるカメラ視野から隣接するカメラ視野へ、他のカメラ視野を経由せずに移動を行い、前者のカメラからはトラックレット  $r_j$ 、後者からはトラックレット  $r_i$  が得られたとき、 $r_j = prev(r_i)$  と表す。 $prev(r_i)$  の候補の集合  $H_i$  を

$$H_i = \{r_j | j < i, (v_i, v_j) \in E, t_{min}(c_i, c_j) < s_i - e_j < t_{max}(c_i, c_j)\} \quad (6)$$

として作成する。この処理では、時空間制約を用いて  $prev(r_i)$  に当たるトラッ

クレットの候補を絞っている。

次に、 $H_i$  内の各  $r_j$  について、 $r_i$  との類似度  $P(r_j, r_i)$  を計算する。そして、最も  $r_i$  との類似度の高い  $r_{\tilde{j}} \in H_i (\tilde{j} = \arg \max_j P(r_j, r_i))$  を選出する。その後、前もって定められた閾値  $\theta_{sim}$  について  $P(r_{\tilde{j}}, r_i) \geq \theta_{sim}$  が成立するならば、 $r_{\tilde{j}}$  を  $r_j$  と置いて  $r_i$  と対応付けする。もし  $P(r_{\tilde{j}}, r_i) < \theta_{sim}$  ならば、 $r_i$  はある人物が初めてカメラに観測されたときに得られたトラックレットなので  $prev(r_i)$  は存在しないと判断し、対応付けを行わない。また、 $H_i$  が空集合である場合も、同様に対応付けを行わない。

以上の処理を、 $i$  の小さい  $r_i \in R$  から順に行い、各  $r_i$  について行われた対応付けに基づいて行列  $M$  の  $i\hat{j}$  要素を更新する。最後に、 $M$  からラベル割り当て  $\mathbf{h}$  を作成すれば、解が得られる。

この手法におけるトラックレット同士の類似度  $P(r_j, r_i)$  は、**時空間特徴**の類似度  $P_{app}(r_j, r_i)$  と**外見特徴**の類似度  $P_{st}(r_j, r_i)$  を用いて、

$$P(r_j, r_i) = P_{app}(r_j, r_i)P_{st}(r_j, r_i) \quad (7)$$

と定義されている。

あるトラックレット  $r_j$  と  $r_i$  の外見特徴の類似度  $P_{app}(r_j, r_i)$  は

$$P_{app}(r_j, r_i) = \sqrt{1 - \rho(p(j), p(i))} \quad (8)$$

と定義されている。ここに、 $p(i)$  は  $r_i$  の色ヒストグラムを表し、 $\rho$  は Bhattacharyya 係数を表す。

一方、あるトラックレット  $r_j$  とそれより撮影開始時刻の遅い  $r_i$  の時空間特徴の類似度  $P_{st}(r_j, r_i)$  は

$$P_{st}(r_j, r_i) = \zeta_{c_j c_i}(s_i - e_j) \quad (9)$$

と定義されている。

## 2.4 従来手法の問題点

Chen らの手法では、人物はブラインドエリアでは移動速度も見えも大きく変化しないという状況を想定していたため、それぞれの人物に対して順に最適化を行うことによって、ラベル割り当て問題を解くことができた。しかし、本研究で想定するシーンでは、問題 (i) : 人物のブラインドエリアでの滞留、問題

(ii) : カメラ間での人物の見えの変化, の2つの問題が発生するため, 真の値以外のラベル割り当てにおいてコスト関数が最小値を取ることがあり, 前述のコスト関数の最小化問題を解いても真の解を得られない可能性がある.

以下では, 両問題について詳述する. ただし, ある人物があるカメラの視野から隣接するカメラの視野への移動を行い, 前者のカメラからはトラックレット  $r_j$ , 後者からは  $r_i (i \neq j)$  が得られたとして, 両トラックレット間で問題 (i)(ii) が発生したという状況設定のもとで議論を行う.

問題 (i) : 人物のブラインドエリアでの滞留とは, ブラインドエリア内の小売店・便所・展示物などの立ち寄り箇所の存在が原因となって, 従来手法の最長遷移時間を超える滞留が発生するという問題である.  $c_j$  から  $c_i$  への遷移の間にこの問題が発生すると, 両トラックレット間での遷移時間  $s_i - e_j$  が最長遷移時間  $t_{max}(c_i, c_j)$  を上回る. この場合,  $r_j$  が  $r_i$  の対応相手の候補集合  $H_i$  に含まれなくなり,  $r_j$  と  $r_i$  が対応付けされなくなる. この問題に対処するには, 時空間制約を使用しない, 即ち  $t_{max}$  を十分大きくすればよいが, この場合には候補集合内のトラックレット数も大幅に増加するという問題が起こる. 対応付け相手の選択肢が増加することで, 候補集合  $H_i$  内に  $P(r_i, r_k) > P(r_j, r_i)$  なる, 別人ではあるものの類似した特徴を持つ人物のトラックレット  $r_k$  が含まれる可能性が高まり, 誤った対応付けがなされる可能性が高くなる.

問題 (ii) : カメラ間での人物の見えの変化とは, 人物に対するカメラの向きが一定でないことが原因となって人物の見えの変化が生じるという問題である. 人物の見えは, 着用する服や所持する荷物の構造によっては撮影方向の変動とともに大きく変化する.  $r_j$  と  $r_i$  で人物の見えが大きく異なる場合,  $P(r_j, r_i)$  の値は小さくなり, ある別人のトラックレット  $r_k (\in H_i)$  について  $P(r_j, r_i) < P(r_k, r_i)$  を満たすほど値が小さくなった場合には, 誤って  $r_i$  と  $r_k$  の対応付けがなされる. 従来研究ではこの問題に対して, 人物がブラインドエリアを通過する時間は時空間制約を用いて限定できるという仮定に基づいて対処してきたが, 本研究の環境下ではその仮定を置くことはできない.

上記の問題 (i) への説明において, 従来手法では時空間制約を使用するか否かにより問題にどの程度対処できるかが変わることを述べた. よって, 時空間制約を使用する場合としない場合における対応付け精度の差について議論の必要性がある. 本稿では, 時空間制約を用いる従来手法を「Feat+STconstraint」, 時空間制約を用いない従来手法を「Feat」と呼ぶ. 以下ではこの2つの手法が,



問題 (i)(ii) に対処できるか否かについて説明する。

まず手法 Feat+STconstraint を用いる場合について述べる。この手法では、問題 (i) が発生した場合、カメラ  $c_j$  の視野からカメラ  $c_i$  の視野への遷移時間  $s_i - e_j$  が最長遷移時間  $t_{max}(c_i, c_j)$  を上回り、 $r_j$  が  $r_i$  の対応相手の候補集合  $H_i$  に含まれなくなるため、正しく対応付けができない。一方、問題 (ii) が発生した場合、時空間制約を用いて対応付けの候補集合  $H_i$  の大きさを小さくすることにより、 $P(r_j, r_i) < P(r_k, r_i)$  なる  $r_k \in H_i$  の存在確率を低下させることで、高い精度での人物対応付けができる。問題 (i) と (ii) が同時に発生した場合は、問題 (i) 発生時と同様に、正しく対応付けができない。

次に、手法 Feat を用いる場合について述べる。この場合、問題 (i) が発生したときは、時空間制約を用いていないため  $t_{max} > s_i - e_j$  となり、 $r_j \in H_i$  となるため対応付けは可能である。但し、時空間制約を用いないことにより全ての対応付け候補が  $H_i$  に含まれるため、 $P(r_j, r_i) < P(r_k, r_i)$  なる  $r_k \in H_i$  が存在する可能性があり、対応付けの誤りが起こりうる。一方、問題 (ii) が発生した場合、 $P(r_j, r_i)$  の値が小さく、かつ  $|H_i|$  の値が大きいため、 $|H_i|$  の値が小さい時と比べて  $P(r_j, r_i) < P(r_k, r_i)$  なる  $r_k \in H_i$  の存在確率が高く、正しく対応付けを行うことは難しい。問題 (i) と (ii) が同時に発生した場合は、問題 (ii) 発生時と同様に、正しく対応付けができない。

以上の両手法の特性を整理したものを、表 1 に示す。

表 1: 従来手法の対応付け精度

トラックレット組	問題発生無し	問題 (i) のみ発生	問題 (ii) のみ発生	問題 (i) と (ii) 発生
Feat	○	△	×	×
Feat+STconstraint	○	×	○	×

但し、表 1 では、「○」は精度良く対応付けできるトラックレット組を、「△」は条件によっては対応付けが可能なトラックレット組を、「×」は対応付け精度が極めて低くなるトラックレット組を表している。

## 第3章 時空間制約の選択的緩和と経路制約を用いた人物対応付け

本節では、第2章で述べた2つの問題に対し、時空間制約の選択的緩和と経路制約の導入により対処する方法を述べる。

3.1節で、従来手法の問題に対処する、時空間制約の選択的緩和と経路制約の概要と原理を述べる。3.2節では、提案手法の各段階における詳細な処理を述べる。

以下では、ある人物  $P$  があるカメラからあるカメラの視野への遷移を行い、前者のカメラからはトラックレット  $r_j$ 、後者からはトラックレット  $r_i (i \neq j)$  が得られたとして、両トラックレット間で問題 (i)(ii) が発生したという仮定のもと、議論を進める。

### 3.1 時空間制約の選択的緩和と経路制約の導入

まず、問題 (i) のみが発生した場合に従来手法が失敗することについては、全ての人物の遷移が時空間制約に従うものとして画一的な処理を行っている点が原因である。時空間制約を満たす遷移については従来手法を適用すれば正しいラベル付けが可能であるが、ブラインドエリアでの滞留により時空間制約を満たさない遷移に関しては、時空間制約を緩和する、つまり  $t_{max}$  を十分大きな値にすればよい。これにより、トラックレット  $r_j = prev(r_i)$  が  $H_i$  に含まれるようになる。但し、時空間制約を用いないことにより、全てのトラックレットが  $r_i$  の対応付け候補となるため、 $P(r_j, r_i) < P(r_k, r_i)$  なる別人のトラックレット  $r_k$  が存在する可能性が高まり、対応付けの誤りが起こりうる。そのため、閾値を厳しく設定することで、外見的な類似度が極めて高いトラックレットのみを対応付けすべきであると考えられる。

一方、問題 (ii) のみが発生した場合には、何らかの方法で候補集合  $H_i$  の要素数を少なくできれば、正しい対応付けが行える可能性が高まる。そのために、本研究では、複数カメラの隣接関係を用いる。その方法として、 $P$  がカメラ  $c_j$  で観測される以前に別のカメラで観測されて得られたトラックレット  $r_k$  があるとき、 $r_k$  を用いて  $r_j$  に関する制約式を導く、ということを考える。これを用いれば、 $h^{(k)} = h^{(i)} (k < i)$  かつ  $(v_k, v_i) \notin E$  なる  $r_k$  が発見されたとき、 $s_j$  と  $e_j$  に関

する不等式

$$\begin{cases} e_k + t_{\min}(c_k, c_j) < s_j \\ e_j + t_{\min}(c_j, c_i) < s_i \end{cases} \quad (10)$$

が成立し,  $s_j$  と  $e_j$  の値の取りうる範囲を限定できる. 同時に,  $(v_k, v_i) \notin E$  なる場合, カメラ  $c_j$  はカメラ  $c_k, c_i$  の視野を連結するような地点に存在しなくてはならない. つまり, 必ず以下の条件を満たすトラックレット集合  $Y = \{r_{y_i}\}_{i=1}^{N_Y}$  が存在しなくてはならない.

$$\begin{cases} r_j \in Y \subseteq \{r_l | s_k \leq s_l \wedge e_l \leq e_i\} \\ r_{y_1} = r_k, r_{y_{N_Y}} = r_i \\ (c_{y_l}, c_{y_{l+1}}) \in E \\ e_{y_l} + t_{\min}(c_{y_l}, c_{y_{l+1}}) < s_{y_{l+1}} \end{cases} \quad (11)$$

この2条件を満たすトラックレット  $r_j$  のみをトラックレット  $r_i$  の対応付け候補とすることにより, 問題 (ii) 発生下でも対応付けの正答率が上昇する. なお, トラックレット  $r_k$  と  $r_i$  が同一人物であることは, 両トラックレットの外見特徴の類似度が大きいならば, 問題 (i) に対処する処理で検知できる. 以上の, 「 $h^{(k)} = h^{(i)} (k < i)$  かつ  $(v_k, v_i) \notin E$  なる  $r_k$  と  $r_i$  が存在するなら, 式 (10) を満たすトラックレット  $r_j$  で, 式 (11) を満たす集合  $Y$  の要素であるものが得られている必要がある」という制約を, 経路制約と呼ぶ.

以上をまとめると, 前節の各問題に対処するアイデアは次の2点となる.

- アイデア (i) : コスト関数の最適化の前に, 特に外見が類似しているトラックレット組に対して, ブラインドエリアを含むカメラ間での時空間制約を選択的に緩和した対応付けを行うことにより, 問題 (i) に対処
- アイデア (ii) : 経路制約を用いた対応付けを予め行うことにより, 問題 (ii) に対処

問題 (i) と (ii) が同時に発生した場合は, 次の2つの条件が成立するときに限って, アイデア (ii) を用いて正しい対応付けができる. 第1の条件は,  $r_j$  と同一人物のものであり, かつ  $s_k < s_j$  であり, かつトラックレット  $r_i$  との外見特徴の類似度が高いトラックレット  $r_k$  が存在することである. 第2の条件は, 経路制約を満たすトラックレットの中に  $P(r_j, r_i) < P(r_l, r_i)$  なる別人のトラックレット  $r_l$  が存在しないことである.

以上より, 提案手法を用いた人物対応付け, は従来手法に比べて様々なトラッ

クレットの組を正しく対応付けできることが分かる．本節で述べた提案手法の特性を整理したものを表 2 に示す．

表 2: トラックレット組ごとの対応付け精度

トラックレット組	問題発生無し	問題 (i) のみ発生	問題 (ii) のみ発生	問題 (i) と (ii) 発生
Feat	○	△	×	×
Feat+STconstraint	○	×	○	×
提案手法	○	△	○	△

### 3.2 対応付け手順

前節のアイデア (i)(ii) を導入した対応付けを，従来手法によって誤った対応付けがなされる前に行うことで，対応付け精度を向上させる．また，アイデア (ii) を用いた対応付けを実行するには，事前に位置の離れたカメラから同一人物のトラックレット ( $r_k, r_i$  とする) を発見しておき，その後で  $c_k, c_i$  の視野を繋ぐ位置のカメラから同一人物のトラックレット  $r_j$  を発見する，という手順を踏む必要がある．従って，アイデア (ii) を用いた対応付けの前に，時空間制約を緩和した対応付け，即ちアイデア (i) を用いた対応付けをしておくべきである．

上記の 2 つの対応付けの前に，明らかに同一人物と判別できるようなトラックレット組 ( $r_l, r_m$  とする) に関しては同一人物としておく，という処理を加えれば，上記の 2 つの対応付けを行う際の情報が増える．簡単に同一人物と判別できるのは，トラックレット  $r_l, r_m$  間で問題 (i) も (ii) も発生していないときである．これは，トラックレット  $r_l$  と  $r_m$  の時空間特徴と外見特徴の類似度がともに大きい値になることから容易に判断できる．従って，ラベルの割り当て手順は以下の 4 段階に分けられる．

- 手順 1 まず，時空間特徴と外見特徴の両方の類似度が極めて大きいトラックレットの組を隣接カメラで探し，同一人物としてラベル付けする．この処理は 3.2.1 節で説明する．
- 手順 2 時空間制約を緩和し，全トラックレットから外見特徴の類似度が大きいトラックレットの組があれば，それらを同一人物としてラベル付けする．この処理は 3.2.2 節で説明する．

手順3 手順2までで同一人物とされたトラックレット同士の間から経路制約を満たすトラックレットを探し、最適なラベル付けをする。この処理は3.2.3節で説明する。

手順4 全てのトラックレットにラベルを割り当てる。このとき、手順1~3で与えられた対応付けの組を固定したまま、残りの対応付けを最適化する。この処理は3.2.4節で説明する。

上記の手順を実行するためには、ある2つのトラックレットが同一人物のものであるかを判定する基準が必要である。これには、トラックレット同士の距離を定義し、その値が一定以下ならば同一人物と見なす、というようにすればよい。ここに、2つのトラックレット  $r_a, r_b (a < b)$  間の距離  $d(r_a, r_b)$  と、2つのトラックレットが同一人物のものである尤度  $P(r_a, r_b)$  との間に、 $d(r_a, r_b) = 1 - P(r_a, r_b)$  なる関係式が成り立つとする。

本研究では、2つのトラックレット  $r_a, r_b (a < b)$  間の距離を、外見特徴の距離と時空間特徴の距離の加重平均として、以下のように定義する。

$$d(r_a, r_b) = \alpha d_{app}(r_a, r_b) + (1 - \alpha) d_{st}(r_a, r_b) \quad (12)$$

$$\text{where } 0 \leq \alpha \leq 1$$

ここに外見特徴の距離  $d_{app}(r_a, r_b)$  は、以下のように定義する。

$$d_{app}(r_a, r_b) = \min_{i,j} B_c(\mathbf{f}_a^i, \mathbf{f}_b^j) \quad (13)$$

但し、 $B_c$  はベクトル同士の Bhattacharyya 距離を表す関数であり、値域は  $[0, 1]$  区間である。

一方、時空間特徴の距離の定義を以下で述べる。本研究では、式(1)で述べた、辺  $(v_a, v_b) \in E$  の移動時間の分布  $\zeta_{c_a c_b}(t)$  として、ある事象が1回発生するまでの時間の分布を表すのに使用される、ガンマ分布を仮定する。ガンマ分布の密度関数  $\gamma$  は以下で与えられる。

$$\gamma(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k) \theta^k} \quad (14)$$

ここに、 $k$  はガンマ分布の形状母数、 $\theta$  は尺度母数である。この両変数の値は辺ごとに異なるため、以下では辺  $(v_a, v_b) \in E$  に対する形状母数と尺度母数を  $k_{c_a c_b}$ 、尺度母数  $\theta_{c_a c_b}$  と表す。また、辺を通過する最短遷移時間の推定値  $\hat{t}_{min}(c_a, c_b)$ 、最

長遷移時間の推定値  $\hat{t}_{max}(c_a, c_b)$  を辺ごとに統計的に求めておく．これより，滞留が発生しない場合の移動時間の分布を，以下のように推定する．

$$\hat{\zeta}_{c_a c_b}(t) = \begin{cases} \gamma(x - \hat{t}_{min}(c_a, c_b); k_{c_a, c_b}, \theta_i) & (\text{if } \hat{t}_{min}(c_a, c_b) \leq t \leq \hat{t}_{max}(c_a, c_b)), \\ 0 & (\text{if } 0 < t < \hat{t}_{min}, \hat{t}_{max} < t), \end{cases} \quad (15)$$

ここに， $t$  はトラックレット同士の時間差  $s_b - e_a$  を表す．

加えて，滞留発生時の移動時間の分布の推定値  $\hat{\eta}$  を求める．移動時間の分布を求めるには，滞留が発生したかどうかを判断し，適切な分布関数を選ぶという方法があるが，滞留が発生したか否かの判断は一般に困難である．そこで， $c_a$  と  $c_b$  の間の移動時間の分布を表す関数  $\hat{\xi}$  を，滞留非発生時の移動時間  $t$  の分布関数  $\hat{\zeta}$  と，滞留発生時の分布関数  $\hat{\eta}$  の線形結合とする．

$$\hat{\xi}_{c_a c_b}(t) = (1 - \beta)\hat{\zeta}_{c_a c_b}(t) + \beta\hat{\eta}_{c_a c_b}(t) \quad (16)$$

この関数を，時間差が  $t$  である2つのトラックレットが同一人物のものである時間的尤度と考える．この関数の値域を  $[0, 1]$  区間に補正し，トラックレット  $r_a$  と  $r_b ((v_a, v_b) \in E)$  の時空間特徴の距離関数  $d_{st}(r_a, r_b)$  を以下の形で得る．

$$d_{st}(r_a, r_b) = 1 - \frac{\hat{\xi}_{c_a c_b}(s_b - e_a)}{\max_{t>0} \hat{\xi}_{c_a c_b}(t)} \quad (17)$$

これを用いて，以下の節では提案手法の4つの手順を説明する．

### 3.2.1 手順1：強い時空間制約と外見制約を用いたラベル割り当て

明らかに時空間制約と外見制約をともに満たしているトラックレット同士を対応付けする．データセット内のあるトラックレットの組  $(r_a, r_b)$  が明らかに時空間制約と外見制約をともに満たしている，ということは，即ち  $(v_a, v_b) \in E$  であり，かつある閾値  $\theta_{step1}$  に対して  $d(r_a, r_b) < \theta_{step1}$  が成立している，という条件を満たすことである．

従って，本手順では，2つのトラックレット  $r_a, r_b$  に対して上記の2つの条件が満たされているかを計算し，満たされているならば同一人物，と判断する操作を，あらゆる  $a, b$  について行う．対応付けの結果に応じて対応付け結果  $M$  を更新することで，本手順の解が得られる．

### 3.2.2 手順2：時空間制約を緩和したラベル割り当て

外見特徴の距離が極めて小さいトラックレットの組を対応付ける．ある2つのトラックレット  $r_a, r_b$  の外見特徴の距離が極めて近いということを判定するには，ある閾値  $\theta_{step2}$  を定め， $d_{app}(r_a, r_b) < \theta_{step2}$  が成立しているかを調べればよい．但し，トラックレット  $r_a, r_b$  と類似する外見の人物が他にもいる場合には，対応付けの誤りが発生しやすくなるため，正しく対応付けを行うには，他のレコードとの外見特徴との距離に比べ  $d_{app}(r_a, r_b)$  が突出して小さいことが必要である．この判断を，ある閾値  $\theta_{diff}$  を用い，以下の不等式(18)によって行う．

$$\frac{d_{app}(r_a, r_b)}{\min\{\min_{i \neq a} d_{app}(r_i, r_b), \min_{j \neq b} d_{app}(r_a, r_j)\}} > \theta_{diff} \quad (18)$$

従って，本手順では，あらゆるトラックレットの組  $r_a, r_b$  に対して上記の2つの条件が満たされているかを計算し，満たされているならばそれらを対応付けする．

### 3.2.3 手順3：経路制約を用いたラベル割り当て

手順2で対応付けられた各々のトラックレットについて，経路制約を用いて同一人物のトラックレットを選び出す．ここで，手順2までで対応付けされたトラックレットの組を  $U^i = (r_{a_i}, r_{b_i})$  と表す．但し，各組について  $s_{a_i} < s_{b_i}$  なる順番でトラックレットが並んでいるものとする．また，手順2で得られた全ての組の集合を  $U = \{U^i\}_{i=1}^{N_R}$  と表す．

本研究では，動的計画法を用いて経路制約を満たすトラックレットを選択する．その手順を以下に示す．

まず，ある  $U^i \in U$  に注目し，中間レコード集合  $Z^i = \{r_k | s_{a_i} \leq s_k \wedge e_k \leq e_{b_i}\}$  を作成する．次に，トラックレット  $r_1^i, r_2^i, \dots, r_{|Z^i|}^i \in Z^i (j < k \Rightarrow s_j^i < s_k^i)$  について， $r_{a_i}$  との最小距離  $\delta(r_{a_i}, r_i^j)$  を求める．この  $\delta(r_{a_i}, r_i^j)$  は，以下のように再帰的に求めることができる．

$$\delta(r_{a_i}, r_i^j) = \min_{1 \leq q \leq j, (v_q^i, v_j^i) \in E} \{\delta(r_{a_i}, r_i^q) + d(r_i^q, r_i^j)\} \quad (19)$$

また， $r_1^i = r_{a_i}$  であることから， $\delta(r_{a_i}, r_1^i) = 0$  と定義する．これより， $\delta(r_{a_i}, r_i^j)$  の値を  $i = 1, 2, \dots, |Z^i|$  について求める．ここで，関数  $\text{before}(r_i^j)$  を

$$\text{before}(r_i^j) = r_i^q \text{ (for } q = \arg \min_{1 \leq q \leq j, (v_q^i, v_j^i) \in E} \{\delta(r_{a_i}, r_i^q) + d(r_i^q, r_i^j)\}) \quad (20)$$

と定義する.  $\delta(r_{a_i}, r_i^j)$  の値を算出するごとに,  $\text{before}(r_i^j)$  も求めておく.

各  $\text{before}(r_i^j)$  を求めれば, 経路制約を満たすトラックレット列  $Y_i = [r_{y_{i1}}, r_{y_{i2}}, \dots, r_{y_{i|Y_i|}}]$  を求めることができる. ここに,  $Y_i$  の各要素は

$$r_{y_{ij}} = \begin{cases} r_{b_i} & (\text{if } j = |Y_i|), \\ \text{before}(r_{y_{ij+1}}) & (\text{if } 1 \leq j < |Y_i|), \end{cases} \quad (21)$$

として, 帰納的に定義されている. これより,  $Y_i$  に含まれる全てのトラックレットに同じラベルを割り振る.

以上の処理を全ての  $U^i$  について行う. このとき, 経路制約を満たすトラックレットの候補が少ない組, 即ち組に含まれる2つのトラックレットの時間差  $s_{b_i} - e_{a_i}$  の小さい  $U^i$  から順に処理を行う. これは, 候補が少ない組に対する処理では誤りが起こりにくく, また誤りが起こりにくい処理を先に行うことにより, 候補の多い組の候補を絞ることができるからである. 但し, 時間差  $s_{b_i} - e_{a_i}$  が負の値である場合や最短遷移時間に満たない場合など,  $Y_i$  を求めることが不可能な場合には, その  $U_i$  については処理を行わない. 結果として作成されたラベル割り当てを手順2以前のラベル割り当てに統合すれば, 本手順の結果が得られる.

#### 3.2.4 手順4: 局所的最適化によるラベル割り当て

手順3までのラベル付け結果を固定したうえで, 従来手法を用いてトラックレットごとに局所的なラベル割り当ての最適化を行えば, まだラベルが割り当てられていない全てのトラックレットにラベルを割り振ることができる. 本研究では, このラベル割り当てに2.3節で説明したChenらの多人数多カメラ間対応付け手法を使用する.

## 第4章 実験・結果

3章では, 人物の移動時間と見えの変化に対処できる, 多人数多カメラ間対応付け手法を提案した. 本章ではこの手法を用いた実験と考察について述べる.

### 4.1 利用データ

本実験では, 人物画像として公開データセットであるShinpuhkan 2014 dataset [11]を用いた. このデータセットは24人の人物の画像を含んでおり, 各人物を視野重複の無い16台の固定カメラで撮影したものとなっている. ここで, デー



タセット内の各人物画像列  $\tilde{I}_i$  には、予め撮影カメラ ID  $\tilde{c}_i$ 、観測開始時刻  $\tilde{s}_i$ 、観測終了時刻  $\tilde{e}_i$ 、人物 ID  $\tilde{p}_i$  の情報が付与されており、これらを組にして正解トラックレット  $\tilde{r}_i = (\tilde{f}_i, \tilde{s}_i, \tilde{e}_i, \tilde{c}_i, \tilde{p}_i)$  と表す。ここに、 $\tilde{f}_i$  は人物画像列  $\tilde{I}_i$  から抽出した特徴量列である。また、トラックレット集合  $\tilde{R} = \{\tilde{r}_i\}$  とカメラネットワーク  $\tilde{G}$  を組にしたものを、データセット  $\tilde{D} = (\tilde{R}, \tilde{G})$  と表す。

## 4.2 実験用データセットの生成

本実験では人物の対応付け精度を統計的に求めるために、多数のデータセットを用意して評価を行う必要がある。そこで、データセット生成器を作成し、それを用いて  $\tilde{D}$  をもとに、カメラネットワーク  $G$  や滞留発生率  $\beta$  を変更させた実験用データセット  $D_n = (R_n, G_n)$  ( $n \in \mathbb{N}$ ) を複数生成した。その生成手順を以下に示す。

### 4.2.1 カメラネットワーク $G$ の生成法

まず、データセット生成器は使用するカメラに対応する頂点集合  $V \in \tilde{V}$  を入力として受け取る。次に、枝集合を  $E = \{(\mathbf{v}_i, \mathbf{v}_j) | \forall \mathbf{v}_i, \mathbf{v}_j \in V\}$  とする。そして、 $G$  のある部分グラフのハミルトン路を1つ作成する、という操作を予め定められた回数  $N_{path}$  だけ行う。ここで作成されたハミルトン路を  $\Pi_1, \Pi_2, \dots, \Pi_{N_{path}}$  とし、以下ではこれをパスと呼ぶ。各人物はこのパスに沿って移動するものとする。最後に、 $E$  から全ての要素を取り除いた後、作成されたパスに含まれる全ての辺を辺集合  $E$  に追加する。以上より  $G = (V, E)$  とする。

### 4.2.2 トラックレット集合 $R$ の生成法

生成したカメラネットワーク  $G$  は、ある仮想的な環境の中に構築されていると見なすことができる。この仮想的な環境の中を各人物に歩かせることにより、トラックレット集合  $R$  の生成を行う。 $R$  を生成するには、各人物の移動経路と、人物がカメラで観測される時刻の2つを定めればよい。

各人物の移動経路としては、前節で生成したパスのうちから1つをランダムで選択するものとする。人物は、選択されたパスの上を移動する。

各人物の観測時刻を定めるためには、まず各人物がカメラネットワーク  $G$  内に進入する時刻を定める。これには、まずデータセット内の24人の  $G$  への進入順をランダムで決定し、その後である人物の  $G$  への進入から次の人物の進入までの時間差を決定する。この時間差は、事象が1回発生するまでの時間を表

すのに使用されるガンマ分布に従う乱数で与える.

次に, 各人物にパス上を仮想的に移動させる処理を行う. いま, 人物  $p$  に対し, あるパス  $\Pi \in \{\Pi_i\}_{i=1}^{N_{path}}$  が選ばれたとする. ここではパスを, 通過する頂点の列として,  $\Pi = (v^1, v^2, \dots, v^{N_{path}})$  と表す. この人物の観測時刻を決定するには, 頂点  $v^i$  での人物  $p$  の観測開始時刻  $s^i$  と観測終了時刻  $e^i$  を定めればよい. その定め方を以下で述べる.

まず, 観測開始時刻  $s^i$  については, この人物がカメラネットワークに進入した時刻が  $\sigma$  であった場合に, 以下の定義に従って帰納的に設定する.

$$s^i = \begin{cases} \sigma & (\text{if } i = 1), \\ e^{i-1} + t_{random} & (\text{if } i > 1), \end{cases} \quad (22)$$

ここに,  $t_{random}$  は辺を通過する所要時間であり, 確率密度関数  $(1-\beta)\zeta_{c^{i-1}c^i}(t) + \beta\eta_{c^{i-1}c^i}(t)$  に従う乱数で与えられる.  $\beta$  は所与の滞留発生率である. 一方, 観測終了時刻  $e^i$  は, 人物  $p$  が頂点  $v^i$  を通過する所要時間を  $s^i$  に加算すればよい. その所要時間は,  $\tilde{D}$  内の値を参考にして決める. 即ち,  $\tilde{p}_k = p$  かつ  $\tilde{c}_k = c^i$  であるような正解トラックレット  $\tilde{r}_k \in \tilde{R}$  を1つ選べば, その所要時間は  $\tilde{e}_k - \tilde{s}_k$  と表される. 従って,  $e^i$  は以下の式で決定される.

$$e^i = s^i + (\tilde{e}_k - \tilde{s}_k) \quad (23)$$

以上で述べたトラックレット作成操作を全ての人物に対して行い, 得られたすべてのトラックレットの集合を  $R$  とする.

### 4.3 実験方法

4.2節で述べた方法で作成したデータセットを入力として, 各人物の移動履歴の推定を行い, 解としてラベル割り当て  $\mathbf{h}$  を出力する, という実験を行った. このとき, 人物の移動速度や見えの変化が, 対応付けの精度に与える影響を評価するため, 滞留発生率  $\beta$  と見えの変化の発生率の双方のパラメタを変化させながら複数のデータセットを生成し, それに対して移動履歴の推定を行った.

見えの変化の起こりやすさを変化させる手段として, 本実験では日陰と日向の照明条件の相違を利用した. 公開データセット  $\tilde{D}$  には, 日陰を観測するカメラと日向を観測するカメラの画像が含まれている. 一般的には, 日向である物体を観測する場合, 日陰で観測するよりも, 日光の影響により輝度が増して見

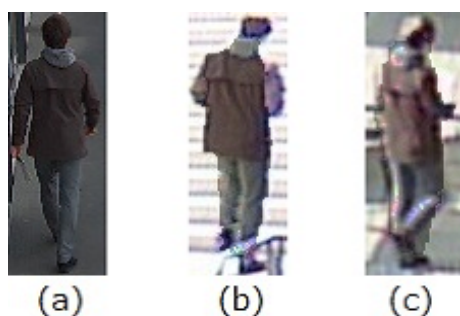


図 1: 照明条件の相違による見えの変化

える. この例を図 1 に示す. 図 1 の 3 枚の画像はいずれも同一人物の画像であり, (a) は日陰を観測するカメラで, (b)(c) は日向を観測するカメラで撮影されたものである. この図から, (a) に比べて (b)(c) は輝度が大きく, 両者の間で見えの変化が発生していることが分かる. 加えて (b) と (c) では互いに人物に対して日光のあたる方向が異なるため, 陰影の付き方が違い, やはり両者の間で見えの変化が発生していることが分かる.

以上より, データセットに含まれるカメラの内, 日向を観測するカメラの割合が高いほど, 見えの変化が発生しやすいデータセットであるといえる. 本実験では, データセット作成時に, 日向を観測するカメラに対応する頂点を何個頂点集合  $V$  に含めるかを変更することで, 見えの変化の発生率を調節した.

上記の 2 つのパラメタの変化に対する精度の変化を評価するため, 次の 2 種類の実験を行った. 第 1 に, 本手法が問題 (i) に対処できていることを示すため, 日向を観測するカメラの割合を固定し, 滞留発生率のみを変化させて複数のデータセットを生成し, それぞれに対して移動履歴の推定を行った. これを実験 A とし, 4.3.1 節で詳述する. 第 2 に, 本手法が問題 (ii) に対処できていることを示すため, 滞留発生率を固定し, 日向を観測するカメラの割合のみを変化させて複数のデータセットを生成し, それぞれに対して移動履歴の推定を行った. これを実験 B とし, 4.3.2 節で詳述する.

#### 4.3.1 実験 A : 滞留発生率の変化に対する対応付け精度の測定

実験 A ではデータセット生成に使用するカメラ台数  $N_{cam}$  を 5 とし, そのうちの日向を観測するカメラの台数  $N_{sun}$  を固定して移動履歴推定を行った. 滞留発生率  $\beta$  は 0 % から 100 % まで 10 % 刻みで変化させた. また, 使用するカメラネットワークは試行ごとにランダムに生成するものとした. このとき, パスの数  $N_{path}$  を 4 とし, うち 2 本のパスはランダムに生成, 残り 2 本のパスは前者の

2本を逆にしたものとした。その他のパラメタは経験的に決定した。

移動履歴推定を行う従来手法として、2.4節で述べたFeat, Feat+STconstraintの2種類を用いた。それに対応して、提案手法も以下の2種類を用いた。

- Feat+Route : 時空間制約を用いない提案手法。即ち、手順4において  $t_{max} = \infty$  とする。最短遷移時間  $t_{min}$  の値は変更しない。
- Feat+STconstraint+Route : 時空間制約を用いる提案手法。手順4で使用の各  $t_{max}$  の値は、実験用データセット生成時に用いた  $t_{max}$  の値と同一とした。

以上の計4種類の手法で人物の移動履歴の推定を行い、解として手法・滞留発生率ごとにラベル割り当て  $\mathbf{h}$  を算出した。

#### 4.3.2 実験B：日向のカメラの割合の変化に対する対応付け精度の測定

実験Bでは  $N_{cam} = 6$  とし、 $N_{sun}$  を0から3まで変化させることで、日向を観測するカメラの割合を変化させ、移動履歴推定を行った。この変化を実現するために、カメラネットワークを図2のように固定して実験を行った。各人物の

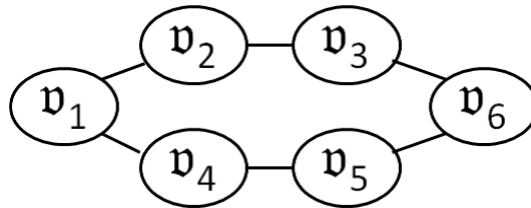


図2: 実験Bに用いるカメラネットワーク

移動する経路は、図2のカメラ番号を用いて  $(v_1, v_2, v_3, v_6)$ ,  $(v_1, v_4, v_5, v_6)$ , およびその逆順の計4通りに固定した。また、 $N_{sun} = 0$  のときには6台のカメラ全てに日陰を観測するカメラを指定し、 $N_{sun}$  の値を増やすごとに、順に  $v_2$ ,  $v_5$ ,  $v_3$ ,  $v_4$  を日向を観測するカメラに置き換えた。

以上の条件のもとで、実験Aと同じ4種類の手法で人物の移動履歴の推定を行い、解として各手法・各  $N_{sun}$  の値についてラベル割り当て  $\mathbf{h}$  を算出した。

## 4.4 評価方法

ラベル割り当ての結果の評価のため、解のラベル割り当て  $\mathbf{h}$  の、真のラベル割り当て  $\hat{\mathbf{h}}$  に対する正しさの指標として、本研究では情報検索の分野で精度評価に用いられるF値を用いた。F値は適合率と再現率という2つの指標の調和平均

均として定義されており，この値を算出するためには，適合率と再現率を定義する必要がある．そこで，本研究では，あるラベル割り当て  $\mathbf{h}$  の適合率を「行われた対応付けのうち正解であったものの割合」，再現率を「行われるべき全ての対応付けのうち実際に行われたものの割合」，と定めた．

両指標の定義を次に示す．まず，ラベル割り当て  $\mathbf{h}$  に対するリンク集合  $L(\mathbf{h})$  を以下のように定義した．

$$L(\mathbf{h}) = \{x = (g_{qj}, g_{q(j+1)}) \mid q < \max_i h^{(i)} \wedge (g_{qj}, g_{q(j+1)}) \in G_q\} \quad (24)$$

但し， $G_q = \{g_{q1}, g_{q2}, \dots, g_{q|G_q|}\}$  (但し  $i < j$  なら  $g_{qi} < g_{qj}$ ) である．

次に，真正集合  $TP(\mathbf{h})$  と偽正集合  $FP(\mathbf{h})$  を以下のように定義した．

$$TP(\mathbf{h}) = \{x \mid x \in L(\mathbf{h}) \wedge x \in L(\hat{\mathbf{h}})\} \quad (25)$$

$$FP(\mathbf{h}) = \{x \mid x \in L(\mathbf{h}) \wedge x \notin L(\hat{\mathbf{h}})\} \quad (26)$$

そして，適合率 *precision* と再現率 *recall* を次のように定義した．

$$precision(\mathbf{h}) = \frac{|TP(\mathbf{h})|}{|L(\mathbf{h})|} \quad (27)$$

$$recall(\mathbf{h}) = \frac{|TP(\mathbf{h})|}{|L(\hat{\mathbf{h}})|} \quad (28)$$

以上より，F 値 *F-measure* は以下の式で表される．

$$F\text{-measure}(\mathbf{h}) = \frac{2 \cdot precision(\mathbf{h}) \cdot recall(\mathbf{h})}{precision(\mathbf{h}) + recall(\mathbf{h})} \quad (29)$$

以上の定義に基づいて，F 値を提案手法と従来手法の計 4 手法の解について求め，それらを比較することにより，得失を評価した．

## 4.5 実験結果

### 4.5.1 実験 A：滞留発生率の変化に対する対応付け精度の測定

実験 A の， $N_{cam} = 5, N_{sun} = 0$  の場合と， $N_{cam} = 5, N_{sun} = 2$  の場合における評価結果を，それぞれ図 3，図 4 に示す．いずれも，1 条件につき 50 個のデータセットを生成し，それらに対して移動履歴推定を行って得られたラベル割り当ての F 値の平均を図に示している．グラフは，横軸に滞留率  $\beta$  を，縦軸に F 値を取っている．

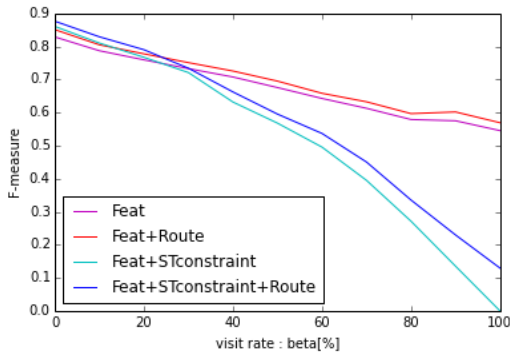


図 3:  $N_{cam} = 5, N_{sun} = 0$  の結果

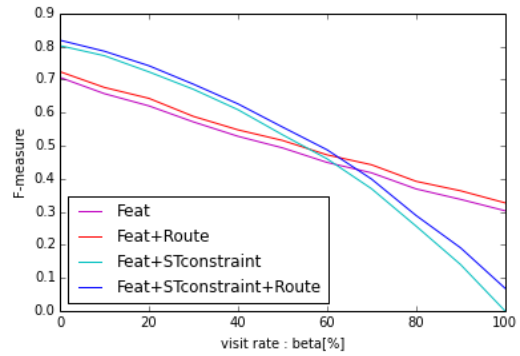


図 4:  $N_{cam} = 5, N_{sun} = 2$  の結果

いずれの結果においても、滞留発生率 $\beta$ の値に関らず、提案手法を用いた場合のF値が従来手法を上回っている。また、滞留発生率 $\beta$ の上昇とともに、提案手法と従来手法のF値の差異が拡大し、特に図3の手法Feat+STconstraint+Routeは、 $\beta$ が100%の時点で従来手法のF値を13.1%上回っている。このことから、提案手法を用いた人物対応付けが、問題(i)、即ち滞留が発生する問題に対処できていることが確認できた。

#### 4.5.2 実験 B : 日向のカメラの割合の変化に対する対応付け精度の測定

実験 A の、滞留発生率 $\beta$ が0%の場合、50%の場合、80%の場合における評価結果を、それぞれ図5、図6、図7に示す。いずれも、1条件につき50個のデータセットを生成し、それらに対して移動履歴推定を行って得られたラベル割り当てのF値の平均を図に示している。グラフは、横軸に日向を観測するカメラの台数 $N_{sun}$ を、縦軸にF値を取っている。

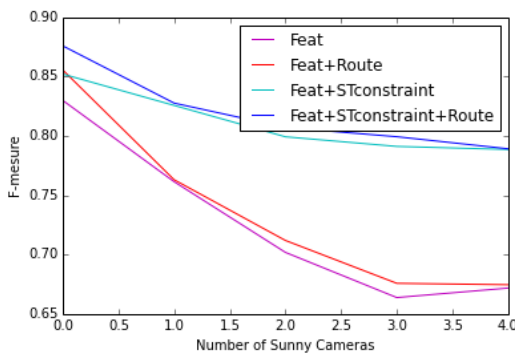


図 5: 滞留発生率0%での結果

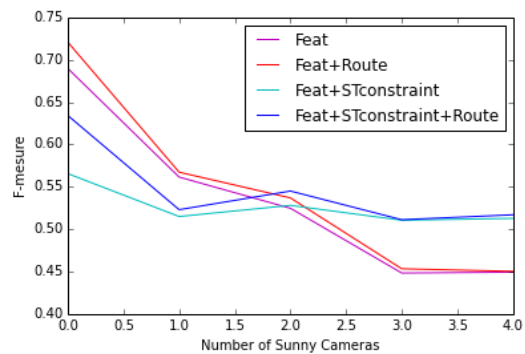


図 6: 滞留発生率50%での結果

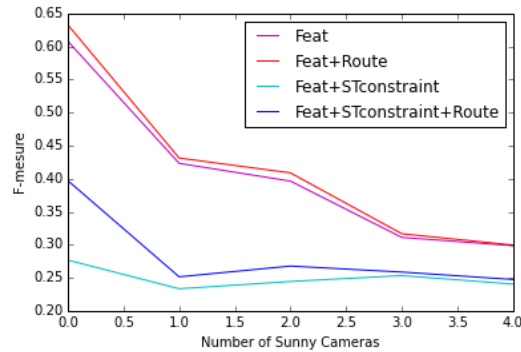


図 7: 滞留発生率 80 %での結果

いずれの結果においても、日向を観測するカメラの台数  $N_{sun}$  の値に関らず、提案手法を用いた場合の F 値が従来手法を上回っている。このことから、提案手法を用いた人物対応付けが、問題 (ii)、即ち人物の見えの変化が発生する問題に対処できていることが確認できた。

また、この実験では日向を観測するカメラの台数  $N_{sun}$  の上昇とともに、提案手法と従来手法の F 値の差異が縮小した。これは、同一人物のトラックレットでも、日向で得られたものと日陰で得られたものでは見えが異なることが多いため、 $N_{sun}$  の上昇とともに手順 2 で正しく対応付けが行われるトラックレットの数が減少し、結果として手順 3 の処理対象も減ったためであると考えられる。

## 第 5 章 結論

本研究では、複数の固定カメラからの画像を用いて人物の対応付けを行い、複数の人物の移動履歴を獲得する手法を提案した。

人物の対応付けに関する従来手法では人物の移動速度や見えが大きく変化しないとの仮定を置いている。しかし、ブラインドエリアでの滞留が発生すると人物の移動速度は大きく変化し、また撮影条件が変動すると人物の見えは大きく変化しうる。本稿では、時空間制約の緩和と経路制約の導入によって、この問題に対処する手法を提案した。そして、公開データセット Shinpuhkan2014dataset を用いて人物の移動履歴推定の実験を行い、人物の移動速度と見えが変化する状況下での対応付け精度を従来手法と比較することで、その有効性を示した。

今後の課題としては、人物のブラインドエリアでの滞留時間を統計的にモデル化することが挙げられる。本研究では、滞留発生時の移動時間の分布を一様

分布と仮定したが，実際の滞留発生時の移動時間の統計を取って分布を得ることができれば，より対応付けの精度が向上すると考えられる。

また，時空間制約の緩和の程度の基準を設定することも挙げられる．本研究では，手順2で最長遷移時間を無限大と置いたが，より長時間の観測データに対する移動履歴推定を行う場合には，この値の設定によって計算時間や精度が大きく変化すると考えられるため，設定基準を設けることが重要となりうる。

## 謝辞

本研究を進めるにあたり，多くの御教示，熱心な御指導を賜りました美濃導彦教授，棕木雅之准教授に深く感謝致します．そして，日頃より親身なご指導を賜り，また本報告書の作成において多くの助言をいただきました川西康友博士に心より感謝致します．また，グループミーティングで多くの助言を頂きました歩行者センシンググループの皆様，ならびに美濃研究室の皆様にも深く感謝いたします。

## 参考文献

- [1] 美濃導彦: センシングウェブ：概念と課題 (特集：センシングウェブ), 人工知能学会誌, Vol. 24, No. 2, pp. 179–184 (2009).
- [2] 武田真人, 矢田紀子, 長尾智晴: パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡, 電子情報通信学会論文誌, Vol. J94-D, pp. 1631–1639 (2011).
- [3] 小林貴訓, 杉村大輔, 平澤宏祐, 鈴木直彦, 鹿毛裕史, 佐藤洋一, 杉本晃宏: 映像監視のための環境に適応的な異常検知ネットワーク, 電子情報通信学会論文誌, Vol. J94-D, pp. 1631–1639 (2011).
- [4] 西岡潔郁, 山崎俊彦, 相澤清晴: 広域・分散マルチサーベイランス映像における人物追跡の基礎検討, 電子情報通信学会論文誌, Vol. J94-D, pp. 2029–2031 (2008).
- [5] 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 (一般セッション, 実世界文字認識と理解), 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 111, No. 317, pp. 117–124 (2011).



- [6] Farenzena, M., Bazzani, L., Perina, A., Murino, V. and Cristani, M.: Person re-identification by symmetry-driven accumulation of local features., *CVPR*, IEEE, pp. 2360–2367 (2010).
- [7] Javed, O., Rasheed, Z., Shafique, K. and Shah, M.: Tracking Across Multiple Cameras With Disjoint Views., *ICCV*, IEEE Computer Society, pp. 952–957 (2003).
- [8] Lian, G., Lai, J.-H. and Zheng, W.-S.: Spatial-temporal consistent labeling of tracked pedestrians across non-overlapping camera views., *Pattern Recognition*, Vol. 44, No. 5, pp. 1121–1136 (2011).
- [9] Song, B. and Chowdhury, A. K. R.: Robust Tracking in A Camera Network: A Multi-Objective Optimization Framework., *J. Sel. Topics Signal Processing*, Vol. 2, No. 4, pp. 582–596 (2008).
- [10] Chen, K.-Y., Huang, C.-L., Hsu, S.-C. and Chang, I.-C.: Multiple Objects Tracking across Multiple Non-Overlapped Views., *PSIVT (2)* (Ho, Y.-S.(ed.)), Lecture Notes in Computer Science, Vol. 7088, Springer, pp. 128–140 (2011).
- [11] Kawanishi, Y., Wu, Y., Mukunoki, M. and Minoh, M.: Shinpuhkan2014: A Multi-Camera Pedestrian Dataset for Tracking People across Multiple Cameras, *20th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pp. 128–140 (2014).