

NAIST-IS-MT1151107

修士論文

分散カメラ間における グループ属性を用いた人追跡

守口 裕介

2013年2月7日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

守口 裕介

審査委員：

萩田 紀博 教授 (主指導教員)

加藤 博一 教授 (副指導教員)

浮田 宗伯 准教授 (副指導教員)

分散カメラ間における グループ属性を用いた人追跡*

守口 裕介

内容梗概

本研究では、視野に重畳のない分散カメラ群によって、広域を移動する人をそれぞれ追跡するために、新たに各人のグループ特徴量を利用する追跡法を提案する。この特徴量は、従来広く研究されてきた人のアピアランス以外の情報を表現している。このようにアピアランスとは独立な特徴量によって、見た目が類似している人の識別精度を向上させる。本手法では、カメラ間の移動においてグループ（行動を共にする1人以上の人の集まり）の関係は変化しないという仮説の下、その人が属しているグループの情報（グループ属性）をカメラ間人同定に用いるために、まず各カメラにおける追跡結果からグループと非グループそれぞれの軌跡同士の間関係を Support Vector Machine (SVM) によって学習、及び新たに得られた軌跡同士の間関係を求めグループ検出を行う。そして、グループ検出の結果から得られるグループ属性からグループ特徴量を算出し、色情報やカメラ間の移動時間を反映させた既存の特徴量と併用することによって、カメラ間で人を同定する。実験では、まず軌跡同士の間関係からグループ検出が平均83%の精度で行えることを確認した。また、グループ検出結果から得たグループ属性によってカメラ間人同定の精度が37%増すことを確認した。

キーワード

人追跡, 分散カメラ, グループ検出, Support Vector Machine

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1151107, 2013年2月7日.

Grouping People for Tracking across Non-overlapping Cameras*

Yusuke Moriguchi

Abstract

In this work, the groups of people are detected in each camera and then utilized as an additional clue for improving the robustness of people identification across non-overlapping cameras. It is assumed that if two persons belong the same group in a camera view, they may be in the same group also in another camera view. The grouping process is achieved by discriminatively classifying the trajectories of grouped people and non-grouped people with their spatio-temporal features of the trajectories. And then the spatio-temporal features are utilized with other features of each person (e.g. color histogram, transit time between cameras) for people identification across non-overlapping cameras. The experimental results demonstrate the improvement in people grouping and people identification by the proposed methods with a public dataset; the rate of group detection was 83 % on average, and the successful rate of people identification across cameras was improved by 37 %.

Keywords:

tracking, non-overlapping cameras, grouping, Support Vector Machine

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1151107, February 7, 2013.

目次

1. はじめに	1
2. 関連研究	4
3. 処理概要	8
4. 単眼カメラ視野内での人追跡	11
4.1 HOG 特徴量を利用した人検出	11
4.2 検出結果の大域的最適連結による人追跡	13
5. 単眼カメラ内のグループ検出	16
5.1 軌跡から得られる時空間特徴量	16
5.2 時空間特徴量を用いた各フレームにおけるグループ内ペア判定	20
5.3 各フレームにおけるグループ内ペア判定の結果を用いたグループ内ペア検出	21
6. グループ属性を用いたカメラ間人同定	22
6.1 カメラ間人同定に用いる従来特徴量	22
6.1.1 人の色ヒストグラム	22
6.1.2 カメラ間の移動時間	25
6.2 人のグループ属性	25
6.3 グループ属性と従来の特徴量を併用したカメラ間人同定	27
7. 実験と結果	28
7.1 単眼カメラ視野内での人追跡実験	32
7.1.1 HOG 特徴量を利用した人検出実験結果	32
7.1.2 検出結果の大域的最適連結による人追跡実験結果	34
7.2 単眼カメラ内のグループ検出実験	39
7.2.1 正解軌跡を用いたグループ検出実験結果	41
7.2.2 人追跡実験の結果の軌跡を用いたグループ検出実験結果	45

7.3	グループ属性を用いたカメラ間人同定実験	48
7.3.1	色ヒストグラム計算時の最良フレーム選択法	48
7.3.2	カメラ間人同定におけるグループ属性の効果	50
7.3.3	グループ検出結果を用いたカメラ間人同定実験結果	52
7.3.4	追跡結果を用いたカメラ間人同定実験結果	54
8.	おわりに	57
	謝辞	58
	参考文献	59
	付録	64
A.	本研究における用語集	64
B.	SVMについての説明	67

目 次

1	一般的な分散カメラの配置例.	2
2	視野の重畳があるカメラ群における三角測量による人同定の概念図.	3
3	監視カメラの映像例. 広域人追跡の実現には, 各カメラの視野内で撮影された人の追跡 (各破線) と, 各カメラで撮影された人の間で同一人物判定を行うカメラ間人同定 (各実線) が必要. 各枠はそれぞれグループを示しており, 赤枠内の2人は同一グループ.	3
4	色ヒストグラム取得の概念図.	5
5	視野の重畳があるカメラ群における三角測量による人同定の概念図.	5
6	処理全体の流れ. 赤色の枠で表示されている処理が提案手法にて新たに追加した処理.	10
7	ポジティブ・ネガティブそれぞれの学習画像の例.	12
8	人検出結果を使った人追跡の概念図. 各矩形が各フレームにおける検出結果, 各矢印が検出結果の連結を表す.	14
9	大域的最適連結の概念図 (文献 [6] より参照). 丸印が各フレームにおける検出結果, 青矢印が検出結果の連結を表す.	14
10	文献 [32] で提案されている時空間特徴量の中で, 2人の位置関係や軌跡の類似性を表した時空間特徴量. v_i と v'_i は並行, v_i と v''_i の長さは等しい.	17
11	文献 [32] におけるグループ内ペアの検出法. $F_1 \sim F_4$ をそれぞれ m ビンに分割されたヒストグラムにし, $F_1 \sim F_4$ をそれぞれ m 次元の特徴ベクトルで表したものと F_5 を組み合わせた $4m + 1$ 次元の特徴ベクトル f° を算出する.	18

12	3人以上のグループにおけるグループ内ペアの概念図. 学習時には赤矢印で示されたような距離の離れているグループ内ペアは学習しない. また, 黒矢印で示されたグループ内ペアが検出された場合, 赤矢印で示されたグループ内ペアもグループ内ペアとして検出する.	18
13	各フレームにおける時空間特徴量を用いたグループ内ペア判定の概念図. 6つの時空間特徴量を6次元特徴ベクトル $f_1 \sim f_N$ として各フレームにおいて算出し, SVMを用いて各フレームにおけるグループ内ペア判定を行う.	20
14	グループ内ペア検出の概念図. 各フレームにおける6次元特徴ベクトル $f_1 \sim f_N$ を判定した結果, グループ内ペアと判定されたフレームがある閾値を超えていた場合, その軌跡ペアをグループ内ペアとして検出する.	21
15	カメラ間人同定における Bhattacharrya 係数 L の比較例.	23
16	カメラ視野間の移動時間の分布図の例. 横軸がカメラ間の移動に費やした秒数, 縦軸がカメラ間の移動にその秒数費やした人数を表す.	24
17	グループ内人数に関するグループ特徴量導出の概念図. 人同士の間グループ内人数の差に対する特徴量を表す. 赤線がグループ内人数が一致した際の特徴量, 青線がグループ内人数の差が1人の際の特徴量.	26
18	実験に用いるカメラの配置図と映像例 (Learning, Recognition and Surveillance のホームページ [38] より参照).	29
19	Person Re-ID 2011 dataset の軌跡データの消失例. 人と人が重なって撮影されているフレームの軌跡データが保存されていないことが分かる.	30
20	各カメラの画像における人検出の例. 各矩形は検出結果を示す.	33
21	オクルージョンによる人の未検出の例.	33

22	人追跡の成功例. 矩形は追跡結果を, 矩形左上の数字は各軌跡に割り振られた ID を示す. 図 20 の人検出結果から検出評価値ノイズが低減できており, 人追跡が成功していることが分かる.	36
23	人検出の失敗による人追跡の失敗例. 左側が検出評価値ノイズによる誤追跡, 右側がオクルージョンによる未追跡を示す.	37
24	追跡結果の 1 軌跡が複数に分割される失敗例. ID41 の軌跡が, 次のフレームで ID48 に変化している.	38
25	複数軌跡の接近時に ID が入れ替わる失敗例. ID3 の軌跡が, グレーの服の人から赤色の服の人へ入れ替わっていることが分かる.	38
26	グループ内ペアの検出成功例 (ID1140 と ID1141) と非グループ内ペア (ID1140 と ID1329) の例. 左のグラフが同時に軌跡ペアが観測されたフレームにおける人の位置 (矢印の始点) と速度ベクトル (矢印) を示す. 右のグラフがその時刻における時空間特徴量とグループ内ペア判定の結果を示す.	42
27	グループ内ペアの誤検出例 (ID60 と ID1013) . 局所的にこの二人が接近しているため, グループ内ペアとして誤検出されている.	43
28	グループ内ペアの未検出例 (ID62 と ID73) . 局所的に二人が離れているために, グループ内ペアとして検出されていない.	43
29	一方のカメラ視野内のみでグループ内ペアと判定できる例. Camera a の画像のみで ID98 と ID1017 がグループ内ペアと判定するのは困難.	44
30	図 24 のように軌跡が複数の追跡結果に分割された場合の, グループ内ペア検出の概念図. 歩行者 i と歩行者 j がグループ内ペアであり, 歩行者 i に対する人追跡結果が軌跡 2 と軌跡 3, 歩行者 j に対する人追跡結果が軌跡 1 となっている場合, 軌跡 1 と軌跡 2, 軌跡 1 と軌跡 3 の両方がグループ内ペアとして検出された時, 歩行者 i と歩行者 j についてグループ内ペア検出が成功したとする.	46
31	追跡結果の誤検出が原因の, グループ内ペアの誤検出例 (ID18 と ID42)	47

32	6.1.1 の 3 つの手法から算出した色特徴量のスコア R^c を用いたカメラ間人同定の結果図 (手法 3 におけるフレーム数 $N = 5, 10, 20$) . 横軸は本人同士の間で得られたスコアの全候補 884 人中の順位を示し, 縦軸はその順位以内に入っている 342 人中の人数の割合を示す. 全順位における結果を左側のグラフに, 100 位以内の部分に着目した結果を右側のグラフに示す.	49
33	人追跡やグループ検出について正解データを用いた理想環境中において, 色特徴量とグループ特徴量を併用した際の, カメラ間人同定の結果図. 全順位における結果を左側のグラフに, 30 位以内の部分に着目した結果を右側のグラフに示す.	51
34	人追跡やグループ検出について正解データを用いた理想環境中において, 2 種類の従来特徴量 (色特徴量・時間特徴量) とグループ特徴量を併用した際の, カメラ間人同定の結果図. 全順位における結果を左側のグラフに, 30 位以内の部分に着目した結果を右側のグラフに示す.	51
35	色特徴量とグループ検出結果によるグループ特徴量を併用した際の, カメラ間人同定の結果図. 全順位における結果を左側のグラフに, 30 位以内の部分に着目した結果を右側のグラフに示す.	53
36	2 種類の従来特徴量 (色特徴量・時間特徴量) とグループ検出結果によるグループ特徴量を併用した際の, カメラ間人同定の結果図. 全順位における結果を左側のグラフに, 30 位以内の部分に着目した結果を右側のグラフに示す.	53
37	追跡結果を用いたグループ特徴量と色特徴量を併用した際の, カメラ間人同定の結果図. 全順位における結果を左側のグラフに, 30 位以内の部分に着目した結果を右側のグラフに示す.	56
38	追跡結果を用いたグループ特徴量と人の色特徴量・時間特徴量を併用した際の, カメラ間人同定の結果図. 全順位における結果を左側のグラフに, 30 位以内の部分に着目した結果を右側のグラフに示す.	56

39	入力パターンが特徴空間中で線形分離可能な場合の線形識別関数 $l(x)$ の算出の概念図.	69
40	入力パターンが特徴空間中で線形分離可能でない場合の線形識別関数 $l(x)$ の算出の概念図.	69
41	入力パターンが特徴空間中で線形分離可能でない場合に用いるヒンジ関数.	69

表 目 次

1	各カメラにおける正解軌跡を用いたグループ内ペア検出結果. それぞれ, 検出成功 (True-Positive), 誤検出 (False-Positive), 未検出 (False-Negative) のグループ内ペア数	41
2	7.1 章の人追跡の実験結果を用いたグループ内ペア検出結果. 括弧内の数値はグループ内ペア検出結果の失敗数の中で, 7.1 章の実験結果で見られた誤追跡や未追跡が原因となり, グループ内ペア検出が誤検出や未検出となってしまう数を示している.	46

1. はじめに

人々が安全・安心な生活を営むために、セキュリティに関する技術は必要不可欠である。その中でも、監視カメラを使った広域にわたる人追跡技術は、犯人の逃走経路の特定や車両や人などの流量解析に用いられる重要な技術である。

一般的に設置されている監視カメラは視野が固定されており、一台のカメラが撮影できる範囲は限定されているため、広域人追跡のためには複数台のカメラが連携しなければいけない。複数台のカメラを使って人追跡を行う際、各カメラの視野内での人追跡に加え、あるカメラに映った人物が別のカメラで映った人物と同一人物か判断する人同定を行う必要がある。一般的に監視カメラはコストの面から視野に重なるの無い配置 (図1) になっていることが多い。そのため、図2のような視野の重なる部分で同時に人物が観測されることを利用した安定な人同定ができない。また、一般的な監視カメラの画像は図3のように解像度が低いため、顔認識などの解像度が高くないと利用できない手法を用いるのは困難であり、色情報などの解像度が低くても利用できる情報から得た特徴量を用いてカメラ間人同定を実現しなければいけない。以上の問題点を解決するために、カメラ間人同定についての先行研究として個人のアピアランスに着目し、精度向上を図っているものが多数報告されている (サーベイ論文 [1] 参照)。しかし、精度の面で頑健なカメラ間人追跡技術は確立されておらず、更なる精度向上が望まれている。

そこで本研究では、カメラ間人同定において個人のアピアランス以外に利用できる特徴量として、例えば家族や友達などの、「位置関係や軌跡の類似性、同時観測性に着目することで検出される1人以上の人の集まり」を「グループ」と定義し、このグループ (図3内の各枠参照) に着目した特徴量を提案する。このグループは移動軌跡上において共に移動することが多く、カメラ間の移動においてもその関係 (位置関係や軌跡の類似性、同時観測性など) が変化しないことが多い。そこで、カメラ間の移動においてグループの関係が変化しないという仮定の下、その人が属しているグループの人数やグループ内全員の人のアピアランス情報といった情報を「グループ属性」と定義し、グループ属性から得られる特徴量 (これを「グループ特徴量」と定義する) を算出する。

このグループ特徴量は、個人のアピアランスから得られる特徴量と独立に定義

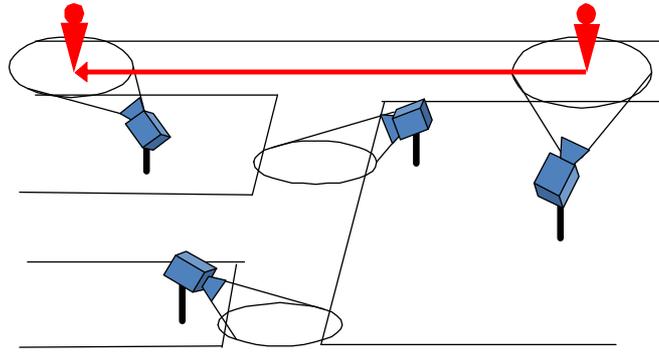


図 1 一般的な分散カメラの配置例.

できる．これは，例えば「赤色の服を着た人は2人組である可能性が高い」という事象は，一般的に起こらないと考えられるためである．そこで，グループ特徴量と従来手法の特徴量を併用することでカメラ間人同定の精度の向上を目指す．

提案手法を実現するためには，誰と誰がグループなのかを判定するグループ検出を行う必要がある．そこで，同一グループに属する人たちは似た軌跡を描いて移動する性質がある [2] ため，各カメラの追跡結果を用いてグループと非グループそれぞれの軌跡同士の関係を Support Vector Machine (SVM) [3] によって学習し，新たに得られた軌跡同士の関係を求めグループ検出を行う．以上のようなグループ検出手法を用いることで得られるグループ特徴量と従来手法の特徴量を併用しカメラ間人同定の精度向上を図る．

2章では，カメラ間人同定とグループ検出について関連研究を述べる．3章では提案手法の概要を説明し，この処理を3段階に分け，4章では単眼カメラ視野内での人追跡，5章では単眼カメラ内のグループ検出，6章ではグループ属性を併用したカメラ間人同定についての手法を提案する．7章では以上の提案手法についての実験を行い，8章で結論を述べる．

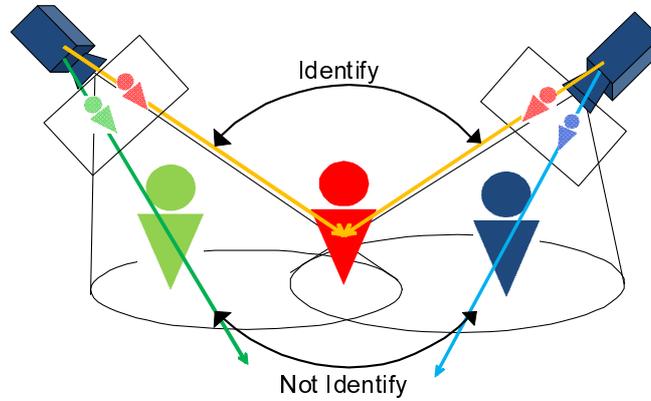


図 2 視野の重畳があるカメラ群における三角測量による人同定の概念図.

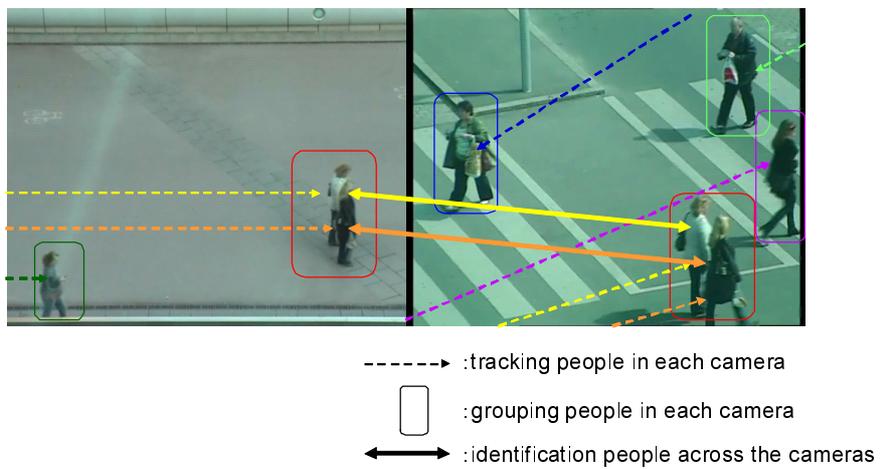


図 3 監視カメラの映像例. 広域人追跡の実現には, 各カメラの視野内で撮影された人の追跡 (各破線) と, 各カメラで撮影された人の中で同一人物判定を行うカメラ間人同定 (各実線) が必要. 各枠はそれぞれグループを示しており, 赤枠内の 2 人は同一グループ.

2. 関連研究

複数台の監視カメラを使った広域人追跡を実現するためには、単眼カメラ視野内での人追跡技術と複数台のカメラ間での人同定技術が必要となる。

単眼カメラ視野内での人追跡技術についての研究には文献 [4][5][6] などがある。このように単眼カメラ視野内での人追跡技術には多数の研究報告がなされているため、本研究ではこれらの研究で提案されている手法を元に、1台のカメラ視野内での人追跡を行う。

複数台のカメラ間での人同定について、個人のアピアランスに着目することで精度を向上させた研究が多数報告されている。個人のアピアランスの取得方法として、Hamdoun ら [7] の特徴点ベース、Wang ら [8] の HOG ベースなどの手法が提案されている。個人のアピアランスの中でも、人の向きの変化やカメラの視点変化に頑健なものとして色ヒストグラムの取り出し方を工夫する研究がされている。その中には、Farenzena ら [9] の HSV ヒストグラムを利用したもの、Gray ら [10] の服の模様のヒストグラムの取得など、様々な色ヒストグラムを取り出す研究が存在する。本研究でもこれらの研究と同様に、図 4 で示されるような人の色ヒストグラムを、カメラ間人同定に用いる特徴量の一つとして利用する。また、個人のアピアランスに着目した研究には他にも、各カメラの照明条件などによる色の変化を補正する手法を提案した研究 [13][14] や、SVM や Adaboost などの識別手法を工夫した研究 [11][12] など様々な研究が存在する。

カメラ間人同定において個人のアピアランスの特徴量だけでなく、カメラの位置関係などのシーンの情報を導入することで、人のカメラ間移動に費やした時間を表す時空間特徴量を併用できるため、カメラ間人同定の精度が向上できる（例えば、図 5 のように大半の人がカメラ間の移動に 60 秒費やすシーンを 10 秒で移動する人はいないという情報を用いることが出来る）。カメラ間の関係をモデリングし、時空間特徴量を得ることで人同定の精度を向上させた研究も多数されており、カメラ間追跡結果を元にカメラ間の関係をモデリングした研究 [15][16] や、追跡結果なしで関係モデリングを行う研究 [17][18][19]、また Wang ら [20] のように逐次的にカメラ間の関係をモデリングする研究などが挙げられる。これらの研究の中には、Ukita[19] の研究で行われているような、カメラ間の経路が複数ある

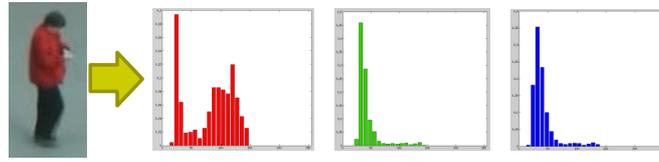


図 4 色ヒストグラム取得の概念図.

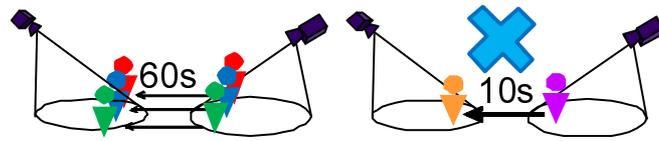


図 5 視野の重畳があるカメラ群における三角測量による人同定の概念図.

環境においてそれら全ての経路の移動時間をモデリングする研究も存在する．本研究では，監視カメラのように設置後に移動することのないカメラを用いたカメラ間人同定を行うので，これらの研究で提案されているようなカメラ間の関係モデリングの結果得られる時間特徴量を用いることで人同定の精度を向上する．

以上のようにカメラ間人同定の研究として個人のアピアランスに着目した研究やカメラ間の関係モデリングの研究が行われているが，アピアランスの特徴量や時空間特徴量を使ったカメラ間人同定技術は精度が十分とは言えず，更なるカメラ間個人同定の精度向上が望まれている．そこで本研究では，以上の研究で挙げられていた人同定に用いられる情報に加えて，新たにその人が属しているグループの情報であるグループ属性という情報に着目する．

「グループ」には様々な定義があり，例えば，社会学の分野で最も用いられている定義として，「複数行為者のあいだの相互行為や相互関係に規則性と持続性が見られ，彼らのあいだにある程度共通の志向が分有されている集合」[21] という定義が存在する．この定義は，学校や企業といった広い意味でのグループも含んでいるため，グループの中でも Cooley[22] の提唱した第一次集団（直接的接触に基づく親密な関係の社会集団，例：家族，友達等）に着目した．この第一次集団は，相互関係のある複数の人の集まりを指すが，本研究では各人の属する「グ

グループ」の情報を、カメラ間人同定における新たな情報として用いるため、「複数の人の集まり」という情報だけでなく、「一人で行動する人」という情報も、複数人のグループと区別できる点でカメラ間人同定に有用な情報であると考えられる。また、この「グループ」を図3のような監視カメラの映像を用いて検出するためには、同一グループに属する人達の位置関係や軌跡の類似性、同時観測性に着目することで、他グループとの判別できると考えられる（例えば、図3のように、他グループの人と比べて、同一グループの人は近距離で移動する）。したがって、本研究では「位置関係や軌跡の類似性、同時観測性に着目することで検出される1人以上の人の集まり」を「グループ」と定義した。そして、この各人の属するグループの情報を「グループ属性」と定義し、カメラ間人同定に用いる新たな情報とした。

グループ属性をカメラ間人同定に用いるためには、グループ検出を行う必要がある。グループ検出に関する研究には静止画を利用した研究[23]もあるが、本研究では監視カメラから得られる映像を用いるため、動画を利用したグループ検出手法に着目する。動画を利用することで、各フレームにおける画像の情報や時系列に沿った人の動きをグループ検出に利用できる一方で、一般的に静止画と比べて監視カメラの動画は人が小さく撮影されるため、顔認識による年齢の推定などが難しく、取り出すことのできる特徴が限定される。グループ検出の研究で重要となる考え方に、Social Force Model[24]と呼ばれるモデルがある。これは、移動時に他人に衝突しないようにするなどの人同士が与える影響[25]をモデル化したものであり、その中には同一グループに属する人々が互いに与えている影響（他グループの人との距離と比べ同一グループ内の人との距離は近い等）も存在し、この影響についての研究も報告されている[2][4][26]。このSocial Force Modelを使ってグループ検出を行った研究もある[27][28]。その中でも、Yucelら[2]は文献[29][30][31]にて提案されているLaser Range Finder(LRF)を使った追跡技術によって得られた軌跡を用いて、グループ検出を行った。この研究では、同一グループに属する軌跡同士の関係を学習させておき、新たに得られた軌跡がグループらしいかどうか判定した。Yamaguchiら[32]は、文献[4][5][6]で提案されているようなカメラを使った追跡結果から得られる軌跡の関係（位置関係や類似性、

同時観測性) を特徴量として用いてグループ検出を行った。

本研究では、文献 [32] で提案された特徴量の問題を解決した時空間特徴量を用いてグループ検出を行い、その結果から得られたグループ特徴量と従来研究の特徴量を併用しカメラ間人同定の精度向上を図る。

3. 処理概要

本研究で提案する手法の概要についてまとめる．図6にて紫色の枠で示しているように，カメラ間人同定における処理は大別して，単眼カメラ視野内で各人の情報を取得する処理と，複数カメラ間にて取得した情報を比較し人同定を行う処理となる．提案手法は，従来のカメラ間人同定手法に図6の赤色の枠で表示されている処理を追加したものとなっている．この図のように，従来手法で利用されている人固有の特徴量と観測領域固有の特徴量に加え，新たにグループ特徴量という従来の特徴量とは独立な特徴量を利用し，カメラ間人同定の精度を向上する．本研究では，この提案手法における全体の処理を，「単眼カメラ視野内での人追跡(4章)」，「単眼カメラ内のグループ検出(5章)」，「グループ属性を用いたカメラ間人同定(6章)」の3処理に分ける．

単眼カメラ視野内での人追跡

まず，単眼カメラ視野内での人検出及び人追跡を行う．各フレームの画像から人物の映っている箇所を取り出すために，各カメラに固有の人・背景の画像と各カメラ共通の人の画像を学習サンプルとし，そのHOG特徴量[33]を計算しSVMを用いて学習させることで，人検出を各カメラ独立に行う(4.1章)．この人検出によって各フレームの人の位置や色情報が得られるため，これらの情報から文献[6]で提案されている手法を用いて，各フレームにおける検出結果を大域的最適的に連結することで人追跡を行う(4.2章)．

単眼カメラ内のグループ検出

本研究で，カメラ間人同定における新たな特徴量として提案しているグループ特徴量を求めるために，単眼カメラ視野内での人追跡により得られる軌跡の位置・時間情報を使ってグループ検出を行う．このグループ検出では，同時に単眼カメラ視野内に存在する全軌跡集合中の任意の軌跡ペアが，同一グループ内に存在するか否かを調べる．この処理を，以下ではグループ内ペア検出と呼ぶ．そのために，まず文献[32]で提案されている特徴量を改良し，軌跡ペアに対して位置関係や軌跡の類似性，同時観測性に着目した時空間特徴量を各フレームにおいて取得する(5.1章)．この時空間特徴

量をグループ内ペアとそれ以外のペアの軌跡について求め、SVMによる学習、及びテストデータから得られた軌跡ペアの各フレームにおけるグループ内ペア判定を行い (5.2 章)、グループ内ペアと判定されたフレーム数が閾値以上となった軌跡ペアをグループ内ペア検出の結果とする (5.3 章)。

グループ属性を用いたカメラ間人同定

単眼カメラ視野内で得られた各人の情報を用いて特徴量を算出し、異なるカメラで撮影された任意の 2 人の間で特徴量の比較を行うことで、その 2 人の同一人物らしさを表す人同定用のスコアを算出し、カメラ間人同定を行う。まず、カメラ間人同定の手法として従来利用されている特徴量である、人の色ヒストグラムという人固有の特徴量と、カメラ間の移動時間という観測領域固有の特徴量から、人同定用のスコアを算出する。人の色ヒストグラムの導出には、人追跡の結果から得られる検出枠内の色情報を使い、色ヒストグラムの類似性を表した Bhattacharyya 係数を用いて、人の色分布の類似度を表すスコアを求める (6.1.1 章)。カメラ間の移動時間の導出には人追跡の結果得られるカメラ視野への入出時間を使い、シーン間の移動時間を確率的に求めておくことでスコアを求める (6.1.2 章)。以上の従来カメラ間人同定に利用されている特徴量に加え、提案手法では新たにグループ特徴量から人同定用のスコアを求める。そこで、5.3 章で求めたグループ検出の結果から、その人の属するグループの情報であるグループ属性を求め、同一グループ内の人数 (グループ内人数特徴量) やグループ内に属する全員の色情報 (グループ内色特徴量) から、人同定用のスコアをそれぞれ求める (6.2 章)。以上 4 種類の特徴量 (色特徴量・時間特徴量・グループ内人数特徴量・グループ内色特徴量) から算出した人同定用のスコアを使って、カメラ間人同定を行う (6.3 章)。

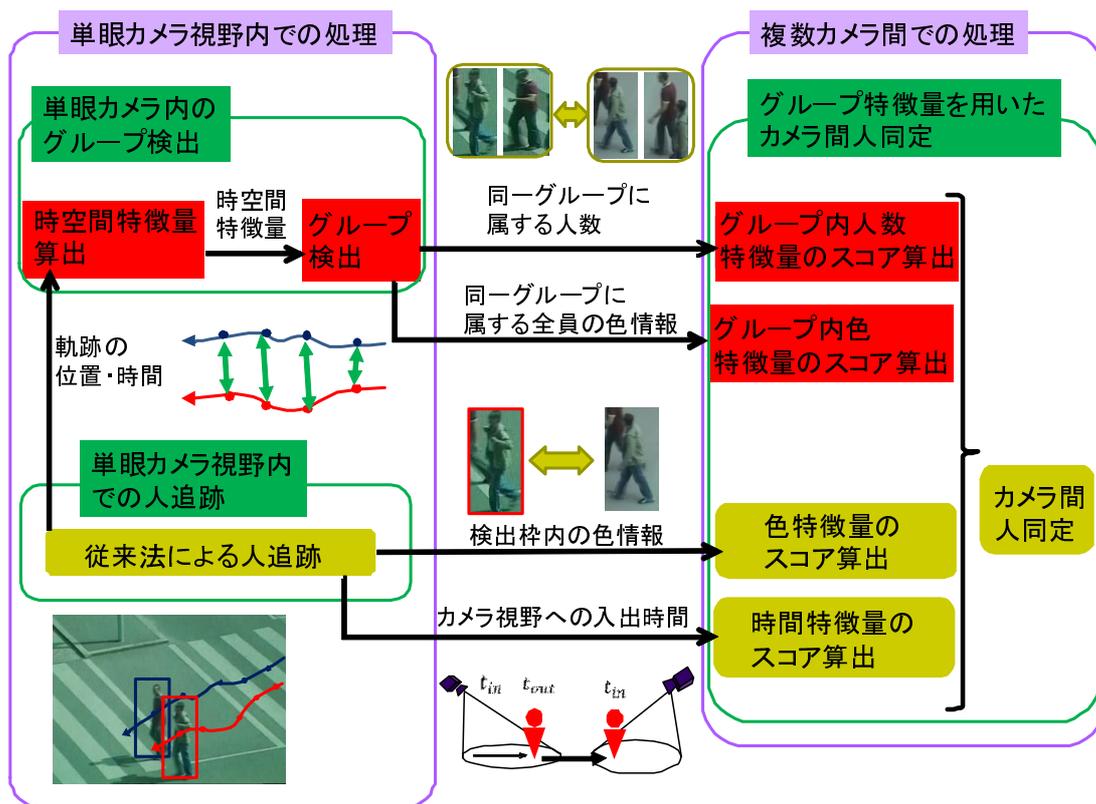


図 6 処理全体の流れ. 赤色の枠で表示されている処理が提案手法にて新たに追加した処理.

4. 単眼カメラ視野内での人追跡

4.1 HOG 特徴量を利用した人検出

単眼カメラ視野内で人追跡を行うためには、各フレームにおける画像（例：図 3）から人を検出する必要がある。

本研究では、この人検出を行うために、一般物体認識の研究分野で用いられている特徴量の中で HOG 特徴量に着目した。HOG は入力画像を局所領域に分割し、その領域内の輝度の勾配方向をヒストグラム化した特徴量である。この HOG 特徴量を使った一般物体認識には以下の特徴がある。

- 入力画像全体に対する特徴量を記述するため、SIFT 特徴量などの特徴点ベースの手法と比べ、物体全体の大まかな形状を記述することが可能
- ヒストグラム化の際に局所領域ごとに正規化を行うため、計算領域内の局所的な照明変動などの変化に頑健

本研究は監視カメラの映像を使った人追跡を対象としている。ここで、一般的に監視カメラの映像は、図 3 のように観測領域に出入りする人の全体を撮影していることから、人全体の大まかな形状が分かる点や、また設置後長期間に渡って撮影し続けるため、撮影時間によっては局所的な照明変動が起これる点を考慮し、本研究では人検出に HOG 特徴量を用いる。

画像全体の中から、大きさを変化させた矩形領域をラスタースキャンにより取り出し、その矩形領域内の HOG 特徴量を計算、その HOG 特徴量の人らしいかどうかを判定する。HOG 特徴量の人らしいかを判定する識別器として本研究では SVM¹を用いる。この SVM は汎化性能が高い識別器であり、本研究で用いる監視カメラのように、未学習の人物が多数撮影視野に入出する場合においても有効であるため、本研究では識別器に SVM を用いた。

また、本研究では SVM のライブラリとして、一般的に広く用いられている LIBSVM[34] を用いた。

¹SVM についての詳細は、付録 B で説明する

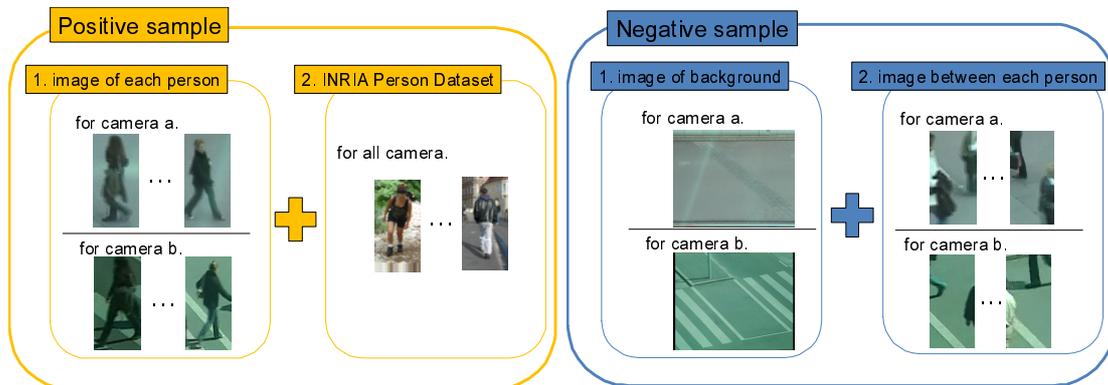


図 7 ポジティブ・ネガティブそれぞれの学習画像の例.

ここで、人の映り方がカメラによって異なる点を考慮し、本研究ではSVMに学習させるHOG特徴量を取得するための画像（以下、学習画像とする）を各カメラ独立に用意した。この学習画像はポジティブサンプル、ネガティブサンプルそれぞれ以下に挙げたものを利用した（図7参照）。

ポジティブサンプル

1. 各カメラの学習用画像中の人を囲んだ矩形領域内の画像（図7中のポジティブサンプルの左側）
2. INRIA Person Dataset[33]内のポジティブサンプル用画像（図7中のポジティブサンプルの右側）

ネガティブサンプル

1. 各カメラの撮影視野内に人が映っていない背景画像の検出用矩形領域を当てはめる箇所の画像（図7中のネガティブサンプルの左側）
2. 各カメラの学習用画像中で人と人之間を囲んだ矩形領域内の画像（図7中のネガティブサンプルの右側）

ポジティブサンプルとして、各カメラの撮影環境に特化した情報としてそのカメラ視野内固有の人の映り方を学習させるために、各カメラ視野内で撮影された画像の中から人の映っている部分の矩形領域を抽出し、その領域の画像を学習さ

せた。また、オクルージョンの影響により、検出すべき人の一部が隠蔽されたり、背景に人や物が映りこむことで、背景が複雑になった場合の汎化性能を向上させるために、各カメラ共通のポジティブサンプルとして INRIA Person Dataset のポジティブサンプル用画像も学習させた。

ネガティブサンプルとして、監視カメラの性質として設置された後に位置や角度の変化が生じにくいいため、各カメラの撮影環境に特化した背景画像を学習させるために、検出用の矩形領域を当てはめる箇所には人が映っていない際の画像を学習させた。また、複数の人が接近して移動する時に、その複数人を個別に検出するために、人と人との間を囲んだ矩形領域内の画像もネガティブサンプルとして学習させることで、その領域の人らしさを表す SVM の評価値を下げた。

このような学習画像を SVM に学習させ、検出用画像中の矩形ごとに、Maji ら [35] の提案した Fast Intersection Kernel SVM による評価値（ポジティブサンプルらしさが高い場合に高くなる評価値）を算出する。この評価値には、HOG 特徴量や評価値の算出過程でノイズ（これを「検出評価値ノイズ w 」と定義する）が生じる。この検出評価値ノイズ w を取り除くために平滑化を行い、その後平滑化した評価値に対してピーク検出し、検出されたピークの矩形の位置とサイズを各フレームにおける人検出結果（図 8 中の矩形）とする。

4.2 検出結果の大域的最適連結による人追跡

4.1 章の人検出結果によって、各フレームの画像中での人の位置とサイズが得られている。この各フレームにおける人の位置と矩形領域内の色情報から、文献 [6] で提案されている手法を用いて、各フレームにおける検出結果を大域的最適的に連結することで人追跡を行う（図 8 参照）。文献 [6] で提案されている手法の概要を以下に示す（図 9 参照）。

1. 全フレームにおける全検出結果を隣接フレーム間で連結して、離散非環グラフを生成する。このグラフはノード（図 9 中の丸印）が検出結果であり、リンク（図 9 中の青矢印）がその検出結果の連結コスト（連結されている検出結果が類似しているほど小さなコスト）を持つ。

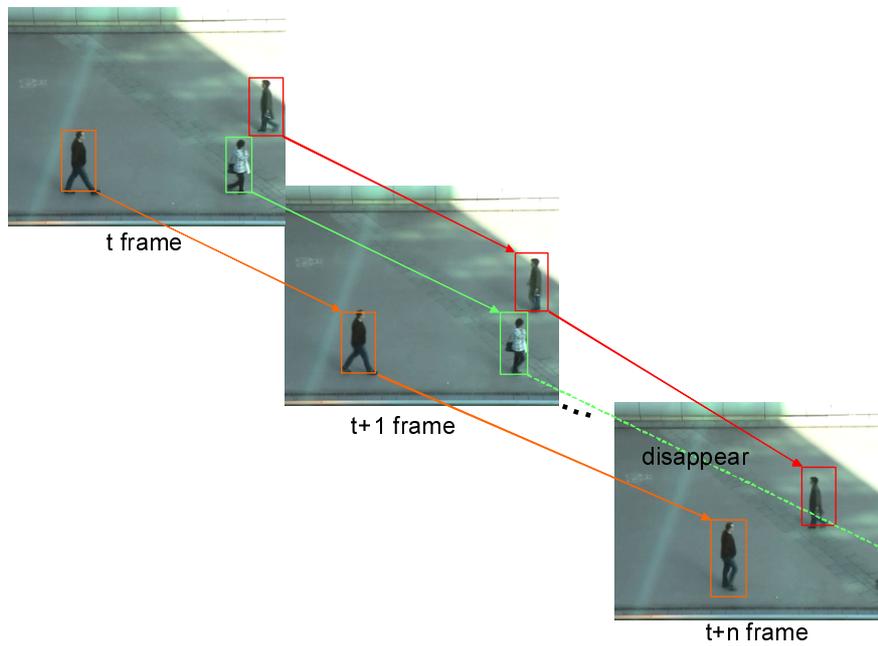


図 8 人検出結果を使った人追跡の概念図。各矩形が各フレームにおける検出結果，各矢印が検出結果の連結を表す。

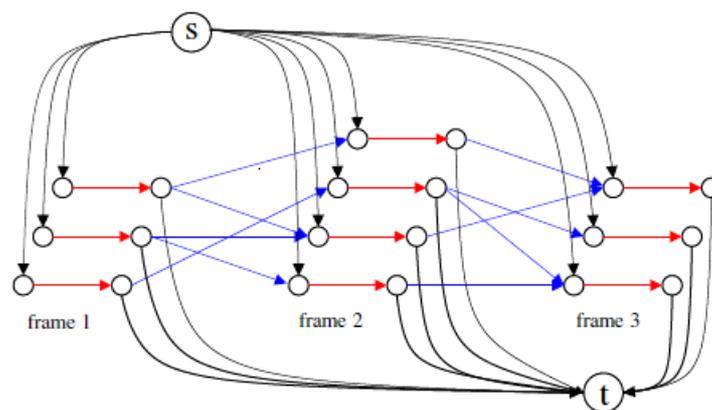


図 9 大域的最適連結の概念図（文献 [6] より参照）。丸印が各フレームにおける検出結果，青矢印が検出結果の連結を表す。

2. この離散非環グラフにおいて、反復的に最短経路探索を適用して、各追跡結果を抽出することによって、全検出結果における大域的最適性を保障した複数人追跡を行う。

本手法では最短経路探索にダイクストラ法を用いた。

5. 単眼カメラ内のグループ検出

5.1 軌跡から得られる時空間特徴量

本研究では、グループ検出を行うために、グループ内ペア検出を行う。このグループ内ペアとは、「グループ（一人のグループを除く）の構成員中の任意の2人」を指しており、例えば図12のような3人のグループの場合、 p_1 と p_2 、 p_2 と p_3 、 p_3 と p_1 の3つのグループ内ペアが存在する。こういった3人以上のグループ検出する際にも、 p_1 と p_2 、 p_2 と p_3 、 p_3 と p_1 の3つのグループ内ペアが検出されると、 p_1 と p_2 と p_3 の3人は同一グループに属すると分かるため、本研究においては、問題を単純化するために、まず任意の2人に対してグループ内ペアであるかを判定し、グループ内ペアの検出結果から、構成員の重複を考慮することで3人以上のグループを検出する手法をとる。

軌跡のグループ内ペア判定をする指標として、文献[32]では同一グループに属する人同士が他グループの人同士より接近して移動することや、描く軌跡が類似すること、同時に観測領域内に入出することに着目し、5つの時空間特徴量を提案した。歩行者 i の平面上での位置と速度をそれぞれ p_i 、 v_i 、 N_i フレーム間で観測領域内に存在した人が観測されたフレーム t の集合を $T_i = [t_{i,1}, t_{i,2}, \dots, t_{i,N_i}]$ とした時、歩行者 i と歩行者 j について5つの時空間特徴量($F_1 \sim F_5$)を以下のように定義する(図10参照)。

(F_1) 2人間の距離： $|p_i - p_j|$

(F_2) 2人の速度の絶対値の差： $||v_i| - |v_j||$

(F_3) 2人の速度の向きの差： $|\arctan(v_i) - \arctan(v_j)|$

(F_4) 1人の速度の向きと2人の相対位置の向きの差： $|\arctan(p_i - p_j) - \arctan(v_i)|$

(F_5) 軌跡ペアの同時観測割合： $|T_i \cap T_j| / |T_i \cup T_j|$

文献[32]では、この5つの時空間特徴量を各フレームにおいて算出し、 $F_1 \sim F_4$ をそれぞれ m ビンに分割されたヒストグラムにし、 F_5 をそれぞれ m 次元

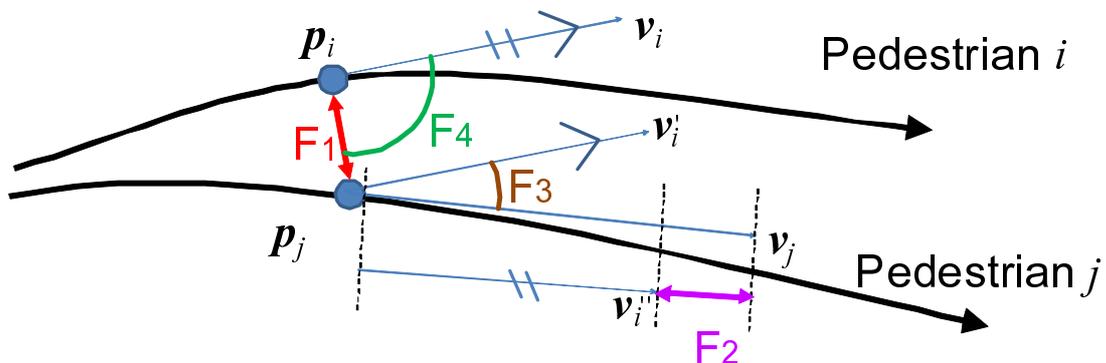


図 10 文献 [32] で提案されている時空間特徴量の中で、2 人の位置関係や軌跡の類似性を表した時空間特徴量. v_i と v_i' は並行, v_i と v_i'' の長さは等しい.

の特徴ベクトルで表す. こうして算出される特徴ベクトルと F_5 を組み合わせた $4m + 1$ 次元の特徴ベクトル f^o を, 各軌跡ペアに対して求める (図 11 参照). 学習時には, グループ内ペアとそうでないペアの軌跡に対して f^o を求め, それぞれをポジティブサンプル・ネガティブサンプルとして学習する. そしてグループ内ペア検出時には, その学習結果を元にグループ内ペア検出を行った. しかし, この時空間特徴量には次の問題点があった.

1. 各フレームにおける $F_1 \sim F_4$ の結果を全フレーム間でヒストグラム化してしまうため, 各フレームにおける時空間特徴量間の共起性 (例: 一定の距離を保つために, 距離が離れると近づこうとする等) がグループ内ペア検出結果に反映されない.
2. F_4 の計算に速度の向きを使用する際に, 軌跡ペアのうち一方の速度しか使用しないため, 使用する速度によって F_4 のスコアが変動する.
3. 一般的にグループ内ペアの F_1 は小さくなるのが期待されるが, 3 人以上のグループにおいてグループ内ペアの時空間特徴量を計算する際, 距離の離れた軌跡ペア (例: 図 12 の赤矢印のペア) の F_1 が大きくなってしまうため, グループ内ペア検出が失敗してしまう. また, 学習時においても, 値

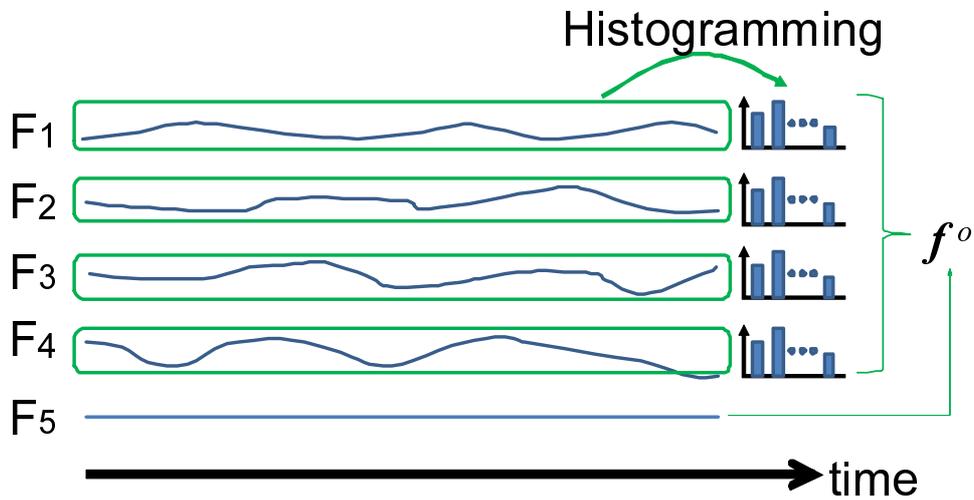


図 11 文献 [32] におけるグループ内ペアの検出法. $F_1 \sim F_4$ をそれぞれ m ビンに分割されたヒストグラムにし, $F_1 \sim F_4$ をそれぞれ m 次元の特徴ベクトルで表したものと F_5 を組み合わせた $4m + 1$ 次元の特徴ベクトル f^o を算出する.

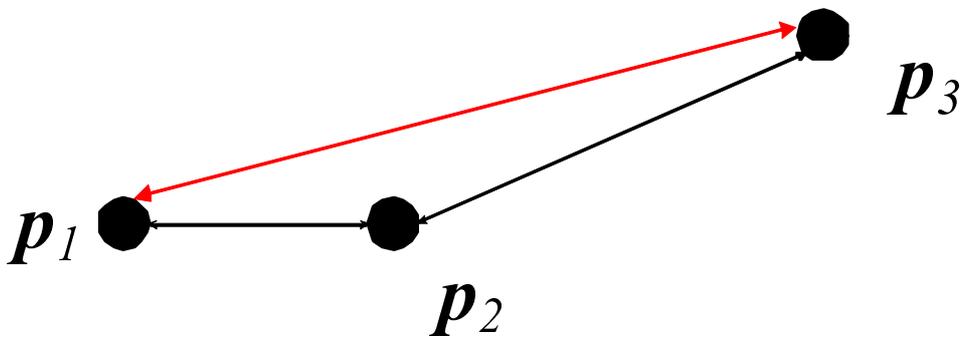


図 12 3人以上のグループにおけるグループ内ペアの概念図. 学習時には赤矢印で示されたような距離の離れているグループ内ペアは学習しない. また, 黒矢印で示されたグループ内ペアが検出された場合, 赤矢印で示されたグループ内ペアもグループ内ペアとして検出する.

の大きい F_1 をグループ内ペアの F_1 として学習するため、そういった値が SVM の判別性能に悪影響を与える可能性がある。

4. 静止している人の速度を求め F_3 や F_4 を導出する際に、速度の向きが安定しない。

以上の問題点を解決するために、本研究ではそれぞれの問題点を以下の手法を用いて解決した。

1. 各フレームにおける時空間特徴量間の共起性をグループ内ペア検出結果に反映するために、各フレームにおける時空間特徴量を1つの特徴ベクトルとして算出する (図 13 参照)。また、 $F_1 \sim F_4$ と同様に各フレームにおける F_5 を求める方法として、 t フレームでの F_5 の計算に用いる T_i を前後 T^f フレーム間に限定する方法をとる。つまり $t - T^f$ フレームから $t + T^f$ フレームの間で F_5 の計算を行う。
2. F_4 を2つの特徴量に分け、 v_i の進行方向を計算に用いる $F_{4,i}$ と v_j の進行方向を計算に用いる $F_{4,j}$ のどちらも時空間特徴量として算出する。
3. 3人以上のグループに対して、学習時は最も近距離にいる軌跡ペアの特徴量のみを学習させる (例：図 12 の黒矢印のペアの特徴量のみを学習させる)。グループ内ペアの検出時には、検出されたグループ内ペアの構成員間に重複があった場合、それらのグループ内ペアの構成員全員を同一グループに属すると判定することで、3人以上のグループの検出も行う (例：図 12 の黒矢印で示された2つの軌跡ペアがグループ内ペアと検出された場合、赤矢印の軌跡ペアもグループ内ペアとして検出する)。
4. あるフレームにおける速度が閾値以下となる人については、そのフレームにおけるグループ内ペア判定を行わない。

以上の解決法を用いて、新たに定義された5種類6つの時空間特徴量 ($F_1, F_2, F_3, F_{4,i}, F_{4,j}, F_5$) を用いてグループ内ペア検出を行う。

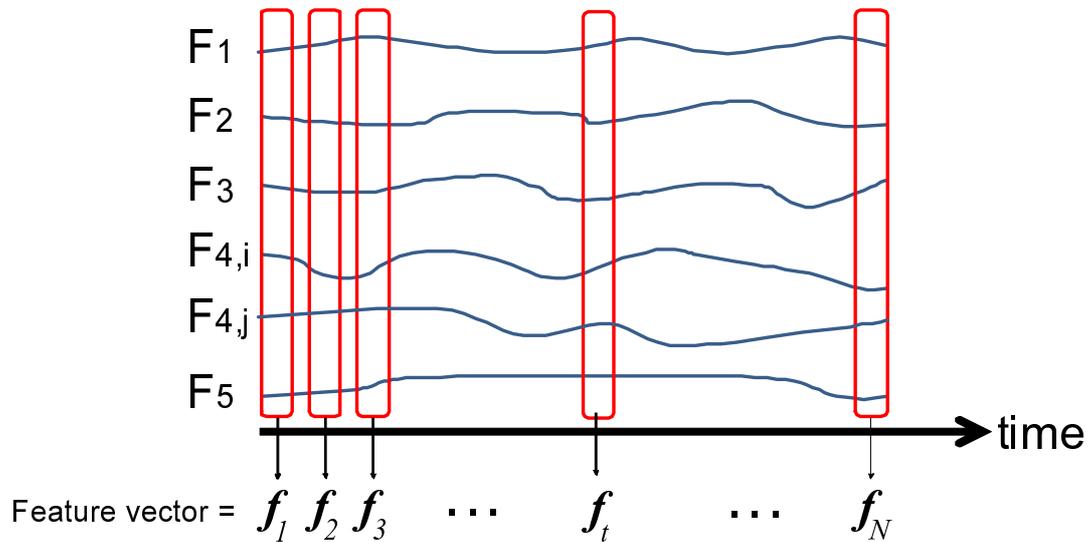


図 13 各フレームにおける時空間特徴量を用いたグループ内ペア判定の概念図. 6つの時空間特徴量を6次元特徴ベクトル $f_1 \sim f_N$ として各フレームにおいて算出し, SVM を用いて各フレームにおけるグループ内ペア判定を行う.

5.2 時空間特徴量を用いた各フレームにおけるグループ内ペア判定

5.1章で定義した時空間特徴量を用いて, 軌跡ペアの各フレームにおけるグループ内ペア判定を行う (図 13 参照). この判定には, 識別器として SVM を用いる.

学習時には, グループ内ペアとそうでないペアの軌跡の各フレームにおける時空間特徴量を6次元の特徴ベクトル (図 13 中の $f_1 \sim f_N$) とし, それぞれポジティブサンプル・ネガティブサンプルとして SVM に学習させる. 検出時には, 任意の軌跡ペアに対して各フレームにおける時空間特徴量を求め, 学習済みの SVM により各フレームのグループ内ペア判定結果を得る. このグループ内ペア判定は, グループ内ペアか否かの二値判定となる.

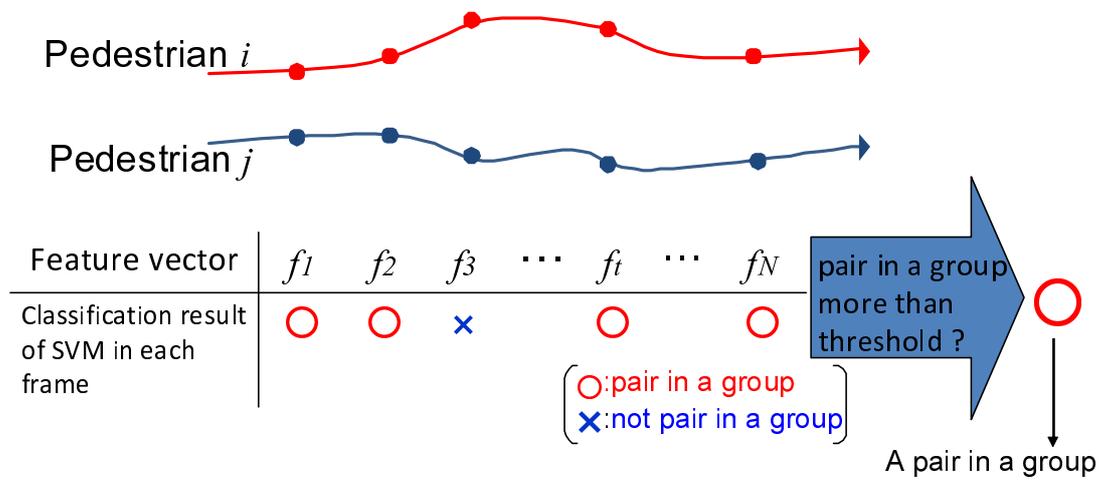


図 14 グループ内ペア検出の概念図. 各フレームにおける 6 次元特徴ベクトル $f_1 \sim f_N$ を判定した結果, グループ内ペアと判定されたフレームがある閾値を超えていた場合, その軌跡ペアをグループ内ペアとして検出する.

5.3 各フレームにおけるグループ内ペア判定の結果を用いたグループ内ペア検出

各フレームにおけるグループ内ペア判定結果を用いて, 軌跡ペア全体に対してグループ内ペア検出を行う.

5.2 章の手法によって各フレームでのグループ内ペア判定した結果, グループ内ペアと判定されたフレーム数が閾値以上となった軌跡ペアをグループ内ペア検出の結果とする. その後, 5.1 章で述べたように, 検出されたグループ内ペアの構成員間の重複を調べ, 重複が発見されたグループ内ペアの構成員全員を同一グループに属すると判定することで, 3 人以上のグループについても検出する.

6. グループ属性を用いたカメラ間人同定

単眼カメラ視野内で得られた各人の情報を用いて特徴量を算出し、異なるカメラで撮影された任意の2人の間で特徴量の比較を行うことで、その2人の同一人物らしさを表す人同定用のスコア R を算出し、カメラ間人同定を行う。

6.1 カメラ間人同定に用いる従来特徴量

カメラ間人同定の手法として従来利用されている特徴量である、人の色ヒストグラムという人固有の特徴量と、カメラ間の移動時間という観測領域固有の特徴量から、人同定用のスコアをそれぞれ求める。

6.1.1 人の色ヒストグラム

文献 [9][10][11][12][13][14] 等の多数の研究で行われているように、人追跡の結果得られる矩形領域内の色情報を使い、複数のカメラで検出された人の色ヒストグラムの類似度を算出する。2つの色ヒストグラム間の類似度の算出には、Mean-Shift法を用いた非剛体対象追跡の研究 [36] 等の多数の研究でヒストグラムの類似度の算出に用いられている、Bhattacharrya 係数を用いた。任意の2人の色ヒストグラムの類似度を求めるために、人の色ヒストグラムを RGB の各チャンネルに対して8階調に分割して求め、得られた24 (= 8 × 3) 階調のヒストグラムを正規化したものをそれぞれ A^c と B^c とした時、 A^c と B^c の類似度を表す Bhattacharrya 係数 L は、以下の式により算出される。

$$L = \sum_{u=1}^{24} \sqrt{A_u^c B_u^c} \quad (1)$$

この Bhattacharrya 係数 L は 0~1 の値をとり、値が高いほどヒストグラム間の類似度が高いということを表す。

人追跡の結果得られる各人に対して、各追跡結果の検出矩形領域内に映りこんでいる背景の色情報を取り除くために検出矩形領域の中心部分を取り出し、その部分の色ヒストグラムを計算し、Bhattacharrya 係数を求める。カメラ間人同定に

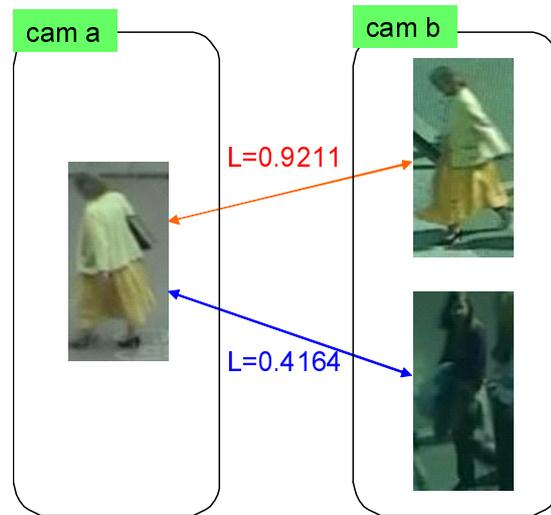


図 15 カメラ間人同定における Bhattacharrya 係数 L の比較例.

において、異なるカメラで撮影された場合においても、同一人物間の Bhattacharrya 係数は別人との間の Bhattacharrya 係数と比べて高くなる（図 15 参照）ことが期待されるため、これを色特徴量のスコア $R^c(=L)$ とする。また、人の色ヒストグラムを取得する際に、人追跡の結果を用いるため、複数フレームにおける人の色情報を用いることができるが、この色情報から人の色ヒストグラムを取得する方法として、本研究では次の 3 手法を挙げた。

1. その人が検出されたフレーム全ての色ヒストグラムを平均し正規化する。
2. その人の色ヒストグラムとその人が検出されたフレーム付近で検出された人物の色ヒストグラムの間で Bhattacharrya 係数を計算し、Bhattacharrya 係数が最小となるフレームにおける色ヒストグラムを正規化する。
3. その人の色ヒストグラムとその人が検出されたフレーム付近で検出された人物の色ヒストグラムの間で Bhattacharrya 係数を計算し、Bhattacharrya 係数の小さいものから n フレーム分だけ取り出し、その n フレームにおける色ヒストグラムを平均し正規化する。

手法 1 は、軌跡全体の色情報を平均することで、人が移動する際の様々な見え

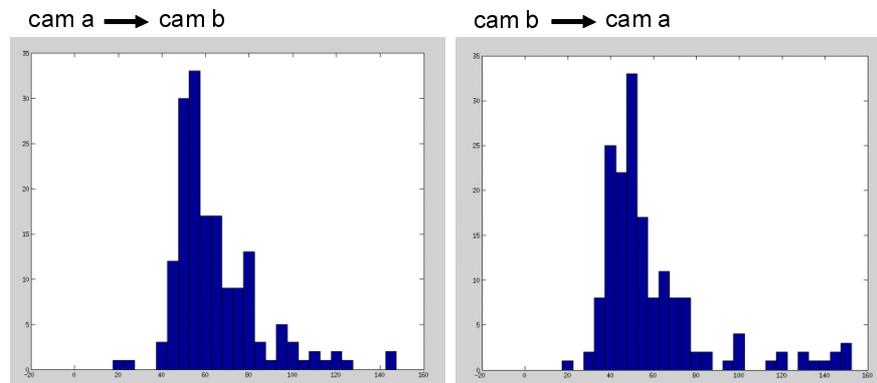


図 16 カメラ視野間の移動時間の分布図の例. 横軸がカメラ間の移動に費やした秒数, 縦軸がカメラ間の移動にその秒数費やした人数を表す.

方をまとめた総合的な色ヒストグラムを表現することを目的としている. 手法2,3は, カメラ間の移動において, あるカメラ視野内でその人と同じタイミングに検出された人物が, 他方のカメラで同時に検出される可能性が高いという仮定のもと, カメラ視野内でその人の前後で検出された人物との Bhattacharyya 係数を算出し, その類似度が低くなるフレームの色ヒストグラムを用いる事で, カメラ間人同定における他者との識別性能を向上させることを目的としている. また, 手法2は他者との類似度が最小になる1フレームのみの色ヒストグラムを用いる事で他者との識別性能を重視しているが, 他者の映り方と最も類似度の低い1フレームの画像のみを利用するため, そのフレームの色情報が取得失敗していたために他者間の類似度が低くなっていた場合, 自身との類似度も低くなってしまふ恐れがある. そこで, 手法3では類似度が低い n フレーム分のヒストグラムを平均することで, 手法2と比べ他者との識別性能は下がるが, 色情報の取得失敗による自身との類似度の低下を軽減し, 自身との間で計算される R^c を安定して高くすることを目的としている.

6.1.2 カメラ間の移動時間

文献 [15][16][17][18][19][20] 等の多数の研究でカメラ間の関係をモデリングし、その結果得られたカメラ間の移動にかかる時間の情報を用いることでカメラ間人同定の精度を向上させている。これらの研究で行われているように、人追跡の結果得られるカメラ視野への入出時間を使い、カメラ間の移動時間という観測領域固有の特徴量のスコア R^t を求める。

カメラ間の移動時間の分布図の例を、図 16 に示す。また、本研究ではカメラ間の経路が一つだけ存在するシーンで実験を行っており、そのため図 16 から分かるように、カメラ間の移動時間の分布は1つのガウス分布で近似できるような分布となる。そこで、カメラ間の移動時間の期待値 m^t と分散 δ^{t^2} を求め、確率的にカメラ間の移動時間のガウス分布 $g^t(t; m^t, \delta^{t^2})$ を求めておく。

カメラ間人同定においては、各カメラにおけるカメラ視野への出入りの時間から、各人のカメラ間の移動時間を求めることができる。ここで、あるカメラ a で検出された歩行者 i の視野からの退出時間を $t_i^{a,out}$ 、別カメラ b で検出された歩行者 j の視野への進入時間を $t_j^{b,in}$ とした時、以下の式のように、カメラ間移動時間 $t_j^{b,in} - t_i^{a,out}$ を算出済みのガウス分布 $g^t(t; m^t, \delta^{t^2})$ に代入することで、観測領域固有の特徴量のスコア R^t を求める。

$$R_{i,j}^t = g^t(t_j^{b,in} - t_i^{a,out}; m^t, \delta^{t^2}) \quad (2)$$

6.2 人のグループ属性

グループ検出の結果から、その人が誰とグループなのかという情報であるグループ属性が分かる。その人が誰ともグループでない（1人で行動している）という情報もグループ属性とする。ここで、カメラ間の移動においてこのグループ属性は変化しないという仮定の下、その人のグループ属性をカメラ間人同定に用いる。このグループ属性の中でも、本研究では以下の2点に着目して、カメラ間人同定における新たな特徴量であるグループ特徴量を求め、人同定用のスコアを算出する。

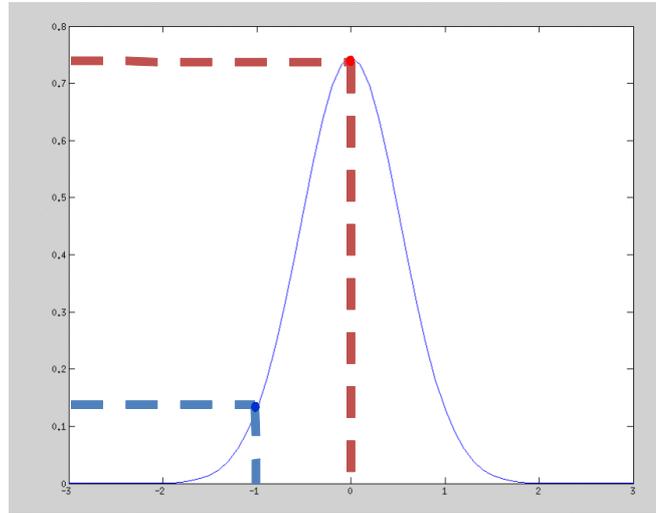


図 17 グループ内人数に関するグループ特徴量導出の概念図. 人同士のグループ内人数の差に対する特徴量を表す. 赤線がグループ内人数が一致した際の特徴量, 青線がグループ内人数の差が 1 人の際の特徴量.

1. 同一グループに属する人数
2. 同一グループに属する全ての人の色情報

1つ目のグループ属性として, その人の属するグループ内の人数という情報を用いて, グループ内人数が等しい人同士のスコアが高くなるようなグループ内人数特徴量のスコア R^{gp} を設定する. そのために, カメラ間人同定において歩行者 i と歩行者 j のグループ内人数の差 $p_{i,j}^{gp}$ を計算し, 図 17 のように表されるガウス分布 $g^{gp}(p^{gp}; m^{gp}, \delta^{gp2})$ に代入することで得られた値を, グループ内人数特徴量のスコア $R_{i,j}^{gp}$ とする. このスコア算出にガウス分布を用いたのは, 人数が一致したときにスコアが高くなり, かつグループ検出の失敗や, グループ内人数の変化にも対応できるように, グループ内人数が異なる場合にも多少のスコアを与えるためである.

$$R^{gp} = g^{gp}(p^{gp}; m^{gp}, \delta^{gp2}) \quad (3)$$

2つ目のグループ属性として、その人の属するグループ内の全員分の色情報を用いて、色ヒストグラムの類似度をグループ内色特徴量のスコア R^{gc} として算出する。グループの色情報を算出するために、その人を含めたグループの構成員全員 (N 人とする) のヒストグラム $A_1^g, A_2^g, \dots, A_N^g$, を加算し、正規化したグループ内ヒストグラム A^g を算出する。以上のように求めたグループ内ヒストグラム A^g, B^g を用いて、6.1.1章と同様に、ヒストグラム間の類似度を表す Bhattacharyya 係数 (6.1.1章参照) を求め、これをグループ内色特徴量のスコア R^{gc} とする。また、一人で行動している人に対しても、その人の色ヒストグラムをグループ内ヒストグラムとすることで、グループ内色特徴量のスコアを求める。

$$R^{gc} = L^g = \sum_{u=1}^{24} \sqrt{A_u^g B_u^g} \quad (4)$$

以上で算出したグループ内人数特徴量のスコア R^{gp} とグループ内色特徴量のスコア R^{gc} をカメラ間人同定に用いる。

6.3 グループ属性と従来の特徴量を併用したカメラ間人同定

6.1章、6.2章で求めた特徴量のスコアから、カメラ間人同定を行う。

カメラ間人同定用のスコア R は、6.1章、6.2章で求めた4種類の特徴量のスコア R^c, R^t, R^{gp}, R^{gc} 全てのスコアが高くなった場合に R が高くなるように、以下の式で表すように、4種類のスコアを掛け合わせて算出する。

$$R = R^c R^t R^{gp} R^{gc} \quad (5)$$

この R の値が高いものから、カメラ間人同定の候補として挙げることで、カメラ間人同定を実現する。

7. 実験と結果

実験に用いるデータセットは、Graz University of Technology と Austrian Institute of Technology による Person Re-ID 2011 dataset[37] に含まれる 2つのカメラ画像によってなる動画を利用した。これら 2つの動画は図 18 のような配置のカメラで撮影されたものであり、本研究において対象としている監視カメラのように上方から見下ろすようなカメラ配置となっている。また、データセット内には動画のほかに、各人の ID 付き追跡結果の正解データが与えられており、この正解データは、2つのカメラで共に撮影された人と一方のカメラでのみ撮影された人が ID のナンバリングによって判別できるようなデータとなっている。よって、本研究ではこのデータセットの動画を使用して、提案手法に関する実験を行った。この Person Re-ID 2011 dataset の詳細を以下に示す。

Camera a

フレーム数 92825 フレーム

フレームレート 25 フレーム/秒

追跡人数 385 人

Camera b

フレーム数 99997 フレーム

フレームレート 25 フレーム/秒

追跡人数 749 人

2つのカメラ間の関係

カメラ間移動時間 平均 65 秒

カメラ間で対応付けされている人数 200 人

ただし、このデータセットに保存されている追跡結果の正解データは、人の色情報が明確に分かるように、人が単独で映っているフレームにおける追跡結果だ

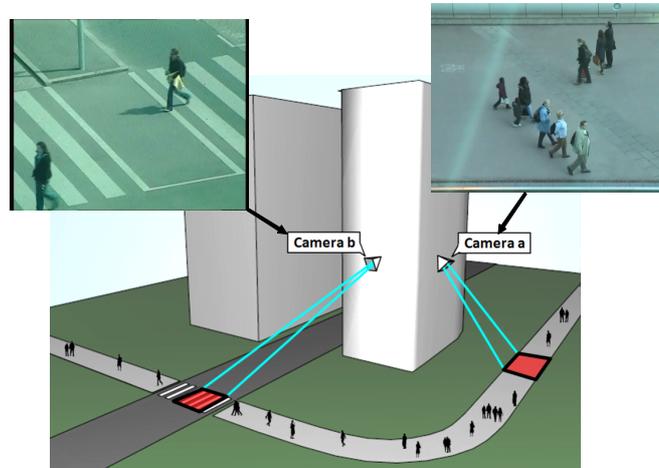


図 18 実験に用いるカメラの配置図と映像例 (Learning, Recognition and Surveillance のホームページ [38] より参照) .

けが保存されているので，他の人や物体が人の撮影領域に含まれるフレームにおいて人の追跡結果が与えられておらず (図 19 参照)，軌跡全体においてオクルージョンが生じている人については ID の割り振りが行われていない．本研究ではオクルージョンが起こっているフレームにおいても人の位置を求め，軌跡を得ることが必要なので，追跡結果の正解データが与えられていないフレームについては，手動で正解データを与えることで，正解データの補完を行った．補完後の正解データにおける，各カメラの検出人数は Camera a に 573 人，Camera b に 925 人であり，その中で双方のカメラで検出された人数は 342 人である．

各人のグループ属性 (誰と誰がグループ内ペアかという情報) の正解データは，動画中に現れる人物について，各動画中の位置関係や軌跡の類似度，また親子等の社会的関係などに着目し手動で与えた．また，本研究では 2 つの動画からグループ属性の正解データを与える際に，片方のカメラでのみグループと判別できるような軌跡ペアも正解データとして与えている．グループ内ペアの総数は，Camera a の動画に 134 対，Camera b の動画に 152 対である．

以下では，4，5，6 章で提案した手法について以下の実験を行った．

単眼カメラ視野内での人追跡 (4 章)



図 19 Person Re-ID 2011 dataset の軌跡データの消失例. 人と人が重なって撮影されているフレームの軌跡データが保存されていないことが分かる.

1. 4.1 章で提案した人検出手法の精度を確認するための、各フレームの画像を用いた人検出実験
2. 4.2 章で説明した人追跡手法の精度を確認するための、人検出結果を用いた人追跡実験

単眼カメラ内のグループ検出 (5 章)

1. 5 章で提案したグループ検出手法の有用性を確認するための、正解軌跡を用いたグループ検出実験
2. 追跡結果からの検出精度を調べるための、7.1 章の実験により得た追跡結果の軌跡を用いたグループ検出実験

グループ属性を用いたカメラ間人同定 (6 章)

1. 色特徴量の算出に必要な色ヒストグラムの有効な取得法を調べるための、6.1.1 章で述べたフレーム選択法の比較実験
2. カメラ間人同定におけるグループ属性の効果を確認するための、正解データから算出したグループ特徴量と従来特徴量を併用したカメラ間人同定実験
3. グループ検出精度がカメラ間人同定におけるグループ属性に与える影響を調べるための、7.2.1 章の実験結果から算出したグループ特徴量と

正解データによる従来特徴量を併用したカメラ間人同定実験

4. 実環境中で本手法のシステムを実行した際のカメラ間人同定精度を調べるための、7.1章の実験により得た追跡結果から算出したグループ特徴量と従来特徴量を併用したカメラ間人同定実験

本実験で用いるデータセットについて、以下にまとめる。

Camera a

フレーム数 92825 フレーム

フレームレート 25 フレーム/秒

追跡人数 573 人

グループ内ペア数 134 対

Camera b

フレーム数 99997 フレーム

フレームレート 25 フレーム/秒

追跡人数 925 人

グループ内ペア数 152 対

2つのカメラ間の関係

カメラ間移動時間 平均 65 秒

カメラ間で対応付けされている人数 342 人

7.1 単眼カメラ視野内での人追跡実験

7.1.1 HOG 特徴量を利用した人検出実験結果

4.1 章で提案した人検出手法の精度を確かめるために、2つのカメラで撮影された動画の各フレームにおける画像を用いた人検出を行った。

人検出の結果の例を図 20 に示す。図 20 の Camera a の画像中の黒服を着た人の例などから、一人に対して複数の検出結果が表示されており、誤検出が発生していることが分かる。これらは、検出評価値ノイズ w の影響によって、人付近で小さなピークが多数発生しているためだと考えられる。また、Camera a の画像中心の親子の例など、人と人との間に誤検出が発生している例もある。これは、平滑化によって2人のピーク値が周辺の評価値を上昇させ、ピークを出現させやすくしているためだと考えられる。

人検出が未検出となる例として、オクルージョンによって複数人に対して一つの検出結果しか表示されない例のみ存在する（図 21 参照）。これは、平滑化によって複数人のピークが統合され一つにまとめられたためだと考えられる。

検出評価値ノイズの低減のためには、検出時に行う平滑化をより強くすることで、検出評価値ノイズによる細かなピークを除去できると考えられる。しかし、平滑化を強くすることで接近した複数人を検出する際に、複数人それぞれに表れるピークが消えてしまい、図 21 のような未検出がより増えると考えられる。以上より、平滑化を強くしても複数人それぞれに対して強いピークが検出されるような人検出手法を行う必要がある。

Camera a



Camera b

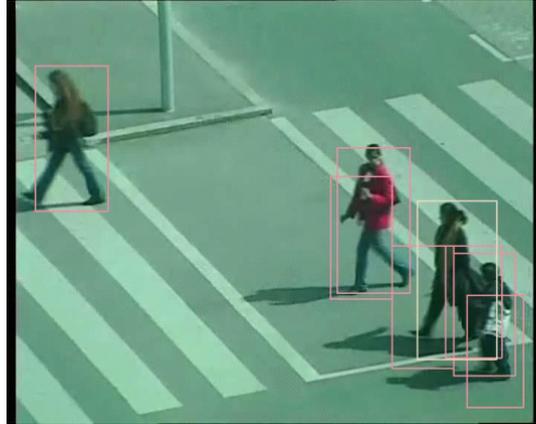


図 20 各カメラの画像における人検出の例. 各矩形は検出結果を示す.

False negative by occlusion

Camera a



Camera b



図 21 オクルージョンによる人の未検出の例.

7.1.2 検出結果の大域的最適連結による人追跡実験結果

4.2章で説明した人追跡手法の精度を確かめるために、7.1.1章の2つのカメラで撮影された各動画における検出結果を用いて、人追跡を行った。

各カメラにおける人追跡の結果の成功例を図22に示す。図22より、1秒後のフレームにおいても各人に同じIDが割り振られていることから、人追跡が成功していることが分かる。さらに、Camera aのID8とID78や、Camera bのID24とID87のように、接近して移動する複数の軌跡について、人それぞれを追跡成功していることが分かる。これは、学習時のネガティブサンプルとして、人と人との間を囲んだ矩形領域内の画像（図7中のネガティブサンプルの右側）を与えているため、人と人との間の領域においてSVMの評価値が下がり、各人のピークをそれぞれ得ることができたためだと考えられる。また、図20のように人検出時に見られた検出評価値ノイズによる誤検出が得られていたが、こういった誤検出結果は近接フレームで連続的に検出されにくいため、追跡時に連結候補が無くなり、その結果検出評価値ノイズをある程度低減できていると考えられる。

7.1.1章で行った人検出実験での失敗に起因する追跡の失敗例を図23に示す。誤追跡の例（人の映っていない部分を人として追跡している例）として、検出評価値ノイズによる人の誤検出が多数残り、それらが連結されたことによる誤追跡の例が存在する（図23左側参照）。また、未追跡の例（人の映っている部分を追跡できていない例）としては、複数人が接近した場合の人の未検出によって、連結すべき検出枠が存在しなくなり、未追跡となる例が存在する（図23右側参照）。これらの誤追跡や未追跡は、7.1.1章で述べた手法を用いた人検出の精度向上により改善できると考えられる。

また、検出結果の連結時の失敗例として、軌跡が複数の軌跡に分割され別IDが割り振られてしまう例（図24参照）が存在する。これは、近接フレームでの人検出結果が離れすぎているために連結コストが高くなり、連結失敗していることが原因だと考えられる。他の失敗例として、別の人との連結が発生しIDが入れ替わってしまう例（図25参照）が存在する。複数軌跡の接近時に、オクルージョンにより複数の人に対して1つの検出結果が得られた場合に、人自身の連結コストより別の人との連結コストが低くなることで、他者との間で連結が起こっ

てしまったためだと考えられる。これらの失敗例の解決策として、本実験では既存の大域的最適連結手法 [6] をそのまま利用したが、連結コストの計算方法を変化させる（例えば、移動方向が同じ人の連結コストを下げることで、同一人物間の連結を起りやすくする等）ことで改善できると考えられる。

Camera a



Camera b

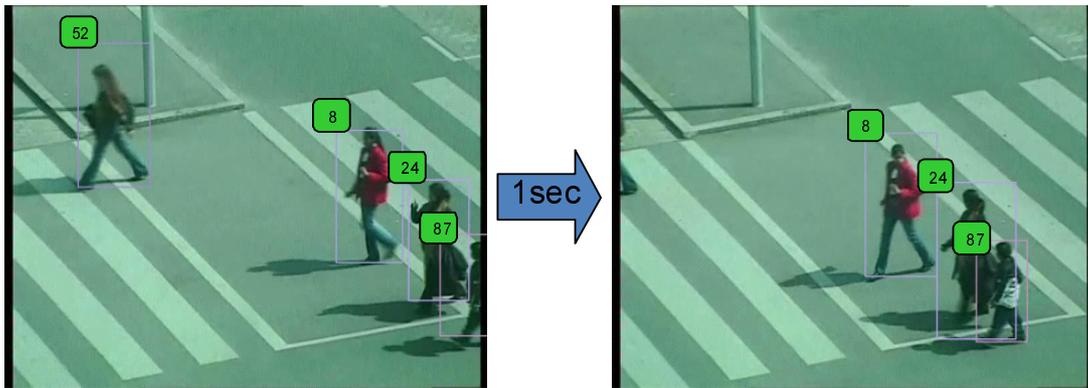
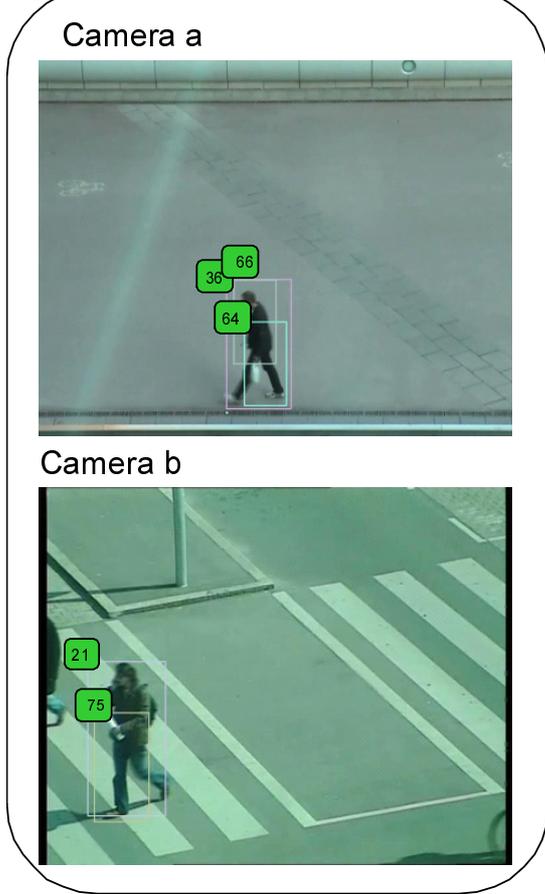


図 22 人追跡の成功例. 矩形は追跡結果を, 矩形左上の数字は各軌跡に割り振られた ID を示す. 図 20 の人検出結果から検出評価値ノイズが低減できており, 人追跡が成功していることが分かる.

False positive by noise of feature



False negative by occlusion

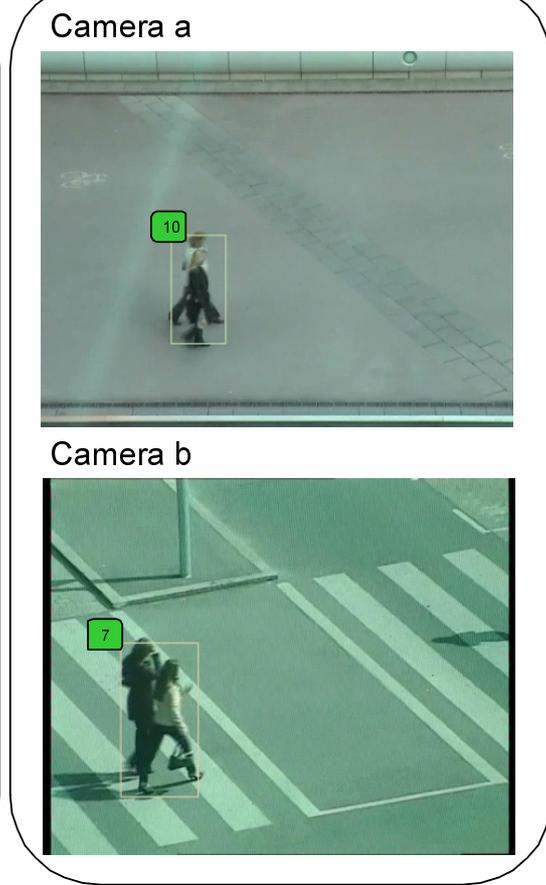


図 23 人検出の失敗による人追跡の失敗例。左側が検出評価値ノイズによる誤追跡，右側がオクルージョンによる未追跡を示す。

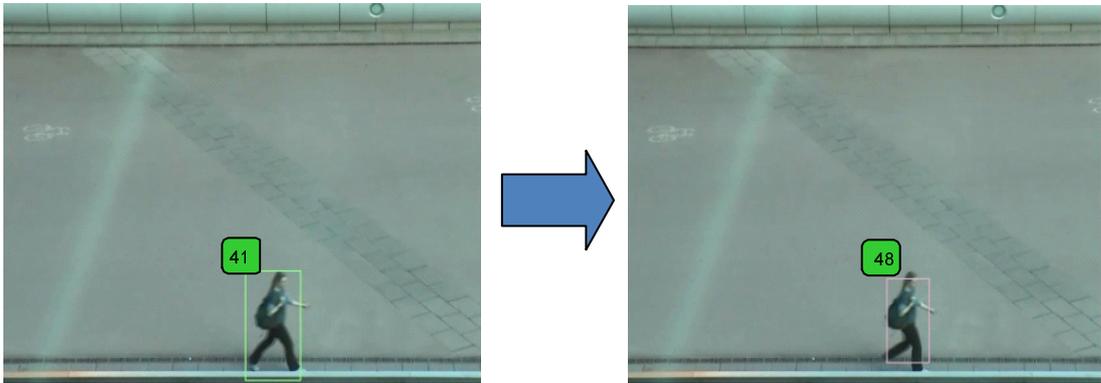


図 24 追跡結果の 1 軌跡が複数に分割される失敗例. ID41 の軌跡が, 次のフレームで ID48 に変化している.

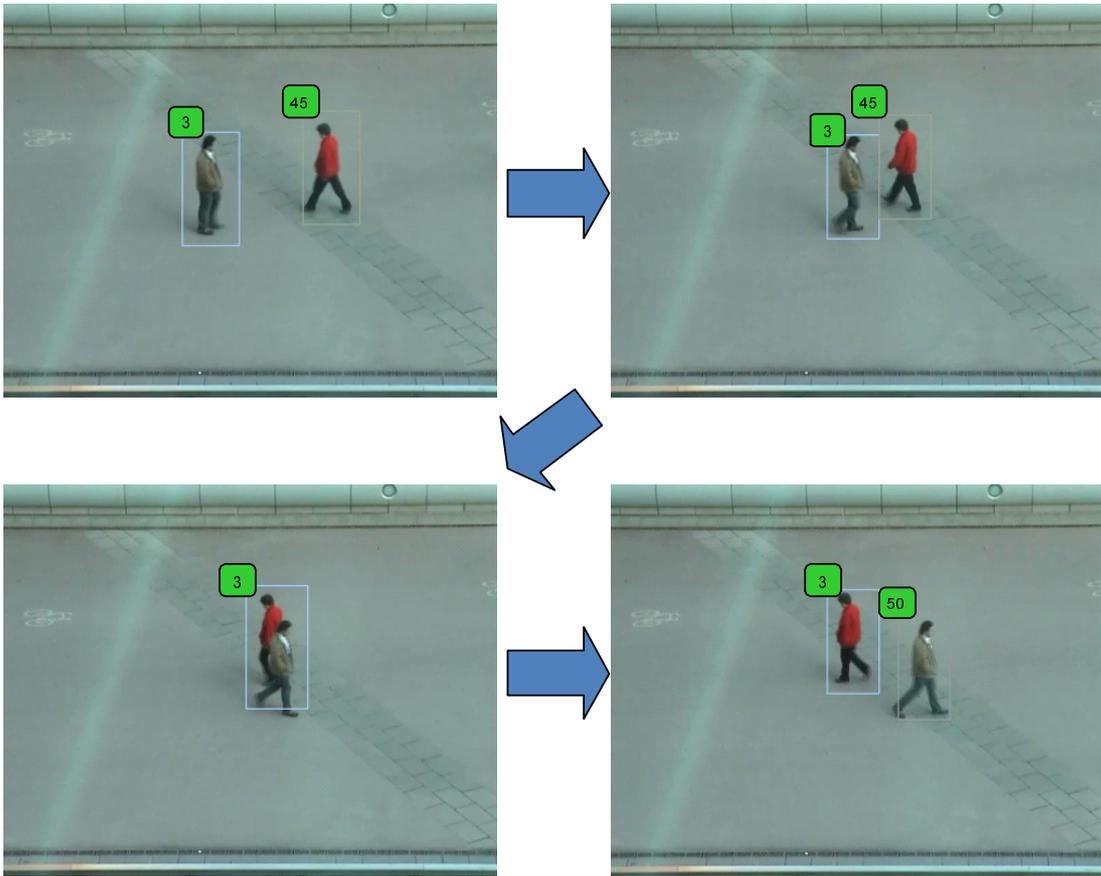


図 25 複数軌跡の接近時に ID が入れ替わる失敗例. ID3 の軌跡が, グレーの服の人から赤色の服の人へ入れ替わっていることが分かる.

7.2 単眼カメラ内のグループ検出実験

本研究で提案しているグループ検出手法の精度を確かめるため、グループ検出の実験を2種類の軌跡データを用いて行う。まず、正解データとしてデータセット内の軌跡データを用いる事で理想環境中でのグループ検出手法の有用性を確かめ(7.2.1章)、次に、7.1章の実験により得た追跡結果の軌跡を用いる事で、人追跡の実験結果に対するグループ検出精度を確かめる(7.2.2章)。

これらの軌跡データは画像中の人の撮影されている位置を示したものであり、実世界上の人の位置を示したものではない。本研究で提案しているグループ検出用の時空間特徴量は、実世界上でのグループ内ペアの軌跡の特徴に着目して定義されたものであるため、画像中での位置を元に時空間特徴量を算出すると、グループ検出の精度が低減してしまう恐れがある。したがって、人の画像平面上の位置 P^c から実世界上のカメラ視野空間中の位置 P^w を求めるために、各カメラに対してホモグラフィ行列 H を求め、以下の式にしたがってホモグラフィ変換を行うことで、画像平面上から実世界上への射影を行う。

$$P^w = HP^c \quad (6)$$

こうして得られた各人の実世界上の位置 P^w を用いて、時空間特徴量を算出する。

グループ内ペア判定用のSVMの学習を行うために、本研究では各カメラで撮影された動画を時系列順に9シーケンスに分割し交差確認法を行った。これは、分割した動画の内の1シーケンスで撮影された軌跡ペアをテスト用に、残りの8シーケンスで撮影された軌跡ペアを学習用に用いるという処理を、9シーケンス全てをテストするまで繰り返しグループ検出精度を確かめる手法である。

また、検出精度を安定させるために、以下のように、軌跡ペアに限定しグループ検出を行った。

1. 各フレームにおけるグループ内ペア判定の結果を一定回数以上算出するために、同時にカメラ視野内に撮影されたフレームが100フレーム以上(グループ内ペア判定回数が100回以上)の軌跡ペアに限定した。

2. 本実験で用いたデータセット内で撮影されているのは大半が人であるため、人以外（例：自転車）の軌跡ペアはサンプル数が少なく十分なSVMによる学習が行えていないと考え、グループ検出を行う軌跡ペアを歩行者同士の軌跡ペアに限定した。

また、5.3章で述べたグループ内ペア検出時に用いる閾値は、予備実験の結果から、軌跡ペアが同時に観測された全フレーム中 $2/3$ 以上のフレームにおいてグループ内ペアと判定された場合にグループ内ペアとして検出した時に、最もグループ内ペアの検出精度が高かったため、本研究ではこの値を閾値とする。

表 1 各カメラにおける正解軌跡を用いたグループ内ペア検出結果. それぞれ, 検出成功 (True-Positive), 誤検出 (False-Positive), 未検出 (False-Negative) のグループ内ペア数

	True-Positive	False-Positive	False-Negative
Camera a	108	14	26
Camera b	130	38	22

7.2.1 正解軌跡を用いたグループ検出実験結果

まず, 提案しているグループ検出手法の有用性を確かめるために, 理想的な環境 (人追跡が完全に成功した場合) でのグループ検出実験を行う. そこで, 本実験では, 精度が保障されているデータセット内の軌跡データを用いる.

各カメラにおけるグループ内ペア検出結果を表 1 に示す. この結果から, グループ内ペア全体の 80~86% が検出成功しているということが分かる. グループ内ペア検出の成功例として, グループ内ペアが検出成功している例と非グループ内ペアがグループ内ペアとして検出されていない例を図 26 に示す. 図 26 に代表される結果から, グループ内ペア判定に F_1 , つまり 2 軌跡の位置関係が強く影響を与えていることが分かる.

グループ内ペアの誤検出例を図 27 に示す. カメラ視野内においてのみ近距離で移動した軌跡ペアは, F_1 の値が小さくなるため, 誤検出されていることが分かった. 特に Camera b は, 横断歩道という時間によって人通りの多くなる場所を撮影しているため, 局所的な軌跡ペアの接近が多くなるので, その結果 F_1 の影響によってグループ内ペアの誤検出が増加していると考えられる.

グループ内ペアの未検出例を図 28 に示す. カメラ視野内において離れて移動していた軌跡ペアは, F_1 の値が大きくなり, 未検出となることが分かった. また, 正解データ作成時には, 図 29 のような一方のカメラ視野ではグループ内ペアと判定できないようなペアもグループ内ペアとしているため, そういった軌跡ペアはグループ内ペア検出時に未検出とされる. そのような軌跡ペアは社会的関係の推定を行うことで検出成功となることが期待される.

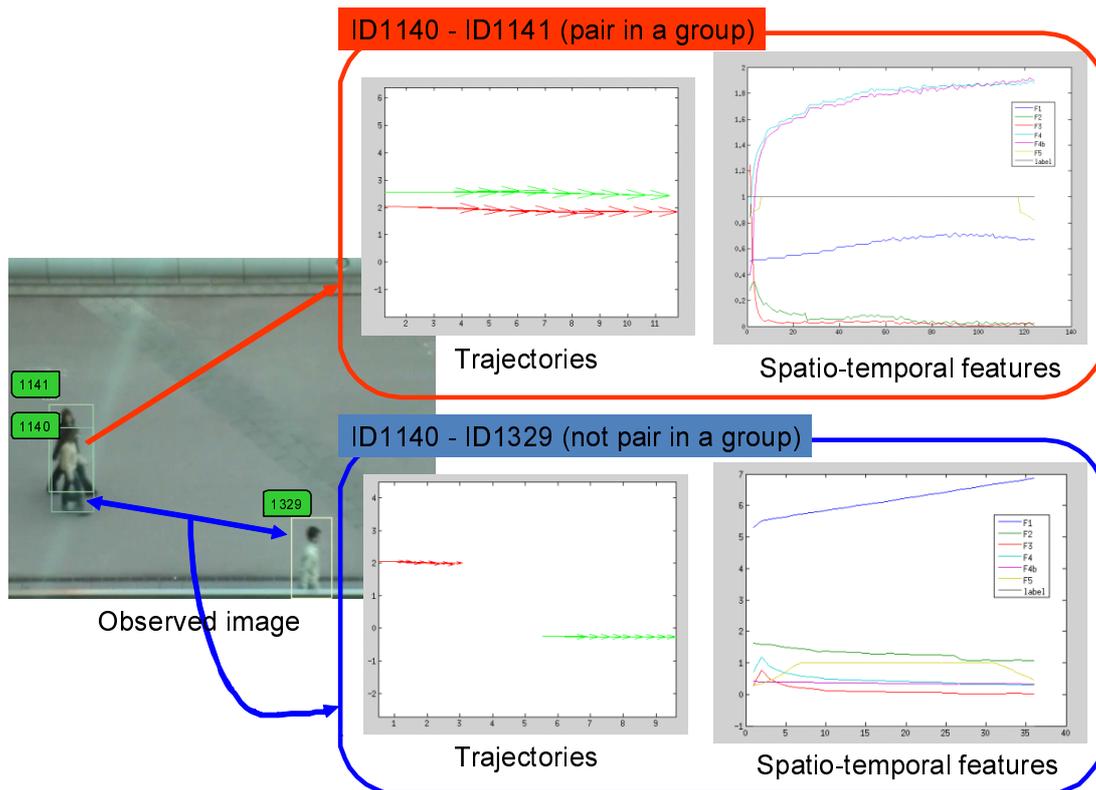


図 26 グループ内ペアの検出成功例 (ID1140 と ID1141) と非グループ内ペア (ID1140 と ID1329) の例. 左のグラフが同時に軌跡ペアが観測されたフレームにおける人の位置 (矢印の始点) と速度ベクトル (矢印) を示す. 右のグラフがその時刻における時空間特徴量とグループ内ペア判定の結果を示す.

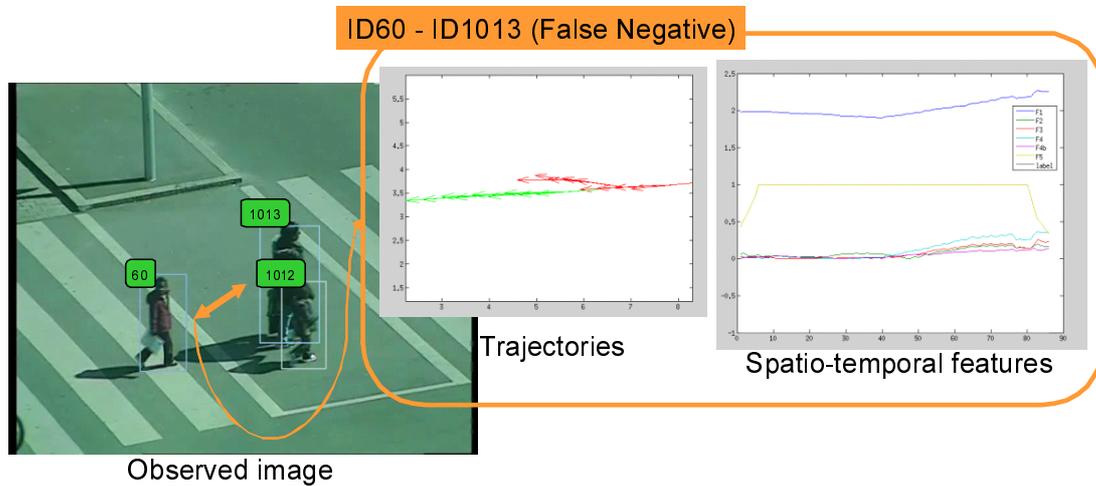


図 27 グループ内ペアの誤検出例 (ID60 と ID1013) . 局所的にこの二人が接近しているため、グループ内ペアとして誤検出されている.

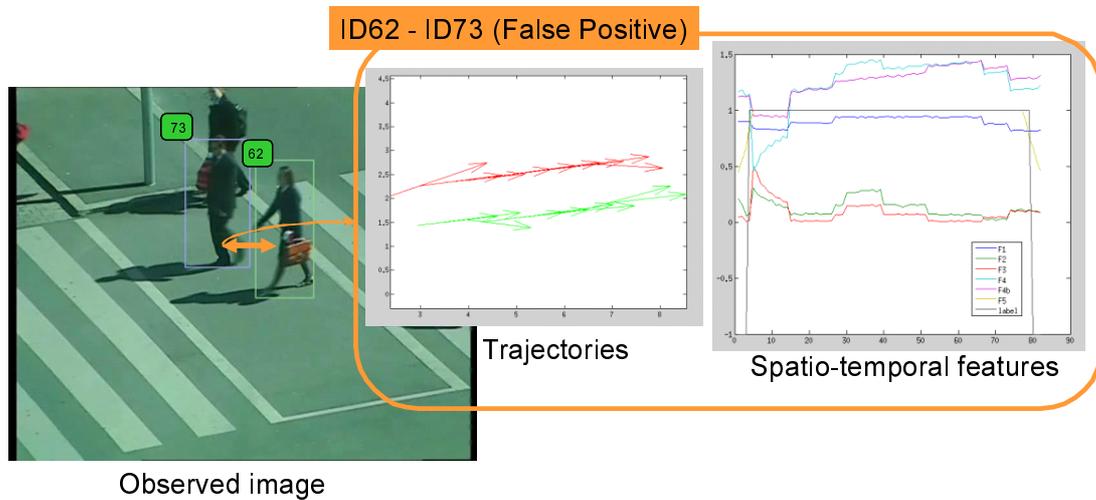


図 28 グループ内ペアの未検出例 (ID62 と ID73) . 局所的に二人が離れているために、グループ内ペアとして検出されていない.

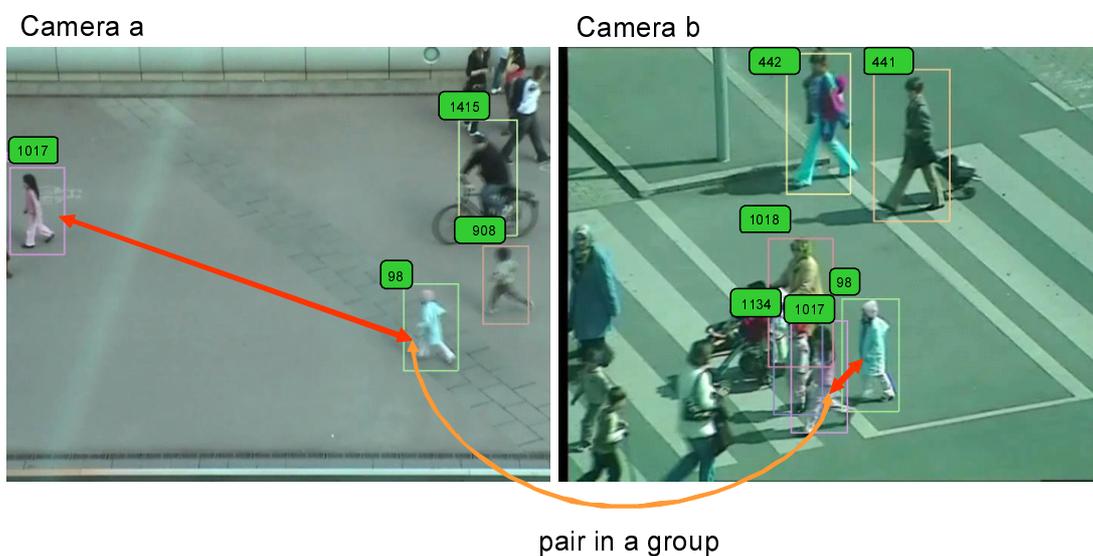


図 29 一方のカメラ視野内のみでグループ内ペアと判定できる例. Camera a の画像のみで ID98 と ID1017 がグループ内ペアと判定するのは困難.

7.2.2 人追跡実験の結果の軌跡を用いたグループ検出実験結果

次に、7.1章で求めた人追跡の結果を用いた時の、グループ内ペア検出手法の精度を確かめる。事前準備として、各追跡結果と正解軌跡を比較し、各追跡結果に対して正解軌跡のIDを振り当てる。そして、追跡結果に対してグループ内ペア検出を行い、ペア検出結果に対して振り当てられた正解軌跡のIDを確かめ、グループ内ペアの正解データ内に存在する場合に検出成功とする。また、図24のように一人の歩行者に対して複数の追跡結果が得られた場合、その複数の追跡結果全てに対してグループ内ペア検出が成功した時に、そのIDのグループ内ペア検出が成功したとする（例：図30の歩行者*i*と歩行者*j*がグループ内ペアの時、軌跡1と軌跡2、軌跡1と軌跡3がグループ内ペアとして検出された場合に検出成功とする）。

グループ内ペア検出結果を表2に示す。表2中の括弧内の数値は、グループ内ペア検出結果の失敗数の中で、7.1章の実験結果で見られた誤追跡や未追跡が原因となり、グループ内ペア検出が誤検出や未検出となっている数を示している。このグループ内ペア検出の実験結果から、グループ内ペア検出失敗の原因の大半が人追跡の失敗であることが分かる。特にグループ内ペアの未検出例が多数発見されているのは、グループ内ペアは近距離で移動することが多く、図23のようにオクルージョンの影響により未追跡となり、その結果グループ内ペア検出が行えていないためだと考えられる。また、グループ内ペアの誤検出は、図31のように追跡結果とその付近に生じる誤追跡結果が、 F_1 の値が小さい軌跡ペアとして認識され、その結果グループ内ペアとして誤検出されているためだと考えられる。

以上のように、単眼カメラ視野内での追跡精度を向上させることで、グループ内ペア検出精度は向上できると考えられる。

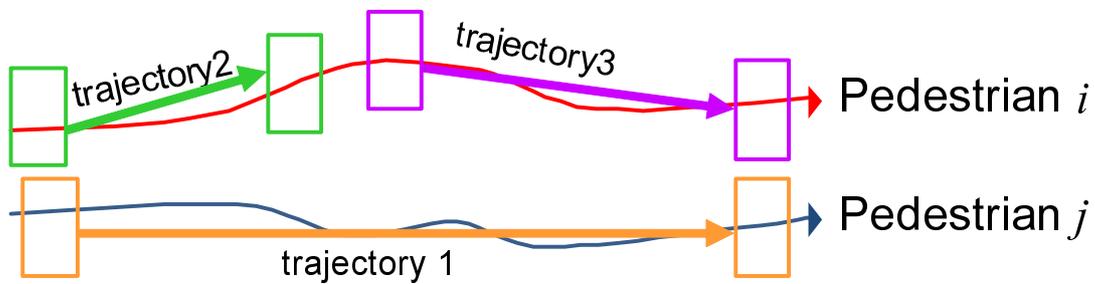


図 30 図 24 のように軌跡が複数の追跡結果に分割された場合の、グループ内ペア検出の概念図. 歩行者 i と歩行者 j がグループ内ペアであり, 歩行者 i に対する人追跡結果が軌跡 2 と軌跡 3, 歩行者 j に対する人追跡結果が軌跡 1 となっている場合, 軌跡 1 と軌跡 2, 軌跡 1 と軌跡 3 の両方がグループ内ペアとして検出された時, 歩行者 i と歩行者 j についてグループ内ペア検出が成功したとする.

表 2 7.1 章の人追跡の実験結果を用いたグループ内ペア検出結果. 括弧内の数値はグループ内ペア検出結果の失敗数の中で, 7.1 章の実験結果で見られた誤追跡や未追跡が原因となり, グループ内ペア検出が誤検出や未検出となっている数を示している.

	True-Positive	False-Positive	False-Negative
Camera a	6	30 (11)	128 (117)
Camera b	3	11 (4)	149 (138)

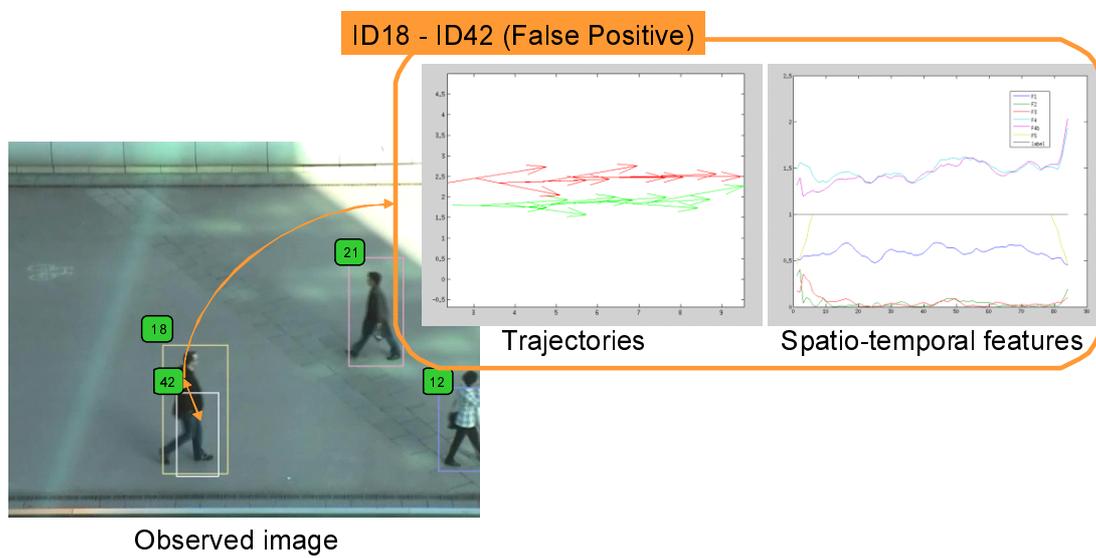


図 31 追跡結果の誤検出が原因の、グループ内ペアの誤検出例 (ID18 と ID42)。

7.3 グループ属性を用いたカメラ間人同定実験

本研究で提案しているカメラ間人同定におけるグループ属性の有用性を確かめる実験を行った。Camera a の視野内で検出されたある人 i に対して、Camera b で検出された全人 884 人との間でカメラ間人同定用のスコア $(R_{(i,1)}, R_{(i,2)}, \dots, R_{(i,884)})$ を式 (5) により求め、スコア R が高いものから人同定の候補として挙げる。この処理を、双方のカメラ視野内で検出成功している 342 人に対して行い、全候補 884 人中での本人同士の間スコア $R_{(i,i)}$ の順位を調べることで、人同定の精度を計算する。カメラ間人同定における精度を表すために、一般的に広く用いられている表現法である CMC 曲線 (Cumulative Match Characteristic curve) を用いる。また、本実験のシーンにおけるグループ内人数特徴量のスコアの算出に用いるガウス分布の平均 m^{gp} と分散 δ^{gp^2} は、予備実験の結果から $m^{gp} = 0$ 、 $\delta^{gp^2} = 0.3$ とした。

以上の人同定の精度の算出を、6 章で提案したグループ特徴量を併用した場合と併用しなかった場合、それぞれについて求め、本研究で提案するカメラ間人同定におけるグループ属性の有用性を調べる。

7.3.1 色ヒストグラム計算時の最良フレーム選択法

まず、6.1.1 で述べた色特徴量の取得時のフレーム選択法の中で、色特徴量のみでカメラ間人同定を行う際に最も精度の良くなるフレーム選択法を調べる。

6.1.1 で述べた 3 手法を用いて人の色特徴量のスコア R^c を計算し、 R^c のみを用いてカメラ間人同定を行った結果を図 32 に示す。図 32 から、各手法において色特徴量のみを用いた場合の精度に大きな差は見られなかった。このスコアは実験を行う環境によって変化すると考えられる。本実験の環境においては、候補の上位において手法 1 のフレーム選択法を利用した際の結果が最も良いことが分かる。

また、手法 3 において $N = 5, 10, 20$ と変化させたが、 $N = 20$ において手法 1 とほとんど変わらない結果が得られたことから、人の平均的な色情報は 20 フレーム程度の色情報を用いる事で得ることができると考えられる。

したがって、以降の実験における色特徴量のスコア R^c は、本研究の実験環境

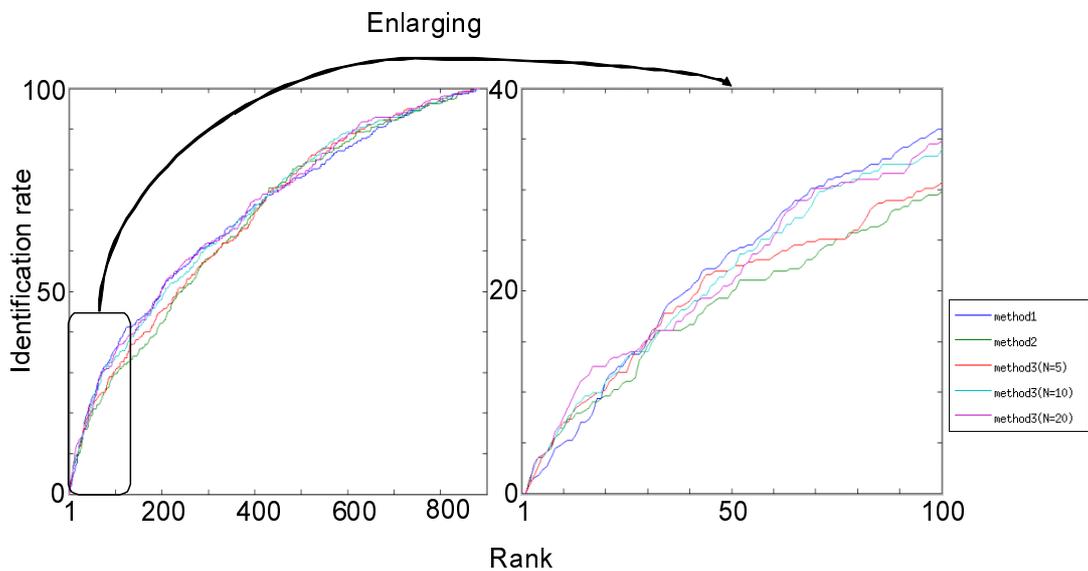


図 32 6.1.1 の 3 つの手法から算出した色特徴量のスコア R^c を用いたカメラ間人同定の結果図（手法 3 におけるフレーム数 $N = 5, 10, 20$ ）．横軸は本人同士の間で得られたスコアの全候補 884 人中の順位を示し，縦軸はその順位以内に入っている 342 人中の人数の割合を示す．全順位における結果を左側のグラフに，100 位以内の部分に着目した結果を右側のグラフに示す．

において，候補の上位での精度が最良となった手法 1 のフレーム選択法を用いて算出することとする．

7.3.2 カメラ間人同定におけるグループ属性の効果

まず、カメラ間人同定におけるグループ属性の効果を調べるために、理想的な環境での実験を行う。そこで、人同定用のスコア R を求める際に、人のグループ属性や軌跡データは精度が保障されている正解データ（7.1章、7.2章の実験で用いたもの）を使用して R を算出する。

まず、色特徴量と2種類のグループ特徴量を併用した際のカメラ間人同定の結果を図33に示す。図33から、グループ特徴量の中で、特にグループ内人数特徴量によって、カメラ間人同定の精度が向上していることが分かる。一方、グループ内色特徴量も、 R^c と R^{gp} を使用した時の精度と R^c 、 R^{gp} と R^{gc} を併用時の精度を比較すると、多少カメラ間人同定の精度向上に貢献していることが分かる。しかし、グループ内人数特徴量による効果と比較するとグループ内色特徴量の効果は小さく、グループ内人数特徴量を併用しなかった場合は有用性が得られなかったため、グループ内色特徴量のスコア R^{gc} の算出法については、新たな算出法を考える必要がある。

次に、2種類の従来特徴量（色特徴量・時間特徴量）とグループ特徴量を併用した際のカメラ間人同定の結果を図34に示す。図34と図33の左側のグラフ同士を比較すると、時間特徴量によるカメラ間人同定の精度向上が顕著であることが分かる。また、30位以内の部分に着目した右側のグラフを見ると、2種類のグループ特徴量それぞれによってカメラ間人の精度が向上していることが分かり、2種類のグループ特徴量をどちらも併用した際に最も精度向上が見られた。

また、グループ特徴量を併用した時と併用しなかった時について、本人同士の間で得られたスコアが10位以内に入った人数を比較することで、グループ特徴量の効果を定量的に評価する。結果として、2種類の従来特徴量のみを使ったときは135人、従来特徴量と2種類のグループ特徴量を併用した時は195人となり、グループ特徴量によってカメラ間人同定における精度に44%の精度向上が見られた。

以上より、カメラ間人同定において、グループ特徴量が精度向上に貢献していることが分かり、カメラ間人同定におけるグループ特徴量の有用性が示された。

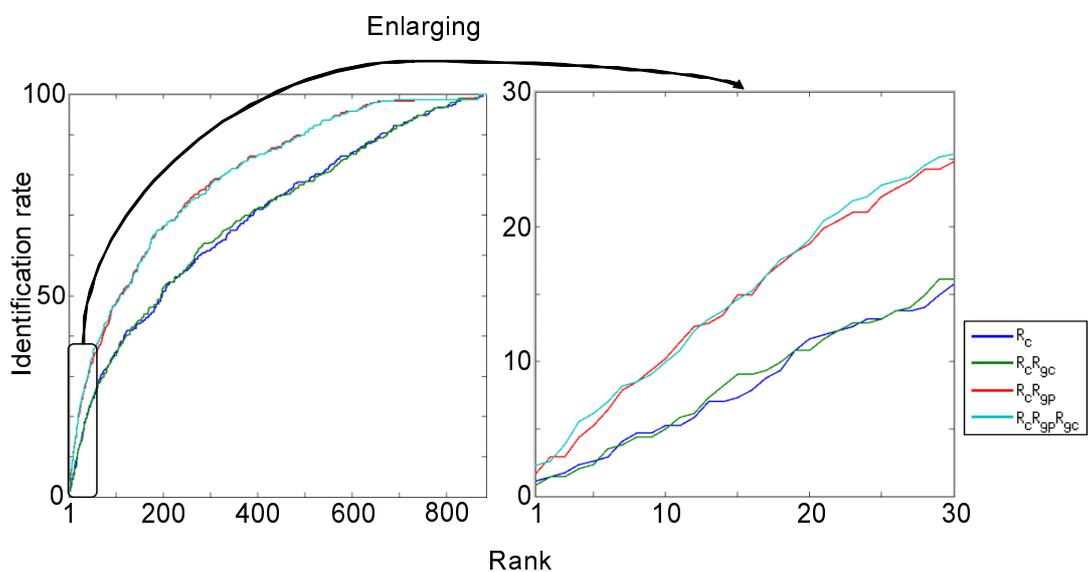


図 33 人追跡やグループ検出について正解データを用いた理想環境中において、色特徴量とグループ特徴量を併用した際の、カメラ間人同定の結果図。全順位における結果を左側のグラフに、30 位以内の部分に着目した結果を右側のグラフに示す。

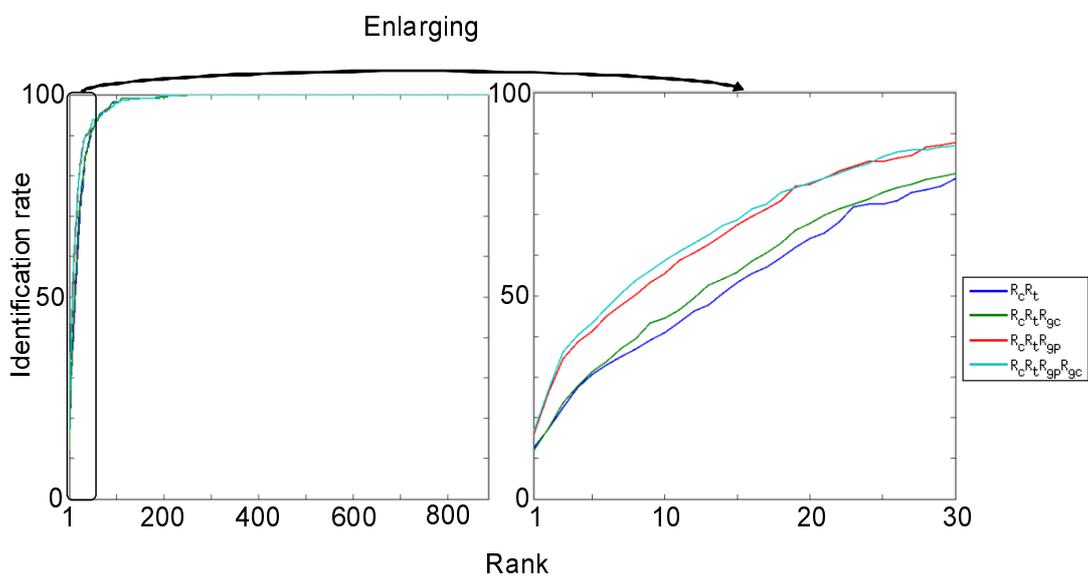


図 34 人追跡やグループ検出について正解データを用いた理想環境中において、2 種類の従来特徴量（色特徴量・時間特徴量）とグループ特徴量を併用した際の、カメラ間人同定の結果図。全順位における結果を左側のグラフに、30 位以内の部分に着目した結果を右側のグラフに示す。

7.3.3 グループ検出結果を用いたカメラ間人同定実験結果

グループ検出精度が、カメラ間人同定におけるグループ属性に与える影響を調べるために実験を行った。そこで、グループ検出の結果として7.2.1章で行った正解軌跡に基づくグループ検出結果を用いる事で、各人のグループ属性を求める。そして、導出されたグループ特徴量と従来特徴量を併用し、カメラ間人同定の精度を確かめ、7.3.2章の実験結果と比較する。

まず、色特徴量とグループ検出の結果から得たグループ特徴量を併用した際の、カメラ間人同定の結果を図35に示す。図35から、グループ検出の結果から得たグループ特徴量を用いてもカメラ間対象同定の精度が向上していることが分かる。図35と図33の比較から、グループ内人数特徴量の効果の低減が見られた。これは、グループ検出の失敗によるものであり、カメラ間人同定精度への影響が大きいグループ内人数特徴量が特に影響を与えていると考えられる。

次に、2種類の従来特徴量（色特徴量・時間特徴量）とグループ検出の結果から得たグループ特徴量を併用した際の、カメラ間人同定の結果を図36に示す。図34の右側のグラフの比較から、時間特徴量を用いない場合と同様に、グループ検出結果から得られたグループ特徴量を用いることによって、カメラ間人同定の精度向上が見られた。また右側のグラフ同士の比較から、これも時間特徴量を用いない場合と同様に、グループ検出の失敗によるグループ特徴量の効果の低減が見られた。

また、7.3.2と同様に、グループ検出から得られたグループ特徴量の効果を定量的に評価する。結果として、従来特徴量のみを使ったときは135人、従来特徴量と2種類のグループ特徴量を併用した時は185人となり、グループ特徴量によってカメラ間人同定における精度に37%の精度向上が見られた。

以上より、カメラ間人同定において、グループ検出の正解データを用いたときと比べ、グループ属性の導出に本研究で提案したグループ検出手法を用いた際には、精度に多少の低減はあるものの、全体的に同定精度が向上していることが分かった。このことより、単眼カメラ視野内における人追跡が成功すると、グループ検出結果から得たグループ属性を用いる事で、カメラ間人同定の精度が向上することが分かった。

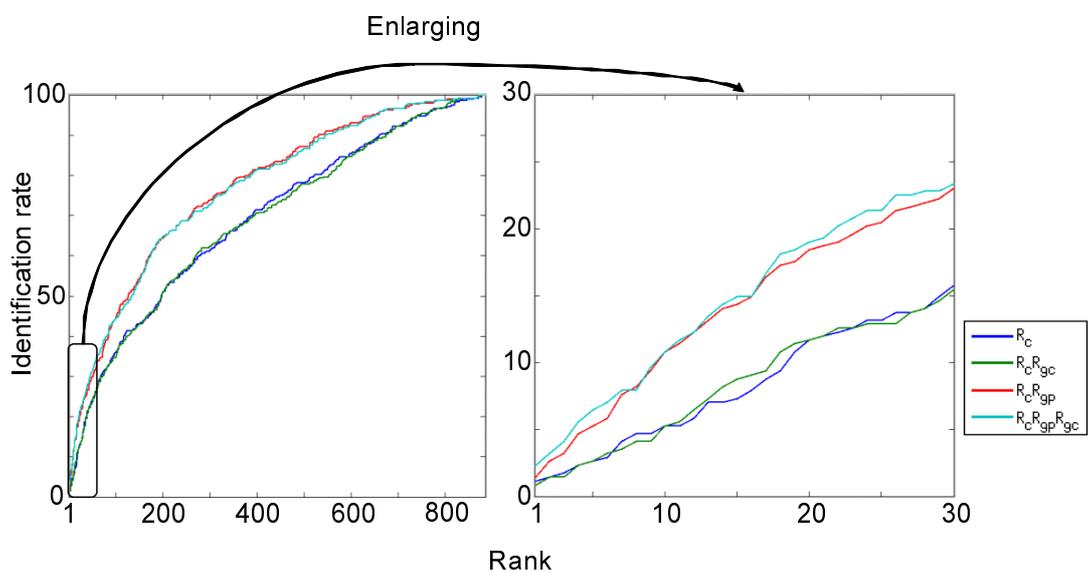


図 35 色特徴量とグループ検出結果によるグループ特徴量を併用した際の、カメラ間人同定の結果図. 全順位における結果を左側のグラフに、30 位以内の部分に着目した結果を右側のグラフに示す.

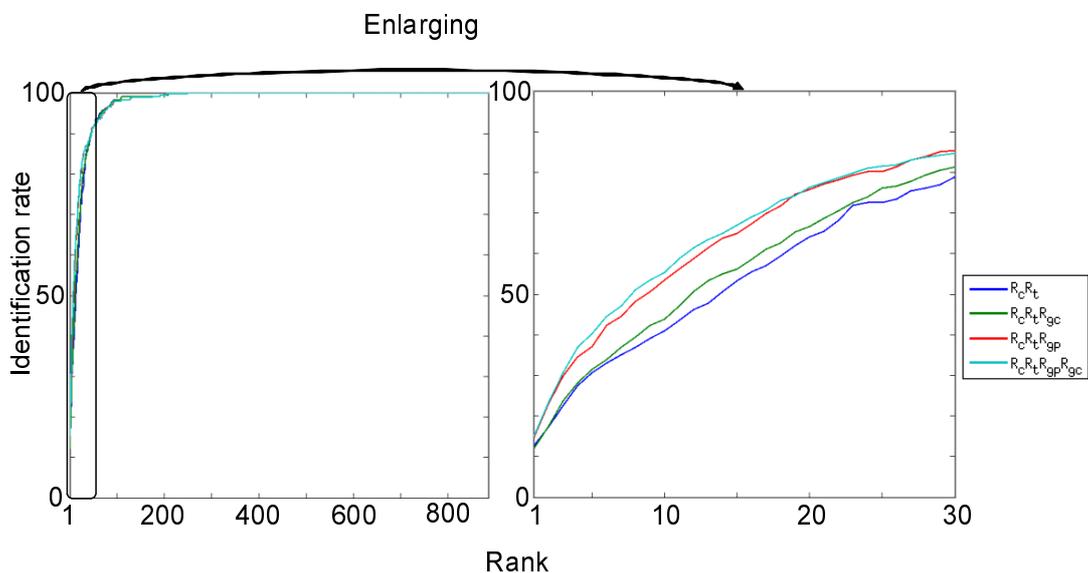


図 36 2 種類の従来特徴量（色特徴量・時間特徴量）とグループ検出結果によるグループ特徴量を併用した際の、カメラ間人同定の結果図. 全順位における結果を左側のグラフに、30 位以内の部分に着目した結果を右側のグラフに示す.

7.3.4 追跡結果を用いたカメラ間人同定実験結果

実環境中で本手法のシステムを実行した際のカメラ間人同定精度を調べるために、実験を行った。双方のカメラで検出されている342人中、7.1章の実験で双方のカメラ視野内で追跡成功している272人について、各カメラ視野内の追跡結果に基づくカメラ間人同定を行う。つまり、7.1.2章の実験による追跡結果と、7.2.2章の実験による追跡結果から得たグループ属性を用いて、カメラ間人同定用のスコア R を求め、カメラ間人同定を行う。また、図30のように一つの正解軌跡に複数の追跡結果が含まれる場合、実用面を考えると、異なるカメラで撮影された人の中で同定が行えれば良いので、複数の追跡結果それぞれについてカメラ間人同定を行い、その中で最高の順位を求めることで、その人に関するカメラ間人同定の結果とする（最高順位の追跡結果同士によって、その人に関するカメラ間人同定ができたと判定する）。

まず、追跡結果から得たグループ特徴量と色特徴量を併用した際の、カメラ間人同定の結果を図37に示す。図37から、追跡結果から得たグループ特徴量を用いても、グループ内人数特徴量によってカメラ間人同定の精度が向上していることが分かる。精度向上の原因は、図30のような複数の追跡結果が含まれた軌跡の順位が向上していたためである。これは、図30のような各追跡結果に対してグループ検出を行う際、同時観測時間が100フレーム以下の場合にグループ検出を行わないこと等が原因で、各追跡結果のグループ検出結果が異なるという事態が生じ、複数の種類のグループ特徴量を持った追跡結果それぞれに対してカメラ間人同定を行うため、偶然グループ内人数が一致した人のグループ内人数特徴量が高くなり、その結果がその軌跡における順位を引き上げたためだと考えられる。

また、図37と図35の比較から、追跡結果を用いることによるカメラ間人同定の精度の低減が見られた。この原因として、7.2.2章の実験での追跡失敗によるグループ検出の失敗による影響と、追跡時に得られる矩形が対象を中心としたものになっておらず、色情報に背景情報が大量に含まれてしまった影響の2点が考えられる。

次に、グループ検出の結果から得たグループ特徴量と2種類の従来特徴量を併用した際の、カメラ間人同定の結果を図38に示す。図38の右側のグラフの比較

から、時間特徴量を用いない場合と同様に、追跡結果から得られたグループ特徴量を用いることによって、カメラ間人同定の精度向上が見られた。また図 36 との比較から、これも時間特徴量を用いない場合と同様に、人追跡の失敗によるグループ検出の失敗と色情報への背景色の混入から、精度低減が見られた。

また、これまでの実験と同様に、追跡結果から得られたグループ特徴量の効果を定量的に評価する。結果として、従来特徴量のみを使ったときは 43 人、従来特徴量と 2 種類のグループ特徴量を併用した時は 57 人となり、グループ特徴量によってカメラ間人同定における精度に 33% の精度向上が見られた。しかし、この精度向上は前述の通り、誤追跡によるものであり、また従来特徴量と 2 種類のグループ特徴量の併用時でも、本人同士のスコアが 10 位以内に含まれるのは全 272 人中 16% である。ここで、7.3.3 章の正解軌跡を用いた場合の実験結果は全 342 人中 54% であることから、追跡結果を用いる事による精度の低下が著しいことが分かる。本システムを実環境中で利用するには、更なる単眼カメラ視野内での人追跡精度の向上が必要である。

以上より、カメラ間人同定において、実環境中で本システムを使用するためには、単眼カメラ視野内における人追跡精度を向上させ、カメラ間人同定の精度を向上させる必要がある。

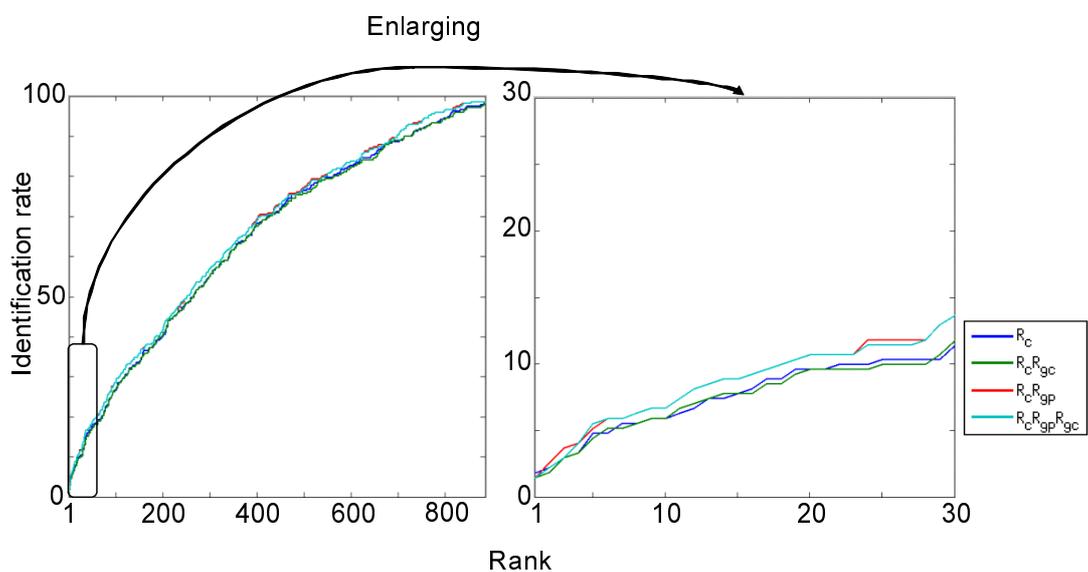


図 37 追跡結果を用いたグループ特徴量と色特徴量を併用した際の、カメラ間人同定の結果図。全順位における結果を左側のグラフに、30 位以内の部分に着目した結果を右側のグラフに示す。

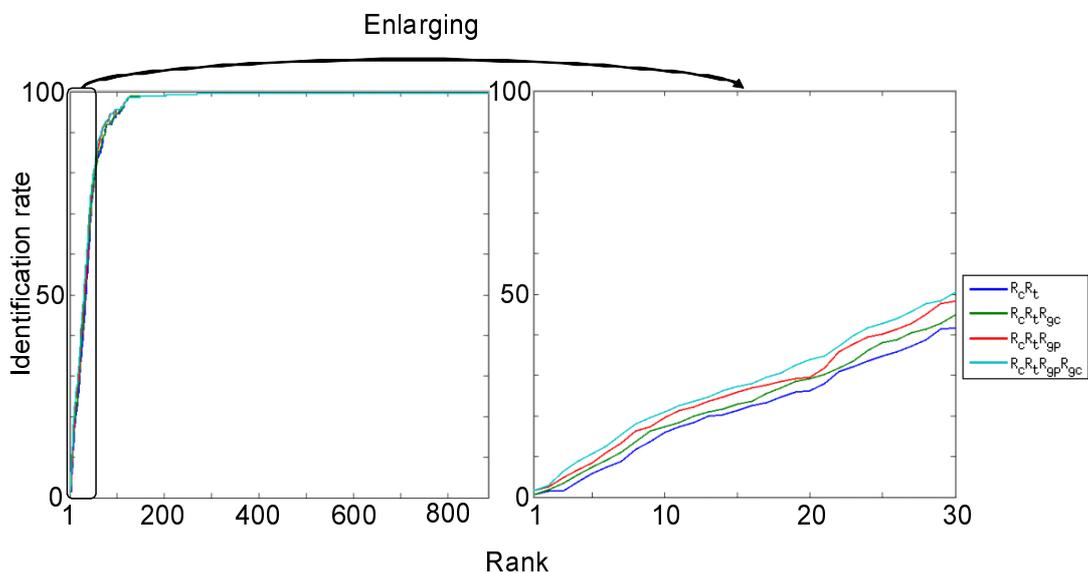


図 38 追跡結果を用いたグループ特徴量と人の色特徴量・時間特徴量を併用した際の、カメラ間人同定の結果図。全順位における結果を左側のグラフに、30 位以内の部分に着目した結果を右側のグラフに示す。

8. おわりに

本研究では、カメラ間人同定における新たな情報としてグループ属性を定義し、グループ属性から得たグループ特徴量と従来特徴量を併用することで、カメラ間非と同定の精度を向上する手法を提案した。グループ属性を得るために、単眼カメラの視野内における軌跡ペアの各フレームにおける時空間特徴量からグループ内ペア検出の実験を行い、正解軌跡を用いた実験において80~86%の精度でグループ内ペア検出を行えた。そして、グループ属性を用いたカメラ間人同定の実験において、軌跡とグループ属性の正解データを与えた場合、従来特徴量のみを用いたときと比較して、グループ特徴量を併用することで、カメラ間人同定の精度を44%向上させた。また、グループ検出の実験結果を用いた場合においても、カメラ間人同定の精度が37%向上し、単眼カメラ視野内での人追跡が成功すれば、その軌跡データからグループ属性を導出し、グループ特徴量を併用することでカメラ間人同定の精度を向上できることが分かった。

今後の課題として、カメラ間人同定に用いるグループ特徴量として、本研究ではグループ内色特徴量とグループ内人数特徴量の2つの特徴量を提案したが、グループ内色特徴量の効果が弱いため、本研究でのグループ内色特徴量の取得法を検討する必要がある。また、この2種のグループ特徴量以外の新たなグループ特徴量を加える事で、カメラ間人同定の精度を向上させる可能性を検討する。例えば、グループ検出を行う際に、誰と誰がグループかという情報以外に、親子関係などの社会的関係を推定することで、新たなグループ特徴量の取得が可能となると考えられる。また、本研究において、単眼カメラ視野内における人追跡を行ったが、検出評価値のノイズによる誤検出やオクルージョンによる未検出が多発し、精度としては十分といえる結果が得られなかった。その結果、追跡結果を用いたグループ検出・カメラ間人同定の実験において十分な精度を確かめることができなかった。よって、さらに頑健な単眼カメラ視野内の追跡手法を用いて、追跡結果からグループ検出・カメラ間人同定を行うことで、実環境中で本システムを実行できるように改良をする必要がある。

謝辞

本研究は様々な方々の御指導，御助言，御協力によって行うことができました。

研究全般において，議論を重ね，数多くの御指導，御助言，御鞭撻を賜りました環境知能学研究室萩田紀博教授に心より感謝いたします。

研究において，貴重な御助言を賜りましたインタラクティブメディア設計学研究室加藤博一教授に心より感謝いたします。

研究生生活全般において，新たな分野において右も左も分からない私に，懇切な御指導，御助言，御鞭撻を賜りました浮田宗伯准教授に心より感謝いたします。

研究において，日々の生活やミーティングの中で，数々の貴重な御助言，御鞭撻を賜りました神原誠之准教授に心より感謝いたします。

実験において，貴重な実験データをいただいた ATR 知能ロボティクス研究所の塩見昌裕様並びに研究員の皆様，Graz University of Technology と Austrian Institute of Technology の皆様に心より感謝いたします。

研究生生活において，同期として共に励ましあい，共に支えあった，環境知能学研究室の足立麻衣子さん，柴門恒希君，高橋達君，高砂雄介君に心より感謝いたします。

また，研究室内ミーティングにおける様々なコメントや，研究室生活における様々な話題を出していただいた環境知能学研究室の皆様に感謝いたします。特に，実験データやプログラムの作成においてご協力をいただいた M1 の岡田亜沙美さんに心より感謝いたします。

最後に，2年間の奈良での生活を大阪の実家からサポートしていただいた両親と弟に心より感謝いたします。

参考文献

- [1] 井尻喜久, 川西康友, 美濃導彦, 村瀬洋, “サーベイ論文：視野を共有しない複数カメラ間での人物照合” *電子情報通信学会技術研究報告*, 111(317), pp.117–124, 2011.
- [2] Z. Yucel, T. Ikeda, T. Miyashita, and N. Hagita, “Identification of mobile entities based on trajectory and shape information,” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.
- [3] B. Scholkopf, K. K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, “Comparing support vector machines with Gaussian kernels to radial basis function classifiers,” *IEEE Transactions on Signal Processing*, Vol.45, No.11, pp.2758–2765, 1997.
- [4] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, “You’ll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking,” *IEEE International Conference on Computer Vision*, 2009.
- [5] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, “Tracking Multiple People under Global Appearance Constraints,” *IEEE International Conference on Computer Vision*, 2011.
- [6] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, “Globally-Optimal Greedy Algorithms for Tracking a Variable Number of Objects”, *IEEE Conference on Computer Vision and Pattern Recognition*, 2011
- [7] O. Hamdoun, F. Moutarde, B. Stanculescu, and B. Steux, “Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences,” *ACM/IEEE International Conference on Distributed Smart Cameras*, 2008.
- [8] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, “Shape and Appearance Context Modeling,” *IEEE International Conference on Computer*

Vision, 2007.

- [9] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person Re-Identification by Symmetry-Driven Accumulation of Local Features,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [10] D. Gray and H. Tao, “Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features,” *European Conference on Computer Vision*, 2008.
- [11] C. H. Kuo, C. Huang and R. Nevatia, “Inter-camera Association of Multi-target Tracks by On-Line Learned Appearance Affinity Models,” *European Conference on Computer Vision*, 2010.
- [12] W. S. Zheng, S. Gong and T. Xiang, “Person Reidentification by Probabilistic Relative Distance Comparison,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [13] O. Javed, K. Shafique and M. Shah, “Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [14] A. Gilbert and R. Bowden, “Tracking Objects Across Cameras by Incrementally Learning Inter-camera Colour Calibration and Patterns of Activity,” *European Conference on Computer Vision*, 2006.
- [15] V. Kettner and R. Zabih, “Bayesian Multi-Camera Surveillance,” *IEEE Conference on Computer Vision and Pattern Recognition*, 1999.
- [16] O. Javed, Z. Rasheed, K. Shafique and M. Shah, “Tracking Across Multiple Cameras with Disjoint Views,” *IEEE International Conference on Computer Vision*, 2003.
- [17] D. Makris, T. Ellis and J. Black, “Bridging the gaps between cameras,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

- [18] K. Tieu, G. Dalley and W. L. Grimson, “Inference of Non-Overlapping Camera Network Topology by Measuring Statistical Dependence,” *IEEE International Conference on Computer Vision*, 2005.
- [19] N. Ukita, “Probabilistic-Topological Calibration of Widely Distributed Cameras,” *Machine Vision and Applications*, Vol.18, No.3, pp.249–260, 2007.
- [20] X. Wang, K. Tieu and W. E. L. Grimson, “Correspondence-Free Activity Analysis and Scene Modeling in Multiple Camera Views,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.32, No.1, pp. 56–71, 2010.
- [21] 浜嶋朗, 石川晃弘, 竹内郁郎, “社会学小辞典,” 有斐閣, 2005.
- [22] C. H. Cooley, “Social Organization: A Study of the Larger Mind,” pp.25-31, Charles Scribner’s Sons, 1909.
- [23] A. C. Gallagher and Tsuhan Chen, “Understanding Images of Groups of People,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [24] D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” *Physical Review E*, Vol.51, No.5, pp.4282-4286, 1995.
- [25] R. Mehran, A. Oyama, and M. Shah, “Abnormal Crowd Behavior Detection using Social Force Model,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [26] P. Scovanner and M. F. Tappen, “Learning Pedestrian Dynamics from the Real World,” *IEEE International Conference on Computer Vision*, 2009.
- [27] W. Ge, R. T. Collins, and R. B. Ruback, “Vision-based Analysis of Small Groups in Pedestrian Crowds,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.34, No.5, pp.1003–1016, 2012.

- [28] S. Pellegrini, A. Ess, and L. van Gool, “Improving Data Association by Joint Modeling of Pedestrian Trajectories and Groupings,” *European Conference on Computer Vision*, 2010.
- [29] K. O. Arras, O. Mozos, and W. Burgard, “Using Boosted Features for the Detection of People in 2D Range Data,” *IEEE International Conference on Robotics and Automation*, 2007.
- [30] J. Cui, H. Zhao, and R. Shibasaki, “Fusion of Detection and Matching Based Approaches for Laser Based Multiple People Tracking,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [31] O. M. Mozos, R. Kurazume, and T. Hasegawa, “Multi-Part People Detection Using 2D Range Data,” *International Journal of Social Robotics*, Vol.2, No.1, pp.31–40, 2010.
- [32] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg, “Who are you with and Where are you going?” *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [33] N. Dalal and B. Berg, “Histograms of Oriented Gradients for Human Detection” *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [34] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, Vol.2, Issue.3, pp.27:1–27:27, 2011.
- [35] S. Maji, A. C. Berg and J. Malik, “Classification using Intersection Kernel Support Vector Machines is Efficient” *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [36] V. Ramesh and P. Meer, “Real-Time Tracking of Non-Rigid Objects using Mean Shift” *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

- [37] M. Hirzer, C. Beleznai, P. M. Roth and H Bischof “Person Re-Identification by Descriptive and Discriminative Classification” *Scandinavian Conference on Image Analysis* , 2011.
- [38] Learning, Recognition and Surveillance,
“<http://lrs.icg.tugraz.at/datasets/prid/index.php>”

付録

A. 本研究における用語集

本研究では、数多くの用語が定義されているため、それらの定義とその説明を纏める。

グループ属性に関して

グループ

「位置関係や軌跡の類似性、同時観測性に着目することで検出される1人以上の人の集まり」

これは、Cooley[22]の提唱した第一次集団（直接的接触に基づく親密な関係の社会集団、例：家族、友達等）を1人で行動する人も含むよう拡大解釈した定義である。

グループ属性

「その人が属しているグループの情報」

例えば、グループ内の人数やグループに属している全ての人の色情報など。グループの定義を1人以上に拡大したことによって、1人で行動している人についてもグループ属性が得られる。

グループ内ペア

「グループ（一人のグループを除く）の構成員中の任意の2人」

単眼カメラ視野内における各軌跡から、時空間特徴量を計算することでグループ内ペアの検出を行う。

グループ検出

「グループ内ペア検出結果から、そのカメラ視野内においてグループを検出する処理」

検出されたグループ内ペアの構成員間の重複を調べ、重複が発見されたグループ内ペアの構成員全員を同一グループに属すると判定することで、3人以上のグループを検出する。

カメラ間人同定における特徴量

本研究においてカメラ間人同定を行う際に用いる特徴量を挙げ、カメラ a で撮影された歩行者 i とカメラ b で撮影された歩行者 j の間で算出される人同定用のスコア算出方について説明する。

従来特徴量

「カメラ間人同定の分野において、既存研究で用いられている特徴量である色特徴量と時間特徴量」

色特徴量

「人の色情報を用いた人固有の特徴量」

各カメラで検出された人に関して、色ヒストグラム間の類似度である Bhattacharya 係数を求め、これを色特徴量のスコア R^c とする。

時間特徴量

「カメラ間の移動時間を用いた観測領域固有の特徴量」

カメラ間の移動時間のガウス分布 $g^t(t; m^t, \delta^{t^2})$ を求めておき、このガウス分布に各人のカメラ間移動時間 $t_j^{b,in} - t_i^{a,out}$ を代入した値を、時間特徴量のスコア R^t とする。

グループ特徴量

「カメラ間人同定において、本研究で提案するグループ属性から算出される特徴量であるグループ内人数特徴量とグループ内色特徴量」

グループ内人数特徴量

「グループ属性の内、同一グループに属する人数の情報を用いて算出される特徴量」

同一グループに属する人数が一致したときにスコアが高くなるようなガウス分布 $g^{gp}(p^{gp}; m^{gp}, \delta^{gp^2})$ を求めておき、このガウス分布にグループ内人数の差 $p_{i,j}^{gp}$ を代入した値を、グループ内人数特徴量のスコア R^{gp} とする。

グループ内色特徴量

「グループ属性の内、同一グループに属する全ての人の色情報を

用いて算出される特徴量」

同一グループに属する人全員のヒストグラム $P_1^g, P_2^g, \dots, P_N^g$, を加算し, 正規化したグループ内ヒストグラム P^g を算出し, グループ内ヒストグラム間で Bhattacharyya 係数を求め, これをグループ内色特徴量のスコア R^{gc} とする.

B. SVMについての説明

SVM (Support Vector Machine) は、クラス分類の分野においてよく用いられている識別関数の一種であり、マージン最大化を行うことで汎化性能を上げる手法である。本研究において人検出やグループ検出においてSVMを用いているため、ここではSVMの理論について説明する。

図39のように特徴空間中に存在するN個の入力パターン (x_1, \dots, x_N) が2クラスに分類されている。入力パターン i のクラスを $c_i \in \{-1, 1\}$ とおく時、これらを学習することで、未知のデータ点 x を代入した時に、その符号からクラス分類を行うことができる線形識別関数 $l(x)$ (図39中の黒太線) を設計する。

$$l(x) = wx + b \quad (7)$$

この線形識別関数 $l(x)$ の w, b をN個の入力パターンを用いた学習によって算出する。ここで、この線形識別関数の汎化性能 (未知のデータに対する識別性能) を上げる為に、マージン最大化を行う。このマージンとは、図39で表されているように、他クラスとの距離が最も近い位置にいる入力パターンにおける識別関数との距離を示している。このマージンを最大化することで汎化性能の向上を目指している。このときマージンを最大化する式は以下の式で表される。

$$\min_w \frac{1}{2} \|w\|^2 \quad (8)$$

また、この線形識別関数によって、N個のパターンが全て分類されるために、N個のパターンすべてについて以下の制約式が成立する。

$$c_n(wx_n + b) \geq 1, \quad n = 1, 2, \dots, N \quad (9)$$

この制約付き最適化問題を解くために、(9)の制約式ごとにラグランジュ乗数 $a_n \geq 0$ を導入すると、次のラグランジュ関数が得られる。

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{n=1}^N a_n \{c_n(wx_n + b) - 1\} \quad (10)$$

この式(10)を w および b で偏微分した式を 0 とすることで、 w および b の値が求まり、線形識別関数 $l(x)$ が求まる。

以上の式は図 39 のような特徴空間中において入力パターンが線形分離可能な配置の場合 $l(x)$ の算出法であるが、図 40 のように N 個の入力パターンが線形分離が可能でないように場合もある。こういった場合では、汎化誤差 $\xi_n (\geq 0)$ を許すことで、式(8)を以下のように変化させる ($C > 0$)。

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{n=1}^N \xi_n \quad (11)$$

この誤差は ξ_n は図 41 のようなヒンジ関数で表すことができ、式(11)は次のように変形できる。

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{n=1}^N \max(0, c_n(wx_n + b)) \quad (12)$$

この式(12)を使って、線形分離可能な場合と同様に、制約式を満たしつつ最適化問題を解くことで、線形分離が可能でない場合における線形識別関数 $l(x)$ を求める。

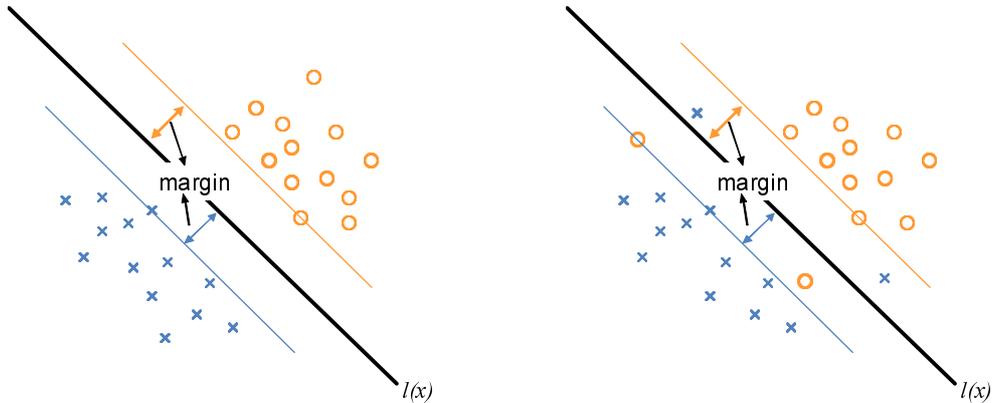


図 39 入力パターンが特徴空間中で線形分離可能な場合の線形識別関数 $l(x)$ の算出の概念図.
 図 40 入力パターンが特徴空間中で線形分離可能でない場合の線形識別関数 $l(x)$ の算出の概念図.

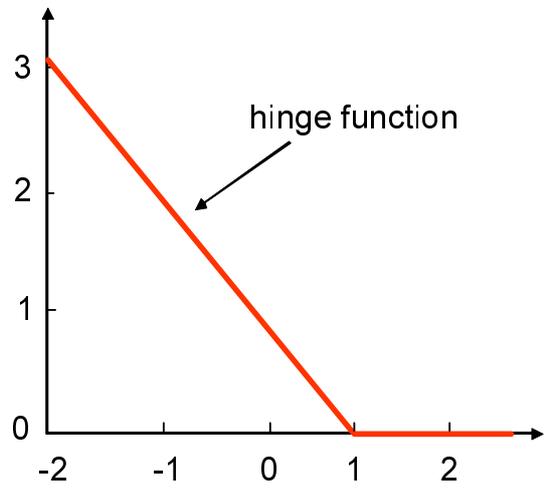


図 41 入力パターンが特徴空間中で線形分離可能でない場合に用いるヒンジ関数.