

特別研究報告書

カメラごとの顔検出数の比率に着目した
特定人物の所在推定

指導教員 椋木 雅之 准教授

京都大学工学部情報学科

藤野 拓海

平成 27 年 1 月 30 日

カメラごとの顔検出数の比率に着目した特定人物の所在推定

藤野 拓海

内容梗概

近年, 犯罪の予防を図るために街頭の防犯カメラの設置台数が増加している. そのため, 人々が日常生活の中で防犯カメラに写る機会は多くなっている. 防犯カメラ映像は犯罪抑止の目的以外に, 人物の行動解析にも用いることができる. 多数の人物が生活している街中の広範囲な地域において特定人物の日常的な行動範囲を知ることができれば, 人物の搜索や都市計画に応用することができる.

本研究では複数のカメラが設置されている地域において, 多数のカメラ映像から特定人物を探しだし, 特定人物の所在を推定することを目的とする. ここで, 所在とはある人物の日常的な行動範囲のこととする. 多くの人物は通勤・通学など毎日ほぼ同じように行動するので, 長期間観測すれば日常的に観測されるカメラが存在する. カメラが密に設置されていれば, ある特定人物が日常的に観測されている複数のカメラを知ること, 各カメラの位置から大まかな所在がわかる.

特定人物が観測されているカメラを求めるために, カメラ映像から人物の顔を検出し, 顔認識を用いて特定人物を探す. 現在, 顔認識の性能は非常に良くなっている. しかし, 顔認識の性能がどれほど高くても誤認識は起こる. 所在推定の対象となる地域の人口が多くなるほどカメラ映像に写る人物も多くなり, 特定人物の正認識よりも誤認識のほうが多くなってしまおうという問題が生じる.

以上の問題に対して本研究では, 各カメラでの誤認識数を予測し, 実際の特定人物の顔認識数と誤認識の予測数の差を取ることで所在を推定する手法を提案する. 特定人物が観測されるカメラは全カメラに対して一部である. よって多くのカメラで顔認識はすべて誤認識である. また全カメラで他人受入率が一定であれば, あるカメラでの誤認識数は他人受入率とそのカメラでの顔検出数の積に近づく. このときカメラごとの全人物の顔検出数と特定人物の顔認識数を用いて他人受入率を推定し, 各カメラでの特定人物の誤認識数を予測することができる. 一方で, 特定人物が日常的に観測されるカメラでは長期間観測することにより顔認識数とその分多くなる. よって特定人物の顔認識数と誤認識の予測数の差が大きいカメラを特定人物の所在とし, 誤認識の影響を抑える.

実際には他人受入率はカメラごとに多少の変動がある。他人受入率の変動により、誤認識の予測数と真の誤認識数の誤差が大きくなると、上記の手法では精度が低くなる。

これに対処するため、本研究ではさらに移動制約を導入し誤認識の影響を低減する。特定人物は移動に基づき、一定以上時間を空けずに順次カメラに観測される。これを移動制約とする。特定人物が観測されるカメラは特定人物の移動に基づくので、特定人物の正認識結果は移動制約を満たす。一方、誤認識は多くの異なる人物から得られたものであり移動制約を満たさないものが多い。したがって移動制約を導入することにより特定人物の正認識を残したまま、誤認識を除去することができる。実際には誤認識にも移動制約を満たしてしまうものがあるが、このような誤認識結果は全人物からランダムに選択した顔検出結果からシミュレートできる。シミュレーションを繰り返し行うことで、移動制約を満たす誤認識結果の数を予測し除外する。

提案手法の有効性を示すために評価実験を行った。ある地域にカメラを多数設置し、不特定多数の人物の顔画像データを収集することは困難であるので、シミュレーションにより特定人物の顔認識結果を生成する。顔認識の精度は所在推定に大きく影響すると考えられるので、実際に多数の顔画像を用いて顔認識した結果から人物ごとの認識率をモデル化した。このシミュレータを用いて、所在推定の対象となる地域の人物数と観測日数が変化したときの各手法での効果を調べた。実験結果から、カメラごとの顔検出数の比率に基づいて誤認識数を推定し誤認識の影響を抑えることで長期間観測により所在推定の精度が向上すること、移動制約を導入することにより少ない観測日数での所在推定の精度低下を抑えられることを示した。

今後の課題としては、実環境における実験を行うことが挙げられる。また、顔認識の方法は本研究では検討していない。誤認識を後処理である程度除去するなど、所在推定に適した顔認識手法を検討することも必要である。

Living Area Estimation based on the Ratio of the Number of Face Detection Results per Camera

Takumi FUJINO

Abstract

In recent years, many sites in a city are observed by surveillance cameras, thus people have many opportunities to be observed by them in daily life. Hence, videos taken by surveillance cameras are also utilized for human behavior analysis.

We aim to estimate a living area of a specific person by utilizing videos taken by surveillance cameras in a city. Each person walks the same route nearly every day, thus he/she is routinely observed by the specific cameras. If surveillance cameras are densely installed, the cameras by which the specific person is routinely observed indicate the living area of the specific person.

To determine the cameras by which the specific person is routinely observed, we detect faces in videos and distinguish facial images of the specific person from those of others by face recognition. Currently the accuracy of face recognition has become very high, however false recognition can occur. As the number of people observed by the cameras increases, the number of false recognition results becomes larger than that of true recognition results. In this case, we cannot find the living area of the specific person.

This paper proposes a method for living area estimation by subtracting the expected number of false recognition results from the actual number of face recognition results per camera to suppress the influence of false recognition.

We suppose that false acceptance rates (FAR) per camera is equal. In this case, ratios between the number of face detection results and the number of false recognition results are constant among the cameras. Since the number of cameras by which the specific person is observed is small, the number of false recognition results and the number of face recognition results are equal in many cameras. Therefore, FAR can be estimated by using ratios between the number of face detection results and the number of face recognition results per camera, namely, the number of false recognition results is predictable. As

the specific person is routinely observed by some cameras, the number of true recognition results of such cameras becomes large. Therefore a difference between the expected number of false recognition results and the actual number of face recognition results becomes large on these cameras, which indicate the living area of the specific person.

Actually, FAR of each camera is different from those of the other cameras, which causes errors of the expected number of false recognition results. Due to these errors, the accuracy of the above method may be degraded.

To deal with this problem, we introduce a transition constraint to suppress the influence of false recognition. The specific person is observed by cameras within a certain time interval based on the transition, which we call the transition constraint. The true recognition results of the specific person meet the transition constraint. On the other hand, the false recognition results, which consist of the recognition results of many different people, seldom meet the transition constraint. Therefore a ratio of the true recognition results becomes high by eliminating recognition results which do not meet the transition constraint. In fact, several false recognition results meet the transition constraint accidentally. We predict and eliminate the influence of such false recognition results.

We conducted experiments for evaluating our proposed method. Since it is difficult to collect many face images taken by surveillance cameras in a city from the point of view of privacy, we generated face recognition results from simulation. The accuracy of living area estimation of the specific person is affected by recognition rate per person. Hence, we designed the simulation based on the actual face recognition results. The result shows that the accuracy of living area estimation of the specific person is increased by our method as the observation period becomes long. It also shows that the transition constraint keeps the accuracy of the living area estimation of the specific person under the short-term observation.

To apply our method to real data is left for future work.

カメラごとの顔検出数の比率に着目した特定人物の所在推定

目次

第1章	序論	1
第2章	特定人物の所在推定	3
2.1	問題設定	3
2.2	所在推定の基本的な処理の流れ	4
2.3	カメラごとの顔認識数に基づく所在推定手法	4
2.4	顔認識の関連研究と所在推定における問題点	5
第3章	カメラごとの顔検出数の比率に着目した特定人物の所在推定手法	6
3.1	カメラごとの顔検出数の比率に着目した所在推定	6
3.2	移動制約を用いた所在推定	7
3.3	カメラごとのスコア決定	9
3.4	所在推定の手順	9
3.4.1	カメラごとの顔検出数の比率に着目した所在推定でのスコア決定	9
3.4.2	移動制約を用いた所在推定でのスコア決定	11
3.4.3	特定人物の所在の推定	13
第4章	人物の行動と特定人物の顔認識結果のシミュレーション	13
4.1	シミュレータの概要	13
4.2	シミュレータの処理の流れ	14
4.3	実際の顔認識結果に基づく顔認識の性能のモデル化	16
4.3.1	顔認識による特定人物の判定方法	16
4.3.2	顔認識の性能のモデル化	17
第5章	シミュレーションによる特定人物の所在推定の実験	20
5.1	方法	20
5.2	結果	22
5.3	考察	22
第6章	結論	26
	謝辞	27

第1章 序論

近年, 犯罪の予防を図るために街頭の防犯カメラの設置台数が増加している. そのため, 人々が日常生活の中で防犯カメラに写る機会は多くなっている. 通常防犯カメラは犯罪抑止や事件が起きたときの証拠確保の目的で設置されるが, カメラ映像は人物の行動解析にも用いることができる. 大西ら [1] は商業施設における顧客行動解析を目的として, 大型施設における長期間にわたる人流比較解析手法を提案している. 街頭に設置された様々な地点のカメラを用いれば, より広範囲で人物の行動傾向を知ることができる.

多数の人物が生活している街中の広範囲な地域において個人の日常的な行動範囲を知ることができれば, 様々な応用ができる. 例えば, 人物の搜索の際, あらかじめ探したい人物が日常的に生活している地域を絞り込むことができれば, 搜索の効率化ができる. また, 個人の日常的な行動範囲を知るとは, 都市の構造や特徴を理解する上でも重要である. Cranshaw ら [2] はソーシャルメディアを活用して, 個人の行動パターンを反映した都市の構造を発見するモデルを提案した. ここで, 行動解析の対象となる個人を特定人物とよぶ. 特定人物の日常的な行動範囲は特定人物の持つスマートフォンの位置情報を利用したり, Cranshaw らのように特定人物の位置情報を蓄積しているソーシャルメディアを活用することによって獲得することができる. しかし, この方法ではスマートフォンを持たない人やソーシャルメディアを活用しない人の行動は知ることができない. 一方で, 街頭の防犯カメラ映像は, 特定人物に特別な条件を必要としない.

本研究では複数のカメラが設置されている地域において, 多数のカメラ映像から特定人物を探しだし, 特定人物の所在を推定することを目的とする. ここで, 所在とは個人の日常的な行動範囲のこととする. 多くの人物は通勤・通学など毎日ほぼ同じように行動するので, 長期間観測すれば日常的に観測されるカメラが存在する. カメラが密に設置されていれば, ある特定人物が日常的に観測されている複数のカメラを知ることによって, 各カメラの位置から大まかな所在がわかる. よって, 所在はカメラの集合で表す. 人物の服装は毎日一定ではないので, カメラ映像から人物の顔を検出し, その中から顔認識により特定人物が観測されているカメラを求める.

現在, 顔認識の精度は非常に良くなっている. したがって単純な所在推定手法として, 特定人物の顔認識をカメラごとに行い, 顔認識数が多い順にカメラを選

択して所在とする方法が考えられる。しかし、顔認識の精度がどれほど高くても特定人物と異なる人物を誤認識する。所在推定の対象となる地域の人口が多くなるほど、カメラ映像に写る人物も多くなり、特定人物の誤認識数のほうが正認識数と比べて多くなる。例えば、1日にカメラ映像に写る回数をどの人物も100回とする。また顔認識の精度を本人拒否率 (False Rejection Rate ; FRR) と他人受入率 (False Acceptance Rate ; FAR) で表す。FRRは本人が認識されない誤りの割合、FARは他人が誤って認識される割合である。このときFRRが10%、FARが0.1%であるとすると、カメラ映像に映る人物が特定人物の他に1日10000人いれば、特定人物の正認識は90件、誤認識は1000件生じる。したがって単純な所在推定手法では、カメラ映像に写る人物が多くなるほど精度が低くなるという問題がある。

以上の問題に対して本研究では、各カメラごとに誤認識数を予測し、実際の特定人物の顔認識数と誤認識の予測数の差を取ることで所在を推定する手法を提案する。

所在推定の対象となる地域は十分広く、人口も十分多いものとする。この時、特定人物が日常的に観測されるカメラは全カメラに対してごく一部である。更に、FARはどのカメラでも変動は小さく一定であるとする。カメラごとにFARが完全に同じであれば、どのカメラにおいても特定人物の誤認識数は全人物の顔検出数とFARの積に近づく。特定人物が観測されないカメラは多く、そのカメラでの顔認識結果は全て誤認識である。よってこれら多くのカメラからFARを推定できる。このFARの推定値から、特定人物の誤認識数をカメラごとに予測することができる。一方で、特定人物が日常的に観測されるカメラでは長期間観測することにより顔認識数が多くなり、その顔認識数と誤認識の予測数の差は大きくなる。よって長期間観測し特定人物の顔認識数と誤認識の予測数の差が大きいカメラの集合を特定人物の所在とすることで、誤認識の影響を抑えることができる。この手法をカメラごとの顔検出数の比率に着目した所在推定手法とよぶ。

実際にはFARは多少の変動がある。FARの変動により、誤認識の予測数と真の誤認識数の誤差が大きくなると、カメラごとの顔検出数の比率に着目した所在推定手法では精度が落ちてしまう。

これに対処するために、本研究ではさらに移動制約を導入し、誤認識の影響を低減する。特定人物は移動に基づき、一定以上時間を空けずに順次カメラに観測

される。これを移動制約とする。FRR が小さい場合、特定人物が検出されていれば高い割合で顔認識される。このとき特定人物が観測されるカメラは特定人物の移動に基づくので、特定人物の正認識結果は移動制約を満たす。一方、誤認識は多くの異なる人物から得られたものであり移動制約を満たさないものが多い。よって移動制約を満たさない顔認識結果を除外することにより、所在推定の精度を高めることができる。実際には誤認識にも移動制約を満たしてしまうものがある。それらは実際の移動に基づいているわけではなく、多くの異なる人物の顔認識結果が各カメラでの顔検出数やカメラ間の距離に依存して偶然移動制約を満たしたものである。人物の移動に関わらず移動制約を満たすこのような顔認識結果は、全人物からランダムに選択した顔検出結果からシミュレートできる。シミュレーションを繰り返し行うことで、移動制約を満たす誤認識結果が生じる比率を推定する。カメラごとの顔検出数の比率に着目した所在推定手法と同様に、推定した比率から移動制約を満たす誤認識の発生数を予測し、実際に得られた移動制約を満たす顔認識結果との差が大きいカメラの集合を所在とする。

本稿の構成は以下の通りである。2章では特定人物の所在推定の問題設定について述べる。3章では、本研究で提案するカメラごとの顔検出数の比率に着目した特定人物の所在推定手法について述べる。4章では人物の行動と特定人物の顔認識結果のシミュレーションについて述べる。5章では提案手法を評価するために行ったシミュレーションによる特定人物の所在推定の実験の詳細について述べる。最後に6章で結論を述べる。

第2章 特定人物の所在推定

2.1 問題設定

所在推定の対象となる地域

多くの人物が生活している市内や町内のような地域を対象とする。対象の地域内の様々な地点に設置されているカメラが路上の人物を撮影する。カメラの設置台数を N_c とし、設置されているカメラの集合を $\mathcal{C} = \{c_i\}_{i=1}^{N_c}$ とする。 $i \neq j$ のとき、カメラ c_i とカメラ c_j の視野は重複していないと仮定する。この地域を N_d 日間観測することで、カメラで撮影された映像が複数得られる。

カメラ映像からの顔検出による人物の検出結果の取得

映像の各フレームの画像から、顔検出技術と物体追跡技術によってある人物が映像に写り始めてから写り終わるまでの顔画像列 I_{d_i} が得られる。顔画像は、ある人物の顔の領域が矩形で切りだされたものである。顔画像列 I_{d_i} と、人物が写り始めた日時 t_{d_i} 、観測されたカメラ c_{d_i} の3つ組を1つの顔検出結果 $d_i = (I_{d_i}, t_{d_i}, c_{d_i})$ とする。これを得られた全ての映像に対して行い、全人物の顔検出結果の集合 $\mathcal{D} = \{d_i\}_{i=1}^{|\mathcal{D}|}$ を得る。

特定人物の所在推定

本研究では、特定人物 p_λ の顔画像と全人物の顔検出結果の集合 \mathcal{D} を入力として、特定人物 p_λ が日常的に観測されるカメラの集合 $\mathcal{C}_{p_\lambda} \subset \mathcal{C}$ を得る問題を扱う。

2.2 所在推定の基本的な処理の流れ

所在推定問題を検討する上で、まず所在推定の基本的な処理の流れを説明する。

step1

各顔検出結果 $d_i \in \mathcal{D}$ に対して、顔画像列 I_{d_i} と特定人物 p_λ の顔画像を入力とし、顔認識技術を用いて d_i が特定人物 p_λ の顔検出結果であるかどうかを判定する。特定人物 p_λ であると判定した顔検出結果を特定人物 p_λ の顔認識結果とし、特定人物 p_λ の顔認識結果の集合を $\mathcal{D}_{p_\lambda} = \{d_{p_\lambda i}\}_{i=1}^{|\mathcal{D}_{p_\lambda}|} \subset \mathcal{D}$ とする。 \mathcal{D}_{p_λ} には特定人物 p_λ の正認識結果と特定人物 p_λ とは異なる人物の誤認識結果が含まれている。特定人物 p_λ の正認識結果の集合を $\mathcal{D}_{p_\lambda}^+$ 、特定人物 p_λ とは異なる人物の誤認識結果の集合を $\mathcal{D}_{p_\lambda}^-$ とする。 $\mathcal{D}_{p_\lambda} = \mathcal{D}_{p_\lambda}^+ \cup \mathcal{D}_{p_\lambda}^-$ である。また $\mathcal{D}_{p_\lambda}^+$ 及び $\mathcal{D}_{p_\lambda}^-$ は未知である。

step2

\mathcal{D} 及び \mathcal{D}_{p_λ} を入力とし、特定人物 p_λ の所在の推定結果 \mathcal{C}_{p_λ} を求める。

本研究では step1 の処理は既存の方法を用いて行うとして、特に step2 の処理手法について検討する。

2.3 カメラごとの顔認識数に基づく所在推定手法

2.2 節の step2 において、まず「カメラごとの顔認識数に基づく所在推定手法」が単純な方法として考えられる。

多くの人物は通勤・通学など毎日ほぼ一定の行動がある。このとき長期間観測することによって、特定人物 p_λ が日常的に観測されるカメラが複数台存在する。

カメラ c_i での全人物の顔検出結果の集合を $\mathcal{D}_{c_i} = \{d_j | d_j \in D \text{ かつ } d_j \text{ はカメラ } c_i \text{ で観測された検出結果}\}$ とし, カメラ c_i での特定人物 p_λ の顔認識結果の集合を $\mathcal{D}_{p_\lambda c_i} = \mathcal{D}_{p_\lambda} \cap \mathcal{D}_{c_i}$ とする. カメラ c_i が特定人物を日常的に観測するカメラであれば, カメラ c_i での顔認識数 $|\mathcal{D}_{p_\lambda c_i}|$ はその分多くなる. したがって閾値を定め, カメラ c_1, \dots, c_{N_c} でのそれぞれの認識数 $|\mathcal{D}_{p_\lambda c_1}|, \dots, |\mathcal{D}_{p_\lambda c_{N_c}}|$ が閾値以上のカメラを選択する. 得られたカメラの集合を \mathcal{C}_{p_λ} とする.

FRR が大きければ $|\mathcal{D}_{p_\lambda}^+|$ が小さくなり, 特定人物が日常的に観測されるカメラ c_i での顔認識数 $|\mathcal{D}_{p_\lambda c_i}|$ は多くなり, FAR が大きければ $|\mathcal{D}_{p_\lambda}^-|$ が大きくなり, 特定人物が観測されないカメラ c_j での顔認識数 $|\mathcal{D}_{p_\lambda c_j}|$ が多くなる. したがって, このカメラごとの顔認識数に基づく所在推定手法によって所在を推定するためには, FRR と FAR がともに小さいことが必要条件である. しかしそれだけでは所在は推定できない. 次の 2.4 節で現在の顔認識の精度は高く FRR と FAR がともに小さいこと, 一方で顔認識の精度が高くてもカメラごとの顔認識数に基づく所在推定手法では所在推定できないことを述べる.

2.4 顔認識の関連研究と所在推定における問題点

まず現在の顔認識の性能について述べる. 顔認識は所在推定だけでなくセキュリティ・認証システムや人物が写った写真の自動整理など幅広い応用が考えられ, さかんに研究が行われている. Malek ら [3] は顔の見た目の変化に強い顔認証手法を提案した. Malek らの研究では人物 60 人で作成した顔の向きが様々な顔画像 180 枚のデータベースを用いた実験を行い, 顔認証のパラメータを変化させて FRR と FAR の値が等しくなるように設定したとき, その FRR, FAR の値は平均で 5.17% になることが示されている. また, Taigman ら [4] はディープラーニングを用いた顔認識手法を提案している. Taigman らの手法では人間が実際に認識を行ったのに近い精度で顔認識が可能であることが示されている.

以上より顔認識の性能は現在非常に高くなっている. よって特定人物 p_λ の正認識結果は高い精度で \mathcal{D}_{p_λ} に含まれる. しかし, どれだけ顔認識の性能が良くても, 撮影された顔の向きや照明などの撮影条件によって異なる人物を誤認識する. 所在推定の対象となる地域の人口が多い場合, 観測される人物は数千人から数万人になる. このとき誤認識される人物も多くなり, 特定人物 p_λ の正認識結果 $\mathcal{D}_{p_\lambda}^+$ に比べて特定人物 p_λ の誤認識結果 $\mathcal{D}_{p_\lambda}^-$ が多くなる. この $\mathcal{D}_{p_\lambda}^-$ の影響で, カメラ c_i で顔検出数 $|\mathcal{D}_{c_i}|$ が大きければ特定人物 p_λ の顔認識数 $|\mathcal{D}_{p_\lambda c_i}|$ も大きく

なり, $|\mathcal{D}_{c_i}|$ が小さければ $|\mathcal{D}_{p_\lambda c_i}|$ も小さくなる. つまり, 特定人物 p_λ の所在に関わらず各 $|\mathcal{D}_{p_\lambda c_i}|$ は各 $|\mathcal{D}_{c_i}|$ に依存するという問題が生じる. この問題のため 2.3 節で述べたカメラごとの顔認識数に基づく所在推定手法では正しい所在は得られない.

第3章 カメラごとの顔検出数の比率に着目した特定人物の所在推定手法

本章では 2.4 節で述べた問題に対処するための所在推定手法について述べる. その前提として, 本章では 2.4 節で述べたように顔認識の性能は良いと仮定する. つまり FRR と FAR は比較的小さいとする. また, FRR と FAR は小さい範囲でも人物ごとあるいはカメラごとに変動がある. 本研究ではカメラごとの FRR と FAR の変動はごく小さいと仮定する. また特定人物 p_λ に対する顔認識の性能は未知である.

3.1 カメラごとの顔検出数の比率に着目した所在推定

2.4 節で述べた問題に対処するため, カメラ c_i での特定人物 p_λ の誤認識数がカメラ c_i での全人物の検出数 $|\mathcal{D}_{c_i}|$ に依存するという性質を利用して, カメラごとの特定人物 p_λ の誤認識数を予測する. 図 1 にカメラごとの顔検出数の比率に着目した所在推定手法の概要を示す. カメラ c_i での特定人物の誤認識結果の集合を $\mathcal{D}_{p_\lambda c_i}^- = \mathcal{D}_{p_\lambda}^- \cap \mathcal{D}_{c_i}$ とする. $FAR = \frac{|\mathcal{D}_{p_\lambda}^-|}{|\mathcal{D}|}$ である. FAR が比較的小さくどのカメラでも一定である場合, 任意の $i = 1, \dots, N_c$ において $|\mathcal{D}_{p_\lambda c_i}^-| = FAR \times |\mathcal{D}_{c_i}|$, すなわち式 (1) が成り立つ.

$$\frac{|\mathcal{D}_{p_\lambda c_i}^-|}{|\mathcal{D}_{p_\lambda}^-|} = \frac{|\mathcal{D}_{c_i}|}{|\mathcal{D}|} \quad (1)$$

これは, 特定人物の誤認識数 $|\mathcal{D}_{p_\lambda}^-|$ に対する各カメラごとの誤認識数 $|\mathcal{D}_{p_\lambda c_i}^-|$ の比率が, 全人物の顔検出数 $|\mathcal{D}|$ に対する各カメラごとの顔検出数 $|\mathcal{D}_{c_i}|$ の比率と一致していることを表す. 所在推定の対象となる地域が広い場合, 特定人物 p_λ が日常的に観測されるカメラは全カメラに対してごく一部であるため, 多くのカメラ c_i で $|\mathcal{D}_{p_\lambda c_i}^-| = |\mathcal{D}_{p_\lambda c_i}|$ である. よって式 (1) より $\frac{|\mathcal{D}_{p_\lambda c_i}^-|}{|\mathcal{D}_{c_i}|}$ は一定値となるので FAR が予測できる. FAR の予測値を \widehat{FAR} とする. このときカメラ c_i での誤認識の予測数を $|\widehat{\mathcal{D}_{p_\lambda c_i}^-}|$ とすると, $|\widehat{\mathcal{D}_{p_\lambda c_i}^-}| = \widehat{FAR} \times |\mathcal{D}_{c_i}|$ である. 一方で特定人物 p_λ が日常的に観測されるカメラでは $|\mathcal{D}_{p_\lambda c_i}^-| < |\mathcal{D}_{p_\lambda c_i}|$ である. したがって各 c_i に

において $|\mathcal{D}_{p_\lambda c_i}|$ と $|\widehat{\mathcal{D}}_{p_\lambda c_i}^-|$ の差を取ると、この差は特定人物 p_λ が日常的に観測されるカメラで他のカメラよりも大きくなる。この差が大きいカメラを特定人物の所在とすることで、 $\mathcal{D}_{p_\lambda}^-$ の影響を抑えることができる。

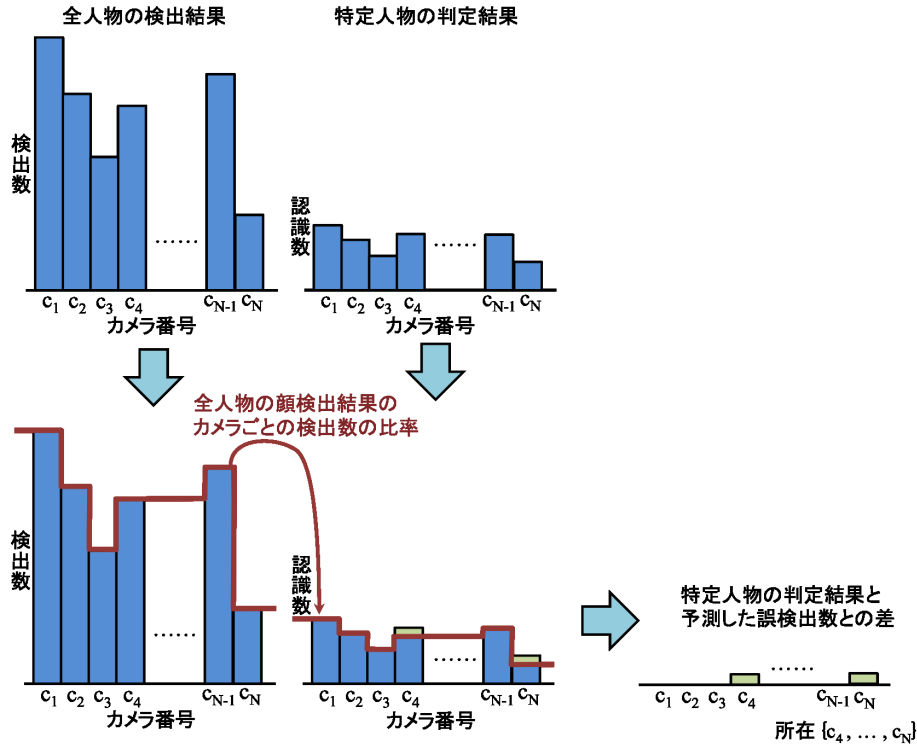


図 1: カメラごとの顔検出数の比率に着目した所在推定

3.2 移動制約を用いた所在推定

カメラごとに FAR は完全に一定ではなく多少変動がある。 $|\mathcal{D}_{p_\lambda c_i}|$ が大きいカメラ c_i で FAR の予測値が少しずれると $|\widehat{\mathcal{D}}_{p_\lambda c_i}^-|$ が大きく変化するため、所在推定の誤りの原因となる。これに対処するために、移動制約を導入して誤認識の影響を抑える。

特定人物は移動に基づきその経路上のカメラで順次観測される。任意の 2 つの顔認識結果 $d_{p_\lambda i}, d_{p_\lambda j}$ が一人の人物の移動から得られるとき、カメラ $c_{d_{p_\lambda i}}, c_{d_{p_\lambda j}}$ の距離に対して観測日時 $t_{d_{p_\lambda i}}, t_{d_{p_\lambda j}}$ の時間間隔は一定の範囲に収まる。これを 2 つの顔認識結果間の移動制約とする。複数の顔認識結果を観測日時順に並べたものを、顔認識結果の時系列とよぶ。ある顔認識結果の時系列を $\mathcal{D}_{seq} = \{d_{p_\lambda k_1}, \dots, d_{p_\lambda k_n}\} (i <$

$j \Rightarrow t_{d_{p_\lambda k_i}} \leq t_{d_{p_\lambda k_j}}$) とする. すべての $i \in 1, \dots, n-1$ において $d_{p_\lambda k_i}$ と $d_{p_\lambda k_{i+1}}$ が移動制約を満たすとき, 認識結果の時系列 \mathcal{D}_{seq} もまた移動制約を満たしているとする. ここで, C の要素数 2 以上の部分集合の集合を $\{\{c_1, c_2\}, \{c_1, c_3\}, \dots, \{c_1, c_{N_c}\}, \{c_2, c_3\}, \dots, \{c_1, c_2, \dots, c_{N_c}\}\} = \{\mathcal{C}_1, \dots, \mathcal{C}_{N_c}\}$ とする. このとき \mathcal{D}_{seq} からカメラの集合 $\mathcal{C}_{seq} \in \{\mathcal{C}_1, \dots, \mathcal{C}_{N_c}\}$ への写像 ϕ を式 (3) 定義する.

$$\mathcal{C}_{seq} = \phi(\mathcal{D}_{seq}) \quad (2)$$

$$= \{c_i \mid \text{ある } d_{p_\lambda k_j} \in \mathcal{D}_{seq} \text{ が存在して } c_i \text{ は } c_{d_{p_\lambda k_j}} \text{ と等しい}\} \quad (3)$$

FAR が小さいとき誤認識結果は異なる多数の人物から発生するので移動制約を満たさない場合が多い. 一方で, FRR が小さいとき正認識結果は特定人物の移動から得られるので他の顔認識結果と移動制約を満たす. よって, 他どの顔認識結果とも移動制約を満たさない顔認識結果を \mathcal{D}_{p_λ} から除外することにより, 誤認識の影響を抑えることができる.

\mathcal{D}_{p_λ} から移動制約を満たす顔認識結果の時系列 $\mathcal{D}_{seq_1}, \dots, \mathcal{D}_{seq_m}$ を求める. これらを求めることにより, 他どの顔認識結果とも移動制約を満たさない顔認識結果を除くことができる. さらに特定人物 p_λ の毎日の一定の行動から, これらの時系列をそれぞれ ϕ によって $\mathcal{C}_{seq_1}, \dots, \mathcal{C}_{seq_m}$ に写すと, この中に特定人物 p_λ が日常的に観測されるカメラを要素とするカメラ集合が多く存在する. 各 \mathcal{C}_{seq_i} は $\mathcal{C}_1, \dots, \mathcal{C}_{N_c}$ のいずれかと等しい. 各 $\mathcal{C}_{seq_1}, \dots, \mathcal{C}_{seq_m}$ において, $\mathcal{C}_1, \dots, \mathcal{C}_{N_c}$ の中の等しいものにそれぞれ一票ずつ投票する. 各カメラ集合 \mathcal{C}_i への投票数を $m_{p_\lambda}(\mathcal{C}_i)$ とする. 特定人物 p_λ が日常的に観測されるカメラを要素とするカメラ集合ではこの投票数が多くなる.

実際には多数の誤認識により特定人物の所在と関係のないカメラ集合も多く出現する. 誤認識の影響を抑えるために, 誤認識によって起こる各カメラ集合への投票数の比を推定する. 誤認識結果は未知であるが, 異なる多数の人物から発生する. そのため誤認識による各カメラ集合への投票数は, 人物に依存せずカメラ間の距離やカメラごとの顔検出数の影響により多くなる. よって \mathcal{D} からランダムに顔検出結果を選び作成した集合で誤認識による各カメラ集合への投票をシミュレートできる. このシミュレーションを繰り返し行うことで, 全投票数に対する $\mathcal{C}_1, \dots, \mathcal{C}_{N_c}$ への投票数の比率を推定する. この比率を用いて 3.1 節と同様に, カメラ集合 \mathcal{C}_i への投票の予測数 $\hat{m}_{p_\lambda}(\mathcal{C}_i)$ を求める. 特定人物 p_λ が日常的に観測されるカメラ集合では $\hat{m}_{p_\lambda}(\mathcal{C}_i) < m_{p_\lambda}(\mathcal{C}_i)$ である. したがって各 \mathcal{C}_i にお

いて $m_{p_\lambda}(\mathcal{C}_i)$ と $\hat{m}_{p_\lambda}(\mathcal{C}_i)$ の差を取ると、この差は特定人物 p_λ が日常的に観測されるカメラ集合で他のカメラ集合よりも大きくなる。この差が大きいカメラ集合に含まれるカメラを特定人物の所在とすることで、誤認識の影響をさらに抑えることができる。

3.3 カメラごとのスコア決定

FAR の変動は小さいが $|\mathcal{D}|$ は大きいので、FAR の少しの変動で $|\mathcal{D}_{p_\lambda}^-|$ は大きく変わる。3.1 節で述べたカメラごとの顔検出数の比率に着目した所在推定手法では、FAR が比較的大きいとき特定人物 p_λ の認識結果の集合 \mathcal{D}_{p_λ} における $\mathcal{D}_{p_\lambda}^+$ の割合が $\mathcal{D}_{p_\lambda}^-$ の割合に対して小さくなり、3.1 節で述べたカメラごとの顔検出数の比率に着目した所在推定手法では精度が低い。また、FRR にも多少変動がある。FRR が比較的大きいとき特定人物 p_λ の正しい認識結果の時系列が含まれる割合が小さく、3.2 節で述べた移動制約を用いた所在推定手法では精度が低い。しかし特定人物 p_λ に対して顔認識の精度は未知であり、どちらの手法が有効であるかはわからない。以上を踏まえて、カメラごとの顔検出数の比率に着目した所在推定手法と移動制約を用いた所在推定手法で得られたそれぞれのカメラ集合に含まれるカメラに対してスコアを与え、合計スコアにより所在を推定する。

3.4 所在推定の手順

提案する所在推定の手順を図 2 に示す。各カメラ $c_i \in \mathcal{C}$ でのカメラごとの顔検出数の比率に着目した所在推定手法のスコア $s_1(c_i)$ 、移動制約を用いた所在推定手法のスコア $s_2(c_i)$ をそれぞれで算出する。最終的にそれぞれのスコアを足しあわせてカメラのスコア $s(c_i)$ とする。閾値を定め、スコアが閾値以上のカメラの集合を特定人物 p_λ の所在推定結果 \mathcal{C}_{p_λ} とする。

3.4.1 カメラごとの顔検出数の比率に着目した所在推定でのスコア決定

全人物の検出数 $|\mathcal{D}|$ に対する各カメラごとの検出数 $|\mathcal{D}_{c_i}|$ の比率 $r(c_i)$ を式 (4) で表す。

$$r(c_i) = \frac{|\mathcal{D}_{c_i}|}{\sum_{j=1}^{N_c} |\mathcal{D}_{c_j}|} = \frac{|\mathcal{D}_{c_i}|}{|\mathcal{D}|} \quad (4)$$

\mathcal{D}_{p_λ} の誤認識数に対する各 $|\mathcal{D}_{p_\lambda c_i}^-|$ の比率は各 $r(c_i)$ と一致する。よって $|\mathcal{D}_{p_\lambda c_i}|$ を $r(c_i)$ で割ったとき、 c_i で特定人物 p_λ が観測されないならば一定値になり、日常的

入力:
 全人物の検出結果の集合 \mathcal{D}
 特定人物の判定結果の集合 \mathcal{D}_{p_λ}
 出力:
 特定人物の所在推定結果 \mathcal{C}_{p_λ}

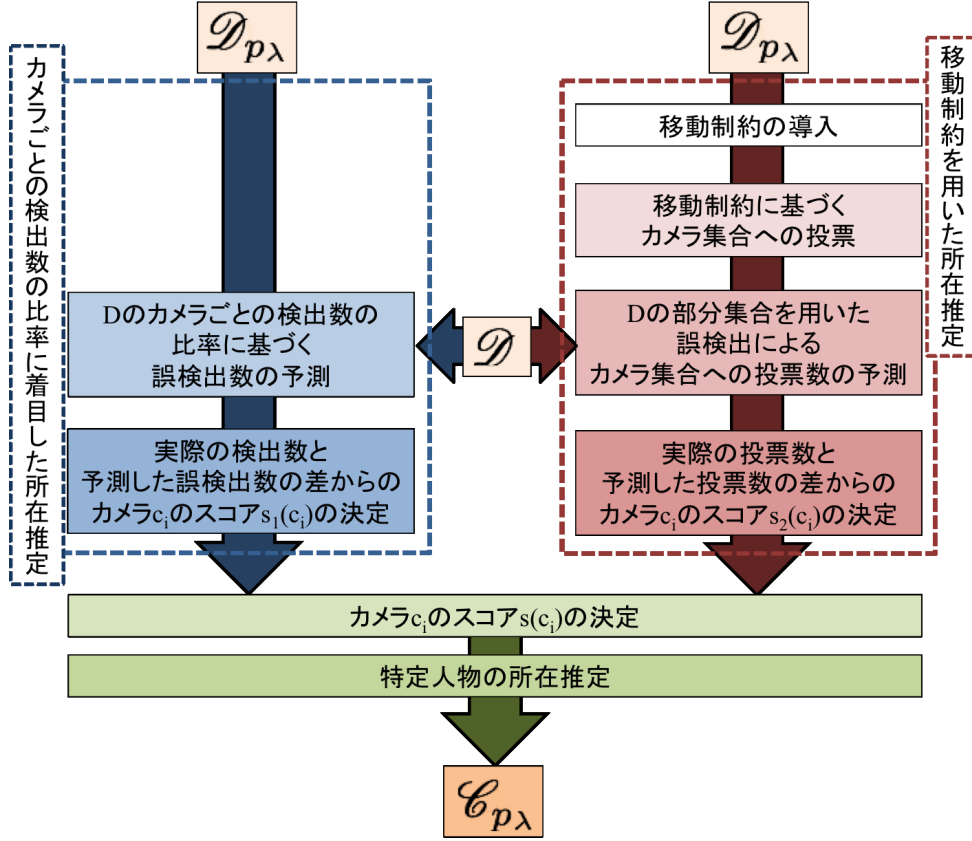


図 2: 提案する所在推定手法の手順

に観測されるならばその一定値よりも高い値になる. 全カメラ台数に対して特定人物 p_λ が観測されるカメラは一部なので, この一定値を $\frac{|\mathcal{D}_{pc_1}|}{r(c_1)}, \dots, \frac{|\mathcal{D}_{pc_{N_c}}|}{r(c_{N_c})}$ の中央値によって求める. このとき, $|\mathcal{D}_{p_\lambda c_i}^-|$ の予測値 $|\widehat{\mathcal{D}_{p_\lambda c_i}^-}|$ は, 式 (5) によって求めることができる. ただし $median \left\{ \frac{|\mathcal{D}_{pc_1}|}{r(c_1)}, \dots, \frac{|\mathcal{D}_{pc_{N_c}}|}{r(c_{N_c})} \right\}$ は $\frac{|\mathcal{D}_{pc_1}|}{r(c_1)}, \dots, \frac{|\mathcal{D}_{pc_{N_c}}|}{r(c_{N_c})}$ の中央値である.

$$|\widehat{\mathcal{D}_{p_\lambda c_i}^-}| = r(c_i) \times median \left\{ \frac{|\mathcal{D}_{pc_1}|}{r(c_1)}, \dots, \frac{|\mathcal{D}_{pc_{N_c}}|}{r(c_{N_c})} \right\} \quad (5)$$

よって各カメラ c_i で特定人物 p_λ が日常的に観測されるかどうかを判定するためのスコア $s_1(c_i)$ は式 (6) のように $|\mathcal{D}_{p_\lambda c_i}|$ と $|\widehat{\mathcal{D}}_{p_\lambda c_i}^-|$ の差で定める.

$$s_1(c_i) = \begin{cases} |\mathcal{D}_{p_\lambda c_i}| - |\widehat{\mathcal{D}}_{p_\lambda c_i}^-| & (|\mathcal{D}_{p_\lambda c_i}| > |\widehat{\mathcal{D}}_{p_\lambda c_i}^-|) \\ 0 & (|\mathcal{D}_{p_\lambda c_i}| \leq |\widehat{\mathcal{D}}_{p_\lambda c_i}^-|) \end{cases} \quad (6)$$

3.4.2 移動制約を用いた所在推定でのスコア決定

特定人物の認識結果に対する移動制約の導入

Shu ら [5] の手法をもとに, 任意の 2 つの認識結果 $d_{p_\lambda i}$ と $d_{p_\lambda j}$ の観測日時の時間間隔に対して, カメラ間の距離と人物の歩行速度から観測日時の時間間隔に下限と上限の制約を定める. ここで時間間隔の下限はあるカメラで観測されてからすぐに別のカメラで観測されないという制約, 上限はあるカメラで観測されてから遠くのカメラで観測されるまでにそのカメラよりも近い他のカメラで必ず観測されるという制約を表す. この下限と上限の制約をまとめて移動制約とする.

下限 (Speed limit)

顔認識結果 $d_{p_\lambda i}$ と $d_{p_\lambda j}$ において, カメラ $c_{d_{p_\lambda i}}$ からカメラ $c_{d_{p_\lambda j}}$ に移動するための最短距離を $distance(c_{d_{p_\lambda i}}, c_{d_{p_\lambda j}})$, 人の歩行速度の最大値を v_{max} とする. このとき, カメラ $c_{d_{p_\lambda i}}, c_{d_{p_\lambda j}}$ 間を移動するための最小時間 $\tau_{min}(c_{d_{p_\lambda i}}, c_{d_{p_\lambda j}})$ を式 (7) で定義する.

$$\tau_{min}(c_{d_{p_\lambda i}}, c_{d_{p_\lambda j}}) = \frac{distance(c_{d_{p_\lambda i}}, c_{d_{p_\lambda j}})}{v_{max}} \quad (7)$$

上限 (Temporal gap)

他のカメラで一切観測されずに突然遠くのカメラで観測されることを防ぐために, 2 つの認識結果の観測日時の時間間隔の上限 τ_{max} を定める. 所在推定対象地域の面積を A^2 , 人の歩行速度の最小値を v_{min} とする. 本研究ではカメラの設置されている密度をもとに式 (8) で定義する.

$$\tau_{max} = \sqrt{\frac{A^2}{N_c}} \cdot \frac{1}{v_{min}} \quad (8)$$

\mathcal{D}_{p_λ} の部分集合の一つを $\mathcal{D}_{seq_l} = \{d_{p_\lambda k_1}, \dots, d_{p_\lambda k_N}\}$ とする. ただし $N \geq 2, i < j \Rightarrow t_{d_{p_\lambda k_i}} \leq t_{d_{p_\lambda k_j}}$ とする. このときすべての $i = 1, \dots, N - 1$ において, 式 (9) を満たすならば, \mathcal{D}_{seq_l} は移動制約を満たすとする.

$$\tau_{min}(c_{d_{p_\lambda k_i}}, c_{d_{p_\lambda k_{i+1}}}) < t_{d_{p_\lambda k_{i+1}}} - t_{d_{p_\lambda k_i}} < \tau_{max} \quad (9)$$

移動制約に基づくカメラ集合への投票 \mathcal{D}_{p_λ} の要素数 2 以上のすべての部分集合に対して, それぞれ式 (9) により移動制約を満たすかどうか判定する. 移動制約を満たす \mathcal{D}_{p_λ} の各部分集合 \mathcal{D}_{seq_l} に対してカメラ集合 \mathcal{C}_{seq_l} を式 (3) の写像 ϕ で求める. $\mathcal{C}_{N_\mathcal{C}}$ の中で \mathcal{C}_{seq_l} と等しいものに対して一票ずつ投票する.

全人物の顔検出結果からの各カメラ集合への投票数の予測

\mathcal{D}_{p_λ} の誤認識は多くの異なる人物から得られる. よって \mathcal{D}_{p_λ} の誤認識の集合から得られるカメラ集合は, \mathcal{D} からランダムサンプリングによって得た顔検出結果の集合から得られるカメラ集合でシミュレートできる. \mathcal{D} からランダムに R 個の要素数 k の部分集合 $\mathcal{D}_k^{(1)}, \dots, \mathcal{D}_k^{(R)}$ を得て, これらの部分集合を用いて \mathcal{D}_{p_λ} の誤検出の組み合わせにより得られる各カメラ集合 $\mathcal{C}_1, \dots, \mathcal{C}_{N_\mathcal{C}}$ への投票数の比率を予測する. $j = 1, \dots, R$ において各 $\mathcal{D}_k^{(j)}$ を入力とし, \mathcal{D}_{p_λ} に対して行ったのと同じ方法で各カメラ集合へ投票する. $\mathcal{D}_k^{(j)}$ から得られるあるカメラ集合 \mathcal{C}_i への投票数を $m(\mathcal{C}_i, \mathcal{D}_k^{(j)})$ とする. $\mathcal{D}_k^{(j)}$ から \mathcal{C}_i が得られないとき $m(\mathcal{C}_i, \mathcal{D}_k^{(j)}) = 0$ である. カメラ集合 \mathcal{C}_i への投票数の比率 $r(\mathcal{C}_i)$ を式 (10) で表す.

$$r(\mathcal{C}_i) = \frac{\sum_{j=1}^R m(\mathcal{C}_i, \mathcal{D}_k^{(j)})}{\sum_{\mathcal{C}_l \in \{\mathcal{C}_1, \dots, \mathcal{C}_{N_\mathcal{C}}\}} \sum_{j=1}^R m(\mathcal{C}_l, \mathcal{D}_k^{(j)})} \quad (10)$$

3.4.1 節と同様に, この比率から予測される投票数よりも多く投票されたカメラ集合は, 特定人物 p_λ の所在に含まれると考えられる. \mathcal{C}_i への投票の予測数 $\hat{m}_{p_\lambda}(\mathcal{C}_i)$ は式 (11) で求める. ただし $mednz \left\{ \frac{m_{p_\lambda}(\mathcal{C}_1)}{r(\mathcal{C}_1)}, \dots, \frac{m_{p_\lambda}(\mathcal{C}_{N_\mathcal{C}})}{r(\mathcal{C}_{N_\mathcal{C}})} \right\}$ は $\frac{m_{p_\lambda}(\mathcal{C}_1)}{r(\mathcal{C}_1)}, \dots, \frac{m_{p_\lambda}(\mathcal{C}_{N_\mathcal{C}})}{r(\mathcal{C}_{N_\mathcal{C}})}$ の 0 でない値の中央値である.

$$\hat{m}_{p_\lambda}(\mathcal{C}_i) = r(\mathcal{C}_i) \times mednz \left\{ \frac{m_{p_\lambda}(\mathcal{C}_1)}{r(\mathcal{C}_1)}, \dots, \frac{m_{p_\lambda}(\mathcal{C}_{N_\mathcal{C}})}{r(\mathcal{C}_{N_\mathcal{C}})} \right\} \quad (11)$$

カメラごとのスコアの決定

式 (6) と同様に, カメラ集合 \mathcal{C}_i に対してスコア $S(\mathcal{C}_i)$ を式 (12) のように実際の投票数 $m_{p_\lambda}(\mathcal{C}_i)$ と予測数 $\hat{m}_{p_\lambda}(\mathcal{C}_i)$ の差で定める.

$$S(\mathcal{C}_i) = \begin{cases} m_{p_\lambda}(\mathcal{C}_i) - \hat{m}_{p_\lambda}(\mathcal{C}_i) & (m_{p_\lambda}(\mathcal{C}_i) > \hat{m}_{p_\lambda}(\mathcal{C}_i)) \\ 0 & (m_{p_\lambda}(\mathcal{C}_i) \leq \hat{m}_{p_\lambda}(\mathcal{C}_i)) \end{cases} \quad (12)$$

各カメラ c_i で特定人物 p_λ が日常的に観測されるかどうかを判定するためのカメラ

ラ間の関係性を考慮したスコア $s_2(c_i)$ は式 (13),(14) のように c_i を含む全てのカメラ集合のスコアの和で定める.

$$w(c_i, \mathcal{C}_j) = \begin{cases} 1 & (c_i \in \mathcal{C}_j) \\ 0 & (c_i \notin \mathcal{C}_j) \end{cases} \quad (13)$$

$$s_2(c_i) = \sum_{j=1}^{N_{\mathcal{C}}} w(c_i, \mathcal{C}_j) S(\mathcal{C}_j) \quad (14)$$

3.4.3 特定人物の所在の推定

式 (6),(14) よりあるカメラ c_i で特定人物 p_λ が日常的に観測されるかどうかを判定するためのスコアを式 (15) で定める.

$$s(c_i) = \alpha s_1(c_i) + (1 - \alpha) s_2(c_i) \quad (15)$$

ここで $0 \leq \alpha \leq 1$ は, $s_1(c_i)$ と $s_2(c_i)$ のどちらを重視するかを決めるパラメータである. 閾値 th を定め, このスコアを用いて特定人物 p_λ の所在推定結果 \mathcal{C}_{p_λ} を式 (16) で求める.

$$\mathcal{C}_{p_\lambda} = \{c_i | s(c_i) > th\} \quad (16)$$

第4章 人物の行動と特定人物の顔認識結果のシミュレーション

4.1 シミュレータの概要

ある地域にカメラを複数台設置し, 不特定多数の人物の大量の顔画像データを収集することは困難である. よって所在推定の精度を検証するために, 所在推定の対象となる地域, 所在推定の対象となる地域にいる人物の日常的な行動, 顔検出結果, 特定人物の顔認識結果をシミュレートする. 顔画像を用意することはできないので, 顔検出結果は d_i の代わりに $\tilde{d}_i = (\theta_{\tilde{d}_i}, t_{\tilde{d}_i}, c_{\tilde{d}_i})$ を用いる. $\theta_{\tilde{d}_i}$ は観測された人物のラベル, $t_{\tilde{d}_i}$ は観測日時, $c_{\tilde{d}_i}$ は観測されたカメラを表す. これを用いてシミュレートした全人物の顔検出結果の集合 $\tilde{\mathcal{D}} = \{\tilde{\mathcal{D}}_i\}_{i=1}^{|\tilde{\mathcal{D}}|}$ を生成する.

人物によって FRR や FAR は異なる. また, 特定人物はどの他人とも同じように誤認識されるわけではなく, 誤認識の起こりやすい人物と誤認識の起こりにくい人物がいる. 誤認識の起こりやすい人物によって, 誤認識が移動制約を満たす場合が多くなるので, 所在推定に影響する. したがって, 人物ごとの FRR と

FAR の違いはなるべく現実に即してシミュレートする必要がある。

所在推定の対象となる地域の人物数を P とし, 各人物を p_1, \dots, p_P とする. 所在推定の性能を確かめるため, p_1, \dots, p_P の中から P_s 人を選び, $p_{\lambda_1}, \dots, p_{\lambda_{P_s}}$ とする. 選んだ人物に対して, それぞれ異なる FRR, FAR で顔認識結果 $\tilde{\mathcal{D}}_{p_{\lambda_1}}, \dots, \tilde{\mathcal{D}}_{p_{\lambda_{P_s}}} \subset \tilde{\mathcal{D}}$ を生成する. そのために人物 p_j が特定人物 p_{λ_i} であると認識される確率を顔認識率 $R_{p_{\lambda_i}}(p_j)$ として定める. 様々な人物の顔認識の性能をモデル化するために, 実際に多数の顔画像を用いて顔認識による特定人物の判定を行い, 得られた結果から各特定人物 p_{λ_i} に対して $R_{p_{\lambda_i}}$ を定めた. 詳細は 4.3 節で述べる.

4.2 シミュレータの処理の流れ

シミュレータの処理の流れを図 3 に示す. シミュレーションにおいて与えるパラメータは所在推定の対象となる地域の大きさ, カメラ台数, 観測される人物数, 観測日数, 顔検出率, 特定人物のラベルである. まず所在推定の対象となる地域は 2 次元平面で表し, その平面上にカメラを配置する. そして人物の移動をモデル化するために, 配置されたカメラを頂点として道路網を表すグラフを作成する. 人物の移動は頂点から頂点への辺を通過しての移動としてモデル化する. グラフのそれぞれの辺の重みをカメラ間の距離とし, 移動時間を歩行速度から求めて人物が各カメラで観測される時刻を定める. カメラで観測された人物が顔検出によって検出されるかどうかを顔検出率に従って決定し, 全人物の検出結果 $\tilde{\mathcal{D}}$ を得る. 最後に, 顔認識の結果を顔認識率 $R_{p_{\lambda_i}}$ によってシミュレートし特定人物 p_{λ_i} の認識結果 $\tilde{\mathcal{D}}_{p_{\lambda_i}}$ を出力する.

所在推定の対象となる地域のカメラ配置

[パラメータ]

- 所在推定の対象となる地域の大きさ A
- カメラ台数 N_c

[処理]

所在推定の対象となる地域として A_{km} 四方の 2 次元空間を生成する. この所在推定対象地域内にカメラを N_c 台ランダムに配置する. 人物が移動する道路を想定し, 各カメラ位置を頂点として, いくつかの頂点を辺で繋ぐことによってグラフ $G = (\mathcal{V}, \mathcal{E})$ を作成する. \mathcal{V} は頂点集合, \mathcal{E} は辺集合である. ただしグラフは連結グラフかつ平面グラフである. 人物は隣接する頂点に対応するカメラ位置間のみ, 他のカメラで観測されずに移動できる. 辺の重みは頂点に対応するカメ

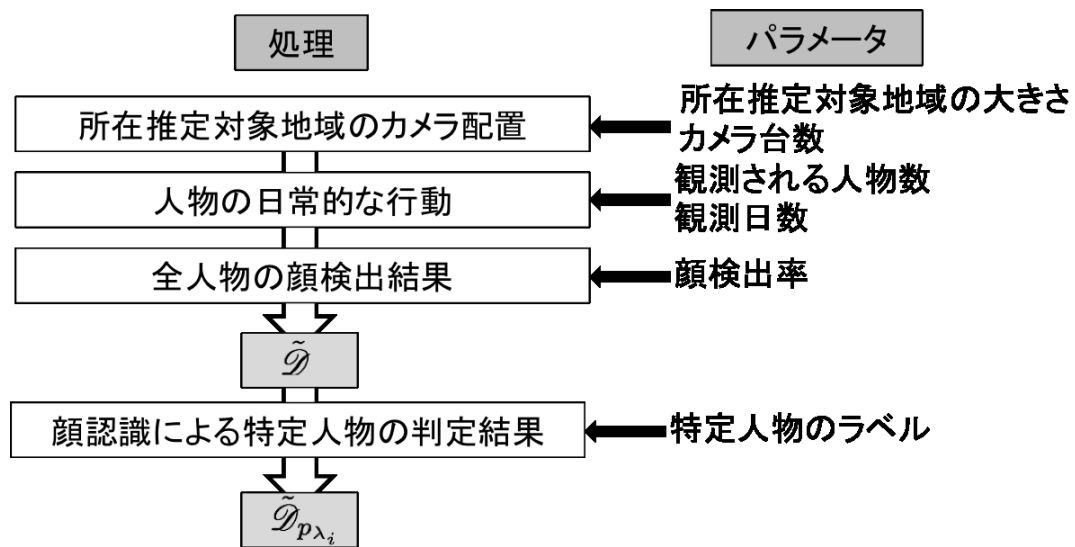


図 3: シミュレータの処理の流れ

ラ位置間の距離とする。隣接する頂点に対応するカメラ位置間の距離はそれぞれの座標に対するユークリッド距離で定める。隣接しない頂点に対応する 2 カメラ間の距離は、その 2 カメラに対応する 2 頂点を結ぶグラフ上の最短経路を求め経路上の辺の重みの和によって定める。

人物の日常的な行動

[パラメータ]

- 人物数 P
- 観測日数 N_d

[処理]

各人物の一日の行動を次の手順で生成する。まず、人物に対して自宅と目的地を想定しグラフ G から頂点を 2 つ選択する。人物は自宅から目的地に向かってこの 2 頂点を結ぶグラフ上の最短経路を移動する。最初の頂点で観測される時間はランダムである。そして人物が頂点から頂点に移動するたびに、移動速度をランダムに定め、頂点に対応するカメラで順次観測されていく。本研究では最小移動速度 $v_{min} = 2km/h$ 、最大移動速度 $v_{max} = 8km/h$ として、この範囲で移動速度をランダムに与えた。目的地に達したあと、ランダムな時間滞在してから今度は自宅に向かって同じように移動する。

人物は毎日行動するわけではなく N_d 日間の中から行動する日をランダムに選択する。ただし行動する日数の最小値は観測日数の半分とした。このように

して人物の日常的な行動を P 人分生成する.

全人物の顔検出結果

[パラメータ]

- 顔検出率 O

[処理]

生成された人物の日常的な行動から, 人物が観測されるカメラ, 日時が決まる. ただし, 顔の向きや障害物による隠れなどにより頂点を通過した人物が必ずしも顔検出されるわけではない. よって人物がカメラで観測されるごとに顔検出されるかどうかを顔検出率 O に従って定め, 全人物の顔検出結果 $\tilde{\mathcal{D}}$ を生成する.

顔認識による特定人物の判定結果

[パラメータ]

- 特定人物のラベル $\theta_{p_{\lambda_i}}$

[処理]

生成した $\tilde{\mathcal{D}}$ から, 特定人物 p_{λ_i} の顔認識結果 $\tilde{\mathcal{D}}_{p_{\lambda_i}}$ を生成する. $\tilde{\mathcal{D}}_{p_{\lambda_i}} \subset \tilde{\mathcal{D}}$ なので, 各 $d_j \in \tilde{\mathcal{D}}$ に対して顔認識率 $R_{p_{\lambda_i}}(d_j)$ に従い, $\tilde{\mathcal{D}}_i$ が $\tilde{\mathcal{D}}_{p_{\lambda_i}}$ に含むかを定める.

4.3 実際の顔認識結果に基づく顔認識の性能のモデル化

本節では実際に多数の顔画像を用いて特定人物 p_{λ_i} の顔認識を行い, 顔認識率 $R_{p_{\lambda_i}}$ を定める方法について述べる.

4.3.1 顔認識による特定人物の判定方法

顔認識には OKAOVision[6] を用いた. OKAOVision によって各顔画像から特徴ベクトルが得られる. 2つの特徴ベクトルの比較にはユークリッド距離を用いた. 特徴ベクトルの距離が小さいほど, 顔画像の類似度が大きいことを表す.

顔画像は一人あたりおよそ 500 枚 ~1500 枚の画像を 627 人分を用意した. 一人あたりの顔画像枚数を揃えるために, この中から 520 枚以上の顔画像を含む 449 人を選択した. そして 449 人からそれぞれ 500 枚ずつ顔画像を選択した. 人物 i の j 番目の顔画像を I_{ij} とし, 顔画像集合 $\Gamma_G = \{\{I_{ij}\}_{j=1}^{500}\}_{i=1}^{449}$ を定める. この Γ_G に含まれる特定人物 p_{λ_i} の顔画像を顔認識により判定した.

特定人物 p_{λ_i} を 449 人の中から 1 人選び, 特定人物 p_{λ_i} の顔画像を Γ_G に含まれているものとは別に 20 枚選択した. それぞれの顔画像を $I_{Q_1}, \dots, I_{Q_{20}}$ としこの顔画像の集合を $\Gamma_Q = \{I_{Q_k}\}_{k=1}^{20}$ とする. この Γ_Q を用いて, Γ_G の各顔画像が特定人物 p_{λ_i} であるかどうか判定した. 顔画像 $I_{ij} \in \Gamma_G$ に対して Γ_Q の全顔画像の

特徴ベクトルとの距離を計算し、得られた 20 個の距離の中で最小値を顔画像 I_{ij} と顔画像の集合 Γ_Q との距離とした。

OKAOVision で得られる特徴ベクトルは、特定人物 p_{λ_i} との距離が他の人物の顔画像との距離と比べて近いことを利用して顔認識を行う。よって特定人物 p_{λ_i} の顔画像との距離のみを用いるだけでは判定は難しい。そこで、 Γ_G に含まれない 150 人から各 20 枚ずつの顔画像を特定人物の判定用に用いた。これを $\Gamma_{Q'} = \{\{I_{i'j'}\}_{j'=1}^{20}\}_{i'=1}^{150}$ とする。顔画像 $I_{ij} \in \Gamma_G$ と、 $\Gamma_{Q'}$ の全顔画像の特徴ベクトルとの距離をそれぞれ計算し、最小値を求める。これを顔画像 I_{ij} と顔画像の集合 $\Gamma_{Q'}$ との距離とした。顔画像 I_{ij} において $\Gamma_{Q'}$ との距離が Γ_Q との距離よりも小さければ、特定人物 p_{λ_i} の顔画像よりも特徴ベクトルの距離が近い顔画像があるとして顔画像 I_{ij} は特定人物 p_{λ_i} の顔画像ではないと判定した。顔画像 I_{ij} において $\Gamma_{Q'}$ との距離よりも Γ_Q との距離が小さいとき、顔画像 I_{ij} の顔画像と特定人物 p_{λ_i} との距離が閾値以下であれば、その顔画像を特定人物 p_{λ_i} の顔画像であると判定した。

Γ_G の顔画像のうち、特定人物 p_{λ_i} と判定された顔画像の枚数を $N_{I,p_{\lambda_i}}$ 、この中で特定人物 p_{λ_i} の正しい顔画像の枚数を $N_{I,p_{\lambda_i}}^{(t)}$ 、誤認識の枚数を $N_{I,p_{\lambda_i}}^{(f)}$ とする。このとき特定人物 p_{λ_i} の本人拒否率 $FRR_{p_{\lambda_i}}$ 、他人受入率 $FAR_{p_{\lambda_i}}$ はそれぞれ式 (17)、(18) で表される。

$$FRR_{p_{\lambda_i}} = 1 - \frac{N_{I,p_{\lambda_i}}^{(t)}}{500} \quad (17)$$

$$FAR_{p_{\lambda_i}} = \frac{N_{I,p_{\lambda_i}}^{(f)}}{(449 - 1) \times 500} \quad (18)$$

Γ_G に含まれる 449 人に対して上記の方法で顔認識を行った。このとき誤認識が全く起こらなかった人物が 7 人いた。顔認識の性能をモデル化するために、この 7 人を除く 442 人の結果を用いた。この結果は、 442×449 の混同行列 M で表すことができる。 M の i 行 j 列の値を $M(i, j)$ とする。特定人物 p_{λ_i} に対して人物 p_j と誤認識が起こった枚数を $N_{I,p_{\lambda_i},p_j}^{(f)}$ とすると、 $M(\lambda_i, j) = \frac{N_{I,p_{\lambda_i},p_j}^{(f)}}{500}$ である。また、 $M(\lambda_i, \lambda_i) = FRR_{p_{\lambda_i}}$ である。

4.3.2 顔認識の性能のモデル化

人物数 P が変化した時、顔認識率を表す $P_s \times P$ の混同行列が得られれば、 P_s 人の顔認識結果をシミュレートすることができる。しかし、この混同行列を得ることは難しい。

人物ごとの FRR, FAR の違いをシミュレートするため、まず実際に顔認識を行った結果において特定人物 p_{λ_i} に対して他の人物の認識率の順位を認識率が高い順に定める。すなわち特定人物 p_{λ_i} に対する人物 p_j の順位を $x_{p_{\lambda_i}}(p_j)$ とすると、 $x_{p_{\lambda_i}}(p_{j_1}) < x_{p_{\lambda_i}}(p_{j_2})$ のとき $M(\lambda_i, j_1) \geq M(\lambda_i, j_2)$ である。この順位と認識率の関係を 442 人分示したものが図 4 である。横軸が順位、縦軸が認識率を表し、一人の特定人物に対して一つの折れ線グラフを描いている。人物ごとに FRR の値は異なるので、各人物の顔認識の結果得られた折れ線グラフをその人物の FRR ごとに色分けしている。本研究では人物数が増えても全人物数に対する順位と認識率の関係は変わらないと仮定する。よって図 4 の各折れ線グラフに対して最小二乗フィッティングによる回帰によって顔認識率を表す関数 $g_{p_{\lambda_i}}$ を推定し、 $R_{p_{\lambda_i}}(p_j) = g_{p_{\lambda_i}}(\frac{x_{p_{\lambda_i}}(p_j)}{P-1})$ とすることで P の変化に対応する。 $\mathcal{D}_{p_{\lambda_i}}$ を生成する際に $P-1$ 人中の人物 p_j の顔認識率の順位をランダムに定め、このモデル化によって特定人物 p_{λ_i} に対する人物 p_j の顔認識率 $R_{p_{\lambda_i}}(p_j)$ を定める。

回帰を行うための $g_{p_{\lambda_i}}$ のモデル化について述べる。図 4 は点の打ち方からそれぞれ単調減少のグラフであり、それぞれの形状からモデルとして式 (19) を設計した。

$$g'_{p_{\lambda_i}}(\frac{x_{p_{\lambda_i}}}{P-1}; a, b) = b \cdot e^{-a \frac{x}{P-1}} \quad (19)$$

回帰によって式 (19) の a, b を推定した結果を、図 4 の折れ線グラフと比較すると、認識率の順位が低い部分の誤差は十分小さかったが、認識率の順位が高い部分の誤差は大きかった。そこでモデルとして式 (20), (21) を設計した。ここで特定人物 p_{λ_i} に対して一枚でも誤認識が起こった人数を P_f とする。

$$G_{p_{\lambda_i}}(P, P_f, FAR_{p_{\lambda_i}}, a, b) = (P-1) \times (FAR_{p_{\lambda_i}} - \int_1^{P_f} g'_{p_{\lambda_i}}(\frac{x_{p_{\lambda_i}}}{P-1}; a, b) dx_{p_{\lambda_i}}) \quad (20)$$

$$g_{p_{\lambda_i}}(\frac{x_{p_{\lambda_i}}}{P-1}; P_f, FAR_{p_{\lambda_i}}, a, b, c) = \begin{cases} g'_{p_{\lambda_i}}(\frac{x_{p_{\lambda_i}}}{P-1}; a, b) + \frac{c \cdot G_{p_{\lambda_i}}(P, P_f, FAR_{p_{\lambda_i}}, a, b)}{(1 - e^{-c P_f})} e^{-c \frac{x_{p_{\lambda_i}}}{P-1}} & (G_{p_{\lambda_i}}(P, P_f, FAR_{p_{\lambda_i}}, a, b) > 0) \\ g'_{p_{\lambda_i}}(\frac{x_{p_{\lambda_i}}}{P-1}; a, b) & (G_{p_{\lambda_i}}(P, P_f, FAR_{p_{\lambda_i}}, a, b) \leq 0) \end{cases} \quad (21)$$

式 (19) を使って a, b を推定したあと、回帰によって式 (21) の c を推定した。(21) の上の式の第 2 項により、式 (19) では誤差が大きい認識率の順位が高い部分を表している。

式 (20) は次のように定めた. まず, 式 (18) より式 (22) が成り立つ. ただし $x_{p_{\lambda_i}}^{-1}(j)$ は特定人物 p_{λ_i} に対する顔認識の順位が j である人物を表す.

$$(P - 1) \times FAR_{p_{\lambda_i}} = \frac{N_{I, p_{\lambda_i}}^{(f)}}{(P - 1)500} = \sum_{j=1}^{P_f} \frac{N_{I, p_{\lambda_i}, x_{p_{\lambda_i}}^{-1}(j)}^{(f)}}{500} \quad (22)$$

式 (22) の右辺は特定人物 p_{λ_i} の顔認識結果により得られた折れ線グラフと両軸で囲まれる領域の面積を表す. したがって関数 $g_{p_{\lambda_i}}(x_{p_{\lambda_i}}, P)$ は式 (23) を満たすべきである.

$$\int_1^{P_f} g_{p_{\lambda_i}}\left(\frac{x_{p_{\lambda_i}}}{P-1}\right) dx_{p_{\lambda_i}} = (P - 1) \times FAR_{p_{\lambda_i}} \quad (23)$$

よって式 (20) は, 式 (19) を積分したとき式 (23) の条件を満たすのに足りない値を表している.

442 人分の $g_{p_{\lambda_i}}\left(\frac{x_{p_{\lambda_i}}}{P-1}\right)$ の推定結果を図 5 に示す. 図 4 と同様に, 人物の FRR ごとに色分けしている. 図 4 から, FRR が大きくなるほど認識率の大きい人物も少なくなるという特徴が見られる. 図 5 から推定した関数はこの特徴を反映していることがわかる.

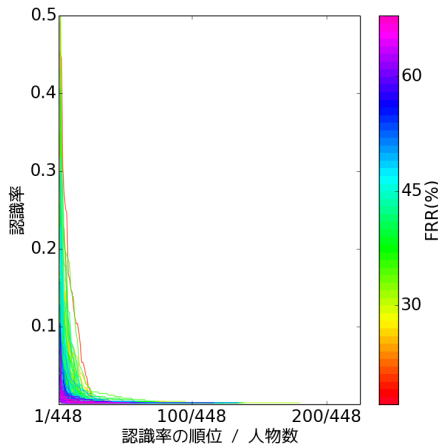


図 4: 人物ごとの顔認識率

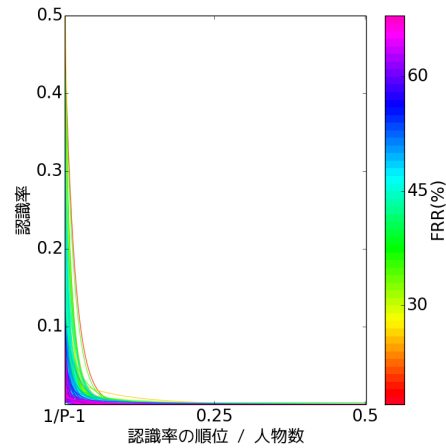


図 5: 顔認識率を表す関数の推定結果

第5章 シミュレーションによる特定人物の所在推定の実験

第3章で述べたカメラごとの顔検出数の比率に着目した所在推定手法, 及び移動制約を用いた所在推定手法の有効性を確かめた.

5.1 方法

第4章で述べたシミュレーションにより実験を行った. パラメータは表1のように設定した. 所在推定の対象となる地域の大きさ, カメラ台数, 顔検出率は固定し, 人物数と観測日数の組を変えて全人物の顔検出結果の集合 $\tilde{\mathcal{D}}$ を3パターン生成した. 特定人物 p_{λ_i} の顔認識結果の集合 $\tilde{\mathcal{D}}_{p_{\lambda_i}}$ を442人分生成し実験を行った.

表1: シミュレーションのパラメータ設定

所在推定対象地域の大きさ	5km		
カメラ台数	100台		
(人物数, 観測日数)	(50000人, 15日)	(50000人, 30日)	(100000人, 30日)
顔検出漏れ率	0.25		

各 $\tilde{\mathcal{D}}, \tilde{\mathcal{D}}_{p_{\lambda_i}}$ に対して, 次の4つの手法で所在を推定し比較した.

- 手法(1): 2.3節で述べたカメラごとの顔認識数に基づく所在推定
- 手法(2): 3.1節で述べたカメラごとの顔検出数の比率に着目した所在推定
- 手法(3): 3.2節で述べた移動制約を用いた所在推定
- 手法(4): 手法(2)と手法(3)のスコア合計による所在推定(提案手法)

手法(1)では $\tilde{\mathcal{D}}$ における各カメラ c_i での顔認識数 $|\tilde{\mathcal{D}}_{c_i}|$ をそのままスコア $s(c_i)$ とした. 手法(2),(3),(4)では, 式(15)においてそれぞれ α を $\alpha = 1, \alpha = 0, \alpha = 0.5$ とした. 式(16)の特定人物の所在推定結果 $\mathcal{C}_{p_{\lambda_i}}$ を求める際の閾値 th はそれぞれカメラごとのスコアの最大値の半分とした.

移動制約を用いた所在推定手法において, $\tilde{\mathcal{D}}_{p_{\lambda_i}}$ の全ての部分集合に対して移動制約を満たすかどうかの判定を行うと計算量が膨大になる. そこで, 要素数7以下の $\tilde{\mathcal{D}}_{p_{\lambda_i}}$ の部分集合のみを用いた. また, FARがある値より大きい人物は誤認識数が多すぎて提案手法を適用することができなかった. よって $|\tilde{\mathcal{D}}_{p_{\lambda_i}}|$ が一

日あたり 700 以下の人物のみ所在推定を行い評価した。

各人物の顔認識の性能の分布を図 6 に示す。図 6 の 1 点が 1 人の特定人物に対応し、その座標が特定人物の顔認識における FRR と FAR を表している。FRR は 20% から 70% に分布していた。また人物数 50000 人の場合、FAR は 0% から 0.2% の範囲に、人物数 100000 人の場合、FAR は 0% から 0.1% の範囲に分布していた。3.3 節で述べたように、FRR と FAR によってそれぞれの手法の有効性が異なると考えられる。よって、FRR と FAR の違いによって所在推定の精度を評価するために、特定人物を FRR と FAR の値によって 4 つにわけると。FRR の閾値を 45%、FAR の閾値を人物数 50000 人の場合 0.1%、人物数 100000 人の場合 0.05% と定め、(FRR 小, FAR 小), (FRR 大, FAR 小), (FRR 小, FAR 大), (FRR 大, FAR 大) の範囲にわけ、この範囲ごとにも評価を行う。

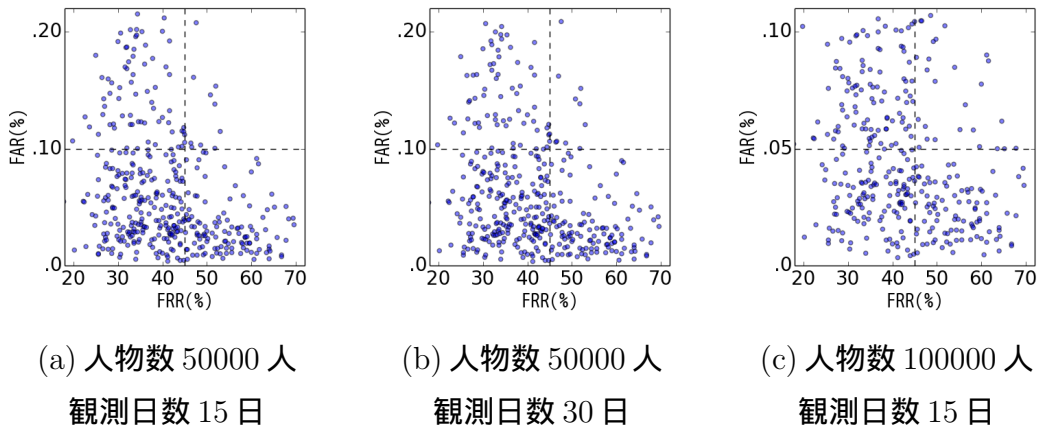


図 6: 顔認識の性能の分布

特定人物 p_{λ_i} の真の所在を $\mathcal{C}_{p_{\lambda_i}}^{(t)}$ とする。このとき特定人物 p_{λ_i} に対する適合率 $precision_{p_{\lambda_i}}$ 、再現率 $recall_{p_{\lambda_i}}$ 、F 値 $Fmeasure_{p_{\lambda_i}}$ はそれぞれ式 (24), (25), (26) で求められる。

$$precision_{p_{\lambda_i}} = \frac{|\mathcal{C}_{p_{\lambda_i}} \cap \mathcal{C}_{p_{\lambda_i}}^{(t)}|}{|\mathcal{C}_{p_{\lambda_i}}|} \quad (24)$$

$$recall_{p_{\lambda_i}} = \frac{|\mathcal{C}_{p_{\lambda_i}} \cap \mathcal{C}_{p_{\lambda_i}}^{(t)}|}{|\mathcal{C}_{p_{\lambda_i}}^{(t)}|} \quad (25)$$

$$Fmeasure_{p_{\lambda_i}} = \frac{2 \cdot precision_{p_{\lambda_i}} \cdot recall_{p_{\lambda_i}}}{precision_{p_{\lambda_i}} + recall_{p_{\lambda_i}}} \quad (26)$$

以上の定義に基づいて各人物のそれぞれの手法における F 値を求め, 手法ごとの平均値を比較し評価した.

5.2 結果

表 2 から表 4 にそれぞれ (人物数 50000 人, 観測日数 15 日), (人物数 50000 人, 観測日数 30 日), (人物数 100000 人, 観測日数 15 日) の F 値の平均の結果を示す. データ数とは 442 人中, FAR が小さく提案手法を適用できた人物数である. いずれの場合でも顔認識の性能が低い (FRR 大, FAR 大) の範囲ではデータ数が少ないので以後議論しない.

表 2 より, (人物数 50000 人, 観測日数 15 日) の場合は手法 (4) の提案手法が最も高い F 値を示した. 手法 (2) と手法 (3) を比べると, (FRR 大, FAR 小) の範囲では手法 (3) と比べて手法 (2) が高く, (FRR 小, FAR 大) の範囲では手法 (2) と比べて手法 (3) が高い. (FRR 小, FAR 大) の範囲で移動制約が有効に働いていると考えられる.

表 3 より, (人物数 50000 人, 観測日数 30 日) の場合は手法 (4) よりも手法 (2) の方が高い F 値を示した. 表 2 と比べると, 全体的に F 値が向上した. また, (FRR 小, FAR 大) の範囲でも手法 (2) が手法 (3) に近い値を示した. これは観測日数が増えたことによる効果と考えられる.

表 4 より, (人物数 100000 人, 観測日数 15 日) の場合は手法 (2) がどの範囲でも最も高い F 値を示した. しかし, 全体的な F 値は表 2, 表 3 と比べて低い. 人物数が増えたことにより誤認識が多くなり, 手法 (2) における誤認識数の推定も, 手法 (3) における移動制約も効果がなくなっていると考えられる. また, データ数も少なく所在推定可能な範囲が狭くなっている. 人物数 100000 人において所在推定を行うには, 今回用いた顔認識の性能では不十分であった.

5.3 考察

観測日数の効果

表 2 と表 3 の比較により, 手法 (2) において観測日数の変化により F 値が向上していることがわかる. この効果を考察するために, 各 $\tilde{\mathcal{S}}$ に対する手法 (2) でのスコアを考察する. 手法 (2) でのカメラ c_i のスコア $s(c_i)$ は, 理想的には c_i が特定人物の真の所在 $C_p^{(t)}$ に含まれていれば顔認識された回数は増加し, c_i が $C_p^{(t)}$ に

表 2: 人物数 50000, 観測日数 15 の場合の F 値

範囲	データ数	手法 (1)	手法 (2)	手法 (3)	手法 (4)
全範囲	406	0.309	0.350	0.277	0.359
(FRR 小, FAR 小)	210	0.350	0.385	0.310	0.390
(FRR 大, FAR 小)	116	0.267	0.385	0.223	0.364
(FRR 小, FAR 大)	70	0.259	0.219	0.268	0.270
(FRR 大, FAR 大)	10	0.260	0.125	0.276	0.276

表 3: 人物数 50000, 観測日数 30 の場合の F 値

範囲	データ数	手法 (1)	手法 (2)	手法 (3)	手法 (4)
全範囲	402	0.318	0.450	0.300	0.423
(FRR 小, FAR 小)	211	0.352	0.514	0.328	0.459
(FRR 大, FAR 小)	116	0.292	0.451	0.269	0.447
(FRR 小, FAR 大)	70	0.262	0.275	0.276	0.278
(FRR 大, FAR 大)	10	0.269	0.222	0.251	0.252

表 4: 人物数 100000, 観測日数 15 の場合の F 値

範囲	データ数	手法 (1)	手法 (2)	手法 (3)	手法 (4)
全範囲	338	0.272	0.319	0.250	0.298
(FRR 小, FAR 小)	121	0.279	0.386	0.269	0.349
(FRR 大, FAR 小)	95	0.294	0.321	0.232	0.303
(FRR 小, FAR 大)	97	0.248	0.271	0.251	0.251
(FRR 大, FAR 大)	25	0.243	0.167	0.220	0.219

含まれていなければ観測日数が増えても 0 になる. $\mathcal{C}_p^{(t)}$ に含まれるカメラでの平均スコアを s^t , $\mathcal{C}_p^{(t)}$ に含まれないカメラでの平均スコアを s^f とする. これらを全人物で計算し平均したものを表 5 に示す. 表 5 により, 観測日数が倍になることによって特定人物の正認識数も増え, s^t の値は 7.9 増えている. 一方で s^f の値は 2.1 しか増えていない. s^t と s^f の差は 4.8 から 10.6 になっており, この差が開くことによって顔認識の性能の変動による誤りに強くなり F 値が向上していると考えられる. よって観測日数を増やすことにより所在推定の精度の向上が可

能である。

表 5: 真の所在により区別したカメラごとのスコアの平均

人数	50000	50000
日数	15	30
s^t	8.5	16.4
s^f	3.7	5.8

移動制約の効果

人物数や観測日数の変化によって、顔認識の性能ごとに手法 (2) と手法 (4) の有効性が変わっている。よって移動制約の効果を検討するために、顔認識の性能ごとに移動制約によって除去されている顔認識結果を求める。

特定人物の正認識結果の集合 $\tilde{\mathcal{D}}_p^+$ のうち、他の顔認識結果と全く移動制約を満たさないものの集合を $\tilde{\mathcal{D}}_{travel}^+$ 、誤認識の結果の集合 $\tilde{\mathcal{D}}_p^-$ のうち、他の顔認識結果と全く移動制約を満たさないものの集合を $\tilde{\mathcal{D}}_{travel}^-$ とする。顔認識結果が移動制約によって除去される割合をそれぞれ $r^+ = \frac{|\tilde{\mathcal{D}}_{travel}^+|}{|\tilde{\mathcal{D}}_p^+|}$, $r^- = \frac{|\tilde{\mathcal{D}}_{travel}^-|}{|\tilde{\mathcal{D}}_p^-|}$ とする。顔認識の性能ごとの r^+, r^- の結果を表 6 に示す。

まず人物数 50000 人、観測日数 15 日の場合において顔認識の性能ごとに考察する。FRR が大きい場合、正認識結果の数が減る。このとき正認識結果が移動制約を満たさなくなり r^+ が大きくなる。(FRR 大, FAR 小) の場合、正認識結果は 52% も除去されており移動制約の効果は低い。よって表 2 の (FRR 大, FAR 小) の場合、手法 (3) の F 値は小さく、手法 (2) の方が大きい。また FAR が大きい場合、誤認識数が増える。このとき誤認識結果が偶然移動制約を満たすことが多くなり r^- が小さくなる。(FRR 小, FAR 大) の場合、誤認識結果は 23% しか除去されていない。しかし、FRR が小さいので正認識結果も 19% しか除去されずに残っているため移動制約を満たす。よって表 2 の (FRR 小, FAR 大) の場合、手法 (2) では誤認識の影響で F 値が小さい。一方で手法 (3) では移動制約が有効に働き手法 (2) より F 値が高くなっている。以上より FRR が大きい場合移動制約を満たさなくなり手法 (3) は有効に働かず、FAR が大きい場合誤認識の影響を強く受ける手法 (2) は有効に働かない。よって両者の有効性を補っている手法 (4) が最も F 値が高い。

次に人物数 50000 人で観測日数が 15 日と 30 日の場合を比べると、顔認識結果が除去される割合はほとんど変わっていない。移動制約は観測時刻に近い顔認識結果に対して影響があるので、観測日数が変わっても効果は変わらない。表 2 と表 3 の手法 (3) の F 値を比べても、観測日数が少なくても手法 (2) ほど F 値は低下しないことがわかる。

最後に人物数が 100000 人の場合、誤認識が多くなることによって移動制約を満たすものが多くなり、FAR が小さくても除去できない誤認識が多くなっている。このため、表 4 において手法 (3) の F 値は極めて低く、移動制約の効果は全くない。

以上より移動制約は、十分な観測日数が得られないとき効果を発揮する。映像の保存期間の制限があることや人物の所在が長期観測の間が変わってしまうことが考えられるので、観測日数が少ない時に精度低下を抑えられることは有用である。

表 6: 移動制約によって除去される顔認識結果の割合

人数	50000		50000		100000	
日数	15		30		15	
除去される割合	r^+	r^-	r^+	r^-	r^+	r^-
全範囲	0.38	0.50	0.38	0.50	0.34	0.40
(FRR 小,FAR 小)	0.36	0.53	0.36	0.53	0.36	0.47
(FRR 大,FAR 小)	0.52	0.62	0.53	0.62	0.43	0.52
(FRR 小,FAR 大)	0.19	0.23	0.19	0.23	0.24	0.23
(FRR 大,FAR 大)	0.32	0.26	0.36	0.25	0.26	0.23

精度向上の余地

今回は、特定人物 p の真の所在 $\mathcal{C}_p^{(t)}$ に含まれるカメラのスコアは高くそれ以外のスコアは低いとし、式 (16) の閾値 th を最大スコアの $\frac{1}{2}$ とした。この閾値を評価するため、特定人物に対し最も F 値が高くなるようにスコアが高い順にカメラを選んだときの F 値の平均を表に示す。括弧内は表 2,3,4 での結果である。いずれの場合でも精度向上の余地があり、閾値の決め方について検討が必要である。

表 7: 各手法における最良の閾値を選択した場合の F 値

人数	50000	50000	100000
日数	15	30	15
手法 (2)	0.503(0.350)	0.599(0.450)	0.472(0.319)
手法 (3)	0.425(0.277)	0.451(0.300)	0.392(0.250)
手法 (4)	0.519(0.359)	0.578(0.423)	0.451(0.298)

第 6 章 結論

本研究では、多数のカメラ映像から顔検出、顔認識によって特定人物が日常的に観測されるカメラを推定する所在推定の問題を扱った。顔認識の性能は現在非常に良くなっている。しかし、顔認識の性能がどれだけ良くても、誤認識は起こる。したがって、観測される人物が一日に数千人から数万人いる場合、誤認識が圧倒的に多くなってしまおうという問題が生じる。

本研究ではこの問題に対して、観測される人物が十分多い場合特定人物の顔認識数と全人物の顔検出数のカメラごとの比が一致することを用いて誤認識数を推定し、誤認識の影響を抑えて所在を推定する手法を提案した。さらに移動制約を導入し、誤認識の影響を低減して精度の向上を試みた。

シミュレーションによる実験で、カメラごとの顔検出数の比率に基づく所在推定により、長期間観測することによって所在推定の精度の向上が可能であることを示した。また十分な観測日数が得られない場合でも、移動制約を導入することにより所在推定の精度低下を抑えられることを示した。

今後の課題としては、スコアによる所在の決定時の閾値の検討がある。提案手法によって特定人物の真の所在に含まれるカメラのスコアが高くなっているが、特定人物の所在を正確に求める点においては改善の余地がある。本研究ではスコアの最大値の半分を閾値としたが、各カメラをスコアを用いて所在に含まれるか含まれないかを定める 2 クラス分類問題に当てはめ、特定人物ごとに最適な閾値を検討することが考えられる。また実環境における実験を行うことが挙げられる。本研究ではカメラごとの FRR, FAR の値はほぼ一定とした。しかし実際のカメラ映像を用いる場合、カメラごとの解像度の違いや照明により、カメラごとの FRR, FAR は変動しやすく所在推定は難しくなる。一方で、人物の全身の画像や顔認識における類似度などを用いることでより精度の高い所在推定を行う

ことができる可能性がある。また、顔認識の方法にも所在推定に適したものと考えると考えられる。例えば、実験により FRR が大きいことよりも FAR が大きいことの方が精度の下がる要因となっていることがわかる。よって、FRR を保ったまま FAR を抑える処理を加えることができれば精度の向上は可能であると考えられる。

謝辞

本研究を進めるにあたり、多くのご教示と熱心なご指導を賜りました美濃導彦教授、椋木雅之准教授に深く感謝いたします。また、日頃より多くの助言を頂きました川西康友氏に厚く御礼申し上げます。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深く感謝いたします。

参考文献

- [1] 依田育士, 大西正輝: 大型複合施設における長期間にわたる人流解析と可視化手法, 電子情報通信学会論文誌, Vol. J93-D, No. 4, pp. 486–493 (2010).
- [2] Cranshaw, J., Hong, J. I. and Sadeh, N. M.: The Livelihoods Project : Utilizing Social Media to Understand the Dynamics of a City, *The 6th International AAAI Conference on Weblogs and Social Media*, pp. 58–65 (2012).
- [3] Malek, O., Venetsanopoulos, A., Androutsos, D. and Zhao, L.: Subspace State Estimator for Facial Biometric Verification, *2014 International Conference on Computational Science and Computational Intelligence*, Vol. 1, pp. 137–143 (2014).
- [4] Taigman, Y., Yang, M., Ranzato, M. A. and Wolf, L.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708 (2014).
- [5] Shu, H., Hsu, Y. F., Yu, S. S. and Chen, T.: Wide area video surveillance with spatial-temporal constraints, *19th IEEE International Conference on Image Processing*, pp. 33–36 (2012).
- [6] OKAOVision: <http://www.omron.co.jp/ecb/products/mobile/>.