

Haar-like 特徴を使った顔検出と Mean-Shift トラッカによる 複数視点人物追跡システム

立花 智哉 山下 淳 金子 透

静岡大学工学部機械工学科 〒432-8561 静岡県浜松市城北 3-5-1

E-mail: {f0210084, tayamas, tmtkane}@ipc.shizuoka.ac.jp

あらまし ビジョンセンサを使って環境を自動的に監視する場合、人物を追跡することが重要な課題となる。しかし画像の中から人物のような姿勢変化の大きな対象を特定することは難しい。そこで本研究では Haar-like 特徴を使った顔検出と Mean-Shift トラッカを組み合わせた人物追跡を自動的に行う手法を提案する。また、単一の視点からでは追跡能力に限界がある。そこで複数の視点を利用した人物の追跡手法を提案する。予備実験により提案方式の有効性を確認した。

キーワード Haar-like 特徴, Mean-Shift トラッカ, 複数視点

A Multi-Viewpoint Person Tracking System Based on Face Detection Using Haar-like Features and Mean-Shift Tracking

Tomoya TACHIBANA Atsushi YAMASHITA and Toru KANEKO

Department of Mechanical Engineering, Shizuoka University

3-5-1 Johoku Hamamatsu-shi, Shizuoka 432-8561, Japan

E-mail: {f0210084, tayamas, tmtkane}@ipc.shizuoka.ac.jp

Abstract Person tracking is an important function to an automatic surveillance system using a vision sensor. However, it is difficult to identify a person exactly in an image due to the variety of poses. This paper describes a method for automatic person tracking based on face detection using Haar-like features and mean-shift tracking. The method increases its trackability by using multi-viewpoint images. A preliminary experiment showed the validity of the method.

Keyword Haar-like Features, Mean-Shift Tracker, Multi-Viewpoint

1. 序論

セキュリティニーズの高まりにより、カメラを用いた様々な環境を監視する機会が増加してきた。しかし監視作業に用いることができる人的資源には限りがあるため、この作業を自動化するための研究が盛んに行われてきた。近年はコンピュータビジョンの分野で画像を認識する技術が進展し、視覚情報を用いることが監視システムの自動化に際して有効な手段となりつつある。

1.1. 研究背景

カメラを用いた監視の目的として人物の行動の解析がある。例えば人物の動いた軌跡をたどることでその行き先を記録する、環境内の危険な領域への接近を感知して自動的にその人物に対して警報を出す等が挙げられる。

人物の行動を知るためには人物追跡を行う必要がある。自動的に人物の追跡を行うためには追跡の前段階

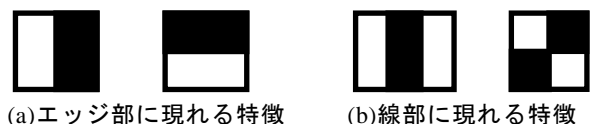


図 1 : Haar-like 特徴の例

として、認識対象が画像中に占める領域とそうでない背景領域とを分離（人物検出）する必要がある。

近年 Viola ら[1]のフレームワークを元に、Lienhart ら[2]によって改良された Haar-like 特徴（図 1）を用いる物体検出法が提案されている。これは統計学習[3]を使って作成した判別器を用いた物体検出手法で、高速かつ安定した物体検出を実現している。

また移動体を追跡する手法の 1 つとして Mean-Shift を用いたもの[4,5]がある。これは画像内で追跡対象モデルの色情報と類似した領域を計算し、求められた類似度の分布に対して探索を行うものである（図 2）。色情報が類似していれば追跡が可能のため、追跡対象の

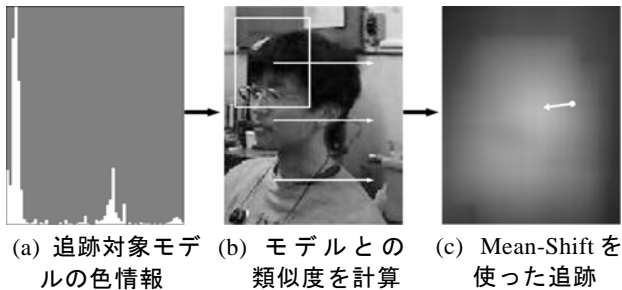


図 2 : Mean-Shift を使った追跡

形状変化や部分的なオクルージョンに対して頑健である。

上に挙げた手法の応用例として、Jaffre の Video Content Indexing に関する研究[6]では Lienhart らの物体検出手法[3]で顔を検出し、Mean-Shift トラッカを用いたコスチュームの追跡を行っている。この研究では、テレビ番組のようなカメラに対して正面を向けている人物を追跡している。

画像認識技術に関する研究では複数視点に関するものも多い。複数視点を用いる目的は認識すべき対象を様々な角度から捉え、認識精度を向上させようというアイデアに基づいている[7]。

1.2. 研究目的

Haar-like 特徴と統計学習を用いて構成した判別器を使った顔検出は複雑な背景下でも適用でき、監視に用いる場合の検出精度に問題はない。しかし人物を追跡する際、上記の Haar-like 特徴を使った検出手法のみでは不十分である。顔を検出する場合、対象の向きや角度によって見え方は変化する。一般的に Haar-like 特徴を使った物体検出は特定の姿勢に限定され、例えば正面顔用の判別器は側面顔を検出することはできない。監視の場合は記念写真のように人物の正面顔ばかりを得られるとは限らず、この手法のみでは人物を見失ってしまう。そこで顔を検出できない間を補完する方法が必要となる。

一方物体追跡手法の 1 つである Mean-Shift トラッカは事前に追跡対象モデルの色情報を得る必要があるが、形状変化に頑健という性質がある。そこで本研究では、Haar-like 特徴を使った顔検出と Mean-Shift トラッカによる顔追跡を組み合わせ、安定して人物の追跡を行う手法を提案する。

しかし Haar-like 特徴を使った顔検出と検出できない間を Mean-Shift トラッカで補完する方法では顔が検出されるまでは追跡が開始されないという問題がある。これは実際の環境では人物がカメラの方向を向いているとは限らず、従って単視点では画像に人物が出現し

てから正面顔が得られるまでに時間がかかる可能性が大きいのである。

そこで本研究では、複数視点を用いることで人物を様々な角度から捉え、検出率を向上させる手法を提案する。具体的には、人物が単一のカメラから得られた画像では検出できない姿勢の人物も複数視点を用いて別方向から捉えることで検出を行い、人物を検出できなかった視点に反映させる。

2. 人物追跡処理の流れ

本研究は Haar-like 特徴を使った顔検出と Mean-Shift トラッカの 2 つの手法の組み合わせで人物の追跡を行う。まずカメラから得られた画像に対して Haar-like 特徴を使った顔検出で人物の正面顔領域を抽出する。検出された人物が追跡中の人物でない場合は新しく追跡対象にする。すでに追跡中の人物であれば顔領域の位置とサイズを更新する。フレームごとにこの更新続けてゆくことで人物を追跡することができる。追跡中の人物の中で、正面顔の検出ができなかった未更新の人物が存在していれば Mean-Shift トラッカを用いて近傍の色情報が類似した領域を探索し、顔領域の位置を更新する。つまり基本的には Haar-like 特徴を用いた顔検出で人物の位置とサイズを更新するが、それが失敗した場合は Mean-Shift トラッカを用いて人物の位置を更

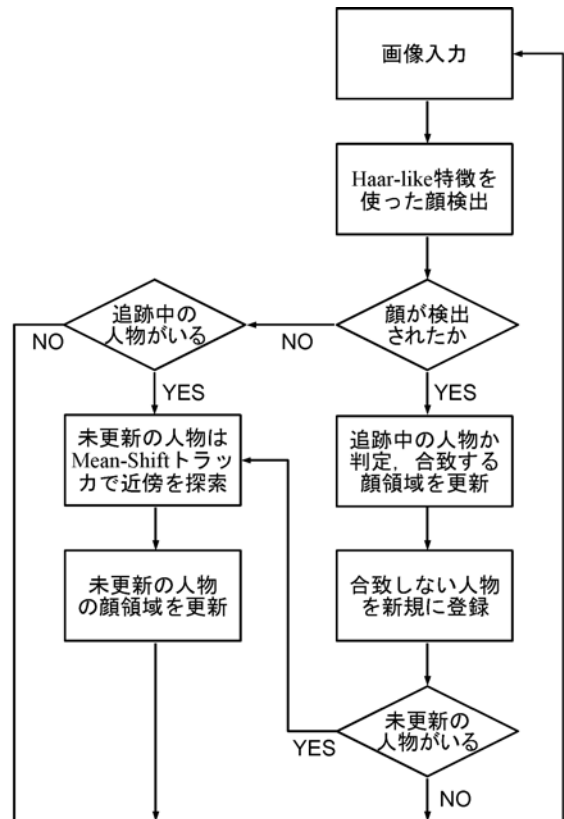


図 3 : 人物追跡処理の流れ

新する．以上の処理の流れを図3に示す．

2.1. 顔検出処理

入力された画像全てに対して[2]の Haar-like 特徴を使った顔検出手法で正面顔矩形領域を検出する．正面顔を検出した結果は図4(a)のようになり，この矩形領域を追跡対象として新規に登録，または過去のフレームで近傍に追跡中の矩形領域が存在していれば位置とサイズの更新を行う．

本研究では顔が検出されると，顔領域の色相分布を計算する．ここで求められた色相分布は Mean-Shift トラッカを使う際，追跡対象モデルの色情報として用いられる．また， Haar-like 特徴を使った顔検出はグレースケールの画像を用いているため，図4(b)のように明暗のパターンが似ていれば顔でないものも顔として検出される．誤検出の多くは図4(d)の色相分布のように(c)の正面顔領域の色相分布と類似していないものが多い．そのため検出された正面顔候補のうち，肌色が主でないものをこの段階で除去することで検出精度を向上させることができる．

2.2. Mean-Shift トラッカ

2.1 節の顔検出処理で顔を検出できず，顔領域の位置の更新ができない場合は人物の追跡を続けることができない．このような追跡中の人物については， Mean-Shift トラッカを用いて近傍の色情報が類似した領域が探索される．類似度は正面顔の検出時に求めた色相分布と現在の画像内の矩形領域の色相分布の間で，(1)式の Bhattacharyya 係数を用いて求められる．ただし， p , q は比較対象となる正規化された色相分布 ($\sum_{u=1}^m p_u = 1, \sum_{u=1}^m q_u = 1$) , u は色相の成分番号， m は色相の成分数を表す．

$$\rho = \sum_{u=1}^m \sqrt{p_u q_u} \tag{1}$$

色相分布を用いた追跡手法は[5]に基づいている．この手法では顔を追跡するために彩度や明度の低い領域の色相は無視される．これは実環境において彩度や輝度の低い部分は照明の影響を受けやすく，追跡するときの不安定要因となるためである．しかしこのままでは黒い髪の人物がカメラに対して反対方向を向いた場合に顔領域が明度の低い部分ばかりとなり，追跡できなくなる．そこで本研究では明度の低い部分に限っては本来の色相分布に割り当てず，色相分布の外側に拡張した特別な色として評価することで，カメラに対して反対方向を向いた人物の顔も追跡できるようにしている．

以上の手順で色の類似度分布が求められた後，

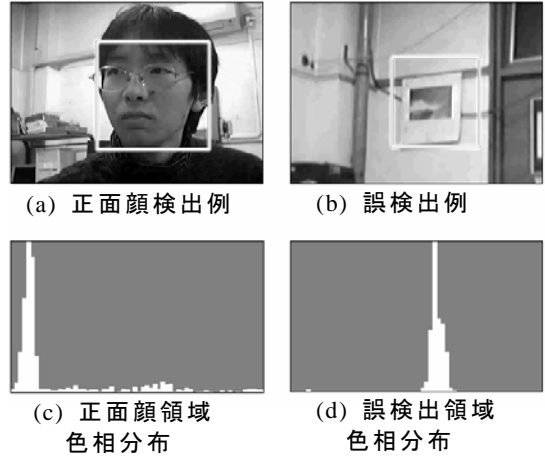


図4：検出された顔領域

Mean-Shift によって最も追跡対象らしい場所が探索され，この結果を用いて顔領域の位置の更新を行う．

3. 複数視点画像の利用

本研究ではカメラ同士の相対的な位置関係が既知である2台のカメラを用いて人物追跡を行う．この場合，一方の画像中の追跡対象の位置からもう一方で追跡対象が存在する領域を限定できる．今回は人物の顔が2台のカメラから全く見えなくなる可能性を少なくするため，カメラ同士を向かい合わせて配置する(図5)．

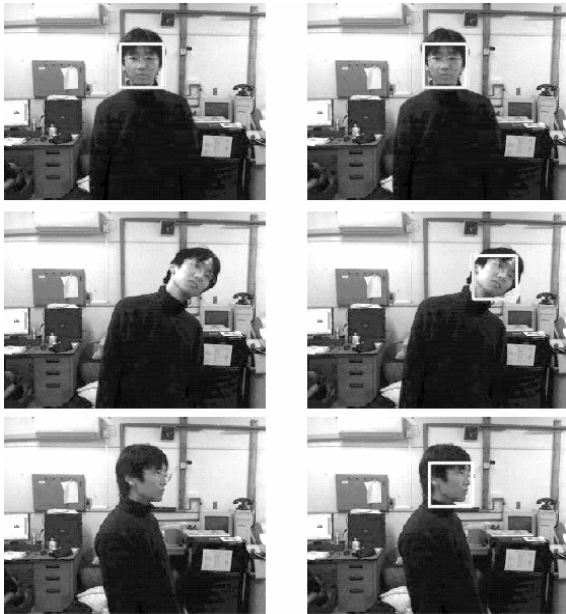
一方のカメラで人物が検出された場合，もう一方の画像中で同一人物の顔の位置を推定する．顔領域の位置情報のみではもう一方の画像中の人物の顔の位置の特定はできない．しかしカメラ同士が向かい合わせて配置されている場合，一方の画像で正面顔が検出されればもう一方の画像の同一人物は後頭部が映っていると予想される．そこで人物の毛髪が黒であると仮定し，限定された候補領域から黒色の占める割合が大きい領域を探索することで顔領域を検出する．



図5：カメラの配置

4. 実験

まず1台のカメラを用いて Haar-like 特徴を使った顔検出と Mean-Shift トラッカの組み合わせによる人物追跡実験を行った．処理結果を図6に示す．画像中の四角形は追跡中の人物の顔領域である．(a)は Haar-like 特徴による顔検出のみによる追跡，(b)は Haar-like 特



(a) Haar-like特徴を使った顔検出のみによる人物追跡結果
 (b) Haar-like特徴を使った顔検出とMean-Shiftトラッカを組み合わせた人物追跡結果

図 6: 単視点人物追跡結果

徴による顔検出と Mean-Shift トラッカとを組み合わせた手法である. 上から 1 枚目の画像では(a)と(b)双方とも成立の正面顔は正しく追跡できているが, 2, 3 枚目のように顔の姿勢が正面顔から変化した際, Haar-like 特徴を使った顔検出のみによる追跡では追跡できないことがわかる.

次に 2 台のカメラを用いて人物追跡の実験を行った. 処理結果を図 7 に示す. (a), (b)は 2 つのカメラそれぞれから得られた画像に人物追跡処理を行ったものである. (a)の上から 1 枚目の画像では正面顔が検出され, この位置情報と黒い領域の探索で(b)の上から 1 枚目で人物の頭部領域が検出できている. ここで両方の画像で対応している人物に同じラベルがつけられる. 以下, 視点間で同一人物同士が対応したまま追跡できていることがわかる.

以上の結果より, 提案手法の有効性が示された.

5. 結論

Haar-like 特徴を使った顔検出と Mean-Shift トラッカの組み合わせによる人物追跡手法を構築し, その有効性を確認した. 具体的には本研究では顔の姿勢変化に対応した.

また, 複数視点から得られる画像を用いることで単一視点では検出困難な姿勢の人物を追跡する手法を提案し, 実験により有効性を確認した.

しかし今回は人物の毛髪が黒色であるという前提



(a) 視点1
 (b) 視点2

図 7: 複数視点人物追跡結果

があったため, 今後は黒色以外の毛髪の色の人物的にも検討する必要がある.

文 献

- [1] Paul Viola and Michel J. Jones: "Rapid Object Detection Using a Boosted Cascade of Simple Features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.511-518, (2001)
- [2] Rainer Lienhart and Jochen Maydt: "An Extended Set of Haar-like Features for Rapid Object Detection", Proceedings of the 2002 IEEE International Conference on Image Processing, Vol.1, pp.900-903, (2002)
- [3] Yoav Freund and Robert E. Schapire: "Experiments with a New Boosting Algorithm", Proceedings of the Thirteenth International Conference on Machine Learning, pp.148-156, (1996)
- [4] Dorin Comaniciu, Visvanathan Ramesh and Peter Meer: "Real-Time Tracking of Non-Rigid Objects Using Mean Shift", Proceedings of the 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, pp.142-149, (2000)
- [5] Gary R. Bradski: "Real Time Face and Object Tracking as a Component of a Perceptual User Interface", Proceedings of the 4th IEEE Workshop on Applications of Computer Vision, pp.214-219, (1998)
- [6] Gael Jaffre and Philippe Joly: "Costume: A New Feature for Automatic Video Content Indexing", Proceedings of RIAO2004, pp.314-325, (2004)
- [7] 福井 和広: "複数視点画像を用いた顔画像認識", 精密工学会 画像応用技術専門委員会研究報告, Vol.19, No.3, pp.1-10, (2004)