

任意姿勢の顔検出と Mean-Shift トラッカによる 複数視点人物追跡

竹内 大悟[†] 伊東 祐[†] 山下 淳[†] 金子 透[†]

[†] 静岡大学工学部機械工学科 〒432-8561 静岡県浜松市中区城北 3-5-1

E-mail: [†] {f0510083, f0730011, tayamas, tmtkane}@ipc.shizuoka.ac.jp

あらまし ビジョンセンサを用いて環境を自動的に監視する場合、画像中の人物の追跡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。この問題に対して我々は Haar-like 特徴と Mean-Shift トラッカを組み合わせた人物追跡手法を提案したが、正面顔が検出されるまで人物追跡が行われないという問題があった。そこで、本研究では判別器を生成し、任意姿勢の顔検出を行うことでこの問題を解決する。

キーワード AdaBoost アルゴリズム, Haar-like 特徴, Mean-Shift トラッカ, 複数視点

Multi-Viewpoint Person Tracking Based on Face Detection of Arbitrary Pose and Mean-Shift Algorithm

Daigo TAKEUCHI[†] Yu ITO[†] Atsushi YAMASHITA[†] and Toru KANEKO[†]

[†] Department of Mechanical Engineering, Shizuoka University

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

E-mail: [†] {f0510083, f0730011, tayamas, tmtkane}@ipc.shizuoka.ac.jp

Abstract Person tracking in images 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. We previously proposed a person tracking method by Haar-like features and mean-shift algorithm. However, there is a problem that a person is not tracked until his/her frontal face is detected. This problem is solved in this study by face detection of arbitrary poses using trained cascades.

Keyword AdaBoost Algorithm, Haar-like Features, Mean-Shift Algorithm, Multi-Viewpoint

1. 序論

セキュリティニーズの高まりにより、カメラを用いて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。

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。

物体検出手法の1つとして Haar-like 特徴を用いたもの[1]がある。これは統計学習[2]を使って生成した検出対象か否かを判別するための判別器を用いた物体検出手法である。これにより高速かつ安定した物体検出を実現している。

統計学習による判別器生成の手法の1つとして Haar-like 特徴と AdaBoost アルゴリズムを組み合わせで用いたものがある[3]。AdaBoost アルゴリズムとは判別器を生成する（学習）アルゴリズムのことである。これらを用いて学習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。

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、例えば図1のように正面顔用の判別器は下向きの顔を検出することはできない（検出が成功した場合矩形を表示）。我々は正面顔を検出できない間を補完する方法として Mean-Shift トラッ

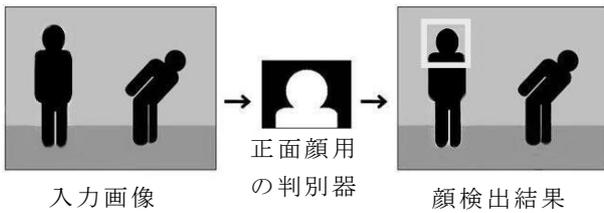


図1：顔検出の例

カ[4][5]を併用して人物追跡を行う手法を提案している[6]. 更に単視点では人物が物体に隠れた場合追跡ができないという問題があるため、複数視点を用いて解決している.

以上の手法を用いることで正面顔の検出後は安定した追跡を行うことが可能である.

2. 研究目的

文献[6]では正面顔の判別器のみを使用した場合、正面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. しかし、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、従って正面顔の判別器のみ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.

そこで本研究では、Haar-like 特徴と AdaBoost アルゴリズムにより学習した判別器を用いることで人物の複数の姿勢に対応することにより、検出率を向上させる手法を提案する. 具体的には我々の手法[6]に加え、正面顔の判別器では検出できない側面顔や下向きの顔の判別器を共に用いることで顔検出を行う.

3. 人物追跡処理の流れ

判別器による顔検出と Mean-Shift トラッカの2つの手法の組み合わせによる人物追跡処理を説明する. まずカメラから得られた画像に対して判別器を使った顔検出で人物の顔領域を検出する. 基本的には判別器を用いた顔検出を続けることで追跡を行う. また、1度検出に成功すれば、以降検出に失敗しても Mean-Shift トラッカにより検出した顔の色情報を用いて類似する領域を探索することで追跡を行う.

4. 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2).

一方のカメラで人物が検出された場合、もう一方の



図2：カメラの配置

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が難しい. 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・正面下向きの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な領域を探索することで顔領域を検出する. また、一方の画像で右側面顔が検出されたらもう一方の画像では左側面顔が映っていると予測される. そのため事前に左側面顔の色情報を与えておくことで類似する領域を探索することで顔領域を検出する. 逆に左側面顔が検出された場合は右側面顔の色情報を事前に用意して類似する領域を探索することで顔領域を検出する.

5. 判別器生成

Haar-like 特徴(図3)と AdaBoost アルゴリズムを用いた判別器生成処理の流れを説明する[3].

まず、判別器生成に使用する検出対象の画像と非検出対象の画像を用意する. 次に、これらの画像内に存在する Haar-like 特徴ごとに弱判別器を生成する. ここで、弱判別器とはある Haar-like 特徴が画像内に存在するか否かを判別するものである.

AdaBoost アルゴリズムを用いて得られた弱判別器から判別に有効なものを選択し保存する. 選択と保存を繰り返すことにより判別器を生成する. 以上の工程を経て、判別器は実際の処理に用いることができる.



(a)エッジ部に現れる特徴 (b)線部に現れる特徴

図3：Haar-like特徴の例

5.1. 弱判別器の選択と保存

画像から得られる Haar-like 特徴の数は膨大であるため、弱判別器の数も膨大になる. そのためこれらの弱判別器から判別に有効なものを選択する必要がある.

そこで、判断基準としてエラー値を用いる. エラーとは検出対象の画像であるにもかかわらず特徴が抽出

されない、または非検出対象の画像であるにもかかわらず特徴が抽出されることをいう。これを値で表したものがエラー値であり、(1)式で表される。

$$\varepsilon_k = \sum_i \omega_i |h_k(x_i) - y_i| \quad (1)$$

ただし ε はエラー値、 i は画像番号、 k は弱判別器の番号、 ω は画像の重み、 $h_k(x)$ は弱判別器を表し、入力画像 x に対応する Haar-like 特徴が存在すると判別した場合、弱判別器は 1 を出力し、存在しないと判別した場合、弱判別器は 0 を出力する。 y は入力画像が検出対象の場合 1、非検出対象の場合 0 を出力する。(1)式を用いて全ての弱判別器のエラー値を求め、エラー値が最小の弱判別器を選択し保存する。エラー値の算出と弱判別器の選択・保存の処理を繰り返し行うことで複数の弱判別器を得る。

5.2. 判別器

得られた複数の弱判別器により判別器を生成する。この判別器は(2)式により表される。

$$H(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

ただし、 H は判別器を表し、 T は繰り返し処理の回数、つまり保存された弱判別器の数を表し、 t は保存された弱判別器の番号である。 $h_t(x)$ は弱判別器を表し、入力画像 x に保存された弱判別器に対応する Haar-like 特徴が存在すると判別した場合、弱判別器は 1 を出力し、存在しないと判別した場合、弱判別器は 0 を出力する。 α は弱判別器の重みを表し、これは保存された弱判別器に対応するエラー値により求まる。

(2)式について説明する。まず、入力された画像に対して保存された弱判別器に対応する Haar-like 特徴が存在するか否かの判別を行う。全ての弱判別器に対してこの処理を行い特徴が存在すると判別した弱判別器の数が多ければ判別器は検出対象、そうでなければ非検出対象という結果を返す。

6. 実験

一般的に監視カメラは頭上に存在するため正面下向きの顔を検出する必要がある。そこで正面下向きの顔を検出するために 2 章で示した手法により判別器を生成し、その評価を行った。また、正面顔・側面顔・正面下向きの顔の判別器を用いて顔検出・人物追跡の実験を行った。なお、正面・側面顔の判別器は既存のもの (OpenCV に存在する判別器) を使用した。また、使用した計算機の CPU は Core 2 Duo の 3GHz、メモリ

は 2GB である。顔検出・人物追跡実験での画像の解像度は 320×240 pixels である。

6.1. 判別器の精度評価

正面下向きの判別器の生成には正面下向きの画像 4800 枚、非顔画像 7800 枚を用いた。この判別器の精度評価には、生成に用いた画像とは別に、正面下向きの顔画像 1500 枚と非顔画像 1500 枚を用いた。結果を表 1 に示す。ただし正面下向きの顔画像による評価の場合、成功とは正面下向きの顔が正しく検出された場合で、失敗とはそれ以外の場合のことである。非顔画像による評価の場合、何も検出されなければ成功であり、それ以外は失敗とする。

表 1: 精度評価の結果

	成功
正面下向きの顔画像	89%
非顔画像	100%

表 1 より非顔画像を用いて精度評価を行った結果、成功が 100%であった。この結果から誤って正面下向きの顔以外を検出することは少ないことが分かる。また、正面下向きの顔画像の成功は 89%という結果が得られた。この結果から検出に成功する可能性が高いことが分かる。

以上のことから正面下向きの顔検出には生成した判別器を使用する。

6.2. 顔検出実験

顔検出の結果を図 4 に示す。図 4 の左の画像 3 枚では正面顔の判別器のみを使用した結果を示し、右の画像 3 枚では正面・側面・正面下向きの顔の判別器を使用した結果を示す。画像内の矩形は顔検出が成功したことを表す。

(c)では左右の画像から正面顔が検出されていることが分かる。ただし、(a)、(b)では側面・正面下向きの顔の判別器を使用した場合のみ側面・正面下向きの顔が検出されていることが分かる。

6.3. 複数視点人物追跡

2台のカメラを用いた人物追跡の結果を図 5 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。

画像内の矩形は人物追跡が成功したことを表している。(b)の上から 1 枚目で正面下向きの顔が検出され、この位置情報と黒い領域の探索で(a)の上から 1 枚目で同一人物の後頭部が検出されている。また、(a)の上から 2 枚目で右の人物の側面顔が検出され、(b)の上から



(a)側面顔画像
(b)正面下向きの顔画像
(c)正面顔画像

図4：顔検出実験の結果

2 枚目で同一人物が追跡されていることが分かる。以降、視点間で同一人物同士が対応したまま追跡が行われている。

6.2 節, 6.3 節の結果により, 提案手法の有効性が示された。

7. 結論

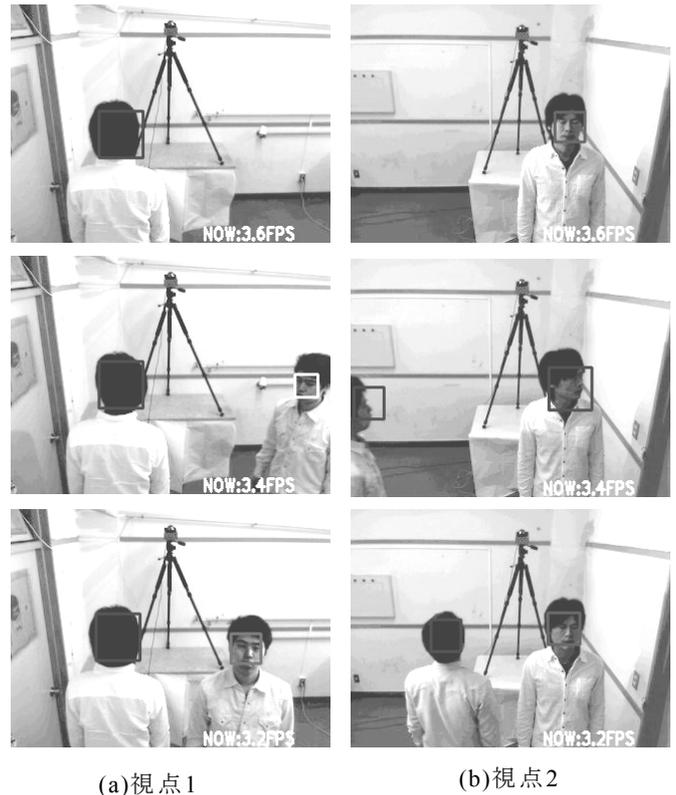
Haar-like 特徴と AdaBoost アルゴリズムにより学習した判別器を用いて任意姿勢の顔検出による人物追跡を行い, その有効性を確認した。具体的には正面顔の判別器による顔検出と Mean-Shift トラッカを用いた複数視点人物追跡に加え, 側面顔・正面下向きの顔の判別器による顔検出を行うことで, 検出率を向上させた。

ただし複数の判別器を共に使用し, 更に視点数を増加することで処理速度は低下する。これにより人物の追跡が困難となる可能性が大きいいため, 今後は処理速度の向上を検討する必要がある。

謝辞

本研究の一部は財団法人スズキ財団科学技術研究助成の補助を受けた。

また, 判別器の生成に使用した顔画像については財団法人ソフピアジャパンから使用許諾を受けたものであり, 権利者に無断で複写・利用・配布等を行うこ



(a)視点1 (b)視点2

図5：人物追跡実験の結果

とは禁じられている。

文 献

- [1] 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)
- [2] Yoav Freund and Robert E. Schapire: "Experiments with a New Boosting Algorithm", Proceedings of the 13th International Conference on Machine Learning, pp.148-156, (1996)
- [3] 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)
- [4] 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)
- [5] 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)
- [6] 立花 智也, 山下 淳, 金子 透: "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", 映像情報メディア学会技術報告, Vol.31, No.14, pp.71-74, (2007)