

Haar-like 特徴と Mean-Shift トラッカによる 複数視点人物監視システム*

伊東祐^{*1}, 山下淳^{*2}, 金子透^{*2}

A Multi-Viewpoint Human Surveillance System Using Haar-like Features and Mean-Shift Tracker

Yu Ito^{*3}, Atsushi Yamashita, Toru Kaneko

^{*3}Department of Mechanical Engineering, Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka 432-8561, Japan

Visual surveillance systems using vision sensor increase in various environments by the rise of security needs. Human tracking is an important function for an automatic surveillance system using a vision sensor. However, it is difficult to detect a human in an image with stability under general environments and to track a human exactly in an image due to the variety of poses. The face detection method using Haar-like features can realize and stable detection under general environments because it uses classifiers made by statistical learning. On the other hand, the tracking method using a mean-shift tracker has the advantage that it is robust against the shape change of tracking object. This paper describes a method for automatic human tracking based on face detection using Haar-like features and mean-shift tracking. The method increases its trackability by using multi-viewpoint images. Experimental results show the validity of the method.

Key Words : Human Tracking, Multi-Viewpoint, Haar-like Features, Mean-Shift Tracker

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いる。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いる。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、画像を人物領域と人物以外の背景領域に分離（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

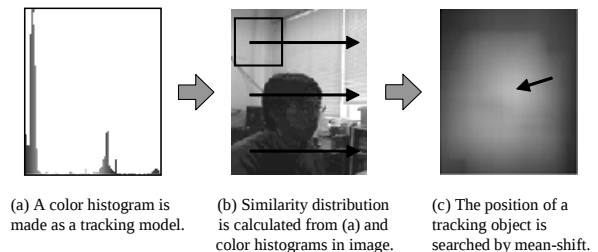


Fig. 1 Mean-shift tracker

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。これら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1）。色情報が類似していれば追跡が可能のため、追跡対象の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。

また、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⁾。

*原稿受付 2008 年 10 月 24 日

^{*1}学生員，静岡大学大学院
(〒432-8561 静岡県浜松市中区城北 3-5-1)

^{*2}正員，静岡大学
E-mail: {f0730011, tayamas, tmtkane}@ipc.shizuoka.ac.jp

カメラを用いて画像から監視を行う手法は、上記に述べたように数多く研究されているが、実用化の上では人物検出および人物追跡の精度が重要になり、その基準を満足する手法は少ない。監視をする上で求められる精度は、時系列の画像中で確実に人物を検出し、正確に追跡することと検出対象以外を間違えて検出しないことの2つの項目である。そこで、この2つ項目を満足する人物追跡手法を考案する必要がある。

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、この手法のみでは顔を見失ってしまう可能性がある。そこで顔を検出できない間を補完する方法として Mean-Shift トラッカを用いる。

一方、物体追跡手法の1つである Mean-Shift トラッカは、形状変化に頑健という性質があるが、事前に追跡対象モデルが必要になる。そこで、追跡対象モデルには Haar-like 特徴により検出した顔領域の色情報を用いる。

以上のことを踏まえて、本研究では Haar-like 特徴を使った顔検出と Mean-Shift トラッカによる人物追跡を組み合わせ、安定した人物の追跡を行う。しかし、Haar-like 特徴を使った顔検出と Mean-Shift トラッカによる人物追跡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の顔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ためである。

そこで我々は、複数視点を用いることで顔を様々な角度から捉え、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顔も、複数視点を用いて別方向から捉えることで検出を行う手法を提案する。

また、Mean-Shift トラッカにより追跡を行うと、追跡対象の動きなどにより誤追跡につながる場合が大きく分けて2つある。まず、1つ目は画像中で追跡対象の大きさが変化する場合に起こる誤追跡で、Mean-Shift の使用時にトラッカのサイズが変更されないことに原因がある。従来手法として、CAMSHIFT 法というトラッカのサイズ変更も考慮した Mean-Shift トラッカが提案されている⁽⁸⁾。この手法は、画像内の情報だけ

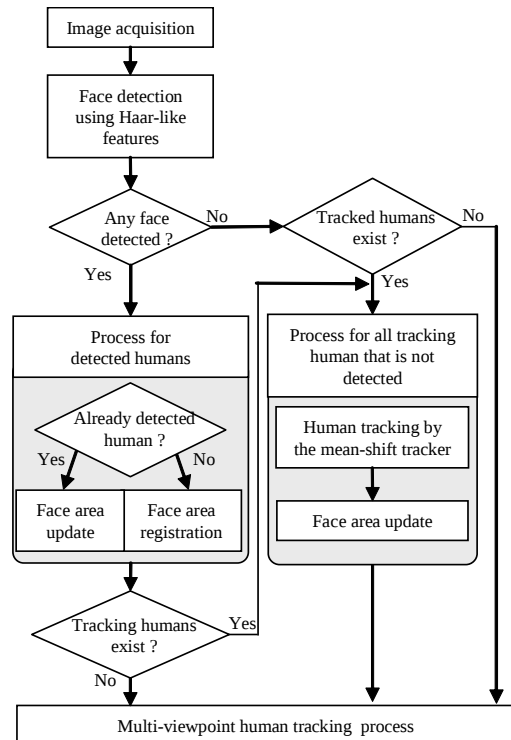


Fig. 2 A human tracking process from a single viewpoint

でトラッカのサイズを変更しているため、追跡対象と同色の物体がトラッカに隣接すると正しく変更されないことがある。そこで本研究では、複数台のカメラから算出した追跡対象の3次元座標を利用し、より正確にトラッカのサイズを変更することを目指す。そして、トラッカサイズを適切に変更することで Mean-Shift トラッカの追跡精度の向上を図る。

2つ目の誤追跡は追跡対象モデルとなる色相分布に対象以外の色が含まれてしまう場合に起こる。これは、Mean-Shift トラッカが近傍の最も類似した点を追跡対象とすることと、逐次追跡対象モデルとなる色相分布を更新することの2つが合わさることで起こりうる。ここで、逐次色相分布を更新するのは、追跡対象である人物は正面顔や後頭部などでその色相分布が異なるためである。これに対して本研究では、Mean-Shift トラッカに用いる類似度に肌と毛髪の色占有率を新たに加えることで誤追跡を低減させる。

3. 処理の流れ

始めに、単視点人物追跡処理の流れを説明する(図2)。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 特徴を使って人物の正面顔領域を検出する。検出された顔が追跡中の人物の顔でない場合は新しく追跡対象とし、顔領域の位置とサイズを記憶する。すでに追跡中の人物の顔であれば、顔領域の位置とサイ

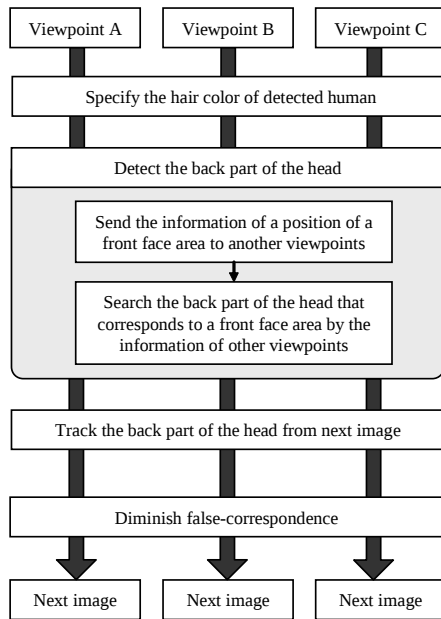


Fig. 3 A human tracking process from a multi-viewpoint

ズを更新する．検出した顔が追跡中の人物の顔と判断するのは，検出した顔領域の位置の近傍に前フレームで追跡している人物の顔領域が存在するときである．前フレームで追跡している人物のうち，正面顔が検出できずに未更新でいる人物が存在すれば 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新する．

また，Mean-Shift トラッカを使用する際，1つ前および2つ前のフレームで人物の3次元座標が算出されている場合は，人物の移動方向が予想できる．これにより Mean-Shift で使用するトラッカのサイズの変更を行う．単視点での処理が終了後，複数視点の処理を行う．

次に複数視点人物追跡処理の流れを説明する（図3）．単視点人物追跡により正面顔が検出されたら，画像中での正面顔の位置とサイズの情報を他視点に送信する．他視点から画像中での正面顔の位置とサイズの情報を受け取ると，その情報を基に正面顔に対応した後頭部を探索し，検出された後頭部を新規に登録する．検出した後頭部の領域（以後説明する上で顔領域と呼ぶ）は次のフレームから各視点で追跡する．また，視点間の対応付けを行った後，その情報を元に視点間の誤対応を削減する．

4. 単視点人物追跡

4.1 顔検出処理 入力された画像全てに対して文献⁽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)

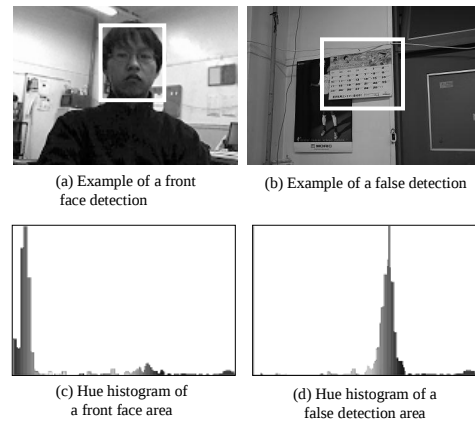


Fig. 4 Detected face region

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．

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行う．その際，正面顔に存在する明暗のパターンが多数存在する場所では，図4(b)のように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（横軸：色相，縦軸：画素数）のように図4(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．

4.2 Mean-Shift トラッカ 4.1 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を元に探索する．探索を行う際には，類似度と占有率を足し合わせたものを使用する．

類似度は前のフレームでの対象の検出結果から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，式(1)の Bhattacharyya 係数⁽⁹⁾を用いて求められる．ここで式(1)において 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．

$$\rho = \sum_{u=1}^m \sqrt{p_u q_u} \quad (1)$$

色相分布を用いた追跡手法は文献⁽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を無視する．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，追跡中の人物はカメラに対して反対方向に向くことがあ

り，その場合に追跡中の顔領域の大部分は毛髪が占め，黒などの明度の低い髪は追跡できなくなる．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，色相分布に新たに明度の低い成分を加えることで，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．

占有率は現在の画像内の矩形領域の色相分布から式(2)を用いて算出する．式(2)において q は色相分布， u は色相の成分番号， m は色相の成分数， f は肌の色に対応する色相の成分番号， h は毛髪の色に対応する色相の成分番号を表す．

$$\lambda = \frac{q_f + q_h}{\sum_{u=1}^m q_u} \quad (2)$$

類似度と占有率をそれぞれ算出した後，式(3)で合成する．ここで， α は0～1の間の値を取り，実験的に決定する．

$$\gamma = \alpha\rho + (1 - \alpha)\lambda \quad (3)$$

類似度と占有率を合成した値 γ を，計算範囲全体で算出することで色の確率分布を作成する．以上の手順で色の確率分布が求められた後，Mean-Shiftによって最も追跡対象らしい場所を探索し，この結果を用いて顔領域の位置の更新を行う．

4.3 過去の移動情報の利用 5章で説明する複数視点人物追跡により，追跡中の人物の3次元座標が1つ前と2つ前のフレームで算出された場合，人物の過去の移動情報が算出できる．過去の移動情報を利用するにあたって，現在のフレームで人物が存在するであろう位置 $\mathbf{p}(t)$ を式(4)により予測する．

$$\mathbf{p}(t) = \mathbf{p}(t-1) + \mathbf{v}(t-1, t-2)\Delta t \quad (4)$$

ここで， $\mathbf{p}(t-1)$ は1つ前のフレームでの人物の3次元座標を， $\mathbf{v}(t-1, t-2)$ は1つ前のフレームと2つ前のフレームの人物の3次元座標から求めた速度ベクトルを， Δt は1つ前のフレームから現在のフレームまでの時間を表す．また，処理を行うにあたって人物の3次元座標は，画像の縦方向，横方向，奥行方向(カメラの光軸方向)を軸とするように変換する．

カメラの光軸方向の移動情報により，Mean-Shiftで使用するトラックのサイズを変更する．これは，Mean-Shiftにより追跡している人物がカメラの光軸方向に移動した際に正しく追跡できないことがあるために行う．式(4)により求めた現在のフレームでの人物の予測位置 $\mathbf{p}(t)$ から，現在のフレームでMean-Shiftによる探索に用いるトラックのサイズを求める．1つ前のフレームにおいて光軸方向のカメラから人物までの距離を $p_z(t-1)$ ，現在のフレームでの人物の予測位置において光軸方向のカメラから人物までの距離を

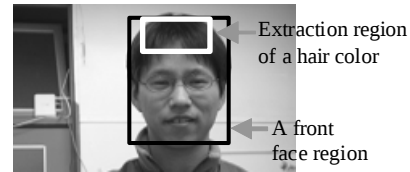


Fig. 5 Extraction of a hair color

$p_z(t)$ ，1つ前のフレームでのトラックのサイズを $S(t-1)$ ，現在のフレームでMean-Shiftによる探索に用いるトラックのサイズを $S(t)$ とすると，トラックのサイズ $S(t)$ は式(5)により算出できる．

$$S(t) = \frac{p_z(t-1)}{p_z(t)} S(t-1) \quad (5)$$

トラックのサイズを変更を行った後，4.2節のMean-Shiftトラックにより人物の追跡を行うことで，Mean-Shiftトラックの追跡精度を向上させる．

5. 複数視点人物追跡

5.1 毛髪の色の特定 5.2節で説明する視点間の同一人物の対応付けを行うためには，対応付けを行う人物の毛髪の色が必要となる．視点間の同一人物の対応付けは正面顔を検出した際に行うため，正面顔領域から毛髪の色を抽出する．正面顔の領域は4.1節で説明したHaar-like特徴を用いた顔検出手法により検出される(図5の黒枠)．検出した正面顔の領域において，その領域の上部に毛髪が写っていると予想できるため，正面顔領域の上部を毛髪の色特定領域に指定する(図5の白枠)．次に，毛髪の色特定領域において最も頻度の高い色を特定し，毛髪の色とする．頻度の高い色の特定は，毛髪の色特定領域に存在する全ての色をHSV空間に投票することで算出する．

5.2 視点間の対応付け 顔検出精度の向上を目的とした複数視点人物追跡は，位置関係が既知である3台のカメラを用いて行う．3台のカメラの配置は，カメラを配置する環境により異なるが，今回は三角形に配置し，カメラの光軸が三角形の中心に向くようにする．これにより，人物の顔が3台のカメラから全く見えなくなる可能性を少なくする．

あるカメラで人物の正面顔を検出した場合，その他の画像中での同一人物の顔の位置をエピソード拘束により限定し，推定する．ただし，画像中に映る正面顔の位置情報だけでは，その他の画像で同一人物が存在する位置は特定できない．しかし，3台のカメラを三角形に配置しているため，1台のカメラで正面顔を検出すると，その他のカメラには同一人物を斜め後方から見た画像が映っていると予想できる．そこで，5.1節で求めた毛髪の色から人物の斜め後方から見た画像のヒストグラムを作成し，そのヒストグラムと類似度の高い領域を探索することで顔領域を検出する．



Fig. 6 The experimental environment

5.3 誤対応の削減 5.2 節で説明した視点間の同一人物の対応付けだけでは、視点間で間違っ
て対応付けすることがある．そこで、あるカメラで人物の正面顔検出し、残り2台のカメラにおいても正面顔に
対応した領域を検出できた場合、視点間で行う同一人物の対応付けが適切かどうか判断する．

視点間の対応付けが適切かどうかの判断は、正面顔を検出したカメラとその他2台のカメラにより同一人物の3次元座標値を2組算出し、算出した3次元座標値を比較することで行う。視点間の対応付けが適切な場合は、2組の3次元座標値は同じになる。もし、2組の3次元座標値が異なる場合は、誤った対応付けを行ったことになる。そして、誤対応が存在する場合は、誤対応を検出・除去する。

誤対応の検出は、2台のカメラにより算出した3次元座標値を残りのカメラで撮影した画像に投影し、投影した位置を中心とした矩形領域のヒストグラムが顔(後頭部)ヒストグラムかどうかで判断する。矩形領域のヒストグラムが顔(後頭部)ヒストグラムの場合、投影した3次元座標値は正しいと判断する。もし、矩形領域のヒストグラムが顔(後頭部)ヒストグラムでない場合は、投影した3次元座標値が間違っていることを示し、3次元座標値を算出した画像で誤対応があることが分かる。

誤対応の修正は，誤対応の検出により正しいと判断された3次元座標値を投影した点を中心とした矩形領域を正しい顔領域とすることで行われる．

6. 実験

実験は先に述べたように3台のカメラを正三角形に配置して行った(図6).

まず，3視点での人物追跡を行った結果を図7に示す．画像中の四角形は追跡中の人物の顔領域の検出結果を示す．また，人物の移動軌跡を図8に示す．

図 7(a-1) では正面顔が検出され、人物 1 とラベル付けされている．図 7(a-1) に対応した図 7(b-1)・(c-1) では、正しく視点間の対応付けができていることが分かる．図 7(a-1)・(a-2) と正面顔が検出できている人物 1 は、正面顔から側面顔へと見える姿勢が変化する図 7(a-3)・(a-4) でも Mean-Shift により正しく追跡できていることが分かる．次に、図 7(b-2) でも正面顔が検出され、人物 2 とラベル付けされている．そして、

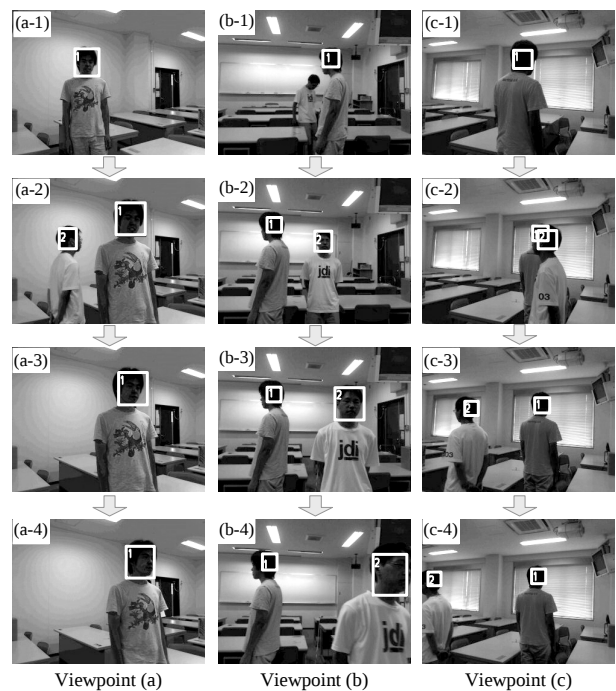


Fig. 7 The experimental result

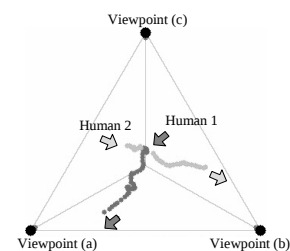


Fig. 8 Trajectory of tracking humans

図 7(a-2)・(c-2) では、正しく視点間の対応付けができていたことが分かる．図 7(c-2) では人物 1 と人物 2 が重なっているが、その場合でも、図 7(b-3) のように正しく追跡ができていたことが分かる．

また、図8に、図7で検出した人物1と人物2の移動軌跡を表示した結果を示す。図8では、図7の視点(b)から見て人物1は画像の左側へ、人物2はカメラのほうに近づいていることが分かり、その他の視点からも同様に移動軌跡が正しく出ていることが分かる。

2つ目に、トラックのサイズが変更する状況で追跡精度の向上を実験により確認した(図9)。

図 9(a) のトラックサイズを変更しない場合は、トラック内に対象以外が含まれ最終的には追跡に失敗していることが分かる．これに対して、3 次元座標を利用してトラックサイズを変更した本手法（図 9(b)）では、正しく追跡できていることが分かる．

3つ目に、Mean-Shiftトラッカで追跡する際に占有率を導入したことによる追跡精度の向上を実験により確認した(図10)。

図 10(a) は従来の類似度分布のみを使った場合で、最初は後頭部を追跡していたが、途中から誤って服を

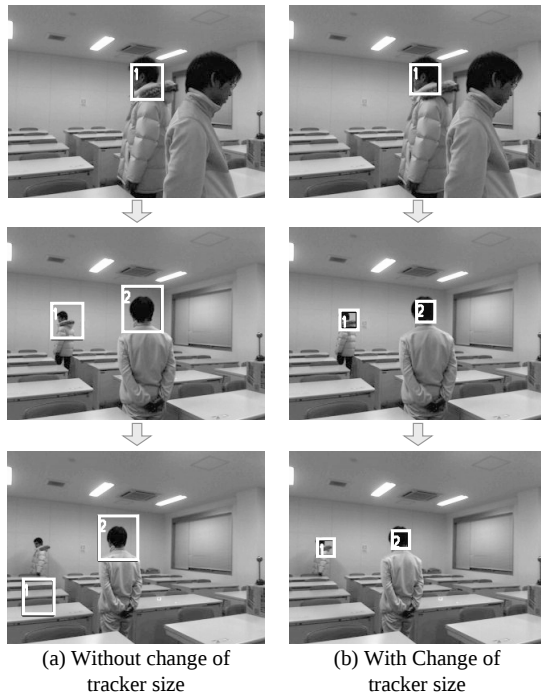


Fig. 9 Differenc of tracking accuracy by change of tracker size

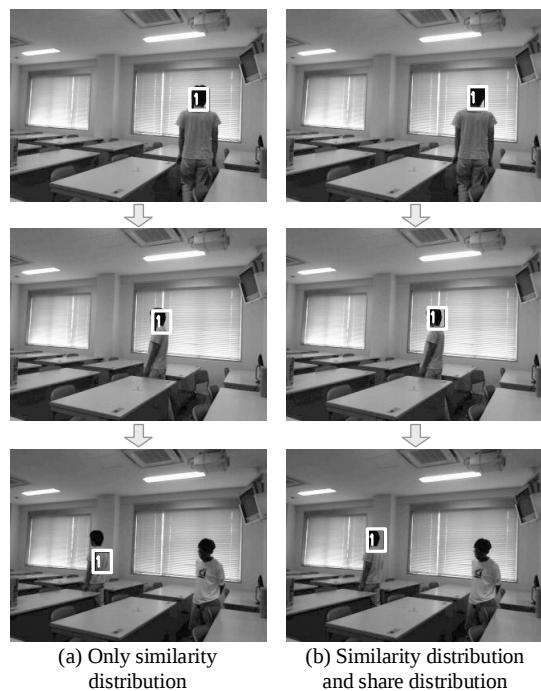


Fig. 10 Differenc of tracking accuracy by share distribution

追跡している．それに対して占有率を導入した本手法では，図 10(b) に示すように終始後頭部が追跡できていることが見て取れる．

最後に，処理速度について言及する．処理速度は Core 2 Duo 3.0GHz の CPU，2GB のメモリを搭載したパソコンで 3 視点人物追跡を行う場合，約 9fps である．この処理速度は実際に使う上では問題ないが，今後，検出する姿勢を増やす場合は処理速度の向上を

考える必要が出てくる．

以上の結果より，提案手法の有効性が確認できた．

7. 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．また，複数視点と占有率の導入により Mean-Shift トラッカの追跡精度の向上を図り，その有効性も確認した．

今後の課題として，正面以外の顔の検出や毛髪の色の特定を毛髪のない人物にも対応できるようにする，トラッカのサイズや位置の算出にカルマンフィルタを用いることなどが考えられる．

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．

文 献

- (1) P. Viola and M. 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) R. Lienhart and J. 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) Y. Freund and R. E. Schapire: "Experiments with a New Boosting Algorithm", *Proceedings of the 13th International Conference on Machine Learning*, pp.148-156, (1996).
- (4) D. Comaniciu, V. Ramesh and P. 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) G. 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) G. Jaffre and P. Joly: "Costume: A New Feature for Automatic Video Content Indexing", *Proceedings of RIAO2004*, pp.314-325, (2004).
- (7) P. Braud, M. Dhome, J. T. Laprest and N. Daucher: "Modelled Object Pose Estimation and Tracking by a Multi-Camerasystem", *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.976-979, (1994).
- (8) G. R. Bradski: "Computer Vision Face Tracking For Use in a Perceptual User Interface", *Intel Technology Journal*, No.Q2, p.15, (1998).
- (9) G. Xuan, P. Chai and M. Wu: "Bhattacharyya Distance Feature Selection", *Proceedings of the 13th International Conference on Pattern Recognition*, Vol.2, pp.195-199, (1996).