Haar-like 特徴を用いた顔検出と Mean-Shift トラッカによる人物追跡
—3 視点を利用した顔検出精度の向上—

○ 伊東祐* 山下淳** 金子透**
*静岡大学大学院工学研究科機械工学専攻
**静岡大学工学部機械工学科

ビジョンセンサを用い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が, 画像の中から人物のような姿勢変化の大きな対象を安定して追跡することは困難である．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ることにより，自動的に人物追跡を行う．しかし，単一の視点からでは顔検出精度に限界があり，顔の向きや姿勢により検出できないことがある．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．また，3 視点を利用することで，視点間で同一人物の対応付けをする際に発生する誤対応の削減を行う．実験により提案手法の有効性を確認した．

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[1]. 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[2]のフレームワークを元に，Lienhart ら[3]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[4]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.その他にも，人物の監視を行ううえで多くの情報を持つ，顔の検出や認識に関する研究もある[5].

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[6][7]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．色情報が類似していれば追跡が可能なため，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[8]では Lienhart らの物体検出手法[4]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[9].

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．

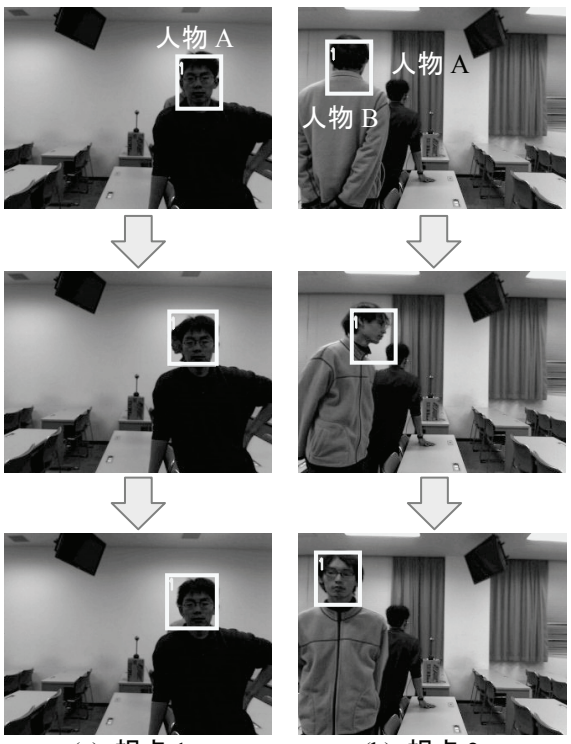
一方，物体追跡手法の 1 つである Mean-Shift トラッカは，形状変化に頑健という性質があるが，事前

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している[10].

文献[10]では，人物の顔を検出できなかった視点に正面顔を検出した視点の情報を反映させることで，顔検出精度の向上を実現している．具体的には，2 台のカメラを向かい合わせに配置し，一方のカメラで正面顔を検出したときに，もう一方のカメラに映っている後頭部を探すことで，カメラ間の同一人物の対応付けを行う．しかし，後頭部の探索に毛髪の色が黒という条件を用いているため，探索範囲内に同一人物の毛髪の色以外の黒い領域が存在するとき，その黒い領域を誤って対応付けすることがある（図 1）．図 1 では，視点 1 で検出した人物 A の正面顔に対して，視点 2 で人物 A の後頭部（黒い領域）の探索を行っている．しかし，視点 1 で検出した正面顔の人物 A とは別の人物 B の後頭部を誤って対応付けしていることが分かる．これにより，画像中には人物 A・B が映っているにもかかわらず，1 人だけいると判断されており，間違った情報を扱うことになる．そこで，視点間の対応付けをより正確に行うことが必要になる．



(a) 視点 1　　(b) 視点 2

図 1　2 視点による視点間の誤対応

以上のことを踏まえて，本研究では 3 台のカメラを利用する．これにより，カメラ間の同一人物の対応付けで起こる誤対応を削減する．また，視点数を 2 台から 3 台に増加させることにより，正面顔の検出精度の向上をはかる．

2．人物追跡処理の流れ

まず始めに，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

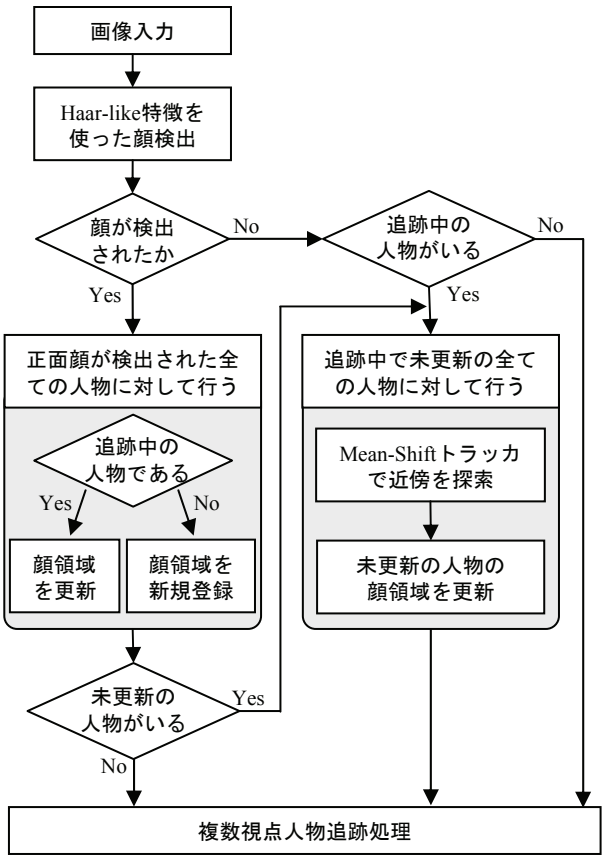


図 2　単視点人物追跡処理の流れ

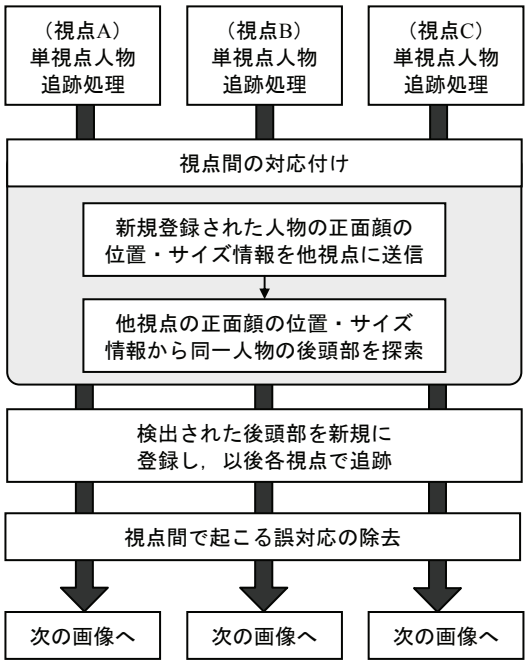


図 3　複数視点人物追跡処理の流れ

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にし，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．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，正面顔の位置とサイズの情報を他視点に送信する．他視点から正面顔の位置とサイズの情報を受け取ると，その情報を元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，視点間の対応付けが行ったあと，その情報を元に視点間の誤対応を削減する．

2.1　顔検出処理

入力された画像全てに対して[3]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．

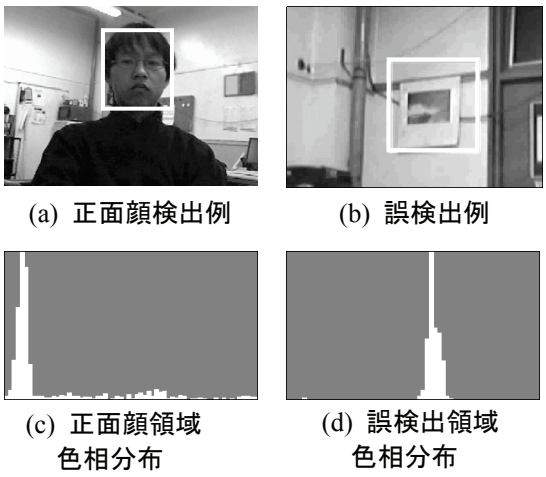


図 4　検出された顔領域

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．

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

色相分布を用いた追跡手法は[7]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，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. 複数視点の利用

3.1 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，今回は図 5 のように三角形に配置し，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．

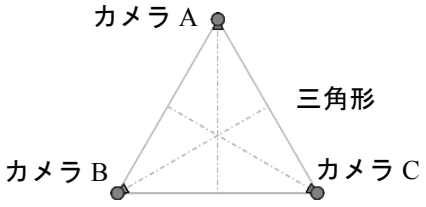
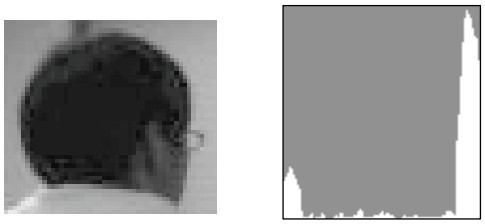


図 5　カメラ配置（上視点）

あるカメラで人物の正面顔を検出した場合，その他の画像中で同一人物の顔の位置をエピポーラ線により限定し，推定する．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（図 6）．そこで，毛髪の色を黒と仮定し，人物の斜め後方から見た画像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．



(a) 顔領域　　(b) ヒストグラム

図 6　斜め後方から見た顔領域とそのヒストグラム

3.2 誤対応の削減

3.1 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．以後説明を行う上で，正面顔を検出したカメラをカメラ A とし，その他のカメラをカメラ B・カメラ C とする．また，カメラ A・カメラ B・カメラ C で撮影した画像を，画像 A・画像 B・画像 C とする．誤対応の削減処理の流れを図 7 に示し，以下で説明する．

視点間の対応付けが適切かどうかの判断は，カメラ A・カメラ B より算出した 3 次元座標値と，カメラ A・カメラ C より算出した 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 つの 3 次元座標値が同じになる．もし，2 つの 3 次元座標が異なる場合は，視点間の同一人物の対応付けをする際に，誤った対応付けを行ったことになる．そして，誤った対応付けを行った場合は，誤対応を検出し，除去する．

まず始めに，カメラ A・カメラ B で算出した 3 次元座標値を，画像 C に投影する．そして，画像 C 上に投影された点を中心にして，ある大きさを持った矩形領域のヒストグラムを作成する．次に，矩形領域のヒストグラムと人物の斜め後方から見た画像のヒストグラムの類似度 *Sim*(A,B)を(1)式を用いて計算する．2 つのヒストグラムを比較して類似度 *Sim*(A,B)が高い場合は，カメラ A・カメラ B で算出した 3 次元座標値を真とする．逆に，2 つのヒストグラムを比較して類似度 *Sim*(A,B)が低い場合は，カメラ A・カメラ B で算出した 3 次元座標値を偽とする．3 次元座標値が偽と判断されると，カメラ A では正面顔を検出しているため，カメラ B での同一人物の対応付けが誤っていたことを示している．同様に，カメラ A･カメラ C で算出した 3 次元座標値も，画像 B に投影し，2 つのヒストグラムの類似度 *Sim*(A,C)から 3 次元座標の真偽を判断する．

最後に，カメラ A・カメラ B で算出した 3 次元座標値とカメラ A・カメラ C で算出した 3 次元座標値の真偽を確かめることで，誤った対応付けを修正する．ここで，各 3 次元座標の真偽の組み合わせには 4 通り存在する．

1 つ目として，カメラ A・カメラ B の 3 次元座標

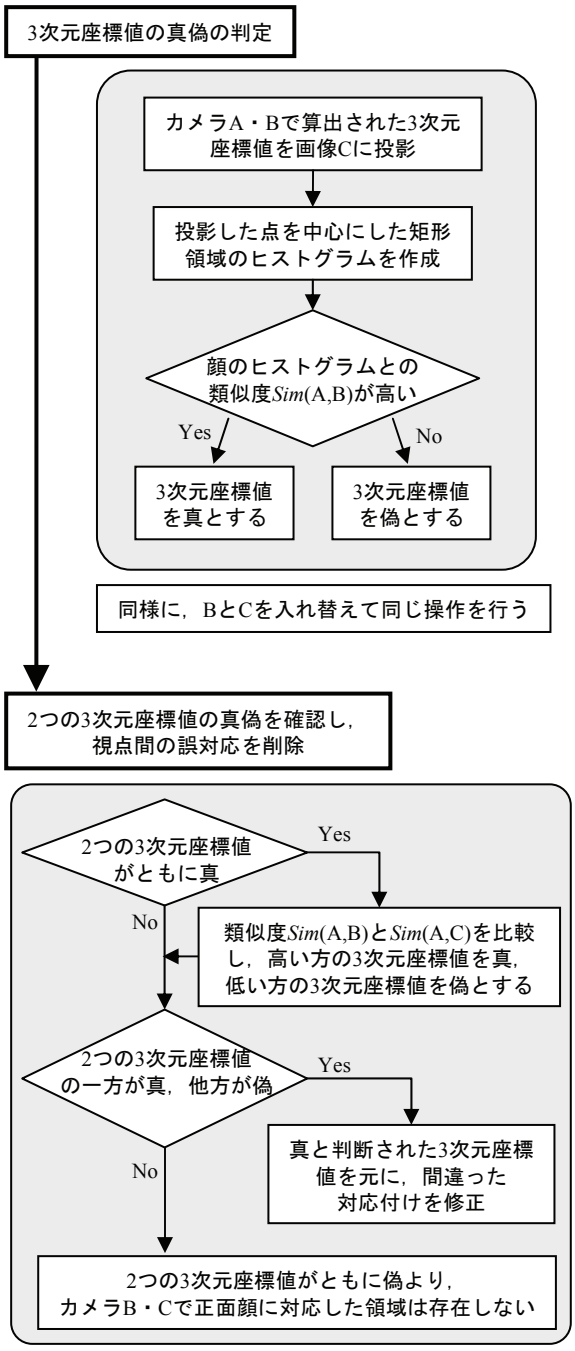


図 7　視点間で起こる誤対応の削減処理の流れ

値が真と判断され，カメラ A・カメラ C の 3 次元座標値が偽と判断された場合は，カメラ A・カメラ B の 3 次元座標値を画像 C に投影することで，カメラ C の対応付けを修正する．

2 つ目として，カメラ A・カメラ B の 3 次元座標値が偽と判断され，カメラ A・カメラ C の 3 次元座標値が真と判断された場合は，カメラ A・カメラ C の 3 次元座標値を画像 B に投影することで，カメラ B の対応付けを修正する．

3 つ目として，2 つの 3 次元座標値がともに真と判断された場合は，3 次元座標値の真偽を確かめる際に使用した *Sim*(A,B)と *Sim*(A,C)を比較し，類似度がより高い方の 3 次元座標値を真，もう一方の 3 次元座標値を偽と変更する．そして，上で述べた方法により誤対応を修正する．

4 つ目として，2 つの 3 次元座標値がともに偽と判断された場合は，カメラ B・カメラ C で正面顔に対応した領域は存在しないとする．

以上の処理により，視点間で行う同一人物の対応付けで誤対応が発生した場合，誤対応を削減できる．

4. 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 8）．

図 8　実験環境

1 つ目の実験として，3 視点での人物 1 人が映る状況で追跡を行った．その結果を図 9 に示す．画像中の四角形は追跡中の人物の顔領域である．

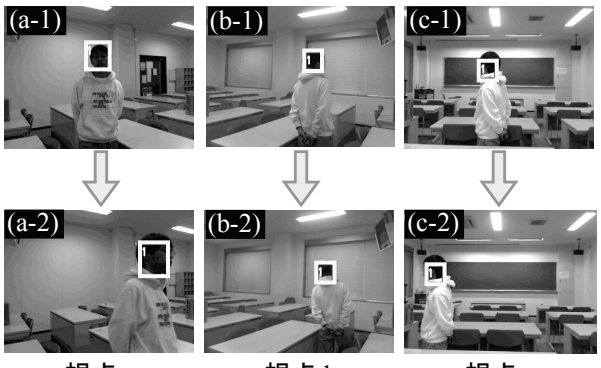


図 9　3 視点による人物追跡結果 1

図 9(a-1)では正面顔が検出され人物とラベル付けされている．そして，図 9(b-1)・(c-1)では図 9(a-1)で検出された顔の位置を基に人物が検出されている．また，図 9(a-2)では図 9(a-1)から姿勢が変化しても追跡できていることがわかる．視点 b・c のように正面顔が 1 度も検出できない場合でも，図 9(b-1)・(c-1)で検出した顔領域の情報から図 9(b-2)・(c-2)でも顔領域が正しく追跡ができていることがわかる．

2 つ目の実験として，人物が 2 人いる状況で行った. これにより, 視点間で誤対応が発生する状況で, 誤対応の削減ができているかを確認した．その結果を図 10 に示す.

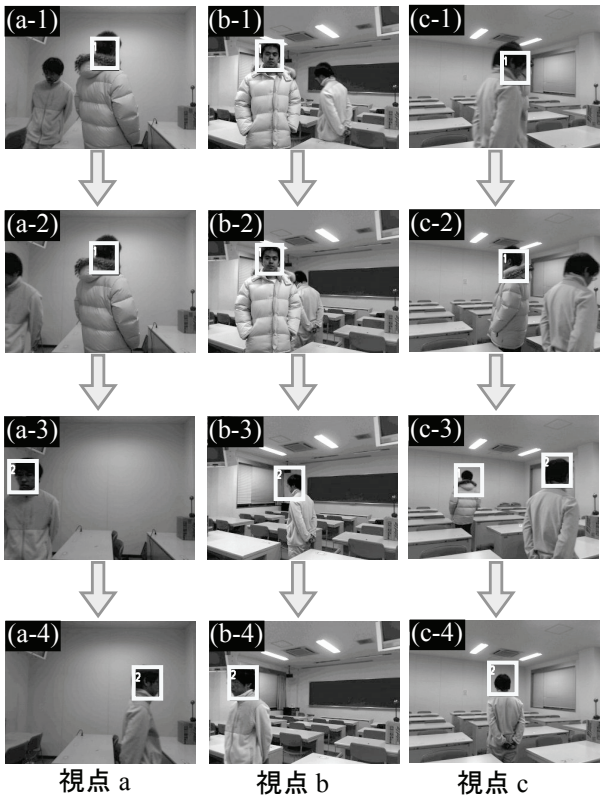


図 10　3 視点による人物追跡結果 2

図 10(b-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 10(b-1)に対応した図 10(a-1)・(c-1)のうち，図 10(a-1)では正しく視点間の対応付けができているが，図 10(c-1)では他の人物が重なっているため，視点間で誤った対応付けを行っている．しかし，時間が経過することにより図 10(c-1)で重なっていた 2 人の人物も，図 10(c-2)ではそれぞれ映っている．これにより，図 10(c-1)で誤っていた視点間の対応付けも，図 10(c-2)では修正されていることが分かる．次に，図 10(a-3)で正面顔が検出され，人物 2 とラベル付けされている．そして，図 10(a-3)に対応した図 10(b-3)・(c-3)では正しく視点間の対応付けが行われている．以後，各視点で人物 1・2 が見えている限り正しく追跡が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を，3 視点で行うことにより顔検出精度の向上をはかる手法を構築し，その有効性を確認した．具体的には，2 通りの方法で算出した人物の 3 次元座標値を比べることにより, 視点間の対応付けが正確に行えているか判断し，誤対応の際には視点間の対応付けの修正が行えたことを実験により確かめた.

しかし今回は人物の毛髪が黒色であるという前提があったため，今後は黒色以外の毛髪の人物に関しても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.

参考文献

[1] 橋本 学: “映像監視における CV 実用技術の現状と課題”, 情報処理学会研究報告(2005-CVIM-151), Vol.2005, No.112, pp.55-62, (2005)

[2] 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)

[3] 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)

[4] 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)

[5] 岩井 儀雄, 勞 世竑, 山口 修, 平山 高嗣: “画像処理による顔検出と顔認識”, 情報処理学会研究報告 (2005-CVIM-149), Vol.2005, No.38, pp.343-368, (2005)

[6] 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)

[7] 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)

[8] Gael Jaffre and Philippe Joly: “Costume: A New Feature for Automatic Video Content Indexing“, Proceedings of RIAO2004, pp.314–325, (2004)

[9] 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)

[10] 立花 智哉, 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)