計測自動制御学会東北支部 第199回研究集会 (2001.12.15)
資料番号 199-7

CONDENSATION を用いた多視点画像の融合による3次元動的環境の理解

Understanding three-dimensional dynamic environment fusing multiple images by CONDENSATION algorithm

中島平*，○浜崎浩二*, 岡谷貴之*, 出口光一郎*

Taira Nakajima*,○Kouji Hamasaki*, Takayuki Okatani*, Koichiro Deguchi*

*東北大学大学院

*Tohoku University

キーワード： コンピュータビジョン(Computer vision) 輪郭からの形状復元(Shape from silhouettes), (Condensation), 複数カメラ (Multiple Cameras), 物体追跡 (motion tracking),

連絡先： 〒980-8579 仙台市青葉区荒巻字青葉01 東北大学大学院 情報科学研究科 システム情報科学専攻 出口研究室

中島 平，Tel.: (022)217-7017，E-mail: nakag@fractal.is.tohoku.ac.jp

1. はじめに

3 次元環境で行動する複数人物の位置を知ることは，セキュリティカメラによる監視や，サッカーなどのスポーツ中継におけるフォーメーションの表示などの，様々な応用が期待される興味深い問題である．これまでに，環境中の単一人物の追跡に関しては非常に多くの研究がなされてきているが，複数人物の追跡は，移動する人物同士の相互干渉等のために困難な課題であり，最新のサーベイにおいてもわずかな例しか見られない[1]．

これまでに提案されている複数人物追跡法のうちの多くは人物同士の相互干渉を軽減するために，複数のカメラを用いている[2][3]．しかし，それらの手法でも，撮影画像をカメラごとに独立して扱うために，相互干渉の軽減効果が少なかったり，異なるカメラで撮影された同一人物の対応を取るための余分な処理を必要とするなどの問題があった．

本研究では，平面上を移動する複数人物の位置を知ることを目的とし，複数のカメラによって多方向から撮影された動画像を，空間内の仮想的な水平面上で融合し，その仮想平面上に現れる人物の像を追跡することを試みる．本方法では常に複数のカメラが人物を捉える．そのため，既存の手法に比べて目標を追跡できる範囲が狭くなるという欠点を持つ一方で，単一のカメラに由来するノイズや，複数人物の相互干渉を大幅に低減し，よりロバストな追跡を行うことが期待される．

以下，第 2 章では，物体の輪郭から形状復元を行う Shape from silhouettes 法[4] を応用した，2 次元平面への多視点画像の融合手法について述べる．第 3 章では，融合された画像を元に複数人物の時系列追跡を行うための手法として，CONDENSATION 法[5]を説明し，本論文で仮定した人物の運動モデ

ルについて述べる．第 4 章は，提案方法のシミュレーション結果と考察，そして，現実のシーンへの適用例を示す．第 5 章は結論である．

2. 多視点画像の 2 次元平面への融合

2.1 Shape from silhouettes 法

Shape from silhouettes 法(以下 SS法) は，物体の輪郭から形状復元を行う手法である．SS法は多数のカメラを用いて多方向から対象を撮影し，各カメラから得られる物体の輪郭画像を逆投影した錐体内部の積で物体の形状を表現する(図1)．

この手法は，物体の凹凸に起因する隠蔽によって必ずしも正確な形状は得られないという欠点があるものの，複数の画像の対応点のマッチングを必要とせず，カメラの位置，角度などのパラメータが不正確であることによって，形状が大幅に欠落することもない．さらに，形状を復元するために用いる輪郭画像は，背景差分法により，容易に得ることができる．本研究では，実用上の観点からSS法の簡易性に着目した．

さらに，本研究の目的は物体の正確な 3 次元形状を得ることではなく，動く物体の位置を知ることにあるので，通常のSS法では多視点画像を 3 次元空間に逆投影するのに対して，本手法では，多視点画像を 2 次元の仮想的な水平面に逆投影し，その水平面における物体の断面形状を得ることとする(図2)．このことにより，人体は 2 次元上の楕円として近似的に表現することができる．これは，人体を 3 次元で表現するのと比較して，人物の姿勢に制限が加わるものの，よりシンプルに問題を表現することができるために，処理速度の向上とシステムのロバスト性が見込まれる．

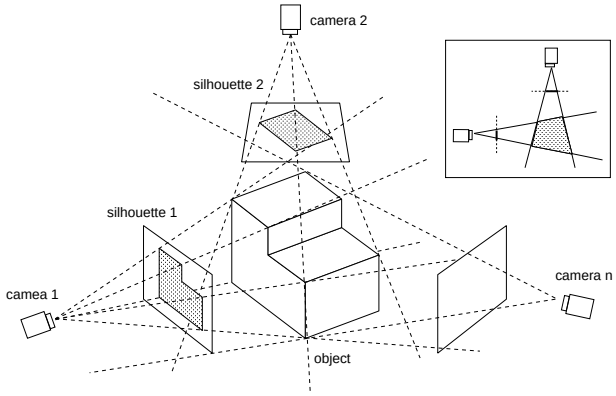


Fig. 1　輪郭からの形状復元(3次元)

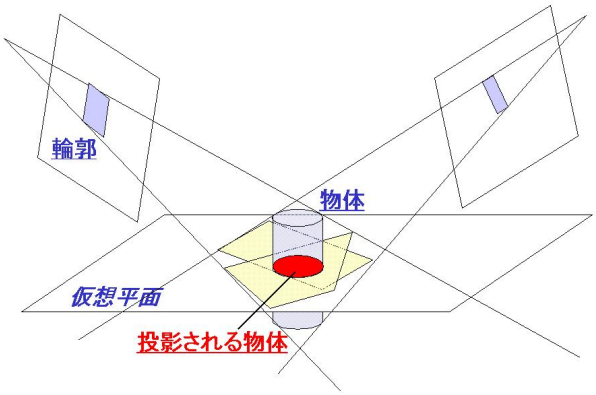


Fig. 2　輪郭からの形状復元(2次元)

2.2 輪郭画像の水平面への融合

ここでは，実際に複数のカメラからの輪郭画像から空間内の水平面に物体の断面形状を復元する方法を述べる．まず，背景差分法により，2値化された輪郭画像を得る．SS法は直接的には，輪郭画像を水平面に逆投影した領域内部の積として形状を得る．しかし，逆投影をするためには，カメラからの距離に比例して画像の拡大や補完が必要となり，処理が複雑になる．

本論文では，目的とする水平面上に仮想的な点を考え，その点を各カメラからの輪郭画像へと投影する．全てのカメラ画像で輪郭内に入っている場合にのみ，その点が輪郭画像を逆投影した積の内部となる．実際には，水平面上にピクセルの集

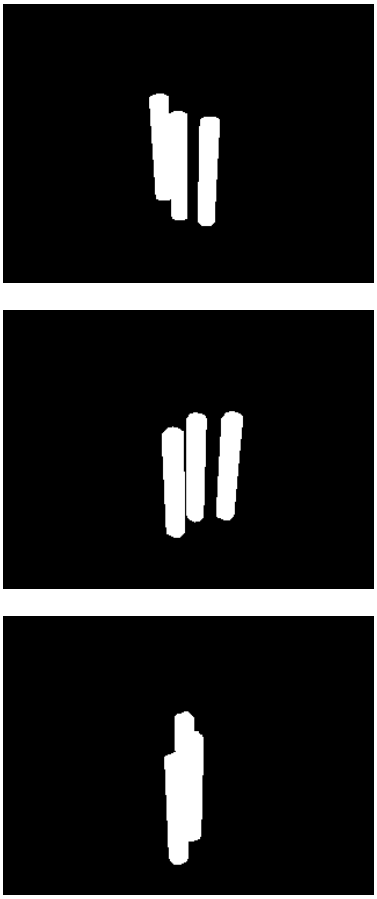


Fig. 3　3 枚の輪郭画像

合を仮定し，積の内部に相当するピクセルには 0 以外の画素値を与え，積の外部のピクセルには画素値0を与えることによって，物体の有無を区別する．

次に，投影の仕組みを説明する．ワールド座標上の点$(x, y, z)^\top$ と，i番目のカメラ上の点$(u_i, v_i)^\top$ は，式(1)で結び付く[6]．

$$w_i \begin{pmatrix} u_i \\ v_i \end{pmatrix} = P_i \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \qquad (1)$$

ここで，w_iは0でない定数であり，P_i は，透視投影行列である．本研究はカメラのモデルとして，理想的なピンホールカメラモデルを仮定する．この場合，透視投影行列はカメラの焦点距離とワールド座標上の位置，姿勢(角度)によって定まる．よって，平面上の点 (x, y, z_0) (z_0は定数)を決めれば，

Fig. 4　輪郭からの二次元形状復元例

カメラi上の点は式から w_iを消去することによって求まる．

後述する人工データを用いたシミュレーションで得られた，3つの円柱を3台のカメラで捉えた3枚の輪郭を図 3に，その平面上への逆投影像を図 4に示す．

3.　複数人物の追跡

フレーム間の輝度変化等に起因する特徴抽出の誤りや画像ノイズ等によって，不完全な観測データしか得られない状況で標的物体を追跡する場合，時系列フィルタリングが有効であることが知られている[7]．

時系列フィルタリングは，真値に対する過去の観測データに基づいた推定値である予報値と現在の観測値から最適な推定値を求める手法であり，代表的なものにカルマンフィルタがある．カルマンフィルタを適用するためには，予報値，観測値，推定値がガウス分布に従っている必要がある．ところが，図 4に示されるように，複数物体の追跡では観測データが非ガウス型(多峰性)になる場合があり，カルマンフィルタの適用は困難である．このような非ガウス型の分布に対して，時系列フィルタリングを行う手法として，CONDENSATIONが提案されている．

3.1 CONDENSATION

CONDENSATIONは逐次モンテカルロ法として総称される統計的状態推定手法の一種である．CONDENSATIONでは，モンテカルロ法により生成した多数の粒子によって，予報値，推定値の分布を近似し，ベイズの法則(式(2))を利用することにより，非ガウス型の分布に対して時系列フィルタリングを行う．

$$p(\boldsymbol{x}|\boldsymbol{z}) = kp(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x}) \qquad (2)$$

ここで$\boldsymbol{x}$は 追跡物体の位置や速度等の状態量， $\boldsymbol{z}$は画像から得られる観測データであり，kは定数である．図 5に Condensation の単位時間ステップにおける処理を示す．理解を容易にするため，通常多次元ベクトルである$\boldsymbol{x}, \boldsymbol{z}$はスカラ$x, z$として説明する． 処理はさらに三つのサブステップから構成される．

1) 入力として，時刻$(t-1)$における 観測z_{t-1}が得られたときの状態量x_{t-1} の分布$p(x_{t-1}|z_{t-1})$が，N個のサンプル$s_{t-1}^{(n)}, n = 1, \cdots, N$ と，その確率$\pi_{t-1}^{(n)}, n = 1, \cdots, N$ によって近似的に与えられる．

2) あらかじめ与えてある状態量の時間遷移モデル $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ に基づいて，$p(x_t|z_{t-1})$ に相当するN個の新サンプル$s_t^{(n)}$を計算する．(時間遷移モデルの詳細例は後述する)

3) 観測を行い得られたz_t を，あらかじめ与えてある$p(z_t|x_t)$に用いて，新しいサンプル$s_t^{(n)}$の確率$p(x_t|z_t)$を求める．得られる出力は時刻tにおける 観測z_tが得られたときの状態量x_t の分布$p(x_t|z_t)$の近似表現，$s_t^{(n)}, \pi_t^{(n)}$ である．

上記の説明で分かるように，CONDENSATION を利用するためには，

初期状態 時刻 0 の状態量 $\boldsymbol{x}_0$ の分布 $s_0^{(n)}, \pi_0^{(n)}$

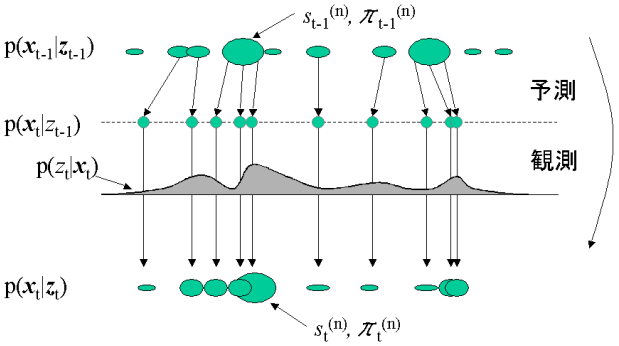


Fig. 5 CONDENSATION

状態量分布の時間更新 時刻 $(t-1)$ の状態量 $\boldsymbol{x}_{t-1}$ が得られたときの $\boldsymbol{x}_t$ の分布$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$

状態量分布の観測による更新 時刻 t の状態量$\boldsymbol{x}_t$ が得られたときの 観測 $\boldsymbol{z}_t$ の分布$p(\boldsymbol{z}_t|\boldsymbol{x}_t)$

が必要となる．次節では本提案手法における，上記3つのモデルについて述べる．

3.2 モデル

平面上に投影された人物の像の運動を考える．研究が初期段階であることから，できるだけシンプルなモデルを構築する．まず，人物を半径rの円でモデル化する．人物の運動に関しては，ある初期速度$(\dot{x}_0, \dot{y}_0)$ を持った運動で，毎フレーム5% 程度の速度変化を仮定する．

人物一人につき一つの状態量を持つとする．状態量のモデルとしては人物の平面上での位置(x_t, y_t)と，その速度$(\dot{x}_t, \dot{y}_t)$ を組にした 4組の量 $\boldsymbol{x}_t = (x_t, y_t, \dot{x}_t, \dot{y}_t)^\top$ を定める．初期状態として，人物の初期位置を既知，初期速度は未知とした．

状態量の時間更新は，以下の式によってモデル化した．

$$\boldsymbol{x}_t = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{x}_{t-1} + \begin{pmatrix} 0 \\ 0 \\ v \\ v \end{pmatrix} \qquad (3)$$

ここで，Δt は，フレーム間の時間，vは 平均0, 標準偏差σ_v のガウシアンシステムノイズである．

観測モデルとしては，まず，観測値zとして，複数の輪郭が投影された平面画像の画素値の集合を考える．例えば，平面の投影領域が200×200 ピクセルならば，zは40000次元のベクトルとなる．(x, y) を中心とする半径rの円を考え，その円内の画素値が0でないピクセル数に比例した確率を与えるものとする．

4. 実験

提案手法の有効性を検討するため，人工データ及び実データを用いた実験結果を示す．

4.1 人工データによるシミュレーション

幅200cm，奥行き200cm，高さ200cm の仮想空間内で，人間モデルとして半径10cm，高さ150cmの円柱が移動する場合を想定する．人間モデルの初期位置は，空間の縁で，2cm/フレームの初速をもって中心に向かって移動を開始する．毎フレーム5% 程度の速度の変化を仮定し，全ての人物が空間外部に移動するまで，追跡を続ける．また，人物の衝突判定は行わない．

CONDENSATIONのパラメタとしては，初期速度を平均0 標準偏差2.0 のガウス分布に従うとし，単位フレームの時間 $\Delta t = 1$, システムノイズの標準偏差$\sigma_v = 1.0$, 観測モデルの観測半径$r = 10$とした．

基準となる水平面を 高さ 120 cm にとり， 平面の中心を原点に取る．カメラは複数台を想定するが，全てのカメラは原点からの距離500cmとし，高さ320cm の水平面内の円周上に均等に配置する．全てのカメラは基準面上にある原点を向き，同じ焦点距離(400ピクセル)を持つ．

カメラ数を2から5まで，人物(オブジェクト)数を2から4までそれぞれ変化させ，追跡実験を行った．追跡の評価として，失敗(Fail)，低精度(Partial)，成功(Success)の3種類を定める． 追跡途中に目標が空間内に存在するにも関わらず，目標の推定位置が空間外になってしまった場合を失敗とする．また，失敗はしていないが，目標との距離が20cm以上離れてしまった場合，つまり，真のモデルと推定モデルが重ならなくなってしまった場合を低精度とする．実験は各カメラ数と人物数の組合わせで50回づつ行い，結果をパーセントで表示した(図 6)．

人物数が1の時は，カメラ台数に関わらず追跡に成功している．また，本実験では人物同士が完全に重なる可能性があるにもかかわらず，人物数が4になっても，カメラ台数が多い(5,6台)場合には，70% 程度の追跡成功率を達成している．

人物数が増えるにしたがって成功率は下がる傾向にあるが，特にカメラ数が少ないときにその傾向が大きい．例えば，カメラ台数が2台の時は人物数が3以上だと，人物を分離することが大変困難となり，成功率は5%以下までに落ち込んでしまう．また，カメラ台数が偶数の時と奇数の時で結果の傾向が異なっている．特に興味深い結果として，カメラ台数が4と6の時のみに追跡に失敗していることがあげられる．カメラ台数が2,4,6台の時は2つのカメラが向かい合って配置される．人物の投影面の高さは120cm であるため，人物一人を撮影したときにも，互いに向かい合った 2 つのカメラの水平面への投影像の積集合が大きくなってしまう傾向がある．そのため，4台以上のカメラで複数人物を撮影したときに，本来は人物が存在しない位置に投影像ができてしまう可能性が高く，その幻の投影像を追跡してしまうために，目標を見失い，追跡に失敗してしまうと考えられる．この問題を避けるためには，どの2つのカメラも一直線上で向かい合って配置されないようにする必要があると思われる．

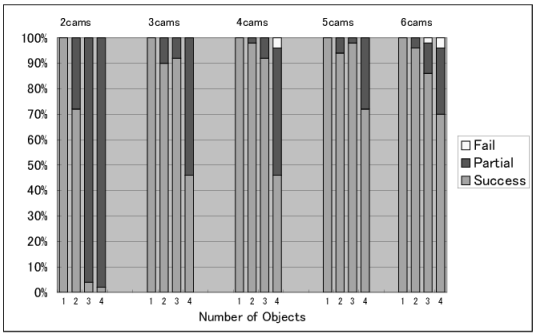


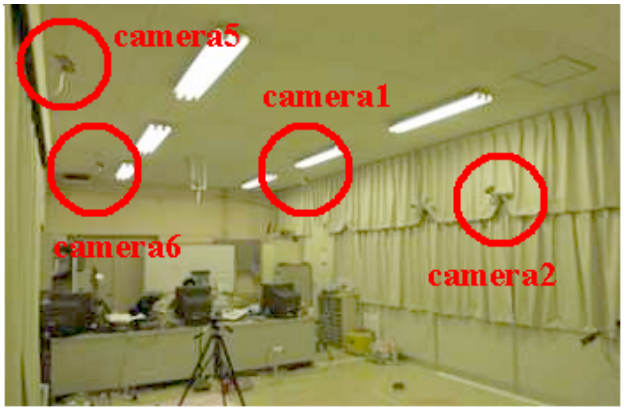
Fig. 6　カメラ台数と人物数を変化させたときの追跡結果

4.2　現実のシーンへの適用

本提案手法を現実のシーンへと適用した．CONDENSATIONのパラメータは人工データの時と同じものを用いた．図 7，8 , 9にそれぞれ，実験環境，複数のカメラから投影された水平面上の逆投影像，追跡結果を示す．図中の白い部分が追跡可能範囲であり，追跡結果は半径10cmの円で示されている．途中で2つの人物像が融合しているにも関わらず，再び像が分離したときに，それぞれ元の像を正しく追跡している．

5.　おわりに

本研究では，平面上を移動する複数人物の位置を知ることを目的とし，複数のカメラによって多方向から撮影された動画像を，空間内の仮想的な水平面上で融合し，その仮想平面上に現れる人物の像を追跡することを試みた．複数の画像を融合することにより，それぞれの単一のカメラに由来するノイズが軽減され，非常に単純な背景差分法を用いているのにも関わらず，ロバストな特徴量の抽出が達成された．また，特徴量の観測データは非ガウス的振る舞いをするために，複数人物の追跡にはCONDENSATION法を利用したが，人工



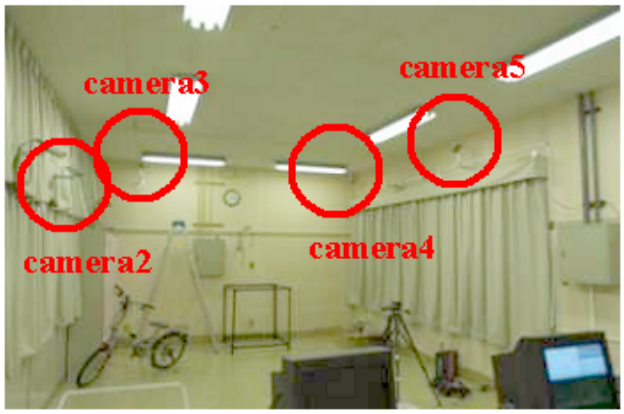


Fig. 7　実験環境

データと実データを用いた実験によって，良好な追跡が可能であることを示した．また，任意の2台のカメラが，直線上で互いに向かい合って配置される場合に追跡性能が落ちる場合があることから，カメラの配置に配慮する必要があることを示した．

今後の課題を以下に示す．現在の段階では，人物の追跡初期位置は既知のものとして扱っているが，実用化のためには，人物の追跡初期位置を自動で求める必要がある．また，追跡範囲を広げるために，効果的なカメラの配置や台数を知ることも求められる．

参考文献

1) T.B. Moeslund and E.Granum: A Survay of Computer Vision-Based Human Motion Capture, Computer Vision and Image Understanding, to appear. (2001)

2) Q. Cai and J.K. Aggarwal: Tracking Human Motion Using Multiple Cameras, International Conference on Computer Vision and Pattern Recognition (1998)

3) A. Nakazawa, H. Kato, and S. Inokuchi: Human Tracking Using Distributed Video Systems, International Conference on Pattern Recognition (1998)

4) 吉沢 徹: 光によるヒトの 3 次元形状計測, 計測と制御, **39**-4, 267/272 (2000)

5) M. Isard and A. Blake: CONDENSATION - conditional density propagation for visual tracking, International Journal of Conputer Vision 29(1), 5/28 (1998)

6) 出口 光一郎: ロボットビジョンの基礎, コロナ社 (2000)

7) 市村 直幸: 自己組織型状態空間モデルを用いた運動軌跡のフィルタリング, コンピュータビジョンとイメージメディア, 128-2, 9/16 (2001)

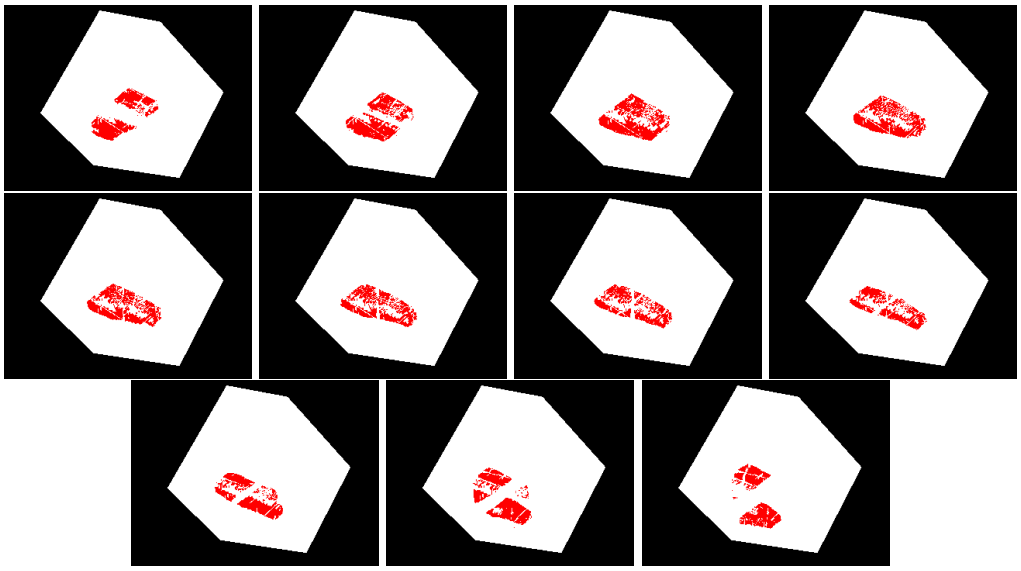


Fig. 8　逆投影画像

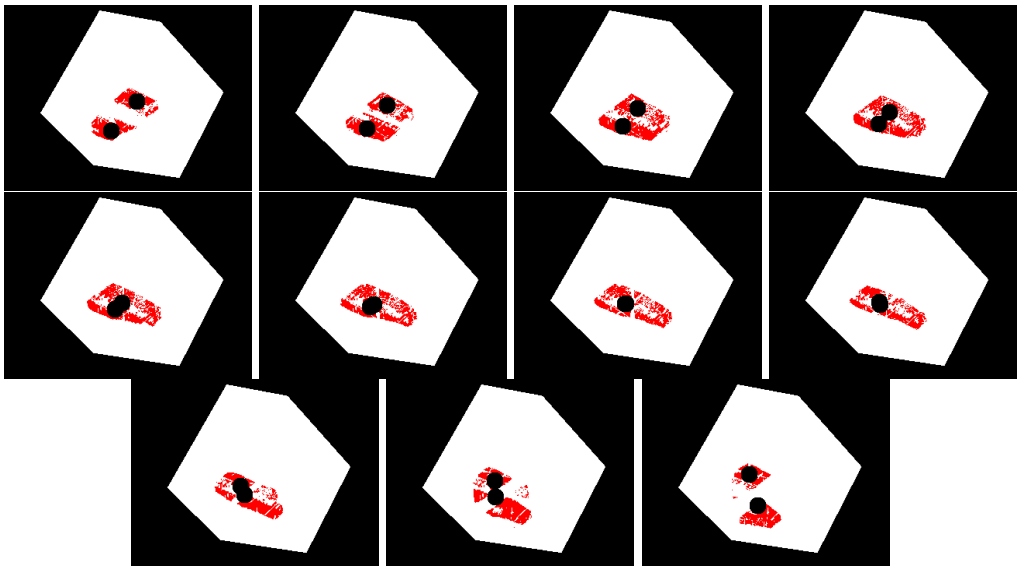


Fig. 9　追跡結果