

複数運動ビデオデータベース Hopkins155 の誤追除去

菅谷 保之^{1,a)} 松下 裕一^{1,b)} 金谷 健一^{2,c)}

概要: ビデオ画像上で追跡した特徴点軌跡を異なる運動に分類する多くの数学的手法が発表されているが、ほとんどは Hopkins155 データベースに適用して性能を評価している。本論文では Hopkins155 に問題があり、必ずしも正しい評価がなされないことを指摘し、Hopkins155 の誤追跡を除去したデータベースを作成する。誤追跡を検出する原理は、個々の運動が画像面内で並進であれば特徴点の軌跡は高次元空間の平行な 2 次元アフィン空間上に載るという事実である。平行な 2 次元アフィン空間を含む 3 次元空間が存在するので、軌跡の高次元空間を 3 次元空間に射影すれば、ビデオ画像の長さに関わらずすべての点の軌跡がディスプレイ上で可視化される。運動が厳密に並進でなくても、並進に近ければその 3 次元空間では正しい軌跡は平行に近い曲面上に載っているため、追跡の誤りが一目で分かる。並進でない運動もビデオ画像系列を短いフレーム数の小区間に分割すれば、各小区間ではそれぞれの運動はほぼ並進とみなせるので、小区間ごとの 3 次元空間で RANSAC による平面当てはめを繰り返せば誤追跡が自動的に検出できる。同時にどのフレームの近辺で誤りが生じたかも検出できる。そして、従来の RANSAC が有効でない理由が明らかになる。我々のデータベースを用いれば運動分離手法の性能を正しく評価できる。

Removing Mistracking of Multibody Motion Video Database Hopkins155

YASUYUKI SUGAYA^{1,a)} YUICHI MATSUSHITA^{1,b)} KENICHI KANATANI^{2,c)}

Abstract: Many mathematical techniques have been presented for classifying feature point trajectories over multibody motion video sequences into different motions, and most are applied to the Hopkins155 database for evaluating their performance. In this paper, we point out that Hopkins155 has problems and that correct performance evaluation is not necessarily done using it. We create a new database by removing incorrect trajectories from Hopkins155. The basic principle of mistracking removal is the fact that correct trajectories all belong to parallel 2-D affine spaces in a high-dimensional space if all motions are translational. Because parallel 2-D affine spaces are included in a 3-D affine space, all trajectories can be visualized on a display, irrespective of the video sequence length, by projecting the high-dimensional trajectory space onto that 3-D space. Even if the motions are not exactly translational, correct trajectories are represented as points on nearly parallel surfaces in that 3-D space, and hence incorrect trajectories are easily discerned by visual inspection. We automate this process by noting that if the image sequence is divided into short intervals, individual motions can be regarded as approximately translational in each interval. We detect incorrect trajectories by repeated plane fitting in the 3-D space by RANSAC. We point out why conventional RANSAC voting does not work and demonstrate that our method allows us to tell in which frames incorrect trajectories occurred. The performance of multibody motion segmentation can be correctly evaluated using our database.

¹ 豊橋技術科学大学情報・知能工学系
Department of Information and Computer Sciences, Toyohashi University of Technology, Japan

² 岡山大学特命教授
Okayama University

a) sugaya@iim.cs.tut.ac.jp

b) matusita@iim.cs.tut.ac.jp

c) kanatani2013@yahoo.co.jp

1. はじめに

ビデオ画像上を移動する特徴点の軌跡を独立な運動に分類するさまざまな方法が提案されている。Costeira ら [2] は因子分解法の計算で用いられる「作用行列」の要素の非零判定によって運動軌跡を分類した。Gear [6] は行列の標

準形を用いて運動軌跡の分離問題をグラフのマッチングに帰着させ、市村 [8] は行列の QR 分解を用い、市村 [9] は大津の判別規準を適用した。井上ら [10] はファジクラスタリングを用いた。黒澤ら [11], [12] は幾何学的 AIC によるモデル選択と投票によるロバスト推定を組み合わせた。Wu ら [27] は部分空間の直交分解を適した。Gruber ら [7] は因子分解法に EM アルゴリズムを用いた。菅谷ら [18] は複数のモデルを組み合わせる多段階最適化を提案した。Vidal ら [25][26] は多項式の当てはめによって高次元空間の点集合を複数の部分空間に分類する「一般主成分分析 (GPCA)」を提案した。Fan ら [4] や Yan ら [28] は高次元空間の点を投票によって部分空間に分類する方法を提案した。Schindler ら [16] や Rao ら [15] は最小記述長 (MDL) 原理によるモデル選択を用いている。田中ら [21] は菅谷ら [18], [19], [20] の多段階最適化と Vidal ら [25][26] の一般主成分分析 (GPCA) を組み合わせた手法を提案した。最近各運動軌跡を少数の軌跡の線形結合で表す「疎部分空間クラスタリング (SSC)」 [1], [3], 「低ランク表現 (LRR)」 [13], 「最小二乗回帰 (LSR)」 [14] などを用いてグラフの分割に帰着する方法がいろいろ検討されている。

2. 従来手法の問題点

前記の方法はどれも、撮像をアフィンカメラで近似すると、同一の剛体運動に属する特徴点の軌跡は高次元空間の 4 次元部分空間または 3 次元アフィン空間に含まれるという原理に基づいている。このため、既存の方法で必ずしも正しい分離が行われないの要因は次のように整理できる。

- (1) 撮像に強い透視投影効果があり、アフィンカメラで近似できない。
- (2) 運動が特殊で、高次元の軌跡空間に退化が生じている。
- (3) 各フレームで特徴点が正しい位置に抽出されていない。
- (4) フレーム間の異なる特徴点がマッチングされ、追跡が正しくない。

しかし、従来手法のほとんどは実験によって正解率を評価するだけであり、正しく分離されない原因が十分追求されていない。特に Johns Hopkins 大学の Tron ら [24] が Hopkins155 データベースを作成して以来その傾向が強まり、最近では高度に抽象的な数学的手法を考案しては Hopkins155 によって正解率を他の手法と比較するのが普通になっている。このため、性能の差の原因がどこにあるのか、あるいはどういう場合にその手法が有効であるかなどが不明である。今後は単に新しい手法を考案して正解率を比較するのではなく、正しく分離されない原因を個別に検討することが必要である。

筆者らの経験によると、上記の要因 (1) はほとんど考慮する必要はないと思われる。それは複数物体の運動を連続的にビデオ撮影する状況では移動物体がカメラから十分離れているからである。筆者らはこれまで要因 (2) に集中

して研究を行ってきた。退化にはさまざまなタイプがあるが、物体が平面上を移動している、回転せずに平行移動をしているなど実際に生じやすいものが多い。そこで、さまざまな退化を仮定して、特殊な運動から一般的な運動へと階層的に最適化を行ってこれを解決する手法を提案した [18], [19], [20], [21]。

しかし、要因 (3), (4) については十分研究されているとは言えない。従来は特徴点位置ずれも対応の誤りも一括して「アウトライア」と呼ばれ、「アフィンカメラの拘束を満たしていない」という観点から RANSAC による投票法がとられた。本論文ではこれはデータの発生メカニズムを考慮すると有効でないことを指摘する。しかし、従来は特徴点位置を正規乱数誤差でずらしたり、一様分布に従う大きな誤差を加えるなど、データの発生メカニズムを反映しないデータで性能評価されてきた。実用的な手法を開発するためにはデータの発生メカニズムをより詳細に検討して、それを反映した手法を導入すべきである。

3. Hopkins155 の吟味

本論文では、多くの研究者が性能評価に利用している Hopkins155 データベースを吟味する。Hopkins155 のビデオ画像ではどれも特徴点が正しく追跡されているように見えるが、詳細に観察すると微妙な誤りがかかなりある。しかし、そのような誤りはフレームごとに画像を見たのでは分かりにくい。そこで本論文では誤りを発見する作業を支援する方法を導入する。

原理は次の通りである。一つの特徴点の M フレームに渡る軌跡は $2M$ 次元空間の 1 点で表されるが、アフィンカメラの仮定が満たされるときは、同一の剛体運動に属する点の軌跡はその $2M$ 次元空間の 3 次元アフィン空間に拘束される [11], [12]。しかし、現実には運動が画像内の 2 次元的な並進、回転であることが多い。このとき 3 次元アフィン空間は 2 次元アフィン空間に退化する。複数の物体や背景に回転がない場合、すなわち互いに異なる方向に並進している場合、それぞれに対する 2 次元アフィン空間は互いに平行になる。平行な 2 次元アフィン空間は一つの 3 次元アフィン空間に含まれる。したがって、軌跡の $2M$ 次元空間をその 3 次元空間に投影すれば、正しい軌跡はすべてその 3 次元空間内の平行な平面上に載っている。たとえ各運動が完全な並進でなくても、平行な平面に近い曲面上に載っているため、その様子を目で見る事ができる。このとき誤った軌跡はそれらから飛び離れているので、フレームごとに見ると見落とすような微妙な誤りも一目で分かる。

従来の RANSAC によるアウトライア除去は $2M$ 次元空間に 4 次元部分空間や 3 次元アフィン空間を当てはめているので [17]、このような 3 次元空間内での異常を検出することはできない。本論文ではさらに上記の作業も自動化し、3 次元データの解析により、各特徴点の軌跡の正しさを

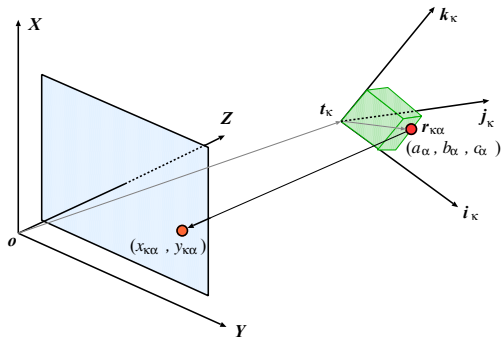


図 1 アフィンカメラモデル.

の指標を付与できることを示す. これによって人が介在しなくても誤った軌跡が除去できる. 本論文では実際のビデオ画像にどのような問題があるかを具体的な事例によって検証するとともに, Hopkins155 の 誤りを除去したデータベースを作成する.

4. アフィンカメラモデルと軌跡の空間

剛体運動する点の軌跡が $2M$ 次元空間の 3 次元アフィン空間に含まれることを示す. N 個の特徴点 $\{p_{\alpha}\}$ を M 枚の画像に渡って追跡し, 第 κ 画像における α 番目の特徴点 p_{α} の画像座標を $(x_{\kappa\alpha}, y_{\kappa\alpha}), \kappa = 1, \dots, M, \alpha = 1, \dots, N$ とする. その運動履歴を次の $2M$ 次元ベクトルで表し, 単に「軌跡」と呼ぶ.

$$p_{\alpha} = (x_{1\alpha}, y_{1\alpha}, x_{2\alpha}, y_{2\alpha}, \dots, x_{M\alpha}, y_{M\alpha})^T \quad (1)$$

これにより各特徴点の軌跡は $2M$ 次元空間の 1 点とみなせる. 本論文では, 図 1 のようにカメラの光軸を Z 軸とする XYZ カメラ座標系をとり, これに相対的にシーンが運動すると解釈する. シーン中に物体座標系を任意に固定し, 特徴点 p_{α} のその物体座標系に関する座標を $(a_{\alpha}, b_{\alpha}, c_{\alpha})$ とする. 第 κ フレームでの物体座標系の原点を t_{κ} とし, 各座標軸の基底ベクトルをカメラ座標系で表したものを $\{i_{\kappa}, j_{\kappa}, k_{\kappa}\}$ とする. 特徴点 p_{α} の第 κ フレームにおける 3 次元位置 $r_{\kappa\alpha}$ はカメラ座標系では次式で表せる.

$$r_{\kappa\alpha} = t_{\kappa} + a_{\alpha}i_{\kappa} + b_{\alpha}j_{\kappa} + c_{\alpha}k_{\kappa} \quad (2)$$

平行投影や弱透視投影, 疑似透視投影を抽象化した「アフィンカメラ」を仮定すると, 3 次元点 $r_{\kappa\alpha}$ の画像面への投影は次のようになる.

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = A_{\kappa}r_{\kappa\alpha} + b_{\kappa} \quad (3)$$

ここで, A_{κ}, b_{κ} は第 κ フレームでのカメラの位置や内部パラメータによって定まる 2×3 行列および 2 次元ベクトルである. 式 (2) を代入すると式 (3) は次のように書ける.

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \tilde{m}_{0\kappa} + a_{\alpha}\tilde{m}_{1\kappa} + b_{\alpha}\tilde{m}_{2\kappa} + c_{\alpha}\tilde{m}_{3\kappa} \quad (4)$$

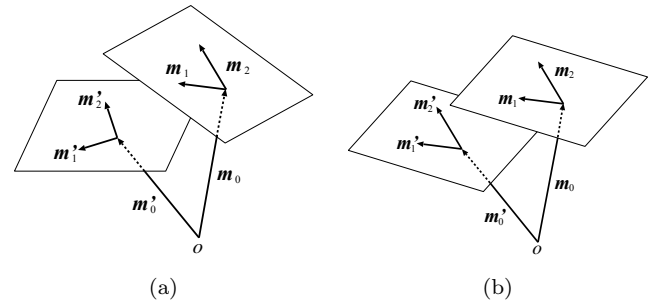


図 2 軌跡ベクトルの含まれる空間. (a) 平面運動では物体と背景の軌跡ベクトルはそれぞれ 2 次元アフィン空間に含まれる. (b) 物体も背景も回転せずに単に並進すると二つのアフィン空間は互いに平行になる.

ただし $\tilde{m}_{0\kappa}, \tilde{m}_{1\kappa}, \tilde{m}_{2\kappa}, \tilde{m}_{3\kappa}$ は第 κ フレームでのカメラ位置や内部パラメータによって決まる 2 次元ベクトルである. これをすべてのフレーム $\kappa = 1, \dots, M$ に渡って式 (1) のように縦に並べると, 式 (1) の軌跡 p_{α} が次のように書ける.

$$p_{\alpha} = m_0 + a_{\alpha}m_1 + b_{\alpha}m_2 + c_{\alpha}m_3 \quad (5)$$

ここで $m_i, i = 0, \dots, 3$ は $\tilde{m}_{i\kappa}$ をすべてのフレーム $\kappa = 1, \dots, M$ に渡って縦に並べた $2M$ 次元ベクトルである.

式 (5) は同一の剛体運動をする特徴点 p_{α} の軌跡が $2M$ 次元空間中の 4 本のベクトル $\{m_0, m_1, m_2, m_3\}$ が張る 4 次元部分空間に含まれることを表す. しかし, m_0 の係数はすべての α に共通に 1 である. このため, 実際には軌跡 p_{α} は m_0 を通り, $\{m_1, m_2, m_3\}$ の張る 3 次元アフィン空間に含まれる.

5. 軌跡の可視化

現実のシーンでは物体や背景が画像内で 2 次元的な運動をし, 回転が Z 軸 (光軸) 回りのみであることが多い. このとき物体座標系の基底ベクトル k_{κ} を Z 軸方向にとれば, アフィンカメラのもとでは画像面に投影されないから, 式 (5) の m_3 が 0 となり, 軌跡ベクトルが m_0 を通り, $\{m_1, m_2\}$ の張る 2 次元アフィン空間に含まれる (図 2(a)).

さらに, 物体も背景も回転しなければ, 式 (2) の i_{κ}, j_{κ} をそれぞれ X 方向, Y 方向の基底 i, j に固定してよい. これは物体と背景に共通であるから, 式 (5) の m_1, m_2 も物体と背景に共通になり, それぞれの 2 次元アフィン空間は平行な 2 次元アフィン空間になる (図 2(b)). そして平行な 2 次元アフィン空間を含む 3 次元アフィン空間が存在する. したがって, $2M$ 次元空間のこの 3 次元アフィン空間を通常の 3 次元に射影すれば, $2M$ 次元空間のすべての特徴点軌跡が平行な平面上の点として「目で見える」ことができる. たとえ運動が厳密な並進でなくても, 並進に近ければ, 3 次元空間中の平行な平面に近い曲面上の点として可視化される. これを主成分分析により次のように行う. (1) 軌跡 $p_{\alpha}, \alpha = 1, \dots, N$ の重心 p_C からの差 \tilde{p}_{α} を計算

する.

$$p_C = \frac{1}{N} \sum_{\alpha=1}^N p_\alpha, \quad \tilde{p}_\alpha = p_\alpha - p_C \quad (6)$$

(2) 次のように $2M \times N$ 行列の特異値分解を行う.

$$(\tilde{p}_1, \dots, \tilde{p}_N) = U \text{diag}(\sigma_1, \dots, \sigma_r) V^T \quad (7)$$

ただし, $r = \min(2M, N)$ であり, U は r 本の正規直交系の列を持つ $2M \times r$ 行列, V は r 本の正規直交系の列を持つ $N \times r$ 行列である. そして, $\sigma_1 \geq \dots \geq \sigma_r (\geq 0)$ が特異値であり, $\text{diag}(\sigma_1, \dots, \sigma_r)$ はそれらを対角要素とする対角行列である.

(3) 行列 U の第 i 列を u_i として, 次の3次元ベクトル r_α , $\alpha = 1, \dots, N$ を計算する.

$$r_\alpha = \begin{pmatrix} (\tilde{p}_\alpha, u_1) \\ (\tilde{p}_\alpha, u_2) \\ (\tilde{p}_\alpha, u_3) \end{pmatrix} \quad (8)$$

これは重心 p_C を原点とする座標系をとり, そのアフィン空間を張る3本の正規直交系(行列 U の列ベクトル)を計算して, 各軌跡をその3本の正規直交基底の線形結合で表すことに相当する.

6. 誤追跡の検出

前節のように可視化すると, 物体や背景が並進していれば正しい軌跡は3次元空間の平行な平面上にあり, 運動が厳密な並進ではなくても並進に近ければ, ほぼ平行な曲面上に載る. しかし, 追跡に誤りがあれば, その軌跡はその平面や曲面から飛び離れ, その3次元空間をディスプレイ上で回転しながら観測すると明瞭に識別できる. このようにして入力ビデオ画像を見たのでは分からなかった誤りが発見できる.

これは目視によるもので, 非常に効果的であるが, これを自動化することを考える. そのアイデアは, 物体やシーンの運動がビデオ画像上で厳密な並進ではなくても, ビデオ画像を互いに重複する短い小フレーム区間に分割すると, 各小区間では運動はほぼ並進運動とみせるということである. その小区間の軌跡を3次元空間に射影すれば, 各運動の属する軌跡がそれぞれほぼ平面上に載るので, 平面を当てはめることによって平面からはずれた点を検出できる. この操作にはRANSACが適している. すなわちRANSACのよって一つの平面を当てはめ, その平面に近い点を除いた残りの点にRANSACによって平面を当てはめる. このようにして複数の平面を当てはめ, 各軌跡の最も近い平面からの距離によってその軌跡の正しさの指標とする. これを分割したすべての小区間について行い, その結果を統合することによって最終的な判断を行う. この手順を図3に示す. 以下, 各ステップの詳細を説明する.

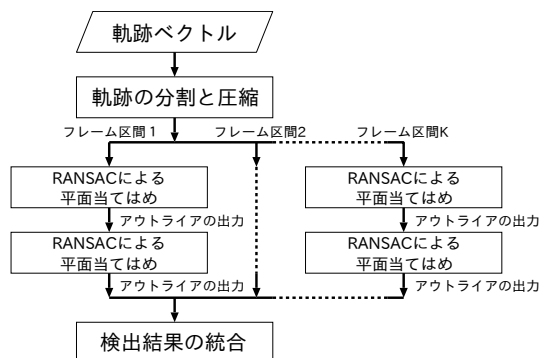


図3 提案手法のフローチャート

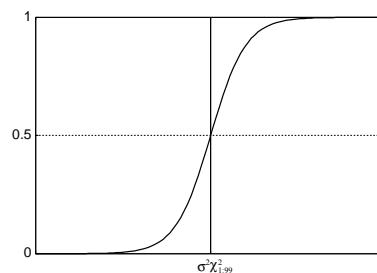


図4 シグモイド関数 $y = 1 / (1 + e^{-(x - \sigma^2 \chi_{1,99}^2)})$.

6.1 平面当てはめ

3次元点 $(x_\alpha, y_\alpha, z_\alpha)$, $\alpha = 1, \dots, N$ に平面

$$Ax + By + Cz + Df_0 = 0 \quad (9)$$

を当てはめる. f_0 は計算を安定化するためのスケール定数である. ベクトル ξ, θ を

$$\xi = (x, y, z, f_0)^T, \theta = (A, B, C, D)^T \quad (10)$$

と定義すると, 式(9)は次のように書ける.

$$(\xi, \theta) = 0 \quad (11)$$

ただし, 以下ベクトル a, b の内積を (a, b) と書く. そして θ のスケールの不定性を除くために $\|\theta\| = 1$ と正規化する. データ点に平面上にあるとは限らない場合は $(\xi_\alpha, \theta) \approx 0$, $\alpha = 1, \dots, N$ となる単位ベクトル θ を求める. このためによく知られた方法は次の J を最小化する最小二乗法である.

$$J = \sum_{\alpha=1}^N (\xi_\alpha, \theta)^2 \quad (12)$$

これは次のような2次形式で表せるので, 解は行列 M の最小固有値に対する単位固有ベクトルである.

$$J = \sum_{\alpha=1}^N (\xi_\alpha, \theta)^2 = (\theta, \underbrace{\sum_{\alpha=1}^N \xi_\alpha \xi_\alpha^T}_{\equiv M} \theta) = (\theta, M\theta) \quad (13)$$

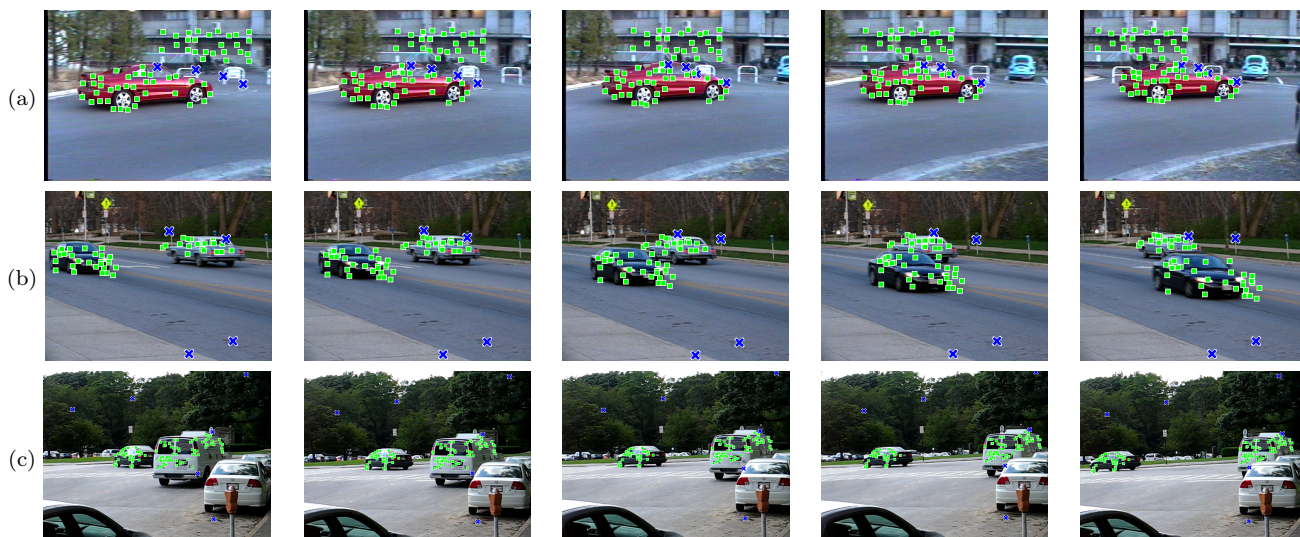


図 5 Hopkins155 データベースからの 3 例 (特徴点追跡を追加している).

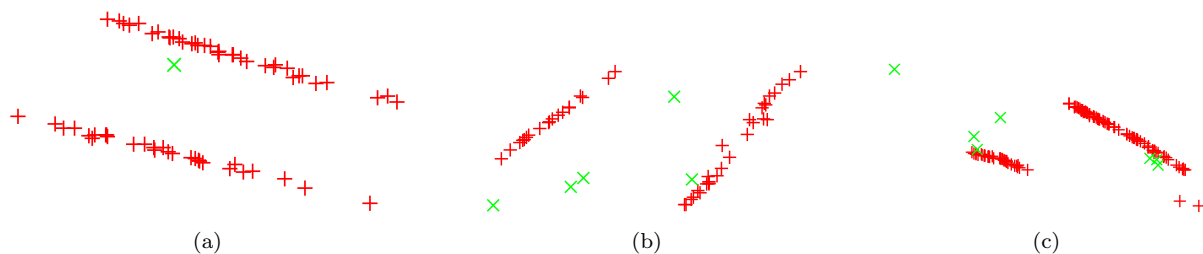


図 6 図 5 のそれぞれの最初の 5 フレームの特徴点軌跡の 3 次元表示による可視化.

6.2 RANSAC

当てはめる平面が複数あり、かつ各点が厳密にはそれらの上に乗っているとは限らない場合に複数の平面を当てはめるには次に示す RANSAC による平面当てはめを繰り返す.

- (1) 投影した 3 次元空間の軌跡 \mathbf{r}_α , $\alpha = 1, \dots, N$ からランダムに 3 点を選択する.
- (2) 選択した 3 点から最小二乗法により平面 θ を当てはめる.
- (3) すべての軌跡 $\{\mathbf{r}_\alpha\}$ のうち

$$\frac{(\mathbf{r}_\alpha, \theta)^2}{\theta_1^2 + \theta_2^2 + \theta_3^2} \leq \sigma^2 \quad (14)$$

となるものの個数を S とする. ただし, θ_i はベクトル θ の第 i 成分である. 左辺は点 \mathbf{r}_α と当てはめた平面との二乗距離である. また σ は特徴点の検出誤差の標準偏差であり, 経験的に設定する.

- (4) 上記のステップを反復し, S を最大とする平面パラメータ θ を求める.
- (5) 次式を満たす \mathbf{r}_α を除去する.

$$\frac{(\mathbf{r}_\alpha, \theta)^2}{\theta_1^2 + \theta_2^2 + \theta_3^2} > \sigma^2 \chi_{1.99}^2 \quad (15)$$

ただし $\chi_{r,a}^2$ は自由度 r の χ^2 分布の $a\%$ 点である. これは有意水準 1% でその平面から期待値 0, 標準偏差

σ の正規分布の誤差によってずれたとは見なせない点を除去するということである.

6.3 軌跡の信頼性の指標

各区間での結果を次のように統合する. そのために次のようにシグモイド関数 (図 4) を用いて, 誤りである可能性を評価する

$$P(\mathbf{p}_\alpha^{(i)}) = \frac{1}{1 + e^{-(d_\alpha^{(i)} - \sigma^2 \chi_{1.99}^2)}} \quad (16)$$

ただし $\mathbf{p}_\alpha^{(i)}$ は軌跡 \mathbf{p}_α の第 i 区間での部分軌跡を表すベクトルであり, $d_\alpha^{(i)}$ は第 i 区間の式 (15) の左辺である. ただし, 各平面に対して式 (15) が成立する部分軌跡は誤りがないとして, $P(\mathbf{p}_\alpha^{(i)}) = 0$ とする. そして全区間に対する信頼性を次式で評価する (大きいほど誤っている可能性が高い).

$$L(\mathbf{p}_\alpha) = \prod_{i|P(\mathbf{p}_\alpha^{(i)}) \neq 0} P(\mathbf{p}_\alpha^{(i)}) \quad (17)$$

ただし K はビデオ画像の分割数である. この式を用いてしきい値処理により最終的な判定を行う.

7. 実験

世界中の研究者がベンチマークとして用いている Johns Hopkins 大学の Hopkins155 データベース*1[24] を検討し

*1 <http://www.vision.jhn.edu/data/hopkins155>

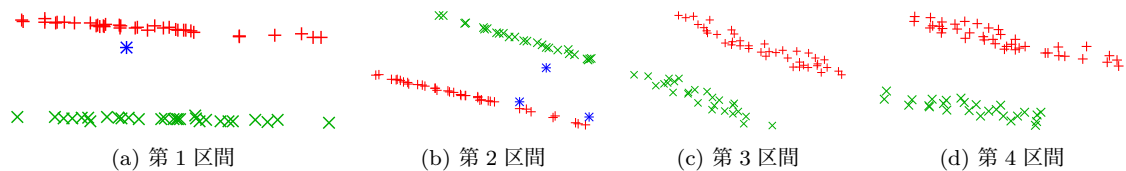


図 7 図 5(a) の特徴点の部分軌跡の 3 次元可視化.

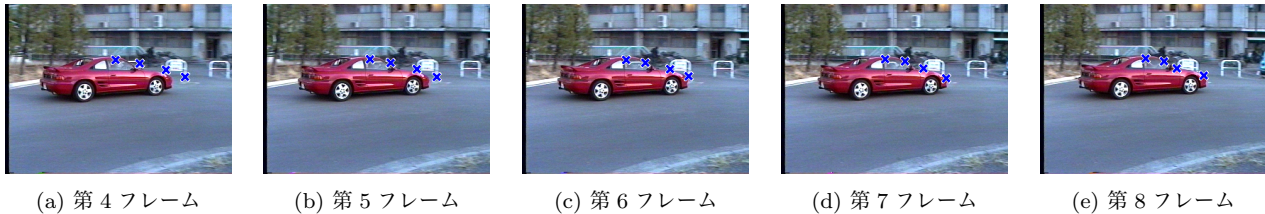


図 8 図 5(a) の特徴点の追跡誤り.

た. まず, このデータベースから 3 例を選び, 新たな特徴点軌跡を追加したデータを使用して, 提案手法の効果を観察する. 実験ではそれぞれの実ビデオ画像について 1 フレームが重なり合う 5 フレーム刻みで小区間に分割して前節の方法を適用した.

実験に使用した 3 例からそれぞれ 5 フレームを抜き出したものを図 5 に示す. 図中には追跡した特徴点を正しいと判定されたものに □ 印, 誤りと判定されたものに × 印を付けている. 特徴点の誤差の標準偏差は経験的に $\sigma = 1.0$ として実験を行った. 図 6 はそれぞれの画像例の初めの 5 フレームの部分軌跡を 3 次元空間に可視化したものである. これらが平面上に分布していること, および誤った追跡がその平面から離れていることが見て分かる.

次に図 5 のビデオ画像 (a) について各分割フレームごとにどのような判定がなされたかを調べた. 図 7 は 4 つの分割フレームでの部分軌跡を 3 次元空間に可視化したものである. 赤い + 印は 1 回目の平面当てはめで正しいと判定された点, 緑の × 印は 2 回目の平面当てはめで正しいと判定された点, 青の * 印は誤りと判定された点を表す.

実際にビデオ画像を注意深く調べると, 運動する車上の特徴点 (図 8 左端の特徴点) の追跡位置が第 5 フレームでずれていることがわかる. 第 6 フレーム以降は第 5 フレームでずれた位置を正しく追跡しているため, 誤りとは判定されなかった. 第 6 フレームから第 10 フレームまでの区間で検出された誤り (図 8 右端から三つの特徴点) については, このフレーム区間で三つの背景点が生じて隠蔽されて, それ以降は車上的特徴点として追跡されている. このことから第 1 フレームから第 5 フレームの区間では正しく背景点として追跡されていたため正しいと判定されていたが, 第 6 フレームから第 10 フレームの区間で追跡位置がずれたため誤りと判定され, それ以降のフレームでは車上的特徴点として追跡が進んだため正しいと判定された.

このように追跡が正しいかどうかを直接にビデオ画像を眺めても判定するのは非常に困難である. 我々の手法は単

に軌跡が正しいか正しくないかを自動的に判定するだけでなく, どのフレームの近辺で誤りが生じたのかも知ることができる. 比較のために, 従来からよく行われている軌跡の 2M 次元空間に 4 次元部分空間を当てはめる RANSAC を行うと我々が検出した誤りの 50% は検出できず, 3 次元アフィン空間を当てはめる RANSAC によっても同様に 50% は検出できなかった. これからも 3 次元空間に限定した RANSAC が不可欠であることが分かる.

次に, Hopkins155 データベースの中の自然のシーンを撮影したビデオ画像 (CG で作成した人工画像は今回は考慮しない) のうち 2 個の運動 (移動する物体と移動する背景) から成るもの 35 個について, 提案手法による誤追跡の検出と目視による確認により誤追跡を除去した. 誤追跡を除去する前の 35 個のデータに対して, Sugaya-Kanatani の方法 [20] によって運動分離を行うと, 100% の正解率が得られなかったシーンが 3 個あった. この 3 例に対する分離結果を図 9 に示す. また, この 3 例から誤追跡を除去したデータを入力として, 運動分離を行った結果が図 10 である. 図 9, 図 10 の括弧内の数値は運動分離の正解率を表し, 3 例中の 2 例 (図 10(a), (c)) については誤追跡を除去することで 100% の運動分離が行えた. すなわち, これらについてはアルゴリズムに問題はなく, 低性能評価の原因はデータベースにあったことが分かる.

一方, 図 10(b) に対して運動分離が正しく行えない原因を調べたところ, 初期分類の結果が悪く, それ以降の多段階最適化によっても初期解が更新されなかったことがわかった. Sugaya-Kanatani の多段階最適化で用いられている EM アルゴリズムの精度は初期値に大きく依存することから, 初期分類手法を改良することで, さらに運動分離の精度が向上する余地があることがわかる. Sugaya-Kanatani の方法の初期分類は, Vidal ら [25][26] の一般主成分分析 (GPCA) を用いた 2 平面の同時当てはめを用いているが, 試しにこれを本研究で用いた RANSAC による 1 平面ずつの当てはめに置き換えたところ, 100% の運動分離結果が得

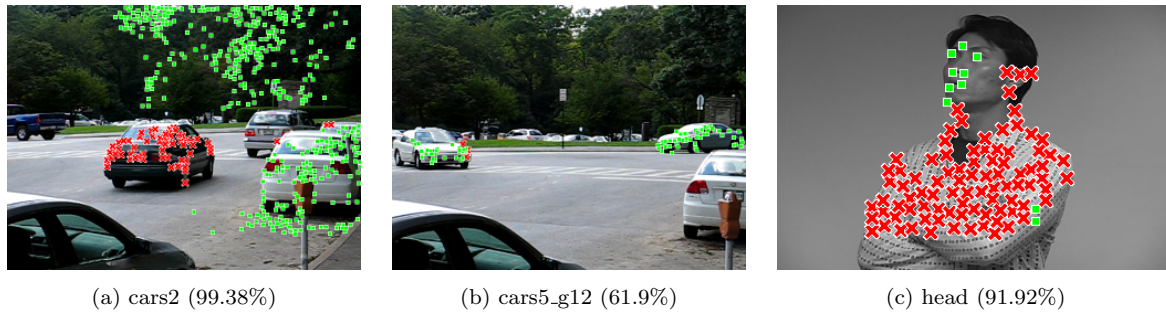


図 9 誤追跡の除去前の運動分離結果. 括弧内は運動分離の正解率.



図 10 誤追跡の除去後の運動分離結果. 括弧内は運動分離の正解率.

られることを確認した. すなわち, この場合にはデータに問題はなく, アルゴリズムが不十分であったことがことがわかる. そして, その考察からアルゴリズムの性能を向上させることができた.

最近新しい手法を提案しては Hopkins155 によって全体的な正解率でその性能を評価, 比較する傾向が見られる. しかし, それでは問題がデータにあるのかアルゴリズムにあるのかが明らかでなく, アルゴリズムの向上につながらない. 正しい評価には信頼できるデータが必要である. そこで, 本論文の方法で 誤追跡を除去した Hopkins155 の 35 個のデータをデータベース化した*2.

8. まとめ

本論文では, ビデオ画像上で追跡した複数運動物体の特徴点軌跡から誤追跡を検出する方法を提案した. その原理は個々の運動が画像面内で並進であれば特徴点の M フレーム間の軌跡は $2M$ 次元空間の平行な 2 次元アフィン空間上に載るという事実である. 平行な 2 次元アフィン空間を含む 3 次元空間が存在する. したがって $2M$ 次元空間をその 3 次元空間に射影すれば, ディスプレイ上で $2M$ 次元空間の軌跡が可視化される. 運動が厳密に並進でなくても, 並進に近ければ (通常のシーンではほとんどそうである), 可視化した 3 次元空間では軌跡は平行に近い曲面上に載っているのだから, 追跡の誤りが一目で分かるという利点がある. これが本論文に示した第 1 点である.

従来の運動分離の研究は抽象的な数学的手段で新たな手法を提案しては Hopkins155 データベースに対する正解率

のみによって性能を評価, 比較する傾向があり, なぜ誤りの原因に関する個別の検討はほとんどなされていなかった. 本論文では現実のデータ生成のメカニズムを考慮し, 従来の $2M$ 次元空間での RANSAC によるアウトライア除去は有効ではなく, 3 次元空間内での検討が必要ことが明らかにした.

さらに, 並進でない運動もビデオ画像系列を短いフレーム数の小区間に分割すれば, 各小区間ではそれぞれの運動はほぼ並進とみなせることに着目して, 小区間ごとに可視化した 3 次元空間で RANSAC による平面当てはめを繰り返すことによって追跡の誤りを自動的に検出する方法を示した. これにより, 直接にビデオ画像を眺めても判定するのが困難な誤りが検出できる. さらに, どのフレームの近辺で誤りが生じたかも知ることが出来る. そして, このようにして Hopkins155 の画像から誤追跡を除去し, それをデータベース化した. これは今後の運動分離の研究に不可欠なものになると期待される.

謝辞 本研究の一部は日本学術振興会科学研究費 (若手研究 B 23700202, 挑戦的萌芽研究 24650086) の助成によった.

参考文献

- [1] B. Cheng, J. Yang, S. Yan, Y. Fu and T. S. Huang, Learning with l^1 -graph for image analysis, *IEEE Trans Patt. Anal. Mach. Intell.*, **19**-4, pp. 858–866, April 2010.
- [2] J. P. Costeira and T. Kanade, A multibody factorization method for independently moving objects, *int. J. Computer Vision*, **29**-3, pp. 159–179, Sept. 1998.
- [3] E. Elhamifar and R. Vidal, Sparse subspace clustering,

*2 <http://www.iim.cs.tut.ac.jp/T-Hopkins/>

- Proc. IEEE Conf. Comput. Vision Patt. Recog.*, pp. 2790–2797, June 2009, Miami FL, U.S.A.
- [4] Z. Fan, J. Zhou and Y. Wu, Multibody grouping by inference of multiple subspace from high-dimensional data using oriented-frames, *IEEE Trans Patt. Anal. Mach. Intell.*, **28**-1 pp. 91–105, 2006.
- [5] M. A. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM*, Vol. 24, No. 6, pp. 381–395, 1981.
- [6] C. W. Gear, Multibody grouping from motion images, *Int. J. Comput. Vision*, **29**-2, pp. 133–150, 1998.
- [7] A. Gruber and Y. Weiss, Multibody factorization with uncertainty and missing data using the EM algorithm, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, Vol. 1, pp. 769–775, June–July 2004, Washington DC, U.S.A.
- [8] 市村直幸, 形状空間への直交射影行列と判別基準を用いた複数運動の分割, 情報処理学会研究報告, 2000-CVIM-120-3, pp. 17–24, Jan. 2000.
- [9] 市村直幸, 富田文明, 形状行列からの特徴選択に基づく動きの分割, 電子情報通信学会論文誌 D-II, J81-D-II-12, pp. 2757–2766, Dec. 1998.
- [10] 井上光平, 浦浜喜一, クラスタリングによる動画画像中の複数物体の分離, 電子情報通信学会技術研究報告, PRMU2000-45, pp. 29–36, July 2000.
- [11] 黒澤典義, 金谷健一, “部分空間分離法とモデル選択による運動物体の分離”, 情報処理学会研究報告, 2000-CVIM-124-4, pp. 25–32, Nov. 2000.
- [12] 黒澤典義, 金谷健一, “アフィン空間分離法による運動物体の分離”, 情報処理学会研究報告, 2001-CVIM-125-3, pp. 25–32, Mar. 2001.
- [13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu and Y. Ma, Robust recovery of subspace structure by low-rank representation, *IEEE Trans Patt. Anal. Mach. Intell.*, **35**-1, pp. 171–184, 2013.
- [14] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang and S. Yan, Robust and efficient subspace segmentation via least squares regression, *Proc. Euro. Conf. Comput. Vision.*, Vol. 7, pp. 347–360, October 2011, Firenze, Italy.
- [15] S. R. Rao, R. Tron, R. Viadl and Y. Ma, Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, June 2008, Anchorage, AK, U.S.A.
- [16] K. Schindler, D. Suter and H. Wang, A model-selection framework for multibody structure-and-motion of image sequences, *Int. J. Comput. Vision*, **79**-2 pp. 159–177, 2008.
- [17] Y. Sugaya and K. Kanatani, Outlier removal for motion tracking by subspace separation, *IEICE Trans. Information and Systems*, Vol. E86-D, No. 6 pp. 1095–1102, 2003.
- [18] 菅谷保之, 金谷健一, 複数運動分離の教師なし学習による多段階最適化, 情報処理学会研究報告, 2003-CVIM-138-25, pp. 185–192, May 2003.
- [19] Y. Sugaya and K. Kanatani, Multi-stage optimization for multi-body motion segmentation, *IEICE Transactions on Information and Systems*, Vol. E87-D, No. 7, pp. 1935–1942, July 2004.
- [20] Y. Sugaya and K. Kanatani, Improved multistage learning for multibody segmentation, *Proc. 5th Int. Conf. Computer Vision Theory Applications (VISAPP 2010)*, May 2010, Angers, France, Vol. 1, pp. 199–206.
- [21] 田中慎也, 田中勇桂, 原裕貴, 菅谷保之, 金谷健一, 一般主成分分析による複数運動分離の多段階最適化, 情報処理学会研究報告, 2009-CVIM-168-8, 2009.
- [22] B. Tordoff and D. W. Murray, Guided-MLESAC: Faster image transform estimation by using matching priors, *IEEE Trans. Patt. Anal. Mach. Intell.*, **27**-10, pp. 1523–1535, 2005.
- [23] P. H. S. Torr and A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, *Computer Vision and Image Understanding*, **78**-1, pp. 138–156, 2000.
- [24] R. Tron and R. Vidal, A benchmark for the comparison of 3-D motion segmentation algorithms, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, June 2007, Minneapolis, MN, U.S.A.
- [25] R. Vidal, Y. Ma and S. Sastry, Generalized principal component analysis (GPCA), *IEEE Trans. Patt. Anal. Mach. Intell.*, **27**-12, pp. 1945–1959, 2005.
- [26] R. Vidal, R. Tron and R. Hartley, Multiframe motion segmentation with missing data using PowerFactorization and GPCA, *Int. J. Comput. Vision*, **79**-1 pp. 85–105, 2008.
- [27] Y. Wu, Z. Zhang, T. S. Huang and J. Y. Lin, Multibody grouping via orthogonal subspace decomposition, sequences under affine projection, *Proc. IEEE Conf. Computer Vision Pattern Recog.*, Vol. 2, pp. 695–701, December 2001, Kauai, HI, U.S.A.
- [28] J. Yan and M. Pollefeys, A general framework for motion segmentation: Independent, articulate, rigid, non-rigid, degenerate and nondegenerate, *Proc. Euro. Conf. Comput. Vision.*, Vol. 4, pp. 94–104, May 2006, Graz, Austria.