情報処理学会研究報告 CVIM, 2002-CVIM-134-2, Sept. 2002, pp. 9–16.

運動物体分離のためのカメラモデルの自動選択

菅谷 保之　　金谷 健一

岡山大学工学部情報工学科

{sugaya,kanatani}@suri.it.okayama-u.ac.jp

移動カメラで撮影したビデオ画像上で特徴点を追跡して得た軌跡データからシーン中の背景や移動物体を分離する各種のアルゴリズムが提案されているが，その性能は仮定したカメラモデルの妥当性に依存する．本論文では幾何学的 AIC を用いて最も適したモデルを自動的に選択する方法を提案する．そして実シーンのビデオ画像を用いて，選ばれたモデルに基づく分離を行えば分離の精度が向上することを確認する．また，特徴点の軌跡データを選択したモデルに対応する部分空間に射影することによって，それらをフレーム数に無関係な低次元ベクトルに変換できることを示す．これは長いビデオ系列に対する計算時間の削減に極めて有効である．

キーワード: 運動物体の分離, 部分空間分離法, アフィン変換分離法, カメラモデル, モデル選択, 幾何学的 AIC

Automatic camera model selection for multibody motion segmentation

Yasuyuki Sugaya and Kenichi Kanatani

Department of Information Technology, Okayama University, Okayama 700-8530 Japan

{sugaya,kanatani}@suri.it.okayama-u.ac.jp

We study the problem of segmenting independently moving objects from moving backgrounds in a video sequence by tracking feature points over consecutive frames. Several algorithms exist for classifying the trajectories of the feature points into independent motions, but the performance depends on the validity of the underlying camera imaging model. In this paper, we present a scheme for automatically selecting the best model by using the geometric AIC. Using video images of real scenes, we confirm that the segmentation accuracy is indeed improved by the use of the alogorithm based on the selected model. We also show that we can compress the trajectory data into low-dimensional vectors, independently of the frame number, by projecting them onto the subspace associated with the selected model. This is very effective for reducing the computation time for a long video sequence.

Key words: motion segmentation, subspace separation, affline space separation, camera model, model selection, geometric AIC

1. まえがき

ビデオ画像からシーン中を移動する物体が存在するかどうかを判定し，存在すればそれを背景から分離することは多くのビデオ画像処理の基礎である．カメラが静止している場合には背景差分やフレーム間差分に基づく各種の手法が開発されているが，移動する偵察車両からの撮影のようにカメラ自身が移動する場合は，移動する背景から独立に運動する物体を検出する必要があり，処理が極めて困難となる．

これに対してオプティカルフローや色やテクスチャなどに基づく多くの手法が研究されているが，Costeira・Kanade [1] は特徴点の軌跡の分類をその画像座標の履歴から計算されるある行列の要素の非零判定に帰着させた．その後，これをもとに種々の応用や改良が研究された [2, 5, 9, 13, 14, 17] ．例えば Gear [2] は行列のある標準形を用いてグラフのマッチングに帰着させた．市村・富田 [5] は行列の QR 分解を用い，市村 [4] は大津の判別規準 [18] を適用した．井上・浦浜 [9] はファジクラスタリングを用いた．

黒澤・金谷 [13, 14] は Costeira-Kanade の方法が「アフィンカメラのもとでは剛体運動する物体上の特徴点の画像上の運動履歴がある 4 次元部分空間に拘束される」という事実の帰結であること指摘し，幾何学的 AIC によるモデル選択 [11] と最小メジアン法によるロバスト推定を組み合わせた「部分空間分離法」[13] および「アフィン空間分離法」[14] と呼ぶ分離手法を示した．牧ら [16, 15] はこの考え方を同じ数学的構造を持つ運動物体の照度変化の解析に適用している．

運動物体の分離において物体個数の推定は困難な

問題であり，ほとんどの研究では発見的なしきい値処理が用いられている．黒澤・金谷 [13, 14] の方法では物体の個数を既知としているが，幾何学的 AIC および幾何学的 MDL を用いたモデル選択による推定も試みられている [12]．

特徴点の追跡には Kanade-Lucas-Tomasi の方法 [21] がよく用いられるが，誤った追跡を防ぐために市村ら [6, 7, 8] は特徴点追跡の改良や特徴点軌跡の非線形フィルタリングを試みている．Huynh ら [3] や菅谷・金谷 [20] は特徴点軌跡に適切な部分空間を当てはめることによって誤まった追跡を除く方法を示している．

本論文では物体の個数が与えられ，誤った追跡も除去されているという条件のもとで黒澤･金谷 [13, 14] の分離方法の性能をさらに向上させる手法を提案する．

黒澤・金谷 [13, 14] によれば，同じ物体に属する特徴点の軌跡は 4 次元部分空間に含まれるだけでなく，その中のより狭い 3 次元アフィン空間に含まれる．また物体が画像内を 2 次元的に移動するとさらに狭い 3 次元部分空間および 2 次元アフィン空間に含まれる．理論上は強い拘束を利用するほど分離の性能が向上するはずであるが，実際には画像の誤差やカメラのモデル化の誤差のために必ずしもそうでないことが指摘されている [14]．

そこで本論文ではどの空間を用いるのがよいかを幾何学的 AIC によるモデル選択で判定する試みを示す．そして実シーンのビデオ画像を用いて，選択された空間に基づく分離を行えば精度が実際に向上することを確認する．さらに，選ばれた空間にデータを射影することによってデータの次元が削減できることを示す．この結果，計算コストはビデオフレーム数に無関係に特徴点数にのみ依存し，長いビデオ系列に対する計算速度が著しく向上する．

以下，提案手法の基本原理を述べ，ビデオ画像を用いた実験によりその有効性を示す．

2. 特徴点の軌跡

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 p_α の画像座標を $(x_{\kappa\alpha}, y_{\kappa\alpha})$, $\kappa = 1, ..., M$, $\alpha = 1, ..., N$ とする．そして α 番目の特徴点の運動履歴を次の $2M$ 次元ベクトルで表す．

$$\boldsymbol{p}_\alpha = (x_{1\alpha}\ y_{1\alpha}\ x_{2\alpha}\ y_{2\alpha} \cdots\ x_{M\alpha}\ y_{M\alpha})^\top \quad (1)$$

これによって各特徴点の軌跡を $2M$ 次元空間の 1 点と同一視できる．

シーン中の物体に任意の 3 次元物体座標系を固定し，特徴点 p_α のその物体座標系に関する 3 次元座標を $(a_\alpha, b_\alpha, c_\alpha)$ とする．時刻 κ での 3 次元物体座標系の原点の位置と各座標軸の 3 次元基底ベクトルをそれぞれ $\boldsymbol{t}_\kappa$, $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ とすると，特徴点 p_α の時刻 κ における 3 次元位置 $\boldsymbol{r}_{\kappa\alpha}$ は次式で表せる．

$$\boldsymbol{r}_{\kappa\alpha} = \boldsymbol{t}_\kappa + a_\alpha \boldsymbol{i}_\kappa + b_\alpha \boldsymbol{j}_\kappa + c_\alpha \boldsymbol{k}_\kappa \quad (2)$$

平行投影や弱透視投影，疑似透視投影を一般化したアフィンカメラ [10] のもとでは，3 次元点 $\boldsymbol{r}_{\kappa\alpha}$ は画像上の次の 2 次元位置に投影される．

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \boldsymbol{A}_\kappa \boldsymbol{r}_{\kappa\alpha} + \boldsymbol{b}_\kappa \quad (3)$$

ただし $\boldsymbol{A}_\kappa, \boldsymbol{b}_\kappa$ はそれぞれ時刻 κ 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 [14]．式 (2) を代入すると，式 (3) は次のように書ける．

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

$\tilde{\boldsymbol{m}}_{0\kappa}, \tilde{\boldsymbol{m}}_{1\kappa}, \tilde{\boldsymbol{m}}_{2\kappa}, \tilde{\boldsymbol{m}}_{3\kappa}$ は時刻 κ でのカメラの位置や内部パラメータで決まる 2 次元ベクトルである．これを $\kappa = 1, \cdots, M$ に渡って式 (1) のように縦に並べると，特徴点の軌跡を表す $2M$ 次元ベクトル $\boldsymbol{p}_\alpha$ は次のように書ける．

$$\boldsymbol{p}_\alpha = \boldsymbol{m}_0 + a_\alpha \boldsymbol{m}_1 + b_\alpha \boldsymbol{m}_2 + c_\alpha \boldsymbol{m}_3 \quad (5)$$

$\boldsymbol{m}_i$, $i = 0, 1, 2, 3$ は $\tilde{\boldsymbol{m}}_{i\kappa}$ を $\kappa = 1, \cdots, M$ に渡って縦に並べた $2M$ 次元ベクトルである．

3. 運動観測のカメラモデル

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 $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$ の張る「4 次元部分空間」に含まれることを意味する．ゆえに，観測した特徴点を異なる剛体物体の運動に分離するには，それらの軌跡データ $\{\boldsymbol{p}_\alpha\}$ を互いに異なる 4 次元部分空間に分類すればよい．これが部分空間分離法 [13] の原理である．

しかし，式 (5) をよく見ると，$\boldsymbol{m}_0$ の係数はすべての α に共通に 1 である．したがって $\boldsymbol{p}_\alpha$ は $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$ の張る 4 次元部間内のある「3 次元アフィン空間」に含まれる．ゆえに特徴点の運動を分離するには軌跡データ $\{\boldsymbol{p}_\alpha\}$ を互いに異なる 3 次元アフィン空間に分類すればよい．これがアフィン空間分離法 [14] の原理である．

理論的にはより強いアフィン空間の拘束条件を用いるほうがよい分離ができるはずであるが，式 (5) は

アフィンカメラモデルに基づいている．実際のカメラは透視投影であり，これをアフィンカメラで近似するモデル化誤差はアフィン空間分離法のほうが大きい．一般に，用いる拘束条件が強いほどモデル化誤差の影響が大きくなり，拘束条件が弱いほどモデル化誤差にロバストになる．黒澤・金谷 [14] によれば，部分空間分離法とアフィン空間分離法のどちらがよいかはカメラのモデル化誤差と特徴点軌跡の誤差とのバランスで定まり，透視効果が強く特徴点軌跡の精度が高い場合は部分空間分離法が，透視効果が弱く特徴点軌跡の精度が低い場合はアフィン空間分離法がそれぞれ適している．しかし，撮影したビデオ画像がどちらに属するかは事前には未知である．

一方，物体が 2 次元的な運動する場合，すなわち画像内での並進や拡大縮小や画像に垂直な軸の周りに回転する場合は $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ の一つを恒等的に $\boldsymbol{0}$ にとってよいから，$\boldsymbol{p}_\alpha$ はある 3 次元部分空間に含まれる．しかし $\boldsymbol{m}_0$ の係数が恒等的に 1 であるから，さらにその中の 2 次元アフィン空間に含まれる．ゆえに 2 次元的運動する物体の運動を分離するには，軌跡データ $\{\boldsymbol{p}_\alpha\}$ を互いに異なる 3 次元部分空間あるいは 2 次元アフィン空間に分類すればよい．しかし，物体の運動が 2 次元的かどうか，およびアフィンモデルが適しているかどうかは事前には未知である．

以上の考察から，撮影したビデオ画像に対してどの空間 (以下カメラモデルと呼ぶ) を仮定するのがよいかを判定できれば，そのモデルに基づく分離を行うことによって分離の精度が向上すると期待される．本論文ではこれを幾何学的 AIC によるモデル選択 [11] を適用し，その効果を確認する．

4. カメラモデルの選択

シーン中を 1 個の物体が運動すれば，独立な運動は背景と移動物体の 2 種類であり，特徴点軌跡を表す $2M$ 次元ベクトルは $2M$ 次元空間の二つの 4 次元部分空間または 3 次元アフィン空間に含まれる．したがって全特徴点の軌跡データ $\{\boldsymbol{p}_\alpha\}$, $\alpha = 1, ..., N$ はそれらを含む 8 次元部分空間 $\mathcal{L}^8$ または 7 次元アフィン空間 $\mathcal{A}^7$ に属す[1]．

さらに背景も移動物体も 2 次元的な運動をする場合は，特徴点軌跡が二つの 3 次元部分空間または 2 次元アフィン空間に属するから，全特徴点軌跡はそれらを含む 6 次元部分空間 $\mathcal{L}^6$ または 5 次元アフィン空間 $\mathcal{A}^5$ に属す．

[1] n_1 次元部分空間と n_2 次元部分空間を含む最小の部分空間の次元は $n_1 + n_2$ であるが，m_1 次元アフィン空間と m_2 次元アフィン空間を含む最小のアフィン空間の次元は $m_1 + m_2 + 1$ である．

以上より次の包含関係をもつ 4 個のカメラモデルが候補となる (← の右側が左側の部分集合)．

$$\begin{array}{ccc} & \mathcal{L}^6 & \\ \swarrow & & \searrow \\ \mathcal{L}^8 & & \mathcal{A}^5 \\ \nwarrow & & \swarrow \\ & \mathcal{A}^7 & \end{array} \tag{6}$$

運動が m 種類ある場合は $\mathcal{L}^8$, $\mathcal{A}^7$, $\mathcal{L}^6$, $\mathcal{A}^5$ がそれぞれ $\mathcal{L}^{4m}$, $\mathcal{A}^{4m-1}$, $\mathcal{L}^{3m}$, $\mathcal{A}^{3m-1}$ となる ($m = 2$ のときが上記の場合)．

5. 幾何学的 AIC のよるモデル選択

モデル選択の素朴な考えは，得られた特徴点軌跡に各空間を当てはめ，残差 (データ点と当てはめた空間との距離の二乗和) が最小となるモデルを選ぶことである．しかし，これでは自由度 (その空間を指定するためのパラメータの最小個数) が最も大きい空間の残差が必ず最小になる．したがって残差と自由度のバランスを考慮しなければならない．これを評価するのに幾何学的 **AIC**[11] を用いる．これは次のように計算される．以下 $n = 2M$ と置く．

n 次元空間の N 点 $\{\boldsymbol{p}_\alpha\}$ に d 次元部分空間 $\mathcal{L}^d$ を最適に (すなわち残差が最小になるように) 当てはめたとき，その残差 $J_{\mathcal{L}^d}$ は，$n \times n$ モーメント行列

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} \boldsymbol{p}_\alpha \boldsymbol{p}_\alpha^\top \tag{7}$$

の固有値を $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ とすると次のように計算できる [14]．

$$J_{\mathcal{L}^d} = \sum_{i=d+1}^{n} \lambda_i \tag{8}$$

このとき幾何学的 AIC は次式で与えられる [11]．

$$\text{G-AIC}_{\mathcal{L}^d} = J_{\mathcal{L}^d} + 2d(N + n - d)\epsilon^2 \tag{9}$$

ただし ϵ は各フレームの各特徴点の各座標に入る誤差の標準偏差であり，ノイズレベルと呼ぶ [11]．

n 次元空間の N 点 $\{\boldsymbol{p}_\alpha\}$ に d 次元アフィン空間 $\mathcal{A}^d$ を最適に当てはめたとき，残差 $J_{\mathcal{A}^d}$ は，N 点 $\{\boldsymbol{p}_\alpha\}$ の重心を $\boldsymbol{p}_C$ とし，その周りの $n \times n$ モーメント行列

$$\boldsymbol{M}' = \sum_{\alpha=1}^{N} (\boldsymbol{p}_\alpha - \boldsymbol{p}_C)(\boldsymbol{p}_\alpha - \boldsymbol{p}_C)^\top \tag{10}$$

の固有値を $\lambda'_1 \geq \lambda'_2 \geq \cdots \geq \lambda'_n$ とすると，次のように計算できる [14]．

$$J_{\mathcal{A}^d} = \sum_{i=d+1}^{n} \lambda'_i \tag{11}$$

このとき幾何学的 AIC は次式で与えられる [11]．

$$\text{G-AIC}_{\mathcal{A}^d} = J_{\mathcal{A}^d} + 2\Big(dN + (d+1)(n-d)\Big)\epsilon^2 \quad (12)$$

各モデルに対して幾何学的 AIC を計算し，それが最も小さいものが最も妥当なモデルとみなす．

6. データの次元圧縮

運動物体の分離にはさまざまな行列演算が必要となる．特徴点軌跡のデータベクトルは M フレームに対して $n\ (=2M)$ 次元ベクトルであり，例えば 100 フレームを追跡すると 200 次元ベクトルとなる．したがってフレーム数 M が大きくなるほどデータの次元が高くなり，それに伴って計算コストも増加する．

しかしデータベクトルはフレーム数 M とは無関係に，運動の個数 m によって決まるある d 次元部分空間 $\mathcal{L}^d$ に含まれ，d はごく小さい値である．実際には誤差のために厳密のその空間に含まれるとは限らないが，分離はその空間内で行われる．したがってすべてのデータをその d 次元部分空間 $\mathcal{L}^d$ に射影し，空間の座標変換を行って，最初の d 本の基底ベクトルが $\mathcal{L}^d$ に含まれるようにすれば，すべてのデータがフレーム数 M に無関係に d 次元ベクトルで表される．例えば $m=2$ の場合はすべてのデータが 8 次元ベクトルとなる．

このようにしてよい理由は，運動物体の分離が「部分空間の分離」という関係のみに基づいているためであり，全空間をどのように線形変換しても部分空間の構造は不変であり，どの座標系で計算しても結果が同じになるためである．

具体的には式 (7) の $n \times n$ モーメント行列の固有値を $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ とし，対応する固有ベクトルの正規直交系を $\{\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_n\}$ とするとき，各 n 次元ベクトル $\boldsymbol{p}_\alpha$ を次の d 次元ベクトル $\tilde{\boldsymbol{p}}_\alpha$ に置きかえればよい（$(\boldsymbol{p}_\alpha, \boldsymbol{u}_i)$ は $\boldsymbol{p}_\alpha$ と $\boldsymbol{u}_i$ の内積を表す）．

$$\tilde{\boldsymbol{p}}_\alpha = \begin{pmatrix} (\boldsymbol{p}_\alpha, \boldsymbol{u}_1) \\ (\boldsymbol{p}_\alpha, \boldsymbol{u}_2) \\ \vdots \\ (\boldsymbol{p}_\alpha, \boldsymbol{u}_d) \end{pmatrix} \quad (13)$$

同様に考えれば，d 次元アフィン空間 $\mathcal{A}^d$ が選ばれたとき，全空間にアフィン変換を施し，座標原点を $\mathcal{A}^d$ 上にとり，最初の d 本の座標基底が $\mathcal{A}^d$ 内にあるようにすれば，各データを d 次元ベクトルで表せるはずである．なぜなら「アフィン空間の分離」という関係は全空間のアフィン変換に対して不変だからである．しかし黒澤・金谷 [14] のアフィン空間分離法はその計算過程で部分空間分離法を併用しているので，アフィン変換によって部分空間の構造が変化してしまう．したがって，黒澤・金谷 [14] のアフィン空間分離法を使う限り，これはできない．

そこで d 次元アフィン空間 $\mathcal{A}^d$ に対しては，それを含む $d+1$ 次元部分空間 $\mathcal{L}^{d+1}$ を用いてデータを $d+1$ 次元ベクトルに圧縮し，それを $\mathcal{A}^d$ に射影したものをデータとする．具体的には $d+1$ 次元部分空間 $\mathcal{L}^{d+1}$ 内で式 (10) の形の $(d+1)\times(d+1)$ モーメント行列 $\boldsymbol{M}'$ を計算し，その固有値を $\tilde{\lambda}'_1 \geq \tilde{\lambda}'_2 \geq \cdots \geq \tilde{\lambda}'_d$，対応する固有ベクトルの正規直交系を $\{\tilde{\boldsymbol{u}}'_1, \tilde{\boldsymbol{u}}'_2, ..., \tilde{\boldsymbol{u}}'_d\}$ とする．そして次の $(d+1)\times(d+1)$ 次元射影行列 $\tilde{\boldsymbol{P}}'_d$ を計算する．

$$\tilde{\boldsymbol{P}}'_d = \sum_{i=1}^{d} \tilde{\boldsymbol{u}}'_i \tilde{\boldsymbol{u}}'^{\top}_i \quad (14)$$

これを用いて各 $d+1$ 次元ベクトル $\tilde{\boldsymbol{p}}_\alpha$ を次の $d+1$ 次元ベクトル $\hat{\boldsymbol{p}}'_\alpha$ に置き換える．

$$\hat{\boldsymbol{p}}'_\alpha = \tilde{\boldsymbol{p}}_C + \tilde{\boldsymbol{P}}'_d(\tilde{\boldsymbol{p}}_\alpha - \tilde{\boldsymbol{p}}_C) \quad (15)$$

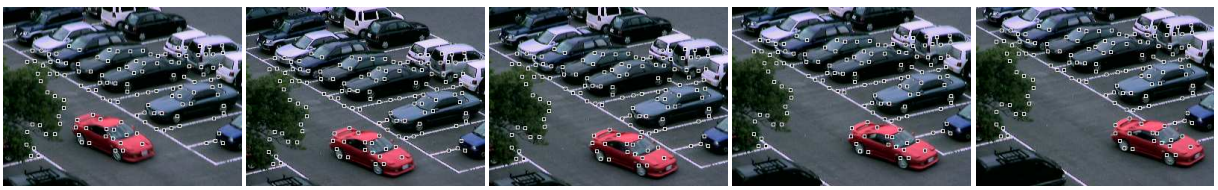
なお，黒澤・金谷の部分空間分離法 [13] とアフィン空間分離法 [14] の計算過程で幾何学的 AIC によるモデル選択が行われ，そこで「余次元」が計算されているが，上記のようにデータを d 次元部分空間 $\mathcal{L}^d$ または d 次元アフィン空間 $\mathcal{A}^d$ に射影した場合は，余次元はもとの n 次元空間ではなくその射影した空間を基準にして計算する (詳細省略)．

7. 計算手順のまとめ

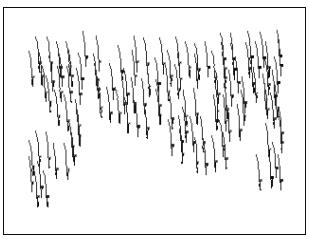
以上を組み合わせてビデオ画像から運動物体を分離する手順をまとめると次のようになる．ただし運動の個数 m は既知とする．

1. Kanade-Lucas-Tomosi の方法 [21] によりビデオ画像から特徴点の追跡を行う．
2. 菅谷・金谷の方法 [20] によって誤った追跡データを除去する[2]．
3. 残った特徴点軌跡に対して $4m$ 次元部分空間 $\mathcal{L}^{4m}$，$4m-1$ 次元アフィン空間 $\mathcal{A}^{4m-1}$，$3m$ 次元部分空間 $\mathcal{L}^{3m}$，および $3m-1$ 次元アフィン空間 $\mathcal{A}^{3m-1}$ を最適に当てはめ，幾何学的 AIC が最小になるモデルを選択する．
4. 選択した空間にデータを射影して次元を圧縮する．
5. 圧縮したデータを用いて，選択したモデルに適合する部分空間分離法 [13] またはアフィン空間分離法 [14] により運動物体の分離を行う[2]．

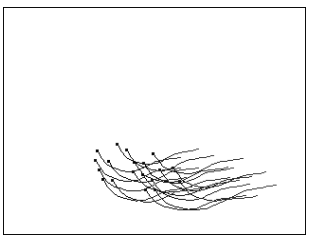
[2]以下にプログラムを公開している．
http://www.suri.it.okayama-u.ac.jp/program.html



モデル	$\mathcal{L}^8$	$\mathcal{A}^7$	$\mathcal{L}^6$	$\mathcal{A}^5$
G-AIC	836.9	779.1	688.9	**631.1**



背景点の軌跡



物体点の軌跡

手法	Costeira-Kanade	市村	Shi-Malik	$\mathcal{L}^8$	$\mathcal{A}^7$	$\mathcal{L}^6$	$\mathcal{A}^5$
正解率 (%)	85.3	92.6	86.8	75.0	86.0	97.7	**100**

図 1: 上段：入力ビデオ画像 (1, 8, 15, 22, 30 フレーム) と追跡した特徴点 (136 個)．中段：各モデルの幾何学的 AIC，および選ばれたモデルによって分離した背景点の軌跡と物体点の軌跡．下段：各分離手法による分離の正解率．

8. 実ビデオ画像実験

実ビデオ画像を用いて前節のカメラモデルの選択実験を実行した例を示す．

図 1 の上段はカメラを移動しながら 30 フレームに渡って撮影したビデオ画像 (320×240 画素) から 5 フレームを抜き出したものである．画像中に最終フレームまで正しく追跡された 136 個の特徴点を □ 印で示す．

この 136 個の特徴点軌跡に対して 8 次元部分空間 $\mathcal{L}^8$，7 次元アフィン空間 $\mathcal{A}^7$，6 次元部分空間 $\mathcal{L}^6$，5 次元アフィン空間 $\mathcal{A}^5$ をそれぞれ最適に当てはめ，幾何学的 AIC を計算したものが中段左の表である．

式 (9), (12) の幾何学的 AIC の計算にはノイズレベル ϵ が必要である．これは各特徴点の各フレームでの位置の誤差が独立なら，最も一般のモデル $\mathcal{L}^8$ を当てはめ残差から推定できる [11]．しかし実際にはフレーム間の相関が強く，どのフレームでも正確に追跡される明確な特徴点どのフレームでも多少の変動を伴うあいまいな特徴点の二種類に分かれてしまう [20]．このため通常の推定方法が適用できないので，ここでは経験値として $\epsilon = 0.5$(画素) とし，以下の実験例についてもこの値を用いた[3]．そして，この ϵ の値を 0.1 から 1.0 の範囲で変化させても選ばれるモデルは変化しないことが確認された．

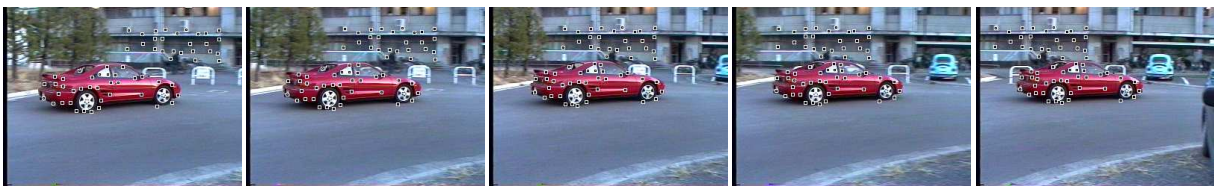
表からわかるように，この場合は 5 次元アフィン空間 $\mathcal{A}^5$ が選ばれる．この例は遠景をズームカメラで撮影したものであり，運動物体と背景の動きは画像内での剛体運動に近いので，$\mathcal{A}^5$ が選ばれたことは妥当といえる．このモデルを仮定してアフィン空間分離法で分離した背景点の軌跡と物体点の軌跡を中段右に示す．

下段はそれ以外の方法で分離した結果の正解率 (= 正しく分離された特徴点数/ 全特徴点数) との比較である．「Costeira-Kanade」は Costeira・Kanade [1] にあるように，「作用行列」(付録 A) の行と列を絶対値の大きいの要素がまとまるように入れ換えて近似的にブロック対角行列に変換する方法である．「市村」は市村 [4] に従って作用行列の各行に「大津の判別規準」[18] を適用し，最も高い判別規準を与える行を用いて分離したものである．「Shi-Malik」は特徴点を頂点とする完全グラフを考え，作用行列の各要素の絶対値を対応する枝の重みとし，Shi・Malik [19] の「正規化カット最小化」(付録 B) によって分離したものである．井上・浦浜 [9] のファジクラスタリングも類似の考え方によっている．$\mathcal{L}^8$, $\mathcal{A}^7$, $\mathcal{L}^6$, $\mathcal{A}^5$ はそれぞれのモデルに対応する部分空間分離法およびアフィン空間分離法を表す．予想通り，$\mathcal{A}^5$ のよるアフィン空間分離法のみが完全に正しい分離を与えた．

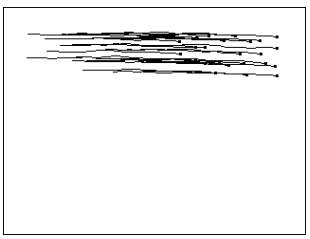
図 2 は別の例であり，17 フレームに渡って 63 個の特徴点を追跡したものである．このシーンも遠景をズームカメラで撮影したものであり，運動物体と背景の動きが画像内での剛体運動に近いので，モデルとして 5 次元アフィン空間 $\mathcal{A}^5$ が選ばれている．そして，このモデルによる分離が他の手法に比べて最もよい結果を与えている．

図 3 は物体に近い位置から物体を回り込みように

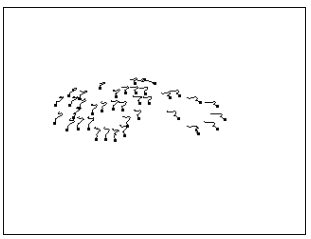
[3] 菅谷・金谷 [20] のアウトライア除去でもこの値が用いられている．



モデル	$\mathcal{L}^8$	$\mathcal{A}^7$	$\mathcal{L}^6$	$\mathcal{A}^5$
G-AIC	434.6	413.6	398.7	**379.8**



背景点の軌跡



物体点の軌跡

手法	Costeira-Kanade	市村	Shi-Malik	$\mathcal{L}^8$	$\mathcal{A}^7$	$\mathcal{L}^6$	$\mathcal{A}^5$
正解率 (%)	57.1	57.1	57.1	92.0	61.9	61.9	**100**

図 2: 上段 : 入力ビデオ画像 (1, 5, 9, 13, 17 フレーム) と追跡した特徴点 (63 個) . 中段 : 各モデルの幾何学的 AIC , および選ばれたモデルによって分離した背景点の軌跡と物体点の軌跡 . 下段 : 各分離手法による分離の正解率 .

表 1: 実行時間の比較

	図 1	図 2	図 3
フレーム数	30	17	100
特徴点数	136	63	73
実行時間 (秒)	373	5	12
短縮率 (%)	94.7	71.4	15.2

カメラを移動して 100 フレームを撮影し , 73 個の特徴点を追跡したものである . この場合は 8 次元部分空間 $\mathcal{L}^8$ が選ばれている . これは物体を近距離から撮影したことにより透視効果が強くなったためと考えられる . そしてこのモデルに基づく部分空間分離法によって最もよい分離結果が得られている .

データの次元の圧縮による計算コストの変化を調べたものが表 1 である . ここでは図 1, 2, 3 の例において , 選択されたモデルにかかわらず共通にデータを 8 次元ベクトルに変換して実行時間を比較した .「短縮率 (%)」は (次元の圧縮した場合の実行時間)/(次元を圧縮をしなかった場合の実行時間) を%で表示したものである .

図 1 ではフレーム数はあまり多くないが特徴点数が非常に多いので短縮率は 94.7%に留まっている . それに対して特徴点の少ない図 2 では 71.4%となり , フレーム数の多い図 3 で 15.2%もの短縮になっている . このようにフレーム数が多いほど次元の圧縮の効果が顕著に現れ , 実行時間はほぼ特徴点数 N のみに依存する . この N への依存を厳密に解析するのは困難であるが , 表 1 のデータからは $O(N^5)$ ないし $O(N^6)$ と予想される .

9. まとめ

移動カメラで撮影したビデオ画像上で特徴点を追跡して得た軌跡データからシーン中の背景や移動物体を分離する問題に対して , 複数のカメラモデルを候補とし , 幾何学的 AIC を用いて最も適したモデルを自動的に選択する方法を提案した . そして実シーンのビデオ画像を用いて , 選ばれたモデルに基づく分離を行えば分離の精度が向上することを確認した .

また特徴点の軌跡データを得られたカメラモデルに対応する部分空間に射影すれば , それらをフレーム数に無関係な低次元ベクトルに変換できることを示した . これは長いビデオ系列に対する計算時間の削減に極めて有効である .

謝辞: 実験に協力して頂いた岡山大学大学院生の譲田賢治氏に感謝する . 本研究の一部は文部科学省科学研究費基盤研究 C (2) (No. 13680432) , テレコム先端技術研究支援センター , 栢森情報科学振興財団の助成によった .

参考文献

[1] J. P. Costeira and T. Kanade, A multibody factorization method for independently moving objects, *Int. J. Computer Vision*, **29**-3, 159–179, Sept. 1998.

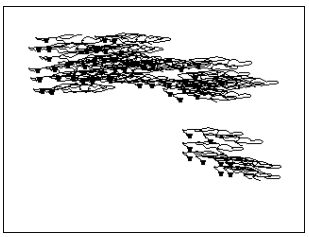
[2] C. W. Gear, Multibody grouping from motion images, *Int. J. Comput. Vision*, **29**-2, 133–150, Aug./Sept. 1998.

[3] D. Q. Huynh and A. Heyden, Outlier detection in video sequences under affine projection, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pp. 695–701, Dec. 2001.

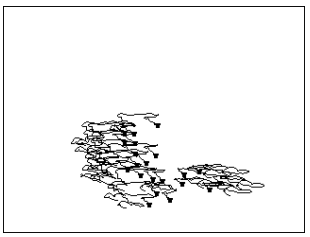
[4] 市村直幸, 形状空間への直交射影行列と判別基準を用いた複数運動の分割, 情報処理学会研究報告, 2000-CVIM-120-3, 17–24, Jan. 2000.

[5] 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, 2757–2766, Dec. 1998.

モデル	$\mathcal{L}^8$	$\mathcal{A}^7$	$\mathcal{L}^6$	$\mathcal{A}^5$
G-AIC	**2117.9**	2281.5	3158.5	3340.1



背景点の軌跡



物体点の軌跡

手法	Costeira-Kanade	市村	Shi-Malik	$\mathcal{L}^8$	$\mathcal{A}^7$	$\mathcal{L}^6$	$\mathcal{A}^5$
正解率 (%)	76.7	58.9	76.7	**93.1**	60.2	57.5	89.0

図 3: 上段：入力ビデオ画像 (1, 25, 50, 75, 100 フレーム) と追跡した特徴点 (73 個)．中段：各モデルの幾何学的 AIC，および選ばれたモデルによって分離した背景点の軌跡と物体点の軌跡．下段：各分離手法による分離の正解率．

[6] 市村直幸, 自己組織化型状態空間モデルを用いた運動軌跡のフィルタリング, 情報処理学会研究報告, 2001-CVIM-128-2, 9–16, July 2001.

[7] 市村直幸, フレーム毎の特徴点抽出に基づく特徴点の追跡, 情報処理学会研究報告, 2001-CVIM-130-5, 31–38, Nov. 2001.

[8] 市村直幸, 生駒哲一, 非ガウス型状態空間状態モデルを用いた特徴点位置系列のフィルタリング, 情報処理学会研究報告, 2000-CVIM-122-3, 17–24, May 2000.

[9] 井上光平, 浦浜喜一, クラスタリングによる動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, 29–36, July 2000.

[10] 金出武雄, コンラッド・ポールマン, 森田俊彦, 因子分解法による物体形状とカメラ運動の復元, 電子情報通信学会論文誌 D-II, **J74-D-II**-8, 1497–1505, Aug. 1993.

[11] 金谷健一, 幾何学的当てはめにおけるモデル選択, 電子情報通信学会論文誌 A, **J84-A**-11, 1385–1393, Nov. 2001.

[12] 金谷健一，黒澤典義，松永力，モデル選択によるランク推定と複数運動の分離, 情報処理学会研究報告，2001-CVIM-126-3, 17–24, Jan. 2001.

[13] 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-124-4，25–32, Nov. 2000.

[14] 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-125-3，25–32, Mar. 2001.

[15] 牧淳人, 服部寛, 輝度部分空間を用いた運動物体の分離, 電子情報通信学会技術研究報告, PRMU2001-134, 147–153, Nov. 2001.

[16] 牧淳人, 渡邊睦, C. Wiles, Geotensity 拘束による 3 次元形状獲得, 電子情報通信学会論文誌 D-II, **J83-D-II**-8, 1714–1752, Aug. 2000.

[17] 長崎健, 川嶋稔夫, 青木由直, 因子分解法に基づく運動画像列解析による多関節物体の構造推定, 電子情報通信学会論文誌 D-II, **J81-D-II**-3, 483–492, March 1998.

[18] 大津展之, 判別および最小 2 乗規準に基づく自動しきい値選定法, 電子通信学会論文誌 D, **J63-D**-4, 349–356, 1980.

[19] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Patt. Anal. Machine Intell.*, **22**-8, 888–905, Aug. 2000.

[20] 菅谷保之, 金谷健一, 部分空間分離法による特徴点追跡のアウトライア除去, 情報処理学会研究報告, 2002-CVIM-133-24, 177–184, May 2002.

[21] C. Tomasi and T. Kanade, Detection and Tracking of Point Features, CMU Tech. Rep. CMU-CS-91-132,Apr. 1991; `http://vision.stanford.edu/~birch/klt/`.

付録 A: 作用行列

n 次元空間 $\mathcal{R}^n$ の N $(> n)$ 点 $\{\boldsymbol{p}_\alpha\}$, $\alpha = 1, ..., N$ が m 個の独立な r 次元部分空間 $\mathcal{L}_i^r$, $i = 1, ..., m$ 内に拘束されているとする．各 $\mathcal{L}_i^r$ には r 個以上の点が含まれているとし，$d = rm$ と置く．

$N \times N$ 計量行列

$$G_{\alpha\beta} = (\boldsymbol{p}_\alpha, \boldsymbol{p}_\beta) \tag{16}$$

の固有値を $\lambda_1 \geq \cdots \geq \lambda_N$，対応する固有ベクトルの正規直交系を $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_N\}$ とするとき，次の $N \times N$ 行列 を作用行列と呼ぶ．

$$\boldsymbol{Q} = \sum_{i=1}^{d} \boldsymbol{v}_i \boldsymbol{v}_i^\top \tag{17}$$

【定理】 点 $\boldsymbol{p}_\alpha$, $\boldsymbol{p}_\beta$ が異なる部分空間に属せば $\boldsymbol{Q}$ の $(\alpha\beta)$ 要素 $Q_{\alpha\beta}$ が 0 となる．

$$Q_{\alpha\beta} = 0, \quad \boldsymbol{p}_\alpha \in \mathcal{L}_i^r, \ \ \boldsymbol{p}_\beta \in \mathcal{L}_j^r, \ \ i \neq j \tag{18}$$

これは Costeira・Kanade [1] の結果を一般化したものであり，次の事実に基づいて証明できる [13]．n 次元空間 $\mathcal{R}^n$ の N $(> n)$ 点は線形従属であるから $\sum_{\alpha=1}^{N} c_\alpha \boldsymbol{p}_\alpha = \boldsymbol{0}$ となるすべてが 0 ではない係数 $\{c_\alpha\}$ が無数に存在する．しかし $\{\boldsymbol{p}_\alpha\}$ が $\mathcal{L}_1 \oplus \mathcal{L}_2 = \mathcal{R}^n$ であるような二つの部分空間 $\mathcal{L}_1$, $\mathcal{L}_2$ に分離されていれば ($\oplus$ は直和)，そのような係数 $\{c_\alpha\}$ はすべて

$\sum_{\mathbf{p}_\alpha \in \mathcal{L}_1} c_\alpha \boldsymbol{p}_\alpha = \mathbf{0}$ となるものと $\sum_{\mathbf{p}_\alpha \in \mathcal{L}_2} c_\alpha \boldsymbol{p}_\alpha = \mathbf{0}$ となるものとから生成される (完全な証明は [13]) .

計量行列 $\boldsymbol{G}$ の固有値を計算するのに式 (7) の $N \times N$ モーメント行列

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} \boldsymbol{p}_\alpha \boldsymbol{p}_\alpha^\top \tag{19}$$

の固有値 $\lambda_1 \geq \cdots \geq \lambda_N$ と対応する固有ベクトルの正規直交系 $\{\boldsymbol{v}_u, ..., \boldsymbol{u}_N\}$ を計算してもよい . $\boldsymbol{G}$ と $\boldsymbol{M}$ は同じランクの半正値対称行列であり , 0 でない固有値は等しく , 0 でない固有値 λ_i に対応する単位固有ベクトルは次のように互いに変換される ($v_{i\alpha}$ は $\boldsymbol{v}_i$ の第 α 成分) .

$$\boldsymbol{v}_i = \frac{1}{\sqrt{\lambda_i}} \begin{pmatrix} (\boldsymbol{p}_1, \boldsymbol{u}_i) \\ \vdots \\ (\boldsymbol{p}_N, \boldsymbol{u}_i) \end{pmatrix}, \quad \boldsymbol{u}_i = \frac{1}{\sqrt{\lambda_i}} \sum_{\alpha=1}^{N} v_{i\alpha} \boldsymbol{p}_\alpha \tag{20}$$

あるいは $n \times N$ 観測行列

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_N \end{pmatrix} \tag{21}$$

を次のように特異値分解[4]してもよい ($\mathrm{diag}(\cdots)$ は $\cdots$ を対角要素とする対角行列) .

$$\boldsymbol{W} = \boldsymbol{U}_{n\times n} \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_n) \boldsymbol{V}_{N\times n}^\top \tag{22}$$

この $\sigma_1^2, ..., \sigma_n^2$ が $\lambda_1, ..., \lambda_n$ に等しく , $\boldsymbol{V}_{N\times n}, \boldsymbol{U}_{n\times n}$ がそれぞれ $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ および $\boldsymbol{u}_1, ..., \boldsymbol{u}_N$ を列とする $N \times n$ および $n \times n$ 行列となっている . N, n の大きさに応じて最も効率的な計算法を選べばよい[5] .

付録 **B: Shi-Maklik** の正規化カット最小化

重みつき無向グラフの N 個の頂点を二つのグループ A, B に分類する問題を考える . 辺の重みを頂点間の類似度とみなし , グループ内の頂点の類似度は大きく , グループ間の頂点の類似度は小さくなるように分割したいとする .

頂点 α, β を結ぶ辺の重みを $W_{\alpha\beta}$ とし , 頂点 α から出ている辺の重みの総和を d_α $(= \sum_{\beta=1}^{N} W_{\alpha\beta})$ とする . 頂点 α がグループ A に属すとき $x_\alpha = 1$, グループ B に属すとき $x_\alpha = -1$ とするとき , Shi-Malik の方法 [19] は次の正規化カットを最小にするように分割するものである .

$$\mathrm{Ncut} = \frac{\sum_{x_\alpha=1, x_\beta=-1} W_{\alpha\beta}}{\sum_{x_\kappa=1} d_\kappa} + \frac{\sum_{x_\alpha=-1, x_\beta=1} W_{\alpha\beta}}{\sum_{x_\kappa=-1} d_\kappa} \tag{23}$$

グループ間の類似度を小さくするにはグループ間の辺の重みの総和 (これをカットと呼ぶ) を最小にすればよいが , 出ている辺の重みの小さい一つの頂点のみが一つのグループとして分割されるなど , 分割のバランスが不自然になる . 式 (23) はグループ間の類似度は小さく , かつ各グループ内の類似度は大きくなるようにカットを正規化したものである .

Shi・Malik [19] はこの正規化カットが次の計算によって最小化できることを示した .

1. 次のように $N \times N$ 対角行列 $\boldsymbol{D}$ を定義する .
$$\boldsymbol{D} = \mathrm{diag}(d_1, ..., d_N) \tag{24}$$
2. $W_{\alpha\beta}$ を $(\alpha\beta)$ 要素とする $N \times N$ を $\boldsymbol{W}$ とする .
3. 次の一般固有値問題の 2 番目に小さい一般固有値 λ に対する N 次元一般固有ベクトル $\boldsymbol{y} = (y_1, ..., y_N)^\top$ を計算する .
$$(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{y} = \lambda \boldsymbol{D}\boldsymbol{y} \tag{25}$$
4. $y_1, ..., y_N$ の最大値を $y_{\max}$, 最小値を $y_{\min}$ とし , 区間 $[y_{\min}, y_{\max}]$ を適当に等分割する . そして各分点 y_* に対して $y_\alpha > y_*$ なら $x_\alpha = 1$, $y_\alpha \leq y_*$ なら $x_\alpha = -1$ とする . これをすべての分点 y_* について調べて , 式 (23) の正規化カットが最小になる y_* を定める .
5. 定めた y_* によってベクトル $\boldsymbol{y} = (y_1, ..., y_N)^\top$ の各要素を ± 1 に 2 値化した N 次元ベクトル $\boldsymbol{x} = (x_1, ..., x_N)^\top$ を返す .

上記のステップ 3 の計算は ,

$$\boldsymbol{D}^{-1/2} = \mathrm{diag}(\frac{1}{\sqrt{d_1}}, ..., \frac{1}{\sqrt{d_N}}) \tag{26}$$

と置き , $N \times N$ 対称行列

$$\boldsymbol{A} = \boldsymbol{D}^{-1/2}(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{D}^{-1/2} \tag{27}$$

の 2 番目に小さい固有値に対する N 次元単位固有ベクトル $\boldsymbol{z}$ を計算し , これに対して

$$\boldsymbol{D}^{1/2} = \mathrm{diag}(\sqrt{d_1}, ..., \sqrt{d_N}) \tag{28}$$

を掛けた次の N 次元ベクトル $\boldsymbol{y}$ を返せばよい .

$$\boldsymbol{y} = \boldsymbol{D}^{1/2}\boldsymbol{z} \tag{29}$$

[4] これが Costeira・Kanade [1] の示した形であるが , 証明が理解しにくい . 理論的には計量行列 $\boldsymbol{G}$ で考えて , 特異値分解は「数値計算の手段」とみなすのがよいと思われる .

[5] よく設計されたライブラリツールなら特異値分解が最も効率的なはずだが , インプレメントの仕方にもよる .