

複数運動分離の教師なし学習による多段階最適化

菅谷 保之 金谷 健一

岡山大学工学部情報工学科

ビデオ画像上を移動する特徴点の軌跡から独立な運動を分離するさまざまな方法が提案されているが、物体や背景は一般の3次元運動をすると仮定するので、平面運動に近い実際のビデオシーンでは分離の精度が劣化する。本論文では退化した運動を仮定する教師なし学習 (EM アルゴリズム) によって分離を最適化し、その結果をさらに一般の運動を仮定する教師なし学習で最適化する。この多段階最適化により、実際のシーンでよく生じる単純な運動に対して極めて高精度であるだけでなく、かなり複雑な3次元運動に対しても精度が保持される。シミュレーションおよび実ビデオ画像実験によれば、提案手法が従来のどの手法よりも優れ、この問題の最終的な解法と考えられる。

Multi-stage Optimization of Multi-body Motion Segmentation by Unsupervised Learning

Yasuyuki Sugaya Kenichi Kanatani

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

Many methods have been proposed for separating feature point trajectories tracked through a video sequence into independent motions, but objects and backgrounds are assumed to be in general 3-D motions in the scene. As a result, the separation accuracy considerably deteriorates in realistic video sequences in which the motions of objects and backgrounds are nearly planar. In this paper, we introduce optimization by unsupervised learning (an EM algorithm) based on a degenerate model followed by unsupervised learning based on the general motion model. This multi-stage optimization allows us not only to separate simple motions that we frequently encounter with high precision but also to preserve the high performance for considerably general 3-D motions. According to simulations and real video experiments, our method is superior to all existing methods and can be a definitive method for this type of problem.

1. まえがき

ビデオ画像上を移動する特徴点の軌跡から独立な運動を分離するさまざまな方法が提案されている。

Costeira・Kanade [1] は軌跡の分類をその画像座標の履歴から計算される「作用行列」の要素の非零判定に帰着させた。その後、これに基づく種々の応用や改良が研究された。Gear [3] は行列の標準形を用いてグラフのマッチングに帰着させ、市村・富田 [5] は行列のQR分解を用い、市村 [4] は大津の判別規準を適用した。井上・浦浜 [6] はファジクラスタリングを用いた。黒澤・金谷 [8, 9] 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させた。Wuら [14] は部分空間の直交分解を適用している。

しかし、これらはどれも物体や背景がそれぞれ異なる一般の3次元運動をすることを仮定している。物体や背景が平面的な運動をしたり、共通の方向の軸回りに回転したりする場合は「退化」と呼ばれ、分離の計算が破綻することが知られている。もちろん実際には厳密な退化が起こることはないが、退化に

近いほど分離の精度が劣化する。

従来の研究では、提案手法の性能を複雑に3次元運動をするシミュレーション画像を用いて評価することが多かった。ところが実際のシーン中を移動する人物や車両を撮影すると、背景や物体が極端な3次元運動をすることはなく、極めて単純な運動であることが多い。

素朴には単純な運動ほど扱いやすいと思われるが実際は反対であり、動きが複雑なほど個々の運動の区別が容易である。それに対して単純な運動ほど退化に近く、正しい分離が困難となる。著者らは前報 [12] でこのような運動の性格の相違を幾何学的AICによるモデル選択によって判別し、それに適した手法を当てはめる試みを行なったが、精度の向上は限定されたものであった。

本論文では、このような実際シーンによく生じる単純な運動に対して、退化した運動を仮定する教師なし学習 (EM アルゴリズム [2] の一種) [10] によって分離を最適化し、その結果をさらに一般の運動を仮定する教師なし学習で最適化する。この多段階最適化により、単純な運動のみならず、相当に複雑な運動に対しても精度が保持される。シミュレーションおよび実ビデオ画像実験によれば、提案手法が従来

†700-8530 岡山市津島中 3-1-1, 岡山大学工学部情報工学科
{sugaya,kanatani}@suri.it.okayama-u.ac.jp

のどの手法よりも優れ、この問題の最終的な解法と考えられる。

以下、2節では画像上の運動の軌跡の満たす拘束条件を示し、3節では教師なし学習による最適化を「非ベイズ型」と「ベイズ型」に分けて説明する。4章では通常起こりやすい退化の運動のモデル化とそれを用いる多段階最適化の方法を述べる。5章ではシミュレーションおよび実ビデオ画像による実験例を示し、6章でまとめる。

2. 軌跡の空間

2.1 軌跡ベクトル

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。そしてその運動履歴を次の $2M$ 次元ベクトルで表し、軌跡ベクトルと呼ぶ。

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

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。便宜上、画像番号 κ を「時刻 κ 」と呼ぶ。

本論文ではカメラの光軸を Z 軸とするカメラ XYZ 座標系をとり、これを世界座標系とする。そして静止したカメラに相対的にシーンが運動すると解釈する。シーン中に 3 次元物体座標系を任意に固定し、特徴点 p_α のそのシーン座標系に関する 3 次元座標を $(a_\alpha, b_\alpha, c_\alpha)$ とする。時刻 κ でのシーン座標系の原点と各座標軸の基底ベクトルを世界座標系 (= カメラ座標系) で表したものをそれぞれ $t_\kappa, \{i_\kappa, j_\kappa, k_\kappa\}$ とすると、特徴点 p_α の時刻 κ における 3 次元位置 $r_{\kappa\alpha}$ は世界座標系では次式となる。

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

2.2 アフィンカメラモデル

平行投影や弱透視投影や疑似透視投影を抽象化したアフィンカメラ [7] は、3 次元点 $r_{\kappa\alpha}$ が次のように画像上に投影されると仮定するものである。

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

ここに A_κ, b_κ 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 [9]。式 (2) を代入すると、式 (3) は次のように書ける。

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

$\tilde{m}_{0\kappa}, \tilde{m}_{1\kappa}, \tilde{m}_{2\kappa}, \tilde{m}_{3\kappa}$ 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。これを時刻 $\kappa = 1, \dots, M$ に渡って式 (1) のように縦に並べると、式 (1) の軌跡ベクトル p_α は次のように書ける。

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

$m_i, i = 0, 1, 2, 3$ は $\tilde{m}_{i\kappa}$ を時刻 $\kappa = 1, \dots, M$ に渡って縦に並べた $2M$ 次元ベクトルである。

2.3 軌跡の拘束条件

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 $\{m_0, m_1, m_2, m_3\}$ の張る「4 次元部分空間」に含まれることを意味する。したがって、観測した特徴点を異なる剛体運動に分離するには、それらの軌跡ベクトル $\{p_\alpha\}$ 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 [8] の原理である。

しかし、式 (5) において m_0 の係数はすべての α に共通に 1 である。このため p_α は $\{m_0, m_1, m_2, m_3\}$ の張る 4 次元部分空間内のある「3 次元アフィン空間」に含まれる。したがって、特徴点の運動を分離するには軌跡ベクトル $\{p_\alpha\}$ 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 [9] の原理である。

2.4 部分空間・アフィン空間分離法

前節の原理に基づく分離のプログラムは Web 上に公開されている¹。これらは Costeira・Kanade [1] の作用行列の持つ情報に幾何学的 AIC によるモデル選択と最小メジアン法によるロバスト推定を組み合わせたものである [8, 9]。

理論的には部分空間拘束条件よりも強いアフィン空間拘束条件を用いるほうが分離の精度が高く、シミュレーション実験を行なっても、誤差がほとんどなく透視効果が極端に強い場合を除けばアフィン空間分離法のほうが部分空間分離法よりも高精度である [9]。しかし実ビデオシーンに適用すると、透視効果がほとんどない場合でもアフィン空間分離法の精度が部分空間分離法とほぼ同じかそれ以下となる場合もある [12]。これは理論やシミュレーション結果と明らかに矛盾する。本論文の解析はその理由を与えるものでもある。

¹<http://www.suri.it.okayama-u.ac.jp/program.html>

3. 教師なし学習による最適化

3.1 非ベイズ型

部分空間分離法またはアフィン空間分離法による分離結果は必ずしも正しいものではない。そこで、その方法で分離された各クラスにそれぞれ4次元部分空間または3次元アフィン空間を最適に当てはめ、各軌跡ベクトルを最も近い空間に帰属させる。そしてこれを収束するまで反復する(図1(a))。

この操作は、画像上で追跡した各特徴点の座標の各成分に独立に、期待値が0で特徴点やフレームによらない一定の分散を持つ正規分布に従う誤差が入ると仮定した場合の最尤推定に基づく教師なし学習と見なせる。なぜなら、このような誤差モデルでは、当てはめた空間までの距離を比較することは尤度(の対数)を比較することと等価だからである。

3.2 ベイズ型

さらに当てはめた空間の内部のデータの分布までモデル化することも考えられる(図1(b))。これも標準的な教師なし学習[10]と同じ考え方であるが、部分空間・アフィン空間拘束条件を仮定すると、確率の計算が複雑になる。アフィン空間拘束条件の場合は次のようになる。

$n = 2M$ と置き、 n 次元空間の N 個の軌跡ベクトル $\{p_\alpha\}$ が m 個のクラスに分類されているとする。各軌跡 α の各クラス k に対する重み $W_\alpha^{(k)}$ を次のように初期化する。

$$W_\alpha^{(k)} = \begin{cases} 1 & \text{軌跡 } \alpha \text{ がクラス } k \text{ に属するとき} \\ 0 & \text{それ以外} \end{cases} \quad (6)$$

そして次のA, Bを $\{W_\alpha^{(k)}\}$ が収束するまで反復する。

A. 各クラス $k = 1, \dots, m$ について次の計算を行なう。

1. クラス k の占有率 $w^{(k)}$ を次のように計算する。

$$w^{(k)} = \frac{1}{N} \sum_{\alpha=1}^N W_\alpha^{(k)} \quad (7)$$

2. クラス k の重心 $p_C^{(k)}$ を次のように計算する。

$$p_C^{(k)} = \frac{\sum_{\alpha=1}^N W_\alpha^{(k)} p_\alpha}{\sum_{\alpha=1}^N W_\alpha^{(k)}} \quad (8)$$

3. クラス k のモーメント行列

$$M^{(k)} = \frac{\sum_{\alpha=1}^N W_\alpha^{(k)} (p_\alpha - p_C^{(k)})(p_\alpha - p_C^{(k)})^\top}{\sum_{\alpha=1}^N W_\alpha^{(k)}} \quad (9)$$

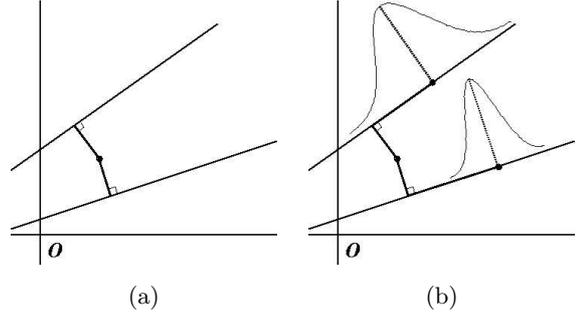


図 1: (a) 非ベイズ型モデル. (b) ベイズ型モデル.

の大きい3個の固有値に対応する単位固有ベクトル $u_1^{(k)}, u_2^{(k)}, u_3^{(k)}$ を計算する。

4. クラス k への射影行列 $P^{(k)}$ とその外側方向への射影行列 $P_\perp^{(k)}$ を次のように計算する (I は単位行列)。

$$P^{(k)} = \sum_{i=1}^3 u_i^{(k)} u_i^{(k)\top}, \quad P_\perp^{(k)} = I - P^{(k)} \quad (10)$$

5. クラス k のアフィン空間の外側に垂直な方向の二乗平均誤差を次のように推定する。

$$\hat{\sigma}^2 = \max\left[\frac{\text{tr}[P_\perp^{(k)} M^{(k)} P_\perp^{(k)}]}{n-3}, \sigma^2\right] \quad (11)$$

ただし $\text{tr}[\cdot]$ は行列のトレース(対角和)であり、 σ は特徴点追跡の誤差である²。

6. クラス k の共分散行列 V を次のように計算する。

$$V^{(k)} = P^{(k)} M^{(k)} P^{(k)} + \hat{\sigma}^2 P_\perp^{(k)} \quad (12)$$

B. 各軌跡 $\alpha = 1, \dots, N$ について次の計算をする。

1. 軌跡 α の各アフィン空間に対する尤度 $P(\alpha|k)$, $k = 1, \dots, m$ を次のように計算する。

$$P(\alpha|k) = \frac{e^{-(p_\alpha - p_C^{(k)}, V^{(k)-1} (p_\alpha - p_C^{(k)})) / 2}}{\sqrt{\det V^{(k)}}} \quad (13)$$

2. 軌跡 α の重み $W_\alpha^{(k)}$, $k = 1, \dots, m$ を次のように更新する(ベイズの定理)。

$$W_\alpha^{(k)} = \frac{w^{(k)} P(\alpha|k)}{\sum_{l=1}^m w^{(l)} P(\alpha|l)} \quad (14)$$

収束したら各軌跡 α を $W_\alpha^{(k)}$, $k = 1, \dots, m$ が最大となるクラス k に分類する。

3.3 教師なし学習

前節の手順では、軌跡の各3次元アフィン空間の内部での分布に期待値 $p_C^{(k)}$, ランク3の共分散行列

²実画像実験では $\sigma = 0.5$ (画素) とした。

$P^{(k)} M^{(k)} P^{(k)}$ の正規分布を当てはめ、その外側へ分布は期待値が0で、分散が場所や方向によらずに一定の正規分布とし、これを軌跡 α がクラス k に分類されたという条件のもとでの「条件付き確率」 $P(\alpha|k)$ としている。そして、クラス k の占有率 $w^{(k)}$ をそのクラスの「事前確率」とし、式 (14) のベイズの定理で各クラスへの分類の「事後確率」 $W_\alpha^{(k)}$ を計算する。これによって分類を更新し、これを収束するまで反復する。分類は確率的であり、軌跡 α は事後確率 $W_\alpha^{(k)}$ (一般に小数) でクラス k に分類される。

一方、アフィン空間内での分布を考えず、外側の分布のみを考えると 3.1 節の非ベイズ型となる。また 3 次元アフィン空間の代わりに原点を通る 4 次元部分空間を用いても手順は同様である。このような反復はクラスタリングの「教師なし学習」として研究され、反復によって尤度が単調の増加し、一つの局所解に収束することが証明される³ [10]。

ただし、この方法で大域的な最適解に収束させるには相当に正しい初期値から反復を開始しなければならない。任意の初期値からこのような教師なし学習のみで正しいクラスタリングを行なうのはほとんどの場合、無理である。

4. 退化モデル

4.1 平面内の運動

通常のシーンでは物体と背景が画像内で 2 次元的な運動をすることが多い。回転が画像面内で起これば、式 (2) の基底ベクトル k_κ を Z 軸方向に取ることができる。これは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、背景も物体も軌跡ベクトルが m_0, m_1, m_2 の張る 3 次元部分空間の中の 2 次元アフィン空間に含まれ、「退化」と呼ばれる。Costeira・Kanade [1] が導入した作用行列に基づく方法はその改良や変形を含めて、すべてそれぞれの運動の軌跡の空間が独立であることを仮定している。したがって、このような退化が生じると、従来の方法では正しい分離を行なうことができない。

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、背景に共通であるから、式 (5) の m_1, m_2 も物体、背景に共通になり、それぞれの 2 次元アフィン空間は互いに「平行」になる (図 2)。

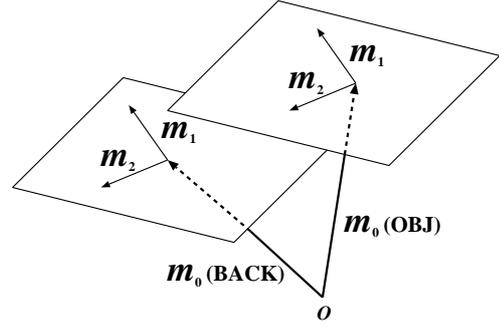


図 2: 物体と背景が独立に並進すると、それぞれの軌跡ベクトルの 2 次元アフィン空間は互いに平行になる。

しかし、二つの平行な 2 次元アフィン空間に対して、両者を含む 3 次元アフィン空間が存在する。アフィン空間分離法は軌跡ベクトルを独立な 3 次元アフィン空間に分離するものであるから、2 種類の運動物体の軌跡ベクトルが同一の 3 次元アフィン空間に含まれれば原理的に分離できない。これが、実ビデオシーンでアフィン空間分離法の分離精度が予想されるほど高くない理由であると考えられる。

4.2 退化モデルによる教師なし学習

実際のシーンではシーンも物体もほぼ並進成分のみで独立な回転成分は少ない。これを考慮すると、前節の退化モデルによる教師なし学習を適用することによってロバストな分離が可能になると期待される。その手順は次のようになる。

まず各軌跡 α の各クラス k に対する重み $W_\alpha^{(k)}$ を式 (6) のように初期化する。そして次の A, B, C を $\{W_\alpha^{(k)}\}$ が収束するまで反復する。

A. 各クラス $k = 1, \dots, m$ について次の計算を行なう。

1. クラス k の占有率 $w^{(k)}$ を (7) のよって計算する。
2. クラス k の重心 $p_C^{(k)}$ を式 (8) のよって計算する。
3. クラス k のモーメント行列 $M^{(k)}$ を式 (9) のよって計算する。

B. 次の計算を行なう。

1. 全特徴点のモーメント行列

$$M = \sum_{k=1}^m w^{(k)} M^{(k)} \quad (15)$$

の大きい 2 個の固有値に対する単位固有ベクトル u_1, u_2 を計算する。

2. 2 次元アフィン空間への射影行列 P とその外側方向への射影行列 P_\perp を次のように計算する。

$$P = \sum_{i=1}^2 u_i u_i^\top, \quad P_\perp = I - P \quad (16)$$

³後に欠落データを補う EM アルゴリズム [2] と数学的に同じ構造を持つことが指摘され、今日では EM アルゴリズムとして紹介されることが多い。

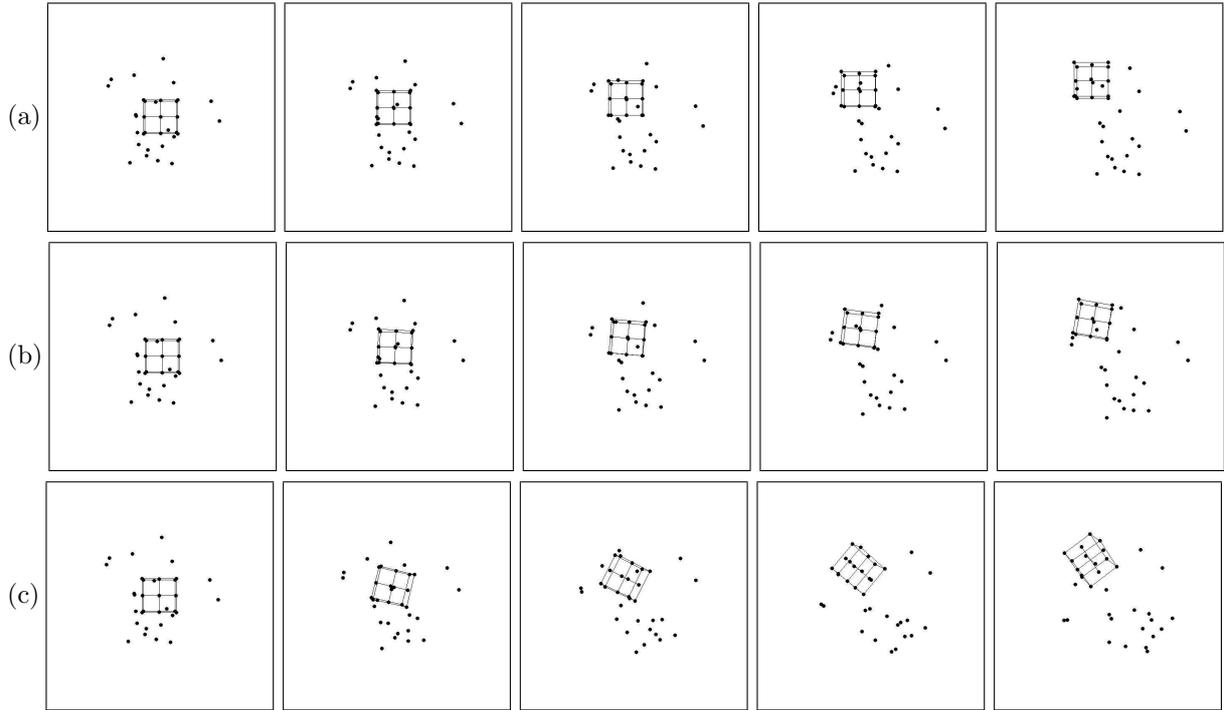


図 3: 背景点 (20 個) と物体点 (14 個) の運動 . (a) 退化した運動 , (b) 退化に近い運動 , (c) 一般の 3 次元運動 .

3. アフィン空間の外側に垂直な方向の二乗平均誤差を次のように推定する .

$$\hat{\sigma}^2 = \max\left[\frac{\text{tr}[\mathbf{P}_\perp \mathbf{M} \mathbf{P}_\perp]}{n-2}, \sigma^2\right] \quad (17)$$

4. クラス $k = 1, \dots, m$ の共分散行列 $\mathbf{V}^{(k)}$ を次のように計算する .

$$\mathbf{V}^{(k)} = \mathbf{P} \mathbf{M}^{(k)} \mathbf{P} + \hat{\sigma}^2 \mathbf{P}_\perp \quad (18)$$

C. 各軌跡 $\alpha = 1, \dots, N$ について次の計算をする .

- 軌跡 α の各部分空間に対する尤度 $P(\alpha|k)$, $k = 1, \dots, m$ を式 (13) によって計算する .
- 軌跡 α の重み $W_\alpha^{(k)}$, $k = 1, \dots, m$ を式 (14) によって更新する .

収束したら各軌跡 α を $W_\alpha^{(k)}$, $k = 1, \dots, m$ が最大となるクラス k に分類する . この手順でアフィン空間内部での分布を考えなければ非ベイズ型の手順が得られる .

4.3 多段階最適化

上記の学習を開始するための初期分類として, 各運動の軌跡ベクトルが 2 次元アフィン空間であることを仮定するアフィン空間分類法を用いる [9] . そして上記の退化モデルに基づく最適化を行えば, 実際に退化モデルが成立しているとき非常に正確な分類が行なわれるであろう . しかし背景や物体の運動

が厳密に平面的な運動であるとは限らない . また面内および面外の回転も独立に起こっているであろう . これを考慮するために, 得られた分類を初期値として 3.2 節に述べた 3 次元アフィン空間に基づく最適化を施す . すなわち, 次の 3 段階の処理を行なう .

- 2 次元アフィン空間に基づくアフィン空間分離法で初期分類を行なう .
- それに対して退化モデルに基づくベイズ型学習を行なう .
- それに対して 3 次元アフィン空間に基づくベイズ型学習を行なう .

こうすれば退化に近い運動では最適な解が得られ, 退化から相当に隔たった運動に対してもかなりの精度が維持されると期待される .

5. 実験

5.1 シミュレーション実験

図 3 はシーン中で 20 個の背景点と 14 個の物体点がそれぞれ独立に移動している 512×512 画素を想定したシミュレーション画像である . 見やすくするために物体点を線で結んでいる . 本論文の理論は奥行き効果のないアフィンカメラに基づいているが, この例ではより現実的な状況を想定して画像フレームを見込む画角が 30° の透視投影を用いている . そして, 背景は画像面に対して角度 10° で遠ざかり, 物体は別の方向に角度 10° で近づいている . 図 3(a) で

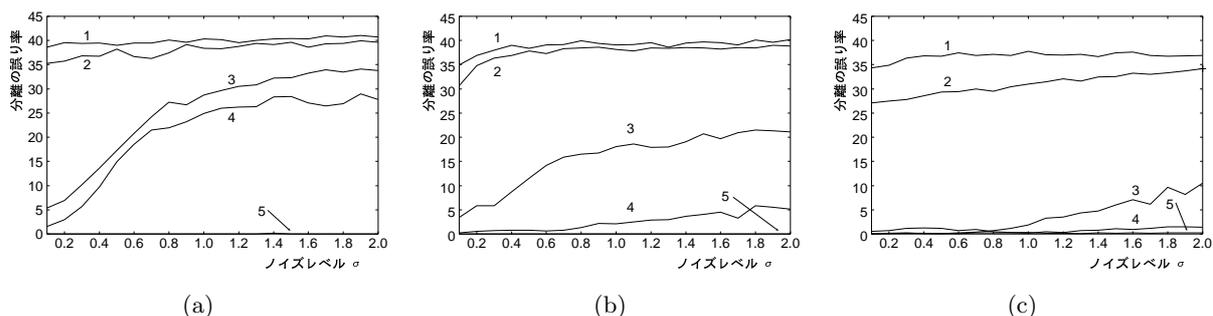


図 4: 図 3 に対する分離の平均正解率．横軸は加えた誤差の標準偏差．1. Costeira・Kanade の方法，2. 市村の方法，3. 最適部分空間分離法，4. 最適アフィン空間分離法，5. 多段階最適化法．

は物体も背景も回転せず平行移動している．図 3(b) では光軸に対してそれぞれ別々に方向に 10° 傾いた回転軸の周りに互いに反対方向に角度 10° の回転をしている．図 3(c) では 60° 傾いた回転軸の周りに互いに反対方向に角度 60° の回転をしている．

この例に対して，各点の位置の x, y 座標に独立に期待値 0，標準偏差 σ (画素) の正規分布に従う誤差を加えて背景点と物体点を分離した．図 4 は横軸の各 σ に対して誤差を変えて 500 回試行した平均の誤り率 (全 34 点中の誤って分類された点の割合) を縦軸にプロットしたものである．分離方法として，Costeira・Kanade [1] に示されている作用行列の行と列を入れ換えて近似的なブロック対角行列に変換する方法，作用行列に大津の判別条件を適用する市村の方法 [4]，黒澤・金谷の部分空間分離法 [8] およびアフィン空間分離法 [9] のそれぞれにベイズ型学習を組み合わせたもの (それぞれ「最適部分空間分離法」，「最適アフィン空間分離法」と呼ぶ)，および 4.3 節の多段階最適化を比較している．

図 3(a) は退化した運動であり，図 4(a) からわかるように，最適部分空間分離法も最適アフィン空間分離法もよく機能せず，4.1 節で考察したようにアフィン空間分離法の優位が失われている．誤差を加えない厳密なデータでも 100% 正しい分離ができないのは透視投影画像のためであり，アフィンカメラのモデル化誤差が画像の誤差と同じ役割を果たすからである (黒澤・金谷 [9] はこれを「実効誤差」と呼んでいる)．これに対して多段階最適化法はこのような退化の運動のモデルに基づいているので，ほとんど誤差の大きさによらずに正確な分離を行う．

図 3(b) は退化ではないので，図 4(b) からわかるように最適部分空間分離法も最適アフィン空間もある程度の精度で分離するが，多段階最適化法は依然としてほとんど誤差の範囲で正しい分離を行う．

図 3(c) は物体も背景もほとんど一般の 3 次元運動

をしているので，図 4(c) からわかるように最適部分空間分離法も最適アフィン空間分離法もかなり正しい分離を行なうが，多段階最適化法はそれよりさらに高い精度を維持している．

多段階最適化法と最適アフィン空間分離法は同じ教師なし学習を行なっているが，多段階最適化法の精度が高いのはその前段階の処理で高い精度の初期値が得られるからである．このように，退化でない場合も退化のモデルに基づく初期値から出発することで高い精度が達成される．

なお，どの場合でも素朴な Costeira・Kanade の方法 [1] は無力であり，ほとんどランダムな分離に近い．作用行列に大津の判別条件を適用する市村の方法 [4] も退化に対してはほとんど効果がないが，図 3(c) のような一般の 3 次元運動ではやや効果を発揮する．しかし最適部分空間分離法，最適アフィン空間分離法，および多段階最適化法にははるかに及ばない．

5.2 学習の効果

図 5 は教師なし学習がどの程度の効果を持つかを調べるために，図 3(c) の運動に対して黒澤・金谷の部分空間分離法 [8] およびアフィン空間分離法 [9] のみを施した場合と，それぞれに非ベイズ型およびベイズ型学習を組み合わせたものに対して図 4 と同様に誤り率をプロットしたものである．これから非ベイズ型およびベイズ型の学習がともに効果的であり，ベイズ型のほうが非ベイズ型学習より優位であることが確認できる．ただし，それでも図 4(c) からわかるように，多段階最適化法には及ばない．

図 6 は 4.3 節に示した 3 段階がどの程度の効果を持つかを調べるために，図 3(c) の運動に対してそれぞれの段階ごとに誤り率を評価したものである．この運動は退化でないにもかかわらず，退化モデルによる学習でかなり精度が向上し，その後の一般のモデルによる学習によって極めて高精度の分離が達成

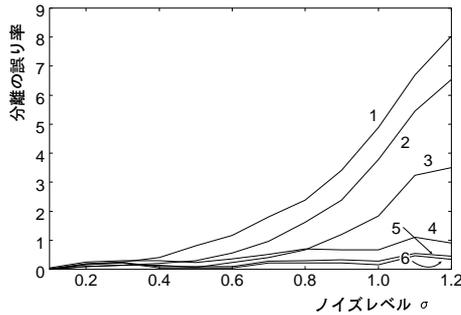


図 5: 図 3(c) の運動に対する学習の効果．横軸は誤差の標準偏差，縦軸は分離の平均正解率．1. 部分空間分離法，2. 部分空間分離法に非ベイズ型学習を加えたもの，3. 部分空間分離法にベイズ型学習を加えたもの，4. アフィン空間分離法，5. アフィン空間分離法に非ベイズ型学習を加えたもの，6. アフィン空間分離法にベイズ型学習を加えたもの．

されている．

5.3 実ビデオ画像実験

3次元シーンを撮影した3種類のビデオ画像列 A, B, C (320 × 240 画素) からそれぞれ Kanade-Lucas-Tomasi の方法 [13] で特徴点を抽出して追跡した．図 7 は特徴点位置をマークした 5 フレームを抜き出したものである．表 1 に最終フレームまで追跡できた軌跡から菅谷・金谷 [11] の方法でアウトライアを除去して得られた軌跡の個数，フレーム数，および多段階最適化法の実行時間 (秒) を示す．この計算では，すべての軌跡ベクトルの次元を 8 次元に圧縮して計算を効率化する菅谷・金谷 [12] の手法を用いている．CPU には Pentium 4 2.4B GHz，主メモリ 1Gb，OS には Linux を用いた．

表 2 に Costeira・Kanade の方法 [1]，市村の方法 [4]，黒澤・金谷の部分空間分離法 [8]，アフィン空間分離法 [9]，それらに本論文のベイズ型学習を施した最適部分空間分離法，最適アフィン空間分離法，および多段階最適化法による分離の正解率 (正しく分類された軌跡の割合) を % で示す．ただし最初の二つ以外は初期推定に黒澤・金谷の部分空間分離法 [8] およびアフィン空間分離法 [9] を用いており，これらは内部でランダム投票によるロバスト推定を行なっているため，試行ごとに結果がやや異なることがある．そこで表では 50 回の試行の平均値を示した．

この結果からわかるように，Costeira・Kanade の方法 [1] ではどれもほとんどランダムな分離しか行なわれず，市村の方法 [4] は画像列 A, B には効果があるが画像列 C には効果が見られない．画像列 A では部分空間分離法 [8] よりアフィン空間分離法 [9] のほうが高精度であり，画像列 B ではどちらも同程度に

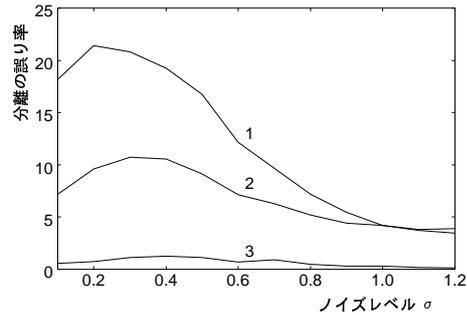


図 6: 図 3(c) の運動に対する学習の効果．横軸は誤差の標準偏差，縦軸は分離の平均正解率．1. 2次元アフィン空間に基づくアフィン空間分離法，2. それに退化モデルに基づくベイズ型学習を加えたもの．3. さらに 3次元アフィン空間に基づくベイズ型学習を加えたもの．

高精度であるが，画像列 C では逆に部分空間分離法のほうがアフィン空間分離法より高精度である．これらのことから画像列 C の運動はかなり退化に近いと考えられる．

この部分空間分離法とアフィン空間分離法にベイズ型学習を施すと画像列 A では大きな効果が見られるが，画像列 B, C では効果はわずかである．このことから教師なし学習の効果は初期値の精度に大きく依存することがわかる．それに対して多段階最適化法はどの場合も 100% の正解率を達成している．

6. まとめ

ビデオ画像上を移動する特徴点の軌跡から独立な運動を分離する方法として，本論文では退化した運動を仮定する教師なし学習 (EM アルゴリズム) によって分離を最適化し，その結果をさらに一般の運動を仮定する教師なし学習で最適化する多段階最適化の方法を提案した．そしてシミュレーションおよび実ビデオ画像を用いて，多段階最適化法は従来のどの手法よりも優れていることを実証した．

多段階最適化法がこのように優れている理由は，図 7 のような実ビデオ画像では物体や背景が極端な 3次元運動をすることはなく，回転が少なく平面運動に近いにもかかわらず，従来の Costeira・Kanade [1] に由来する作用行列を基にするどの方法も，物体と背景がともに一般の 3次元運動をするという前提で構築されていたためである．このため実際に起きやすい単純な運動にはかえって性能が悪化する．

それに対して多段階最適化法は現実に生じやすい退化を仮定したモデルによって学習し，さらに一般の 3次元運動のモデルによる学習を加えているので，通常起こり得るさまざまな 3次元運動に対して高い性能が発揮され，この問題の最終的な解法と考えら



図 7: 実ビデオ画像上の特徴点の追跡 (5 フレームを抜き出したもの) .

表 1: 図 7 のビデオ画像列に対する多段階最適化手法の実行時間 .

	A	B	C
フレーム数	30	17	100
軌跡の個数	136	63	73
実行時間 (秒)	2.50	0.51	1.49

れる⁴ .

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] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM Algorithm, *J. Roy. Statist. Soc.*, **B-39**, 1–38, 1977.
- [3] C. W. Gear, Multibody grouping from motion images, *Int. J. Comput. Vision*, **29**-2, 133–150, Aug./Sept. 1998.
- [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.
- [6] 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.
- [7] 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

⁴実験に用いたプログラムソースは注 1) と同じ Web サイトに公開している .

表 2: 図 7 のビデオ画像に対する分離の正解率 (%) .

	A	B	C
Costeira-Kanade 法	60.3	71.3	58.8
市村の方法	92.6	80.1	68.3
部分空間分離法	59.3	99.5	98.9
アフィン空間分離法	81.8	99.7	67.5
最適部分空間分離法	99.0	99.6	99.6
最適アフィン空間分離法	99.0	99.8	69.3
多段階最適化法	100.0	100.0	100.0

- 通信学会論文誌 D-II, **J74-D-II-8**, 1497–1505, Aug. 1993.
- [8] 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.
- [9] 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.
- [10] M. I. Schlesinger and V. Hlaváč, *Ten Lectures on Statistical and Structural Pattern Recognition*, Kluwer, Dordrecht, The Netherlands, 2002.
- [11] 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.
- [12] 菅谷保之, 金谷健一, 運動物体分離のためのカメラモデルの自動選択, 情報処理学会研究報告, 2002-CVIM-134-2, 9–16, Sept. 2002.
- [13] 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/klf/>.
- [14] 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, Kauai, Hawaii, U.S.A., Dec. 2001.