

粒子フィルタによる三次元復元

Structure from Motion by using Particle Filter

吉岡 和志†
Kazushi Yoshioka

末松 伸朗†
Nobuo Suematsu

林 朗†
Akira Hayashi

1 はじめに

本研究では粒子フィルタを使用し、対応付けが未知の場合の三次元復元の手法を提案する。

三次元復元とは、動画列のような画像列からカメラ動きを推定し、撮影対象の三次元構造を復元する問題である。三次元構造の復元には画像間における特徴点の対応付けが必要となる。しかし、オクルージョンなどが存在し特徴点を安定して検出することができない等の理由により、対応付けを正確に行うことは困難である。従来の研究では特徴点検出や対応付けに対して制約を必要としている。カルマンフィルタを用いた手法では環境中の特徴点と画像観測点の対応付けを既知としている。EM アルゴリズムを用いたものは、観測される特徴点の数が一定であると仮定をおいている。

本研究では、粒子フィルタを用いて、特徴点の検出が安定しない場合、また、環境中の特徴点と画像上の観測点との対応付けが未知な場合においても有効な三次元復元の手法を提案する。粒子フィルタは複数の対応付けの可能性を考慮することができ、対応付けが不明な場合においてもロバストなアルゴリズムとなる。

2 問題設定

本研究では、移動する単眼カメラで三次元環境中に散在する多数の特徴点を撮影した画像列から、カメラの位置・姿勢と環境中の三次元構造を復元する事を目的とする。ただし、環境中の特徴点は静的であるとす。図1は移動するカメラと環境中の特徴点の様子を表したものである。

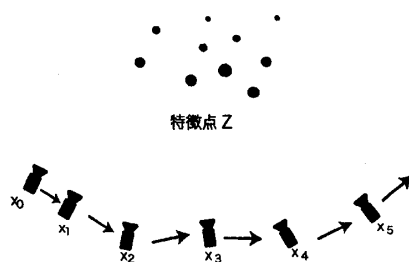


図1 カメラと環境中の特徴点

時刻 t において、撮影した時のカメラ位置・姿勢を $x_t(x_0$ のカメラ座標系をワールド座標系とする)、時刻 t に見つかった画像上の N_t 個の観測点集合を $Y_t = \{y_1^t, \dots, y_{N_t}^t\}$ (y_i^t は画像座標系

である)、環境中の M 個の特徴点の集合を $Z = \{z_1, \dots, z_M\}$ 、環境中の特徴点と画像上での観測点との対応付けを J_t とする。

本研究では観測値を画像上の観測点 Y_t とし、被推定値はカメラの位置・姿勢 x_t 、環境中の特徴点 Z_t 、画像上の観測点と環境中の特徴点の対応付け J_t である。つまり、問題は次の確率密度関数を推定することである。

$$p(x_{1:t}, J_{1:t}, Z|Y_{0:t}) \quad (1)$$

(1) にベイズの定理を用いて式変形を行う。

$$\begin{aligned} & p(x_{1:t}, J_{1:t}, Z|Y_{0:t}) \\ & \propto p(Y_t, J_t|Y_{0:t-1}, x_{0:t}, J_{0:t-1}, Z)p(x_{1:t}, J_{1:t-1}, Z|Y_{0:t-1}) \\ & = p(Y_t, J_t|Y_{0:t-1}, x_{0:t}, J_{0:t-1}, Z) \\ & \quad \times p(x_t|x_{1:t-1}, J_{1:t-1}, Z, Y_{0:t-1})p(x_{1:t-1}, J_{1:t-1}, Z|Y_{0:t-1}) \end{aligned}$$

$p(Y_t, J_t|Y_{0:t-1}, x_{0:t}, J_{0:t-1}, Z)$ は観測モデル $p(Y_t, J_t|x_t, Z)$ であり、 $p(x_t|x_{1:t-1}, J_{1:t-1}, Z, Y_{0:t-1})$ はカメラモーションモデル $p(x_t|x_{t-1})$ であるとの仮定をおくと

$$\propto p(x_{1:t-1}, J_{1:t-1}, Z|Y_{0:t-1}, x_0)p(Y_t, J_t|x_t, Z)p(x_t|x_{t-1}) \quad (2)$$

となる。状態空間モデルとして適当なカメラモーションモデル $p(x_t|x_{t-1})$ と観測モデル $p(Y_t, J_t|x_t, Z)$ を指定すれば、(2) 式は逐次的なベイズフィルタの形になる。

(2) 式に粒子フィルタを適用し、分布推定を行う。

3 手法

本研究では粒子フィルタを用いて三次元復元を行う。粒子フィルタの状態ベクトルを $s_t = \{x_t, J_t, Z\}$ とする。カメラの位置・姿勢 x と対応付け J の推定は粒子フィルタで行う。粒子フィルタの仮説の一つとして与えられたカメラの位置・姿勢 $x^{(i)}$ に対しての対応付け $J^{(i)}$ の決定は The Scott and Longuet-Higgins algorithm[2] により行う。カメラの位置・姿勢 $x^{(i)}$ と対応付け $J^{(i)}$ が決定するとステレオ計測により三次元点 $Z^{(i)}$ の計算が可能になる。画面上の観測誤差 (二次元ガウス分布) に対してステレオ計測を行うことにより、 $Z^{(i)}$ は三次元ガウス分布で表現される.[1]。三次元点 $Z^{(i)}$ の平均と分散の更新はカルマンフィルタを用いて行う。

3.1 粒子フィルタ

粒子フィルタは観測系列に基づいて、状態ベクトルの確率密度分布の推定を行う手法である。粒子フィルタは非線形・非ガウス型状態空間モデルに対してもフィルタリングが可能であり、観測に異常値が存在する場合の問題も扱うことができる。

その基本的なアイデアは状態ベクトルの確率密度分布を多数の粒子によりモンテカルロ近似を行うというものである。 N 個

† 広島市立大学大学院情報科学研究科 〒731-3194 広島市安佐南区大塚東 3-4-1
Email: k-yoshi@robotics.im.hiroshima-cu.ac.jp

の粒子がそれぞれ状態と重み $\{s_t^{(i)}, w_t^{(i)}\} (i=1 \dots N)$ を持ち分布を近似している。それは、複数の粒子によって複数の仮説を追っていると考えられる。よって、ある粒子において対応付けを誤った場合にも、複数の仮説を追っているためロバストな推定が可能であると考えられる。

粒子フィルタのアルゴリズムは次のようになる。

1. 時刻 $t-1$ の状態: 時刻 $t-1$ において分布 $p(x_t, J_t, Z|Y_{0:t-1}, x_0)$ が N 個の粒子 (状態 $s_{t-1}^{(i)}$ と重み $w_{t-1}^{(i)}$) で近似されているとする。 ($i=1 \dots N$)
2. 伝播: カメラモーションモデル $p(x_t|x_{t-1} = x_{t-1}^{(i)})$ により、時刻 t における N 個のカメラ位置・姿勢の状態 $x_t^{(i)}$ を生成する。 $x_t^{(i)}$ に対して、後に述べる The Scott and Longuet-Higgins algorithm を用いて Y_{t-1} と Y_t の対応付け $J_t^{(i)}$ を決定する。見つかった対応付け $J_t^{(i)}$ に従ってステレオ計測を行い三次元点 $Z^{(i)}$ を計算する。
3. 重み付け: 観測モデル $w_t^{(i)} = p(y_t|x_t = x_t^{(i)})$ を用いて粒子 $x_t^{(i)}$ の重み $w_t^{(i)}$ を決定する。重みは総和が 1 となるように正規化を行う。
4. リサンプリング: 重み $w_t^{(i)}$ に従って、 N 個の粒子を抽出する。重みの大きな粒子からは多くの粒子が抽出され、重みの小さな粒子は抽出されずに消滅する。

以上の処理を逐次的に行う。

3.2 The Scott and Longuet-Higgins algorithm による対応付け

The Scott and Longuet-Higgins algorithm は画像間の観測点に対して一対一の対応付けを見つけるアルゴリズムである。これを用いて対応付け J_t を決定する。The Scott and Longuet-Higgins algorithm は次の処理を行う。

1. 距離行列 G をつくる。
時刻 $t-1$ の観測点集合 $Y_{t-1} = \{y_1^{t-1}, \dots, y_k^{t-1}\}$ と時刻 $t+1$ の観測点集合 $Y_t = \{y_1^t, \dots, y_l^t\}$ のそれぞれの点について、対応するエピソード線との距離を考える。位置・姿勢が x_{t-1} のカメラから見た y_n^{t-1} の光線と観測点 y_m^{t-1} とのユークリッド距離を d_{1mn} 、位置・姿勢が x_t のカメラから見た y_m^{t-1} の光線と観測点 y_n^t とのユークリッド距離を d_{2kl} とする。

$$\text{cost}_{mn} = d_{1mn} + d_{2mn} \quad (3)$$

$$g_{mn} = \exp(-\text{cost}_{mn}/\sigma^2) \quad (4)$$

ガウス重み付けされた $g_{m,n} (m=1 \dots k, n=1 \dots l)$ を要素に持つ行列を G とする。

2. 行列 G に対し特異値分解を実行する。

$$G = UDV^t \quad (5)$$

3. 対角成分に特異値が並んだ行列 D の特異値を 1 に置き換えた行列を E をつくり、次の計算を行う。

$$P = UEV^t \quad (6)$$

4. 行列 P の要素 p_{mn} が m 行、 n 列において共に最大の値であるとき、 Y_{t-1} の m 番目の観測点と Y_t の n 番目の観測点には一対一の対応付けがあるとする。

以上のアルゴリズムを用いて対応付け J_t を決定する。

3.3 カルマンフィルタによる三次元点の更新

時刻 t の三次元点の平均・分散が観測として与えられたとき、時刻 $t-1$ まで維持されてきた三次元点の平均・分散の更新をカルマンフィルタを用いて行う手法を示す。特徴点の三次元計測を行うにはステレオ計測が必要なので、観測値は 2 回分を組として扱うことにし、 $\eta_t^m = (y_{t-1}^m, y_t^m), \Delta x_t = x_t - x_{t-1}$ とする。ここで z_m は η_t^m に対応する三次元特徴点である。

このとき、

$$p(z_m|x_{0:t}, Y_{0:t}, J_{0:t}) = p(z_m|\Delta x_{1:t}, \eta_{1:t}^m) \quad (7)$$

と書き換えられ、

$$\begin{aligned} p(z_m|\Delta x_{1:t}, \eta_{1:t}^m) &\propto p(\eta_t^m|z_m, \Delta x_{1:t})p(z_m|\Delta x_{1:t}, \eta_{1:t-1}^m) \\ &= p(\eta_t^m|z_m, \Delta x_t)p(z_m|\Delta x_{1:t-1}, \eta_{1:t-1}^m) \end{aligned} \quad (8)$$

となる。また、

$$p(\eta_t^m|z_m, \Delta x_t) \propto p(z_m|\eta_t^m, \Delta x_t)p(\eta_t^m|\Delta x_t) \quad (9)$$

$p(\eta_t^m|\Delta x_t)$ はステレオで観測される二点間の拘束条件のみを表しており、 $p(\eta_t^m|\Delta x_t)$ を定数として扱う。従って、

$$p(z_m|\Delta x_{1:t}, \eta_{1:t}^m) = p(z_m|\eta_t^m, \Delta x_t)p(z_m|\Delta x_{1:t-1}, \eta_{1:t-1}^m) \quad (10)$$

となる。 $p(z_m|\eta_t^m, \Delta x_t)$ と $p(z_m|\Delta x_{1:t-1}, \eta_{1:t-1}^m)$ は共にガウス分布なので、カルマンフィルタによって三次元点の更新が可能である。

4 実験

シミュレーションにおいて本手法の有効性を検証する。図 1 に示すように、環境中に特徴点を配置し、カメラで撮影するという状況をシミュレーションにより再現した。オクルージョンが発生する状況を考慮し、環境中の特徴点の数と画面上の観測点の数は必ずしも一致せず、時刻 t の観測点の数 N_t は増減する。

粒子フィルタでの推定値の評価法として粒子の重み平均を考える。カメラの位置・姿勢の重み平均推定値と真のカメラ位置・姿勢の平均二乗誤差で評価を行う。

5 終わりに

本研究では粒子フィルタを使用した対応付けが未知の場合の三次元復元の手法を提案した。粒子フィルタを用いることにより、特徴点検出に信頼性がない場合でも、ロバストに機能する手法を提案し、シミュレーションにおいて実験を行った。

今後の課題として、実画像を用いての実験、検証が必要であると考えられる。

参考文献

- [1] Olivier Faugeras. *Three-Dimensional Computer Vision - A geometric viewpoint*. The MIT Press, third edition, 1999.
- [2] Maurizio Pilu. Uncalibrated stereo correspondence by singular value decomposition, August 1997.
- [3] 宇江田浩之. パーティクルフィルタを利用した structure from motion. Master's thesis, 広島市立大学, Jan 2005.