

## 見えの生成モデルに基づく物体追跡

米川 達 弥<sup>†1</sup> 川本 一 彦<sup>†2</sup>  
井宮 淳 <sup>†2</sup> 杉本 晃 宏<sup>†3</sup>

画像上での物体の見えの変化を記述する確率的な生成モデルを導入し、見えの変化に対して追従性の高い物体追跡法を提案する。提案手法では、物体の見え方変化をパラメトリック固有空間法を用いて事前に学習し、学習画像から構成した固有部分空間上における学習多様体として表現する。この学習多様体上で次時刻での物体の見えを確率モデルによって生成し、観測画像と比較して尤度評価する。この生成的な方法は、ある1方向から撮影した物体の固定的なテンプレートをを用いる方法に比べて、テンプレートを逐次更新していることに相当するため、見えの変化に対して追従性を高めることができる。ただし、一般に、学習多様体は複雑な形状の曲線や曲面となり、物体の向き変化とその投影画像との関係は非線形になる。そのため、学習多様体上の時系列モデルの解析的な扱いは困難になる。そこで、パーティクルフィルタを用いて追跡アルゴリズムを構築する。実画像を用いた実験によりその有効性を示す。

### Object Tracking Based on Generative Appearance Model

TATSUYA YONEKAWA <sup>†1</sup> KAZUHIKO KAWAMOTO <sup>†2</sup>  
ATSUSHI IMIYA <sup>†2</sup> and AKIHIRO SUGIMOTO<sup>†3</sup>

We propose a method for tracking objects in image sequences using a generative appearance model on learned manifolds. The appearance model is defined as a stochastic model on the learned manifold and is used to predict time-varying appearances of objects in image sequences. Unlike the classical template matching, the proposed method generates appearance templates of objects in successive images and updates them in an online manner. The learned manifolds are constructed by the parametric eigenspace method and the appearance templates on the manifolds are generated by a particle filter. Each particle has its own appearance template and updates it in a framework of sequential Bayesian estimation. In experiments with real image sequences, we show the effectiveness of the proposed method.

#### 1. はじめに

コンピュータビジョンの分野において、画像上の物体追跡は動画処理の基礎技術の一つとして、様々な研究が行われている。それらの研究の多くは、ロボットの制御や監視システム、車載カメラなどに利用されている。実世界に存在する物体の多くは、周りの状況や物体の状態により、見えの変化を起こす。そのため、物体追跡では、物体の見えの変化に対する追従性を高めることが課題としてある。

画像上の物体が見えの変化を起こす要因としては、主に以下のことが考えられる。

- (1) 物体の姿勢変化
- (2) カメラの位置変化
- (3) 照明の変化

物体の姿勢変化について考えると、ある物体を鉛直方向の1軸を中心に1回転させたときに、固定されたカメラから得られる画像は図1のようになる。このような多様な見えの変化をする物体を、単純なテンプレートマッチングによる物体追跡では追跡できない。したがって、見えの変化をどのように学習するかが問題になってくる。

そこで、物体の見えの変化をパラメトリック固有空間法<sup>3)</sup>を用いて事前に学習することを考える。この手法は、物体の見えの変化を画像の固有ベクトルからなる固有部分空間上での多様体によって表現する。本論文中では、この多様体のことを学習多様体と呼ぶ。この学

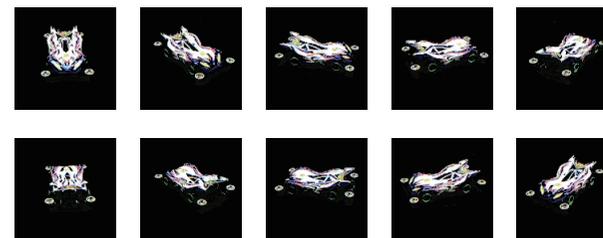


図1 見えの変化の例

<sup>†1</sup> 千葉大学大学院 融合科学研究科

Graduate School of Advanced Integration, Chiba University

<sup>†2</sup> 千葉大学総合メディア基盤センター

Institute of Media and Information Technology, Chiba University

<sup>†3</sup> 国立情報学研究所 コンテンツ科学研究系

Digital Content and Media Sciences Research Division, National Institute of Informatics

習多様体は、見えの変化する連続した画像集合を主成分分析して得られる固有部分空間に存在している。そのため、見えの変化を学習多様体で表現することで、より少ない計算量で物体を認識することができる。

しかし、一般に学習多様体は非常に複雑な形状の曲面となる。さらに、見えの変化の要因とそれに対応した学習多様体上のパラメータの関係は非線形になる。したがって、学習多様体上での時系列モデルもまた非線形となる。そこで、非線形・非ガウス型の時系列モデルを扱うことのできるパーティクルフィルタ<sup>4),5)</sup>を用いる。

本研究では物体の見えの時間変化を記述する時系列モデルを導入し、見えの変化を考慮した物体追跡を提案する。そして、提案手法を用いた追跡アルゴリズムを構築することを目的とする。

## 2. 学習多様体による見えの表現

パラメトリック固有空間法を用いて物体の見えの変化を学習多様体として表現する。パラメトリック固有空間法とは、3次元物体の向きや光源の変化に対し、連続的に変動する2次元画像の変化を、画像の固有ベクトルから構成される固有部分空間上での多様体によって表現する手法である。

本研究では、対象の物体は地面と接地した状態で移動し、鉛直方向の1軸を中心とした回転による姿勢変化のみを扱う。まず、対象物体をターンテーブルに乗せて一定角度ずつ回転させながらカメラで撮影を行う。カメラと物体の距離、カメラの仰角、照明の位置は固定しておく。次に撮影した画像から対象物体のみを抜き出し物体以外の部分に0の値を代入した画素値  $N$  の画像を生成する。この画像を学習画像とする。

学習画像をラスタースキャンし、画素値を要素とするベクトルを

$$\tilde{\mathbf{a}} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N)^\top \quad (1)$$

で表す。次に、このベクトルを大きさが1となるように次式を用いて正規化する。

$$\mathbf{a} = \frac{\tilde{\mathbf{a}}}{\|\tilde{\mathbf{a}}\|} \quad (2)$$

このように、明るさの正規化をすることで照明による明暗の変化や、カメラの特性による影響を軽減することができる。

ここで、物体を刻み数  $R$  で一回転させた画像の集合を  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R\}$  とする。全学習サンプル画像から平均の画像  $\mathbf{c}$  を求める。

$$\mathbf{c} = \frac{1}{R} \sum_{r=1}^R \mathbf{a}_r \quad (3)$$

さらに、学習画像から平均画像を差し引いた行列  $\mathbf{A}$  を作る。

$$\mathbf{A} = (\mathbf{a}_1 - \mathbf{c}, \mathbf{a}_2 - \mathbf{c}, \dots, \mathbf{a}_R - \mathbf{c}) \quad (4)$$

このときの画像集合の共分散行列  $\mathbf{Q}$  は

$$\mathbf{Q} = \mathbf{A}\mathbf{A}^\top \quad (5)$$

によって計算される。よって固有部分空間は次の固有方程式

$$\lambda_i \mathbf{e}_i = \mathbf{Q}\mathbf{e}_i \quad (6)$$

を解くことで求まる。例えば  $k$  次元の固有部分空間であれば  $k$  個の大きい固有値  $\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_N$  に対応する固有ベクトル  $\mathbf{e}_1 \dots \mathbf{e}_k$  を基底ベクトルとすることによって得られる。

最後に学習画像から平均画像を引いたベクトルを式

$$\mathbf{g}_r = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k)^\top (\mathbf{a}_r - \mathbf{c}) \quad (7)$$

により固有部分空間へ投影する。すると1枚の画像は固有部分空間上の1点に対応し、学習画像を示す点と点の間を補間することによって多様体を得られる。このとき、学習多様体はその回転角度  $\theta$  でパラメータ化された曲線となる。例として4度ずつ回転させて撮影した90枚の学習サンプル画像から構成した3次元での学習多様体を図2に示す。点と点の間

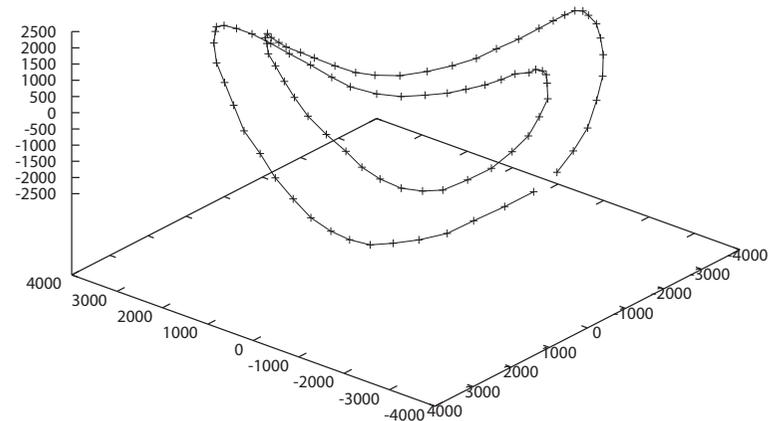


図2 3次元での学習多様体の例

は直線で補間している．

物体の向き  $\theta$  を与えると，学習多様体上の対応する点が決まる．その点は固有部分空間上での物体の向き  $\theta$  に対応する物体の画像となる．この画像を  $g(\theta)$  と書くことにする．

### 3. 学習多様体上のパーティクルフィルタ

学習多様体上の一般状態空間モデルを導入しパーティクルフィルタによって状態を推定する．

#### 3.1 一般状態空間モデルとベイズフィルタ

時刻  $t$  の画像上での物体の位置を  $\mathbf{r}_t = (x_t, y_t)^\top$ ，物体の向きを  $\theta_t$  で表す．これらを並べて状態ベクトル  $\mathbf{x}_t = (\mathbf{r}_t^\top, \theta_t)^\top$  とする．本研究における物体追跡とは，この状態ベクトル  $\mathbf{x}_t$  を求めることである．そのために， $\mathbf{x}_t$  の状態推移と観測過程を一般状態空間モデルで記述する．

一般状態空間モデルは状態ベクトル  $\mathbf{x}_t$  の時間変化を記述するシステムモデルと状態ベクトルから観測ベクトルを生成する観測モデルからなる．問題は観測ベクトルに基づいて状態ベクトルを推定することである．つまり，事後分布  $p(\mathbf{x}_t | \mathbf{Y}_t)$  の逐次推定を行う．ただし， $\mathbf{Y}_t = \{y_1, \dots, y_t\}$  である．事後分布の最適な逐次推定は，ベイズフィルタとして以下のように導出できる<sup>5)</sup>．

$$P(\mathbf{x}_t | \mathbf{Y}_{t-1}) = \int P(\mathbf{x}_t | \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} | \mathbf{Y}_{t-1}) d\mathbf{x}_{t-1} \quad (8)$$

$$P(\mathbf{x}_t | \mathbf{Y}_t) = \frac{P(y_t | \mathbf{x}_t) P(\mathbf{x}_t | \mathbf{Y}_{t-1})}{P(y_t | \mathbf{Y}_{t-1})} \quad (9)$$

ただし，初期分布  $p(\mathbf{x}_0)$  は与えられているとする．

#### 3.2 学習多様体上でのモデル

本研究では，状態ベクトルの時間変化は滑らかであるとし，状態推移の確率モデルを次のように仮定する．

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{x}_t - \mathbf{x}_{t-1}, \Sigma) \quad (10)$$

ただし， $N(m, \Sigma)$  は，平均  $m$  と共分散行列  $\Sigma$  のガウス分布とする．状態ベクトルの各成分の時間変化はそれぞれ独立と仮定し，共分散行列  $\Sigma$  は対角行列であるとする．

$$\Sigma = \text{diag}(\sigma_x, \sigma_y, \sigma_\theta) \quad (11)$$

ここで，向き  $\theta_t$  の推移は，学習多様体上の画像  $g(\theta_t)$  の推移となることに注意が必要である．

本研究では，物体とカメラの間の距離は大きく変化せず画像上でのその大きさは一定と仮定する．このとき，位置  $\mathbf{r}_t$  が与えられると，その大きさの画像上の領域が決まる．この領

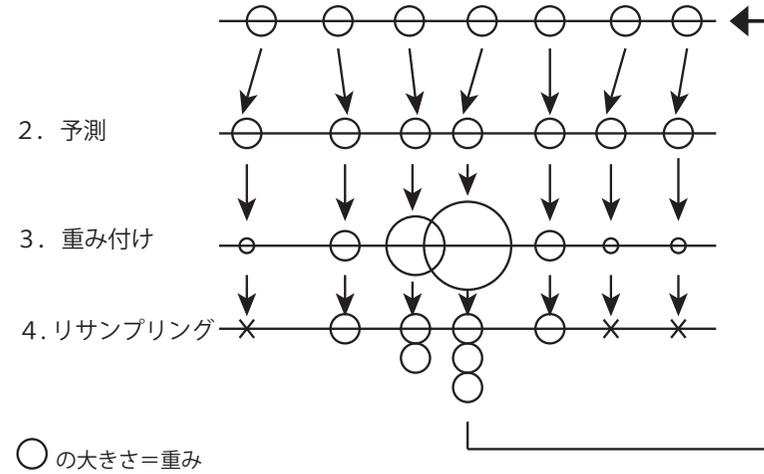


図3 パーティクルフィルタの模式図

域を明るさの正規化後，固有部分空間上へ投影した画像を  $y_t$  とする．観測過程の確率モデルを次のように仮定する．

$$p(y_t | \mathbf{x}_t) \propto \exp\left(-\frac{\|y_t - g(\theta_t)\|^2}{2\sigma^2}\right) \quad (12)$$

#### 3.3 パーティクルフィルタによる状態推定

本研究では，式(10)と(12)の確率モデルをもとに，物体を追跡する．式(12)は状態ベクトル  $\mathbf{x}_t$  に関して非線形になるため，解析的に取り扱うことは困難である．そこで，パーティクルフィルタを導入する．パーティクルフィルタは式(10)と(12)の確率モデルを  $N$  個の状態ベクトルの重み付きランダムサンプル  $\{\mathbf{x}_t^{(i)}\}_{i=1}^N$  でモンテカルロ近似する．この重み付きランダムサンプルのことをパーティクルと呼ぶ．図3にはパーティクルフィルタの模式図を示す．また，図4にパーティクルフィルタのアルゴリズムを示す．

- (1) 初期化 ( $t \leftarrow 0$ )
- $N$  個のパーティクル  $\{\mathbf{x}_0^{(i)}\}_{i=1}^N$  を初期分布  $p(0)$  から生成する .  

$$\mathbf{x}_0^{(i)} \sim p(0) \quad (13)$$
  - $t \leftarrow 1$
- (2) 予測 :
- $N$  個のパーティクル  $\{\tilde{\mathbf{x}}_t^{(i)}\}_{i=1}^N$  を式 (10) のシステムモデル  $p(\mathbf{x}_t|\mathbf{x}_{t-1})$  にしたがって生成する .  

$$\tilde{\mathbf{x}}_t^{(i)} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad (14)$$
- (3) 重み付け:
- 各パーティクル  $\tilde{\mathbf{x}}_t^{(i)}$ ,  $i = 1, \dots, N$  に関して式 (12) にしたがって重み  $\tilde{w}_t^{(i)}$  を計算する .  

$$\tilde{w}_t^{(i)} = p(\mathbf{y}_t|\tilde{\mathbf{x}}_t) \quad (15)$$
  - 各パーティクル  $\tilde{\mathbf{x}}_t^{(i)}$ ,  $i = 1, \dots, N$  の重みを正規化する .  

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{i=1}^N \tilde{w}_t^{(i)}} \quad (16)$$
- (4) リサンプリング
- $N$  個のパーティクル  $\{\tilde{\mathbf{x}}_t^{(i)}\}_{i=1}^N$  を対応する重みに従って復元抽出し,  $\{\mathbf{x}_t^{(i)}\}_{i=1}^N$  を生成する .
  - 各パーティクルの重みを  $w_t^{(i)} = 1/N$ ,  $i = 1, \dots, N$  とする .
  - $t \leftarrow t+1$  とし, 2 に戻る .

図 4 パーティクルフィルタのアルゴリズム

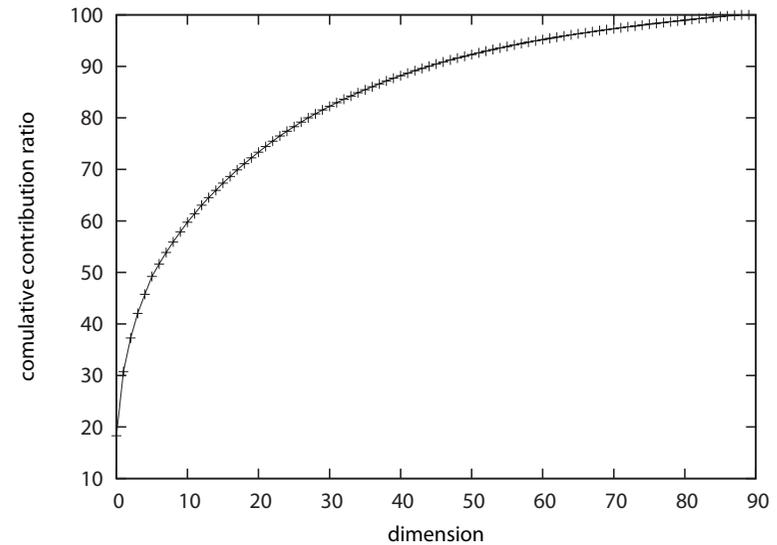


図 5 累積寄与率

## 4. 実 験

コンピュータビジョン向けのオープンソースのライブラリである OpenCV を用いて, C++ でプログラムを作成した. 実験の準備として車型のプラモデルを雲台に乗せ, 正確に 4 度ずつ回転させながら学習画像 90 枚を撮影した. 学習画像の画素数は  $64 \times 64$  である. それらの学習画像を主成分分析し学習多様体を構成した.

固有部分空間の次元は, 図 5 に示す通り累積寄与率が 8 割を超える  $k = 30$  とした. 式 (12) の分散は  $\sigma = 0.1$  とし, パーティクルの数は  $N = 200$  とした.

実験は全て学習画像を撮影した時と同じ設定で行った. つまり, カメラの姿勢と照明の位置は固定とする.

結果の出力には入力動画上にそれぞれのパーティクルが示す領域を緑枠で表した. また, 動画の左上にはそれぞれのパーティクルが推定した角度の平均値を表示した.

### 4.1 回転のみ

まず, 回転のみによって見え変化をする動画に対して実験を行った. 実験動画には雲台に

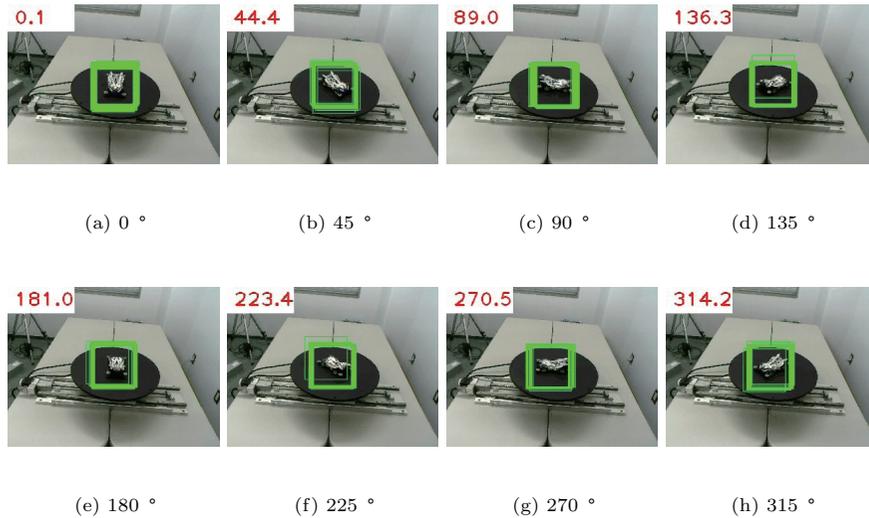


図 6 回転のみの動画に対する実験結果

載せた物体を回転させる動画を使用した．実験結果は図 6 のようになった．図の下の角度は雲台から得た真値である．

実験結果より，回転による見えの変化に影響されずに追跡を行えていることが分かる．また，物体の姿勢推定も同時に行えている．

#### 4.2 回転 + 平行移動

次に回転に加えて平行移動も行った動画に対して実験を行った．実験動画には雲台に載せた物体を回転させたり，平行移動を行った動画を使用した．この実験の結果は，図 7 に示す．

実験結果より，平行移動に対しても本手法が有効であることを示した．しかし，この実験では色が黒い雲台を使用して実験を行ったため，対象物体の周りの背景が黒で統一されてしまっている．そのため，対象物体は理想に近い状態で固有部分空間に投影される．これでは，提案手法の汎用性が検討できないので，対象物体の周りの背景を変えて同様な実験を行った．その結果，図 8 のように背景によっては追跡がうまく行かない場合も発生した．これは，固有部分空間に投影する際に対象物体の周りの画素も含めて投影されていることが原因である．

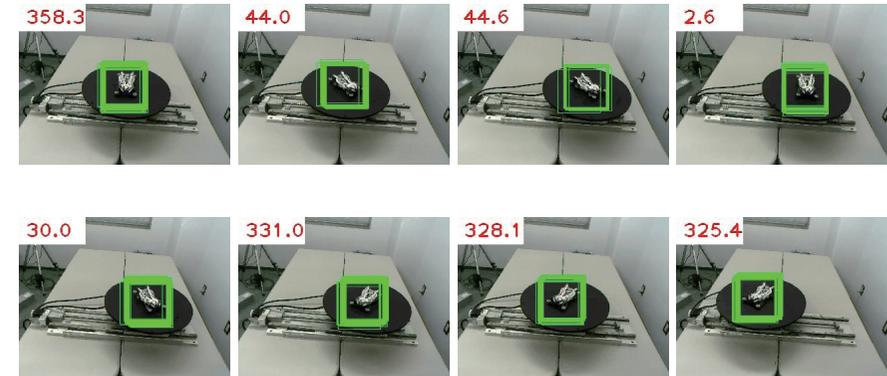


図 7 回転と移動を行う動画に対する実験結果

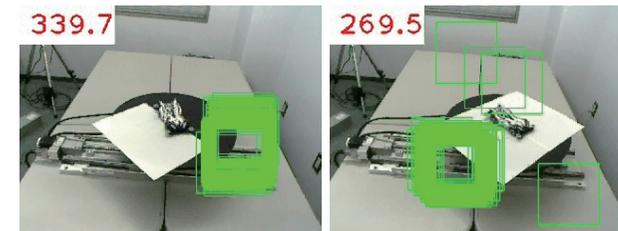


図 8 追跡失敗例

### 4.3 最小矩形での切り出し

そこで、背景の影響を最小限にとどめるために、対象の物体を囲む最小の矩形で画像を切り出し、固有部分空間に投影する方法を導入した。学習多様体を作る際に、角度だけではなく最小の矩形の高さと幅も一緒に記憶しておく。実験には、対象物体の角度が0度から4度刻みで回転させた90枚の画像を使用した。この90枚の画像に対して全画素をそのまま投影した場合と、最小矩形で切り出してから投影した場合の学習多様体との距離と尤度を調べた。理想的には、学習多様体との距離は0となる。実験結果は距離の変化を図9に、尤度の変化を図10に示した。横軸の角度は対象物体の向きを示している。使用した次元は $k=8$ 、尤度計算に使用した分散は $\sigma=100$ である。図中のall projectionは最小矩形を用いずに全画素を投影した場合のデータを、rectangle projectionは最小矩形を用いた投影方法でのデータを示している。

実験結果から分かるように最小矩形で切り出したほうが、全ての方向に対して全画素を投影した場合よりも尤度が高くなっていることが分かる。特に、角度が0, 90, 180, 270度の際は、学習多様体との距離がかなり小さくなっている。これは、最小矩形と対象物体の領域がほぼ同じになり、理想的な状態で投影されているからだと考えられる。しかし、それ以外の角度では理想と比べるとまだ背景の影響を受けているといえる。

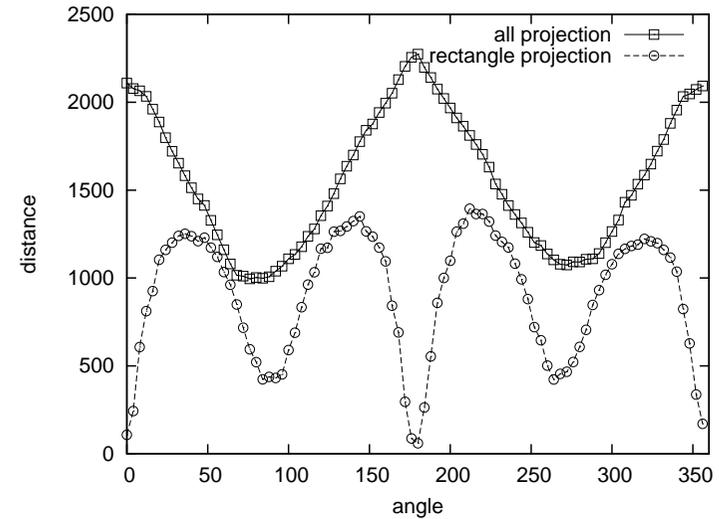


図9 距離の変化

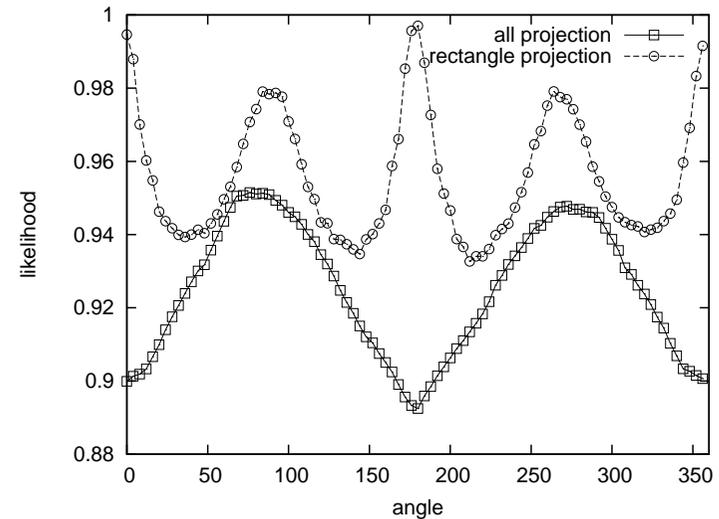


図10 尤度の変化

## 5. ま と め

本研究では、見えの変化する物体を追跡するためのアルゴリズムを提案し、その実験検証を行った。まず、物体の鉛直方向の1軸回転による見えの変化をパラメトリック固有空間法を用いて学習多様体で表現した。さらに、その多様体上の一般状態空間モデルを導入しパーティクルフィルタを用いた物体追跡を行った。そして、実験結果より本手法が見えの変化に対して有効なことを示した。同時に姿勢推定も行えることも示した。姿勢推定を行えたことにより、その情報を追跡にフィードバックすることが可能になると考えられる。例えば、車は向きによって進む方向が限られてくる。そのため、向きを推定出来ることで、進むはずのない方向への状態推移の確率モデルの分散を小さくできる。そうすることで、より安定した追跡を行えることが望める。さらには、パーティクルの数を減らせることにもつながり計算量の減少も望むことができる。

しかし、背景のパターンによっては追跡できない場合も確認された。この背景問題を解決するために、今回は最小矩形での切り出しを導入した。この方法によってある程度背景の影響を取り除くことができたが、理想的には画像の領域をマスクなどを用いて切り出すことが望ましい。

また、本研究で学習多様体を構成するために使用したパラメトリック固有空間法は、学習時に姿勢パラメータを正確に制御する必要がある。したがって、例えば、雲台に乗せられないほどの大きな物体には適用するのが難しい。そこで、パラメトリック固有空間法以外の学習多様体構成法を導入し、学習画像の作成をより簡便にする方法について今後検討する。

## 謝 辞

本研究の一部は、科学研究費補助金課題番号 20300061 の補助を受けて行った。

## 参 考 文 献

- 1) M.ISARD. Condensation-conditional density propagation for visual tracking. *International Journal of Computer Vision*, Vol.29, No.1, pp. 5-28, 1998.
- 2) 加藤丈和. パーティクルフィルタとその実装法 (チュートリアル). 情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア], Vol. 2007, No.1, pp. 161-168, 2007.
- 3) 村瀬洋, シュリーナイヤー. 2次元照合による3次元物体認識: パラメトリック固有空間法. 電子情報通信学会論文誌., Vol.77, No.11, pp. 2179-2187, 1994.

- 4) 樋口知之. 粒子フィルタ. 電子情報通信学会誌, Vol.88, No.12, pp. 989-994, 2005.
- 5) 北川源四郎. モンテカルロ・フィルタおよび平滑化について (特集 計算統計学の発展). 統計数理, Vol.44, No.1, pp. 31-48, 1996.