On-line Variational PCA for Adaptive Visual Tracking

Torazo Date* Takashi Bando* Tomohiro Shibata* Shin Ishii*

* Graduate School of Information Science
Nara Institute of Science and Technology (NAIST)
Takayama 8916-5, Ikoma, Nara 630-0192, Japan

Abstract

In the research area of visual tracking as well as recognition, there have been numerous attempts to use appearance-based object representation which does not require explicit knowledge or precise geometric representation of the object. In particular, the eigenspace representation has been widely used for visual tracking of a target, because of its rich representational power based on clear mathematical properties. There is, however, a big problem that the appearance of a target object may vary in time due to various factors such as changes in the lighting condition, and in location, pose, scale, and shape of the object. On-line algorithms that incrementally update eigenvectors (basis images) can be an answer to this problem. In this study, we propose an adaptive visual tracking method by combining an on-line variational Bayesian version of principal component analysis (PCA) and particle filtering. Computer simulations demonstrate our method enables robust visual tracking of a target whose appearance varies in noisy environments.

1 Introduction

In the research area of visual tracking as well as recognition, there have been numerous attempts to use appearance-based object representation which does not require explicit knowledge or precise geometric representation of the object. In particular, the eigenspace representation has been widely used (e.g., [6, 5]) because of its computational efficiency based on clear and established mathematical properties. There is, however, a serious problem that the conventional eigenspace method is not robust against changes in the appearance of a target, such as lighting condition, pose, scale, shape, and so on. This problem can be mediated if every appearance variation of the target object is prepared as basis images, but this is quite impractical. An algorithm which updates eigenvectors (basis images) on-line can be an answer to this problem. Lim et al. (2004) proposed such a visual tracking algorithm [2] that employed a sequential algorithm based on singular value decomposition (SVD), called R-SVD method, and demonstrated robust face tracking of a person walking in cluttered backgrounds.

In this article, we propose an adaptive visual tracking method by combining an on-line variational Bayesian version of principal component analysis (PCA) and particle filtering [4]. Computer simulations demonstrate our method enables robust visual tracking of a target whose appearance varies in noisy environments.

2 Algorithm

2.1 Overview

Our visual tracking algorithm is basically combination of on-line variational PCA and particle filter. Particle Filter (PF) [7] is an approach to making Bayesian estimation of intractable posterior distributions from time-series observation signals disturbed by non-Gaussian noises. The effectiveness of PFs has been reported in various research area, such as real-time visual tracking in Computer Vision (e.g., [4]). Under an assumption that the target dynamics form temporal Markov chains and observations are independent, incremental Bayesian estimation of the hidden target state $\mathbf{X}_t$ at time t is computed by

$$p(\mathbf{X}_t|\mathcal{I}_t) \propto p(\mathbf{I}_t|\mathbf{X}_t) \int p(\mathbf{X}_t|\mathbf{X}_{t-1}) p(\mathbf{X}_{t-1}|\mathcal{I}_{t-1}) d\mathbf{X}_{t-1} \quad (1)$$

where $\mathcal{I}_t = \{\mathbf{I}_1, \ldots, \mathbf{I}_t\}$ is a set of observed images.

In this study, smoothness is assumed for target motions: $\mathbf{X}_t - \mathbf{X}_{t-1} = \mathbf{X}_{t-1} - \mathbf{X}_{t-2} + \xi_t$, where ξ denotes a process noise. In the next subsection, we describe the likelihood term $p(\mathbf{I}_t|\mathbf{X}_t)$, based on on-line variational PCA.

2.2 On-line Variational PCA

Principal component analysis (PCA) is a well-established method of multivariate analysis to perform feature extraction from a data point or a set of data. A data space is reduced by a projection to a comparatively low-dimensional subspace, a feature space. In PCA, this projection is linear. PCA can be reformulated into a probabilistic model including latent variables, which is called Probabilistic Principal Component Analysis (PPCA).

Our proposed method is an on-line updating method whose generative model is given by PPCA, to construct the basis images. This method assumes observed data include noise, thereby robust basis images can be obtained in a real-world environment which includes various distracters or occlusion.

An observed image vector $\mathbf{I}_t$ is given by

$$\begin{aligned}\mathbf{I}_t =& \boldsymbol{\mu} + W\mathbf{s}_t + \mathbf{n}_t \\ =& \tilde{W}\tilde{\mathbf{s}}_t + \mathbf{n}_t, \quad \mathbf{n}_t \sim \mathcal{N}_{Nd}(\mathbf{n}_t|\mathbf{0}, \Sigma_t),\end{aligned} \tag{2}$$

where $\boldsymbol{\mu} \in \mathcal{R}^{Nd}$ is the mean of observed images, $W \in \mathcal{R}^{Nd\times Nb}$ is the basis images (the eigen-images), $\tilde{W} = (W, \boldsymbol{\mu})$, $\mathbf{s}_t \in \mathcal{R}^{Nb}$ is basis score, $\tilde{\mathbf{s}} = (\mathbf{s}_t^T, 1)^T$, Nd is the number of pixels, and Nb is the number of bases. $\mathcal{N}_D(\mathbf{I}|\boldsymbol{\mu}, \Sigma)$ denotes a D-dimensional Gaussian distribution with the mean $\boldsymbol{\mu}$ and the variance Σ. Then, we obtain the following formulation:

$$p(\mathbf{I}_t|\tilde{W}, \mathbf{s}_t) = \mathcal{N}_{Nd}(\mathbf{I}_t|\tilde{W}\tilde{\mathbf{s}}_t, \Sigma_t). \tag{3}$$

We also assume the prior distribution of $\mathbf{s}_t$ is given by

$$p(\mathbf{s}_t) = \mathcal{N}_{Nb}(\mathbf{s}_t|\mathbf{0}, E_{Nb}), \tag{4}$$

where E_{Nb} is a $Nb \times Nb$ unit matrix. Using the Bayes theorem, we have

$$p(\mathcal{I}_t, \mathcal{S}_t|\tilde{W}) = \prod_{\tau=1}^{t} p(\mathbf{s}_\tau)p(\mathbf{I}_\tau|\mathbf{s}_\tau, \tilde{W}). \tag{5}$$

For the prior distribution of the eigen-vectors, we assume natural conjugate:

$$p(\tilde{W}) = \prod_{d=1}^{D} \mathcal{N}_{N_b}(\boldsymbol{\theta}_d|\mathbf{e}_d, \Sigma_\theta), \tag{6}$$

where $\boldsymbol{\theta}_d = (w_d^{(1)}, ..., w_d^{(\tilde{N}b)})^T \in \mathcal{R}^{\tilde{N}b}$, $\mathbf{e}_d = (\delta_{1,d}, ..., \delta_{\tilde{N}b,d})^T \in \mathcal{R}^{\tilde{N}b}$, $\Sigma_\theta = \gamma^{-1}E_{\tilde{N}b} \in \mathcal{R}^{\tilde{N}b\times\tilde{N}b}$, $\tilde{N}b = Nb + 1$, and then $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_D)^T = \tilde{W}$. $\delta_{i,j}$ is the Kronecker delta:

$$\delta_{i,j} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j). \end{cases} \tag{7}$$

Then, the posterior distribution of the basis scores and the parameter (basis images) is given by

$$q(\tilde{W}, \mathcal{S}_t) = q(\mathcal{S}_t)q(\tilde{W}), \quad q(\mathcal{S}_t) = \prod_{\tau=1}^{t} q(\mathbf{s}_\tau), \tag{8}$$

where $\mathcal{S}_t = \{\mathbf{s}_1, ..., \mathbf{s}_t\}$.

In the variational Bayes, the free energy is defined and maximized. The following is an on-line version of the free energy in which a forgetting factor is introduced in order to be insensitive to inaccurate past data, and to cope with sudden changes of a target appearance or changes in the environment.

In our on-line variational PCA, the forgetting factor $\lambda_s (0 \leqq \lambda_s \leqq 1)$ is introduced in order to forget (discount) gradually the influence of the past incorrect estimation. $\eta_t = \left(\sum_{\tau=1}^{t} \Lambda(\tau, t)\right)^{-1}$ is a normalization coefficient, $\Lambda(\tau, t) = \prod_{s=\tau+1}^{t} \lambda_s$, and then the on-line free energy is given by

$$\begin{aligned}\mathcal{F}_t^\lambda[q] =& \eta_t \sum_{\tau=1}^{t} \Lambda(\tau, t)\{\log p(\mathcal{I}_\tau) \\ &- KL(q(\tilde{W}, \mathcal{S}_\tau)||p(\mathcal{S}_\tau, \tilde{W}|\mathcal{I}_\tau))\}.\end{aligned} \tag{9}$$

The maximization of the free energy is equivalent to the minimization of the discounted KL divergence between $q(\mathcal{S}_t, \tilde{W})$ and $p(\mathcal{S}_t, \tilde{W}|\mathcal{I}_t)$.

On variational Bayes estimation, $q(\mathcal{S}_t)$ and $q(\tilde{W})$ that maximize the free energy are computed by repeating a VB-E step and a VB-M step. After convergence, we obtain approximated solutions of factor score matrix S and eigen-vector matrix $\tilde{W}$. Then, eigen-vector matrix $\tilde{W}$ can be used as basis images.

The tracking is achieved by particle filter governed by the observation model $p(\mathbf{I}_t|\mathbf{X}_t)$, which is the likelihood of $\mathbf{X}_t$ given $\mathbf{I}_t$, and the dynamics model between two states $p(\mathbf{X}_t|\mathbf{X}_{t-1})$. The observation model is given by using the estimated basis $\tilde{W}$ as

$$p(\mathbf{I}_t|\mathbf{X}_t) \propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{I}_t(\mathbf{X}_t) - \tilde{W}\tilde{\mathbf{s}}_t\|^2\right), \tag{10}$$

where $\mathbf{I}_t(\mathbf{X}_t)$ is an observed image at the target state (position) $\mathbf{X}_t$, and σ^2 denotes the measurement noise variance assumed to be a constant for simplify. In this study, the motion of the target object between

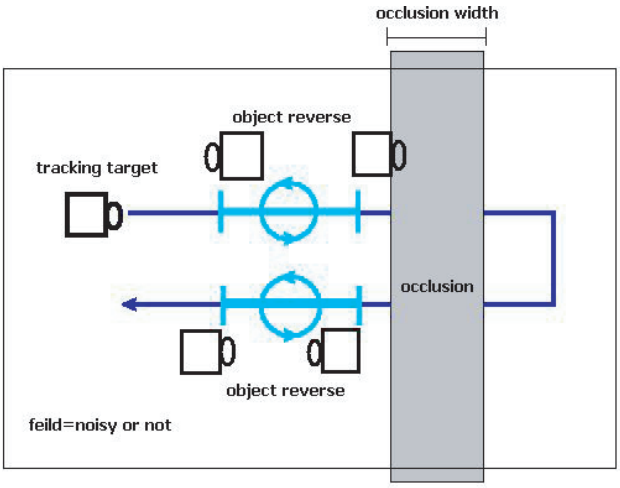

Figure 1: Simulation environment

two consecutive frames is assumed to be linear. The target state is defined as $\mathbf{X}_t = (x_t, y_t)$, where x_t and y_t denote the two-dimensional position on the image plane at time t.

3 Experiments

3.1 Setup

We conducted tracking experiments through computer simulations to examine the effectiveness of our proposed method, that is, adaptive basis image updates and robust visual tracking.

Figure 1 illustrates the simulation environment. A target, a coffee cup, moved from the left side of the image toward the right side with a velocity of 2 [pixels/frame], and turned back to the right side, as depicted by an allow. During this course, the appearance of the target was reversed twice with specific intervals indicated by gray lines. There were four experimental conditions; normal condition, noise condition in which pixel-wise Gaussian noise was added, occlusion condition in which a stationary occluder depicted as a shade was added, and noise-occlusion condition in which both noise and the occluder were added (cf. Figure 2). In each condition, 10 times of simulations were conducted to estimate statistically the performance of our method.

200 particles were used. The image stream was 110 frames of 120×160 gray scale images. The image of the target coffee cup was of 20×20 pixels. The width of occlusion was 15 pixels. The pixel-wise Gaussian noise was zero-mean with variance 30^2. The maximum number of basis images was set to 10.

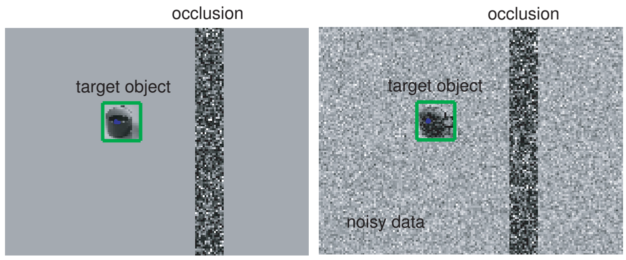

Figure 2: Sample images in simulation

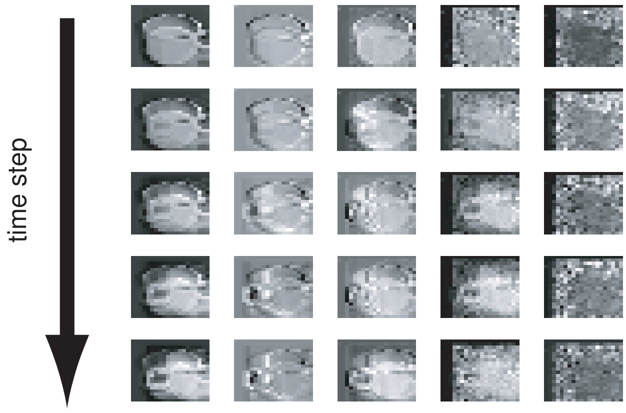

Figure 3: Time course of basis images when the target object was reversed (for the normal condition).

3.2 Results

Figure 3 demonstrates how the basis images were updated. This figure shows the time course of five basis images of the tracked object every three flames just after the object was reversed. Each basis was quickly and successfully adapted to the reversed images, in which the handgrip of the cup was a salient feature in the appearance of the cup. This result suggests three or four bases would be enough to represent this specific target object.

Figure 4 presents an example time course of the online free energy. Shaded areas and unshaded areas correspond to the two different appearances of the coffee cup (see Figure 1). The free energy sharply dropped down just after the target appearance was reversed, followed by quick recoveries. Because the free energy strongly correlates with the reliability of the estimation of basis images, this result indicates our successful implementation of the on-line variational PPCA. The time course shown in Figure 4 was for the normal con-

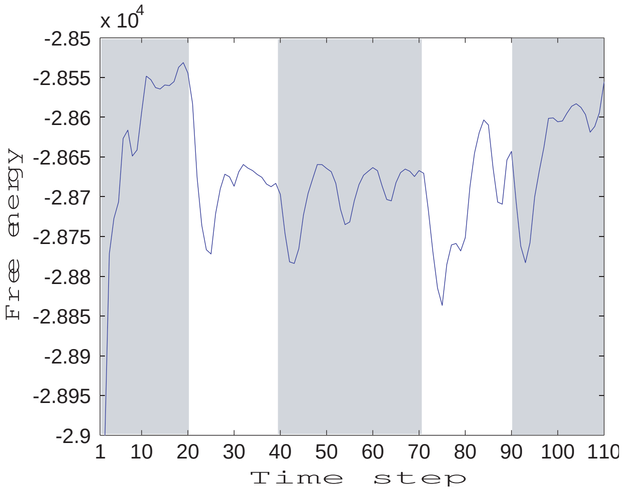

Figure 4: Example time course of on-line free energy

Table 1: Average tracking error over 10 trials

Task	error(MSE)
normal	4.29±2.68
noisy	4.42±2.63
occlusion	4.39±2.67
noisy and occlusion	4.52±2.62

dition, but we confirmed that profiles in other cases were almost consistent to this time course.

The tracking performance in the four conditions is summarized in Table 1. The tracking errors are similar. Increasing the number of particles from 200 should decrease the errors.

4 Conclusion and Future Work

In this study, we proposed an adaptive visual tracking method by combining the on-line variational Bayesian PCA and particle filtering. Simulation experiments demonstrated that the algorithm realized accurate and robust visual tracking of a target whose appearance was suddenly reversed, even in the case where Gaussian noise was artificially added and an occluder was placed in the pathway of the target.

There are many issues opened for future work. First of all, the performance comparison with existing methods like R-SVD [2] is necessary. Our preliminary experiments using the same simulation environment suggests that the tracking performance of R-SVD is as high as our method. Further exploration is required for the performance comparison, and for making the difference salient.

One of the most interesting and useful extensions would be on-line control of the forgetting factor. Hirayama et al. proposed a method in which the forgetting factor was controlled by the free energy calculated based on their probabilistic model, and demonstrated the effective detection of abrupt changes in face images [1]. As shown in their study, probabilistic models different from the conventional PCA model can be considered to improve the tracking performance in real-world environments.

References

[1] J. Hirayama and J. Yoshimoto and S. Ishii, "Bayesian representation learning in the cortex regulated by acetylcholine", *Neural Netw*, Vol.**17**(10), pp.1391-1400, 2004.

[2] J. Lim and D. Ross and R.-S. Lin and M.-H. Yang, "Incremental Learning for Visual Tracking", *Advances in Neural Information Processing Systems*, 2004.

[3] Z. Khan and T. Balch and F. Dellaert, "A Rao-Blackwellized Particle Filter for EigenTracking", *Proc Comput Vis Pattern Recogn*, Vol.**2**, pp.980-986, 2004.

[4] M. Isard and A. Blake, "CONDENSATION-Conditional Density Propagation for Visual Tracking", *Int J Comput Vis*, Vol.**29**(1), pp.5-28, 1998.

[5] M.J. Black and A.D. Jepson, "EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation", *Int J Comput Vis*, Vol.**26**(1), pp.63-84, 1998.

[6] H. Murase and S. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance", *Int J Comput Vis*, Vol.**14**, pp.5-24, 1995.

[7] N.J. Gordon and D.J. Salmond and A.F.M Smith, "Novel approach to nonlinear non-Gaussian Bayesian state estimation", *IEEE Proc Radar Signal Processing*, Vol.**140**, pp.107-113, 1993.