ENHANCING MEMORY-BASED PARTICLE FILTER WITH DETECTION-BASED MEMORY ACQUISITION FOR ROBUSTNESS UNDER SEVERE OCCLUSION

Dan Mikami, Kazuhiro Otsuka, Shiro Kumano and Junji Yamato
NTT Communication Science Laboratories, NTT, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 243-0198, Japan

Keywords: Pose Tracking, Face Pose, Memory-based Prediction, Memory Acquisition.

Abstract: A novel enhancement for the memory-based particle filter is proposed for visual pose tracking under severe occlusions. The enhancement is the addition of a detection-based memory acquisition mechanism. The memory-based particle filter, M-PF, is a particle filter that predicts prior distributions from past history of target state, which achieved high robustness against complex dynamics of a tracking target. Such high performance requires sufficient history stored in memory. Conventionally, M-PF conducts online memory acquisition which assumes simple target's dynamics without occlusions for guaranteeing high quality histories. The requirement of memory acquisition narrows the coverage of M-PF in practice. In this paper, we propose a new memory acquisition mechanism for M-PF. The key idea is to use a target detector that can produce additional prior distribution of the target state. We call it M-PFDMA for M-PF with detection-based memory acquisition. The detection-based prior distribution well predicts possible target position/pose even in limited visibility conditions caused by occlusions. Such better prior distributions contribute to stable estimation of target state, which is then added to memorized data. As a result, M-PFDMA can start with no memory entries but soon achieve stable tracking even under severe occlusions. Experiments confirm M-PFDMA's good performance in such conditions.

1 INTRODUCTION

Visual object tracking has been acknowledged as one of the most important techniques in computer vision (Comaniciu et al., 2003), and is required for a wide range of applications such as automatic surveillance, man-machine interfaces (Bradski, 1998; Tua et al., 2007), and communication scene analysis (Otsuka et al., 2008).

For visual object tracking, Bayesian filter-based trackers have been acknowledged as a promising approach; they represent a unified probabilistic framework for sequentially estimating the target state from an observed data stream (Gordon et al., 1993). At each time step, the Bayesian filter computes the posterior distribution of the target state by using observation likelihood and the prior distribution. One implementation, the particle filter (Isard and Blake, 1998), has been widely used for target tracking. It represents probability distributions of the target state by a set of samples, called particles. Particle filter, in short PF, can potentially handle non-Gaussian, nonlinear dynamics/observation processes; this contributes to robust tracking. However, most particle filter-based visual trackers are rather constrained since they employ linear, Gaussian, and time invariant dynamics for simplicity.

Mikami et al. focused on the issue of simplicity degrading PF robustness in real-world situations including abrupt movements and occlusions. To deal with the target's non-Markov, non-Gaussian, and time-varying dynamics, they proposed a memory-based particle filter, called M-PF, as an extension of the particle filter (Mikami et al., 2009). M-PF eases the Markov assumption of PF and uses past history of the target's states to predict the prior distribution on the basis of the target's long-term dynamics. M-PF offers robustness against abrupt object movements and quick recovery from tracking failure. However, such high performance can be achieved only if target history in memory is voluminous and high quality. The first implementation of M-PF in (Mikami et al., 2009) includes an online memory acquisition period which requires the capture of simple dynamics without occlusions for assuring stable tracking. Demanding memory acquisition in this manner narrows the

Mikami D., Otsuka K., Kumano S. and Yamato J..
ENHANCING MEMORY-BASED PARTICLE FILTER WITH DETECTION-BASED MEMORY ACQUISITION FOR ROBUSTNESS UNDER SEVERE OCCLUSION.
DOI: 10.5220/0003808302080215
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 208-215
ISBN: 978-989-8565-04-4

coverage of M-PF.

To acquire memory more stably in a wider range of real-world situations, this paper proposes a new memory-acquisition mechanism for the M-PF-based tracker. The key idea is combining M-PF with a target detector. The target detector can find the target position/pose even in cluttered conditions, and the detection result is used for creating an additional prior distribution of target position/pose. This detection-based prior distribution and the original memory-based prior distribution are combined and provided to the posterior distribution estimation step. Such combined prior distribution prediction contributes to more stable estimation of the target state even in limited visibility conditions. This estimated target state is then added to memory, and is used for creating the memory-based prior distribution in future steps. This cycle of detection, combined prior distribution prediction, posterior distribution estimation, and memory accumulation is highly synergistic in terms of boosting M-PF performance in real-world environments. We name it M-PFDMA for M-PF with detection-based memory acquisition. M-PFDMA has the following advantages; stable initial tracking without memory, quick recovery after occlusion, and wider pose range recoverability.

To verify the effectiveness of M-PFDMA, we implement a facial pose tracker. Facial pose tracking should yield attributes of the face such as position and rotation. As the object detector, we use the "joint probabilistic increment sign correlation face detector," in short JPrISC face detector, the multi-view face detector proposed by Tian et al. (Tian et al., 2010). The JPrISC face detector can detect faces from frontal view to near profile view, and can output ten face pose classes. Facial pose tracking experiments verify that M-PFDMA can acquire target's state history even under severe occlusion and achieves accurate tracking and high recoverability.

In the context of human pose tracking, the idea of *tracking-by-detection* has been gaining attention in recent years as a possible alternative to the traditional target tracking approach (Murphy-Chutorian and Trivedi, 2008; Ozuysal et al., 2006; Andriluka et al., 2010); it reflects the rapid progress in object detectors. However, the current human facial pose detectors are inadequate for realizing mature tracking-by-detection since they are not fast enough for real-time tracking, not accurate enough for determining precise target position/pose, and not able to well handle target dynamics. Rather than relying on just the detector, combining the detector with a tracker has been seen as a reasonable solution. One example was proposed by (Kobayashi et al., 2006), (Ba and Odobez, 2008). They combined a PF-based tracker with a face detector for observation (Kobayashi et al., 2006) and prior distribution prediction (Ba and Odobez, 2008). In (Kobayashi et al., 2006), a multi-pose class face detector provides multiple choices with regard to the observation function but the tracking pose resolution is limited to detectable pose classes. In (Ba and Odobez, 2008), a head position detector yields a uniform prior distribution over all pose ranges, which limits pose accuracy in tracking. Their usage of detectors improved PF-based tracking, but they targeted only a simple environment with no occlusions. On the contrary, our target is more practical and so occlusions are considered. M-PFDMA aims at high recoverability from occlusions. The main feature of M-PFDMA is integrating the object detector into the M-PF framework; the object detector contributes to fast and stable acquisition of target state memory. Eventually, the improved memory contents leads to fast and reliable recovery from occlusions. The synergetic effects of M-PFDMA separate it from previous combination methods.

The remainder of this paper is organized as follows; Sect. 2 briefly reviews memory-based particle filter and its recoverability from tracking failure. Section 3 proposes our new facial pose tracker. Section 4 details the experimental environment and results. Finally, Sect. 5 concludes and discusses our proposal.

2 MEMORY-BASED PARTICLE FILTER AND ITS TRACKING RECOVERABILITY

This section overviews the memory-based particle filter and addresses the recovery problem from tracking failure, which has, up to now, been an unsolved problem in the field of tracking.

2.1 Memory-based Particle Filter

M-PF (Mikami et al., 2009) realizes robust target tracking without explicit modeling of the target's dynamics even when the target moves quickly. Figure 1 outlines M-PF. M-PF keeps the temporal sequence of past state estimates $\widehat{\boldsymbol{x}}_{1:T} = \{\widehat{\boldsymbol{x}}_1, \cdots \widehat{\boldsymbol{x}}_T\}$ in memory. Here, $\widehat{\boldsymbol{x}}_{1:T}$ denotes a sequence of state estimates from time 1 to time T, and $\widehat{\boldsymbol{x}}_t$ denotes a pose estimate at time t. M-PF assumes that the subsequent parts of past similar states provide good estimates of the current future. M-PF introduced Temporal Recurrent Probability (TRP), which is a probability distribution

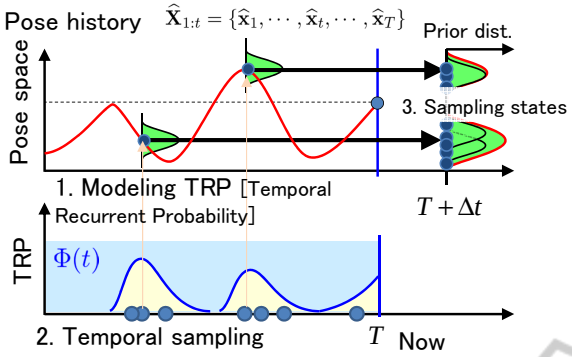


Figure 1: M-PF employs past state sequences to predict a future state. First, it calculates the reoccurrence possibility of past state estimates (TRP). Past time steps are then sampled on the basis of TRP. Past state estimates corresponding to the sampled time steps are combined to predict prior distribution. M-PF enables the implicit modeling of complex dynamics.

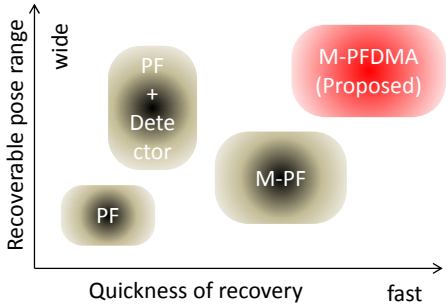


Figure 2: Properties of facial pose tracker with regard to recovery speed and recoverable pose range; PF is only able to rediscover the target if it takes a pose similar to the pose prior to tracking loss. By integrating a detector, PF+Detector enables recovery if the target can be detected. M-PF enables recovery if the target takes a stored pose. M-PF-based recovery can find the target faster than the detector. M-PFDMA supports memory-based quick recovery and detection-based wide pose range recovery.

defined in the temporal domain that indicates the possibility that a past state will reappear in the future. To predict the prior distribution, M-PF starts with TRP modeling. It then conducts temporal sampling on the basis of TRP. The sampled histories are denoted by blue dots in Fig. 1. It retrieves the corresponding past state estimates for each sampled time step, denoted by pink dots in Fig. 1. After that, considering the uncertainty in the state estimates, each referred past state is convoluted with kernel distributions (light green dist. in Fig. 1), and they are mixed together to generate the prior distribution (green dist. in Fig. 1). Finally, a set of particles is generated according to the prior distribution (blue dots in right part of Fig. 1). The M-PF-based face pose tracker in (Mikami et al., 2009) estimates the position and rotation at each time step. M-PF uses the same observation process as traditional PF, which uses a single template built at initialization. This yields the 50 degree face rotation limit noted in (Mikami et al., 2009).

2.2 Recovery from Tracking Failure

The conventional visual trackers track a target with the assumptions of simple dynamics and excellent visibility. Severe occlusions remained a challenging problem, i.e. how can a tracker rediscover the lost target under severe occlusion. This occlusion recovery problem can be viewed from two aspects; *quickness of recovery* and *recoverable pose range*.

The quickness of recovery indicates how rapidly a tracker can rediscover a target after the target reappears after being lost due to occlusion. The recoverable pose range indicates the pose range within which the tracker can rediscover the face, we must expect the target pose to change significant during an occlusion. Conventional methods can be mapped as in Fig. 2 according to these aspects.

The conventional PF-based tracker tries to rediscover a lost target by simply broadening prior distribution frame by frame according to random walk dynamics. It may be able to rediscover a lost target if the occlusion period is short and pose changes is small. However, as time passes and a pose changes significantly, rediscovery probability falls dramatically.

Original M-PF (with simple online memory acquisition) can rediscover the lost target if the target takes a pose that is stored in memory by memory-based prior distribution prediction. Such memory-based rediscovery is faster than detection-based rediscovery or PF-based rediscovery. This is because memory-based prior distribution can well predict possible poses/positions of the target after occlusion. Therefore, M-PF provides more rapid recovery and wider recoverable pose range.

Combinations of particle filtering and detector, denoted by PF+Detector in Fig. 2, are able to rediscover the lost target if the target takes detectable poses. Though the detectable poses and the required time for detection depend on the detector's performance, generally speaking, detectors have much higher computational costs than trackers. To detect targets that take a greater variety of poses, computational costs become even higher. This also means that the recovery speed tends to be rather slow.

M-PFDMA aims at achieving faster recovery in wide pose ranges by integrating detection-based memory acquisition into the memory-based particle filter. Detection-based prior distribution helps rediscovery of targets that take previously unobserved

poses. Rediscovered and tracked positions/poses are stored in history. Acquired history can be used to improve memory-based prior distribution prediction. This synergetic combination of detection-based memory acquisition and memory-based prior distribution prediction enables faster recovery in wide pose ranges.

3 MEMORY-BASED PARTICLE FILTER WITH DETECTION-BASED MEMORY ACQUISITION: M-PFDMA

This section proposes an enhancement of M-PF for object tracking called M-PFDMA. It stands for M-PF with detection-based memory acquisition. It achieves quick recovery in wide pose range due to its synergistic combination of an object detector and tracker.

As reviewed in Sect. 2.1, the basic assumption of M-PF was that the target repeats similar movements again and again. On the basis of this assumption, M-PF introduced the temporal recurrent probability (TRP), which indicated the tendency of past similar states reappearing in the future. M-PF replaced the simple dynamics model employed in conventional particle filters, such as the random walk model, by the temporal recurrent probability, for predicting prior distribution, called memory-based prior distribution. M-PFDMA yields detection-based prior distribution in addition to the memory-based prior distribution. The detection-based prior distribution is folded into memory-based prior distribution. This integration of an effective detector into a unified M-PF framework is the key contribution of M-PFDMA.

The remainder of this section first overviews M-PFDMA, and then, describes the prior distribution prediction formulation. Finally, we describe its implementation in a facial pose tracker.

3.1 System Overview

M-PFDMA integrates an object detector into the M-PF framework. Figure 3 illustrates the block diagram of M-PFDMA; the differences from M-PF are hatched. M-PFDMA has four main components; initialization, prior distribution prediction, posterior distribution prediction, and tracking result estimation. In the initialization step, the tracker detects a target and makes a target model. In the prior distribution calculation step, it calculates prior distribution on the basis of two clues. One is memory-based prior distribution prediction, which is described in Sect. 2.1. The other is detection-based prior distribution estimation; detection-based probability distribution is generated on the basis of the detection results. The memory-based and detection-based prior distributions are then combined, which is described in Sect. 3.2 in more detail. The posterior distribution calculation step calculates observation likelihood, and then, by using likelihood and prior distribution, calculates posterior distribution. Finally, a pose estimate is obtained by weighted averaging and stored in memory. The steps from determining the prior distribution to pose estimation are repeated in each frame.

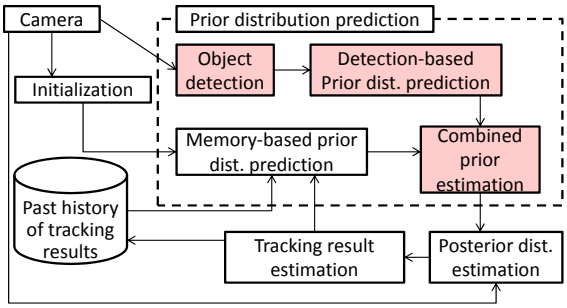


Figure 3: Block diagram of our facial pose tracker. It has four main components; initialization, prior distribution prediction, posterior distribution prediction, and tracking result estimation. The key differences of M-PFDMA from M-PF are hatched, i.e. integration of detection-based prior distribution prediction into memory-based prior distribution prediction. It enables rediscovery of a lost target even if the target takes a position/pose which has not been stored in memory while occlusions. As the more position/pose are added into the memory, the better the memory-based prior distribution prediction becomes. The synergetic effect between detection and tracking is the key contribution of the proposed M-PFDMA.

3.2 Formulation of Prior Distribution Prediction of M-PFDMA

M-PFDMA's key extension from M-PF resides in its prior distribution prediction parts. Unlike M-PF, M-PFDMA employs an object detector and combines memory-based prior distribution and detection-based prior distribution. This section formulates the prior distribution of M-PFDMA.

Bayesian filters, including particle filter, calculate prior distribution by multiplying the observation likelihood by the motion dynamics of target state as in

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{Z}_{1:t}) = \int p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}) \cdot p(\boldsymbol{x}_{1:t}|\boldsymbol{Z}_{1:t}) d\boldsymbol{x}_{1:t}, \quad (1)$$

where $\boldsymbol{x}_t$ denotes the state vector indicating position and rotation, $\boldsymbol{x}_{1:t} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_t\}$ denotes the state sequence of state vector from time 1 to t, and $\boldsymbol{Z}_{1:t} =$

$\{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_t\}$ denotes the sequence of observations from time 1 to t.

As the dynamics model, conventional particle filters assume a short term Markov model as in

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}) \approx p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t). \quad (2)$$

The memory-based particle filter assumes the tendency of repeating past positions/poses, and introduced the temporal recurrent probability $\Phi(\cdot)$. It replaced the dynamics model with the temporal recurrent probability given by

$$p(\boldsymbol{x}_{t+\Delta t}|\boldsymbol{x}_{1:t}, \Delta t) = \sum_{\tau=1}^{t} \Phi(t|\widehat{\boldsymbol{x}}_{1:t}, \Delta\tau) \cdot K(\boldsymbol{x}_{t+\Delta\tau}|\widehat{\boldsymbol{x}}_\tau),$$

where Δt denotes the time offset between current time and prediction target and $\widehat{\boldsymbol{x}}$ denotes a point estimate stored in memory. $K()$ is the kernel distribution that represents uncertainty in the stored state estimate.

M-PFDMA replaces the memory-based prior distribution by a combination of memory-based prior distribution and detection-based prior distribution calculated based on object detection results $\widetilde{\boldsymbol{x}}_j(z_t)$ as

$$p(\boldsymbol{x}_{t+\Delta t}|\boldsymbol{x}_{1:t}, \Delta t, z_{t+\Delta t}) = \alpha \sum_{j=1}^{N_d} q(\boldsymbol{x}_t|\widetilde{\boldsymbol{x}}_j(z_t))$$

$$+(1-\alpha) \sum_{\tau=1}^{t} \Phi(t|\widehat{\boldsymbol{x}}_{1:t}, \Delta\tau) \cdot K(\boldsymbol{x}_{t+\Delta\tau}|\widehat{\boldsymbol{x}}_\tau), \quad (3)$$

where α denotes the mixing weight between memory-based prior distribution (first part of (4)) and detection-based prior distribution $q()$ (latter part of (4)), and N_d denotes the number of detected objects. The detection result $\widetilde{x}_j(z_t)$ includes estimated position and pose, and $q(x|\widetilde{\boldsymbol{x}}_j(z_t))$ denotes Gaussian distributions with mean for each position/pose. In this paper, a static predefined value $\alpha = 0.15$, which is independent of the number of detected objects, is employed for the mixing weight α.

The combined prior distribution realizes high recoverability and accuracy, which yields quick and stable memory acquisition.

3.3 Implementation in a Facial Pose Tracker

We implement a facial pose tracker on the basis of M-PFDMA. As the detection method, the multi-view face detector proposed by Tian et al. (Tian et al., 2010), called a JPrISC face detector, is employed [1]. The multi-view face detector is able to detect faces and to output ten pose classes including frontal view and near profile view. It achieved relatively fast detection without degrading detection accuracy by using a calculation sequence determined by entropy.

[1]Though we used the JPrISC as the face pose detector, it is not specialized in face pose. By collecting training images, it can be applied for detecting other objects.

4 EXPERIMENT

To confirm M-PFDMA's performance, we focus on severe occlusion cases because stable tracking is already possible with conventional methods if the target is not occluded, even if it exhibits complex dynamics.

4.1 Experimental Settings

Video capturing environment is as follows. We used PointGreyResearch's FLEA, a digital color camera, to capture 1024 x 768 (pixels) images at 30 frames per second. Note that the tracking processes uses only grayscale images converted from the color images. The CPU of the PC used was an Intel Core2Extreme 3.0GHz (Quad Core) and the GPU was NVIDIA GeForce GTX480. All experiments used 2000 particles.

Our tracker was implemented on the basis of STC-Tracker (Lozano and Otsuka, 2008), which accelerates particle filtering by GPU implementation. It can run at 30 fps, and our M-PFDMA-based tracker also can run at 30 fps.

4.2 Typical Example of Proposed Facial Pose Tracker in Action

We compare M-PFDMA to three other methods. First one is the memory-based particle filter (M-PF) (Mikami et al., 2009), the second one is a combination of particle filtering and face detector called the JPrISC face detector (PF+Detector). Note that the JPrISC face detector can output ten pose classes in addition to position. Therefore, PF+Detector is expected to show higher performance than similar method in (Ba and Odobez, 2008) especially in terms of speed of recovery. And the third one is FaceAPI (Seeingmachines,), which is known as the best face tracker that is commercially available. The FaceAPI can detect a face in near frontal view, and then can sequencially estimate position and pose of the face. Note that because FaceAPI is a commercial facial pose tracker, the detailed algorithm is not apparent.

We prepared two videos with severe occlusions. In the first video, objects horizontally and vertically cross at the camera's centerline as in Fig. 8, and cause occlusions. Tracking starts in the top-right area, and the subject moves up-down and left-right while

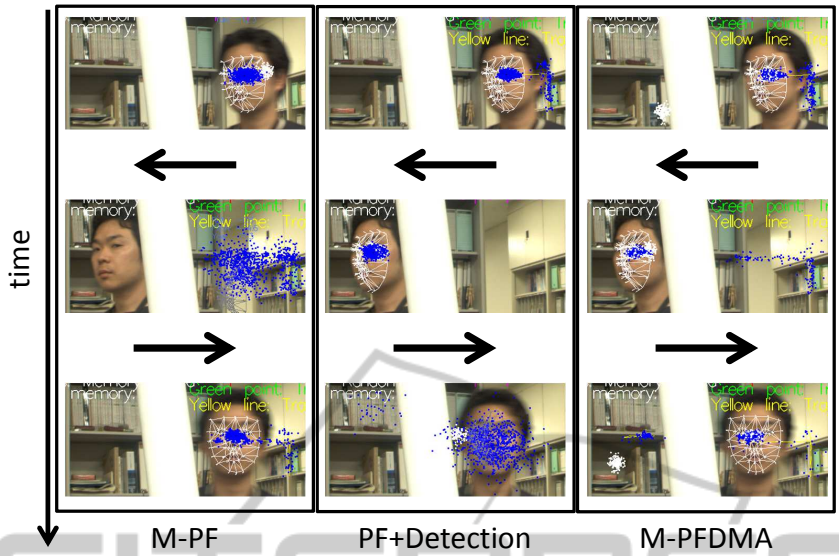


Figure 4: Tracking results of M-PF, PF+Detector, and M-PFDMA (proposed). White mesh denotes estimated position/pose. While tracking is unstable, the mesh turns gray. The left column shows M-PF's output, the middle column shows PF+Detector's output, and the right column shows M-PFDMA's output. In each column, figures are listed in time order.

changing his pose. The second video simulates video conference situation; the subject makes a presentation in front of a camera, interacting with participants on the other side.

Figure 4 shows snapshots of the tracking behavior of M-PF, PF+Detector, and M-PFDMA, from the first video. Only upper half images are shown. In Fig. 4, left column, middle column, and right column show the result of M-PF, PF+Detector, and M-PFDMA, respectively. In each column, figures are listed in time order. From first row to second row, the target face moved from right to left; and from second row to third row, it moved from left to right. Second row shows that M-PFDMA and PF+Detector successfully detected the target face while M-PF didn't, because the target's pose had not been stored. It is one example of improvement in recoverable pose range. This recovery confirms the effectiveness of the detector. However, these recoveries were achieved by the detector, so they took rather a long time. Third row shows that M-PF and M-PFDMA found the lost target, while PF+Detector took much longer to rediscover it. In this situation, past state history of right side had already been stored, so the memory-based methods quickly rediscovered the lost target.

FaceAPI is able to detect a face in near frontal view, so it successfully rediscovered the tracking target after occlusions if the target was near frontal view and was observed without blur. While a target is moving, the observed image becomes blurred. Therefore, FaceAPI failed to rediscovered the lost target while it was moving even if it had not been occluded. On the contrary, M-PFDMA quickly rediscover the lost target under blurred observation. Thus, M-PFDMA outperformed other trackers including PF-based trackers and a commercial non-PF-based tracker.

Figure 5 shows snapshots of the tracking behavior of M-PFDMA from the second video. The snapshots are listed in time order. The second video includes self-occlusion, e.g., turning back while moving left to right (Fig. 5(b)-(e)) and right to left (Fig. 5(g)-(h)). It also includes scale changes (Fig. 5(e)-(f)) and non-rigid deformations (Fig. 5(g)), e.g., facial expression changes. During natural behaviors as depicted in Fig. 5, our M-PFDMA successfully estimates the position and rotation of the target while the target is not occluded.

The above comparisons verified that M-PFDMA supports both memory-based quick recovery and detection-based recovery and so covers the cases in which the target deviates from stored poses.

4.3 Quantitative Evaluations

For quantitative evaluations, the same videos were used. First, tracking success ratio was examined. The tracking success ratio is the ratio of frames in which the tracker estimated pose and position correctly as confirmed by manual observation. The video includes numerous occlusions. Therefore, recovery performance, i.e. speed of recovery and recoverable pose range directly impacts the tracking success ratio. Comparison targets are M-PF, PF+Detector, and FaceAPI.

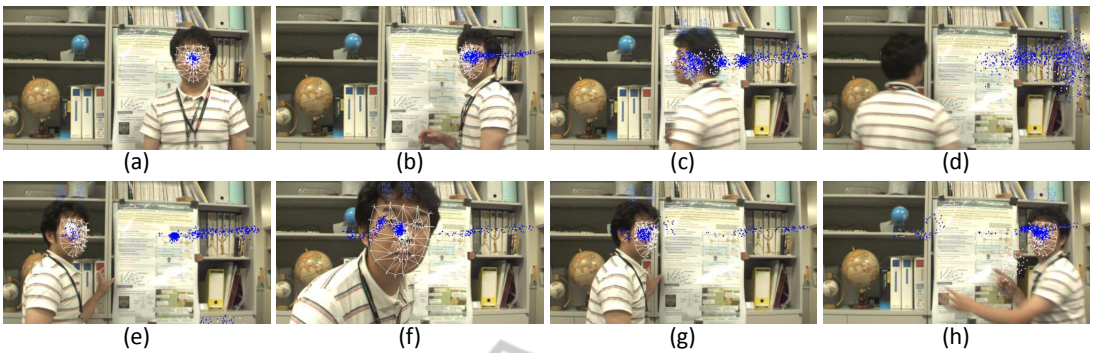


Figure 5: Tracking results for M-PFDMA (proposed). White mesh denotes estimated position/pose. While tracking is unstable, the mesh turns gray. (a) Initialization. (b)-(e) The subject moves from right to left while rotating; (e)-(g) the subject gets close to and backs away it; (g)-(h) the subject moves from left to right while rotating. During such natural behavior of the subject, the proposed M-PFDMA occasionally missed the target. However, it rediscovered the target soon after occlusions.

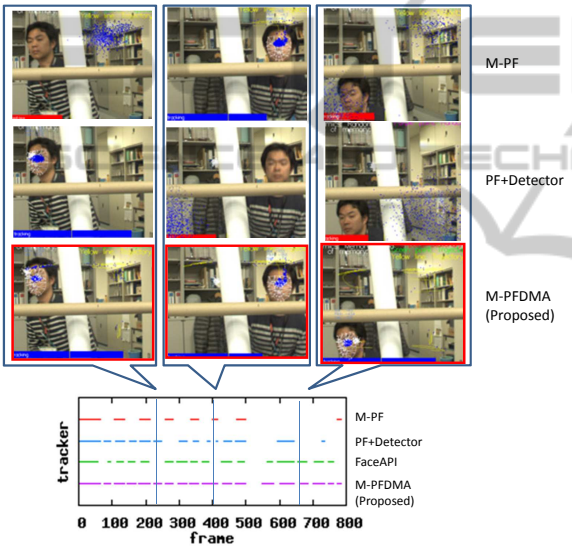


Figure 6: Tracking results of M-PF, PF+Detector, FaceAPI, and M-PFDMA are shown. M-PFDMA achieved faster recovery of lost target than the others. In frame 225, PF+Detector and M-PFDMA successfully estimated the position/pose of the target, whereas M-PF didn't because history wasn't stored close to the position/pose (left column). In frame 400, M-PF and M-PFDMA successfully estimated the position/pose of the target, but PF+Detector didn't because detection requires large computation time (middle column). In frame 655, only M-PFDMA successfully estimated the position/pose of the target (right column).

The results, shown in Table 1, confirm that M-PFDMA successfully estimated the target's position and rotation in 81.5%/90.1% (video 1/video 2) of frames, whereas M-PF, PF+Detector, and FaceAPI achieved only 37.5%/36.9%, 52.4%/62.2%, and 64.2%/53.7%, respectively. Figure 6 shows a part of the tracking results for video 1 in more detail. In Fig. 6, the horizontal axis denotes frame number. Each line denotes a tracking result. The discontinuities indicate the frames in which target position/pose were not correctly estimated. Figure 7 shows tracking results for video 2. As shown in Figs. 6 and 7, M-PFDMA combined the frames correctly estimated by PF+Detector and M-PF while avoiding their deficiencies.

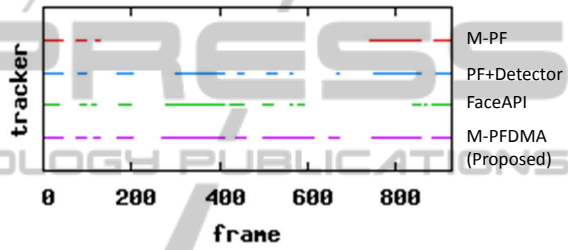
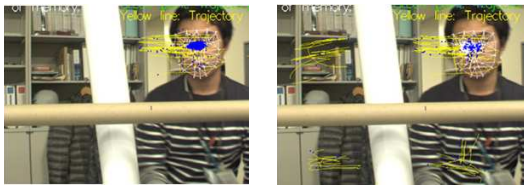


Figure 7: Tracking results of M-PF, PF+Detector, FaceAPI, and M-PFDMA are shown. M-PFDMA achieved faster recovery of lost target than the others. Compared with Fig. 6, a similar property is observed.



(a) M-PF (b) M-PFDMA (proposed)

Figure 8: Stored history by each method; (a) M-PF and (b) M-PFDMA (proposed method). Positions of stored history are shown by yellow lines. M-PF obtained history only in right-top area, where the target face was initialized. Contrary, M-PFDMA obtained hitory in wider area.

Next, the memory acquisition results of the previous M-PF and the proposed M-PFDMA from video 1 are shown in Fig. 8. Yellow lines denote the position of the history stored in memory. So, target's movements along the yellow lines are supposed to be tracked stably on the basis of memory-based prior

Table 1: Comparisons of tracking successful ratio among M-PF, PF+Detector, FaceAPI, and M-PFDMA (proposed). When we calculated the tracking successful ratio, the frames which were not able to track due to occlusion were excluded.

	successful tracking ratio	
Method	video 1	video 2
M-PF	37.5%	36.9%
PF+Detector	52.4%	62.2%
FaceAPI	64.2%	53.7%
M-PFDMA (proposed)	81.5%	90.1%

distribution prediction. The left figure shows the output of M-PF, and the right figure shows that of M-PFDMA. As shown in Fig. 8, M-PF stored history covered only the top-right area, where tracking started. This means that the tracker failed to rediscover target when the target moved to other areas after occlusions due to large changes in position/pose while occlusions. On the contrary, the stored memory of M-PFDMA covered the entire field of view. The numbers of stored memories of this sequence by M-PF and M-PFAP are 455 and 808, respectively. The memory acquisition performance of M-PFDMA under severe occlusion was confirmed.

5 CONCLUSIONS

A memory-based particle filter with detection-based memory acquisition, in short M-PFDMA, was proposed for vision-based object tracking. M-PFDMA offers robust memory acquisition under severe occlusion since it creates a synergistic combination of detection-based memory acquisition and the memory-based approach. M-PFDMA was shown to achieve high accuracy and quick recovery in real-world situations. We verified its effectiveness in facial pose tracking experiments.

Future works include memory management. M-PFDMA stores the correctly estimated target state in memory. The correctness is automatically judged by using the maximum likelihood among particles. Though it works well in most cases, the quality of stored data is very important for memory-based prior distribution prediction. Therefore, we will consider more precise ways of judging tracking correctness. Future works also include automatic determination of the mixing weight α of detection-based and memory-based prior distribution. Results on this paper employed static α. However, the ideal α varies by conditions such as current tracking stability. We would like to reveal factors affecting α, and then, would like to tackle automatic setting of α according to the factors.

REFERENCES

Andriluka, M., Roth, S., and Schiele, B. (2010). Monocular 3D pose estimation and tracking by detection. In *CVPR 2010*, pages 623–630.

Ba, S. and Odobez, J.-M. (2008). Probabilistic head pose tracking evaluation in single and multiple camera. *Multimodal Technologies for Perception of Humans*, 4625/2008:276–286.

Bradski, G. R. (1998). Computer vision face tracking for use in a perceptual user interface. In *Proc. IEEE Workshop Applications of Computer Vision*, pages 214–219.

Comaniciu, D., Ramesh, V., and Meer, P. (2003). Kernel-based object tracking. *IEEE Trans. PAMI*, 25:564–577.

Gordon, N., Salmond, D., and Smith, A. F. M. (1993). Novel approach to non-linear and non-gaussian bayesian state estimeation. *IEE Proc.F:Communications Rader and Signal Processing*, 140(2):107–113.

Isard, M. and Blake, A. (1998). Condensation - conditional density propagation for visual tracking. *IJCV*, 29(1):5–28.

Kobayashi, Y., Sugimura, D., Sato, Y., Hirasawa, K., Suzuki, N., Kage, H., and Sugimoto, A. (2006). 3D head tracking using the particle filter with cascaded classifiers. In *BMVC*.

Lozano, O. M. and Otsuka, K. (2008). Real-time visual tracker by stream processing. *Journal of VLSI Signal Processing Systems*.

Mikami, D., Otsuka, K., and Yamato, J. (2009). Memory-based particle filter for face pose tracking robust under complex dynamics. In *Proc. CVPR*, pages 999–1006.

Murphy-Chutorian, E. and Trivedi, M. M. (2008). Head pose estimation in computer vision: A survey. *IEEE Trans. PAMI*.

Otsuka, K., Araki, S., Ishizuka, K., Fujimoto, M., Heinrich, M., and Yamato, J. (2008). A realtime multimodal system for analyzing group meeting by combining face pose tracking and speaker diarization. In *Proc. ACM ICMI*, pages 257–264.

Ozuysal, M., Lepetit, V., Fleuret, F., and Fua, P. (2006). Feature harvesting for tracking-by-detection. In *LNCS 3953(ECCV2006)*, pages 592–605.

Seeingmachines. FaceAPI. http://www.seeingmachines.com/product/faceapi/.

Tian, L., Ando, S., Suzuki, A., and Koike, H. (2010). A probabilistic approach for fast and robust multi-view face detection using compact local patterns. In *Proc. the IIEEJ Image Electronics and Visual Computing Workshop*.

Tua, J., Taob, H., and Huang, T. (2007). Face as mouse through visual face tracking. *CVIU*, 108(1-2):35–40.