

Paper Title :

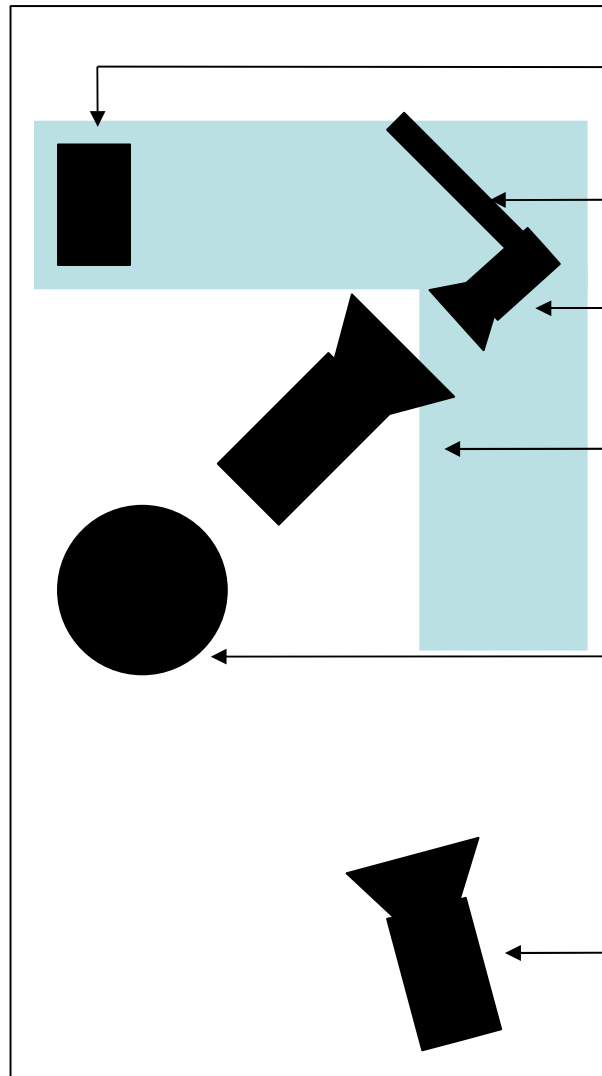
“Memory-based Particle Filter for Face Pose Tracking
Robust under Complex Dynamics”

The “movies.zip” file includes demonstration movies as follows;

- README.pdf: this file. Capturing environment (P.2), explanations of each video file (P.3 – P. 10) .
- movie1.wmv: description of our method,
- movie2.wmv: comparisons of our method and previous one described in Section 5.2,
- movie3.wmv: demonstrations of the head motion of various people; includes abrupt motion tracking and recovery from tracking failure ,
- movie4a.wmv, movie4b.wmv: face tracking of unseated person

All videos are encoded in WMV (Windows Media Video) format ver. 9.

Description of Capture Environment.



PC: runs face pose tracker. Tracking results are presented on display, and saved to HDD in the PC.

Display: shows tracking results.

Camera1: captures the person being tracked. It sends capture image to PC.

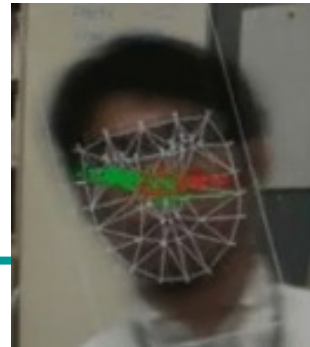
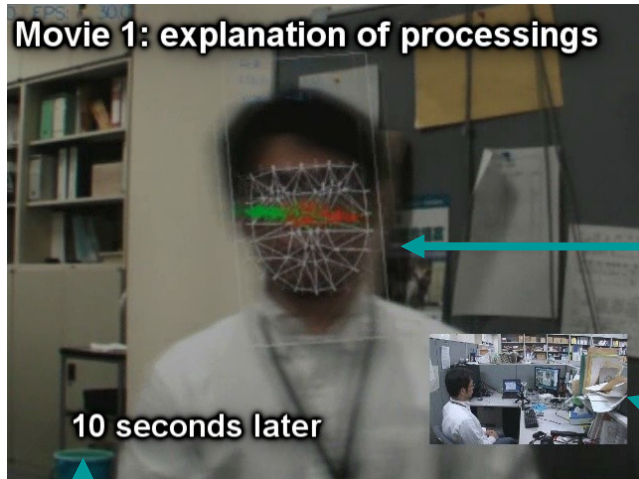
Camera2: captures the display to create demo videos.
(Used for movies 1, 3, 4a, and 4b)

Chair: with Target person.

Camera 3: captures the experimental environment for demo videos.
(Used for movies 1, 3, 4a, and 4b)

Movie 1 : Explanation of processing.

This video explains how the proposed tracker behaves, including initialization, data accumulation, and prior predictions.



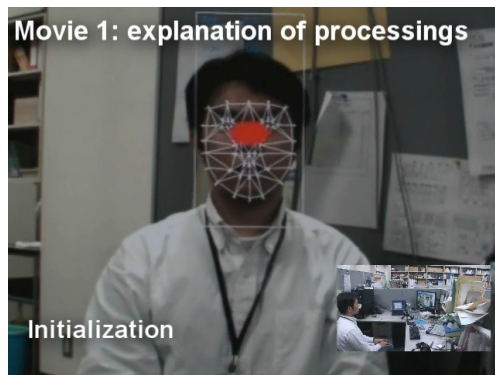
White mesh shows the face model; its position and aspect indicate per tracking result.

Green and red points indicate positions of sampled particles, **green ones** indicate the trajectory-similarity-based samples, and **red ones** indicate the stationary-property-based samples.

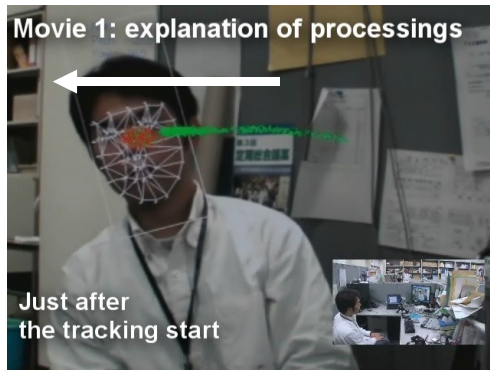
View from Camera 3

Screenshot of PC display captured with Camera 2

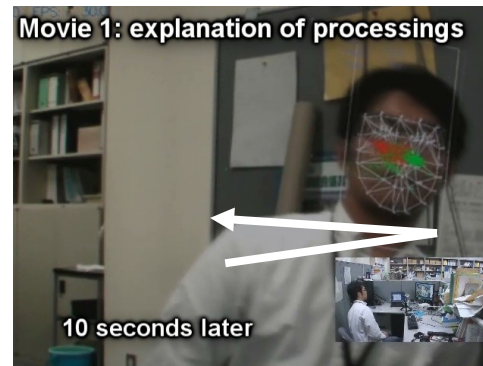
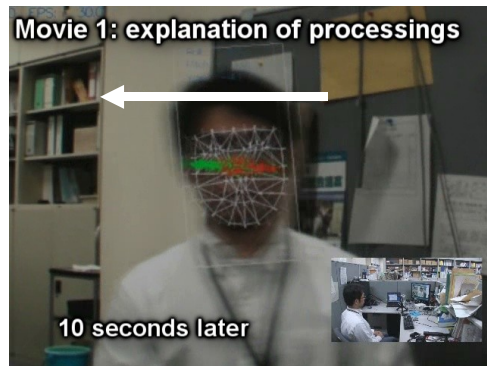
1) The proposed tracker automatically finds the frontal face in the image, and creates face model (face template). When there is no previous data available, our memory-based tracker starts without past data. However, it can track the face as effectively as the previous tracker, but also accumulates the estimates.



2) Just after the tracker starts, the trajectory-based prediction is not effective because there is not enough data to create appropriate predictions.

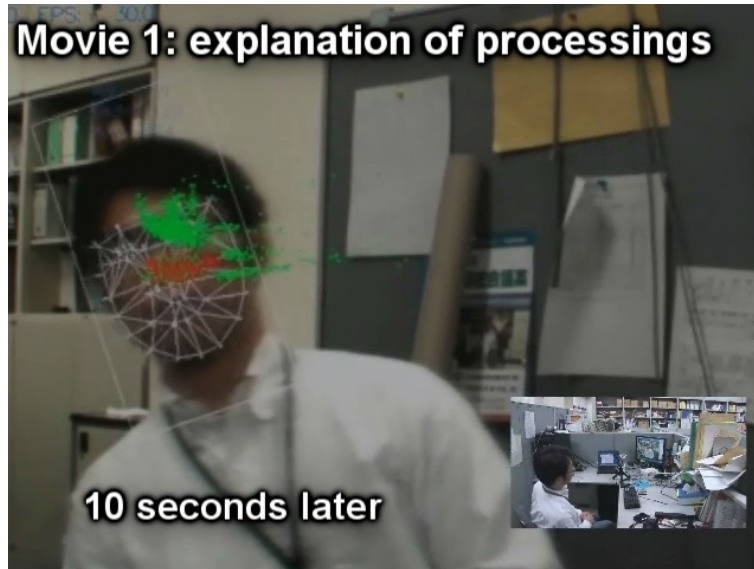


3) After about 10 seconds, the stationary-based samples (red) and the trajectory-based samples (green) develop different distributions. In the case shown here, the person moved his head from left to right, and right to left, repeatedly (swinging behavior). The below picture (left) is a snapshot of a right-to-left movement. The red samples (stationary-based) follow the face, and the green samples (trajectory-based) precede it. This means that the trajectory-based prior prediction well identified the actual motion.



At the moment of change in moving direction (Right above picture), red samples (stationary-based) became a good prediction. Some green samples (trajectory-based) predicted the change in moving direction, but other green samples overshoot the change point.

4) A couple of minutes after initialization, the tracker had accumulated a certain amount of data. In the below picture, green samples (trajectory-based) exhibit complex distributions. Some are curved, some are widely distributed. These distributions reflect past trajectories that the face went through, and though to be possible trajectories in the future. Such prediction based on long-term dynamics (nonlinear, time-variant, and non-Markov) is the main feature of our memory-based PF and tracker. No other tracker can provide comparable read-ahead capability.



Movie 2 : Comparisons of our method to previous one.

These videos and tracking results were used to create the quantitative comparison in Section 5.2 (and 5.3) of our paper. Both trackers were applied to the same pre-recorded videos. The right pane shows the result of the proposed tracker and the left shows that of the previous tracker.

Type I (moderate motion)



Both trackers provide approximately the same tracking performance.

Type II (abrupt motion)



The proposed tracker can handle abrupt motions while the previous tracker cannot.

Type III (occlusion cases)



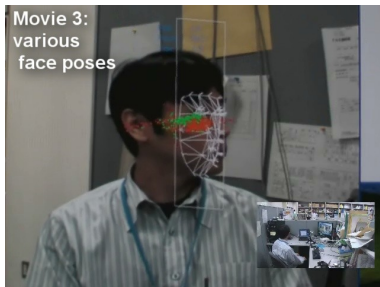
The proposed tracker successfully tracked the face and recovered the face after occlusions. The rate of recovery was 100%.

On the other hand, the previous tracker often lost the face. We had to repeatedly restart the previous tracker after it lost the face. This video shows just four recovery sequences because of file size limitation.

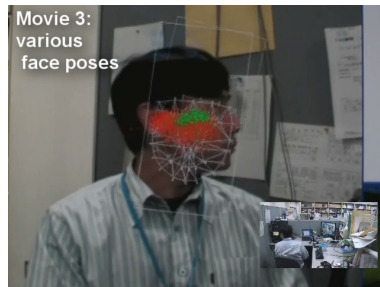
Movie 3: head motion of various people.

This video demonstrates the REALTIME performance of the proposed tracker for various people and motions, including turning back (self-occlusion case) and covering face with hand(s) or objects (mutual-occlusion cases). This video confirms the behavior of the proposed tracker in various situations (not described in our paper), and its effectiveness.

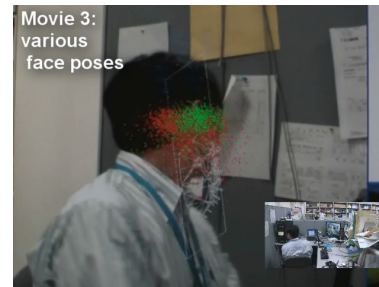
Example sequence of **self-occlusion**



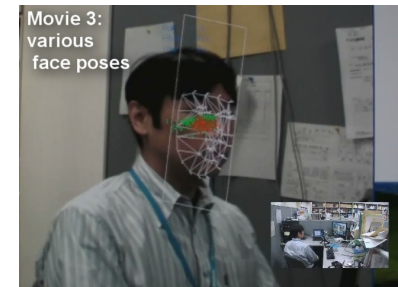
Frame = 362



Frame = 386

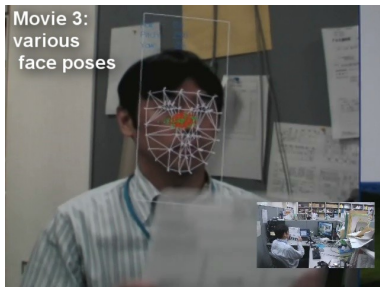


Frame=405

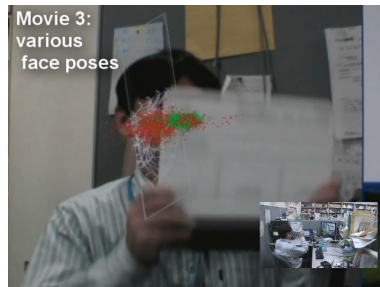


Frame=431

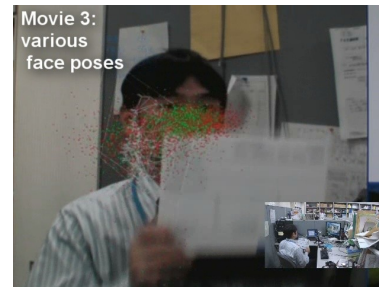
Example sequence of **mutual-occlusion**



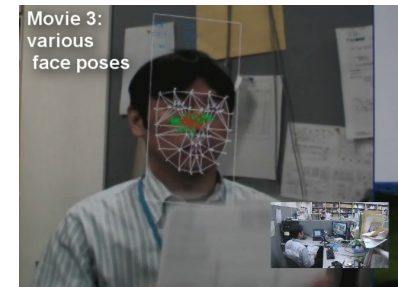
Frame = 549



Frame = 558



Frame=587

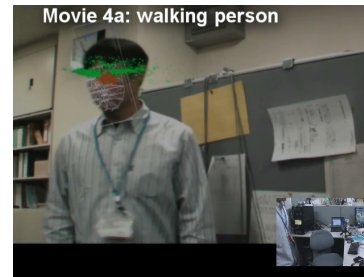
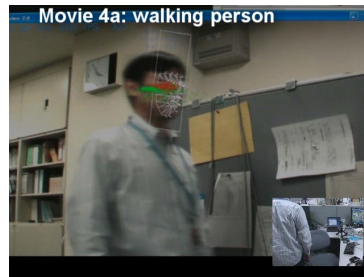
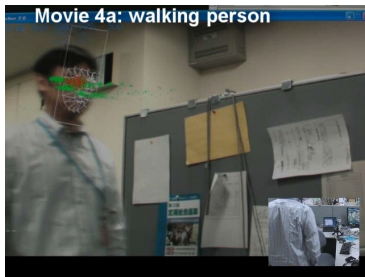


Frame=595

Movie 4 : Standing situations

Our face tracker has the potential to track faces in other situations, such as when the person is standing. The movies demonstrate some examples for further discussion (partially mentioned in Conclusions and Discussion section of the paper). The tracking was done in REALTIME.

Case 4-1) Walking person. Movie4a



Walking person is well tracked. Sometime, occlusions happen, but face is soon recovered.

Supplemental discussion about standing situations

As mentioned in Section 3, the memory-based particle filter is suitable for targets that are physically constrained. For a seated person, his/her waist is fixed on the seat of the chair while the upper body can exhibit articulated motion. In standing situations, the physical constraints are the person's height, the ground, and gravity. It means, that face position is roughly determined by his/her height, and its position in the image by the person's position in world coordinates and relative camera coordinates.

For the case of Movie 4, we used the same tracker as used in the paper. This movie indicates that our tracker, originally designed for seated people works quite well. However, it would be preferable to separately handle the global position of the person (standing position) and head motion relative to the body. If this is possible, the head motion data captured at one place can be reused when the same head motion happens in a different place. This is a promising strategy to extend the targets covered by our memory-based tracker.

Case 4-2) Sudden movement (squat) Movie4b

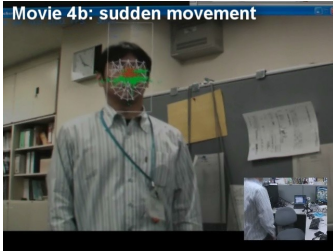


Fig. 4b-1
First squat (fast)

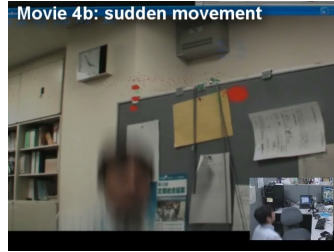


Fig. 4b-2

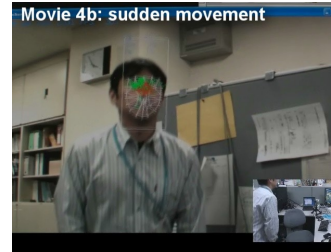


Fig. 4b-3

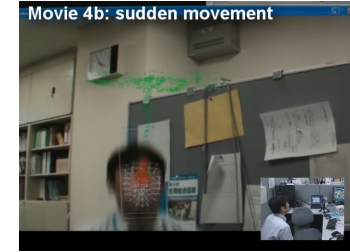


Fig. 4b-4
Second squat (slow)

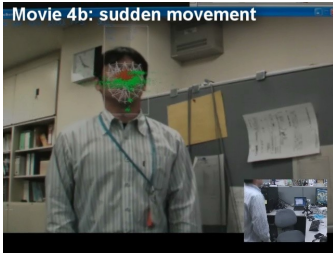


Fig. 4b-4
Third squat (fast)

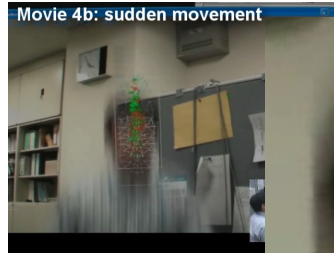


Fig. 4b-5

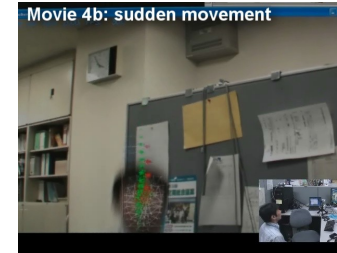
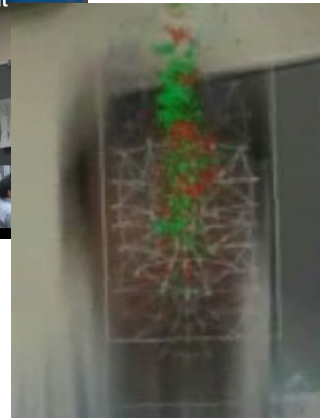


Fig. 4b-6

First, a person suddenly squats very quickly. At the first occasion, the tracker could not follow the motion, because it did not have such a motion in its database. The second time the person performed the motion he slowed down so that the tracker can follow and gather the history. Finally, the person squatted down as quickly as the first time. At this occasion, the tracker successfully followed the moving face.

To make this happen, we introduced *temporal scaling* into the trajectory similarity-based sampling. That is, the current speed of the target is compared to the past speed to scale the rate of the temporal development. This compensates the difference in speed between current and past events. Thus slow motion history can be used to create the prior distribution of the faster motion. This example confirmed that the proposed tracker can handle extremely fast motion and the effectiveness of its online data accumulation process.

Also, this movie verifies that our tracker is extremely robust against degraded image quality. In the final case (Fig. 4b-5), the face image in the motion is quite blurred. Existing feature-point-based trackers like AAM-based tracking can not handle such situations.