3D Head Tracking using the Particle Filter with Cascaded Classifiers

Yoshinori Kobayashi, Daisuke Sugimura, Yoichi Sato
Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JAPAN
{yosinori,sugimura,ysato}@iis.u-tokyo.ac.jp

Kousuke Hirasawa, Naohiko Suzuki, Hiroshi Kage
Advanced Technology R&D Center, Mitsubishi Electric Corporation
8-1-1, Tsukaguchi-honmachi, Amagasaki City, Hyogo 661-8661, JAPAN
Hirasawa.Kosuke@cs.MitsubishiElectric.co.jp
Suzuki.Naohiko@dw.MitsubishiElectric.co.jp
Kage.Hiroshi@dc.MitsubishiElectric.co.jp

Akihiro Sugimoto
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
sugimoto@nii.ac.jp

Abstract

We propose a method for real-time people tracking using multiple cameras. The particle filter framework is known to be effective for tracking people, but most of existing methods adopt only simple perceptual cues such as color histogram or contour similarity for hypothesis evaluation. To improve the robustness and accuracy of tracking more sophisticated hypothesis evaluation is indispensable. We therefore present a novel technique for human head tracking using cascaded classifiers based on AdaBoost and Haar-like features for hypothesis evaluation. In addition, we use multiple classifiers, each of which is trained respectively to detect one direction of a human head. During real-time tracking the most suitable classifier is adaptively selected by considering each hypothesis and known camera position. Our experimental results demonstrate the effectiveness and robustness of our method.

1 Introduction

Tracking people by using camera images has become an important task in computer vision applications along with the spread of surveillance cameras. In fact, tracking people in the real world using camera images is a challenging problem due to changes in appearance caused by their motions, the necessity for low resolution in broad monitoring, changes in color that come from varying light conditions and clutters observed in the background.

A variety of tracking algorithms have been proposed. In particular, over the last few years, the particle filter framework is reported to be effective for tracking people[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13]. The particle filter is a Bayesian sequential importance sampling technique, which recursively approximates the posterior distribution using a finite set of weighted samples. A set of samples can be used to approximate non-Gaussian distribution and they are propagated by a state transition model for each recursion. It thus allows us to realize robust tracking against observation noise and abrupt changes of target's motion.

Most of particle filter based people tracking algorithms use contour similarity or color histogram for hypothesis evaluation[1, 3, 4, 6, 8, 9, 10, 12]. While these simple evaluation methods are relatively less affected by a variety of appearance, they are still not sufficient enough to track a human head in a real environment. This is because evaluating appearance contour with low resolution is difficult in a clutter scene or appearance color may be also affected by a varying light. Therefore, to improve the robustness and accuracy of human head tracking based on the particle filter, we need a more sophisticated hypothesis evaluation method.

On the other hand, numerous methods for detecting faces in general images have been proposed. Among them, the AdaBoost-based face detector using Haar-like features has become popular because of its accuracy and robustness against observation with low resolution or varying illumination conditions. The AdaBoost-based classifier consists of linearly connected weak classifiers. Viola and Jones arranged the classifiers in a cascade structure and proposed an efficient computation technique for Haar-like features[11]. Though the training of AdaBoost-based cascaded classifiers (hereafter referred to in this paper as the 'cascaded classifier') requires huge amount of time, the cascaded classifier rapidly detects a face because most of non-face target regions are rejected in an early stage of the cascade. However, in tracking objects with multiple cameras, it is not an efficient way to search for the various-size objects over a whole image from each camera sequentially and exhaustively. So, it is a straightforward and efficient idea to incorporate cascaded classifiers into the particle filter framework to specify the search region.

In this paper, we present a novel method for 3D human head tracking based on the particle filter framework incorporating cascaded classifiers for hypothesis evaluation. In the particle filter framework, a state of a human head is represented with its 3D position and orientation. A likelihood of a human head is evaluated by multiple cameras with overlapping field of view by using cascaded classifiers.

Okuma et al.[7] succeeded in combining two algorithms: particle filter and cascaded classifier to produce a mixed importance sampling. In contrast, we apply cascaded classifiers to hypothesis evaluation in the particle filter framework. Though Yang et al.[13] used a cascaded classifier in a part of hierarchical process of hypothesis evaluation, the output of cascaded classifier is only used for a binary decision (true or false) to cut off the hierarchical process for the negative. It is only recently proposed by Thierry et al.[2] that the score of classifier is used as the likelihood in the particle filter. However, these methods aimed for 2D tracking using one camera without considering camera's point of view. In contrast, our goal is to track people in 3D using multiple cameras by considering the differences in appearance among these camera images.

In the context of 3D people tracking, methods using multiple cameras are proposed. These methods focus on integrating evaluations by each camera[5, 12]. Therefore, in these methods, each camera employs relatively simple evaluation methods such as color or contours. Nickel et al.[5] succeeded in tracking a speaker in front of audience using a multi-camera based particle filter. This method employs cascaded classifiers partly in an evaluation phase. However, the output of cascaded classifiers is only used for a binary decision and human head orientation or camera's point of view are not concerned. In contrast, we apply cascaded classifiers to hypothesis evaluation where classifiers are adaptively selected depending on human head orientation and camera's point of view.

The contributions of this work are summarized as below. 1)AdaBoost-based cascaded classifiers are used for hypothesis evaluation. 2)Multiple cascaded classifiers trained respectively to detect one direction of a human head are adaptively used in hypothesis evaluation by considering the relationship between human head orientation and camera position. Consequently, Tracking a human head robustly and accurately is realized even in the case of observation with low resolution or under varying light conditions.

2 Particle filtering

In this section we briefly give an overview of the particle filter framework.

Suppose that a state of a target at time t is denoted by the vector $\mathbf{x}_t$, and that the observation of camera image at time t is denoted by the vector $\mathbf{z}_t$. Then all the observations up to time t is $\mathbf{Z}_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$. Assuming the Markov process enables us to describe a prior probability $P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ at time t by

$$P(\mathbf{x}_t \mid \mathbf{Z}_{t-1}) = \int P(\mathbf{x}_t \mid \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1}) d\mathbf{x}_{t-1}, \tag{1}$$

where $P(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1})$ is a posterior probability at time $t-1$, and $P(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ is a state transition probability from $t-1$ to t. Assuming that $P(\mathbf{z}_t \mid \mathbf{Z}_{t-1})$ remains constant, a posterior probability $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ at time t is described by

$$P(\mathbf{x}_t \mid \mathbf{Z}_t) \propto P(\mathbf{z}_t \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{Z}_{t-1}), \tag{2}$$

where $P(\mathbf{z}_t \mid \mathbf{x}_t)$ is a likelihood and $P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ is a prior probability at time t. Tracking is then achieved by calculating the expectation of posterior probability $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ at each time.

In the particle filter framework, the probability distribution is approximated by a set of samples $\{\mathbf{s}_t^{(1)}, \ldots, \mathbf{s}_t^{(N)}\}$. Each sample $\mathbf{s}_t^{(n)}$ representing a hypothesis has the weight $\pi_t^{(n)}$ representing a corresponding discrete sampling probability. The hypothesis evaluation, which is also called as the sample evaluation, is to compute the weight $\pi_t^{(n)}$ by considering the observation likelihood corresponding to the sample $\mathbf{s}_t^{(n)}$. A set of samples is then updated by the following procedures at each time.

1. Sampling:

   Select samples $\{\mathbf{s}'^{(1)}_{t-1}, \ldots, \mathbf{s}'^{(N)}_{t-1}\}$ in proportion to weight $\{\pi_{t-1}^{(1)}, \ldots, \pi_{t-1}^{(N)}\}$ corresponding to sample $\{\mathbf{s}_{t-1}^{(1)}, \ldots, \mathbf{s}_{t-1}^{(N)}\}$.

2. Propagation:

   Propagate samples $\{\mathbf{s}'^{(1)}_{t-1}, \ldots, \mathbf{s}'^{(N)}_{t-1}\}$ with state transition probability $P(\mathbf{x}_t \mid \mathbf{x}_{t-1} = \mathbf{s}'^{(n)}_{t-1})$ and generate new samples $\{\mathbf{s}_t^{(1)}, \ldots, \mathbf{s}_t^{(N)}\}$ at time t.

3. Weight computation:

   Compute weights $\pi_t^{(n)} \approx P(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}_t^{(n)})$ corresponding to sample $\mathbf{s}_t^{(n)}$ by evaluating a likelihood through camera images$(n = 1, 2, \ldots, N)$.

3 AdaBoost-based cascaded classifier

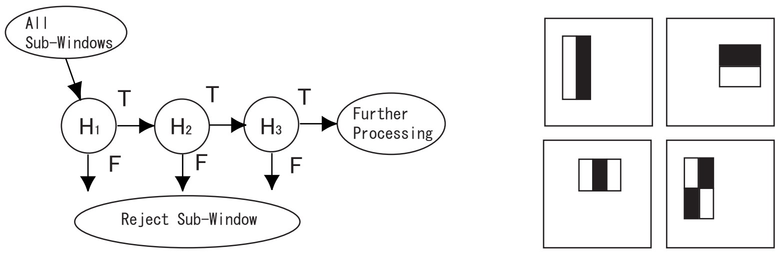
The AdaBoost-based face detector proposed by Viola and Jones[11] has a cascade structure to reduce detection time (Fig.1(a)). This cascade is effective in the evaluation phase even in the particle filter framework.

In Fig.1(a), $\mathbf{H}_i$ represents a strong classifier. Each strong classifier classifies an input image into a positive or a negative. Only positive images are used as the input of the next strong classifier. At each stage, a strong classifier is trained to detect almost all face images while rejecting a certain fraction of non-face images. For instance, the classifier at each stage is trained to eliminate 50% of the non-face images while falsely eliminating is only 0.1% of the face images. After passing 40 stages, we can then expect a false alarm rate about $0.5^{40} \approx 9.1 \times 10^{-13}$ and a hit rate about $0.999^{40} \approx 0.96$. Thus the face detector detects almost all the face images and rejects almost all the non-face images.

A strong classifier $\mathbf{H}_i(x)$ at each stage of the cascade consists of many weak classifiers $h_t(x)$ (Fig.1(b)). This can be described as follows:

$$\mathbf{H}_i(x) = \mathbf{sgn}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right), \qquad (3)$$

where T is the number of weak classifiers and $\alpha_t = \log \frac{1-\varepsilon_t}{\varepsilon_t}$. We note that ε_t is an error rate specified in the training phase. Each weak classifier $h_t(x)$ evaluates a target image region by using Haar-like features. The weak classifier performs that the sum of the intensity of pixels located within the black rectangles is subtracted form the sum of the intensity of pixels located within the white rectangles. The AdaBoost algorithm selects efficient features to classify the target image region among a huge variety of features.



(a) Cascade of classifiers

(b) Examples of features

Figure 1: Cascaded classifier

4 Incorporating cascaded classifier into the particle filter framework

In this section we describe details about incorporating cascaded classifiers into the particle filter framework. In our method, cascaded classifiers are used for sample evaluation. Multiple cascaded classifiers are first trained respectively to detect one direction of a human head and then adaptively used. This leads to robust and accurate human head tracking.

4.1 Model of human head and projection to camera

A human head is modeled as an ellipsoid. 3D position of a human head is represented with center coordinates of ellipsoid $[x,y,z]$. The coordinate system is represented with their X and Y axes aligned on the ground plane and the Z-axis representing the vertical direction from the ground plane. We assume that a person walks in a room without tilting his/her head. Therefore the orientation of a human head can be identified by only θ that represents the rotation around the Z-axis.

The n-th sample at time t denoted by $\mathbf{s}_t^{(n)} = \left[x_t^{(n)}, y_t^{(n)}, z_t^{(n)}, \theta_t^{(n)}\right]^\top$ is projected by the i-th calibrated camera as follows.

$$\mathbf{p}_{i,t}^{(n)} = F_i\left(\mathbf{s}_t^{(n)}\right), \tag{4}$$

where $\mathbf{p}_{i,t}^{(n)}$ is the projection of a sample $\mathbf{s}_t^{(n)}$ onto the image plane of the i-th camera.

A direction of a human head relative to the i-th camera is denoted as below.

$$\theta_{i,t}^{(n)} = \theta_t^{(n)} - \mathbf{tan}^{-1}\left(\frac{\left[\mathbf{C}_i - \mathbf{K}\mathbf{s}_t^{(n)}\right]^y}{\left[\mathbf{C}_i - \mathbf{K}\mathbf{s}_t^{(n)}\right]^x}\right), \tag{5}$$

where $\theta_{i,t}^{(n)}$ is the direction of a human head relative to the i-th camera. $\mathbf{C}_i$ represents the XY coordinates of the i-th camera. $\mathbf{K}$ is a matrix that extracts the XY coordinates of a human head from a sample $\mathbf{s}_t^{(n)}$. $[\,]^x$ denotes the extraction of the X coordinate.

A projected size of a human head l_i corresponding to the i-th camera can be calculated by projection of the ellipsoidal head model.

In the next section, we evaluate the image region of the i-th camera corresponding to a sample $\mathbf{s}_t^{(n)}$ by using these parameters: the image coordinates $\mathbf{p}_{i,t}^{(n)}$, the direction of a human head $\theta_{i,t}^{(n)}$ and the size of a human head $l_{i,t}^{(n)}$.

4.2 Evaluation of samples using cascaded classifiers

Here, we consider to evaluate a likelihood of an image region $g_{i,t}^{(n)}$.

A classifier in a later stage provides more sophisticated performance in detection since much more weak classifiers $h_t(x)$ are already applied, and thus the image region that passed a later stage of the cascade has much more features of a human head. This implies that we can employ as the likelihood of a human head the number of stages that the image region $g_{i,t}^{(n)}$ passed. This also involves an advantage of computational cost by rejecting the non-target sample in an early stage of the cascade.

The concrete evaluation procedure of a sample $\mathbf{s}_t^{(n)}$ is given bellow. Note that each cascaded classifier is already trained respectively to detect one direction of a head such as front of head, 90° left of head, 90° right of head. The number of stages in our cascade is tuned enough to observe great difference of the number of passed stages between head and non-head image regions.

1. For the i-th camera, at time t, obtain the image coordinates $\mathbf{p}_{i,t}^{(n)}$, the direction of a human head $\theta_{i,t}^{(n)}$ and the size of a human head $l_{i,t}^{(n)}$ by projecting a sample $\mathbf{s}_t^{(n)}$ to the i-th camera.

2. Extract the square image region corresponding to a sample $\mathbf{s}_t^{(n)}$ with center coordinates $\mathbf{p}_{i,t}^{(n)}$, square size $l_{i,t}^{(n)}$. When the projected region is out of the field of view of the i-th camera, the evaluation process is terminated and evaluation score is not used in step 6.

3. Resize the extracted image region to obtain an image region $g_{i,t}^{(n)}$ as the input of the cascaded classifier. This is because in our current system the size of an input image for cascaded classifiers is constant (eg.24×24 pixel).

4. Select a cascaded classifier by considering the direction of a human head $\theta_{i,t}^{(n)}$ relative to the i-th camera. For example, if we use three classifiers such as the classifier for front, the classifier for 90° left, the classifier for 90° right, then the classifier for front is selected in the case of $-45^\circ \leq \theta_{i,t}^{(n)} \leq 45^\circ$, the classifier for 90° left is selected in the case of $45^\circ < \theta_{i,t}^{(n)} \leq 135^\circ$ and the classifier for 90° right is selected in the case of $-135^\circ \leq \theta_{i,t}^{(n)} < -45^\circ$.

5. Apply the selected cascaded classifier to the image region $g_{i,t}^{(n)}$ and obtain the number of passed stages as the likelihood of a human head. This likelihood is used for the weight $\pi_{i,t}^{(n)}$ of the i-th camera corresponding to sample $\mathbf{s}_t^{(n)}$. For example, when the image region is rejected at the first stage of the cascade, the weight is 1. Note that the smallest score of likelihood is 1 not 0 in this scenario. When the region of an image passes all the stages, the weight is 40 if the cascade has 40 stages of classifiers.

6. Repeat the procedures from step 1 to 5 for each camera. The weights $\pi_{i,t}^{(n)}$ corresponding to the i-th camera are integrated by

$$\pi_t^{(n)} = \prod_i \pi_{i,t}^{(n)}. \tag{6}$$

These procedures above are repeated for each sample. The estimation of the state at time t is calculated as the expectation of the weights over the sample set.

5 Experimental results

To demonstrate the effectiveness and robustness of the proposed method, we performed experiments on tracking of a human head in an indoor environment.

We used two cameras on the ceiling with overlapping field of view. Image sequences are captured with the size of VGA at a video frame rate and processed by one PC (Pentium4 3.2GHz, Memory 1GByte). We used three different types of cascaded classifiers, all of which are trained for detecting front of head, 90° left of head and 90° right of head respectively. Each cascaded classifier has 40 stages and was trained by 24×24 pixel-sized images. We employ as our state transition model an average speed of last few frames with Gaussian noise. We gave the initial position of a human head manually. The number of samples was 200 that was much fewer than the case of searching over a whole image sequentially and exhaustively. In this configuration, we processed 200 samples within 30ms, so that the tracking of a human head was performed in real time.

We show the result of tracking in the following figures. A subject walked, bended and stretched in an observed area with changing orientations of his/her head. Fig.2 shows the example images of the tracking result. In each image, the tracking result is drawn by the colored rectangle with dots corresponding to samples. Fig.3 shows the trajectories of the tracked human head with the ground truth. The ground truth was reconstructed by stereo triangulation using same pair of cameras. The coordinates of a human head in each camera image were extracted manually. The error measurements shown in Table 1 are the average Euclidean distance between the estimated position and the ground truth both on XY plane and the Z-axis. The standard deviations are also shown in Table 1. Figs.2, 3 and Table 1 show that the human head was successfully tracked with high-accuracy.

(a) #450 (b) #530 (c) #650

Figure 2: Tracking results

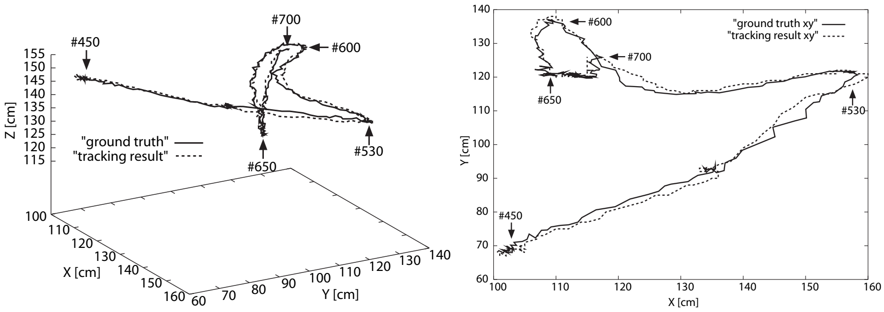


Figure 3: Trajectories of a human head position

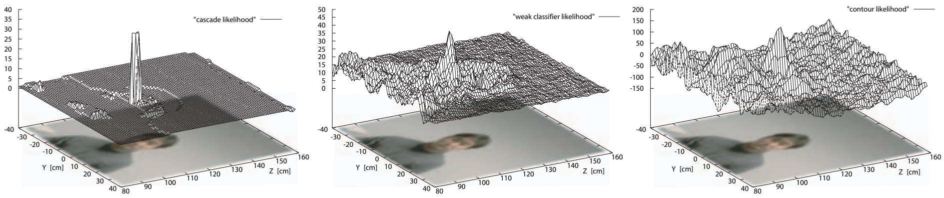
Table 1: Tracking errors

	Average error [cm]	Standard deviation [cm]
Z-axis	1.02	0.65
XY plane	1.98	1.46

6 Discussion

In the particle filter framework, precise evaluation of samples plays a key role for improving accuracy of tracking. The ideal evaluation method has to have a sharp peak around the ground truth of a head position. To observe values of our evaluation method, which outputs the number of passed stages as the likelihood, we calculated the likelihoods at every points at 1cm square mesh around the ground truth of a head position (Fig.4(a)). As a reference, we also calculated the likelihoods at the same point by using different types of evaluation method. The evaluation method based on the total score of weak classifiers calculates the sum in Eq.3 as the likelihood (Fig.4(b)). This evaluation method employs the single strong classifier that is trained to achieve the same detection rate as the cascaded classifier of our evaluation method. The evaluation method based on the contour orientation similarity calculates the sum of the inner products of contour gradients and normal vectors on the contour of upper head model (Fig.4(c)).

Fig.4 shows that our evaluation method has the sharper peak around the ground truth of a head position than the other methods. Our evaluation method also involves an advantage of computational cost. In the case of using our method, all scores of weak classifiers are not calculated because samples with low scores of likelihood are rejected in an early stage of the cascade. However, in the case of using the method based on the total score of weak classifiers, all scores of weak classifiers in Eq.3 are calculated in each sample.



(a) Number of passed stages (Our method)

(b) Total score of weak classifiers

(c) Contour orientation similarity

Figure 4: Likelihood distribution

Depending on the relationship between a human head orientation and a camera position, a human head is observed with variety appearance. Therefore we adaptively use several different types of cascaded classifiers. To validate our selection of cascaded classifiers we observed outputs of cascaded classifiers where each classifier was respectively trained in advance to detect one direction of a human head. Fig.5 shows likelihood variations of each cascaded classifier while a subject turns his head from left to right in the scene. Likelihoods are depicted as the sum of the evaluation scores at 266 mesh points. Rough directions of his head are also shown in Fig.5.

In the case where an observed head direction is left, the score of the classifier for the 90° left head is relatively high. In the case where a frontal head is observed, the classifier for the front head has the highest score. In the case where an observed head direction is right, on the other hand, the score of the classifier for the 90° right head is relatively high. In this way, the appropriate cascaded classifier always gives the highest score. We can therefore use different types of cascaded classifiers for evaluating samples.

To confirm the robustness against occlusion in a multi-camera based tracking, we also conducted experiments on tracking with 4 cameras. Typical frames in our experiment are shown in Fig.6. Fig.6 shows that tracking was stable even though one of the four cameras lost the human head due to occlusion (Fig.6(a)). This is caused that the observation with occlusion does not give negative effects to the likelihood integration because our evaluation gives uniformly low likelihoods (not zero) except around the ground truth of a head position. Fig.6(b) also shows stable tracking even when the subject was in an area where two of the four cameras cannot observe the subject. This is because the evaluation result

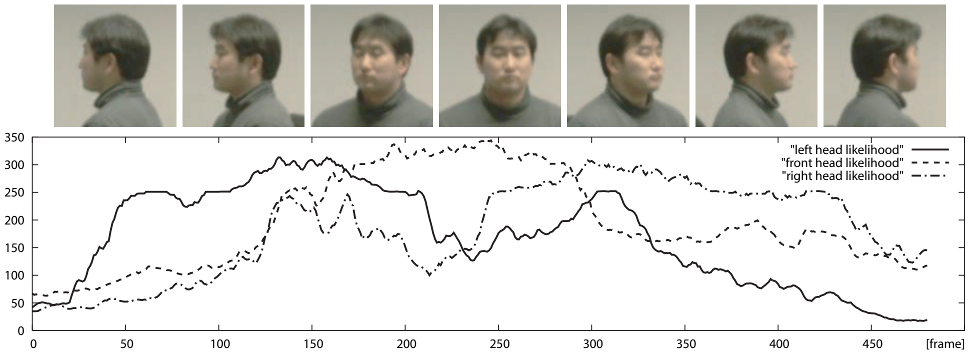


Figure 5: Relationship between likelihood and direction of a head

is not integrated if the sample projected region is out of the field of view of the camera.

(a) Robust tracking when the one camera cannot observe the subject due to occlusion

(b) Robust tracking when the subject is in an area where two cameras cannot observe the subject

Figure 6: Robustness of multi-camera tracking

It is also possible to estimate the human head orientation by using this proposed method. In our experiments, in the case of using cascaded classifiers trained for detecting three directions with 90° angular interval, roughly around 45° error has been observed. It is believed that the higher accuracy estimation of human head orientation can be performed by using additional classifiers trained to detect precise direction of a human head. This framework of adaptive selection of classifier can employ additional classifiers without increasing computational cost.

7 Conclusion

In this paper, we proposed a method for 3D human head tracking based on the particle filter framework incorporating cascaded classifiers into hypothesis evaluation. The efficiency of adaptive selection of cascaded classifiers have been also presented. We have shown the improvement of reliability for likelihood calculation by using cascaded classifiers. This realizes robust and accurate human head tracking. We confirmed the effectiveness of our method by experiments on tracking of a human head in an indoor environment.

In the future, we extend our work to the multiple people tracking. This framework of multiple camera tracking allows us to track multiple objects without establishing correspondences among objects observed in each camera. Since the output of our likelihood function does not represent a probability of likelihood theoretically, we plan to adopt the likelihood calibration technique described in [2] to improve the performance of tracking with cascaded classifiers. The initial detection of human heads and employing additional classifiers to improve estimation accuracy of head orientation are also left for future works.

References

[1] Santa Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *Proc. of International Conference on Computer Vision and Pattern Recognition*, pages 232–237, 1998.

[2] Thierry Chateau, Vincent Gay-Belille, Frederic Chausse, and Jean-Thierry Lapreste. Real-time tracking with classifiers. In *Proc. of International Workshop on Dynamical Vision in conjunction with ECCV*, 2006.

[3] Michael Isard and Andrew Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[4] Gareth Loy, Luke Fletcher, Nicholas Apostoloff, and Alexander Zelinsky. An adaptive fusion architecture for target tracking. In *Proc. of International Conference on Automatic Face and Gesture Recognition*, pages 261–266, 2002.

[5] Kai Nickel, Tobias Gehrig, Rainer Stiefelhagen, and John McDonough. A joint particle filter for audio-visual speaker tracking. In *Proc. of International Conference on Multimodal Interfaces*, pages 61–68, 2005.

[6] Katja Nummiaro, Esther Koller-Meier, and Luc Van Gool. An adaptive color-based particle filter. *Image and Vision Computing*, 21(1):99–110, 2003.

[7] Kenji Okuma, Ali Taleghani, Nando De Freitas, James J. Little, and David G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *Proc. of European Conference on Computer Vision*, pages 28–39, 2004.

[8] Patrick Prez, Jaco Vermaak, and Andrew Blake. Data fusion for visual tracking with particles. *Proc. of the IEEE*, 92(3):495–513, 2004.

[9] Jamie Sherrah and Shaogang Gong. Fusion of perceptual cues for robust tracking of head pose and position. *Pattern Recognition*, 34(8):1565–1572, 2001.

[10] Akihiro Sugimoto, Kiyotake Yachi, and Takashi Matsuyama. Tracking human heads based on interaction between hypotheses with certainty. In *Proc. of Scandinavian Conference on Image Analysis*, pages 617–624, 2003.

[11] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of International Conference on Computer Vision and Pattern Recognition*, pages I:511–518, 2001.

[12] Ya-Dong Wang, Jian-Kang Wu, and Ashraf A. Kassim. Particle filter for visual tracking using multiple cameras. In *Proc. of IAPR Conference on Machine Vision Applications*, pages 298–301, 2005.

[13] Changjiang Yang, Ramani Duraiswami, and Larry Davis. Fast multiple object tracking via a hierarchical particle filter. In *Proc. of International Conference on Computer Vision and Pattern Recognition*, pages I:212–219, 2005.