Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Robust Object Tracking with Continuous Data Association based on Artificial Potential Moving Horizon Estimation

Ryoya Abe * **Tomoya Kikuchi** * **Kenichiro Nonaka** *
Kazuma Sekiguchi *

* *Graduate School of Integrative Science and Engineering, Mechanics, Tokyo City University, Japan (e-mail: knonaka@tcu.ac.jp).*

Abstract: In this paper, a novel object tracking method based on moving horizon estimation (MHE) is introduced that integrates data association into numerical optimization on Bayesian state estimation. Object tracking is a classical problem that often appears in radar and vision systems for which either deterministic or probabilistic approach has been applied. While the former often encounter a failure of association, the latter avoids it by computing the expectation concerning the observations, but it requires the prior knowledge of the probabilistic distribution and may suffer from outliers. In this paper, a partially deterministic approach is built on MHE to resolve these concerns. Data association is realized by a potential function comprising the superposition of Gaussian distributions centered at each observed feature. The potential function is embedded into the stage cost of MHE; the optimal data association is conducted using the sequence of observations within the horizon. Thus robust object tracking is achieved utilizing multi-sampling data and integration of both dynamics and explicit constraints reflecting physical limitations. The advantage of the proposed method is verified through object tracking experiments on the crowded environment, where occlusion and misrecognition frequently occur.

Keywords: object tracking, data association, moving horizon estimation, artificial potential.

1. INTRODUCTION

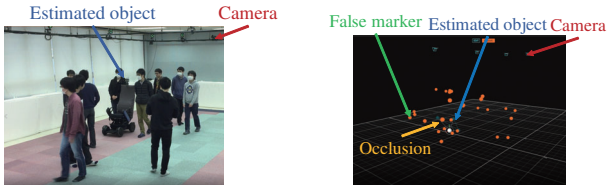
1.1 Visual Object Tracking

Object tracking is a crucial technology for systems which requires information of moving objects. It first developed in the application for Radars and has extended its application area to vision-based systems later. Most algorithms extract features associated with moving objects. Especially in robot vision systems, various feature-based methods has been proposed in Bar-Shalom et al. (2009), Kragic and Kyrki (2008) and Li and Jilkov (2003). Scale-Invariant Feature Transform (SIFT) or Speeded Up Robust Features (SURF) are widely utilized for extracting features from images. Gupta et al. (2016) utilized SURF matching to track humans. Izatt et al. (2017) fused point cloud data from an RGBD sensor with tactile information to realize object tracking. Wu et al. (2017) proposed a object tracking algorithm for a monocular camera system utilizing a saliency map integrated on the Kalman filter framework. Other vision sensors are utilized in localization and estimation of nearby objects, as presented in Bernini et al. (2014) and Lagisetty et al. (2013).

1.2 Filters for object tracking

To track features throughout consecutive samplings from vision sensors, data association is essential in which each similar features from different frames are associated. Starting from the classical methods which utilize deterministic combinational optimization like global nearest neighbor (GNN), probabilistic methods are intensively developed and now widely utilized. In order to deal with the motion of an object, a model based stochastic filter is applied so that it can cope with the occlusion and misrecognition caused by neighboring objects. Considering practical application, it is crucial to consider robustness for occlusion and misrecognition, because it is an inevitable problem of object tracking in the actual environment like a crowd. Thus, it is important to realize continuous and robust object tracking. Camuñas-Mesa et al. (2017) uses a kind of event-triggered cameras for stereo vision to measure feature points, in which the stereo camera detects the feature points which satisfy the constraints of distance to epipolar line or direction toward the edge while occlusion of feature points are considered. In Wang et al. (2018), occlusion is detected, and the failure of object tracking is avoided by one of the phase-only correlation functions called Peak-to-Correlation Energy (PCE). Yang et al. (2016) considered occlusions in multiple frames by performing continuous synthetic aperture of the image. Bertozzi et al. (2015) utilized a soft cascade and classifier of ACF to distinguish between people and objects. Assa and Janabi-Sharifi (2013) improved the robustness based on sensor fusion utilizing estimated noise parameters.

In the observation of features, misrecognition is caused by similar targets. To resolve this issue, data association and estimation method using the filtering technique is proposed. Allodi et al. (2016) associated the vision sensor with laser sensor based on a classifier, and estimated the states using UKF. Beinhofer et al. (2013) suppressed the



(a) Crowded environment (b) Motion capture image

Fig. 1. Occlusion and misrecognition in a crowd

error of association on SLAM using actor-critic Monte Carlo reinforcement learning. Zhu et al. (2017) improved robustness against misrecognition using an EM algorithm utilizing maximum likelihood estimation of parameters.

For motion capture systems, Kikuchi et al. (2019) utilized a potential function related to markers to realize continuous data association, however, their estimation may become a local optimum, because the data association and model-based filter are separated and processed serially.

1.3 Proposed method

In this paper, we merge a deterministic data association into a model-based filter to overcome occlusion and misrecognition. We utilize moving horizon estimation (MHE), which obtains an optimal estimation by minimizing an objective function, which incorporates a various form of criteria comprised of prediction error and consideration of system and observation noise, while satisfying constraints. To integrate data association into MHE framework, we introduce a potential function which reflects the observation of markers to evaluate prediction error, while it prevents from being effected by outliers due to occlusion and ambiguous features. The advantage of the proposed method is verified through a simulation under an adverse condition. It is also applied to a motion capture system to track an electric wheelchair surrounded by pedestrians to verify its robustness and real-time feasibility.

2. MODEL BASED OBJECT TRACKING

2.1 Process and output model

Consider a discrete time stochastic system described by

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{v}_k \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{g}(x_k) + \boldsymbol{w}_k \tag{2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^n$ are state vectors, $\boldsymbol{u}_k \in \mathbb{R}^l$ is an input vector, and $\boldsymbol{y}_k \in \mathbb{R}^{mp}$ is an output vector, in which we assume that the motion of the system is observed through p features of m-dimensional vectors. $\boldsymbol{f}(\cdot,\cdot) \in \mathbb{R}^n$ and $\boldsymbol{g}(\cdot) \in \mathbb{R}^{mp}$ are vector valued functions. $\boldsymbol{v}_k \in \mathbb{R}^n$ and $\boldsymbol{w}_k \in \mathbb{R}^{mp}$ are process and observation noise, respectively, where they are assumed to be Gaussian with zero mean and covariance $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$: $\boldsymbol{v}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k)$ and $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_k)$.

2.2 Observation model

On the object tracking problem, $\boldsymbol{y}_k$ may not be directly measured due to occlusion and misrecognition. We assume that it is accessible through an observation vector $\boldsymbol{z}_k \in \mathbb{R}^{mq_k}$ defined as the following observation model:

$$\boldsymbol{z}_k = \boldsymbol{C}_k \boldsymbol{y}_k + \boldsymbol{D}_k \boldsymbol{r}_k \tag{3}$$

where $\boldsymbol{r}_k \in \mathbb{R}^{mr}$ is a normally distributed random vector, $\boldsymbol{C}_k \in \mathbb{R}^{mq_k \times mp}$ and $\boldsymbol{D}_k \in \mathbb{R}^{mq_k \times mr}$ are *unknown* coefficient matrices, in which each row vector of $[\boldsymbol{C}_k, \boldsymbol{D}_k]$ is one of the standard basis vector (1 in one element and 0's in elsewhere) of $m(p+r)$ dimensional space which is independent with each other. We also assume that q_k is variable with respect to k so that the number of observed feature varies for each sampling.

2.3 Data association

Since $\boldsymbol{C}_k$ is unknown, it is necessary to associate $\boldsymbol{z}_k$ with plausible $\hat{\boldsymbol{y}}_k$. The simplest method is known as global nearest neighbor (GNN) which selects the nearest candidate. In the cluttered environment, probabilistic data association is utilized which takes the mean of candidates weighted by probability. In addition, to reduce the effort on implausible candidates, *validation region* is also used which limits the candidates of $\boldsymbol{z}_k^j$ to the ones in the neighborhood of prediction. Typical criteria (Wang et al. (2002)) is

$$\left(\boldsymbol{z}_k^j - \boldsymbol{H}\hat{\boldsymbol{x}}_k^-\right)^{\mathrm{T}} \boldsymbol{S}_k^{-1} \left(\boldsymbol{z}_k^j - \boldsymbol{H}\hat{\boldsymbol{x}}_k^-\right) \leq \gamma. \tag{4}$$

3. MOVING HORIZON ESTIMATION

In this section, moving horizon estimation (MHE) is briefly reviewed. MHE is an optimal model based filter, which estimates the state using the past inputs and outputs of within *horizon*. The optimal estimation is computed by minimizing the index function which evaluates the uncertainty due to the process and observation noise, and the arrival cost. A general form of index function is

$$\begin{aligned} J = &\sum_{t=k-H}^{k-1} ||\hat{\boldsymbol{x}}_{t+1} - \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_t)||^2_{\boldsymbol{Q}_t^{-1}} \\ &+ \sum_{t=k-H}^{k} ||\boldsymbol{D}_t^{\perp}(\boldsymbol{z}_t - \boldsymbol{C}_t \boldsymbol{g}(\hat{\boldsymbol{x}}_t))||^2_{\boldsymbol{R}_t^{-1}} \\ &+ ||\hat{\boldsymbol{x}}_{k-H} - \hat{\boldsymbol{x}}^p_{k-H}||^2_{\boldsymbol{P}^{-1}_{k-H}} \end{aligned} \tag{5}$$

where H is the horizon, $\hat{\boldsymbol{x}}$ is the posterior state estimation, $\hat{\boldsymbol{x}}^p_{k-H}$ is the state estimation computed in the previous sampling time. MHE can also incorporate them as a constrained optimization problem.

Since $\boldsymbol{C}_k$ and $\boldsymbol{D}_k$ are unknown, (5) cannot be directly used; data association should be conducted to determine them. But identification of them can be considered as a kind of optimization problem, if it can be transformed into a gradient based numerical optimization problem, we can integrate data association into the framework of MHE.

4. MHE EMBEDDED WITH DATA ASSOCIATION

To transform data association into a gradient based numerical optimization problem, we utilize a potential function based optimization in Kikuchi et al. (2019), where the potential function is comprised of the Gaussian dis-

tributions associated with observed features. Consider the following Gaussian distribution function:

$$L_k^{ij} = \frac{1}{\sqrt{(2\pi)^n |\Sigma_k|}} \exp\left\{ -\frac{1}{2} (\hat{\boldsymbol{y}}_k^i - \boldsymbol{z}_k^j)^{\mathrm{T}} \Sigma_k^{-1} (\hat{\boldsymbol{y}}_k^i - \boldsymbol{z}_k^j) \right\} \tag{6}$$

where $\hat{\boldsymbol{y}}_k^i$ is sub-vector which corresponds to the estimation of i-th feature, $\boldsymbol{z}_k^j$ is the sub-vector corresponds to the observation of j-th feature, and Σ_k is the innovation covariance matrix. This function evaluates an estimated point $\hat{\boldsymbol{y}}_k^i$ with respect to the Gaussian distribution with mean $\boldsymbol{z}_k^j$, corresponding to each observation. We define a potential function $L_k(\hat{\boldsymbol{y}}_k)$ as a sum of L_k^{ij}:

$$L_k(\hat{\boldsymbol{y}}_k) = \sum_{i=1}^{p} \sum_{j=1}^{q_k} L_k^{ij}. \tag{7}$$

It is noted that the gradient of this potential function around the marker is steep while it becomes flat at the residual area where occlusion possibly occurs. This potential function (7) is incorporated into the objective function of MHE as follows:

$$\begin{aligned} J_{\mathrm{DA}} = & \sum_{t=k-H}^{k-1} ||\hat{\boldsymbol{x}}_{t+1} - \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_t)||^2_{\boldsymbol{Q}_t^{-1}} \\ & + \sum_{t=k-H}^{k} L_t^{ij} + ||\hat{\boldsymbol{x}}_{k-H} - \hat{\boldsymbol{x}}_{k-H}^p||^2_{\boldsymbol{P}_{k-H}^{-1}}. \end{aligned} \tag{8}$$

The difference between J in (5) and J_{DA} in (7) is the evaluation of innovation at the second term. It is extended to incorporate all possible markers while making it insensitive at the area where markers do not exist. In Kikuchi et al. (2019), the latest observations were associated with the predicted features using an artificial potential function, and then the state was estimated using MHE based on the association result. The essential difference from Kikuchi et al. (2019) is that evaluation utilizing (7) is incorporated into the objective function of MHE, which realizes the optimal estimation through the horizon.

5. NUMERICAL COMPARISON

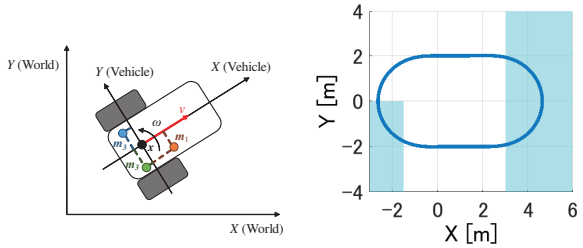
In this section, the proposed method is applied to motion tracking of a vehicle surrounded by false features. The performance of the proposed object tracking method is verified through a simulation comparing with probabilistic data association filter (PDAF) with validation region.

5.1 Vehicle model

We deal with a differential drive vehicle depicted in Fig. 2(a). (x, y) is the coordinates of the center of wheels, and θ is the heading angle. The translational velocity is v_k, and the angular velocity is ω_k. We take v_k and ω_k as control input. To derive a process model in (1), we define the state vector as $\boldsymbol{x}_k = [x_k,\, y_k,\, \theta_k]^{\mathrm{T}}$ and the input vector as $\boldsymbol{u}_k = [v_k,\, \omega_k]^{\mathrm{T}}$. Then $\boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k)$ is described as

$$\boldsymbol{f}(x_k, u_k) := \boldsymbol{x}_k + \begin{bmatrix} v_k \cos\theta_k \\ v_k \sin\theta_k \\ \omega_k \end{bmatrix} \Delta \tag{9}$$

where Δ is the sampling time.



(a) a differential drive vehicle. (b) running trajectory.

Fig. 2. (a) a differential drive vehicle model where three markers are attached to the body, and (b) its running trajectory where occlusion area is colored in blue.

For the output model in (2), we take three markers geometrically fixed to the vehicle represented as

$$\boldsymbol{g}(\boldsymbol{x}_k) := \begin{bmatrix} \boldsymbol{g}_1(\boldsymbol{x}_k) \\ \boldsymbol{g}_2(\boldsymbol{x}_k) \\ \boldsymbol{g}_3(\boldsymbol{x}_k) \end{bmatrix} \tag{10}$$

with

$$\boldsymbol{g}_i(\boldsymbol{x}_k) := \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \boldsymbol{m}_i \tag{11}$$

for $i = 1, 2, 3$, where $\boldsymbol{m}_i \in \mathbb{R}^2$ is the pose of features represented by the Cartesian coordinates fixed to the vehicle. We set $\boldsymbol{m}_i$ as follows: $\boldsymbol{m}_1 = [0.24, -0.1]^{\mathrm{T}}, \boldsymbol{m}_2 = [-0.12, 0.2]^{\mathrm{T}}, \boldsymbol{m}_3 = [-0.12, -0.1]^{\mathrm{T}}$.

5.2 Vehicle trajectory and marker occlusion

Fig. 2(b) depicts the moving path of the vehicle; it starts from $(0, -2)$ and moves around a track in counterclockwise direction at the velocity 0.3 m/s. To set false features $\boldsymbol{r}_k$ in the observation, we set the following additional markers. Three hundred static markers are randomly placed under uniform distribution in the region of $-3 \le x \le 5$ and $-3 \le y \le 3$. Fifteen groups of markers comprising fifteen markers randomly move in the field. In blue rectangular area, Marker #2 is occluded, and random markers appear in the region within 0.3 m radius from the occluded marker. Parameters are summarized in Table 1.

Table 1. Parameters.

Parameter	Value
Sampling period	50 ms
W	0.275 m
$R_{\mathrm{r}}, R_{\mathrm{l}}$	0.128 m
H	20
$\boldsymbol{w}_{\mathrm{s}}$	$[9.5\times10^{-2}, 9.5\times10^{-2}, 1.5\times10^{-3}]^{\mathrm{T}}$
$\boldsymbol{w}_{\mathrm{o},i}$	$[6.716, 7.058]^{\mathrm{T}} \times 10^{-6}$

5.3 Probabilistic data association filter

To evaluate the estimation performance, we also apply a conventional tracking method based on a probabilistic data association filter (Grinberg (2017)). The rate parameter λ of Poisson distribution is set to be 0.8835, computed from the number of static false markers for validation region with 0.1 m radius per the range of distribution.

5.4 Simulation result

Figure 3 depicts the errors of PDAF and MHE, respectively, for the same process and observation noise, $\boldsymbol{v}_k$ and

$\boldsymbol{w}_k$, and random false markers $\boldsymbol{r}_k$. Both methods successfully tracked the vehicle, but a larger error appeared during occlusion. The error of MHE was slightly larger than PDAF in the section without occlusion, but the error of PDAF was far larger than MHE during occlusion.

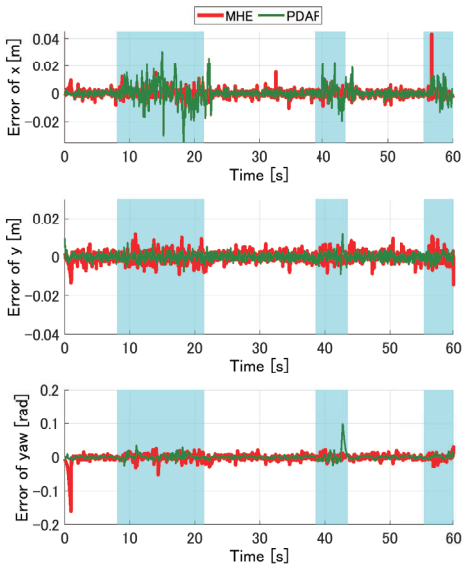


Fig. 3. Estimation errors of MHE / PDAF for x, y and θ.

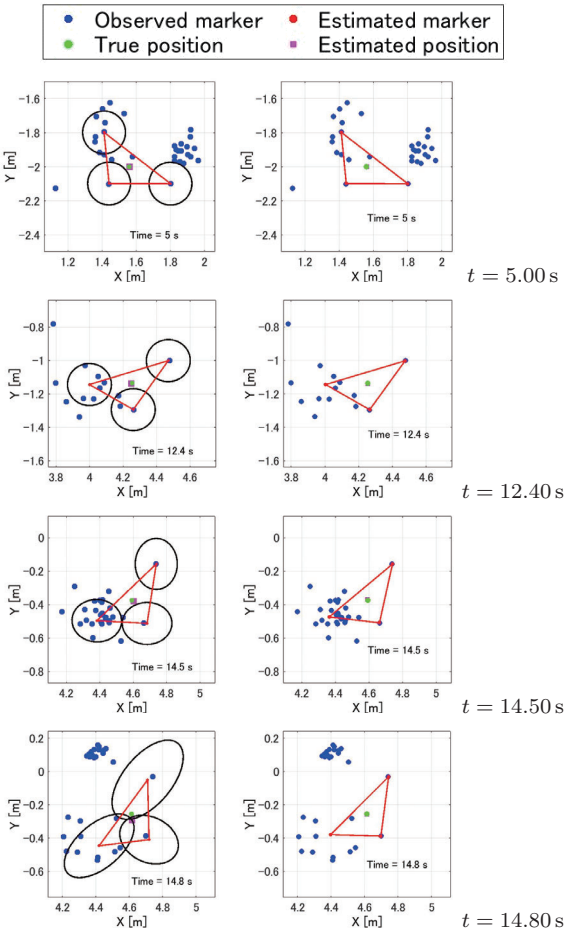


Fig. 4. Snapshots describing the allocation of markers and the estimates for PDAF (left) and MHE (right).

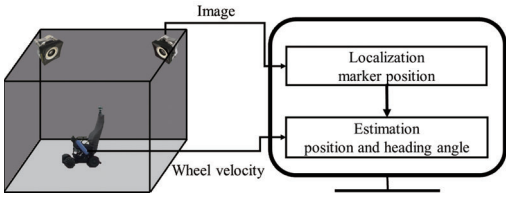


Fig. 5. Motion capture based vecual feedback system.

Fig. 6. Left: electric wheelchair with attached markers, right: measurement of markers in this experiment system. Red markers are attached to the estimation target and green markers are used for a reference.

Table 2. Parameters.

Parameter	Value
W	0.275 m
R	0.128 m
H	5
$\boldsymbol{w}_{\mathrm{s}}$	$[8.0\times10^{-5}, 8.0\times10^{-5}, 7.5\times10^{-4}]^{\mathrm{T}}$
$\boldsymbol{w}_{\mathrm{o},i}$	$[6.716, 7.058]^{\mathrm{T}}\times10^{-5}$

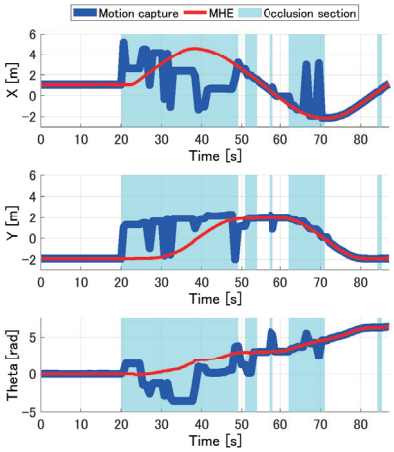


Fig. 7. The trajectories of x, y and θ obtained by a commercial software (blue) and MHE (red). Durations of occlusion are colored by cyan. MHE successfully tracked the vehicle even under occlusion.

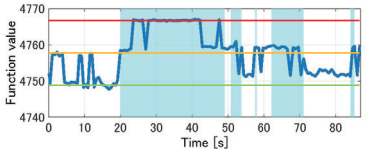


Fig. 8. The time variation of the potential function: the number of occluded markers are approximately evaluated by the level indicated by green (zero), yellow (one), and red (two).

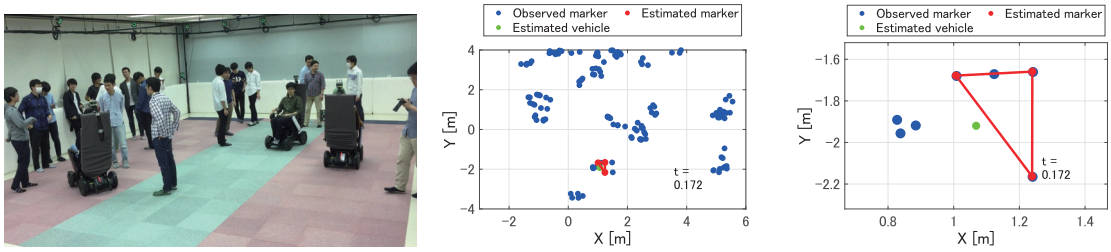


Fig. 9. Left: experimental landscape, ceter: plot of markers and estimated ones, right: enlarged view of plots (0.01 s).

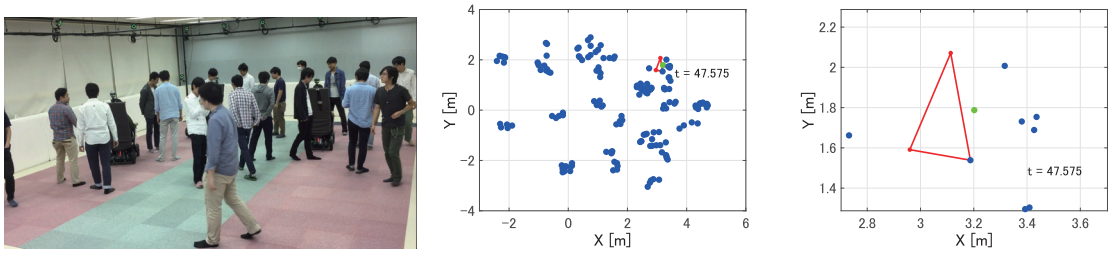


Fig. 10. Pedestrians surrounds the electric wheelchair and two markers is occluded (47.6 s).

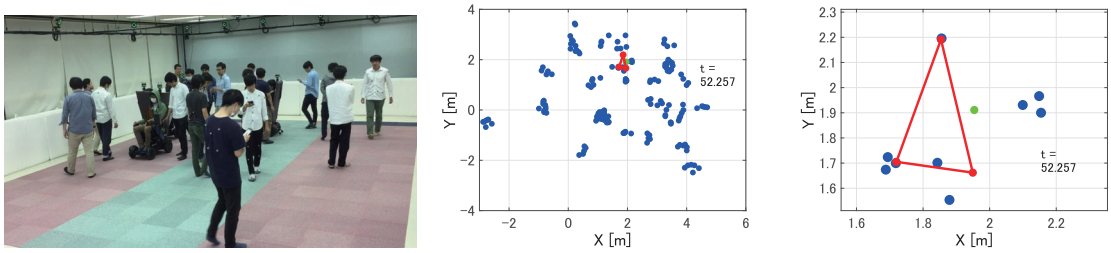


Fig. 11. MHE recovers to track the missing marker and continue to estimate the right position of the wheelchair (52.3 s).

To understand the process, snapshots of markers and the estimation are depicted in Fig. 4 for PDAF (left) and MHE (right), respectively. A triangle connecting three estimated markers are drawn in each figure to highlight the posture of markers. The ellipses are the validation regions for PDAF. At $t = 5.00$ s, all markers were successfully identified, and the estimated position coincided with the true position. At $t = 12.40$ s, one marker was occluded, but both methods still tracked the vehicle. But at $t = 14.50$ s, the error of PDAF increased due to the false markers in the validation regions of occluded markers. Consequently PDAF failed tracking at $t = 14.80$ s. MHE also tracks plausible markers, even when one feature was occluded; it continued to track the vehicle position accurately until the end of the simulation, as the figures at $t = 14.50$ s and $t = 14.80$ s depict MHE correctly tracked the vehicle utilizing observations of multiple frames.

6. IMPLEMENTATION ON MOTION CAPTURE

In this section, we implement MHE into a motion capture system tracking an electric wheelchair in crowded area to demonstrate both practical robustness against occlusion and computational feasibility in real-time system.

6.1 Experimental system

Figure 5 depicts a schematic of experimental system. Marker positions attached to a wheelchair are localized by a motion capture system comprised of 20 cameras which can detect and compute the three dimensional position of observed markers. Three markers for estimation were attached to the estimation target, which frequently lost by camera in a crowd. Actual wheel velocities are transmitted to a computer, and used to estimate the pose of the vehicle. Tracking control input $\boldsymbol{u}_k$ is computed using the estimated state. The wheelchair used in this experiment is depicted in Fig. 6. To validate the estimation accuracy, additional eight markers for evaluation were attached to the wheel chair. The pause measured from these markers was used as the reference in the evaluation of experiments. The parameter used in this experiment is summarized in Table 2. In this experiment, we sets $H = 5$ to reduce the calculation cost in the optimization.

6.2 Result of experiment

Figure 7 depicts the trajectories of x, y and θ estimated by MHE (red line) together with one computed by a

commercial software supplied with motion capture system (blue line). The shaded areas (painted in cyan) indicates duration of occlusion, where the motion capture system failed to capture the markers which resulted in stagnation of measurement or abrupt change of value to a biased one. On the other hand, MHE successfully tracked the actual location continuously. Figure 8 depicts the evaluation value of output estimation error, which indicates the number of tracked markers. Figures 9 through 11 depicts experimental landscapes (left), plots of detected and estimated markers connected by triangle and estimated vehicle position (center), and enlarged plot (right).

Figure 9 shows pedestrians do not exist around the target, and target markers are successfully measured. Figure 10 shows pedestrians staying around the target which renders a marker occluded, but the target was still tracked. Figure 11 shows the tracking was successful even in the environment where many false observations exist around the target. We can see that under occlusion and cluttered environment surrounded by false markers, MHE could continue to track the vehicle position successfully.

7. CONCLUSION

In this paper, for object tracking, moving horizon estimation (MHE) is introduced which integrates data association into numerical optimization on state estimation. Data association is realized as the minimization of a potential function comprised of the superposition of Gaussian distribution centered at each observed feature. This potential function is embedded into the stage cost of MHE; data association is conducted with respect to both the current and the past observation data within the horizon. The performance of MHE was quantitatively evaluated through the simulation to show that it outperformed conventional PDAF based object tracking. MHE was also applied to the actual motion capture object tracking system which showed that it could track a vehicle surrounded by people. Our future work is to apply the proposed method into other vision systems that use the image information and features, and we validate its effectiveness.

REFERENCES

Allodi, M., Broggi, A., Giaquinto, D., Patander, M., and Prioletti, A. (2016). Machine learning in tracking associations with stereo vision and lidar observations for an autonomous vehicle. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, 648–653. IEEE.

Assa, A. and Janabi-Sharifi, F. (2013). A robust vision-based sensor fusion approach for real-time pose estimation. *IEEE transactions on cybernetics*, 44(2), 217–227.

Bar-Shalom, Y., Daum, F., and Huang, J. (2009). The probabilistic data association filter. *IEEE Control Systems Magazine*, 29(6), 82–100.

Beinhofer, M., Kretzschmar, H., and Burgard, W. (2013). Deploying artificial landmarks to foster data association in simultaneous localization and mapping. In *2013 IEEE International Conference on Robotics and Automation*, 5235–5240. IEEE.

Bernini, N., Bertozzi, M., Castangia, L., Patander, M., and Sabbatelli, M. (2014). Real-time obstacle detection using stereo vision for autonomous ground vehicles: A survey. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 873–878. IEEE.

Bertozzi, M., Castangia, L., Cattani, S., Prioletti, A., and Versari, P. (2015). 360 detection and tracking algorithm of both pedestrian and vehicle using fisheye images. In *2015 IEEE intelligent vehicles symposium (iv)*, 132–137. IEEE.

Camuñas-Mesa, L.A., Serrano-Gotarredona, T., Ieng, S.H., Benosman, R., and Linares-Barranco, B. (2017). Event-driven stereo visual tracking algorithm to solve object occlusion. *IEEE transactions on neural networks and learning systems*, 29(9), 4223–4237.

Grinberg, M. (2017). *Feature-based Probablistic Data Association for Video-Based Multi-Object Tracking*. KIT Scientific Publishing.

Gupta, M., Kumar, S., Behera, L., and Subramanian, V.K. (2016). A novel vision-based tracking algorithm for a human-following mobile robot. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7), 1415–1427.

Izatt, G., Mirano, G., Adelson, E., and Tedrake, R. (2017). Tracking objects with point clouds from vision and touch. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 4000–4007. IEEE.

Kikuchi, T., Tsuno, K., Nonaka, K., and Sekiguchi, K. (2019). Continuous marker association utilizing potential function for motion capture systems. In *2019 IEEE/SICE International Symposium on System Integration (SII)*, 578–583. IEEE.

Kragic, D. and Kyrki, V. (2008). *Unifying perspectives in computational and robot vision*, volume 8. Springer Science & Business Media.

Lagisetty, R., Philip, N., Padhi, R., and Bhat, M. (2013). Object detection and obstacle avoidance for mobile robot using stereo camera. In *2013 IEEE International Conference on Control Applications (CCA)*, 605–610. IEEE.

Li, X.R. and Jilkov, V.P. (2003). Survey of maneuvering target tracking. part i. dynamic models. *IEEE Transactions on aerospace and electronic systems*, 39(4), 1333–1364.

Wang, M., Liu, Y., Su, D., Liao, Y., Shi, L., Xu, J., and Miro, J.V. (2018). Accurate and real-time 3-d tracking for the following robots by fusing vision and ultrasonar information. *IEEE/ASME Transactions on Mechatronics*, 23(3), 997–1006.

Wang, X., Challa, S., and Evans, R. (2002). Gating techniques for maneuvering target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3), 1087–1097.

Wu, Y., Sui, Y., and Wang, G. (2017). Vision-based real-time aerial object localization and tracking for uav sensing system. *IEEE Access*, 5, 23969–23978.

Yang, T., Ma, W., Wang, S., Li, J., Yu, J., and Zhang, Y. (2016). Kinect based real-time synthetic aperture imaging through occlusion. *Multimedia Tools and Applications*, 75(12), 6925–6943.

Zhu, A.Z., Atanasov, N., and Daniilidis, K. (2017). Event-based feature tracking with probabilistic data association. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 4465–4470. IEEE.