https://doi.org/10.20965/jrm.2020.p0537

Paper:

Moving Horizon Estimation with Probabilistic Data Association for Object Tracking Considering System Noise Constraint

Tomoya Kikuchi, Kenichiro Nonaka, and Kazuma Sekiguchi

Tokyo City University
1-28-1 Tamazutsumi, Setagaya, Tokyo 158-8857, Japan
E-mail: {g1881019, knonaka, ksekiguc}@tcu.ac.jp

[Received December 24, 2019; accepted March 26, 2020]

Object tracking is widely utilized and becomes indispensable in automation technology. In environments containing many objects, however, occlusion and false recognition frequently occur. To alleviate these issues, in this paper, we propose a novel object tracking method based on moving horizon estimation incorporating probabilistic data association (MHE-PDA) through a probabilistic data association filter (PDAF). Since moving horizon estimation (MHE) is accomplished through numerical optimization, we can ensure that the estimation is consistent with physical constraints and robust to outliers. The robustness of the proposed method against occlusion and false recognition is verified by comparison with PDAF through simulations of a cluttered environment.

Keywords: moving horizon estimation, probabilistic data association, object tracking, occlusion, misrecognition

1. Introduction

1.1. Object Tracking

Object tracking is indispensable in automation technology. Early implementations occurred in radar systems and now it is widely utilized in vision-based systems. Motion capture systems typically measure the position and orientation of target objects with a fast sampling rate. They are commonly used for motion analysis of humans and robots [1–3], as well as for visual feedback motion control of unmanned aerial vehicles, mobile robots and humanoid robots [4–7]. In automobile and traffic systems, typical examples are the fixed-point observation of the flow of traffic [8–10] as well as the on-board detection and tracking of preceding vehicles, lane lines, road signals, and traffic participants [11–13].

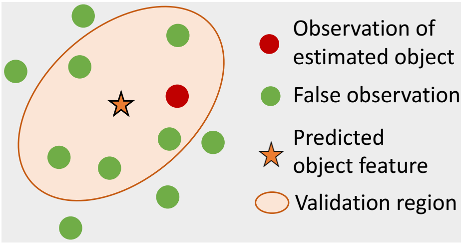
1.2. Occlusion and Misrecognition

Occlusion is a major issue for such vision-based measurement, in which some observations are lost because they are hidden by other objects. Occlusion will cause deterioration of object tracking performance of, for example, traffic monitoring systems as well as on-board detection and tracking systems of automobiles.

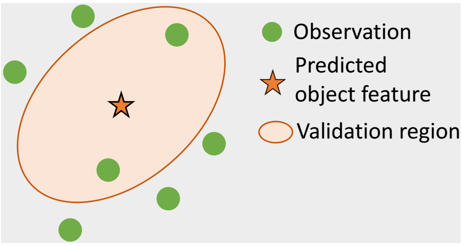
To address this issue, various studies have been conducted [14, 15]. Multiple sensors or robots are fused utilizing model-based estimation to compensate for the duration of occlusion [16–18]. Wang et al. [19] combined a monocular camera with an ultrasonic sensor utilizing an extended Kalman filter (EKF) to develop a three-dimensional tracking system. Yun et al. also applied an EKF to measure the quick motion of fingers, which frequently occlude each other [20]. Lee used deep learning-based vehicle detection and speeded up robust feature (SURF) matching for position estimation [21]. Takahashi et al. [22] proposed *moving horizon estimation* (MHE) for vehicle tracking utilizing only some of the features available during occlusion, in which the weight of the occluded feature is set to be zero to exclude the missing features.

To consider association ambiguity, *domain knowledge-aided moving horizon estimation* (DMHE) has hitherto been combined with *multiple hypothesis tracking* (MHT) [23, 24]. This approach realized accurate tracking performance and robustness using a physical road constraint. Ishikawa et al. [25] suppressed the influence of outliers on observations due to false observations by dynamically changing the weight of MHE based on predictions. When multiple-sensors are utilized, the measurement cycles of sensors are different, but unmeasured observations can be deemed to be occlusion. Liu et al. proposed *multi-rate moving horizon estimation* (MMHE) [26] to incorporate missing observations. In these methods, data association and state estimation were separated; the accuracy of data association strongly affects estimation performance.

Another issue is false recognition, in which other observed features are wrongly associated with the target feature. **Fig. 1** depicts two typical situations with many observations, in which the *validation region* computed by prior estimation is introduced to narrow down the observations. In **Fig. 1(a)**, several features, including the target feature, are observed. In **Fig. 1(b)**, the target feature is occluded, and only false features are observed. Both cases may cause erroneous associations, which would impair estimation performance. To address this issue, the authors hitherto proposed an MHE utilizing an artificial potential function [27]. The proposed estimation method



(a) The validation region contains many false observations as well as observation of the estimated object.



(b) Observation of the estimated object is occluded and the validation region only shows false observations.

Fig. 1. Possible relationships between the validation region and observations.

is robust against occlusion and false recognition, but still strongly relies on association accuracy.

1.3. Probabilistic Data Association

For algorithms in which object tracking and Bayesian estimation are separated, the failure of data association causes deterioration of the subsequent estimation. To address this issue, Abe et al. [28] embedded an artificial potential field for each observation into the objective function of MHE, in which the simultaneous optimization of data association and state estimation was realized; however, a priori probabilistic distribution was not incorporated. *Probabilistic data association filter* (PDAF) is one of the most reliable methods for mitigating false recognition and occlusion [29, 30]. PDAF utilizes a stochastic association measure, which weights the probability for each observation with respect to prediction, to cope with both occlusion and false recognition. Chen et al. [31] proposed a *combined interactive multiple-model probabilistic data association algorithm* (C-IMM-PDA) in which distance weighting was utilized together with a Kalman filter. However, using this filter means that it is generally difficult to explicitly include physical constraints. Yang et al. [32] combined PDAF with a *particle filter* (PF) to deal with multiple objects subject to non-Gaussian noise. It can incorporate nonlinearity and constraints; however, PF generally requires a huge number of particles to achieve an optimal estimation of high accuracy.

1.4. Proposed Method

To achieve robust estimation against misrecognition and occlusion, in this paper, *moving horizon estimation - probabilistic data association* (MHE-PDA) is proposed,

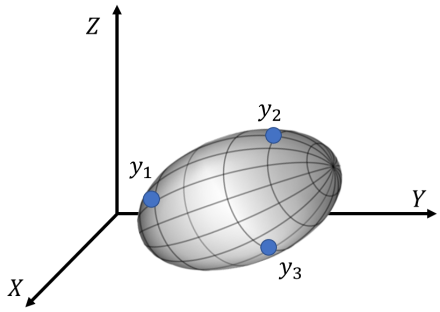


Fig. 2. The relationship between the center of gravity and each object feature.

in which probabilistic data association is integrated into the evaluation function of the MHE, so that the maximum likelihood estimation is obtained through numerical optimization.

Since PDAF assumes a Gaussian distribution on system and observation noise, non-Gaussian noise or outliers will reduce the accuracy of estimation. Since the optimal state estimate is numerically obtained in MHE-PDA, it can naturally include constraints. Thus, we can expect that appropriate constraints might improve estimation performance by alleviating the influence of noise which does not adhere to a Gaussian distribution.

This paper is based on our preliminary report [33], in which PDA was incorporated into the framework of MHE to reflect the probability on each observation. Importantly, herein, constraints on system noise are introduced which mitigate the effects of large deviations and misrecognition caused by false observations which do not obey a Gaussian assumption.

2. Object Model [33]

This section describes the state space representation of an object to be tracked.

2.1. State Equation

We consider a discrete time system, with a state space model represented by Eq. (1):

$$\boldsymbol{x}[k+1] = \boldsymbol{f}\left(\boldsymbol{x}[k], \boldsymbol{u}[k]\right) + \boldsymbol{G}\boldsymbol{v}[k], \quad \ldots\ldots\ldots \quad (1)$$

where $\boldsymbol{x}[k] \in \mathbb{R}^n$ is a state, $\boldsymbol{u}[k] \in \mathbb{R}^m$ is an input, $\boldsymbol{f} \in \mathbb{R}^n$ is a vector valued function, $\boldsymbol{v}[k] \in \mathbb{R}^n$ is a system noise vector subject to $\boldsymbol{v}[k] \sim \mathscr{N}(0, \boldsymbol{Q})$ for a covariance matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, and $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ is a coefficient matrix of system noise.

2.2. Output Equation

We consider a tracking problem for an object on which feature points are attached. **Fig. 2** depicts an object with three feature points. Since the position of a feature moves along with the translation and rotation of the object, a feature point is described as a vector-valued function representing the coordinate transformation. Let s be the number of feature points of the object to be tracked, and $\boldsymbol{y}_i[k]$ ($i=1,\ldots,s$) be the observed position vector of a feature point, described as a function of the state $\boldsymbol{x}[k]$ as follows:

$$\boldsymbol{y}_i[k] = \boldsymbol{h}_i(\boldsymbol{x}[k]) + \boldsymbol{w}_i[k], \quad \ldots \ldots \ldots \quad (2)$$

where $\boldsymbol{h}_i(\boldsymbol{x}[k])$ is the position vector of an observed feature point, and $\boldsymbol{w}_i[k]$ is an observation noise vector subject to $\boldsymbol{w}_i \sim \mathcal{N}(0, \boldsymbol{R}_m)$, and $\boldsymbol{R}_m$ is a covariance matrix.

Let $\boldsymbol{y}[k] \in \mathbb{R}^q$ be an observation vector comprising $\boldsymbol{y}_i[k]$ as follows:

$$\boldsymbol{y}[k] = \left[\boldsymbol{y}_1[k]^{\mathrm{T}}, \ldots, \boldsymbol{y}_s[k]^{\mathrm{T}}\right]^{\mathrm{T}} \ldots \ldots \ldots \quad (3)$$

Then, the output equation is represented as a lumped form by Eq. (4):

$$\boldsymbol{y}[k] = \boldsymbol{h}(\boldsymbol{x}[k]) + \boldsymbol{w}[k]. \quad \ldots \ldots \ldots \quad (4)$$

where $\boldsymbol{h}(\boldsymbol{x}[k])$ is a lumped vector valued function represented by

$$\boldsymbol{h}(\boldsymbol{x}[k]) = \left[\boldsymbol{h}_1(\boldsymbol{x}[k])^{\mathrm{T}}, \ldots, \boldsymbol{h}_s(\boldsymbol{x}[k])^{\mathrm{T}}\right]^{\mathrm{T}}, \quad \ldots \quad (5)$$

$\boldsymbol{w} \in \mathbb{R}^q$ is an observation noise vector subject to $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{R})$ and described as

$$\boldsymbol{w}[k] = \left[\boldsymbol{w}_1[k]^{\mathrm{T}}, \ldots, \boldsymbol{w}_s[k]^{\mathrm{T}}\right]^{\mathrm{T}}, \quad \ldots \ldots \quad (6)$$

and $\boldsymbol{R} \in \mathbb{R}^{q \times q}$ is a covariance matrix.

3. Probabilistic Data Association Filter [33, 34]

This section presents a probabilistic data association filter (PDAF), which can accommodate multiple feature observations and cope with occlusion.

3.1. Prediction Step

The predicted state estimate $\hat{\boldsymbol{x}}^-[k]$ is represented by Eq. (7):

$$\hat{\boldsymbol{x}}^-[k] = \boldsymbol{f}(\hat{\boldsymbol{x}}[k-1], \boldsymbol{u}[k-1]). \quad \ldots \ldots \quad (7)$$

For the extended linearization of $\boldsymbol{f}$,

$$\boldsymbol{A} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}, \quad \ldots \ldots \ldots \ldots \quad (8)$$

the predicted error covariance matrix $\boldsymbol{P}^-[k]$ is represented by

$$\boldsymbol{P}^-[k] = \boldsymbol{A}[k]\boldsymbol{P}[k]\boldsymbol{A}^{\mathrm{T}}[k] + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}} \quad \ldots \ldots \quad (9)$$

Similarly, for the extended linearization of $\boldsymbol{h}$,

$$\boldsymbol{C} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}, \quad \ldots \ldots \ldots \ldots \quad (10)$$

and the innovation covariance matrix for the prediction of observation vector is

$$\boldsymbol{S}[k] = \boldsymbol{C}[k]\boldsymbol{P}^-[k]\boldsymbol{C}^{\mathrm{T}}[k] + \boldsymbol{R}. \quad \ldots \ldots \ldots \quad (11)$$

For each feature point $i = 1, \ldots, s$, let $\boldsymbol{C}_{\mathrm{m},i}$ be the Jacobian matrix of the i-th feature point of the object represented by $\boldsymbol{C}_{\mathrm{m},i} = \partial \boldsymbol{h}_i / \partial \boldsymbol{x}$. The innovation covariance matrix of the i-th feature point is represented by Eq. (12):

$$\boldsymbol{S}_{\mathrm{m},i}[k] = \boldsymbol{C}_{\mathrm{m},i}[k]\boldsymbol{P}^-[k]\boldsymbol{C}^{\mathrm{T}}_{\mathrm{m},i}[k] + \boldsymbol{R}_{\mathrm{m}} \quad \ldots \ldots \quad (12)$$

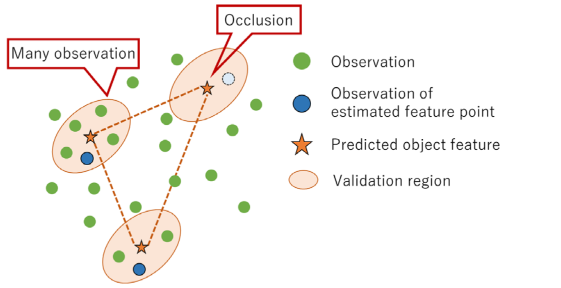


Fig. 3. The relationship between validation regions and observations.

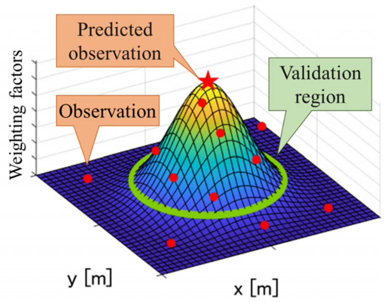


Fig. 4. Weighting factors distribution.

3.2. Validation Region and Weighting Factors

To reduce the computational burden, PDAF extracts observations inside the validation region, which is the interior of an ellipsoid centered at a predicted output estimate as depicted in **Fig. 3**. The size of the ellipsoid is described by Mahalanobis's generalized distance.

The output prediction $\hat{\boldsymbol{y}}_i^-$ corresponding to the i-th feature point on the object is calculated by

$$\hat{\boldsymbol{y}}_i^-[k] = \boldsymbol{h}_i(\hat{\boldsymbol{x}}^-[k]) \quad \ldots \ldots \ldots \ldots \quad (13)$$

Let $\boldsymbol{z}_{i,p}[k]$ be the p-th observation extracted from the validation region for the i-th feature. Then, the error $\tilde{\boldsymbol{y}}_{i,p}^-$ between $\boldsymbol{z}_{i,p}[k]$ and the output prediction $\hat{\boldsymbol{y}}_i^-[k]$ is defined by

$$\tilde{\boldsymbol{y}}_{i,p}^-[k] = \boldsymbol{z}_{i,p}[k] - \hat{\boldsymbol{y}}_i^-[k]. \quad \ldots \ldots \ldots \quad (14)$$

To extract object observation candidates, $\boldsymbol{z}_{i,p}$, corresponding to the i-th feature $\boldsymbol{y}_i[k]$, the gating region is defined as $M_{i,p} \leq \gamma$ for a positive γ and Mahalanobis distance $M_{i,p}$ defined by

$$M_{i,p} = \sqrt{\tilde{\boldsymbol{y}}_{i,p}^-[k]^{\mathrm{T}} \boldsymbol{S}_{\mathrm{m},i}[k]^{-1} \tilde{\boldsymbol{y}}_{i,p}^-[k]} \quad \ldots \ldots \quad (15)$$

Fig. 4 depicts a Gaussian distribution of the weighting factor. The region drawn inside the ellipse is the validation region. Only the observations $\boldsymbol{z}_{i,p}$ satisfying $M_{i,p} \leq \gamma$ are utilized in the innovation of the i-th feature of the object. Let $m_i[k]$ be the number of observations in the validation region of the i-th feature point. The weighting factor for each observation inside the validation region is defined in the following. The association probability of each observation comprises the occlusion probability $\beta_{i,0}$ and weighting factors $\beta_{i,j}$ $(j = 1, \dots, m_i[k])$ corresponding to each observation $\boldsymbol{z}_{i,j}$ in the validation region. Weighting factors $\beta_{i,j}$ $(j = 1, \dots, m_i[k])$ are represented by

$$\beta_{i,j} = \begin{cases} \dfrac{1 - P_D P_G}{1 - P_D P_G + \sum\limits_{j=1}^{m_i[k]} L_{i,j}[k]} & (j = 0) \\ \dfrac{L_{i,j}[k]}{1 - P_D P_G + \sum\limits_{j=1}^{m_i[k]} L_{i,j}[k]} & (j = 1, \dots, m_i[k]) \end{cases} \quad (16)$$

where P_D is the target detection probability and P_G is the gate probability. Let λ be the rate of the Poisson distribution. Then, the likelihood ratio $L_{i,j}$ is represented by Eq. (17):

$$L_{i,j} = \frac{\mathcal{N}\left[\boldsymbol{z}_{i,j}[k] : \hat{\boldsymbol{y}}_i^-, \boldsymbol{S}_{\mathrm{m},i}[k]\right] P_D}{\lambda}. \quad (17)$$

3.3. Filtering Step

The posterior covariance matrix is calculated in the filtering step. Kalman gain $\boldsymbol{W}[k]$ is represented by

$$\boldsymbol{W}[k] = \boldsymbol{P}^-[k]\boldsymbol{C}[k]\boldsymbol{S}[k]^{-1}. \quad (18)$$

The combined innovation for each feature point is defined as

$$\boldsymbol{\nu}_i[k] = \sum_{j=1}^{m_i[k]} \beta_{i,j} \boldsymbol{\nu}_{i,j}[k], \quad (19)$$

where $\boldsymbol{\nu}_{i,j}$ is the innovation of each feature point represented by

$$\boldsymbol{\nu}_{i,j}[k] = \boldsymbol{z}_{i,j}[k] - \boldsymbol{h}_i(\hat{\boldsymbol{x}}[k]). \quad (20)$$

Further, the innovation comprising all features belonging to validation regions is represented by Eq. (21):

$$\boldsymbol{\nu}[k] = [\boldsymbol{\nu}_1[k]^{\mathrm{T}}, \dots, \boldsymbol{\nu}_s[k]^{\mathrm{T}}]^{\mathrm{T}}. \quad (21)$$

The posterior covariance matrix is calculated by

$$\boldsymbol{P}[k] = \boldsymbol{P}^-[k] - \boldsymbol{W}[k]\boldsymbol{S}[k]\boldsymbol{W}[k] + \tilde{\boldsymbol{P}}[k]. \quad (22)$$

Let $\boldsymbol{\mu}_j := [\sqrt{\beta_{1,j}}\boldsymbol{\nu}_{1,j}^{\mathrm{T}}, \sqrt{\beta_{2,j}}\boldsymbol{\nu}_{2,j}^{\mathrm{T}}, \dots, \sqrt{\beta_{s,j}}\boldsymbol{\nu}_{s,j}^{\mathrm{T}}]^{\mathrm{T}}$. The correction term of the posterior covariance matrix $\tilde{\boldsymbol{P}}$ is represented by Eq. (23):

$$\tilde{\boldsymbol{P}}[k] = \boldsymbol{W}[k]\boldsymbol{E}_\nu[k]\boldsymbol{W}[k]^{\mathrm{T}}, \quad (23)$$

with

$$\boldsymbol{E}_\nu[k] = \sum_{j=1}^{\bar{m}[k]} \boldsymbol{\mu}_j[k]\boldsymbol{\mu}_j[k]^{\mathrm{T}} - \boldsymbol{\nu}[k]\boldsymbol{\nu}[k]^{\mathrm{T}}, \quad (24)$$

where $\bar{m}[k]$ is the maximum number of $m_i[k]$. PDAF deals with the occlusion in the filtering step by considering occlusion probability as well as by changing the size of the vector and matrix of $\boldsymbol{W}[k]$, $\boldsymbol{S}[k]$, $\boldsymbol{\nu}[k]$ and $\boldsymbol{\mu}[k]$, adapting to the number of measurable feature points. This is applicable to nonlinear systems, but it is not straightforward to include constraints in such contexts because of the incremental update formulation of the Kalman filter.

4. Moving Horizon Estimation

Moving horizon estimation (MHE) is a model-based estimation method based on real-time optimization. It utilizes the observation and the input in a moving window representing the range from the current time to a finite time past. The optimal state estimate is obtained as a local minimum solution by numerically minimizing an objective function. A general objective function comprises the norms of the process error and the output error, and the arrival cost. It is generally represented by

$$\begin{aligned} \boldsymbol{J} = & \sum_{n=k-T}^{k-1} \|\hat{\boldsymbol{x}}[n+1] - \boldsymbol{f}(\hat{\boldsymbol{x}}[n], \boldsymbol{u}[n])\|^2_{(\boldsymbol{GQG}^T)^{-1}} \\ & + \sum_{n=k-T}^{k} \|\boldsymbol{y}[n] - \boldsymbol{h}(\hat{\boldsymbol{x}}[n])\|^2_{\boldsymbol{R}^{-1}} \\ & + \left\|\hat{\boldsymbol{x}}[k-T] - \hat{\boldsymbol{x}}_{\mathrm{p}}[k-T]\right\|^2_{\boldsymbol{P}^-[k-T]^{-1}}, \end{aligned} \quad (25)$$

where $\hat{\boldsymbol{x}}_{\mathrm{p}}[k-T]$ is the state estimation obtained in the previous sampling time. Error is evaluated as the Mahalanobis distance. The optimal estimation of MHE is obtained by solving an optimization problem for a set of state estimations $\hat{\boldsymbol{x}}[k], \hat{\boldsymbol{x}}[k-1], \dots, \hat{\boldsymbol{x}}[k-T]$. Since MHE is computed through numerical optimization, it can naturally include constraints. In addition, by virtue of recursive optimization, an optimal estimation can be obtained for nonlinear systems, while the extended Kalman filter provides a local solution. However, the calculation cost tends to be larger since it requires a nonlinear optimization problem to be solved for multiple unknown vectors of the current and past state estimates.

5. MHE-PDA

5.1. Configuration

Moving horizon estimation with probabilistic data association (MHE-PDA) is employed in which we incorporate PDAF into MHE so that multiple feature point candidates are stochastically associated, and the maximum likelihood estimation is obtained through optimization. **Fig. 5** depicts a block diagram of MHE-PDA. First, in the validation step, observations inside the validation region are extracted and weighted by a Gaussian probability of PDAF. The evaluation function of MHE-PDA comprises extracted observations and weight factors. Then, the optimal estimation is obtained in the MHE-PDA block by minimizing the evaluation function explained below. The validation regions on each horizon step are updated by re-

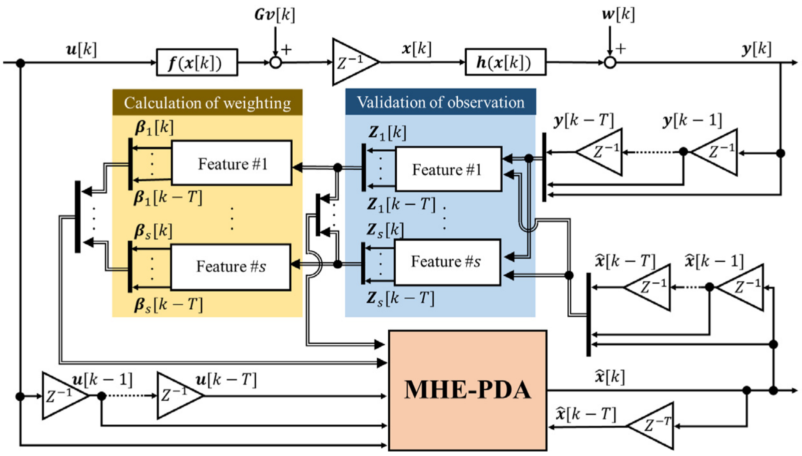


Fig. 5. Block diagram of MHE-PDA: $\boldsymbol{Z}_i$ is the set of observations of features for the i-th object defined as $\boldsymbol{Z}_i = \{\boldsymbol{z}_{i,j}\}_{j=1}^{m_i[k]}$. $\boldsymbol{Z}$ is the total set of objects represented by $\boldsymbol{Z} = \{\boldsymbol{Z}_i\}_{i=1}^{s}$. $\boldsymbol{B}_i$ is the set of weighting factors of the i-th object feature defined as $\boldsymbol{B}_i := \{\beta_{i,j}\}_{j=1}^{m_i[k]}$. $\boldsymbol{B}$ is the total set of weighting factors represented by $\boldsymbol{B} = \{\boldsymbol{B}_i\}_{i=1}^{s}$. In the validation step, observations inside the validation region are extracted and weighted by a Gaussian probability of PDAF. The evaluation function of MHE-PDA comprises extracted observations and weight factors. Then, the optimal estimation is obtained in the MHE-PDA block by minimizing the evaluation function explained below.

flecting the previous state estimations of MHE-PDA so that the influence of outliers in the past observations is reduced.

5.2. Objective Function

MHE-PDA conducts state estimation and probabilistic data association simultaneously by minimizing an evaluation function composed of the weighted prediction errors for each possible observation, the error of the motion model and the arrival cost. The evaluation function of MHE-PDA is represented by

$$\begin{aligned}\boldsymbol{J} = &\sum_{n=k-T}^{k-1} \|\hat{\boldsymbol{x}}[n+1] - \boldsymbol{f}(\hat{\boldsymbol{x}}[n], \boldsymbol{u}[n])\|^2_{(\boldsymbol{GQG}^T)^{-1}} \\ &+ \sum_{n=k-T}^{k} \sum_{i=1}^{s} \sum_{j=1}^{m_i} \|\beta_{i,j}(\boldsymbol{z}_{i,j}[n] - \boldsymbol{h}_i(\hat{\boldsymbol{x}}[n]))\|^2_{\boldsymbol{R}_m^{-1}} \\ &+ \|\hat{\boldsymbol{x}}[k-T] - \hat{\boldsymbol{x}}_{\mathrm{p}}[k-T]\|^2_{\boldsymbol{P}^-[k-T]^{-1}} \quad . \quad . \quad . \quad (26)\end{aligned}$$

The difference from Eq. (25) appears in the second term, in which the output prediction error is evaluated for multiple observations for each feature using the weight factor $\beta_{i,j}$ calculated by Eq. (16). When the observation does not exist in the validation region, the weight factor $\beta_{i,j}$ becomes zero, and the optimal estimate is solely determined by the motion model and the arrival cost. On the other hand, when observations exist in the validation region, each observation is associated with a probability depending on the likelihood. Thus, MHE-PDA is the maximum likelihood estimation in which $\hat{\boldsymbol{x}}$ is obtained by minimizing the evaluation function represented by Eq. (26).

5.3. Constraint on System Noise

Under the assumption that the noise obeys a normal distribution, the state equation and the output equation

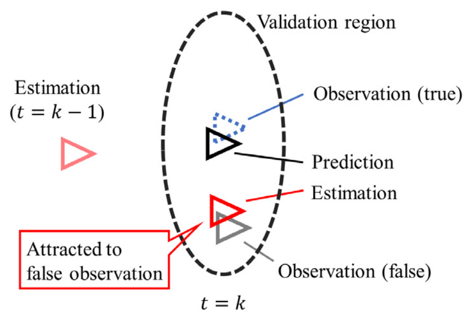


Fig. 6. When the true feature is occluded, the output estimate tends to be attracted toward the direction of the false observation.

are occasionally disturbed by substantial noise. In object tracking, false observations also disturb the estimation. The validation region can exclude such observations far from its prediction on both PDAF and MHE-PDA. However, false observations inside the validation region may jeopardize the output estimate, especially when the true feature is occluded, as depicted in **Fig. 6**.

Since the MHE framework can naturally incorporate constraints, MHE-PDA can also specify the condition restricting the possible region where the state estimate may exist. Let us consider the probability distributions depicted in **Fig. 7**. The dotted curve depicts a normal distribution, and the solid curve depicts a truncated distribution, which restricts the magnitude of noise. We introduce such a magnitude constraint into the system noise in Eq. (1). Constraints reflecting this type of amplitude restriction are represented by

$$-\boldsymbol{d} \preceq \boldsymbol{G}^{-1}(\hat{\boldsymbol{x}}[n] - \boldsymbol{f}(\hat{\boldsymbol{x}}[n-1], \boldsymbol{u}[n-1])) \preceq \boldsymbol{d} \quad . \ (27)$$

for $n = k, \ldots, k-T$, where $\boldsymbol{d} \in \mathbb{R}^n$ indicates the bound on

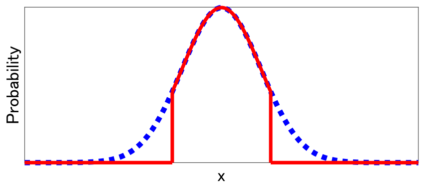
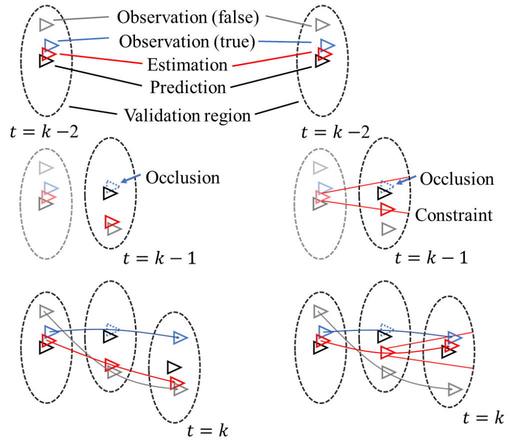


Fig. 7. The distribution with a system noise constraint.



(a) MHE-PDA without constraint (b) MHE-PDA with constraint

Fig. 8. Tracking in a multi-sampling context.

the magnitude, and " $\preceq$ " indicates the inequality between the corresponding elements.

Figure 8 depicts tracking behaviors for (a) MHE-PDA without constraint and (b) MHE-PDA with constraint. Both figures consider the situation where another object crosses the path when the true feature is occluded. In (a) MHE-PDA without constraints, the estimate is attracted toward the false feature of another object and the estimate deviates from the target object. In (b) MHE-PDA with constraints, the magnitude restriction on system noise limits the possible region and it prevents the estimate from tracking the false feature and later finds the true feature.

6. Simulation

We verify the effectiveness of the proposed method through numerical simulations, in which we assume that observations are disturbed by multiple false features and occlusions. Performance is evaluated by comparing with the conventional PDAF and MHE-PDA with and without constraints on the magnitude of the system noise.

6.1. Simulation Model

In this simulation, we consider a differential drive two wheeled vehicle depicted in **Fig. 9**. It moves on the horizontal plane, and its coordinates are represented by three variables: the translational position x_g and y_g, and the orientation θ. Let r be the wheel radius, W be the distance

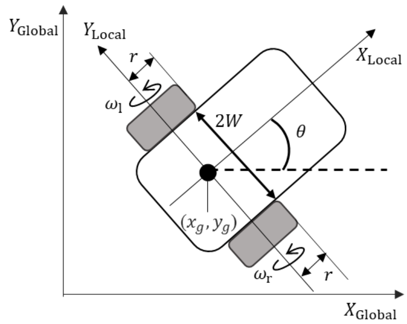


Fig. 9. Model of a differential drive vehicle.

between each wheel, and the rotational velocities of the left and the right wheel are ω_l and ω_r, respectively. The state $\boldsymbol{x}$ is defined by Eq. (28) as follows:

$$\boldsymbol{x}[k] = \left[x_\mathrm{g}[k], y_\mathrm{g}[k], \theta[k]\right]^\mathrm{T} \quad \dots \dots \dots (28)$$

The input $\boldsymbol{u}$ is defined by

$$\boldsymbol{u}[k] = [\omega_\mathrm{l}[k], \omega_\mathrm{r}[k]]^\mathrm{T}, \quad \dots \dots \dots (29)$$

the translational velocity is represented by

$$V[k] = \frac{r\omega_\mathrm{l}[k] + r\omega_\mathrm{r}[k]}{2}, \quad \dots \dots \dots (30)$$

and the rotational velocity is represented by

$$\omega[k] = \frac{-r\omega_\mathrm{l}[k] + r\omega_\mathrm{r}[k]}{2W} \quad \dots \dots \dots (31)$$

Then in the discretized state Eq. (1), $\boldsymbol{f}(\boldsymbol{x}[k], \boldsymbol{u}[k])$ is represented by

$$\boldsymbol{f}(\boldsymbol{x}[k], \boldsymbol{u}[k]) = \boldsymbol{x}[k] + \begin{bmatrix} V[k]\cos\theta[k] \\ V[k]\sin\theta[k] \\ \omega[k] \end{bmatrix} \Delta \quad \dots (32)$$

and the coefficient matrix $\boldsymbol{G}$ is described by

$$\boldsymbol{G} = \mathrm{diag}(\Delta, \Delta, \Delta). \quad \dots \dots \dots (33)$$

where Δ indicates the sampling time.

Let $\boldsymbol{H}$ be a rotation matrix between the object fixed coordinate system and the global coordinate system represented by

$$\boldsymbol{H}(\theta[k]) = \begin{bmatrix} \cos\theta[k] & -\sin\theta[k] \\ \sin\theta[k] & \cos\theta[k] \end{bmatrix} \quad \dots \dots (34)$$

Let $\boldsymbol{x}_{\mathrm{m},i}[k]$ $(i = 1, \dots, s)$ be the feature point position described on the coordinates fixed to the target object. Then the output function $\boldsymbol{h}_i$ for the i-th feature point is represented by

$$\boldsymbol{h}_i(\boldsymbol{x}[k]) = \boldsymbol{H}(\theta[k])\boldsymbol{x}_{\mathrm{m},i} + \begin{bmatrix} x_\mathrm{g}[k] \\ y_\mathrm{g}[k] \end{bmatrix}. \quad \dots \dots (35)$$

6.2. Simulation Environment

In this simulation, we assume that a vehicle moves through a crowded environment, and the features are observed by an external camera. **Fig. 10** indicates the driv-

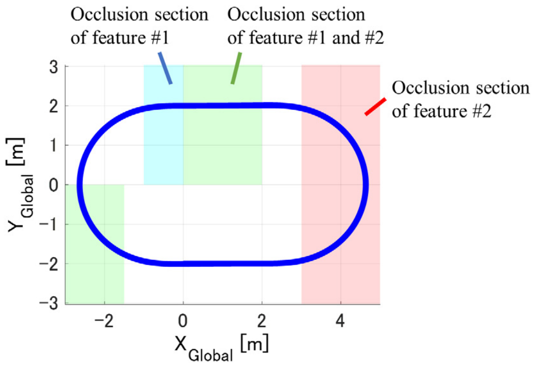


Fig. 10. Vehicle path: in each colored rectangular region, features are occluded. Feature #1 is occluded in the top left region, feature #2 is occluded in the right hand side region, and features #1 and #2 are occluded in the top middle and left bottom regions. Feature #3 is always observable.

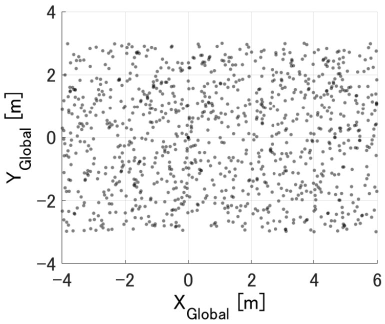


Fig. 11. The distribution of static false features.

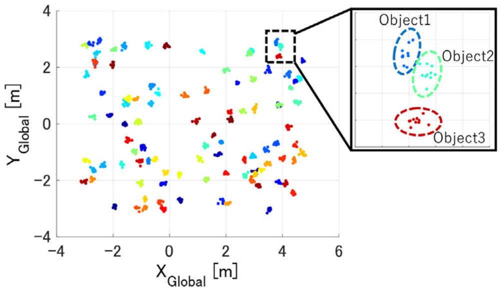
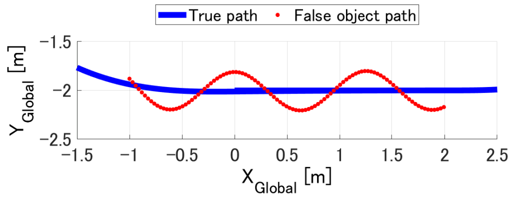
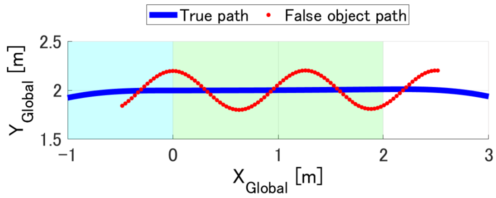


Fig. 12. Fake dynamic observations.



(a) Section 1.



(b) Section 2.

Fig. 13. The false object path arranged ad per the estimated object.

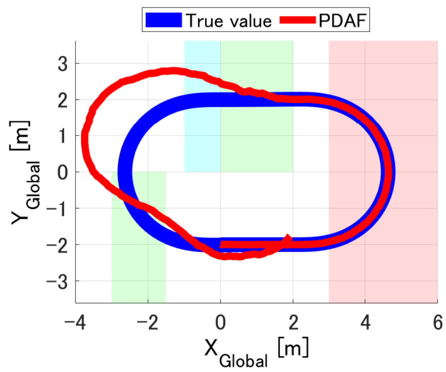
ing path of the vehicle. To evaluate effectiveness when faced with occlusion, one or two features are occluded in the colored rectangular region. Feature #1 is occluded in the top left region, feature #2 is occluded in the right hand side region, and features #1 and #2 are occluded in the top middle and left bottom regions. In addition, to simulate misrecognition, two kinds of false features are set. One is static features, and the other is a dynamic object which moves across the target object. The static features represent feature points attached to static objects. They are placed as a uniform distribution in the range $-4 \le x \le 6$ and $-3 \le y \le 3$ as depicted in **Fig. 11**. The dynamic group is composed of ten false features as depicted in **Fig. 12**, where the same color features belong to the same object. The dynamic groups are assumed to be feature points of dynamic objects such as pedestrians and other mobile objects. Each feature moves with a constant velocity in the section $-4 \le x \le 6$ and $-3 \le y \le 3$. Moreover, as shown in the sinusoidal path in **Fig. 13**, a false object which has the same assignment of features with the estimated object exists in Section 1 ($-1.0 \le x \le 2.0$ and $y \le 0$) and Section 2 ($-0.5 \le x \le 2.5$ and $0 \le y$). In Sections 1 and 2, the estimated object and the false dynamic object converge, where the false dynamic object hovers around the target object as the sinusoidal curve depicted in **Fig. 13**. They move along a sinusoidal wave around the path. In Section 2, two features are occluded; thus, false recognition can easily occur. The simulation parameters are summarized in **Table 1**. The vehicle parameters and $\boldsymbol{Q}$, $\boldsymbol{R}$ were taken from the experimental system [35]. λ, P_D and P_G, were determined by calibration of PDAF to track the target under the environment with static false features depicted in **Fig. 11**. γ was selected to be small so that the validation regions for each feature point did not overlap. $\boldsymbol{d}$ was determined by calibration reflecting the actual motion constraint of the vehicle.

6.3. Simulation Results and Discussion

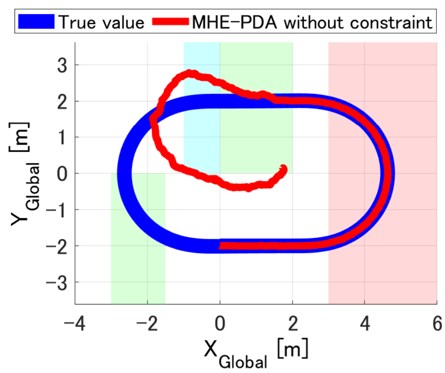
The estimated path of PDAF, MHE-PDA without constraint, and MHE-PDA with constraint are depicted in **Figs. 14(a)**, **14(b)**, and **15(a)**, respectively. PDAF and MHE-PDA without constraint could track the target object until Section 2. However, these methods failed in Section 2, where the object contained features with the same allocation as the estimated object while the features of the target were occluded. On the other hand, the proposed method succeeded in tracking the estimated object in each occlusion section. **Fig. 15(b)** depicts the estimate of the heading angle θ, which also tracked the actual object. **Table 2** indicates the resulting root-mean-square error (RMSE) of PDAF for $t = 0$–16 s, MHE-PDA with con-

Table 1. Definition of variables.

Parameter	Value
Simulation time [s]	40
Cycle time [s]	0.05
W [m]	0.492
r [m]	0.128
v [m/s]	0.6
T	30
γ	100
λ	30
P_D	0.8
P_G	0.8
s	3
$\boldsymbol{d}$	$[0.0071, 0.0071, 0.0068]^{\mathrm{T}}$
$\boldsymbol{Q}$	$\mathrm{diag}(8.0, 8.0, 7.5) \times 10^{-2}$
$\boldsymbol{R}$	$\mathrm{diag}(6.7, 7.1) \times 10^{-5}$
$\boldsymbol{x}_0$	$[0.0, -2.0, 0.0]^{\mathrm{T}}$
$\boldsymbol{x}_{\mathrm{m},1}$	$[0.3, -0.12]^{\mathrm{T}}$
$\boldsymbol{x}_{\mathrm{m},2}$	$[-0.15, 0.24]^{\mathrm{T}}$
$\boldsymbol{x}_{\mathrm{m},3}$	$[-0.15, -0.12]^{\mathrm{T}}$
N_{s}	1000
N_{m}	100

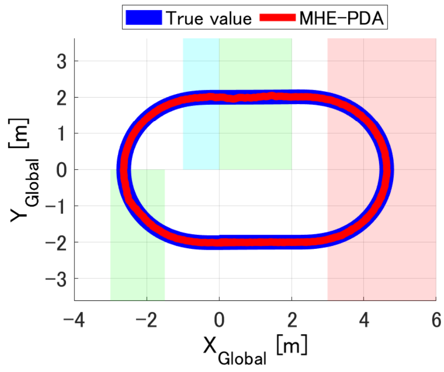


(a) PDAF.

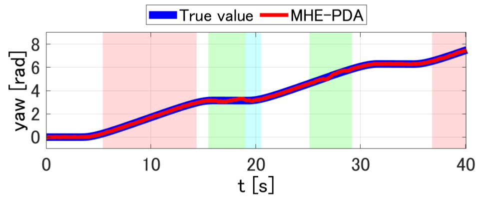


(b) MHE-PDA without constraint.

Fig. 14. Simulation results using the comparison method.



(a) Trajectory of MHE-PDA.



(b) The time transition of θ.

Fig. 15. Simulation results using MHE-PDA.

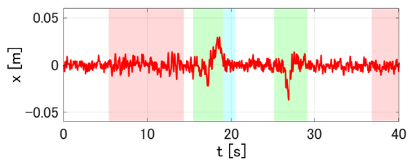
Table 2. RMSE of each estimated method.

	Section	x [mm]	y [mm]	θ [rad]
PDAF	$t \leq 16$ s	5.3	4.8	0.014
MHE-PDA	$t \leq 16$ s	5.0	5.0	0.013
MHE-PDA	All	6.6	9.1	0.034

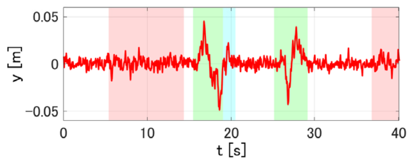
straint for $t = 0$–16 s, and MHE-PDA with constraint for $t = 0$–40 s, respectively. This shows that the deviation of each method was sufficiently small until $t = 16$ s. MHE-PDA outperformed PDAF after features were occluded. The RMSE of the proposed method increased, but it was still small.

In the following, we focus on the MHE-PDA with constraints. **Fig. 16** depicts the time transition of estimation error. The estimation error increased in the region where two features are occluded, thus the estimation is affected by the occlusion. However, the proposed method prevented the tracking of false features by satisfying the constraints as shown in **Fig. 17**.

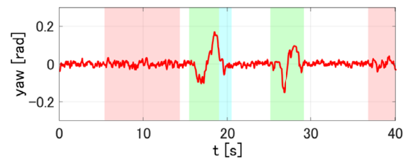
Figure 18 depicts the time transition of the number of observations which existed in the validation region. **Fig. 19** depicts snapshots of features of both the target and the false objects together with the validation region. The green, red, and blue points are the estimated features, the object feature points, and the validation regions, respectively. The purple points are the observations of static object features, dynamic object features, and estimated object features shown in **Figs. 11** and **12**, respectively. **Fig. 19(a)** shows that features #1 and #2 had many observations in each validation region at $t = 15.75$ s. **Fig. 19(b)** shows that features #1 and #2 did not have an observation in each validation region at $t = 25.5$ s. **Fig. 19(c)** shows



(a) x.

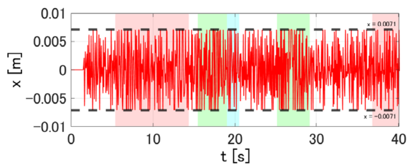


(b) y.

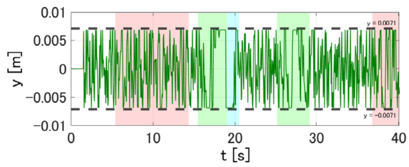


(c) θ.

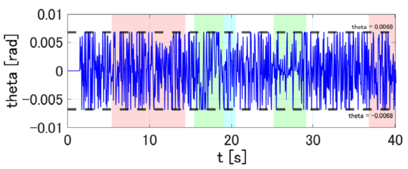
Fig. 16. The time transition of state estimation error.



(a) x.



(b) y.

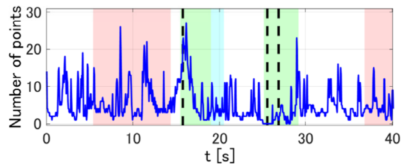


(c) θ.

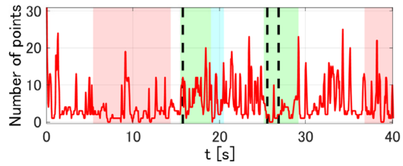
Fig. 17. The error of estimated value and predicted value at $k = T$.

that features #1 and #2 were occluded and false observations were predicted at $t = 26.8$ s.

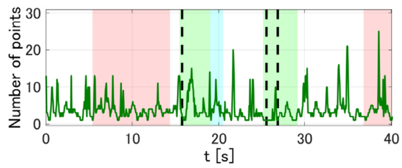
Figure 20 depicts the time transition of the sum of the weighting factors, $\sum_{j=1}^{m_i} \beta_{i,j}$, for each object feature. Feature #3 was not occluded in this simulation; thus, its weighting factor was kept high. However, the weights of features #1 and #2 decreased when features #1 and #2 were occluded. Thus, the proposed method could suppress false recognition, due to the multiple observations and occlusion by weighting factors.



(a) Feature #1.



(b) Feature #2.



(c) Feature #3.

Fig. 18. The number of feature points inside each validation region. The dotted lines indicate $t = 15.75$ s, $t = 25.5$ s, and $t = 26.8$ s, respectively.

7. Conclusion

In this paper, we proposed MHE-PDA, which incorporates probabilistic data association into the framework of MHE, as a method in which nonlinear estimation can be accomplished via numerical optimization. The proposed state estimation method utilizes a system noise constraint and is robust against occlusion and false recognition. In the simulation, we verified the effectiveness of the proposed method against occlusion and false recognition by comparing with PDAF. Future work in this domain requires implementing the proposed method into on-board cameras to validate its effectiveness.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 19H02098.

References:

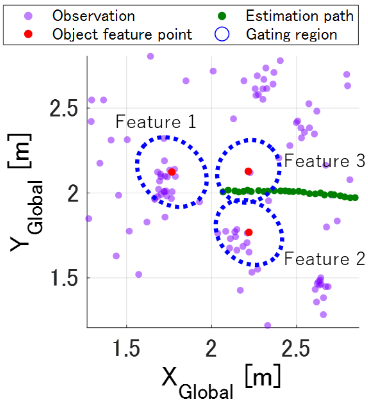
[1] F. Chenavier and J. L. Crowley, "Position estimation for a mobile robot using vision and odometry," Proc. of 1992 IEEE Int. Conf. on Robotics and Automation, Vol.3, pp. 2588-2593, 1992.

[2] K. Sakai, Y. Maeda, S. Miyoshi, and H. Hikawa, "Visual feedback robot system via fuzzy control," Proc. of SICE Annual Conf. 2010, pp. 3264-3267, 2010.

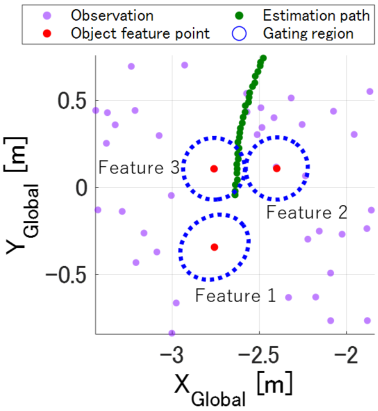
[3] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," Int. J. of Computer Vision, Vol.87, No.1, pp. 4-27, 2009.

[4] S. A. Habsi, M. Shehada, M. Abdoon, A. Mashood, and H. Noura, "Integration of a vicon camera system for indoor flight of a parrot ar drone," Proc. of 2015 10th Int. Symp. on Mechatronics and its Applications (ISMA), pp. 1-6, 2015.

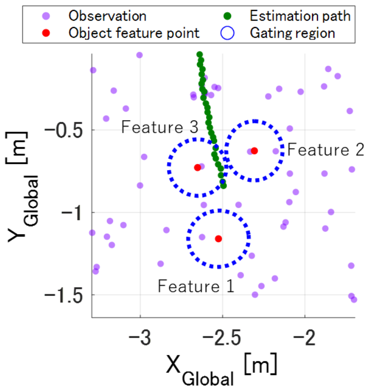
[5] Z. T. Dydek, A. M. Annaswamy, and E. Lavretsky, "Adaptive control of quadrotor uavs: A design trade study with flight evalua-



(a) $t = 15.75$ s.



(b) $t = 25.5$ s.



(c) $t = 26.8$ s.

Fig. 19. Features of the target vehicle and false objects.

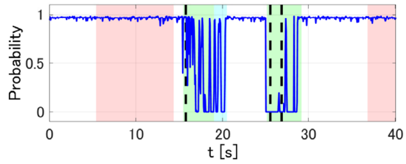
tions," IEEE Trans. on Control Systems Technology, Vol.21, No.4, pp. 1400-1406, 2013.

[6] J. Gall, B. Rosenhahn, T. Brox, and H.-P. Seidel, "Optimization and filtering for human motion capture," Int. J. of Computer Vision, Vol.87, No.1, pp. 75-92, 2008.

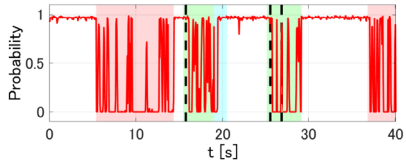
[7] K. Ohno, T. Nomura, and S. Tadokoro, "Real-time robot trajectory estimation and 3d map construction using 3d camera," 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 5279-5285, 2006.

[8] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur, "A survey of vision-based traffic monitoring of road intersections," IEEE Trans. on Intelligent Transportation Systems, Vol.17, No.10, pp. 2681-2698, 2016.

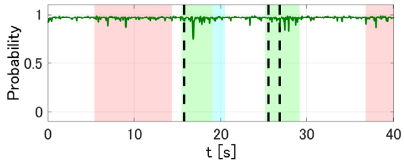
[9] Z. Liu and Z. You, "A real-time vision-based vehicle tracking and traffic surveillance," Proc. of 8th ACIS Int. Conf. on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007), Vol.1, pp. 174-179, 2007.



(a) Feature 1.



(b) Feature 2.



(c) Feature 3.

Fig. 20. The sum of weighting factors in the validation region of each feature. The dotted lines indicate $t = 15.75$ s, $t = 25.5$ s, and $t = 26.8$ s, respectively.

[10] J. Sochor, R. Juránek, and A. Herout, "Traffic surveillance camera calibration by 3d model bounding box alignment for accurate vehicle speed measurement," Computer Vision and Image Understanding, Vol.161, pp. 87-98, 2017.

[11] Y. Yuan, Z. Xiong, and Q. Wang, "An incremental framework for video-based traffic sign detection, tracking, and recognition," IEEE Trans. on Intelligent Transportation Systems, Vol.18, No.7, pp. 1918-1929, 2017.

[12] V. D. Nguyen, H. Van Nguyen, D. T. Tran, S. J. Lee, and J. W. Jeon, "Learning framework for robust obstacle detection, recognition, and tracking," IEEE Trans. on Intelligent Transportation Systems, Vol.18, No.6, pp. 1633-1646, 2017.

[13] G. Prabhakar, B. Kailath, S. Natarajan, and R. Kumar, "Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving," 2017 IEEE Region 10 Symp. (TENSYMP), pp. 1-6, 2017.

[14] L. A. Camñas-Mesa, T. Serrano-Gotarredona, S. Ieng, R. Benosman, and B. Linares-Barranco, "Event-driven stereo visual tracking algorithm to solve object occlusion," IEEE Trans. on Neural Networks and Learning Systems, Vol.29, No.9, pp. 4223-4237, 2018.

[15] A. Kainuma, H. Madokoro, K. Sato, and N. Shimoi, "Occlusion-robust segmentation for multiple objects using a micro air vehicle," Proc. of 2016 16th Int. Conf. on Control, Automation and Systems (ICCAS), pp. 111-116, 2016.

[16] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, "Object tracking with multi-view support vector machines," IEEE Trans. on Multimedia, Vol.17, No.3, pp. 265-278, 2015.

[17] C. H. Kuo, S. W. Sun, and P. C. Chang, "A skeleton-based pairwise curve matching scheme for people tracking in a multi-camera environment," Proc. of 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf., pp. 1-5, 2013.

[18] T. Umeda, K. Sekiyama, and T. Fukuda, "Vision-based object tracking by multi-robots," J. Robot. Mechatron., Vol.24, No.3, pp. 531-539, 2012.

[19] M. Wang, Y. Liu, D. Su, Y. Liao, L. Shi, J. Xu, and J. Valls Miro, "Accurate and real-time 3-d tracking for the following robots by fusing vision and ultrasonar information," IEEE/ASME Trans. on Mechatronics, Vol.23, No.3, pp. 997-1006, 2018.

[20] Y. Yun, P. Agarwal, and A. D. Deshpande, "Accurate, robust, and real-time estimation of finger pose with a motion capture system," Proc. of 2013 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 1626-1631, 2013.

[21] E. S. Lee and D. Kum, "Feature-based lateral position estimation of surrounding vehicles using stereo vision," Proc. of 2017 IEEE Intelligent Vehicles Symp. (IV), pp. 779-784, 2017.

[22] M. Takahashi, K. Nonaka, and K. Sekiguchi, "Moving horizon estimation for vehicle robots using partial marker information of motion capture system," J. of Physics: Conf. Series, Vol.744, No.1, 012049, 2016.

[23] R. Ding, M. Yu, and W.-H. Chen, "A multiple target tracking strategy using moving horizon estimation approach aided by road constraint," Proc. of 24th Int. Technical Conf. on the Enhanced Safety of Vehicles (ESV), pp. 1-13, 2015.

[24] R. Ding, M. Yu, H. Oh, and W. Chen, "New multiple-target tracking strategy using domain knowledge and optimization," IEEE Trans. on Systems, Man, and Cybernetics: Systems, Vol.47, No.4, pp. 605-616, 2017.

[25] Y. Ishikawa, N. Takahashi, T. Takahama, and K. Nonaka, "Moving horizon estimation of lateral and angular displacement for steering control," Trans. of the Society of Automotive Engineers of Japan, Vol.50, No.5, pp. 1487-1493, 2019.

[26] A. Liu, W. Zhang, M. Z. Q. Chen, and L. Yu, "Moving horizon estimation for mobile robots with multirate sampling," IEEE Trans. on Industrial Electronics, Vol.64, No.2, pp. 1457-1467, 2017.

[27] T. Kikuchi, K. Tsuno, K. Nonaka, and K. Sekiguchi, "Continuous marker association utilizing potential function for motion capture systems," Proc. of 2019 IEEE/SICE Int. Symp. on System Integration (SII), pp. 578-583, 2019.

[28] R. Abe, T. Kikuchi, K. Nonaka, and K. Sekiguchi, "Robust object tracking with continuous data association based on artificial potential moving horizon estimation," 21st IFAC World Congress, 2020 (in press).

[29] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter," Proc. of the IEEE, Vol.92, No.3, pp. 536-557, 2004.

[30] S. B. Colegrove and S. J. Davey, "The probabilistic data association filter with multiple nonuniform clutter regions," Record of the IEEE 2000 Int. Radar Conf., pp. 65-70, 2000.

[31] X. Chen, Y. Li, Y. Li, J. Yu, and X. Li, "A novel probabilistic data association for target tracking in a cluttered environment," Sensors (Basel), Vol.16, No.12, 2180, 2016.

[32] T. Yang and P. G. Mehta, "Probabilistic data association-feedback particle filter for multiple target tracking applications," J. of Dynamic Systems, Measurement, and Control, Vol.140, pp. 1-14, 2017.

[33] T. Kikuchi, K. Nonaka, and K. Sekiguchi, "Visual object tracking by moving horizon estimation with probabilistic data association," Proc. of 2020 IEEE/SICE Int. Symp. on System Integration (SII), pp. 115-120, 2020.

[34] M. Grinberg, "Feature-based Probablistic Data Association for Video-Based Multi-Object Tracking," KIT Scientific Publishing, 2017.

[35] K. Shibata, K. Nonaka, and K. Sekiguchi, "Model predictive obstacle avoidance control suppressing expectation of relative velocity against obstacles," 2019 IEEE Conf. on Control Technology and Applications (CCTA), pp. 59-64, 2019.

Name:
Tomoya Kikuchi

Affiliation:
Graduate Student, Division of Mechanics Engineering, Graduate School of Integrative Science and Engineering, Tokyo City University

Address:
1-28-1 Tamazutsumi, Setagaya, Tokyo 158-8857, Japan

Brief Biographical History:
2018 Received Bachelor degree from Tokyo City University
2018- Graduate Student, Division of Mechanics Engineering, Graduate School of Integrative Science and Engineering, Tokyo City University

Main Works:
- State estimation, object tracking, vision sensor

Membership in Academic Societies:
- The Society of Instrument and Control Engineers (SICE)

Name:
Kenichiro Nonaka

Affiliation:
Department of Mechanical Systems Engineering, Faculty of Engineering, Tokyo City University

Address:
1-28-1 Tamazutsumi, Setagaya, Tokyo 158-8857, Japan

Brief Biographical History:
1997 Received Doctor of Engineering from Tokyo Institute of Technology
1997- Research Assistant, Musashi Institute of Technology
2000- Lecturer, Musashi Institute of Technology
2007- Associate Professor, Musashi Institute of Technology
2013- Professor, Tokyo City University

Main Works:
- Model predictive control, moving horizon estimation, vehicle control

Membership in Academic Societies:
- The Society of Instrument and Control Engineers (SICE)
- The Institute of Electrical and Electronics Engineers (IEEE)
- The Japan Society of Mechanical Engineers (JSME)

Name:
Kazuma Sekiguchi

Affiliation:
Department of Mechanical Systems Engineering, Faculty of Engineering, Tokyo City University

Address:
1-28-1 Tamazutsumi, Setagaya, Tokyo 158-8857, Japan

Brief Biographical History:
2010 Received Doctor of Engineering from Tokyo Institute of Technology
2010- Assistant Professor, Tokyo Institute of Technology
2013- Assistant Professor, Tokyo City University
2015- Lecturer, Tokyo City University
2019- Associate Professor, Tokyo City University

Main Works:
- Nonlinear systems control, underactuated system control, UAV control

Membership in Academic Societies:
- The Institute of Electrical and Electronics Engineers (IEEE)
- The Society of Instrument and Control Engineers (SICE)
- The Institute of Systems, Control and Information Engineers (ISCIE)