

Multiple-Hypothesis Chance-Constrained Target Tracking Under Identity Uncertainty

Yoonseon Oh and Songhwai Oh

Abstract—We propose a robust target tracking algorithm for a mobile robot under identity uncertainty, which arises in crowded environments. When a mobile robot has a sensor with a fan-shaped field of view and finite sensing region, the proposed algorithm aims to minimize the probability of losing a moving target. We predict the next position of a moving target in a crowded environment using a multiple-hypothesis prediction algorithm which combines the motion model and appearance model of the target. When the distribution of the target's next position follows a Gaussian mixture model, the proposed tracking algorithm can track a target with a guaranteed tracking success probability. If the tracking success probability is sufficiently good, the method minimizes the moving distance of the mobile robot. The performance of the method is extensively validated in simulation and experiments using a Pioneer robot with a Microsoft Kinect sensor.

I. INTRODUCTION

An ability to reliably track a moving target in crowded environments is an enabling technology for many service robot applications, such as hospital monitoring, surveillance, and museum guidance [1]. In a crowded environment, there can be false detections from a crowd and identity uncertainty about the target can make the tracking problem more challenging. It is important to guarantee the performance of a mobile robot against noises and identity uncertainties. In addition, if a mobile robot is used in a domestic environment, such as hospitals, nursing homes, and homes, it is desirable to reduce the moving distance of a mobile robot since frequent movements draw attention from users.

A target tracking problem can be considered as a navigation problem by interpreting the position of a target as a goal position and steering a robot to the goal, and this approach is taken by [2]–[4]. It is often assumed that a target can be always detected and there is no identity uncertainty, i.e., there is no false detection. Target tracking is treated as a problem of maintaining the visibility of a moving target using a sensor with a finite sensing region [5], [6] or minimizing a certain risk function [7]–[9].

When tracking a target, it is important to consider the uncertainty in our estimate about the target's position. In [10] and [11], the proposed controllers minimize the uncertainty of the predicted location of the target. In practice, however, it is more desirable to minimize the tracking failure probability

rather than minimizing the prediction uncertainty in order to provide a guaranteed performance. Tracking algorithms in [12], [13] maximize the probability of future visibility and suggest a robust algorithm against uncertainty. However, the suggested algorithms cannot handle false detections, hence, they cannot be applied in crowded environments.

A probabilistic data association filter (PDAF) can provide more robust pedestrian tracking in cluttered environments [14]. In [15], the authors applied a PDAF to track a target and improved detection performance by combining color information about the target with laser data. However, the suggested algorithm does not handle false detections systematically and operates as a nearest-neighbor filter.

This paper proposes a tracking algorithm such that a mobile sensor with a finite and fan-shaped sensing region maintains visibility of a moving target in crowded environments. Our method predicts the target's next position as a Gaussian Mixture model (GMM) based on multiple hypotheses [14], when there are multiple detections. For a given predictive distribution, we propose a chance-constrained target tracking method which can guarantee the tracking performance under uncertainty. When the tracking success probability is expected to be good enough, we minimize the moving distance of the robot. Compared to [12], new motion and appearance models are proposed to handle false detections in cluttered environments and a completely new optimization method is developed to provide a tighter upper bound on the tracking failure probability.

We formulate the described target tracking problem as an optimization problem to minimize the upper bound on the tracking failure probability and the moving distance of a robot. While the proposed optimization problem is complex, we solve the problem in real-time by finding solutions for different heading directions and choosing the best one. We have validated the performance of the proposed method extensively from a number of simulations and experiments using a Pioneer robot with a Kinect sensor for tracking a pedestrian in a crowd.

The remainder of this paper is structured as follows. The problem of target tracking under identity uncertainty is formulated in Section II. A multiple-hypothesis prediction algorithm is described in Section III. The proposed scheme to estimate the tracking failure probability is described in Section IV. A solution to the proposed target tracking problem is presented in Section V. Results from simulation and real-world experiments are presented in Section VI.

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2009348) and by ICT R&D program of MSIP/IITP (14-824-09-013, Resilient Cyber-Physical Systems Research).

Y. Oh and S. Oh are with the Department of Electrical and Computer Engineering, ASRI, Seoul National University, Seoul 151-744, Korea (e-mail: {yoonseon.oh, songhwai.oh}@cpslab.snu.ac.kr).

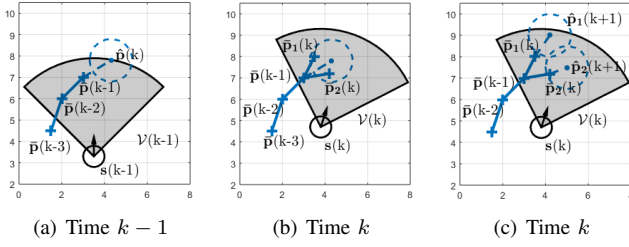


Fig. 1. An illustration of the target tracking problem considered in this paper. Gray regions are the sensing region of a mobile robot. (a) A target is detected at time $k-1$ and is located at $\bar{\mathbf{p}}(k-1)$. The predicted new location of the target is $\hat{\mathbf{p}}(k)$ at time k . The blue dashed line represent the variance of the prediction. (b) Measured positions at time k are $\bar{\mathbf{p}}_1(k)$ and $\bar{\mathbf{p}}_2(k)$. (c) Predicted positions based on $\bar{\mathbf{p}}_1(k)$ and $\bar{\mathbf{p}}_2(k)$ are $\hat{\mathbf{p}}_1(k)$ and $\hat{\mathbf{p}}_2(k)$, respectively. The mobile robot moves to $\mathbf{s}(k+1)$ to make sure that the target is within the sensing range with a guaranteed probability.

II. PROBLEM FORMULATION

We consider a mobile robot (or mobile sensor) and a target moving on a 2D plane (see Figure 1). The position of the mobile robot at time k is denoted by $\mathbf{s}(k) = [x_s(k) \ y_s(k)]^T$. The heading of the robot is the angle from the x -axis and denoted by $\varphi_s(k)$. We assume that a sensor is rigidly attached to the mobile robot and its direction is the same as the heading of the robot. The sensor has a finite and fan-shaped sensing region and the sensing region at time k is denoted by $\mathcal{V}(k)$ (shaded regions in Figure 1). Its angular field of view is θ_s and its maximum range is R_s .

The dynamic model of the mobile robot is assumed to be a discrete-time unicycle model, where control $\mathbf{u}(k)$ consists of a directional velocity u_v and an angular velocity u_w . For a unit interval of length T , the transition function $f(\mathbf{s}(k), \varphi_s(k), \mathbf{u}(k))$ is

$$\begin{aligned} \mathbf{s}(k) &= \mathbf{s}(k-1) + \mathbf{f}_r(u_v)u_v, \\ \varphi_s(k) &= \varphi_s(k-1) + u_w T, \end{aligned} \quad (1)$$

where

$$\mathbf{f}_r(u_v) = \begin{bmatrix} \int_0^T \cos(\varphi_s(k-1) + u_w t) dt \\ \int_0^T \sin(\varphi_s(k-1) + u_w t) dt \end{bmatrix}.$$

Its moving distance is derived as $d(\mathbf{u}) = |u_v|T$. The admissible ranges of control are

$$V_{\min} \leq u_v \leq V_{\max} \quad \text{and} \quad W_{\min} \leq u_w \leq W_{\max}. \quad (2)$$

The position of the target at time k is denoted by $\mathbf{p}(k) = [x_T(k) \ y_T(k)]^T$. If a target is in a crowded environment, there can be multiple detections. Let $Z(k) = \{z_1(k), \dots, z_{m_k}(k)\}$ denote measurements from the target and false detections at time k , where m_k is the number of detections (or measurements) at time k . Cumulative measurements until k are denoted by $Z^k = \{Z(1), \dots, Z(k)\}$. Each measurement is defined as $z_i(k) = \{\bar{\mathbf{p}}_i(k), \mathbf{g}_i(k)\}$, where $\bar{\mathbf{p}}_i(k)$ is a measured position and $\mathbf{g}_i(k)$ is a feature vector from the appearance model of the i th detected object.

Figure 1 shows an illustration, where a mobile sensor has detected a target from time $k-2$ to $k-1$ (Figure 1(a)) and moves to a new location to make sure that the target is

visible at time k (Figure 1(b)). When a single trajectory of the target is measured like Figure 1(a), the next position of the target is available using a motion prediction algorithm. The predicted position has a Gaussian distribution like a dashed line in Figure 1(b). However, an ambiguity can occur if there are more than one measurements as shown in Figure 1(b), i.e., we cannot be sure which measurement is from the designated target. For example, in Figure 1(c), if $\bar{\mathbf{p}}_1(k)$ is a true detection, the next predicted position will be $\hat{\mathbf{p}}_1(k+1)$. If $\bar{\mathbf{p}}_2(k)$ is a true detection, the next predicted position will be $\hat{\mathbf{p}}_2(k+1)$. Therefore, we combine all predictions from m_k measurements using a Gaussian mixture model. The mixture weight is determined by the likelihood function which is based on the appearance and motion models. The prediction algorithm is described in Section III.

Our goal is to find the control $\mathbf{u}(k)$ such that the target is located in the sensing region of the mobile robot at time k based on all measurements up to time $k-1$. Because of the uncertainty about the prediction, it is difficult to satisfy the deterministic visibility constraint, $\mathbf{p}(k) \in \mathcal{V}(k)$. For this reason, our controller guarantees the tracking failure probability, $P(\mathbf{p}(k) \notin \mathcal{V}(k))$. In addition, if the tracking failure probability is sufficiently small, we minimize the moving distance of a mobile sensor.

III. PREDICTION UNDER IDENTITY UNCERTAINTY

We propose a motion prediction algorithm which consistently predicts the position of a target in the presence of false detections from cluttered environments by applying the multiple-hypothesis framework [14].

A. Multiple-Hypothesis Prediction

We define a hypothesis as a set of measurements up to the current time, such that, for each measurement time, only one measurement is included in the set. If there are M_k hypotheses at time k , the j th hypothesis considers $z_{\mathcal{I}_j(k)}(k)$ as a true measurement about the target among $Z(k)$, where $\mathcal{I}_j(k)$ is the index of a measurement selected by the j th hypothesis. For each hypothesis, a motion prediction algorithm is applied to predict $\mathbf{p}(k+1)$. The j th hypothesis can be denoted by $\xi_j^k = \{\hat{\mathbf{p}}_j(k+1), \Sigma_j(k+1), Z_j^k\}$. Then, by considering the contribution of all hypotheses, the distribution of the next position $\mathbf{p}(k+1)$ follows

$$P(\mathbf{p}(k+1)|Z^k) = \sum_{j=1}^{M_k} P(\mathbf{p}(k+1)|\xi_j^k, Z^k) P(\xi_j^k|Z^k), \quad (3)$$

where $P(\mathbf{p}(k+1)|\xi_j^k, Z_j^k)$ is the predictive posterior distribution of the target's next position, which has the Gaussian distribution as described below. Hence, the distribution of $\mathbf{p}(k+1)$ given Z^k is a Gaussian mixture model (GMM) and $P(\xi_j^k|Z^k)$ is a mixture weight.

The first term on the right hand side of (3) is determined by a human motion model. We adopt an autoregressive Gaussian process motion model (AR-GPMM), a nonparametric motion model, as a motion model since it has been shown that an AR-GPMM can better predict the motion of a human than

parametric models [12], [16]. An AR-GPMM predicts the target's position $\mathbf{p}(k+1)$ as a Gaussian distribution using Gaussian process regression applied to past measurements. Then, the first term can be easily computed as follows:

$$\mathbf{p}(k+1)|\xi_j^k, Z^k \sim \mathcal{N}(\hat{\mathbf{p}}_j(k+1), \Sigma_j(k+1)).$$

In order to determine the last term of (3), we apply Baye's rule,

$$\begin{aligned} P(\xi_j^k|Z^k) &\propto P(Z(k)|\xi_j^k, Z^{k-1}) \\ &\times P(\xi_j^k|\xi_{m(j)}^{k-1}, Z^{k-1})P(\xi_{m(j)}^{k-1}|Z^{k-1}), \end{aligned} \quad (4)$$

where $\xi_{m(j)}^{k-1}$ is a parent hypothesis of ξ_j^k . When we assume that the position and appearance of the target are independent, the first term on the right hand side of (4) becomes

$$P(Z(k)|\xi_j^k, Z^{k-1}) = \prod_{i=1}^{m_k} P(\bar{\mathbf{p}}_i(k)|\xi_j^k)P(\mathbf{g}_i(k)|\xi_j^k).$$

The measured position $\bar{\mathbf{p}}_{\mathcal{I}_j(k)}(k)$ follows a Gaussian distribution whose mean and covariance are predicted by AR-GPMM based on measurements from time $k-p$ to $k-1$ in ξ_j^k , where p is the order of AR-GPMM. We assume that the measured positions of false detections are uniformly distributed in the sensing region with a volume of V_m , i.e., $P(\bar{\mathbf{p}}_{i \neq \mathcal{I}_j(k)}(k)|\xi_j^k) = 1/V_m$. Consequently, the conditional probability of measured positions is

$$\prod_{i=1}^{m_k} P(\bar{\mathbf{p}}_i(k)|\xi_j^k) = \left(\frac{1}{V_m}\right)^{m_k-1} P(\bar{\mathbf{p}}_{\mathcal{I}_j(k)}(k)|\xi_j^k). \quad (5)$$

Similarly, the conditional probability of feature vectors is

$$\prod_{i=1}^{m_k} P(\mathbf{g}_i(k)|\xi_j^k) = \left(\frac{1}{V'_m}\right)^{m_k-1} P(\mathbf{g}_{\mathcal{I}_j(k)}(k)|\xi_j^k), \quad (6)$$

where V'_m is the volume of the feature space. The distribution $P(\mathbf{g}_i(k)|\xi_j^k)$ represents the appearance similarity between the target and the i th detection.

The distribution $P(\xi_j^k|\xi_{m(j)}^{k-1}, Z^{k-1})$ in (4) is the prior distribution of hypothesis ξ_j^k which depends only on m_k [14]. Therefore, it is constant at time k and can be ignored. The last term $P(\xi_{m(j)}^{k-1}|Z^{k-1})$ in (4) is available from the previous iteration. Then, the posterior probability (4) can be computed as follows:

$$\begin{aligned} P(\xi_j^k|Z^k) &\propto P(\bar{\mathbf{p}}_{\mathcal{I}_j(k)}(k)|\xi_j^k)P(\mathbf{g}_{\mathcal{I}_j(k)}(k)|\xi_j^k) \\ &\times P(\xi_{m(j)}^{k-1}|Z^{k-1}). \end{aligned} \quad (7)$$

In our framework, the number of hypotheses is $M_k = M_{k-1}m_k$ and it increases exponentially. Hence, we prune hypotheses with low probabilities. At each step, we only allow at most $M_{\max}$ hypotheses with highest posterior probabilities. In addition, we delete a hypothesis if $P(\xi_j^k|Z^k)$ is less than some small number (1×10^{-10} is used in simulations and experiments).

B. Appearance Model for a Kinect Sensor

This section describes an appearance model for computing (6) when an RGB image and the skeleton of a pedestrian are available using a Kinect sensor¹. We generate five image patches around a torso, a hip, legs, shoulders, and arms, since each body part can have distinctive color information. The proposed appearance model $\mathbf{g}$ consists of five color histograms, i.e., $\mathbf{g} = \{\mathbf{g}^1, \mathbf{g}^2, \mathbf{g}^3, \mathbf{g}^4, \mathbf{g}^5\}$. Each $\mathbf{g}^j$ corresponds to a different body part.

We assume that the appearance model of the target is $\mathbf{g}_0 = \{\mathbf{g}_0^1, \mathbf{g}_0^2, \mathbf{g}_0^3, \mathbf{g}_0^4, \mathbf{g}_0^5\}$, which is acquired when tracking starts. Then, the distance between $\mathbf{g}$ and $\mathbf{g}_0$ is measured as

$$d_{sim} = \sum_{j=1}^5 w_j \|\mathbf{g}_0^j - \mathbf{g}^j\|,$$

where w_j is a weight of the j th body part, such that $\sum_{j=1}^5 w_j = 1$. Since a Kinect sensor does not detect all joints every time, the weights of undetected parts are set to zero. If $\mathbf{g}$ is from the target, the likelihood of $\mathbf{g}$ is

$$P(\mathbf{g}|\xi_j^k) \propto \frac{1}{\sigma_g} \phi\left(\frac{d_{sim}}{\sigma_g}\right),$$

where ϕ is a probability density function of the standard normal distribution and σ_g^2 is the variance. Notice that $P(\mathbf{g}|\xi_j^k)$ gets larger as d_{sim} approaches to zero.

IV. TRACKING FAILURE PROBABILITY

The sensing region at time k is a convex region bounded by l_1 , l_2 , and l_r as shown in Figure 2. To employ the chance-constrained method [17], we approximate the sensing region by an N -sided polygon bounded by lines from l_1 to l_N , which is denoted by $\bar{\mathcal{V}}$ (the darker region in Figure 2 for $N = 5$). Then we define $\mathbf{a}_i$ as the normal vector of l_i and b_i as the shortest distance between l_i and the origin as follows:

$$\begin{aligned} \mathbf{a}_1 &= \begin{bmatrix} -\sin(\varphi_s + \theta_s/2) \\ \cos(\varphi_s + \theta_s/2) \end{bmatrix}, & b_1 &= \mathbf{a}_1^T \mathbf{s}, \\ \mathbf{a}_2 &= \begin{bmatrix} \sin(\varphi_s - \theta_s/2) \\ -\cos(\varphi_s - \theta_s/2) \end{bmatrix}, & b_2 &= \mathbf{a}_2^T \mathbf{s}. \end{aligned}$$

$$\begin{aligned} \text{For } i &= 3, 4, \dots, N, \quad \mathbf{a}_i = \begin{bmatrix} \cos(\varphi_s + \frac{\theta_s(2i-5-N)}{2N}) \\ \sin(\varphi_s + \frac{\theta_s(2i-5-N)}{2N}) \end{bmatrix}, \\ b_i &= \mathbf{a}_i^T \mathbf{s} + R_s \cos(\theta_s/2(N-2)). \end{aligned}$$

We consider the problem of tightly bounding the tracking failure probability. We define the upper bound on the tracking failure probability as ϵ , such that $P(\mathbf{p} \notin \mathcal{V}) \leq \epsilon$. Since we consider a one-step look-ahead motion strategy, notations can be simplified by representing all variables and constraints relative to $\mathbf{s}(k-1)$. Without loss of generality, we assume that $\mathbf{s}_0 = [0 \ 0]^T$ and $\varphi_s(k-1) = \varphi_0$, where $\mathbf{s}_0 = \mathbf{s}(k-1)$. Let $\mathbf{s} = [x_s \ y_s]^T$ be the position of the sensor at time k . The relative position of the target from $\mathbf{s}_0$ at time k is

¹Xbox skeleton grab API (<http://msdn.microsoft.com/en-us/library/hh973074.aspx>).

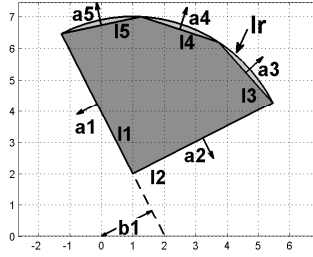


Fig. 2. The light gray region is the sensing region of a mobile robot located at $\mathbf{s}(k)$. The dark gray region is the approximated sensing region. $\mathbf{a}_i$ is the normal vector of line l_i .

denoted by $\mathbf{p}$. The position $\mathbf{p}$ is predicted as a GMM with M_k components. Each component is parameterized by mean $\hat{\mathbf{p}}_j$ and covariance Σ_j for hypothesis ξ_j . To simplify the notation further, we omit the time index k in our discussion below.

From (3), the tracking failure probability is²

$$P(\mathbf{p} \notin \mathcal{V}) = \sum_{j=1}^{M_k} P(\xi_j) P(\mathbf{p} \notin \mathcal{V} | \xi_j). \quad (8)$$

The conditional tracking failure probability is bounded by

$$P(\mathbf{p} \notin \mathcal{V} | \xi_j) \leq P(\mathbf{p} \notin \bar{\mathcal{V}} | \xi_j).$$

The tracking failure region of the approximated sensing region is the union of half-spaces of hyperplanes from l_1 to l_N and $P(\mathbf{p} \notin \bar{\mathcal{V}} | \xi_j)$ can be written as:

$$\begin{aligned} P(\mathbf{p} \notin \bar{\mathcal{V}} | \xi_j) &= P\left(\bigcup_{i=1}^N \{\mathbf{a}_i^T \mathbf{p} > b_i\} | \xi_j\right) \\ &\leq \sum_{i=1}^N P(\mathbf{a}_i^T \mathbf{p} > b_i | \xi_j) =: \epsilon_j. \end{aligned} \quad (9)$$

By combining (8) and (9), the upper bound ϵ on the tracking failure probability is:

$$P(\mathbf{p} \notin \mathcal{V}) \leq \sum_{j=1}^{M_k} P(\xi_j) \epsilon_j =: \epsilon. \quad (10)$$

The margin of the tracking failure probability is denoted by $\Delta\epsilon$. Since the projection of $\mathbf{p}$ given ξ_j on $\mathbf{a}_i$ follows the Gaussian distribution with mean $\mathbf{a}_i^T \hat{\mathbf{p}}_j$ and variance $\sigma_{ij}^2 = \mathbf{a}_i^T \Sigma_j \mathbf{a}_i$, the tracking failure probability given ξ_j of the constraint l_i is as follows:

$$P(\mathbf{a}_i^T \mathbf{p} > b_i | \xi_j) = 1 - \Phi\left(\frac{\mathbf{a}_i^T \mathbf{f} \mathbf{r} u_v - \bar{c}_{ij}}{\sigma_{ij}}\right),$$

where $\mathbf{f} \mathbf{r}$ denotes $\mathbf{f} \mathbf{r}(\mathbf{u}_w)$ and $\bar{c}_{ij} =$

$$\begin{cases} \mathbf{a}_i^T \hat{\mathbf{p}}_j & \text{for } i = 1, 2 \\ \mathbf{a}_i^T \hat{\mathbf{p}}_j - R_s \cos(\theta_s/2(N-2)) & \text{for } i = 3, \dots, N \end{cases}$$

²It is assumed that all distributions below are conditional on measurements Z^k .

and Φ is the cumulative distribution function of a standard normal random variable (see Theorem 1 in [12] for the proof). Then the upper bound ϵ can be computed as:

$$\epsilon = \sum_{j=1}^{M_k} \sum_{i=1}^N P(\xi_j) \left(1 - \Phi\left(\frac{\mathbf{a}_i^T \mathbf{f} \mathbf{r} u_v - \bar{c}_{ij}}{\sigma_{ij}}\right)\right). \quad (11)$$

V. MOTION STRATEGIES

We propose a single-step motion strategy which minimizes the upper bound on the tracking failure probability given in (11). Since the objective function is simpler as a function of u_v than as a function of u_w , we solve the problem for a fixed value of u_w . Then, the desired control is a solution of the following optimization problem:

$$\begin{aligned} \Pi_1 : \quad & \min_{u_v} \quad \epsilon \\ \text{subject to} \quad & 0 \leq \epsilon \leq 1, \\ & V_{\min} \leq u_v \leq V_{\max} \\ & [\mathbf{s} \ \varphi_s]^T = f(\mathbf{s}_0, \varphi_0, u_v). \end{aligned}$$

We assume that if the optimized ϵ is less than a threshold E_1 , the tracking failure probability is sufficiently good. For such case, we minimize the moving distance under the constraint $\epsilon \leq E_1$, i.e., we solve the following problem:

$$\begin{aligned} \Pi_2 : \quad & \min_{u_v} \quad \frac{1}{2} u_v^2 \\ \text{subject to} \quad & 0 \leq \epsilon \leq E_1, \\ & V_{\min} \leq u_v \leq V_{\max} \\ & [\mathbf{s} \ \varphi_s]^T = f(\mathbf{s}_0, \varphi_0, u_v). \end{aligned} \quad (12)$$

This procedure is repeated for each $u_w \in \mathcal{A}$, where $\mathcal{A}$ is a set of candidate angular velocities, to find the optimal control which minimizes ϵ and $d(\mathbf{u})$.

A normalized gradient descent method [18] is used to solve Π_1 . The step size is searched by the backtracking line search for fast convergence. These procedures are repeated until ϵ converges to the optimal ϵ^* or $\epsilon \leq E_1$. If ϵ is less than E_1 , we solve the problem Π_2 . Otherwise, we do not solve Π_2 and return the optimal control u_v^* as a solution. Since the solution can be a local optimal point, the initial value should be carefully selected. A set of candidates for an initial value can be chosen from $[V_{\min}, V_{\max}]$. It is denoted by $\Upsilon = \{v_1, \dots, v_{|\Upsilon|}\}$. For each value from Υ , we compute $\epsilon(v_i)$ using (11) and the value with the lowest $\epsilon(v_i)$ is chosen as the initial value for Π_1 .

The objective function of Π_2 is convex and $u_v = 0$ is the optimal solution if constraints are satisfied. Otherwise, we use the gradient descent method in order to find a feasible solution.

VI. EXPERIMENTS

We have performed multi-step target tracking simulations and pedestrian tracking experiments. Multi-step simulations are designed to analyze the overall framework, which consists of motion prediction and tracking. In experiments, a mobile sensor is a mobile robot equipped with a Kinect sensor. The parameters for a sensor are set to $\theta_s = 57^\circ$

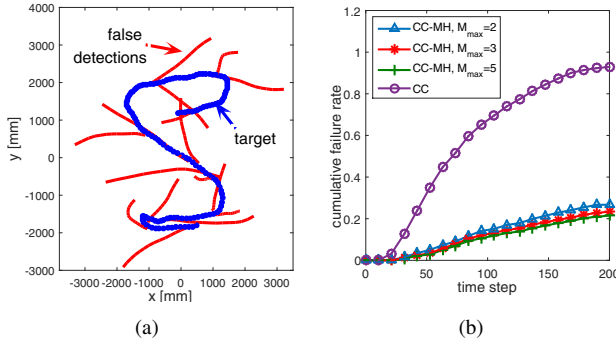


Fig. 3. (a) Simulation settings. Part of trajectories of a target and false detection are plotted. The target is moving along the blue line and objects causing false detection are moving along red lines. (b) Cumulative histograms of tracking failures from simulation using human trajectories with two tracking algorithms, CC-MH and CC. Histograms are plotted along the time step of a trajectory. There are 20 moving objects.

and $R_s = 3500\text{ mm}$ in simulations and $\theta_s = 50^\circ$ and $R_s = 3000\text{ mm}$ in field experiments as the Xbox software specifies.

We approximate the sensing region as a four-sided polygon which has the smallest margin. The computing time of our algorithm is quite small. The threshold E_1 is varying from 10^{-6} to 10^{-1} . The average time required to solve Π_1 and (optionally) Π_2 is 8.673 ms for the case with $N = 4$ and $M = 3$ (in MATLAB). If we increase the number of mixture components to 10, it only takes 27.58 ms on average to solve Π_1 and Π_2 .

A. Simulation: Multi-Step Target Tracking

An example of scenarios used for multi-step target tracking simulation in crowded environments is shown in Figure 3(a). A target moves along real human trajectories collected using the Vicon motion capture system³. There are 86 trajectories, each with a length of 200 steps. A crowded environment is simulated using moving objects which has a linear dynamic model with a constant acceleration. Trajectories of objects are randomly generated and measured, inducing false detections. Since the appearance of objects cannot be measured in simulation, the distance d_{sim} of the target and false detections are generated randomly. Ten scenarios are generated for each trajectory of the target resulting 860 scenarios. The admissible range of control is set to $V_{\min} = -700\text{ mm/s}$, $V_{\max} = 700\text{ mm/s}$, $W_{\min} = -140^\circ/\text{s}$, and $W_{\max} = 140^\circ/\text{s}$, according to the configuration of a Pioneer 3-AT mobile robot.

In Figure 3(b), our method is validated by comparing to [12], which tracks a target using chance-constrained optimization without considering false detections. The method from [12] is labeled as CC tracking and the method proposed in this paper is called as chance constrained target tracking with multiple hypotheses or CC-MH tracking. Cumulative histograms of tracking failures with respect to time steps are plotted in Figure 3(b). CC tracking selects the most probable measurement, i.e., a nearest-neighbor filter, and predicts the

³Vicon MX motion capture system. Available at <http://www.vicon.com/>.

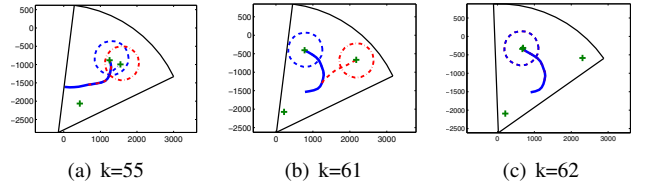


Fig. 4. Target tracking with a multi-hypothesis prediction algorithm. The black line represents the sensing region of the sensor. The blue line is the trajectory of the target and green plus markers are measurements. The blue dashed line shows prediction from the trajectory of a target. Red dashed line shows prediction from the false detection. Dashed circles show covariance of the estimated position.



Fig. 5. Experimental setup. (a) The trajectory of a target is shown in a blue dashed line. Other pedestrians (red dashed boxes) generate false detection. (b) A TV, chairs, and sitting persons in red boxes generate static false detections.

target's next position. The tracking failure probability is much lower with CC-MH tracking than with CC tracking as false detections are properly handled by multiple-hypothesis prediction. We can also observe that the performance improves as more hypotheses are allowed.

Figure 4 shows the process of tracking. At time $k = 55$, two hypotheses are generated by two measurements (Figure 4(a)). The mobile sensor does not make a decision which measurement is true and tracks both predictions instead (Figure 4(a) and 4(b)). At time $k = 62$, the sensor tracks only the true target because the mixture component weight of a false detection has decayed (Figure 4(c)). We have also analyzed the relationship between E_1 and the moving distance. There is a trade-off between the moving distance and the tracking failure rate. The property is similar to our prior findings in [12].

B. Pedestrian Tracking Experiments

We used a Pioneer 3-AT and a Pioneer 3-DX mobile robot equipped with a Microsoft Kinect camera. We ran our algorithm at 5Hz. The position of a person is detected using the skeleton grab API of Xbox software. Note that a Kinect camera can detect up to five humans.

We performed experiments similar to the multi-step simulation in Section VI-A. We marked waypoints on the floor as shown in Figure 5(a). A target moves along markers and five other pedestrians move near the target. Trajectories of other pedestrians are pre-defined and we have repeated the scenario 10 times using two tracking algorithms: the proposed CC-MH tracking algorithm and the CC tracking algorithm from [12]. The maximum number of hypotheses is set to $M_{\max} = 3$ and $E_1 = 5 \times 10^{-5}$. As shown in Figure 6, CC-MH tracking

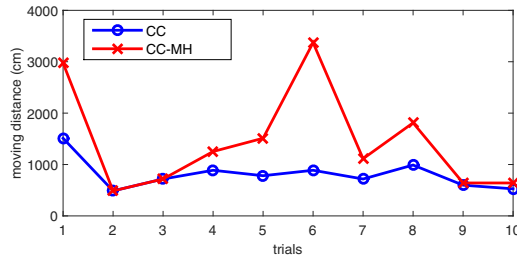


Fig. 6. Moving distances of a target for ten trials. The target is moved until tracking is failed or the target is arrived at the final position.

TABLE I
TRACKING SUCCESS RATES

Number of moving persons near the target	0	1	2	3
CC-MH	100%	90%	95%	90%
CC	95%	65%	55%	75%

tracks the target for longer distances. The average tracking distance of the proposed method was 14.520 m while the average tracking distance for CC tracking was 8.107 m . The benefit of systematically handling false detections can be seen from experiments. Results from this and other field experiments are included in our video submission.

We have also conducted experiments with less crowded environments and shorter trajectories like Figure 5(b). It is to show tracking success rates of our method. In this environment, a number of false detections are reported from sitting persons, chairs, and a TV. When the target is the only object moving, the tracking success rates of CC-MH tracking and CC tracking are 100% and 95%, respectively, from 20 trials. Since static false detections are distant from the target, both tracking algorithms show good performance. However, when there are moving pedestrians near the target, the success rate of CC tracking drops to 55% while that of CC-MH tracking stays at above 90%. The tracking success rates at different numbers of moving pedestrians are shown in Table I. The result shows that the proposed method is extremely robust in crowded environments and false detections have to be treated properly.

VII. CONCLUSION

In this paper, we have presented a target tracking algorithm which can handle identity uncertainty occurring in crowded environments for a mobile sensor with a fan-shaped field of view and finite sensing region. Multiple detections about the target are properly treated with the proposed multiple-hypothesis prediction algorithm and the distribution of the future position of a target is approximated by a Gaussian mixture model. The proposed method can minimize the upper bound on the tracking failure probability. In addition, the proposed method minimizes the moving distance of the robot under the guaranteed tracking failure probability. While the sensing region is nonlinear and complex, we have derived an optimization scheme, which can be solved in real-time. A simulation study shows that the proposed tracking

algorithm is robust in crowded environments. The method is also validated in physical environments using a Pioneer with a Kinect sensor to track a moving pedestrian in crowded environments.

REFERENCES

- [1] P. Trahanias, W. Burgard, A. Argyros, D. Hahnel, H. Baltzakis, P. Pfaff, and C. Stachniss, "Tourbot and webfair: Web-operated mobile robots for tele-presence in populated exhibitions," *Robotics Automation Magazine*, IEEE, June 2005.
- [2] F. Belkhouche, B. Belkhouche, and P. Rastgoufard, "Parallel navigation for reaching a moving goal by a mobile robot," *Robotica*, vol. 25, 1 2007.
- [3] E. Masehian and Y. Katebi, "Sensor-based motion planning of wheeled mobile robots in unknown dynamic environments," *Journal of Intelligent and Robotic Systems*, vol. 74, no. 3-4, 2014.
- [4] Y. Kanayama, Y. Kimura, F. Miyazaki, and T. Noguchi, "A stable tracking control method for a non-holonomic mobile robot," in *IEEE/RSJ International Workshop on Intelligent Robots and Systems*, Nov 1991.
- [5] T. Bandyopadhyay, Y. Li, J. Ang, M.H., and D. Hsu, "A greedy strategy for tracking a locally predictable target among obstacles," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2006.
- [6] T. Muppilala, S. Hutchinson, and R. Murrieta-Cid, "Optimal motion strategies based on critical events to maintain visibility of a moving target," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, April 2005.
- [7] C.-Y. Lee, H. Gonzalez-Banos, and J.-C. Latombe, "Real-time tracking of an unpredictable target amidst unknown obstacles," in *7th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Dec 2002.
- [8] H. Gonzalez-Banos, C.-Y. Lee, and J.-C. Latombe, "Real-time combinatorial tracking of a target moving unpredictably among obstacles," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2002.
- [9] T. Bandyopadhyay, N. Rong, M. Ang, D. Hsu, and W. S. Lee, "Motion planning for people tracking in uncertain and dynamic environments," *Workshop on people detection and tracking, IEEE International Conference on Robotics and Automation (ICRA)*, 2009.
- [10] E. W. Frew and S. M. Rock, "Trajectory generation for constant velocity target motion estimation using monocular vision," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, Sept 2003.
- [11] K. Zhou and S. Roumeliotis, "Multirobot active target tracking with combinations of relative observations," *IEEE Transactions on Robotics*, vol. 27, no. 4, Aug 2011.
- [12] Y. Oh, S. Choi, and S. Oh, "Chance-constrained target tracking for mobile robots," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2015.
- [13] S. LaValle, H. Gonzalez-Banos, C. Becker, and J.-C. Latombe, "Motion strategies for maintaining visibility of a moving target," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, Apr 1997.
- [14] I. Cox, "A review of statistical data association techniques for motion correspondence," *International Journal of Computer Vision*, vol. 10, no. 1, pp. 53–66, 1993.
- [15] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia, "People tracking and following with mobile robot using an omnidirectional camera and a laser," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, May 2006, pp. 557–562.
- [16] S. Choi, E. Kim, and S. Oh, "Real-time navigation in crowded dynamic environments using gaussian process motion control," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2014.
- [17] L. Blackmore and M. Ono, "Convex chance constrained predictive control without sampling," in *Proc. of the AIAA Guidance, Navigation and Control Conference*, 2009.
- [18] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.