International Journal of Control, Automation, and Systems (2015) 13(3):530-538 ISSN:1598-6446 eISSN:2005-4092
DOI 10.1007/s12555-014-0192-3 http://www.springer.com/12555

Visual Odometry based on a Bernoulli Filter

Feihu Zhang*, Daniel Clarke, and Alois Knoll

Abstract: In this paper, we propose a Bernoulli filter for estimating a vehicle's trajectory under random finite set (RFS) framework. In contrast to other approaches, ego-motion vector is considered as the state of an extended target while the features are considered as multiple measurements that originated from the target. The Bernoulli filter estimates the state of the extended target instead of tracking individual features, which presents a recursive filtering framework in the presence of high association uncertainty. Experimental results illustrate that the proposed approach exhibits good robustness under real traffic scenarios.

Keywords: Bernoulli filter, ego-motion vector, random finite set.

1. INTRODUCTION

Using cameras for vehicle navigation is the current trend for intelligent vehicles. The basic idea is to find associated features and calculate the displacement in consecutive frames. Much work has been done utilizing Structure-from-Motion (SfM) technique [1]. It refers to the process of estimating three dimensional information from two dimensional images. Stereo camera provides high qualities in 3D construction to calculate the camera motion. Furthermore, RANdom SAmple Consensus (RANSAC) enables the SfM to overcome a large number of outliers [2]. However, there are still open issues:

- Features which are utilized to estimate the ego- motion vector may contain falsely associated pairs between consecutive frames. Robust matching techniques are required to avoid false matching.
- Unevenly distributed features which are aggregated in a small region may influence the performance since they are not uniformly distributed throughout the whole space. Effective extraction techniques are required to overcome this challenge.
- Algorithms for visual odometry are typically based on features from stationary objects. However, typical road scenes may contain a large number of features stemming from moving objects.
- Features are often considered as individual targets, such that an ego-motion vector is acquired by calculating the average states of the group targets. However, as time passes, the number of targets may become huge.

Various algorithms have been investigated for eliminating influences from the issues. One approach to visual odometry applies the Structure-from-Motion technique [3-6]. The idea is to find high quality features between consecutive frames and estimate the corresponding displacements. SfM technique reduces the complexity of dealing with the whole image by solely relying on features, which makes the computation realistic [7].

The dense motion algorithm, also known as optical flow, has been studied which focuses on changing in brightness regions [8,9]. The computed flow fields are typically useful for obstacle avoidance, however, tough for the global geometry. Although optical flow is computationally cheaper, the precision is not guaranteed.

Corke *et al*. compared the above approaches with the conclusion that Structure-from-Motion allows higher precision on computational requirements opposed to the optical flow [10].

Probability Hypothesis Density (PHD) filter for visual odometry is first proposed in our previous work [11,12]. The random finite set (RFS) paradigm is a mathematically principled and elegant approach to multi-target filtering which has already attracted considerable attentions in recent years, whereas the PHD filter is a predict and correct framework for recursive Bayesian filtering in such RFS formulation [13-15]. Features are treated as set-valued observations as random finite set allows solving the problem of dynamically estimating multiple-targets in the presence of clutter and association uncertainty. The overall group state, also known as the ego-motion vector, is acquired by calculating the average state in the states set.

Although PHD implementation provides a high precision localization, another important but lesser known Bayes filter is the Bernoulli filter [16]. In this paper, we expand the previous work by investigating the visual odometry based on a Bernoulli filter. Unlike PHD recursion which propagates moments distribution, the Bernoulli filter propagates Bernoulli distribution which approximates the posterior density. Implementation of

Manuscript received May 19, 2014; revised September 2, 2014; accepted October 12, 2014. Recommended by Associate Editor Hongbo Li under the direction of Editor Fuchun Sun.

Feihu Zhang and Alois Knoll are with the institute of Robotics and Embedded Systems, Technische Universität München, 85748 Garching bei München, Germany (e-mails: feihu.zhang@tum.de, knoll@in.tum.de).

Daniel Clarke is with the Cranfield University, Defense Academy of the United Kindom, Shrivenham SN6 8LA, United Kingdom (e-mail: d.s.clarke@cranfield.ac.uk).

* Corresponding author.

the Bernoulli filter is involved a few approximations: the Sequential Monte Carlo approximation [17] is proposed for the estimation of the posterior spatial probability density function (PDF). In this paper, ego-motion vector is considered as the state of an extended target which generates multiple measurements (features) per frame. The Bernoulli filter is utilized to estimate the corresponding state instead of tracking individual features.

The benefits of the proposed approach are concluded as following: Firstly, the matching process is eliminated since the Bernoulli filter avoids the data association issue. Modeling set-valued states and set-valued observations allows solving the problem of multiple targets tracking in the presence of association uncertainty. Second, ego-motion vector is solely estimated. In PHD implementation, ego-motion vector is calculated by averaging the corresponding states of the whole targets. The number of targets may become huge as time passes. Regarding to this paper, features are treated as the multiple measurements originated from the extended target. The overloaded phenomenon is therefore eliminated. Third, with the particle or Sequential Monte Carlo (SMC) implementation, the Bernoulli filter is advantageous in a non-linear environment whereas the PHD filter requires an additional clustering step.

An off-the-shelf platform provides data under real traffic scenarios for evaluating the proposed approach [18]. Experimental results exhibit the performance in a large scale urban environment.

The structure of this paper is organized as follows: Section 2. describes details about the preprocessing phase. Section 3. introduces the mathematic background of the Bernoulli filter. Section 4. presents experimental results. Finally, the paper is concluded in Section 5.

2. PREPROCESSING PHASE

2.1. Feature extraction

For SfM based visual odometry techniques, features are extracted under changes in lighting and viewpoint, also fast to detect. Various features are investigated based on their properties, e. g. Harris corner feature [19], SIFT feature [20], SURF feature and Kanade-Lucas-Tomasi feature [21,22].

In this paper, SURF feature descriptor is utilized due to its computation properties under real time requirements. Features are considered as the multiple measurements which originated from the extended target in vehicle coordinates.

2.2. Transformation between image coordinates and vehicle coordinates

In this step, we determine the mapping between the image coordinates $[u,v]^T$ of a tracked pixel and the vehicle coordinates $[x,y,z]^T$ of the corresponding point. First, we introduce the used coordinate systems (see Fig. 1). The vehicle coordinates $[x,y,z]^T$ are defined in a three-dimensional Cartesian coordinate system with origin in the middle of the rear axle; the camera coordinate $[x_c,y_c,z_c]^T$ is described in a three-dimensional Cartesian coordinate system with origin in the optical center of the camera, and the image coordinates $[u,v]^T$ are defined in a two-dimensional Cartesian coordinate system with origin in the upper left corner of the image. For an easy implementation of the transformation, we will use homogeneous coordinates hereafter.

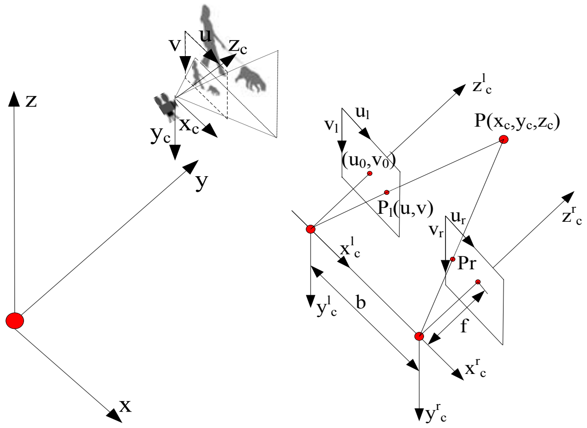

Fig. 1. Different coordinates systems.

2.2.1 Transformation from vehicle coordinates to camera coordinates

Considering the object in the image as a rigid body, the transformation from the vehicle coordinates to camera coordinates is represented by six independent extrinsic parameters. These are the three translation parameters within the translation vector $T=[x_t,y_t,z_t]^T$, and the three rotation parameters within the rotation matrix $R=R_xR_yR_z$. R is the rotation matrix corresponding to the product of rotations around the z-axis (rotation matrix R_z), y-axis (rotation matrix R_y) and x-axis (rotation matrix R_x). In summary, the transformation between vehicle and camera coordinates is

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{1}$$

2.2.2 Transformation from camera coordinates to image coordinates

For stereo camera system, we have two camera coordinates: left camera coordinates $[x^l_c,y^l_c,z^l_c]^T$ and right camera coordinates $[x^r_c,y^r_c,z^r_c]^T$ (see Fig. 1). For an easy implementation of the transformation we will use the camera coordinate $[x_c,y_c,z_c]^T$ instead of the left camera coordinates $[x^l_c,y^l_c,z^l_c]^T$. We calculate the 3D coordinates of the pixel from both stereo images in the camera coordinates (left camera coordinates). Three-dimensional objects of a scene are projected onto the two-dimensional surface of the camera sensor. From a standardized projection of a point $P=[x_c,y_c,z_c]^T$ in vehicle coordinates system based on pinhole model, we can get its image coordinates $[u,v]^T$.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 + \dfrac{fx_c}{z_c} \\ v_0 + \dfrac{fy_c}{z_c} \end{bmatrix}, \tag{2}$$

where $[u_0, v_0]$ is the point of camera optical axis and image plane intersection, f is the camera focal length. This formula can be expressed in homogeneous coordinates as

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \tag{3}$$

From (1) and (3), we can express the relationship between the vehicle coordinates $[x, y, z]^T$ of a point in the space to its image coordinates $[u, v]^T$ as follows:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{4}$$

Rearranging the above equation, we can transform the pixels from image coordinates to vehicle coordinates.

3. TRACKING PHASE

3.1. Standard Bayes filter

The standard formulation of the Bayes filtering framework is described by two equations:

$$\begin{aligned} x_k &= F(x_{k-1}, w_{k-1}), \\ z_k &= H(x_k, v_k) \end{aligned} \tag{5}$$

referred to as the motion process and the measurement process, respectively. *F* is a nonlinear transition function defining the evolution of the state as a Markov process. *H* defines the relationship between the state and the measurement. w_k, v_k are the process and observation noises with covariance Q_k and R_k.

3.2. Overview on RFS

The RFS is a hidden Markov chain model with set-valued states and set-valued observations. The RFS approach to multiple-target tracking is an emerging and promising alternative to the traditional association methods. It is an approximation to alleviate the computational intractability of the optimal multi-target Bayes filter, proposed by Mahler [23].

A RFS X is specified by its cardinality distribution $\rho(n) = P\{|\mathrm{X}| = n\}$ with a joint symmetric distributions $\rho_n = (x_1, \ldots, x_n)$ where $n \in \mathbb{N}, \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$, Furthermore, as a finite-set valued random variable, the PDF and moments of the RFS X are also defined. Regarding to Mahler's finite set statistics (FISST) theory, the PDF of a RFS X is denoted by $f(\mathrm{X})$ as following:

$$f(\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}) = n! \cdot \rho(n) \cdot p_n(\mathbf{x}_1, \cdots, \mathbf{x}_n), \tag{6}$$

while the set integral is defined as

$$\int f(\mathbf{X}) \delta \mathbf{X} = f(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} f(\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}) d\mathbf{x}_1 \ldots d\mathbf{x}_n, \tag{7}$$

where the integration of $f(\mathrm{X})$ is equal to 1. Furthermore, the following RFS components are also related to the proposed approach.

- Bernoulli RFS

For Bernoulli RFS, $\rho(n)$ is considered as a Bernoulli distribution which is either empty (with probability $1-q$) or have one element (with probability q), distributed over the whole state space $\mathcal{X}$ based on PDF $p(\mathbf{x})$. The PDF is as following:

$$f(\mathbf{X}) = \begin{cases} 1-q, & \text{if} \quad \mathbf{X} = \varnothing \\ q \cdot p(\mathbf{x}) & \text{if} \quad \mathbf{X} = \{\mathbf{x}\}. \end{cases} \tag{8}$$

- IID Cluster RFS

With known cardinality $|\mathbf{X}|$, the RFS **X** is consisted of IID random variables. The PDF is defined as:

$$f(\mathbf{X}) = |\mathbf{X}|! \cdot \rho(|\mathbf{X}|) \cdot \prod_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x}). \tag{9}$$

- Poisson RFS

A Poisson RFS is a special case of the IID cluster RFS whose PDF is defined as:

$$f(\mathbf{X}) = e^{-\lambda} \prod_{\mathbf{x} \in \mathbf{X}} \lambda p(\mathbf{x}). \tag{10}$$

- Binomial RFS

The cardinality distribution of RFS **X** is considered as binomial distribution with parameter *L* (binary experiments total numbers) and p_0 (each experiment success probability), while the PDF is defined as following:

$$f(\mathbf{X}) = \frac{L!}{(L - |\mathbf{X}|)!} p_0^{|\mathbf{X}|} (1-p_0)^{L-|\mathbf{X}|} \prod_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x}). \tag{11}$$

3.3. Mathematical background of Bernoulli filter

The targets in a multi-target scenario at time *k* are represented as a finite set of vectors $\mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,n(k)}$ which take values from the state space $\mathcal{X} \in \mathbb{R}^{n_\mathbf{x}}$. Under RFS framework, not only the number of targets n_k but also the individual states are random and time-varying. Similarly, the observations are represented as a finite set of vectors $\mathbf{z}_{k,1}, \ldots, \mathbf{z}_{k,m(k)}$ which take values from the observation space $\mathcal{Z} \in \mathbb{R}^{n_\mathbf{z}}$. These finite sets are known as the multi-targets state and observation:

$$\mathbf{X}_k = \{\mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,n(k)}\} \in \mathcal{F}(\mathcal{X}), \tag{12}$$

$$\mathbf{Z}_k = \{\mathbf{z}_{k,1}, \ldots, \mathbf{z}_{k,m(k)}\} \in \mathcal{F}(\mathcal{Z}), \tag{13}$$

where $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Z})$ denote the sets of all finite subsets $\mathcal{X}$ and $\mathcal{Z}$, respectively. Assuming the target state is a Markov process with transitional density $\phi_{k|k-1}(\mathbf{X}_k \mid \mathbf{X}_{k-1})$ whereas the probability density $\varphi_k(\mathbf{Z}_k$

$|\mathbf{X}_k)$ denotes the likelihood function, the stochastic filtering problem is elegantly addressed under the RFS framework.

Suppose at time $k-1$ the posterior PDF of target state $f_{k-1|k-1}(\mathbf{X}_{k-1}|\mathbf{Z}_{1:k-1})$ is known, the predicted and updated posterior densities are expressed as following:

$$f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{1:k-1}) = \int \phi_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1}) \cdot f_{k-1|k-1}(\mathbf{X}_{k-1}|\mathbf{Z}_{1:k-1})\delta\mathbf{X}_{k-1}, \tag{14}$$

$$f_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k}) = \frac{\varphi_k(\mathbf{Z}_k|\mathbf{X}_k) f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{1:k-1})}{\int \varphi_k(\mathbf{Z}_k|\mathbf{X}) f_{k|k-1}(\mathbf{X}|\mathbf{Z}_{1:k-1})\delta\mathbf{X}}. \tag{15}$$

Integrals in (14) and (15) are set integrals and the expressions for transitional density and likelihood function are also set integrals. In a special case for Bernoulli RFS by assuming $\mathbf{X}_k$ is a singleton $(f(\mathbf{X})=0,$ if $|\mathbf{X}|>1)$, the set integral simplifies to:

$$\int f(\mathbf{X})\delta\mathbf{X} = f(\varnothing) + \int f(\{\mathbf{x}\})d\mathbf{x} = 1-q+q\int p(\mathbf{x})d\mathbf{x} = 1. \tag{16}$$

The binary random variable $\epsilon_k \in \{0,1\}$ is introduced for modeling target appearance and disappearance during the whole period, where $\epsilon_k = 1$ illustrates that the target existing on time *k*. Dynamics of ϵ_k is modeled by the first order Markov chain with a transitional probability matrix Ξ, which is defined as $\Xi_{ij} = P\{\epsilon_k = j-1 | \epsilon_{k-1} = i-1\}$ for $i,j \in \{1,2\}$. The matrix is structured as following:

$$\Xi = \begin{bmatrix} 1-p_b & p_b \\ 1-p_s & p_s \end{bmatrix},$$

where $p_b := P\{\epsilon_{k+1} = 1 | \epsilon_k = 0\}$ is the probability of target birth whereas $p_s := P\{\epsilon_{k+1} = 1|\epsilon_k = 1\}$ is the probability of target survival. The PDF $b_{k|k-1}(\mathbf{x})$ also denotes the birth density once the target appears during the time interval $t_k - t_{k-1}$. Thus the Bernoulli RFS $\mathbf{X}_k$ is described by PDF $\phi_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$ as following:

$$\phi_{k|k-1}(\mathbf{X}|\varnothing) = \begin{cases} 1-p_b, & \text{if } \mathbf{X} = \varnothing \\ p_b \cdot b_{k|k-1}(\mathbf{x}) & \text{if } \mathbf{X} = \{\mathbf{x}\}, \end{cases}$$

$$\phi_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1}) = \begin{cases} 1-p_s, & \text{if } \mathbf{X}_k = \varnothing \\ p_s \cdot \pi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) & \text{if } \mathbf{X}_k = \{\mathbf{x}\}. \end{cases} \tag{17}$$

The set observation $\mathbf{Z}_k$ is considered as the union of two random finite sets:

$$\mathbf{Z}_k = \mathbf{C}_k \cup \mathbf{W}_k, \tag{18}$$

where $\mathbf{W}_k$ represents the measurements which originated from the targets. $\mathbf{C}_k$ represents the set of false detections, modeled by a Poisson distribution:

$$Pr\{|\mathbf{C}_k| = v\} = \frac{e^{-\lambda}\lambda^v}{v!}, \qquad v = 0,1,2,\cdots,$$

$$\kappa(\mathbf{C}_k) = e^{-\lambda} \prod_{\mathbf{z}\in\mathbf{C}_k} \lambda c(\mathbf{z}), \tag{19}$$

where false detections are considered as IID random values with PDF $c(\mathbf{z})$, intensity $\kappa(\cdot)$.

$\mathbf{W}_k$ consists of multiple measurements with detected probability p_d. Assuming L_k generating points at time *k*, model the RFS $\mathbf{W}_k = \{\mathbf{w}_{k,1},\cdots,\mathbf{w}_{k,L_k}\}$ as a binomial RFS which treats the individual points tracking issue to the extended target tracking. The PDF of RFS $\mathbf{W}_k$ is given by:

$$\eta(\mathbf{W}_k|\{\mathbf{x}\}) = \frac{L_k!}{(L_k - |\mathbf{W}_k|)!} p_d^{|\mathbf{W}_k|}(1-p_d)^{L_k - |\mathbf{W}_k|} \bullet \prod_{\mathbf{w}\in\mathbf{W}_k} g_k(\mathbf{w}|\mathbf{x}).$$

Similar to (17), the likelihood function for $\mathbf{Z}_k$ is represented as empty and existing as:

$$\varphi(\mathbf{Z}_k|\varnothing) = \kappa(\mathbf{Z}_k), \tag{20}$$

$$\varphi(\mathbf{Z}_k|\mathbf{x}) = \kappa(\mathbf{Z}_k)\{(1-p_d)^{L_k} + \sum_{\Omega\in\mathcal{P}_{1:L_k(\mathbf{Z}_k)}} \frac{L_k!}{(L_k - |\Omega|!)} \cdot p_d^{|\Omega|}(1-p_d)^{L_k-|\Omega|} \prod_{\mathbf{z}\in\Omega} \frac{g(\mathbf{z}|\mathbf{x})}{\lambda c(\mathbf{z})}\}, \tag{21}$$

where $\mathcal{P}_{1:L_k(\mathbf{Z}_k)}$ is the set of whole subsets of $\mathbf{Z}$ with cardinalities equal to $1,\cdots,L_k$.

According to (8), the Bernoulli RFS is therefore specified by:

$$f_{k|k}(\mathbf{X}_k|\mathbf{Z}_k) = \begin{cases} 1-q_{k|k}, & \text{if } \mathbf{X}_k = \varnothing \\ q_{k|k} \cdot s_{k|k}(\mathbf{x}) & \text{if } \mathbf{X}_k = \{\mathbf{x}\}, \end{cases}$$

where the posterior probability of target existence is considered as $q_{k|k} = P\{\epsilon_k = 1|\mathbf{Z}_{1:k}\}$; the posterior spatial PDF of target is considered as $s_{k|k}(\mathbf{x}) = p_k\{\mathbf{x}|\mathbf{Z}_{1:k}\}$.

According to (14), the predict probability of existence and the spatial PDF of the Bernoulli filter are as following:

$$q_{k|k-1} = p_b(1-q_{k-1|k-1}) + p_s q_{k-1|k-1} \tag{22}$$

$$s_{k|k-1}(\mathbf{x}) = \frac{p_b(1-q_{k-1|k-1})b_{k|k-1}(\mathbf{x})}{q_{k|k-1}} + \frac{p_s q_{k-1|k-1}\int \pi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})s_{k-1|k-1}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1}}{q_{k|k-1}}. \tag{23}$$

The above equations specify the prediction of the Bernoulli filter, whereas the update probabilities are based on (15) as following:

$$q_{k|k} = \frac{1-\Delta_k}{1-q_{k|k-1}\Delta_k} q_{k|k-1}, \tag{24}$$

$$s_{k|k}(\mathbf{x}) = \frac{(1-p_d)_k^L + \sum \frac{L_k! p_d^{|\Omega|}}{(L_k - |\Omega|)!(1-p_d)^{-L_k+|\Omega|}} \prod_{\mathbf{z}\in\Omega} \frac{g(\mathbf{z}|\mathbf{x})}{\lambda c(\mathbf{z})}}{1-\Delta_k} \cdot s_{k|k-1}(\mathbf{x}), \tag{25}$$

where

$$\Delta_k = 1-(1-p_d)^{L_k} - \sum \frac{L_k!\, p_d^{|\Omega|}}{(L_k-|\Omega|)!(1-p_d)^{-L_k+|\Omega|}} \cdot \frac{\int \prod_{\mathbf{z}\in\Omega} g(\mathbf{z}\mid\mathbf{x}) s_{k|k-1}(\mathbf{x})d\mathbf{x}}{\prod_{\mathbf{z}\in\Omega} \lambda c(\mathbf{z})}.$$

Using these random finite set models it is possible to construct extended target tracking analogous to the case of single-target tracking.

It is to be noted that neither the predict nor the update equations have closed form solutions. The Sequential Monte Carlo method is proposed to provide a generic implementation of the Bernoulli filter. More details of the SMC Bernoulli filter is given [16].

3.4. Implementation details

Under RFS framework, visual odometry is considered a special case for the Bernoulli filter by assuming $\mathbf{X}_k$ is a singleton ($f(\mathbf{X})=0$, if $|\mathbf{X}|>1$). Features are considered as multiple measurements that originated from the extended target ego-motion vector. Similar to the Bayes filter in (5), the Bernoulli filter also requires a motion process and a measurement process to represent the mapping between the state and the observation at each frame.

In this paper, the state is defined as:

$$\mathbf{x}_k = \left[\Delta x_k, \Delta y_k, \Delta\beta_k, \frac{\mathrm{d}\Delta x_k}{\mathrm{d}t}, \frac{\mathrm{d}\Delta y_k}{\mathrm{d}t}, \frac{\mathrm{d}\Delta\beta_k}{\mathrm{d}t} \right]^T, \tag{26}$$

where $(\Delta x_k, \Delta y_k, \Delta\beta_k)$ is defined as the ego-motion vector at time k, $(\frac{\mathrm{d}\Delta x_k}{\mathrm{d}t}, \frac{\mathrm{d}\Delta y_k}{\mathrm{d}t}, \frac{\mathrm{d}\Delta\beta_k}{\mathrm{d}t})$ is the corresponding velocity. Furthermore, $(\Delta x_k, \Delta y_k)$ is also considered as the translation movement while $\Delta\beta_k$ is the rotation change. Assuming the target has a constant velocity model, the state transition matrix $\mathbf{F}$ is defined:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{27}$$

The process noise is defined by:

$$Q_k = diag\left(\sigma^2_{\Delta x}, \sigma^2_{\Delta y}, \sigma^2_{\Delta\beta}, \sigma^2_{\frac{\mathrm{d}\Delta x_k}{\mathrm{d}t}}, \sigma^2_{\frac{\mathrm{d}\Delta y_k}{\mathrm{d}t}}, \sigma^2_{\frac{\mathrm{d}\Delta\beta_k}{\mathrm{d}t}} \right). \tag{28}$$

Regarding the measurement process, the equation is established based on Euler's rotation theorem as following:

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} \cos\Delta\beta_k & -\sin\Delta\beta_k \\ \sin\Delta\beta_k & \cos\Delta\beta_k \end{bmatrix} \begin{bmatrix} x_k - \Delta x_k \\ y_k - \Delta y_k \end{bmatrix}, \tag{29}$$

which maps the state to the associated features $(x_k, y_k$ is considered as the feature's position in vehicle coordinates at time k).

Algebraic manipulations on (29) obtains a pseudo measurement process as following:

$$\mathbf{0} = \begin{bmatrix} \cos\Delta\beta_k \cdot x_k - \Delta x_k \cdot \cos\Delta\beta_k - \sin\Delta\beta_k \cdot y_k + \sin\Delta\beta_k \cdot \Delta y_k \\ \sin\Delta\beta_k \cdot x_k - \Delta x_k \cdot \sin\Delta\beta_k + \cos\Delta\beta_k \cdot y_k - \cos\Delta\beta_k \cdot \Delta y_k \end{bmatrix} - \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \mathbf{H}(\mathbf{x}_k),$$

where the observation noise is described as:

$$R_k = diag([\sigma_x^2, \sigma_y^2]). \tag{30}$$

In this paper, measurements $(x_k, y_k, x_{k+1}, y_{k+1})$ are randomly generated by combining features pairs on consecutive frames.

The state

$$\mathbf{x}_k = \left[\Delta x_k, \Delta y_k, \Delta\beta_k, \frac{d\Delta x_k}{dt}, \frac{d\Delta y_k}{dt}, \frac{d\Delta\beta_k}{dt} \right]^T$$

is considered as the combination of the ego-motion vector and its velocity. The Bernoulli filter is utilized to track the object centroid instead of tracking individual scattering measurements. Visual odometry is therefore addressed by the Bernoulli filter as an extended target tracking problem as following:

Measurements are considered as originating from the extended target and clutter, while the clutter doesn't satisfy (29). The benefit of the Bernoulli filter is that it models the set-valued observations as RFS and allows the solution to the problem of dynamically estimating target in the presence of clutter and association in a Bayes filtering framework. Regarding to a standard Kalman filter, it needs a one-to-one relation between real measurements and real targets solved by data association. However, the Bernoulli filter overcomes this as it is an n-to-one mapping (here n includes the true measurements and clutter), which is a robust way for tracking an extended target. The clutter influence (falsely associated pairs) is eliminated according to posterior density under Bayes rules.

4. EXPERIMENTAL RESULTS

The proposed approach is evaluated by an off-the-shelf dataset from KITTI [18] which collects data from GPS, gyroscope and stereo camera at 10 frames/s with a resolution of 1344×391. All sequences correspond to the real traffic conditions in urban scenarios where the vehicle has the average speed of 40km/h. Features are detected by SURF descriptor while the tracking phase is processed in vehicle coordinates. A dead reckoning method is utilized to calculate the global trajectory at each frame [24].

Comparative results for visual odometry between the PHD filter and the standard SfM techniques (RANSAC based visual odometry) have already been investigated in our previous work [11,12]. In this paper, the focus is on the performance evaluation of the PHD filter and the Bernoulli filter. In the PHD implementation, a Gaussian Mixture PHD filter [14] is implemented to track individual features under linear Gaussian assumptions with a closed form solution. From the physical model in Euler's rotation theorem, individual features are considered with the same velocity-motion vector while the vehicle's ego-motion vector is therefore acquired by calculating the average velocities of each feature. Regarding the Bernoulli filter; a detector-output measurement model of Bernoulli particle filter is utilized (the pseudo code can be found in [16]). In Bernoulli implementation, the ego-motion vector is considered as an extended target while the features are considered as the multiple measurements originated from the extended target. Although the data association issue is avoided in both solutions, in contrast to the PHD filter, the Bernoulli filter calculates the ego-motion filter solely on the measurements.

Fig. 2 illustrates the tracking phase in vehicle coordinates. There is a small region contains lots of features in Fig. 2, which is a challenge for RANSAC based visual odometry since features are not uniformly distributed throughout the whole space. Furthermore, those features are originated from moving objects that influence the estimation precision. However, under the RFS framework, the above issues are addressed by both PHD and Bernoulli filters. As illustrated in Fig. 2, the final estimations from the PHD filter are uniformly distributed and each state contributes the same importance to calculate the ego-motion vector. Furthermore, since the PHD filter propagates the posterior intensity according to the dynamic system model, the estimation still keeps high precision although those features are stemming from moving objects. Finally, the ego-motion vector is acquired by calculating the average velocities of each state. With respect to the Bernoulli filter, the goal is to track the features' centroid (ego-motion vector) instead of tracking individual features. Therefore, there is no Bernoulli representation in Fig. 2. As an extended target, features from consecutive two frames are considered as the measurements to directly estimate the ego-motion vector. Influences from moving objects are eliminated based on the posterior spatial probability density function.

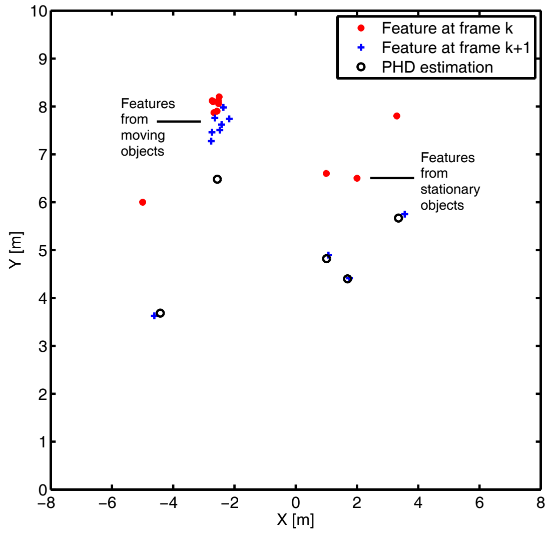

Fig. 2. Features in vehicle coordinates.

A number of ten scenarios are presented to illustrate the potential of the Bernoulli filter for visual odometry. In Fig. 3, the blue line denotes the true trajectories while the red and green lines denote the trajectories acquired by the Bernoulli filter and the PHD filter, respectively. It can be observed that both the PHD and the Bernoulli filters are close to the true trajectories, although the measurements set may contain the false associated features.

Table 1 summarizes the results of the corresponding scenarios. The index illustrates the experiments in Fig. 3 and the distance is the length of the experiment. Odometry error means the distance between the estimated location and the true location in the end. As we can see, the Bernoulli filter provides a more precise result in Figs. 3(a), (d), (e), (i) and Fig. 3(j), an almost equal result in Figs. 3(b), (f), (g) and Fig. 3(h), a worse result in Fig. 3(c). The reason why Fig. 3(c) has a huge bias is still being investigated. However, we can achieve the conclusion that the Bernoulli filter might be more suitable for visual odometry than the PHD filter.

The contributions of utilizing Bernoulli filter for visual odometry are concluded as following:

First, the matching process is avoided since the whole features are considered as the multiple measurements originated from the extended target. For a standard visual odometry algorithm, RANSAC has been widely utilized for motion estimation in the presence of outliers. It achieves its goal by iteratively selecting a random subset of the original data and the matching process is therefore required. However, the Bernoulli filter avoids the matching process by modeling a set-valued state and observation to consider the visual odometry as an extended target tracking problem. Furthermore, an n-to-one mapping relationship between measurements and target is proposed which exhibits high performance in visual odometry. Clutter influence is eliminated according to the posterior density calculated by the Bernoulli filter.

Table 1. Performance of the algorithms.

Index	Distance	Frames	Bernoulli Odometry Error	PHD Odometry Error
a	429m	1423	5.1m(1.2%)	10.3m(2.4%)
b	582m	354	8.1m(1.4%)	8.3m(1.42%)
c	297m	818	7.14m(2.4%)	3.7m(1.24%)
d	233m	966	2.3m(0.99%)	5.6m(2.4%)
e	560m	1248	5.6m(1%)	8.3m(1.5%)
f	400m	601	3.7m(0.92%)	3.61m(0.9%)
g	260m	447	0.5m(0.19%)	0.75m(0.28%)
h	92m	111	0.753m(0.82%)	0.95m(1.1%)
i	196m	341	1.8m(0.91%)	4.5m(2.3%)
j	490m	1401	7.28m(1.49%)	8.1m(1.4%)

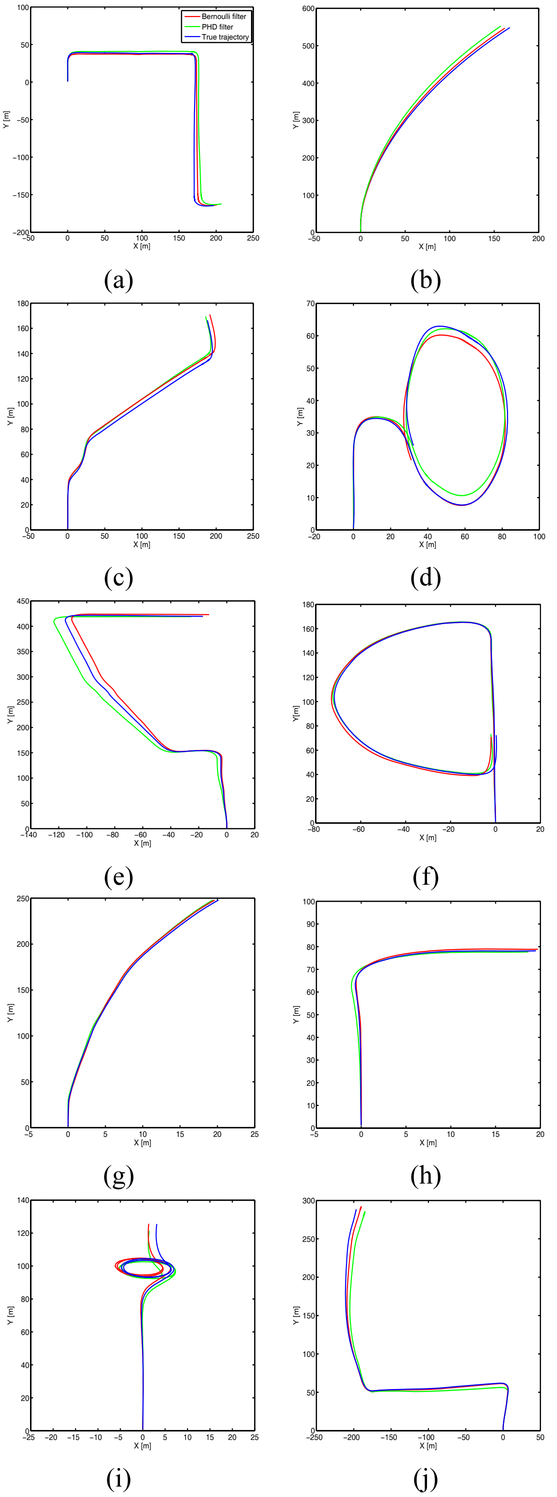

Fig. 3. Visual odometry result. Red line is the Bernoulli filter estimation. Green line is the PHD filter estimation. Blue line is the True trajectory.

Second, unevenly distributed features are considered as multiple measurements from an extended target whereas the effective extracting techniques are therefore avoided. Unevenly distributed features may influence the estimation results both provided by RANSAC and the PHD filter. Regarding to RANSAC approach, the influence exists in the initiation process since it requires a random subset during the iteratively process. On the other hand, although the PHD filter has the merging and pruning techniques to consider aggregated features as a single target, how to calculate the optimal merging threshold is still an issue. With respect to the Bernoulli filter, features are considered as the set-valued observation to update the ego-motion vector which allows the solution to the problem of dynamically tracking target without considering measurements' distribution. Features aggregated in a small region only influences individual targets, in contrast to the extended target tracking.

Third, the Bernoulli filter propagates the posterior intensity to eliminate the influences in condition that the features stemming from the moving objects. For feature based visual odometry approaches, how to remove features originating from moving objects is a challenge. Most researchers utilize the RANSAC approach by considering those features as outliers. However, as features are mostly from moving objects, those from static objects may be considered as the outliers. Within the RFS framework, the Bernoulli filter addresses this challenge by propagating the Bernoulli distribution which approximates the posterior density at each frame. According to the dynamic motion model (equation (5)), features from the moving objects are eliminated as clutter by Bayes filtering.

Last, the computation requirement is guaranteed. Compared to the standard SfM techniques which focus on individual features, the Bernoulli filter only tracks the ego-motion vector. In the PHD implementation, the ego-motion vector is calculated by averaging the corresponding velocities of the whole targets. The number of targets may become huge as time passes. Regarding the Bernoulli filter, features are treated as multiple measurements originated from an extended target. The overloaded phenomenon is therefore eliminated.

5. CONCLUSION

Visual odometry in urban scenarios is challenging due to a large amount of outliers. The PHD filter for visual odometry has already been investigated in our previous work. The average states of features are calculated to approximate the ego-motion vector under RFS framework. In this paper, a Bernoulli filter is proposed to address visual odometry in another way. Features are considered multiple measurements which originated from an extended target while the ego-motion vector is considered as the state of the target. Compared to the PHD solution, the Bernoulli filter estimates the state instead of tracking individual features, which provides a recursive filtering algorithm in the presence of association uncertainty. The overloaded issue is also avoided since the tracking process is solely focused on the extended target itself. Furthermore, the proposed Bernoulli filter avoids the clustering step necessary within the PHD filter, which makes the Bernoulli filter better adaptable in non-linear environments. The effectiveness of the proposed approach has been illustrated through experiments and an improvement in performance was achieved.

REFERENCES

[1] D. Scaramuzza, F. Fraundorfer, M. Pollefeys, and R. Siegwart, "Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints," *Proc. of International Conference on Computer Vision*, pp. 1413-1419, October 2009.

[2] G. Garcia, M. A. Sotelo, I. Parra, D. Fernandez, and M. Gavilan, "2D visual odometry method for global positioning measurement," *Proc. of International Symposium on Intelligent Signal Processing*, pp. 1-6, October 2007.

[3] A. Davison, "Real-time simultaneous localization and mapping with a single camera," *Proc. of International Conference on Computer Vision*, pp. 1403-1410, 2003.

[4] D. Burschka and G. D. Hager, "V-GPS (SLAM): vision-based inertial system for mobile robots," *Proc. of International Conference on Robotics and Automation*, pp. 409-415, May 2004.

[5] K. Konolige, M. Agrawal, and J. Sola, "Large-scale visual odometry for rough terrain," *Proc. of International Symposium on Research in Robotics*, pp. 201-212, November 2007.

[6] D. Scaramuzza and R. Siegwart, "Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 1015-1026, October 2008.

[7] D. Scaramuzza, F. Fraundorfer, and R. Siegwart, "Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC," *Proc. of International Conference on Robotics and Automation*, pp. 4293-4299, May 2009.

[8] C. McCarthy and N. Barnes, "Performance of optical flow techniques for indoor navigation with a mobile robot," *Proc. of International Conference on Robotics and Automation*, pp. 5093-5098, May 2004.

[9] J. Campbell, R. Sukthankar, and I. Nourbakhsh, "Techniques for evaluating optical flow for visual odometry in extreme terrain," *Proc. of International Conference on Intelligent Robots and Systems*, pp. 3704-3711, October 2004.

[10] P. Corke, D. Strelow, and S. Singh, "Omnidirectional visual odometry for a planetary rover," *Proc. of International Conference on Intelligent Robots and Systems*, pp. 4007-4012, October 2004

[11] F. Zhang, H. Stähle, A. Gaschler, C. Buckl, and A. Knoll, "Single camera visual odometry based on Random Finite Set Statistics," *Proc. of International Conference on Intelligent Robots and Systems*, pp. 559-566, October 2012.

[12] F. Zhang, H. Stähle, G. Chen, C. Buckl, and A. Knoll, "Visual odometry based on random finite set statistics in urban environment," *Proc. of Intelligent Vehicles Symposium*, pp. 69-74, June 2012.

[13] B.-T. Vo, C. See, N. Ma, and W. T. Ng, "Multi-sensor joint detection and tracking with the Bernoulli filter," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1385-1402, April 2012.

[14] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4091-4104, November 2006.

[15] W. Yang, Y. Fu, J. Long, and X. Li, "Joint detection, tracking, and classification of multiple targets in clutter using the PHD filter," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3594-3609, October 2012

[16] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on Bernoulli filters: theory, implementation and applications," *IEEE Trans. on Signal Processing*, vol. 61, no. 13, pp. 3406-3430, July 2013.

[17] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174-188, February 2002.

[18] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," *Proc. of Intelligent Vehicles Symposium*, pp. 486-492, June 2010.

[19] C. Harris and M. J. Stephens, "A combined corner and edge detector," *Proc. of Alvey Vision Conference*, pp. 147-152, 1988.

[20] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. of IEEE International Conference on Computer Vision*, pp. 1150-1157, 1999.

[21] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," *Proc. of European Conference on Computer Vision*, pp. 404-417, May 2006.

[22] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*, Carnegie Mellon, April 1991.

[23] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. on Aerospace and Electronic Systems*, no. 4, pp. 1152-1178, October 2003.

[24] N. Houshangi and F. Azizi, "Mobile robot position determination using data integration of odometry and gyroscope," *Proc. of Automation Congress*, pp. 1-8, July 2006.

Feihu Zhang received his B.Sc. degree in Automation and his M.Sc. degree in Control Theory and Application from Graduate School of Xi'an Jiaotong University, Xi'an, China, 2010. He is currently a Ph.D. candidate at Technical University of Munich. His main research interests are intelligent vehicle and robotics.

Daniel Clarke received his Ph.D. from the University of Innsbruck in 2009. Between 2013 and 2014 he joined fortiss as a leader in Data and Sensor Fusion research group. From 2015 he is a lecture at Cranfield University, at the Defense Academy of the United Kingdom. His research interests include the development of the techniques and methodologies necessary to realize the integration of homogeneous and heterogeneous sensor networks.

Alois Knoll received the diploma M.Sc. in Electrical/Communications Engineering from the University of Stuttgart, Germany, in 1985 and his Ph.D. (summa cum laude) in computer science from the Technical University of Berlin, Germany, in 1988. He served on the faculty of the computer science department of TU Berlin until 1993, when he qualified for teaching computer science at a university (habilitation). He then joined the Technical Faculty of the University of Bielefeld, where he was a full professor and the director of the research group Technical Informatics until 2001. Between May 2001 and April 2004 he was a member of the board of directors of the Fraunhofer-Institute for Autonomous Intelligent Systems. At AIS he was head of the research group Robotics Construction Kits, dedicated to research and development in the area of educational robotics. Since autumn 2001 he has been a professor of Computer Science at the Computer Science Department of the Technische Universität München. He is also on the board of directors of the Central Institute of Medical Technology at TUM (IMETUM-Garching); between April 2004 and March 2006 he was Executive Director of the Institute of Computer Science at TUM.