



Extraction of Motion Information from Occupancy Grid Map Using Keystone Transform

Hongqi Fan¹, Dawei Lu^{1,*}, Yanwen Jiang¹ and Achim J. Lilienthal²

¹National Key Laboratory of Automatic Target Recognition (ATR), National University of Defense Technology, Changsha 410073, China

²Perception for Intelligent Systems, Technical University of Munich, Munich, Germany

Abstract

Considering the tractability of OGM (Occupancy Grid Map) and its wide use in the dynamic environment representation of mobile robotics, the extraction of motion information from successive OGMs are very important for many tasks, such as SLAM (Simultaneously Localization And Mapping) and DATMO (Detection and Tracking of Moving Object). In this paper, we propose a novel motion extraction method based on the signal transform, called as S-KST (Spatial Keystone Transform), for the motion detection and estimation from successive noisy OGMs. It extends the KST in radar imaging or motion compensation to 1D spatial case (1DS-KST) and 2D spatial case (2DS-KST) combined multiple hypotheses about possible directions of moving obstacles. Meanwhile, the fast algorithm of 2DS-KST based on Chirp Z-Transform (CZT) is also given, which includes five steps, i.e. spatial FFT, directional filtering, CZT, spatial IFFT and Maximal Power Detector (MPD) merging and its computational complexity is proportional to the 2D-FFT. Simulation test results for the point objects and the extended objects show that SKST has a good performance on the extraction of sub-pixel motions in very noisy environment, especially for those slowly moving obstacles.

Keywords: Mobile robotics, occupancy grid map, moving object, keystone transform, 2DS-KST, velocity estimation.

Citation

Fan, H., Lu, D., Jiang, Y., & Lilienthal, A. J. (2024). Extraction of Motion Information from Occupancy Grid Map Using Keystone Transform. Chinese Journal of Information Fusion, 1(1), 63–78.

© 2024 IECE (Institute of Emerging and Computer Engineers)

Notifications

ι Unit of imaginary number;

l, m Discrete spatial variables for continuous ones x, y respectively;

i, j Discrete spatial frequencies for continuous ones u, v respectively;

n Discrete time variables for continuous one t ;

k Discrete temporal frequency for continuous one f ;

$\Delta_x = \Delta_y = R$ Size of Spatial cell, i.e., sampling interval in space;

$\Delta_t = T$ Frame period of 2D or 1D data, i.e., sampling interval in time;

V_{\max} Maximum speed of interest, known as a prior for a given problem;

L Number of space cells along x or y direction;

N Number of data frames, i.e. the length of 2D or 1D signal along t ;

$f(x, y, t)$ OGM or image at t instant for continuous case, denoted $f_t(x, y)$ as well;

$f(l, m, n)$ OGM or image at t_n instant for discrete case, denoted $f_n(l, m)$ as well;

Academic Editor:

You He

Submitted: 17 March 2024

Accepted: 06 June 2024

Published: 10 June 2024

Vol. 1, No. 1, 2024.

10.62762/CJIF.2024.361892

*Corresponding author:

Dawei Lu

davidloo.nudt@gmail.com

$\mathcal{F}(\cdot, \cdot, \cdot)$ Fourier transform of $f(\cdot, \cdot, \cdot)$ along any number of dimensions.

1 Introduction

Efficient perception of and reasoning about environments is still a major challenge for mobile robots operating in dynamic, densely cluttered or highly populated environments [1–3]. In the context of DAS (Driver Assistance System) [1, 4–6] or industrial field applications [7, 8], environment perception (or monitoring) usually includes three interleaved tasks, i.e., SLAM (Simultaneous Localization And Mapping) [9, 10], DATMO (Detection And Tracking of Moving Objects) [11] and CTM (Cell Transition Mapping) [12], which have already received a wide attention in the society of mobile robotics.

For the three tasks above, one of the most important things is the fast and reliable extraction of motion information from successive sensor observations. It consists of MOD (Moving Object Detection) even as well as their velocity estimation, which directly affects the performance of localization, mapping, moving object tracking [13]. Furthermore, high level tasks such as path planning, collision avoidance, will be affected by it as well [14]. For example, reasoning on behaviors in DAS requires to separate environment into static and dynamic parts [6, 15]. False negatives (dynamic objects) can lead to serious errors in the resulting maps such as spurious objects or misalignments due to localization errors [16]. On the other side, false positives (static objects) will degrade the performance and computability of tracking module, since the complexity of typical object tracking algorithm increases combinatorially with object number. Meanwhile, MOT (Moving Object Tracking) and CTM can directly benefit from the accurate and timely velocity estimation for moving objects in the dynamic environments.

This paper discuss the problem of extracting motion information from successive sensor observations, including MOD and their velocity estimation. In order to solve this problem effectively, the first thing of all is choosing a suitable representation of dynamic environments. Among all the existing environment representations, OGM (occupancy grid map) proposed by Elfes [17, 18] is the most popular one, which maps the environment as an array of probabilistic cells and easily integrates scans from multiple sensors, even from different type of sensors, for instance, sonar, laser range finder, IR camera [6]. Considering this tractability of OGMs and its wide use

in SLAM and DATMO, this paper will take OGMs as inputs, and pay attention on the problem of extracting motion information from successive OGMs.

1.1 Related work

Almost all research so far on extracting motion information from OGMs have been done under the framework of DATMO. In [11], Petrovskaya et al. propose to classify these methods in three categories: Traditional DATMO, Model-based DATMO and Grid-based DATMO. See [11] and the comprehensive survey [19] for more details about DATMO.

In this paper, we focus our attention on the detection of moving grid cells and their velocity estimation, regardless of the tracking problem on object level. According to different ways of understanding motion, we classify those techniques into three categories: Object-Oriented (OO), Cell-Oriented(CO), and OCCupancy-Oriented(OCO).

1.1.1 OO methods

OO approaches attribute the dynamic changes of successive local OGMs to moving objects. As a result, DATMO are divided into two sub-problems, MOD and MMOT (Multiple Moving Object Tracking) [3, 4, 13]. MOD is to isolate the moving objects in dynamic environments and MMOT is to achieve their state estimation and filter false alarms using typical multiple target tracking algorithms, such as MHT [4, 13] and JPDA [3]. For MOD we are interested in, a consistency based approach, which is based on inconsistencies observed for new data by comparing them with maps constructed by SLAM [4, 13], has often been used. In [3], a time-fading static map is used for consistency detector instead of the static map constructed by SLAM. Another important clue whether the object is moving or not is those moving objects detected in the past. To exploiting this information, a local dynamic grid map is also created to store information about previously detected moving objects in [4, 13]. If an occupied grid cell is near an area that was previously occupied by moving objects, it can be recognized as a potential moving object. There exist at least three drawbacks as follows for OO motion information extraction:

- Consistency-based MOD can not effectively detect those slowly moving objects, such as pedestrians on the road, especially in the case of short time interval between consecutive measurements.
- Velocity of every cell can not be extracted from OGMs directly, which usually is estimated by subsequent MMOT algorithm or by fusing MOD

results with other sensor capable of measuring velocity, such as radar. For example, in [4], the author proposed a generic architecture to solve SLAM and DATMO in dynamic outdoor environments, in which the object detection results were fusing with radar local data and provide the detected objects with their velocities.

- The last, perhaps one having most criticisms, is that under the view of OO, MMOT following MOD usually need complex data clustering and association which have combinatorial complexity and whose performance drastically degrade with the number of false alarms of MOD. To suppress the off-road false alarms given by MOD, [20] integrated the road model into their DATMO framework.

1.1.2 CO methods

Different from the viewpoint of OO, CO methods think that the dynamic changes of local OGMs are caused by the motion of grid cells rather than objects. The most remarkable work in this aspect is so called BOF (Bayesian Occupancy Filter), [1] and [21]. The former (4D-BOF) combines the occupancy grid with probabilistic velocity objects and leads to a four dimensional grid representation. However, the latter (2D-BOF) still uses a 2-dimensional occupancy grid but attaches an associated velocity distribution for every cell. It is obviously found that the 4D-BOF can represent overlapping objects with different velocities, while the 2D-BOF has the advantages to be computationally less demanding [6]. [2] evaluated these BOFs respectively through two experiments, the collision danger estimation for 4D-BOF and the human tracking for 2D-BOF. BOFs provide a Bayesian framework for grid-based monitoring of the dynamic environment. It allows us to extract information of motion cells, containing both occupancy and velocity, only based on the sequences of local OGMs. Furthermore, the complex track association operations is no need, as not existing concepts of objects or tracks in BOF model. Thus it is very suit for those applications in which information on the object level is not concerned. However, for those applications in which object level information is concerned, [22] give an easy way to integrate the 2D-BOF in [1, 23] with FCTA (Fast Clustering and Tracking Algorithm). Because BOF estimates the occupancy and velocity values for both static and dynamic parts of the environment, the subsequent FCTA has a dependency of parameters. In order to solve this problem, [6] integrated a novel real time scheme of MOD into

the framework of [22] so that the static parts could be effectively removed from the output of BOF. The MOD in [6] is essentially a consistency-based detector, but it is based on transferring occupancy information between consecutive data grids rather than performing a complete SLAM solution. See [6] for more details about its MOD method.

1.1.3 OCO methods

The key insight of OCO methods is that the dynamic changes of local OGMs are caused by the flowing occupancies in grid cells rather not by the motion of cells themselves. Thus velocity is no longer defined for cells, but for occupancy inside the cell boundaries [5]. Therefore, every occupied cell could have more than one ancestor cell, that is to say, it can be having different velocities simultaneously. Another key fact under the OCO viewpoint is the occupancy preservation, which means the occupancy cannot disappear, unless at the border. From this viewpoint, [5] proposed an improved 2D-BOF, called as BOFUM (BOF Using prior Map). In BOFUM, the prior map information is encapsulated into the reachability matrix and its probability, which describe respectively whether a cell C can be reached from an ancestor cell A, and how the probability of this transition is. BOFUM also uses additional velocity states for dynamic model adaption, which clearly differs in previous approaches [2, 21]. As a result, BOFUM can predict the cell transitions more accurately than those CO methods. Similar to the reachability matrix and its probability in [5], [12] proposed a concept of CTMap (Conditional Transition Map) to model the motion patterns in dynamic environments. However, the transition parameters of CTMap was learned from a temporal signal of occupancy in cells in [12] rather than the method using in [5]. To extract the possible transition, a local neighborhood cross-correlation method was used in [12], which resulted in a large computationally demanding and was constrained with noise in OGMs, such as leaves or hands shaking. Moreover, the cross-correlation method can only give the direction of cell transition, but not give the full information of cell velocity. Nonetheless, it is worth pointing that regarding occupancy in cells as a temporal signal in [12] open a new window for motion information extraction form OGMs. Following this idea and based on the OCO viewpoint, a dual PHD filter was proposed [24], which can separate the dynamic and static cells and estimate the posterior occupancy and its flowing velocity in each cell under the random finite sets filtering framework. It provided the formal and strong

way to deal with the OGMs for multiple purposes, such as BOF, CTMap building. However, like the most of RFS filters, the import modeling process according to the application scenario is relatively hard for the engineers.

1.2 KST in radar signal processing

In this paper, we attempt to propose a different way of extracting motion information from successive OGMs based on a signal transformation, called the keystone transform (KST), which has been popularly used in the radar signal processing community. For example, [25] employed the KST was used to remove the linear component of the range migration for the moving target in the synthetic aperture radar (SAR) imaging. [26] proposed the fast implementation of KST based on Chirp-Z transform (CZT) in order to overcome the range migration in the long time coherent accumulation for detecting the dim moving targets. Recently, the high order KST were proposed for the above applications in [27] [28] and [29] to correct the three order phase migration of the radar echo signal. For extracting the motion information from the OGMs, there exist many similarities to the range migration correction in the radar signal processing, such as focusing or accumulating the moving object with the unknown velocity, and the concepts of range and Doppler in radar signal processing are like the space cell and the velocity in the OGMs. Nevertheless, there also exist several differences with the KST of radar signal processing. One significant difference is that the radar signal in the complex number field and the OGMs is in the real number field. Another important difference is the unknown motion is usually the radial motion along the line of sight of the radar, although it can be high order motion (i.e. with the non-zero acceleration). However, the spatial motion in the OGMs can be one dimensional along the street way, two dimensional along the ground, and three dimensional in the free space. For the ground mobile robotics, we consider one dimensional and two dimensional cases in this paper.

1.3 Scope of this paper

This paper tries to develop a different way of extracting motion information from successive OGMs based on a signal transformation by extending the KST in the radar signal processing community to the 1D and 2D spatial case. The main theoretic idea occurred in our conference paper [30] and this journal article was mainly enlarged from the three aspects, i.e., the more detailed survey, the fast algorithm implementation of

2DS-KST and the experiments for the extended objects. For the sake of integrity, the original point object test results are also kept in this article.

2 One Dimensional Spatial KST

Occupancy grid maps model the environment as an array of cells. Typically, these are layered out in a two-dimensional grid. However, we first discuss the keystone transform for the one dimensional case, which is called as 1DS-KST hereinafter. One reason is that there has a prior straight line constraint with the motion of objects in many applications, for instance, cars moving on the highway or city roads. Detection and estimation the velocity of object moving along straight line themselves have a certain significance for these cases. Another reason is that 1DS-KST, where the concept of fast time is instead by one dimensional spatial grid, has a more direct relationship with KST in radar signal processing. It is helpful to understand the principle of keystone transform, especially for understanding the two dimensional spatial KST introduced in the next section.

For 1DS-KST, the main assumptions are the following:

- The velocities for all moving objects are constant during N successive frames.
- $R \geq 2 \cdot V_{\max} \cdot T$, i.e., $V_{\max} \leq R/(2T)$.
- The sensor is motionless or the motion of it has been compensated by SLAM or other methods.

The first assumption always holds as long as the total length of time window $N \cdot T$ is very short or the amount of velocity change is less than one velocity resolution cell of KST. The second assumption is borrowed from [12], and it is derived from the Nyquist Sampling condition, which ensures the temporal signal of the occupancies in every grid cell are not aliased at a time sampling period T . Meanwhile it ensures the spatial continuity of the motion of objects, which is necessary to filter the non-continuous changes of OGMs. In fact, for typical applications and modern sensors, this condition is easy to meet. For example, if the size of grid cell R is equal to 2 meters and the time sampling period T is equal to 0.1 seconds, the maximum velocity of objects is 10 m/s, which is enough high for most dynamic environment monitoring applications involving pedestrians, industrial robots and vehicles. As for those applications having higher maximal speed, such as automotive application, we can use a larger R or smaller T in order to avoid objects "jumping over" adjacent cells. As the focus of this paper is to extract the

motion information from OGMs, the third assumption is natural and it can make this problem isolated from other problems, such as registration and localization.

Let us first consider the case that there is only one occupied grid cell, called as an ideal point object blow, in the sensor field of view. Assume it is moving at a constant velocity V , $V \in [-V_{\max}/2, +V_{\max}/2)$, and has an initial position l_0R , then for any given time instant t , the obtained OGM has the following form:

$$f_t(l) = \delta_{r_t}(l) \quad (1)$$

where δ denote a unit pulse function, which is defined as

$$\delta_{r_t}(l) = \begin{cases} 1 & \text{if } r_t \in [lR - \frac{R}{2}, lR + \frac{R}{2}) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

And in (1) r_t is the position of this object at time instant t , which can be written as

$$r_t = l_0R + Vt \quad (3)$$

The Discrete Fourier Transform $\mathcal{F}_t(i)$ of $f_t(l)$ can be given as follows

$$\mathcal{F}_t(i) \stackrel{\text{def.}}{=} \sum_{l=0}^{L-1} f_t(l) \cdot \exp\left(-i2\pi \frac{l \cdot i}{L}\right) \quad (4)$$

$$\cong \exp\left(-i2\pi \frac{r_t \cdot i}{LR}\right) \quad (5)$$

$$= \exp\left(-i2\pi \frac{l_0i}{L}\right) \cdot \exp\left(-i2\pi \frac{V \cdot t \cdot i}{LR}\right) \quad (6)$$

Since the signal $f_t(l)$ is a real signal in the case of OGM, $\mathcal{F}_t(i)$ always satisfies conjugate symmetry, i.e.,

$$\mathcal{F}_t(i) = \mathcal{F}_t^*(L - i) \quad (7)$$

Therefore, we only concern the non-negative spatial frequency cells, that is, cells of $i = 0, \dots, L/2 - 1$.

To use KST, we need choose a fixed cell of spatial frequency i_c as a reference. In general, the center frequency within the effective bandwidth of signal is chosen for a reference in the application of radar signal processing. For our case of 1D-OGM, we can multiply $\mathcal{F}_t(i)$ by a window $w(i)$ in the spatial frequency domain, which corresponds to a spatial filtering process and results in a blurred OGM. Without loss of generality, we denote this window $W(i)$ as the following:

$$W(i), \quad i_{\min} \leq i \leq i_{\max} \quad (8)$$

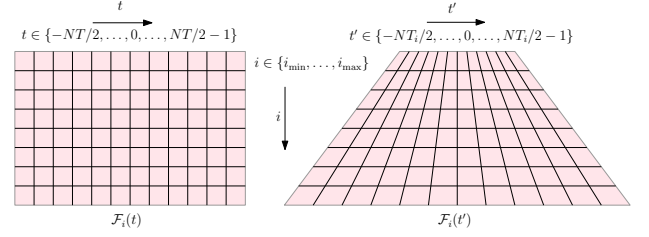


Figure 1. Sampling pattern of Keystone transform.

So $i_c = (i_{\min} + i_{\max})/2$ can be selected as the reference. Thus (6) can be expressed as

$$\mathcal{F}_t(i) = \exp\left(-i2\pi \frac{l_0i}{L}\right) \cdot W(i) \cdot \exp\left(-i2\pi \frac{V i_c}{LR} \cdot \frac{i}{i_c} t\right), \quad i_{\min} \leq i \leq i_{\max}. \quad (9)$$

If let $t' = \frac{i}{i_c} \cdot t$, we can get

$$\mathcal{F}_t(i) = \exp\left(-i2\pi \frac{l_0i}{L}\right) \cdot W(i) \cdot \exp\left(-i2\pi \frac{V i_c}{LR} \cdot t'\right) \quad (10)$$

$$\stackrel{\text{def.}}{=} \mathcal{F}_i(t') \quad (11)$$

Now it is the time to consider a temporal variable t . During the time interval $[-NT/2, NT/2)$, we obtained N successive OGMs at some discrete time instants. Without loss of generality, we can assume that $t = -NT/2, \dots, 0, T, \dots, NT/2 - T$, that is to say, we obtain many signals $\mathcal{F}_i(t')$ at some time instants t' .

Since the scale factor i/i_c between t and t' is variable with i , the sampling period T_i and total time interval NT_i for t' are both different in terms of i . The sampling patterns of $\mathcal{F}_t(i)$ and $\mathcal{F}_i(t')$ are shown as Fig.1. In Fig.1, the sampling pattern of $\mathcal{F}_i(t')$ is like a keystone shape, that's why this corresponding transform is called as KST.

To correct the keystone effect in the sampling pattern of $\mathcal{F}_i(t')$, we need to compute the values at the discrete time $t' = -NT/2, \dots, 0, T, \dots, NT/2 - T$ for every i , which are usually obtained by an interpolate filter $h_i(n')$ in the context of KST. See [31] for more details of the interpolate filter $h_i(n')$. Thus, after the interpolate filtering,

$$\tilde{\mathcal{F}}_i(n) = \mathcal{F}_i(n') \otimes h_i(n') \quad (12)$$

$$\cong \exp\left(-i2\pi \frac{l_0i}{L}\right) \cdot W(i) \cdot \exp\left(-i2\pi \frac{V i_c}{LR} \cdot nT\right),$$

$$n = -N/2, \dots, 0, 1, \dots, N/2 - 1 \quad (13)$$

Then the IDFT transform of $\tilde{\mathcal{F}}_i(n)$ in terms of i will give the following result:

$$\tilde{f}_n(l) = \delta(l - l_0) \otimes w(l) \cdot \exp\left(-\iota 2\pi \frac{V i_c}{LR} \cdot nT\right) \quad (14)$$

$$= w(l - l_0) \cdot \exp\left(-\iota 2\pi \frac{V i_c}{LR} \cdot nT\right) \quad (15)$$

$$= \begin{cases} w(0) \cdot \exp\left(-\iota 2\pi \frac{V i_c}{LR} \cdot nT\right) & \text{if } l = l_0 \\ w(l - l_0) \cdot \exp\left(-\iota 2\pi \frac{V i_c}{LR} \cdot nT\right) & \text{otherwise} \end{cases} \quad (16)$$

where, $w(l)$ is coefficients of spatial filter corresponding with the window $W(i)$.

In general, we must choose an appropriate window type and a suitable width of $W(i)$ so that $w(i)$ has a locally compact main-lobe and a side-lobe low enough. Thus, for those cells of $l \neq l_0$, the non-zero $\tilde{f}_n(l)$ will not affect the analysis of velocity as long as the distance of object is larger than the width of main-lobe of $w(i)$. This is easy to meet if the moving objects are not closely spaced pixel by pixel. Fortunately, this is the fact in OGM case. In fact, even for the superpositional target, if they have different velocities, the non-zero coefficients of $w(l)$ at $l \neq 0$ have no effect to velocity analysis as well. Therefore, let us consider the cell l_0 as the next step,

$$\tilde{f}_n(l_0) = w(0) \cdot \exp\left(-\iota 2\pi \frac{V T i_c}{LR} \cdot n\right) \quad (17)$$

After doing the DFT for $\tilde{f}_n(l_0)$ in terms of n , we can get:

$$\tilde{\mathcal{F}}_{l_0}(k) = \sum_{n=-N/2}^{N/2-1} w(0) \cdot \exp\left(-\iota 2\pi \frac{V T i_c}{LR} \cdot n\right) \cdot \exp\left(-\iota 2\pi \frac{nk}{N}\right) \quad (18)$$

$$= \sum_{n=-N/2}^{N/2-1} w(0) \cdot \exp\left(-\iota 2\pi n \left(\frac{V T i_c}{LR} + \frac{k}{N}\right)\right) \quad (19)$$

As (19) shown, different velocities V will be located at the different frequency grid cells k . The KST method, therefore, allow for different velocities in a single cell, which is similar to 4D-BOF [1] and BOFUM [5]. For an object having the velocity V , we have the following approximation:

$$\frac{k}{N} \cong -\frac{V T i_c}{LR} \Rightarrow V \cong -\frac{kLR}{N T i_c} \quad (20)$$

The corresponding resolution of the above velocity measurement is

$$\Delta V = \frac{LR}{N T i_c} \quad (21)$$

and the normalized one is

$$\Delta V = \frac{\Delta V \cdot T}{R} = \frac{L}{N i_c} \quad (22)$$

We can choose i_c and N according to the dynamic characteristics of environment and the velocity resolution of interest which is related to the minimal detectable velocity.

3 Two Dimensional Spatial KST

A one-dimensional OGM is of limited practical use. For mobile robots, the two-dimensional OGM is the usual case. This section will develop a method of two dimensional spatial KST (denoted as 2DS-KST) for the purpose of extracting motion information from successive OGMs. Besides of those assumptions in section 2, two additional ones needed here is as the following:

- All objects are moving along nearly the same but unknown direction during N frames¹.
- The unknown motion directions for all objects belong to a prior set with finite number of elements and are constant during N frames.

3.1 2DS-KST with multiple hypotheses

Let us still consider only one ideal point object in the sensor field of view. Assume its initial position is $\mathbf{r}_0 = [l_0, m_0]^T$ and it has a constant velocity $\mathbf{V} = [V_x, V_y]^T = [V \cos(\theta), V \sin(\theta)]^T$, in which V_x, V_y satisfy the condition of $V_x, V_y \in [-V_{\max}/2, +V_{\max}/2)$. Then for a given time instant t , the OGM has the following form:

$$\delta_{\mathbf{r}_t}(l, m) = \begin{cases} 1 & \text{if } [r_{xt}, r_{yt}]^T \in \text{Rect}(l, m) \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

where $\text{Rect}(l, m)$ denote the region of the cell (l, m) .

For the 2D spatial case, equation (5) becomes as follows:

$$\mathcal{F}_t(i, j) = \exp\left(-\iota 2\pi \frac{\mathbf{r}_0 \cdot \mathbf{i}}{L}\right) \cdot \exp\left(-\iota 2\pi \frac{V \mathbf{u}_\theta \cdot \mathbf{i}}{LR} t\right) \quad (24)$$

$$= \exp\left(-\iota 2\pi \frac{\mathbf{r}_0 \cdot \mathbf{i}}{L}\right) \cdot \exp\left(-\iota 2\pi \frac{V i_\theta}{LR} t\right) \quad (25)$$

¹This condition may be relaxed according to the future results.

where $\mathbf{i} = [i, j]^T$, $\mathbf{u}_\theta = [\cos \theta, \sin \theta]^T$ and $i_\theta = i \cos \theta + j \sin \theta$.

As the above described, we assume there are finite possible hypotheses for θ , and denote the p th hypothesis as θ_p ($p = 1, \dots, \nu$). But the real case is we don't know the actually moving direction of the object, so we need to do KST for every possible hypothesis, then we can get:

$$\begin{aligned} \mathcal{F}_t^{\theta_p}(i, j) = & \exp\left(-i2\pi \frac{\mathbf{r}_0 \cdot \mathbf{i}}{L}\right) \cdot W_{\theta_p}(\mathbf{i}) \\ & \cdot \exp\left(-i2\pi \frac{V_{\theta_p} i_c^{\theta_p}}{LR} \cdot \frac{i_{\theta_p}}{i_c^{\theta_p}} \cdot t\right) \\ & \cdot \exp\left(-i2\pi \frac{V_{\theta_p^\perp} i_{\theta_p^\perp}}{LR} \cdot t\right) \end{aligned} \quad (26)$$

where

- V_{θ_p} , $i_c^{\theta_p}$ and i_{θ_p} are projections of vectors \mathbf{V} , $\mathbf{i}_c^{\theta_p}$ and \mathbf{i} along the θ_p direction, respectively, while $\mathbf{i}_c^{\theta_p}$ is the reference spatial frequency vector for θ_p hypothesis when doing KST in next step;
- $V_{\theta_p^\perp}$, $i_{\theta_p^\perp}$ are projections of vectors \mathbf{V} , \mathbf{i} perpendicular to the θ_p direction, respectively;
- $W_{\theta_p}(\mathbf{i})$ is the two dimensional window function for θ_p hypothesis which has the same meaning as the window (8) in one dimensional case.

(26) result from the fact that the item $V i_\theta$ in (25) can be rewritten as:

$$V i_\theta = V_{\theta_p} i_{\theta_p} + V_{\theta_p^\perp} i_{\theta_p^\perp} \quad (27)$$

Furthermore, if the hypothesis θ_p is true, that is to say, θ is approximately parallel to θ_p , the last product item in (26) can be ignored further. In this case, for the θ_p hypothesis, (26) can be approximated as the following:

$$\begin{aligned} \mathcal{F}_t^{\theta_p}(i, j) \cong & \exp\left(-i2\pi \frac{\mathbf{r}_0 \cdot \mathbf{i}}{L}\right) \cdot W_{\theta_p}(\mathbf{i}) \\ & \cdot \exp\left(-i2\pi \frac{V_{\theta_p} i_c^{\theta_p}}{LR} \cdot \frac{i_{\theta_p}}{i_c^{\theta_p}} \cdot t\right) \end{aligned} \quad (28)$$

$$\stackrel{\text{def.}}{=} \mathcal{F}_i^{\theta_p}(t') \quad (29)$$

where t' is defined as:

$$t' = \frac{i_{\theta_p}}{i_c^{\theta_p}} t \quad (30)$$

The interpolate filter of Keystone transform is the same as 1D case, so we can get $\tilde{f}_n^{\theta_p}(l_0, m_0)$ as follows:

$$\tilde{f}_n^{\theta_p}(l_0, m_0) = w_{\theta_p}(0, 0) \cdot \exp\left(-i2\pi \frac{V_{\theta_p} T i_c^{\theta_p}}{LR} \cdot n\right) \quad (31)$$

The counterpart of the above equation in 1D case is (17). After doing the DFT for $\tilde{f}_n^{\theta_p}(l_0, m_0)$ in terms of n for every occupied grid cell $[l_0, m_0]^T$, we can get:

$$\begin{aligned} \tilde{\mathcal{F}}_{l_0, m_0}^{\theta_p}(k) = & \sum_{n=-N/2}^{N/2-1} w_{\theta_p}(0, 0) \\ & \cdot \exp\left(-i2\pi n \left(\frac{V_{\theta_p} T i_c^{\theta_p}}{LR} + \frac{k}{N}\right)\right) \end{aligned} \quad (32)$$

Thus, similar to 1D case, V_{θ_p} , the magnitude of the velocity in the grid cell $[l_0, m_0]^T$ can be given as follows:

$$V_{\theta_p} \cong -\frac{kLR}{NT i_c^{\theta_p}} \quad (33)$$

while the direction of the velocity in this cell is given by θ_p , which is the assumption given in the previous, once used in (28).

3.2 Merging multiple hypotheses

The velocity measurement given by (33) is in the case of the θ_p hypothesis is true for this cell. In practice, we don't know which hypothesis is true for any occupied grid cell $[l, m]^T$ in advance, so we must merge the results of those multiple hypotheses so that the correct motion information can be given.

In fact, by 2DS-KST with multi-hypothesis, we can get ν cuboid $\tilde{\mathcal{F}}^{\theta_p}(l, m, k)$, which is the general form of (32) for any arbitrary grid cell $[l, m]^T$ under the hypothesis θ_p . That is to say, we get the two dimensional matrix $\tilde{\mathcal{F}}_{l, m}(\theta_p, k)$ for every grid cell $[l, m]^T$, which represents the possible velocity in this cell, including the magnitude (denoted by k , which can be negative) and the direction (denoted by θ_p).

Here we use a MPD (Maximal Power Detector) based method for the purpose of extracting the motion information from $\tilde{\mathcal{F}}_{l, m}(\theta_p, k)$, which is based on the fact that the power item $|\tilde{\mathcal{F}}_{l, m}(\theta_p, k)|^2$ will be larger when θ_p and k are more matched with the true value of the velocity. It can be described as the following four steps:

- The first step is the maximal merging step:

$$\mathcal{P}(l, m) = \max_{\theta_p, k} |\tilde{\mathcal{F}}_{l, m}(\theta_p, k)|^2 \quad (34)$$

- The second is the following power detector:

$$\mathcal{P}(l, m) \underset{H_0}{\overset{H_1}{\geq}} P_{\min} \quad (35)$$

where H_1 and H_0 denote the event whether the grid cell to be determined is occupied or not, respectively, and where P_{\min} denote the threshold of MPD.

- The third is the following estimator if H_1 holds on:

$$(\hat{k}, \hat{\theta}) = \arg \max_{\theta_p, k} |\tilde{\mathcal{F}}_{l,m}(\theta_p, k)|^2 \quad (36)$$

$$\hat{\mathbf{V}}_{l,m} = \hat{V}_{\hat{\theta}} \cdot \exp(i\hat{\theta}) \quad (37)$$

where $\hat{V}_{\hat{\theta}}$ can be obtained according to (33) if \hat{k} given.

- The last step is the separator as follows:

$$|\hat{\mathbf{V}}_{l,m}| \underset{S}{\overset{D}{\geq}} V_{\min} \quad (38)$$

where D and S denote the event whether the grid cell to be determined is dynamic or not, respectively, and where V_{\min} denote the threshold of the separator, which directly related to the minimal detectable velocity or ΔV in (21).

As the above MPD approach of merging multiple hypotheses only outputs the most likely point estimation for the velocity, it is obvious that it is only appropriate for the case when the occupancies in each grid cell only have one significant velocity during the time interval. Nevertheless, it does not affect the versatility of 2D-KST for extracting motion information. For instance, we can use the more complex GMM (Gaussian Mixture Model) instead of the point estimation in MPD method as the velocity model of occupancy, which would undoubtedly increase the robustness and the compatibility to the complex situations. In fact, which velocity model should be selected is a key problem in practice and it closely depends on the physical application.

For the more complex method of merging multiple hypotheses, it is already beyond the scope of this paper and may be an appropriate topic for the future work.

4 Algorithm Implementation

In the previous two sections, 1DS-KST and 2DS-KST are presented, which are both interpolate filter based KST. In this section, a different implementation of

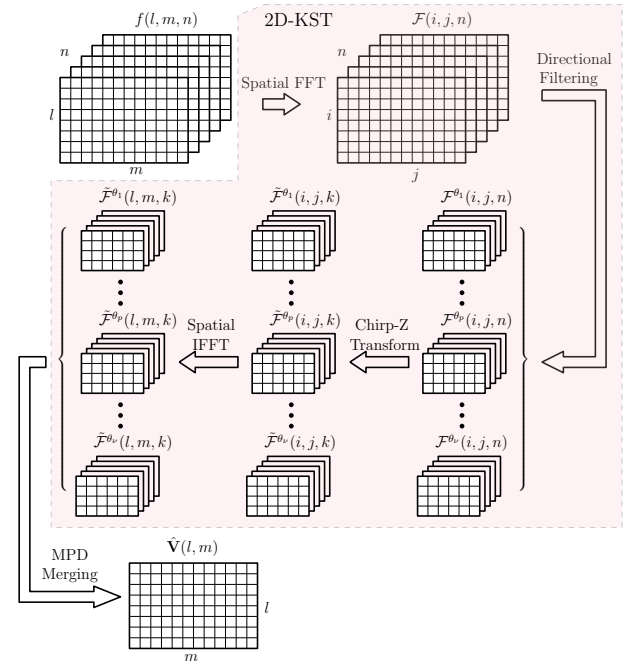


Figure 2. Flow diagram of 2D-KST based on CZT.

2DS-KST, which is more effective for our problem, is presented. The other issues related to implementation will be also discussed.

According to the description in Section 3.2, in order to extract the motion information from OGMs, we need the cuboid of $\tilde{\mathcal{F}}^{\theta_p}(l, m, k)$ under each hypothesis θ_p ($p = 1, \dots, \nu$), rather than the direct result $\tilde{f}^{\theta_p}(l, m, n)$ of the keystone transform for the original input $f(l, m, n)$. In the previous two sections, we used the interpolate filter based keystone transform to obtain $\tilde{f}^{\theta_p}(l, m, n)$ first, then obtained $\tilde{\mathcal{F}}^{\theta_p}(l, m, k)$ by doing the DFT transform for $\tilde{f}^{\theta_p}(l, m, n)$ in terms of n .

Here we introduce another implementation without using interpolate filter, which is based on the so called CZT (Chirp-Z Transform)[32]. It can both correct the keystone effect in the sampling pattern of the data $\mathcal{F}^{\theta_p}(i, j, n)$ and transform them to the k domain simultaneously. That is to say, it outputs $\tilde{\mathcal{F}}^{\theta_p}(i, j, k)$ instead of $\tilde{\mathcal{F}}^{\theta_p}(i, j, n)$ in the interpolate filter based method, so it is more effective for our problem. More importantly, CZT has the fast implementation and can flexibly control the frequency range to be analyzed. See [26, 32] for more details and applications of CZT. The flow diagram of CZT based 2D-KST is shown in Fig.2, where the data pattern in every step is explicitly shown and the meanings of those variables can be seen in section 2, 3 and the notations before the introduction.

The main steps in Fig.2 are explained as follows:

1. *Spatial FFT*: The OGMs are transformed from the spatial domain (l, m) to the spatial frequency domain (i, j) by 2D-FFT in this step, which can be done sequentially for each new OGM, or with batch processing for all N OGMs through parallel computing. In order to use FFT, L is usually a power of 2. If it is not this case, we can pad zeros at the tail. Thus the complexity of this step is $\mathcal{O}(2\alpha NL^2 \log L)$, where α ranges around the interval from 4 to 5 depending on the implementation [33].
2. *Directional filtering*: The window $W_{\theta_p}(i, j)$ in spatial frequency domain under every possible direction θ_p of velocity is multiplied respectively to the previous results, which corresponds to let every OGM pass several bandpass spatial filters, respectively. Here we use the following rectangle window,

$$W_{\theta_p}(i, j) = \begin{cases} 1 & \text{if } i_{\theta_p} \in [i_{\min}^{\theta_p}, i_{\max}^{\theta_p}] \\ 0 & \text{otherwise} \end{cases} \quad (39)$$

where $i_{\min}^{\theta_p}$ and $i_{\max}^{\theta_p}$ are the minimal and the maximal spatial frequency respectively, and the response of the spatial filter are completely determined by them. The proper values for these two parameters are closely dependent on the size of the moving objects. We will discuss this question in Section 5. From the computational point of view, the rectangle window in (39) has two advantages at least. One is that the multiplication operation can be bypassed, the other is that only the data points inside of $W_{\theta_p}(i, j)$ need to be computed in the next step. Because the window can be preset beforehand and there is no multiplication operation, the complexity of this step can be ignored.

3. *Chirp-Z transform*: As described before and shown in Fig.2, CZT is used both to correct the keystone effect in the sampling pattern of the data $\mathcal{F}^{\theta_p}(i, j, n)$ and to transform them to the k domain simultaneously. For our application, the parameters of CZT are set as follows:

$$z_k = A_{\theta_p} \cdot \omega_{\theta_p}^{-k} \quad (40)$$

$$\omega_{\theta_p} = \exp\left(-l \frac{2\pi}{N} \cdot \frac{i_{\theta_p}}{i_c^{\theta_p}}\right) \quad (41)$$

$$A_{\theta_p} = \omega_{\theta_p}^{-K/2} \quad (42)$$

$$K \geq \frac{N \arg_{\theta_p} \max(i_c^{\theta_p})}{L} \quad (43)$$

where K is the number of points of temporal frequency. To keep all hypotheses have the same the number of points of temporal frequency, we use the item $\arg_{\theta_p} \max(i_c^{\theta_p})$ in (43) instead of $i_c^{\theta_p}$. According to the principle of CZT, there is no other constraint to K . However, if K can divide N and N is a power of 2, the complexity of CZT is minimal and given by [33]:

$$\mathcal{C}_{CZT} = 4\alpha N \log K + (4\alpha + 25)N - K \quad (44)$$

For our problem, the typical setting is $K = N/2$ and the area of every window $S(W_{\theta_p}) = L^2/4$, thus the complexity of this step is approximate to $\mathcal{O}(\nu L^2 \cdot (\alpha N \log N + 6))$.

4. *Spatial IFFT*: As shown in Fig.2, this step uses the 2D-FFT to transform every cuboid $\tilde{\mathcal{F}}^{\theta_p}(i, j, k)$ from spatial frequency domain to spatial domain. As the migration across grid cells caused by motion has been compensated, the moving grid cell can be focused to the average, initial, or last position during the NT interval, which depends on the definition of the time index used in the CZT. Obviously, the complexity of this step is $\mathcal{O}(2\alpha \nu NL^2 \log L)$.
5. *MPD merging*: The velocity of any cell identified as moving one is estimated in this step. The complexity of this step is approximately $\mathcal{O}(\nu NL^2/2)$.

According to the above analysis, the total complexity of 2D-KST can be approximated as follows:

$$\mathcal{C}_{\text{Total}} \cong \mathcal{O}((1 + \nu \log_L N/2 + \nu) \cdot 2\alpha NL^2 \log L) \quad (45)$$

(45) means that the computational complexity of the total processing of 2D-KST is about $(\nu \log_L N/2 + \nu)N$ times more than the complexity of 2D-FFT with the side length L .

5 Results

In this section, we design three experiments to evaluate the performance of the proposed method on the extraction of motion information. The first two is based on numerical simulation data, and the last one is based on the real data from mobile robot.

The first experiment, called as point object test, is to demonstrate the validity of our method through point object data, which means all objects only occupy one grid cell in this experiment. Some intermediate results are also shown as figures in this experiment, which is benefit to understanding the process of KST.

Table 1. Simulation Parameters in the Experiment I

	Parm.	1D case	2D case
OGMs	L	128	64
	N	100	40
Objects State l_0, m_0, V, θ	#0	20, 0	10, 10, 0, -
	#1	40, -0.5	20, 15, 0.5, 0°
	#2	60, 0.05	30, 20, 0.1, 90°
	#3	80, -0.2	35, 30, 0.2, 45°
	#4	100, 0.1	40, 40, 0.3, 135°
	#5	—	45, 50, 0.4, 165°
Poisson Rate	λ	16	64
Windows	θ_p	—	$0 : \pi/8 : 7\pi/8$
	$i_c^{\theta_p}$	$L/4$	$L/4\alpha$
	i_{\min}	$L/8$	$i_c^{\theta_p}/2$
	i_{\max}	$3L/8$	$3i_c^{\theta_p}/2$

The second experiment, i.e. extent object test, is to illustrate the responses of the spatial frequency window to the different size of objects. Some useful guidelines about how to choose the appropriate parameters for the spatial frequency window are derived further.

The third experiment, referred to as real data test, is to uncover some potential applications for the methods proposed in this paper. Two datasets are used in this experiment, which are collected by the robot in the indoor and outdoor environments respectively. The difference is that the ground truth is known precisely for the first dataset, while it is not the case of the outdoor dataset.

5.1 Point object test

In this experiment, we use the simulation data of point object to test the 1D-KST and 2D-KST. Considering the sensor noise and the imperfections in the OGM building process, we add to the OGMs some Poisson noise, which is uniformly distributed in the grid cells and whose number obeys the Poisson distribution. The main parameters in the simulation is shown in the Table 1, where the velocities V are represented in the normalized format, and the scale factor α in $i_c^{\theta_p}$ is defined as follows:

$$\alpha = \max(|\cos(\theta_p)|, |\sin(\theta_p)|) \quad (46)$$

The results of the one dimensional test are shown in Fig. 3 through 5. From Fig. 3, we can see five objects

moving at different speeds along positive direction or negative direction. The speed of the fastest one is 0.5, which is the maximal feasible velocity given by the Nyquist sampling theorem, and the speed of the slowest moving object is 0.05, which is very close to the velocity resolution given by (22). Through this setting, we can validate 1D-KST in terms of velocity measurement capability. Meanwhile, we set a stationary object at $l_0 = 20$, which can be used to check the performance on the separation of the dynamic grids from the static ones. After adding the Poisson noise, the OGMs look very noisy.

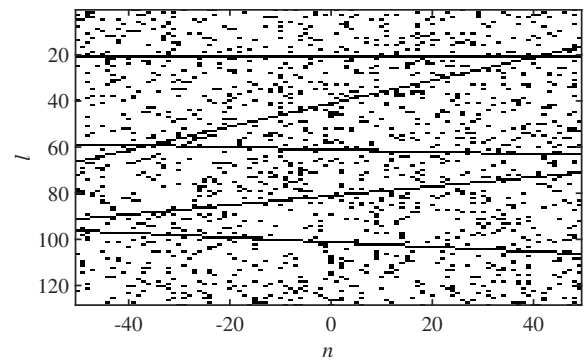


Figure 3. Sequence of the simulated one dimensional OGMs: $f(l, n)$.

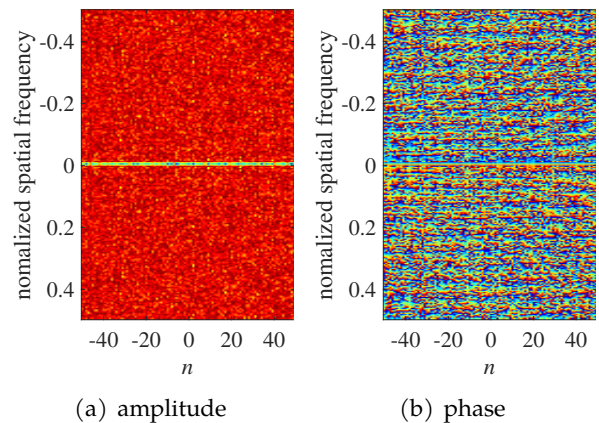


Figure 4. Result of OGM sequence by the spatial Fourier transform: $\mathcal{F}(i, n)$.

The result of OGM sequence by the spatial Fourier transform is shown in Fig. 4. As the noise in the OGMs, it is hardly to see any change information along the time axis from Fig. 4(a). However, it can still be seen from the phase of $\mathcal{F}(i, n)$ in Fig. 4(b). Moreover, it is not hardly to see the symmetry with respect to the spatial frequency axis from Fig. 4 because our OGMs are all real numbers, thus from the information point of view we can only set the spatial frequency window in one side of the spatial frequency axis.

The final result of 1D-KST is shown in Fig. 5. When

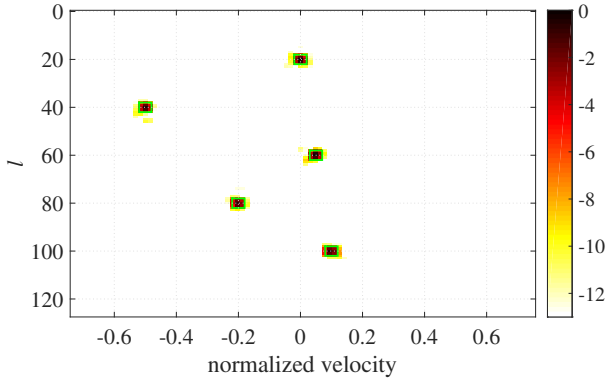


Figure 5. Result of 1D-KST: $\tilde{\mathcal{F}}(l, k)$. The color denotes the total power of the accumulated occupancies in this cell during the N time instants. The green box on every object indicates the corresponding resolution of 1D-KST for the position and the velocity.

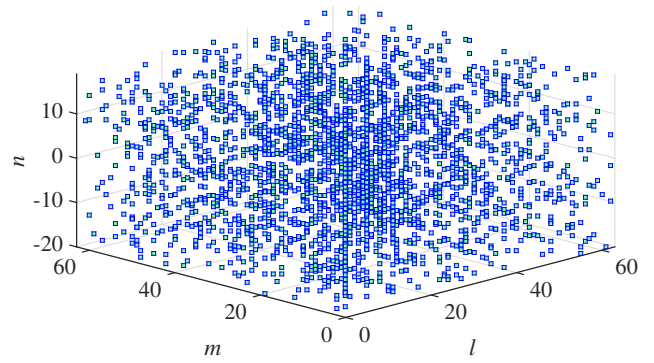


Figure 6. Sequence of the simulated two dimensional OGMs: $f(l, m, n)$.

looked together with Fig. 3, it is clearly shown that all objects are well located in the grid cells at the midpoint time instant and their velocities are also measured with a high precision even for these so noisy OGMs². Moreover, the occupiers of every are well focused in the green box determined by 1D-KST's resolution, which is determined by (22) and the width of the spatial frequency window:

$$\Delta l \cong L / (i_{\max} - i_{\min}) \quad (47)$$

From this result, we can conclude that 1D-KST has a good performance for the point object in terms of OGM filtering and the extraction of motion information.

The results of the two dimensional test are shown in Fig. 6 through 10. Similar to the one dimensional case, we choose the maximal feasible and minimal detectable speed to are set to evaluate the capability of 2D-KST in terms of velocity measurement. Different from one dimensional case, we use eight hypotheses of direction to match the possible moving directions in OGMs. Among all five moving objects, four are moving along the direction in the hypothesis sets, while the last one (#5) is moving near the middle direction between the hypothesis $\theta_p = 7\pi/8$ and the reverse direction of $\theta_p = 0$, but slightly close to $\theta_p = 7\pi/8$. Through this setting, we can evaluate the performance when the true moving direction dose not match any hypotheses. To see the result more clearly, we decrease the map size $L = 64$ in the two dimensional test. The detailed parameter setting used in this simulation can be seen in Table 1.

²If we want to obtain the OGM at any time instant n , we only need multiply $\tilde{\mathcal{F}}(i, k)$ by an item z_k^n before the processing of spatial IFFT, where z_k is defined in (40) and $n = 0, 1, \dots, N - 1$.

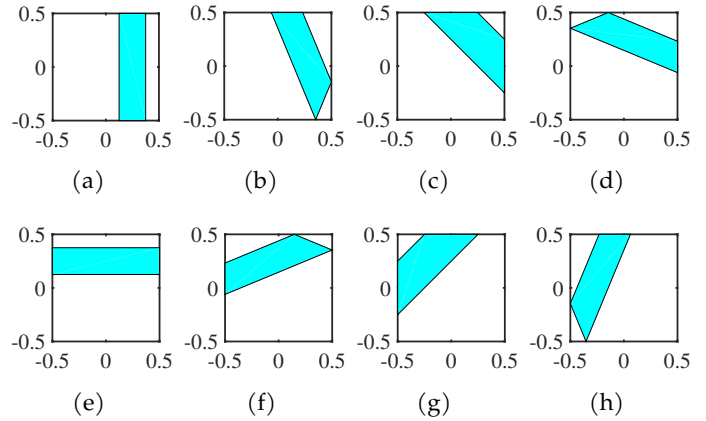


Figure 7. Spatial frequency window $W_{\theta_p}(i, j)$. (a)-(h) are the cases of $\theta_p = 0, \pi/8, \dots, 7\pi/8$, respectively. The x-ticks and y-ticks are both the normalized values of spatial frequency. The value of the point inside the colorful quadrangle is equal to 1, while zero for the outside point.

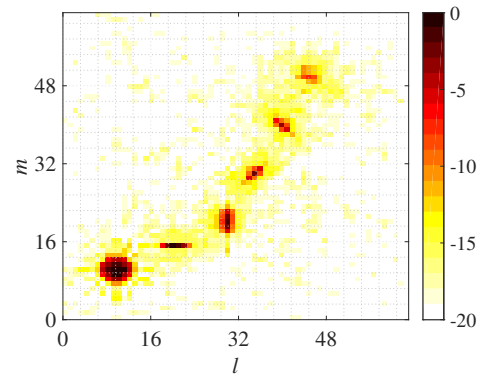


Figure 8. Result of 2D-KST after the first merging step of MPD: $\mathcal{P}(l, m)$. The color denotes the total power of the accumulated occupancies in this cell during the N time instants.

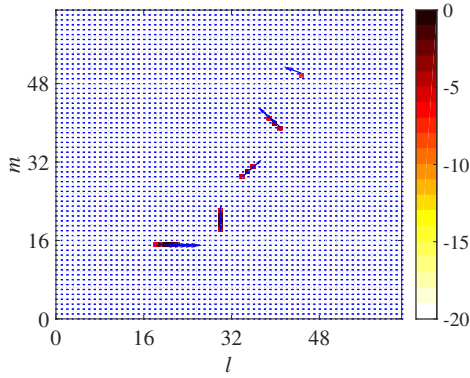


Figure 9. Result of 2D-KST after the processing of MPD. The color has the same meaning as Fig. 8. The blue arrows denote the velocities of the grid cells, whose lengths are proportional to the speed.

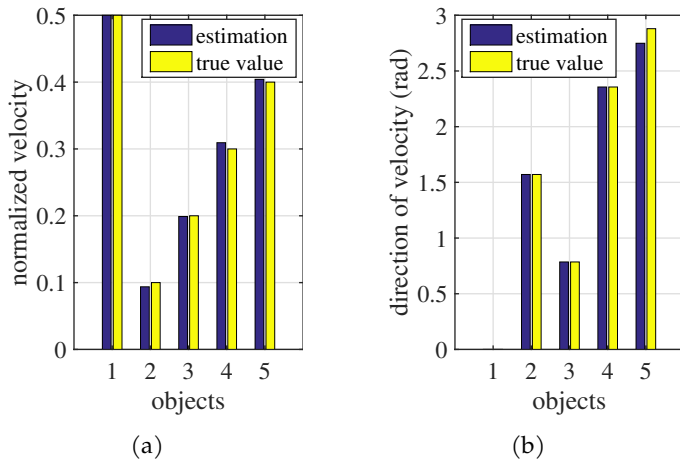


Figure 10. Comparison of the measurement and the true value of the velocity.

Fig. 6 shows the noisy 2D-OGM sequence. It is very difficult to recognize the trajectory of each object by human eyes from this figure. To make the window $W_{\theta_p}(i, j)$ more intuitive than the expression (39) to the reader, we show them in Fig. 7. In principle, the counterpart of the spatial frequency window here is the directional filter in optical signal processing, this is also the reason why we call the second step of 2D-KST as directional filtering.

After the first merging step of MPD, the result $\mathcal{P}(l, m)$ of 2D-KST processing is shown in Fig. 8, which can be regard as the filtered OGM at the time instant $n = 0$. From this figure, we can found that the noise in the original OGM is almost filtered, and the grid cells in the vicinity of objects have a significant occupancy value. The occupancies in the yellow grid cells around each object are the leakage of occupancy in the corresponding object grid cell, which is the consequence caused by the window $W_{\theta_p}(i, j)$. Among

all the six objects, the stationary object at (10, 10) behaves obviously isotropic, while the other moving objects behave the obvious directionality, which means that they have the biggest extent along their moving direction. Moreover, we can found that the grid cells near the mismatched one at (45, 50) have a relatively smaller value than those near the matched objects. Nonetheless, their values are at least twice as large as the values of those yellow grid cells. Thus, if we set the appropriate threshold P_{\min} , see (35), we can remove the leakage occupancies in the yellow cells but maintain those in the vicinity of the mismatched object grid cell. Furthermore, if our requirement is to extract the moving objects, we can filter the stationary one by setting the appropriate threshold V_{\min} , see (38).

The result of 2D-KST after MPD processing is shown in Fig. 9, where $P_{\min} = 0.3981$ (-8dB) and $V_{\min} = 0.085$ are used, and where the blue arrows represent the velocity of the occupies in the corresponding cells, the length and pointing for the amplitude and the direction respectively. As can be seen from Fig. 9, the stationary object has been removed successfully, although it has the maximal occupancy, while the five moving objects all persist in existing. What's more, the velocity measurements are highly in accordance with the ground truth in Table 1. To see it more obviously, the simple plot extractor is used, which extracts the local power maximum grid cells from the MPD results as the candidate detections and computes the occupancy weighted velocity as the velocity measurement of each candidate detection. The measurement results are shown in Fig. 10, which suggest that 2D-KST has a good precision of velocity measurement for the point object and can be easily integrated with the plot extractor or other object clustering algorithm, such as FCTA [22].

5.2 Extend object test

The purpose of this experiment is to illustrate how the parameters of the spatial frequency window affect the output for different object sizes. Since the window parameters have the same effect on the output for the 1D and 2D cases and the 1D case is more easily to explain, we test two typical windows in one dimensional case for five different object sizes firstly. Through this test, some guidelines about how to choose the parameters for the spatial frequency window are derived. Then we validate these guidelines through 2D test for different object sizes. The size of each object used in this experiment is listed in Table 2, where the value means how many grid

cells are occupied by this object. For 2D case, the two values correspond to the numbers of the occupied grid cells parallel and perpendicular to the direction of the velocity respectively. The other parameters are the same as Table 1 if without any special explanations.

Table 2. Sizes of the extent objects

object label	#0	#1	#2	#3	#4	#5
1D case	5	4	1	3	2	-
2D case	6,3	3,3	1,1	2,1	2,2	3,2

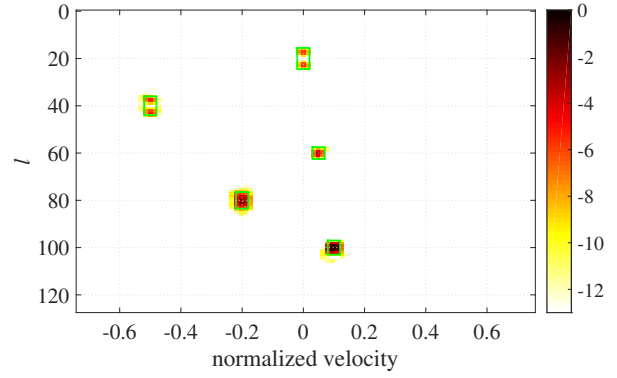
The responses to the two typical windows in one dimensional case to the five object sizes are shown in Fig. 11.

As can be seen from Fig. 11(a) and (b), the wide window has a better resolution whether for the position or for the velocity, because it has a wider window and a higher reference spatial frequency i_c . See (47) and (22) for the formulas about the KST resolution. But the large objects, object #0 at cell 20 and object #1 at cell 40 are split into two parts because the bandpass spatial filter is used in our KST approach. However, the two parts, the head and the tail, are both in the extent of the corresponding object which is represented by the green box. Fortunately, this split outcome may be accepted for many applications, such as CTM building [12] or DATMO using FCTA [22].

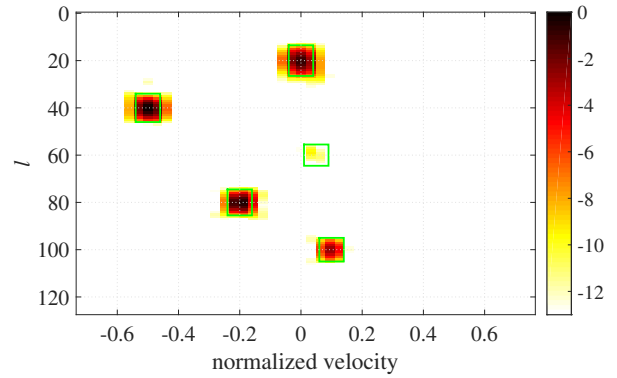
If it can not be accepted, the narrow window can be used instead. As shown in Fig. 11(b), all objects are focused in the corresponding green box. However, the velocity resolution gets worsened as the lower i_c used, and the object #2 at cell 60 has a negligible occupancy inside its green box since only very little proportion of its power can enter the narrow window.

In principle, we can get the following two empirical criteria for selecting the parameters of the spatial frequency:

1. $i_{\max} - i_{\min} \leq i_c$, which ensures that the scaled time for each spatial frequency i shown in Fig. 1 has not too big gap with the one for i_c . As a result, the average resolution of frequency for every cell i can be approximated by the one of i_c .
2. $i_{\min} \cdot E_o < L/2$, where E_o is the possible maximal size of the object along its moving direction. This criterion ensures that the object can be focused along their moving direction, because for the E_o length continuous segment in OGM, its maximal effective spatial frequency is at $L/(2E_o)$. Only if



(a) Wide window



(b) Narrow window

Figure 11. Responses of the two typical windows in one dimensional case to the different sizes of objects. (a) $i_c = L/4, i_{\min} = L/8, i_{\max} = 3L/8$; (b) $i_c = L/8, i_{\min} = L/16, i_{\max} = 3L/16$. The color denotes the total power of the accumulated occupancies in this cell during the N time instants. The length of the green box on each object along l direction denotes the sum of the position resolution of 1D-KST and the extent of this object, while the length along normalized velocity direction denotes the corresponding velocity resolution of 1D-KST.

this condition holds, the information of this object as a whole can be maintained.

In practice, we can choose the appropriate parameters of spatial frequency window to meet the above two criteria. But for those applications in which the moving objects have the significant difference in size, for example, the case of Fig. 11, we can use multiple windows to match different object sizes, or use a uniform window but based on the multiple resolution grid cell maps for different sizes of objects. Along this path, it will result in a multiple resolution KST approach, which is beyond the scope of this paper and maybe the next step work in the future.

Next we validate the wide window in 2D case, the parameters can be seen in Table 1, through those objects in Table 2, where the extensions of all

Table 3. Detections by the plot extractor

	\hat{l}, \hat{m}	\hat{V}	$\hat{\theta}$ (deg)	object label
1	(20,14)	0.5	0	
2	(20,15)	0.5	0	#1
3	(20,16)	0.5	0	
4	(30,20)	0.09	87.1	#2
5	(34,29)	0.20	45	#3
6	(36,31)	0.20	45	
7	(40,40)	0.31	135	#4
8	(47,49)	0.38	161.4	
9	(45,50)	0.35	171.6	#5
10	(43,51)	0.38	161.4	

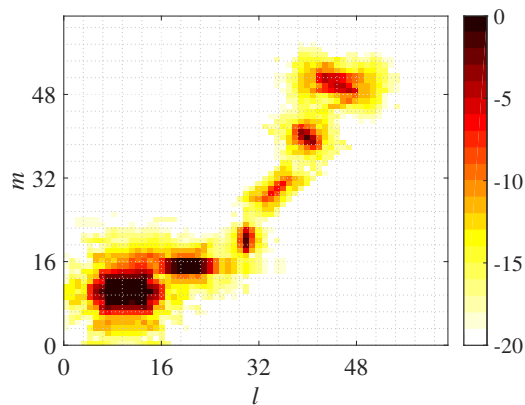
moving objects are no larger than 3 grid cells. The corresponding results are shown in Fig. 12. From these results, we can conclude that when using the appropriate setting of spatial frequency window, the 2D-KST can perform very well for the extend objects when they have no too much difference in size.

Similar to the point object test, in order to demonstrate the potential capability of integrating with FCTA or other extend target tracking algorithms, we give the result of the plot extractor, which is shown in Table 3. As can be seen from Table 3, ten detections are achieved for those five moving objects altogether, and the average position is at the center of extend object, see Table 1 for the ground truth. Moreover, the velocity measurements also have a high precision, even for the mismatch one (#5), the maximal speed error is less than 0.05 and the direction of velocity is no more than 7 degrees among those three detections generated by this object.

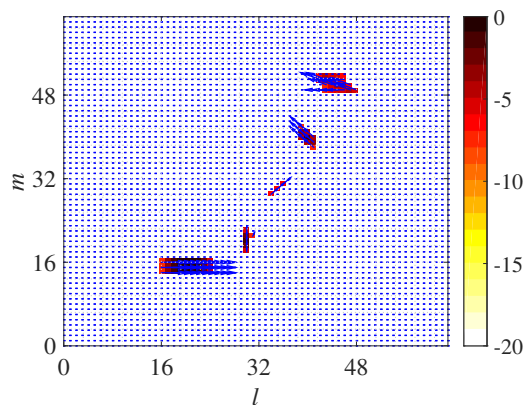
Above all, as the point object test, it seems that 2D-KST has a good precision of velocity measurement for the extend object and can be easily integrated with the plot extractor or any other extend object tracking algorithm.

6 Conclusions

This paper developed a different way of extracting motion information from successive noisy OGMs based on a signal transformation by extending the KST in the radar signal processing community to the 1D and 2D spatial case. And the fast algorithm for the 2DS-KST is also given and has the proportional computational



(a) Result after the first merging step of MPD



(b) Result after the MPD processing

Figure 12. Result of 2D-KST for extend object. The color denotes the total power of the accumulated occupancies in this cell during the N time instants. The blue arrows denote the velocities of the grid cells, whose lengths are proportional to the speed.

complexity with the 2D-FFT. Simulation results show that our method can extract the sub-pixel motions effectively from the sequence of very noisy OGMs, which has a wide use, in many application scenarios, such as the industrial field, airport and other indoor environment.

Further evaluation by real data, multi-resolution KST for complex scenarios, integration with Bayesian Occupancy Filter and CTMAP, and hypothesis merging method based on more complex velocity model, are worth being paid attentions in the next step.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (62303478), ATR Foundation (2035250204) and Key Lab. Foundation (220302).

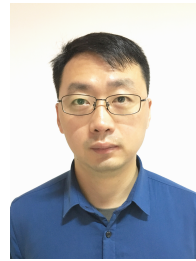
References

- [1] Coué, C., Pradalier, C., Laugier, C., Fraichard, T., & Bessière, P. (2006). Bayesian occupancy filtering for multitarget tracking: an automotive application. *The International Journal of Robotics Research*, 25(1), 19-30. [CrossRef]
- [2] Tay, M. K., Mekhnacha, K., Yguel, M., Coue, C., Pradalier, C., Laugier, C., ... & Bessiere, P. (2008). The Bayesian occupation filter. In *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems* (pp. 77-98). Berlin, Heidelberg: Springer Berlin Heidelberg. [CrossRef]
- [3] Kondaxakis, P., Kasderidis, S., & Trahanias, P. (2008). Tracking multiple targets from a mobile robot platform using a laser range scanner. In *2008 IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications* (pp. 175-184). [CrossRef]
- [4] Vu, T. D., Bulet, J., & Aycard, O. (2008). Mapping of environment, detection and tracking of moving objects using occupancy grids. In *Intelligent Vehicles Symposium* (pp. 684-689). Los Alamitos: IEEE.
- [5] Gindele, T., Brechtel, S., Schroder, J., & Dillmann, R. (2009, June). Bayesian occupancy grid filter for dynamic environments using prior map knowledge. In *2009 IEEE Intelligent Vehicles Symposium* (pp. 669-676). IEEE. [CrossRef]
- [6] Baig, Q., Perrollaz, M., & Laugier, C. (2014). Advances in the Bayesian Occupancy Filter framework using robust motion detection technique for dynamic environment monitoring. *IEEE Robotics and Automation Magazine*.
- [7] Saarinen, J. P., Andreasson, H., Stoyanov, T., & Lilienthal, A. J. (2013). 3D normal distributions transform occupancy maps: An efficient representation for mapping in dynamic environments. *The International Journal of Robotics Research*, 32(14), 1627-1644. [CrossRef]
- [8] Mosberger, R., Andreasson, H., & Lilienthal, A. J. (2014). A customized vision system for tracking humans wearing reflective safety clothing from industrial vehicles and machinery. *Sensors*, 14(10), 17952-17980. [CrossRef]
- [9] Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping: part I. *IEEE robotics & automation magazine*, 13(2), 99-110. [CrossRef]
- [10] Bailey, T., & Durrant-Whyte, H. (2006). Simultaneous localization and mapping (SLAM): Part II. *IEEE robotics & automation magazine*, 13(3), 108-117. [CrossRef]
- [11] Petrovskaya, A., Perrollaz, M., Oliveira, L., Spinello, L., Triebel, R., Makris, A., ... & Bessière, P. (2012). Awareness of road scene participants for autonomous driving. *Handbook of Intelligent Vehicles*, 1383-1432.
- [12] Kucner, T., Saarinen, J., Magnusson, M., & Lilienthal, A. J. (2013, November). Conditional transition maps: Learning motion patterns in dynamic environments. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1196-1201). IEEE. [CrossRef]
- [13] Wang, C. C., Thorpe, C., Thrun, S., Hebert, M., & Durrant-Whyte, H. (2007). Simultaneous localization, mapping and moving object tracking. *The International Journal of Robotics Research*, 26(9), 889-916. [CrossRef]
- [14] Ess, A., Leibe, B., Schindler, K., & Van Gool, L. (2009, May). Moving obstacle detection in highly dynamic scenes. In *2009 IEEE International Conference on Robotics and Automation* (pp. 56-63). IEEE. [CrossRef]
- [15] Baig, Q., Perrollaz, M., Do Nascimento, J. B., & Laugier, C. (2012, December). Using fast classification of static and dynamic environment for improving Bayesian occupancy filter (BOF) and tracking. In *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)* (pp. 656-661). IEEE. [CrossRef]
- [16] Hahnel, D., Triebel, R., Burgard, W., & Thrun, S. (2003, September). Map building with mobile robots in dynamic environments. In *2003 IEEE International Conference on Robotics and Automation* (Cat. No. 03CH37422) (Vol. 2, pp. 1557-1563). IEEE. [CrossRef]
- [17] Elfes, A. (1989). A probabilistic framework for robot perception and navigation. *PhD thesis, Carnegie-Mellon University*.
- [18] Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6), 46-57. [CrossRef]
- [19] Pancham, A., Tlale, N., & Bright, G. (2011). Literature review of SLAM and DATMO. *RobMech* 2011. [CrossRef]
- [20] Baig, Q., & Aycard, O. (2012, June). Improving moving objects tracking using road model for laser data. In *2012 IEEE Intelligent Vehicles Symposium* (pp. 790-795). IEEE. [CrossRef]
- [21] Chen, C., Tay, C., Laugier, C., & Mekhnacha, K. (2006, December). Dynamic environment modeling with gridmap: a multiple-object tracking application. In *2006 9th International Conference on Control, Automation, Robotics and Vision* (pp. 1-6). IEEE. [CrossRef]
- [22] Mekhnacha, K., Mao, Y., Raulo, D., & Laugier, C. (2008, September). Bayesian occupancy filter based" fast clustering-tracking" algorithm. In *IROS 2008*.
- [23] Tay, M. K., Mekhnacha, K., Chen, C., Yguel, M., & Laugier, C. (2008). An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments. *International Journal of Vehicle Autonomous Systems*, 6(1-2), 155-171. [CrossRef]
- [24] Fan, H., Kucner, T. P., Magnusson, M., Li, T., &

- Lilienthal, A. J. (2017). A dual PHD filter for effective occupancy filtering in a highly dynamic environment. *IEEE Transactions on Intelligent Transportation Systems*, 19(9), 2977-2993. [CrossRef]
- [25] Zhu, D., Li, Y., & Zhu, Z. (2007). A keystone transform without interpolation for SAR ground moving-target imaging. *IEEE Geoscience and Remote Sensing Letters*, 4(1), 18-22. [CrossRef]
- [26] Zhao, Y., Wang, J., Huang, L., & Yang, R. (2011, October). Low complexity keystone transform without interpolation for dim moving target detection. In *Proceedings of 2011 IEEE CIE International Conference on Radar* (Vol. 2, pp. 1745-1748). IEEE. [CrossRef]
- [27] Huang, P., Liao, G., Yang, Z., Xia, X. G., Ma, J., & Zheng, J. (2016). Ground maneuvering target imaging and high-order motion parameter estimation based on second-order keystone and generalized Hough-HAF transform. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1), 320-335. [CrossRef]
- [28] Huang, P., Liao, G., Yang, Z., Xia, X. G., Ma, J. T., & Ma, J. (2016). Long-time coherent integration for weak maneuvering target detection and high-order motion parameter estimation based on keystone transform. *IEEE Transactions on Signal Processing*, 64(15), 4013-4026. [CrossRef]
- [29] Zhan, M., Huang, P., Zhu, S., Liu, X., Liao, G., Sheng, J., & Li, S. (2021). A modified keystone transform matched filtering method for space-moving target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-16. [CrossRef]
- [30] Fan, H., Lu, D., Kucner, T. P., Magnusson, M., & Lilienthal, A. (2018, July). 2D spatial keystone transform for sub-pixel motion extraction from noisy occupancy grid map. In *2018 21st International Conference on Information Fusion (FUSION)* (pp. 1-7). IEEE. [CrossRef]
- [31] Richards, M. A. (2014). The keystone transformation for correcting range migration in range-doppler processing. *pulse*, 1000(1).
- [32] Rabiner, L. R., Schafer, R. W., & Rader, C. M. (1969). The chirp z-transform algorithm and its application. *Bell System Technical Journal*, 48(5), 1249-1292. [CrossRef]
- [33] Rajmic, P., Prusa, Z., & Wiesmeyer, C. (2014, September). Computational cost of Chirp Z-transform and Generalized Goertzel algorithm. In *2014 22nd European Signal Processing Conference (EUSIPCO)* (pp. 1004-1008). IEEE.



Hongqi Fan (S'08-M'09-SM'24, IEEE) received his B.S. degree in mechanical engineering and automation from Tsinghua University, Beijing, P.R. China, in 2001, and his Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, Hunan, P. R. China, in 2008. He is currently a professor at the National University of Defense Technology. His research interests include information fusion, target tracking, signal processing, Intelligent Guidance Systems. (E-mail: fanhongqi@nudt.edu.cn)



Dawei Lu (corresponding author) received the M.S. degree information and communication engineering from the National University of Defense Technology, Changsha, Hunan, P. R. China, in 2005. He is currently a lecturer at the National University of Defense Technology. His research interests include Radar system modeling and simulation, Extended target Tracking, Random Finite Set Estimation and Information Fusion, Intelligent Guidance Systems. (Email: davidloo.nudt@gmail.com)



Yanwen Jiang was born in Hengyang, Hunan, China, in 1991. She received the M.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology (NUDT) in 2014 and 2018 respectively. She is currently an Associate Professor with NUDT. Her research interests include array radar imaging and signal processing. (Email: j1991yuwen@163.com)



Achim J. Lilienthal (Senior Member, IEEE) received the Ph.D. degree in computer science from Tübingen University, Tübingen, Germany, in 2005. He is currently a professor of Computer Science at TU Munich, Germany, where he leads the chair "Perception for Intelligent Systems". He is coordinator of the Horizon 2020 project DARKO, manager of the project "KI.Fabrik" at the German Technology Museum in Munich and guest professor at the University of Örebro, Sweden, where he established the Mobile Robotics and Olfaction lab. His core research interest is in perception for intelligent systems. Typically based on approaches that leverage domain knowledge and AI. He has authored or coauthored more than 250 refereed conference papers and journal articles and evaluator for several national funding agencies and the EU. his research work addresses mobile robot olfaction, rich 3-D perception, navigation of autonomous transport robots, human-robot interaction, and mathematics education research. (Email: achim.j.lilienthal@tum.de)