

Article

Multi-Target Tracking Using Windowed Fourier Single-Pixel Imaging

Jinyu Zhang ¹, Taiyang Hu ^{1,*}, Xiaolang Shao ¹, Mengxuan Xiao ¹, Yingjiao Rong ² and Zelong Xiao ¹

¹ School of Electronic Engineering and Optical Technology, Nanjing University of Science and Technology, Nanjing 210094, China; zhangjinyu@njust.edu.cn (J.Z.); xl_shao1994@126.com (X.S.); MengxuanXiao@njust.edu.cn (M.X.); zelongxiao@njust.edu.cn (Z.X.)

² Science and Technology on Near-Surface Detection Laboratory, Wuxi 214035, China; enjoy_rong@163.com

* Correspondence: sun1983hu@126.com

Abstract: The single-pixel imaging (SPI) technique enables the tracking of moving targets at a high frame rate. However, when extended to the problem of multi-target tracking, there is no effective solution using SPI yet. Thus, a multi-target tracking method using windowed Fourier single-pixel imaging (WFSI) is proposed in this paper. The WFSI technique uses a series of windowed Fourier basis patterns to illuminate the target. This method can estimate the displacements of K independently moving targets by implementing $6K$ measurements and calculating $2K$ windowed Fourier coefficients, which is a measurement method with low redundancy. To enhance the capability of the proposed method, we propose a joint estimation approach for multi-target displacement, which solves the problem where different targets in close proximity cannot be distinguished. Using the independent and joint estimation approaches, multi-target tracking can be implemented with WFSI. The accuracy of the proposed multi-target tracking method is verified by numerical simulation to be less than 2 pixels. The tracking effectiveness is analyzed by a video experiment. This method provides, for the first time, an effective idea of multi-target tracking using SPI.



Citation: Zhang, J.; Hu, T.; Shao, X.; Xiao, M.; Rong, Y.; Xiao, Z. Multi-Target Tracking Using Windowed Fourier Single-Pixel Imaging. *Sensors* **2021**, *21*, 7934. <https://doi.org/10.3390/s21237934>

Academic Editor: Ming-Jie Sun

Received: 31 October 2021

Accepted: 25 November 2021

Published: 28 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: single-pixel imaging; multi-target tracking; windowed Fourier transform; fast moving object tracking

1. Introduction

Multi-target detection and tracking have widespread applications in military defense [1,2], area surveillance [3,4], UAV detection [5,6], and other fields. Compared with single target tracking, multi-target tracking is more widely used and more difficult to realize. There are two types of conventional multi-target tracking methods. The first one is an image-free method, such as laser radar and millimeter radar, which emits electromagnetic waves with different beam steering to illuminate the target and receives the back-scattered echoes to estimate the target's position and velocity. The multi-target tracking capability of the radar relies on the construction of multiple beams, and the antenna system is more difficult to design. Moreover the tracking systems usually require multi-channel compounding that is complex and costly. The other one is an image-based method, which estimates the displacements of the moving objects by acquiring successive reconstructed images and analyzes the trajectory of the moving object from a few frames [7,8]. The block matching method is a typical multi-target tracking method [9,10]. Compared to radar systems, image-based target tracking systems are less expensive. However, high-speed cameras are costly, and the data throughput is usually huge in a short amount of time. In addition, high-speed imaging devices in the millimeter waveband are more challenging to be designed.

With the development of optical digital modulating devices (DMDs) and millimeter-wave metasurface materials, the single-pixel imaging (SPI) technique, which can be used to develop an imaging system, has attracted the attention of many scholars in recent years [11–16]. The SPI technique uses different modulated patterns to illuminate the

target and recovers the target image by a computational method. It can be applied to develop imaging systems with a low system complexity in, e.g., millimeter and terahertz wavebands [17–20]. Due to the flexible and broadband system composition of SPI, it is also promising for the detection and tracking of moving objects. In addition, the accuracy of an SPI imaging system is affected by the motion blur of the target. It has been shown that, if the target motion parameter can be estimated accurately, it can also be of great value for the imaging calibration of moving targets [14]. The studies on the use of SPI for imaging and tracking of moving targets are as follows.

Jiao et al. proposed a motion estimation and image quality enhancement scheme for a single reconstructed image [21]. The optimal motion parameters and reconstructed images can be searched based on the connection between the image reconstruction variance and the accuracy of the estimated motion parameters. By performing motion estimation between subsequent frames, another estimation method proposed by S. Monin et al. may be applied to any type of trajectory [22]. Their estimation methods based on one or a few image frames are time-consuming. Moreover, the moving target estimated by their methods is a single target.

The measurement data used for imaging is often redundant for the estimation of motion parameters. From this perspective, scholars have tried to reduce the number of measurements to increase the frame rate of target tracking. Shi et al. proposed a method to obtain the position of the object in the scene from the 1D projected curves [23]. This method demonstrates that reducing redundancy in data acquisition is an effective approach for the fast tracking method. In [24,25], the method proposed by Zhang et al., based on Fourier single-pixel imaging (FSI), uses six Fourier basis patterns for structured light modulation to measure only two Fourier coefficients to estimate the displacement of a moving object. Recently, a faster method was proposed by Zha and Shi et al.; it requires only three geometric moment patterns to illuminate a moving object in one frame [26]. These works fully demonstrate that the trajectory of a moving target in a scene can be estimated with low redundancy using a small amount of measurement data. However, the problem of tracking multiple targets remains unsolved.

Thus, a multi-target tracking method using windowed Fourier single-pixel imaging (WFSI) is proposed in this paper. Inspired by the successful application of the short-time Fourier transform (STFT) in signal processing, we designed a series of window functions that vary with the target positions. The WFSI technique uses windowed Fourier basis patterns to illuminate the target. Combining these different window functions, the problem of tracking multiple targets can be transformed into the problem of tracking a single target. One window function is designed for each target, and two windowed Fourier coefficients are calculated by six measurements to complete the estimation of the displacement of this target. For the K target, $6K$ measurements are required. In addition, considering the scene that two targets are close to each other, a multi-target joint estimation approach is proposed to enhance the capability of the proposed tracking method. In addition, a joint multi-target estimation approach is proposed to enhance the capability of the proposed tracking method in consideration of the two targets in close proximity. Numerical simulation experiments are conducted to verify the multi-target tracking capability of the proposed method. We also analyzed the effectiveness of multi-target tracking through a video experiment. Consequently, the proposed method enables multi-target tracking in real-time and for a long duration.

2. Principle and Method

2.1. Single-Target Tracking Method Using FSI

The FSI technique can reconstruct high-quality images by acquiring their Fourier spectrum [27]. The tracking of a single target can be estimated by the FSI technique by measuring 2 Fourier coefficients [24]. FSI employs a series of Fourier basis patterns to illuminate the target, and each Fourier basis pattern is characterized by its spatial frequency pair (f_x, f_y) and its initial phase ϕ :

$$P(x, y, \phi) = A + B \cdot \cos(2\pi f_x x + 2\pi f_y y + \phi) \quad (1)$$

where (x, y) denotes 2-D coordinates in the spatial domain, A is the average intensity of the pattern, and B denotes the contrast. The Fourier basis patterns can be converted from grayscale to binary by employing image upsampling and the Floyd–Steinberg error diffusion dithering method [28]. Using the binary Fourier basis pattern to illuminate the target image $I(x, y)$, the total light intensity D is equivalent to an inner product of the target image $I(x, y)$ and the binary Fourier basis pattern $P_{bin}(x, y)$:

$$D(f_x, f_y, \phi) = \iint I(x, y) P_{bin}(x, y, \phi) dx dy \quad (2)$$

The Fourier coefficients of the target image can be acquired using the three-step phase-shifting algorithm [24]:

$$F_I(f_x, f_y) = \left[2D(f_x, f_y, 0) - D(f_x, f_y, \frac{2\pi}{3}) - D(f_x, f_y, \frac{4\pi}{3}) \right] + \sqrt{3}j \cdot \left[D(f_x, f_y, \frac{2\pi}{3}) - D(f_x, f_y, \frac{4\pi}{3}) \right] \quad (3)$$

The integral in Equations (2) and (3) implies that the Fourier transform is a global-to-point transformation. The change of the target's position in the spatial domain leads to the change of the entire Fourier coefficients in the spatial frequency domain. Therefore, the presence of a target in space can be determined by detecting one or a few Fourier coefficients in the spatial frequency domain.

The scene image containing a moving target can be represented as the sum of the target image and the background image:

$$I_{scene}(x, y; t) = I_{target}(x, y; t) + I_{background}(x, y) \quad (4)$$

The Fourier spectrum of the target image is equal to the difference between the Fourier spectrum of the scene image and the background image, so background subtraction can be applied to obtain the Fourier spectrum of the target:

$$\begin{aligned} \mathcal{F}\{I_{target}(x, y; t)\} &= \mathcal{F}\{I_{scene}(x, y; t) - I_{background}(x, y)\} \\ &= \mathcal{F}\{I_{scene}(x, y; t)\} - \mathcal{F}\{I_{background}(x, y)\} \end{aligned} \quad (5)$$

where $\mathcal{F}\{\}$ denotes the Fourier transform. For the sake of convenience of expression, if not specifically stated, $I(x, y)$ denotes the target image after background subtraction. For a moving target, the displacement $(\Delta x, \Delta y)$ from t_1 to t_2 in the spatial domain results in a phase shift in the spatial frequency domain:

$$I(x, y; t = t_2) = I(x + \Delta x, y + \Delta y; t = t_1) \quad (6)$$

$$F_I(f_x, f_y; t = t_2) = F_I(f_x, f_y; t = t_1) \cdot e^{-j2\pi(f_x \Delta x + f_y \Delta y)} \quad (7)$$

where $F_I(f_x, f_y)$ denotes the Fourier coefficients of the image $I(x, y)$. According to Equation (12), it can be found that the displacement of the target $(\Delta x, \Delta y)$ can be calculated from the phase term $\Delta\phi = -j2\pi(f_x \Delta x + f_y \Delta y)$. In order to solve for Δx and Δy separately, a system of equations needs to be established using at least two different pairs of spatial frequency (f_x, f_y) . For simplicity, the spatial frequency pairs, $(f_x, 0)$ and $(0, f_y)$ can be used. Therefore, the displacement $(\Delta x, \Delta y)$ can be derived through the following:

$$\Delta x = -\frac{1}{2\pi f_x} \cdot \arg\left\{ \frac{F_I(f_x, 0; t = t_2)}{F_I(f_x, 0; t = t_1)} \right\} \quad (8)$$

$$\Delta y = -\frac{1}{2\pi f_y} \cdot \arg\left\{ \frac{F_I(0, f_y; t = t_2)}{F_I(0, f_y; t = t_1)} \right\} \quad (9)$$

where $\arg\{\}$ denotes the argument operation.

2.2. Multi-Target Tracking Method Using WFSI

It was demonstrated that only 2 Fourier coefficients measurements (6 Fourier basis patterns) are needed to achieve the detection and tracking of a moving target, in the above-mentioned section. However, when there are multiple moving targets in the scene, it is incapable of distinguishing different targets in the spatial frequency domain due to the global-to-point property of the Fourier transform.

Inspired by the STFT, we tried to design different window functions in the spatial domain to transform the multi-target tracking into single-target tracking. If there are K targets, the scene $I(x, y)$ in Equation (2) needs to be modified as the sum of multi-target images $I_i(x, y)$:

$$I(x, y; t) = \sum_{i=1}^K I_i(x, y; t) \quad (10)$$

The spatial window functions $w_i(x, y; t)$ are designed to satisfy the following:

$$I_i(x, y; t) = I(x, y; t) \cdot w_i(x, y; t) \quad (11)$$

Based on the WFSI technique, firstly we design appropriate window functions to transform the multi-target tracking problem into a single-target tracking problem. Afterwards, $6K$ measurements are implemented to calculate $2K$ Fourier coefficients. After that, using the multi-target tracking method including the independent approach and joint approach, the displacements of the multi-target can be estimated. Lastly, according to the estimated result, the window functions are redesigned to continue the multi-target tracking. The sketch of multi-target tracking using WFSI is shown in Figure 1.

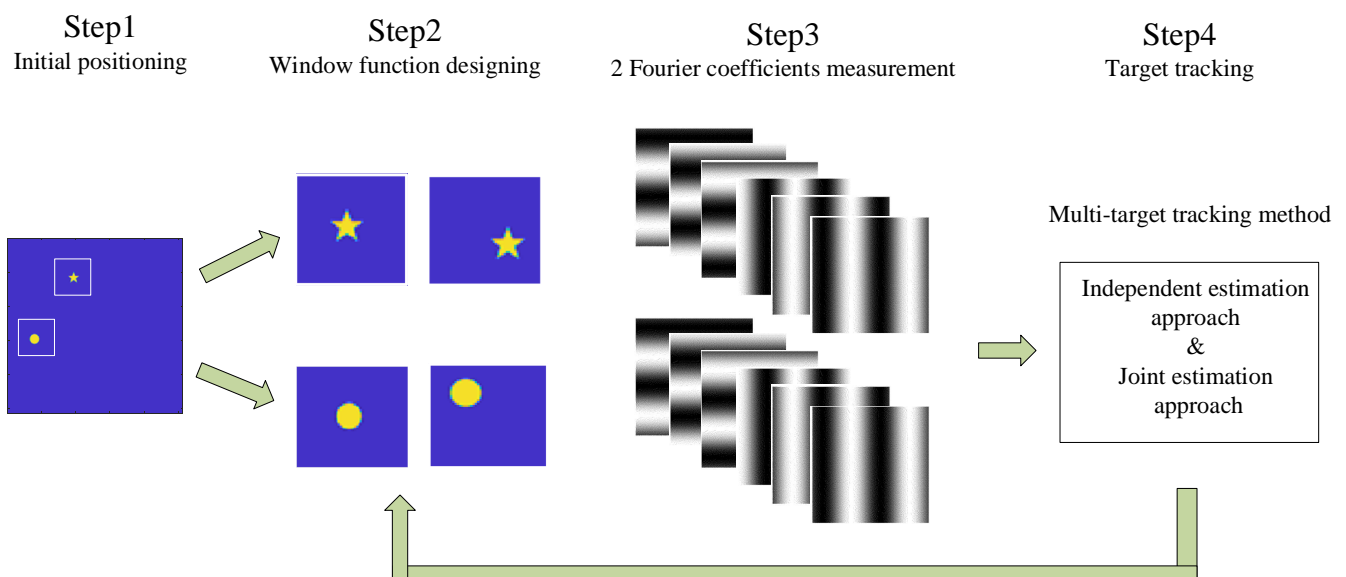


Figure 1. Sketch of multi-target tracking using WFSI. The process is as follows: Firstly, locate the initial multi-target positions. Secondly, design the window functions according to the positions. Thirdly, implement $6K$ measurements to calculate $2K$ Fourier coefficients. After that, using the multi-target tracking method with the independent approach and joint approach, the displacements can be estimated. Then, according to the estimated result, redesign the window functions to continue multi-target tracking.

Two moving targets are illustrated as an example, as shown in Figure 2. The window function is binary that can be regular or special in shape. In this paper, a rectangular window function is selected, the center of the rectangular window is the target center location, and the window length is taken as four times the target size. There are many methods to determine the approximate area of the target center position. We applied the projection curve method proposed in [23] to determine the positions of the targets. This

projection curve method constructs modulation information that satisfies the projection conditions and can transform 2D images into 1D projection curves. The 1D projection curves, which provide the location information of the moving object, can be obtained with a high resolution in real time. In this paper, the scene image is projected on the X axis and Y axis, respectively. For an image with 512×512 pixels, a compressed sensing method is used to select $2 \times 0.3 \times 512$ patterns to obtain the two projection curves. From these two curves, it is easily determined that there are four approximate regions for these two targets $\{(x_1, y_1), (x_2, y_2), (x_1, y_2), \text{ and } (x_2, y_1)\}$. Among these four regions, targets existed in only two of them. The Fourier coefficients for the image of regions where no targets exist are approximately zero. For K moving targets, using the projected curve method, less than K^2 possible regions are obtained. Therefore, the multi-target position can be determined according to the $2K$ Fourier coefficients of the target image after weighting the K window functions. In fact, only one Fourier coefficient is needed to determine the presence of a target in the region. Here, we chose 2 Fourier coefficients because the later tracking methods require 2 Fourier coefficients. It should be noted that the proposed method cannot distinguish between two targets if their spatial locations overlap with each other.

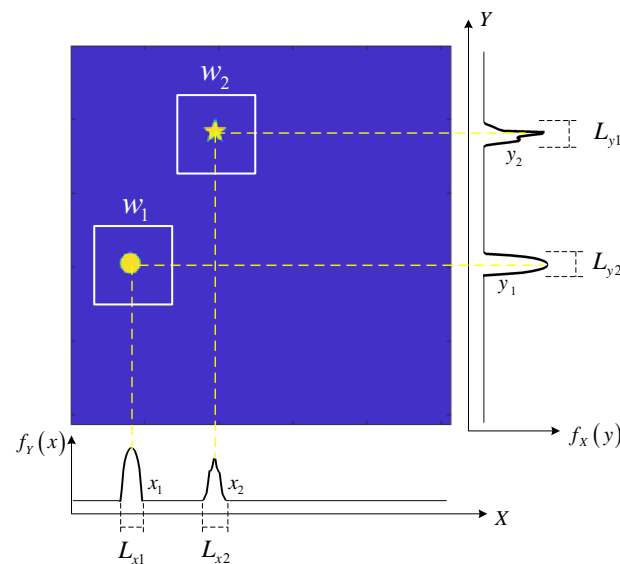


Figure 2. Illustration of multi-target locating via the projected curves method. $f_Y(x)$ and $f_X(y)$ are the projection of the scene image on the X axis and the Y axis. (x_1, x_2) and (y_1, y_2) are the regions where targets exist on the X axis and the Y axis. $(L_{x1}, L_{x2}, L_{y1}, L_{y2})$ are the length of the regions. (w_1, w_2) represents the window functions designed for the two targets.

2.3. Independent Estimation Approach

After designing the spatial window functions, the Fourier coefficients of each target image $I_i(x, y)$ can be denoted as

$$F_{I_i}(f_x, f_y; t = t_2) = F_{I_i}(f_x, f_y; t = t_1) \cdot e^{-j2\pi(f_x \Delta x_i + f_y \Delta y_i)} \quad (12)$$

where $(\Delta x_i, \Delta y_i)$ denote the displacements of each target from t_1 to t_2 . According to Equations (2) and (11), the light intensity $D_i(f_x, f_y, \phi)$ is the integral value of the inner product of the entire image $I(x, y)$, the spatial window function $w_i(x, y)$, and the binary Fourier basis pattern $P_{bin}(x, y)$. The sampling matrix designed by the DMD device is the product of $w_i(x, y)$ and $P_{bin}(x, y)$. The WFSI technique uses windowed Fourier basis patterns to illuminate the target, which can be expressed as Equation (13).

$$D_i(f_x, f_y, \phi) = \iint I(x, y) \cdot w_i(x, y) \cdot P_{bin}(x, y, \phi) dx dy \quad (13)$$

The windowed Fourier coefficients of each target $I_i(x, y)$ can be calculated by applying the three-step phase-shifting algorithm according to Equation (3). Thus, the displacement $(\Delta x_i, \Delta y_i)$ of each target can be obtained by calculating the phase term $\Delta\varphi_i = -j2\pi(f_x\Delta x_i + f_y\Delta y_i)$ according to Equation (12).

The window function $w_2(x, y)$ in Figure 2 is shown again in Figure 3b. The 3 binary Fourier patterns $(P_{bin}(\frac{2}{N}, 0, 0), P_{bin}(\frac{2}{N}, 0, \frac{2\pi}{3}), P_{bin}(\frac{2}{N}, 0, \frac{4\pi}{3}))$ used to calculate the Fourier coefficient $F_{I_2}(\frac{2}{N}, 0)$ are shown in Figure 3c–e.

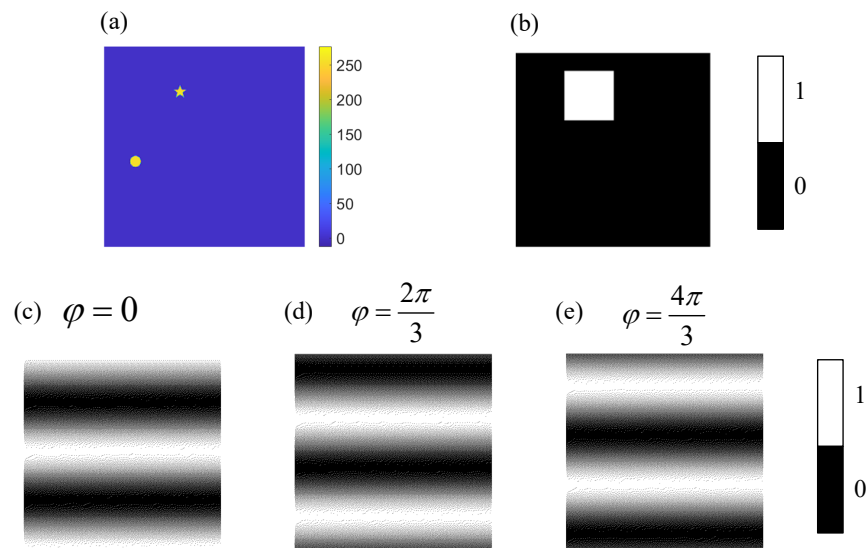


Figure 3. Illustrations of the measurement process. (a) is the target image with 256×256 pixels, (b) is the designed window function with 256×256 pixels. (c–e) are the binary Fourier patterns $P_{bin}(\frac{2}{N}, 0, \varphi)$ with 512×512 pixels used to calculate the Fourier coefficient $F_{I_2}(\frac{2}{N}, 0)$.

The multi-target tracking method described above is limited to an assumption that the target moves within the region of the spatial window function, but this assumption is not easily satisfied over a long duration. Updating the center positions of the window functions corresponding to different targets in real time during the measuring process is one useful approach. Therefore, a strategy to dynamically alter the window position is used in this paper. The center of the i th window function \bar{w}_i changes according to the estimation result of the displacement $(\Delta x_i, \Delta y_i)$ of the i th target each time, which is denoted as

$$\begin{aligned}\bar{w}_i^{(k+1)} &= (x_i^{(k)}, y_i^{(k)}) \\ &= (x_i^{(k-1)} + \Delta x_i^{(k-1)}, y_i^{(k-1)} + \Delta y_i^{(k-1)})\end{aligned}\quad (14)$$

where the superscript k represents the k th measurement.

In this way, if the displacement of a target is estimated accurately and the window length is bigger than the displacement between 2 frames, the target will always be within the window. Therefore, the upper limit of the target velocity that can be estimated by the proposed method is determined by the selected window length. The window length should not be too large. An excessively long window length can lead to overlapping window functions, and the estimated result will produce errors.

2.4. Joint Estimation Approach

When the distance between two targets in the scene is less than the length of the designed rectangular window, the multi-target tracking cannot be transformed into single-target tracking. At this point, the choice of window shape becomes very important if we want to distinguish two targets by the window function. However, the designing of special-shaped windows is a more difficult problem. Thus, a joint estimation approach is

proposed to solve the problem of target discrimination when different targets are within one window. Due to the linear feature of the Fourier transform, the Fourier spectrum of M targets within a spatial region $W(x, y)$ is equal to the sum of the Fourier spectrum of M single target $T_i (i = 1, 2, \dots, M)$, which can be denoted as

$$\mathcal{F}\{I(x, y)W(x, y)\} = \mathcal{F}\left\{\sum_{T_i \in W} I_i(x, y)\right\} \quad (15)$$

which means

$$\begin{aligned} F_{IW}(f_x, f_y; t = t_2) &= \sum_{T_i \in W} F_{I_i}(f_x, f_y; t = t_2) \\ &= \sum_{T_i \in W} F_{I_i}(f_x, f_y; t = t_1) \cdot e^{-j2\pi(f_x \Delta x_i + f_y \Delta y_i)} \end{aligned} \quad (16)$$

In the case where the Fourier spectrum of every target has been determined, there are $2M$ unknown variables in Equation (16). Therefore, the displacements $(\Delta x_i, \Delta y_i)$ of these M targets can be calculated by solving a system of equations corresponding to $2M$ different spatial frequencies (f_x, f_y) . Equation (16) is a nonlinear system of equations, and the computational difficulty of calculating the roots increases rapidly with the increase of M . There are two types of methods for solving nonlinear equations: the analytical method and the iterative search method. The analytical method is faster than the search method and it is generally accepted that nonlinear equations have an analytic solution when $M < 5$. Considering the solution complexity and real-time requirement, the joint estimation of more than 5 targets is hard to achieve. In this paper, we use the analytical method to solve 2 targets within one window.

3. Results

3.1. Simulation

Numerical simulations are employed to study the proposed method. The scene for the simulation is shown in Figure 4, where (a) represents the background image set as a town's sky, and (b,c) are the scene images at moments t_1 and t_2 , respectively. There are four moving targets in the scene, denoted as T_1, T_2, T_3, T_4 . The shape of T_1 and T_2 is a circle, and the shape of T_3 and T_4 is a pentagram. The scene image is 512×512 pixels in size, and the target position at the moment t_1 is set to $T_1(t_1) = (20, 20), T_2(t_1) = (20, 200), T_3(t_1) = (100, 200), T_4(t_1) = (200, 300)$. We assume the moment t_1 as the initial moment of discovering the targets and estimate the displacement $(\Delta x_i, \Delta y_i)$ of each target from moment t_1 to t_2 . The displacements of the four targets are preset to $(5, 5), (20, 0), (-10, 20), (20, -10)$. A Gaussian white noise with $(\mu = 0, \sigma^2 = 0.001)$ is added to the scene image.

3.1.1. Multi-Target Locating

The first experiment located the targets in the scene at moment t_1 , as shown in Figure 4b. The projected curves of the entire target image $I(x, y; t_1)$ under $2 \times 0.25 \times 512$ modulated patterns are shown in Figure 5. These illumination patterns are constructed from a Hadamard matrix according to the method in [23]. The reconstruction method of projected curves is not described in the specific process here.

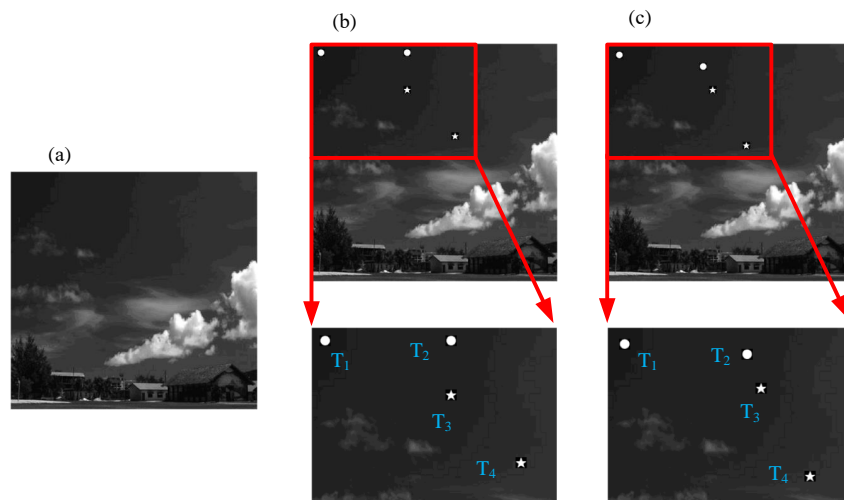


Figure 4. Simulation scenes: (a) a complex background, (b) a scene image at moment t_1 , and (c) a scene image at moment t_2 . There are four moving targets in the scene, denoted as T_1, T_2, T_3, T_4 .

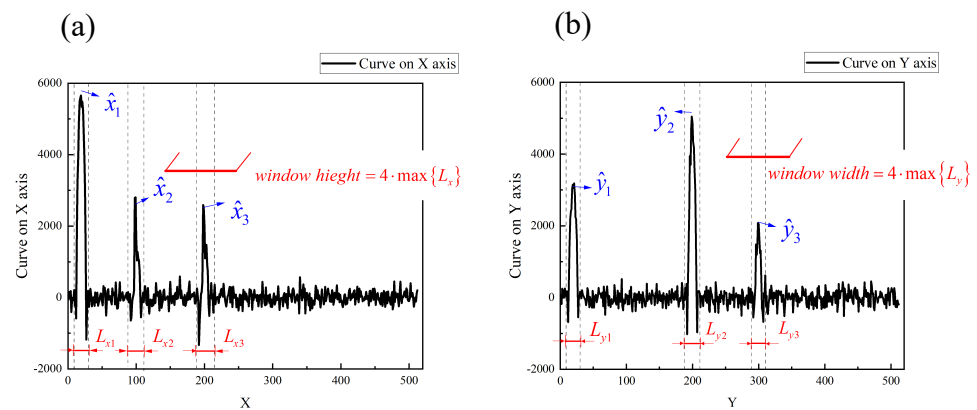


Figure 5. Projected curves of the image with four targets. (a) is the projected curve of image scene on X axis and (b) is the projected curve on Y axis. $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ are the three regions on the X axis where the target may exist, and $(\hat{y}_1, \hat{y}_2, \hat{y}_3)$ are the three regions on the Y axis. The length L_x and L_y of the regions determine the size of window function.

From the projected curve on the X axis shown in Figure 5a, we can determine the four target positions in three regions on the X axis. The coordinates of the center points of these three regions on the X axis are $\hat{x}_1 = 19$, $\hat{x}_2 = 98$, and $\hat{x}_3 = 198$. Similarly, the center points of the four regions on the Y-axis are denoted as $\hat{y}_1 = 20$, $\hat{y}_2 = 199$, and $\hat{y}_3 = 299$. These coordinates (\hat{x}, \hat{y}) define nine regions where targets may exist, that is, nine rectangular window functions. In addition, the size of the spatial window function can be determined by the length L_x and L_y of the regions in the projected curves. Here we take four times the length of the longest region on the projected curve as the window length of the window function in the X or Y axis. The size of window function is defined as 64×64 pixels, since the maximum of (L_x, L_y) is $(16, 16)$ pixels.

The 18 Fourier coefficients are calculated corresponding to these nine rectangular window functions are calculated according to Equation (1) and (2). Thus, we can exclude the regions where no targets existed and obtain the exact locations of the centers of four targets, as shown in Table 1.

Table 1. The estimated results of multi-target locations.

Targets	T_1	T_2	T_3	T_4
Position coordinate	(20, 20)	(20, 200)	(100, 200)	(200, 300)
Estimation result	(19, 20)	(20, 199)	(98, 199)	(198, 299)

3.1.2. Multi-Target Tracking Using the Independent Estimation Approach

The four window functions designed based on the multi-target position are shown in Figure 6. They are centered on $\bar{w}_1 = (19, 20)$, $\bar{w}_2 = (20, 199)$, $\bar{w}_3 = (98, 199)$, and $\bar{w}_4 = (198, 299)$. The parts of w_1 and w_2 that are beyond the boundary are discarded. The spatial windows designed for different targets are shown in Figure 7, where (a–d) represent the images at the moment t_1 , and (e–h) represent the images at the moment t_2 . (a,e) represent target T_1 , (b,f) represent target T_2 , (c,g) represent target T_3 , and (d,h) represent target T_4 .

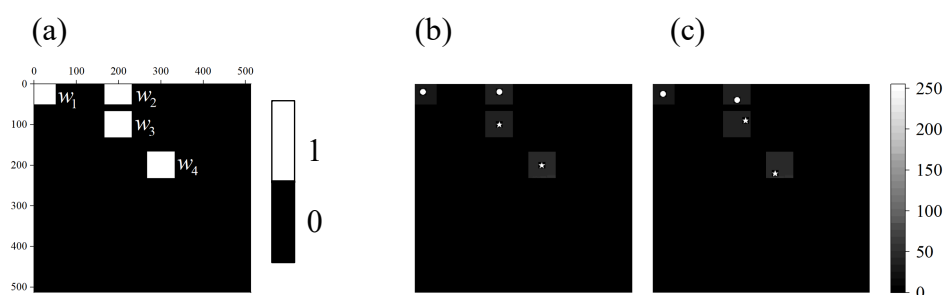


Figure 6. The four window functions for four targets: (a) is the four-window function, (b) is the windowed image at moment t_1 , and (c) is the windowed image at moment t_2 .

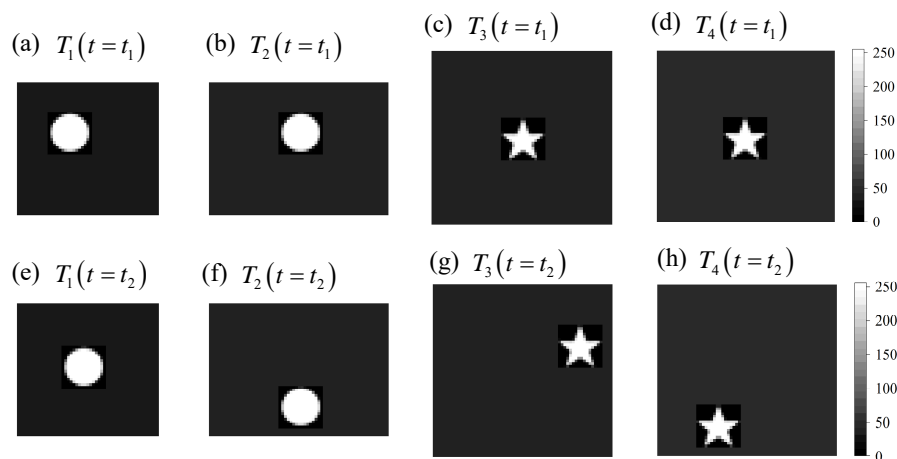


Figure 7. The four window functions: (a–d) are the images at the moment t_1 , and (e–h) are the images at the moment t_2 ; (a,e) represent target T_1 , (b,f) represent target T_2 , (c,g) represent target T_3 , and (d,h) represent target T_4 .

The estimated results of the four targets according to Equation (12) are shown in Table 2. The errors of estimated results are within two pixels. The errors of T_3 and T_4 are larger than those of T_1 and T_2 because the target T_3 and target T_4 have a lower brightness and are more affected by noise.

Table 2. The results using the independent estimation approach.

Targets	T_1	T_2	T_3	T_4
True value	(5,5)	(20,0)	(−10,20)	(20,−10)
Estimation result	(4.82,4.60)	(19.67,0.08)	(−8.89,19.65)	(21.15,−8.55)
Error	0.43	0.34	1.16	1.85

3.1.3. Multi-Target Tracking Using the Joint Estimation Approach

Assuming that the speed of the four targets remains unchanged, the scene image at the moment t_3 is as shown in Figure 8. At this moment, target T_2 and target T_3 are already very close. By updating the center coordinates of the window functions with the result estimated at the moment of t_2 , we can obtain the results of the window function selected for target T_2 and target T_3 at the moment of t_2 and t_3 , as shown in Figure 8, where (a–d) represent the images at moment t_2 , (e–h) represent the images at moment t_3 , (c,g) are the images of target T_2 , and (d,h) are the images of target T_3 .

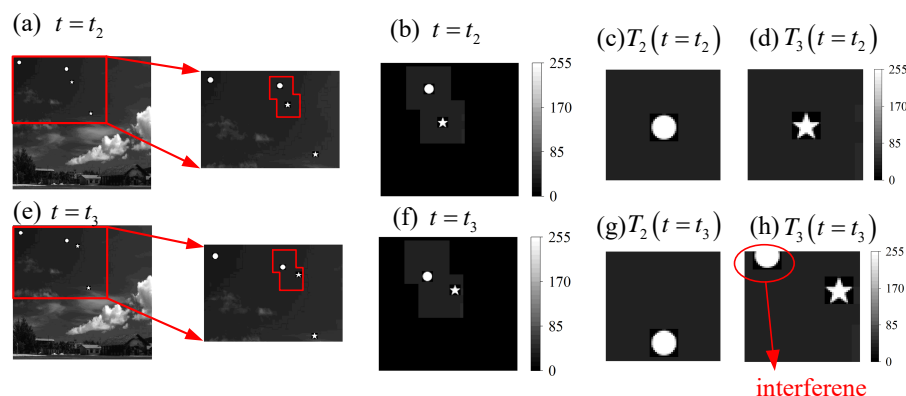


Figure 8. Illustrations of the windowed image at moment t_2 and t_3 : (a–d) represent the images at moment t_2 , and (e–h) represent the images at moment t_3 , (c,g) are the images of target T_2 , and (d,h) are the images of target T_3 . (h) contains the interference caused by target T_2 .

Significant errors will occur when estimating the displacement of target T_3 using the independent estimation method. Using the previously mentioned joint estimation method, the four Fourier coefficients of the windowed target image are calculated at moment t_2 , and the displacement of target T_3 is solved using the analytical method. The results of the two estimation approaches are shown in Table 3.

Table 3. The results using the joint estimation approach.

	Independent Estimation	Joint Estimation	True Value
T2	(19.66, −0.14)	(19.48, 0.15)	(20, 0)
Error of T2	0.37	0.29	
T3	(−23.34, −6.92)	(−11.29, 19.82)	(−10, 20)
Error of T3	30.04	1.30	

The results of T_2 using the different estimation approach are similar. However, there is a significant error in the result of T_3 using the independent estimation approach. It is obvious that there is an interference in the target estimation as shown in Figure 8h.

As the number of targets becomes larger, the computational effort of the joint estimation approach increases significantly. In conclusion, the joint estimation approach is mainly

used when multiple targets are close. The independent estimation approach is preferred when the distance between different targets is long.

3.2. Experiment

In order to verify the capability of the proposed method, a video experiment was performed. Two balls representing two targets are moving irregularly in the scene at the same time, as shown in Figure 9. The experiment captured a video through a cell phone camera and performed target tracking for each image frame in the video. The matrix representing the scene brightness distribution of each frame was multiplied by a different sampling matrix, and the sum of the products was assumed to be the detected light intensity of the single-pixel detector. The video size is 1024×720 pixels and was scaled to 512×512 pixels. There are 636 frames in total, and the video frame rate is 60 Hz. The video was shot at a distance of 60 cm at an angle of 45 degrees. The target scene size is $50 \text{ cm} \times 35 \text{ cm}$, and the diameters of the balls are both 1.5 cm.

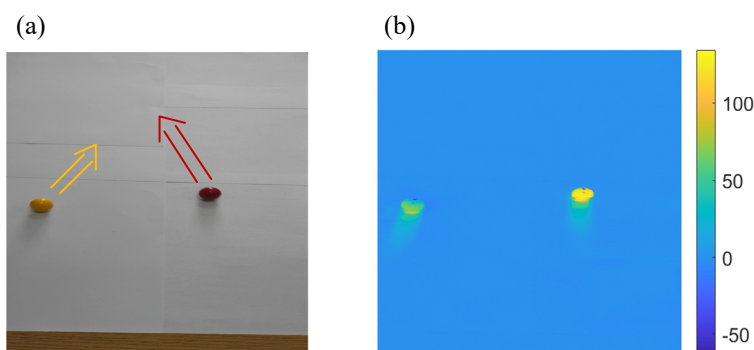


Figure 9. Experiment scene: (a) is the 240th frame in the video, and (b) is the image of 240th frame with background subtraction.

We assume that the 240th frame is the moment when the targets are first detected. Multi-target tracking was performed for a total of 391 frames from the 240th frame to the 630th frame. Estimated results in five different frames are shown Figure 10 from the 260th frame to the 340th frame. The five frames present the motion of the balls and the estimated results. The first row of the Figure 10 shows the scene images, and the second row shows the results of multi-target tracking. The real-time estimated position of target T_1 (representing the yellow ball on the left in the 240th frame) is marked with a red dot in Figure 10, and target T_2 (representing the red ball on the right in the 240th frame) is marked with a red triangle.

As shown in Figure 10, the estimated results and the actual positions of the targets always overlap with small error. It can be concluded from each frame that the estimation error of the proposed multi-target tracking method is less than the radius of the ball. Considering that the ball occupies about 30 pixels in the image, the estimation error is less than 15 pixels. If we manually calibrate these five frames, the average estimation error is about 6 pixels.

The estimated trajectories of the two balls are shown in Figure 11. The yellow points are the representation of the left yellow ball in Figure 9a, and the red points represent the right red ball. The trajectories are displayed every five frames.

The positions of the yellow ball are shown in Figure 12a every two frames, and the red ball is represented in Figure 12b. Between the 240th frame and the 260th frame, the independent estimation approach was used to track the targets. After the 260th frame, the two targets are close and cannot be distinguished by the spatial window function. The joint estimation approach was used from the 260th frame to the 320th frame. After the 320th

frame, the two balls are separated again, and the independent estimation approach was used after the 320th frame.

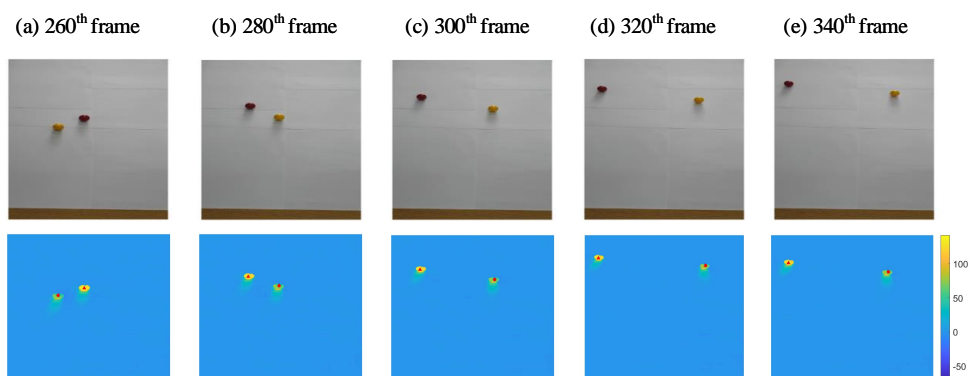


Figure 10. Estimated results in five different frames from the 260th frame to the 340th frame. The first row shows the scene images, and the second row shows the results of multi-target tracking. The real-time estimated position of Target 1 (representing the yellow ball on the left in the 240th frame) is marked with a red dot, and Target 2 (representing the red ball on the right in the 240th frame) is marked with a red triangle.

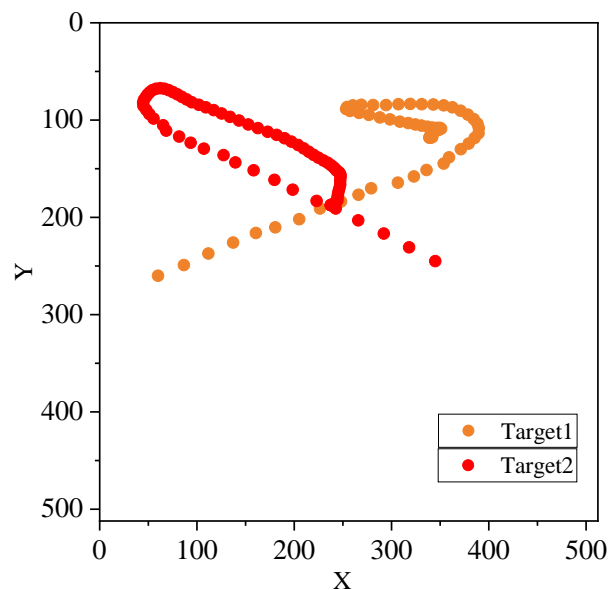


Figure 11. The estimated trajectories.

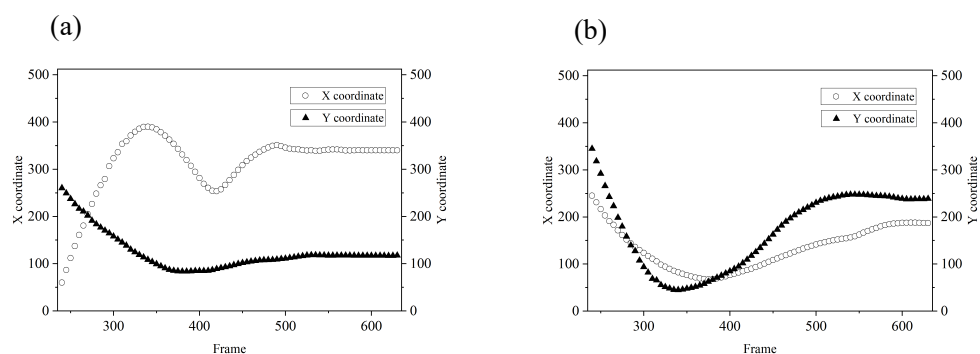


Figure 12. The estimated results of coordinates between the 240th frame and the 630th frame: (a) is the yellow ball, and (b) is the red ball.

4. Discussion

The shape of the window function affects the tracking performance of the proposed method. In this paper, a common rectangular window function was chosen, and the window length was also preset to be a constant. However, in a practical application, the shape and window length of the window function should be related to the target shape, the target motion speed, and other factors. A large window length is required for high-speed targets, and a small window length is required for multi-target estimation. Therefore, a window function with an adjustable shape will be better suited for this method.

In the joint estimation approach, only the estimation of two targets is shown in this paper. However, theoretically, this method can be used to estimate more than two targets simultaneously. Fewer than five targets can be estimated directly according to the analytical method. When used for more than five targets, the time consumption of the search method is more demanding. Subsequent work will further extend the proposed joint estimation approach for more targets.

5. Conclusions

In conclusion, this paper presents a multi-target tracking method using windowed Fourier single-pixel imaging. The proposed method can track K targets by implementing $6K$ measurements using a single-pixel detector. For example, using a DMD with a modulate rate of 10 kHz, the tracking frame rate for five targets was up to 333 Hz. The ability to track multiple targets will greatly broaden the application range of SPI in security and military fields. In particular, a combination with millimeter waveband detectors will further expand SPI's application range due to the reconfigurable antenna patterns.

Author Contributions: Conceptualization, J.Z. and T.H.; methodology, J.Z.; software, J.Z. and X.S.; validation, J.Z., T.H. and M.X.; formal analysis, X.S.; investigation, Z.X.; data curation, M.X.; writing, J.Z.; supervision, T.H.; funding acquisition, Y.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology on Near-Surface Detection Laboratory (Grant No TCGZ2017A005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there is no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SPI	single-pixel imaging
FSI	Fourier single-pixel imaging
WFSI	windowed Fourier single-pixel imaging
STFT	short-time Fourier transform
DMD	digital modulating devices

References

1. Lu, H.; Li, Y.; Li, H.; Lv, R.; Lang, L.; Li, Q.; Song, G.; Li, P.; Wang, K.; Xue, L.; et al. Ship Detection by an Airborne Passive Interferometric Microwave Sensor (PIMS). *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2682–2694. [[CrossRef](#)]
2. Zhang, J.; Hu, T.; Li, W.; Li, L.; Yang, J. Research on Millimeter-wave Radiation Characteristics of Stereo Air Target. In Proceedings of the 2020 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP), Suzhou, China, 29–31 July 2020; pp. 1–3. [[CrossRef](#)]
3. Goodwin, A.; Glancey, M.; Ford, T.; Scavo, L.; Brey, J.; Heier, C.; Durr, N.J.; Acharya, S. Development of a low-cost imaging system for remote mosquito surveillance. *Biomed. Opt. Express* **2020**, *11*, 2560–2569. [[CrossRef](#)] [[PubMed](#)]
4. Liggins, M.; Chong, C.Y. Distributed fusion architectures and algorithms for target tracking. *Proc. IEEE* **1997**, *85*, 95–107. [[CrossRef](#)]

5. Case, E.E.; Zelnio, A.M.; Rigling, B.D. Low-Cost Acoustic Array for Small UAV Detection and Tracking. In Proceedings of the 2008 IEEE National Aerospace and Electronics Conference, Dayton, OH, USA, 16–18 July 2008; pp. 110–113. [[CrossRef](#)]
6. Christnacher, F.; Hengy, S.; Laurenzis, M.; Matwyschuk, A.; Naz, P.; Schertzer, S.; Schmitt, G. Optical and acoustical UAV detection. In Proceedings of the Electro-Optical Remote Sensing X, Edinburgh, UK, 26–27 September 2016; International Society for Optics and Photonics: Bellingham, WA, USA, 2016; Volume 9988, p. 99880B.
7. Magana-Loaiza, O.S.; Howland, G.A.; Malik, M.; Howell, J.C.; Boyd, R.W. Compressive object tracking using entangled photons. *Appl. Phys. Lett.* **2013**, *102*, 231104. [[CrossRef](#)]
8. Sun, S.; Lin, H.; Xu, Y.; Gu, J.; Liu, W. Tracking and imaging of moving objects with temporal intensity difference correlation. *Opt. Express* **2019**, *27*, 27851–27861. [[CrossRef](#)]
9. Zhu, S.; Ma, K.K. A new diamond search algorithm for fast block-matching motion estimation. *IEEE Trans. Image Process.* **2000**, *9*, 287–290. [[CrossRef](#)] [[PubMed](#)]
10. Barjatya, A. Block matching algorithms for motion estimation. *IEEE Trans. Evol. Comput.* **2004**, *8*, 225–239.
11. Zhou, D.; Cao, J.; Cui, H.; Hao, Q.; Chen, B.k.; Lin, K. Complementary Fourier single-pixel imaging. *Sensors* **2021**, *21*, 6544. [[CrossRef](#)]
12. Edgar, M.P.; Gibson, G.M.; Padgett, M.J. Principles and prospects for single-pixel imaging. *Nat. Photonics* **2019**, *13*, 13–20. [[CrossRef](#)]
13. Gopalsami, N.; Liao, S.; Elmer, T.W.; Koehl, E.R.; Heifetz, A.; Raptis, A.P.C.; Spinoulas, L.; Katsaggelos, A. Passive millimeter-wave imaging with compressive sensing. *Opt. Eng.* **2012**, *51*, 091614. [[CrossRef](#)]
14. Jiang, W.; Li, X.; Peng, X.; Sun, B. Imaging high-speed moving targets with a single-pixel detector. *Opt. Express* **2020**, *28*, 7889–7897. [[CrossRef](#)] [[PubMed](#)]
15. Zhou, F.; Shi, X.; Chen, J.; Tang, T.; Liu, Y. Non-imaging real-time detection and tracking of fast-moving objects. *arXiv* **2021**, arXiv:2108.06009.
16. Xu, S.; Zhang, J.; Wang, G.; Deng, H.; Ma, M.; Zhong, X. Single-pixel imaging of high-temperature objects. *Appl. Opt.* **2021**, *60*, 4095–4100. [[CrossRef](#)] [[PubMed](#)]
17. Olivieri, L.; Totero Gongora, J.S.; Pasquazi, A.; Peccianti, M. Time-resolved nonlinear ghost imaging. *ACS Photonics* **2018**, *5*, 3379–3388. [[CrossRef](#)]
18. Olivieri, L.; Gongora, J.S.T.; Peters, L.; Cecconi, V.; Cutrona, A.; Tunesi, J.; Tucker, R.; Pasquazi, A.; Peccianti, M. Hyperspectral terahertz microscopy via nonlinear ghost imaging. *Optica* **2020**, *7*, 186–191. [[CrossRef](#)]
19. Chen, S.C.; Feng, Z.; Li, J.; Tan, W.; Du, L.H.; Cai, J.; Ma, Y.; He, K.; Ding, H.; Zhai, Z.H.; et al. Ghost spintronic THz-emitter-array microscope. *Light. Sci. Appl.* **2020**, *9*, 1–9. [[CrossRef](#)]
20. Zanutto, L.; Piccoli, R.; Dong, J.; Caraffini, D.; Morandotti, R.; Razzari, L. Time-domain terahertz compressive imaging. *Opt. Express* **2020**, *28*, 3795–3802. [[CrossRef](#)]
21. Jiao, S.; Sun, M.; Gao, Y.; Lei, T.; Xie, Z.; Yuan, X. Motion estimation and quality enhancement for a single image in dynamic single-pixel imaging. *Opt. Express* **2019**, *27*, 12841–12854. [[CrossRef](#)]
22. Monin, S.; Hahamovich, E.; Rosenthal, A. Single-pixel imaging of dynamic objects using multi-frame motion estimation. *Sci. Rep.* **2021**, *11*, 7712. [[CrossRef](#)]
23. Shi, D.; Yin, K.; Huang, J.; Yuan, K.; Zhu, W.; Xie, C.; Liu, D.; Wang, Y. Fast tracking of moving objects using single-pixel imaging. *Opt. Commun.* **2019**, *440*, 155–162. [[CrossRef](#)]
24. Zhang, Z.; Ye, J.; Deng, Q.; Zhong, J. Image-free real-time detection and tracking of fast moving object using a single-pixel detector. *Opt. Express* **2019**, *27*, 35394–35401. [[CrossRef](#)] [[PubMed](#)]
25. Deng, Q.; Zhang, Z.; Zhong, J. Image-free real-time 3-D tracking of a fast-moving object using dual-pixel detection. *Opt. Lett.* **2020**, *45*, 4734–4737. [[CrossRef](#)] [[PubMed](#)]
26. Zha, L.; Shi, D.; Jian, H.; Yuan, K.E.; Meng, W.; Yang, W.; Runbo, J.; Yafeng, C.; Wang, Y. Single-pixel tracking of fast-moving object using geometric moment detection. *Opt. Express* **2021**, *29*, 30327–30336. [[CrossRef](#)] [[PubMed](#)]
27. Zhang, Z.; Ma, X.; Zhong, J. Single-pixel imaging by means of Fourier spectrum acquisition. *Nat. Commun.* **2015**, *6*, 6225. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, Z.; Wang, X.; Zheng, G.; Zhong, J. Fast Fourier single-pixel imaging via binary illumination. *Sci. Rep.* **2017**, *7*, 12029. [[CrossRef](#)]