Passive Low SNR Tracking by Spatial-Temporal Fusion of Sliding-window Radon Transforms

Alexander N. Dolia, Scott F. Page, Neil M. White, Chris J. Harris
School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK
{ad, sfp03r, nmw, cjh}@ecs.soton.ac.uk

Abstract – In the presence of multi-path propagation, attenuation of signal spectral components and target motion, the estimates of time delay between random wideband signals in low signal-to-noise (SNR) scenarios are characterized by a high probability of large errors. This results in increase the number of sensors to obtain desired performance levels (and therefore results in a series of further problems of managing the sensors). In this paper we propose a novel approach to time delay estimation (TDE) based on spatial-temporal fusion of local Radon transforms (LRD) in a sliding window, which significantly decreases probability of large errors of TDE. The properties of the cross-correlation function (CCF) in low SNR scenarios for linear and nonlinear filtering of sequences of CCF outputs are considered. The Efficiency of the proposed method in terms of normal error variance and the probability of large errors of TDE using numerically simulated data is discussed.

Keywords: Time delay estimation, nonlinear filtering, local radon transform.

1 Introduction

In the case of sensor signals in low SNR scenarios (that may also be influenced by other destructive factors), tracking in passive sonar systems based on the time delay estimation (TDE) between random wideband signals results in poor accuracy characterized by a high probability of large errors $P_{\tau}$ for location of a target [1,2] i.e. the cases when the time delay estimation greatly differs from its true value $\tau_0$.

For low SNR (SNR<<1), "prewhitening" and correction of the received signal cross-spectrum is necessary. An alternative procedure is to use the elementary cross-correlator output filter. In this case, the function of SNR over the spectra has to be known a priori, or pre-estimated. In some particular cases however, there are optimal techniques and algorithms for TDE, but in passive systems accurate information about a target and noise spectras is often absent [1,3,4]. The function of SNR over spectras can also vary in time, for example because of target movement, propagation and attenuation effects. Therefore, in practice it is difficult to design an optimal real-time passive system and suboptimal methods, for example, such as the Radon transform can be applied [7,11,12].

In the case of lines, the Radon transform is mathematically equivalent to the Hough transform (HT) [11,12,13]. A Matlab implementation of the Hough transform that performs detection of lines, circles and ellipses is given in [13]. The Radon and Hough transforms have been extensively used for the video, radar and sonar tracking and parameter estimation with very good theoretical and practical results [15,16,17,18].

The paper by Illingworth and Kittler [14] provides a list of 136 references about the HT in chronological manner.

“Classical” Radon transform requires the entire image, which increases computational load and time response to obtain TDEs [7,12]. Besides, in passive low SNR tracking performed by sonar systems the bearing-time images [5,6,7,] have different properties than the range-time intensity returns for a given beam position [15]. The windowed Radon or, in other words, LRD has less computational load and time response and it can detect curves (not lines) still using the sliding window with linear sub apertures [7].

In this paper we generalize and modify the method proposed in [7]. Thus, the following solution seems reasonable: let us use the preliminary "prewhitening" of additive noise and subsequently, standard correlation, performing on the basis of a standard or modified FFT processor implementation [5]. We then propose to apply: 1) local Radon transform detectors with a pre-estimated or given decision threshold to the sequence of cross-correlator outputs represented as 2-D data arrays or images; 2) forward-backward propagation of their decisions in a small sliding-window; and 3) spatial-temporal fusion of the results of this propagation obtained in previous time instants.

The efficiency of the proposed approach, based on the spatial-temporal fusion of the sliding-window Radon transforms is quantitatively compared with other methods with respect to normal error variance $\sigma_\tau^2$ and probability of large errors $P_{\tau}$. The methodology of adaptive parameter selection is discussed. The performance of the proposed algorithm and other methods are demonstrated on simulated datasets with different target spectras.
The paper is organized as follows. In Section 2 we discuss the properties of the CCF, signal and noise models for cases of low SNR and influence of different factors. Methods of representing CCF sequences as 2-D images are considered in Section 3. The 2-D filtering of CCFs is introduced in Section 4. The description of the proposed method is given in Section 5. Accuracy of the proposed method is compared with a number of different methods in Section 6 followed by conclusions and acknowledgements.

2 Signal and noise models

Let us consider the signal-noise model for TDE using two spatially displaced sensors in the Fourier domain as the following [6]:

\[
\dot{S}_{u_1}(w) = \dot{S}_{u_1}(w)e^{j\phi_1(w)}e^{-j\alpha(w)\tau_0} + \hat{N}_1(w), \quad (1)
\]

\[
\dot{S}_{u_2}(w) = \dot{S}_{u_2}(w)e^{j\phi_2(w)}e^{-j\alpha(w)\tau_0} + \hat{N}_2(w) , \quad (2)
\]

where \(j^2 = -1\); \(\dot{S}_{u_1}(w)\) and \(\dot{S}_{u_2}(w)\) are the spectra of received signal-noise mixtures from sensors \(u_1(t)\) and \(u_2(t)\), respectively, where \(t\) is time; \(\dot{S}_{u_1}(w)\) is the scaled complex spectrum of the signal irradiated by a target; \(\phi_1(w)\) and \(\phi_2(w)\) - are random, slowly changing functions (phase fluctuations) with zero means and variances that increase according to \(2\)(phase fluctuations) with zero means and quasi constant values. In the general case, \(\tau_0\) and \(\phi_1(w)\) and \(\phi_2(w)\) depend on time \(t\). For a fixed distance \(r\) the function \(\alpha(w, r)\) increases when \(|w|\) becomes greater [3,4].

Without loss of generality we assume that the source is a constant point; in the general case, \(\tau_0\) and both \(\phi_1(w)\) and \(\phi_2(w)\) depend on time \(t\). For a fixed distance \(r\) the function \(\alpha(w, r)\) increases when \(|w|\) becomes greater [3,4].

Let us introduce new notations \(\dot{S}_1(w) = \dot{S}_{u_1}(w)e^{-\alpha(w)\tau_0}\) and \(\Delta\phi(w) = \phi_1(w) - \phi_2(w)\) then we obtain the cross-spectrum as the following

\[
\dot{S}_{12}(w) = \dot{S}_{u_1}(w)\dot{S}_{u_2}^*(w) = \dot{S}_1(w)\dot{S}_1^*(w)e^{j\Delta\phi(w)}e^{j\alpha(w)\tau_0} + \{\dot{S}_1(w)e^{j\phi_1(w)}\hat{N}_2^*(w) + \dot{S}_1(w)e^{-j\phi_2(w)}e^{j\alpha(w)\tau_0}\hat{N}_1(w)\}, \quad (3)
\]

where \(\hat{N}_1(w)\) and \(\hat{N}_2(w)\) are the spectra of received signal-noise mixtures from sensors \(u_1(t)\) and \(u_2(t)\), respectively; \(0\) is a predefined constant; if \(|w|\geq w_u\) then \(S(w)\) is of triangular shape (see Fig. 1) as follows

\[
\left|\dot{S}_{u_1}(w)\right| = k_0 |1 - |w|/w_u|.
\]

where \(SNR = k_0/2\) and \(k_0\) is a predefined constant; if \(|w|\geq w_u\) then \(S(w)\) is of triangular shape (see Fig. 1) as follows

\[
\left|\dot{S}_{u_1}(w)\right| = k_0 |1 - |w|/w_u|.
\]

where \(*\) denotes the complex conjugate.

After “prewhitening” using the function \(N(w)\) the CCF, \(Y(\tau)\), is described as the following [1]

\[
Y(\tau) = \int\left\{\dot{S}_{u_1}(w)/N(w)\right\}e^{j\omega\tau}dw = R_u(\tau - \tau_0) + n(\tau), \quad (4)
\]

\[
R_u(\tau - \tau_0) = \int\{\dot{S}_1(w)\dot{S}_1^*(w)e^{j\Delta\phi(w)}e^{j\alpha(w)\tau_0}/N(w)e^{j\alpha(w)\tau}dw,
\]

where \(\tau\) is the considered delay between signals \(u_1(t)\) and \(u_2(t)\); \(R_u(\tau - \tau_0)\) is the output signal component (the inverse Fourier response of the first term in (3) after prewhitening of the \(\dot{S}_{12}(w)\)); \(n(\tau)\) is the output noise component. In order to find the TDE using the standard cross-correlation method we need to find the TDE, \(\tau_0\), that corresponds to the maximum of the cross-correlation function \(Y(\tau)\), that is

\[
\tau_0 = \text{argmax}(Y(\tau)) . \quad (5)
\]

The factor \(e^{j\Delta\phi(w)}\) and source motion contribute to the output noise component. Usually the average width of the main lobe of the CCF of the signal component \(R_u(\tau - \tau_0)\), \(\delta_S\), is greater than the mean width of other output side lobes, \(\delta_A\). In low SNR case, the value of \(\delta_A\) is mainly determined by the additive noise spectrum band (see (3), (4)) and the other destructive factors. At the same time the multipath effects, signal absorption (increasing with frequency), random medium inhomogeneities and source motion degrade the received signal spatial-temporal coherence [1,3,4]. The consequence is that \(\delta_S\) increases in comparison to \(\delta_A\) due to attenuation.

The other practical considerations that support the assumption that \(\delta_S > \delta_A\) are the following. Usually signal processing is performed for the frequency band \(-w_u < w < w_u\) (\(w_u\) is the upper frequency of the signal processing) for which \(w_u > w_{us}\) where \(w_{us}\) is the upper frequency of the received signals \(u_1(t)\) and \(u_2(t)\). In passive systems the value \(w_{us}\) is often unknown and can vary depending on distance due to attenuation.

Let us illustrate this by numerical simulations. Assume that the ratio of power spectra \(\left|\dot{S}_{u_1}(w)\right|\) is of triangular shape (see Fig. 1) as follows

\[
\int\left|\dot{S}_{u_1}(w)/N(w)\right| = k_0 |1 - |w|/w_u|.
\]

where \(SNR = k_0/2\) and \(k_0\) is a predefined constant; if \(|w|\geq w_u\) then \(S(w)/N(w)\) is of triangular shape (see Fig. 1) as follows

\[
\left|\dot{S}_{u_1}(w)\right| = k_0 |1 - |w|/w_u|.
\]
CCF for cases of normal errors of TDE and large errors the mean values of the ratios \( \delta_S / \delta_A \) have been found for different SNRs, for example, for SNR=0.3 it is 1.45, for 0.25 the ratio is 1.42 and for SNR=0.2 it is 1.38. This supports the assumption introduced above.

The other assumptions concerning the output \( Y(\tau) \) are the following: a) the time interval of the signals \( u_1(t) \) and \( u_2(t) \) that is used to get the estimates of \( S_{u_1}(w) \) and \( S_{u_2}(w) \) is more than 8.12 times greater than the maximal possible time delay \( \tau_{\text{max}} \). If the signals \( u_1(t) \) and \( u_2(t) \) are sampled with the sampling rate \( \Delta \tau \) then the value \( \tau_{\text{max}} \) is determined by the ratio of interferometer size to acoustic wave propagation speed; b) the sampling rate \( \Delta \tau \) of the received signal-noise mixture or, at least, CCF discretization \( \Delta \tau \) is not less than \( 8\tau / w_u \), i.e. data is over sampled. This requirement deals with the necessity to provide sufficient accuracy of \( \delta_S \), \( \delta_A \) and \( \tilde{\tau}_0 \) estimation and to apply the output sequence processing algorithms described below.

\[
|S_u(w)|/|N(w)| = K_0
\]

Fig. 1: Illustration of the ratio of triangular shape power spectra (see (6))

### 3 Sequence of CCFs as 2-D images

The model-noise assumptions and properties of CCF discussed above are put into the basis of the proposed algorithm of CCF sequence post-processing that improves the accuracy of TDE. Denote \( \tau_l \) and \( \tau_{l+1} \) the TDEs of the sampled \( \tau_0 \) at time intervals \( l \) and \( l+1 \), respectively. It is assumed that \( \max(\tau_{l+1} - \tau_l) < 2\Delta \tau \), i.e., TDEs \( \tau_l \) and \( \tau_{l+1} \) vary slowly enough between time intervals \( l \) and \( (l+1) \) (see Fig. 2).

There are various methodologies of CCF sequence representation as 2-D images. We use the following one. Let \( [Y[l,i]] \) be an array of CCFs calculated for the \( l \)-th time segment. Then for the fixed time interval \( l \) true TDE \( \tau_0 \) is \( \tau_0 = i\Delta \tau \), \( i = -i_{\text{max}} \ldots i_{\text{max}} \) (\( i_{\text{max}} = i_{\text{max}}\Delta \tau \) is the maximal physically possible delay). Let us map the real value array \( [Y[l,i]] \) onto \( Y^n[l,i] \in [0,255] \) in the following way

\[
Y^n[l,i] = Y[l,i]K_{\text{nor}}[l] + 128 \, ,
\]

\[
K_{\text{nor}}[l] = \begin{cases} 
127/l & \text{if } |Y_{\text{max}}[l]| \geq |Y_{\text{min}}[l]|, \\
127/l & \text{otherwise}
\end{cases}
\]

where \( Y_{\text{max}}[l] \) and \( Y_{\text{min}}[l] \) are the maximum and minimum of CCF for the \( l \)-th time segment respectively. In this way one first gets the representation of the obtained image as a traditional 8-bit integer value array suitable for image processing applications and second, normalization of \( Y[l,i] \) makes the influence of quantization errors negligible. In case of relatively high SNR the obtained image contains the information fragments visually observed as a continuous curve of mainly vertical orientation (in Fig. 2, 3 horizontal orientation) having several pixel width (for ratio (5) approximately equal to 7.8) and the region corresponding to random noise. When the SNR is low and the probability of large errors is high the curve is hardly observed visually (see Fig 2) and the noise region is characterized by greater fluctuations. It includes significant quasilinear variation and quasiconstant delay. Only the fragments of the simulated data are presented in Fig. 2.

Fig. 2: The cross-correlation function of the single target over time: the white curve indicates location (bearing or TDE) of the target in different time instances.

In order to estimate \( \delta_S \) we can analyze the CCF in the neighbourhood of the maximum of \( Y^n[l,i] \) for the fixed time interval \( l \). In practice, the estimate of \( \delta_A \) can be obtained using the expectation of the interval \( \tau \) that \( Y(\tau) \geq 0 \). From now on we suppose that the cross-correlator operates sequentially, i.e. produces the outputs for equal time subintervals \( l \) following each other. In case of high probability of large errors one can at least estimate the value of \( \delta_A \). This information can be further used for initial selection of the size of an aperture of sliding window of a 1-D or 2-D filter when the tracking of time delay is not yet performed.

In [6] it was shown that 1-D filtering of elementary CCF or the sequence of TDEs obtained from elementary observations by means of even the simple median filter
provide some reduction of TDE large error occurrence but its performance is not acceptable. 2-D filter application offers better results. That is why we highlight the analysis of 2-D filter properties and efficiency. One more advantage of 2-D filter application is that they are widely applied to image processing and therefore the results of their analysis are widely available [10].

4 2-D Filtering of CCFs

There are several approaches to time delay estimation accuracy improvement [1]. Consider now the methods of TDE based on analysis of $Y^n[l,i]$ in the small sliding window. First, we analyse filters that utilize a rectangular weighting mask $W$, with size $M \times L$. For example, in our experiments $L$ and $M$ are equal to 1, 5 or 7.

![Image](image.png)

**Fig. 3:** The estimates of location of the single target over time: white dots indicate the location of the targets, there are 128 rows corresponding to different locations over time and line number 64 relates to the zero TDE.

![Image](image.png)

**Fig. 4 Illustration of transformation $Y[l,i]$ into $Y^n[l,i]$.**

In the simplest case, for $\forall w_{ij} = 1$ (elements of $W$). Then we consider the case when we have a bank of filters $\{W^p\}$, $p = 1,...,P$ ($P$ is the number of filters in filter bank $\{W^p\}$) where some of the elements of $W^p$ are equal to one, $w^p_{i,j} = 1$ and others equal to zero, $w^p_{i,j} = 0$. We then describe the proposed method that fuses the outputs of the filter bank for different positions of sliding window.

The first method performs spatial processing or 1-D linear adaptive filtering of rows of image $Y^n[l,i]$ as follows

$$Y^f[l,i] = 1/K \sum_{j=-(L-1)/2}^{(L-1)/2} Y^n[l,i+j]w_{1,j} \quad (9)$$

where $Y^f[l,i]$ is the output sequence smoothed data array or the result of filtering, and $w_{1,j} = 1$. In this case we assume all elements of $W$ are equal to one, $w_{ij} = 1$.

We assume that $M = 1$ where $M$ is the number of rows in the weighted function $W$. Obviously, if we use a linear transform from $Y[l,i]$ to $Y^n[l,i]$ it equals the product of cross-spectra $S_{12}(w)$ with the corresponding spectrum of weighted function $W$.

It is shown in [6] that for this technique based on $\delta_S$ and $\delta_A$ estimation the adaptive selection of $L$ is the following:

$$L = \begin{cases} 
\frac{(\delta_S + 3\delta_A)}{4\Delta\tau}, & \text{if } \delta_S > \delta_A \\
1, & \text{otherwise}
\end{cases} \quad (10)$$

where $L$ has to be odd.

This recommendation is valid for both 1-D filters to be applied to elementary CCF and for 2-D ones in order to determine their horizontal size $L$. Theoretically, it should provide near optimum reduction in $P_\delta$ and $\sigma_\delta^2$.

If the speed of time delay temporal variation is rather small, significant improvement can be obtained using cross-spectrum recursive averaging for different time instants. If we use a linear mapping from $Y[l,i]$ to $Y^n[l,i]$ it is equivalent to temporal processing of $Y^n[l,i]$ or 1-D linear adaptive filtering of the columns of image $Y^n[l,i]$ in the following fashion

$$Y^f[l,i] = 1/M \sum_{m=-(M-1)/2}^{(M-1)/2} Y^n[l+m,i]w_{m,1} \quad (11)$$

where $w_{m,1} = 1$. This method fails if time delay changes rapidly.

The aforementioned 1-D filters are particular cases of 2-D filtering when $M \geq 1$ and $L \geq 1$. This filter performs spatial-temporal processing and can be written as follows

$$Y^f[l,i] = \frac{1}{ML} \sum_{m=-(M-1)/2}^{(M-1)/2} \sum_{j=-(L-1)/2}^{(L-1)/2} Y^n[l+m,i+j]w_{m,j} \quad (12),$$

where $w_{m,j} = 1$. Obviously, in this case we can perform 1-D spatial filtering and then apply 1-D temporal filtering or vice versa.
5 Proposed method

In order to take into account bearing variation (TDE) over time we can use the filter bank with masks \( W^p \) that match the variation in bearing. Because, in practice \( M \leq 7 \), we can assume linear variation of TDE or bearing over time. For example, when \( M = 3 \) and \( L = 3 \) the filter bank will consist of three masks \( \{ W^1, W^2, W^3 \} \)

\[
W^1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad W^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad W^3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.
\]

(13)

Therefore, we assume that the changing of TDE \( \tau_j \) can be approximated locally with a set of short lines of angle \( \theta \in [\pi/4,3\pi/4] \) that changes from \( \pi/4 \) to \( 3\pi/4 \), \( \theta \in [\pi/4,3\pi/4] \); see (13) where \( M = 3 \) and \( L = 3 \). For the vertical line \( \theta = \pi/2 \) (it corresponds to the horizontal lines for the coordinate system used in Fig. 2,3) and \( \theta = 0 \) for the horizontal line. We now consider the method that consists of two steps [7]:

Step 1, apply the local Radon transform in a sliding window or a convolution \( Y^n \) with the bank of filters \( \{ W^p \} \) (see (13)) as follows

\[
R^f_1 = Y^n \otimes W^1, \ldots, R^f_p = Y^n \otimes W^p, \quad (14)
\]

where \( \otimes \) denotes convolution;

Step 2, for given \( l \) find maximum values among \( \{ R^f_{l,i} \} \),

\[
Y^f[l,i] = \max_{\psi} \{ R^f_{l,i}, R^f_{l,i+1}, \ldots, R^f_{l,i+M-1} \}. \quad (15)
\]

One of the challenges in passive low SNR tracking is how to fuse the result of processing of \( Y^n[l,i] \) in small sliding window, for example 5x5, over time and space in order to improve time delay estimation. In the simple case, we can get the only one estimate \( \hat{\tau}_l \) for the time segment \( l \) and then use association methods in order to combine \( \hat{\tau}_l \) obtained in time interval \( l \). But in more complicated case we can get a set of different estimates \( \hat{\tau}_l \) for the same time segment (the \( l \) row of \( Y^n[l,i] \)). For example, find such values \( i \) for the fixed \( l \) that \( Y^n[l,i] \) is more than a predefined threshold \( T \) and then apply association methods with the removal of false tracks.

Further development of the aforementioned method (see (15)) concerns the processing after the first step. In batch settings (when the entire image \( Y^n[l,i] \) is available) the proposed algorithm can be written as follows

Step 1, perform the local Radon transforms \( R^i_{m,n} \) (\( \theta \in [\pi/4,3\pi/4] \), see (13))

\[
R^i_{m,n} = \sum_{k=-(M-1)/2}^{(M-1)/2} \sum_{l=-(L-1)/2}^{(L-1)/2} w_{k,l}^i Y^n_{m+k,n+l}, \quad (16)
\]

where \( i = 1, \ldots, P \), e.g., \( P = 5 \) if \( L = 5 \) and \( M = 5 \); \( Y^n_{m,n} \) is just a new notation of \( Y^n[m,n] \), that is \( Y^n_{m,n} = Y^n[m,n] \)

Step 2, for every row \( m \) define a threshold \( T_m \)

\[
T_m = \max_{i,n} R^i_{m,n} - T_u, \quad (17)
\]

where \( T_u \) is the predefined constant, for example, in our experiment \( T_u \) is equal to 0 or 11 (see Table 1).

Step 3, detect the short lines using threshold \( T_m \) and propagate the decision back and forward over time using the mask \( f_{m,n}^i \) (in the simplest case, \( f_{m,n}^i = w_{i,m,n}^i \)) followed by spatial-temporal fusion of the results of detection \( \Psi_{m,n} \)

\[
\Psi_{m+k,n+l} = \begin{cases} \Psi_{m+k,n+l} + f_{m+k,n+l}^i & \text{if } R^i_{m,n} \geq T_m \\ \text{not changed} & \text{if } R^i_{m,n} < T_m \end{cases} \quad (18)
\]

\[
k = -(K-1)/2, \ldots, (K-1)/2 \\
l = -(M-1)/2, \ldots, (M-1)/2
\]

Step 4, repeat Step 1-3 for all rows of \( Y^n_{m,n} \);

Step 4, find the estimate of TDE \( \hat{\tau}_m \),

\[
\hat{\tau}_m = \left( \arg \max_n \Psi_{m,n} - \varepsilon \right) \Delta \tau. \quad (19)
\]

where \( \varepsilon \) - is a constant that determines which column of \( \Psi_{m,n} \), \( R^i_{m,n} \) or \( \Psi_{m,n} \) corresponds to zero time delay, \( \tau_m = 0 \).

If the task is to define if the point \((m,n)\) belongs to any detected short line then we need to modify Step 3 of the proposed algorithm,

\[
\Psi_{m+k,n+l} = \begin{cases} f_{m+k,n+l}^i & \text{if } R^i_{m,n} \geq T \\ \text{not changed} & \text{if } R^i_{m,n} < T \end{cases} \quad (20)
\]
or simply check if \( \Psi_{m,n} > 0 \) (the point \((m,n)\) belongs to at least to the one short line) or \( \Psi_{m,n} = 0 \) (there is no short line detected in this point). \( \Psi_{m,n} \) can be interpreted as a measure of confidence about the location of the target at time instant \( m \) and is equal to the number of all short lines detected by the algorithm that intersect at the point \((m,n)\). If \( k \) and \( l \) do not change and are equal to 0 in the third step (in the second proposed algorithm) and \( T_n = 0 \) then both proposed methods give the same TDEs.

If the 1-D filtering is required to perform for every lines of \( Y^m_n \) as in (9) then it can be incorporated into the proposed algorithm by the changing filter mask \( w_{m,n} \) but not the decision propagation function \( f_{m,n}^u \). In this case, we can detect more thicker short lines. Although it increases the number of operation that is necessary to perform a convolution of \( Y^m_n \) with \( w_{m,n} \), because the size of the \( w_{m,n} \) should be \( M \times (2L) \) it could improve the accuracy of TDEs for the triangular-like spectra (see Fig. 1).

In future work, the filter bank \( \{ W^p \} \) based on Gabor wavelets [13] (see Fig. 5) and motion detector using, for example, spatio-temporal energy models [19] can be applied for TDE utilising the proposed approach. In this case the propagation of the results of motion detection in the small sliding window using the mask \( f_{m,n}^u \) could be dependant on the orientation of Gabor wavelets and its size.

Consider \( \hat{\tau}_0 \) as an estimate of TDE with large error if the following inequality is valid \( |\hat{\tau}_0 - \tau_0| > \delta \tau S / 2 \) where \( \tau_0 \) is the true value of TDE. Note \( P_a + P_n = 1 \) where \( P_n \) is the probability of the normal estimate or the probability that the inequality \( |\hat{\tau}_0 - \tau_0| \leq \delta \tau S / 2 \) is valid, \( P_n = P(|\hat{\tau}_0 - \tau_0| \leq \delta \tau S / 2) \).

The task to reduce \( P_a \) is more important than to decrease \( \sigma^2 \). By reducing \( P_a \) we minimize the probability that the variance of the TDE exceeds a certain threshold. In other words, it minimizes the upper bound of variance of the TDE.

To show the necessity of filter parameter adaptation (for instance aperture size updating), we also analysed the other case of power spectra ratio:

\[
\frac{\hat{S}_{\tau}(w)}{N(w)} = \text{const},
\]

(21)

if \(|w| \geq w_a \), then \( \frac{\hat{S}_{\tau}(w)}{N(w)} = 0 \).

The results of numerical simulations for methods 1-7 above for time delay sinusoidal behaviour, \( \tau_1 = 64 + 33 \sin(2\pi / 399 \cdot t) \), are presented in Table 1. For spectrum ratio (6) \( \delta_S = 8 \Delta \tau \) and \( \delta_A = 4 \Delta \tau \); and \( \delta_S = 4 \Delta \tau \), \( \delta_A = 4 \Delta \tau \) for (21).

In the experiments we simulate \( \hat{S}_{12}(k) \) with 2048 elements where \( \hat{S}_{12}(k) \) is a discrete counterpart of the spectrum \( \hat{S}_{12}(w) \). The 10 elements of \( \hat{S}_{12}(k) \) (255 elements for the positive and 255 ones for negative values of \( w \)) were used to simulate the spectrum of the signal emitted by the target. After applying the 2048 element inverse Fourier transform (see (4)) to \( \hat{S}_{12}(k) \) the 128 elements of \( Y(\tau) \) were used for further analysis.

The first method in the Table 1 is the standard one that finds the location of the maximum of the cross-correlation function \( Y(\tau) \) (see (4), (5) and Fig.4). It can be seen that the probability of large measurements is very high, for example, if \( \text{SNR} = 0.19 \) then \( P_a = 0.6 \) (see the method N1 in Table 1). If we calculate the median of the TDEs obtained by the method N1 using the sliding window framework (let call this method N2, see Table 1) then it is possible to reduce the probability of estimates with large errors but \( P_a \) is still high [8]. Much better results in comparison with methods N1,2 can be obtained by averaging the cross-spectra (method N3, see Table 1).

1-D median filtering of \( Y^m_n \) using the sliding window approach is slightly better than the standard CCF approach for triangular-like power spectra but worse than method N2 when the spectra is rectangular-like.

It is seen that the best results with respect to reduction of \( P_a \) are provided by proposed method 7 (local Radon transform with fusion of results of line detection) and

**Fig. 5:** Spatial-temporal cosine (left) and sine (right) simple cell receptive fields for the perception of motion [19] or Gabor functions [13].

### 6 Experiments

In case of low SNR or when there is simultaneous influence of several destructive factors mentioned earlier in the paper the accuracy of TDE is characterized by two main parameters [2,6]: 1) the variance of normal errors \( \sigma^2 \); 2) the probability of large errors \( P_a \).

Consider \( \hat{\tau}_0 \) as an estimate of TDE with large error if the following inequality is valid \( |\hat{\tau}_0 - \tau_0| > \delta \tau S / 2 \) where \( \tau_0 \) is the true value of TDE. Note \( P_a + P_n = 1 \) where \( P_n \) is the probability of the normal estimate or the probability that the inequality \( |\hat{\tau}_0 - \tau_0| \leq \delta \tau S / 2 \) is valid, \( P_n = P(|\hat{\tau}_0 - \tau_0| \leq \delta \tau S / 2) \).

The task to reduce \( P_a \) is more important than to decrease \( \sigma^2 \). By reducing \( P_a \) we minimize the probability that the variance of the TDE exceeds a certain threshold. In other words, it minimizes the upper bound of variance of the TDE.

To show the necessity of filter parameter adaptation (for instance aperture size updating), we also analysed the other case of power spectra ratio:

\[
\frac{\hat{S}_{\tau}(w)}{N(w)} = \text{const},
\]

(21)

if \(|w| \geq w_a \), then \( \frac{\hat{S}_{\tau}(w)}{N(w)} = 0 \).

The results of numerical simulations for methods 1-7 above for time delay sinusoidal behaviour, \( \tau_1 = 64 + 33 \sin(2\pi / 399 \cdot t) \), are presented in Table 1. For spectrum ratio (6) \( \delta_S = 8 \Delta \tau \) and \( \delta_A = 4 \Delta \tau \); and \( \delta_S = 4 \Delta \tau \), \( \delta_A = 4 \Delta \tau \) for (21).

In the experiments we simulate \( \hat{S}_{12}(k) \) with 2048 elements where \( \hat{S}_{12}(k) \) is a discrete counterpart of the spectrum \( \hat{S}_{12}(w) \). The 10 elements of \( \hat{S}_{12}(k) \) (255 elements for the positive and 255 ones for negative values of \( w \)) were used to simulate the spectrum of the signal emitted by the target. After applying the 2048 element inverse Fourier transform (see (4)) to \( \hat{S}_{12}(k) \) the 128 elements of \( Y(\tau) \) were used for further analysis.

The first method in the Table 1 is the standard one that finds the location of the maximum of the cross-correlation function \( Y(\tau) \) (see (4), (5) and Fig.4). It can be seen that the probability of large measurements is very high, for example, if \( \text{SNR} = 0.19 \) then \( P_a = 0.6 \) (see the method N1 in Table 1). If we calculate the median of the TDEs obtained by the method N1 using the sliding window framework (let call this method N2, see Table 1) then it is possible to reduce the probability of estimates with large errors but \( P_a \) is still high [8]. Much better results in comparison with methods N1,2 can be obtained by averaging the cross-spectra (method N3, see Table 1).

1-D median filtering of \( Y^m_n \) using the sliding window approach is slightly better than the standard CCF approach for triangular-like power spectra but worse than method N2 when the spectra is rectangular-like.

It is seen that the best results with respect to reduction of \( P_a \) are provided by proposed method 7 (local Radon transform with fusion of results of line detection) and
especially if it is incorporated with 1-D linear prefiltering for triangular shape spectrum ratios. However, rectangular spectrum ratio 1-D prefiltering decreases the accuracy and this proves necessity of CCF analysis and filter parameter adaptation.

The method N6 that combines linear and nonlinear data processing algorithms [9,10] (see (14) and (15)) and method N5 with appropriate choice of M and L also gives good results. At the same time, output sequence post-processing can also reduce $\sigma^2_\tau$ in comparison to the primary TDE obtained by method N1.

7 Conclusions

A novel approach to time delay estimation, based on spatial-temporal fusion of local Radon transform in a sliding window that significantly decreases the probability of estimation with large errors is proposed. Some possibilities of parameter adaptation (horizontal size of aperture $L$) are considered. Numerical simulation results permit to compare the algorithms’ efficiency and demonstrate good performance of the proposed methods. In future work the main focus will be analysis of the filter bank based on Gabor functions and integration of mobile TDE systems in unified multi-sensor active management framework.

Acknowledgements

The first author would like to thank Alexander Zelensky and Vladimir Lukin for helpful discussions and suggestions. This research was supported by the Data Information Fusion Defence Technology Centre, United Kingdom, project 8.1 “Multi-sensor active management.”

References

Table 1: Quantitative comparison of TDE methods for the two types of the spectra

<table>
<thead>
<tr>
<th>№</th>
<th>Methods</th>
<th>Size of sliding window</th>
<th>The triangular-like spectra SNR=0.19</th>
<th>The rectangular-like spectra SNR=0.19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M$</td>
<td>$L$</td>
<td>$P_T$</td>
</tr>
<tr>
<td>1</td>
<td>Standard cross-correlation time delay estimation</td>
<td></td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td>2</td>
<td>Median of TDEs obtained by first method</td>
<td>5</td>
<td>1</td>
<td>0.52</td>
</tr>
<tr>
<td>3.1</td>
<td>Cross-spectrum averaging</td>
<td>5</td>
<td>1</td>
<td>0.12</td>
</tr>
<tr>
<td>3.2</td>
<td></td>
<td>7</td>
<td>1</td>
<td>0.086</td>
</tr>
<tr>
<td>4.1</td>
<td>1-D median filtering of $Y^a[l,i]$ using sliding window approach</td>
<td>1</td>
<td>5</td>
<td>0.62</td>
</tr>
<tr>
<td>4.2</td>
<td></td>
<td>1</td>
<td>7</td>
<td>0.65</td>
</tr>
<tr>
<td>5.1</td>
<td>Linear filter (see (9)-(11)),</td>
<td>1</td>
<td>5</td>
<td>0.59</td>
</tr>
<tr>
<td>5.2</td>
<td></td>
<td>1</td>
<td>7</td>
<td>0.6</td>
</tr>
<tr>
<td>5.3</td>
<td></td>
<td>5</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>5.4</td>
<td></td>
<td>7</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>5.5</td>
<td></td>
<td>5</td>
<td>5</td>
<td>0.094</td>
</tr>
<tr>
<td>5.6</td>
<td></td>
<td>5</td>
<td>7</td>
<td>0.12</td>
</tr>
<tr>
<td>5.7</td>
<td></td>
<td>7</td>
<td>5</td>
<td>0.038</td>
</tr>
<tr>
<td>5.8</td>
<td></td>
<td>7</td>
<td>7</td>
<td>0.059</td>
</tr>
<tr>
<td>6.1</td>
<td>Nonlinear filter (see (14), (15)) with 1-D linear pre-filtering, 1x5</td>
<td>5</td>
<td>5</td>
<td>0.16</td>
</tr>
<tr>
<td>6.2</td>
<td></td>
<td>5</td>
<td>5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LRT</th>
<th>Threshold</th>
<th>$T_u = 0$</th>
<th>$T_u = 11$</th>
<th>Size of sliding window of the 1-D linear pre-filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>7.2</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>7.3</td>
<td>+</td>
<td>-</td>
<td>1x5</td>
<td>5</td>
</tr>
<tr>
<td>7.4</td>
<td>-</td>
<td>+</td>
<td>1x5</td>
<td>5</td>
</tr>
<tr>
<td>7.5</td>
<td>+</td>
<td>-</td>
<td>1x7</td>
<td>5</td>
</tr>
<tr>
<td>7.6</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>7.7</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>7.8</td>
<td>+</td>
<td>-</td>
<td>1x5</td>
<td>7</td>
</tr>
<tr>
<td>7.9</td>
<td>-</td>
<td>+</td>
<td>1x5</td>
<td>7</td>
</tr>
<tr>
<td>7.10</td>
<td>+</td>
<td>-</td>
<td>1x7</td>
<td>7</td>
</tr>
</tbody>
</table>