Maximum Likelihood Identification of Glint Noise

WEN-RONG WU
National Chiao Tung University

If the non-Gaussian distribution function of radar glint noise is known, the Masreliez filter can be applied to improve target tracking performance. We investigate the glint identification problem using the maximum likelihood (ML) method. Two models for the glint distribution are used, a mixture of two Gaussian distributions and a mixture of a Gaussian and a Laplacian distribution. An efficient initial estimate method based on the QQ-plot is also proposed. Simulations show that the ML estimates converge to truths.

Manuscript received March 4, 1993; revised April 3, 1994.

IEEE Log No. T-AES/32/1/00761.

Author's address: Dept. of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

I. INTRODUCTION

The foundation of a radar tracker is a Kalman filter. The Kalman filter is optimal if the system is linear and measurement noise is Gaussian. However, in reality, the system is often nonlinear and noise may have non-Gaussian components. Masreliez and Martin [2] demonstrated that non-Gaussian noise can severely degrade the performance of the Kalman filter, in particular for noise with a heavy-tailed distribution. In a radar system, there exists this heavy-tailed non-Gaussian noise and it is known as glint noise.

A number of researchers have considered the problem of Kalman filtering in the non-Gaussian environment. One of the most effective schemes was proposed by Masreliez [3–4]. He introduced a nonlinear score function as the correction term in the state estimate and the results are often nearly optimal. While this approach seems promising, it encounters the difficulty of implementing the convolution operation involved in the evaluation of the score function. Wu and Kundu in [6] described an efficient way to implement the score function. Wu [7] applied the method to tackle the tracking problem when glint noise is present.

To use the Masreliez filter, the glint model has to be known. Glint models have been investigated for years. A common approach is considering the physics that describes electromagnetic wave scattering. Delano [23] found that the glint probability density can be described by the student t distribution with two degrees of freedom. Gubonin [24–25] argued that the glint density is described by generalized hypergeometric functions. Recently, Sandhu and Saylor [26] generalized Gubonin's results. A key feature of their work is that the mean glint estimator is unbiased and this enables model parameters to be estimated from the first-order measured glint data. They demonstrated that the simulated target signature is statistically identical with the actual measured data. Although Sandu's method can model glint noise well, it is not suitable for the use in the Masreliez filter. The main reason is that hypergeometric functions may not have moment generating functions. This violates the assumptions in [6]. In addition, as we see in Section III, hypergeometric functions are complicated and it is difficult to deal with.

We consider some simplified models here. Two glint noise models, a mixture of two Gaussian distributions and a mixture of a Gaussian and a Laplacian distribution are studied. The Gaussian mixture was originally proposed by Hewer and Martin in [1]. They analyzed QQ-plots of empirical glint data and claimed that this model works well. However, no identification method was considered. The Gaussian and Laplacian mixture was used in [7] for robust consideration. No model identification was mentioned in [7] either.

These two models are rather simple; however, as we show, they can describe glint noise almost as well as hypergeometric functions.

The history of identifying a Gaussian mixture is a long one. Pearson first considered the problem by using the method of moments [8], which was studied later by Cohen [9]. Extensions to more general cases were recently considered by Fukunaga and Flick [9], and Derin [11]. During the early nineteenth century, attention was turned to graphical techniques for mixture identification. Basically, these techniques explore the relationship between the shape of the QQ-plot of the mixture with unknown parameters. This was originated by Harding [12] extended by Cassie [13] and studied later by Bhattacharya [14]. Recently, it was reconsidered by Postaire and Vasseur [15]. The maximum likelihood (ML) method was first used by Rao [16]. Extension to an arbitrary number of mixtures was made by Hassleblad [17]. Hosmer [18] investigated ML identification when the sample size is small. Day [19] compared the ML method with a number of other methods and concluded that the ML method was superior. This same result was also observed by Fowlkes [20].

One thing worth noting is that most of the studies consider the case where the mixture has different means. In such cases, the distribution is frequently bimodal, however, in our applications, the mixture is unimodal and identification of this mixture is somewhat different. The aforementioned works all consider the identification of a Gaussian mixture. The identification of our second model, i.e., the Gaussian and Laplacian mixed distribution has not been considered before. Existing methods may not be adequate to solve the problem making further study necessary. We explore the glint identification problem using the ML method. Since the ML is an iterative method and many local minima may exist, good initial estimates are important. We also propose a very effective initial estimate method.

The organization of the paper is as follows. In Section II we briefly review the score function approach for completeness. In Section III, we discuss the glint models. In Section IV, we describe the ML identification problem and the initial estimate technique. In Section V, some simulations are carried out. The conclusion is drawn in Section VI.

II. SCORE FUNCTION APPROACH

A. General Filtering Problem

The general filtering problem can be formulated as the state estimate given all the history of the observation. Consider a linear system described as follows:

$$x_{k+1} = \phi_k x_k + w_k \tag{1}$$

$$z_k = H_k x_k + v_k \tag{2}$$

where x_k is the state vector, w_k and v_k represent white noise sequences and are assumed to be mutually independent. The basic problem is to estimate the state x_k from the noisy observation $(z_1,\ldots,z_k)$. The probability density of the state conditioned on all the available observation data is called the *a posteriori* density. If this density is known, an estimate for any type of performance criterion can be found. Thus, the estimation problem can be viewed as one of determining the *a posteriori* density. Denote $f(\cdot)$ as a density and $Z^k = \{z_0, z_1, \ldots, z_k\}$. The *a posteriori* density can be described by the following relations

$$f(x_k \mid Z^k) = \frac{f(x_k \mid Z^{k-1}) f(z_k \mid x_k)}{f(z_k \mid Z^{k-1})} \tag{3}$$

$$f(x_k \mid Z^{k-1}) = \int f(x_{k-1} \mid Z^{k-1}) f(x_k \mid x_{k-1}) dx_{k-1} \tag{4}$$

where the normalizing constant $f(z_k \mid Z^{k-1})$ is given by

$$f(z_k \mid Z^{k-1}) = \int f(x_k \mid Z^{k-1}) f(z_k \mid x_k) dx_k. \tag{5}$$

The $f(z_k \mid x_k)$ in (3) is determined by the observation noise density $f(v_k)$ and (2). Similarly, $f(x_k \mid x_{k-1})$ in (4) is determined by the state noise density $f(w_k)$ and the (1). Theoretically, knowing these densities, we can determine the *a posteriori* density $f(x_k \mid Z^k)$. However, it is generally impossible to carry out the integration in (4) for every instant. Consequently, the *a posteriori* density cannot be determined for most applications. The only one exception is when the initial state and all the noise sequences are Gaussian. In this case, (3)–(4) are reduced to the standard Kalman filter equations.

B. Score Function Approach

Consider a linear system described in (1)–(2). The variables w_k and v_k can be non-Gaussian. The density of z_k conditioned on the *a prior* observations is denoted by $f(z_k \mid Z^{k-1})$. We name $f(z_k \mid Z^{k-1})$ the *observation prediction density* and assume that it is twice differentiable. Similarly, $f(x_k \mid Z^{k-1})$ is the density of x_k conditioned on *prior* observations and is named the *state prediction density*. Assuming that $f(x_k \mid Z^{k-1})$ is a Gaussian density with mean $\overline{x}_k$, and covariance matrix M_k, Masreliez has shown that the minimum variance state estimate $\hat{x}_k$, and its covariance matrix $P_k = E\{(x_k - \hat{x}_k)(x_k - \hat{x}_k)^t \mid Z^k\}$ can be recursively calculated as follows [3]:

$$\hat{x}_k = \overline{x}_k + M_k H_k^t g_k(z_k) \tag{6}$$

$$P_k = M_k - M_k H_k^t G_k(z_k) H_k M_k \tag{7}$$

$$\overline{x}_{k+1} = \phi_k \hat{x}_k \tag{8}$$

$$M_{k+1} = \phi_k P_k \phi_k^t + Q_k \tag{9}$$

where $g_k(\cdot)$ is a column vector with components:

$$\{g_k(z_k)\}_i = -\left[\frac{\partial f(z_k \mid Z^{k-1})}{\partial (z_k)_i}\right][f(z_k \mid Z^{k-1})]^{-1} \tag{10}$$

and $G_k(z_k)$ is a matrix with elements

$$\{G_k(z_k)\}_{ij} = \frac{\partial\{g_k(z_k)\}_i}{\partial (z_k)_j}. \tag{11}$$

The function $g_k(\cdot)$ is called the score function of $f(z_k \mid Z^{k-1})$. The following procedure summarizes the implementation of the filter.

Step 0 Assume that at stage $k-1$, $\hat{x}_{k-1}$ and P_{k-1} are known.

Step 1 Calculate $M_k = \phi_{k-1}P_{k-1}\phi_{k-1}^t + Q_{k-1}$.

Step 2 Approximate the state prediction density $f(x_k \mid Z^{k-1})$ by a Gaussian distribution with mean $\overline{x}_k = \phi_{k-1}\hat{x}_{k-1}$ and covariance matrix M_k.

Step 3 Find the observation prediction density $f(z_k \mid Z^{k-1})$ by convolving $f(H_k x_k \mid Z^{k-1})$ with $f_{v_k}(\cdot)$.

Step 4 Find $g_k(z_k)$ and $G_k(z_k)$.

Step 5 Apply (6)–(7) to find $\hat{x}_k$ and P_k.

Step 6 Let $k \to k+1$ and start all over from Step 1.

III. GLINT MODELS

We first discuss the glint model proposed by Sandhu and Saylor in [26]. They assumed that the scatters of a radar target has a dominant deterministic component which is located at the instantaneous radar center. Let $\{a_k\}_{k=1}^N$ be a set of samples of zero mean random variables with a_k representing the magnitude of the kth scatter in a particular aspect interval such that $E[\sum_{k=1}^N a_k^2] = 2\sigma^2$. The quantity a_0 corresponds to the dominant, deterministic scatter. Let (x_k, y_k, z_k) be the coordinate of the kth scatter (with respect to the target center). $\{(x_k, y_k, z_k)\}_{k=1}^N$ are assumed to be uniformly distributed sequences of three independent random variables within the target volume with means μ_x, μ_y, μ_z and variancs $\sigma_x^2, \sigma_y^2, \sigma_z^2$. Since the three axes are independent, we only consider the glint in x axis in the following. It is shown in [26] that the glint can be modeled as

$$h_x = \mu_x + \sigma_x \beta_x \tag{12}$$

where

$$\beta_x = \frac{(r_1 + \sqrt{2}s\cos\phi_0)r_2 + (r_3 + \sqrt{2}s\sin\phi_0)r_4}{(r_1 + \sqrt{2}s\cos\phi_0)^2 + (r_3 + \sqrt{2}s\sin\phi_0)^2} \tag{13}$$

where ϕ_0 is the phase delay of the dominant scatter, r_1, r_2, r_3, r_4 are zero mean, unit variance, normal random variables, and $s = a_0/\sqrt{2}\sigma$. The distribution of β_x is shown to be

$$f(\beta_x) = \frac{\exp(-s^2)}{4\sqrt{2}b^3}\,{}_1F_1\left[-1/2, 1, \frac{-s^2}{2b^2}\right] \tag{14}$$

where $b = (1+\beta_x^2)/2$, and ${}_1F_1$ is a generalized hypergeometric function which can be expanded into modified Bessel functions as

$$f(\beta_x) = \frac{\exp(-s^2)}{4\sqrt{2}b^3}\left[\left(1 + \frac{s^2}{2b^2}\right)I_0\left(\frac{s^2}{4b^2}\right) + \frac{s^2}{2b^2}I_1\left(\frac{s^2}{4b^2}\right)\right]\exp\left(\frac{s^2}{4b^2}\right). \tag{15}$$

As we can see, the glint distribution is rather complicated. It is difficult to incorporate this model into the filtering process.

The variable μ_x corresponds to the mean glint and varies slowly. It can be effectively removed in the tracking loop [1]. Without loss of generality, we can assume that $\mu_x = 0$. In [1], the glint is modeled as a Gaussian mixture which is described by

$$\begin{aligned} f(x) &= \epsilon f_1(x) + (1-\epsilon)f_2(x) \\ &= \epsilon N(x; 0, \sigma_1) + (1-\epsilon)N(x; 0, \sigma_2) \end{aligned} \tag{16}$$

where

$$N(x; \eta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{(x-\eta)^2/2\sigma^2}, \tag{17}$$

$\sigma_2 \gg \sigma_1$, and ϵ is a positive number close to (but less than) one. Note that (16) explicitly assumes that both Gaussian components have zero means. The validity of the Gaussian mixture model is verified by the QQ-plot of the glint noise. A QQ-plot is a plot of the *ordered* data $x_{(i)}$ versus the normal quantiles $x_{p_i} = \Phi^{-1}(p_i)$ where $p_i = (i-1/2)n$, $i = 1, 2, \ldots, n$ and Φ^{-1} is the inverse of the unit Gaussian distribution. It can be shown that the QQ plot of a Gaussian distribution is linear regardless its mean and variance. The QQ-plot of a Gaussian mixture is locally linear. There could be three regions where the QQ-plot is linear; one in the central region and two in the tail. Based on this observation, Hewer, Martin, and Zeh [1] studied glint records and concluded that the Gaussian mixture is a good model. Here, we perform an experiment to test their argument. We used Sandhu's model (12) to generate a glint record and plotted the corresponding QQ-plot. The parameters we used are the typical values from [26] ($s = 1$, $\sigma_x = 5$, $\mu_x = 0$, and $\phi_0 = 0.7$). Fig. 1 shows the glint data and Fig. 2 shows its QQ-plot. Fig. 3 shows the theoretical QQ-plot of a Gaussian mixture. Indeed, comparing Figs. 2 and 3, we can clearly see that the glint distribution does have the Gaussian mixture behavior.

From (13), it is easy to see that when $s \to 0$, $f(\beta_x) \to$ Student t distribution which is very heavy tailed. Very large glint spikes are observable. Since s is time varying, these spikes may not be modeled in the underlying Gaussian mixture. We then are interested to know if the Masreliez filter can tolerate these spikes. This is related to the concept of *robustness*. An estimator is said to be robust if small changes from an

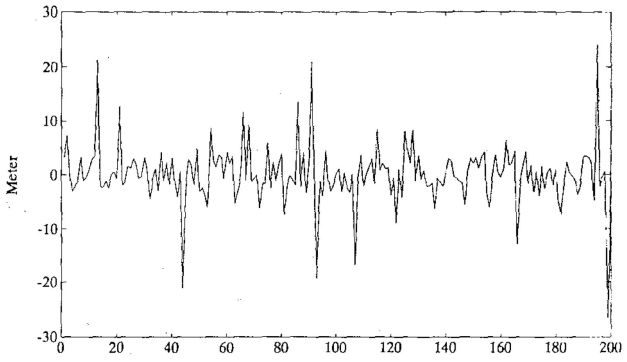


Fig. 1. Generated glint record.

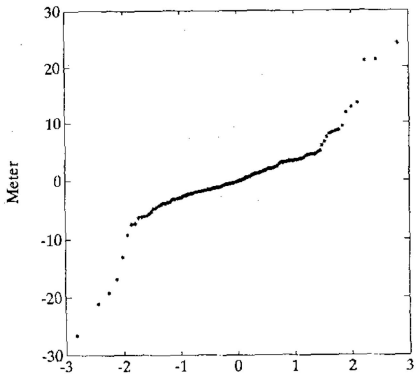


Fig. 2. QQ-plot of glint record.

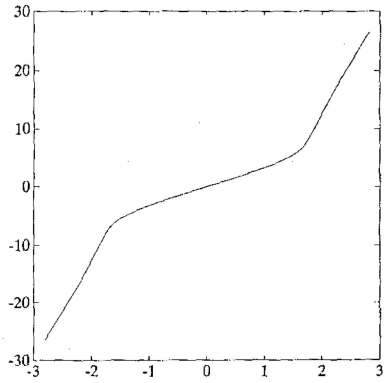


Fig. 3. Theoretical QQ-plot of Gaussian mixture.

assumed nominal model produce only small changes in the estimate. There are many forms of robustness. The most fundamental one, which was introduced by Hampel [27], is *qualitative robustness*. To define the qualitative robustness, we need the concept of the *influence curve*. For a finite sample, an interpretation of the influence curve can be made as follows. Consider the location estimation problem. We are given n observations from the nominal distribution. Let $\hat{\mu}_n$ be an estimate. Now, suppose the nth observation is contaminated and its value becomes x. Let the estimate with the contaminated observation be $\tilde{\mu}_n$. Then the influence curve is defined as

$$IC(x) = \sqrt{n}(\tilde{\mu}_n - \hat{\mu}_n). \tag{18}$$

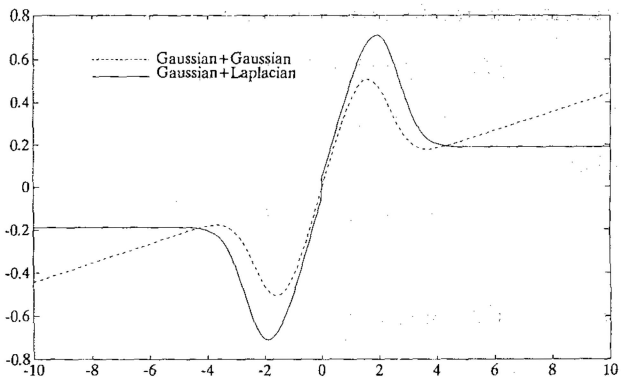


Fig. 4. Score functions of distributions.

From (18), we can realize that the influence curve measures the effect of contamination on the estimate. An estimator is said to be qualitatively robust if its influence curve $IC(\cdot)$ is continuous and bounded. In other words, if an estimator is qualitatively robust, a contaminated observation can only have finite influence on the estimate no matter how large it is. This interpretation is intuitively appealing and we use that in our filtering problem.

We now investigate the qualitative robustness of the Masreliez filter with the Gaussian mixture model. As described in (6), the state estimate is

$$\hat{x}_k = \overline{x}_k + M_k H_k^t g_k(z_k). \tag{19}$$

We now have to ask how $\hat{x}_k$ will be changed if z_k is contaminated. From (19), we find that the amount of change is determined by the shape of the score function of the observation prediction density $g_k(\cdot)$. Thus, we can say that the score function plays the same role as the influence curve does. If the score function is bounded, the contaminated z_k will only have finite influence on $\hat{x}_k$. From (5), we know that the observation prediction density is obtained by the convolution of the state prediction density and observation noise density. Since the state prediction density is assumed to be Gaussian, the shape of the score function of observation prediction density is similar to that of observation noise density. Fig. 4 shows the score functions of the Gaussian mixed and the Gaussian and Laplacian mixed distribution. The score function of a Gaussian mixture is piecewisely linear. The slope in the tail region is smaller. However, since the slope is not zero, the score function is not bounded. Thus, the Masreliez filter with the Gaussian mixed noise model is not qualitatively robust. Now, when we see the score function of the Gaussian and Laplacian mixed distribution, we find that the slope of score function is zero in the tail region. In other words, it is bounded. Thus, we conclude that the Masreliez filter with the Gaussian and Laplacian mixed model is qualitatively robust. This is the reason why the second glint model, i.e., the Gaussian and Laplacian mixed distribution, is proposed in [7]. This model can be

described in the following equation

$$f(x) = \epsilon f_1(x) + (1-\epsilon) f_2(x) = \epsilon N(x;0,\sigma) + (1-\epsilon) L(x;0,\mu) \quad (20)$$

where

$$L(x;\zeta,\mu) = \frac{1}{2\mu} e^{-|x-\zeta|/\mu}. \quad (21)$$

IV. ML GLINT IDENTIFICATION

A. ML Problem

Once we have the models, the next problem is how to identify parameters. Here we employ the ML method to solve this problem. Consider the model one in (16) and let $\theta = (\epsilon, \sigma_1, \sigma_2)$. For a sample with size N, the ML estimates of the parameters θ is found by maximizing the likelihood function $L_m(\theta)$

$$L_m(\theta) = \prod_{i=1}^{N} f(x_i;\theta). \quad (22)$$

For computational efficiency, we can take the logarithm of (22). This results in a log likelihood function

$$\ln[L_m(\theta)] = \sum_{i=1}^{N} \ln[f(x_i;0)]. \quad (23)$$

Taking first partials with respect to θ_i and setting the derivatives equal to zero, we can solve the maximization problem. However, as we can readily see, the equations are nonlinear and a closed-form solution doesn't exist and we must use iterative algorithms to find the solution. One such algorithm is proposed in [17]. Define

$$p_1^{(k)} = \epsilon^{(k)}, \qquad p_2^{(k)} = (1-\epsilon^{(k)}) \quad (24)$$

$$w_{ij}^{(k)} = \frac{p_i^{(k)} f_i^{(k)}(x_j)}{f^{(k)}(x_j)} \quad (25)$$

where $f_i^{(k)}(x_j)$, $i = 1,2$ and $f^{(k)}(x_j)$ are the density functions evaluated at x_j and $\theta^{(k)}$ (the estimate of θ in kth iteration). The estimates at the $k+1$ iterations are

$$\sigma_i^{(k+1)} = \left[\frac{\sum_{j=1}^{N} w_{ij}^{(k)} x_j^2}{\sum_{j=1}^{N} w_{ij}^{(k)}} \right]^{1/2}, \qquad i = 1,2 \quad (26)$$

$$\epsilon^{(k+1)} = \sum_{j=1}^{N} \frac{w_{1j}^{(k)}}{N}. \quad (27)$$

The iterations are continued until some predetermined criteria are met. If the iterations stop at m iteration then the estimate is $\hat{\theta} = \theta^{(m)}$. Day [19] has pointed out that $w_{ij}^{(k)}$ is an estimate, using $\theta^{(k)}$, of the *posteriori* probability that x_j came from the population i. There are singularities in L_m associated with each sample point, i.e., for each x_j, $1 \le j \le N$

$$\lim_{\sigma_1 \to 0} L_m(x_j, \epsilon, \sigma_1, \sigma_2) = \lim_{\sigma_2 \to 0} L_m(x_j, \epsilon, \sigma_1, \sigma_2) = \infty. \quad (28)$$

This kind of singularity has not occurred in our simulations. This is because, as described by [18], if one estimate of the standard deviation did start to converge to zero the estimate of the probability for that component converged to zero as fast or faster. This has the effect of reducing the mixture to a single component. To assure the absence of the singularity problem, we can check the validity of the mixture model during the iteration. This is done as follows. We first define a lower bound of σ. Whenever the $\hat{\sigma}_i$ is smaller than this value, the probability corresponding to σ_i is then forced to zero. The Gaussian mixture model then degenerates to a single component.

For the identification of model two in (20), a similar technique can be applied. However, (26) is different

$$\sigma^{(k+1)} = \left[\frac{\sum_{j=1}^{N} w_{1j}^{(k)} x_j^2}{\sum_{j=1}^{N} w_{1j}^{(k)}} \right]^{1/2} \quad (29)$$

$$\mu^{(k+1)} = \left[\frac{\sum_{j=1}^{N} w_{2j}^{(k)} |x_j|}{\sum_{j=1}^{N} w_{2j}^{(k)}} \right]. \quad (30)$$

In our simulations, the singularity problem has not occurred either. This indicates that the argument made above is still valid in this case.

B. Initial Estimate

As we saw in the preceding section, the cost function that the ML method has to optimize is highly nonlinear. There could be many local minima. Therefore, good initial estimates are important. There is a method proposed in [20], however, we find it is not suitable for our use since our distribution is unimodal. We now propose a scheme, which uses a QQ-plot analysis, to obtain the initial estimate. It is known that the QQ-plot of a Gaussian distribution is linear. This can be shown as follows. Let the error function be defined as

$$\mathrm{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{t^2/2} dt. \quad (31)$$

Then, for an arbitrary Gaussian density with mean η and variance σ^2, its cumulative distribution function (CDF) is $0.5 + \mathrm{erf}((x-\eta)/\sigma)$. Let x and y denote the horizontal and the vertical axis of the QQ-plot. The QQ-plot of the Gaussian distribution can be found by

solving

$$0.5 + \text{erf}\left(\frac{y-\eta}{\sigma}\right) = 0.5 + \text{erf}(x) \tag{32}$$

$$\frac{y-\mu}{\sigma} = x \rightarrow y = \sigma x + \mu. \tag{33}$$

Thus, for a Gaussian distribution, the QQ-plot is a straight line and the slope is the standard deviation σ. We can use this property to obtain the initial estimate of the standard deviation of the Gaussian mixture. As shown in (16), we know that in the central region of the Gaussian mixture, $N(x;0,\sigma_1)$ dominates. By contrast, in the tail regions, $N(x;0,\sigma_2)$ dominates, therefore, the slope of the linear segment (denoted as $L_1(x)$) in the central QQ-plot can be used as the initial estimate of σ_1 and that (denoted as $L_2(x)$) in the tail region as the initial estimate of σ_2

$$\hat{\sigma}_1 = \frac{dL_1(x)}{dx}, \qquad \hat{\sigma}_2 = \frac{dL_2(x)}{dx}. \tag{34}$$

For a Gaussian and Laplacian mixed distribution, we can still use the slope of the central region of its QQ-plot to estimate the initial σ (20). Let the straight line in the central region of the QQ-plot be $L_1(x)$. Thus,

$$\hat{\sigma} = \frac{dL_1(x)}{dx}. \tag{35}$$

However, in the tail region, the Laplacian distribution dominates and the QQ-plot is nonlinear. How to obtain the initial estimate of μ needs further consideration. First, let's find the QQ-plot of the Laplacian distribution. Since the QQ-plot is symmetrical with respect to zero, we have only to consider the region where $x \geq 0$ or $x \leq 0$. For $x \geq 0$, we have

$$0.5 + \int_0^y \frac{1}{2\mu} e^{-t/\mu}\, dt = 0.5 + \text{erf}(x). \tag{36}$$

Performing the integration and rearranging the equation, we can obtain

$$e^{-y/\mu} = 1 - 2\,\text{erf}(x). \tag{37}$$

Now, as we know, there is no closed-form solution for the error function. We cannot find a closed-form expression describing the relationship between x and y, however, we can apply some approximation method. Taking the logarithm of (37), we obtain

$$\frac{y}{\mu} = -\ln[1 - 2\,\text{erf}(x)]. \tag{38}$$

The function in the right-hand side of (38) is plotted and we find it can be well approximated by a quadratic function. Thus

$$-\ln[1 - 2\,\text{erf}(x)] \approx c_1 x^2 + c_2 x + c_3 \tag{39}$$

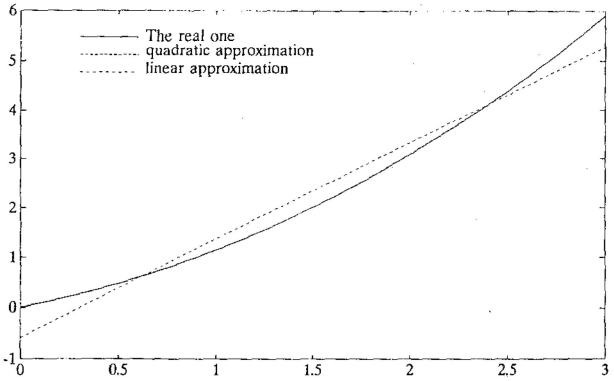


Fig. 5. Real and approximated $-\log[1-2\,\text{erf}(x)]$.

for some constants c_1, c_2, and c_3. The QQ-plot of a Laplacian distribution is then

$$y \approx \mu(c_1 x^2 + c_2 x + c_3). \tag{40}$$

From our experiments, we find that the coefficients of the quadratic function are

$$c_1 = 0.4687, \qquad c_2 = 0.5061, \qquad c_3 = 0.1454. \tag{41}$$

The actual and approximated $-\ln[1-2\,\text{erf}(x)]$ are plotted in Fig. 5. From this figure, we can see that the approximation is so good that we can barely distinguish them. We now conclude that the QQ-plot of a Laplacian distribution is almost a quadratic function. Now, we may use (40) to obtain the initial estimate of μ. For an empirical QQ-plot (Laplacian and Gaussian mixed), we first approximate the curve in the tail region by a quadratic function. Let the coefficients of the quadratic function be c_1', c_2', and c_3'. Since c_1 is shift invariant, μ can be estimated by

$$\hat{\mu} = \frac{c_1'}{c_1}. \tag{42}$$

However, our experiments indicate that this estimate is only good for large sample size. When the quantity of available data is small, the performance is poor. This can be explained by the fact that the quadratic function has more unknowns and the approximation is not good for the small sample size. One way to overcome this problem is to use the linear approximation, i.e., approximate the function $-\ln[1-2\,\text{erf}(x)]$ using a linear function. This leads to

$$-\ln[1 - 2\,\text{erf}(x)] \approx d_1 x + d_2 \tag{43}$$

$$y \approx \mu(d_1 x + d_2). \tag{44}$$

From experiments, d_1 is found to be 1.95. This is also shown in Fig. 5. The way to estimate μ for small sample size now is clear. Let the tail region of an empirical QQ-plot (Gaussian and Laplacian mixed distribution) be approximated by a straight line that

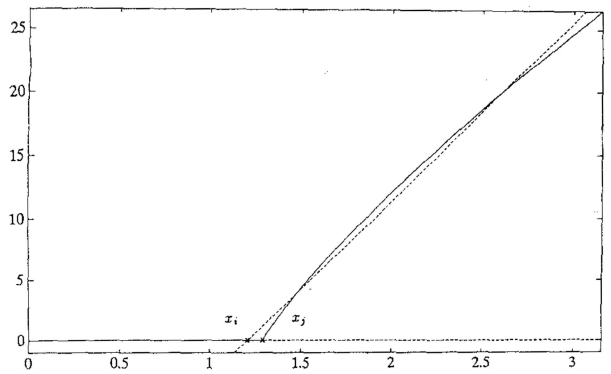


Fig. 6. Theoretical QQ-plot for mixture with infinite variance ratio.

is denoted by $L_2(x)$

$$L_2(x) = d_1'x + d_2'. \tag{45}$$

Then,

$$\hat{\mu} = \frac{d_1'}{d_1}. \tag{46}$$

Simulations show that this method yields very good results. We demonstrate this in the next section.

The final parameter we have to estimate is ϵ. If the Gaussian mixture is bimodal, there is a *reflection coefficient* that can be used to estimate ϵ [20]. In our case, this is not valid. Here, we propose a simple and effective way to do the job. Our method can best be described by an extreme case where $\sigma_1 = 0$. That means one distribution in the mixture is an impulse and the variane ratio (σ_2^2/σ_1^2) is infinite. The CDF of the Gaussian mixture is then discontinuous. The jump occurs at $x = 0$. The CDFs before and after jump are 0.5 and $0.5 + \epsilon$. The quantile of the Gaussian mixture right after jump is 0^+. The corresponding quantile of the unit Gaussian (x-axis of the QQ-plot), denoted as x_j, is obtained by

$$0.5 + \epsilon = 0.5 + \text{erf}(x_j) \tag{47}$$

$$x_j = \text{erf}^{-1}(0.5\epsilon). \tag{48}$$

The QQ-plot (for $x > 0$) is shown in Fig. 6. We see that the QQ-plot is zero until $x = x_j$. From the plot, we can estimate the standard deviations of mixed distributions using the method discussed before. We define the region corresponding to $0 < x < x_j$ is the central region and $x_j < x$ is the tail region. Thus, $L_1(x)$ is just the x-axis. The plot in the tail region is nonlinear. $L_2(x)$ can be obtained via the linear regression on the curve in the region. Now, from Fig. 6, we can see that the x-coordinate of the intersection of $L_1(x)$ and $L_2(x)$, denoted as x_i, is approximately equal to x_j, i.e.,

$$x_i \approx x_j. \tag{49}$$

In general, the slope of $L_1(x)$ is larger than zero. However, when the slope of L_1 increases, the slope of L_2, influenced by L_1, will decrease and vice versa. This leaves the x-coordinate of the intersection almost unchanged. Thus, we conclude that the x-coordinate of the intersection of $L_1(x)$ and $L_2(x)$ is not sensitive to the variation of σ_1 and σ_2 and (49) holds for $\sigma_2 \gg \sigma_1$ (this is also the assumption of the glint model). Based on the property described above, a method for the initial estimate of ϵ is now proposed.

1) Find the x-coordinate of the intersection of $L_1(x)$ and $L_2(x)$ in the empirical QQ-plot, i.e., solving the equation

$$L_1(x_i) = L_2(x_i). \tag{50}$$

2) Inverting (48) to find $\hat{\epsilon}$, i.e.,

$$\hat{\epsilon} = 2\,\text{erf}(x_i). \tag{51}$$

To implement the proposed initial estimate method, we first have to define the central and the tail regions. From the models mentioned above, it is reasonable to assume that $\epsilon > 0.5$. Thus, we can take the lower 50% of the sample (after ordering) and consider it is in the central region. Using the linear regression technique, we can find $L_1(x)$. Similarly, we can take hihger $(1-\epsilon)\times 100\%$ of the sample and consider it is in the tail region. However, ϵ is unknown. To estimate ϵ, $L_2(x)$ has to be found. Thus, $\hat{\epsilon}$ and $L_2(x)$ cannot be determined simultaneously. Here, we propose a simple iterative method to solve the problem. Let $L_2^{(n)}(x)$ and $\hat{\epsilon}^{(n)}$ be the $L_2(x)$ and $\hat{\epsilon}$ at nth iteration. At $(n+1)$th iteration, $L_2^{(n+1)}(x)$ is obtained by the linear regression on the upper $(1-\hat{\epsilon}^{(n)})\times 100\%$ of samples and $\hat{\epsilon}^{(n+1)}$ is obtained by using the intersection of $L_1(x)$ and $L_2^{(n+1)}(x)$ (as listed in (51)). The iteration continues until convergence is observed. The initial $\hat{\epsilon}^{(1)}$ is taken as 0.5.

In the following, we analyze the proposed algorithm to understand why it works. Let $x_i^{0.5}$ be the x-intersection of $L_1(x)$ and $L_2(x)$ when ϵ is 0.5 and $L_2(x)$ is found by using the upper 50% of the sample; $\hat{\epsilon}_{0.5}$ be the estimate of ϵ. Also let x_i^{ϵ} be the x-intersection of $L_1(x)$ and $L_2(x)$ when ϵ is known and $L_2(x)$ is found by using the upper $(1-\epsilon)\times 100\%$ of the sample; $\hat{\epsilon}$ be the estimate of ϵ. As we discussed above, $\hat{\epsilon}_{0.5}$ will be around 0.5 and $\hat{\epsilon}$ around ϵ. Since $\epsilon > 0.5$ and the error function is increasing, $x_i^{0.5} < x_i^{\epsilon}$ and $\hat{\epsilon}_{0.5} < \hat{\epsilon}$. Now, the real situation is that ϵ is unknown and initially we assume that $\epsilon = 0.5$. Thus, in the first iteration, we take the upper 50% of the sample (instead of $(1-\epsilon)\times 100\%$) to find $L_2(x)$. Let $x_i^{(1)}$ be the x-intersection of $L_1(x)$ and $L_2(x)$. From Fig. 7, it is apparent that $x_i^{0.5} < x_i^{(1)} < x_i^{\epsilon}$ and $\hat{\epsilon}_{0.5} < \hat{\epsilon}^{(1)} < \hat{\epsilon}$. Continue this iteration and we conclude that at the nth stage $x_i^{(n-1)} < x_i^{(n)} < x_i^{\epsilon}$ and $\hat{\epsilon}^{(n-1)} < \hat{\epsilon}^{(n)} < \hat{\epsilon}$. When n is large enough, $x^{(n)}$ approaches x_i^{ϵ} and $\hat{\epsilon}^{(n)}$ approaches $\hat{\epsilon}$.

The same procedure can be used to find the initial estimate of ϵ in the Gaussian and Laplacian mixed

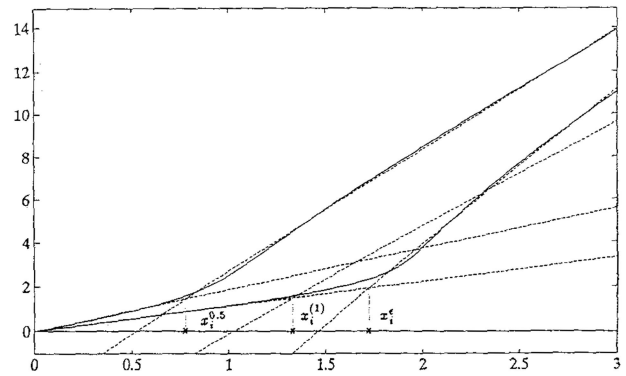


Fig. 7. Iterated initial estimate using theoretical QQ-plot.

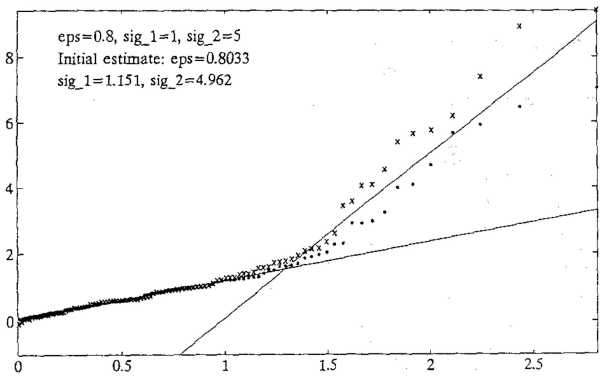


Fig. 8. Initial estimate using empirical QQ-plot (Gaussian + Gaussian).

distribution. If (39) is used, x_i corresponds to the intersection of $L_1(x)$ and the quadratic function. If (45) is used, x_i corresponds to the intersection of $L_1(x)$ and $L_2(x)$. In the following, we provide some simulations to test proposed schemes.

V. SIMULATIONS

In this section, we carried out some simulations to test the performance of the proposed algorithms. Note that the shape of the QQ-plot is only dependent on the variance ratio of two mixed components. The same ratio will give the same shape. Since our distributions are zero mean and symmetrical, the theoretical QQ-plot will be symmetrical with respect to zero. Thus, we can reflect samples with negative value to the positive side and make the empirical QQ-plot all positive.

To test proposed schemes, eight cases were simulated; they correspond to the combinations of two ϵs and four variance ratios. The ϵs used are 0.6 and 0.9, th evariance ratios are 9, 25, 49, and 100, and we performed the Monte Carlo simulation with 100 runs in each case. The sample size was 200. The resulting mean, variance, and mean square error (MSE) were taken as the performance indexes. Tables I and II list the results for the Gaussian mixture. Tables III and IV list the results for the Gaussian and Laplacian mixed distribution. We first examine our initial estimate scheme in Tables I and II. We find the estimate of ϵ was remarkably good, particularly when the variance ratio is high. The results are comparable with the ML estimates because the assumption of our initial estimate scheme is more valid when the variance ratio is high. The mean value of σ_1 estimate is larger than the actual value due to the influence of the $f_2(\cdot)$, however, the estimate variance is small. The mean of σ_2 estimate is also larger than its real value, which is the property of the QQ-plot. Only when x goes to infinity, does the slope of $L_2(x)$ approach σ_2, otherwise, it will be larger than σ. When ϵ is large, the number of samples used in estimating σ_2 is small, which results in the larger estimate variance. In general, the variance of our initial estimate is within three times of the ML estimate. Despite the bias, our initial estimate scheme works very well. Fig. 8 shows a sample run of the initial estimate. Marks with the plus

TABLE I
Performances of Initial and ML Identification for Gaussian and Gaussian Mixed Distribution ($\epsilon = 0.6$)

	Initial estimate			ML estimate		
	Mean	Variance	MSE	Mean	Variance	MSE
$\epsilon = 0.6$	0.69320	0.00509	0.01378	0.59865	0.01093	0.01093
$\sigma_1 = 1$	1.41342	0.01861	0.18952	0.99952	0.02748	0.02748
$\sigma_2 = 3$	3.26376	0.21706	0.28663	2.96747	0.13791	0.13897
$\epsilon = 0.6$	0.65829	0.00257	0.00596	0.60822	0.00281	0.00288
$\sigma_1 = 1$	1.59436	0.01986	0.37313	1.01841	0.01151	0.01185
$\sigma_2 = 5$	5.95573	0.51796	1.43138	4.99993	0.22213	0.22213
$\epsilon = 0.6$	0.62210	0.00157	0.00206	0.59829	0.00138	0.00138
$\sigma_1 = 1$	1.68457	0.02983	0.49846	1.00432	0.01043	0.01045
$\sigma_2 = 7$	8.42606	0.93791	2.97154	6.97367	0.41682	0.04175
$\epsilon = 0.6$	0.60952	0.00126	0.00135	0.59714	0.00092	0.00093
$\sigma_1 = 1$	1.77233	0.02957	0.62607	1.01371	0.01000	0.01019
$\sigma_2 = 10$	12.2714	1.51834	6.67777	9.89850	0.76394	0.77425

TABLE II
Performances of Initial and ML Identification for Gaussian and Gaussian Mixed Distribution ($\epsilon = 0.9$)

	Initial estimate			ML estimate		
	Mean	Variance	MSE	Mean	Variance	MSE
$\epsilon = 0.9$	0.88050	0.00563	0.00601	0.88561	0.00472	0.00492
$\sigma_1 = 1$	1.06229	0.01029	0.01417	0.98496	0.00740	0.00763
$\sigma_2 = 3$	3.15629	1.10535	1.12978	2.95852	0.47649	0.47821
$\epsilon = 0.9$	0.89126	0.00069	0.00077	0.89509	0.00077	0.00080
$\sigma_1 = 1$	1.09940	0.00898	0.01886	0.99457	0.00412	0.00415
$\sigma_2 = 5$	6.09003	2.83989	4.02803	4.83176	1.11505	1.14335
$\epsilon = 0.9$	0.88971	0.00063	0.00074	0.89784	0.00039	0.00040
$\sigma_1 = 1$	1.12037	0.01087	0.02536	1.00391	0.00436	0.00438
$\sigma_2 = 7$	9.41867	4.95128	10.8009	6.83916	1.33671	1.36258
$\epsilon = 0.9$	0.88278	0.00046	0.00076	0.89499	0.00027	0.00029
$\sigma_1 = 1$	1.11735	0.01163	0.02540	1.00422	0.00473	0.00475
$\sigma_2 = 10$	13.1334	7.01397	16.8321	9.30372	2.35633	2.84113

TABLE III
Performances of Initial and ML Identification for Gaussian and Laplacian Mixed Distribution ($\epsilon = 0.6$)

	Initial estimate			ML estimate		
	Mean	Variance	MSE	Mean	Variance	MSE
$\epsilon = 0.6$	0.80187	0.00452	0.04527	0.58116	0.03111	0.03146
$\sigma = 1$	1.26495	0.01383	0.08402	0.98410	0.03578	0.03604
$\mu = 2.1213$	2.46817	0.59264	0.71295	2.21319	0.30882	0.31726
$\epsilon = 0.6$	0.76224	0.00278	0.02910	0.59154	0.00774	0.00781
$\sigma = 1$	1.45330	0.01619	0.22167	1.00470	0.01293	0.01295
$\mu = 3.5355$	4.02912	0.84423	1.08786	3.55265	0.40334	0.40363
$\epsilon = 0.6$	0.73083	0.00151	0.01863	0.59470	0.00360	0.00363
$\sigma = 1$	1.53829	0.01915	0.30891	0.98786	0.01124	0.01138
$\mu = 4.9497$	5.57411	1.21979	1.54940	5.04121	0.59353	0.59522
$\epsilon = 0.6$	0.71695	0.00172	0.01540	0.59500	0.00197	0.00199
$\sigma = 1$	1.66890	0.03000	0.47742	1.00795	0.01126	0.01132
$\mu = 7.0710$	7.83685	2.50788	3.09430	6.94502	0.98091	0.99680

TABLE IV
Performances of Initial and ML Identification for Gaussian and Laplacian Mixed Distribution ($\epsilon = 0.9$)

	Initial estimate			ML estimate		
	Mean	Variance	MSE	Mean	Variance	MSE
$\epsilon = 0.9$	0.88811	0.00905	0.00919	0.84528	0.02490	0.02790
$\sigma = 1$	1.03836	0.00902	0.01049	1.00008	0.00936	0.00936
$\mu = 2.1213$	1.86965	0.69653	0.75987	2.19407	1.23662	1.24191
$\epsilon = 0.9$	0.91361	0.00094	0.00113	0.87231	0.00764	0.00841
$\sigma = 1$	1.07545	0.00808	0.01377	0.99176	0.00670	0.00677
$\mu = 3.5355$	3.88457	2.87216	2.99399	3.55213	2.02840	2.02868
$\epsilon = 0.9$	0.91266	0.00059	0.00075	0.89174	0.00134	0.00140
$\sigma = 1$	1.09512	0.01097	0.02002	0.99752	0.00451	0.00451
$\mu = 4.9497$	5.60137	6.54428	6.90592	4.94402	3.47810	3.48124
$\epsilon = 0.9$	0.91209	0.00049	0.00064	0.89674	0.00064	0.00065
$\sigma = 1$	1.11286	0.01082	0.02355	0.99993	0.00512	0.00512
$\mu = 7.0710$	8.99289	7.95192	11.6453	7.38493	4.28215	4.38066

and the star sign represent samples in the positive and negative side of the original QQ-plot.

Next, we examine the ML results in Tables I and II. It seems that some properties of the ML estimate are similar to those of our initial estimate. When ϵ is larger, the following properties are observed: the estimate of ϵ and the estimate of σ_1 became better (due to the smaller influence of $f_2(\cdot)$); the variance of σ_2 estimate became larger. It seems that for larger variance ratio, the separation of two components is easier and the estimate of ϵ is better. However, because of the large variance, the estimate of σ_2 became poorer. In all cases, the ML estimate produced satisfactory results.

In Tables III and IV, we show the result of the initial and the ML identification of the Gaussian and Laplacian mixed distribution. Note that for comparison purposes, the parameters of μ were chosen such that they have the same variance ratios with previous ones.

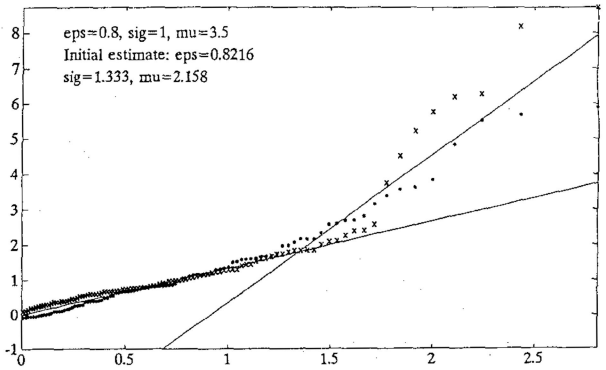


Fig. 9. Initial estimate using empirical QQ-plot (Gaussian + Laplacian).

From the tables, we find that the behaviors of the initial and the ML estimate are similar to those in Gaussian mixture case. Comparing with the results in Gaussian mixture, we find the following values becoming larger: the bias of the initial estimate of ϵ (for smaller ϵ), the variance of μ estimate (compare with σ_2), and the MSE of ϵ estimate. This reflects the fact that the Laplacian distribution is long tailed and more difficult to identify, however, note that in many cases, the increase of MSE is not significant. Fig. 9 shows a sample run of the initial estimate. From the figure, it is difficult to see that the curve in the tail region is a quadratic function, thus, the linear approximation is appropriate. In all the simulations, none of them failed, contrary to the moment method which has a high probability of failure. Also, the convergence of the ML method is rather fast. Thus, we conclude that our initial estimate scheme works effectively and the ML method using the estimate produces stable and superior results.

Finally, we applied the ML method to identify the model parameters using the glint data shown in Fig. 1. For the Gaussian mixture model, the resulting values are

$$\epsilon = 0.8354, \qquad \sigma_1 = 2.9084, \qquad \sigma_2 = 15.617. \tag{52}$$

For the Gaussian and Laplacian mixed model, the resulting values are

$$\epsilon = 0.7480, \qquad \sigma = 2.7366, \qquad \mu = 9.0450. \tag{53}$$

To verify the identification results, we plotted the empirical QQ-plot of the glint data and the theoretical QQ-plot of the identified models in Fig. 10. From this, we can see that these two models fit the data nicely. The spiky character of the glint noise is fully modeled. Note that in terms of QQ-plots, there is little difference between two models. Based on experimental results, we verify that the employed models are indeed appropriate. For the filtering purposes, the use of complicated hypergeometric functions is not justified.

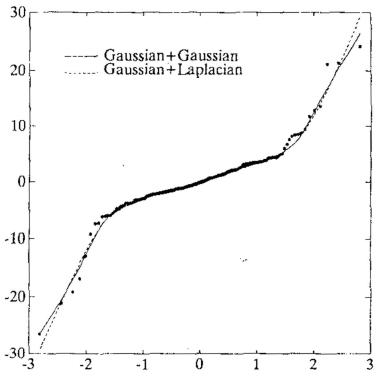


Fig. 10. QQ-plots of identified models.

VI. CONCLUSIONS

In this paper, we have proposed a method for identification of glint noise. The glint noise is highly non-Gaussian and it is present in radar systems. Two models are used for the glint distribution, one is the Gaussian mixture and the other is the Gaussian and Laplacian mixed distribution. The ML method is employed to identify the parameters of the models. An iterative initial estimate method based on the QQ-plot analysis is also proposed. Experiments show that the proposed initial estimate is very effective and the ML method has good performance. Once the glint noise is identified, the algorithm proposed in [7] can be applied to enhance the performance of tracking algorithm when the glint noise is present. Note that there is a growing interest in the non-Gaussian noise processing [22] and identification of this kind of noise is important. Our method is simple and efficient and it may find many applications in other areas.

REFERENCES

[1] Hewer, G. A., Martin, R. D., and Zeh, J. (1987)
Robust preprocessing for Kalman filtering of glint noise.
IEEE Transactions on Aerospace and Electronic Systems, **AES-23** (Jan. 1987), 120–128.

[2] Masreliez, C. J., and Martin, R. D. (1977)
Robust Bayesian estimation the linear model and robustifying the Kalman filter.
IEEE Transactions on Automatic Control, **AC-22** (1977), 361–371.

[3] Masreliez, C. J. (1975)
Approximate non-Gaussian filtering with linear state and observation relations.
IEEE Transactions on Automatic Control, **AC-20** (1975), 107–110.

[4] Masreliez, C. J. (1972)
Robust recursive estimation and filtering.
Ph.D. dissertation, University of Washington, Seattle, 1972.

[5] Tollet, I. H. (1975)
Robust forecasting.
Ph.D. dissertation, University of Washington, Seattle, 1975.

[6] Wu, W. R. (1989)
Kalman filtering in non-Gaussian environment using efficient score function approximation.
In *Proceedings of 1989 IEEE International Symposium on Circuits and Systems*, 1989, 413–416.

[7] Wu, W. R. (1993)
Target tracking with glint noise.
IEEE Transactions on Aerospace and Electronic Systems, **29** (Jan. 1993), 174–185.

[8] Pearson, K. (1984)
Contribution to the mathematical theory of evolution.
Philosophical Transactions of the Royal Society, London, Ser. A, **185** (1984), 71–110.

[9] Cohen, A. C. (1967)
Estimation in mixtures of two normal distributions.
Technometrics, **9** (1967), 15–28.

[10] Fukunaga, K., and Flick, T. E. (1983)
Estimation of parameters of a Gaussian mixture using the method of moments.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-5** (July 1983), 410–416.

[11] Derin, H. (1987)
Estimating components of univariate Gaussian mixtures using Prony's method.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-9** (Jan. 1987), 142–148.

[12] Harding, J. P. (1948)
The use of probability paper for the graphical analysis of polymodal frequency distributions.
Journal of the Marine Biological Association of the United Kingdom, **28** (1948), 141–153.

[13] Cassie, R. M. (1954)
Some uses of probability paper for the graphical analysis of polymodal frequency distributions.
Australian Journal of Marine and Freshwater Research, **5** (1954), 513–522.

[14] Bhattacharya, C. G. (1967)
A simple method for resolution of a distribution into its Gaussian components.
Biometrics, **23** (1967), 115–135.

[15] Postaire, J. G., and Vasseur, C. P. A. (1981)
An approximation solution to normal mixture identification.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-3** (Mar. 1981), 163–179.

[16] Rao, C. R. (1948)
The utilization of multiple measurements in problems of biological classification.
Journal of the Royal Statistical Society, Ser. B, **10** (1948), 159–203.

[17] Hasselbald, V. (1966)
Estimation of parameters for a mixture of normal distributions.
Technometrics, **8** (1966), 431–444.

[18] Hosmer, D. W., Jr. (1973)
On MLE of the parameters of a mixture of two normal distributions when the sample size is small.
Communications in Statistics, **1**, 3 (1973), 217–227.

[19] Day, N. E. (1969)
Estimating the components of a mixture of normal distributions.
Biometrika, **56** (1969), 463–474.

[20] Fowlkess, E. B. (1979)
Some methods for studying the mixture of two normal (lognormal) distributions.
Journal of American Statistical Association, **74** (Sept. 1979), 561–575.

[21] Huber, P. J. (1981)
Robust Statistics.
New York: Wiley, 1981.

[22] Wegman, E. J., Schwartz, S. C., and Thomas, J. B. (Ed.) (1988)
Topics in Non-Gaussian Signal Processing.
New York: Springer, 1988.

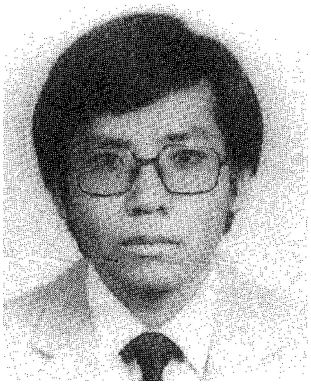
[23] Delano, R. M. (1952)
A theory of target glint or angular scintillation in radar tracking.
Proceedings of the IRE, (Dec. 1952), 1778–1784.

[24] Gubonin, N. S. (1965)
Fluctuation of the phase front of the wave reflected from a complex target.
Radio Engineering and Electronic Physics, **10**, 5 (1965).

[25] Gubonin, N. S., and Chapurskiy, V. V. (1971)
Covariance matrix of coordinate fluctuation of the instantaneous radar target centre of a group of reflectors.
Radio Engineering and Electronic Physics, (1971), 42–48.

[26] Sandhu, G. S., and Saylor, A. V. (1985)
A real-time statistical radar target model.
IEEE Transactions on Aerospace and Electronic Systems, **AES-21** (July 1985), 490–507.

[27] Hampel, F. R. (1974)
The influence curve and its role in robust estimation.
Journal of the American Statistical Association, **62** (1974), 1179–1186.



Wen-Rong Wu was born in Taiwan, R.O.C., in 1958. He received his B.S. degree in mechanical engineering from Tatung Institute of Technology, Taiwan, in 1980, the M.S. degree in mechanical and electrical engineering, and the Ph.D. degree in electrical engineering from State University of New York at Buffalo in 1985, 1986, and 1989, respectively.

Since August 1988, he has been a faculty member in the Department of Communication Engineering in National Chiao Tung University, Taiwan. His research interests include estimation theory, digital signal processing, and image processing.