# Nearest Neighbour Criterion – Analitical Performance Evaluation[1]

Ljudmil Bojilov
*Assoc. prof. in IPP, BAS*
*bojilov@bas.bg*

## Abstract.

*In this paper we consider the capabilities of nearest neighbor as a criterion of data association to ensure correct decision in associating measurement to target. We continue the work started in our previous paper to establish analytical approach end to derive explicit expressions for calculating correct data association probability for different levels of false alarm densities. Recurrent equation for correct association probability is derived. By means of this equation expressions for some particular values of FA densities are received. The correctness of the derivation is proved by the results of exhaustive Monte Carlo simulations.*

**Keywords: target tracking, gating, false returns, noise level.**

## 1. Introduction.

"Nearest neighbor" is the most simple, and in the same time, the most widespread rule (criterion) for data association in target tracking [2, 4]. The most popular suggestion for its implementation is that it is suitable for simple scenarios with low level of false alarms or without false alarms at all [2]. In our previous work [1] an attempt has been made to define more rigorously different 'levels of false alarms', used as a term very often in the target tracking literature. We have investigated how different number of false alarms per gate (as expected values) affects the filtration process when *nearest neighbor* as data association criterion is used. Relying on this investigation we have proposed four levels of false alarms: low level, medium level, high level and unacceptably high level.

The nearest neighbor approach will associate correctly the true measurement only in the case when this measurement lies closer to the predicted position of the target than the false return. In the cited above previous work we have investigated by Monte Carlo simulation probability of correct association in the cases when besides the true measurement in the target gate are fallen one, two, three *et cetera* false returns. In parallel, using calculus we have received the value of the probability for correct association in the case of two returns in the gate – true and false, and for one particular value of the gate threshold – $G = 9.21$. Value of this probability, calculated theoretically, has fully coincided with the value received from Monte Carlo experiments. In the current work we continue our efforts for derivation explicit expressions for mentioned above probability. In such expressions the gate threshold has to be an independent variable.

## 2. Problem formulation.

In target tracking gating [2, 3, and 4] is an important part of data association stage of the tracking. Gating is a technique for eliminating unlikely measurement-to-track pairings and very often is referred to as a coarse association [2]. By using predetermined gate threshold a gate is formed around the predicted track position (a solid line ellipse on figure 1). Assuming single target if a single measurement is fallen in the gate area than this measurement will be associated with the track for updating its track filter. If more then one return is fallen in the track gate (as in figure 1) the nearest neighbor correlation logic can be implemented. The problem posed for consideration is how can we calculate particular probability for correct association $P_{CA}$ if besides the true measurement in the target gate are fallen one, two and so on false returns.

Following the frame of the linear Kalman filter (KF) the measurement at scan $k$ is given by the equation

[1] This work was partially supported by Center of Excellence BIS21++, 016639, and MI-1506/05

$$z(k) = Hx(k) + w(k), \qquad (1)$$

where $H$ is a measurement matrix, $x(k)$ is a target state vector and $w(k)$ is zero mean, white Gaussian measurement noise with covariance $R$. Every cycle of KF starts with state prediction vector calculation

$$\hat{x}(k+1|k) = F(k)\hat{x}(k|k) \qquad (2)$$

where $F(k)$ is transition matrix and $\hat{x}(k|k)$ is updated state vector at the previous scan. Next step is to calculate the difference between the correlated measurement and its predicted position

$$\nu(k+1) = z(k+1|k) - H(k+1)\hat{x}(k+1|k) \qquad (3)$$

referred to as residual vector or innovation with residual covariance $S = HPH' + R$, where $P$ is one step prediction covariance matrix. Now an important measure for closeness of a given measurement to a particular target can be obtain: $d^2 = \nu' S^{-1} \nu$. Taking in consideration the $\chi^2$ (kai-square) distribution of this quantity an ellipsoidal region can be formed around the predicted state position

$$d^2 \le G \qquad (4)$$

with threshold $G$ chosen to insure predetermined gate probability $P_G$.

Every measurement fallen in the gate area at a *statistical distance* $d$ from predicted state position forms an ellipse (dashed line in fig. 1). All points on this ellipse lie at the same statistical distance $d$ from the predicted position. It is obvious that every other return fallen inside the dashed ellipse (return $f_1$ on the figure) will lie closer to the predicted position than measurement $m$. In this case the nearest neighbor rule will give incorrect association – it will correlate with the target the false return $f_1$ instead the true measurement $m$.

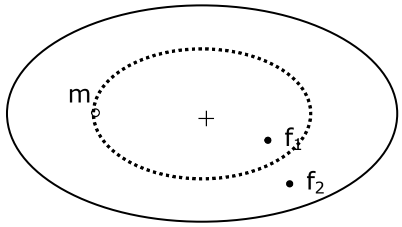


**Figure 1. Gate area around the predicted position with three returns in it**

All this considerations suggest the simple way to express the probability that given a true measurement fallen in the gate the false return fallen in the gate too will lie at a greater distance [1]. Remind that the false returns are independent and uniformly distributed in the surveillance region, the probability that false return will be closer to the predicted position is exactly equal to the ratio of the volumes of two curves – dashed and solid line ellipses. The solid line and dashed line ellipses are defined by the equations

$$\nu' S^{-1} \nu = G, \quad \text{and} \quad \nu' S^{-1} \nu = d^2.$$

Reminding that the volume of an arbitrary hyperellipsoid [4] is given by

$$V = c_n G |S|^{1/2},$$

where $c_n$ is the volume of unit hypersphere, for our 2D-case the two volumes are

$$V_s = \pi G |S|^{1/2} \quad \text{and} \quad V_d = \pi d^2 |S|^{1/2}.$$

And now, for the ratio of the volumes of the two curves and, hence, for incorrect association probability we have

$$P_{IA} = \frac{d_{ij}^2}{G}$$

and, on the contrary, the probability that the false return will fall outside the dashed ellipse and so, the nearest neighbor rule will give correct association, is

$$P_{CA} = \left(1 - \frac{d_{ij}^2}{G}\right). \qquad (5)$$

Having in mind that the event '*the false return is fallen outside the dashed line ellipse*' is independent, correct association probability when $k$ false returns have been fallen in the gate is

$$P_{CA} = \left(1 - \frac{d_{ij}^2}{G}\right)^k. \qquad (6)$$

## 3. Probability expressions derivation.

We can express now statistical distance squared by its variables

$$d_{ij}^2 = \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2},$$

where $x$ and $y$ are innovation's coordinates (3) and $\sigma_x^2$ and $\sigma_y^2$ – their variances (the main diagonal of the matrix $S$). As a common assumption, the true measurements are normally distributed around the predicted position and so, the probability density function of the vector $(x, y)$ is two-dimensional Gaussian distribution

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)}. \quad (7)$$

Here for simplicity we assume uncorrelated $x$ and $y$, i.e. matrix $S$ is diagonal. Now we will try to derive an expression for the overall probability for correct association given a particular value of the gate threshold. For arbitrary number of false alarms derivation will start with the next integration

$$P_{CA}(G) = \frac{1}{2\pi\sigma_x\sigma_y} \iint_{x,y\in E} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} \left(1 - \frac{\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}}{G}\right)^k dxdy \quad (8)$$

The integration is performed over the ellipsoidal volume $E$ (4) of the target gate. Implementing variable substitution $u = \frac{x}{\sqrt{2}\sigma_x}$ and $v = \frac{y}{\sqrt{2}\sigma_y}$ we have

$$P_{CA}(G) = \frac{1}{\pi} \iint_{C_G} e^{-(u^2+v^2)} \left[1 - \frac{2}{G}(u^2+v^2)\right]^k dudv. \quad (9)$$

Now integration is performed over the circle $C_G$ with radius $R = \sqrt{\frac{G}{2}}$ (equation (4)). The most straightforward way to continue integration is by coordinate transformation from Cartesian to Polar coordinate system with coordinates $(r, \varphi)$:

$u = r\cos\varphi$; $v = r\sin\varphi$ with $\left(0 < r < \sqrt{G/2}; 0 < \varphi < 2\pi\right)$.

Having in mind that Jacobian of this transformation is

$$J = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r$$

we have for the new integral

$$P_{CA}(G) = \frac{1}{\pi} \int_{r=0}^{\sqrt{G/2}} \int_{\varphi=0}^{2\pi} re^{-r^2} \left(1 - \frac{2}{G}r^2\right)^k d\varphi dr$$

and after direct integration over $\varphi$ we obtain

$$P_{CA}(G) = 2\int_{r=0}^{\sqrt{G/2}} re^{-r^2} \left(1 - \frac{2}{G}r^2\right)^k dr = I_k. \quad (10)$$

We will now derive recurrent expression for the above integral. As a first step to calculate the probability from (10) for the case $k = 0$, i.e. when no false alarms there are in the gate area

$$I_0 = 2\int_{r=0}^{\sqrt{G/2}} re^{-r^2} dr = -e^{-r^2}\Big|_0^{\sqrt{G/2}} = 1 - e^{-\frac{G}{2}}. \quad (11)$$

There is no surprise – the correct association probability for this case is exactly equal to the probability for a measurement to fall in the gate (gate probabilty $P_G$), given the gate threshold $G$.

And for the recurrent expression one obtains

$$I_k = 2\int_{r=0}^{\sqrt{G/2}} re^{-r^2} \left(1 - \frac{2}{G}r^2\right)^k dr =$$

$$= -\int_{r=0}^{\sqrt{G/2}} \left(1 - \frac{2}{G}r^2\right)^k de^{-r^2} = \cdots$$

$$\cdots = 1 - \frac{2k}{G} 2\int_{r=0}^{\sqrt{G/2}} re^{-r^2} \left(1 - \frac{2}{G}r^2\right)^{k-1} dr \quad (12)$$

But in the above expression after the last sign of equality the integral has the same structure as $I_k$ with the only difference – the power of the expression in parentheses is $k-1$, and so, the equation (12) can be rewritten as

$$I_k = 1 - \frac{2k}{G} I_{k-1}. \quad (13)$$

Using recurrent expression (13) and the result (11) we can now calculate the correct association probability for any integer as a value of $k$. Thereby, for a single false return besides the true measurement ($k = 1$) the result is

$$P_{CA}(G) = 1 - \frac{2}{G}\left(1 - e^{-\frac{G}{2}}\right). \qquad (14)$$

For a two false alarms and a true measurement ($k = 2$) in the target gate as is on the figure 1 the correct association probability is

$$P_{CA}(G) = 1 - \frac{4}{G} + \frac{8}{G^2}\left(1 - e^{-\frac{G}{2}}\right). \qquad (15)$$

And the last result is for three false returns besides the true measurement ($k = 3$)

$$P_{CA}(G) = 1 - \frac{6}{G} + \frac{24}{G^2} - \frac{48}{G^3}\left(1 - e^{-\frac{G}{2}}\right). \qquad (16)$$

The last four expressions (from 13 to 16) and the result (11) are the main achievements of the presented paper. The explicit expressions for correct association probability may be used in different kinds of investigations for replacing exhaustive Monte Carlo simulations. Even though exhaustive, results of Monte Carlo simulation give us filling of certainty. So, following the opposite logic we can try to verify correctness of our derivation by estimating the probability of correct association for some particular cases implementing Monte Carlo simulation and comparing these results with the corresponding results directly calculated with the above expressions.

## 4. Numerical experiments.

For our experiments we chose five different values of gate thresholds and for every one value we investigate the probability for correct association with one, two and three false returns besides the true measurement. The values of gate thresholds and corresponding gate probabilities are:

$$G_1 = 9.21(P_{G_1} = 0.99);$$
$$G_2 = 7.824(P_{G_2} = 0.98);$$
$$G_3 = 7.013(P_G = 0.97);$$
$$G_4 = 6.44(P_G = 0.96);$$
$$G_5 = 5.99(P_{G_3} = 0.95).$$

When the frame of numerical experiments is constructed we have to be careful to not miss to model the event '*the measurement is fallen in the gate*' with corresponding gate probability. In our experiments we have used simplified version of the Matlab routine given in [5] and realizing Desert-Musso algorithm for generating uniformly distributed points in hyperellipsoid (in our case – ellipse) for generating false alarms in gate area. In addition, for equal base for comparison, the number of false alarms generated was not random Poisson value but exact quantity.

**Table 1. $P_{CA}$ according derived expressions**

| | **Gate threshold (Gate probability)** | | | | |
|---|---|---|---|---|---|
| **m_fa** | **9.21 (0.99)** | **7.824 (0.98)** | **7.013 (0.97)** | **6.44 (0.96)** | **5.99 (0.95)** |
| **1** | 0.785 | 0.7495 | 0.7234 | 0.7018 | 0.683 |
| **2** | 0.659 | 0.6168 | 0.5874 | 0.564 | 0.544 |
| **3** | 0.5706 | 0.527 | 0.4974 | 0.4744 | 0.455 |

The Table 1 gives results for correct association probability calculated with the expressions derived above. For comparison the table with the same stricture is filled up (Table 2) with the results of correct association probability obtained from exhaustive Monte Carlo experiments. Every one value in the table is averaged over 100000 Monte Carlo runs.

**Table 2. $P_{CA}$ obtained from MC simulation**

| | **Gate threshold(Gate probability)** | | | | |
|---|---|---|---|---|---|
| **m_fa** | **9.21 (0.99)** | **7.824 (0.98)** | **7.013 (0.97)** | **6.44 (0.96)** | **5.99 (0.95)** |
| **1** | 0.7845 | 0.7493 | 0.7235 | 0.70199 | 0.6829 |
| **2** | 0.6586 | 0.6168 | 0.5873 | 0.564 | 0.5438 |
| **3** | 0.5703 | 0.5269 | 0.4976 | 0.4744 | 0.455 |

In Monte Carlo experiments every loop includes measurement generation using Gaussian generator. Next, Mahalanobis distance is compared with the corresponding gate threshold *G*. If the measurement occurs outside the gate area, we add unity to the counter of bad cases. Else, predetermined number of false alarms are generated and if even one of them is closer to the predicted position of the target in comparison with the true measurement, once again the bad cases counter is increased by one. At the end of the cycle the correct association probability is calculated dividing the total number of MC runs (T_runs) minus bad cases (B_runs) by T_runs

$$P_{CA} = \frac{T\_runs - B\_runs}{T\_runs}.$$

## 5. Discussion.

Comparison of the results from the two tables indicates the fully statistical coincidence for the corresponding correct association probabilities. Constructed recurrent frame (equations 11 & 13) give us opportunity, besides the equations (14-16), to derive an expression for correct association probability for arbitrary number of false alarms.

It is interesting to be pointed out, that according the results from MC experiments cited in table 1 of [1] association probability for, say, $m\_fa = 1$ as an exact value is less then the corresponding probability for $m\_fa = 1$, but accepted as an expected value for a Poisson generator. So, the probabilities calculated by the help of (14-16) are conservative estimation association probabilities of the real practice.

**Table 3. $P_{CA}$ for random number of false alarms**

| m_fa | G = 9.21 | | G = 7.824 | | G = 7.013 | |
|---|---|---|---|---|---|---|
| | **Appr.** | **MC** | **Appr.** | **MC** | **Appr.** | **MC** |
| **0.1** | 0.969 | 0.97 | 0.955 | 0.952 | 0.952 | 0.946 |
| **0.2** | 0.947 | 0.95 | 0.93 | 0.934 | 0.924 | 0.923 |
| **0.3** | 0.926 | 0.931 | 0.905 | 0.915 | 0.896 | 0.901 |
| **0.4** | 0.904 | 0.914 | 0.88 | 0.895 | 0.868 | 0.88 |
| **0.5** | 0.883 | 0.896 | 0.855 | 0.876 | 0.84 | 0.859 |

Derived expressions (11) and (14-16) give probabilities for predetermined exact values of the false alarms per gate, per scan. But more realistic picture is to accept the given false alarm value as an excepted value of some distribution (e.g. Poisson or Bernoulli). We have noticed, however, that by the means of (11) and (14) one can calculate good approximations of association probabilities corresponding to the values of $k$ between $0.1 \le k \le 0.5$, defined as excepted values of Poisson Distribution.

Table 3 contains the results of MC experiments (the right column for every one gate threshold value) compared with the corresponding probabilities (the columns headed with *Appr.* - approximated), calculated by the heuristic formula

$$I_k = I_0\left(1 - \frac{2k}{G}\right), \text{ for } 0.1 \le k \le 0.5 . \quad (17)$$

The results in this table, although not so precise as in the previous two tables, are comparatively good. Differences between analytical and experimental values are, in the worst case, in $1{,}5\% \div 2{,}5\%$ interval.

## 6. Conclusions.

In the presented paper an analytical approach is established for investigation and evaluation the association probabilities in the case, when nearest neighbor rule is used. Derived expressions (11) and (14-16), as well as the recurrent formula (13) give us possibility to estimate the implementation boundaries when using mentioned above rule in different specific cases replacing the time consuming exhaustive Monte Carlo simulation. Even though not so precise as expressions (14-16) heuristic formula (17) is derived for calculating association probabilities in the cases when the false alarms are not exact whole numbers.

## References:


1. Bojilov, L., "False Alarms Level in Target Tracking - Analysis and Definition" , In Proc. of *International Conference on Automatics and Informatics'2004*, Sofia, Bulgaria, October, 2004, pp. 45-48.
2. Blackman, S., *Multy-Target Tracking with Radar Applications*, Artech House, Norwood, MA, 1986.
3. Bar-Shalom, Y., Li, X.-R. *Estimation and Tracking – Principles, Techniques and Software*, Artech House, New York, 1993.
4. Bar-Shalom, Y., Li, X.-R, *Multytarget-Multysensor Tracking – Principles and Techniques*, 1995.
5. Dezert J., Musso C., "An Efficient Method for Generating Points Uniformly Distributed in Hyperellipsoids", *Proceedings of the Workshop on Estimation, Tracking and Fusion: A Tribute to Yaakov Bar-Shalom*, Naval Postgraduate School, Monterey, CA, May 17, 2001.