

Estimation of sample size influence to lengthened objects detection algorithm

Artem K. Sorokin* and Vladimir G. Vazhenin

¹Ural Federal University, Ekaterinburg, Mira str., 32, Russian Federation

Abstract. This paper is devoted to the important topic of choosing the sample size for the designed algorithm, which allows us to detect lengthened objects of the underlying terrain by the signal of the pulse radar altimeter. At first, here is given the short description of the algorithm and its distinctions from other existing algorithms. Then the abilities of the given algorithm are discussed. After all, the principle of choosing the number of samples in the train is performed. At last, the recommendations about choosing the lengthened objects, which could be detected, are presented.

1 Introduction

Nowadays there is a common problem of autonomous navigation for unmanned airborne vehicles [1]. Many decisions were implemented in real systems [2–4], but none of them works in all cases. So it is necessary to invent a new system, which will work in this exact case. As the result this problem has not been solved for numerous cases. On the other hand, the miniaturization process in many technical devices requires implementation of multifunctional systems. In this case existing systems get new possibilities. In accordance to the mentioned above, we designed a new correction system for the navigation system, which is based on the typical pulse radar altimeter. In the base of this system lies the process of lengthened objects detection. To the lengthened objects were attributed such objects of the underlying terrain, which exist on the flat landscape and composed from typical homogenous terrains. These objects should have borders, which could be presented by the straight line or the chain of straight lines. In the base of the detection algorithm lies the discrimination criterion, which finds the similarity between current distribution of the reflected amplitude values and the typical distribution of the reflected amplitude values, which was obtained from the known typical terrain, such as "forest", "meadow" and etc. Then we compare these distributions and compare the intersection area with threshold level. We do so for each typical terrain [5, 6], then we choose the most suitable terrain type or decide that none of the terrain could be corresponded to the current distribution. As the result, we are able to detect the changing of the terrain during the flight. In the base of this criterion we designed the algorithm, which can detect the border between typical terrains. Also this algorithm can identify types of these terrain to classify the lengthened object, which is composed from two or more typical terrain types. As the result, we can correct the position of the unmanned airborne vehicle. Obviously, not each combination of typical

* Corresponding author sorokinak@urfu.ru

terrains could be detected. This combination should have minimum contrast. In most cases, as it was obtained during the exploration, this contrast should be at least 10dB. Also we used another parameter, i.e. backscattering pattern width. The terrains with different backscattering patterns could be discriminated easier by this algorithm than the terrains with the same backscattering pattern. It is because we use histogram (numerical equivalent of distribution) as the parameter.

The important question is the robustness and the accuracy of the designed algorithm. The answer for these questions is the main purpose of this paper.

2 Criterion and algorithm

The important question is the robustness and the accuracy of the designed algorithm. The answer for these questions is the main purpose of this paper. At first we perform a criterion (1, 2), which was designed especially for the case of autonomous navigation system by the radar signal.

$$\Phi_{j,q}(t) = \sum_i \min[f_j(u); f_q(u)] \tag{1}$$

$$\Lambda(q,t) = \operatorname{argmax}_j [\Phi_{j,q}(t)] \tag{2}$$

Here $\Phi_{j,q}(t)$ – the comparison functional; $f_j(u)$, $f_q(u)$ – the current and etalon histograms of the reflected signal amplitude; j, q – indexes of j -th etalon and current signal probability density correspondently. $\Lambda(q,t)$ – decision rule; $\operatorname{argmax}[*]$ – chooses the argument of maximum coincidence; t – time, during which the counts were accumulated.

The comparison functional [7] allows us to detect the equivalent of coincidence probability of current and etalon probability density [8–10]. This functional have a parameter: the coincidence probability. Such a parameter can be a numerical measurement of coincidence of current and etalon histogram. The decision rule uses this parameter to choose the probability density with maximum intersection area. This probability density is corresponded to some typical terrain. We assume, that the current probability density could be referred to the terrain with maximum a posteriori probability, i.e. the probability density which provided maximum intersection area. In case of insufficient a posteriori probability, the algorithm outputs zero. It means, that no decision could be obtained. We can imagine this in accordance to the following scheme (see Fig. 1):

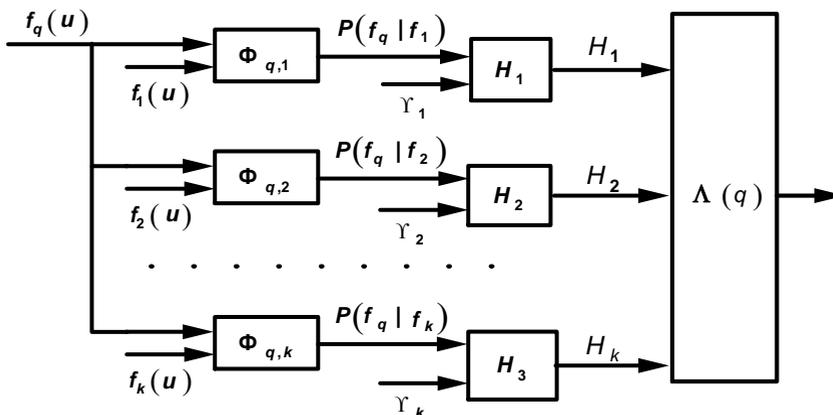


Fig. 1. Scheme of the decision rule work logic.

This scheme contains comparison functional, threshold detector and decision rule. Here the threshold detector blocks the output in case if value of a posteriori probability is lower, than threshold level, which can be chosen experimentally for each typical terrain. At this stage, when we applied decision rule, we have a decision about the typical terrain, which suits for the current train of the reflected signal amplitudes or zero in other case.

Now it is time to classify the detected lengthened object type. We use rather natural classification. Here it is necessary to mention, that the flight speed is constant or near to the constant value. If so, we can use time equivalent to the flight distance. In case of our algorithm it means the train of counts, which used in the current histogram. In other words, we can use number of counts instead of flight distance. After this note we can formulate the rule according to that we divide the lengthened objects to the stripe objects and borders. To the stripe object we can refer objects, which width is less than exposure spot's size, otherwise an object will be classified as the border. Such a classification allows us to get estimations of necessary number of counts, which we need to take in order of obtaining the histogram of the reflected signal amplitude. It is because in case of stripe objects we have limited number of counts, which is proportional to the lengthened object width. Also we have to notice, that in case of our algorithm we have to choose for the correction parts of the flight trajectory where the lengthened objects could be found. I.e. these parts should meet following requirements: they should have the geometry of the lengthened object in accordance to the mentioned above, they should have enough contrast to detect them.

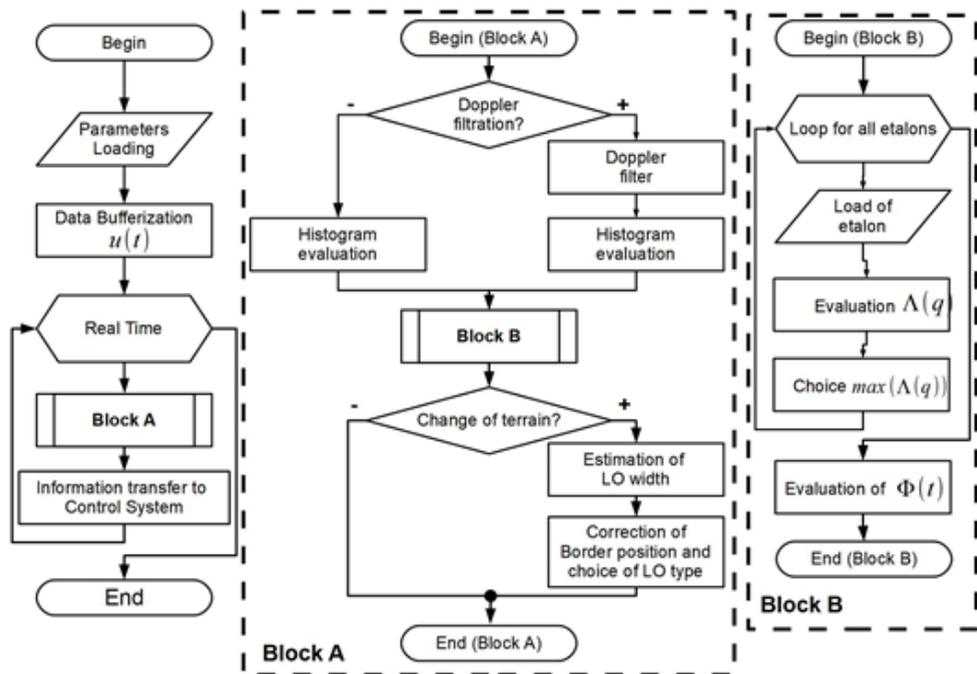


Fig. 2. Structural scheme of lengthened objects detection algorithm.

Another thing about which we have to say is the Doppler filtration. In the context of the suggested algorithm we can refer to the Doppler filtration such processing methods as Doppler beam sharpening and multiple band Doppler filtration. In each case we use Doppler frequency shift for filtration. This frequency shift appears as the result of vehicle moving relatively to the ground surface. But this shift is rather small to detect it inside of the single pulse, as the result we have to process whole pulse train. Such a processing allows us to increase the sensitivity of the developed algorithm to the terrain type, also it leads to the sharpening of obtained histogram of the reflected signal amplitude. But this

process has one distinction. If we increase the sensitivity, we obtain less robustness decision about terrain type especially in case of small number of counts and high level of pulse noise. Especially this notice is actual in case of highly noised terrains as "forest" or "bushes", but the Doppler filtration works well in case of such terrains as "asphalt" or "concrete" terrain. So we need to combine these two methods to obtain better result, in the model experiments we obtained twice more detection of narrow lengthened objects (about 0.2 of the exposure spot width) of "water" and "asphalt" type. In model experiments we obtained, that better (about 0,3 increase of detection probability) results of the Doppler filtration could be gained in case of nadir direction filtration.

After the given description we can build the scheme of the algorithm (see Fig. 2). This scheme can operate in the real-time. At first, we load parameters, such as pulse train, antenna, vehicle parameters and etc. Then in real-time we collect and process data. We choose the processing type: it can be as Doppler filtration as typical algorithm without Doppler filtration. Then accumulated data could be processed to evaluate current histogram of the reflected signal amplitude. Then obtained histogram could be processed in accordance to the designed algorithm. In a loop this histogram compared to the etalon histograms and algorithm evaluates an intersection area. This area used by as the parameter of the algorithm. If the evaluated square is less than the threshold level, the zero is passed to the output. After comparing with all the etalons algorithm choses the most suitable one, i.e. with maximum a posteriori probability. This etalon is accepted as the terrain type of current histogram. During time the algorithm replaces counts in accordance to their order in time. Then algorithm fixate the moment, when the terrain was changed and its width. After that it classify the lengthened object (is it stripe or border object) and terrain types, which combines this object. This information is passed to the navigation system for the following correction of the unmanned airborne vehicle.

3 Sample size and accuracy

Now we can go to the subject of this paper, i.e. estimation of the necessary number of counts, which provides the desirable accuracy. This problem presents main interest for the most known nonparametric algorithms. Also this problem has no any universal analytical decision. In most cases the algorithm designer should use numeric methods to obtain necessary estimations of important parameters. In our exploration we also used numeric methods to get various estimations. In case of our algorithm we can divide errors to some groups.

The first group are errors, which are connected to the histogram building. They are shown in fig. 3.

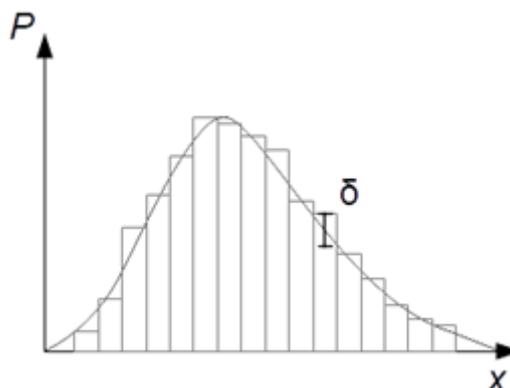


Fig. 3. The probability density of the reflected signal amplitude.

Here we can see, that errors of histogram approximation could be of two types: the ambiguity in the horizontal direction (the signal amplitude ambiguity) and ambiguity in the vertical direction (the histogram approximation ambiguity). Both of these errors are connected to two questions: how many counts we should use for each bin of the histogram and what is the accuracy of the histogram approximation. To answer for these questions, we refer to work from 70th of D.W. Scott [11]. There we can find the formula of the necessary number of counts, which are needed for the histogram approximation (3).

$$n_{\text{bins}} = c\sqrt[3]{N} \tag{3}$$

In this formula c – the coefficient, which value in our case can be from 1 up to 3; N – the number of counts, which were used for the histogram; n_{bins} – the number of the histogram bins. Also in this paper we can reveal that fluctuation error could be approximated by the following formula (4):

$$\delta = \frac{c_\delta}{\sqrt[3]{N}} \tag{4}$$

Here δ – the fluctuation error; c_δ – the coefficient, which value in our case lies between 1 and 3. On the one hand, this fluctuation error depends on the number of counts, which were taken for histogram. On the other hand, the number of bins also depends on the number of counts. So, in accordance to Scott's paper we just need know only the number of counts to predict our accuracy of histogram approximation.

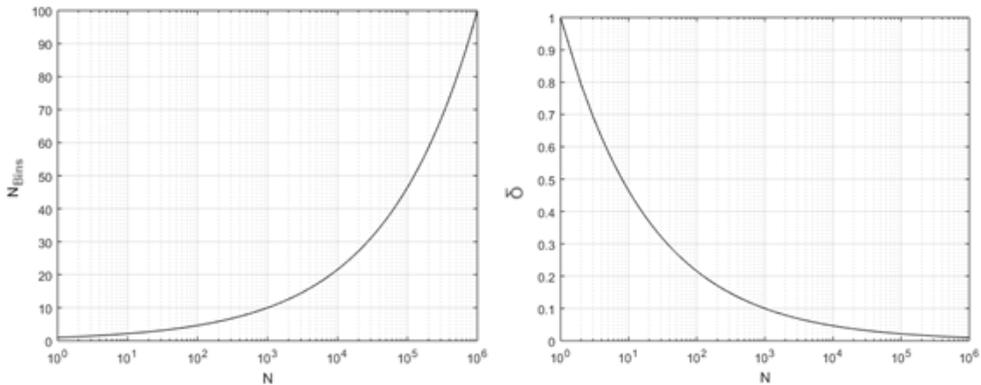


Fig. 4. The dependences $\delta(N_p), n_{\text{bins}}(N)$ for choosing the necessary accuracy.

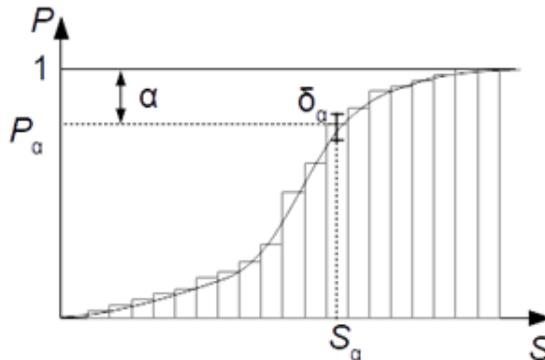


Fig. 5. The probability distribution of the reflected signal amplitude.

In fig. 4 we performed these dependencies. Nevertheless, if we go to other papers, we can reveal some other dependencies, and these parameters will be proportional to the square

root of 2 – 5 degree of N with some complex coefficients. But according to the resent researches most of these complex dependencies designed for special cases and they don't give significant accuracy increase in our case.

Other group of errors we can reveal arises when we come to the question: what is the confidence level of our decision? Here in accordance to traditional analysis method we can draw a cumulative distribution (or integral distribution) of our algorithm parameter, i.e. a posteriori probability, which numerical equivalent could be obtained as intersection area of two histograms: current and etalon. An error, which presents the most interest, is the error of cumulative distribution approximation. It is presented in fig. 5.

Here S – an intersection area; α – the confidence level; S_α – the intersection area at α confidence level; P_α – the a posteriori probability at this confidence level; δ_α . This distribution allows us to get estimation of the confidence level. Also there arises the question: what is the number of necessary counts (intersection areas)? During our exploration we revealed semi-empirical formula (5):

$$\frac{c_\alpha}{\sqrt{N_p}} \leq \delta_\alpha \leq \frac{2c_\alpha}{\sqrt{N_p}} \tag{5}$$

Here N_p – the number of intersection areas, which were taken for the cumulative distribution; c_α – the coefficient, which value lies between 1 and 1.5. Also this dependency with $c_\alpha = 1$ is shown in fig. 6.

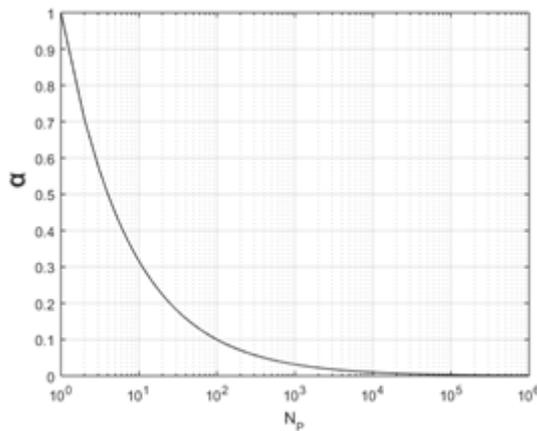


Fig. 6. The dependency $\alpha(N_p)$ for choosing the necessary accuracy.

In accordance to the obtained dependencies we can predict necessary amount of counts which provides desirable accuracy. For example, for accuracies $\delta = 0.2$ and $\alpha = 0.1$ we need in $N_{sum} = N \cdot N_p = 12500$ counts. Another example, for accuracies $\delta = 0.3$ and $\alpha = 0.3$ we need only $N_{sum} = 412$ counts and etc.

Other important question is how many counts we can obtain per 1m of the track? For this algorithm important to have independent counts, therefore, to answer this question helps following formula (6):

$$I = \frac{\lambda}{2\Theta_{ef}} \tag{6}$$

Here λ – the wavelength; Θ_{ef} – the effective antenna pattern width; I – the trajectory correlation interval. In accordance to this formula, correlation interval depends on the wavelength and antenna parameters. The effective antenna pattern could be narrowed by

the Doppler filtration or by the changing the antenna geometry. The wavelength could be decreased only if the hardware changed. So, for the case of our paper, we use 7cm wavelength and only Doppler filtration can reduce the correlation interval and as the result, the number of counts can be increased. In the case of the unsharpened antenna beam for the antenna pattern width $\Theta_{ef} = 40^\circ$ we have $I = 0.07\lambda = 0.05$ m. I.e. we have 20 counts per 1m of the track. Similar conclusion could be obtained for other wavelength. As the result we can evaluate the needed width of the lengthened object, which could be detected with desirable accuracy. For example, if we have wavelength $\lambda = 0.07$ cm, $\delta = 0.2$, $\alpha = 0.1$, $\Theta_{ef} = 40^\circ$, we can detect lengthened objects with width 60m. The same calculation could be done for each typical terrain objects. There it is necessary to notice, that this calculation could be implemented for objects, which are possible to detect (see [12]); the altitude is from 100m up to 500m (this parameter influences indirectly to the Θ_{ef}). So, in this paper we discussed important question of necessary sample size for obtaining desirable accuracies and needed width of lengthened objects, which should be detected by the designed algorithm.

4 Conclusion

This paper is devoted to the creation of correction system for the autonomous navigation system of unmanned airborne vehicle. Such a system could be implemented in case, when navigation by landform is impossible or have significant difficulties. One distinction of the designed correction algorithm is that this algorithm does not need any additional hardware, it uses standard pulse radar altimeter as the sensor. As correction objects this system uses lengthened objects, which can be as artificial as natural objects. As the algorithm parameter we use the histogram of the reflected signal amplitude. That allows us to enhance functionality of existing on-board systems. At first we presented the designed algorithm, which can detect lengthened objects of two types ("border" and "stripe") and reveal types of terrain, which composes this lengthened object. The main attraction of the paper is concentrated on the sample size influence to the accuracy of the designed algorithm. We discussed in detail the dependencies, which allow us to choose necessary number of counts to provide needed accuracy. Also we made some recommendations about choosing the lengthened objects as correction objects. Thus it allows us to estimate abilities of algorithm application in real systems for further implementation of the correction system.

The research was executed by the grant of the Ministry of education and science of the Russian Federation (project № 8.2538.2017/PCh).

References

1. M.I. Skolnik *Radar Handbook* (3rd edn. The McGraw-Hill Companies, 2008)
2. Y. Zhigui and L. ChuanJun, in 2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST), pp. 453–457 (2017)
3. Zhu Yunfeng, Sun Yongrong, Huang Bin, Li Rongbing, and Wu Ling, in 2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC), pp. 2076–2081 (2016)
4. Zhuoning Dong, Wenbin Li, and Yanxing Zhou, in 2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC), pp. 982–987 (2016)
5. F.T. Ulaby, M.C. Dobson *Handbook of radar scattering statistic for terrain* (Artech house, USA, 1989)

6. A.K. Fung, K.S. Chen *Microwave Scattering and Emission Models for Users* (Artech house, USA, 2010)
7. T.W. Anderson *An Introduction to Multivariate Statistical Analysis*. (John Wiley & Sons, Inc., 3rd edn., 2003)
8. H. Yang, Y. Fu, and F. Qian, in 2018 37th Chinese Control Conference (CCC), pp. 292–296 (2018)
9. Y. A. Kropotov and A. A. Belov, in 2016 13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), pp. 69–72 (2016)
10. R. B. Sinitsyn, in 2006 European Radar Conference, pp. 178–181 (2006)
11. D. W. Scott, in *Biometrika* **66** pp. 605–610 (1979)
12. A. K. Sorokin, V.G Vazhenin in International Conference on Signal Processing and Communication (ICSC 2018) Lecture Notes in Electrical Engineering pp. 199–209 (2019)