A METHOD FOR DETERMINING THE PARAMETERS OF AN OPTICAL SYSTEM DURING THE SPATIAL SCANNING OF A BIOLOGICAL SOLUTION

Y. E. Megel, I. I. Kalimanova, V. V. Shlyakhov, S. I. Kovalenko, A. V. Rybalka, S. N. Kovalenko

Abstract: In research, control of biological processes for determining the motility of microorganisms, (estimation of the impact of chemical compounds, electrophoretic mobility, evaluation of severity of pathology, etc.) using a scanning optoelectronic system as the most effective means of solving applied problems, it is necessary to determine the objects number in the optical system field of vision since the measurement accuracy of the solution concentration and the objects motion speed in the solution depend on their number. In the article a probabilistic method for estimation the number of objects and the number of frames during frame-by-frame filming for providing a given accuracy with a given probability is presented.

Key-Words: biological object, optoelectronic system, measurement, motility, possibility.

1. Introduction

To measure the objects linear dimensions and their speed optical methods have been widely disseminated. The main advantages of which are non-contact way, high measurement accuracy and high spatial resolution, obtaining information about the form of the object. Using optical methods of measurement of specified parameters, it is possible to solve a wide range of problems related to research in microbiology, medicine, biology, agriculture [1]. The purpose of optoelectronic tracking system is the monitoring of the relative displacements of the object located in the system field of vision and emitted or reflected electromagnetic waves in the optical range. The result of the study of the parameters of such objects is the measurement of the speed, trajectory, and rhythm of the motion of individual cells by frame-by-frame scanning of the distance travelled by the cell for a certain time interval [2, 3]. A necessary condition for finding these parameters is the accuracy of estimating the measurements of such a system, while an important element is the determination of the number of objects and the number of measurements, which ensures a given accuracy of such a system and limits of its measurement.

2. Statement of the problem of determining the number of observed objects during scanning a solution

Set a problem of determining the sample size so that the point estimation of probability of emergence of an object of a specified type of motility will be registered by the system.

Having fixed some kind of motility (in this case it makes no differences which one), it is possible to imagine the situation in the form of a Bernoulli scheme. It means that in a single test the occurrence of an event A, when the object has a specified type of motility, is recorded. It is known [4] that an unbiased, consistent and effective point estimate of the probability of occurrence $\widetilde{P}(A)$ of an event A is

$$\widetilde{P}(A) = k_1 / k_2 \ , \tag{1}$$

where k_1 – the number of occurrences of an event A, and k_2 – the total number of tests. However, with this solution it is impossible to speak about the accuracy of this estimation depending on k_2, except the fact, that it converges in probability to the true value at an unlimited increase of the total number of tests. Therefore, it makes sense to operate with interval estimates, i.e. with the specified confidence interval (it is determined by required accuracy) and with specified confidence probability to find an interval (a,b), which covers the true value with the specified probability.

Hence it is necessary that an estimate of a form (1) possesses sufficient accuracy with a high probability. Using the central limit theorem for the Bernoulli scheme, we can write:

$$\begin{aligned} &P\{|k_1 / k_2 - p| > \varepsilon\} = \\ &= P\{|(k_1 - k_2 p) / \sqrt{k_2 pq}| > \varepsilon\sqrt{k_2} / \sqrt{pq}\} \approx \\ &\approx 2\Phi(-\varepsilon\sqrt{k_2} / \sqrt{pq}). \end{aligned} \tag{2}$$

where $\Phi(X)=\left(1/\sqrt{2\pi}\int_{-\infty}^{X}e^{-t^2/2}dt\right)$ – is the error function.

Taking into account that $p+q=1$, then $pq=p(1-p)$ and $f(p)=p(1-p)$ reaches max in the point $p=1/2$, i.e. the maximum value of $\sqrt{pq}$ is equal to $1/2$. Then

$$-\varepsilon\sqrt{k_2}/\sqrt{pq}\le -2\varepsilon\sqrt{k_2}\,. \qquad (3)$$

Since the error function is increasing one, then

$$\Phi\left(-\varepsilon\sqrt{k_2}/\sqrt{pq}\right)\le\Phi\left(-2\varepsilon\sqrt{k_2}\right)$$

or

$$P\{|k_1/k_2-p|>\varepsilon\}\le 2\Phi\left(-2\varepsilon\sqrt{k_2}\right). \qquad (4)$$

Thus, by setting the accuracy ε and α – the confidence probability, with which it is necessary to provide the specified accuracy, we obtain equality:

$$\Phi\left(-2\varepsilon\sqrt{k_2}\right)=\alpha/2\,. \qquad (5)$$

Since often the error function in tables is given in the form

$$\Phi^*(X)=\left(1/\sqrt{2\pi}\right)\int_0^X e^{-t^2/2}dt, \qquad (6)$$

we obtain the equality (5) by the following sequence of operations

$$\Phi(X)=1/2+\Phi^*(X),$$
$$1/2+\Phi^*\left(-2\varepsilon\sqrt{k_2}\right)=\alpha/2\,. \qquad (7)$$
$$\Phi^*\left(2\varepsilon\sqrt{k_2}\right)=(1-\alpha)/2$$

Denote by γ a quantile grade level $(1-\alpha)/2$ of function Φ^*, i.e.

$$\Phi^*(\gamma)=(1-\alpha)/2\,.$$

Then, having chosen $\alpha=0.1;\ 0.05;\ 0.01;\ 0.001$, we will get the value for γ, presented in the table 1.

Table 1. The arguments of the error function when the function takes the values $(1-\alpha)/2$

α	0.1	0.05	0.01	0.001
γ	1.65	1.96	2.58	5

For k_2 from (7) we will have

$$k_2=\left[\frac{\gamma^2}{4\varepsilon^2}\right], \qquad (8)$$

where k_2 – the number of objects in a test, ε – the accuracy of the deviation of the frequency of emergence of the motile objects from the true value, γ – the quantile of confidence probability.

It is clear from (8), that an increase in accuracy by an order of magnitude leads to an increase in the total number of tests by two order of magnitude. Give below a table of values k_2, depending on α and ε.

Table 2. The minimum number of observed objects during scanning a single frame

ε \ α	0.1	0.05	0.01	0.001
0.1	67	96	166	625
0.01	6724	9604	16641	62500
0.001	672400	960400	1664100	6250000

This table shows that the minimum number of objects that must get into count. Having recognized among them a certain number of motile ones, it is possible to obtain an estimate of the probability of the emergence of an motile object in the frame in the form of frequency (7). In this case, the deviation of the frequency from the true value should be less than ε and the event occurs with probability $1-\alpha$.

For example, to make the estimate in form of frequency differ from the true value by less than 0.01 (1%) with the probability of $\alpha=0.95$, a minimum of 9604 objects should be observed. On the other hand, if $\alpha=0.5$, then to within 0.1 (10%) in half of the cases, it is possible to obtain an estimate when observing 11 objects.

For the automatic monitoring system for estimation the motility of microbiological objects at specified values $\alpha=0.9$ and accuracy of 0.1 (10%) close to the optimal value [5], 96 objects should be monitored.

3. The method of solving the problem of determining the number of scanned frames for a specified accuracy and a specified probability

Formulate the statement of such a problem as a follows: it is necessary to evaluate the objects number per unit of volume of analysed solution. By taking a sample, it is possible to find in it the total number of objects, that are in the system field of vision and this number will be random. The distribution of this random value, strictly speaking, is unknown, but it can be considered as normal one, since a lot of independent random factors impact. It should be noted, that the problem of checking the statistical hypothesis of the distribution of this random value can be set and solved. Since the number of observed objects in the system field of vision is large, we specify the normal distribution law.

Based on the problem statement, where ξ – the objects number in a frame and random value is normally distributed with parameters α and γ, for determining the concentration it is on interest to estimate its expected value.

The statistics it is possible to collect by getting realisations of the random value ξ in *m* samples. Denote these realisations by $x_1,...,x_m$. It is known that an interval estimate α with a given reliability γ and with a known variance is as follows

$$x - t\sigma/\sqrt{m} < \alpha < x + t\sigma/\sqrt{m}, \qquad (9)$$

where $t\sigma/\sqrt{m} = \delta$ – an accuracy of the estimate; *m* – the sample size (the number of frames); *t* – quantile of the error function $\Phi(t)$,

$\left(\Phi(t) = \left(1/\sqrt{2\pi}\right)\int_0^t e^{\frac{-x^2}{2}}dx\right)$ of level $\gamma/2$, i.e.

$\Phi(t) = \gamma/2$ and $x = \sum_{i=1}^{m} x_i/m$; in case of an unknown variance

$$x - t(\alpha,k)S/\sqrt{m} < \alpha < x + t(\alpha,k)S/\sqrt{m}, \qquad (10)$$

where *S* – the so-called standard (unbiased variance estimation)

$S = \sqrt{\sum_{i=1}^{m}(x_i - x)^2/(m-1)}$, and $t(\alpha,k)$ – the critical point of the Student distribution with $k = m-1$ - the number of degrees of freedom and corresponding to the level of significance $2 = 1-\gamma$.

The ratios (9, 10) allow determining the minimum number of frames, providing the required with a specified probability.

In this case, the following options are possible. In the first one, assume that the variance is unknown to us, but it does not depend on the level of concentration. A variation from the true value of concentration depends on equipment parameters and on arbitrary motion of objects in the frame, but not on the solution concentration. Therefore it is possible to proceed as follows. Find a point estimate of variance, e.g. in the form of a standard S, for some biological solution, composed of randomly located motile bio objects, (taking, in so doing, once a sufficient large number of measurement $\approx 10^2$). Then during measurement the concentration for other solutions we assume the obtained earlier estimate as the true value of the variance σ^2. In this case we come to the first possible option in this case.

The variance σ^2 is known and it is necessary to obtain an estimation of objects concentration in the frame in the form of averaging objects numbers over samples. This estimation of objects concentration should have a specified accuracy δ with a given probability γ [2]. The ratio (10) allows finding the minimum number of scanning frames providing the conditions set out above.

$$m = t^2\sigma^2/\delta^2. \qquad (11)$$

Thus, at a concentration level of objects in the solution equal to about 10^2, we obtain the following values for different reliability levels (Table 3).

Table 3. Estimation of the number of frames for a specified accuracy

γ	0.95	0.975	0.99	0.999	0.9999
t	1.96	2.24	2.58	3.4	5
m	≈ 4	≈ 5	≈ 6	≈ 10	25

Based on the results of table 3, if the variance is commensurable with the required accuracy, m grows as a square of the quantile t^2 of the error function $\Phi(X)$ of level $\gamma/2$. In case of reliability coincides with 1 to within a fourth digit, we have $m = 25$. It means that estimate of 25 frames is practically guarantees us the specified accuracy in these conditions. Now assume that a second option takes place.

Now, suppose that the second option takes place, where the variance σ^2 is unknown and we do not have possibility to obtain its point estimate in a form of a standard. Then to find a minimal number of frames, that provides the estimate of the concentration of specified accuracy δ with the specified reliability γ, we can use the successive approximation method. From (11) it is clear, that accuracy determined by the ratio $t(\alpha,k)/\sqrt{m}$. Then carry out the following procedure. First for two measurement (two frames), i.e. $m = 2$, find $S = \sqrt{[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2]}$ and $t_2(1-\gamma,1)$. Hence we get $\delta_2 = t(1-\gamma)S_2\sqrt{2}$. Compare this value to δ. If $\delta_2 \le \delta$ the process should be stopped and the answer is $m = 2$, if $\delta_2 > \delta$ we take one more measurement (take one more frame) and find

$$S_3 = \sqrt{\sum_{i=1}^{3}(x_1 - \overline{x})^2 / 2}, \quad t_3(1-\gamma,2) \text{ and}$$

$$\delta_3 = t_3(1-\gamma,2)S_3/\sqrt{3}$$

Compare δ_3 to δ. If $\delta_3 \le \delta$ then $m = 3$ is an answer, otherwise take one more measurement (frame) etc. In some k-th step we will obtain:

$$\delta_k = t_k(1-\gamma, k-1)\cdot \cdot\sqrt{\sum_{i=1}^{k}(x_i - \overline{x})^2 /(k-1)}/\sqrt{k} \le \delta. \quad (12)$$

As soon as it happens, we can assume that resulting estimation of the number of measurements (frames) has the required accuracy for a given probability [4].

This accuracy, according to tables 2, 3, is optimal for the system under consideration [5].

4. Conclusions

The method of probabilistic estimation of the number of moving microbiological objects in the optoelectronic system field of vision was offered. It is shown that when scanning 25 samples (frames) with the concentration of a solution containing 96 objects, a specified accuracy with a high probability (to four decimal places close to one) is guaranteed while ensuring an appropriate resolution and a field of vision of the optical system.

6. References

[1] **R. P. Amann, D. Waberski.** Computer-assisted sperm analysis (CASA): capabilities and potential developments. *Theriogenology*, 2014, vol. 81, no. 1, pp. 5-17.

[2] **Y. E. Megel.** Metod otsenki maksimalnogo chisla ob'ektov pri avtomaticheskom analize opticheskoy informatsii. *Elektricheski izmervaniya, Dopovidi NAN Ukrayiny.* 2002, no. 10, pp. 62-67.

[3] **Y. E. Megel.** Model i metod resheniya zadachi diskriminatsii traektoriy pri skanirovanii biologicheskih ob'ektov. *Systemy obrobky ynformatsiyi: Zb. nauk. pr. Kharkiv: «Naukova dumka»,* 2001, V. 2 (12), pp. 117-121.

[4] **E. S. Venttsel** *Teoriya veroyatnostey.* Moskva: Vyisshaya shkola, 2006, 575 p.

[5] **Y. Megel, A. Rudenko, S. Kovalenko, M. Tarulli, G. Venkov.** *Operations Research, Calculus of Variations and Optimal Control*, Part II, TU-Sofia, 2016, p. 188.

Information about the Authors:

Yury Evgenevich Megel. MSc in radio-electronics (1972). DSc in Biological and Medical Devices and Systems (2004), Professor (2005) in the Department of Cybernetics, Educational and Scientific Institute of Business and Management, Kharkiv Petro Vasylenko National Technical University of Agriculture, Ukraine. Scientific field and interests: Digital Automation of Biotechnological Processes. Workplace: Kharkiv Petro Vasylenko National Technical University of Agriculture, Educational and scientific institute of business and management, Department of Cybernetics, Mailing address: 44, Alchevskikh st., Kharkiv, Ukraine, 61002.

Web address: *http://www.khntusg.com.ua;*
e-mail address: megel_ye@mail.ru

Ivanka Ivanova Kalimanova. MSc in Physics/

Physical optics (1969). PhD – Technical University of Sofia (1986), Associate Professor in the Department of Precision Engineering and Measuring Devices, Faculty of Mechanical Engineering, TU-Sofia, (1991). Scientific field and interests: Optical, Electro-Optical and Laser Devices, Optical Measurements. Workplace: Technical University Sofia, Mechanical Engineering Faculty, Laboratory "CMM", Mailing address: 8 "Kl. Ohridski" Blvd, 1756 Sofia, Bulgaria.

Web address: *http://www.tu-sofia.bg;*
e-mail: ikaliman@tu-sofia.bg

Vladislav Vladimirovich Shlyakhov. MSc in Mathematics (1979); DSc in Mathematics (2010); Professor (2015) in the Department of Higher Mathematics, Faculty of Information and Analytical Technologies and Management, Kharkiv National University of Radio Electronics, Ukraine. Scientific field and interests: Mathematical Modelling of Sensor Technical Systems. Workplace: Kharkiv National University of Radio Electronics, Faculty of Information and Analytical Technologies and Management, Department of Higher Mathematics, Mailing address: 14, Nauki ave., Kharkiv, Ukraine, 61166; Web address: *http://nure.ua;*
e-mail: divl@ukr.net

Antonina Ivanovna Rybalka. MSc in Physics (1969). PhD in Optics (1978), Associate Professor (1989) in the Department of Physics, Faculty of Automation and Computer Technologies, Kharkiv National University of Radio Electronics, Ukraine. Scientific field and interests: Optical systems, Modelling of Optical Systems. Workplace: Kharkiv National University of Radio Electronics, Faculty of Automation and Computer Technologies, Department of Physics, Mailing address: 14, Nauki ave., Kharkiv, Ukraine, 61166/

Web address: *http://nure.ua;*
e-mail: airybalka@mail.ru

Sergey Vladimirovich Kovalenko. MSc in Applied Mathematics (1996), PhD in Control Processes Automation (2015), Assoc. prof (2016) in Department of Systems Analysis and Information-Analytical Technologies, Faculty of Computer Science and Program Engineering, National Technical University "Kharkiv Polytechnic Institute", Ukraine. Scientific field and interests: Digital Image Processing, Pattern Recognition. Workplace: National Technical University "Kharkiv Polytechnic Institute", Faculty of Computer Science and Program Engineering, Department of Systems Analysis and Information and Analytical Technologies, Mailing address: 2, Kyrpychova str., Kharkiv, Ukraine, 61002.

Web address: *http://www.kpi.kharkov.ua*
e-mail: adbc@ukr.net

Svetlana Nikolayevna Kovalenko. MSc in Applied Mathematics (1996), PhD in Mathematical Modelling and Computational Methods (2009), Associate Professor (2011) in the Department of Cybernetics, Educational and Scientific Institute of Business and Management, Kharkiv Petro Vasylenko National Technical University of Agriculture, Ukraine. Scientific field and interests: Mathematical Modelling, Optimisation. Workplace: Kharkiv Petro Vasylenko National Technical University of Agriculture, Educational and Scientific institute of Business and Management, Department of Cybernetics, Mailing address: Alchevskikh st., 44, Kharkiv, Ukraine, 61002.

Web address: *http://www.khntusg.com.ua,*
e-mail: adbc@ukr.net