

# Intelligent High-Performance Computing for Big Data Processing in Fiber Optical Measuring Networks

Elena V. Zakasovskaya<sup>1</sup>, Valentin S. Tarasov<sup>1,2</sup>, Nadezhda I. Denisova<sup>3</sup>

<sup>1</sup>Vladivostok State University of Economics and Service, Vladivostok, Russia, elena.zakasovskaya@vvsu.ru

<sup>2</sup>Far Eastern Federal University, Vladivostok, Russia, valentin.tarasov@vvsu.ru

<sup>3</sup>Saint Petersburg University, St Petersburg, Russia, denisovanadezda0@gmail.com

## Abstract

The paper deals with the problem of reconstructing the parameters of physical fields using distributed information and measurement systems for cases of incomplete laying of measurement lines. High-performance computing is typically used for solving advanced problems and performing research activities through computer modeling. The rise of Big Data has changed the entire perspective of data and data handling. Ever growing analytical needs for Big Data can be satisfied with extremely high-performance computing models. A new combined algorithm is presented, which is concluded in the “optimization of the geometry” of the measuring network with a view to further applying the complex of neural networks. The possibility of choosing and using the appropriate neural network from a complex of several pre-trained.

## 1 Introduction

Computerization of almost all areas of modern life has a great impact on a human, on the majority of his activities and it contributes to development of information technology. In accordance with the modern tendency in the development of measuring instruments, in case of large amount of collected and processed information it is necessary to use not a lot of measuring instruments, but rather complex devices such as information and measuring systems [1].

Information measuring systems (IMS) is used to solve a wide range of applied problems however the main purpose is to provide continuous monitoring of large-scale and spatially inhomogeneous multidimensional physical fields [2].

What is important in the work of the IMS is the process of collecting information, on which the type of measuring network depends. Topology of the communication system depends on the choice of network technology and as a consequence the scope of application, types of input signals, types of measurements and functional properties of components. Examples of IMS with fundamentally different network topology are:

- Fiber-optic measuring networks (FOMN) based on fiber-optic information and measuring networks;
- Information-measuring systems based on wireless sensor networks (WSN).

Present day science intensive production cannot do without constant monitoring and control over the behavior dynamics of parameters range of distributed physical fields (PFs). The distributed information-measuring systems are called upon to solve this problem.

Information and measuring systems based on wireless sensor networks have great potential. This sort of low-power communication devices can be deployed over the entire area of almost any physical space, ensuring continuous monitoring of physical phenomena in real time, processing and transferring collected information, and coordinating actions with other nodes of the network. However, it is impossible to fully deploy intelligent distributed IMS using wireless technologies in critical infrastructures, primarily due to the lack of network technologies that meet information security requirements [3].

FOMN are of greatest interest. These systems have various topologies and organization, and are constructed, for example, on fiber-optical element base [1]. One of the fundamental parts of the distributed fiber optical measuring system (DFOMS) is distributed fiber optical measuring network responsible for collecting measuring information regarding the PF parameters under study [2].

---

*Copyright © 2019 for the individual papers by the papers' authors. Copyright © 2019 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).*

In: Sergey I. Smagin, Alexander A. Zatsarinnyy (eds.): V International Conference Information Technologies and High-Performance Computing (ITHPC-2019), Khabarovsk, Russia, 16-19 Sep, 2019, published at <http://ceur-ws.org>

FOMN represents a set of fiber-optical measuring lines (MLs) [1, 2] stacked in accordance with a certain setup on the surface studied. Thus, reconstructing distributed PFs' parameters against characteristics of optical radiation passing through FOMN assigns this mathematical problem to tomography [4]. To restore distributed function of PFs by means of FOMN, MLs were stacked along 2-4 directions.

In case of full data, in other words, with sufficient quality projections on all 180 degrees of angular range, high quality reconstructions are known to be obtained [4-6]. For comparison, to receive quality images in industry tomography, the necessary number of directions is  $p = 102-103$ .

Standard analytical methods are unacceptable for a fiber-optical tomography as direct application of inverse operator does not provide a unique stable solution. It is characteristic of any low angle problem in tomography. So, there is a good reason to consider other algorithms, perhaps, their synthesis as well. Restoring PFs' functions by using FOMN can be broken into several steps: sampling, receiving and processing projection data, and back projecting.

The existing success and great prospects for the development of information and measuring systems are largely due to the fact that sensor networks built on the same principles can be used in completely different areas of human activity. However, wireless sensors have a number of limitations that have a negative impact on the provision of information security in the transmission of data within and outside the network.

## 2 Notifications and Standard Definitions

Let's  $f(x_1, x_2)$  is the function of distributed PFs' parameter on a planar surface  $\Omega$ . Throughout this paper we will assume that  $f$  is infinitely differentiable and has a compact support. The 2D Radon transform  $\mathfrak{R}$  maps a density function  $f$  as its line integrals. The objective of tomography is to produce an accurate image of an object interior based on a finite number of scanned views.

Mathematically, the problem is to reconstruct  $f$ , given the measurements of  $g(\varphi, s)$  on  $\Omega$ . The short-term objective will be focused on a comprehensive description of a projection function  $g = \mathfrak{R}f$ .

Let index  $i$  determine an  $i$ -th direction of scanning  $\varphi_i$ , and index  $j$  determine the samples  $s_{ij}$  in the selected  $i$ -th direction. In this case a pair of indexes  $(i, j)$  corresponds to a straight line  $L_{ij}$  along which the area is scanned. Then a projection value along the straight line  $L_{ij}$  can be written as

$$g_{ij} = \mathfrak{R}_{ij} f = \int_{L_{ij}} f(x_1, x_2) dl, \quad (1)$$

where  $\mathfrak{R}$  is the Radon transform of function  $f$ , and  $dl$  is a gain along the straight line  $L_{ij}$ . Pairs of numbers  $(\varphi_i, s_{ij})$  determine the parallel setup of scanning on a plane.

Let's break the area of research  $S \subset R^2$  into smaller sites, so as

$$S = \bigcup_{k=1}^N S_k.$$

We consider the function  $f$  constant in each cell  $S_k$  and equal to  $f_k$ , the symbol  $f$  also denotes the matrix corresponding to this partition:

$$f = \begin{pmatrix} f_1 & f_2 & \dots & f_m \\ f_{m+1} & f_{m+2} & \dots & f_{2m} \\ \dots & \dots & \dots & \dots \\ f_{n(m-1)+1} & f_{n(m-1)+2} & \dots & f_{nm} \end{pmatrix} \Leftrightarrow F = (f_1 \ f_2 \ \dots \ f_m \ f_{m+1} \ \dots \ f_{2m} \ \dots \ f_{nm})^T. \quad (2)$$

Elementary cells  $S_k$  are referred to as image elements. Let's assume function  $f$  as constant in each site  $S_k$  and equal to  $f_k$ . Symbol  $f$  also denotes a matrix corresponding to this decomposition. Then the integrated equations (1) are transformed into a system of linear algebraic equations whose matrix forms looks as following:

$$AF = G \quad (3)$$

## 3 Optimization of FOMN Geometry

The specificity of the fiber-optic tomography tasks is the presence of an FOMN ultra-small survey data acquisition scheme. As a rule, in such FOMN, the number of measuring lines is less than the number of monitored areas. SLAE is underdetermined here (3).

Due to the fact that the input data has a large dimension, it is necessary to perform processing that allows you to select the most significant parameters by reducing the number of free variables in SLAE (3).

In this context, FOMN optimization consists in deleting rows and columns along the edges of the matrix  $f$  (in the "trimming" of matrix  $f$ ), the sum of the elements of which has values equal to zero. Knowing the size of the matrix  $f$  and the values of the column of the projection data, one can always check whether the matrix  $f$  has such a row or column. Then the rows and columns with this feature are removed from the matrix  $f$ . Next, the matrix itself (2) and the column of projection data are modified.

As a result of the Trimming Algorithm, a new matrix  $f$  of size  $n' \times m'$  is formed, with  $n' \leq n$ ,  $m' \leq m$ . Thus, when executing the algorithm described above, candidate areas in which the required “objects” are located are selected.

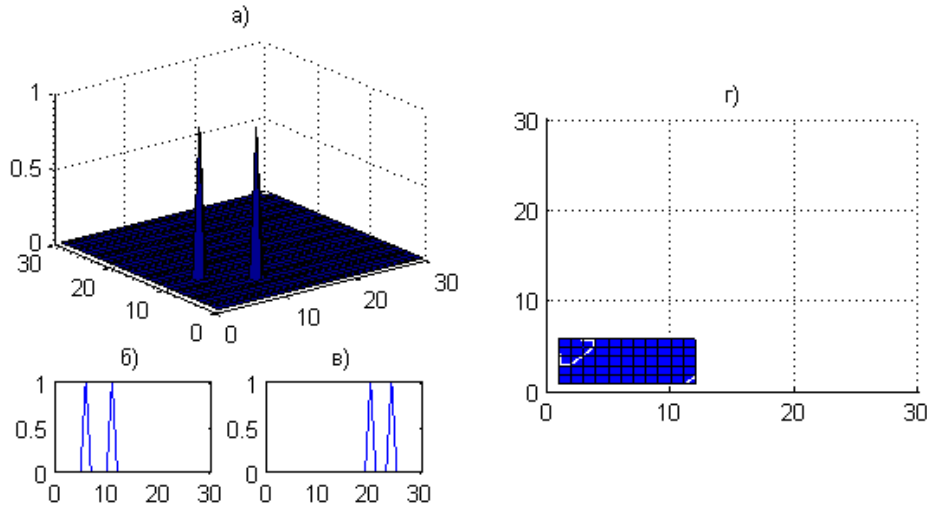


Figure 1: a) investigated function  $z(x, y) = e^{-0.5((x-6)^2 + (y-11)^2)} + e^{-0.5((x-16)^2 + (y-7)^2)}$ , b) and c) projection graphs, d) the subdomain obtained as a result of applying the “trimming” procedure.

Let us give an example of applying the above method to a specific distribution of a parameter of a physical field, which in analytical form is given by a function:

$$z(x, y) = e^{-0.5((x-6)^2 + (y-11)^2)} + e^{-0.5((x-16)^2 + (y-7)^2)}$$

Figure 1 shows projection data for two mutually perpendicular scanning directions. From the graphs it can be seen that the first and last values of the projection data in both cases have quite a lot of zero values and, therefore, it is advisable to do a preprocessing procedure in the form of trimming the area at the edges.

As a result of applying the FOMN Trimming Algorithm, the measured measurement network with dimensions  $n \times m = 30 \times 30$  is transformed into a network with dimensions  $n' \times m' = 7 \times 11$  (Fig. 1).

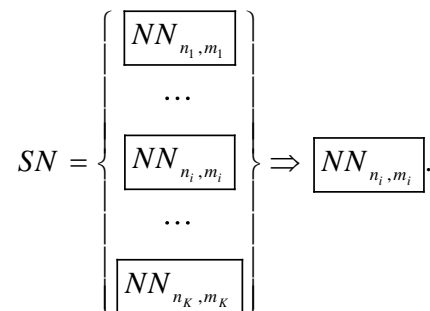
After the Trimming Algorithm described above for an area (matrix  $f$ , respectively), you can apply both regular recovery procedures, such as FBP [4], ART [7], and special algorithms developed by the authors UQC [8-9], as well as neural network algorithms to restore the functions under study [8, 10].

After the recovery of functions for the trimmed  $n' \times m'$  region is completed, a procedure for restoring the original  $n \times m$  dimensions are performed using a list of the surface layer containing information on deleted rows and columns of the matrix  $f$ .

#### 4 Complex of Neural Networks

High-performance computing technology focuses on developing parallel processing algorithms and systems by incorporating both administration and parallel computational techniques.

The next item of projection processing is the neural network processing of projection data obtained as a result of optimization of FOMN geometry:



Let FOMN, after applying the procedure described in clause 2, have dimensions  $n' \times m'$ ,  $n' \leq n$ ,  $m' \leq m$ . In general, the sizes  $n$  and  $m$  should decrease ( $n' \leq n$ ,  $m' \leq m$ ). This happens in most cases, because the spatial frequency  $b = \pi$  imposes restrictions on the size of the objects under study. In extreme cases, you will have to use the neural network for the entire area.

It is not known in advance what sizes will be there. Therefore, the question naturally arises: what exactly is the size of the neural network to use?

The answer to the question posed is contained in the approach proposed in this paper. It consists in the following:

1. We will train in parallel (independently of each other) several neural networks of different sizes.

Denote by  $NN(n_i, m_i)$  a neural network of size  $n_i \times m_i$ , i.e. neural network, which is intended for FOMN processing of the appropriate size.

Through  $SN$  denote we set of all  $K$  pre-trained neural networks of the form  $NN(n_i, m_i)$ :

$$SN = \{NN(n_1, m_1), \dots, NN(n_i, m_i), \dots, NN(n_K, m_K)\}, \quad (4)$$

$$n_1 \leq \dots \leq n_i \leq \dots \leq n_K, m_1 \leq \dots \leq m_i \leq \dots \leq m_K, \\ (n_i, m_i) \neq (n_j, m_j); i \neq j, 1 \leq i, j \leq K. \quad (5)$$

2. For processing the projection data of the  $n' \times m'$ ,  $n' \leq n$ ,  $m' \leq m$  posting from FOMN, we choose in the  $SN$  set of the form (4) a neural network of a suitable size, i.e.  $NN(n_i, m_i)$ , for which

$$n_{i-1} \leq n' \leq n_i, m_{i-1} \leq m' \leq m_i, 1 \leq i, j \leq K. \quad (6)$$

From conditions (5), (6), it obviously follows that the neural network  $NN(n_i, m_i)$  is a network of the smallest dimension, with which it is possible to process the projection data of a measuring network of size  $n' \times m'$ .

## 5 Using RBF Networks

In the work, neural networks of radial basis type are used in the work of neural networks. Earlier in the article [10], the authors have already investigated the possibility of using radial basis neural networks (RBFNN).

The information generated by the network, represented by vector  $G$ , was a set of topographic data for which the neural network must reconstruct the vector  $F$ . Thus, the neural network must perform the transformation  $F = A^{-1}(G)$ , having previously been trained on a set of training pairs  $\{(G, F)\}$ .

To create a training page, the author used Reinforcement method of selecting training pairs in which pairs of the form  $(G_i, F_i)$  were considered, where  $AF_i = G_i$ .

When creating RBFNN training pairs in [10], Gaussian-type functions were used, and the parameters were selected as lattice points of the corresponding scanning scheme and Gaussian pairs.

For example, for the  $5 \times 5$  field, a training page was created, consisting of 3325 training pairs, on which the RBFNN network was trained. It was experimentally shown that the constructed network makes it possible to restore the functions of the spatial distribution of the studied physical quantity with an error at a single point of no more than 2%, and has good predictive capabilities. However, it was noted that with this method of recovery in high-dimensional tasks using FOMN, there are serious difficulties in training the network due to the very large amount of training pages. Therefore, it became necessary to search for optimal paths when using neural networks.

One of the ways to optimize the processing of information is the use of a set of previously-trained neural networks of various dimensions.

The choice of reference functions should depend on the width of the spectrum  $b$  of the function  $f(x, y)$  under study. Functions of a Gaussian type

$$z(x, y) = e^{-a_i((x-c_i)^2 + (y-b_i)^2)} \quad (7)$$

can be used here, since they take non-zero values only in the zone around a certain center.

To analyze the neural network method for solving the problem using RBFNN, this work considered the tomographic task of restoring the functions of the FP according to the information coming from the information-measuring system of size  $30 \times 30$ .

It was assumed that the reference effect on the field has the form of a smooth function with a limited effective width of the spectrum  $b$  equal to the conditional spectral unit  $p$ . It is considered that all values of the function are non-negative and normalized.

In this work, the same three types of reference distributions of a physical quantity are used as in [10]. The first and second types refer to the regular method, and the third refers to the random method. We describe them in more details.

Type I. The reference field distributions in this case are single Gaussians of the form (7), whose centers are located at the nodes of the measuring network. It was found that the optimal parameters for learning are  $a_i$  parameters, which take values.

Type II. Analytically, these functions can be represented as:

$$z(x, y) = e^{-a_1((x-c_1)^2 + (y-b_1)^2)} + e^{-a_2((x-c_2)^2 + (y-b_2)^2)} \quad (8)$$

provided that the carriers are at least  $2\pi/b$ . These are Gaussian couples with non-intersecting carriers.

Type III. Reference distributions of this type were obtained using a randomization process with normalization. Each integer random set  $a_1, a_2, b_1, c_1, b_2, c_2=1, N$  was assigned a function of the form (8). Each vector before inclusion in the training page was normalized.

## 6 Numerical Modeling

To analyze the neural network method of solving the problem using the RBFNN complex, the information-measuring network 30x30 was considered.

In Table 1 for each neural network of a radial-basic type belonging to the set of SN, the following characteristics are presented:

- dimensions  $(n_i, m_i)$  correspond to the geometry of the measuring network, which is processed by a neural network of radial basis type  $NN(n_i, m_i)$ ,
- the total amount of the training page (TP) includes types I - III,
- the average training time for the results of a series of several (from 10 to 15) computational experiments,
- values of the normalized mean square error (MSE) across the entire training page,
- the number of impacts recognized by the neural network  $NN(n_i, m_i)$ ,
- predicting capabilities, i.e. recognition by a network of types of effects that do not belong to the training page,
- the quality of training is averaged characteristic, associated, including with the presence of artifacts as a result of insufficient amount of the training page.

From the results in Table 1, it can be seen that as the size of the network grows, the quality of learning decreases.

Table 1: Characteristics of radial basis type neural networks belonging to the set of SN

	Size ( $n_i, m_i$ )	Volume TP	Studying time	Mse error throughout the TP	Impact number	Predictive properties	Quality of education
1.	3×3	1390	5 sec	$2 \times 10^{-27}$	3	+	high
2.	5×5	3325	2 min	$2,8 \times 10^{-16}$	4	+	high
3.	7×7	4850	3-4 min	$1,3 \times 10^{-27}$	4	+	high
4.	10×10	8500	22 min	$1,8 \times 10^{-27}$	3	+	medium
5.	15×15	9102	22 min.	$3,5 \times 10^{-27}$	3	±	medium
6.	20×20	9264	30 mmin.	$3,4 \times 10^{-29}$	3	±	medium
7.	30×30	9000	25 min.	$10^{-25}$	2	-	low

Table 2. The results of data processing by the RBFNN complex for the reference functions of the form (7)

	$a_i$	$n' \times m'$	Size in %	Time of processing	MSE error
1.	0,5	3×3	1 %	0,0554	$1,9 \times 10^{-4}$
2.	0,1	5×5	2,77 %	0,0823	0,0062
3.	0,05	7×7	5,44 %	0,0879	0,0102

Table 3. Results of data processing by the RBFNN complex for reference functions of the form (8)

Distance between centers Gaussian	$n' \times m'$	Size in %	MSE error
4	3×5	1,6 %	$1,9 \times 10^{-4}$
5	4×5	2,2 %	$1,87 \times 10^{-4}$
11	7×13	10%	0,0181
20	14×14	22 %	0,0017
34	25×25	69,5 %	0,0024

From the above results, it follows that processing with the help of Trimming Algorithm an area with subsequent collective processing by neural networks yields a large gain in accuracy. This is explained by the localization of the site of impact on the network and processing using a neural network, as a rule, of a lower dimension, which is trained more efficiently and quickly. At the same time, the error of the standard deviation for the elements from the training page drops from 15 to 20 times.

Figure 2 shows the results of processing the species exposure  $z(x, y) = e^{-0.5((x-6)^2 + (y-19)^2)}$ .

After localization of the impact, a  $3 \times 3$  size similar region was obtained, which is processed by a well-trained neural network  $NN(3.3)$ . Finally, the original dimensions of the area were restored (Fig. 3b). In fig. 3c shows the result of the restoration of the function under study using a neural network of maximum size  $NN(30.30)$ . The quality

of training NN (30,30) is low, which explains the appearance of artifacts even with the restoration of a single impact on the measuring network.

Tables 2 and 3 present the results of the work of the proposed method for the reference effects on the measuring network of single effects (7) and double effects (8), respectively.

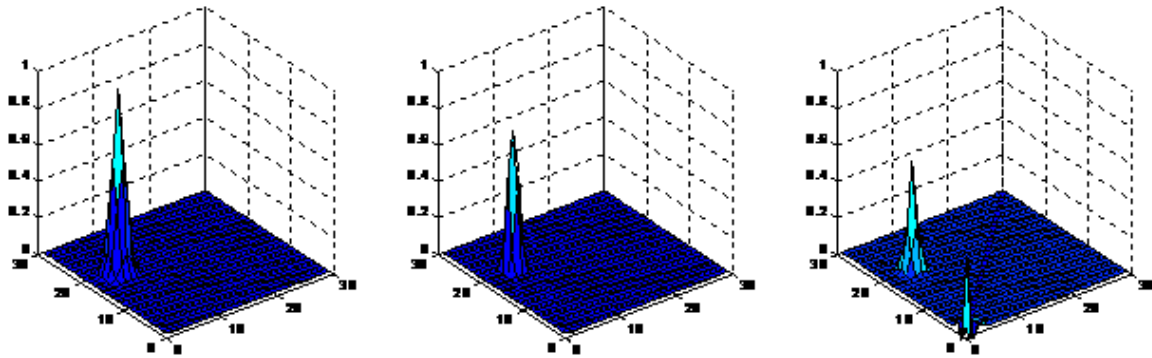


Figure 2: a) investigated function,  $z(x, y) = e^{-0.5((x-6)^2 + (y-19)^2)}$  b) the result of recovery using the complex of neural networks *SN*, c) the result of restoration using *NN (30, 30)*

## 7 Conclusion

Information driven economy relies on the actionable insights extracted from data analytics. The era of data revolution has paved way to the need of convergence of paradigms like High Performance Computing and Big Data processing. The amalgamation of these paradigms is a herculean task involving various aspects like data management and computing efficiency. This has given rise to evolution of the data storage technologies and computing models.

The article presents a new combined projection data processing algorithm for reconstructing information received from fiber-optic measurement lines distributed by FOMN.

This algorithm consists in the sequential execution of two processes:

1. Pre-processing of measurement information by localizing the locations of impact on FOMN,
2. Application of a complex of neural networks for processing measuring systems of various geometries.

From the above results it follows that:

1. Processing using the area cropping procedure with subsequent collective processing by neural networks gives a gain in accuracy largely due to the localization of the site of impact on the network and processing using the neural network, as a rule, of lower dimensionality, which is trained more qualitatively and quickly. At the same time, the value of the mse error for the elements from the training page drops 15 to 20 times.

2. Reducing the mse error and shortening the processing time mainly depends on how radically the computational process has been optimized as a result of the preprocessing and on the complexity of the function being restored.

## References

1. Kulchin, Yu.N.: Distributive Fiber Optical Measuring System. Fizmatlit, Moscow, 272 p. (2001)
2. Kulchin, Yu.N., Vitrik, O.B., Kirichenko, O.V., Petrov, Yu.S.: Multidimensional signal processing by using fiber optic distributed measuring network // Quantum Electronics, Vol.20, No.5, 711-714 (1995)
3. Zakasovskaya E.V., Tarasov V.S., Glushchenko A.A.: Information security issues in the distributed information measurement system // ICIEAM, S.-Petersburg, Russia. May 16-19 (2017)
4. Natterer, F.: Mathematics of Computerized Tomography, John Wiley & Sons Ltd., N. Y., 288 p. (1986)
5. Herman, G. T.: Projections-Based Image Reconstruction. In: "Basics of Reconstructive Tomography", Moscow, Mir, 352 p. (1983) (in Russian).
6. Mel'nikov, V.I., Meshkov, S.V.: Theory of activated rate processes: Exact solution of the Kramers problem, J. Chem. Phys. 85:1018–1027 (1986)
7. Zakasovskaya E.V., Fadeev, V.V.: Restoration of point influences by the fiber-optical network in view of a priori information. SPIE Proc. APCOM, V. 6675 (2007)

8. Zakasovskaya, E.V., Tarasov, V.S.: Optical fiber imaging based tomography reconstruction from limited data // Computer Methods in Applied Mechanics and Engineering. Vol. 328, pp. 542-543 (2018)
9. Kulchin, Yu.N., Zakasovskaya, E.V.: Artifacts suppression in limited data problem for parallel fiber optical measuring systems // Optical Memory & Neural Networks. – Vol. 18, № 3. – pp. 171-180 (2009)
10. Kulchin, Yu.N., Zakasovskaya, E.V.: Application of Radial Basis Function Neural Network for Information Processing in Fiber Optical Distributed Measuring Systems, Optical Memory & Neural Networks (Information Optics), Vol. 17, № 4, pp. 317-327. (2008)