Surveillance of Super-Extended Objects: Bimodal Approach

Andrey V. Timofeev, Dmitry Egorov

Abstract**—**This paper describes an effective solution to the task of a remote monitoring of super-extended objects (oil and gas pipeline, railways, national frontier). The suggested solution is based on the principle of simultaneously monitoring of seismoacoustic and optical/infrared physical fields. The principle of simultaneous monitoring of those fields is not new but in contrast to the known solutions the suggested approach allows to control super-extended objects with very limited operational costs. So-called C-OTDR (Coherent Optical Time Domain Reflectometer) systems are used to monitor the seismoacoustic field. Far-CCTV systems are used to monitor the optical/infrared field. A simultaneous data processing provided by both systems allows effectively detecting and classifying target activities, which appear in the monitored objects vicinity. The results of practical usage had shown high effectiveness of the suggested approach.

Keywords**—**Bimodal processing, C-OTDR monitoring system, LPboost, SVM.

I. Introduction

THE problem of complex monitoring of the super-extended objects has always represented a practical value. For example, oil and gas pipelines, railways, national frontier are examples of typical super-extended objects. Complex monitoring provides solutions for the following tasks: 1) telemetric status check of technological equipment on the monitoring objects; 2) unauthorized activities detection (tie-in to a pipeline, excavation in the monitoring object vicinity, pedestrian activity on railways, etc.); 3) timely detection of technogenic or natural disasters (oil leaks, train derails, damage of railway tracks), which appear in the monitoring objects vicinity. Solution of these problems is based on detection of certain precursors that signal the emergence of targeted events or processes. We will call those precursors a "target events" (TE). Examples of TE: a seismoacoustic vibration accompanying the oil spill from the pipe, a seismoacoustic noise and other symptoms associated with unauthorized attempts of tie-in to pipeline or excavation near the railways, seismoacoustic signals associated with pedestrian activity in vicinity of railways). Existing multimodal solutions for monitoring of extended objects use networks of seismic sensors for control of the seismic field. This approach is becoming expensive if the object's perimeter exceeds 10 km with a system resolution of 10 m. A cost of such system will be 4-5 times more expensive then cost of the C-OTDR-system. This is due to a need to provide electrical power and radiocommunication for each sensor of the network. The system described in this report is intended to provide a complex monitoring based on concurrent observations of seismoacoustic and optical/IR fields. In this case the TE's are detected and classified by both C-OTDR system (observation of seismic field) and long-range surveillance system (Far CCTV: observation of the optical/IR field). The combined data analysis from these two systems will significantly improve the monitoring reliability.

A. V. Timofeev and D. Egorov are with the LLP "EqualiZoom", Astana, Kazakhstan (phone: +7-911-191-42-67, e-mail: timofeev.andrey@gmail.com).

II. Basic Idea of the Bimodal Monitoring System

In the last years, C-OTDR monitoring systems are recognized as a most effective method for monitoring of the super-extended objects. A principle of operation of these systems is based on the infrared stream vibrosensitivity, which was pumped inside of a fiber-optic cable by means of a semiconductor laser. The fiber-optic cable has to be buried in the monitoring object vicinity into depth ~ 50-100 sm. In fact, this fiber is a supersensitive sensor ("distributed microphone") measuring the seismic-acoustic field fluctuations. We will call this fiber-optic cable the fiber-optic sensor (FOS). The semiconductor laser generates an infrared energy with long wave 1550nm in form of impulses with 10 ns duration at a repetition frequency of 2000 kHz. The target information about seismic-acoustic events which were appeared in the FOS is contained inside the backscattered infrared energy stream which has been reflected from the FOS microimpurities. The local refraction coefficient of the FOS dramatically changes under impact of the seismoacoustic vibrations, which were generated by the TE. This change cardinally influences the structure of the chaotic interference backscattered radiation. This structure is called a speckle-structure and it corresponds to a particular FOS part, which is approximately 10-15m long (C-OTDR systems resolution). Further, these FOS parts will be called "C-OTDR-channels". Thus a speckle-structure change means that the seismic-acoustic emission source has appeared in a corresponding C-OTDR-channel (near the corresponding FOS part). Simply put, time-frequency structure of the speckle will match with the time-frequency structure of the signal from the seismic acoustic emission source. The analysis of the speckle structure changes identifies types of the detected seismoacoustic events (Fig. 1) and leads to a decision whether the detected event is a noise or a TE? The place of TE emergence is determined with accuracy of a virtual ellipsoid (10x15-50 m). The ellipsoid size depends on the TE type (Fig. 2). This virtual ellipsoid we will call a target virtual ellipsoid (TVE). One laser is able to serve the FOS with length ~50 km. Here FOS is a conventional monomode fiber-optic cable of type SMF-28 (or ITU-TG.652, ITU-TG.654, ITU-TG.655). FOS can be one or several fiber-optic cores of a multicore fiberoptic cable, which is already deployed near monitoring object and used for an ordinary data transfer. In this case there is an obvious saving of resources that would have been spent for the dedicated FOS installation. However, C-OTDR systems lose their effectiveness in places where soil has high seismic-acoustic impedance (sand, gravel) or powerful natural noise sources (underwater rivers, highways). These are high-risk places for both C-OTDR and energetically autonomous Far CCTV (FCCTV) systems used together. The FCCTV systems obtain high quality pictures of the optic and infrared bands in distances up to 4-4.5 km. The C-OTDR system is the primary target designation source for the FCCTV system. C-OTDR provides coordinates of the TVE and sizes of its axes and FCCTV focuses long-range surveillance cameras on the TVE (Fig. 3).

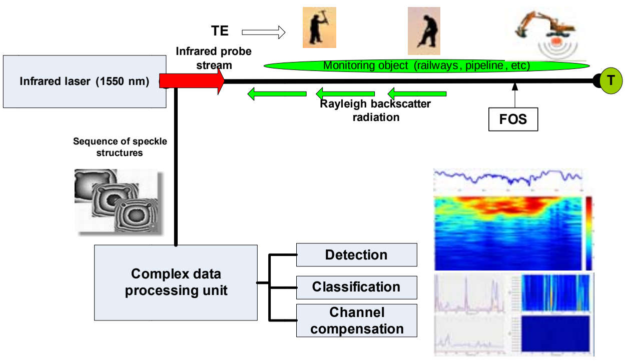

Fig. 1 The general scheme of the C-OTDR monitoring system

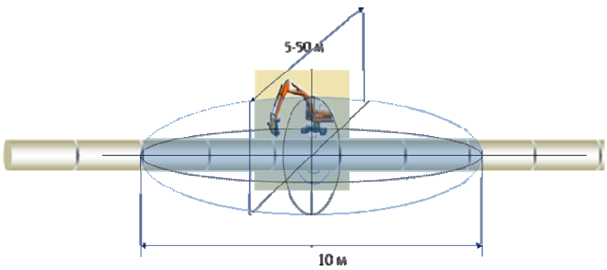

Fig. 2 The target virtual ellipsoid

Further follows a fully automatic procedure of express analysis of obtained images and extraction of relevant image features for classifications. For the next step the image and C-OTDR features are processed together in order to classify a detected TE within a bimodal classification system framework. The parameters of the bimodal monitoring system must be configured in a way that dramatically increases the TE detection classification accuracy to achieve low values of errors I and II types. Concluding, the bimodal-processing ensures a significant error reduction of the TE detection and classification.

III. The Principles of the TE Detection and Classification

The length of extended monitoring objects (oil and gas pipelines, railways, sections of a state borders) is in the level of hundreds or even thousands of kilometers. Some of these objects are located in uninhabited areas with poorly developed transport and telecommunications infrastructure. In this regard, every call to the emergency services will incur additional costs including fuel costs, staff time and amortization. Therefore the operational response cost to each case of a false alarm would be very expensive. In order to reduce operational cost the monitoring system development was based on a well-known Neyman-Pearson principle: minimizing value of the type I errors with guaranteed upper bound of type II errors. The system has designed to work in two basic modes: “mode A” – monitoring object only by using C-OTDR-system, “mode B” – monitoring object by using both FCCTV and C-OTDR systems. In “mode A” the TE detection/classification tasks are solved by using measurements of a seismoacoustic field in the FOS vicinity. The C-OTDR system provides high TE detection reliability but moderate classification accuracy. From one side the C-OTDR system provides cost-effective monitoring solution of super-extended objects. On another hand, FCCTV’s are effective for local monitoring using a 3d-Scenes Analysis. Accordingly the FCCTV needs a reliable target designation, which contains information about TE area coordinates. In “mode B” the FCCTV gets from the C-OTDR area coordinates where a TE had been detected. In this case, the bimodal-processing provides an efficient solution for both TE detection and classification tasks.

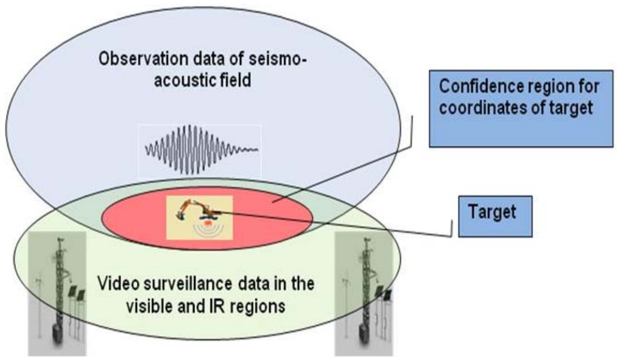

Fig. 3 Using bimodal approach to get TVE

IV. TE Classification Features in a Bimodal System

In practice, TE leaves traces in both seismoacoustic and optical (and IR) fields of the earth. FCCTV system design took into account fact that shape and color features are more efficient in contrast with dynamics features [1]. The shape/color characteristics: Scale Invariant Feature Transform (SIFT) [2], Color SIFT [3], [5], Histogram of Oriented Gradient (HOG) [13], Shape Context (SC) [11], PCA-SIFT [12], RGB-SIFT and HSV-SIFT [4]. The dynamics features: Space-Time Interest Points (STIP) [6], Dense Trajectories [7].

It must be also taken into account that the dynamic features lose their efficiency on super-long distances (more 500 m) because angular velocity of TE strives to zero with increasing distance from TE to the FCCTV system. During C-OTDR development process we found that tandem LFCC (Linear-Frequency Spaced Filterbank Cepstrum Coefficients, [8])-GMM (Gaussian mixture model, [9, 10]) is the most effective feature for the TE classification. Here LFCC's are defined for speckle-structures of particular C-OTDR channels. Thus LFCC-GMM-vectors with dimension 1024 were used as C-OTDR features.

As a result, in the bimodal monitoring system the following TE classification features (hereinafter j - index feature) were used:

- φ_{SIFT} - SIFT (dimension - 128, codebook size 250); j=1;
- φ_{HOG} HOG (dimension - 3780, codebook size 2000); j=2.
- $\varphi_{LFCC-GMM}$ LFCC-GMM (GMM 1024-vectors), j= 3.

V. The Algorithm of TE Classification and a Method of System Training

To provide functionality of the FCCTV / C-OTDR system in “mode B” the C-OTDR and FCCTV subsystems were trained together as three independent classifiers on the same labeled data. In this case φ_{SIFT} , φ_{HOG} and φ_{HOG} features were used. To provide functionality FCCTV / C-OTDR system in the “mode A” only of C-OTDR subsystem data was used and only one classificator which uses the $\varphi_{LFCC-GMM}$ feature. The training sample includes real patterns of the following classes: 1) the pedestrian; 2) a group of pedestrians; 3) car; 4) truck; 5) hand earthworks; 6) earthworks produced by using excavation equipment. Thus number of target classes equals six; indexes of those classes form a set **I**. Each class is represented with N samples video and C-OTDR data. The video and C-OTDR data were obtained in two geographically separated points with a sampling size N=10. At the first point was the clay soil and the video was obtained in low visibility conditions (day, light wind with dust). At the second point the soil was sandy and the video was obtained in good visibility conditions (day, no wind, a thin mist). Let us denote x^I – image from FCCTV subsystem; x^S – seismic field observations received from C-OTDR subsystem; $\left\{\left(x^I_k, y_k\right) \mid k=1,...,N\right\}$ - training sample for FCCTV subsystem (φ_{SIFT} and φ_{HOG}); $\left\{\left(x^S_k, y_k\right) \mid k=1,...,N\right\}$ - training sample for C-OTDR subsystem ($\varphi_{LFCC-GMM}$). Here and further, $y_k \in \mathbf{I}$.

Following the conclusions of [14] as algorithm of TE classification in the "mode B" was used by so called multiclass ν-LPBoost [15], built as a linear convex hull of Lipschitz classifiers. This method steadily works even at a small training sample size [14]. As the Lipschitz classifiers have used conventional SVM (Support Vector Machine) [16].

To solve the multiclass TE classification problem those SVM-classifiers were trained by well-known scheme **one-against-all** (for according feature spaces). In FCCTV subsystem a SIFT and HOG feature spaces were used. Herein SVM classifier $f_{1,i}(\cdot \mid \alpha_{1,i}, b_{1,i}) \equiv f_{1,i}(\cdot)$ corresponds to the SIFT-feature space and SVM classifier $f_{i,2}(\cdot \mid \alpha_{i,2}, b_{i,2}) \equiv f_{i,2}(\cdot)$ corresponds to the HOG-feature space. In C-OTDR subsystem was used SVM classifier $f_{i,3}(\cdot \mid \alpha_{i,3}, b_{i,3}) \equiv f_{i,3}(\cdot)$. Here *i*- index of target class, $i \in \mathbf{I}$; $\left(\alpha_{i,j}, b_{i,j}\right)$ - parameters of *j*-th SVM classifier. Those parameters are subject to setting up on a training stage. According to the concept **one-against-all** every class *i* is separated from the other classes by use the corresponding classifier $f_{i,j}(\cdot \mid \alpha_{i,j}, b_{i,j})$ in the relevant feature space. All these SVM-classifiers $f_{i,j}(\cdot \mid \alpha_{i,j}, b_{i,j})$ are built based on the product Bhattacharya kernels [17]. Optimization of the classifiers parameters $\left(\alpha_{i,j}, b_{i,j}\right)$ was made by use of the usual cross-validation (CV) scheme [18]. A bimodal discriminant function of ν-LPBoost-classifier [15], has following simple form:

$$F(x^I, x^S) = \arg\max_{i \in \mathbf{I}} \left(\beta_1 f_{i,1}(x^I) + \beta_2 f_{i,2}(x^I) + \beta_3 f_{i,3}(x^S)\right).$$

The training phase comes down to an optimal choice of parameters $\left\{\beta_j\right\}$. This choice is performed by using standard optimization method (linear programming) according to the following scheme:

$$\min_{\beta, \xi, \rho} \left(-\rho + \frac{1}{\nu N}\left(\sum_{k=1}^{N} \xi_k\right)\right),$$

under the condition:

$$\begin{aligned}
& y_k\left(\beta_1 f_{y_k,1}(x^I_k) + \beta_2 f_{y_k,2}(x^I_k) + \beta_3 f_{y_k,3}(x^S_k)\right) - \\
& \arg\max_{y_p \neq y_k}\left(\beta_1 f_{y_p,1}(x^I_k) + \beta_2 f_{y_p,2}(x^I_k) + \beta_3 f_{y_p,3}(x^S_k)\right) + \xi_k \geq \rho, \\
& k = 1,...N, \ \sum_{j=1}^{3} \beta_j = 1, \beta_j \geq 0, \ j = 1,..3.
\end{aligned}$$

Here ξ - slack variables, ν - regularization constant, which is chosen using CV. In “mode A” is used only C-OTDR data. Because of in this case the classifier has the following form: $F(x^S) = \arg\max_{i \in \mathbf{I}}\left(f_{i,3}(x^S)\right)$, where $f_{i,3}(x^s)$ is usual SVM-classifier in the space of feature $\varphi_{LFCC-GMM}$.

VI. FCCTV / C-OTDR System Specifications and Peculiarities of Installation

The bimodal FCCTV / C-OTDR monitoring system is designed for installation in deserted places with potential electrical power supply and communications problems. Therefore, the system must be capable of fully autonomous operation including power supply and connectivity with the monitoring control center. If FCCTV / C-OTDR monitoring system installation location is provided with electricity and has mobile communications then system is connected to the centralized sources of electricity and for communication uses either fiber optic channel which is laid along the monitoring object or a mobile internet 3G/4G. In this case the solution is cheaper. In case when it is necessary to equip the FCCTV / C-OTDR system with an autonomous power supply (diesel-generator, battery (up to 1600 A), solar panels (600 watts) or wind power (1000 W)) - the solution cost increases significantly.

The system works in the intelligent energy saving mode. When it is impossible to use a fiber optic connection to a Control Center, the system is equipped with a radio-relay systems (range of 250-800 MHz, range up to 80 km, the data rate 48 Mbit / s). The quality of the video stream: 25 fps at a resolution of 704x576 pixels. An optic-block and C-OTDR processing units are arranged in special housings, which are situated at the FCCTV towers installation points. For electric supply of the C-OTDR and FCCTV is used one and the same power supply, Fig. 4.

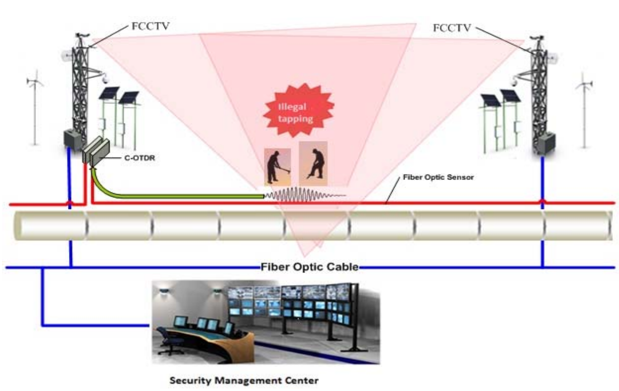

Fig. 4 Structural diagram of FCCTV / C-OTDR

VII. Performance Characteristics of the Bimodal FCCTV / C-OTDR System

Table I summarizes the results of natural experiments which to some extent characterize the quality of the operation of an algorithmic system core. In the experiment, the control unit of C-OTDR subsystem was located at a distance of 20 km from the place where TE's were implemented, i.e., seismic-acoustic environment was monitored at a distance of 20 km. FCCTV sensors were located 2000 and 3000 m from the TE implementation. Symbol α (2) denotes a value of I Type error (false reject), and symbol β (2) - a value of II Type error (false alarm) for "mode B". Accordingly, symbols α (1) and β (1) denote values of I and II errors types for "Mode A". The values α (1), β (1), α (2), β (2) were obtained experimentally for different types of TE. The data which presented in Table I demonstrates an acceptable accuracy of TE classification when using the bimodal FCCTV / C-OTDR system. Expectedly, in the mode "B", the accuracy of the system is significantly higher compared with the mode "A". It should be noted that the weak-energy-small-size TE, which occurring at great distances from the FCCTV-sensors are classified less effectively. For example: TE of type a "pedestrian"

TABLE I
Performance Characteristics

Type of TE	Distance from FCCTV -sensors (m)	$\alpha(2)$	$\beta(2)$	$\alpha(1)$	$\beta(1)$
«hand digging soil»	2000	0.02	0.04	0.1	0.12
	3000	0.03	0.04		
«group of pedestrians»	2000	0.15	0.01	0.13	0.11
	3000	0.21	0.02		
"pedestrian"	2000	0.07	0.04	0.16	0.1
	3000	0.1	0.05		
«car»	2000	0.06	0.03	0.09	0.1
	3000	0.1	0.04		
«truck»	2000	0.07	0.01	0.07	0.08
	3000	0.23	0.02		
«digging soil by a heavy excavator»	2000	0.02	0.01	0.06	0.09
	3000	0.11	0.01		

VIII. Conclusion

Comprehensive monitoring of super-extended objects is becoming increasingly important task that requires using of modern methods to obtain and process relevant information. High efficiency of solving the monitoring problem by using a bimodal approach had been proven during pilot operation of the FCCTV / C-OTDR system. A bimodal system uses two types of physical fields - seismoacoustic and optical/IR, which reflected the current state of one and the same object. To obtain seismic-acoustic data the vibrosensitive properties of an infrared flux were used (the infrared flux had been injected into the fiber optic cable using a semiconductor laser). Fiber optic cable is laid near the monitoring object in depth of 50-100 cm. Analysis of the backscattered IR flux allows identifying and locating threats with high accuracy. In general, control of the object state can only be based on the C-OTDR information. However, in those cases where the damage from possible threats is very significant, the effectiveness of using only C-OTDR-data may be insufficient and because of it an additional source of information is used. Such an additional source is the information obtained from the FCCTV-system. This system provides high-quality images at distance up to 4 - 4.5 km away from the camera in optic/IR ranges. Performed tests and results of the pilot operation confirmed a high efficiency of the bimodal monitoring system.

Acknowledgment

This study has been produced under the project "Development of a remote monitoring system to protect backbone communications infrastructure, oil and gas pipelines and other extended objects (project code name – OXY)", financed under the project "Technology Commercialization", supported by the World Bank and the Government of the Republic of Kazakhstan.

REFERENCES

[1] M. Elhoseiny, A. Bakry, and A. Elgammal, “MultiClass Object Classification in Video Surveillance Systems - Experimental Study”. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW '13). IEEE Computer Society, Washington, DC, USA, 2013, pp. 788-793.
[2] D. G. Lowe, “Distinctive image features from scale-invariant key-points”. IJCV, 60(2), 2014, pp.91–110.
[3] M. Tan, L. Wang, and I. W. Tsang, “Learning sparse svm for feature selection on very high dimensional datasets”, ICML, 2010, pp. 1047-1054.
[4] A. Bosch, A. Zisserman, and X. Muoz, “Scene classification using a hybrid generative/discriminative approach,” *IEEE Trans. Pattern Analysis and Machine Intell*, 30(04), 2008, pp.712-727.
[5] A. E. Abdel-Hakim and A. A. Farag“CSIFT: A SIFT Descriptor with Color Invariant Characteristics,” *Computer Vision and Image Processing Laboratory*. (CVPR'06), 2006, pp.1978-1983.
[6] I. Laptev, “On space-time interest points”, IJCV, 64(2-3), 2005, pp.107–123.
[7] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu, “Dense trajectories and motion boundary descriptors for action recognition’, IJCV,103(1), 2013, pp.60–79.
[8] P. Merlmestein and S. Davis, “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously. Spoken Sentences‖”, *IEEE Trans. On ASSP*, Aug, 1980. pp. 357-366.
[9] D. Titterington, A. Smith, U. Makov, *Statistical Analysis of Finite Mixture Distributions*. Wiley. ISBN 0-471-90763-4, 1985.
[10] M.A. Figueiredo, A.K. Jain, "Unsupervised Learning of Finite Mixture Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (3), 2002, pp.381–396.
[11] S. Belongie, J. Malik, and J. Puzicha. “Matching shapes”, *The 8th ICCV*, Vancouver, Canada, pp. 454-461, 2001.
[12] Y. Ke and R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors”, CVPR, Washington, DC, USA, 2004, pp. 66-75.
[13] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05), vol. 1, 2005, pp.886-893.
[14] V. Gehler, and Sebastian Nowozin, “On feature combination for multiclass object classification”, *Peter. ICCV, IEEE*, 2009, pp. 221-228.
[15] G. Ratsch, B. Scholkopf, A.J. Smola, S. Mika, T. Onoda, and K-R. Muller, “Robust ensemble learning”, In A.J. Smola, P. L. Bartlett, B. Scholkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, MIT Press, 1999, pp. 208–222.
[16] M.A. Hears, S.T. Dumais, E. Osman, J. Platt, and B. Scholkopf, “Support Vector Machines”, *IEEE Intelligent Systems*, vol. 13(4), 1998, pp.18-28.
[17] T. Jebara and R. Kondor, “Bhattacharyya and expected likelihood kernels”, In Proc.16th Annual Conference on Learning Theory (COLT 2003), 2003.
[18] M. Stone, “Asymptotics for and against cross-validation”, *Biometrika*, 1977, 64 (1), pp. 29–35.

Timofeev Andrey V. was born in Chita (Russia). He received Dr. habil. sc. ing.in Computer and Information Sciences from Tomsk State University of Control Systems and Radioelectronics, Russia, in 1994. A number of research publications in the International journals (JKSS, Stat.Methodology., Automation and Remote Control etc.) and International/National conferences are at his credit. He is on the editorial board of several journals and conferences and a referee of several others. His research interests include non-asymptotic nonlinear methods of confidence estimation of multidimensional parameters of stochastic systems; machine learning, large margin classification in Banach Spaces; confidence Lipschitz classifiers; technical diagnostics, C-OTDR systems; data mining; change-point problem; alpha-stable laws; statistical classification in application to biometrics and seismics.

Dmitry Egorov was born in Leningrad (Russia). His research interests include C-OTDR and security systems.