Proceedings of the International Conference on Machine Vision and Machine Learning
Prague, Czech Republic, August14-15, 2014
Paper No. 32

Multichannel Classification of Target Signals by Means of an SVM Ensemble in C-OTDR Systems for Remote Monitoring of Extended Objects

Andrey Timofeev, Dmitry Egorov
JSK "EqualiZoom"
18A Akzholov Str, Astana, 010000, Kazakhstan
timofeev.andrey@gmail.com, edv017@gmail.com

Abstract - The report proposes a new method for multichannel classification of C-OTDR signals. The method is based on the use of an ensemble of SVM-classifiers and it is robust to the types of environments of seismoacoustic signals propagation. The report presents the results of practical application of the proposed method in systems for remote monitoring of extended (distributed) objects.

***Keywords*:** SVM ensemble, Multichannel classification, C-OTDR monitoring systems.

1. Introduction

Application of the C-OTDR (Coherent Optical Time Domain Reflectometer) technology to solve complex problems of remote monitoring of extended objects is currently being evaluated as a very promising approach. In particular, this technology can be effectively used to monitor oil and gas pipelines, controlling technological processes and identifying unauthorized activities in close proximity of the monitored objects. Simplistically, an C-OTDR -system consists of an infrared laser, an optical fiber and a processing unit. The laser sends the probing signals through the optical fiber which is buried in the vicinity of the monitoring object. The processing unit is designed for comprehensive processing of the backscattered signals, which are called speckle-structures. The main item of the C-OTDR technology is a comprehensive analysis of the Rayleigh backscattered radiation characteristics, which transforms into an energetically weakened pulse and propagates constantly in the direction opposite to the direction of a pulsed laser flow. The reflected signal is created by the presence of static impurities in the optical fiber body and defects in the microstructure. Signals scattered by the centers coherently and randomly interfere with each other, forming so-called speckle patterns. Speckle patterns corresponding to different sections of the optical fibers are recorded and accumulated in the data center. The slightest change of the reflectance index value of the fiber, which occurred in a particular place, radically changes the speckle pattern corresponding exactly to this place of the fiber. These changes are reliably detected by the data center. The local changes in refractive index occur under the impact of temperature or due to mechanical action on the optical fiber surface. Let us call the optical fiber buried in the soil to a depth of 50-100 cm, an optical sensor (OS). Mechanical stress on the OS surface is caused by seismic acoustic waves. These waves are generated of the sources of elastic vibrations (SEV) which located in vicinity of laying optical fiber. Upon reaching the OS, seismoacoustic wave causes a local longitudinal microstrain on the surface of OS. Those microstrains in turn, cause a change in the local refractive index of light in a relatively small sector of the OS. As a result, the speckle pattern, which corresponds to this sector, changes significantly. Thus, the OS quite accurately reflects the state of the seismoacoustic field in its vicinity. The seismo-acoustic field contains information about events that occur in the surface layers of the ground near the OS. This field is created by structural waves, which generated due to mechanical effects on the soil or as a result of a seismic activity. Walking or running man, traffic, earthworks, including hand digging are typical sources of the acoustic emission (structural acoustic wave). In this case, the frequency range of the seismoacoustic waves is in the interval of 0.1 Hz to 1000 Hz. The information, which is required for correct identify the type of SEV, is concentrated in the frequency range of 0.2 Hz to 600 Hz, while 95% of the meaningful information is in even the more narrow range of 0.3 Hz - 350 Hz. The spectral characteristics of the target signals which lie above and below this frequency range carry information only about the individual characteristics of the SEV. The SEV, which are subjects of interest for remote C-OTDR monitoring will be called a target SEV. For the convenience of data processing, the entire OS length is broken to successive portions (sites) each has length around 10-15 m. The data from those sites is processed separately. These sites will be called C-OTDR channels or just channels. Width of the channel Δ_w depends on the probe pulse length. It should be noted that there is no physical possibility to significantly reduce the channel width Δ_w, thereby increasing the accuracy and spatial resolution while maintaining a significant range of the OS sensitivity. It is known that the optimum channel width depends on a number of physical factors jointly taken into account when designing a semiconductor laser, which is main C-OTDR component. In practice, target SEV has its own small size and assumed point. Due to the nature of the elastic oscillation, the wave from a point source of seismoacoustic emission is usually detected simultaneously in several C-OTDR channels. At the same time, due to strongly anisotropic medium of the elastic vibrations propagation, the structure of the oscillations (speckle patterns) varies considerably between different channels. In each channel a time-frequency characteristics of the speckle pattern are largely reflect a time-frequency structure of the SEV, which occur in vicinity of the corresponding channel. The oscillation energy is considerably attenuated during propagation in the environment. Intensity of attenuation depends on the average absorption factor of the medium and on the distance from the oscillation point to the location of channel. Accordingly there is always one channel, wherein vibrations of the SEV reflected substantially intense than in other channels. We will call this channel dominant. C-OTDR systems perform three major tasks in the following sequence: a)Task “D” (Detection) – detection of the SEV; b) Task “E” (Estimation) – estimate of the location of the SEV; c) Task “C” (Classification) – classification of detected SEV by means of assigning it to one of N priori given classes. This work describes a new approach to classification of the SEV. This approach has high classification reliability, is stable in terms of the location SEV relative to the OS and is easily implemented in practice.

2. Multichannel C-OTDR Monitoring System

C-OTDR monitoring systems of extended objects are designed to monitor seismoacoustic phenomena near controlled objects. Generally, C-OTDR-systems are composed of three subsystems: “D”, “E” and “C”. Those systems perform tasks D, E and C respectively. The “D” subsystem is based on models of guaranteed change-point estimation of statistical properties of the observed processes. The “E” subsystem is based on triangulation method of estimating the signal source coordinates. However, description of «D» and «E» subsystems is not the purpose of this report. We only mention that joint operation of these subsystems creates data for the subsystem "C". The subsystem “C” gets coordinates of detected SEV (vector $x_0 \in R^2$), and set $Ch_d = \left\{ j - s_l, j - s_l + 1, ... j, ..., j + s_u - 1, j + s_u \right\}$. This set contains sequential channel numbers, in which SEV was detected. Here j is a dominant channel number, $s_l, s_u \geq 0$. Other channels $Ch_d \setminus j$ are called adjacent to the dominant channel. Mathematical model of the SEV is composed as follows:

- **Stage 1. Feature Vectors Extraction**. For each speckle pattern obtained in the probing period T for each of the Z channel are built Linear-Frequency Spaced Filterbank Cepstrum Coefficients (LFCC). In our case these features are based on 19 linear filter-banks (from 200 to 3000 Hz) derived cepstra. Thus, 19 static and 19 first-order delta coefficients were used, giving the feature order m = 38.
- **Stage 2. Approximation of the probability distribution function of the feature vectors** by semi-parametric multivariate probability distribution model, so-called Gaussian Mixture Models (GMM). GMM is one of the principal methods of modeling broadband acoustic emission sources (including SEV) for their robust identification now. The GMM of SEV feature vectors distribution is a weighted

sum of J components densities and given by the equation $P\left(x|\lambda_s\right)=\mathbf{w}_s\mathbf{B}_s^T(x),$ where x is a random m-vector, $\mathbf{w}_s=\left(w_{s1},...w_{sJ}\right)\in R^J, \mathbf{B}_s(x)=\left(B_{s1}\left(x\right),...B_{sJ}\left(x\right)\right)\in R^J,$

$$\underset{s,i}{\forall} B_{si}\left(x\right)=\left(\left(2\pi\right)^{m/2}\left|\Sigma_{si}\right|^{1/2}\right)^{-1}\exp\left(-\frac{1}{2}\left(x-\mu_{si}\right)^T\Sigma_{si}^{-1}\left(x-\mu_{si}\right)\right),\lambda_s=\left\{\left(w_{si},\mu_{si},\Sigma_{si}\right)|\,i=1,J\right\}$$

In general, diagonal covariance matrices Σ_{si} are used to limit the model size. The model parameters λ_s characterize a SEV in the form of a probabilistic density function. During training, those parameters are determined by the well-known expectation maximization (EM) algorithm (Blimes et al.,1998). In this experiment approximation value J was equal to 1024.

We denote:

- SEV is the object of classification.
- λ is the model parameters of SEV.
- $\sigma\left(\lambda\right)$ is index of class, which belongs SEV with parameters λ . Thus $p^0=\sigma\left(\lambda\right)\in\mathbf{P}=\left\{1,...,D\right\}$, p^0 - index of the class, which belongs SEV.
- Given: cortege of coordinates of all channels in the system $X_{\mathbf{Z}}=\left(x_1,x_2,...,x_Z\right)$, where $\mathbf{Z}$ is the set of indexes of the channels, $\mathbf{Z}=\left\{1,...,Z\right\}\subseteq N$;
- T – channel probing period (speckle-structures renewal period in system channels);
- $\left(\mathrm{M},\|\cdot\|_{\mathrm{M}}\right)$ is feature space, M is the set of feature values, $\|\cdot\|_{\mathrm{M}}$ - space norm, $Diam(\mathrm{M})=\sup\limits_{\lambda_1,\lambda_2\in\mathrm{M}}\|\lambda_1-\lambda_2\|_{\mathrm{M}}$
- $\left|X\right|$ - cardinal value of the set X;
- $\left\langle\mathbf{x}\right\rangle_i$ - i-th component of the vector $\mathbf{x}\in R^m$ $,1\le i\le m$.

Thus, for identification of class SEV, each SEV is modeled by a GMM and is referred to as his model parameters $\lambda\in\mathrm{M}$. Priori defined target classes of SEV, which collectively makes up a finite set $\mathbf{P}$, $\left|\mathbf{P}\right|=D$. For example, in case of oil and gas pipeline monitoring the array $\mathbf{P}$ consists of the following classes: "hand digging the soil", "chiseling ground scrap ", "walking man", "running man", "passenger car", "truck", "heavy equipment excavator", "easy excavation equipment", "spilling fluid pressure", "liquid spills without pressure", "movement of the cleaning scraper", "normal process noise from operating the pipeline," "abnormal process noise from operating the pipeline, "cutting frozen soil", "large ungulate", "small ungulate", " large ungulate", "shrew digging the ground”.

3. Classification Solution in the Multi-channel C-OTDR Monitoring System

Classification solution in this case is considerably complicated because target SEV may appear at different distances from the OS. Therefore, time-frequency and energy characteristics of elastic vibrations will reach the OS being significantly weakened by the passage through absorbing, anisotropic medium (soil). Thus, if classification sample and a model correspond to different distances to the OS, then their time-frequency characteristics can vary drastically. This may cause quality degradation of the classification system. To mitigate this problem, a special approach was used. In this approach, the system was trained on data set, which included collection of measurements at varying distances from OS and with different values of the absorption coefficient. Accordingly, following concepts were introduced:

- Area of position (AP) of the SEV relative to the OS is a square with sides equal to the channel width Δ_w 10 m$\le\Delta_w\le$20 m . The center of the square (AP) is located at $r\cdot\Delta_w-0.5\cdot\Delta_w$ from the OS, r=1,2,3... Here r - number of AP for a particular channel. Thus, for particular channel the first AP is located at distance $0.5\cdot\Delta_w$, the second at distance $1.5\cdot\Delta_w$ etc. Let us denote the square located at distance $\Delta_w \cdot (r-0.5)$ from the OS by symbol $\Omega(r \mid \Delta_w)$. Maximum distance between SEV and OS, in which SEV can still be registered by the C-OTDR system is comparatively small and in practice does not exceed the value $u\Delta_w$, where $u \le 10$. Thus, all of the possible AP belong to the following set: $\mathbf{\Omega} = \left\{ \Omega(r \mid \Delta_w) \middle| r = 1,...u \right\}$.

- Generalized slots for values of the absorption coefficient of the acoustic waves. For practical reasons, two generalized slots have been allocated for the absorption coefficient. The first one $\alpha(1) = [10^{-5}, 10^{-3}]$ corresponds to cemented strains; the second $\alpha(2) = [10^{-3}, 0.5]$ corresponds to loose strains. Naturally, the number of absorption coefficient gradations can be increased to increase the model adequacy. We denote: v as number of selected intervals and set $\mathrm{A} = \left\{ \alpha(i) \middle| i = 1,..v \right\}$.

Denote $\mathbf{\Lambda} = \mathbf{\Omega} \otimes \mathrm{A}$ - as Cartesian product of sets $\mathbf{\Omega}$ and A. Each element $\kappa = (\Omega(r \mid \Delta_w), \alpha) \in \mathbf{\Lambda}$ is a reference to the external conditions of learning (distance and the absorption coefficient). During the training reference SEV(p), belonging to the class $p \in \mathbf{P} = \{1,...,D\}$, is placed in the center of area $\Omega(r \mid \Delta_w) \in \mathbf{\Omega}$. The absorption coefficient $\Omega(r \mid \Delta_w)$ belongs to the interval $\alpha \in \mathrm{A}$. This pair corresponding to element $\kappa = (\Omega(r \mid \Delta_w), \alpha) \in \mathbf{\Lambda}$. In the future, element $\kappa \in \mathbf{\Lambda}$ will be called a parameter of external conditions. In the k-м experiment with the reference SEV(p), the elastic vibrations of the reference SEV (p) are fixed in the group of system channels whose indexes are collected in the set $Ch(k \mid p,\kappa) = \left\{ j_k - s_k^l, j - s_k^l + 1, ..., j_k, ..., j_k + s_k^u - 1, j + s_k^u \right\},\ s_k^l, s_k^u \ge 0$. Here j_k - is an index of the dominant channel in the k-th experiment.

The number of experiments with reference SEV (p), at fixed environmental conditions parameter $\kappa \in \mathbf{\Lambda}$, is N. That is $k \in \{1,...N\}$, k is number of experiment. It should be noted that the j_k-th channel is dominant with respect to other channels $Ch(k \mid p,\kappa) \setminus j_k$.

Consider a binary function $d_j,\ j \in \mathbf{Z}$, which takes the value 1 when the SEV was detected in j-th channel, and the value 0 otherwise, $(j \notin \mathbf{Z}) \Rightarrow (d_j = 0)$.

For a particular $\kappa \in \mathbf{\Lambda}$ a priori fixed finite value $\mathbf{s}_\kappa > 0$ such that $\forall k, p : s_k^l, s_k^u \le \mathbf{s}_\kappa < \infty$. Consider the function $\mathbf{Ch}(j \mid \kappa)$, mapping set of indexes of channels $\mathbf{Z}$ on the set of its subsets such that $\mathbf{Ch}(j \mid \kappa) = \left\{ \chi(j - \mathbf{s}_\kappa), .., j-1, j+1, ..., \chi(j + \mathbf{s}_\kappa) \right\}$. Here $\mathbf{Ch}(j \mid \kappa) \subseteq \mathbf{Z}, \chi(a) = \delta(a) \cdot d_a$, $\delta(a) : (a < 0 \Rightarrow \delta(a) = 0, a > Z \Rightarrow \delta(a) = Z)$. Thus for dominant channel $j \in \mathbf{Z}$, function $\mathbf{Ch}(j \mid \kappa)$ maps it to the set of adjacent with it channels $\left\{ \chi(j - \mathbf{s}_\kappa), .., j-1, j+1, ..., \chi(j + \mathbf{s}_\kappa) \right\}$. The resulting channel set does not depend on the class of signal and is redundant. Redundancy is reflected in the fact that set $\mathbf{Ch}(j \mid \kappa)$ includes those channels in which signals of some classes may never be discovered. For example, at one and the same value κ, the elastic vibrations generated by a pedestrian, who is at a distance of 10 m from the OS, cannot be detected in the channel, located 30m from the dominant channel. However elastic vibrations that caused by a jackhammer will easily be detected in this channel, even though these vibrations are made at the same distance of 10 m from the OS. This kind of redundancy is intentional and created to standardize the data processing. Not knowing the SEV type, the system is a priori configured to use larger number of channels. If the SEV was not detected in the channel $k \in \mathbf{Ch}(j \mid \kappa)$, this channel is automatically excluded from the procedure of the multichannel classification. According to $\mathbf{Ch}(j \mid \kappa)$ definition, in case where the SEV was not detected in the channel $j \in \mathbf{Z}$, the corresponding element $k \in \mathbf{Ch}(j \mid \kappa)$ is zero. Let's call function $\mathbf{Ch}(j \mid \kappa)$ a channel template for a parameter κ. The argument j of the channel template is an index of the dominant channel.

The output of this template will be a set of channel indexes, which includes all possible channel indexes, in which the SEV may be detected, while it situated in the area of position $\Omega(r \mid \Delta_w) = \langle \kappa \rangle_1$. This area is located near the dominant channel j and has an absorption coefficient $\bar{\alpha} \in \alpha = \langle \kappa \rangle_2 \in \mathrm{A}$. The result of training (learning) is a set of classifiers $\mathbf{C} = \{ f_p^{(j,\kappa)}(\square) \mid p \in \{1,...,D\}, \kappa \in \mathbf{\Lambda}, j \in \bigcup_l \mathbf{Ch}(l \mid \kappa)\}$. Here $f_p^{(j,\kappa)}(\square)$ - discriminating function SVM of the j-th channel classifier, under conditions κ for the class p; index $\kappa \in \mathbf{\Lambda}$ refers to the learning environment (area and absorption coefficient) of the corresponding ensemble of classifiers $f_p^{(j,\kappa)}(\square)$, $p \in \mathbf{P}$. For the j-th channel vector model parameters corresponding to the SEV(p) in this channel is denoted $\lambda_{pk}^{(j,\kappa)}$. Thus for fixed $\kappa \in \mathbf{\Lambda}$, the training sample size N for the j-th channel and class p is the following set: $\mathrm{M}^{(p,j,\kappa)} = \left\{\left(\lambda_{pk}^{(j,\kappa)}, \tilde{y}_k(p \mid j,\kappa)\right)\right\}_{k=1}^{N}, \tilde{y}_k(p \mid j,\kappa) \in \mathbf{P}, p \in \mathbf{P}$. On all channels, all classes and all values $\kappa \in \mathbf{\Lambda}$, the training set is: $\mathbf{M} = \bigcup_{p,j,\kappa} \mathrm{M}^{(p,j,\kappa)}$. For fixed $\kappa \in \mathbf{\Lambda}$, consider $\mathrm{M}(\kappa) \subseteq \mathbf{M} : \mathrm{M}(\kappa) = \bigcup_{p,j} \mathrm{M}^{(p,j,\kappa)}$. Set $\mathrm{M}(\kappa)$ is called $\kappa-$stratum of $\mathbf{M}$. In other words, the training sample $\mathbf{M}$ is stratified according to $\kappa \in \mathbf{\Lambda}$. This is a very important system feature. The $\kappa \in \mathbf{\Lambda}$ will be defined based on data from subsystems «D» and «E» at the sample classification stage. After that, based on $\kappa \in \mathbf{\Lambda}$, in the overall set of classifiers $\mathbf{C}$ will be a distinguished subset of classifiers $\mathrm{C}(\kappa) = \{ f_p^{(j,\kappa)}(\square) \mid p \in \mathbf{P}, j \in \bigcup_l \mathbf{Ch}(l \mid \kappa)\} \subseteq \mathbf{C}$, used for classifying depending on the dominant channel index j.

The solution of the classification problem was obtained by using the multi-class **Support Vector Machine (SVM) method**. The SVM is a state-of-the-art learning machine based on the structural risk minimization induction principle (Hearst et al.,1998). The one-against-all method was used to decompose the D-class problem into series of two-class problems and construct D binary SVM-classifiers, where each of which separates one class from all the rest. Thus, the p-th SVM is trained with all the training examples of the p-th class with positive labels and all the others with negative. In this case we have the training sets $\mathrm{M}^{(p,j,\kappa)} = \left\{\left(\lambda_{pk}^{(j,\kappa)}, \tilde{y}_k(p \mid j,\kappa)\right)\right\}_{k=1}^{N}, \tilde{y}_k(p \mid j,\kappa) \in \mathbf{P}$. Here N=950, j – number of the C-OTDR channel, $\kappa \in \mathbf{\Lambda}$, p – number of class, $p \in \mathbf{P}$. For each class $p \in \mathbf{P}$, j-th C-OTDR channel, $\kappa \in \mathbf{\Lambda}$, SVM classifier was developed with decision function:

$$f_p^{(j,\kappa)}(\lambda) = w_{p/j,\kappa}^T \phi(\lambda) + b_{p/j,\kappa} = \sum_{k=1}^{N} a_k^{p/j,\kappa} \tilde{y}_k(p \mid j,\kappa) K(\lambda_k, \lambda) + b_{p/j,\kappa} ,$$

where

$$\tilde{y}_k(p \mid j,\kappa) = \begin{cases} 1, \text{ if } \sigma(\lambda_k) = p, \\ -1, \text{ if } \sigma(\lambda_k) \neq p. \end{cases}$$

$a_k^{p/j,\kappa}$ are the Lagrange multipliers, $0 \le a_k^{p/j,\kappa} \ge C$, i = 1, . . . , N,

$\sum_{k=1}^{N} a_k^{p/j,\kappa} \tilde{y}_k(p \mid j,\kappa) = 0$. The parameters $w_{p/j,\kappa}^T, b_{p/j,\kappa}$ were defined so that

$$\underset{w_{p/j,\kappa}, b_{p/j,\kappa}, \xi}{\mathrm{Minimize}} : L(w_{p/j,\kappa}, \boldsymbol{\xi}^{p/j,\kappa}) = \begin{cases} 0.5\left\|w_{p/j,\kappa}\right\|^2 + C\sum_{i=1}^{N}\left\langle \xi^{p/j,\kappa} \right\rangle_i \\ s.t.\ \tilde{y}_k(p \mid j,\kappa)\left(w_{p/j,\kappa}^T \phi(\lambda) + b_{p/j,\kappa}\right) \ge 1\text{-} \left\langle \xi^{p/j,\kappa} \right\rangle_k ,\ \left\langle \xi^{p/j,\kappa} \right\rangle_k > 0 \end{cases}$$

$\xi^{p/j,\kappa} = \left(\left\langle \xi^{p/j,\kappa} \right\rangle_1, ..., \left\langle \xi^{p/j,\kappa} \right\rangle_N\right) \in R^N$. Here, C is a trade-off parameter, so-called soft margin parameter. The effectiveness of SVM is based on the choice of soft margin parameter C. Different values of C were tried and the one with the best cross-validation accuracy picked. We used the Bhattacharyya divergence to measure the degree of similarity between two probability distributions (GMMs). Following (Kailath ,1967), the Bhattacharyya-kernel $K\left(\lambda_s, \lambda_f\right)$ for two GMMs λ_s and λ_f is given by

$$K\left(\lambda_s, \lambda_f\right) = \frac{1}{8}\sum_{i=1}^{J}\left\{\left(\mu_{si} - \mu_{fi}\right)^T \left[\frac{\left(\Sigma_{si} + \Sigma_{fi}\right)}{2}\right]^{-1} \left(\mu_{si} - \mu_{fi}\right)\right\} +$$

$$\frac{1}{2}\sum_{i=1}^{J}\ln\left(\frac{\frac{\left(\Sigma_{si} + \Sigma_{fi}\right)}{2}}{\sqrt{\left|\Sigma_{si}\right|\left|\Sigma_{fi}\right|}}\right) - \frac{1}{2}\sum_{i=1}^{J}\ln\left(w_{si} w_{fi}\right).$$

The parameter $\kappa \in \mathbf{\Lambda}$ is defined during classification stage of sample λ, based on «D» и «E» data. Then based on $\kappa \in \mathbf{\Lambda}$ subset is defined: $\mathrm{C}(\kappa) \subseteq \mathbf{C}$. Next define subset $\mathrm{C}(j_d \mid \kappa) \subseteq \mathrm{C}(\kappa)$ corresponding to the dominant channel index j_d : $\mathrm{C}(j_d \mid \kappa) = \left\{ f_p^{(j,\kappa)}(\cdot) \middle| p \in \{1,...,D\}, j \in \mathbf{Ch}(j_d \mid \kappa)\right\} \subseteq \mathrm{C}(\kappa)$.

The ensemble of classifiers $\mathrm{C}(j_d \mid \kappa)$ corresponds to external conditions of observation $\kappa \in \mathbf{\Lambda}$ of the object SEV, which is the subject to classification. The SEV described by the set of parameters $\boldsymbol{\lambda}(j_d \mid \kappa) = \left\{\lambda_j^{(\kappa)} \middle| j \in \mathrm{C}(j_d \mid \kappa)\right\}$. Here $\lambda_j^{(\kappa)}$ is the model parameter, which has been calculated for pattern from the j-th channel at a fixed $\kappa \in \mathbf{\Lambda}$.

If classifier determines that observed vector parameters λ corresponding to SEV from class p, we note $s(p \mid \lambda)$. Otherwise note - $s(\neg p \mid \lambda)$. Overall processing of classification information of channels $\mathrm{C}(j_d \mid \kappa)$ is implemented using Decision-Level Fusion Scheme (ISO,2006):

$$p^*(\boldsymbol{\lambda}(j_d)) = \arg\max_{p \in \mathbf{P}} F_p^{(\kappa)}\left(\boldsymbol{\lambda}(j_d \mid \kappa)\right), F_p^{(\kappa)}\left(\boldsymbol{\lambda}(j_d \mid \kappa)\right) = \sum_{j \in \mathbf{Ch}(j_d|\kappa)} \psi_j\left(\kappa\right)\tilde{y}_p^{(j,\kappa)}(\lambda_j^{(\kappa)}),$$

$$\tilde{y}_p^{(j,\kappa)}(\lambda_j^{(\kappa)}) = sign\left(f_p^{(j,\kappa)}\left(\lambda_j^{(\kappa)}\right)\right), \left(\tilde{y}_p^{(j,\kappa)}(\lambda_j^{(\kappa)}) = 1\right) \Rightarrow s(p), \left(\tilde{y}_p^{(j,\kappa)}(\lambda_j^{(\kappa)}) = -1\right) \Rightarrow s(\neg p). \quad (1)$$

Weighting functions $\Psi = \left\{\psi_j\left(\kappa\right) \middle| j \in Z, \kappa \in \mathbf{\Lambda}\right\}$ are chosen to provide minimal posteriori loss or Bayes risk. Thus, $p^*(\boldsymbol{\lambda}(j_d))$ - is an index of class to which classifier (1) assigns the subject of classification SEV based on $\boldsymbol{\lambda}(j_d)$. Denote the posterior realization SEV in j-th channel, $j \in \mathrm{C}(j_d \mid \kappa) : y_p^{(j,\kappa)} = \begin{cases} 1, if \ \ \sigma(\lambda_j^{(\kappa)}) = \mathrm{p} \\ -1, if \ \ \sigma(\lambda_j^{(\kappa)}) \neq \mathrm{p} \end{cases}$, and for ensembles $\mathrm{C}(j_d \mid \kappa)$ $: y_p^{(\kappa)} = \begin{cases} 1, if \ \ \sigma(\boldsymbol{\lambda}(j_d)) = \mathrm{p} \\ -1, if \ \ \sigma(\boldsymbol{\lambda}(j_d)) \neq \mathrm{p} \end{cases}$.

Let us consider the following events: $\tilde{\omega}_{j,p}\left(\lambda_j^{(\kappa)}\right) : \left\{y_p^{(j,\kappa)} \neq \tilde{y}_p^{(j,\kappa)}(\lambda_j^{(\kappa)})\right\}$. Further, $\varepsilon\left(j \middle| p, \kappa\right) = \mathbf{E}_{\lambda, y_p^{(j,\kappa)}}\left(\tilde{\omega}_{j,p}\left(\lambda_j^{(\kappa)}\right) \middle| p\right)$ is the average total error of the j-th classifier with condition of fixed $p \in \mathbf{P}, \kappa \in \Lambda$. Given the hypothesis of independence of classification decisions on $\mathbf{P}$, $\boldsymbol{\varepsilon}\left(j\middle|\kappa\right)=\sum_{p\in\mathbf{P}}\varepsilon\left(j\middle|p,\kappa\right)D^{-1}$ is on average total error of the $j-th$ classifier on the $\mathbf{P}$.

Theorem 1. *For the exponential loss function, at fixed* p, j, κ *, selection of functions* Ψ *as* $\psi_j\left(\kappa\right)=\ln\left(\left(1-\varepsilon(j\mid\kappa)\right)/\varepsilon(j\mid\kappa)\right)\left(\sum_i \ln\left(\left(1-\varepsilon(i\mid\kappa)\right)/\varepsilon(i\mid\kappa)\right)\right)^{-1}$ *provides minimum value of posterior loss function* $\mathbf{E}_{\lambda,y_p^{(j,\kappa)}}\left(\exp\left(y_p^{(j,\kappa)}\tilde{y}_p^{(j,\kappa)}(\lambda)\right)\middle|p,j,\kappa\right)$.

Proof of Theorem 1 is based on the ideas outlined in (Friedman et al., 2000).

Consider definition of discriminating function on ensemble of classifiers, different from (1). In this case we have

$$\bar{F}_p^{(\kappa)}\left(\boldsymbol{\lambda}(j_d\mid\kappa)\right)=\sum_{j\in\mathbf{Ch}(j_d\mid\kappa)}\psi_j\left(\kappa\right)f_p^{(j,\kappa)}\left(\lambda_j^{(\kappa)}\right),\ \bar{p}^*(\boldsymbol{\lambda}(j_d))=\arg\max_{p\in\mathbf{P}}\bar{F}_p^{(\kappa)}\left(\boldsymbol{\lambda}(j_d\mid\kappa)\right). \quad (2)$$

Herein $\bar{p}^*(\boldsymbol{\lambda}(j_d))$ - an estimate for the real value of class $p_0 \in \mathbf{P}$, to which belongs the SEV. Consider an obvious representation: $f_p^{(j,\kappa)}\left(\lambda_j^{(\kappa)}\right)=\mathbf{E}_{p^0}\left(f_p^{(j,\kappa)}\left(\lambda_j^{(\kappa)}\right)\middle|\kappa,j\right)+\eta^{(\kappa,j)}\left(p\right)$, value $U\left(\mathrm{M}\right)=Diam\left(\mathrm{M}\right)\sqrt{\sum_i\psi_i^2\left(\kappa\right)}$ and a set of indexes $\Xi\left(P_c\right)=\{p\in\mathbf{P}\mid\bar{F}_{\bar{p}^*}^{(\kappa)}\left(\boldsymbol{\lambda}(j_d\mid\kappa)\right)-\bar{F}_p^{(\kappa)}\left(\boldsymbol{\lambda}(j_d\mid\kappa)\right)\le c(P_c)\}$, for some $c(P_c)>0$.

Theorem 2. *Let*

1. Random variables $\left\{\eta\left(p\middle|\kappa,j\right)\middle|p\in\mathbf{P}\right\}$ *are mutually independent and* $\forall p,j,\kappa:\mathbf{E}_{p^0}\left(\eta^{(\kappa,j)}\left(p\right)\right)=0$ *;*

2. $\exists L\in]0,\infty[:\ \forall_{\lambda_1,\lambda_2\in\mathrm{M}}\left|f_p^{(j,\kappa)}\left(\lambda_1\right)-f_p^{(j,\kappa)}\left(\lambda_2\right)\right|\le L\left\|\lambda_1-\lambda_2\right\|_{\mathrm{M}}$ *;*

3. $\exists\vartheta^*\left(\mathbf{P}\right)=Inf\left\{\vartheta\in R^1\middle|\forall p^0\in\mathbf{P}:\underset{p}{Max}\left[\mathbf{E}_{p^0}f_p^{(j,\kappa)}\left(\lambda\right)-\mathbf{E}_{p^0}f_{p^0}^{(j,\kappa)}\left(\lambda\right)\right]\le\vartheta\right\}$;

Then $P\left(p^0\in\Xi\left(P_c\right)\right)\ge P_c$ *, if* $c(P_c)=\vartheta^*\left(\mathbf{P}\right)+L\left(1-P_c\right)^{-0.5}U\left(\mathrm{M}\right), P_c\in]0,1[$.

The proof is similar to the proof of Theorem 1 in (Timofeev, 2012). The set $\Xi\left(P_c\right)$ with a probability of not less than prescribed value P_c contains index p^0 of the class, which the object SEV belongs. Value $\left|\Xi\left(P_c\right)\right|$ characterizes the quality of classification the object SEV for a fixed value of the confidence coefficient P_c.

4. Practical Application of the Suggested Method

The approach described in this report is used for the analysis of C-OTDR data to classify types of SEV. Parameters of the C-OTDR system: a) duration of the probe pulse - 10 ns; b) frequency sensing - 8 kHz; c) power of the probe signal - 15 mW; d)laser wavelength - 1550 nm . Tests were performed for all types of events from the set $\mathbf{P},\left|\mathbf{P}\right|=13$. In the process of testing the algorithms targeted SEV were detected at distances of 2 to 50 m from the OS, depending on the type of SEV. The best performance of the classification accuracy was obtained for the class "heavy equipment excavator": $P_I \approx 0.02, P_{II} \approx 0.01$ for distance 30m, and $P_I \approx 0.11, P_{II} \approx 0.12$ for distance 60m. Here, P_I is probability of the Type I errors; P_{II} - probability of the Type II errors. Worst performance of classification quality was achieved for the class "hand digging the soil": $P_I \approx 0.09, P_{II} \approx 0.08$ for distance 5m, and $P_I \approx 0.18, P_{II} \approx 0.22$ for distance 8m. The power of the confidence set $\Xi(0.9)$ for $P_c = 0.9$ does not exceed the value of 2. The tests were conducted in an area where the prevailed the clay soil. It is important to note that all SEV were detected and classified, when they were at a distance of 30 km from the location of the laser point data. From a practical standpoint, the results should be considered a success.

5. Conclusion

The suggested method is robust to the types of environments of seismoacoustic signals propagation. The results of the practical application of this method for multichannel classification of the C-OTDR signals proved high efficiency of this approach.

References

Blimes J.A., Gentle (1998) A Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Tech. Rep., Int'l Computer Science Institute, Berkeley CA. 97-021.

Chang Huai You, Kong Aik Lee, Haizhou Li (2009) A GMM supervector Kernel with the Bhattacharyya distance for SVM based speaker recognition, icassp, IEEE International Conference on Acoustics, Speech and Signal Processing: 4221-4224.

Friedman J., Hastie T., and Tibshirani R. (2000) Additive logistic regression: A statistical view of boosting (with discussions). The Annals of Statis- tics, 28(2):337–407.

Hearst M.A., Dumais S.T., Osman E., Platt J., and Scholkopf B. (1998) Support Vector Machines, IEEE Intelligent Systems, vol. 13(4):18-28.

ISO/IEC JTC 1/SC 37 N 1506, (2006) Biometrics, 2006-02-28.

Kailath T. (1967) The Divergence and Bhattacharyya Distance Measures in Signal Selection, IEEE Transactions on Communication Technology 15 (1): 52–60.

Timofeev A.V. (2012) The guaranteed estimation of the Lipschitz classifier accuracy: Confidence set approach. Journal Korean Stat. Soc. 41(1). P. 105-114.