International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:9, No:5, 2015

Multi-Channel Information Fusion in C-OTDR Monitoring Systems: Various Approaches to Classify of Targeted Events

Andrey V. Timofeev

***Abstract*—**The paper presents new results concerning selection of optimal information fusion formula for ensembles of C-OTDR channels. The goal of information fusion is to create an integral classificator designed for effective classification of seismoacoustic target events. The LPBoost (LP-β and LP-B variants), the Multiple Kernel Learning, and Weighing of Inversely as Lipschitz Constants (WILC) approaches were compared. The WILC is a brand new approach to optimal fusion of Lipschitz Classifiers Ensembles. Results of practical usage are presented.

***Keywords*—**Lipschitz Classifier, Classifiers Ensembles, LPBoost, C-OTDR systems, ν-OTDR systems.

I. Introduction

APPLICATION of the C-OTDR (Coherent Optical Time Domain Reflectometer) technology to decide various problems of extended objects remote monitoring is currently being evaluated as a very promising approach [1], [2]. 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, a 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, a fiber optical sensor (FOS). Mechanical stress on the FOS surface is caused by seismic acoustic waves. These waves are generated of the sources of elastic vibrations (SEV) which located in vicinity of laying the FOS. Upon reaching the FOS, seismoacoustic wave causes a local longitudinal microstrain on its surface. Those microstrains in turn, cause a change in the local refractive index of light in a relatively small sector of the FOS. As a result, the speckle pattern, which corresponds to this sector, changes significantly. Thus, the FOS quite accurately reflects the state of the seismoacoustic field in its vicinity.

An alternative to the C-OTDR approach is an approach based on the use of photon counting. Those systems are named as Photon-counting OTDR or ν-OTDR. Systems of this kind provide the good spatial resolution and long FOS. But ν-OTDR systems provide just greater size of monitoring cycle in contrast to C-OTDR monitoring systems. However this drawback is gradually eliminated.

The seismoacoustic field contains information about events that occur in the surface layers of the ground near the FOS. 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 seismoacoustic emission (structural acoustic wave). In this case, the frequency range of the seismoacoustic waves is in the interval of 0 Hz to 1000 Hz. The information, which is required for correct identify the type of SEV, is concentrated in the frequency range of 0 Hz to 500 Hz, while 95% of the meaningful information is in even the more narrow range of 0 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 (TSEV). For the convenience of data processing, the entire FOS 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 depends on the probe pulse length. In practice, TSEV 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

Andrey V. Timofeev is with the LPP “EqualiZoom”, Astana, 010000, Kazakhstan (phone: +7-911-191-42-67; e-mail: timofeev.andrey@gmail.com).

time, due to strongly anisotropic medium of the elastic vibrations propagation, the structure of the oscillations (speckle patterns) varies considerably between different C-OTDR 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 and distorted during propagation in the environment. Intensity of attenuation and distortion depend on the average absorption factor of the medium and on the distance from the oscillation point to the location of channel. C-OTDR monitoring systems perform three major tasks in the following sequence: a)vTask "D" (Detection) – detection of the TSEV; b) Task "E" (Estimation) – estimate of the location of the TSEV; c) Task "C" (Classification) – classification of detected TSEV by means of assigning it to one of *D* priori given classes. In the multichannel case, the task "C" has to be solved by creation the method of effective multichannel data fusion. There are number of various approaches to effective multichannel data fusion for task "C" (classification). This paper describes results of a comparing various multichannel data fusion approaches for TSEV classification including a brand new approach which based on weighing of inversely as Lipschitz Constants (WILC) and it allows to improve the generalization ability of the classification system.

II. Designations and Research Objective

Let us denote

- **C-OTDR channels.** $Ch(\kappa_k)$ is k-th C-OTDR channel, where a tuple $\kappa_k = (\mathrm{A}_k, R_k)$, here A_k is an absorption coefficient of k-th channel, R_k is a length of k-th channel.
- **Feature**. A tuple $(\underline{Z}, d)$ is a compact feature space where $\underline{Z}$ is a set of feature values, d is a metric of $\underline{Z}$, data of all channels belongs to $\underline{Z}$;
- **Set of SEV classes**. A set Θ is a finite set of indexes of SEV-classes, $|\Theta| = D$.
- **Training Set**. $\mathbf{Z}_T = \{(Z_i, \theta_i) \mid i = 1,...N\}, |\{(Z_i, \theta_i)\}| = N$, $Z_i = \{\mathbf{z}_{1i}, \mathbf{z}_{2i}, ..., \mathbf{z}_{mi}\}$, $\theta_i \in \Theta$, each of $\mathbf{z}_{ki} \in \underline{Z}, k \in \{1,...,m\}$, corresponds to $Ch(\kappa_k)$, and to θ_i;
- **True index of SEV class**. A $\theta^* \in \Theta$ is a **true index** of the SEV-class to which the samples $\mathbf{z}_k$ belong, thus θ^* is an index of a **target class**.
- **Samples to classify**. A set $Z = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m\}$ is feature sample set; each of $\mathbf{z}_k \in \underline{Z}, k \in \{1,...,m\}$, corresponds to $Ch(\kappa_k)$; in another words, we obtain the feature sample $\mathbf{z}_k$ from *k*-th channel $Ch(\kappa_k)$.
- **Lipschitz Margin Classifier**. Let $\mathrm{f}_k(\theta \mid \mathbf{z}_k), k \in \{1,...,m\}; \theta \in \Theta,$ be a binary Lipschitz Margin classifier (**LMC**) [7], [8] with Lipschitz Constant (LC) L_k; $\mathrm{f}_k(\theta \mid \mathbf{z}_k): Z \to \{\theta, \Theta \setminus \theta\}$ (concept: **one against all**); so, classifier $\mathrm{f}_k(\theta \mid \cdot)$ divides the feature space (Z, d) into two classes θ and $\Theta \setminus \theta$; $\mathrm{f}_k(\theta \mid \mathbf{z}_k) = (f_k(\theta \mid \mathbf{z}_k), R_k)$ here $f_k(\theta \mid \mathbf{z}_k) \in R^1$ is **discriminate (**stochastic**) functions** (so-called score-parameters, which shows similarity degree of a sample $\mathbf{z}_k$ regarding to class $\theta \in \Theta$; discriminant function $f_k(\theta \mid \mathbf{z}_k)$ explicitly dependent on the index hypothesis to be tested θ and implicitly on the index of the target class θ^*; R_k is the classification decision-making rule $R_k : \tilde{\theta}_k = Arg \underset{\theta \in \Theta}{Max}(f_k(\theta \mid \mathbf{z}_k))$; let us denote ϑ - set of LMC $f_k(\theta \mid \mathbf{z}_k, \vartheta)$ parameters, which needs to be tuned during of training process; otherwise, set ϑ will be denoted as LMCP or LMCP ϑ;
- **Ensemble of LMC**. $\mathbf{F}(\theta \mid Z) = \{\mathrm{f}_k(\theta \mid \mathbf{z}_k) \mid k = 1,...,m\}$ is an ensemble of the LMC;
- $\mathbf{F}_q : Z \to \{q, \Theta \setminus q\}$ is an integral classifier on the ensemble $\mathbf{F}(\theta \mid Z)$; $\mathbf{F}_\theta = (F(\theta | \mathbf{F}(\theta \mid Z)), \mathbf{R})$ $\mathbf{R} : \tilde{\theta} = Arg \underset{\theta \in \Theta}{Max}(F(\theta | \mathbf{F}(\theta \mid Z)))$ is output of integral classifier $F(\theta | \mathbf{F}(\theta \mid Z))$;
- $F(\theta | \mathbf{F}(\theta \mid Z))$ - **discriminate function** on the classifiers ensemble $\mathbf{F}(\theta \mid Z)$, $F(\theta | \mathbf{F}(\theta \mid Z)) = \sum_k \beta_k f_k(\theta \mid \mathbf{z}_k)$, where $\sum_k \beta_k = 1, \forall_k \beta_k \geq 0$; coefficients $\{\beta_k\}$ are determined by various methods, which are object of our investigation.

So, there exist *m* statistical independent C-OTDR channels. $\mathbf{Ch} = \{Ch(\kappa_k) \mid k = 1,...,m\}$. Each of those channels depends of external (environmental) parameters tuple $\kappa_k = (\mathrm{A}_k, R_k)$. Simply speaking, these channels transmit signals from sources of elastic vibrations (SEV) to FOS. Thus signals $\mathbf{z}_k \in Z_k, k \in \{1,...,m\}$ are outputs of C-OTDR channels $\mathbf{Ch}$. The tuple κ_k defines the effectiveness of channel $Ch(\kappa_k)$ for signal transmission.

The signals Z are contain relevant information about SEV time-frequency parameters. Every two channels $Ch(\kappa_k)$ and $Ch(\kappa_p)$ distort the SEV time-frequency parameters by differently because of external parameters κ_k and κ_p are different. Accordingly we suppose every two different samples $\mathbf{z}_k$ and $\mathbf{z}_p$ are statistically independent if $k \neq p$. For each C-OTDR channel $Ch(\kappa_k)$ are used appropriate *D* binary classifiers $\mathrm{f}_k(\theta_i \mid \mathbf{z}_k)$, $\theta_i \in \Theta$. Each LMC $\mathrm{f}_k(\theta_i \mid \mathbf{z}_k)$ is binary classifier, which divides the feature space (Z, d) into two classes θ_i and $\Theta \setminus \theta_i$.

So, we need to classify of the SEV type using observation Z of C-OTDR channels $\mathbf{Ch}$. An obvious approach to solving this problem is to use the ensemble of LMC ($\mathbf{F}(\theta \mid Z)$). But the problem of effective multichannel data fusion arises. There are number of various approaches to multichannel data fusion.

The goal of this paper is to compare some data fusion methods effectiveness. A number of known approaches and one a brand new method were studied. The brand new method is based on use of Lipschitz constants of LMC's.

III. Some Approaches to C-OTDR Multichannel Data Fusion for Multiclass Classification of TSEV

So the classification problem TSEV is reduced to the task of creating an effective multiclass classificator (MC) which is based on a LMC classifiers ensemble. We remark that an ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by weighted or unweighted voting) to classify new examples [6]. At any rate a MC learning method choice is a dominant problem. Usually the problem of learning a MC from training data is often addressed by means of kernel method (KM) [3], [4]. In this case each kernel corresponds to an appropriate channel of the set $\mathbf{Ch}$. For brevity we will not describe the baseline of this well-known method but we are going to pay attention to some KM modifications which are designed to work with LMC classifiers ensembles.

For the sake of simplicity, we will consider as a LMC a classic SVM [9]. By definition a SVM discriminant function $f_k(\theta \mid \mathbf{z}_k, \vartheta)$ depends on the parameters $\alpha \in R^N$ (N is a power of the training set $\mathbf{Z}_T$) and $b \in R^1$. Here α is a normal vector to the hyperplane, $|b|/\|\alpha\|$ is the perpendicular distance from the hyperplane to the origin, thus we have $\vartheta = \{\alpha, b\}$, and each tuple ϑ defines the hyperplane in the feature space.

A. Multiple Kernel Learning (MKL)

In contrast to baseline kernels selection ("averaging kernels" and "product kernels" [5], [10]), MKL kernel selection is to learn a kernel combination during the training phase of the algorithm. So, the MKL objective is to optimize jointly over a linear combination of kernels

$$\mathbf{k}\left(Z^{(i)}, Z^{(j)}\right) = \sum_{k=1}^{m} \beta_k \, \mathrm{k}_k\left(\mathbf{z}_{ki}, \mathbf{z}_{kj}\right) \text{ with LMCP } \vartheta = \{\alpha, b\}.$$

Here $Z^{(i)} = \{\mathbf{z}_{1i}, \mathbf{z}_{2i}, ..., \mathbf{z}_{mi}\}$, $Z^{(j)} = \{\mathbf{z}_{1j}, \mathbf{z}_{2j}, ..., \mathbf{z}_{mj}\}$, $\sum_{k=1}^{m} \beta_k = 1, \beta_k \geq 0$. MKL was originally introduced in [8]. Let us denote $K_i(Z) = \left(\mathrm{k}_i(Z, Z_1), \mathrm{k}_i(Z, Z_2), ..., \mathrm{k}_i(Z, Z_N)\right) \in R^N$, $i = 1, ..., m$. The final decision has form $F_{MKL}(Z) = Arg \underset{\theta \in \Theta}{Max}\left(\sum_{k=1}^{m} \beta_{k,\theta}\left(K_k(Z)^T \alpha_\theta + b_\theta\right)\right)$. The choice of parameters MKL is made by using for each θ the following scheme:

$$\min_{\alpha, b, \beta}\Big[0.5\left(\sum_{k=1}^{m} \beta_{k,\theta} \alpha_\theta^T K_k \alpha_\theta\right) +$$
$$C\sum_{i=1}^{N}\left(D\left(\theta_i, b_\theta + \sum_{k=1}^{m} \beta_{k,\theta} K_k^T(Z_i)\alpha_\theta\right)\right)\Big]$$
$$\text{sb.t.} \sum_{k=1}^{m} \beta_{k,\theta} = 1, \beta_{k,\theta} \geq 0$$

$D(\theta, t) = \max(0, 1 - \theta t)$, $(Z_i, \theta_i) \in \mathbf{Z}_T$. In other words, in MKL case we optimize jointly the convex hull of kernels. Here for each θ we have the same LMCP $\vartheta_\theta = \{\alpha_\theta, b_\theta\}$ for different k.

B. LP-Boost (LP-β)

So, we will consider a case when classifiers $f_k(\theta \mid \mathbf{z}_k)$ of ensemble $F(\theta | \mathbf{F}(\theta \mid Z))$ are not trained jointly, but coefficients $\{\beta_k\}$ are determined jointly. Here we have a situation where LMCP tuples ϑ_θ are different for different k. This method is called the β-LP-Boost [11], and here the final decision has the form

$$F_{LP\beta}(Z) = Arg \underset{\theta \in \Theta}{Max}\left(\sum_{k=1}^{m} \beta_k\left(K_k(Z)^T \alpha_{\theta,k} + b_{\theta,k}\right)\right).$$

The training phase comes down to an optimal choice of parameters $\{\beta_k\}$. This choice is performed by using standard optimization method (linear programming - LP) according to:

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

under the condition:

$$\sum_{k=1}^{m} \beta_k\left(\mathrm{k}_k(Z, Z_i)\alpha_{\theta',k} + b_{\theta',k}\right) -$$
$$\arg\max_{\theta^* \neq \theta'} \sum_{k=1}^{m} \beta_k\left(\mathrm{k}_k(Z, Z_i)\alpha_{\theta^*,k} + b_{\theta^*,k}\right) + \xi_i \geq \rho,$$
$$i = 1, ..., N, \ \sum_{k=1}^{m} \beta_k = 1, \beta_k \geq 0, k = 1, ..m.$$

here ξ - slack variables, ν - regularization constant, which is chosen using Cross Validation (CV). In frame of this approach not need provide the normalization of kernels $\mathrm{k}_k(\cdot)$. Moreover, features for which $\beta_k = 0$ need not to be computed for the final decision function.

C. LP-Boost (LP-B)

Another version of LP approach to choice $\{\beta_k\}$ was called B-LP-Boost [12]. In this case, each class has its own weight vector. So, we have (m x D) weighting matrix B. The final decision has the form:

$$F_{LPB}(Z) = Arg \ \underset{\theta\in\Theta}{Max}\left(\sum_{k=1}^{m} B_k^{\theta}\left(K_k\left(Z\right)^T \alpha_{\theta,k} + b_{\theta,k}\right)\right)$$

Choice of parameters $\{\beta_k\}$ we make in such way:

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

under the condition:

$$\sum_{k=1}^{m} B_k^{\theta'}\left(\mathrm{k}_k\left(Z,Z_i\right)\alpha_{\theta',k}+b_{\theta',k}\right)-$$
$$\sum_{k=1}^{m} B_k^{\theta''}\left(\mathrm{k}_k\left(Z,Z_i\right)\alpha_{\theta'',k}+b_{\theta'',k}\right)+\xi_i\geq\rho,$$
$$i=1,...,N;\theta'\neq\theta'',\ \ \forall\theta,\mathrm{m}:\left(\sum_{k=1}^{m}B_k^{\theta}=1,B_k^{\theta}\geq 0,k=1,..m\right).$$

Here ξ are slack variables, ν - regularization constant, ν is chosen using CV. Here we have a linear programming problem too, but this problem is more expensive because of dimension increasing.

D.MKL Weighing of Inversely as Lipschitz Constants (WILC- MKL)

Let us consider the brand new modification of the MKL that differ from classical MKL by method choice of linear combination parameters. The motivation of this approach is using some intrinsic properties of LMC. The fact is that value of Lipschitz Constant significantly determines of the LMC properties. Simply speaking, the Lipschitz classifier decision function has to a small Lipschitz constant. This feature comes from well-known regularization principle, which recommends avoid using discriminative functions with a high variation. So, LMC's with small LC are more preferable for providing of stable classification process. In other words, classifiers with small LC provide the greater generalization ability of classification system. Hence, in formula of $F\left(\theta\middle|\mathbf{F}(\theta\mid Z)\right)$ LMC's with small LC must get weight coefficients with bigger value. Let us call this approach to modification of MKL as Weighing of Inversely as to value of the Lipschitz Constant (WILC) or WILC-MKL. In frame of WILC-MKL approach to LMC-ensemble $\mathbf{F}(\theta\mid Z)$ we have the following discriminative function:

$$F_{WILC}(Z) = Arg \ \underset{\theta\in\Theta}{Max}\left(\sum_{k=1}^{m}\sigma_{\theta,k}\beta_{k,\theta}\left(K_k\left(Z\right)^T\alpha_{\theta}+b_{\theta}\right)\right).$$

Here

$$\sigma_{\theta,k}=L_{\theta,k}^{-1}\left(\sum_{j=1}^{m}L_{\theta,j}^{-1}\beta_{j,\theta}\right)^{-1},k=1,...m\,,\sigma_{\theta,k}\beta_{k,\theta}>0\,,$$

$\sum_{k=1}^{m}\sigma_{\theta,k}\beta_{k,\theta}=1$, $L_{\theta,k}$ is Lipschitz Constant of discrimination function $K_k\left(Z\right)^T\alpha_{\theta}+b_{\theta}$. Thus, using WILC-MKL, we make attempt to improve the generalization ability of MKL by considering information about variation characteristics of classifiers discrimination functions. As it was shown in series of practical experiments, usage of WILC-MKL allows considerably improve the performance of LMC-ensemble in some practical cases.

IV. Results of Practical Usage

All of above described methods were used for multichannel TSEV classification in C-OTDR system of railways monitoring. This system was successfully installed on the railways test area (RTA) of Kazakhstan Railways Company (JSC NC "KTZ") in September of 2014, and this system continues to operate. The RTA is located at a distance of 10 km from Astana City. The FOS which has been installed on RTA has length around 2000 m, depth of the FOS laying is 50 cm approximately, and FOS offset from rails is 5 m. Parameters of the C-OTDR system: a) duration of the probe pulse is 50-200 ns; b) period of probe pulse ~ 50-300 µs; c) laser wavelength - 1550 nm.

In this case, the main problem is to fusion of multichannel data to classify the TSEV with maximum accuracy. As was said above, for each C-OTDR channel $Ch(\kappa_k)$ are used appropriate D binary classifiers $\mathrm{f}_k\left(\theta_i\mid\mathbf{z}_k\right)$, $\theta_i\in\Theta$. Each LMC $\mathrm{f}_k\left(\theta_i\mid\mathbf{z}_k\right)$ is binary classifier, which divides the feature space $\left(\underline{Z},d\right)$ into two classes θ_i and $\Theta\setminus\theta_i$. Each LMC $\mathrm{f}_k(\cdot)$ was trained independently, and each LMC uses the same set of features in the space $\left(\underline{Z},d\right)$. The $\left(\underline{Z},d\right)$ is the ordinary GMM-vector space [13]. We describe the procedure for calculation of the GMM-vectors very briefly. On feature extraction phase for each speckle pattern obtained in the probing period T for each of the channel are built Linear-Frequency Spaced Filterbank Cepstrum Coefficients (LFCC). In our case these features are based on 10 linear filter-banks (from 0.1 to 500 Hz) derived cepstra. Thus, 10 static and 10 first-order delta coefficients were used, giving the feature order m = 20. Further, approximation of the probability distribution function of the feature vectors (LFCC) by semi-parametric multivariate probability distribution model, so-called Gaussian Mixture Models (GMM), was carried. Presently, the GMM is one of the principal methods of modeling broadband acoustic emission sources (including TSEV) for their robust identification. The GMM of TSEV feature vectors distribution is a weighted sum of J components densities [13] and given by the equation $P\left(x\middle|\lambda_s\right)=\mathbf{w}_s\mathbf{B}_s^T(x)$, where x is a random m-vector, $\mathbf{w}_s=\left(w_{s1},...w_{sJ}\right)\in R^J$,

$$\mathbf{B}_s(x)=\left(B_{s1}\left(x\right),...B_{sJ}\left(x\right)\right)\in R^J,$$
$$\underset{s,i}{\forall}B_{si}\left(x\right)=\left(\left(2\pi\right)^{m/2}\left|\Sigma_{si}\right|^{1/2}\right)^{-1}$$
$$\exp\left(-\frac{1}{2}\left(x-\mu_{si}\right)^T\Sigma_{si}^{-1}\left(x-\mu_{si}\right)\right)$$
$$\lambda_s=\left\{\left(w_{si},\mu_{si},\Sigma_{si}\right)\mid i=1,J\right\}.$$

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 [13]. In the described experiments value J was equal to 1024. Thus, for identification of TSEV class, each TSEV is modeled by a GMM-vector and is referred to as his model parameters $\lambda \in \underline{Z}$. The classic SVM with Bhattacharyya-kernel [2] was used as the LMC.

TABLE I
THE PRACTICAL DETECTION RESULTS

Method	Type of TSEV	Accuracy	Volume of training set
MKL	"hand digging the soil"	76%	60
	"chiselling ground scrap"	79%	60
	"pedestrian"	78%	80
	"group of pedestrians "	79%	30
	"passenger car"	79%	50
	"train"	100%	150
	"heavy equipment excavator"	81%	20
	"easy excavation equipment"	83%	20
	"shrew digging the ground"	81%	30
LP-β	"hand digging the soil"	81%	60
	"chiselling ground scrap"	83%	60
	"pedestrian"	81%	80
	"group of pedestrians "	83%	30
	"passenger car"	80%	50
	"train"	100%	150
	"heavy equipment excavator"	85%	20
	"easy excavation equipment"	86%	20
	"shrew digging the ground"	84%	30
LP-B	"hand digging the soil"	82%	60
	"chiselling ground scrap"	85%	60
	"pedestrian"	79%	80
	"group of pedestrians "	84%	30
	"passenger car"	81%	50
	"train"	100%	150
	"heavy equipment excavator"	86%	20
	"easy excavation equipment"	88%	20
	"shrew digging the ground"	88%	30
WILC-MKL	"hand digging the soil"	81%	60
	"chiselling ground scrap"	82%	60
	"pedestrian"	78%	80
	"group of pedestrians "	83%	30
	"passenger car"	84%	50
	"train"	100%	150
	"heavy equipment excavator"	84%	20
	"easy excavation equipment"	87%	20
	"shrew digging the ground"	86%	30

Priori defined target classes of TSEV, which collectively makes up a finite set Θ. For example, in case of railways monitoring the array Θ consists of the following TSEV classes: "train", "hand digging the soil", "chiseling ground scrap", "pedestrian", "group of pedestrian", "passenger car", "heavy equipment excavator", "easy excavation equipment". Five alternative approaches for multichannel data fusion were compared on stage of TSEV classification. In particular, MKL, LP-β, LP-B, and WILC-MKL approaches were used. The results of using these methods as parts of the C-OTDR system are presented in Table I.

In the process of using the method WILC-MKL values of Lipschitz Constants were evaluated numerically for each LMC from ensemble $\mathbf{F}(\theta \mid Z)$. The volumes of training sets were equal for each of various data fusion approaches, but those volumes were different for various TSEV types. Presented results prove that LP (β and B) are more effective with respect to MKL, and WILC-MKL approaches. At the same time, WILC-MKL is more effective compared to MKL, but the LP-B is the best approach for a fusion of multichannel data in C-OTDR monitoring systems. It is important: the LP-B approach requires more computing resources than the WILC-MKL approach, wherein the accuracies of those methods are close. That is why the WILC-MLK approach is preferable from the practical point of view.

V. CONCLUSION

This paper describes results of comparison of various multichannel data fusion approaches for TSEV classification including MKL, LP-β, LP-B, and WILC-MKL. The practical usage of these approaches proves better effectiveness (in sense of accuracy) of LP-B approach to fusion of multichannel data for classification of TSEV type. A brand new approach, WILC-MKL, was suggested for multichannel data fusion. This approach is simple to use (it requires less computing resources than LP-B) and performs well in a C-OTDR classification subsystem.

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] K. N. Choi, J. C. Juarez, and H. F. Taylor (2003), Distributed fiber-optic pressure/seismic sensor for low-cost monitoring. SPIE 5090, pp.134-141.

[2] A.V. Timofeev, D.V. Egorov (2014), Multichannel classification of target signals by means of an SVM ensemble in C-OTDR systems for remote monitoring of extended objects, MVML-2014 Conference Proceedings V.1, Prague.

[3] T. Hofmann, D. Sholkopf, J. Smola (2008) Kernel Methods in machine Learning, Annals of Statistics, V.36, No. 3, pp. 1171-1220.

[4] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan. (2004) Learning the kernel matrix with semidefinite programming. JMLR, 5, pp. 27–72.

[5] Jebara, T., Kondor, I., Howard, A. (2004) Probability product kernels. Journal of Machine Learning Research, 5, pp. 819–844.

[6] M. Narasimha Murty, V. S. Devi (2011) Combination of Classifiers in Pattern Recognition (Undergraduate Topics in Computer Science), Vol. 0, (pp. 188-206), London: Springer.

[7] A.V. Timofeev (2012) The guaranteed estimation of the Lipschitz classifier accuracy: Confidence set approach. Journal Korean Stat. Soc. 41(1). pp. 105-114.
[8] U., Luxburg, O. Bousquet (2004) Distance-based classification with Lipschitz functions, Journal of Machine Learning Research Vol. 5:, pp. 669-695.
[9] M. A. Hears, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, Support Vector Machines, IEEE Intelligent Systems, vol. 13(4), pp.18-28, 1998.
[10] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan (2004) Multiple kernel learning, conic duality, and the SMO algorithm. In ICML.
[11] A. Demiriz, K.P. Bennett, and J. Shawe Taylor (2002) Linear programming boosting via column generation. Machine Learning, vol. 46 (1-3), pp. 225-254.
[12] P. Gehler and S. Nowozin (2009) On feature combination for multiclass object classification. In Proc. ICCV.
[13] J. A. Blimes, 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. pp. 97-021

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 seismic.