*SOCIETY. INTEGRATION. EDUCATION*
*Proceedings of the International Scientific Conference. Volume V, May 25th-26th, 2018. 475-483*

# APPROACHES AND SOLUTIONS FOR SIGN LANGUAGE RECOGNITION PROBLEM

**Aleksejs Zorins**

**Pēteris Grabusts**
Rezekne Academy of Technology, Latvia

***Abstract**. The goal of the paper is reviewing several aspects of Sign Language Recognition problems focusing on Artificial Neural Network approach. The lack of automated Latvian Sign Language has identified and proposals of how to develop such a system have made. The authors use analytical, statistical methods as well as practical experiments with neural network software. The main results of the paper are description of main Sign Language Recognition problem solving methods with Artificial Neural Networks and directions of future work based on authors' previous expertise.*
***Keywords:** Sign language recognition, artificial neural networks, Latvian sign language.*

## Introduction

One of very important people's communication components is a gesture (sign language), which allows expressing emotions and provides comprehensible information in addition to spoken language. For hearing majority the gesture is an additional method for communication, for deaf community this is the only way to express themselves. Deaf people are being integrated into society through the sign language, the part of which is the representation of the national alphabet gestures.

There is a lack of computerized Latvian sign language recognition system and there is a strong necessity for it. The goal of the authors is to develop the recognition system of Latvian Sign Language based on Artificial Neural Networks (ANN) to help deaf people to integrate into society.

Numerous applications of Artificial Neural Networks (ANN) exist at the present time with different learning algorithms, topologies etc. It is strongly believed that ANN is built using human brain's functioning principles but still ANN is a tricky way for real problem solving, because in any application should be found answers to the following questions (Fausett, 1994):

Is the network complex enough to be capable to encode a solution?

Is it possible to find solution in a reasonable amount of time?

How can we guarantee that a trained network is matching closely enough our problem domain and hidden regularities in the data?

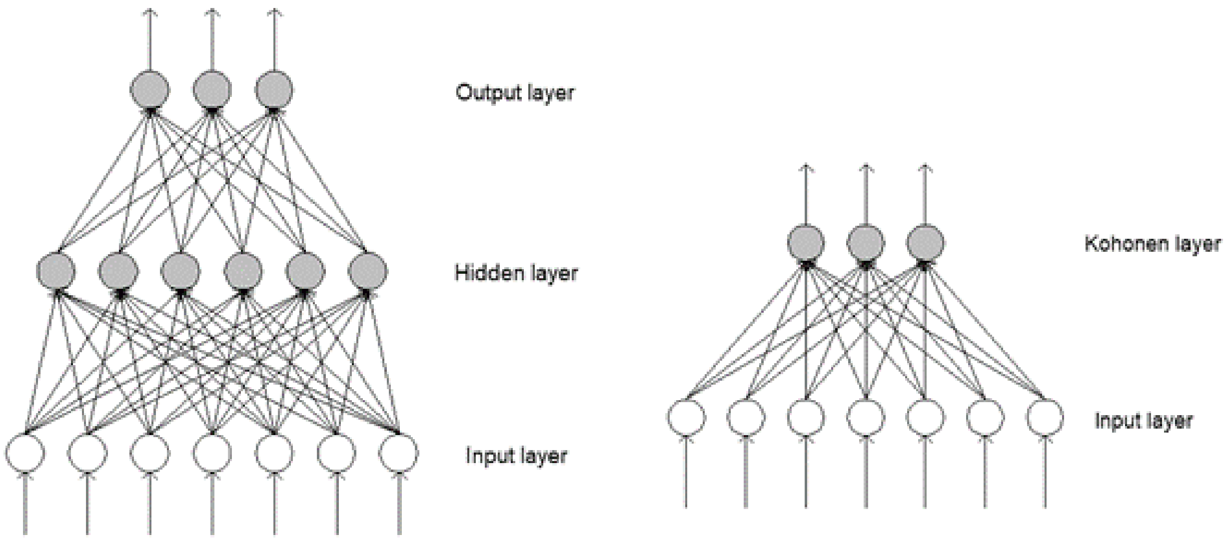

*http://dx.doi.org/10.17770/sie2018vol1.3082*

In the recent years modern science advanced a lot in understanding human brain functions and structure (Cooper, 2011). Each learning algorithm and each network topology should be carefully developed to solve more or less complex problem in real life. One may say that almost each serious application requires its own network topology, algorithm and data pre-processing. The same case is connected with sign language recognition problem.

## The Essence of Artificial Neural Network

A neural network is a set of interconnected simple processing elements, or neurones. Neural networks are potentially useful for studying the complex relationships between inputs and outputs of a system. There are two neural network models investigated in this research: backpropagation networks and Kohonen self-organizing maps.

A multilayer feed forward network with an appropriate pattern of weights can be used to model some mapping between sets of input and output variables. Figure 1a shows an example of feed forward network architecture, with three output units and one hidden layer, which can be trained using backpropagation. The shaded nodes in figure 1a are processing units. The arrows connecting input and hidden units and connecting hidden units and the output units represent weights.




**a) Backpropagation network,** **b) Kohonen self-organizing map**

Fig. 1. **The architecture of two types of networks.**

The backpropagation learning algorithm (Rojas, 1996) is formulated as a search in the space of the pattern of weights, *W*, in order to find an optimal configuration, *W**, which minimizes an error or cost function, *E(W)*. The pattern of weights will then determine how the network will respond to any arbitrary input. The error or cost function is defined by (1):

$$E = \frac{1}{2}\sum_{i}\sum_{p}(t_{ip} - o_{ip})^2 \tag{1}$$

This function compares an output value $o_{ip}$ to a desired value tip over the set of *p* training vectors and i output units. The gradient descent method is used to search for the minimum of this error function through iterative updates:

$$W(k+1) = W(k) - \eta\nabla E \tag{2}$$

where $\eta$ is the learning rate, and $\nabla E$ is an estimate of the gradient of *E* with respect to *W*.

The algorithm is recursive and consists of two phases: forward-propagation and backward-propagation. In the first phase, the input set of values is presented and propagated forward through the network to compute the output value for each unit. In the second phase, the total-squared error calculated in the first phase is propagated from the output units to the input units. During this process, the error signal is calculated recursively for each unit in the network and weight adjustments are determined at each level. These two phases are executed in each iteration of the backpropagation algorithm until the error function converges.

The main difference between them and conventional models is that the correct output cannot be defined a priori, and therefore a numerical measure of the magnitude of the mapping error cannot be used (Rojas, 1996). However, the learning process leads to the determination of well-defined network parameters for a given application.

The self-organizing networks assume a topological structure among the cluster units. This property is observed in the brain, but is not found in other artificial neural networks (Konar, 2005). There are m cluster units, arranged in a one- r two-dimensional array: the input signals are n-dimensional. Figure 1b shows architecture of a simple self-organizing network, which consists of input and Kohonen or clustering layer. The shadowed units in the figure 1b are processing units. This simplified network may cluster the data into three classes, but in the real problem domains one clustering unit for each class is not enough, therefore we should understand each Kohonen layer neurone in the Figure as a number of units (cluster of neurones).

When a self-organizing network is used, an input vector is presented at each step. These vectors constitute the "environment" of the network. Each new input produces an adaptation of the parameters. If such modifications are correctly controlled, the network can build a kind of internal representation of the environment.

Consider the problem of charting an n-dimensional space using a one-dimensional chain of Kohonen units (Fausett, 1994). The units are all arranged in sequence and are numbered from 1 to m (see figure 2).

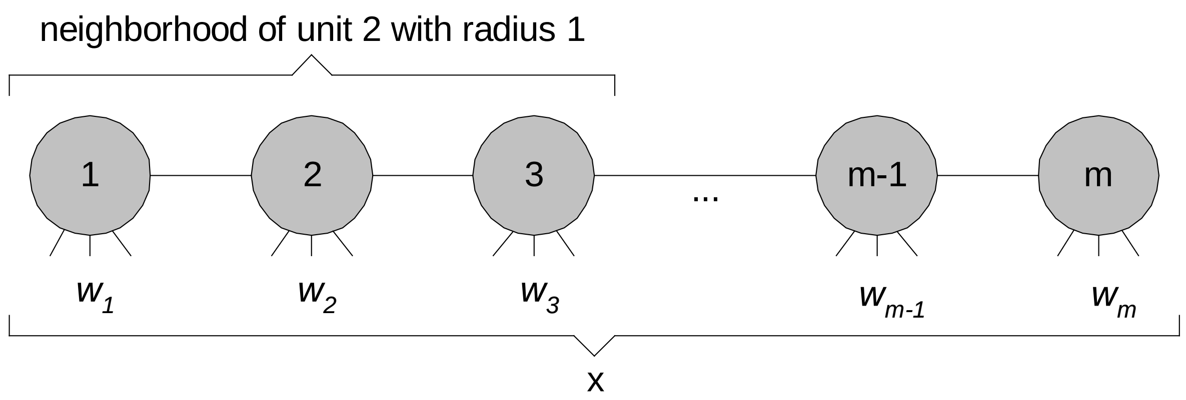



Fig. 2. **A one-dimensional lattice of computing units**

The *n*-dimensional weight vectors $w1$, $w2$, …, $w_m$ are used for the computation. The objective of the charting process is that each unit learns to specialize on different regions of input space. When an input from such a region is fed into the network, the corresponding unit should compute the maximum excitation. Kohonen's learning algorithm is used to guarantee that this effect is achieved.

A Kohonen unit computes the Euclidian distance (the dot product metric can also be used) between an input $x$ and its weight vector $w$. In the Kohonen one-dimensional network, the neighbourhood of radius 1 of a unit at the $k$-th position consists of the units at the positions $k-1$ and $k+1$. Units at both ends of the chain have asymmetrical neighbourhoods. Kohonen learning uses a neighbourhood function $\phi$, whose value $\phi(i, k)$ represents the strength of the coupling between unit $i$ and unit $k$ during the training process. The complete description of Kohonen learning algorithm can be found in (Rojas, 1996) and (Fausset, 1994).

## Sign Language Recognition Task

Let us consider the sign language alphabet recognition task. These alphabets mainly consist of static signs; however, the Latvian sign language (LSL) additionally has several signs, which are shown in motion (see figure 3).

In Latvia there is a website of Latvian Deaf People Rehabilitation, which has a Sign Language Interpreters' Department. The main goal of this organization is to "facilitate the client's social integration, availability of necessary information and services, provide sign language interpreter's services for communication with other individuals and legal entities according to the client's perception and communication abilities" (*The Latvian Sign Language Development Department*).

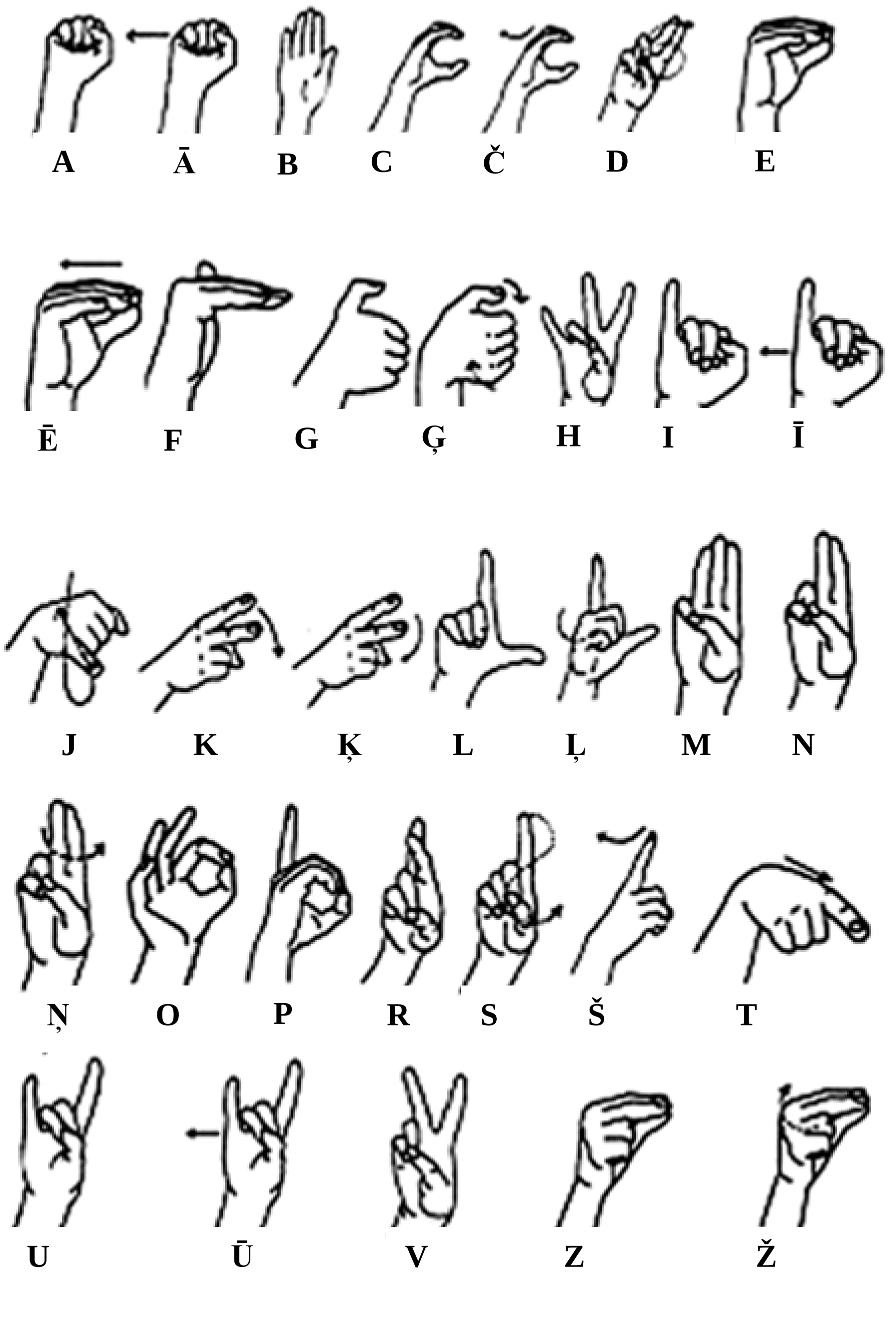



Fig. 3. **The symbols of Latvian sign language**
(*The Latvian Sign Language Development Department*)

There is still a lack of computerized LSL recognition in our country and that is why it is worth developing the recognition system of Latvian Sign Language based on Artificial Neural Networks (ANN) to help people in social rehabilitation and integration into society.

Scientific resources provide us a wide range of sign language recognition (SLR) methods, which could be classified using input data and sensor technology (Cooper, 2011):

- Using different hand markers;
- Specially developed data gloves;
- Infra-red sensor technology;
- Visual (video or photo) methods.

The last two methods allow recognizing the sign language in a real time, for example, through web cameras or Leap Motion technology, which price is now affordable. There is a large number of research works, using mathematical and statistical methods for SLR task and numerous applications with artificial neural networks, genetic algorithms, hidden Markov models etc.

Wide range of classification and recognition algorithms raises difficulties in choosing appropriate method for specific SLR task. The analysis of current situation in the field allows concluding that this issue is still in progress and no final solution has been proposed yet.

The situation is the same with different approaches for SLR in real time, using video streaming data.

Collection and proper pre-processing of the data is a crucial step for successful SLR problem solution. The most popular approach now is the use of web or video cameras due to affordable costs of hardware for this application (Cooper, 2011). The data glove technology has several disadvantages – the low cost gloves provide little information about the gesture while the more precise ones are more expensive, for instance, Myo Gesture Control Armband is two times more expensive than LeapMotion controller. There are also some difficulties with putting on and off this device, which may not be appropriate for fast and easy use in public facilities and organizations.

It should be concluded that the most perspective technologies now are video streaming information for real time language processing, the static pictures for an alphabet only (not appropriate for Latvian alphabet thou) and the LeapMotion technology, which is the most affordable device on the market today. The authors of the paper have not found any information on using LeapMotion for SLR task; therefore, it is one of the most perspective future research directions.

**Artificial Neural Network Approaches for Sign Language Recognition**

After a proper data preparation the neural network has been trained on the test samples of signs and then this trained network is used for earlier unknown data recognition.

Dogic and Karli have used back-propagation training algorithm with sigmoid activation function and two hidden layers. Their ANN has 15 input neurons and 90 training samples. As a result they obtained 84 % correctly classified patterns with a 40 neurons in a hidden layer (Dogis, 2014).

Byeongkeun Kang used Convolutional Neural Networks consisting of 5 convolution layers, 3 max-pooling layers, and 3 fully connected layers (Kang, 2015). The author extracted 4096-dimensional feature vector and trained the network with 5 different settings and achieved very impressive 99 % correctly classified cases.

Mekala has used combinational ANN which is based on the cache search memory concept of a CPU hen all blocks of a system have dual information exchange bus (Mekala, 2011; Mekala, 2013). The recognition part is also based on error back-propagation ANN. The algorithm detects all letters from A to Z with 100 % accuracy.

Another interesting approach is Modular Neural Networks presented in (Zorins, 2009). An important advantage of modular approach is an improvement of generalization (ability to perform well on the test data) due to decomposition of complex function into simpler ones. The modular neural networks have been proposed to solve this problem. The main idea is a natural decomposition of a function of large complexity into simple functions and realization of each function by a separate neural network (Rojas, 1996).

There are other options to use in recognition part. The authors are going to implement their own ANN learning algorithms and architectures developed earlier.

One of these examples could be Kohonen “freezing” learning algorithm (Zorins, 2007). The standard Kohonen self-organizing maps may be trained in unsupervised (in most cases) and in supervised manner. This type of network uses grid of neurons or a topological structure among the cluster units. The modified “freezing” algorithm developed by the author allows splitting network learning process into some stages, when each part of the network is trained individually. The algorithm flowchart is shown in figure 4.

In the first learning phase, neural network is split into number of clusters of neurons, where each of the clusters is associated with dataset class. In this way, we obtain training with teacher. In the second stage, each cluster is trained accordingly to standard Kohonen learning algorithm. Each of neuron clusters is trained individually, while others are “frozen” and do not take part in the training.

After completion of individual cluster training the network is "de-frozen" and learning process ends.

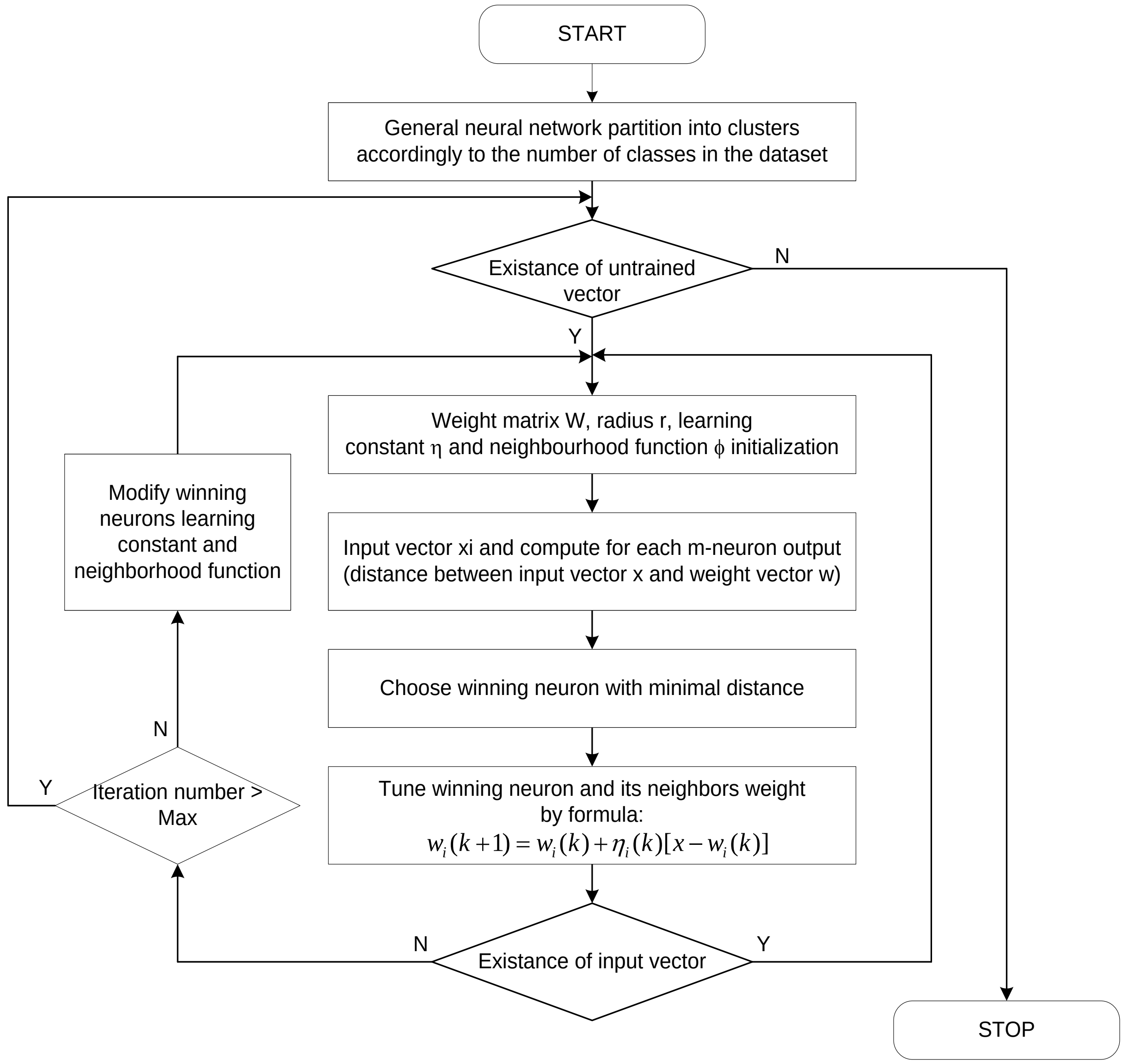



Fig. 4. **Kohonen neural network "freezing" learning algorithm flowchart**

## Conclusions

There is a number of successful sign language recognition applications using Artificial Neural Networks, however, not for Latvian sign language. The goal of the authors is to develop Latvian Sign Language recognition system based on artificial neural networks. In order to do that it necessary to:

- Develop a database of Latvian sign language using infra-red sensor technology;

- Develop and implement appropriate data pre-processing method;
- Analyze, choose and implement the most appropriate Artificial Neural Network architecture and training algorithm for the sign language recognition task;
- Design software for Latvian deaf community.

We are in the beginning of our research hoping to help people with special needs to integrate into society and advance in a very interesting field of Artificial Intelligence.

**References**

Cooper, H., Holt, B., & Bowden, R. (2011). *Sign Language Recognition, Chapter in Visual Analysis of Humans*: Looking at People, Springer.

Dogic, S., & Karli, G. (2014). Sign Language Recognition using Neural Networks. *TEM Journal 3* (4), 296-301.

Fausett, L. (1994). *Fundamentals of Neural Networks. Architectures, algorithms and applications*, Prentice Hall.

Kang, B., Tripathi, S., & Nguyen, T. (2015). Real-time Sign Language Fingerspelling Recognition using Convolutional Neural Networks from Depth map. In: *3rd IAPR Asian Conference on Pattern Recognition, Kuala Lumpur, Malaysia.*

Konar, A. (2005). *Computational intelligence: principles, techniques and applications*. - Springer-Verlag, London.

Mekala, P. etc. (2013). Gesture Recognition Using Neural Networks Based on HW/SW Cosimulation Platform. *Advances in Software Engineering, Volume 2013.*

Mekala, P. etc. (2011). Real-time Sign Language Recognition based on Neural Network Architecture. *IEEE 43rd Southeastern Symposium on System Theory*, 195-199.

Rojas, R. (1996). *Neural networks. A systematic approach*, Springer, Berlin.

*The Latvian Sign Language Development Department*. Retrieved from http://rc.lns.lv/index.php.

Zorins, A. (2007). *Improvement possibilities of interval value prediction using Kohonen neural networks*. *Riga Technical University conference proceedings, Vol. 31*, 8.-16.

Zorins, A. (2009). Retail turnover prediction using modular artificial neural networks. *Proceedings of 7th International Scientific Practical Conference "Environment. Technology. Resources"*, Rezekne, Latvia.