Using of a Convolutional Neural Network with Changing Receptive Fields in the Tasks of Image Recognition

R.M. Nemkov, O.S. Mezentseva and D. Mezentsev

Abstract In the article synthesis procedure of mathematical model features of convolution neural network (CNN) is described. In order to improve the generalization capability of the network the training set is generated by adding a distorted image with changing of CNN receptive fields. This fact differs given procedure from the known procedures. We propose the reduction algorithm of an extended training set and the synthesis algorithm of features for CNN with non-standard receptive fields. The experiments results of the developed algorithms were shown in the article in order to assess of generalization capability changes of the convolution neural network. The experiments were performed with the hardware platform of the "Mechatronics" stand (SPA "Android techniques", Russia).

Keywords Convolutional neural network · Pattern recognition · Training with noise

1 Introduction

For invariant pattern recognition there are various methods: potential function, Bayesian networks, Markov networks, artificial neural networks, different types of associative memory and others. Problem of invariant images recognition is still not solved in general [3, 4].

Nowadays, the subclass of artificial neural networks (ANN) also known as convolution neural network (CNN) dominates in image recognition [1]. CNN

R.M. Nemkov (✉) · O.S. Mezentseva · D. Mezentsev
Department of Information Systems & Technologies, North-Caucasus Federal University, 2, Kulakov Prospect, Stavropol, Russian Federation
e-mail: nemkov.roman@yandex.ru

O.S. Mezentseva
e-mail: omezentceva@ncfu.ru

D. Mezentsev
e-mail: terrakoto@me.com

A. Abraham et al. (eds.), *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Advances in Intelligent Systems and Computing 451, DOI 10.1007/978-3-319-33816-3_2

shows results more better than conventional ANN at 10–15 % [2]. However, training of ANN is an incorrect inverse problem [3]. Incorrect problem means that even large data set can't give us full information content about current task. Therefore in the synthesis of ANN mathematical model parameters key part belongs to training data. Making representative training data is one of the most difficult problems in machine learning [3].

The complex results of theoretical and experimental researches of a new type of CNN with changing receptive fields (RP) were presented in the article. With a help of such networks we can create distortions in relation to the current pattern (through the changing of pattern perception) by changing the internal parameters of the hierarchical model of ANN (receptive fields), thereby we obtain new patterns and expanding the training set [6, 7, 9].

2 Mathematical Modeling of CNN with Extended Training Set

Well-known generic synthesis parameters algorithm for CNN includes the following steps:

1. Forward propagation implementation, i.e. calculating signals propagation from the input to the output layer. For convolutional layer neurons values were obtained by the formula (1)

$$y_{m,n} = C^{i}_{m,n} = \varphi(p) = \varphi(b + \sum_{q \in Q_i} \sum_{k=0}^{K_C-1} \sum_{l=0}^{K_C-1} X^{q}_{m+k,n+l} * Z^{q}_{k,l}), \tag{1}$$

   where $C^{i}_{m,n}$—neuron output located at the ith card of C-layer in m, n, $\varphi(\cdot) = A * tanh(B * p)$—position where A = 1.7159, B = 2/3, p—weighted sum, b—bias, Q_i—set of indices cards of previous layer associated with the card C^i, K_C—size of square RP for the neuron $C^{i}_{m,n}$, $Z^{q}_{k,l}$—qth part of the adjustable features, which is responsible for interaction with qth map of previous layer.
2. Back propagation implementation, i.e. calculating signals propagation from the output layer to the input. For convolutional layer (2, 3)

$$\delta^{\lambda}_{m,n} = \sum_{i \in D} \delta^{\lambda+1}_{i} * w^{\lambda+1}_{i}[m,n] * \varphi'(p^{\lambda}_{m,n}), \tag{2}$$

   where D—is the set of neurons on subsequent map ($\lambda + 1$ layer) associated with neuron n, m, $w^{\lambda+1}_{i}[m,n]$—$u^{\lambda+1}_{i}$ index of the S-layer card, which is connected with the card of C-layer, $\varphi^{'}(\cdot)$ is a derivative of $\varphi(\cdot)$, $\delta^{i}_{m,n}$—a residual was gathered for a neuron with coordinates m, n within the map of layer λ.

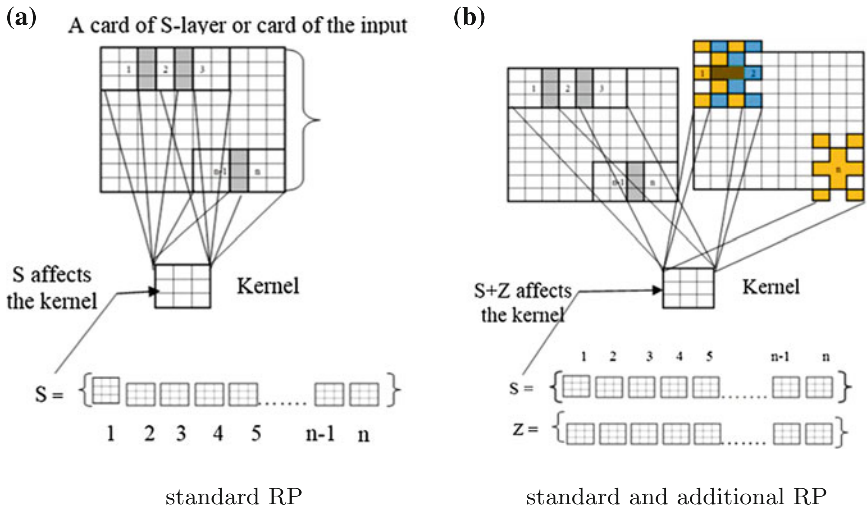

Fig. 1 **a** Standard RP, **b** standard and additional RP

$$\frac{\partial E}{\partial (Z^{\lambda}_{k,l})^{q}} = \sum_{m=0}^{SizeC} \sum_{n=0}^{SizeC} \delta^{\lambda}_{m,n} * y^{\lambda-1}_{m+k,n+l}, \quad (3)$$

where q—the part of the kernel of configurable features for which we receive the components of the gradient, SizeC—the size of the C-layer card.

The same pattern can be perceived by the network in different ways if for some layers of CNN we change set of RP. Due to this, we can extend the training set. The algorithm, which changes RP of neurons lying on any combination of S-layers before feeding pattern, was developed.

It is known, that the canonical form of RP in CNN is a square [11]. We offer to use a template for obtaining a non-standard RP, which elements are indices indicating their neighbors within two discrete steps from them on matrix of pixels. If you change all RP which lying on a map, a lot of information will impact on the configurable parameters that will lead to obtaining a better invariant (Fig. 1).

The algorithm which changes RP neurons lying on any combination of C-layers before feeding pattern is offered.

Each C-layer can have several types of RP: $SetRP_i \in \{RP_1, RP_2, \ldots, RP_n\}$, where $SetRP_i$—many types of RP for the ith C-layer, RP_1—a square RP, n_i—the quantity of RP types for the ith C-layer.

We offer to use two algorithms in order to assign neurons RP for ith C-layer. With the help of algorithm $Alg_1(RP_t \in SetRP_i; C_i)$ to all neurons of layer we assign C_i to RP with index RP_t, t = 1…T—measure number, $T = n_i$. With the help of algorithm $Alg_2(RP \in SetRP_i; C_i)$ to each neuron of layer we assign C_i random field RP of $SetRP_i$. These algorithms can ensure invariance to affine transformations.

Invariance to scale, texture background, a position of an object, luminance levels is provided with the original training sample.

Marking algorithms of a layer can be used by marking algorithms of all C-layers because of any C-layers may have their RP. Here suggests two algorithms:

$$Strategy_1(Alg_1(\cdot), Q_j) = \begin{cases} Alg_1(RP_1; C_i), & if\ x_i = 0 \\ Alg_1(RP_t; C_i), & if\ x_i = 1 \end{cases}, \quad (4)$$

$$Strategy_2(Alg_1(\cdot), Alg_2(\cdot), Q_j) = \begin{cases} Alg_1(RP_1; C_i), & if\ x_i = 0 \\ Alg_2(RP \in SetRP_i; C_i), & if\ x_i = 1 \end{cases}, \quad (5)$$

where i = 1 is the number of C-layers, $x_i \in Q_j, 1 \le t \le T = n_i, Q_j$—specific combination of C-layers, which has or hasn't non-standard shape RP.

The method of synthesis of CNN mathematical model parameters with extended training set based on the proposed algorithm changes of a shape of RP for neurons on an arbitrary combination of C-layers is proposed. This method adapts the back propagation algorithm to the non-standard form of RP.

It includes the following steps:

1. Change RP of neurons at the required combination of C-layers, using the algorithm of RP forms changes (Fig. 2), in the process of training before feeding a next pattern to the input of CNN.
2. Obtain neurons output values of C-layer during forward propagation according to 6 instead of 1.

$$y_{m,n} = \varphi(b + \sum_{q \in Q_i} \sum_{k=0}^{K_C-1} \sum_{l=0}^{K_C-1} Z_{k,l}^q X_{m+k+F_i(RP_{m,n};k;l),n+l+F_j(RP_{m,n};k;l)}^q), \quad (6)$$

where $F_i(RP_{m,n}; k; l), F_j(RP_{m,n}; k; l)$—the functions returns a shifts of row and column of RP template for neuron m, n at position k, l in this template. $index_{k,l}$ is an element of a template $RP_{m,n}$ at the position k, l, $index_{k,l} = 0...24$. These functions are defined as:

$$\begin{aligned} F_i(\cdot) &= \begin{cases} 0; & index_{k,l} \in \{0, 4, 5, 16, 17\} \\ 1; & index_{k,l} \in \{6, 7, 8, 18, 19\} \\ 2; & index_{k,l} \in \{20, 21, 22, 23, 24\} \\ -1; & index_{k,l} \in \{1, 2, 3, 14, 15\} \\ -2; & index_{k,l} \in \{9, 10, 11, 12, 13\} \end{cases} \\ F_j(\cdot) &= \begin{cases} 0; & index_{k,l} \in \{0, 2, 7, 11, 22\} \\ 1; & index_{k,l} \in \{3, 5, 8, 12, 23\} \\ 2; & index_{k,l} \in \{13, 15, 17, 19, 24\}, \\ -1; & index_{k,l} \in \{1, 4, 6, 10, 21\} \\ -2; & index_{k,l} \in \{9, 14, 16, 18, 20\} \end{cases} \end{aligned} \quad (7)$$

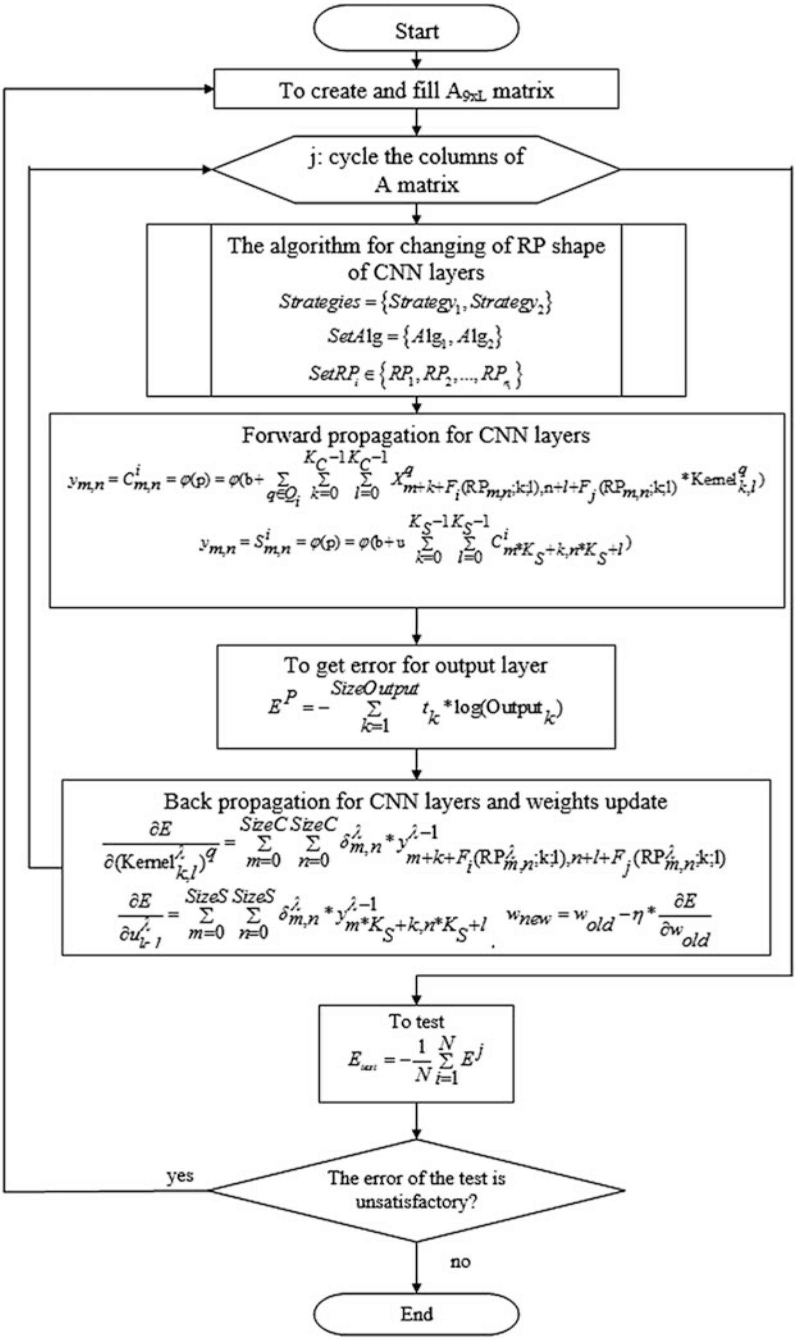

Fig. 2 The generalized synthesis features algorithm of CNN mathematical model with extended training set generated by the change of its RP

3. Obtain a local gradient for C-layer according to 8 instead of 3 in the process of back propagation.

$$\frac{\partial E}{\partial (Z_{k,l}^{\lambda})^{q}} = \sum_{m=0}^{SizeC} \sum_{n=0}^{SizeC} \delta_{m,n}^{\lambda} * y_{m+k+F_i(RP_{m,n}^{\lambda};k;l),n+l+F_j(RP_{m,n}^{\lambda}k;l)}^{\lambda-1}, \tag{8}$$

where $C_{m,n}^{i}$—output of the neuron located on the ith card C-layer in position m, n, $\varphi(\cdot) = A * tanh(B * p)$ when A = 1.7159, B = 2/3, p—weighted sum, b—bias, Q_i—set of cards indices of the previous layer associated with the C^i card, K_C—size of square RP for $C_{m,n}^{i}$ neuron, $X_{m+k,n+l}^{q}$—input to the $C_{m,n}^{i}$ neuron, $Z_{k,l}^{q}$—qth part of the adjustable parameters, which is responsible for interaction with the qth card of the previous layer.

Features synthesis algorithm of CNN mathematical model with extended training set (Fig. 2).

Numerical method for reduction of the extended training set generated by the change of CNN receptive fields was developed [9]. With increasing degree of network training more and more of patterns have weak contagion on the network weights adjustments. So to speed up the training without loss of quality, statistical information about the correct recognition of distorted patterns accumulated in the previous period is used. On the basis of this information, you can skip the part forward propagations, resulting in a reduction in training time without loss of quality recognition.

Matrix A reduction is carried at the beginning of the next period basis on convolution of comparing patterns list to be removed. At the end of period the list is created. During the period statistical information about patterns recognition is accumulated. For this purpose two matrices X_{N_0X7} and Y_{N_0X7} are used, where $x_{i,j}, y_{i,j}$—natural numbers. If the pattern distortions is correctly recognized, then the value of the matrix X is increased by 1in the row i (which is equal to the index of the donor-pattern, on which distortions are superimposed) and in column j (which is equal to the index of a particular combination of C-layers with non-standard RP). If it is recognized incorrectly, the value increases in the Y matrix at the same positions.

Realization algorithm of this method is shown in Fig. 3.

3 Experimental Research of CNN Generalization Capability with Changing Receptive Fields

Experiments for evaluating generalization capability of the proposed CNN mathematical model and quality of objects different classes recognition were carried out. Structure diagram of design techniques objects recognition software from the mobile robot camera is shown.

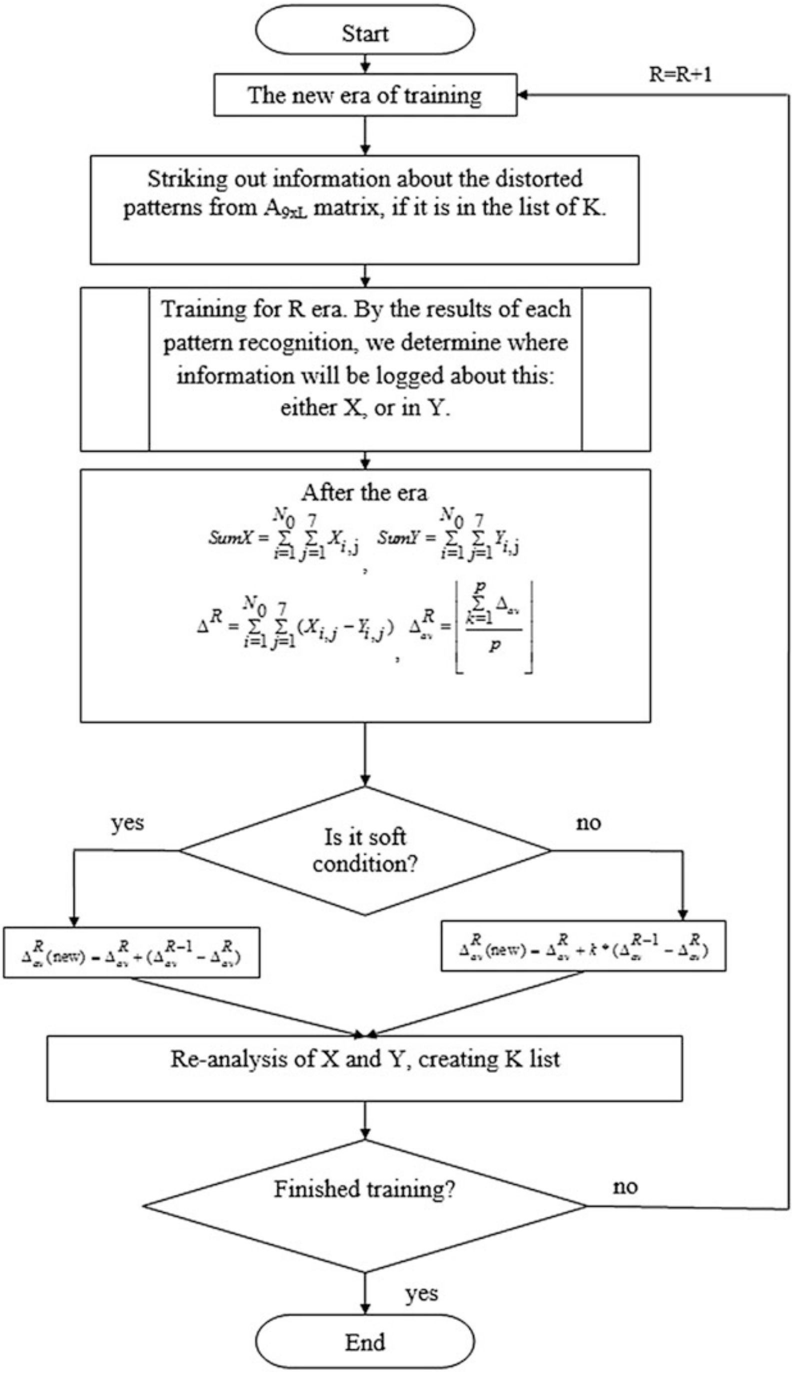

Fig. 3 The generalized extended training set reduction algorithm

Table 1 The results of computational experiments to assess the generalization capability of the learning algorithm of CNN with the help of proposed method and the impact of the reduction in the learning process

Generalization error with using of the proposed method of synthesis features mathematical model of CNN, %	Generalization error without using of the proposed method of synthesis features mathematical model of CNN, %	Error generalizations of the best analogue, %	Training time without reduction, h	Training time with reduction, h
MNIST				
0.6	2.8	0.23	46.15	31.2
Small NORB				
4.3	8.4	6.6	15.6	9.8
Training set for “Mechatronics” stand				
0.3	1.8	–	17.3	10.1

Experiments for evaluating generalization capability of the proposed CNN mathematical model with extended training set were carried out using three training sets: MNIST (handwritten digits) [5], Small NORB (five classes of objects) [8], the training set is created for recognition of 10 classes of objects from the “Mechatronics” stand (vision system: AXIS M1054 camera, 320 × 240 pixels (HDTV), horizontal viewing angle 80, the allowable exposure range from 0.9 to 105 lx, frame rate 30 fps (H. 264/M-JPEG)) and providing scale invariance, object position on the plane, the texture backgrounds and the illumination.

In Table 1 the results of evaluating experiments are presented. Training of CNN with a help of mathematical model improves average generalizing capability of the network to 2.5 % compared with the results without it and to 1.25 % in comparison with the best results achieved analogues as follows from the table [10]. The numerical algorithm of extended training set reduction allows to reduce training

Table 2 Percent of correct recognition for the experiments at “Mechatronics”—stand when using the proposed method of training of CNN and without it

The object/the distance from camera	50 cm (%)	1 m (%)	1.5 m (%)	2 m (%)
Bike	91 (85)	87 (79)	88 (74)	75 (69)
Car	89 (81)	85 (77)	82 (71)	68 (67)
Flashlight	94 (82)	90 (71)	92 (65)	72 (59)
Frog	95 (79)	91 (78)	89 (60)	69 (59)
Plane	91 (84)	90 (72)	87 (67)	74 (65)
Player	90 (79)	89 (77)	90 (63)	75 (58)
Pumpkin	85 (81)	83 (71)	80 (68)	63 (63)
Soldier	96 (85)	92 (76)	90 (64)	73 (60)
Stapler	92 (78)	89 (64)	88 (62)	71 (59)
Empty class	98	98	98	96

time by an average of 9 h, which is 37 % of the initial training time without reduction [9].

There was developed software which allows to recognize objects with the camera of mobile robot for objects recognition from “Mechatronics” stand.

The experiments for objects recognition on “Mechatronics” stand for different distances from the camera (AXIS M1054) are shown in Table 2.

According to results we see that the recognition accuracy for objects up to 96 % when using the method of synthesis of mathematical model features CNN with extended training set. When distance to the camera is increase the recognition accuracy is decrease due to inferior quality of input image. In this article presented the results of applying of convolutional neural network with changed receptive fields towards image recognition tasks.

References

1. Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009)
2. Bengio, Y., LeCun, Y.: Scaling learning algorithms towards AI. In: Large Scale Kernel Machines, MIT Press (2007)
3. Bishop, C.M.: Neural Networks for Pattern Recognition, pp. 482. Great Clarendon Street, USA (1995)
4. Cruz, V., Cristobal, G., Michaux, T., Barquin, S.: Invariant image recognition using a multi-network neural model. In: Proceedings International Joint Conference Neural Networks, vol. II, pp. 17–21 (1989). Electronic Neurocomputers
5. Modified National Institute of Standards and Technology (MNIST) (30.01.2016). http://yann.lecun.com/exdb/mnist/
6. Nemkov, R.: Dynamical change of the perceiving properties of neural networks as training with noise and its impact on pattern recognition. In: Young Scientists—International Workshop on Trends in Information Processing (YSIP) (2014) (30.01.2016). http://ceur-ws.org/Vol-1145/paper4.pdf/
7. Nemkov., R., Mezentseva, O.: The use of convolutional neural networks with non-specific receptive fields. In: The 4th International Scientific Conference: Applied Natural Sciences, pp. 284–289. Novy Smokovec (2013)
8. NYU Object Recognition Benchmark (NORB) (30.01.2016). www.cs.nyu.edu/ylclab/data/norb-v1.0/
9. Nemkov, R.M., Mezentseva, O.S.: Dynamical change of the perceiving properties of convolutional neural networks and its impact on generalization. Neurocomputers: Development and Application, vol. 2, pp. 12–18 (2015)
10. Nemkov, R.M.: Synthesis method of mathematical model parameters of the convolutional neural network with extended training set (30.01.2016). http://www.science-education.ru/125-19867/
11. Simard, P., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Proceeding of the Seventh International Conference on Document Analysis and Recognition (ICDAR03), vol. 2, pp. 958–964 (2003)

http://www.springer.com/978-3-319-33815-6

Proceedings of the First International Scientific
Conference "Intelligent Information Technologies for
Industry" (IITI'16)
Volume 2
Abraham, A.; Kovalev, S.; Tarassov, V.; Snášel, V. (Eds.)
2016, XV, 504 p. 196 illus., 120 illus. in color., Softcover
ISBN: 978-3-319-33815-6