

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



New Principles in Algorithm Design for Problems of Face Recognition

Vitaliy Tayanov

*Vyacheslav Chornovil State Institute of Modern Technologies and Management of Lviv
Ukraine*

1. Introduction

This chapter is devoted to two main problems in pattern recognition. First problem concerns the methodology of classification quality and stability estimation that is also known as classification reliability estimation. We consider general problem statement in classification algorithms design, classification reliability estimation and the modern methods solution of the problem. On the other hand we propose our methods for solution of such kind of a problem. In general this could be made by using of different kind of indicators of classification (classifier) quality and stability. All this should summarize everything made before with our latest results of solution of the problem. Second part of the chapter is devoted to new approach that gives the possibility to solve the problem of classifier design that is not sensitive to the learning set but belongs to some kind of learning algorithms.

Let us consider recognition process, using the next main algorithms, as some parts of the some complicated recognition system: algorithm for feature generating, algorithms for feature selection and classification algorithms realizing procedure of decision making. It is really important to have good algorithms for feature generating and selection. Methods for feature generating and selection are developed a lot for many objects to be recognized. For facial images the most popular algorithms use 3D graph models, morphological models, the selection some geometrical features using special nets, etc. From other hand very popular algorithms for feature generating and selection are those which use Principle Component Analysis (PCA) or Independent Component Analysis (ICA). PCA and ICA are good enough for a number of practical cases. However there are a lot of different deficiencies in classifier building. For classifiers using learning the most essential gap is that all such classifiers can work pretty well on learning stage and really bad for some test cases. Also they can work well enough if classes are linearly separated e.g. Support Vector Machines (SVM) in linear case. For non-linear case they have a number of disadvantages. That is why it is important to develop some approaches for algorithms building that are not sensitive for the kind of sample or complexity in classification task.

All classification algorithms built for the present could be divided almost into 5 groups: algorithms built on statistical principles, algorithms, built on the basis of potential functions and non-parametrical estimation, algorithms, using similarity functions (all kinds of metrical classifiers like 1NN and kNN classifiers) algorithms, using logical principles like decision lists, trees, etc, hierarchical and combined algorithms. A lot of recognition systems use a number of algorithms or algorithm compositions. For their optimization and tuning

one uses special algorithms like boosting. These compositions can be linear or non-linear as well.

To build any effective classifier we need to use some algorithms that allow us to measure the reliability of classification. Having such algorithms we can find estimates of optimal values of classifier parameters. Accuracy of these estimates allows us to build reliable and effective classification algorithms. They perform the role of indicators of measurement of different characteristics and parameters of the classifier.

We propose the new approach for object classification that is independent of the learning set but belongs to some kind of learning algorithms. Methods, using for new classification approach concerns the field of results combination for the classification algorithms design. One of the most progressive directions in this area assumes using of the consensus in recognition results, produced by different classifiers.

Idea of usage of consensus approach is the following. One divides all objects to be recognized into three groups: objects, located near separation hyperplane (ambiguity objects), objects, located deeply inside the second class and belong to the first one (misclassified objects) and objects that are recognized correctly with large enough index of reliability (easy objects). The group of ambiguity objects is the largest one that can cause errors during recognition due to their instability. Because of that it is extremely important to detect such kind of objects. Next step will be detecting of the true class for every of ambiguity objects. For this it is planned to use apparatus of cellular automata and Markov models. It is important to mark that such an approach allows us to reduce the effect of overestimation for different recognition tasks. This is one of the most principle reasons for using such kind of algorithms.

The practical application of consensus approach for the task of face recognition could be realized by the following way. If we use one of the following classifiers e.g. 1NN, kNN, classifier, built on the basis of potential functions, SVM, etc. we can have some fiducial interval of the most likely candidates. Fiducial interval in general is the list of the candidates that are the most similar to the target object (person). If we use decision making support system controlled by operator result of the system work could be one or two candidates, given to the operator for the final decision or expertise. If we use an autonomous system the final decision should be made by the system that has to select one of the most likely candidates using some special verification algorithms that analyse the dynamics of behaviour of the object in the fiducial interval under the verifying external conditions and parameters of the algorithm.

2. Estimations building in pattern recognition

2.1 Some important tasks of machine learning

The modern theory of machine learning has two vital problems: to obtain precise upper bound estimates of the overtraining (overfitting) and ways of it's overcoming. Now the most precise familiar estimates are still very overrated. So the problem is open for now. It is experimentally determined the main reasons of the overestimation. By the influence reducing they are as follows [Vorontsov, 2004]:

1. The neglect of the stratification effect or the effect of localization of the algorithms composition. The problem is conditioned by the fact that really works not all the composition but only part of it subject to the task. The overestimation coefficient is from several tens to hundreds of thousands;

2. The neglect of the algorithms similarity. The overestimation coefficient for this factor is from several hundreds to tens of thousands. This factor is always essential and less dependent from the task than first one;
3. The exponential approximation of the distribution tail area. In this case the overestimation coefficient can be several tens;
4. The upper bound estimation of the variety profile has been presented by the one scalar variety coefficient. The overestimation coefficient is often can be taken as one but sometimes it can be several tens.

The reason of overtraining effect has been conditioned by the usage of an algorithm with minimal number of errors on the training set. This means that we realize the one-sided algorithms tuning. The more algorithms are going to be used the more overtraining will be. It is true for the algorithms, taken from the distribution randomly and independently. In case of algorithm dependence (as rule in reality they are dependent) it is suggested that the overtraining will be reduced. The overtraining can be in situation if we use only one algorithm from composition of two algorithms. Stratification of the algorithms by the error number and their similarity increasing reduces the overtraining probability.

Let us consider a duplet algorithm-set. Every algorithm can cover a definite number of the objects from the training set. If one uses internal criteria [Kapustii et al., 2007; Kapustii et al., 2008] (for example in case of metrical classifiers) there is the possibility to estimate the stability of such coverage. Also we can reduce the number of covered objects according to the stability level. To cover more objects we need more algorithms. These algorithms should be similar and have different error rate.

There is also interesting task of redundant information decrease. For this task it is important to find the average class size guaranteeing the minimal error rate. The reason in such procedure conditioned also by the class size decrease for the objects interfering of the recognition on the training phase.

The estimation of the training set reduction gives the possibility to define the data structure (the relationship between etalon objects and objects that are the spikes or non-informative ones). Also the less class size the less time needed for the decision making procedure. But the most important of such approach consists in possibility to learn precisely and to understand much deeper the algorithms overtraining phenomenon.

In this paper we are going to consider the metrical classifiers. Among all metrical classifiers the most applied and simple are the k NN classifiers. These classifiers have been used to build practical target recognition systems in different areas of human's activity and the results of such classification can be easily interpreted. One of the most appropriate applications of metrical classifiers (or classifiers using the distance function) concerns the biometrical recognition systems and face recognition systems as well.

2.2 Probabilistic approach to parametrical optimization of the k NN classifiers

The most advanced methods for optimization composition algorithm, informative training set selection and feature selection are bagging, boosting and random space method (RSM). These methods try to use the information containing in the learning sample as much as they can. Let us consider the metrical classifier optimization in feature space, using different metrics. The most general presentation of the measure between feature vectors \mathbf{x} and \mathbf{y} has been realized through Manhattan measure as the simple linear measure with weighted coefficients a_i [Moon & Stirling, 2000]:

$$d(x, y) = \sum_{i=1}^n a_i |x_i - y_i|, \quad (1)$$

where $d(x, y) = \sum_{i=1}^n a_i |x_i - y_i|$ is the arbitrary measure between vectors \mathbf{x} and \mathbf{y} .

Minkovski measure as the most generalized measure in pattern recognition theory can be presented in form of

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^n a_i |x_i - y_i| \right)^{\frac{1}{p}} = C(p) \sum_{i=1}^n a_i |x_i - y_i|, \quad (2)$$

where parametrical multiplier $C(p)$ have been presented in form of

$$C(p) = \left(\sum_{i=1}^n a_i |x_i - y_i| \right)^{\frac{1-p}{p}}; a_i = |x_i - y_i|^{p-1}; p > 0 \quad (3)$$

One can make the following conclusions. An arbitrary measure is the filter in feature space. It determines the weights on features. The weight must be proportional to the increase of one of indexes when it has been added to general feature set used for class discrimination procedure. Such indexes are: correct recognition probability, average class size, divergence between classes, Fisher discriminant [Bishop, 2006]. One can use another indexes, but the way of their usage should be similar. If one of the features does not provide the index increase (or worsen it) the value of such feature weight should be taken as zero. So by force of supplementary decrease of feature number one can accelerate the recognition process retaining the qualitative characteristics. The feature optimization problem and measure selection has been solved uniquely. This procedure has been realized using weighted features and linear measure with weighted coefficients. Feature selection task at the same time has been solved partially. First the feature subset from general set is determined. Such set has been determined by some algorithm (for example by the number of orthogonal transforms). Such algorithm should satisfy the definite conditions like follows: class entropy minimization or divergence maximization between different classes. These conditions have been provided by the Principle Component Analysis [Moon & Stirling, 2000]. The last parameter using in the model is the decision function or decision rule. Number of decision functions can be divided into functions working in feature space and the functions based on distance calculation. For example the Bayes classifier, linear Fisher discriminant, support vector machine etc. work in feature space. The decision making procedure is rather complex in multidimensional feature space when one uses such decision rules. Such circumstance is especially harmful for continuous recognition process with pattern series that have been recognized. Thus realizing the recognition system with large databases in practice one uses classifiers based on distance function. The simplest classifier is 1NN. But this classifier has been characterized by the smallest probability indexes. Therefore one should use k NN one. So the task consists in selection of k value that is optimal for decision making procedure in bounds of fiducial interval. This interval corresponds to the list of possible candidates. Unlike the classical approach k value has upper bound by class size. In classical approach the nearest neighbor value should be taken rather large, approximating Bayes classifier.

Let us consider RS with training. The calculation and analysis of the parameters of such systems is carried out on the basis of learning set. Let there exists the feature distribution in linear multidimensional space or unidimensional distribution of distances. We are going to analyse the type of such distribution. The recognition error probability for $\mu = 0$ could be presented as $\int_{|x| \geq \theta} p(x) dx$, where θ is the threshold. According to the Chebyshev inequality

[Moon & Stirling, 2000] we obtain $\int_{|x| \geq \theta} p(x) dx \leq \frac{\sigma^2}{\theta^2}$.

Let us consider the case of mean and variance equality of $p(x)$ distribution. The upper bound for single mode distributions with mode $\mu = 0$ with help of Gauss inequality [Weinstein, 2011] is:

$$P(|x - \mu| \geq \lambda \tau) \leq \frac{4}{9\lambda^2} \quad (4)$$

where $\tau^2 \equiv \sigma^2 + (\mu - \mu_0)^2$.

Let $\mu = \mu_0 = 0$ and $\tau \equiv \sigma$. Then the threshold θ is $\theta = \lambda \tau = \lambda \sigma$ and $\lambda = \frac{\theta}{\sigma}$. Thus the Gauss inequality for the threshold θ could be presented in form of:

$$\int_{|x| \geq \theta} p(x) dx \leq \frac{4\sigma^2}{9\theta^2}. \quad (5)$$

As seen from (5), the Gauss upper bound estimate for the single-mode distribution is better in 2.25 times then for the arbitrary distribution. So the influence of the distribution type on the error probability is significant. The normal distribution has equal values of mode, mean and median. Also this distribution is the most popular in practice. On the other hand the normal distribution has been characterized by the maximum entropy value for the equal values of variance. This means that we obtain the minimal value of classification error probability for the normally distributed classes. For the algorithm optimization one should realize the following steps:

- to calculate the distance vector between objects for the given metric;
- to carry out the non-parametrical estimation of the distance distribution in this vector by the Parzen window method or by the support vector machines;
- to estimate the mean and variance of the distribution;
- on the basis of estimated values to carry out the standardization of the distribution ($\mu = 0, \sigma = 1$);
- to build the distributions both for the theoretical case and estimated one by the non-parametrical methods;
- to calculate the mean square deviation between the distributions;
- to find out the parameter space, when deviation between the distributions less then given δ level.

2.2.1 Probability estimation for some types of probability density functions

Let us consider some probability density functions (pdfs) that have a certain type of the form (presence of the extremum, right or left symmetry). If pdf have not one of such types of

structure one can use the non-parametrical estimation. As the result of such estimation we get the uninterrupted curve describing pdf. This function can be differentiated and integrated by the definition. Because the Gaussians have been characterized by the minimal error of the classification for the given threshold θ and does not exceed $\frac{4\sigma^2}{9\theta^2}$ (see eq.5) for the unimodal and symmetric pdf or pdf with right asymmetry, the double-sided inequality for the given value of recognition error can be presented in form of :

$$0.5(1 - \operatorname{erf}(\frac{\theta}{\sigma})) \leq \varepsilon \leq \frac{4\sigma^2}{9\theta^2} \quad (6)$$

where $\mu = 0$.

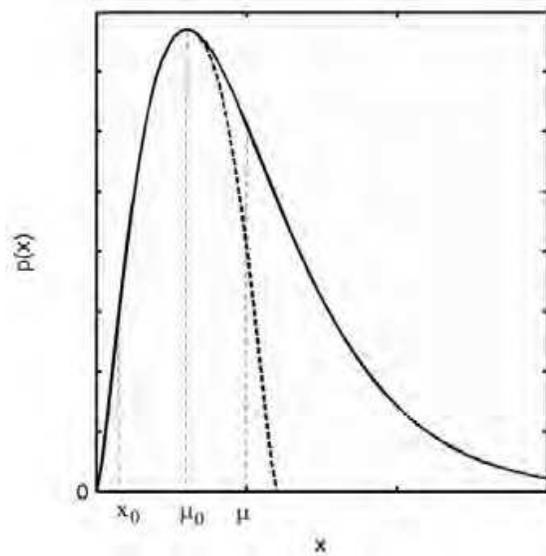


Fig. 1. Right asymmetry of pdf

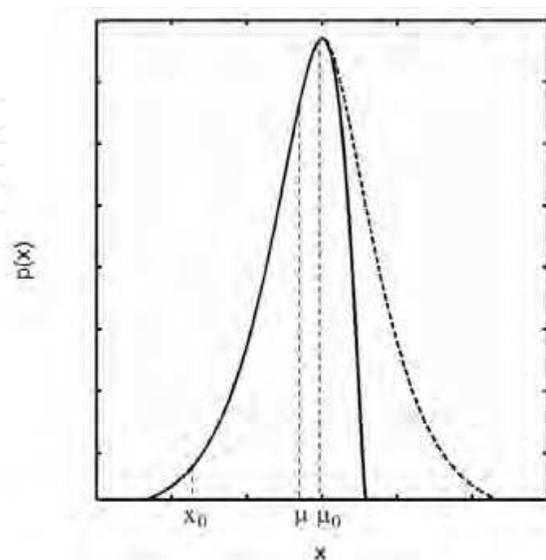


Fig. 2. Left asymmetry of pdf

Let us analyse the form of potentially generated pdfs of distances between objects. All of the distributions will have extremum. This will be conditioned by following facts. All of the pdfs have been determined on the interval $[0, \infty)$ and the density near zero and for the large distances is not high because these values are mostly unlikely. The right asymmetry is much more likely because pdf of distances is limited by zero and from the other side it has no strictly determined limitations.

Let's consider a widespread problem of classification in the conditions of two classes. We will denote the size of classes as s_1 and s_2 correspondingly. Then if the probability of replacement of object of a class having size s_1 within a fiducial interval is equal ε_1 the probability of no replacement of objects from the same class by objects from a class s_2 in this interval is equal to $(1 - \varepsilon_1)^{s_2}$ under the condition of independence of objects [Kapustii et al, 2008; Kyrgyzov, 2008; Tayanov & Lutsyk, 2009]. For other class at corresponding changes this probability is equal to $(1 - \varepsilon_2)^{s_1}$. If now one selects some virtual class and admits that replacement of any object of this class by objects from the mentioned two classes is authentic event it is possible to write down a following equation:

$$\gamma((1 - \varepsilon_1)^{s_2} + (1 - \varepsilon_2)^{s_1}) = 1, \quad (7)$$

where the proportionality multiplier is calculated trivially.

Sometimes there are situations when distances between objects are equal to 0. Thus non-parametrically estimated distribution of one of the classes can have a maximum in a point corresponding to zero distance. Let density of distributions are equal $p_1(0)$ and $p_2(0)$ in a zero point. The estimation of relation between probabilities can be set in a form of $\frac{p_1(0)^{s_2}}{p_2(0)^{s_1}}$ or

$\ln \frac{p_1(0)^{s_2}}{p_2(0)^{s_1}}$. Thus it is necessary to make boundary transition from cumulative density

function (cdf) to pdf as they are connected among themselves by differentiation operation.

The relation $\ln \frac{p_1(0)^{s_2}}{p_2(0)^{s_1}}$ or generally $(\ln \frac{p_2(0)^{s_1}}{p_1(0)^{s_2}})$ can be used for construction of the following classifier

$$\begin{aligned} \ln \frac{p_1(\theta)^{s_2}}{p_2(\theta)^{s_1}} > \gamma_1; & \quad \ln \frac{p_2(\theta)^{s_1}}{p_1(\theta)^{s_2}} > \gamma_2; \\ \text{or} & \\ \ln \frac{p_1(\theta)^{s_2}}{p_2(\theta)^{s_1}} < \gamma_1, & \quad \ln \frac{p_2(\theta)^{s_1}}{p_1(\theta)^{s_2}} < \gamma_2, \end{aligned} \quad (8)$$

where values $\ln \frac{p_1(\theta)^{s_2}}{p_2(\theta)^{s_1}} = 0$ or $\ln \frac{p_2(\theta)^{s_1}}{p_1(\theta)^{s_2}} = 0$ have no influence on classification results and

the decision can be accepted for benefit of any class. In case of non-parametric estimation the probability of such value is almost equal to 0. This approach is especially useful for the recognition tasks with similar objects i.e. objects that are weak separated in the feature space. It should be noted that such type of algorithms have been oriented on the tasks with

high level of class overlapping. Face recognition belongs to the tasks that have sufficiently a lot of objects that could not be separated so easy.

2.3 Combinatorial approach

Let us present the recognition results for k NN classifier in form of binary sequence:

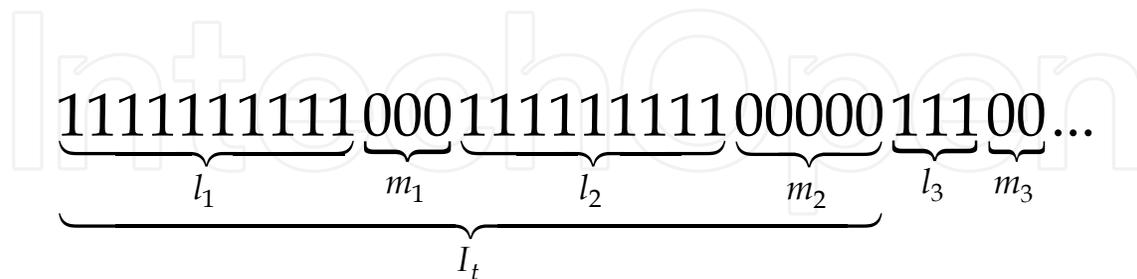


Fig. 1. The recognition results in form of binary sequence for k NN classifier

Using k NN classifier it is important that among k nearest neighbours we have the related positive objects majority or the absolute one. Let us consider the simpler case meaning the related majority. The k NN classifier correct work consists in fact that for k nearest neighbours it has to be executed the condition

$$\left| \bigcup_i \tilde{l}_i \right| > \left| \bigcup_i \tilde{m}_i \right|, \quad i = 1, 2, 3, \dots, \quad (9)$$

where \tilde{l}_i, \tilde{m}_i are the groups that appear after class size decrease. Under the group one understands the homogeneous sequence of elements. In such sequence (see Fig.1) there exist patterns of all classes. In general case there is no direct conformity between the group number and the class number although.

Let us consider the case of non-pair k value in k NN classifier only. This means that we have the case of synonymous classification. Such univocacy could disappear in case of pair k value and votes equality for different classes.

Let us estimate the effect of class size reduction in case of k NN classifier. Note that reduced class sizes are equal to each other and equal s^* . Let us consider the k NN classifier correct

work condition: $ENT\left(\frac{k}{2}\right) + 1 \leq s^*$. In contradistinction to 1NN classifier there is no such an

importance of the first nearest patterns of the true class. Thus all such sequences one could denote as l_i . Let us determine the probabilities that it will be selected s^* patterns from the true class by the combinatorial approach. These probabilities have fiducial sense. This means that for the given part of positive objects there will be no selections among the patterns of the false classes by the correspondent combinatorial way. The multiplication of pointed two probabilities determines the probability of k NN classifier correct work. Let assign q_j as the recognition error probability for the corresponding m_i groups:

$$\begin{aligned}
 q_1 &= P\left(\inf\left(\left|\bigcup_i m_i\right|\right) \geq ENT\left(\frac{k}{2}\right) + 1\right); \\
 q_2 &= P\left(\inf\left(\left|\bigcup_i m_i\right|\right) + |m_{i+1}| \geq ENT\left(\frac{k}{2}\right) + 1\right); \\
 q_3 &= P\left(\inf\left(\left|\bigcup_i m_i\right|\right) + |m_{i+1}| + |m_{i+2}| \geq \right. \\
 &\quad \left. \geq ENT\left(\frac{k}{2}\right) + 1\right); \dots \\
 q_j &= P\left(\inf\left(\left|\bigcup_i m_i\right|\right) + \left|\bigcup_j m_{i+j-1}\right| \geq \right. \\
 &\quad \left. \geq ENT\left(\frac{k}{2}\right) + 1\right); \dots
 \end{aligned} \tag{10}$$

The combinatorial expression for q_j probability could be written in form of:

$$q_j = \frac{\sum_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^j_{\left|\bigcup_{i,j} m_{i+j-1}\right|} C^{s^*-j}_{\left|s-\left|\bigcup_{i,j} m_{i+j-1}\right|\right|}}{C_s^{s^*}}, \left|\bigcup_{i,j} m_{i+j-1}\right| \geq ENT\left(\frac{k}{2}\right) + 1. \tag{11}$$

The fiducial probability for arbitrary true pattern sequence is equal:

$$P_{q_j} = \frac{\sum_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^j_{\left|\bigcup_i l_i\right|} C^{s^*-j}_{\left|s-\left|\bigcup_i l_i\right|\right|}}{C_s^{s^*}}. \tag{12}$$

Thus the correct recognition probability for k NN classifier has been determined by probability (12) and addition to probability (11):

$$\begin{aligned}
 P_j = P_{q_j} (1 - q_j) &= \frac{\sum_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^j_{\left|\bigcup_i l_i\right|} C^{s^*-j}_{\left|s-\left|\bigcup_i l_i\right|\right|}}{C_s^{s^*}} - \\
 &\quad \frac{\left(\sum_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^j_{\left|\bigcup_i l_i\right|} C^{s^*-j}_{\left|s-\left|\bigcup_i l_i\right|\right|}\right) \left(\sum_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^j_{\left|\bigcup_{i,j} m_{i+j-1}\right|} C^{s^*-j}_{\left|s-\left|\bigcup_{i,j} m_{i+j-1}\right|\right|}\right)}{\left(C_s^{s^*}\right)^2}.
 \end{aligned} \tag{13}$$

It is modelled the recognition process with different sequences of patterns of true and false classes for the 1NN and k NN classifiers in case of absolute majority. For modelling the face recognition system has been taken. The class size (training set) has been taken as 18

according to the database that it was made. On the Fig.1 the results of modelling of the training set decrease influence on the recognition results for the 1NN classifier have been presented. On the Fig.2 the similar results for k NN classifier under condition $ENT\left(\frac{k}{2}\right)+1=s^*$ have been presented.

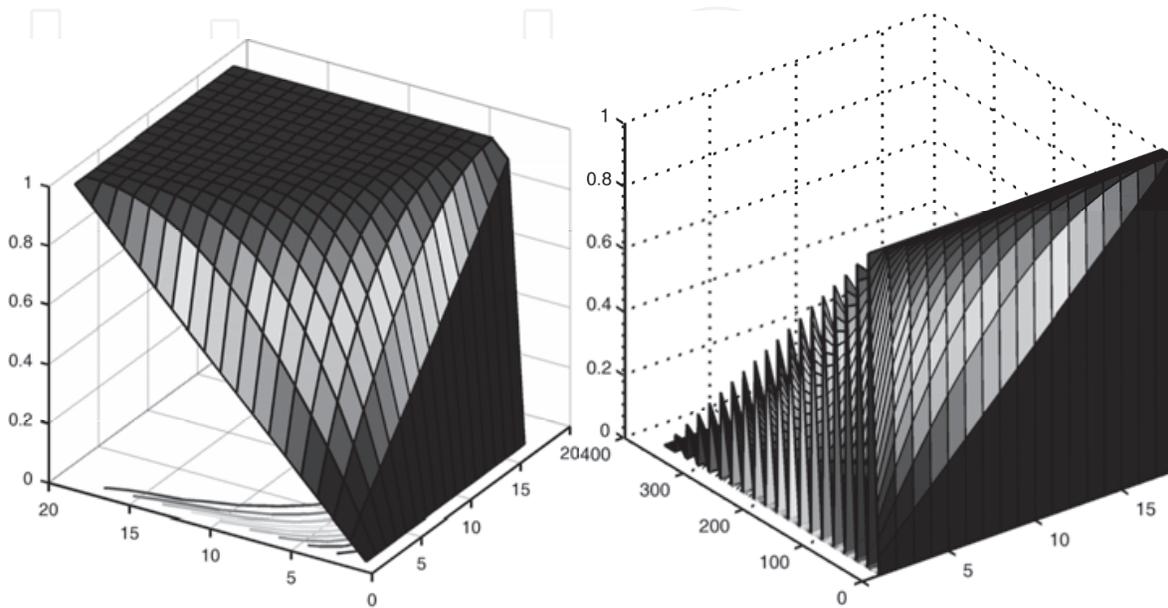


Fig. 2. The probability of correct recognition as function of training set (x axis) and number of true/false objects in the target sequence (y axis) for the 1NN classifier

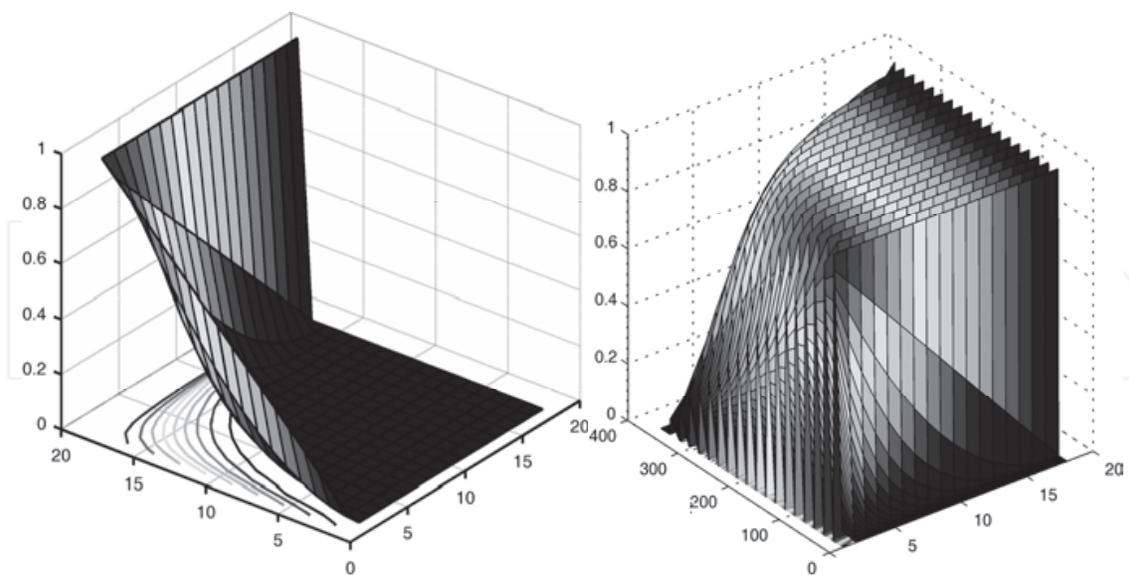


Fig. 3. The probability of correct recognition as function of training set (x axis) and number of true/false objects in the target sequence (y axis) for the k NN classifier

On the Fig.1,2 x axis means the size of the training set and the y axis means the size of the true patterns sequence (left picture) and sequence of both true and false patterns (right

picture). The y axis has been formed by the following way. We organized 2 cycles where we changed the number of true and false patterns. For every combination of these patterns and different class sizes we calculate the probability of correct recognition.

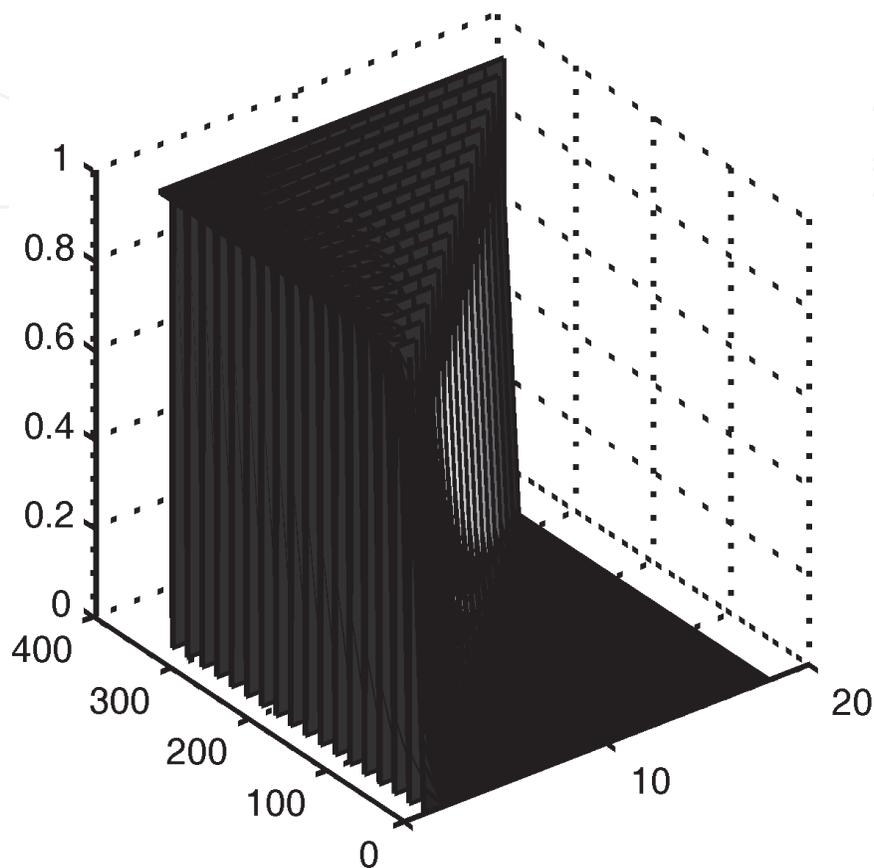


Fig. 4. The probability of correct recognition as function of training set (x axis) and $ENT\left(\frac{k}{2}\right)+1$ value (y axis)

On the Fig.4,5 the results of k NN classifier modelling have been presented. Here it has been satisfied the following condition: $ENT\left(\frac{k}{2}\right)+1 \leq s^*$. On the Fig.5 the fiducial probability as

function of training set size (x axis) and $ENT\left(\frac{k}{2}\right)+1$ value (y axis).

The probability part of proposed approach is based in following idea. Despite of combinatorial approach, where the recognition results were determined precisely, we define only the probability of the initial sequence existence. Due to low probability of arbitrary sequence existing (especially for the large sequences) it has been determined the probability of homogeneous sequences existing of the type $\{0\}$ or $\{1\}$. This probability has been determined on the basis of the last object in given sequence as probability of replacing this object (the object from the true class $\{1\}$ by the others objects of the false classes from the

database. This means that the size of homogeneous sequence has been determined by the most "weak" object in the homogeneous pattern sequence. The probability of existing of the non-homogeneous sequences is inversely proportional to the $2^{|l+m|}$ value, where $|l+m|$ is the sequence size. This procedure could be realized using distribution function (fatigue function) of the distances between the objects. This approach has been developed for metrical classifiers and classifiers on the basis of distance function in [Kapustii et al., 2008; Tayanov & Lutsyk, 2009]. Thus we need to calculate the probability of sequence with true patterns existing that has definite size or for the given probability rate we need to calculate the maximal size of the sequence that satisfies this probability. For the binary sequence the sum of the weights of the lower order bits is always less than the next most significant bit.

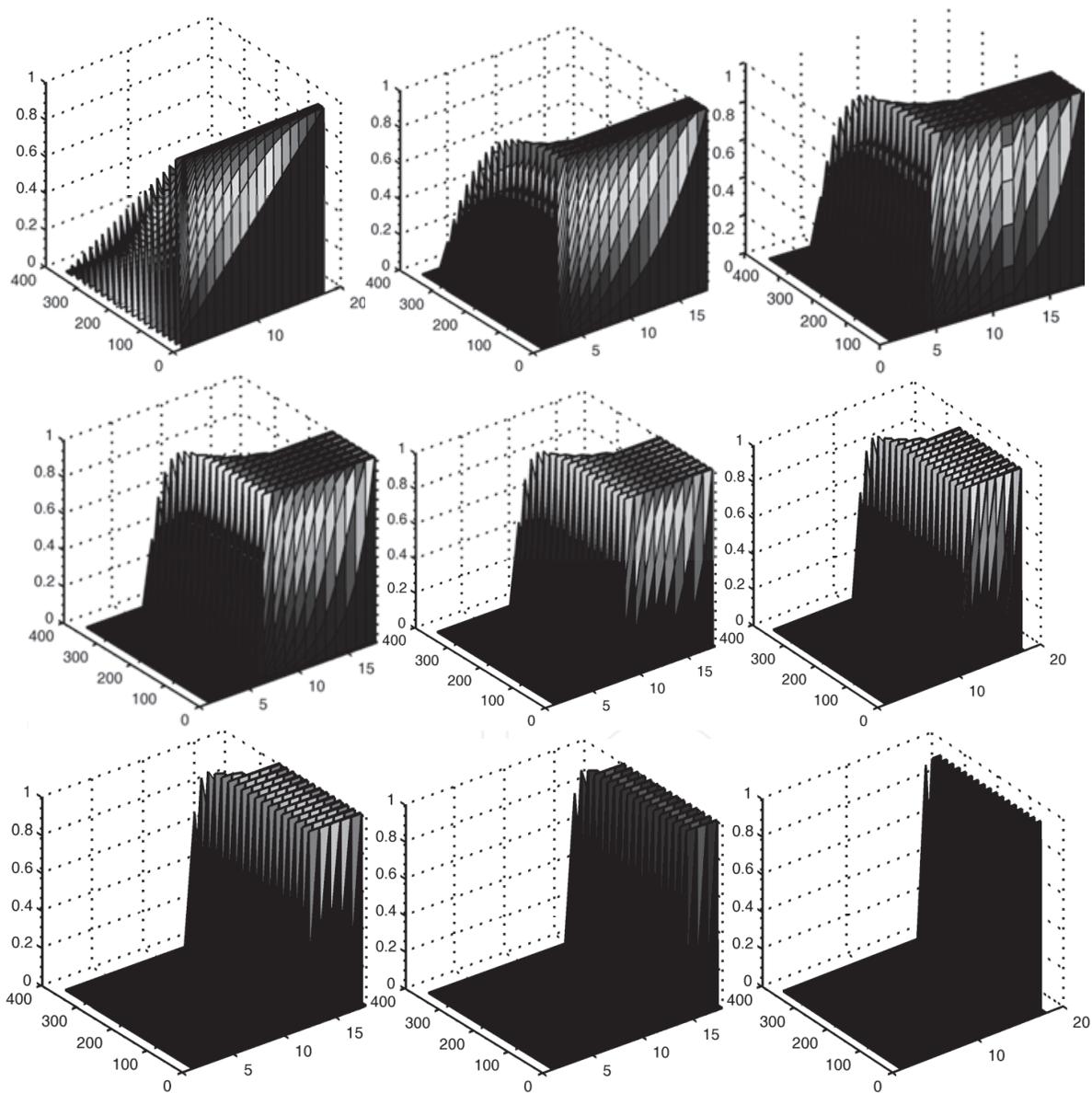


Fig. 5. The probability of correct recognition as function of $ENT\left(\frac{k}{2}\right)+1$ (x axis) and number of true/false objects in the target sequence (y axis)

The difference is equal to 1. This means that arbitrary pattern replacement of the true class in the fiducial interval is equivalent to the alternate replacement of the previous ones. The minimal whole order of the scale of notation that has such peculiarity is equal to 2. Thus we need to calculate the weights of the true patterns position and compare them with binary digit. Such representation of the model allows us to simplify the probability calculation of the patterns replacement from the true sequences by the patterns of false classes. On the other side the arbitrary weights can be expressed through the exponent of number 2 that also simplifies the presentation and calculation of these probabilities. So the probability of the homogeneous sequence of the true patterns existence has been calculated on the basis of distance distribution function and is the function of the algorithm parameters. We should select the sequence of the size that has been provided by the corresponding probability. We after apply the combinatorial approach that allows us to calculate the influence effect of the class size decrease on the recognition probability rate. Thus the probabilistic part of the given approach has been determined by the recognition algorithm parameters. So the integration of both probabilistic and combinatorial parts allows us to define more precise the influence of the effect of the training set reduction.

Let us consider step by step the example of fast computing of replacement of true pattern probability from the sequence where relation between weights of the objects is whole exponent of number 2. Thus for example the weights can be presented by the following way: $w = \{2^9, 2^6, 2^4, 2^3, 2^2, 2^1, 2^0\}$. As known the probability of replacement of true object from the sequence by the false one when it is known that replacement is true event is inversely proportional to the weights of these objects. Let define the probability of replacement of the object having the 2^9 weight comparatively to the object with 2^6 weight. As far as we do not know what object has been replaced the total weight of the fact that there will not be replaced the objects with 2^6 weight and lower is equal: $w = \{2^9, 2^6, 2^4, 2^3, 2^2, 2^1, 2^0\}$. This weight can be expressed through 2^6 weight accurate within 1 by following way: $2^6(1+0.5) = 1.5 * 2^6$. In case of large sequences this one has weak influence on the accuracy. The relation between 2^9 and 2^6 is equal to 8. In case of divisible group of events we obtain the $8\lambda + 1.5\lambda = 1$ equation, where the proportional coefficient λ approximately equal to 0.11. So the probability of non-replacement of the object with 2^9 weight is equal $8 * 0.11 = 0.88$. The object with 2^6 weight has the corresponding probability equal to $1 - 0.88 = 0.12$. Since we know exactly that replacement is the true event and the last object has weight equal to 1 the accuracy correction that equal to 1 makes the appropriate correction of probability calculation.

3. Classification on the basis of division of objects into functional groups

Algorithms of decision making are used in such tasks of pattern recognition as supervised pattern recognition and unsupervised pattern recognition. Clustering tasks belong to unsupervised pattern recognition. They are related to the problems of cluster analysis. Tasks where one provides the operator intervention in the recognition process belong to the learning theory or machine learning theory. The wide direction in the theory of machine learning has the name of statistical machine learning. It was founded by V.Vapnik and Ja. Chervonenkis in the sixties-seventies of the last century and continued in nineties of the same century and has the name of Vapnik-Chervonenkis theory (VC theory) [Vapnik, 2000].

It should be noted that classification algorithms built on the basis of training sets are mostly unstable because learning set is not regular (in general). That is why it has been appeared the idea of development of algorithms that partially use statistical machine learning but have essentially less sensitivity to irregularity of the training sets.

This chapter focuses on tasks that partially use learning or machine learning. According to the general concept of machine learning a set is divided into general training and test (control) subsets. For the training subset one assumes that the class labels are known for every object. Using test subsets one verifies the reliability of the classification. The reliability of algorithms has been tested by methods of cross-validation [Kohavi, 1995; Mullin, 2000].

Depending on the complexity of the classification all objects can be divided into three groups: items that are stable and are classified with high reliability ("easy" objects), objects belonging to the borderline area between classes ("ambiguous" objects) and objects belonging to one class, and deeply immersed inside another one ("misclassified" objects). Among those objects that may cause an error the largest part consists of terminal facilities. Therefore it is important to develop an algorithm that allows one to determine the largest number of frontier facilities. The principal idea of this approach consists in preclassification of objects by dividing them into three functional groups. Because of this it is possible to achieve much more reliable results of classification. This could be done by applying the appropriate algorithms for every of obtained groups of objects.

3.1 The most stable objects determination

The idea of the model building is as follows. The general object set that have to be classified is divided on three functional groups. To the first group of objects the algorithm selects the objects with high level of classification reliability. The high level of reliability means that objects are classified correctly under the strong (maximal) deviations of the parameters from optimal ones. From the point of view of classification complexity these objects belongs to the group of so called "easy" objects. The second group includes objects, on which there is no consensus. If one selects two algorithms in a composition of algorithms, they should be as dissimilar as possible and they should not be a consensus. If one uses larger number of algorithms, the object belongs to the second group if there is no consensus in all algorithms. If consensus building uses intermediate algorithms, parameters of which are within the intervals between the parameters of two the most dissimilar algorithms, this makes it impossible to allocate a larger number of objects, on which there is no consensus. Dissimilarity between algorithms is determined on the basis of the Hamming distance between results of two algorithms defined as binary sequences [Kyrgyzov, 2008; Vorontsov, 2008]. In practice this also means that in general it will not be detected the new objects, if one uses composition of more than two algorithms, on which one builds the consensus. The third group consists of those objects, on which both algorithms have errors, while they are in consensus. The error caused by these objects can not be reduced at all. Thus the error can not be less than the value determined by the relative amount of objects from the third group. The next step will be the reclassification of the second group of objects. This special procedure allows us to determine the true class, to which a particular object belongs to. Reclassifying the second group of objects we can also have some level of error. This error together with the error caused by the third group will give the total error of all proposed algorithms.

The research carried out in this paper concerns the analysis of statistical characteristics of the results of a consensus generating by two algorithms. The objective of the task analysis is a statistical regularity of characteristics of various subsets taken by division of the general set into blocks of different size. Probability distribution by the consensus for three groups of objects has been carried out by nonparametric estimation using Parzen window with Gauss kernels.

3.1.1 Experimental results

Figs 6-11 show the parametrically estimated pdfs for the probability of a correct consensus, the probability of incorrect consensus and the probability that consensus will not be reached.

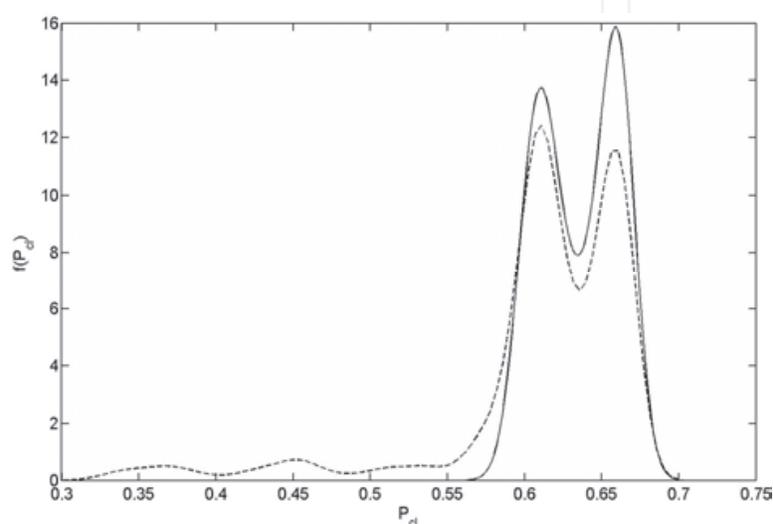


Fig. 6. Task "pima" from UCI repository: non-parametrical estimation of the pdf of the correct consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly) .

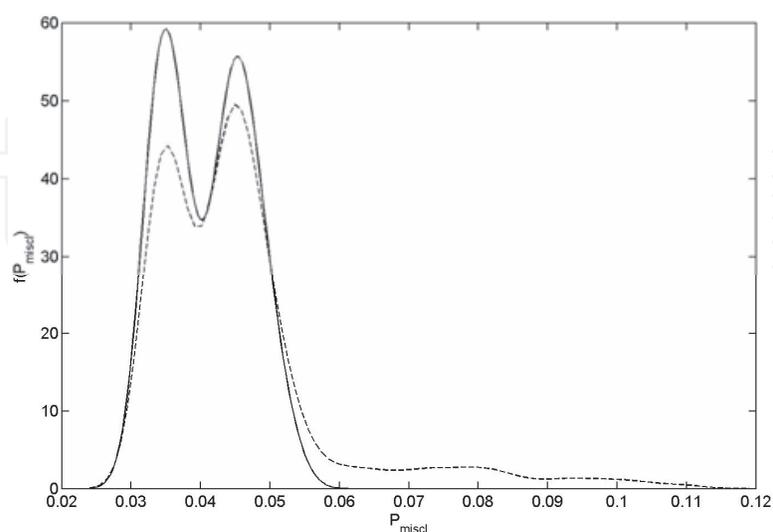


Fig. 7. Task "pima" from UCI repository: non-parametrical estimation of the pdf of the incorrect consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

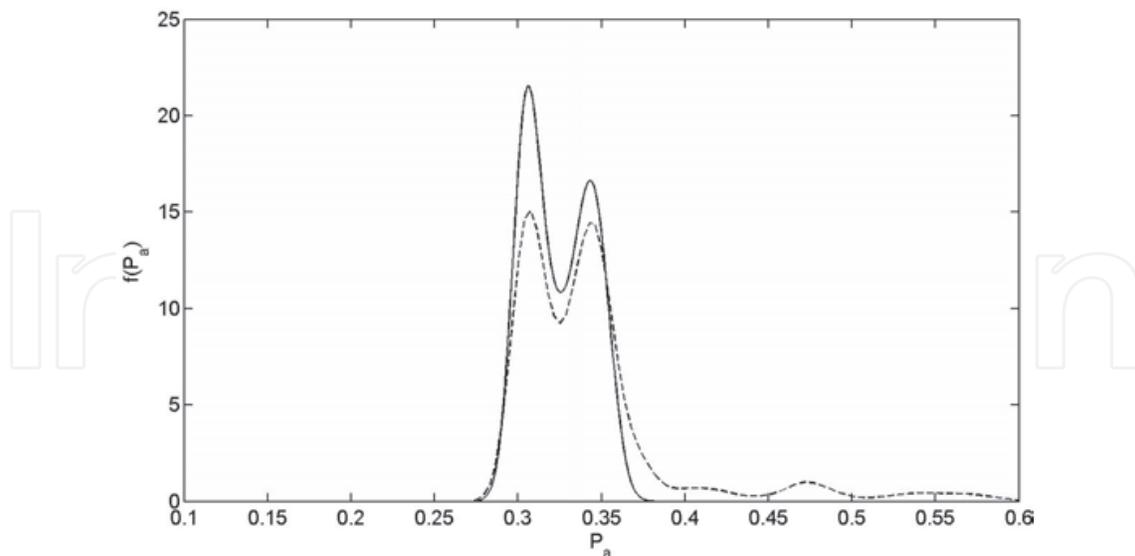


Fig. 8. Task "pima" from UCI repository: non-parametrical estimation of the pdf of no consensus between two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

As can be seen from the figures these distributions can be represented using one-component, two-component or multicomponent Gauss mixture models (GMM). In multicomponent GMM weights determined according to their impact factors. Distribution parameters (mean and variance) and weights of impact in the model are estimated using EM algorithm. Estimation of corresponding probability values was carried out by blocks with a minimal size of $Q = 30$ and $Q = 200$ elements. The size of these blocks has been driven by a small sample size which according to various criteria ranges from 30 to 200 items. According to the standard definition of a small sample it is assumed that sample is small when it is characterized by irregular statistical characteristics.

As seen from all figures the estimates obtained by blocks with a minimal size of 30 elements and some more are irregular. This means that for these tasks the sub-sample size of 30 items and some more is small. This has been indicated by long tails in the corresponding probability distributions. The maximum in zero point for two-component model is characterized by a large number of zero probabilities. This can be possible if there are no mistakes in the consensus of two algorithms. Estimates of probabilities on the basis of average values and the corresponding maximum probability distributions (for maximum likelihood estimation (MLE)) are not much different, which gives an additional guarantee for the corresponding probability estimates. Significance of obtained consensus estimates of probabilities of correct consensus, incorrect consensus and probability that consensus will not be achieved, provides a classification complexity estimate. Problems and algorithms for the complexity estimation of classification task is discussed in [Basu, 2006]. For example, tasks "pima" and "bupa" are about the same level of complexity because values of three probabilities are approximately equal. Tab. 1 shows that all algorithms excepting proposed new one have large enough sensitivity to the equal by the classification complexity tasks they work with. Mathematical analysis of composition building of algorithms has been considered in details in [Zhuravlev, 1978].

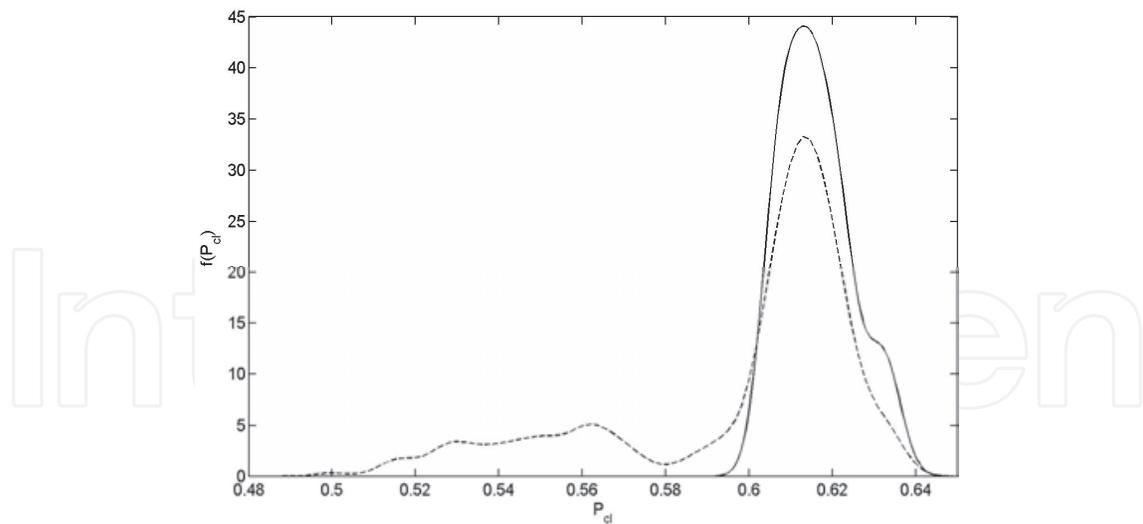


Fig. 9. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of the correct consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

Figs 6-11 show graphic dependencies of consensus results for problems taken from repository UCI. This repository is formed at the Irvin's University of California. The data structure of the test tasks from this repository is as follows. Each task is written as a text file where columns are attributes of the object and rows consist of a number of different attributes for every object. Thus the number of rows corresponds to the number of objects and the number of columns corresponds to the number of attributes for each object. A separate column consists of labels of classes, which mark each object. A lot of data within this repository has been related to biology and medicine. Also all these tasks could be divided according to the classification complexity. In the data base of repository there exists a number of tasks with strongly overlapped classes. Some of them will be used for research.

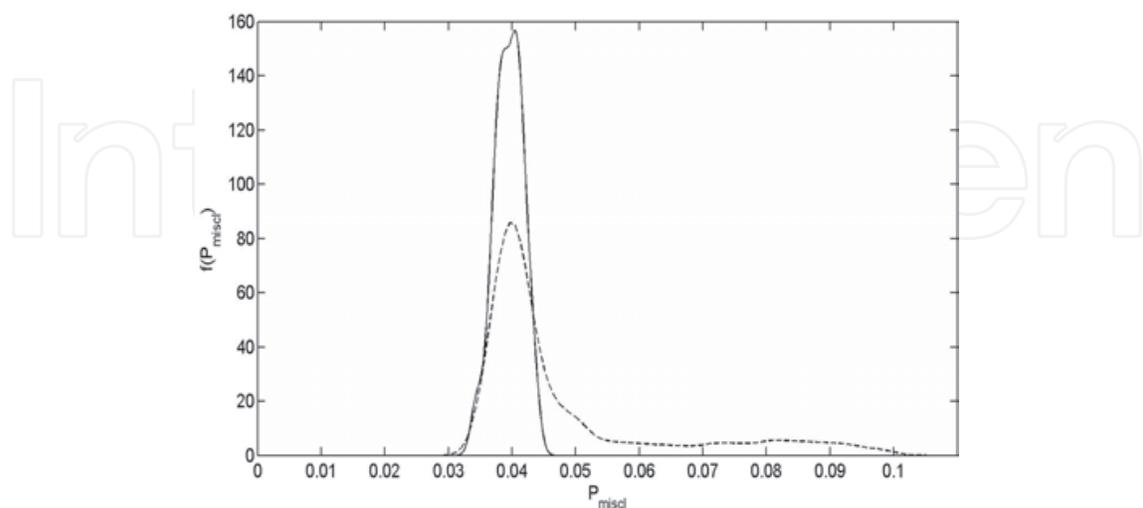


Fig. 10. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of the incorrect consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

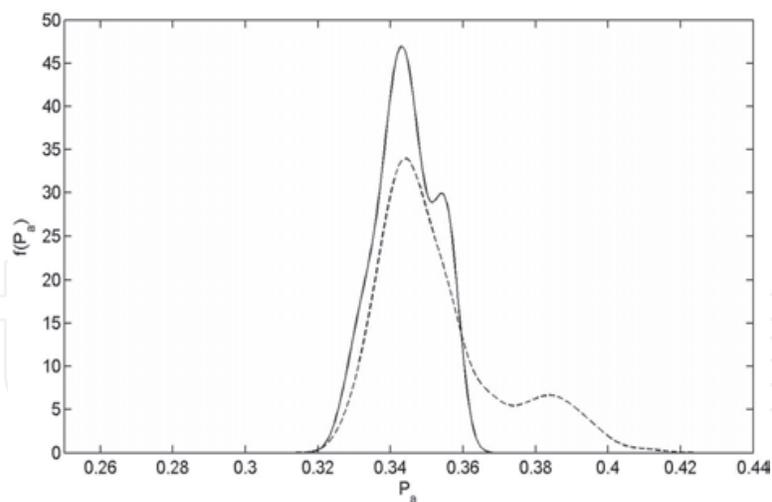


Fig. 11. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of no consensus between two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

In Tab. 1 one gives the probabilities of errors obtained on the test data for different classifiers or classifier compositions. All these algorithms were verified on two tasks that are difficult enough from the classification point of view. For the proposed algorithm it has been given the minimal and maximal errors that can be obtained on given tested data.

In Tab. 1 the value of minimal error is equal to consensus error for the proposed algorithm. The value of maximal error has been calculated as sum of minimal error and the half of the related amount of objects, on which there is no consensus (fifty-fifty principle). As seen from the table the value of maximal error is much less than the least value of error of all given algorithms for two tasks from UCI repository. In comparison with some algorithms given in the table the value of minimal error is approximately 10 times less for the proposed algorithm than the error of some other algorithms from the table. The proposed algorithms are characterized by much more stability of the classification error in comparison with other algorithms. It can be seen from corresponding error comparison for two tasks from the UCI repository.

Algorithm	Task	bupa	pima
Monotone (SVM)		0.313	0.236
Monotone (Parzen)		0.327	0.302
AdaBoost (SVM)		0.307	0.227
AdaBoost (Parzen)		0.33	0.290
SVM		0.422	0.230
Parzen		0.338	0.307
RVM		0.333	-
Proposed algorithm (min/max)		0.040/0.212	0.041/0.203

Table 1. Error of classification for different algorithms

	Q=200		Q=30	
	μ	σ	μ	σ
P_c	0.635	0.024	0.611	0.064
P_e	0.041	0.006	0.046	0.013
$P_{\bar{c}}$	0.324	0.019	0.344	0.052

Table 2. Task "pima" from UCI repository

	Q=200		Q=30	
	μ	σ	μ	σ
P_c	0.635	0.024	0.611	0.064
P_e	0.041	0.006	0.046	0.013
$P_{\bar{c}}$	0.324	0.019	0.344	0.052

Table 3. Task "bupa" from UCI repository

In tabs 2-3 the estimates of probability of belonging of every object from the task of repository UCI to every of three functional groups of objects have been given. In this case the objects, on which consensus of the most dissimilar algorithms exists (P_c), belong to the class of so called "easy" objects. Then objects, on which both of algorithms that are in consensus make errors (P_e), belong to the class of objects that cause uncorrected error and this error can not be reduced at all. The last class of objects consists of objects, on which there is no consensus of the most dissimilar algorithms ($P_{\bar{c}}$). This group of objects also belongs to the class of border objects. In the tables one gives variances of corresponding probabilities too. Minimal size of the blocks, on which one builds estimates using algorithms of cross-validation changes from 30 to 200.

3.1.2 Case of three classifiers in the consensus composition

In the previous case we analysed the classifier composition that consists of two the most dissimilar algorithms. Now we are going to build the classifier composition that consists of three algorithms. The third algorithms we choose considering the following requirements. These algorithms have to be exactly in the middle of two the most dissimilar algorithms. This means that the Hamming distance between the third algorithm and one of the most dissimilar algorithms is equal to the distance between the "middle" algorithm and the second algorithm in the consensus composition of two the most dissimilar algorithms. In Tabs 4 and 5 the results of comparison of two consensus compositions have been given. The first composition consists of two algorithms and the second one consists of three algorithms correspondingly. As in the previous case we used "pima" and "bupa" testing tasks from UCI repository.

	consensus of two classifiers		consensus of tree classifiers	
	μ	σ	μ	σ
P_c	0.635	0.024	0.607	0.021
P_e	0.041	0.006	0.0347	0.006
$P_{\bar{c}}$	0.324	0.019	0.358	0.017

Table 4. Task "pima" from UCI repository

As seen from the both tabs there is no big difference between two cases. Consensus of two algorithms can detect a bit larger quantity of correctly classified objects that means a bit more reliable detection of correctly classified objects. Consensus of three algorithms can detect a bit larger quantity of objects on which we have no consensus (the third group of objects). But if we will use the "fifty-fifty" principle for detection objects from the third group the general error of classification will be the same. We can also note that the variances of two consensus compositions have no large differences between each other.

	consensus of two classifiers		consensus of tree classifiers	
	μ	σ	μ	σ
P_c	0.616	0.008	0.586	0.012
P_e	0.040	0.002	0.037	0.002
$P_{\bar{c}}$	0.344	0.008	0.377	0.013

Table 5. Task "bupa" from UCI repository

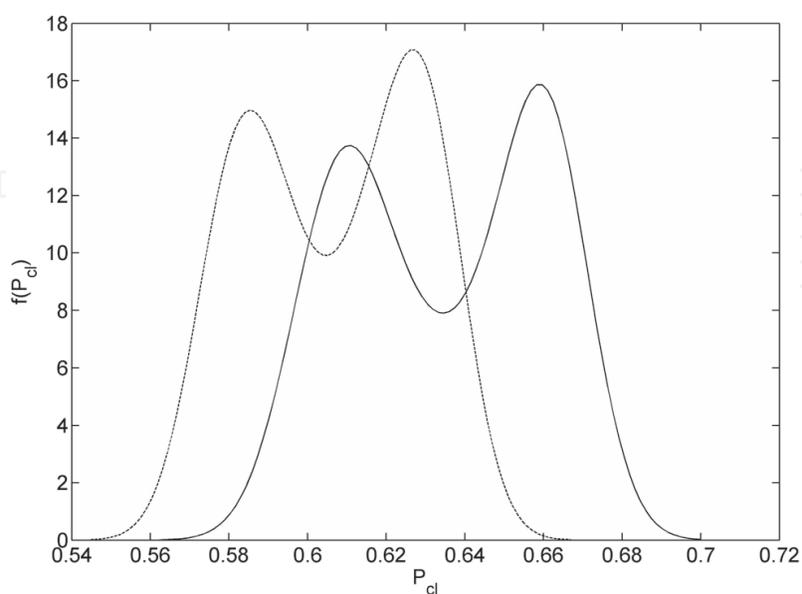


Fig. 12. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of correct consensus between two (solid line) and tree (dot-line) algorithms

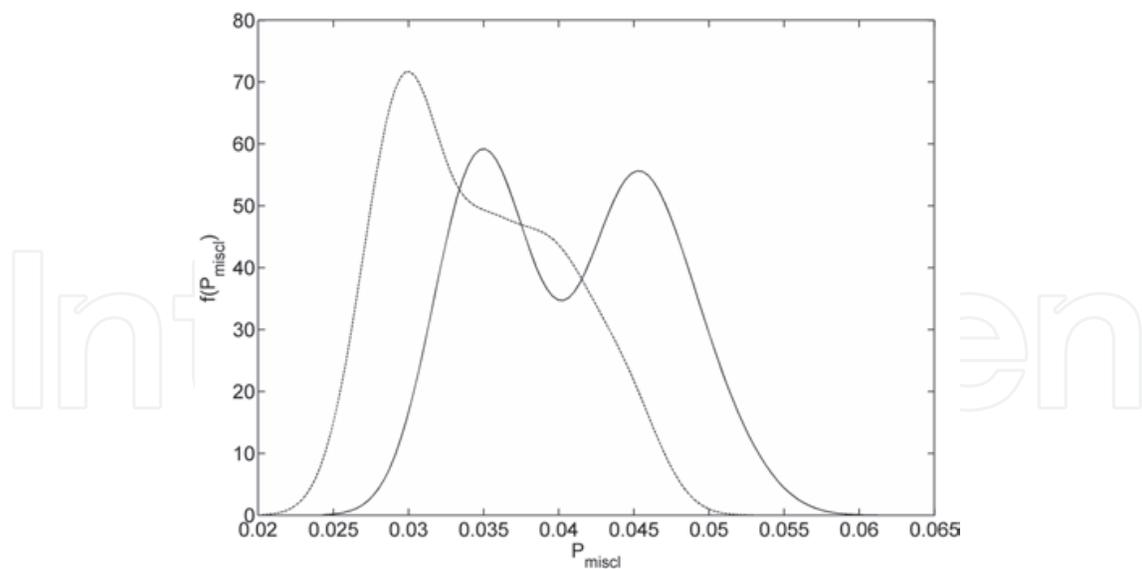


Fig. 13. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of incorrect consensus between too (solid line) and tree (dot-line) algorithms

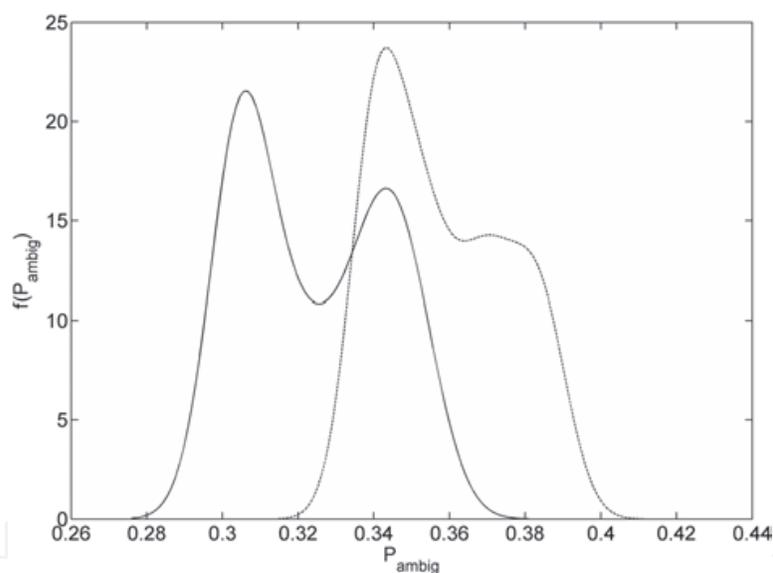


Fig. 14. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of no consensus between too (solid line) and tree (dot-line) algorithms

On figs 12-17 the results of consensus building for three algorithms have been given. Here we also use two tasks from the UCI repository as in the case of two algorithms. According to figs corresponding to the task of "bupa" we can make the following conclusions. In comparison to the case of two algorithms we can see that for the number of preclassified groups of objects we have just shifts between the corresponding pdfs and the form of curves is approximately the same. We can also note that relative value of the shift is rather small (about 5% for the pdf of correct probability). This shift is almost conditioned by the statistical error of determining of the most different algorithms.

According to figs corresponding to the task of "pima" we can mark that differences in forms of pdfs are more essential than in previous task. This circumstance could be used for

comparison of the task complexity using the value of overtraining as stability to learning. Using such approach it is possible to obtain much more precise and informative estimations of the complexity from the learning process point of view.

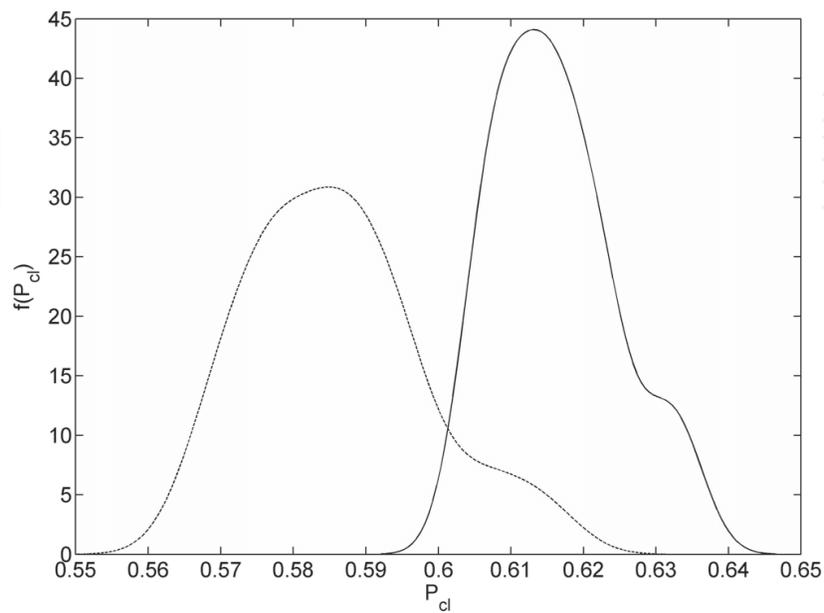


Fig. 15. Task "pima" from UCI repository: non-parametrical estimation of the pdf of correct consensus between too (solid line) and tree (dot-line) algorithms

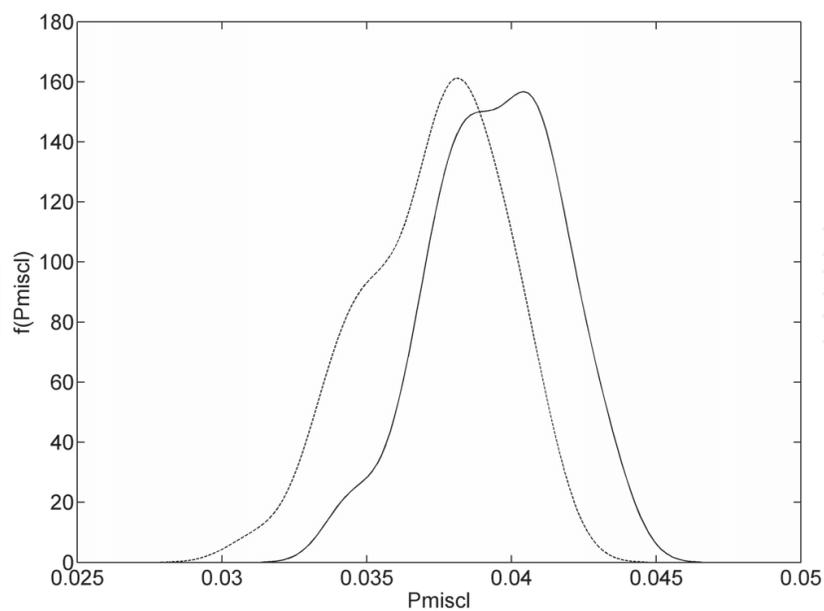


Fig. 16. Task "pima" from UCI repository: non-parametrical estimation of the pdf of incorrect consensus between too (solid line) and tree (dot-line) algorithms

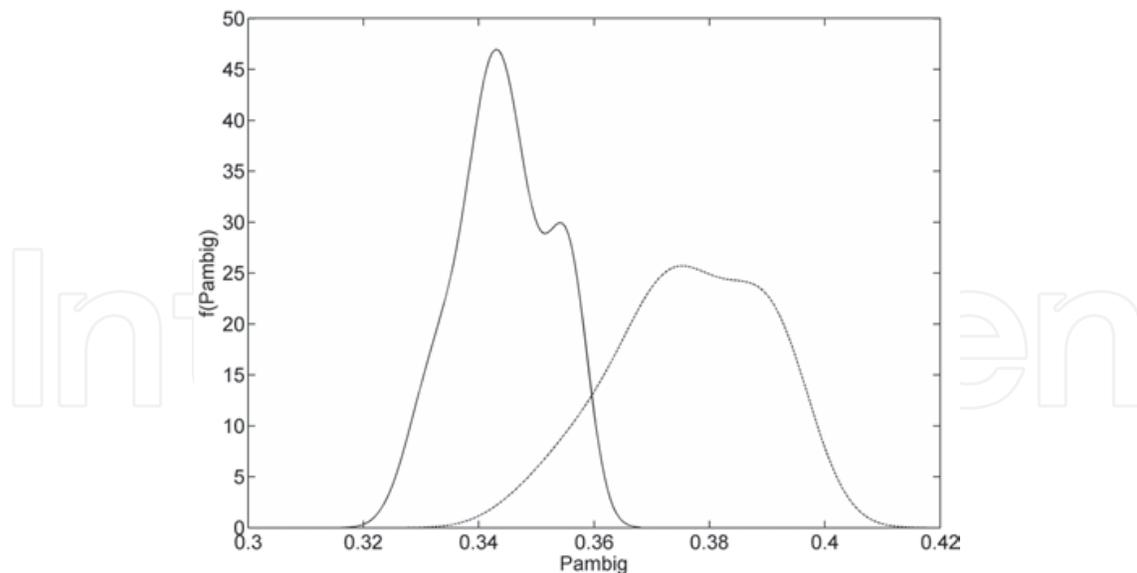


Fig. 17. Task "pima" from UCI repository: non-parametrical estimation of the pdf of no consensus between too (solid line) and tree (dot-line) algorithms

4. Specific of usage of the proposed approach for problems of face recognition

The problem of face recognitions is one of the principle tasks of the large project connected with determining of human behaviour and psychoanalysis of the human, based on the face expression and body movement. Such type of systems belongs to the class of no contact systems. Unlike to the human recognition systems based on fingerprints or images of iris these systems do not require human to keep finger of eyes near (or on) the scanner. This is also very important from law point of view. It is impossible to force a human to put the finger on the scanner if he does not want to do this and if this is not a criminal case. The same fact is concerned the case of iris recognition systems. To take a picture of somebody is not forbidden and this or that person could not be familiar with the fact that somebody took already picture of the face of such person. This is really important when creating the training and test databases. Face recognitions systems can be joined with hidden video cameras installed in shops, supermarkets, banks and other public places. Here it is important to hide the fact of video surveillance. This could be done with help of no contact recognition systems only. On other hand the facial information and mimicry could be used for the human behaviour determination and psychophysical state of the human. This is important to avoid and predict of acts of terrorism. Here it is very important information about dynamics of face expression and movement of the separate parts of the face.

In spite of the fact that face recognition systems has larger value of error of both of the types than finger print recognition systems, iris recognition systems and others, they find a lot of different applications because of their flexibility of installation, training and testing. In this situation it is very important to make research in the field of recognition probability estimation, overtraining estimation, model parameters estimation, etc. to find the most optimal parameters of the face recognition systems. To build very reliable recognition systems it is important to use proposed approach that allows us to build hierarchical recognition on the basis of objects division into functional groups and due to this to use the effect of preclassification.

For the procedure of decision making one proposes to use the notation of fiducial interval. By the fiducial interval one understands the list of possible candidates for the classification. Usage of the fiducial interval is very useful for the decision making support systems with presence of an operator. The result of the system work is the list of candidates that are the most similar to the object to be recognized. In this case the final decision about the object will be made by operator. The system can work as completely autonomous one with using of the fiducial interval for the decision making. In fiducial interval there exist several group of objects that belong to their own class. Our task is to find the group of objects that corresponds to the object to be recognised or to make decision that there is no corresponding objects in fiducial interval. The idea of fiducial interval consists in following concepts. The size of the fiducial interval (the number of possible candidates) has to be enough to be sure that if the corresponding objects are in the database of the recognition system they will drop into this interval. The size of the fiducial interval corresponds to the fiducial probability. The larger is fiducial interval the larger is fiducial probability. That is why it is convenient to use the notation of fiducial interval for the probability of the fact that corresponding objects will drop in the list of possible candidates. The second paragraph of this chapter has been devoted to the problems of forming of some types of fiducial intervals.

5. Discussion and future work

In this chapter we shortly considered some approaches for solution of such important problems as recognition reliability estimation and advanced classification on the basis of division of objects into three functional groups. In domain of reliability estimation there exist two principal problems. First problem concerns the tasks of statistical estimation of the probability of correct recognition especially for small training sets. This is very important when we can not achieve additional objects so fast and make our training set more representative. That could be in situations when we work with data slowly changing in time.

Another important problem concerns the effect of overestimation in pattern classification. The value of overestimation could be found as difference between the recognition results on training and test sets. In the beginning of the chapter one mentioned the main problems of the statistical learning theory and overestimation as one of the most principal problems. One did not pay attention to this problem in this chapter but it is planned to do in future research. The attention has been payed to the problems of recognition reliability estimation. In this chapter the results of both combinatorial and probabilistic approach to recognition reliability estimation have been presented. As seen from the figures there was realized the advanced analysis and estimation of the recognition results when the training set is decreased. So we can make the prognosis of the recognition probability for reduced training sets using combinatorial approach. The reliability of such approach can be provided on the basis of probabilistic approach.

It was considered some methods of the reliability estimation for some types of classifiers. Such of the classifiers belongs to the group of so called metrical classifiers or classifiers on the basis of dissimilarity functions or distance functions. It will be interesting to consider the proposed methods in case of other types of classifiers e.g. classifiers using separating hyperplane, classifiers built on logic functions and others. It will be interesting to consider the idea of how to express one classifier through another or to build relations between the different types of classifiers. All this could give us the possibility to use one approach to reliability estimation for any type of the classifiers.

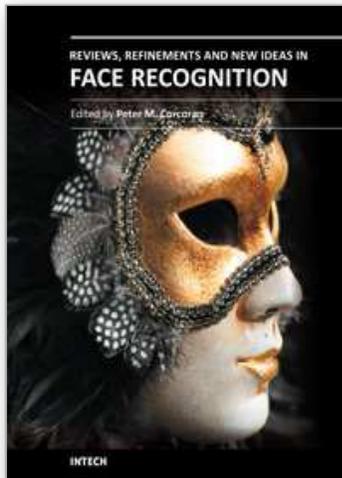
In the second part of the chapter the probability of belonging of every object to each of the three groups of objects: a group of "easy" objects, on which it is reached the correct consensus of two algorithms, a group of objects, on which two the most dissimilar algorithms have an incorrect consensus and a group of objects, on which one does not achieve consensus have been considered. The analysis shows that there are probability distributions of data that can be presented as a multicomponent models including GMM. All this makes it possible to analyze the proposed algorithms by means of mathematical statistics and probability theory. From the figures and tables one can see that the probability estimations using methods of cross-validation with averaged blocks of 30 and 200 elements minimum differ a little among themselves, which makes it possible to conclude that this method of consensus building, where consensus consists in the most dissimilar algorithms, is quite regular and does not have such sensitivity to the samples as other algorithms that use training. As seen from the corresponding tables, the minimum classification error is almost less by order of magnitude than error for the best of existing algorithms. The maximal error is less of 1.5 to 2 times in comparison with other algorithms. Also, the corresponding errors are much more stable both relatively to the task, on which one tests the algorithm and the series of given algorithms where the error value has significantly large variance. Moreover, since the minimal value of error is quite small and stable, it guaranties the stability of receipt of correct classification results on objects, on which consensus is reached by the most dissimilar algorithms. Relatively to other algorithms such a confidence can not be achieved. Indeed, the error value at 30–40% (as compared to 4%) gives no confidence in results of classification. The fact that the number of ambiguous objects selected by two the most different algorithms is less than the number of objects selected by three algorithms conditioned by the overtraining of two the most dissimilar algorithms. So the future research in this domain should be devoted to the problem of overtraining of the ensemble of two the most dissimilar algorithms. This means that it should be reduced the overtraining of the preclassification that allows us to reduce the error of classification gradually and due to this to satisfy much more reliable classification.

6. References

- Basu, M. & Ho, T. (2006). *Data complexity in pattern recognition*, Springer-Verlag, ISBN 1-84628-171-7, London
- Bishop, C. (2006). *Pattern recognition and machine learning*, Springer-Verlag, ISBN 0-387-31073-8, New York
- Kapustii, B.; Rusyn, B. & Tayanov, V. (2007) Mathematical model of recognition systems with small databases. *Journal of Automation and Information Sciences*, vol. 39, No. 10, pp. 70–80, ISSN 1064-2315
- Kapustii, B.; Rusyn, B. & Tayanov, V. (2008). Features in the Design of Optimal Recognition Systems . *Automatic control and computer sciences*, vol. 42, No. 2, pp. 64–70, ISSN 0146-4116
- Kapustii, B.; Rusyn, B. & Tayanov, V. (2008). Estimation of the Influence of Information Class Coverage on Generalized Ability of the k-Nearest-Neighbors Classifier. *Automatic control and computer sciences*, vol. 42, No. 6, pp. 283–287, ISSN 0146-4116

- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proceedings of 14th International Joint Conference on Artificial Intelligence*, pp. 1137-1145, Palais de Congres Montreal, Quebec, Canada, 2010
- Kyrgyzov, I. (2008). *Recherche dans les bases de donnees satellitaires des paysages et application au milieu urban :clustering, consensus et categorisation: Ph.D. thesis, l'ecole Nationale Superiere des Telcommunications, Paris*
- Moon, T., Stirling, S. (2000). *Mathematical methods and algorithms for signal processing*, Prentice-Hall, ISBN 0-201-36186-8, New Jersey.
- Mullin, M., Sukthankar, R. (2000) Complete cross-validation for nearest neighbour classifiers, *Proceedings of International Conference on Machine Learning*, pp. 639-646, 2000
- Tayanov, V. & Lutsyk, O. (2009). Classifier Quality Definition on the Basis of the Estimation Calculation Approach. *Computers & Simulations in Modern Science, Mathematical and Computers in Science and Engineering, A series of Reference Books and Textbooks*, pp. 166-171, ISSN 1790-2769
- Vapnik, V. (2000). *The Nature of Statistical Learning Theory (2)*, Springer-Verlag, ISBN 0-387-98780-0, New York
- Vorontsov, K. (2010). Exact combinatorial bounds on the probability of overfitting for empirical risk minimization. *Pattern Recognition and Image Analysis*, vol.20, No. 3 pp. 269-285 ISSN 1054-6618
- Vorontsov, K. (2008) On the influence of similarity of classifiers on the probability of overfitting pattern recognition and image analysis: new information technologies, *Proceedings of International Conference on Pattern Recognition and Image Analysis: new information technologies (PRIA-9)*, Volume 2., Nizhni Novgorod, Russian Federation, pp. 303-306, 2000
- Weinstein, E. (March 2011). *Gauss's Inequality*, In: *A Wolfram Web Resource*, 16.03.2011, Available from <http://mathworld.wolfram.com>
- Zhuravlev, J. (1978). An Algebraic Approach to Recognition and Classification Problems. *Problems of cybernetics*, vol.33, pp.5-68 (in Russian)

IntechOpen



Reviews, Refinements and New Ideas in Face Recognition

Edited by Dr. Peter Corcoran

ISBN 978-953-307-368-2

Hard cover, 328 pages

Publisher InTech

Published online 27, July, 2011

Published in print edition July, 2011

As a baby one of our earliest stimuli is that of human faces. We rapidly learn to identify, characterize and eventually distinguish those who are near and dear to us. We accept face recognition later as an everyday ability. We realize the complexity of the underlying problem only when we attempt to duplicate this skill in a computer vision system. This book is arranged around a number of clustered themes covering different aspects of face recognition. The first section on Statistical Face Models and Classifiers presents reviews and refinements of some well-known statistical models. The next section presents two articles exploring the use of Infrared imaging techniques and is followed by few articles devoted to refinements of classical methods. New approaches to improve the robustness of face analysis techniques are followed by two articles dealing with real-time challenges in video sequences. A final article explores human perceptual issues of face recognition.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Vitaliy Tayanov (2011). New Principles in Algorithm Design for Problems of Face Recognition, Reviews, Refinements and New Ideas in Face Recognition, Dr. Peter Corcoran (Ed.), ISBN: 978-953-307-368-2, InTech, Available from: <http://www.intechopen.com/books/reviews-refinements-and-new-ideas-in-face-recognition/new-principles-in-algorithm-design-for-problems-of-face-recognition>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen