Construction of the model of recognition operators in the large dimensional feature space

Sh Fazilov, S Radjabov, O Mirzaev, S Mirzaeva
Tashkent University of Information Technologies, 108, Amir Temur ave., Tashkent, 100200, Uzbekistan
E-mail: s_radjabov@yahoo.com

Abstract. In this paper, the problem of constructing a modified model of recognition operators based on potential functions is considered. The main advantage of the proposed recognition operators is the improvement of accuracy and a significant reduction in the amount of computational operations for the recognition of unknown objects, which makes it possible to apply them in the construction of real-time recognition systems. To test the efficiency of the proposed model, experimental research was carried out to solve the generated model problem and the problem of face recognition.

1. Introduction
It is known that the specialists of applied mathematics and computer science have been focusing their research works on pattern recognition problem. As indicated in [1,2], the first works on pattern recognition include the works of R. Fisher, A. Kolmogorov and A.Ya. Khinchin, performed in the first half of the 20th century.

In recent years, an increasing number of specialists have been paying attention to the problem of pattern recognition, and the number of scientific publications on this subject is constantly growing. This is due to the fact that in recent years the recognition of images has been increasingly applied in science, technology, production and everyday life. Mainly with their help the problems of geological forecasting, medical and technical diagnostics, face recognition, speech recognition and other applied problems are being solved.

The analysis of literary sources on recognition shows that different authors, in particular [3-14], give a different classification of methods of pattern recognition. For example, in [5], existing methods are divided into three categories: heuristic, mathematical and linguistic. In [6] the recognition methods are classified into two groups. The first group consists of methods based on operations with characteristics, which are called intentional methods of recognition. This group includes algorithms based on: estimates of the distribution densities of characteristic values; assumptions about the class of decision functions; the apparatus of mathematical logic; structural method. The methods of the second group are called extensional methods of recognition. These methods are based on operations with objects. These include algorithms based on: the method of comparison with the prototype; the kNN method; the method of voting; the team of decisive rules.

A similar classification of problems and methods of recognition with a certain level of detail is given in many literature sources. However, in our opinion, one of the most successful
classifications of recognition methods belongs to Academician of RAS Yu.I.Zhuravlev, which distinguishes recognition models based on historically established schools and directions in this area [1,3,4].

The development of the pattern recognition theory is divided into two stages. The first stage of development was the nature of projects of various technical devices or algorithms for solving specific applied problems. The value of the developed methods of pattern recognition is determined primarily by the obtained experimental results. The second stage of development is characterized by the transition from individual algorithms to the construction of models a family of algorithms for a single description of methods for solving classification problems. By now, a number of models of recognition algorithms have been thoroughly developed and studied in detail: models based on separating functions [5,6,10,14]; models based on mathematical statistics and probability theory [5,6,10,11]; models based on the potential principle [5,6]; models based on the calculation of estimates [5,6]. However, the analysis of these models shows that they are focused on solving problems, where objects are described in a space of features of a small dimension. In this regard, the issues of development and research of models of recognition algorithms oriented to solving problems of diagnostics, prediction and classification of objects in conditions of large dimensionality of the feature space are relevant.

The purpose of this paper is to build a model of recognition operators, taking into account the large dimensionality of the feature space. To achieve this purpose, a modified model of recognition operators has been developed, the key point of which is the selection of preferred features. As the initial model, a model of recognition operators based on potential functions was chosen [3,5,15].

It should be noted that when solving applied recognition problems, where objects are specified in the space of features of small dimension, the proposed model does not give very good results. This is explained by the fact that in such a situation most of the features are correlated rather weakly, which does not allow us to build a sufficiently simple recognition operator. However, in the case of a large dimensionality of the feature space, the probability that many features are strongly correlated increases. The experience of solving a number of model and applied problems shows that in this case the proposed recognition operator works more efficiently than the original one. Thus, the available priori information about the problem under consideration prompts us which of the developed models of recognition operators is more suitable for solving it.

2. Basic concepts and notations

Relying on [1,3], we introduce some notions and notations. Consider the set of admissible objects \( \mathcal{S} \), which is covered by \( l \) subsets (classes): \( K_1, K_2, \ldots, K_l \), \( K_i \cap K_j = \emptyset, i \neq j, i,j \in \{1, \ldots, l\} \).

It is assumed that the partition of the \( \mathcal{S} \) is not completely defined, but there is some initial \( J_0 \) information about the classes.

Let the objects \( S_1, \ldots, S_i, \ldots, S_m (\forall S_i \in \mathcal{S}, i = 1, \ldots, m) \) given in the space of initial features of \( X (X = (x_1, \ldots, x_i, \ldots, x_n)) : S_1 = (a_{11}, \ldots, a_{1i}, \ldots, a_{1n}), \ldots, S_m = (a_{m1}, \ldots, a_{mi}, \ldots, a_{mn}) \).

We introduce the following notations:

\[
\tilde{S}^m = \{S_1, \ldots, S_i, \ldots, S_m\}, \tilde{K}_j = \tilde{S}^m \cap K_j, C\tilde{K}_j = \tilde{S}^m \setminus \tilde{K}_j.
\]

Then the initial information \( J_0 \) can be given in the form

\[
J_0 = \{S_1, \ldots, S_i, \ldots, S_m; \tilde{a}(S_1), \ldots, \tilde{a}(S_i), \ldots, \tilde{a}(S_m)\},
\]
where $\tilde{\alpha}(S_i)$ is the information vector of the object, which is given in the form

$$\tilde{\alpha}(S_i) = (\alpha_{i1}, \ldots, \alpha_{ij}, \ldots, \alpha_{il}),$$

$$\alpha_{ij} = \begin{cases} 
1, & \text{if } S_i \in \tilde{K}_j; \\
0, & \text{if } S_i \notin \tilde{K}_j.
\end{cases}$$

The set of information vectors corresponding to the objects forms the information matrix $\|\alpha_{ij}\|$.

A description (numerical characteristic) of the object $I(S_i) = (a_1, \ldots, a_i, \ldots, a_n)$ corresponds to each object $S \in \tilde{S}$ in the space of the initial features $X = (x_1, \ldots, x_i, \ldots, x_n)$.

For arbitrary recognition algorithms, the following assertion is relevant [3]: any recognition algorithm $A$ can be represented as a composition of two operators $B$ and $C$:

$$A = B \circ C,$$

$$B(J_0, \tilde{S}^q) = \|b_{ij}\|_{q \times l},$$

$$C\left(\|b_{ij}\|_{q \times l}\) = \|\beta_{ij}\|_{q \times l},$$

where $b_{ij}$ are real numbers, $\beta_{ij}$ is the value of the predicate $P_j(S'_i) = "S'_i \in K_j" \quad (\beta_{ij} = P_j(S'_i)), \beta_{ij} \in \{0, 1, \Delta\}$, where $\beta_{ij}$ is interpreted as in the works of Yu.I. Zhuravlev.

It follows that each algorithm $A \in \{A\}$ can be divided into two consecutive stages: 1) the problem $Z$ is translated into a numerical matrix $\|b_{ij}\|$ of size $q \times l$ ($q$ is the number of rows equal to the number of objects, $l$ is the number of columns equal to the number of classes); 2) according to this numerical matrix, the decision rule determines the belonging of the objects to the classes.

The paper considers only the threshold decision rules, in which the decision is made element-by-element. Let $b \in \{b_{ij}\}$ and $c_1, c_2$ are certain threshold numbers ($0 < c_1 < c_2$). Then the decision rule is defined as follows [1]:

$$C(b) = \begin{cases} 
0, & \text{if } b < c_1; \\
1, & \text{if } b > c_2; \\
\Delta, & \text{if } c_1 \leq b \leq c_2.
\end{cases}$$

Due to the fact that all recognition algorithms can be represented as a composition of two operators, then only recognition operators are considered.

3. Problem statement

Lets consider an arbitrary collection of objects of $\tilde{S}^q = \{S'_1, \ldots, S'_q\} (\tilde{S}^q \subset \tilde{S})$. Let $\tilde{S}^q$ be objects in $X$. It is assumed that the dimensionality of the space of features is quite large. Under these conditions, many features can be correlated, and this makes it difficult to use most recognition algorithms without significantly modifying them [4]. The task is to construct a recognizing operator $B$ that allows us to calculate the predicate $P_j(S'_i)$ values from the initial information $J_0$ using the decision rule $C$ (see formula 1). In other words, the required operator $B$ (with application of $C$) takes the set $(J_0, \tilde{S}^q)$ to the matrix $\|b_{ij}\|$ [1].
4. Solving Method

The paper considers a new approach to solving the problem of constructing a recognition operator in conditions of large dimensionality of the feature space. Based on this approach, a model of modified recognition operators based on potential functions is proposed. The main idea of the proposed model of recognition operators is to form a space of independent and preferred features with subsequent recognition of objects specified in this space. The task of these algorithms includes the following main steps.

1. Formation of subsets of strongly correlated features. At the first stage, a system \( W_A \) of "independent" subsets of strongly correlated features, whose composition will depend on the parameter \( n' \) (the number of "independent" subsets), is defined. Let \( \Xi_q (q = 1, n') \) be a subset of strongly correlated features. The measure of the proximity \( L(\Xi_p, \Xi_q) \) between the subsets \( \Xi_p \) and \( \Xi_q \) can be specified in various ways, for example:

\[
L(\Xi_p, \Xi_q) = \frac{1}{N_p \cdot N_q} \sum_{x_i \in \Xi_p} \sum_{x_j \in \Xi_q} \eta(x_i, x_j),
\]

where \( N_p, N_q \) is the number of features belonging respectively into the subsets of \( \Xi_p, \Xi_q \); \( \eta(x_i, x_j) \) is the function characterizing the strength of the pair connection between the features of \( x_i \) and \( x_j \).

Depending on the method of assigning a measure of proximity between subsets of strongly correlated features (\( \Xi_p \) and \( \Xi_q \)) and the quality functional of the classification analysis, it is possible to obtain various procedures for extracting independent sets of strongly correlated features [16-17].

2. Selection of a set of representative features. The second step in defining the model of algorithms is to select a set of representative features. The basic idea of choosing representative features lies in their difference (dissimilarity) in the formed set of representative features [16-18]. In the process of forming a set of representative features, it is required that each distinguished feature is a typical representative of its subset of strongly correlated features. As a result of this stage, we obtain an abbreviated feature space, whose dimension is much smaller than the original \((n' < n)\) [16,17]. Further, the formed feature space is denoted by \( X' (X' = (x'_1, \ldots, x'_i, \ldots, x'_n')) \). It should be noted that in the space \( X' \) each object \( S \) corresponds to an \( n' \)-dimensional vector \((a_1, \ldots, a_i, \ldots, a_n')\).

3. Determination of preferred features. Consider a set of representative features \( \{x'_1, \ldots, x'_i, \ldots, x'_n'\} \), defined in the previous step. The choosing of the \( X' \) as a preferred feature is based on an assessment of the dominance of the feature in question, which divides the objects belonging to the set into two subsets \( K_j \) and \( C \bar{K}_j \) [18]:

\[
\Psi_{ij} = \frac{\Theta_{ij}}{\Lambda_{ij}},
\]

\[
\Theta_{ij} = \left( \sum_{s \in K_j} \sum_{s_u \in C \bar{K}_j} (a_i - a_{iu})^2 \right) \tilde{N}_{1j},
\]

\[
\Lambda_{ij} = \left( \sum_{s \in K_j} \sum_{s_u \in C \bar{K}_j} (a_i - a_{iu})^2 \right) \tilde{N}_{2j},
\]

where \( \tilde{N}_{1j} = m_{1j} \times m_{2j}, \tilde{N}_{2j} = (m_{1j}(m_{1j} - 1) + m_{2j}(m_{2j} - 1)) \), \( m_{1j} = |K_j|, m_{2j} = |C \bar{K}_j| \).

The larger the value of \( \Psi_{ij} \), the greater the preference is given to the corresponding feature when dividing objects belonging to \( K_j \). If two or more features receive the same preference, then
either one is selected. When calculating $R_{ij}$, it is assumed that $S$ and $S_u$ are different objects (i.e. $S \neq S_u$).

For each subset $K_j$ at this stage, the preferred feature is determined and it is denoted by $\chi_j$. As a result, a set of preferred features $\chi_j = (\chi_1, \chi_2, \ldots, \chi_{n^j})$ is formed. Each set of preferred features characterizes only one subset (class) of objects. Only the preferred features are discussed below.

4. Determination of the difference function $d(S_u, S_v)$ between objects $S_u$ and $S_v$. At this stage, difference function is defined that characterizes the differences between the objects $S_u$ and $S_v$ in the new space of the preferred features $\chi_j \ (\chi_j = \bigcup_{j=1}^l \chi_j)$. It should be noted that the $\chi_j$ space is formed as a result of reducing the dimension of the feature space $X'$. The following principle is used in constructing the function $d(S_u, S_v)$: "the larger the value function $d(S_u, S_v)$, the greater the difference between these objects".

Let two objects $S_u$ and $S_v$ be given in space $\chi_j$:

$$S_u = (a_{u1}, \ldots, a_{uk}) \quad \text{and} \quad S_v = (a_{v1}, \ldots, a_{vk}).$$

The difference between these objects is determined by one of the following ways:

a). Let the distance between the objects and with respect to the $j$-th feature be given in the form

$$d(S_u, S_v) = (a_{ui} - a_{vi})^2.$$ 

Then the difference function between objects is given in the form

$$d(S_u, S_v) = \sum_{i=1}^k d_i(S_u, S_v).$$

b). It is known that not all the features describing the objects under consideration are the same importance when solving practical problems of pattern recognition. This difference in the importance of features is taken into account by the introduction of a new parameter $\lambda_i (i = 1, k)$, which characterizes the importance of the feature $\chi_i \ (\chi_i \in \chi)$. In this case, the $d(S_u, S_v)$ function is defined as follows:

$$d(S_u, S_v) = \sum_{i=1}^k \lambda_i d_i(S_u, S_v).$$

5. Giving the proximity function $\phi(S_u, S_v)$ between objects $S_u$ and $S_v$. At this stage, the proximity function between objects $S_u$ and $S_v$ is determined by the potential functions $\phi(S_u, S_v)$ [15].

We introduce a positive function $\phi(S_u, S_v)$ ($S_u, S_v \in \chi$) that decreases as the object $S_v$ moves away from the object $S_u$. To each object $S_v$ in the space $\chi$ there corresponds one point, which we also denote by $S_v$. The function $\phi(S_u, S_v)$ plays the role of the charge potential located at the point $S_v$ for a fixed $S_u$. Typical examples of potential functions can be [5,15]:

1) $\phi(S_u, S_v) = e^{-\xi d(S_u, S_v)}$,

2) $\phi(S_u, S_v) = \frac{1}{1 + \xi d(S_u, S_v)}$.

where $\xi$ is the parameter of the algorithm.

6. Calculation of the membership score for an object $S$ by class $K_j$. At this stage, the evaluation (in the form of the total potential) for $S$ relative to objects belonging to the class $K_j$ is calculated. In this case, each class of objects is characterized by its total potential [15,17].

We suppose that the objects $S_{m_j-1+1}, S_{m_j-1+2}, \ldots, S_{m_j}$ belong to the class $K_j$ 

$$\left\{ S_{m_j-1+1}, S_{m_j-1+2}, \ldots, S_{m_j} \right\} = \hat{S}^n \cap K_j.$$ 

We consider the total potential for
an object $S$ for objects of class $K_j$. Let the values of the potential functions 
\[\phi(S_{m_{j-1}+1}, S), \phi(S_{m_{j-1}+2}, S), \ldots, \phi(S_{m_j}, S)\] 
be calculated. The total potential for a class is assumed to be the function 
\[\phi(S) = \sum_{S_u \in K_j} \gamma_a \phi(S_u, S), \quad \tilde{S}^m \cap K_j,\]
where $\gamma_a$ is the parameter of the algorithm.

Thus, we define a model of modified recognition operators based on the potentials principle. When the operator $B(S')$ is applied successively to $S'_1, \ldots, S'_i, \ldots, S'_q$ objects, we obtain the matrix $\|b_{ij}\|_{q \times l}$, $b_{ij} = \mu_j(S'_i)$.

It should be noted that any operator $B$ from this model is completely determined by specifying a set of parameters $\pi = \left( n', \{\tilde{w}\}, n'', \{\tilde{\rho}\}, \{\lambda_i\}, \xi, \{\gamma_u\} \right)$. We denote the set of all recognizing operators from the proposed model by $B(\pi, S)$. The search for the best algorithm is performed in the parameter space of $\pi$ taking into account the value of the $c_1, c_2$ parameters of the threshold decision rule (1). Undoubtedly, the construction of the best algorithm can be performed in the parameter space of $\pi$, and their effectiveness (as well as of all heuristic algorithms) is determined by the results of their application in solving a number of applied problems.

These operators are slightly different from known recognition operators based on potential functions [5,15], especially for the first two stages. However, it should be noted that without these stages it is quite difficult to solve many applied problems.

5. Experiments
An experimental study of the operability of the proposed model of recognizing operators was carried out on the example of the generated model problem and the person identification problem on the base of face images.

The initial data of the recognized objects for the model problem are generated in the space of the correlated features. The number of classes in this experiment is two ($l = 2$). The size of the initial sample consists of 200 implementations ($V = 200$). There are 100 objects in each class, i.e. $|\tilde{K}_j|$. The number of features in the model example is 120. The number of subsets of strongly correlated features is 5.

The initial data of recognizable objects for face recognition problem consists of 360 face images that make up four disjoint subsets (class, i.e. $l = 4$). There are 90 ($|\tilde{K}_j| = 90$) different faces for one person, they are photographed at different times, but the shooting conditions are approximately the same.

As test models of recognition operators, we choose: the classical model of recognition operators of the type of potential functions $\left(B_1\right)$ [15], the model of recognition operators based on the evaluation of estimates $\left(B_2\right)$ [16], and the model $\left(B_3\right)$ of recognition operators proposed in this paper. Comparative analysis of the above models of recognition operators in solving the problems examined is based on the accuracy of recognition of the objects of the control sample and the time spent for recognizing objects: the time spent on training; time spent on recognizing objects from the control sample.

To determine the evaluation by these criteria the method of sliding control is used in order to exclude the successful (or unsuccessful) partitioning of the initial sample $V$ into two parts $V_1$ and $V_k$ ($V = V_0 \cup V_k, V_0$ is the sample for training, $V_k$ is the sample for control) [19]. In this method, the initial selection of objects is randomly divided into 10 non-overlapping blocks, each of which include $\xi$ (in model problem $\xi = 20$, in the face recognition problem $\xi = 36$) objects. In this case, it is required that the proportion by the number of objects belonging to different
classes in all blocks is preserved. As a result, it turns out that each block includes \( h \approx \left( \frac{t}{l} \right) \), (in model problem \( h = 10 \), in the face recognition problem \( h = 9 \)) objects of each class. The process of sliding control of these blocks includes several steps: at each step, 8 of 10 blocks are selected as training sample, and recognition operators with given parameters are trained on this sample. The trained recognition operator is then checked on the remaining 2 blocks (control sample).

As a result of each checking, an assessment of the quality of the recognition operators is determined and recorded according to the specified criteria. When each next step is performed to evaluate the quality of the recognition operators from the control and training samples, one block is selected and replaced by their places. In order to avoid re-use of the objects of the control sample, the corresponding blocks are marked, and they do not participate in the selection of candidates for inclusion in the control sample. After the completion of the sliding examination procedure, the accuracy of recognition and time scores were determined as averages. The experiments were carried out on a Pentium IV Dual Core 2.2 GHz computer with 1 Gb of RAM.

6. Results and discussion
In order to test the operability of the proposed model of operators, an experimental study was first carried out in solving of the model problem, and then when solving the problem of person recognition from the face images. The results of solving the model problem with the use of recognition operators \( B_1 \), \( B_2 \) and \( B_3 \) are given in Table 1.

| Recognition operator | Time for training (sec.) | Time for recognition (sec.) | Accuracy (%) |
|----------------------|--------------------------|-----------------------------|--------------|
| \( B_1 \)            | 4,898                    | 0,049                       | 81,4         |
| \( B_2 \)            | 16,763                   | 0,022                       | 94,6         |
| \( B_3 \)            | 14,924                   | 0,020                       | 96,7         |

Table 1. The results of solving a model problem using various recognition operators.

Table 2 shows the results of solving the problem of person recognition using the same recognition operators. It should be noted that the original images were preliminarily processed using the algorithms given in [20,21].

| Recognition operator | Time for training (sec.) | Time for recognition (sec.) | Accuracy (%) |
|----------------------|--------------------------|-----------------------------|--------------|
| \( B_1 \)            | 3,141                    | 0,024                       | 80,5         |
| \( B_2 \)            | 8,607                    | 0,011                       | 89,7         |
| \( B_3 \)            | 9,729                    | 0,009                       | 91,4         |

Table 2. The results of solving the problem of person recognition with the help of various recognition operators.

A comparison of these results shows (see Tables 1 and 2) that the proposed model of recognition operators \( B_3 \) made it possible to improve the accuracy of recognition of objects
described in the space of correlated features (10% higher than with $B_1$, and 2% higher than with $B_2$). This is because the model $B_1$ does not take into account the correlations of the features, and the model $B_2$ has a larger VC-dimension [13] than $B_3$. For model $B_3$, the time spent on object recognition is much less than the same indicator of models $B_1$ and $B_2$ (see Tables 1 and 2). This is due to the fact that in the proposed recognition operators, only the preferred features are used to recognize the objects, which has caused an increase in the speed of object recognition. Therefore, this model can be used in the development of real-time recognition systems. At the same time, it is necessary to note the fact that the time spent on learning the algorithm has increased, because for constructing an optimal recognition operator, it is required to optimize a greater number of parameters than that of using traditional models of recognition operators, in particular $B_1$.

7. Conclusion
Currently solving the problems of pattern recognition, described in large dimensionality of feature spaces is associated with significant computational difficulties. To reduce the number of computational operations, it may be effective to isolate representative features in the context of the features correlations within the framework of any sample from the general population. The developed model can be used in the compilation of various software complexes focused on solving problems of pattern recognition and object classification.

References
[1] Zhuravlev Y and Gurevich I 1988 Recognition, classification, prediction 2 5
[2] Potapov A S 2007 Pattern recognition and machine perception (St. Petersburg: Politechnic) p 548
[3] Zhuravlev Y I et al. 2006 Recognition. Mathematical methods. Software system. Practical application (Moscow: Fazis) p 159
[4] Fazilov S, Mirzaev O and Radjabov S 2015 Problems of computational and applied mathematics 2 99
[5] Tou J and González R C 1978 Principles of pattern recognition (Moscow: Mir) p 414
[6] Simankov V S and Lutsenko E V 1999 Adaptive management of complex systems based on the theory of pattern recognition (Krasnodar: Publishing House of Kuban State Technical University) p 318
[7] Kudryavtsev V B, Andreev A E and Gasanov E E 2007 Theory of test recognition (Moscow: Fizmatlit) p 320
[8] Merkovich A B 2011 Pattern Recognition: An Introduction to Statistical Learning Methods (Moscow: Editorial URSS) p 256
[9] Haykin S 2006 Neural Networks: A Comprehensive Foundation (Moscow: Williams) p 1104
[10] Theodoridis S and Koutroumbas K 2003 Pattern Recognition (New York: Elsevier) p 711
[11] Dougherty G A 2013 Pattern Recognition and Classification (New York: Springer Science + Business Media) p 196
[12] Murty M N and Devi D V S 2015 Introduction to Pattern Recognition and Machine Learning (New Jersey: World Scientific) p 383
[13] Vapnik V N and Chervonenkis A Y 1974 Theory of Pattern Recognition (Moscow: Nauka) p 416
[14] Jayadeva, Khemchandani R and Chandra S 2017 Twin Support Vector Machines. Models, Extensions and Applications (New York: Springer) p 211
[15] Aizerman M A, Bravermann E M and Rozonoer I L 1970 The method of potential functions in the theory of machine learning (Moscow: Nauka) p 348
[16] Mirzaev N 2010 Proc. Int. Conf. on Computer Technology and Applications (Vladivostok) vol 1 (Vladivostok: IEEE) p 394
[17] Fazilov S, Mirzaev O and Mirzaev N 2016 Radioelectronics, informatics, management 1 58
[18] Kamilov M, Fazilov S, Mirzaev N and Radjabov S 2012 Proc. Int. Conf. on Problems of Cybernetics and Informatics (Baku) vol 1 (Baku: IEEE) p 184
[19] Braga-Neto U M and Dougherty E R 2016 Error Estimation for Pattern Recognition (New York: Springer) p 312
[20] Tukhtasinov M, Mirzaev N and Narzulloev O 2016 Proc. Int. Conf. on Dynamics of Systems, Mechanisms and Machines (Onsk) vol 1 (Onsk: IEEE) p 1
[21] González R C and Woods R E 2008 Digital Image Processing (New Jersey: Prentice-Hall) p 976