XVI IMEKO World Congress
Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future
Vienna, AUSTRIA, 2000, September 25-28

THE INFLUENCE OF FEATURES SEQUENCE ON RECOGNITION

J. Volf
Czech Technical University in Prague, Faculty of Mechanical Engineering
Technická 4, 166 07 Prague 6, Czech Republic

Abstract: This paper deals a method of tactile pattern recognition. It is based on discriminatory analysis. The computation goes through separate pattern classes. The Ivanovic deviation is used as a discriminatory function in this case. The m dimensional vector is input parameter. The method can be classified into n groups of patterns. Because the classification function Ivanovic deviation is probability function, influence of features sequence on right pattern recognition appears here. This problem is study in this paper.
Keywords: Discriminatory analysis, features, pattern recognition

1 INTRODUCTION

This paper deals with a method of tactile pattern recognition. The computation goes through separate pattern classes. The Ivanovic deviation is used as a discriminatory function in this case. The m dimensional vector is input parameter. The method can be classified into n groups of patterns. Pattern recognition can work at real time. The discriminatory analysis is a quick and effective method. It allows the solution of complex problems with regard to the possible number of pattern classes and of pattern features. Because the classification function Ivanovic deviation is probability function, influence of features sequence on right pattern recognition appears here. This problem is study in this paper.

2 THE MATHEMATICAL PROBLEM FORMULATION

Let us have the pattern classes marked $T_1,...,T_i,...,T_n$ and m of features marked $X_1,..., X_j,..., X_m$. The aim of problem classification is to insert an unknown pattern into appropriate pattern classes.

Let p_{ij_l} be the value of j features in l pattern (etalon) advanced to learning of i class. The arithmetical average $\bar{x}_{ij}$ feature value of X_j for pattern class T_i may be computed by the following relation:

$$\bar{x}_{ij} = \frac{1}{P_i}\sum_{l=1}^{P_i} p_{ijl} \tag{1}$$

P_i - number of pattern (etalons) for pattern class T_i

We compute matrix $\underline{X} = \left\| \bar{x}_{ij} \right\|$ dimension L x m x n, where $L = \max(P_i)$, on the basis of this data. This matrix we can disassemble on n matrixes by dimension of L x m. We may ones express for example by n matrix table, where <u>every</u> table corresponds to <u>one</u> pattern class. Arrangement of this matrix shows Fig. 1. Some examples of these matrixes are depicted in Table 1 - Table 3.

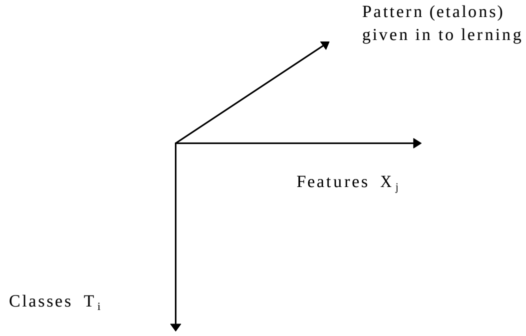

Figure 1. Arrangement of pattern matrixes

Table 1 to Table 3. Examples of pattern matrixes:

Table 1.

Feature / Etalon for class T_1	$\mathbf{X_1}$	$\mathbf{X_2}$	**...**	$\mathbf{X_j}$	**...**	$\mathbf{X_m}$
1	$\overline{x}_{11}$	$\overline{x}_{12}$	...	$\overline{x}_{1j}$	...	$\overline{x}_{1m}$
.	.	.		.		.
.	.	.		.		.
i	$\overline{x}_{i1}$	$\overline{x}_{i2}$	...	$\overline{x}_{ij}$	...	$\overline{x}_{im}$
.	.	.		.		.
.	.	.		.		.
P_1	$\overline{x}_{P_1 1}$	$\overline{x}_{P_1 2}$	...	$\overline{x}_{P_1 j}$	...	$\overline{x}_{P_1 m}$

Table 2.

Feature / Etalon for class T_i	$\mathbf{X_1}$	$\mathbf{X_2}$	**...**	$\mathbf{X_j}$	**...**	$\mathbf{X_m}$
1	$\overline{x}_{11}$	$\overline{x}_{12}$	...	$\overline{x}_{1j}$	...	$\overline{x}_{1m}$
.	.	.		.		.
.	.	.		.		.
i	$\overline{x}_{i1}$	$\overline{x}_{i2}$	...	$\overline{x}_{ij}$	...	$\overline{x}_{im}$
.	.	.		.		.
.	.	.		.		.
P_i	$\overline{x}_{P_i 1}$	$\overline{x}_{P_i 2}$	...	$\overline{x}_{P_i 2}$	...	$\overline{x}_{P_i m}$

Table 3.

Feature / Etalon for class T_n	$\mathbf{X_1}$	$\mathbf{X_2}$	**...**	$\mathbf{X_j}$	**...**	$\mathbf{X_m}$
1	$\overline{x}_{11}$	$\overline{x}_{12}$	...	$\overline{x}_{1j}$	...	$\overline{x}_{1m}$
.	.	.		.		.
.	.	.		.		.
i	$\overline{x}_{i1}$	$\overline{x}_{i2}$	...	$\overline{x}_{ij}$	...	$\overline{x}_{im}$
.	.	.		.		.
.	.	.		.		.
P_n	$\overline{x}_{P_n 1}$	$\overline{x}_{P_i 2}$	...	$\overline{x}_{P_n 2}$	...	$\overline{x}_{P_n m}$

The pattern classes present m dimensional statistic sets in this problem.

By formula (2) we compute dispersion σ^2 and standard deviation σ of the feature X_j for i pattern class (T_i):

$$s_{ij}^{2} = \frac{1}{P_i - 1} \sum_{a=1}^{P_i} (p_{ij_a} - \overline{x}_{ij})^2 \qquad (2)$$

i = 1, ... , n ; j = 1, ... ,m

Then we compute the symmetrical covariance matrixes $\underline{W}_i$ and the symmetrical correlation matrix $\underline{R}_i$:

The coefficients of the covariance matrix, which described covariances of features X_j and X_k in the pattern class T_i, are defined by these forms:

$$w_{ijk} = \frac{1}{P_i - 1} \sum_{a=1}^{P_i} (p_{ija} - \overline{x}_{ij})(p_{ika} - \overline{x}_{ik}) \qquad j \neq k$$

$$w_{ijj} = s^{2}_{ij} \qquad i = 1, \ldots, n \tag{3}$$
$$w_{ijk} = w_{ikj} \qquad j = 1, \ldots, m$$

and the coefficients of the correlation matrix are defined by these forms:

Let's $T_i = (\overline{x}_{i1}, \overline{x}_{i2}, \ldots, \overline{x}_{im})$ be the values of arithmetical averages of features for class T_i and UNKN = (z_1, ... , z_j, ... ,z_m) the values of i an unknown pattern.

Let us mark the value differences of j feature of i class and j feature of unknown patterns:

$$|d_{ij}| = |\overline{x}_{ij} - z_j| \tag{4}$$
$$i = 1, \ldots, n \qquad j = 1, \ldots, m$$

A deviation of m features is observed between class T_i and an unknown pattern UNKN. This deviation presents the above expressed value of set of discrimination effects for all m features and in discrimination analysis is called deviation and marked D.

A set of formulations of these deviations exists, but not all formulations comply with conditions described in [1], [2]. The Ivanovic deviation formulated in [1] best answers this purpose. This one for an unknown pattern of class T_i is defined by the formula:

$$D_i = \sum_{j=1}^{m} \frac{|d_{ij}|}{s_j} \prod_{k=1}^{j-1} (1 - r_{kj}) \tag{5}$$
$$\text{for} \quad i=1, \ldots, n \qquad j=1, \ldots, m$$

where
d_{ij} - is defined by relation (4)
σ_j - is standard deviation of features X_j
r_{kj} - is coefficient of correlation between features X_k and X_j

This deviation fully describes not only linear, but also even stochastic dependencies and non-dependencies. At the same time it enables us to compare individual classes with quantitatively incomparable functions, as this case.

3 PATTERN RECOGNITION SYSTEM

Eight 8 classes (n=8) of tactile pattern were chosen: circles (T_1), right triangles (T_2), equilateral triangles (T_3), isosceles triangles (T_4), pentagons (T_5), hexagons (T_6), squares (T_7) and rectangles (T_8).

Eight features (m=8), also were selected: rectangularity (X_1), squarity (X_2), see (Fig. 1a), circularity (X_3), see (Fig. 1b), number of vertex (X_4), symmetry by axis x (X_5) and y (X_6), incompatibility 1 (X_7), see (Fig. 3a) and incompatibility 2 (X_8), see (Fig. 3b). These two features are different by form of computing of border. The incompatibility INK is computed by this formula:

$$\text{INK} = \frac{\text{BOR}}{\text{AR}} \tag{6}$$

where
BOR - border length
AR - tactile pattern area

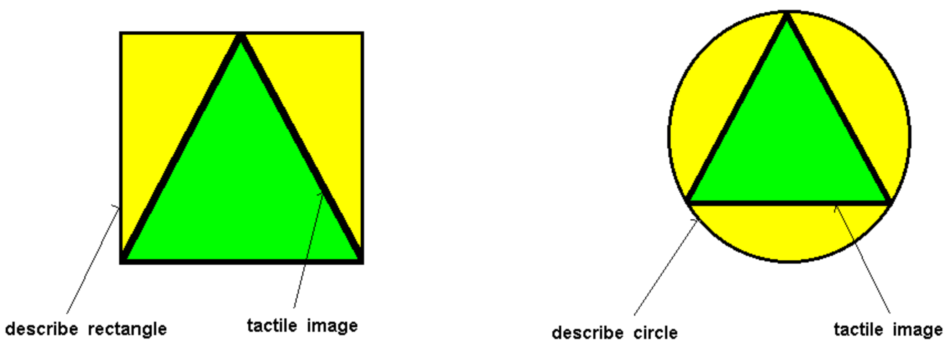

Figure 2. Example of rectangularity and circularity for a triangle

Fig. 2a shows squarity and Fig. 2b shows circularity. The squarity is defined as the quotient of tactile image area and of the described square area (in percents). The circularity and rectangularity are defined analogously.

Fig. 3a and Fig. 3b show two possible ways of computing tactile image border length (thick outline).

a) b)

Figure 3. Two methods for computing of border

Pattern recognition is resolved in two phases:

1. learning
2. self recognition

In the learning phase we will advance to system known tactile images (etalons) and upon this base we compute the matrix of etalons (for example Tab. 1 and Tab. 2) and all other statistical parameters (correlation and covariantial matrixes dimension 8x8). These computations are applicable in all cases (if we don't wish to change classes of pattern).

In the phase of pattern recognition the tactile image vector should be computed from eight features. In this way the tactile pattern is acquired. After this Ivanovic deviations of unknown pattern from etalons should be computed. By self-classification the pattern is inserted into this class, for which the Ivanovic deviation is minimal.

4 RESULTS OF PATTERN RECOGNITION

Work problem was: Verification of hypothesis that success of pattern recognition depends not only on selection and *also on sequence* of features, which input to calculation of Ivanovic deviation. The calculations verified this hypothesis. This hypothesis was verified on number of 360 000 tactile images. The system was learned by simulation data file on PC. Resolutions are showed in Tab. 4. and Tab. 5. For lucidity only indexes i of features X_i are written in first column of both Tables..

Table 4. Success of pattern recognition of tactile images (first 10 the best)

Features sequence in features vector	Success of pattern recognition of tactile images (first 10 the best)								Total success
Pattern classes	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	
[-]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
7,8,3,5,6,2,4,1	100	100	100	100	100	100	100	100	100
1,7,8,4,5,6,3,2	100	100	100	98,36	100	100	100	100	99,8
1,7,8,5,6,3,2,4	100	100	100	98,53	100	100	100	100	99,8
4,1,7,8,5,6,3,2	100	100	100	98,31	100	100	100	100	99,8
7,8,3,2,5,6,1,4	100	98,04	100	100	100	100	100	100	99,8
7,8,3,2,5,6,4,1	100	100	100	98,36	100	100	100	100	99,8
7,8,3,4,5,6,2,1	100	100	100	98,25	100	100	100	100	99,8
7,8,3,5,6,1,4,2	100	100	98,41	100	100	100	100	100	99,8
5,6,1,7,8,4,3,2	100	98,11	100	98,57	100	100	100	100	99,6
7,8,3,5,6,4,2,1	100	100	100	97,22	100	100	100	100	99,6

Table 5. Success of pattern recognition of tactile images (last 10 the worst)

Features sequence in features vector / Pattern classes	Success of pattern recognition of tactile images (last 10 the worst)								Total success
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	
[-]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]
4,1,3,2,5,6,7,8	100	95,24	43,84	94,92	100	100	100	100	90,6
7,8,1,2,3,4,5,6	100	100	38,6	81,54	100	100	100	100	90,6
7,8,4,5,6,1,2,3	100	98,25	33,82	98,36	100	100	100	100	90,6
7,8,2,1,35,6,4	100	98,28	35	85	100	100	100	100	90,2
7,8,2,4,3,1,5,6	100	98,15	35,09	83,58	100	100	100	100	90,2
7,8,4,2,3,1,5,6	100	100	43,28	80,36	100	100	100	100	90,2
2,7,8,5,6,1,4,3	100	98,36	38,81	93,65	92,73	100	100	100	90
7,8,2,3,4,1,5,6	100	98,18	38,1	80,7	100	100	100	100	89,8
2,1,7,8,34,5,6	100	100	29,51	84,62	100	100	100	100	89,4
7,8,2,3,1,5,6,4	100	100	35,29	83,87	100	100	100	100	89,2

Classes of pattern:

- T_1 circles
- T_2 right triangles
- T_3 equilateral triangles
- T_4 isosceles triangles
- T_5 pentagons
- T_6 hexagons
- T_7 squares
- T_8 rectangles

Features:

1 rectangularity
2 squarity
3 circularity
4 number of vertex
5 symmetry by axis x
6 symmetry by axis y
7 incompatibility 1
8 incompatibility 2

100% full success upon features sequence was only for this sequence: 7,8,3,5,6,2,4,1 and again the worst 89,2% success for features sequence: 7,8,2,3,1,5,6,4. Total success was in interval from 89,2 % to 100 %.

Maximal 100% probability right classification in class for all 360 000 tactile pattern was reached for four pattern classes: circles, hexagons, squares and rectangles. Average fruitfulness by classification was 99,8% for pentagons and 98,9% for right triangles. Program makes mostly of errors by classification equilateral triangles (64,4%), which wrong inserts to isosceles triangles (it's point to discussion, because every equilateral triangles is isosceles triangles, too) and by isosceles triangles (93,8%), which inserts again to equilateral triangles.

Generally the most problem by pattern classification comes on triangles classification. Here is smallest success by pattern recognition. The feature squarity is offender of this problem. By definition of this features (see Fig. 2) is not possible right to describe equilateral triangle. The rectangularity feature is next marked feature, which intervenes to right classification of triangles.

It was established that replacement of squarity feature has the most influence on success of pattern recognition and replacement of number of vertex and incompatibilities have least most influence on success of pattern recognition.

5 CONCLUSION

Work problem was the verification of hypotheses that success of pattern recognition depends on selection and also sequence of features, which input to calculation of Ivanovic deviation. The calculations verified this hypotheses.

The system was learned by simulated data file on PC AT. The results are shown in Tab. 4 and Tab. 5 for 360000 tested tactile images. The dates were also simulated by a PC. The Ivanovic deviation was used.

Total success was in interval from 89,2 % to 100 %.

It was established that replacement of squarity feature has the most influence on success of pattern recognition and replacement of number of vertex and incompatibilities have least most influence on success of pattern recognition.

ACKNOWLEDGEMENTS

This research has been supported by Czech Technical University research grant CEZ J04/98:212200008.

REFERENCES

[1] B. Ivanovic: *„Diskriminiciona analyza“.* Naucna knjiga. Beograd 1963
[2] F. Volf: *„Hodnotová analýza ve stavebnictví (Value Analysis in Civil Engineering)“.* SNTL, Praha 1982 (in Czech)
[3] F. W. Anderson: *„Vedenie v mnogorozmernyj statisticeskij analyz“.* Moskva, Fizmatgiz 1963 (translation to Russian)
[4] J. Volf: *„Classification of Tactile Images by Discrininatory Analysis“.* In: EUROSENSORS X. , Belgien, Leuven 1996, pp. 1369 - 1372
[5] J. Volf: *„Using Discrininatory Analysis II for Classification of Tactile Images“.* In: Fifth International Symposium on Measurement and Control in Robotics ISMCR’96, Belgien, Brusel 1996, pp.

AUTHOR: Jaromir VOLF, Czech Technical University, Faculty of Mechanical Engineering, Technicka 4, 166 07 Prague 6, Czech Republic, Phone +4202 24352737, Fax +4202 24310292
E-mail: volf@fsid.cvut.cz