and Cybernetics
VOL. 2 (1973) No. 1/2

# Linguistic analysis and description of two--dimensional patterns

by

M. I. IANCHEVA

Institute of Engineering Cybernetics,
Bulgarian Academy of Sciences

In dealing with the problem of diagnostic classification of the electronic schemes, this paper discusses the analysis and description of two dimensional patterns of electronic curves. Linguistic approach to pattern recognition is used for the description of this class of pictures. A method and an algorithm for the analysis are given. An example is given to illustrate the results of the algorithm.

## 1. Nomenclature

$F$ — feature space
$Z$ — pattern space
$N$ — pattern class space
$\{F\}$ — geature set, $\{F\}=\{M\} \cup \{W\}$
$\{M\}$ — subset of characteristic points
$\{W\}$ — subset of line features
$p_i$ — number of characteristic points in class $i$, $i=1, ..., N$
$M_{p_i}$ — $p$-th characteristic point of the $i$-th class
$P_j$ — point of the curve pattern
$P'_j$ — point of the discrete pattern of the curve
$\{w_q\}$ — set of labels, $q=1, ..., 8$; $\{w_q\}=A, B, C, D, E, F, G, H$
$l_{p_i}$ — length of the $p_i$-th line feature
$t_{w_q}$ — threshold in the direction $w_q$
$P_s$ — starting point of the curve pattern
$P_f$ — final point
$P_{max}$ — total maximum point
$P_{min}$ — total minimum point
$x, y$ — attributes of $\{M\}$ — members
$k$, $\operatorname{tg} \alpha$ — attribute of $\{W\}$ — members

## 2. Introduction

Pattern recognition can be interpreted as a process of mapping the pattern space $Z$ on to the pattern class space $N$. In order to reduce the required amount of information, the correlation between the two spaces is established indirectly via a feature space $\{F\}$.

In the present paper a method and an algorithm for analysis and extraction features of electronic curve visual characteristics are proposed. The method is based on linguistic theory. The descriptions of the patterns which are the results of the algorithm are to be used for automatic diagnostic classification of electronic schemes based on their analogue-type characteristics.

Many papers have been published in the last ten years, treating the problems of pattern recognition from a linguistic point of view. [1, 2, 3]. Linguistic approach in the theory of pattern recognition is characterized by:

— a formal system, consisting of a set of primitives (features) and their attributes;

— rules for composing more complex categories (phrases, sentences),

The list of rules forms the formal system of descriptive grammar. The latter together with the set of features (words) form the language in terms of which a given class of pictures can be described. The patterns are assumed to be the linguistic descriptions, thus enabling the problems of diagnostic classification to be cast into the format of the classification and identification of their descriptions. Linguistic methods are quite appropriate for the recognition and analysis of two dimensional patterns, and this is the reason for using them in solving the problems of diagnostic classification of electronic schemes.

## 3. Formulation of the problem

It is assumed that the visual patterns of curves $y_i=f_i(x)$ are given. Assuming that $N$ classes of curves are available for the recognition, it is necessary to construct the algorithm for analyzing the curve patterns and for generating the sample descriptions of all the $N$ classes.

In order to solve these two problems it is necessary to find for each curve class sample the following:

1. The set of features $\{F\}=\{M\}\cup\{W\}$ where $M=\bigcup_{i=0}^{N}\bigcup_{p=0}^{p_i} M_{p_i}$, and $W=\bigcup_{i=0}^{N}\bigcup_{p=0}^{p_i} W_{p_i}$. The number $p_i$ of the line features of the $i$-th class can be equal or not of the $p_j$ — the number of same features of $j$-th class.

2. The attributes and the attribute parameters of the features. The attribute parameters of the members of $M$ are the values of $x_i$ and $y_i$ for every $M_{p_i}$ in the Cartesian co-ordinate system. The attribute parameters of $W_{p_i}$ is the meaning of the length $k_i$, obtained as the Euclid distance between the initial and the final point of the line:

$$k_i=\sqrt{(x_{i-1}-x_i)^2+(y_{i-1}-y_i)^2}\,. \tag{1}$$



The direction is determined by the tg $\alpha_i$ where $\alpha_i$ is the angle between $W_{p_i}$ and the co-ordinate axis $x$.

3. The relationship between the feature and their attributes of diferent classes.

4. The graph-model of the curve pattern.

5. The rules of grammar for generating the linguistic description of the images.

Two properties — independence and nonredundance — are desirable in a feature set. The independence ensure that each pattern can be reconstructed by a unique combination of features. Nonredundance means that each feature is eventually utilized. Formally a set $\{F\}=\{F_1, F_2, ..., F_k\}$ of vectors is independent if for any subset $s$ and $t$ of $\{1, 2, ..., k\}$ $\sum_{s\in k} F_k=\sum_{k\in t} F_k$ implies $s=t$. (The feature set $\{F\}=\{M\}\cup\{W\}$ can be treated as a set of vectors provided that the latter are determined in $m$-dimensional space of feature attributes). A feature set $\{F\}$ of a given pattern set $Z$ is nonredundant if no proper subset of $\{F\}$ is a feature set for $Z$.

## 4. Solution of the problems

### 4.1. Basic definitions

Each image of the curves is assumed to be discretized by the step of discretisation $\Delta h$. According to this the image is regarded to be a number of points $P'_j$ which are called primitives (primitive categories). These categories are characterized by only two levels of darkness-black and white and $x_j$- and $y_j$-parameters. The black points belong to the curve, and the white do not.

The $x'$- and $y'$-parameters of the points are

$$x'_i=\frac{x_i}{\Delta h}, \quad y'_i=\frac{y_i}{\Delta h}, \tag{2}$$

where $x_i$ and $y_i$ are the parameters before the discretisation.

It is assumed that only the skeleton of the patterns (definition in [4]) is under consideration.

Definition 1. For each point $P'_j$ of the curve image there is a neighborhood of $V_r$ points, $r=1, 2, ..., 8$ (Fig. 1).

Definition 2. The sequence $r_1, r_2, ..., r_q$ is called list of directions.

The list of eight directions, illustrated in Fig. 1b is needed and sufficient for the describing of the images, provided these descriptions are used for the classification and recognition purposes (according to the classification of the descriptions of O. Firschein and M. A. Fischler in [5]).

Definition 3. The black points $P'_j$ are labeled with the labels from $\{w_q\}$ according to the direction of the line passing through this point.

Definition 4. It is said that a single line passes in a direction $w_q$ if and only if $\frac{l}{d} \geq t$, where $d$ is the width of the line in this direction.

Definition 5. The final point in the direction $w_q$ is the point labeled with the symbol $w_i$ if the other neighbouring points provided the Boolean expressions:

$$
\begin{aligned}
&P_0 \in M^A \text{ if } [(P_0 \leftarrow A)] \wedge [(P_1 \leftarrow \bar{A}) \vee \bar{P}_1] \wedge [(P_2 \leftarrow \bar{A}) \vee \bar{P}_2] \wedge [(P_8 \leftarrow \bar{A}) \vee \bar{P}_8] \\
&P_0 \in M^B \text{ if } [(P_0 \leftarrow B)] \wedge [(P_1 \leftarrow \bar{B}) \vee \bar{P}_1] \wedge [(P_2 \leftarrow \bar{B}) \vee \bar{P}_2] \wedge [(P_3 \leftarrow \bar{B}) \vee \bar{P}_3] \\
&P_0 \in M^C \text{ if } [(P_0 \leftarrow C)] \wedge [(P_2 \leftarrow \bar{C}) \vee \bar{P}_2] \wedge [(P_3 \leftarrow \bar{C}) \vee \bar{P}_3] \wedge [(P_4 \leftarrow \bar{C}) \vee \bar{P}_4] \\
&P_0 \in M^D \text{ if } [(P_0 \leftarrow D)] \wedge [(P_3 \leftarrow \bar{D}) \vee \bar{P}_3] \wedge [(P_4 \leftarrow \bar{D}) \vee \bar{P}_4] \wedge [(P_5 \leftarrow \bar{D}) \vee \bar{P}_5] \\
&P_0 \in M^E \text{ if } [(P_0 \leftarrow E)] \wedge [(P_4 \leftarrow \bar{E}) \vee \bar{P}_4] \wedge [(P_5 \leftarrow \bar{E}) \vee \bar{P}_5] \wedge [(P_6 \leftarrow \bar{E}) \vee \bar{P}_6] \\
&P_0 \in M^F \text{ if } [(P_0 \leftarrow F)] \wedge [(P_5 \leftarrow \bar{F}) \vee \bar{P}_5] \wedge [(P_6 \leftarrow \bar{F}) \vee \bar{P}_6] \wedge [(P_7 \leftarrow \bar{F}) \vee \bar{P}_7] \\
&P_0 \in M^G \text{ if } [(P_0 \leftarrow G)] \wedge [(P_6 \leftarrow \bar{G}) \vee \bar{P}_6] \wedge [(P_7 \leftarrow \bar{G}) \vee \bar{P}_7] \wedge [(P_8 \leftarrow \bar{G}) \vee \bar{P}_8] \\
&P_0 \in M^H \text{ if } [(P_0 \leftarrow H)] \wedge [(P_7 \leftarrow \bar{H}) \vee \bar{P}_7] \wedge [(P_8 \leftarrow \bar{H}) \vee \bar{P}_8] \wedge [(P_1 \leftarrow \bar{H}) \vee \bar{P}_1]
\end{aligned}
\quad (3)
$$

Where sign $\rightarrow$ means "is labeled with". Each expression, for example the first, reads as follows: $P_0 \in M$ as a final point in the direction $A$ if and only if $P$ is marked with $A$, while $P_1$, $P_2$ and $P_8$ are labeled with any other symbol from $\{w_i\}$ but not $A$, or they are the white points i.e. $\bar{P}_1, \bar{P}_2, \bar{P}_3$.

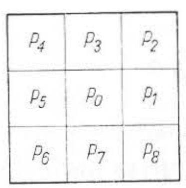


Fig. 1a. The neighbourhood of the point $P_0$

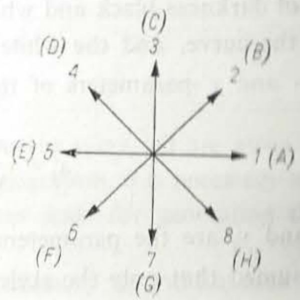


Fig. 1b. The labeling symbols and the direction of the lines

Definition 6. The set of characteristic points $M$ consists of the points $P_i'$ with more than one label and the final points in all directions.

Definition 7. Two characteristic points $M_k$ and $M_{k+1}$ are connected if there is a line connecting them, which does not pass through any other characteristic point.

Definition 8. The lines connecting points $M_i$ and $M_{i+1}$; $M_k$ and $M_{k+1}$ and all the other characteristic points are called line features of the pattern.

### 4.3. Method of analysis

4.3.1. It is assumed that there is a field of pattern on which the curve image is treated as a single pattern. This field is limited by two vertical $\Theta_1 \Theta_2$ and two horizontal lines $\Theta_3$ $\Theta_4$:

$$
\begin{aligned}
&\Theta_1 \rightarrow x = x_s', \quad \Theta_3 \rightarrow y = y_{max}', \\
&\Theta_2 \rightarrow x = x_f', \quad \Theta_4 \rightarrow y = y_{min}',
\end{aligned}
\quad (4)
$$

where $x_s'$ is the $x'$ of the $P_s'$; $x_f'$ is the $x'$ of $P_f'$; $y_{max}'$ is $y'$ of $P_{max}'$ and $y_{min}'$ is $y'$ of $P_{min}'$.

In the case when $P_{max}'$ and $P_{min}'$ do not exist, $y_{max}'$ and $y_{min}'$ are assumed to be:

$$
y_{max}' = \text{supr } y = f(x), \quad y_{min}' = \inf y = f(x). \quad (5)
$$

4.3.2. Every point $P_i'$ of the image is labeled with one of the labels $w_q$, $q = 1, 2, ..., 8$, according to the definitions 3 and 4. It is assumed that the thresholds $t_1, t_3, t_5$, and $t_7$ for the labeling with the labels $A$, $C$, $E$, and $G$ are equal to 4. For the symbols $B$, $D$, $F$ and $H$ the thresholds $t_2, t_4, t_6$ and $t_8$ equal to 2.

The curve points $P_i'$ located on the crossing of the two lines with the directions $w_q$ and $w_s$ are marked with more than one symbol, i.e.with two symbols.

4.3.3. The variations of the patterns of the same class are neglected using the operation of grouping two or more connected characteristic points in one. Let us take two characteristic points $M_i$ and $M_{i+1}$ which are at a distance $\Delta h$. In this case these two points are grouping in one point which parameters are:

$$
x_i = \frac{x_i + x_{i+1}}{2}, \quad y_i = \frac{y_i + y_{i+1}}{2}. \quad (6)
$$

#### 4.3.4. Feature extraction

a. Characteristic points. All the characteristic points, which are members of the subset $\{M\}$ of the feature set $\{F\}$ can be enumerated at random, but in our case it is assumed that the enumeration of the characteristic points begins from the starting point $P_s'$ laying on $\Theta_1$. The first point is named $M_0$. The enumeration continues with the second characteristic point with $x_i \leqslant x_{i+1}$. All the characteristic points $M_{0_i}, M_{1_i}, ..., M_{p_i}$ are listed out with their attribute parameters.

b. Line features. The line features $W_i$ is the connections lines of two point $M_{i-1}$ and $M_i$. Their parameters are: the length $k_i$ obtained from (1); and tg $\alpha_i$.

It is necessary to say that the feature subset $\{M\}$ and $\{W\}$ possess the two properties: independence and nonredundance. But the members of $\{W\}$ and their attributes are not independent from the feature subset $\{M\}$. For this reason the parameters of $W_i$ can not be calculated.

### 4.4. Graph-model of the curve pattern

The graph model of the pattern is a graph with the characteristic points as a node of the graph and the line features as arches. The nodes are marked with $M_i$. The arches are named $W_i$.

### 4.4. Grammar for generation of the descriptions

The curve pattern are very simple and for this reason the grammar is simple too. The descriptive grammar is of Chomsky-type grammars. Only two rules are sufficient for the generation of the descriptions:

the rule for phrases:

$$g_i(Ph) \to M_i(x_i, y_i)\, W_{i+1}(k_{i+1}, \mathrm{tg}\,\alpha_{i+1}), \qquad (7)$$

the rule for sentences:

$$g(S) \to \subset \bigcup_{i=0}^{p_i} g_i(Ph), \qquad (8)$$

Where the symbol $\subset \cup$ is used to denote that the pairs $M_i(x_i, y_i)$ and $W_{i+1}(k_{i+1}, \mathrm{tg}\,\alpha_{i+1})$ forming the phrases $g_i(Ph)$ have all to be connected one after another begining from $i=0$ up to $i=p_i$.

### 4.5. The descriptive language

The formal language for describing patterns (denoted as $E$) consists of the words $M_i \in \{F\}$, $W_i \in \{F\}$; the grammar rules $g_i(Ph)$, $g(S)$; the grammatical symbols, full stop, coma.

$$E \begin{cases} M = \bigcup_{i=0}^{N} \bigcup_{p=0}^{p_i} M_{p_i}, \\ W = \bigcup_{i=0}^{N} \bigcup_{p=0}^{p_i} W_{p_i}, \\ G = g_i(Ph) \cup g(S), \\ A = ./,/;/()/. \end{cases} \qquad (9)$$

### 4.6. Linguistic description of the curve samples

The total expression of the linguistic description of the curve sample is:

$$M_0(x_0, y_0)\, W_1(k_1, \mathrm{tg}\,\alpha_1);\; M_1(x_1, y_1), \ldots, M_{p_i}(x_{p_i}, y_{p_i}). \qquad (10)$$

The expression (10) can be simplified by omitting the parameters $k_i$ and $\mathrm{tg}\,\alpha_i$:

$$M_0(x_0, y_0)\, W_1;\; M_1(x_1, y_1)\, W_2; \ldots; M_{p_i}(x_{p_i}, y_{p_i}). \qquad (11)$$

## 5. The algorithm for analysis and description of the curve images

The programm operations can be outlined as follows:

Step 1. Discrete curve points $P_j'$ are applied to the input of the computer.

Step 2. The points are scanned to obtain $P_s'$, $P_f'$, $P_{max}'$, $P_{min}'$. The boundaries of the picture field $\Theta_1\, \Theta_2\, \Theta_3\, \Theta_4$ are constructed.



Step 3. The points $P_j'$ are labeled according to 4.2.2.

Step 4. The points labeled with two symbols and the final points in each direction are outlined and listed as characteristic point.

Step 5. The connected characteristic points are grouped according to 4.2.4.

Step 7. The line features are extracted from the curve patterns.

Step 8. The graph-model is constructed.

Step 9. The curve pattern description is generated, using the rules $g_i(Ph)$ and $g(S)$.

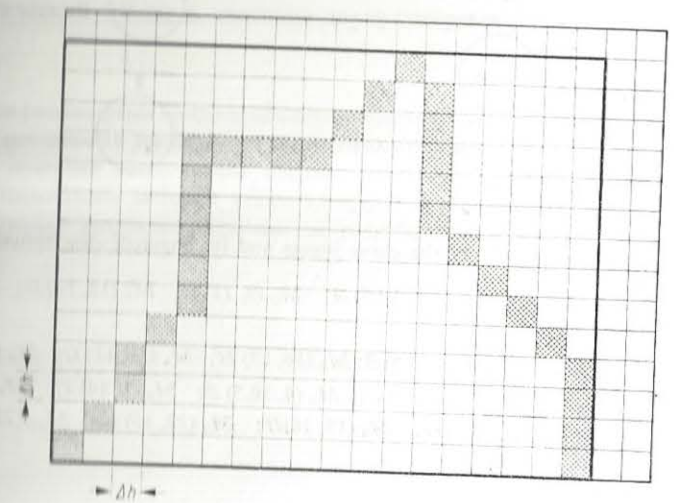


Fig. 2. The input skeleton of the curve image

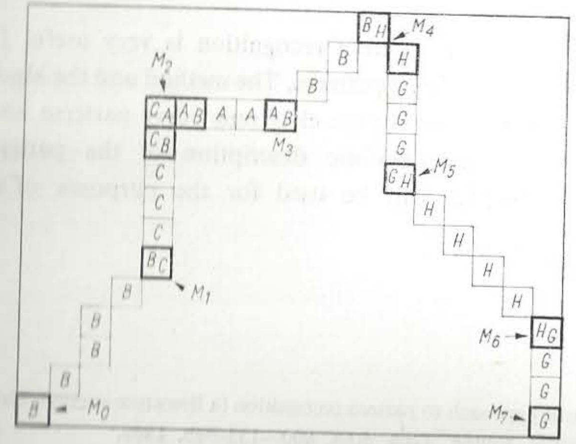


Fig. 3a. The labeled curve points; the characteristic points $M_i$

*Example.* Fig. 2 shows the input skeleton of a curve after discretisation together with the pattern field. Fig. 3a. illustrates the labeled points, obtained characteristic points laying on the crossing of two lines and the final points. Fig. 3b shows the graph-model of the pattern and its linguistic description.

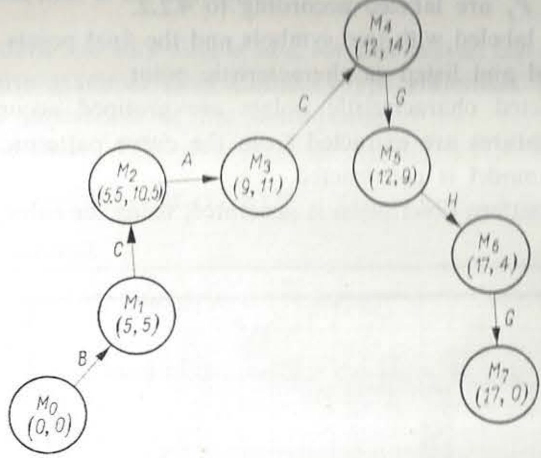


Fig. 3b. The graph model of the curve image and its linguistic description (S)

$(S) \rightarrow M_0\,(0, 0)\,B;\ M_1\,(5, 5)\,C;\ M_2\,(5.5, 10.5)\,A;\ M_3\,(9, 11)\,C;\ M_4\,(12, 14)\,G;\ M_5\,(12, 9)\,H;$
$M_6\,(17, 4)\,G;\ M_7\,(17, 0);$

$S\,(s_1) \rightarrow M_0\,(3, 2)\,B;\ M_1\,(7, 6)\,C;\ M_2\,(8, 9.5)\,B;\ M_3\,(16, 17)\,H;\ M_4\,(21, 11)\,G;\ M_5\,(21, 0);$

$S\,(s_2) \rightarrow M_0\,(0, 0)\,C;\ M_1\,(0, 4)\,B;\ M_2\,(4, 7)\,C;\ M_3\,(4, 10.5)\,B;\ M_4\,(8, 14)\,C;\ M_5\,(8.5, 10.5)\,A;$
$M_6\,(14, 20)\,B;\ M_7\,(18.5, 23.5)\,G;\ M_8\,(19, 16)\,H;\ M_9\,(22, 13)\,G;\ M_{10}\,(22.5, 7.5)\,G;$
$M_{11}\,(23, 4)\,H;\ M_{12}\,(25, 1)$

## 6. Conclusion

The linguistic approach to pattern recognition is very useful for the analysis and description of two dimensional patterns. The method and the algorithm proposed in the paper solve the problems for the electronic curve patterns and for any other class of visual signals. The linguistic description of the pattern obtained as the result of the algorithm can be used for the purposes of their diagnostic classification.

## References


1. Feder J., A linguistic approach to pattern recognition (a literature survey). New York University N.Y., Dept. of Electr. Engng. Tech. Rep. 400—133-Feb. 1969.
2. Narasimhan R., Syntax-directed interpretation of class of pictures. *Comm. ASM* 23 (1966).
3. Watanabe S., Ungrammatical grammar in pattern recognition. *Pattern Recognition* 3, 4 (1971).
4. Narasimhan R. and S. N. Reddy, A syntax-aided recognition scheme for handprinted English letters. *Pattern Recognition* 3, 4 (1971).
5. Firschein O. and M. A. Fischler, A study in descriptive representation of pictorial data. *Pattern Recognition* 4, 4 (1972).


## Analiza lingwistyczna i opis obrazów dwuwymiarowych


Omówiono zagadnienia analizy i opisu dwuwymiarowych krzywych wyświetlanych na ekranie wiążące się z problemem diagnostycznej klasyfikacji układów elektronicznych. Zagadnienie opisu tej klasy obrazów rozwiązano stosując podejście lingwistyczne. Przedstawiono metodę oraz algorytm przeprowadzenia analizy. Podano przykład ilustrujący wyniki zastosowania opracowanego algorytmu.


## Лингвистический анализ и описание двухразмерных образов


В работе рассмотрены вопросы анализа и описания двухразмерных кривых показываемых на экране, связанные с проблемой диагностической классификации электронных схем.

Задача описания этого класса образов решена при использовании лингвистического подхода. Представлен метод а также алгоритм проведения анализа. Приведен опример иллюстрирующий результаты применения разработанного алгоритма.