## PATIERN LINGUISTIC ANALYSIS INVARIANT FOR PLANK TRANSFORNATIONS

J. 0. Simon
and
A. Checroun

Institut de Programmation - 9, quai Saint Bernard PARIS 5e

Pattern analysj s using syntactic methods is discussed briefly to introduce the ideas of the linguistic nature of the attributes and of possible invariance properties versus usual geometrica] t ran sformations.

A linguistic operator invariant for projecti ons from plane to plane is proposed and applied to the analysis of point patterns X. From knowled ge of the operator "result" or "attribute", it is possible to obtain the convex envelope, reconstitute $X$, compare $X$ to another pattern $X^{\prime}$ and more generally, to obtain the possible common subpat terns. Thi s last process showy that a threshold exists, under which a pattern comparison is not reliable.

Linguistic operators invari ant tor plane similarities are also consi dered.

## §1. INTRODUCTION

## J.I. FORMALZATION OF PATTERN RECOGNITION OPERATIONS

Except for some very simple recognitions, many "recogni ti on 1 eve] r," have $i o$ be di st inguished in a pattern recognition problem.

Recognition operations have to be performed to get from one level to the next. Most of the time, they may be described under a common formaliza tion :
a. Let a pattern $X$ he a set of "primitive" patterns, each of which has a name to which are associated numerical values-
b. One or more anatysis operators to operate on $X$ through algorithms ; the result of \& is $u_{x}$.
c. A recognition decision is performed by comparing $u_{\times} w i$ th informat ion previously stored. If this phase is successful, a pattern "feature" is extracted. This feature has EL name ; nunieri cal values may be associated with it.

The recogni $z$ ed features are the rew primitive
patterns. Their set is the pattern on which the operators of the next level will operate.

This scheme is valid obviously for the first $J$ evel extract ion of simple features where $X$ is the set of samples obtai ned $t$ hrough a measuring instru ment such as an optical or acoustical device. Most of the time, the operators $\&$ are linear operators. For example, let $X$ be a one dimensional "signal", $i$ is the rame of the primitive measure or sample, $x$ the munerical value associated with it.

Let $\ell_{\mathrm{p}}$ be the operator such that the result


The operators \&p are defined by the set of numerical values ( $\alpha_{1}, \alpha_{2} . . \alpha_{q}$ ). The result $u^{p}$ is the numerical value obtainea by a translation of $p$ units and a scalar multiplication. Digital filters, neurone nets and Perceptrons male use of such linear operators.

The recogniti on decision is taken according to the numerical value of $u_{A}^{P} \times$. Usually if $u_{x}{ }_{P}$ is greater than a threshold $u_{0}{ }^{A}$, the feature corres ponding to $£$ is recognized.

Such a process is easily extended to images, i.e. two dimensi onal signals.

## T.2. SYNTACTIC OR LINGUISTIC OPBATORS

It has been realized for quite a time that li near operators alone are not able to take into ac count structural relationships in patterns-

At the beginning of the 1960's a number of au thors such as Eden (1), (2), Minsky (5), Naranimhan (6), Kirsch (4), advocated a "description-proees sing" of pattern, thus starting a new "syntactic" approach to pattern recognition problems of higher complexity. Quoting FU and SWAN (3), who recently reviewed the state of the art :
"Syntactic pattern recogni ti on is an attempt to adapt the techniques of formal language theory which provides both a notation (grammar) and an analysis mechanism (parsing) to the problem of re presenting and analysing patterns... Syntactic analysis can proceed only if a grammatical model for the data generation and / or analysis process can be formulated (This does not necessarily im ply that the data is actually generated by a mechanism which operates in the same way as the model). A related practical problem is the development of efficient analys is procedures based on the grammati cal model. The available ljtteraturc concerning syntactic pattern recognition deals al most exclusively with the formulation of grammati cal models, although the development of anaiysi s algorithms is currently receiving more attention.."

This new approach is a subject of consi derable interest among pattern recognition specialists. Nevertheless, the practjcal results in the analysis field do not seem up to the expectatlons-

In fact syntactic techniques are successful, says Kirsch, only if the generati ve grammar is fully known. Rut, as for natural languages, there are some doubts about the existence of such "CHOMSKY-like" grammars for images, even with the improvements provided by the "web grammars" propo sed by Pfaltz and Rosenfeld (7).

We would like to retain two features of the syntactic approach :

- Give up the interpretation of the processed quantities as numeri cal values, consider them rather as words of a language with which nume ri cal values may be associated.
- Try to find "rules" valid for a large class of patterns.
Accordingly, coming, back to our formalism, WE propose to consider analysis operators \& of a lin guistic and invariant nature :
- The result $u_{x}$ is to be one or more words "built on the alphabet consisting of the names of the primitive patterns. For this purpose a formal generative grammar is not absolutely necessary, as long as an algorithm giving an appropriate result can be found. This algorithm, called the analysis operator, should be implemented, of course, by a computer program. The pattern X will be the variable data set of this program.
- The result $u_{x}$ is to be invariant under some transformations of the set $X$, corresponding to some geometrical transformations of the analysed object. This "invariance concept" underlies the syntactic pattern methods, though previous authors do not seen to emphasize its relevance.
The recognition decision will be taken by com paring the $u_{x}$ with some stored information (1.1. phase $c$ ). We will see later examples of how this phase can be implemented.


### 1.3. REMAFKS

1. Our formalism can be compared witn the one pro posed by Fu and Swain (7)- For them a Generalized Syntactic Formalism (GGF) utilizes :

- A Generalized Syntactic Element (GSE), which is a construct consisting of two parts : a na me and an attribute list, which is a set of modifiers or variable properties.
- A Generalized Syntactic Production (GSP), which is a rule specifying how a syntacticelement is composed from or generates other syntactic elements. Usually the syntactic elements generated are "more primitives".

Thus, a set $X$ of primitive patterns can be compared to a GSE, an analysis operator \& to a GSP with the difference that now the results are "Jess primitive" elements, "result" and attribute" being equivalent words.

Our terminology seems more general, being also appropriate for the filtering or sifting processes. One could in fact assert that in terms of the computer states there is no basic difference, only the semantic differs.

To stress the fact that no generative grammar is implied by our operators, we would prefer to call them "linguistic" rather than syntactic.
2. An instrument provides numerical results or samples ; a name is given to each sample and a number of numerical values. For example, in a reti na the cell of name i gives an intensity of numeri cal value $z_{1}$. The coordinates of cell $i$ are the nu merical values $x ., y .$. Thus an "elementary measure" is defined by i, x., y., z.. The set of these measures is the first pattern, each "elementary measure" being a "primitive pattern". This level will be called the first level. It is already a "feature level", if we consider that the "elementary measures" are the "results" of some "physical
operators". But usually we cannot modify the qualities of these physical operators. This property could be used as a way to define the "outer world or universe ...

The second level is reached, most of the ti me by the use of linear operators or filters. The se simple operators have no invariance properties They are thus successful only for invariant sub patterns. Elements of lines, strokes or line cros sings are examples of such subpatterns invariant for many transformations ; thus may be explained the limited success of "Perceptron-like machines" for complex patterns, composed of these simple features.
3. The object of this paper is to propose some "invariant linguistic operators" usually effective for reaching the next level. They apply to two dimensional signals or images. The envisioned geo metrical transformations are plane to plane trans formations. One is for plane projections, the other for plane similarities.

## §TI. LINGUISTIC ANALYSIS OPARATORS INVARIANT FOR PLANE PROJECTIONS

### 11.1. DEFINITIONS

Let the pattern be composed of "primitives" of names A0, A1, A2, ,.. A .... An, to which are
associated numerical values ( $x 0, y 0$ ) , ...( $x ., y.) \ldots$ These primitives may be the "features" of the second level. The numerical values x., y. are the
coordinates of the "center" of features A.. Let these "points A." be in a plane II.

Let us establish one of them as privileged, for example. A. A0 defines a direction. Starting from this direction, let us rotate a vector around $A$. in a given manner. This vector will encounter the other points $A_{.}$in a certain order.

Noting this order we can obtain an n-word on the alphabet $\{A-\}$ of the names, starting by A. :

$$
u_{1}=A_{i}, A_{j_{1}}, A_{j_{2}}, \ldots A_{j_{n}-1}
$$

From all the points except $A$, a word such as $u_{i}$ may be obtained. Let us name $X_{x}$ this set of n words of n letters.

A projection from plane II to plane II' transforms the word A0, A1, A2... A into A' $A^{\prime} A,^{\prime},, .$. A'n. It is easily verified that the set of $n$ words $U^{\prime}$ obtained, from the new pattern $X^{\prime}$ by the same A
algorithm, is derived from $U_{x}$ by substituting
A
A! for A. .
1
Thus the operator $£$ implemented by the propo sed algorithm is invariant for a plane projection it is easy to verify that a plane similarity, a rotation for example, does not modify the $n$ words

## Remarks

1. Let the projection from plane to plane be a "parallel projection", i.e. the center of projec tion be at infinity. Let $G$ be the center of gra vity of the primitive points A., which may be weighted by numerical values $z$..

$$
\sum_{i} z_{i} \overrightarrow{G A} A_{i}=0
$$

$G^{\prime}$, the projection of $G$, is still the center of gravity of the projected points. The quality is invariant for the proposed transformation. Thus for these parallel projections, $G$ may be selected as the "privileged $n+1$ point" $A$ of an n-points picture.
. Let $A_{o}$ be at infinity. This means that the starting direction of the rotating analysis vector is the same for all the $n$ points A.. Let us consider the projections from one plane to another kee ping at infinity the projection of $A$. In other words, this means that $A_{o}$ is the point at infinity on the intersection line of the two considered pla nes. For the new plane, the starting direction of the rotating analysis vector is the same for all projected points. It is the projection of the star ting direction in the former plane. We will ;res trict ourselves to such transformations in the pro posed application examples.

### 11.2. PRACTICAL APPLICATION TO IMAGES

Let us take for the common starting direction of the rotating vector, the direction of the $x$ axes (horizontal right). The rotation sense is taken counter clockwise.

Let $\left(A_{1}, A_{2}, \ldots A_{n}\right\}$ be the alphabet of names of the primitives. The partial result, obtained from pojnt $A$. will be written down as :

n words such as u form an $\mathrm{n} \times \mathrm{n}$ matrix : the "result $U_{x}$ ". This; result is invariant for the envisioned plane projections.

Ary monotone increasing function of the vector angle results in obtaining the $u$.. Let $q$ be the size of the computer register in which is stored the angle of $A$. A. with the horizontal axes. Let $x$ and $y$ be the A. A. projections on the horizontal and vertical axes. An easily computed two's complement function $\varphi$ (A. A ) is :

$$
\begin{aligned}
& 0, \frac{\pi}{2} \quad \ldots\left|z+\frac{\frac{i}{|y|}}{|x|+|y|}\right| 2^{q-2} \\
& \frac{\pi}{2}, \pi \\
& \ldots\left|3+\frac{|y|}{|x|+|y|}\right| 2^{q-2}
\end{aligned}
$$

$$
\begin{aligned}
& \pi, \frac{3 \pi}{2} \ldots\left[\frac{|y|}{|x|+|y|}\right] 2^{q-2} \\
& \frac{3 \pi}{2}, 2 \pi
\end{aligned} \ldots\left[1+\frac{|x|}{|x|+|y|}\right] 2^{q-2} .
$$

Such a function covers the available computer numerical range. It orders the angles with a very high discrimination.
N. B. This precision may be superfluous ; as will be seen later, it may be useful to decide that two points $A_{k}, A$ are seen "in line" from a point $A_{1}$, in other words that the angles of vectors $A$. A, and A. A with a fixed direction are the same, This is achieved by reducing the function pre ci si on.

The figures $1-X$ and $1-Y$ display images $X$ and $Y$ of 8 primitive points. A, B,... H. The corres ponding results are given by Fig. $2-X$ and $2-Y$.

The $Y$ image has points "in line". This is recorded in the $8 \times 8$ matrix by setting a number after the letter. For instance $F$ and $C$ are in line with $A$, also $H$ and $F$; $C$ and $A$ are in line with $E, F$ and $A$ with $H$, cf. Fig. 2-Y.


Fig. 1-Y

| $A$ | $B$ | $G$ | $E$ | $C$ | HIF |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $H$ | $A$ | $G$ | $E$ | $H$ | $F$ | $C$ |
| $D$ | $D$ |  |  |  |  |  |
| $D$ | $A$ | $H$ | $C$ | $E$ | $G$ | $F$ |
| $H$ |  |  |  |  |  |  |

Fig．2－Y

## Remarks

1．A change in the common angular origin of the rotating analysis vector usually transform $\times U$ ：a circular permutation of the $n-1$ letters in a li ne after the first may occur．

An image rotation in the plane would introdu－ ce such permutations．The considered operator is not rotation invariant．Why then use such an ope－ rator rather than the more general one introduced？ Because，by comparing two unknown figures，the cor respondence between the points $A_{0}$ has to be known． In general this is not feasible．

The proposition of taking the center of gravi ty as the point $A_{o}$ is quite interesting，but it restricts the plane projections to parallel pro－ jections．On the other hand by using the parallel origin direction vector we will be able to compa－ re figures with different number of primitive points and also to obtain the convex envelope． This would not be possible by using the＂center of gravity＂origin method．Anyway，the invariance to rotation may be of real value for some problems and this process is worth studying also．

2．If $A_{o}$ is at infinity，the image can be scanned from $A$ ，in the same sense that the rotating vec－ tors＇sense．Here this scanning is made by a hori zontal line，the ordinate of which decreases from $+\infty$ to $-\infty$ ．Thus the $n$ primitive points are put in a certain order．We adopt this order for the first column of the $n X n$ matrix $U_{x}$ ，which is then uni－ quely defined ；cf．Fig．2－X and $2-Y$ ．

## II．3．SOME PROPERTIES DEUOCD RROM $U_{x}$

The $U_{x}$ matrix，invariant for plane projections A
preserving the horizontal direction，is used to re cognize large pattern classes．

Already some properties of the analysed image may be deduced from UX alone．For instance it is possible to state the sense of any three pointed triangle or to decide if any four pointed quadri－ lateral is convex or not

## The convex envelope

A most interesting property is the ability to find the convex envelope，i．e．the points on
which a convex polygon can be built such that all the other image points are inside this polygon． Let two points $A_{k} A_{e}$ be such that ：
a．In the line starting at $A$ ，called the $A_{k}$ line the point immediately after $A_{k}$ is $A_{e}$ ．
b．The $A_{e}$ line terminates with $A_{k}$ ．
It is easy to demonstrate that $A_{k}$ and $A_{e}$ are
elements of the convex envelope，and that $A$ imme diately follows Ak on the envelope for the adop－ ted rotating sense．

The next point $A$ on the envelope is such that it is the point following $A$ ，in the $A$ line． if $A_{k}$ is the last point in the $A^{k}$ line，$A^{e}$ is the second＇circular permutation property）．

All the envelope points are obtained through this algorithm，continuing so as to find again the first point A．

For instance，going back to the eight primiti ve points image of Fig．1－X，the $U_{x}$ of which is displayed in Fig．2－X，it is very easy to obtain any $U$ corresponding to any partial image，dedu－ ced from the original image by suppressing a num－ ber of points．Fig． 3 gives the envelopes of ima－ ges deduced from Fig．1－X．One line gives the points，the line underneath the convex envelope points．For the total image $A B \ldots \mathrm{H}$ ，the envelo－ peisCAEG11F．

$$
\begin{aligned}
& 0.1 .4
\end{aligned}
$$

$$
\begin{aligned}
& \text { D) } 11 \\
& \text { * NよLHPが Fもいト } \\
& \text { Ci At "pto, } \\
& \text { : H.VELUPM C.AHoty } \\
& \text { Hot AC, } \\
& \text { : NVFllop: CAlbtit } \\
& \text { Eralit } \\
& \text { - wVFLIPP! farit } \\
& 6.1 \\
& \text { : ivprglippr Fl.t }
\end{aligned}
$$

Fig． 3
Selected points of Fig．1－X，followed by the con－ vex envelope points．
Rebuilding the image from $U_{x}$
A matrix $U_{x}$ ，being the result of the analysis operator $\&, s$ it possible to obtain again the primitive image $X$ from $U_{X}$ ？

Let us recall that the interest of $U_{X}$ is not A an information reduction from the primitive infor mation of one image，but that a $U_{x}$ is the same for a class of patterns，modulo a＂geometrical＂ transformation $Q$ ，such as has been envisioned．

Even more，some image＂deformations＂do not modify $U^{\wedge}$－Let us consider a pattern $X$ of $n$
points, the $n(n-1)$ half-lines joining two points A., A., the n horizontal half-lines joining a point $A$. with the point $A_{0}$ at infinity. If none of these $n$ half lines crosses an image point $A_{k}$ during a "deformation" transform $\Delta$, obviously the result UX will be unchanged. Such a deformation $A$ may alter the primitive pattern shape considerably. Thus any image rebuilt from UX mey differ from the original image $X$ not only by a transformation $\Theta$, but also by a deformation $A$.

The convex envelope of any subpattern may be obtained from UX. This facility is used to rebuild A
a pattern $X$ ' from $U$. The envelope points are put evenly on a circle in their order. The matrix UX
indicates the regions, defined by a number of straight half-lines, where the other points should be. The set of these other points is ordered in order of decreasing freedom. The first points in this order, thus those with the minimum number of constraints, are put first at the barycenter of the surface where they should be set according to UX. An impossibility may occur before the set is exhausted. The process is then started again after a random variation of the positions of the points.
 $r$

Though we are quite convinced that this is not the best idea, successive iterations of this process usually increases the number of points that can be disposed of according to $U_{Y}$.


Fig. 5'- Reconstitution of figure 5 rotated of $90^{\circ}$

Fig. 4' shows the reconstirtution of the pattern in Fig. 4. All but point $F$ are set of a total of thirteen. The convex envelope is A C J M I B. The matrix UX does not give any indication of the envelope Bhape, the ignorance of which leads then to an important deformation of the two patterns.

Fig. 5( gives the reconstruction of Fig. 5 pat tern from $U$. Out of 38 points, only 30 can be set. This number seems a practical limit.

Though it is interesting to try to rebuild a pattern from $U$, it is not at all the aim of the process, which is principally for pattern identifications. Nevertheless it shows how different two patterns with the same operator attribute UX may be.

## II.4. PATTERN IDENTIFICATION RROM UX

A
The identification of two patterns X and Y , having the same number $n$ of points, is quite simple.

Let

$$
\begin{aligned}
& A_{1} A_{2} \ldots A_{n} \text { be the "points" of } X, \\
& B_{1}, B_{2} \ldots B_{n} \text { be the "points" of } Y .
\end{aligned}
$$

The two patterns $X$ and $Y$ are "equivalent",i-e. may be deduced one from the other by a transformatione and eventually a deformation A , if there exists a bijective application $\quad \mathrm{T}: \mathrm{A} . \mathrm{B}_{\mathrm{J}}$
such that UX is transformed in UY), and inversely : $1^{T-1}: B . \rightarrow A$.

The application is defined readily by comparison of the first columns of $U X$ and $U$. Then the verification of correspondence is made on the matrix lines. For instance, let $X$ and $Y$ be the pat terns of Fig. 1-X and 1-Y.

The application T and T are :

$$
\begin{aligned}
& \text { re } \Gamma^{-1} \quad A \quad C \quad B \quad D \quad F \quad E \quad G \quad H \text { (Pat. X) } \\
& A \quad B \quad D \quad C \quad F \quad E \quad G \quad H \text { (Pat. Y) } \\
& \text { The second lines of } U_{X} \text { and } U_{Y} \text { are : }
\end{aligned}
$$

Using the defined T , a complete correspondence cannot be established between these two lines. We decide that these patterns are not equivalent.

If points are "in line", they may be exchanged during the identification operation.

The discrimination between two patterns of $n$ points is quite high. It depends of course on the angular precision, which may be reduced : two points $A$., $A$. of a pattern are "in line" from a point $A k$ if the angle (A. A , A. A ) is smaller than $€$ in absolute value. Thus classes are defined in the A. line. Points may be interchanged inside a class. A similar process has been tested for the distance operator (cf. paragraph III and
J. C. Simon, A. Checroun and C. Roche (9), (10)) we call it "parenthesizing".

## Finding common equivalent subpatterns

Let $X$ and $Y$ be two patterns, $|X|$ and $|Y|$ be the number of primitive points. The analysis operator \& provides $U_{X}$ and $U_{Y}$. From the knowlenge of $U_{X}$ and $U_{Y}$ is it possible to find two subpat-
terns $Z$ and $Z^{\prime}$ such that $Z \subseteq X$ and $Z \subseteq Y$, $|Z|=|Z|$, and $\Gamma: U_{Z} \rightarrow U_{Z}, \Gamma^{-1}: U_{7,1} \rightarrow U_{Z} ?$ ©n other words, by suppressing a number of points in each pattern is it possible to find two subpat terns satisfying the former correspondence algorithm.?
$F$

C
$H$
A
1
10




1
$\lessdot$
$t$
11

Fig. 6-X
Fig. 6-Y
Let us show the process on an example. Let $X$ be a pattern of 6 points, Fig. $6-X$; $Y$ be a pattern of 8 points, Fig. 6-Y. Let us examine first if $X$ can be found entirely embedded in $Y$, (thus $X=Z$ ).

Refering to UX and UY, the correspondence of A $\quad 1 \begin{aligned} & 1 \\ & \text { (*) For }\end{aligned}$ certain points is obviously impossible (*). For example, $F_{x}, D_{y}$.

Let us test the possible $Y$ points corresponding to the point $C$. From the first column, only A , B , CY may correspond. We shall see that $B_{y}$ cannot be e CX correspondent.

If $T: C \rightarrow B, B$ should be found in all five rectangular domains of UY displayed below : the B position in UY has to be close to the $C$ position in $U$. In the chosen example $B$ is only in four do A mains and tnus $B$ cannot be a $C$ correspondent by
(*) When there is some doubt about the points' origin, an index $X$ or $Y$ will state the origin from the $X$ or $Y$ pattern.

$$
\begin{aligned}
& \mu \rightarrow F \underset{C-1}{C-1} A \quad E \quad B \quad D \\
& t
\end{aligned}
$$

$$
\begin{aligned}
& \mathrm{i} \rightarrow \mathrm{C} \quad \mathrm{~F} \quad \mathrm{~A} \quad \mathrm{~F} \quad \mathrm{~B} \quad \mathrm{D} \quad \mathbf{k} \rightarrow \mathrm{~B} \quad \mathrm{G} \quad \mathrm{~A} \quad \mathrm{D} \\
& \text { B } F D A \quad E \quad D \\
& \begin{array}{llll}
\mathrm{E} & \mathrm{~B} & \mathrm{~F} & \mathrm{C}
\end{array} \mathrm{~A} \quad \mathrm{D} \\
& \text { 1) } F \quad C \text { B } E \quad A \\
& \begin{array}{llllll}
\mathrm{A} & \mathrm{D} & \mathrm{E} & \mathrm{~B} & \mathrm{~F} & \mathrm{C} \\
\hline
\end{array} \\
& \mathrm{U}_{\mathrm{X}} \\
& \mathrm{U}_{\mathrm{Y}}
\end{aligned}
$$

Let us expmine to see if $C$ may be a $B$ correspondent by $\Gamma^{-1}: B \rightarrow C$; the point $C$ should be in five of the seven rectangular domains. As the point $C$ is in six of these domains, $B$ may be a $\Gamma^{-\dagger}$ correspondent of $C$.

$U_{X}$
example, let us give the general $U_{Y}(k, 1)$ be a letter of $U_{Y}$.

1. First column correspondence

$$
\text { If } \quad U_{X}(i, 1) \rightarrow U_{Y}(k, 1)
$$

$\operatorname{Sup}(1,|Z|-|X|+i) \leqslant k \inf (|Y|,|Y|-|Z|+i)$
$(|Z|$ is the number of points of the possible par tial pattern).
2. An $X$ point exists also in other columns of $U_{X}$,
let $\mu$ and $v$ be its coordinates, such that
$U_{X}(i, 1)=U_{X}(\mu, v)$. Each point. $U_{X}(\mu, v)$ determi -
nes a rectangular domain in $U_{Y}$, defined by $t$ and $s$ The conditions on $t$ are :

$$
\begin{aligned}
\mu<i, \text { then } & \sup (1,|Z|-|X|+\mu)<t \\
& \inf (k-1,|Y|-|Z|+\mu) \geqslant t \\
\mu>i, \text { then } & \sup (k+1,|Z|-|X|+\mu)<t \\
& \inf (|Y|,|Y|-|Z|+\mu) \geqslant t
\end{aligned}
$$

And s must satisfy :
$\sup (2,|Z|-|X|+v) \leqslant s \leqslant \inf (|Y|,|Y|-|Z|+v)$

$$
|X|-1 \text { such domains are defined in } U_{Y}
$$

$\Gamma: U_{X}(i, 1) \rightarrow U_{Y}(k, 1)$ is possible if $U_{Y}(k, 1)$
is at least in $|Z|-1$ of these domains.

Correspondence matrices MZ can then be obtained. Let us represent them for the example. A 0 signifies that at least one of the correspondences is impossible, a"'that both are possible according to the preceding rules.

$$
\begin{aligned}
& M_{z} \text { for }|z|=6
\end{aligned}
$$

$M_{Z}$ shows that the correspondence is not possible for any $Z, Z^{\prime}$ such that $|Z|=6$.


Many sets of 5 points in $X$ and $Y$ are candidates for correspondence. Nevertheless this correspondence has to be verified also by the algorithm described for two patterns having the same number of points.

For instance F C 13 D A of pattern $X$ is equiva lent to the $Y$ subpatterns: $G B E C D, G A E C D$, GBEHB, GAEHD, BEHCF, GBHCD, GAHCD, GBEFD.

Of course the process may be applied to a pat tern $X$ itself to find all the equivalent subpatterns $Z$ C X. For example in the pattern $X$ of
Fig. 6-X, F C B D A and FCEI)A are equivalent.
Let $X$ and $Y$ be two patterns, chosen at random but having a fixed number of points : $|x|=n_{1}$, $|Y|=n_{2}$ It is interesting to use the process to find the possible equivalent subpatterns. A sharp threshold in the number $z$ of equivalent subpat tern points is found experimentally. For instance, let $\mathrm{n}=6, \mathrm{n}=8$. Table l gives the test re -
suits ; the first line gives the number $z$ of sub pattern points.

The five other lines give the number of equivalent $z$ subpatterns for five different couples of random patterns $\mathrm{X}, \mathrm{Y}$.

| 6 | 5 | 4 | 3 | 2 | 1 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | 0 | 120 | 572 | 420 | 48 |
| 0 | 5 | 144 | 568 | 420 | 48 |
| 0 | 21 | 249 | 544 | 420 | 48 |
| 0 | 32 | 266 | 536 | 420 | 48 |
| 0 | 18 | 180 | 573 | 420 | 48 |

$$
\text { Table } I, n_{1}=0, n_{2}=8
$$

If $n_{1}=n_{2}=10$, out of 8 couples of random $\mathrm{X}, \mathrm{Y}, \mathrm{no}$ subpattern is found for $z \geqslant 8$. If $z=7$, subpatterns are found only for 3 couples out of 8. If $z \leqslant 7$, many subpatterns are always found.

An approximate value of $\bar{z}$, average of $z$, is given by

$$
z=\left[\begin{array}{l}
|X| \\
|Z|
\end{array}\right] \times\left[\begin{array}{l}
|Y| \\
|Y|
\end{array}\right] /(|Z|-1)!
$$

This has to be compared to what we found, by linear filtering of sampled noisy images, cf. J. C. Siman and J. Camillerapp (8). A sharp threshold exists under which one may consider that the examined pattern is "noise".

## § J I I A UNGUSTIC ANALYSS GFMAVR NNARANT ÆR SIMILARTIES

## J 11.1. DEFNTON

Let $A, A \ldots A$ be the $n$ "points" of a pat _ tern $X$ as defined in $\S 11.1$. Fom one of these, $A$. for- instance, an order on the remaining $n-1$ points may be defined by the euclidian distances between A. and other points A..

Let $v$. be the word obtained thus :

$$
v_{1}=A_{i}, A_{j}, A_{j}, \cdots A_{j_{n}}
$$

Let $v_{x}$ be the set of the $n$ words $v .$. This set may be ordered and be represented by an $\mathrm{n} \times \mathrm{n}$ matrix as before. To introduce an order on the points A themselves may processes may be proposed. For instance let $G$ be the center of gravity, the points A. may be set in the order of their increasing euclidian distance from $G$. This order, which then defines first column of the matrix V , is independant of similarities, as the word $v$. itself. But, if certain points are missing, this order may be altered by displacement of the center of gravity. Another useful order on the points A. is obtained by sweeping a line parallel tu a direc tion ; but then a rotation would induce a circular
permutation on the lines of thematrix $V_{x}$
I 1.2. CAMPAFBN WIH A ÆRMRY RROPAD MEFDD
This operator has to be compared to what already has been proposed, cf. J. C. Siman and al. (9), (10).

In that method all the distances betweeen two points of a pattern $X$ were ordered giving a unique ordered set $D$. The reader is refered to the publication (10) for a detailed description of the comparison between two patterns of the same number of points.

By defining a variable distance precision e, classes are introduced on the ordered set D. They are represented by "parenthesis". During the assignment algorithm between two sets 1) and ]) ${ }^{f}$, letters can be taken indifferently inside a class. This method, called "parenthesizing", allows same flexibility in the compari son between two patterns. Let us summaize the results :

If $e=0$ the assignment is generally impossible ; if e is large enough, the assi gnmerit is undetermined. Tho situations may arise :


In situation (a), the two patterns are said to be comparable to the precision e .

In situation (b), the two patterns are said to be different.
$e_{1}$ is of an order of magnitude smaller than $e_{2}$ and $e_{3}$, which are of the sare order of magni tude. Again a threshold is found, above which a decision cannot be taken with security.

The "parenthesizing process" can also be introduced for the rew operators proposed in this paper. It would introduce sore degree of freedom i the comparison, often necessary from the very nature of the pattern "points". The proposed methods are useful after the first level recogrii tion ; then a "point" represents in fart the center of gravity of a detected characteristic feature, thus it may vary to a certain extent.

A drambadk of the first proposed method, cf. (10), is that the compari son between two pattems of di fferent numbers of points is not practical. Again, according to the nature of the "points", it seams likely that two compared patterns may have a different number of points.

We believe that the optimum solution will be found by a combination of the two ideas : comparison of two patterns having a different number of points and variable precision, thus parenthe sizing. The first performed experiments shaw that a threshold exists and that it is a function of
the angular or distance precision $e$ and of the number $z$ of the possible cormmon subpattern. Under this threshold, comparison is not performed with security.

A pattern $X$ being given and the ordered set $D$ obtained, it is easy to deduce $v$ from $D$. The A
knowledge of $v_{y}$ alone does not allow obtaining $D$. Nevertheless ix is easy to supplement $v_{x}$ for that purpose.

The reconstruction of a pattern from the know ledge of $D$ or $v x$ is possible and easier than the reconstruction performed from the $\S 11 u_{x}$.

The pattern is obtained modulo a possible similarity. Some deformations may also be introdu ced ; these deformations are usually much smaller than for the distance operators.

From the $V_{X}$ results the search for common sub patterns may be performed with algorithms similar to those described in §11.U. They will be published later.

## §TV. CONCLUSIONS

A generalisation of the analysis operators may be proposed. Let $X$ be a "pattern" made of $n$ "elementary patterns",

$$
X=\left\{\begin{array}{cccccc}
A_{1} & \ldots & \ldots & A_{i} & \ldots & \ldots \\
A_{n} \\
x_{1} & \ldots & \ldots & x_{i} & \ldots & \ldots \\
x_{n} \\
y_{1} & \ldots & \ldots & y_{i} & \ldots & \ldots \\
y_{n} & y_{n} \\
z_{1} & \ldots & \cdots & z_{i} & \ldots & \ldots \\
y_{n} \\
\cdots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{array}\right\}
$$

To an elementary pattern of name A., qualificatives $x_{1}, v_{1}, z_{1} \ldots$ may be associated. The analysis operator $£$ delivers $n$ attributes $u$. from $X$. They are obtained with a real function $\varphi$ (A. A.) defined with the $A$. and $A$. qualificatives. The i $J \quad i$ n-1 elementary nattorn namac aro ardarad in $a$ way such that if $\varphi\left(A_{j} \mid A_{i}\right)<\varphi\left(A_{k} \mid A_{i}\right), A_{j}$ is before $A$ in the word $u$. beginning by $A$.. A word $u_{1}$ defines an "order from $A_{f}$.". Orderines the ele menLary patterns names by some other way, an $n \times n$ matrix $U$ is finally obtained. This matrix is usually invariant for many geometrical transforma tions of the image.

Comparison between two patterns $X$ and $Y$ is performed using the matrices $U$ and $U$.

$$
A \quad X
$$

## Prospective remarks

(1) Fiom the wards u., partial order is indured on the set of the elementary pattern couples \{A. A.\}. From this order a "clusterinc hierar chy' in the sense of Jardine or Benzeai is obtained. This hierarchy mey be useful for comparing patterns.
(2) Many $\varphi$ (A. | A.) functions may be used, particularly distances, satisfying the usual distance criteria. For instance, if the $A$.
patterns have messes $\mathrm{m}_{\mathrm{j}}$, then
$d\left(A_{i}, A_{j}\right)=\frac{d^{2}{ }_{i j}}{m_{i} m_{j}}$; if the $A_{i}$ are straight
lines $l_{i}$ of relative angle $\alpha_{i j}$, then
$d\left(A_{i}, A_{j}\right)=\frac{d_{i j}^{2}}{I_{i} l_{j} \cos ^{2} \alpha_{i j}}$ etc...

## Acknowledgement

The authors wish to acknowledge the participa tion of graduate students : Andre, Galland, Gelly, Renard.

## RGTRENCES

(1) M. Eden "On the formalization of Handwri ting". Proceeding of Symposia in Applied math, Jj2, pp. 83-88, Am. Math. Soc., Providence, R. I., 1961.
(2) M. Eden : "Handwriting and Pattern Recogni tion". IRE Trans, on Inf. Theory, 8. pp. 160-166, 1963.
(3) K. S. Fu and P. H. Swain : "On syntactic Pattern Recognition". $3^{\text {rd }}$ Symposia on Computers and Inf. Science, COINS 1969. Bal Harbour, Fla, Dec. 18-20, 1969.
(4 R. A. Kirsch : "Computer Interpretation of English Text and Picture Patterns" IEEE. Trans, on Elect. Computers, EC-13, pp. 363-376, 1964
(5) M. Minsky : "Steps towards Artificial Intelli gence" Proc. IRE, 49, \# 1, Jan. 1961.
(6) R. Narasimhan : "Labeling Schemata and Syntac tic Descriptions of Pictures", In formation and Control, 7,pp. 151179, 1964
(7) J. L. Pfalz and A. Ro serif eld : "Web Grammars". Proc. of the Int. Joint Conf. on Art. Int. (1JCA1) May 1969.
(8) J. C. Simon and J. Camillerapp : "Recherche d'une Forme dans un Fond". C. R. Ac. Sc., Serie A, 267 pp. 946-949, 16 Decembre 1968.
(9) J. C. Simon, A. Checroun and C. Roche : "Com paraison des Formes Independantes des Deplacements, Homotheties et petites Deformations". C. R. Ac. Sc, Series A, 270, pp. 1.607-1.609 15 Juin 1970.
(10) J. C. Simon, A. Checroun and C. Roche : "A Me thod of Comparing two Patterns Inde pendent of possible Transformations and small Distortions". IEEE Sympo sium on syntactic pattern recogni tion. Chicago, oct. 1970. To be published in Pattern Recognition Journal.

