Structural Pattern Analysis
Eds. R. Mohr, Th. Pavlidis & A. Sanfeliu
© World Scientific Publ. Co., 1989, pp. 57 – 66

Singapore

The Pattern Complex

Robert M. Haralick
Department of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195

ABSTRACT

This paper introduces the concept of a pattern complex for the purpose of object recognition and identification. A pattern complex is a set of measurements, one measurement coming from each unit or entity being measured. Some of the measurements in the pattern complex come from entities which may be a part of one of the objects of interest. Other measurements may be the result of uninteresting environmental clutter or noise. Recognition with the pattern complex constitutes the structural problem of simultaneously selecting which units and measurements are associated with an object of interest and then making an identification based on the selected measurements. Classical statistical pattern recognition is the simplest, most special case of recognition with the pattern complex. Likewise, relational matching is a special case of the pattern complex problem.

1. Introduction

In the classical statistical pattern recognition paradigm (Fukunaga, 1972), the object to be recognized has already been isolated from its context or background and is observed. The observation results in a measurement vector. The object is then recognized by a decision rule which associates a category or class with the observed measurement vector.

There are instances in which what we consider as an object is really composed of parts put together in a specific way. Each part has a particular character and the number of parts which objects of the same class can have may vary. The structure of the object is composed of the parts, their properties, and their interrelationships. In this case, the observation of the object results in a set of measurement vectors, one vector per observed part and a superstructure which gives the observed interrelationship between parts. The recognition of such pattern complexes entails more than an assignment of category or class. It entails both the assignment of category and the explanation for the assignment. The explanation is a statement of the correspondence between the parts of the model object and the parts of the observed instance. It should be recognized at this point that recognition in the framework of the pattern complex is not only different than the classical pattern recognition framework, but it is also different than the framework of contextual pattern recognition (Haralick, 1983).

In Section 2 we give five examples of problems which can be posed as instances of the pattern complex problem. In Section 3 we give a formal description of the recognition problem for pattern complexes and give a derivation of the *a posteriori* probabilities whose form then suggests an algorithm to solve the recognition problem.

2. Some Problem Instances

In this section we describe five kinds of recognition problems which are instances of the recognition of a pattern complex. These instances are the multi-level string syntactic pattern recognition problem, the character recognition problem, the speech recognition problem, the relational matching problem, and the affine invariant matching problem.

2.1 The Multi-Level String Recognition Problem

In the multi-level string recognition problem, the model consists of all the strings in the language. We denote the set of characters which can participate in language strings as the set A. We denote the set of all legal language strings by R. The observation consists of an observed signal string $s_1, \ldots, s_N$. The characters of the signal string come from the set S which is not necessarily the same set of characters as those in the language. The relationship between the characters of the signal and the characters of the language is given by a certainty function $c : A \times S \rightarrow (-\infty, \infty)$ which for each language character a and signal character s gives the certainty that a and s are associated in a particular instance. Here certainty can be thought of as $-\log\ Prob\ (a|s)$.

The multi-level string recognition problem requires associating to each position of the signal string a character of the language such that the resulting language string is legal and the certainty of the association is maximized. That is, we must find $g : \{1, \ldots, N\} \rightarrow A$ to maximize $\sum_{n=1}^{N} c(s_n, g(n))$ under the constraint that $(g(1), \ldots, g(n)) \in R$.

In this problem, the observed primitives are the position indexes $\{1, \ldots, N\}$ and the model primitives are the language characters. The measured vector associated with each observed primitive n is just the signal character s_n. The legal model relationships are given by all the length N strings of the language. In this instance, g is a mapping which to the signal primitives associates language primitives which when ordered as the signal primitives are ordered constitute a legal language string and for which the statistical association between the observed signal characters and the corresponding language characters is maximized. The observed signal string is accepted as having arisen from the language, that is, is recognized as associated with a language string if the certainty of the statistical association $\sum_{n=1}^{N} c(s_n, g(n))$ is high enough.

2.2 The Character Recognition Problem

In the character recognition problem, each character can be modeled as a set of strokes. Each stroke has associated with it a feature vector whose components indicate things like kind of stroke, relative center position of stroke, orientation of stroke, curvature of stroke, length of stroke, etc. In this formulation of the character recognition problem, because relative position and orientation is encoded in the feature vector associated with the stroke primitive, the model need not include a separate relation which labels the relation between and among primitives.

Thus the model for each character consists of a pair (W, h) where W is a set of unique identifiers for all kinds of possible strokes which are legal for the given character and h is a function which assigns to each stroke a representative feature vector of its properties. Likewise, the observation consists of a pair (V, t) where V is a set of unique identifiers by which each observed stroke can be indexed and f is a function which associates to each stroke a vector of its measured properties.

Recognition of the observation (V, f) as the character associated with the model (W, h) depends on the finding of a 1-1 single valued relation g, $g \subseteq W \times V$, which associates some of the observed strokes with some of the model strokes and which maximizes the statistical association between the observed feature vectors and the corresponding model feature vectors. We write this maximization as finding the g which maximizes

$$\text{Prob}\left(\sum_{w \in \text{Dom}(g)} \|h(w), f(g(w))\|, \text{Dom}(g) \right).$$

2.3 Speech Understanding

One common approach to speech understanding is to divide the speech signal into possibly overlapping segments, often around 10 ms in length and compute the power spectrum for each segment. Then each segment is associated with a power spectrum frame from which interesting events are extracted. Interesting events can be, for example, spectral peaks whose properties are relative frame number, relative center frequency, relative height, relative width, etc. The model for a given spoken word consists of (W, h) where W is the set of identifiers for all possible events and h is the function which associates to each possible event its prototypical feature vector. The observation consists of the pair (V, f), where V is an index set providing unique identifiers for all the observed events and f is the function which to each event associates the measured feature vector. Recognition of the observation (V, f) as the word associated with the model (W, h) entails finding a relation g which in an ordered way associates some of the possible frame events of the model with some of the observed frame events, and which maximizes the statistical association of the corresponding feature vectors. The relation g, $g \subseteq W \times V$ must be one–one and single–valued. By ordered we mean that we constrain g so that if frame index is the property i, then primitives on the same frame of the model must be associated with primitives on the same frame of the observation and one primitive of the model on a later frame than another primitive of the model must map to observed primitives which stand in the same relation. That is,

(1) if $w_1,\ w_2 \in$ Dom(g) and $h[w_1](i) = h[w_2](i)$, then $f[g(w_1)](i) = f[g(w_2)](i)$,

(2) $w_1, w_2 \in$ Dom(g) and $h[w_1](i) < h[w_2](i)$, then $f[g(w_1)](i) < f[g(w_2)](i)$.

2.4 Relational Matching

The simple structural relational matching problem has an N-ary model relation R over a unit primitive set W and an observed N-ary relation S over an observed primitive set V. The relational matching problem is to find a relation $g, g \subseteq W \times V$, associating some primitives of W with some primitives of V such that the number of model primitives associated with the observed primitives, and the number of model relations preserved in the observation is maximized. In general we seek to maximize $Prob\left(\|R \circ g - S\|, \text{Dom}(g)\right)$ where

$$R \circ g = \{(v_1, \ldots, v_N) \in V^N |\ \text{for some } (w_1, \ldots, w_N) \in W, (w_n, v_n) \in g,\ n = 1, \ldots, N\}.$$

The relation g is constrained to be one–one and single–valued.

2.5 2D Affine Matching

In the 2D affine matching problem, a set W of 2D model points is given. The relationship between any point $w \in W$ and any triple of non-colinear model basis points w_1, w_2, w_3 is given by the affine coordinates of w with respect to w_1, w_2, w_3. In matrix algebra notation, these affine coordinates $\binom{\alpha}{\beta}$ are defined by

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (w_1 - w_3 \quad w_2 - w_3)^{-1} (w - w_3).$$

With this definition for $\binom{\alpha}{\beta}$, w can be represented by

$$w = \alpha(w_1 - w_3) + \beta(v_2 - w_3) + w_3.$$

The observation is a set V of 2D points where some of the points of V are assumed to be some of the model points altered by the same affine transformation. The affine coordinates of a point with respect to an affine basis are invariant under the affine transformation of both the point and the basis points. This invariance permits the following characterization to be given to the affine matching problem. The model R consists of points, their basis, and the affine coordinates of the point with respect to the basis. That is, $R \subseteq W^4 \times E^2$ where

$$\begin{aligned} R = \{(w_1, w_2, w_3, w_4, \alpha, \beta) \in W^4 \times E^2 | w_4 \\ = \alpha(w_1 - w_3) + w_3 \text{ and } |w_1 - w_3 \quad w_2 - w_3| \neq 0\}. \end{aligned}$$

Likewise, the observation consists of $Q \subseteq V^4 \times E^2$ where

$$\begin{aligned} Q = \Big\{(v_1, v_2, v_3, v_4, \alpha, \beta) \in V^4 \times E^2 \Big| \\ v_4 = \alpha(v_1 - v_3) + \beta(v_2 - v_3) + v_3 \text{ and } \Big|v_1 - v_3 \quad v_2 - v_3\Big| \neq 0\Big\}. \end{aligned}$$

For any one–one single valued relation $g \subseteq W \times V$ associating some of the model points with some of the observed points, define

$$\begin{aligned} R \circ g = \Big\{(v_1, v_2, v_3, v_4, \alpha, \beta) | \text{ for some } (w_1, w_2, w_3, w_4, \gamma, \delta) \in R, \\ (w_n, v_n) \in g, \; n = 1, 2, 3, 4 \text{ and } \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \gamma \\ \delta \end{pmatrix} \right\| < \epsilon^2 \Big\}. \end{aligned}$$

An observed (V, Q) is recognized as containing an instance of the model (W, R) if the number of associations established by the largest one–one single–valued relation g satisfying $R \circ g \subseteq Q$ is large enough.

3. The Definition of Pattern Complex

In this section we give the general definition of the pattern complex. Each of the examples in section 2 is an instance of the pattern complex problem.

Let W be an index set for all possible primitives from all possible object classes which may participate in a pattern. Each primitive is the primary entity which can be observed and in our informal initial description the primitive was the object part. Not all primitives participate in the observation of any particular object. Let U be an index set for the primitives or units actually observed in a particular instance. Hence, $U \subseteq W$. Although the primitives or units in U are observed and, therefore, the number of units in U is known, the actual identity of the units in U is not known. For example, suppose the world had only one kind of object: a chair. The unit identities for the primitives of the chair might be left front leg, right front leg, left back leg, right back leg, seat, back, left arm, and right arm. These identities, or unique indexes for them, constitute the set W. The observation of a particular chair might result in some part of the chair with the chair or its environment obscuring some other parts to the chair. So a couple of legs might not be observable. What is observable is observed as a set of coordinated shapes corresponding to the observable legs, one to the seat, one to the back, and one for each arm. However, which observed shape primitive corresponds to a leg or seat, or arm is not known until the complete identity of chair has been established. Before the recognition takes place, the identity of the primitives participating in the description is not known. Therefore, we have said that the identity of the units in U is not known.

Associated with every observed unit u is an observed measurement pattern $f(u)$. If each measurement pattern is an N-tuple of real numbers, then $f(u) \in R^N$. Thus instead of having to classify the observed pattern on the basis of one measurement vector, we must be able to classify the observed pattern on the basis of a variable number of measurement vectors.

For a fixed unit set U we denote the measurement space by $F[U]$. It is the set of all possible functions which assign a measurement vector in R^N to each unit $u \in U$.

$$F[U] = \{f \mid f : U \rightarrow R^N\}$$

In addition to each observed unit u and its measurement pattern $f(u)$, f being a member of $F[U]$, the pattern complex may have a relational superstructure between observed units. Let L be a label set for the different kinds of relationships which units can participate in. The simplest relational structure S can be taken to be is a labeled N-ary relation on U. Hence, $S \subseteq U^N \times L$. A pattern complex can then be defined as a quadruple (U, f, S, L).

To deal with pattern complexes we must define exactly how we model the random process by which the pattern complex is generated. For a fixed class c the observed unit set U is chosen according to a conditional probability distribution $P_W(U \mid c)$, f is chosen according to $P_{F[U]}(f \mid c)$, and S is chosen according to $P_U(S \mid c)$. Finally, we must make an explicit acknowledgement of the fact that the identities of the observed units are not known. We do this by distinguishing between the object unit set U and the observed unit set V. The observed unit set consists of the unrecognized units Some of the units observed in V come from U. Others are unrelated and can be thought of as clutter objects. In our chair example these unrecognized units might be vertical stud 1, vertical stud 2, blob 1, blob 2, etc. They are the detected but as yet unnamed shapes. The identities of the units in V are in essence semantically meaningless indexes, one unique index for each observed unit in V.

Because the identities of the observed units are not known, the one-one correspondence g between a subset of the object unit set U and some of the observed units from set V is not known. Before the observation and without any additional knowledge, all possible correspondences between U and V are considered to be equally likely. Hence the prior probability for g, where $g : U \to V$, is just the constant $1/\#\ \mathrm{Dom}(g)!$, where $\#$ is the number of elements in a set, and ! designates the factorial function.

The unknown correspondence g gives the transformation between the object pattern complex (U, f, S, L) and the observed pattern complex (V, h, T, L) where $h : V \to R^N$ is defined by $h(v) = f\big(g^{-1}(v)\big)$ and $T \subseteq V^N \times L$ is defined by $T = \{(v_1, v_2, \ldots, v_N, l) \in V^N \times L \mid \text{for some } u_1, u_2, \ldots, u_N \in U, (u_1, u_2, \ldots, u_N, l) \in S, g(u_n) = v_n, n = 1, \ldots, N\}$.

In order to perform recognition of the observed pattern complex (V, h, T, L), the *a posteriori* probability $P[c, g\ \mathrm{Dom}(g) \mid (V, h, T, L)]$ must be determined. Recognition proceeds by assigning the observed pattern complex (V, h, T, L) to that class c associated with correspondence g which maximizes $P[c, g\ \mathrm{Dom}(g) | (V, h, T, L)]$. Notice that in our posing of the problem the desired maximization is not just over the unknown class c but is jointly over the unknown class c and the correspondence g which is, in effect, the explanation of why class c is the most likely class. Establishing identity means naming the class and establishing the correspondence between object and model parts. This is very different from the classical pattern recognition paradigm in which only establishing the identity of the class is of interest. By the definition of conditional probability,

$$P\big[c, g\ \mathrm{Dom}(g) | (V, h, T, L)\big] = \frac{P\big[(V, h, T, L) | c, g,\ \mathrm{Dom}(g)\big] P\big[c, g,\ \mathrm{Dom}(g)\big]}{P(V, h, T, L)}.$$

Since

c, g, and $\mathrm{Dom}(g)$ do not appear in the denominator, maximizing $P[c, g,\ \mathrm{Dom}(g)|$

$(V, h, T, L)]$ is equivalent to maximizing $P[(V, h, T, L)|c, g,\ \mathrm{Dom}(g)]P[c, g,\ \mathrm{Dom}(g)]$. Since we assume $P[c, g,\ \mathrm{Dom}(g)]$ to be known, the class conditional probability $P[(V, h, T, L)|c, g,\ \mathrm{Dom}(g)]$ is the probability we must find a way to determine.

To understand how the probability $P[(V, h, T, L)|c, g,\ \mathrm{Dom}(g)]$ can be determined, we first recognize that the key is to translate the observed pattern complex (V, h, T, L) to the object model pattern complex (U, f, S, L) since it is the object model pattern complex for which we have a class conditional distribution.

To translate (V, h, T, L) to (U, f, S, L) we need to know the object unit set U and the correspondence g between $\mathrm{Dom}(g)$, a subset of U and V. It is then natural to begin with

$$\begin{aligned} P\Big((V, h, T, L) \mid c, g,\ \mathrm{Dom}(g)\Big) &= P\Big(h, T|V, c, g,\ \ \mathrm{Dom}(g)\Big)P\Big(V, c, g,\ \mathrm{Dom}(g)\Big) \\ &= P(h, T|V, c, g,\ \ \mathrm{Dom}(g))P(V, g,\ \ \mathrm{Dom}(g)|c)P(c) \end{aligned}$$

Since we assume measurements and relationships are conditionally independent given $V, c, g,$ and $\mathrm{Dom}(g)$,

$$P\Big(h, t|V, c, g,\ \mathrm{Dom}(g)\Big) = P\Big(h|V, c, g,\ \mathrm{Dom}(g)\Big)P\Big(T|V, c, g,\ \mathrm{Dom}(g)\Big)$$

so that

$$\begin{aligned} P\Big((V, h, T, L)|c, g,\ \mathrm{Dom}(g)\Big) &= P\Big(h|V, c, g,\ \mathrm{Dom}(g)\Big) \\ &\quad \cdot P\Big(T|V, c, g,\ \mathrm{Dom}(g)\Big)P\Big(V, g,\ \mathrm{Dom}(g)|c\Big)P(c) \end{aligned}$$

Consider the meaning of $P(V, g,\ \mathrm{Dom}(g)|c)$. Conditioned on class identity it is essentially the probability that evaluates the likelihood of the number of units in $\mathrm{Dom}(g)$ being matched through g to a subset of the observed units V. Since for class c, U designates the set of possible model units, it is natural for $P(V, g,\ \mathrm{Dom}(g)|c)$ to monotonically depend on the difference between the size of U and the size of $\mathrm{Dom}(g)$. Therefore, it is reasonable to take

$$P(V, g,\ \mathrm{Dom}(g)|c) = kq^{\#U - \#\ \mathrm{Dom}(g)},\ \text{for some } k > 0 \text{ and } 0 < q < 1.$$

This probability essentially penalizes for a correspondence g which does not have many matches.

We rewrite $P\Big(h|V, g, c,\ \mathrm{Dom}(g)\Big)$ and $P\Big(T|V, g, c,\ \mathrm{Dom}(g)\Big)$ as follows.

$$\begin{aligned} P\Big(h|V, g, c,\ \mathrm{Dom}(g)\Big) &= P(g \circ h|V,\ \mathrm{Dom}(g), c) \\ P\Big(T|V, g, c,\ \mathrm{Dom}(g)\Big) &= P(T \circ g^{-1}|V,\ \mathrm{Dom}(g), c) \end{aligned}$$

where $g \circ h$ is defined by

$$(g \circ h)(u) = h\big(g(u)\big), \quad u \in \text{Dom}(g)$$

and $T \circ g^{-1}$ is defined by

$$T \circ g^{-1} = \{(u_1, \ldots, u_N) \in U^N | \text{ for some } (v_1, \ldots, v_N) \in T, (v_n, u_n) \in g^{-1}\}.$$

Recall that f is the prototypical measurement function associated with class c on the model units. Then

$$\begin{aligned} P(g \circ h | V, \text{ Dom}(g), c) &= P(g \circ H | V, \text{ Dom}(g), c, f) \\ &= P(\rho(g \circ h - f |_{\text{Dom}(g)}) | c) \end{aligned}$$

where ρ is a suitable metric function. Hence, the farther the measurements of the matched units are from the prototypical measurements fo the corresponding units, the smaller the probability. Recall that S is the relational superstructure associated with class c. Then

$$\begin{aligned} P(T \circ g^{-1} | V, \text{ Dom}(g), c) &= P(T \circ g^{-1} | V, \text{ Dom}(g), c, S) \\ &= P(\#(S - T \circ g^{-1}) | c). \end{aligned}$$

This means that the more tuples the relational superstructure contains that are not matched to observed relational tuples, the lower the probability. Upon making all the substitutions, we obtain

$$\begin{aligned} P\big(V, h, T, L, c, g, \text{ Dom}(g)\big) = {} & P(\rho(g \circ h - f\big|_{\text{Dom(g)}}) | c) \\ & \cdot P(\#(S \circ g - T)\big| c) \cdot kq^{\#U -} - \text{ Dom}(g) P(c) \end{aligned}$$

Determining the most probably class c then corresponds to searching for that association g which maximizes $P[V, h, T, L, c, g, \text{ Dom}(g)]$.

4. Conclusion

In this theoretical note we have given a definition for the pattern complex whose more complicated structure more closely matches the world in which real pattern recognition takes place. The pattern complex includes as immediate special cases, the classical statistical pattern recognition measurement tuple data structure as well as the relational data structure used in structural pattern recognition. The pattern complex makes explicit the fact that in most pattern recognition problems the object being recognized is observed as a number of parts existing in a cluttered environment of extraneous parts and the pattern recognizer does not know ahead of time which of the observed parts go with the object of interest or go with the environment. By making this unknown state of knowledge explicit in the statement of the pattern recognition problem, we come closer to bridging the gap between the segmentation and feature extraction stages and the decision rule construction stage of the pattern recognition problem.

References

1. Fukunaga, K., *Introduction to Mathematical Techniques in Pattern Recognition,* Prentice Hall, New Jersey, 1972.
2. Haralick, R.M., "Decision Making in Context," *IEEE Tranactions on Pattern Analysis and Machine Intelligence,* Vol.PAMI-5, No.4, 1983, pp. 417-428.