LOCATING THE WHOLE PATTERN IS BETTER THAN LOCATING ITS PIECES: A GEOMETRIC EXPLANATION OF AN EMPIRICAL PHENOMENON

by Scott A. Starks and Vladik Kreinovich

NASA Pan-American Center for
Earth and Environmental Studies
University of Texas at El Paso, El Paso, TX 79968
emails {sstarks,vladik}@utep.edu

Abstract. *In many practical problems, we must find a pattern in an image. For situations in which the desired pattern consists of several simple components, the traditional approach is first to look for such components, and then to see whether the relative locations of these components are consistent with the pattern. Recent experiments have shown that a much more efficient pattern recognition can be achieved if we look for the whole pattern (without decomposing it first). In this paper, we give a simple geometric explanation of this empirical fact.*

The practical problem. In many pattern recognition problems, we must locate a known simple pattern in a complicated black-and-white image. For example:

- in automatic analysis of electronic schemes, we must locate symbols of standard electronic components (such as $-|\ |-$);
- in text recognition, we must find letters,
- similar pattern matching problems arise in satellite imaging, etc.

Traditional approach. Most traditional methods for solving this problem are based on the fact that the desired pattern consists of simple geometric components (straight line intervals, arcs, etc.)

For example, the above symbol for capacitor consists of four straight line intervals $-$, $|$, $|$, and $-$.

Traditional methods consist of two stages:

- first, we try to locate each component of the desired pattern;
- after all components are located, we check that their relative locations are close to the relative locations of these components in the desired pattern (to be more precise, we check that the difference between the observed and desired relative locations is within the limits set by the observation inaccuracy of component location).

A new approach turns out to be better. The authors of [Murshed Bortolozzi 1998] propose to recognize the entire pattern (symbol) without first decomposing it into simple components. The resulting algorithm requires more computation time, but leads to much better recognition: namely, if we set up the parameters of both methods in such a way as to avoid false negatives (unrecognized symbols), then for the new method, the number of false positives (false recognitions of a pattern) is much smaller than for the traditional methods.

In this paper, we give a simple geometric explanation of this empirical phenomenon.

Geometric reformulation of the problem. We start with a sample pattern P which consists of several components P_i: $P = P_1 \cup \ldots \cup P_n$. Without losing generality, we can assume that P is a compact set. In the actual image, the actual pattern may be shifted relative to the standard one, so this actual pattern has the form TP for some shift (translation) T.

For simplicity, we will assume that the pattern is surrounded by an empty space, i.e. (at least locally):

- either the actual image I coincides with the shifted pattern TP, in which case the pattern is present,
- or the actual image is different from the shifted pattern, in which case the pattern is not here.

Description of measurement inaccuracy. Due to measurement inaccuracy, the observed image $\widetilde{I}$ is, in general, slightly different from the actual image I. Namely, due to this inaccuracy, for each point p from the original image, the corresponding observed point $\widetilde{p}$ may be different from p.

The observation inaccuracy can be characterized by the largest possible distance $d(\widetilde{p}, p)$ between the actual and the observed points. If this inaccuracy is $\varepsilon > 0$, this means that:

- every point from I is ε-close to some point from $\widetilde{I}$, and
- every point from $\widetilde{I}$ is ε-close to some point from I.

In other words, this means that the *Hausdorff distance* between the actual and observed images does not exceed ε: $d_H(I, \widetilde{I}) \leq \varepsilon$.

New approach reformulated in geometric terms. If the desired pattern P is present in the image, i.e., if $I = TP$, then:

$$\text{There exists a } T \text{ for which } d_H(TP, \widetilde{I}) \leq \varepsilon. \qquad (1)$$

Vice versa, if for an observed image $\widetilde{I}$, this condition holds, this means that there exists a pattern TP which is consistent with the observed image, and therefore, it is quite possible that the observed image contains a desired pattern. Thus, the condition (1) expresses the fact that the observed image $\widetilde{I}$ is consistent with the assumption that the actual image contains the desired pattern.

Hence, if we want to avoid false negatives (i.e., un-recognized patterns), we must check the condition (1). This is what the new approach does.

How good is the new approach.

- If the result of the new approach is *negative*, this means that the observed image *does not contain the pattern*;
- on the other hand, if the result of this approach is *positive*, this means that *it is possible that the observed image contains the pattern* (i.e., that the observed image is consistent with the assumption that it is actually the shifted standard pattern).

We cannot get any better than that. Of course, due to the observation inaccuracy, without additional assumptions, we can never guarantee that the image is actually the desired pattern: the actual image could as well be a slightly distorted pattern, and because of the observation inaccuracy, we do not notice this distortion.

With this comment in mind, we can see that we cannot get any better pattern recognition than by using the new approach.

Traditional approach reformulated in geometric terms. In traditional approach, we first look for *components*, i.e., we look for the possibility for representing the observed image $\widetilde{I}$ as a union of n sets $\widetilde{I}_1, \ldots, \widetilde{I}_n$ such that for every i, the i-th component $\widetilde{I}_i$ of the observed image is consistent with it being actually a shift $T_i P_i$ of i-th component P_i of the desired pattern P.

Similarly to the above argument, we can conclude that the possibility for $\widetilde{I}_i$ to be actually a shift of P_i can be described as follows:

$$\text{There exists a } T_i \text{ for which } d_H(T_i P_i, \widetilde{I}_i) \leq \varepsilon. \qquad (2)$$

Therefore, if we want to avoid false negatives (i.e., if we do not want un-recognized patterns), we should look for a partition $\widetilde{I} = \widetilde{I}_1 \cup \ldots \cup \widetilde{I}_n$ which satisfies the property (2) for all $i = 1, \ldots, n$. This is the first stage of the traditional approach. As a result of this stage:

- If such a partition is impossible, then, based on the observation $\widetilde{I}$, we can conclude that the actual (unknown) image I does not coincide with the desired pattern, and therefore, the desired pattern is not present here.
- On the other hand, if the partition is possible, i.e., if $\widetilde{I} = \widetilde{I}_1 \cup \ldots \cup \widetilde{I}_n$ with $d_H(\widetilde{I}_i, T_i P_i) \leq \varepsilon$ for some shifts T_i, then it is not necessarily true that $\widetilde{I}$ can contain the desired pattern: it may happen that the shifts are too far away from each other.

If the actual image I is indeed a shift of the standard pattern P, i.e., if $I = TP$ for some T, then, due to possible observation inaccuracy, $d_H(\widetilde{I}_i, TP_i) \leq \varepsilon$. Based on the observed components $\widetilde{I}_i$, we select shifts T_i for which $d_H(\widetilde{I}_i, T_i P_i) \leq \varepsilon$. Therefore, we can conclude that if the actual image is indeed the shift of the standard pattern, then $d_H(T_i P_i, TP_i) \leq d_H(T_i P_i, \widetilde{I}_i) + d_H(\widetilde{I}_i, TP_i) \leq 2\varepsilon$.

The Hausdorff distance between two shifts $T_i P_i$ and TP_i of the same set is equal to the distance between $d(T_i, T)$ these shifts, i.e., to the Euclidean distance between the vectors corresponding to these shifts. So, we can conclude that if the pattern is present, then all the shifts T_i generated on the first stage should be 2ε-close

to some (unknown) shift T. This means, in turn, that for every i and j, we have $d(T_i, T_j) \leq d(T_i, T) + d(T, T_j) \leq 4\varepsilon$.

So, on the second stage of the traditional method, we check the following condition:

$$d(T_i, T_j) \leq 4\varepsilon \text{ for all } i \text{ and } j. \tag{3}$$

How good is the traditional approach.
- If the result of traditional approach is *negative*, this means that the observed image *does not contain the pattern*;
- on the other hand, if the result of this approach is *positive*, this does not necessarily mean that it is possible that the observed image contains the pattern; it is quite *possible that the observed image is inconsistent with the* assumption that it is actually the shifted standard *pattern*.

Let us give a simple example explaining why this can happen. Let us consider a 2-component pattern $P = \lfloor\_$ consisting of a vertical component P_1 of length 1 and a horizontal component P_2 of the same length 1. If we take the angle of P as the origin $(0, 0)$ of the coordinate system, then $P_1 = \{0\} \times [0, 1]$ and $P_2 = [0, 1] \times \{0\}$. Let us take $\widetilde{I} = \widetilde{I}_1 \cup \widetilde{I}_2$, where $\widetilde{I}_1 = \{-2\varepsilon\} \times [0, 1]$ and $\widetilde{I}_2 = [2\varepsilon, 1 + 2\varepsilon] \times \{0\}$.

- For this image, the traditional approach can lead to a positive answer: indeed, here:
  - $d_H(\widetilde{I}_1, T_1 P_1) \leq \varepsilon$ for $T_1 = (-\varepsilon, 0)$,
  - $d_H(\widetilde{I}_2, T_2 P_2) \leq \varepsilon$ for $T_2 = (\varepsilon, 0)$, and
  - $d(T_1, T_2) = 2\varepsilon < 4\varepsilon$.
- On the other hand, the image $\widetilde{I}$ is *not* consistent with the pattern P because, as one can easily see, $d_H(\widetilde{I}, TP) \geq 2\varepsilon$ for all possible shifts T.

So, the traditional approach is indeed not perfect.

Open problem. In the above text, we simply gave an *example* of when a traditional method leads to unnecessary false positives. It is desirable to have a *general* numerical *estimate* of the quality of the traditional approach. In precise terms, we have the following problem:

We have n compact sets P_i, and n compact sets $\widetilde{I}_i$. We know that for every i from 1 to n, $d_H(\widetilde{I}_i, T_i P_i) \leq \varepsilon$ for some shifts T_i for which $d(T_i, T_j) \leq 4\varepsilon$ for all i and j. What is the smallest possible value of th Hausdorff distance $d_H(\widetilde{I}, TP)$ between the union $\widetilde{I} = \widetilde{I}_1 \cup \ldots \cup \widetilde{I}_n$ and a shift TP of the union $P = P_1 \cup \ldots \cup P_n$?

Our *guess* is that this smallest possible value is 3ε.

Our *argument* in favor of this guess is as follows: it looks like, since the diameter of the set $\{T_1, \ldots, T_n\}$ is $\leq 4\varepsilon$, that its radius will be $\leq 2\varepsilon$, i.e., that there should exist a shift T for which $d(T_i, T) \leq 2\varepsilon$ for all i. For this shift T, we have

$$d_H(\widetilde{I}_i, TP_i) \leq d_H(\widetilde{I}_i, T_i P_i) + d_H(T_i P_i, TP_i) \leq$$

$$\varepsilon + d(T_i, T) \leq \varepsilon + 2\varepsilon = 3\varepsilon.$$

Reference

N. A. Murshed and F. Bortolozzi, "A fuzzy ARTMAP module for graphics symbol recognition", *Proceedings of the IJCNN'98, IEEE International Joint Conference on Neural Networks*, Anchorage, Alaska, May 4–9, 1998, Vol. 3, pp. 1700–1705.