8/27/18

pp. 164-167.

A317967 A317968

```
100010011010111100
000000000000000000
000000000000000000
000000000000000000
000100110101111000
000100110101111000
000100110101111000
000000000000000000
001101011110001001
000000000000000000
001001101011110001
000100110101111000
011110001001101011
010111100010011010
000000000000000000
010111100010011010
```

ACKNOWLEDGMENT

The author is indebted to one of the referees of a previous version for the method of construction for $h(Z)$.

REFERENCES

[1] B. Elspas, "The theory of autonomous linear sequential networks," *IRE Trans. Circuit Theory*, vol. CT-5, p. 45, 1959.

[2] W. W. Peterson, *Error Correcting Codes.* Cambridge, MA: MIT Press, 1961.

[3] D. Knuth and J. D. Alanen, "Tables of finite fields," *Sankhya, Ser. A.*, vol. 26, p. 305, 1971.

Topoglyphs

MARCEL J. E. GOLAY

Abstract—**A short list is given of two-dimensional line structures which are topologically identifiable one from the other.**

They are listed in the format of the several ways a connected, single contoured area can be subdivided by dividing lines or by "pinching" with junctions of three lines only.

Index Terms—**Hexagonal field, topological alphabet, topological different patterns, topoglyphs.**

INTRODUCTION

After the development of pattern recognition by means of transformations in a hexagonal field had been underway for some time in a long-range project of the Perkin-Elmer Corporation which was aimed primarily at the automatic classification of blood cells [1]–[4], the question was raised by J. G. Atwood (private communication) whether certain special configurations would be particularly amenable to automatic reading by the "cellscan" pattern recognizer. This question implied an about-face of the attack followed until then: in contrast with the development of empirical algorithms for the recognition of relatively complex patterns such as leucocytes, etc., the question asked what kind and what diversity of patterns could be readily identified by means of the several relatively simple algorithms developed early during this work, and by means of chains of such algorithms.

It occurred to the writer that a general approach to the problem suggested by this question could consist in devising and classifying contoured areas (circles and topologically invariant deformations thereof) subdivided in topologically different ways, and the name "topoglyph" was coined for the various configurations envisioned and described in the discussion below. As a result of this approach a veritable topological alphabet was evolved which, unlike standard language alphabets, could be read independently of size, shape, orientation, and position in the field allotted to any one message or address; an alphabet indeed in which the roman letters a, b, d, e, o, p, q, and the numbers 0, 4, 6 and 9 would be considered as (topologically) identical, while all other letters and numbers except g in some fonts and 8 would be identical to a single point, or two for i and j.

In these days of expanding channel capacities and computer speeds and memories the laboriously acquired correction codes and the greater efficiencies they permit have lost some of their former pertinence while new writing schemes such as topoglyphs, which even a small child could learn to place in random orientation and position in spaces left clear within an assigned writing area, may deserve to be looked at for the sake of whatever new fields they might open up. Applications such as the routing of mail and packages, dispatching, forwarding, luggage handling in airports, etc., appeal readily to one's imagination, and teaching large apes to write topoglyphs they can associate with various everyday routines could become an intriguing problem for a zoologist.

DISCUSSION

An elementary knowledge of the various transformations possible in a hexagonal field of points will be assumed, such as the thinning or filling operations in which a ONE may be removed or added whenever its surround consists of all or part of the patterns of index 1 to 5 listed in [1], and a case suitable for description in common language will be chosen as an example. Suppose, accordingly, that a circle be divided in two parts in as many topologically different ways as possible. One obtains the patterns 20, 21 and 23 of Table I, to which pattern 22, the "pinched" circle, is added for the sake of topological completeness and of convenient generalization to division of an area into three, etc., parts by a network of lines in which junctions are always simple—three line junctions—in order to avoid the ambiguities and difficulties which would result from the use of junctions of more than three lines.

Consider now the transformations which will yield the number of patterns 23 in a field which contains also patterns 20, 21, and 22.

The patterns are originally in a field number 1 and are subjected to the following transformations:

1) They are "shrunk" until they are represented by thin lines, i.e., until there are no ONE's in that field with a surround of index 1 through 6. This is accomplished by working with subfields of three, the ONE's of each subfield with surrounds of index 1 through 5 being replaced by ZERO's, and the process being iterated a sufficient number of times for the three subfields in turn.

2) The patterns of field number 1 are marked in a field number 2 where they are blown up without any two patterns coalescing, and without any "cavity" of ZERO's collapsing without a trace. This is again accomplished by working with subfields of three processed in turn, the ZERO's with a surround of index 1 through 5 in each subfield being replaced by a ONE. This results in each singly connected enclosure of ZERO's shrinking to a single ZERO with a surround of index 6, each such ZERO being marked as a ONE in a field number 3.

3) The ONE's of field number 3 being used as "seeds" are blown up by replacing each ZERO of that field with a surround of index

genus

n	A	B
1	1	1
2	4	5
3	23	28
4	186	?

186 + 1

11

186 + 1 + 23 + 6

220

Manuscript received June 18, 1975; revised December 15, 1975.
The author is with the Perkin-Elmer Corporation, Norwalk, CT.

0018-9340/78/0200-0164$00.75

4

TABLE I

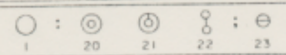
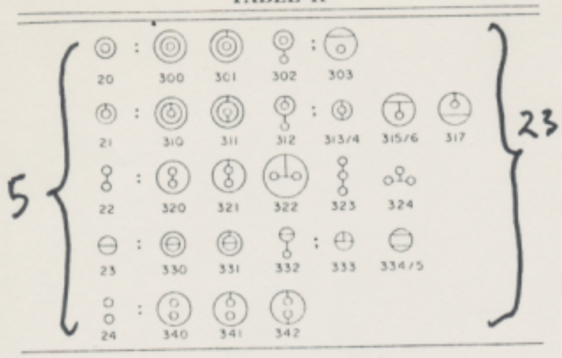


TABLE II



1 through 13 by a ONE, but using the thinned lines of field number 1 to inhibit a ONE being placed in a homologous (corresponding) position of field number 3. This process does not require the use of subfields.

4) The ONE's of fields numbers 1 and 3 are all marked as ONE's in (previously erased) field number 2.

5) The ONE's of field number 1, the homologous positions of which in field number 2 have surrounds of index 6 in that field, are marked as ONE's in (previously erased) field number 3. This process does not require the use of subfields, and it can be easily verified that this results in the original patterns 23 of field number 1 being transformed in simple thinned lines like the line shown as a diameter of the circle illustrating pattern 23; meanwhile the patterns 21 of the original field have been reduced to as many thin closed loops connected each to a line, corresponding to the pattern within the circle of the pattern 21 shown above, and the patterns 20 and 22 have yielded nothing.

6) The ONE's of field number 3 with surrounds of index 1 are replaced by ZERO's. This process is iterated subfield by subfield and when completed results in the simple thin lines being reduced each to a single ONE with a surround of index 0, while the closed loops with appended lines of that same field are reduced to closed loops.

7) The ONE's of field number 3 with a surround of index 0 are marked as ONE's in (previously erased) field number 2 and counted. This yields the number of patterns 23 in the original field.

Consider now the several kinds of topologically different patterns which can be formed by dividing the area encompassed in an original closed loop into three and four areas respectively, including the patterns formed by pinching.

When one tries to devise a logical method of subdividing the original area into several topologically different patterns, and of cataloguing these derived patterns, one finds no completely satisfactory algorithm for the purpose, for the various reasonable algorithms one selects turn out to be either incomplete, or redundant, or even both. The compromise adopted in this study was as follows:

Starting with the four two-area patterns of Table I—the first cipher of the pattern indices refers to the number of areas into which on original single-contoured area has been subdivided—these patterns are transformed into three-area patterns by:

1) Circumscribing them in a circle to which they are not attached.

2) Circumscribing them in a circle to which they are attached in various ways by a connecting line; and

3) Attaching circles to the two-area patterns in various ways.

This triple algorithm is not redundant, while it is incomplete, and served to generate the three-area patterns of Table II shown between the colon (:) and the semicolon (;).

The seven missing patterns were then generated by placing within the two area patterns, but not outside of them, a joining line dividing in two one of these two areas, so that three area patterns were formed.

The four-area patterns shown in Table III were formed similarly, but in three steps. The first step was the circumscribing by circles or appending of circles described above, and performed on the 23 three-area patterns, which was again nonredundant. The second was the addition of unjoined or joined circles within the two area patterns, which yielded most of the missing patterns, after the first semicolon; and the third step was, as for the formation of the missing three-area patterns, the addition of joining lines which could be inserted within, but not outside, the three-area patterns. This last step gave the few missing patterns 4044, 4138–4143, 4335, 4336, 4351, and 4352, for a total of 170 four-area patterns. The 24, 350–353 and 360 patterns, while not legitimately belonging to the regular two and three-area patterns, were included to give certain higher order patterns derivable from them. Likewise, the last six four-area patterns were obtained by joining with one line the six assemblies of two of the three two-area patterns.

The coding of the patterns was done for the three area patterns by replacing the first cipher of the two-area codes by a 3, retaining the second cipher, and adding a third, ordinal cipher. In three cases the third cipher was doubled up (313/4, etc.,) for the sake of providing enough numbers for the four area symmetric patterns derived therefrom and exceeding 10 in number. As for the assymmetric four-area patterns, the code of the twelve pairs of enantiomorphs found was formed by replacing the last cipher by a respective R and L for the two of the pair, and by priming (′) the latter to remove the ambiguity in three cases. Obviously, reliance will have to be had on the differentiation provided by surrounds of index 8 and 9 ([1]) in order to distinguish the R from the L among the enantiomorphs.

The passage from four to five-area patterns, and, more generally, from n to $n + 1$-area patterns can be made exactly as described above for the passage from three to four-area patterns, and the same remarks are valid regarding the obvious absence of any generating scheme simultaneously complete and nonredundant. As to the number of n-area patterns, one's mathematical intuition is that no formula will ever be devised which will yield it, any more than a formula will ever be devised to yield the number of 2-D, 3-D, 4-D, etc. space groups with different symmetries, for we are here in the land of never–never of combinatorial analysis in which unsolvable problems abound.

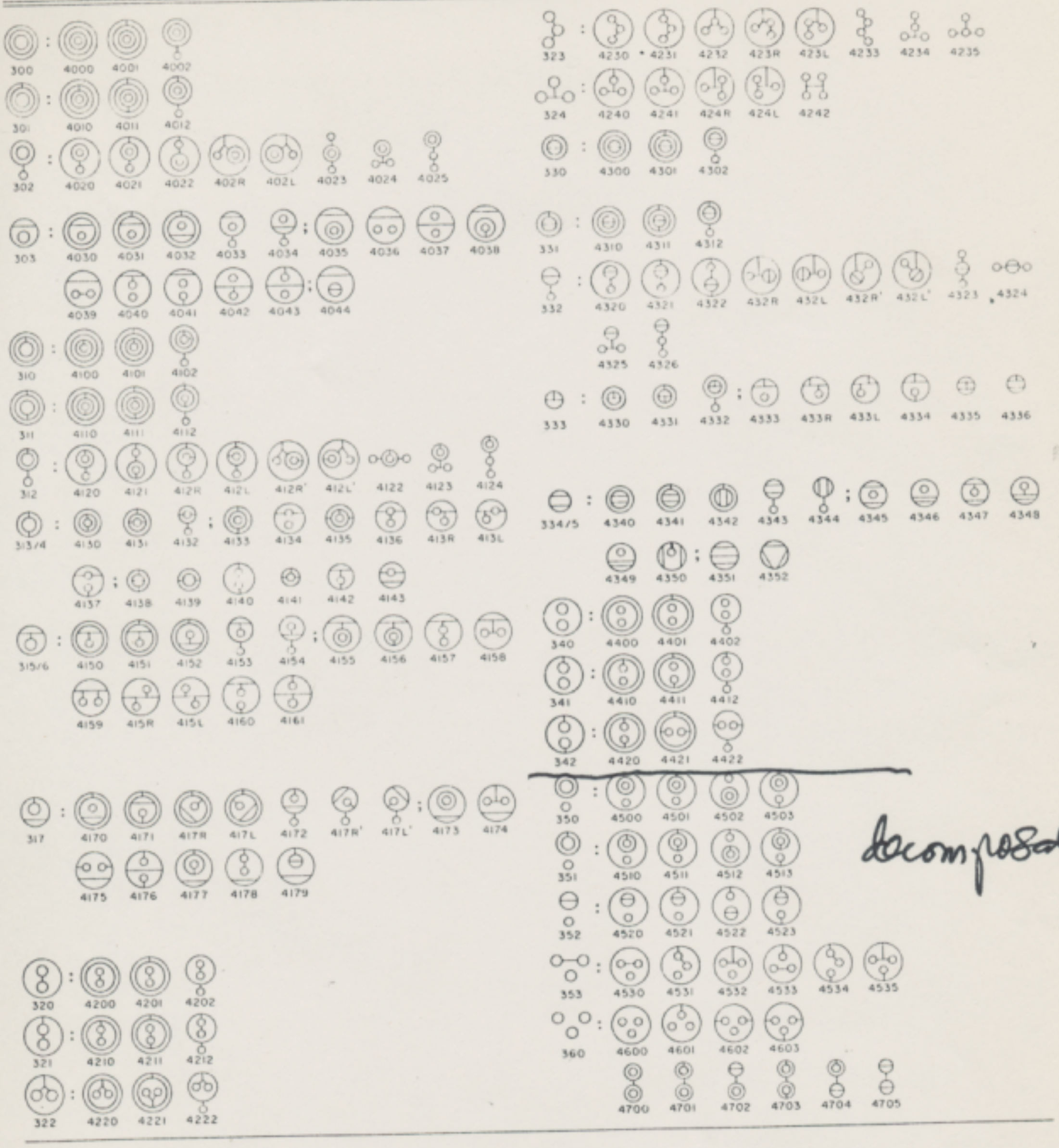
CONCLUSION

While all the patterns of up to four areas listed in the tables are topologically different, and therefore distinguishable from each other by operations based on nearest neighbor logic, such as the example given earlier, except for the last operation in which the number of isolated ONE's in a field must be counted, the task of determining all chains of simple operations required for a large multiplicity of topoglyphs looms formidable and the question is left open whether a general algorithm may exist on the basis of which a computer could be programmed to yield the individual algorithms or chains specific to each pattern.

Two considerations should be mentioned, which concern the overall efficiency of the system, and the amount of information which may be packed by means of topoglyphs in a hexagonal field of, say, 64 × 64 elements. The first has to do with error correction and/or error prevention. Error correction can be accomplished by suitable redundancy in the form of, e.g., the binary, etc., checks

28 rows
186 items

TABLE III



which are called for by the various coding schemes of information theory. And error prevention can be accomplished by having a limited number of initial blowing up operations placing ONE's in all ZERO's with surrounds of index 1 through 13 in order to bridge accidental gaps left in thin lines or remove accidental ZERO inclusions left in thick lines.

The second consideration concerns the "information content to size ratio" of the topoglyphs selected for various tasks, a ratio which is obviously more favorable in the case of small topoglyphs such as 4138 and 4141 than in the case of bulky ones such as 423R and 423L, which contain no more information. This consideration will obviously require that the five or six-area topoglyphs be looked at for ratios more favorable than certain four-area topoglyphs.

These considerations as well as software and hardware considerations indicate that a little more time and a little more effort will be required before every one in the USA, or in the world, be assigned at birth a topoglyphic number to be used as universally as a social security number.

REFERENCES

[1] Marcel J. E. Golay, "Hexagonal parallel pattern transformations," *IEEE Trans. Comput.*, vol. C-18, pp. 803–805, Aug. 1969.
[2] M. Ingram *et al*, "Automatic analysis of blood cells," *Scientific Amer.*, Nov. 1970.
[3] K. Preston, Jr., "Feature extraction by Golay hexagonal pattern transforms," *IEEE Trans. Comput.*, vol. C-20, Sept. 1971.
[4] K. Preston, Jr. *et al.*, "Interactive image processor speeds pattern recognition by computer," *Electronics*, Oct. 1972.

A New Algorithm for the Tangent

FRANK S. PRESTON

***Abstract*—A new mathematical algorithm has been developed for the tangent. The form of the equation guarantees that the error is zero at 0°, 45°, and 90° corresponding to tangents of 0, 1, and infinity. With only one constant the error is brought to zero at two more points and the maximum error is less than one part in 3000. By adding a second constant, the error is reduced to less than one in 720 000. Further terms improve the accuracy geometrically.**

This algorithm is primarily useful in analog computers and analog to digital angle resolution. The technique may be of interest for special purpose digital computers.

***Index Terms*—Analog computing, tangent approximation.**

Existing algorithms for the tangent are used over a limited range of the angle because of limited accuracy. Conventional schemes[1] have difficulty in dealing with the range of values from 0 to infinity. This problem can be avoided by using the following formula:

$$\tan\theta = \frac{\theta(90-\theta+K)}{(90-\theta)(\theta+K)} + \Delta,$$

where Δ is the error of the approximation. Examination will show that $\Delta = 0$ for any value of K at 0°, 45°, and 90°. This allows K to be selected to minimize the error between these three values.

It is more convenient to express the error as the equivalent angle where $\epsilon = \tan^{-1}\Delta$. With $K = 0$, ϵ is about 4° or 1 part in 20. This in itself is not very good but the error is symmetrical about 45° with maximums at about 17° and 73°. A value of 162 for K will reduce the peak error to zero errors at the mid value and two end conditions. Thus the simplest algorithm is given below:

$$\tan\theta \simeq \frac{\theta(252-\theta)}{(90-\theta)(162+\theta)}.$$

This single constant expression is always better than 2 min of arc (or 1 part in 2700) as the reader can easily verify with a few trials.

The accuracy can easily be further improved. By making the substitution of $90A$ for K (where $K = 1.8$) the equation can be simplified to the following:

$$\tan\theta = \frac{X(1-X+A)}{(1-X)(X+A)} \qquad \text{where } X = \theta/90°.$$

A plot of the remaining error looks like a two cycle sine wave crossing zero at the five points 0°, 17°, 45°, 73°, and 90°. Limit values can be found for A to force the slope of the error curve to zero at 0° (and 90°) or at 45°. Any value between these limits will produce a pair of zero error points whereas a value outside these limits will result in a single cycle error curve with zero at only 0°, 45°, and 90°. The two limit values are as follows:

For zero slope at 0° (and 90°) $A_0 = 2/(\pi - 2) = 1.75193998$.

For zero slope at 45° $A_m = \pi/2(4-\pi) = 1.82989618$.

An improvement in accuracy will be achieved by making A_k a variable which smoothly goes from A_0 to A_m to A_0 as θ goes from 0 to 45° to 90°. A parabolic approximation is suitable as follows. Let

$$A_k = A + BX(1-X) \qquad \text{where } X = \theta/90.$$

The constants which minimizes the error are as follows: $A = 1.7523351$, $B = 0.155460$. This always has less than one-half second or 1 part in 720 000 error. The two constants allow forcing the error to zero at two values between 0° and 45° (and two between 45° and 90°). The error curve looks like a sine wave of 3 cycles.

This accuracy exceeds analog mechanization capability and need not be further refined. (Three constants bring the error to less than 1 part in 32 000 000.) A variety of circuits and other analogs are possible using this algorithm. One will be mentioned for illustration. A bridge circuit is convenient for ratios as required by the basic equation. A single potentiometer with a shaft input of θ will produce the ratio $\theta/(90-\theta)$. If two fixed resistors of AR (where R is potentiometer resistance for 90°) are connected between the brush and each end the exact single constant expression is matched. The parabolic expression can be achieved with a shorted potentiometer whose value is BR for 90°.

This technique can be extended to other functions. There appears to be a real advantage of using products rather than sums of series for algorithms for machine and computer solutions of widely varying functions.

Manuscript received November 16, 1976; revised January 20, 1977.

The author is with the NORDEN Division of United Technologies Corporation, Norwalk, CT 06856.

[1] A more detailed discussion, derivation of values and references are contained in a companion manuscript placed in the IEEE Computer Repository.

Inference of Sequential Machines from Sample Computations

L. P. J. VEELENTURF

***Abstract*—This correspondence presents and justifies an algorithm for finding a minimal, though not a smallest, finite sequential machine with a behavior comprising a given finite sample of input–output behavior of some unknown sequential machine. When the number of states n of the machine to be identified is known and the sample contains all input–output pairs of length $2n - 1$, the machine found will be equivalent to the original machine and will be minimal.**

Manuscript received December 4, 1975; revised June 21, 1976.

The author is with the Department of Electrical Engineering, Twente University of Technology, Enschede, The Netherlands.