MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

A.I. MEMO 357 October 1976

Filling in the Gaps:

The Shape of Subjective Contours and a Model for Their Generation

Shimon Ullman

Abstract: The properties of isotropy, smoothness, minimum curvature and locality suggest the shape of filled-in contours between two boundary edges. The contours are composed of the arcs of two circles tangent to the given edges, meeting smoothly, and minimizing the total curvature. It is shown that shapes meeting all the above requirement can be generated by a network which performs simple, local computations. It is suggested that the filling-in process plays an important role in the early processing of visual information.

This paper is also to appear in *Biological Cybernetics*.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Project Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643.

Filling-in the Gaps: The Shape of Subjective Contours and a Model for Their Generation.

Abstract: The properties of isotropy, smoothness, minimum curvature and locality suggest the shape of filled-in contours between two boundary edges. The contours are composed of the arcs of two circles tangent to the given edges, meeting smoothly, and minimizing the total curvature. It is shown that shapes meeting all the above requirement can be generated by a network which performs simple, local computations. It is suggested that the filling-in process plays an important role in the early processing of visual information.

This paper is to appear also in *Biological Cybernetics*.

Introduction

Subjective contours are perceived when the visual system fills in the gap between distinct edges [Brigner and Gallagher 1974, Coren 1972, Coren and Theodor 1975, Gregory 1972, Gregory and Harris 1974, Kanizsa 1976, Schumann 1904]. Other examples of filling-in processes are the perceived trajectory in optimal Beta motion and the continuation of the visual field across retinal scotomatas. The study of these filling-in processes has concentrated to date on the *triggering conditions*. For instance, under what conditions will a complete trajectory be perceived between successively presented stimuli, or what conditions will enhance the generation of subjective contours between distinct edges. Another

important but hitherto neglected problem posed by the filling-in phenomena concerns the *shape* of the filled-in contours and trajectories. Since subjective contours are synthesized by the visual system the examination of their shape might provide clues about the nature of the mechanism that generates them. In this paper I shall first examine the shape of subjective contours and then suggest a network capable of generating these shapes. It will be shown that a network with the local property of trying to keep the contours "as straight as possible" can produce curves possessing the global property of minimizing total curvature.

Section 1: The shape of subjective contours.

Of the infinitely many loci passing through two given edges, which is the one chosen by the visual system? Let me start the quest for an answer by stating three observations and an hypothesis which will serve as guidelines for narrowing down the range of possible shapes to a unique one. The resulting curves will then be compared with actually perceived filled-in contours. The four guidelines are:

1. Isotropy: As far as the filled-in contours are concerned the visual field seems approximately homogeneous and isotropic: The filled-in contours produced by a given figure do not change in shape when the figure is translated or rotated.

2. Smoothness: except for some special cases of filled-in corners, the generated curves are smooth, that is, differentiable at least once.

3. Minimum curvature: This guideline is inspired by the resemblance of the filled-in contours to a thin doubly cantilevered beam or, alternatively, to the curve known in approximation theory as a cubic spline. Both curves are, in some sense, loci of minimum

curvature [Ahlberg, Nilson and Walsh 1967, Sokolnikoff 1956]. The shape of a subjective contour passing through two given edges whose orientation difference is much less than π/2 closely resembles the cubic spline passing through this boundary edges. At this point I only wish to state this resemblance informally; the minimum curvature property will receive a formal treatment in the sequel.

4. The locality hypothesis: The operation by which a boundary edge is extended to generate a subjective contour is assumed to be local in nature. That is, it depends only on the end part of the given edge to be extended and not on the shape of the entire edge. This hypothesis is based in part on experimental observations, and partly on a theoretical consideration that is addressed in the concluding discussion.

The experimental observation is diagrammed in figure 1. Figure 1a depicts a subjective contour b between two boundary edges A and B. In 1b an edge E is added on b, and in 1c a new edge which we shall call A-E occupies the entire portion of b from A to E. The shape of the filled-in contours remains the same in all the three variations. It is assumed here that a contour between edges E1 and E2 is created by generating a set of possible extensions of E1, another set extending E2, and then choosing one of the resulting curves connecting E1 and E2. The fact that the contours produced by E (figure 1b) and by A-E (figure 1c) are identical is readily explained if we adopt the locality assumption, since according to this hypothesis the extension in both cases depends on E alone.

The four guidelines are sufficient to resolve the shapes of the filled-in contours.

Let the set of possible extensions at A be L1, L2,... ...,Lk. One of them, say L1, together with a curve coming from B produce the contour b. If we now replace A in figure

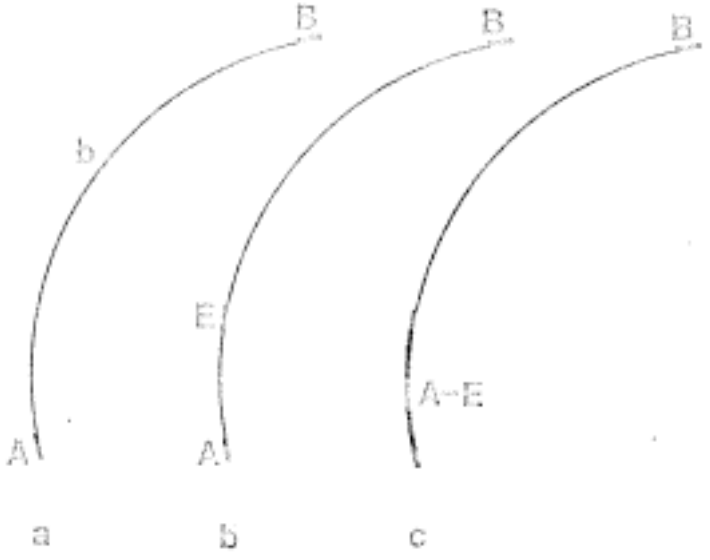


Figure 1

Adding the edges E or A-E does not change the shape of the filled-in contour b.

la by the extended edge A-E (figure 1c) the contour does not change in shape. Since in the second case the contour is produced by E (the locality assumption) we deduce that E-LI, the portion of LI from E onwards, is also one of the "repertoire" of possible extensions of E. From the isotropy property, E-LI is also congruent to some Li, 1≤i≤k. The set of curves L1,... ...,Lk thus possesses the following property. If you take any curve Li and cut it at some point P, then the shape of P-Li, the continuation of Li from P onwards, is congruent to one of the curves L1,... ...,Lk. One can meet this requirement by taking all the Li's to be *arcs of some circles,* (including the limit case of a straight line) since any part of a circle is congruent to any other part of equal length. Under certain conditions, the circles solution is the *only* set of curves obeying the above requirement, however I shall not diverge here to examine these conditions.

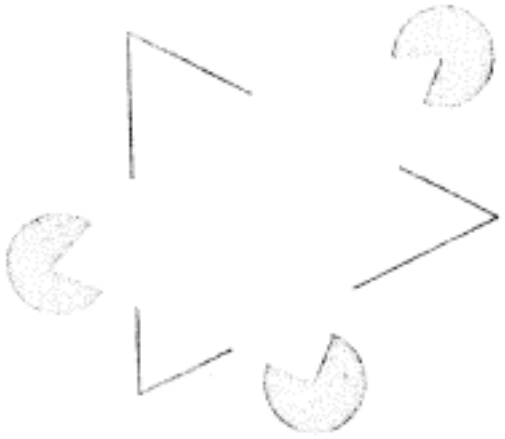
Since the produced curves must all be *tangent* to the edge A to meet the smoothness requirement, the prediction is that the filled-in contour is composed of the arcs of two circles, one tangent to one edge, the other tangent to the second edge. An additional restriction is imposed by the smoothness guideline, which requires that the two arcs share a common tangent at their meeting point. There is still, however, an infinite number of arc-pairs satisfying all the above conditions. Following the minimum curvature hypothesis let us pick, out of all the admissible pairs, the pair which minimizes the total curvature. The total curvature of a curve is defined here as the integral $\int (d\alpha/dl)^2$ over the curve, where α is the slope of the curve. The value of $d\alpha/dl$ at a given point P is known as the local curvature at that point. In the sequel, however, the word "curvature" will refer to total curvature unless otherwise stated.

Since the shape analysis is carried out here while trying also to account for its generating mechanism, an important problem that naturally arises at this point is whether a measure of the total curvature can be computed in a locally connected network. The answer, which is positive and surprisingly easy is discussed in the subsequent section where a network model is presented. Before turning to the model let us compare few examples of contours generated by the outlined method to those generated by the visual system.

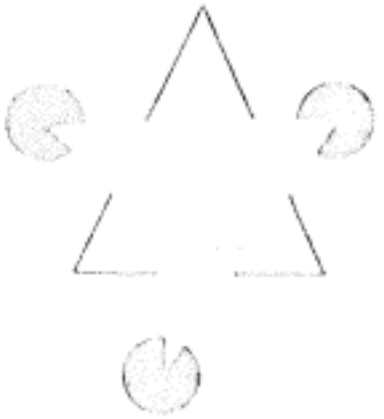
Figure 2 is an example of figures that generate subjective contours. The contours are enhanced by supporting triangles that have a little affect on the perceived shape. (If the supporting edges are too long or too short the subjective contour does not touch them, and it also diminishes in strength.) The shape of the contour can be traced by placing a transparent paper on the figure and marking the points the contour seems to pass through. Figure 3a compares the resulting contour for figure 2a averaged over 10 subjects with the curve generated by the above two-arcs method from the two boundary edges (dashed line). The two-arcs curve is depicted in figure 3b togher with the two generating arcs: arc 1 whose center is at c1 and arc 2 whose center is at c2.

Figure 3b was copied from the display unit of a PDP-10 computer system using a Tektronix 4632 hard-copy unit. It had been computed in the following way: The set of arc-pairs can be described as a one-parameter family where the parameter is β, the angle between one of the edges and the line from this edge to the meeting point of the circles. The minimum curvature pair was found by computing the curvature for 1 degree increments of β.

Figure 4 compares some other actually perceived subjective contours with the ones



a.



b.

Figure 2

Figures that produces subjective contours.

Compare 2a with figure 3 and 2b with curve 1 in figure 4.

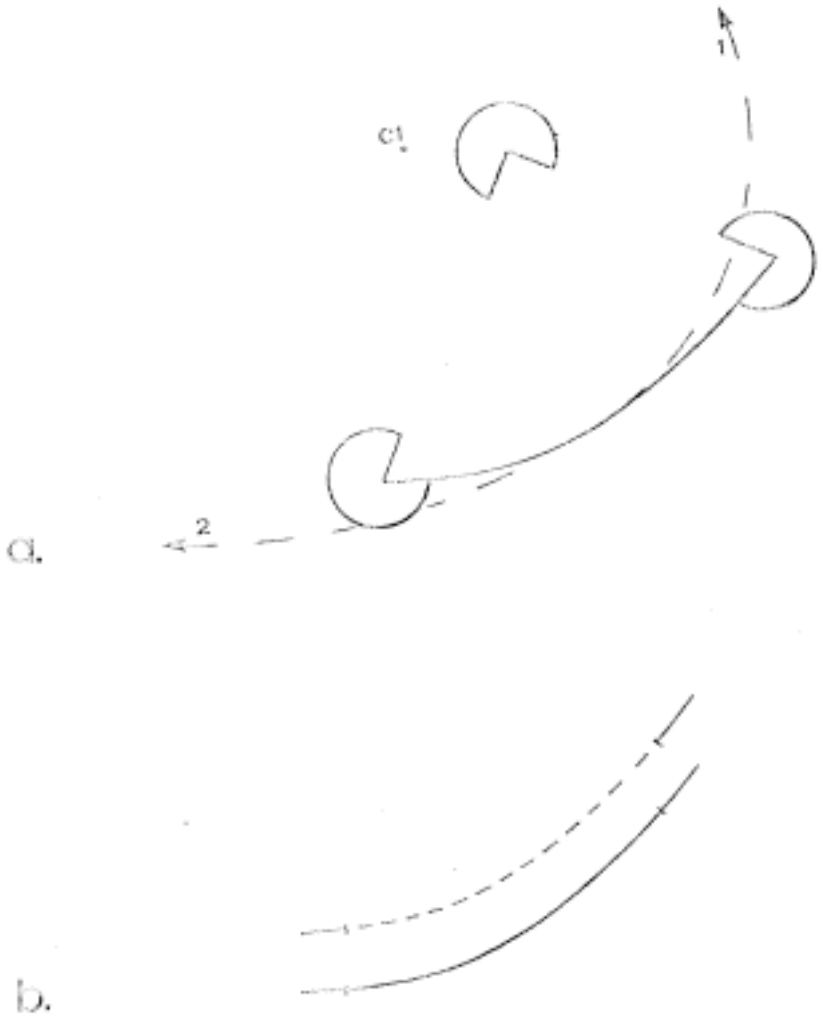


Figure 3

3a: A comparison between an actually perceived subjective contour and a contour generated by the two-arcs method (dashed line).

3b: The two arcs that generate the contour in 3a.

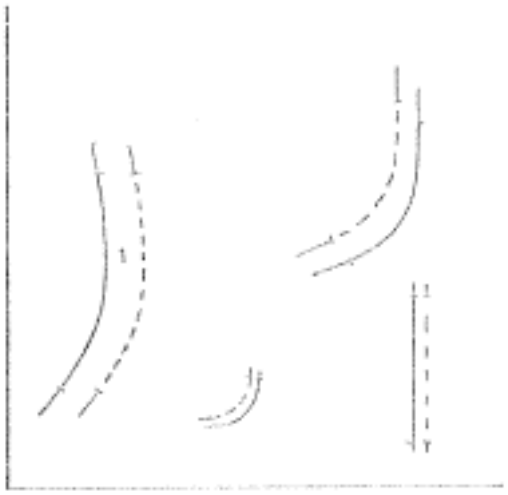


Figure 4

A comparison of actually perceived subjective contours with the ones produced by the two-arcs method (dashed lines).

Figure 5

Both a subjective circle and a subjective square are perceivable.

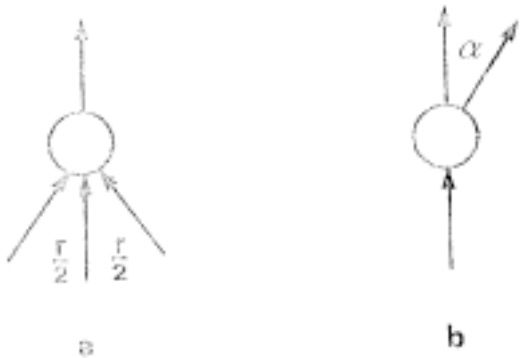
generated by the method under the same circumstances. The contour marked as 1 in figure 4 is produced by the figure in 2b. In figure 5 both a subjective circle and a square are perceivable. The circle is in accordance with the method, the square is addressed in the subsequent section.

Section 2: A network model for generating subjective contours.

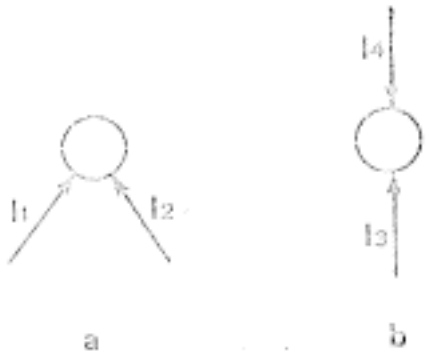
This section outlines a network model capable of generating two-arcs, smooth, minimum-curvature contours between two given boundary edges. As the shape analysis in the preceding section was based in part on the locality assumption, a main goal of the model is to verify that the proposed curves of minimum total curvature can be generated by a net performing only local computations. The net is somewhat serial in nature to facilitate the study of its properties. However less serial networks performing the same computation are not inconceivable.

The network is constructed in three layers. The first two generate the circles tangent to the boundary edges, and the third picks the minimum curvature contour.

The first layer: The building-blocks of the net are *orientation elements*. The net is a grid of points, and at each point P there are a number of orientation elements coming into P from neighboring points, and the same number of orientation elements leaving P to nearby points (The number of the elements at each point depends on the desired angular resolution.) Each output element at P is connected to all the input elements within a range of some r degrees from it, i.e. all the incoming elements whose orientations are between $-r/2$ and $+r/2$ of the output orientation (figure 6a). The basic operation in the network is simply



6.



7.

Figuras 6 and 7

6a: An output element is connected to the input elements within an orientation range of r degrees.

6b: The excitation depends on the angle between the input and the output elements.

7a: The input elements I1 and I2 cannot be active simultaneously.

7b: I3 and I4 lie on the same orientation and can be active simultaneously. They constitute an *orientation pair*.

a local preference for straight lines: each element excites mostly its collinear neighbor. If I is an orientation element entering P from a nearby point, the excitation of the elements radiating from P will decrease with the angle α between the input and the output elements (figure 6b). The decreasing function of α is taken to be $(1 - k\alpha^2)$ for some constant k, that is, the excitaion E(Oi) is $I_0(1 - k\alpha^2_i)$. We shall later see that the exact form of the function is not crucial at all. At a given point P where many orientation elements fan both in and out of P, the excitation of a given output element Oj is the maximum of all the inputs connecting to it; namely: $E(Oj) = \text{Max } I_0(1 - k\alpha^2_i)$, where i ranges over all the elemetns connected to Oj. A boundary edge A serves as an input to the network by exciting from outside the net one of the orientation elements in the first layer. The excitation scheme can then be viewed as a real function E(P,O) which gives the excitation at a point P of some orientation element O.

The second layer: At the second layer each point P accepts input from a *unique* orientation by taking the maximum of the inputs to P in the first layer. If I1 and I2 (fig 7a) are two orientation elements entering P (in the second layer) it is impossible that both will be active. However I3 and I4 (figure 7b) can be active simultaneously as they lie on the same orientation. We shall call such a pair of elements at a point, differing by 180 degrees, an *orientation-pair*. If A is now the input to the network, the excited elements in the second layer will form the shape of *circles tangent to A*. I shall briefly sketch the proof of this claim.

(a). Let O1 be an orientation element at some point P1 in the first layer, and E(O2) be the excitation of some other element O2 at a point P2. Let L denote some arbitrary path (a

sequence of connected elements) from O1 to O2, and let ΠL denote the product $\Pi(1 - \alpha^2 j)$ of all the angles αj between successive elements along L. Then $E(O2) \geq E(O1) \cdot \Pi L$.

Proof: Let B1, B2,... ...Bn be the intermediate elements between O1 and O2 along L. $E(B1) \geq E(O1) \cdot (1 - k\alpha^2_1)$ because E(B1) is the maximum of many terms one of which is $E(O1) \cdot (1 - k\alpha^2_1)$. For the same reason $E(B2) \geq E(B1) \cdot (1 - k\alpha^2_2)$ ... ...$E(O2) \geq E(Bn) \cdot (1 - k\alpha^2_{n+1})$. Hence,

(1) $E(O2) \geq E(O1) \cdot (1 - k\alpha^2_1) \cdot \ldots \cdot (1 - k\alpha^2_{n+1})$, that is,

(2) $E(O2) \geq E(O1) \cdot \Pi L$.

(b) Let L be a locus of excited elements in the second layer radiating from A and passing through O1 (figure 8) (It can be easily verified that from any excited element a unique locus of excited elements can be traced back to A.) Consider two situations. An A-on situation where A is the input to the net, and an A-off situation where A is "turned off" and the new input to the net is obtained by exciting O1 by the same E(O1) it had in the A-on situation. Then, for every element O in the first layer, E(O) in the A-on situation $\geq$ E(O) in the A-off situation. Also, for the elements on L, E is the same in the two situations. The proof of this claim follows from (a) and will not be detailed here.

(c) The loci of excited cells radiating from O1 in the second layer is the same in the A-on and the A-off situations. To verify this claim consider two inputs to a point P along L: an input I along the curve L, and an input I' from outside the curve (figure 8). I being on the curve means that E(I) > E(I') in the A-on situation. From (a) and (b) it follows that E(I) > E(I') in the A-off situation as well. This implies that the curve L between P and O1 remainer the same in the A-off situation, for an arbitrary point P along L.

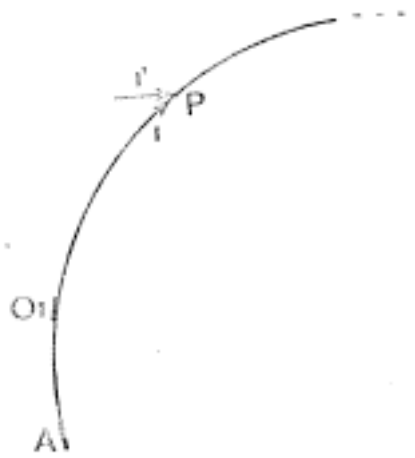


Figure 8

The shape of the curve from O1 onwards is the same in the A-on and the A-off situations.

dL (mm)	0.1	0.25	0.5	1.0	2.0
$nL \times 10^{-3}$	1.52	1.52	1.55	1.56	1.56

The total curvature is 1.514×10^{-3}

Table 1.

Approximating the total curvature by the product $\Omega L = n(1 - \alpha/d)$, as a function of the chord-length d (in millimeters).

(d) The excitation pattern of the net does not depend on the actual magnitude of the excitation A. This is true because the curves are determined only by comparing expressions of the form $A\alpha\|Li$ for different trajectories Li.

(e) The network's structure is homogeneous and isotropic, therefore if A is shifted or rotated the excitation pattern of the entire net will shift and rotate with it.

The conclusions from (a) to (e) are as follows. If A is turned off and E(OI) on, the shape of L from OI onwards is unaltered. The same holds when OI is excited by an amount equal to A instead of E(OI). Together with the isotropy property this entails that the shape of L from OI onwards is congruent to the shape of L from A onwards. Since this is true for an arbitrary OI along L, L must be an arc of a circle (including the limit case of a straight line).

The discussion so far concentrated around single input situations, i.e. A was the only input. When there are two input edges A and B, the only complete trajectories connecting A and P on the second layer will be composed of two arc-circles meeting smoothly. The reson is that only at such meeting points the element belonging to the A-curve, and the element belonging to the B-curve, constitute an orientation pair and can be both copied to the second layer.

The third layer: at the third layer the trajectory which minimizes the total curvature is picked up. It had been claimed in the previous section that the curvature computation can indeed be carried out by the proposed network. In fact, the excitation E(O) itself is inversely proportional to the total curvature of the trajectory from A to O, and can serve as an accurate measure of the total curvature.

Proof. Let L be a phi-degrees arc of some circle of radius R. Devide the arc into n equal chords of length dl each, and let αj denote the angle between elements j and j+1. Then the limit of the product $\Pi(1 - \alpha^2 j/dl)$ when dl → 0 is Exp. -c (e^{-c}) where c is the total curvature of the arc. While this claim holds true for other curves as well it is easily verifiable for a circle where all the αj are equal to phi/n. The logarithm of the above product is

(3) $n\log (1 - \alpha^2/dl)$.

Expanding to a Taylor series one gets:

(4) $-n\ \alpha^2/dl\ +$ error.

The first term is equal to -phi /ndl which converges to -phi/R, compared to the total curvature of the arc which is phi/R. The error term for large n is bounded by 2phi /nR , which vanishes in the limit.

It is instructive to observe how well the curvature-product approximates Exp. -c. For the contour in figure 3 the total curvature is 1.514 ×10 . The minus logarithm of the curvature product $\Pi(1 - \alpha^2/dl)$ is given in table I for different values of dl. It had been mentioned that the excitaion function does not have to be exactly $(1 - k\alpha^2)$. Compare for instance $\pi(1 - k\alpha)$ along two curves from A to B. Expanding once again the logarithm of this product to a Taylot series, the first term is identical for all the curves in question, the second term gives the curvature measure, and the error term is much smaller than the second term if dl and α are small. Note also, that the exact form of the function was immaterial in the circle-generation phase. A final remark concerning the tolerance of the computaion to inaccuracies is that if instead of the minimum curvature contour an arc-pair

with somewhat higher curvature is chosen, the resulting contour will be close spatially to the minimum curvature one.

The minimum curvature trajectory can now be identified. Let the third layer cell at P summate the products E(O1)•E(O2) where O1, O2, are an orientation pair at P in the second layer. The only third layer cells with non-zero activity will be at the points where the two arcs, one from A and the other from B, meet continuously. Furthermore, the third layer activity at P will measure Exp. -c of the *entire* curve connecting A and B and passing through P. What remains to be done is to erase all the contours but the one passing through the most excited third layer cell. Perhaps the most straightforward way of accomplishing this task is by copying the desired contour alone onto a fourth layer by starting at the most excited cell and tracing the two arcs all the way back to A and to B. I shall not concern myself here with such possible copying mechanism as it bears no direct relevance to the problem at hand.

One modification might, however, be required to account for the generations of subjective *corners*. If, for example, the third layer cells accept not only orientation-pairs separated by 180 but also pairs separated by 90 , then when the radius of the minimum curvature curve becomes too small, it will be replaced by a cognitive corner. Since this break-down phenomenon depends on the radius of curvature, it should be expected that in a figure like [5], where both a circular and a square subjective contours are perceivable, the square will predominate with increasing viewing distance (and decreasing radius), as is indeed the case.

Discussion.

Is the filling-in process a special-purpose, rarely activated mechanism, or does it play a coromon role in the visual process? It seems likely to me that it does in fact take an active part in the early processing of visual information. When one actually tries to process an image, one is confronted with the problem that the resulting contours are fragmented, separated by many gaps. This difficulty can be attributed only in part to the imperfectness of the processing techniques. It is an inevitable problem as long as the data and the computation are not noise-free. One of the main problems in early visual processing is thus giving the fragmented contours by correctly filling in the gaps between them [Marr 1975]. The filling-in operation is thus expected to be largely independent of high-level object and figure recognition. Higher level modules seem to interact with the subjective contour generation mainly, though perhaps not solely, via the *triggering conditions*, i.e. the decision which edge should be extended can be affected by such considerations as obscuration and figure completeness. It is also plausible that a major characteristic of the filled-in contours will be a low curvature. At the stage where the filling-in takes place very little is yet known about the contour's shape. It is only known (or assumed) that the contour of a *single* object should extend across the gap. It had been indicated in the past that the partition of an image to sub-parts is largely based on points of high curvature [Attneave 1954, Rosenfeld 1969]. Conversly, "one-objectness", when nothing is known about the object, is indicated by the lack of such points.

Acknowledgements: I thank Dr. D. Marr and Dr. T. Poggio for their comments.

References

Ahlberg, J. H., Nilson, E. N., and Walsh, J. L.: The Theory of splines and Their Applications. New York and London: Academic Press 1967

Attneave, F.: Some informational aspects of visual perception. *Psychological Review, 61*, 183-193 (1954)

Brigner, W. I. and Gallagher, M. B.: Subjective contour: Apparent depth or simultaneous brightness contrast? *Perceptual and Motor Skills*, *38*, 1047-1053, (1974)

Coren, S.: Subjective contours and apparent depth. *Psychological Review, Vol. 79*, No. 4, 359-367 (1972)

Coren, S. and Theodor L. H.: Subjective contour: The inadequacy of brightness contrast as an explanation. *Bulletin of the Psychonomic Society, 6*, 87-89 (1975)

Gregory, R. L.: Cognitive contours. *Nature, Vol. 238*, No. 5358, 51-52 (1972)

Gregory, R. L, and Harris, J. P.: Illusory contours and stereo depth. *Perception and Psychophysics, Vol. 15, No. 3*, 411-416 (1974)

Kanizsa, G.: Subjective contours. *Scientific American, Vol. 234*, No. 4, 48-52 (1976)

Marr, D.: Early processing of visual information. *M.I.T Artificial Intelligence Memo No. 340* (1975)

Rosenfeld, A.: Picture Processing by Computer. New York: Academic Press 1969.

Schumann, F.: Einige Beobachtungen uber die Zusammenfassung von Gesichtseindrucken zu Einheiten. *Psychologische Studien, 1*, 1-32 (1904) (Cited in C. E. Osgood: Method and Theory in Experimental Psychology, New york: Oxford University Press 1971)

Sokolnikoff, I. S.: Mathematical Theory of Elasticity. New York: Wiley, 1956