The Role of Key-Points in Finding Contours *

O. Henricsson and F. Heitger

Communication Technology Laboratory,
Swiss Federal Institute of Technology ETH
CH-8092 Zürich, Switzerland

Abstract. This paper describes a method for aggregating local edge evidences into coherent pieces of contour. An independent representation of corner and junction features provides suitable stop-conditions for the aggregation process and allows to divide contours into meaningful substrings, right from the beginning. The active role of corner and junction points makes the contours converge onto them and greatly reduces the problems associated with purely edge-based methods. A second stage is concerned with completing established contours across regions that are less well-defined by contrast. The algorithm suggested uses the attributes of established structures (e.g. direction of termination) as well as local orientation and edge evidences to constrain possible completions in a rigorous way.

Keywords: edge detection, key-point detection, edge linking, contour completion

1 Introduction

Intensity discontinuities are considered one of the primary image features that allow to segment a scene into meaningful parts. Based on the assumption that object boundaries are generally smooth and mostly contrast defined, much effort has been devoted to design suitable edge detectors (e.g. [3]) that reliably indicate these 1-D intensity discontinuities. The deficiencies of edge-maps, such as fragmentation, gaps at junctions, as well as clutter and faulty connections are well known. Also, object boundaries are not guaranteed to be contrast defined. To obtain more complete and unambiguous boundary definitions, additional processing is needed that accounts for more global relationships among image features.

Perceptual grouping methods have demonstrated a promising potential in this respect. Most approaches use geometrical criteria (e.g. distance, co-curvilinearity, symmetry) to group local edge evidences into larger entities, thus also separating salient structures from clutter [11, 19, 14, 6]. The use of binary edge-maps as input, however, neglects information that could assert the validity of connections on grounds other than geometry.

2-D image features, such as junctions, corners and line-ends represent an-

* The research described in this paper has been supported by the Swiss National Science Foundation, Grant no. 20-36431.92

other important class of image information that can serve the definition of object boundaries. With respect to this task the role of these features is a dual one: First, they reflect prominent events in the course of a boundary that allow to divide it into "natural" parts and to extract meaningful shape decompositions [2, 13]. A corner or a junction together with the directions of their constituent components often characterizes an object more succinctly than edge fragments. Fig. 1 illustrates this aspect. Second, 2-D features occur abundantly in situations of occlusion and within this context they can serve to indicate object contours even if the contrast is vanishing or null. A model of visual cortical contour processing that infers such contours as well as figure ground direction from termination evidences has been presented earlier [9]. Because the 2-D features are so significant, we use the term *key-points*.

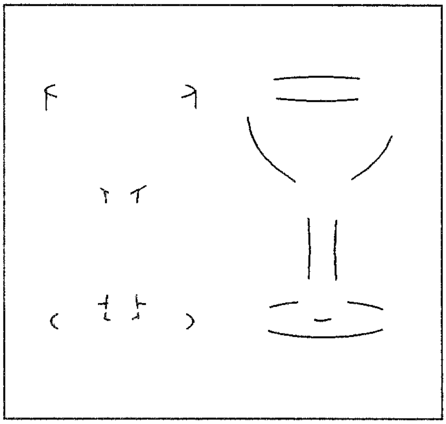

Fig. 1.: The information at corners and junctions (left) is often more important for object definition than edge fragments (right).

In this paper we want to discuss the role that an independent representation of such key-points can take in aggregating local edge evidences to larger, coherent pieces of contour (called *contour-strings* in the sequel). The goal is to represent the contrast defined image features as a collection of meaningful parts, subdivided at locations of high key-point evidence. Of course, this idea is not new, but previous implementations used post-processing of binary edge-maps to achieve segmentation (e.g. [13]), rather than utilizing independent representations of 2-D features. We will show that the complementary nature of edges and key-points can provide a better and more stable definition of image structures.

In a first instance key-points serve as stop-conditions for the aggregation process which is started at points that can most reliably be classified as "edge" points. In a second stage, the key-points that are connected to contour-strings serve as "bridge-heads" for closing gaps across regions of low or vanishing contrast. Connections are only accepted if they satisfy a variety of geometrical constraints, but also provide evidence for residual contrast definition.

For the moment, we deliberately exclude completions of the "illusory contour" type as described in [9]. The boundaries obtained with the present approach are therefore still incomplete. However, this complies with the general philosophy of the present approach: Contour-strings are established in a strictly hierarchical fashion, starting with the most reliable ones and using the information of already established structures to expand into more uncertain regions. Decisions are adapted to the level of uncertainty, with weaker evidences requiring more constraints to satisfy than stronger ones.

2 Filtering and Key-point Detection

2.1 Filters

We convolve the image with filters of even and odd symmetry (6 orientations, channels) and combine their output to response modulus (square-root of oriented energy), an approach similar to [5, 16, 15, 1, 17]. The filters have the following properties (see [8, 18] for a detailed description): They form a quadrature pair, are polar separable in the Fourier domain, and the response modulus yields a unified response to edges and lines.

In this paper we use both the modulus representations and their 2nd derivatives perpendicular to channel direction. The second derivative enhances the negative curvature occurring at modulus peaks and has further shown smaller gaps at junctions with non-maximum suppression. We denote the modulus maps with $\mathcal{M}$ and the second derivative maps $\mathcal{M}^{(2)}$.

Non-maximum suppression is applied on clipped 2nd derivative channels, i.e. local maxima in a direction perpendicular to the dominant channel orientation. This binary map (denoted $\mathcal{N}$) is used as a seed structure for the contour aggregation.

2.2 Key-points

Key-points are defined as strong 2D intensity variations, i.e. the signal not only varies in one direction, but also in other directions. We implemented a detection scheme based upon a model of visual cortical end-stopped cells.

In principle, the 1st and 2nd derivatives in the direction of modulus channels are used, the 1st derivatives being sensitive to the termination of oriented structures (line-ends, corners, junctions) and the 2nd derivatives to blobs and strong curvature. 1st and 2nd derivatives are combined to localize the key-points. A compensation map is used to eliminate spurious responses to 1D structures (straight edges, lines). The key-point detection scheme is described in detail in [8, 18]. We denote the set of detected key-points $\mathcal{K}$ and the 3×3 surround by $\mathcal{K}_s$. Reliable detection as well as accurate localization of key-points is a prerequisite for the contour aggregation as described below.

2.3 Local Orientation and Edge Quality

The response modulus in the six channels is used to determine the local orientation of the underlying structure. We use the real and imaginary coefficients of the first Fourier harmonic to approximate the local orientation, (similar to the approaches of [4, 10]).

$$\theta_{loc} = \frac{1}{2} tan^{-1}\left(\frac{\mathcal{I}m}{\mathcal{R}e}\right), \qquad where \quad \theta_{loc} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$$

Because we use filters that are polar separable in the Fourier domain, the response magnitudes in the different orientation channels are entirely determined by the orientation of an edge/line and the orientation tuning of the filters, defined as $\Omega(\psi) = cos^n(\psi)$.

The residual between the actual response distribution and the edge/line prediction is then used as a measure for edge quality (cf. [18] for details).

$$Residual = \sum_{k=0}^{nori-1} \left(\frac{\mathcal{M}_k}{\mathcal{M}_{max}} - \frac{\Omega(|\theta_{loc} - \theta_k|)}{\Omega(min_{j=0,\dots,nori-1}|\theta_{loc} - \theta_j|)} \right)^2$$

$$Q(\theta_{loc}) = \frac{1}{1 + Residual} \quad , [0 \leq Q(\theta_{loc}) \leq 1] \tag{1}$$

Edge quality will be used for selecting appropriate start points for contour aggregation as will be explained below.

3 Contour Aggregation

The contour aggregation algorithm can be described as a process linking initial edge-markings into coherent contour-strings. The selection of the appropriate track is based upon (1) connectivity, (2) modulus strength, and (3) key-point markings.

3.1 Selecting Start-points

Adequate start points are positions with high edge quality and low influence from surrounding key-points. Such points have per definition (1) a well defined local orientation, as needed for the start condition.

The normalized difference of modulus and key-point value is used to define the key-point influence. The product of edge quality and key-point influence yields a suitable start measure,

$$\mathcal{S}(x,y) = \begin{cases} Q(x,y)^2 \cdot \left(\frac{\mathcal{M}_{max}(x,y) - \mathcal{K}(x,y)}{\mathcal{M}_{max}(x,y) + \mathcal{K}(x,y)} \right) & ,(x,y) \in \mathcal{N} \\ 0 & ,otherwise \end{cases}$$

where $\mathcal{M}_{max}(x,y)$ is the response modulus in the dominant orientation channel, and $\mathcal{K}(x,y)$ the key-point map value. Notice that only points marked by non-maximum suppression ($\mathcal{N}$) are used as start points.

The start points are transferred to a sorted list, the first entry being the point with the highest start value. The contour aggregation algorithm successively picks the currently best start point from the sorted list. After a contour segment has been established, all start points along the segment are eliminated to prevent multiple chaining of the same contour-string.

3.2 Chaining Algorithm

Contour aggregation (pixel chaining) is done by locally evaluating a small set of valid paths within directional masks as depicted in Fig. 2. The evaluation is a two stage process, (1) a priority is assigned to each path, and (2) a path value depending on the priority is assigned, as shown in Table 1. The current position P_0 and the chaining history defines the chaining direction, $\alpha \in [-\pi, \pi]$. At the start point, the chaining direction is initialized as one of the two opposing directions defined by the local orientation map. Each α is associated with a two-level directional mask j, defined in eight possible directions ($\beta(j) = j\frac{\pi}{4}$). Within each mask there are nine distinct paths extending from P_0 through a pixel P_1 in level one to a neighboring pixel P_2 in level two. Each path $i = 0, 1, \ldots, 8$ within the mask j is thus defined by the triplet $[P_0, P_1(i), P_2(i)]$ and evaluated according to the table below. The function f penalizes paths deviating from the current chaining direction α. In a first selection the path(s) with the highest priority is(are) selected. If there is more than one path, the value of $E(i)$ determines the selection. The selected path i' defines the next pixel $P_1(i')$ and an updated chaining direction α'. The chaining continues until one of the following stop conditions is encountered.

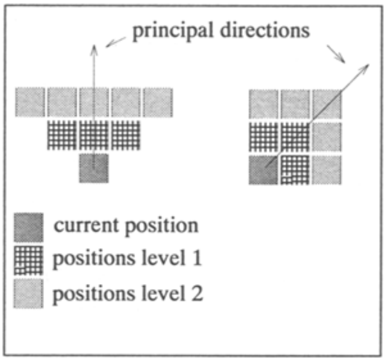

Fig. 2.: Two-level directional masks. The other six masks are only rotated versions of the two.

- $P_0 \in K$, the current position is a key-point. At priority level 3 the chaining algorithm captures the key-point position in a deterministic fashion and the stop condition is set.
- priority 0, termination without key-point marking.
- collision with another, already established, contour-string.

When the chaining algorithm encounters a stop condition it generates a stop marker of the corresponding type.

Priority	Condition	Evaluation/Action
3	$P_0 \in K_s$	Key-Point capture, no path evaluation.
2	$(P_1(i) \in \{\mathcal{N}, \mathcal{K}_s\}) \cap$ $(P_2(i) \in \{\mathcal{N}, \mathcal{K}, \mathcal{K}_s\})$	Normal chaining, evaluating $E(i) = \left(\mathcal{M}^{(2)}_{max}(P_1(i)) + \mathcal{M}^{(2)}_{max}(P_2(i))\right) \cdot f(\beta(j), \gamma(i))$ $\gamma(i)$ is the direction of the path i and $f(\beta(j), \gamma(i)) = cos(\beta(j) - \gamma(i))$.
1	$(P_1(i) \in \mathcal{N}) \cap$ $(P_2(i) \ni \{\mathcal{N}, \mathcal{K}, \mathcal{K}_s\})$	terminating in the next step, evaluating $E(i) = \mathcal{M}^{(2)}_{max}(P_1(i))$
0	$P_1(i) \ni \{\mathcal{N}, \mathcal{K}, \mathcal{K}_s\}$	path terminated

Table. 1. Chaining algorithm.

The chaining algorithm generates contour-strings and stop markers in a connected fashion, inferring a graph-like data structure. Each contour-string is delimited at both ends with a key-point. The order of a key-point is defined as the number of contours strings connected to it. Apart from cross references, semi-global attributes are assigned to contour-strings and key-points (Table 2).

3.3 Post-processing of Established Contour-strings

The established contour-strings can still have strong orientation discontinuities. We therefore divide each contour-string with orientation discontinuities into substrings by using the algorithm suggested by Medioni [13]. The points marked by

feature type	attribute
contour-string	- length, - integrated modulus response (contrast) - type and polarity of contrast, - termination directions (linear and quadratic fits)
key-point	- order = number of connected contours, - termination direction of connected contour, - key-point value (contrast dependent) - key-point type

Table. 2. Attributes of contours and key-points.

the algorithm are further tested by linear fits on the contour-strings to either side of the marked point. If the angular difference of the two opposing directions is large enough ($> \frac{\pi}{4}$), the point is accepted as an additional key-point.

Contour-strings that are connected to a key-point of order ≥ 2 and that have a low integrated modulus value when compared to the remaining contour-strings are pruned. An additional requirement is that the contour-string is not connected to any other structure. Thus, unnecessary high orders of key-points due to these spurious contour-strings are precluded. Furthermore, short and isolated contour-strings are pruned as well if their integrated modulus is below a given threshold (we use 1% of the global average). All pruned structures are transferred to a stack and can be used for later processing. The pruning of spurious contour-strings is important with respect to gap-closing and a robust vertex classification.

4 Bridging Gaps Supported by Contrast Evidences

Recapitulating the process history of the present contour representation, we started with the initial edge map and applied the chaining algorithm as an aggregation process yielding a graph-like representation of the contours. The key-points were used as natural stop-markers during the chaining process. A pruning stage was applied to retain only significant structures.

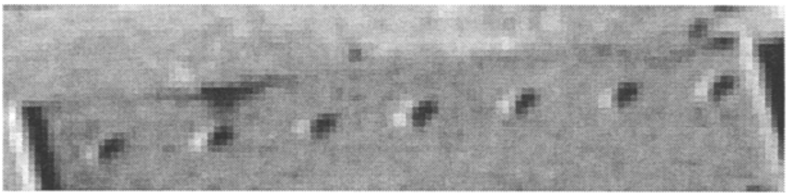

Fig. 3. Cut-out from the aerial image in Fig. 7, showing poor contrast definition.

We can now use the information gained with the contour aggregation process to find other contour strings that are less well defined by contrast, (a typical example is shown in Fig. 3). In other words, already established structures are used to constrain possible extensions across areas of low or vanishing contrast. We have tacitly assumed that all important image structures are in some sense connected to key-points, thus possible contrast defined connections are only allowed between pairs of key-points. However, bridging contrast defined gaps between pairs of key-points extends also to connections that do not comply with co-curvilinearity constraints. Our approach to close contrast defined gaps consists of four stages.

Pre-selection: all connections are checked and those which (1) exceed a predefined distance, (2) are already established or (3) intersect with already established contour string, are eliminated. The pre-selection stage is fast and greatly reduces the number of connections that are further analyzed. This stage is done without analyzing the connection for contrast evidences.

Classification: each remaining connection is classified as either two-, one-, none-sided, depending on distance and angular criteria (see below).

Evaluation: collecting contrast evidences along the connection line. Connections not passing predefined criteria are eliminated.

Selection: the remaining connections compete in a local winners-take-all procedure, leaving only the most significant connections.

4.1 Classification According to Geometrical Criteria

The connections surviving the pre-selection stage are classified as either two-, one- or none-sided depending on angular criteria, as shown in Fig. 4. The classification is done by analyzing the termination directions of the contour strings in relation to a given connection. Analyzing key-point A we have n attached con-

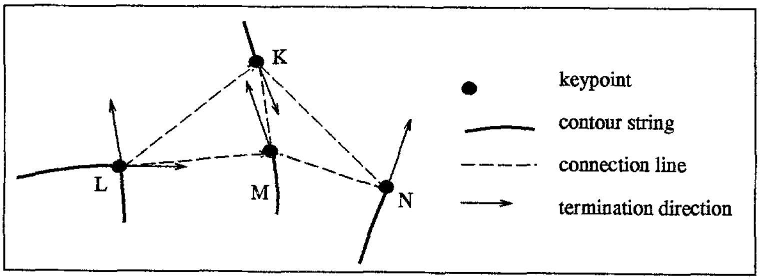

Fig. 4. Examples of connection classification (with parameters below), two-sided KM (or MK), one-sided LM, and none-sided ML, LK, KL, KN, NK, NM, MN.

tour strings $a_1, \ldots, a_n$ with their respective termination angles $\alpha_1, \ldots, \alpha_n$. Each connection from key-point A to B defines a direction, $\beta_{A \to B}$. The connection AB $(A \to B)$ is one-sided if,

$$\min_i \left(|\beta_{A \to B} - \alpha_i|\right) \leq \left(\frac{\pi}{2} - c\right) \cdot e^{-\frac{d^2}{2\sigma^2}} + c \tag{2}$$

where d is the Euklidian distance between A and B. The parameters σ and c control distance and angular criteria. (we used $\sigma = 5.0$ and $c = 15°$). If the connection BA $(B \to A)$ also satisfies (2) we have a two-sided connection. If neither

AB nor BA satisfy (2) the connection is classified as none-sided. This classification is thought to also reflect the level of uncertainty of a given connection, the most certain being the two-sided and the least certain the none-sided.

4.2 Evaluating the Connections for Contrast Evidences

The remaining connections are then tested for contrast evidences. We use two maps for this purpose; (1) the modulus channel that best matches the orientation connection between the key-points and (2) the local orientation map.

If there exists a smooth contrast defined structure between the key-points, we assume that (1) it is best defined by the modulus channel matching the connection orientation and (2) that the average deviation between the connection orientation and the local orientation along the connection line will be small.

Local maxima are searched for along scan-lines orthogonal to the orientation of the connection. As a potential structure between the key-points is expected to have low or vanishing contrast, we also expect the local maxima markings to have a high positional uncertainty (due to noise). Furthermore, the connection is constrained to go through the two key-points but must not necessarily be straight. This suggests the use of a lenticular-shaped region to search for local maxima as depicted in Fig. 5.

Local maxima found along each orthogonal scan line are marked as illustrated in Fig. 5. Not only the number of maxima markings, compared to the number of scan lines (ratio), is important but also their spatial distribution (scatter). Maxima markings are approximated with a second degree polynomial, constraining the fit-curve to go through the key-points. The mean squared error between the maxima markings and the fitted polynomial serves as a measure for scatter. Note however, that this measure does not discriminate between scatter due to noise and interferences stemming from neighboring structures. A connection is only accepted when both, the local maxima analysis (ratio, scatter) and the orientation analysis (average deviation of local orientation) individually pass given thresholds.

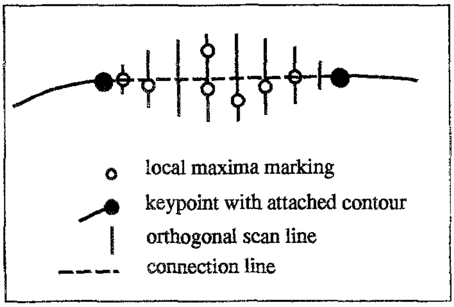

Fig. 5.: Evaluation of a connection.

4.3 Selection Through Competition

The connections remaining at this stage are few and must all have a residual contrast definition. We further reduce the number of connections by local competion, allowing only one connection *from* a given key-point to another. Notice, however, that this winner-takes-all approach still allows a given key-point to *receive* connections from other key-points. We let the competition take place only within the classes two-, one- and none-sided and the selection among the them is strictly hierachical, with the two-sided connection having the highest priority. The competition is based upon geometrical criteria and the evidence for residual contrast definition. A measure reflecting contrast definition is calculated by additively combining the ratio of maxima markings, the scatter and the average deviation of local orientation (see above). The geometrical criterion penalizes deviations from collinearity and applies only to one- and two-sided connections. As a measure we use the cosine of the enclosed angle(s) between the connection line and the termination direction(s) of the contour strings.

5 Results

In this section we show the results that can be obtained with our approach, using two rather complex images, an aerial and a telephone image. In addition we also show feature maps such as the 2nd derivative of response modulus taken in the dominant orientation and key-point map. (Fig. 6, 7, and 8). Image dimensions are 256×256 with 8-bit grey-level resolution. Tests were carried out on SUN Sparc 2 and 10 stations using ANSI-C programming language.

The gap closing algorithm successively reduces the number of connections. We have confirmed this by counting the number of connections in each stage. For the aerial image there were initially 223729 (473^2) connections and remaining after pre-selection 3450, evaluation 995, and selection 89.

6 Conclusions

We have presented a contour aggregation scheme on three distinct levels. The first level is concerned with linking local edge evidences into coherent contour-strings. An independent representation of key-points is used to define appropriate stop-conditions for the linking process. Knowing the location of corners and junctions also alleviates the problem of reconstructing them from edge map evidences, although the latter approach has proven quite successful [12].

The second level is a pruning stage, intended to eliminate spurious contours attached to corners and junctions as well as isolated contour fragments of low contrast. The pruning is important for (1) obtaining more stable classifications of corners junctions etc. and for (2) eliminating spurious contours that may block gap closing in the successive stage.

The third level deals with bridging gaps that are caused by poor contrast definition. The suggested algorithm not only incorporates geometrical information of already established structures, but also residual low-level contrast evidences for making a final decision. We have shown that this strategy effectively selects completions between pairs of key-points that are weakly defined by contrast. Currently we can only deal with fairly straight completions, but we intend to expand the scheme also for curved segments. A distinctive feature of the present approach is that a given completion must not necessarily comply with geometric (e.g. collinearity) constraints, as long as there is sufficient contrast definition. Some examples for this have been shown in Fig. 7.

In general, we believe that before invoking any type of perceptual grouping it is necessary to first find stable representations of the contrast defined features and their connectivity. Having this basis, it seems much easier to infer structures that are not defined by contrast and to discriminate between different completion types (e.g. foreground or background structures).

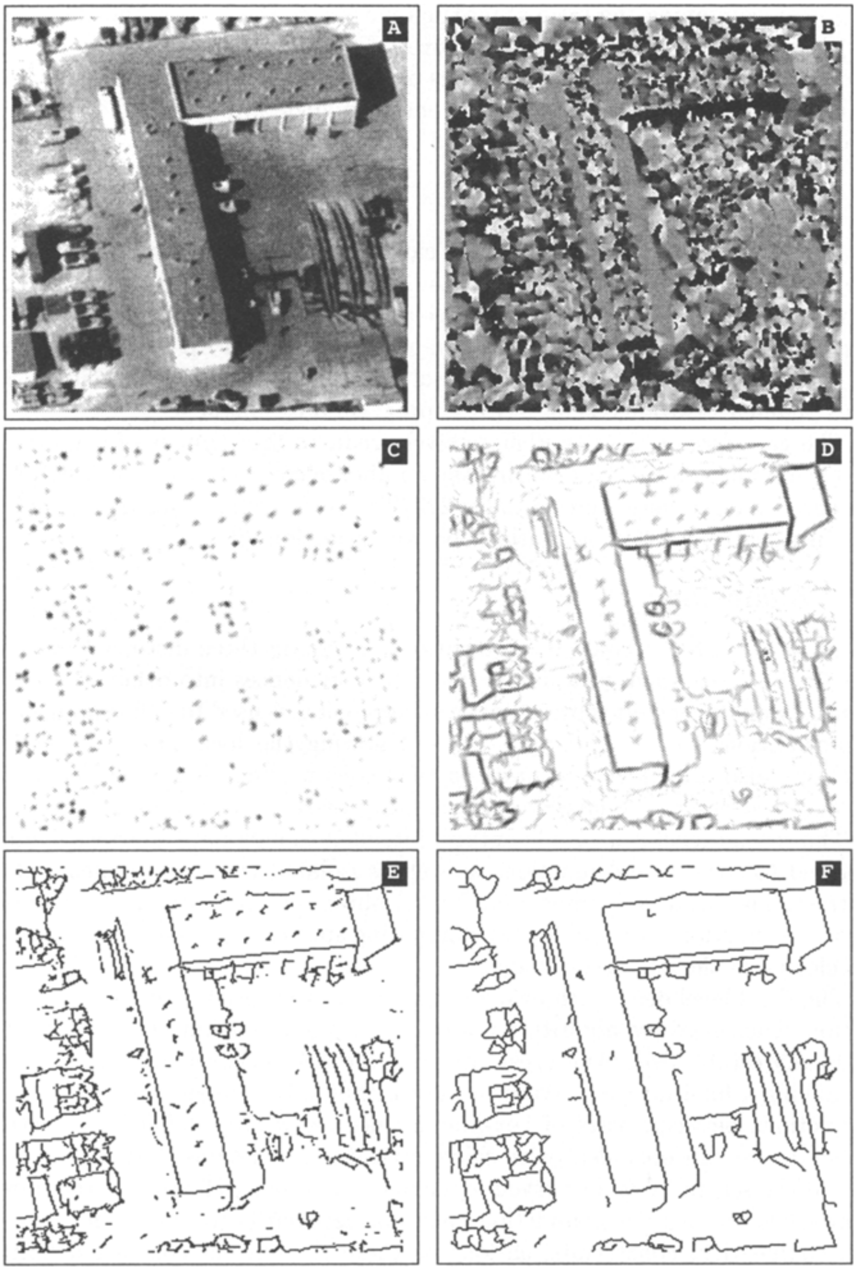

Fig. 6. (A) original aerial image, (B) local orientation coded with grey-values ranging from black 0° via grey 90° to white 180°, (C) key-point map, (D) clipped negative second derivative of modulus, (E) initial edge-map, and (F) the resulting contour representation after the gap closing procedures.

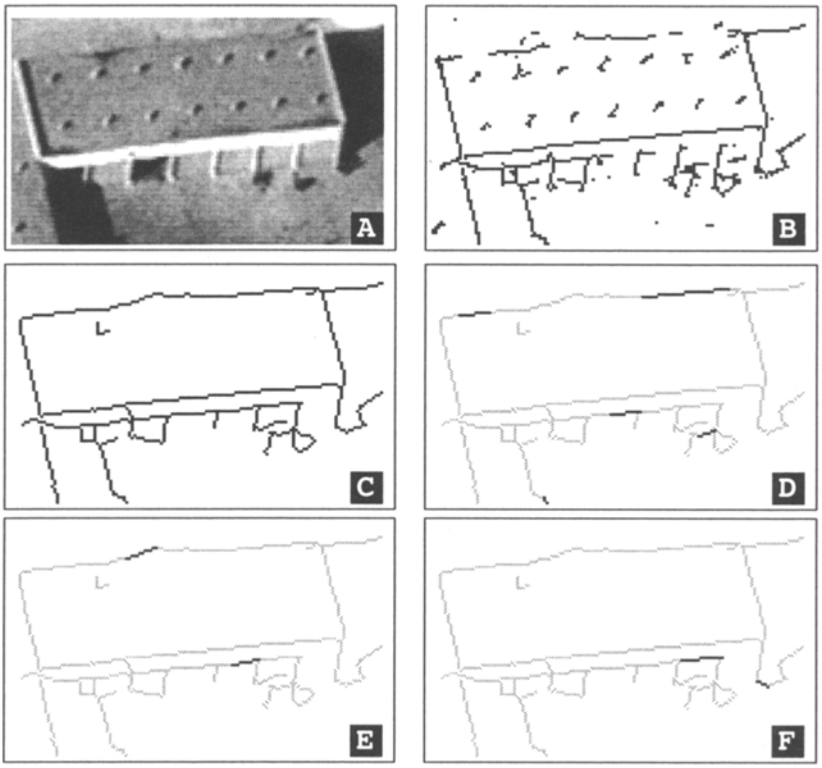

Fig. 7. (A) cut-out from Fig. 6, (B) initial edge-map, (C) the resulting contour representation after the gap closing procedures, (D) two-sided connections bridged by the gap closing algorithm, (E) one-sided connections, and (F) none-sided connections.

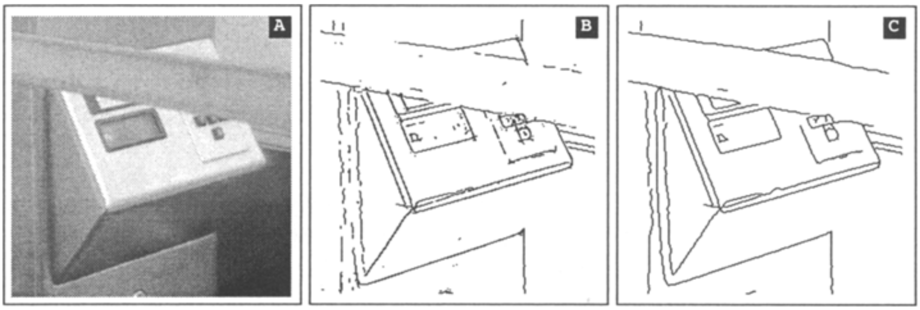

Fig. 8. (A) original telephone image, (B) initial edge-map, and (C) resulting contour representation after the gap-closing procedures

References

1. E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2):pp. 284–299, 1985.
2. H. Asada and M. Brady. The Curvature Primal Sketch. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI–8(1):pp. 2–14, 1986.
3. J. F. Canny. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI–8(6):pp. 679–698, 1986.
4. W. T. Freeman. *Steerable filters and local analysis of image structure.* PhD thesis, MIT, Media Laboratory, Cambridge MA, 1992.
5. G. H. Granlund. In Search of a General Picture Processing Operator. *Computer Graphics and Image Processing*, 8:pp. 155–173, 1978.
6. G. Guy and G. Medioni. Perceptual Grouping Using Global Saliency–Enhancing Operators. In *ICPR'92*, pages 99–103, 1992.
7. F. Heitger, G. Gerig, L. Rosenthaler, and O. Kübler. Extraction of boundary key-points and completion of simple figures. In *SCIA'89*, pages 1090–1097, 1989.
8. F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kübler. Simulation of neural contour mechanisms: From simple to end–stopped cells. *Vision Research*, 32:pp. 963–981, 1992.
9. F. Heitger and R. von der Heydt. A Computational Model of Neural Contour Processing: Figure–Ground Segregation and Illusory Contours. In *ICCV'93*, pages 32–40, Berlin, Germany, 1993.
10. H. Knutsson and G. Granlund. Texture analysis using two-dimensional quadrature filters. In *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, pages 206–213, 1983.
11. D. G. Lowe. *Perceptual organization and visual recognition.* Kluwer Academic Publishers, Boston, MA, 1985.
12. J. Matas and J. Kittler. Junction detection using probabilistic relaxation. *Image and Vision Computing*, 11(4):pp. 197–202, 1993.
13. G. Medioni and Y. Yasumoto. Corner Detection and Curve Representation Using Cubic B-Splines. *Comput. Vision, Graphics and Image Process.*, 39:pp. 267–278, 1987.
14. R. Mohan and R. Nevatia. Perceptual Organization for Scene Segmentation and Dsecription. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI–14(6):pp. 616–635, 1992.
15. M. C. Morrone and D. C. Burr. Feature detection in human vision: a phase–dependent energy model. *Proc. R. Soc. Lond.*, B 235:pp. 221–245, 1988.
16. M. C. Morrone and R. A. Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6:pp. 303–313, 1987.
17. P. Perona and J. Malik. Detecting and localizing edges composed of steps, peaks and roofs. Technical Report UCB/CSD 90/590, Computer Science Division, University of California at Berkley, 1990.
18. L. Rosenthaler, F. Heitger, O. Kübler, and R. von der Heydt. Detection of General Edges and Keypoints. In G. Sandini, editor, *Lecture Notes in Computer Science*, pages 78–86. Springer Verlag, Berlin, 1992.
19. A. Sha'ashua and S. Ullman. Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network. In *ICCV'88*, pages 321–327, 1988.