USING PERCEPTUAL GROUPING FOR ROAD RECOGNITION

G. Forlani*, E. Malinverni*, C. Nardinocchi**
*DIIAR - Politecnico di Milano , Piazza Leonardo da Vinci, 32- 20133 Milano, Italy.
**DScMtTe - Università di Ancona, Brecce Bianche, 60100 Ancona, Italy.
gianfra@ipmtf1.topo.polimi.it
eva @ipmtf1.topo.polimi.it
carla@anvax1.unian.it

Commission III, Working Group 3

KEY WORDS: Vision Sciences, Classification, Recognition, Feature Extraction.

ABSTRACT:

Automatic localization and identification of cartographic object from aerial and satellite images has gained an increasing attention in photogrammetry. The approaches for automatic extraction of man made objects may be grouped into two broad categories: semi-automated methods and fully automatic systems. Here an automatic system oriented to road recognition is presented.The system is based on a three stage procedure: image segmentation by feature extraction, perceptual organization of the geometric attributes of the features and object recognition based on an implicit knowledge base representation.

1. INTRODUCTION

On the way towards automatic mapping and GIS data acquisition and update, automatic localization and identification of cartographic object from aerial and satellite images has gained an increasing attention in photogrammetry; while DTM generation and thematic classification have already reached a high degree of reliability, the automatic extraction of man made objects, which are of major interest in applications, is still an unresolved issue. The approaches which are currently pursued may be grouped into two broad categories: semi-automated methods rely on the intervention of human operators to provide either the initial input to the system (e.g. seed points) or to help the system to bridge loopholes or ambiguities (Vosselman & de Knecht 1995; Gruen et al., 1994); fully automatic systems on the contrary should be able to address the whole complexity of the task: therefore they need to implement a more refined strategy based on a set of rules or assumptions which constitute the knowledge base of the system (Barzhoar & Cooper, 1995; Steger et al., 1995). Apart from the definition of a convenient and effective strategy, there is still a lot to improve in the fundamental stage of feature extraction, since many of the algorithmic problems arise from the poor quality of the extracted edges. At present, only semi-automated systems represent a good compromise between the speed of the procedure and the time and the commitment required to the operator. Still, research should aim towards increasing automation, since, apart from cost reasons, most of the interaction required would be anyway too repetitive and may prove, if the process stops too often, less appealing than manual plotting by the operator.

We are working on the development of an automatic image analysis system oriented to recognition of cartographic objects.The system is based on a three stage procedure:

- image segmentation by feature extraction
- perceptual organization of the geometric attributes of the features
- object recognition based on an implicit knowledge base representation.

In its current stage of development, only road recognition is available and therefore that's what we are going to talk about in the following.

2. IMAGE PROCESSING

2.1 Noise reduction

Before any image segmentation is performed, it is necessary to reduce the amount of image noise. There are two main types of noise in images: impulse noise and distributed noise. The former affects the gray value only in some pixel in the image, but to a large extent: it may be termed as a gross error. As such, its effect on the edges is only local and may be neglected. The latter affects all pixels and may be assumed to be randomly distribuited, therefore appropriate filtering is required. A trade off is to be found between the edge smoothing implied by all low pass filters and the noise reduction. Linear filters give a marked smoothing, so non-linear filter are preferred in edge detection. The most effective, but for the median filter, ideal for treating impulse noise, are the Edge Preserving Smoothing (EPS) and the Conditional Averaging Filter (CAF). In images with a small noise content CAF performs better than EPS, since it maintain more details and shows a more accurate edge localization, while both are equivalent otherwise. We used therefore CAF in our preprocessing stage, setting its threshold by visual inspection.

2.2 Image segmentation

Segmentation groups the image pixels in regions satisfying a given criterium; it may be based on texture or edge properties. Here the latter approach is used, based on two gradient characterists: magnitude and direction. In order to detect linear features in the image and to ease the road recognition stage, continuous lines will be approximated by a sequence of line segments.

To select edge pixels a threshold must be introduced on gradient magnitudes. Moreover, as long as the gradient orientation remains pretty much constant along contiguous pixels, they belong to an edge which is a line segment. There are many alternatives in the way the gradient vector may be computed and different threshold may be fixed for its magnitude. The segmentation output therefore will be dramatically affected by this choices, either making life easy for the algorithms or preventing them from getting any acceptable outcome. We based our image preprocessing on the idea of carrying all information until we are in the

condition to discard what becomes clearly useless.
In the gradient computation large masks tend to increase smoothing, loosing details; small masks instead preserve fine detail, but are very sensitive to noise. We used a small 2x2 mask (see Figure 1) which also gives an invariant response with respect to line rotations (Burns et al., 1986).

-1	-1
1	1

-1	1
-1	1

Figure 1. The mask used for computing the gradient

For each pixel in the image, the gradient magnitude is computed and, if its value is larger than the chosen threshold, the orientation is computed as well. In the following, when speaking of image or image orientation we will always refer to this part of the original image. To proceed with the feature extraction, all contiguous pixels enjoying similar gradient orientation are grouped in regions, because they are likely to belong to the same edge. The space of the orientations [0-2π] is divided into suitable equally spaced intervals, the so-called partitions (see Figure 2).

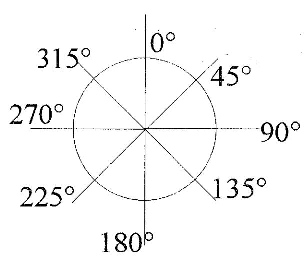


Figure 2. Gradient space partitioning in 8 intervals

2.3 Feature extraction

We look for a description of the image content based on lines. This may be achieved in many ways, e.g. by line following, relaxation, Hough transform etc. (Ballard & Brown, 1982); we opted for an alternative suggested by (Burns et al., 1986), with some minor changes. The concept is the following: we get a line segment from each region where the gradient orientation is in a certain range. The straight line to which the segment belongs is defined by the gravity centre of the area and by the direction perpendicular to gradient direction. The end points of the segment are determined by projecting the points of the area over the straight line.
The segment orientation is computed by a robust method (either Hampel, Huber or the L-1 norm may be selected), which some experiment proved to be better than a total least squares approach, particularly in small regions and with a small number of partitions. The gravity centre is computed as a weighted mean, using the gradient magnitude.
The choice of the number of partitions is critical: if there are too many we get a very fragmented image; on the contrary, large partitions result in a rough approximation of the edge. We found that using either 12 or 24 partitions, depending on the actual image, was appropriate in all cases we processed.
At the end of this stage, we have now a vector representation of the edges, where each line segment is defined by its orientation, its gravity centre and its end points.

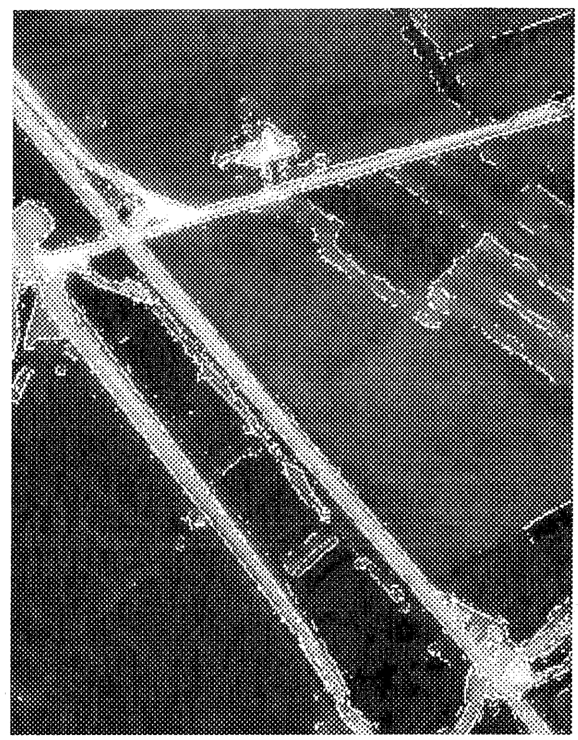


Figure 3. Feature extraction output

Additional information on the goodness of the fit for orientation and location is recorded; moreover, it is always possible to go back to the original image region. Figure 3 shows the feature extraction output superimposed to the original image.

Figure 4. The remaining features after data reduction

2.4 Data reduction

Before moving to a higher level of processing, we try to get segments likely to be grouped and linked in significant structures. This means to connect strictly collinear very close segments in one new longer segment: in other words we simply try to make up for narrow gaps arisen along an edge. In all cases we processed, this amounted to a 15% reduction of the number of lines. After and only after this stage, we remove all short segments(less than 2 pixel long) since they won't play a role in the subsequent stages. Out of the original number of segments, 80% are discarded at the end of this stage (see Figure 4).

3. PERCEPTUAL ORGANIZATION

On the way to image interpretation, we need to move from a description based on gray values to a more abstract level, to identify structures. These may be termed as collections of elements (lines or regions) which the visual human system perceives as connected or interrelated, even without any a priori knowledge of their contents: this process is called perceptual grouping. We look for relations which should be least sensitive to changes to the standpoint and with small probability to arise in the image by chance (i.e. a radiometric edge will truly represent a physical edge), for instance collinearity, proximity, closure, parallelism... (Lowe,1985; Sarkar & Bayer, 1993).

We used proximity to reduce the search space, since features which are far apart in the image are not likely to share significant connections. A search window (see Figure 5) is build around the gravity center of each segment using the value Dmax of the maximum width of the road classes for that image scale.

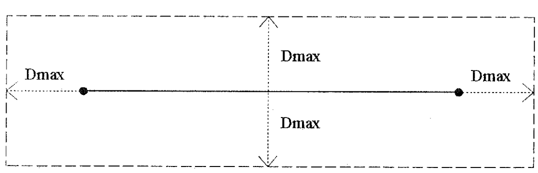


Figure 5. The search window

Within the search window, we look for collinearity and anticollinearity (that is for segments sharing the same direction, but with opposite gradient orientation), cocurvilinearity (and anti-cocurvilinearity), parallelism (and anti-parallelism), junctions (see Figure 6). To correctly label the relations, we must decide to what extent the actual features match an ideal model (e.g. two segments in reality will never be parallel in a strict mathematical sense). This is simply achieved by setting up some threshold values for distances Δr and Δv between the end points and differences in direction Δt for each pair of segments (see Figure 7).

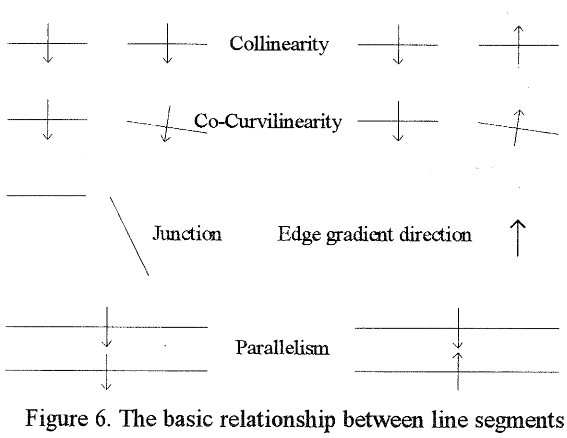


Figure 6. The basic relationship between line segments

Parallel to this process, we classify the attributes of features and relations. Figure 8 shows the attributes recorded for each relation.

We have now completed a relational description of our primitives, that is, of the line segments.

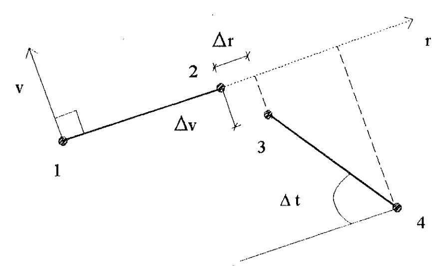


Figure 7. The elements used to define the relation between a pair of segments

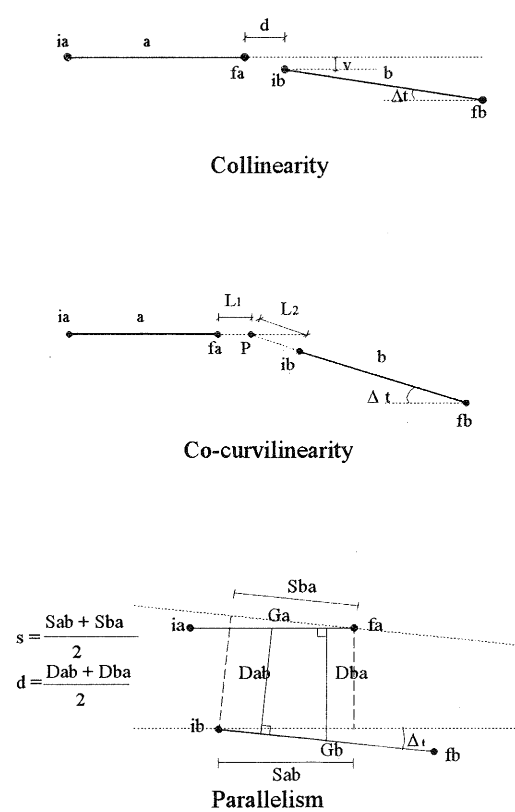


Figure 8. Attributes of relations

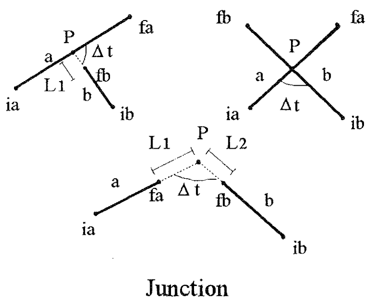


Figure 8. Attributes of relations - continued

4. ROAD RECOGNITION

Our recognition strategy is simply based on three assumptions:

- roads have in most cases antiparallel edges (Nevatia & Babu, 1980);
- roads are elongated objects;
- roads are connected in a network.

In our experience other criteria, such as looking for a homogeneous texture between the road edges, failed rather often.

We proceed now with the analysis of the relations of antiparallism. Perceptual grouping may be applied recursively, generating different levels of virtual features, that is, arrangements of the basic primitives. Here the primitives are pairs of antiparallel segments, grouped in higher level primitives (would-be road legs) based on a continuity criterium, namely collinearity and co-curvilinearity (see Figure 9).

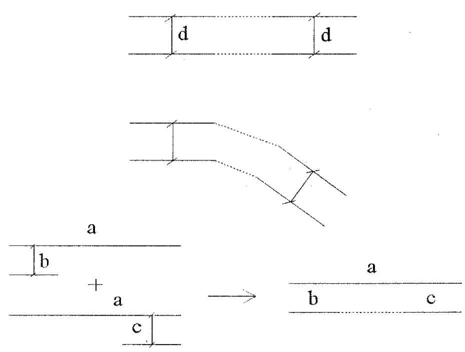
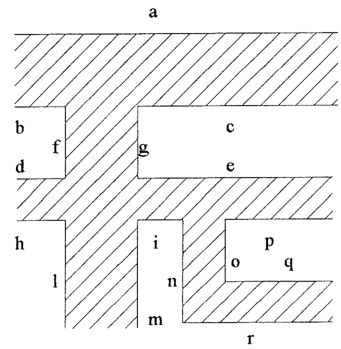


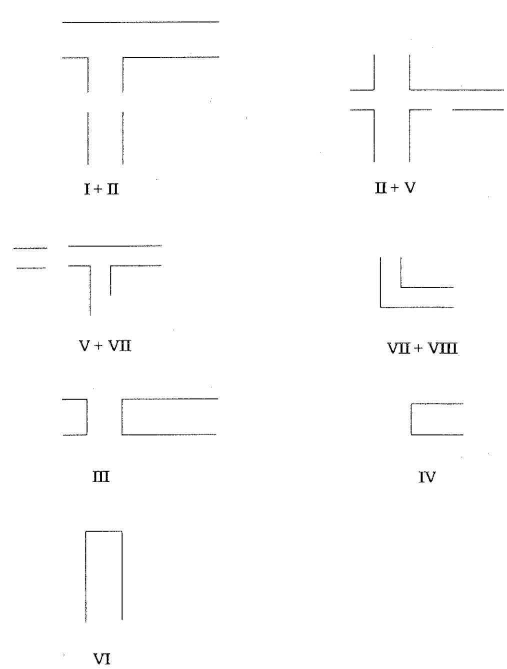
Figure 9. Grouping antiparallel pairs of segments

At the end of this stage we have groups of pairs of antiparallel lines; the same line may nevertheless belong to more than one group: this originates a competition among the groups, which will be solved later on. Looking for a network structure, a last recursive grouping is performed between would-be road legs connected by junctions. Ideally, if there would not be competitions between groups, this process would end up with all (would-be) disjoint road networks in the image. Figure 10 depicts the behaviour the algorithms on an artificial example.



Would-be road-legs	Groups of AP pairs	Ambiguities
I	a - b a - c	2
II	f - g l - m	1
III	b - d c - e	4
IV	p - q	2
V	d - h e - i e - p	3
VI	m - n	2
VII	n - o	1
VIII	q - r	1

Junction relations found between would-be legs.



Would-be nets	Groups of would-be legs
1	I + II + V+VII + VIII
2	III
3	IV
4	VI

Figure 10. Network structure

In order to solve for the competitions (ambiguities), we use the information at hand:

- the total lenght of each (would-be) net;
- the number of groups involved in each net;
- the number of competitions in each net.

At the current stage of our work, for each line with an ambiguity, we evaluate the above informations for the competing nets; we take only the net with highest score in all three items, discarding the other candidates. Figure 11 shows the extracted road network for the original image.

Figure 11. The extracted road network

5. PERSPECTIVES

We are just at the beginning of our experience in road extraction, so be many aspect of our strategy will need a revision or to be studied in more depth.

As far as the choice of the partition is concerned, a possible improvement may be to use a small number of partitions, but to set a procedure capable to further segment the region within, recognizing e.g. a long smooth edge which would be poorly approximated by a line segment only. This seems to be better than increasing the number of partitions, when the outcome may be sensitive to local disturbances of the gradient. In the same line of thought, we plan also to use polynomials or splines, rather than line segments only.

The solution of the ambiguities need to be refined, giving different weights to each information. More important we will have to develop additional likelihood measures to complement the current road extraction strategy in its last stage.

6. REFERENCES

Barzhoar M., D.B. Cooper, 1995.*New geometric stochastic technology for finding and recognizing roads and their features in aerial images*. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images. Edited by A. Gruen, O. Kuebler, P. Agouris. Birkhaeuser Verlag, Basel.

Gruen A., H. Li, 1994.*Semi-automatic road extraction by dynamic programming*. In: International Archives of Photogrammetry and Remote Sensing (IAPRS). Vol. 30 Part 3/1, pp. 324-332.

Ballard D. H., C. M. Brown, 1982.*Computer Vision*. Prentice Hall, Englewood Cliffs NJ.

Burns J.B., A.R. Hanson, E.M. Riseman, 1986. *Extracting straight lines*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8 N.4, pp. 425-455.

Lowe O.G., 1985. *Perceptual Organization and Visual Recognition*. Kluwer, Boston MA.

Nevatia R., K. R. Babu, 1980. *Linear feature extraction and description*. In: Computer Graphics and Image Processing. Vol 13, pp. 257-269.

Sarkar S., Bayer Kim L., 1993. *Perceptual Organization in Computer Vision: A Review and a Proposal for a Classificatory Structure*. In: IEEE Transaction on System, Man and Cybernetics, Vol.23,N.2, pp.382-399.

Steger C., C. Glock, W. Eckstein, H. Mayer, B. Radig, 1995. *Model-based road extraction from images*. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images. Edited by A. Gruen, O. Kuebler, P. Agouris. Birkhaeuser Verlag, Basel.

Vosselman G., J. de Knecht, 1995. *Road tracing by profile matching and Kalman filtering*. In Automatic Extraction of Man-Made Objects from Aerial and Space Images. Edited by A. Gruen, O. Kuebler, P. Agouris. Birkhaeuser Verlag, Basel.