'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

Toward an automatic extraction of the road network by local interpretation of the scene

RENAUD RUSKONÉ, London and SYLVAIN AIRAULT, Saint-Mandé-Cedex

ABSTRACT

This paper deals with the automatic detection of road network from aerial images. Our approach can be characterised by a two-level processing: firstly, a low-level knowledge is used to extract most of the roads, then semantic information is applied to solve interpretation problems. As we aim to implement this system in a production framework, we have been very careful about the reliability of the result. According to this objective, an incomplete but very reliable network is better than a complete one but that contains also many misdetections.
The results are rather fair: we can assess that more than 60% of the road network could be automatically extracted. Nevertheless, some cases will still require a human interpretation or the help of complementary data because of the complexity of some problems.

1. INTRODUCTION

IGN has undertaken for a few years the constituting of a topographic database, the BDTopo®. This database describes the semantic content of the traditional 1:25000 maps with objects localised with a metric accuracy and a three-dimensional description. Its capture is done from panchromatic aerial images which resolution is about 50 centimetres.
Studies have been undertaken at IGN in the field of automatic recognition of buildings, rivers and roads, and relief restitution to speed up the first capture of the BDTopo®. In these domains, the research faces the complexity of natural scene interpretation. Even though the completeness of an automatic aerial image interpretation is still illusory, one can aim to a partial but reliable interpretation of some objects of the scene. Indeed, the use of the extracted network in a production context implies that its reliability would be sufficient not to require controls and manual corrections that would be as costly as a complete manual capture.
The system that is described in this paper is applied to the automatic road network extraction. One of its main characteristics is the auto-evaluation of its outputs to keep the most reliable and eliminate the doubtful ones.

2. AN ARCHITECTURE SOLUTION

Our system is grounded on a loop involving a hypothesis generation phase and the validation of these hypotheses. The first phase uses low level knowledge but the validation requires more complex knowledge. Indeed, even though the "salient" roads can be identified according to rather low level criteria (radiometry, texture, shape), many borderline cases limit the road network full recognition. Their processing can not be solved without using high level knowledge (especially dealing with context Baumgartner, Eckstein, Mayer, Heipke and Ebner, 1997) or function (Garnesson, Giraudon and Montesinos, 1989) of the objects to be identified).
However, we want to restrict the use of interpretative processes. Indeed, given the complexity of the usually processed scenes, it may be more judicious to apply high level knowledge just to solve very localised problems using focusing mechanisms.

Thus, this "two semantic level" extraction may be called a mixed approach:

- bottom-up (data driven) for the low level extraction step that will generate road hypotheses,
- top-down (goal driven) for a checking of the extracted network on higher level criteria. This validation phase is thus focused on the immediate neighbourhood of the road.

3. OUTLINE OF THE DIFFERENT STEPS

Our approach may be summarised by the following figure (see figure 1):

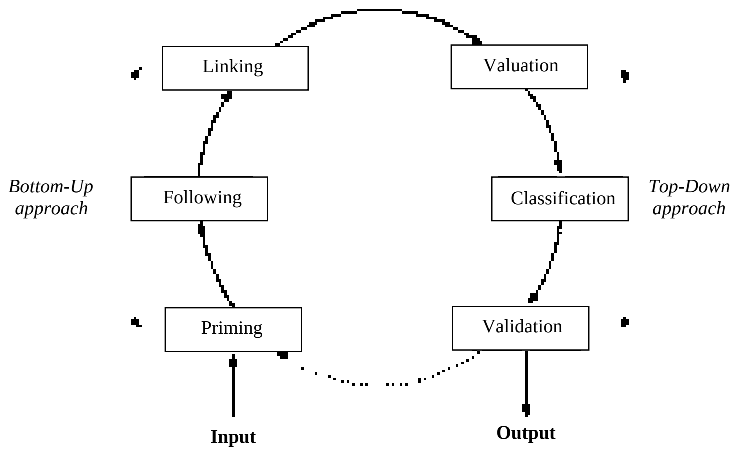

Figure 1: The different steps of the processing.

3.1 Seed point detection

The first step (*Priming* in the figure 1) aims at generating seeds that are the starting points the road following phase will lean on to pursue the detection according to their characteristics (position, orientation and width). This "priming/following sequence" appears to be the method giving the best results (and the most quoted in the bibliography); it is indeed a low level method allowing to use a rather complete road model without increasing the complexity because of the initialisation done during the priming phase.
Firstly, the image is segmented by delimitation of catchment basins in a gradient image ("Watershed" algorithm), the gradient being computed by Canny-Deriche filtering. In a nutshell, this segmentation method considers the image as a topographic relief where the grey levels are equated with the altitudes. The watersheds represent the frontier between the different regions.
Then, the edges of the regions (coming out from the segmentation) are chained and vectorized by polygonal approximation. The parallel edge matching is done one segment after the other, independently of the regions.
Boxes traducing the overlapping of homologous segments are built before to be filtered by an elongation criterion. The seeds on which the road following algorithm will lean on are the box axis extremities.

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

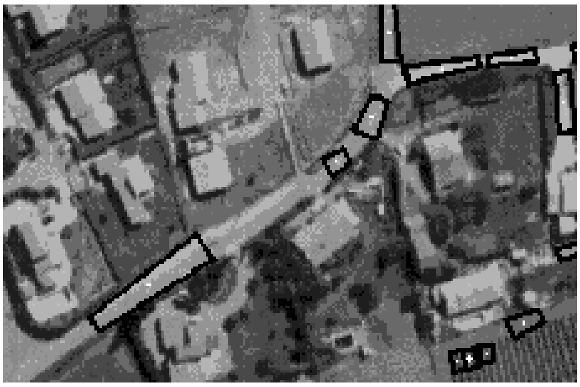

Figure 2: An example of the seed extraction.

3.2 Road following

This phase constitutes the main low level extraction task. It uses the road seeds generated by the former step. It goal is to complete the network by following the road hypotheses (i.e. the seeds) done during the previous step.
At this level, our work is broadly grounded on the results supplied by a semi-automatic approach for the road extraction and especially on the road following algorithm implemented by Sylvain Airault (Airault and Jamet, 1994).
This algorithm is mainly based on the knowledge according to which the road surface has a homogeneous texture bounded by two edges that, more often than not, correspond to radiometric discontinuities. If variances are computed on elongated neighbourhoods, thus the variance in the direction of the road orientation is minimal.
This criterion is implemented through the computation of a tree consisting of all the possible paths long enough to be significant at the scale of a road (50-150 meters). Then, these paths are evaluated according to a cost function measuring the texture homogeneity. Each path is composed of line segments locally chosen given the homogeneity criterion. Optimising this criterion on a set of segments allows us to fit the road shape and to jump over small obstacles.

3.3 Network reconstruction

The network coming from the previous step shows some gaps. They may be due to the encounter with an obstacle that has not been passed round:

- road radiometry variation (horizontal marks, surface nature change, shadows,...)
- screening of the road (vehicles, trees, bridges,...),
- ambiguity about the direction to be followed (crossroads, lack of contrast,...).

This step consists in emitting hypotheses for the connection and for the extension of formerly extracted road portions: the used criteria rely only on the topology and on the geometry and none low level backtrack on the image is done.
The network coming from the previous step shows some gaps. They may be due to the encounter with an obstacle that has not been passed round:

- road radiometry variation (horizontal marks, surface nature change, shadows,...)
- screening of the road (vehicles, trees, bridges,...),
- ambiguity about the direction to be followed (crossroads, lack of contrast,...).

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

This step consists in emitting hypotheses for the connection and for the extension of formerly extracted road portions: the used criteria rely only on the topology and on the geometry and none low level backtrack on the image is done.

Connection hypotheses

Firstly, it is necessary to reconstruct an as complete as possible network by generating connection hypotheses between the already detected arcs. This step is only grounded on a few general pieces of knowledge about the road network topology and geometry:

- the free ends or the isolated segments are rare,
- the very curved segments are a minority,
- two optimal path linking two close points with a similar length are not frequent.

From the two first rules, connection hypotheses are thus done by a very local analysis of the neighbourhood of each potential connection on proximity and rectilinearity criteria. For that, we use the fundamental principles of the perceptual grouping that notices the importance of the organisation in interpretation (Rock, Palmer, 1990): we have supposed that a partially interrupted segment can be completed either by a simple prolongation or by matching with the nearest segment.
Three kinds of connection relations may exist either (see figure 3):

- between free ends or,
- between a free end and the intersection formed with another arc (mixed connection) or,
- between segments continuing each other.

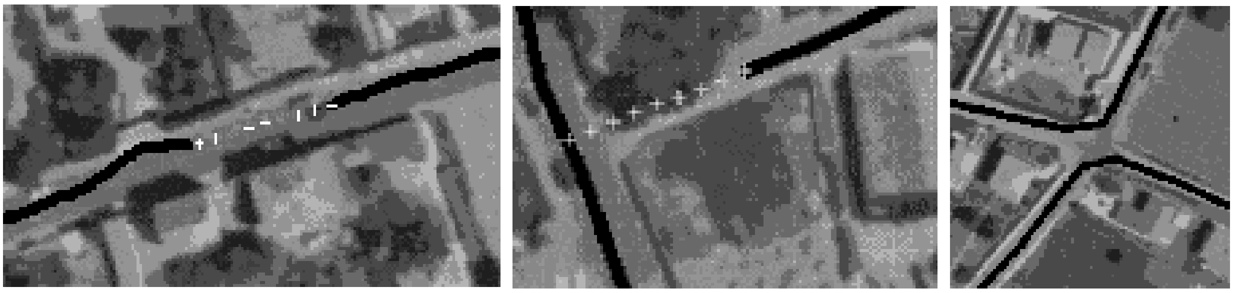

Figure 3: Possible connection types.

As one can notice it from the figure (see figure 3), the points that can be connected to each other can be either free ends or high curvature areas. For what concerns the junction modes, they can be done according to the direction of the last segment or according to the “nearest neighbour”.

Topology validation

The connection hypothesis coherence is then checked by applying the rule that specifies that two paths of similar length seldom link two neighbouring points (Deseilligny, Le Men and Stamon, 1993). The minimal cost path that links two extremities of a connection is searched in the graph.
If a path with a cost very close to the one of the direct connection is found, then the connection is supposed to be wrong and the arcs that compose this connection are erased from the graph; if, on the contrary, the direct connection cost is clearly lower than the path cost, then the connection is supposed to be coherent.
The processing sequence of hypothesis generation/validation is iterated until the process stability (i.e. until no hypothesis can be generated).

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

Prolongation

This reconstruction phase does not only aim to connect the network portions that “match” but also to prolong free ends that have not been connected. Indeed, in this case, it is important to pass beyond the obstacle that has caused an arrest of the following step for the subsequent validation step to interpret the segment overlapping the object.
The free end prolongation is done by adding a constant length segment to any free ends not involved in a connection, in the propagation direction of the free ends.

3.4 Network "valuation"

This is the *descending* phase of our system during which each road hypothesis must be either confirmed or considered as doubtful. Therefore, we firstly label each portion of the network with a measurement vector that allows the transition between a numerical representation and a symbolic representation of the data by deducing from the measurement vector a hypothesis about a portion of the network.
Here, the graph structure that we are using to model the road hypotheses will be modified in two ways:

- splitting of the road arc portions in small constant length arcs,
- computation of a measurement vector attached to each arc,
- computation of a vector of hypotheses about the nature of each arc (actually, on the nature of the objects overlapped by the arcs).

Knowledge model

Firstly, we have made an inventory of the most common objects in the processed images (ground pixel of about 30 to 100 cm). Among these objects, we kept:

- those that may produce modifications of the road characteristics (radiometry near trees or shadows, curvature near crossroads,...),
- those which characteristics are similar to the road.

To identify the object types to be taken into account in the interpretation process, we made a statistic study of the causes of the road following algorithm stops. These failures are, more often than not, caused by crossroads or tree lines (see figure 4). One must however notice that the stop cause analysis is rather complex given that the origin of the error is not always the same obstacle as the one that provokes the arrest. For instance, if a shadow partially masks the road, the following algorithm will avoid this shadow and may lose its way in a neighbouring field.

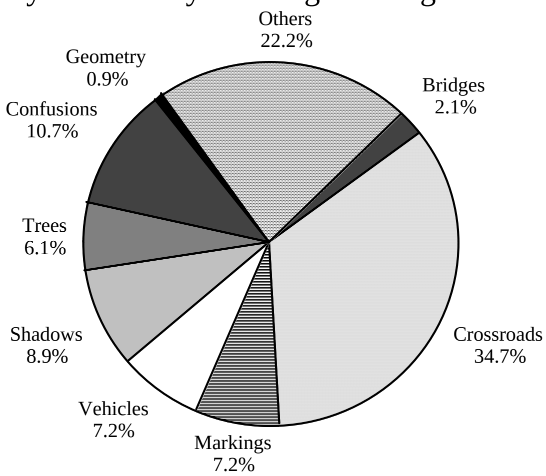

Figure 4: The causes of the road following stops.

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

In this figure, the class *Others* gathers some rare or multi-thematic stop causes (parking lots, dead ends,...). The *Confusion* class refers to places where the road radiometry becomes similar to its neighbourhood. This class appears to be the only one highly depending on the image type, unlike the others that are quite stable.
The analysis of physical stop causes allows us to determine five main themes: the roads, the crossroads, the shadows, the fields and the trees:

- the **roads** are elongated areas with a low curvature; their radiometry is rather low and homogeneous;
- the **crossroads** represent the intersection of more than two road portions; their branches form often wide angles; their characteristics are close to the road ones (even though more noisy);
- the **shadows** are elongated areas with a small curvature (we are only dealing with those that may interrupt the following of a road); their radiometry is low and homogeneous; these areas edge high objects (building, trees,...);
- the **fields** may be characterised by the variability of their radiometry, texture, size and polygonal shape;
- the **trees** have a low and heterogeneous radiometry; they usually cast a shadow; their shape may be either circular (when they are isolated) or elongated (when they form lines) or variable (forest areas); unlike the other objects, their elevation is always higher than the ground.

Measurement choice
We have also listed the characteristics supposed to be distinctive:

- geometric characteristics (surface, width, elongation, curvature,...)(Ruskoné, Airault and Jamet, 1995),
- radiometric characteristics (mean, variance,...)(Airault, Ruskoné and Jamet, 1994),
- altimetric characteristics,
- topological relations (trees may edge the roads but never appear on the middle,...).

Classification
After the valuation, the network is composed of segments labelled by an attribute list. The classification step consists in interpreting these measurements by assigning to each arc a probability for its belonging to one of the possible object classes. Practically, this phase leads thus to the filling of the Nature attribute of the arcs.
We chose to use a classification method derived from the k nearest neighbour classifications. The objective of all these methods is to cluster the points of a set having the most similar characteristics. In the nearest neighbour classifications, the nearest neighbours of a point (nearest in terms of Euclidean distance, in our implementation) define a probable class that is assigned to the considered point.
During a supervision phase, five sets of segments (corresponding to the five considered object classes) are iteratively classified. After each iteration, the belonging probability of each point is used to change its weight; if a point is perfectly identified, its weight is reinforced, If, on the contrary, the point is far from the class centre, it receives a lower weight (because it is not representative of the class).

Results
On the next confusion matrix (see table 1), one notices the rather good reliability of the hypotheses but also the requirement for their careful checking, especially in the class where the confusion is high. The confusion between the roads and the crossroads is strong: so, a road-labelled arc can truly be a road at 74%, but has also 26% chances to be a crossroads. When the arc is identified as a crossroads, the prognosis is much more ambiguous because such a segment has 47% chances to be actually a road. There is also a confusion between the *tree* and *shadow* classes.

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

Y-labelled *are actually: X*

	Road	**Crossroads**	**Shadow**	**Field**	**Tree**	**TOTAL**
Road	**73.7**	25.9	0.1	0.3	0.0	100
Crossroads	46.7	**53.3**	0.0	0.0	0.0	100
Shadow	0.0	1.6	**98.4**	0.0	0.0	100
Field	3.6	0.9	0.0	**84.4**	11.1	100
Tree	0.0	0.0	25.9	3.8	**70.3**	100
Ambiguous	21.4	9.5	7.3	18.6	41.2	100

Table 1: Result of a classification (percentage of Y-labelled arcs being actually X).

By summing the correctly classified segments (diagonal sum), one can estimate that 70.4% of these arcs are identified without errors (i.e. the error rate is 29.6%).

3.5 Validation

The validation step aims to check the hypotheses about the object nature issued from the preceding phase. Therefore, we use modules that are dedicated to the recognition of the considered objects. These modules must be run according to a triggering strategy and their answers must be exploited according to a given resolution strategy.
At this level, the handled objects have four attributes:

- a shape attribute (compact, elongated or extended),
- a measurement vector (issued from the valuation phase),
- a nature attribute (initialised during the valuation phase),
- a reliability (how trustworthy is the nature hypothesis?)

Different modules have already been implemented for the detection of crossroads (Ruskoné and Airault, 1996c), for the confirmation of road presence through the identification of vehicles (Ruskoné, Guigues, Airault and Jamet, 1996a), and for the recognition of shadows or trees. These modules are not dependent of each other and only rely on an internal model of the object they are designed to identify.

Triggering strategy
One can consider with different points of view the choice and the sequencing of the specialist module uses: several possibilities exist between the two extremes that would consist either in checking the most probable hypothesis (*determinist* strategy) or in checking all of them (*exhaustive* strategy). There is no a priori solution. As all hypotheses are more or less reliable, one can consider that, beyond a threshold, some of them may be neglected. Thus, we have chosen to trigger the validation of all the "likely" hypotheses. When the studied arc is labelled as *ambiguous*, all the modules will be triggered.

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

Resolution strategy

Then, once known the nature of any arc portions, it may be useful to make a study in a less local manner by taking into account compatibility relations between the objects. As we want to have a highly reliable result, it could be possible to keep after the *local* validation of the arcs only those identified as roads. However, many obstacles like isolated trees, for instance (even correctly identified) must not cast doubt on the possible presence of a road. The systematic elimination of all the *no-road* arcs after the final interpretation could generate a very sparse network.

Only the matching of the shape and nature attribute can lead to the invalidation (or the confirmation) of a road hypothesis. Thus, we distinguish three kinds of shape likely to have compatibility relations with each of the considered objects:

- a small sized and compact shape (like the one of an isolated tree),
- an elongated shape,
- an extended shape like the one of huge patch objects (like trees or fields).

Colinearity relations exist between elongated objects. Thus, one can estimate that these objects may reinforce a road hypothesis (as for the trees when they form lines or for the elongated shadows of buildings).

On the other hand, extended objects can under no circumstances mask a road (but maybe for the forests where anyhow, the road validation is quite risky). The identification of such obstacles can thus invalidate the presence of a road. One can consider that an arc crossing a forest or field area is very unlikely to be a road.

For what concerns the compact objects, their identification just enables not to invalidate a road hypothesis. Actually, one consider that their size is small enough for the disturbance they provoke to be local and for the road to be identified again in the area.

4. PERFORMANCE EVALUATION

We present here the result of a quality assessment of a road network extraction. This evaluation is not complete because it corresponds to the processing of a unique site. It is a rural area in Brittany (west of France, near a city named Erquy) where the roads are usually well contrasted and the obstacle rather small. Independently from the scene complexity, it is worth evaluating the evolution of the extracted network quality after the different steps of the processing. Actually, the study of each step influence on the final result is motivated by the fact that the best result is not compulsorily the combination of the best intermediate results. One can easily imagine that a very permissive seed extraction step may correspond, at the end of the processing chain, to a best result than the one after a priming that would have generated fewer seeds (but all reliable).

Thus, we have tried to evaluate the quality of any intermediate result, not only according to the reliability (by evaluation of the misdetection rate) but also according to the geometric accuracy. This evaluation is done by using as reference data a usual BDTopo® capture.

Two iterations of the processing chain (priming, following, reconstruction, valuation and validation) have been taken into account:

- a first one where prolongation hypotheses have been emitted then validated,
- a second one beginning by the following of these new seeds.

Evaluation of the processing chain reliability

One tries to measure the *omission*, the *overdetection* and the *misdetection* rate. The measurements are defined according to a matching between the result graph and the reference one as follows:

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

- the *overdetection* measurement qualifies the quantity of objects from the result graph that match the same object of the reference graph. It could be due to the identification of a road side that has been confused with the nearest road) or to the restitution of a road with several tracks (that, of course, have not been generalised to a unique axis);
- the *omission* rate is obviously the proportion of objects of the reference graph that do not have any homologue in the result graph; this indicator measures the completeness of the detection;
- the misdetection measurement is defined as the proportion of objects of the result graph that do not match the reference graph.

The "objects" used for the matching are road portions, i.e. the set of the segments that link two nodes of the graph (the nodes being crossroads or free ends).

	Over-detections	**Omissions**	**misdetections**
Priming	*24.8*	*54.6*	*66.1*
Following	*12.2*	*55.0*	*88.9*
Reconstruction	*17.8*	*39.4*	*49.7*
Validation (iteration 1)	*6.5*	*31.8*	*19.5*
Validation (iteration 2)	*6.5*	*29.0*	*19.8*

Table 2: Evolution of the reliability of the processing chain (in percentages of objects).

One can consider these results (see Table 2) either under point of view of the final result quality or according to the evolution of the results during the processing:

- the final results are rather satisfying as about 70% of the reference road network has been identified with an error rate of about 25%, However, let us remind that this result has been obtained on a rather "easy" image;
- the variation of the results after each phase is interesting because it shows the relevance of each step (noticeable decrease of the errors during the process progress).

About the error nature, one can note that some automatically detected arcs are actually roads but do not exist in the reference. One can also notice that both graphs (reference and result) are not always perfectly overlapping. About the omissions, they can have two different meanings: the simple omissions (due to a failure of any of the steps of the processing) and the hardly visible arcs (that the photointerpret has probably identified thanks to an other data source). This last situation shows on one hand the limit of the automatic interpretation and on the other one the need for a complementary data source.

Evaluation of the geometric quality of the extracted roads

The geometric quality of the extracted network is evaluated by splitting the result graph in similar length segments and by matching each cut point with the reference graph (Airault and Jamet, 1995). Once again, one notices (see Table 3), the very good accuracy of the final result with the same reserve about the nature of the test site. Indeed, the roads have very contrasted edges that make the geometric shifting of the road on its centre rather easy.

'Photogrammetric Week '97' D. Fritsch & D. Hobbie, Eds., Wichmann Verlag, Heidelberg, 1997.

	m_{xy}	emq_{xy}	RMS_{xy}
Priming	1.38	1.62	1.60
Following	0.87	1.13	1.07
Reconstruction	0.91	1.18	1.13
Validation (iteration 1)	0.81	1.05	0.98
Validation (iteration 2)	1.06	1.05	0.99

Table 3: Evaluation of the extracted network geometric accuracy (in meters).

5. CONCLUSION

Even though some improvements may be done to this system, an important part of the of the road network can be identified with a good reliability on rather simple scenes. Moreover, our evaluations confirm the scene model efficiency and the use of knowledge about other objects than the roads to successfully interpret the road scenes.
However, despite a good behaviour for the identification of an obstacle of a given type, this system turns out to be very limited on more complex scenes where an obstacle can be composed of several other objects (vehicles plus ground marks plus shadows plus ...). These problems would require to improve the models by taking into account not only the relations between the road and other objects but also the relations among these objects. Therefore, the autonomy of our specialised modules should be modified by allowing the communication between these different modules.
In a medium/long term perspective, the road network will probably be one of the few themes (among which the buildings, the rivers,...) that could be automatically identified. However, its interpretation will require a considerable amount of knowledge for building the exhaustive model essential to this task. We have shown that a very simple model is sufficient in most of the cases. Thus, the question is: "Must we really endeavour to build and exploit this model to solve only a few border line cases?". It is much more likely that, like the human photo-interpret, we would rather use complementary data sources (scanned maps or databases)(Bordes, Guérin, Giraudon and Maître, 1996).

6. REFERENCES

Airault, S., Jamet, O. (1994): Détection et restitution automatiques du réseau routier sur images aériennes, RFIA proceedings, Paris (France), vol. 1, pp. 519-531.

Airault, S., Ruskoné, R., Jamet, O. (1994): Road detection from aerial images: a cooperation between local and global methods, Image and Signal Processing for Remote Sensing, Satellite Remote Sensing I, SPIE, vol. 2315, pp. 508-518, Rome (Italy).

Airault, S., Jamet, O. (1995): Evaluation of the operationality of a semi-automatic road network capture process, Digital Photogrammetry and Remote Sensing '95, SPIE, vol. 2646, pp. 180-191, St-Petersburg, (Russia).

Baumgartner, A., Eckstein, W., Mayer, H., Heipke, C., Ebner H. (1997): Context-Supported Road Extraction, Automatic Extraction of Man-Made Objects from Aerial and Space Images, Ascona Workshop.

Bordes, G., Guérin, P., Giraudon, G., Maître, H. (1996): Contribution of external data to aerial image analysis, ISPRS, commission III symposium, Vienna (Austria).

Deseilligny, M., Le Men, H., Stamon, G. (1993): Map understanding for GIS data capture: algorithms for road network reconstruction, ICDAR proceedings, Tsukuba (Japan).

Garnesson, P., Giraudon, G., Montesinos, P. (1989): Messie: un système multi-spécialiste en vision. Application en imagerie aérienne, INRIA research report n° 1012.

Rock, I., Palmer, S. (1990): The legacy of Gestalt psychology, Scientific American.

Ruskoné, R., Airault, S., Jamet, O. (1994): Road network interpretation: a topological hypothesis driven system, ISPRS, commission III symposium, pp. 711-717, Munich (Germany).

Ruskoné, R., Airault, S., Jamet, O. (1995): Road network extraction by local context interpretation, Image and Signal Processing for Remote Sensing, Satellite Remote Sensing II, SPIE, vol. 2315, pp. 508-518, Paris (France).

Ruskoné, R., Guigues, L., Airault, S., Jamet O. (1996a): Vehicle detection on aerial images: a structural approach, ICPR proceedings, Vienna (Austria).

Ruskoné, R. (1996b): Road Network Automatic Extraction by local context analysis: application to the production of cartographic data, PhD thesis (english version), Marne-la-Vallée University.

Ruskoné, R., Airault, S. (1996c): Vers une interprétation automatique du réseau routier sur images aériennes: détection et analyse des carrefours, International Journal of GIS and Spatial Analysis, vol.6, 2-3.