A HIERARCHICAL APPROACH TO AUTOMATIC ROAD EXTRACTION FROM AERIAL IMAGERY

C. Heipke[1]

Dep. of Geodetic Science and Surveying, Ohio State University, 1958 Neil Avenue, Columbus, OH, 43210, USA

C. Steger, R. Multhammer

Chair for Image Analysis and Knowledge-Based Systems, Technical
University Munich, Orleansstr. 34, D-81667 Munich, Germany

ABSTRACT

In this paper we describe a new multi resolution approach to automatic road extraction from aerial images. We make use of the fact that different characteristics of objects such as roads can be best detected in different scales. Two different resolutions of the same image are used, a coarse one with 2 m per pixel, and a fine one with 0.25 m per pixel. In the coarse resolution roads are modeled as bright lines and are extracted by a combination of local and global thresholding. In the fine resolution roads are assumed to have two parallel edges, be bright, and have a homogeneous texture. A multi step procedure has been designed to find the roads according to these criteria. Subsequently both outputs are merged using a rule based system.

The developed method has been tested on real imagery, and some preliminary results are reported. Based on the existing experience the multi resolution approach is claimed to be superior to a road extraction in one resolution only.

1. INTRODUCTION

One of the most fascinating promises of digital photogrammetry is a highly automated acquisition and updating of spatial data from images. Notable progress has been made in areas involving image matching like the automatic derivation of digital terrain models (DTM; e.g. Krzystek, Wild 1992), automatic relative orientation (e.g. Schenk et al. 1991; Tang, Heipke 1995), and automatic aerial triangulation (e.g. Tsingas 1992; Schenk, Toth 1993). A number of these developments has already led to commercial products.

Also the recognition and accurate localisation of objects in digital imagery has attracted considerable attention in the past in photogrammetry and computer vision (e.g. McKeown 1991; Quam, Strat 1991). This task is more difficult than the ones mentioned above, because it involves labelling parts of the image with semantic descriptors rather than matching them, and an explicit theory of human vision, which could be followed for this task, does not exist. What is so easy for a human operator to accomplish, namely the identification of houses, roads, and other objects in the depicted scene, is much more difficult for a computer, since the human operator cannot explain, how he/she carries out the job. Evidently, a great deal of training, experience, and background knowledge is involved. Besides combining data from different sensors (e.g. frame or line cameras, imaging and non-imaging lasers, multi spectral scanners) and results from different methods, the representation and utilisation of the knowledge, which require artificial intelligence techniques, have been recognized as a major issue both in digital photogrammetry and in computer vision in order to achieve progress in object recognition and localisation (see e.g. Schenk 1994 for further discussion on this subject).

---

[1] on leave from: Chair for Photogrammetry and Remote Sensing, Technical University Munich, Arcisstr. 21, D-80333 Munich, Germany

In the existing literature on object recognition and localisation from aerial images two ways to attack the problem can be identified: a fully automatic and a semi-automatic approach. In the former a human operator starts a process and after its completion has to check and edit the results, in the latter the operator identifies the objects of interest and provides some starting conditions, before extraction algorithms are invoked, which in turn ask for the help of the operator if necessary. The first approach is more attractive to many researchers, perhaps because it may ultimately provide insight into how the human mind works. The second approach is more promising for short-term and mid-term practical applications and has for instance been adopted for the RADIUS project (Research And Development in Image Understanding Systems), a five year program funded by the Advanced Research Projects Agency (ARPA) in the United States (Edwards et al. 1992). In both cases the approach has to be implemented on a digital photogrammetric workstation (e.g. Heipke 1994; ZPF 1995) in order to be integrated with the whole photogrammetric workflow.

Automatic road extraction from digital aerial and satellite imagery has received considerable attention using both the fully automatic approach (e.g. Barzohar, Cooper 1993; Zerubia, Merlet 1993; Zlotnick, Carnine 1993; Ruskoné et al. 1994) and the semi-automatic one (e.g. Quam 1978; McKeown et al. 1988; Grün, Li 1994; Heipke et al. 1994). Existing methods, however, lack the versality necessary for broader applications.

In this paper we describe ongoing research on multi resolution road extraction from aerial images. We have chosen to follow the fully automatic approach. One goal is to extract roads suitable for inclusion into the German base GIS (geographic information system) ATKIS (Amtliches Topographisch-Kartographisches Informationssystem), more specifically for the DLM25 (Digitales Landschaftsmodell) with a content approximately equivalent to that of a topographic map 1:25 000. Besides extracting ATKIS DLM25 data we are also interested in updating and increasing the accuracy of ATKIS, and in data acquisition for larger scales. In the long run we want to develop a strategy which uses a combination of top-down and bottom-up processes for road extraction based on appropriate object space models for roads. In this paper, however, we report on a bottom-up approach starting with the raw images and leading to a description for the roads in vector format. We have designed a new multi resolution approach, which makes use of the fact that different characteristics of objects can be best detected at different points in scale space. At present we use two different levels of resolution and subsequently combine the obtained results using a rule based system.

In the next chapter we motivate our approach emphasizing the need for a road model. Then the extraction of roads at the two resolution levels using two different road models is described, and the results of each step are discussed. Subsequently, we explain how the two outputs are combined and show that the combination delivers better results than any of the algorithms used in isolation. We close with some concluding remarks and an outlook for further investigations.

2. THE NEED FOR HIERARCHY

In order for a vision system to recognize and localize objects in images it must know what to look for. In other words the system must possess a model of the object of interest and of its appearance in the image. Since we are dealing with objects originating in the three dimensional world the object models should also be three dimensional. For instance, a road is a linear object with a predefined range for width and curvature, and is generally flat. Roads connect houses and urban areas, and therefore exhibit a network structure. If one examines its appearance in images, roads are usually brighter than the surroundings, and the texture enclosed by the road edges is rather homogeneous.

The pixel size on the ground, or ground resolution, determines the size of an object in the image in pixel units. If a film image is scanned, the ground resolution is the product of the average image scale number and the scanning pixel size. The appearance of the object at different resolutions is important for choosing an appropriate object model. For instance, in order to extract a road with a maximum of geometric precision a small pixel size on the ground must be used. Then the road is wide enough to reliably extract the parallel road edges. However, local disturbances such as shadows cast by adjacent buildings or cars on the road can degrade the results considerably. Also adjacent bicycle tracks or walkways might cause a problem, if they cannot be separated from the road. Using larger pixel sizes on the ground the road position is extracted with less geometric accuracy but more easily, because the mentioned disturbances are not so prominent or may even not be visible at all due to averaging of the grey values. Thus, it seems beneficial to extract the desired objects at different ground resolutions based on different object models, and to combine the outputs of the individual levels to obtain refined object descriptions.

Our road models consist of the subgroups geometry and radiometry. Geometry captures such descriptions as location, shape and size of the object. Radiometry deals with the brightness or the colour of the object.

3. ROAD EXTRACTION AT DIFFERENT RESOLUTIONS

3.1. The data for the investigation

In order to study the performance of the suggested approach we have applied it to real imagery. Since we are going to illustrate the algorithm described in this chapter using one of the images we have tested it on, together with the results it produces in intermediate steps, it seems appropriate to introduce the image at this point. It should be noted that the image has been selected to explain the behaviour of the algorithm rather than as a benchmark test.

The image we use is black and white and of scale 1:15 000. It was digitized with a pixel size of 15 µm and 8 bit per pixel, yielding a ground resolution of approximately 0.23 m. The image shows a rural area around Marchetsreut in the Bavarian Forest near Passau/Germany, which includes a village, farm land and a forest (see figure 1). Although it is not essential for the work presented here, we have decided to carry out the road extraction in the orthoimage, because at a later stage of the project this will allow us to easily incorporate information to be updated, stemming e.g. from an existing GIS database, and to make use of the third dimension, e.g. by analyzing the DTM and excluding areas which can’t possibly contain roads. In order to generate the orthoimage a high fidelity DTM was captured manually on an analytical plotter using 1:4 000 imagery of the same area. It was then used to compute the orthoimage with a ground resolution of 0.25 m.

3.2. Low resolution road extraction

The road model at this resolution level is intended for roads with a width of about 3 - 5 pixels when depicted in the image. The goal is to extract the centre lines of the roads with a width of one pixel. In order to meet these requirements the original orthoimage was subsampled by a factor of 8. This operation resulted in a ground resolution of 2 m per pixel. In this resolution roads are modeled as lines brighter than the surroundings.

In order to detect roads the grey value of each pixel is compared to the local Gaussian mean, and pixels significantly brighter than the surroundings as determined by a predefined threshold are marked. Subsequently these pixels are investigated, and only those which have a grey value within a certain range are accepted as road pixels. This range must again be defined prior to the computations, and together with the mentioned threshold describes part of the radiometric road model at this resolution. Next, the skeleton of the selected pixels is computed.

As was to be excepted, besides the pixels thought to represent roads also a number of other pixels was detected. In order to eliminate them 8-connected contours of points are derived. The start and end point of each contour is either a junction point or a point with exactly one neighbour. The points of each contour are subsequently examined whether they are local maxima in the direction perpendicular to that of the contour. Only contours that have more than a certain percentage of points being local maxima are accepted. This step includes a geometric part, namely the grouping of pixels into linear contours, and a radiometric part, namely the fact that a point should be a local maximum to be accepted.

The results of the road extraction at this resolution are shown in figure 2. It can be seen that most of the roads are in fact extracted using this simple algorithm, even within the village. However, some additional features were also selected. Two detailed views of the results are shown in figure 3 a and b, where the extracted centre lines are overlayed with the imagery at the original resolution. The quality of the results can be better assessed there. While the road centres are

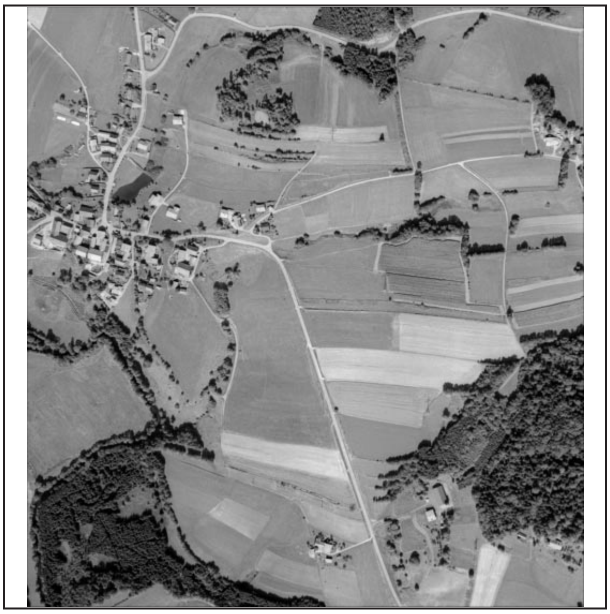

Fig.1: The test image

Fig. 2: Results of low resolution road extraction

Fig. 3a: Low resolution road extraction, detail 1

Fig. 3b: Low resolution road extraction, detail 2

indeed found in many cases, some road parts are not connected, especially at road intersections. Also house boundaries are sometimes selected. In summary it seems prudent to view the results as road hypotheses, which have to undergo additional verification at other resolution levels.

3.3. High resolution road extraction

At the original level of resolution of the orthoimage roads are essentially modeled as bright linear objects with parallel boundaries (geometric model) and a homogeneous area in between (radiometric model). A road may extend from an area with parallel boundaries into an adjacent area, e.g. near an intersection, provided that the criterium of homogeneity is still fulfilled. Deviations from this model exist due to various reasons. For instance a road can appear dark when wet or shortly after construction, and the homogeneity criterium is violated by shadows or occlusions caused by neighbouring 3D objects such as houses or trees, or objects on the road such as road markings or cars. Deviations associated with the neighbouring objects can be overcome by analyzing the 3D object surface surrounding the road. At this stage, however, we have not yet addressed this point.

An algorithm has been developed to find roads according to this model (see Multhammer 1994). To illustrate the various steps, the image of figure 3a will be used. First, edges are extracted from the image via an isotropic version of the Deriche edge detector (Deriche 1987; Lanser, Eckstein 1992). Subsequently a thinning operation is applied, yielding one pixel wide edges. These are then converted into contours. Next, the contours are approximated by polygons to facilitate further processing. We use the algorithm proposed by Ramer (1972). Contours are substituted by connected polygon segments with a limited distance to the approximated contour. Figure 4 shows the selected part of the image together with the computed polygons.

The next step is the selection (perceptual grouping) of the polygon segments into a relation of parallels. This step is of great importance, because a grouping procedure is very powerful in discriminating between object and non-object information (see e.g. Price, Huertas 1992 for further discussion on grouping). First the direction of each line is determined relative to a given reference. In order to be classified as parallel, two segments have to fulfil the following conditions:

- The difference of the two directions must be below a threshold. This threshold depends on the lengths of the segments, since the direction of longer lines is computed more accurately.
- The two segments must overlap. To determine overlap a line in the direction of the bisection of the angle formed by the two segments is computed, and the end points of both segments are projected onto the line. The segments are said to overlap, if and only if the projected lines overlap.
- The distance of the two segments must not exceed a certain threshold corresponding to the maximum width of the roads.

The results of this step are shown in figure 5. It can be seen that most of the parallel road segments have been detected, however, together with a number of segments originating from other objects. As expected, no segments have been selected at road intersections.

As part of the radiometric road model, the area between two parallel polygon segments is investigated next. According to the model this area should be bright and homogeneous. In an attempt to include road markings into the model, we break up the area under consideration into slices parallel to the direction of the bisection of the two segments (see figure 6). Mean and variance of the resampled grey values within each slice are computed. A pair of parallel segments is only accepted, if the mean falls into a predefined range and the variance is smaller than a given threshold. The results of this step are depicted in figure 7 and are called 'modified parallels'.

Fig. 4: Extracted polygons

Comparing the results with figure 6, the algorithm has successfully filtered out many of the non-road parallel segments. Obviously, parts of the road where no parallel segments exist, e.g. road intersections, still pose a problem. In order to overcome this deficiency, segments next to accepted road segments are recursively investigated for homogeneity of the adjacent area (see figure 8). They are accepted if the same homogeneity criteria as above are fulfilled. The results of this step are called 'extended parallels.'

Fig. 5: Found parallels

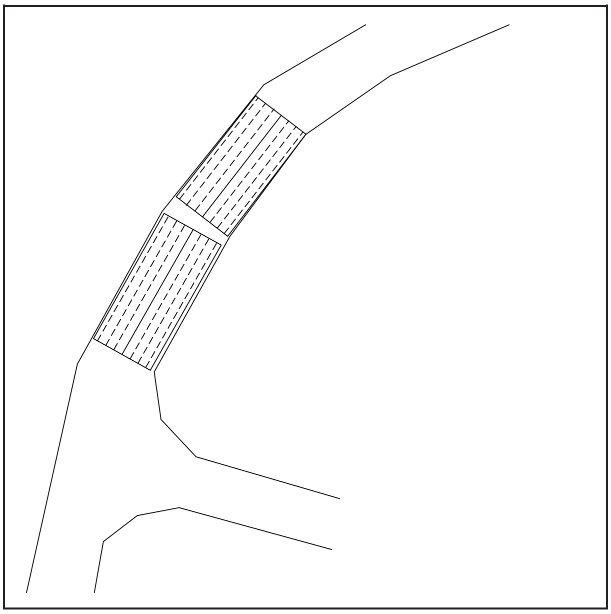

Fig. 6: Checking for homogeneity

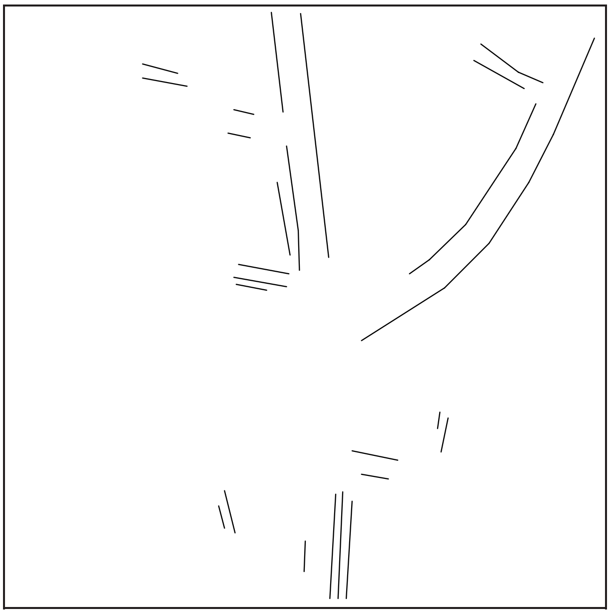

Fig. 7: Modified parallels

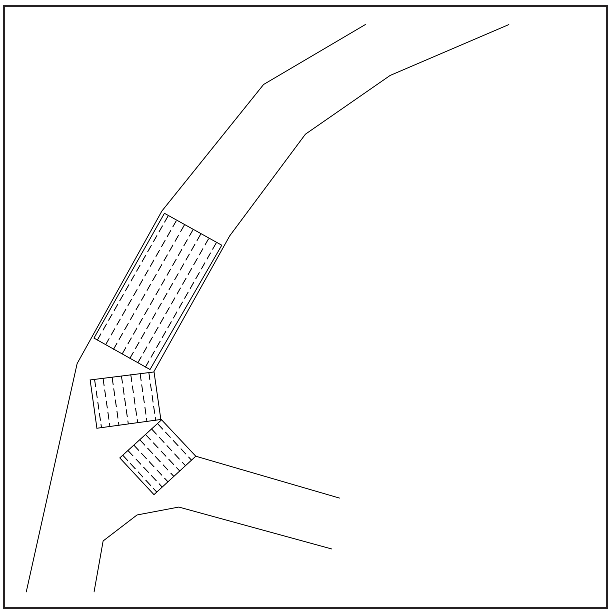

Fig. 8: Checking for homogeneity at road crossings

The final result of the high resolution road extraction algorithm is depicted in figure 9a. Figure 9b shows the results of the other image in figure 3b. Both examples have been computed with the same threshold parameters in order to facilitate a comparison of the results. As can be seen, road extraction was rather successful in these cases. However, some parts such as the lower part of the left road edge running from the top to the bottom in the middle of figure 9a have not been extracted due to lack of parallelism, and local disturbances, such as the shadows of the two trees in the upper part of figure 9b, also lead to incomplete road boundaries.

Fig. 9a: Results for high resolution, case 1

Fig. 9b: Results of high resolution, case 2

3.4. Combination of the results

As has been shown, both of the presented algorithms have advantages and disadvantages. In light of what was discussed in chapter 2 we have combined the results in order to increase their quality. The strategy adopted was to use combinations reinforcing each other to some extend with the risk that some roads only extracted at one of the resolution levels might be lost. The reason behind this strategy is the following: regardless of the quality of any single algorithm it will usually fail in at least some part of the image. Since we have not incorporated a quality measure for the output and thus we can only assess our results by visual inspection, we feel it is save to accept only hypotheses which are as stable as possible according to the chosen model. In order to ensure accurate positioning of the selected polygon segments, only results from the high resolution step are accepted. The low resolution results only serve as a guidance for eliminating false segments and adding missed segments.

We have formulated the combination of the two outputs in terms of a number of rules. They read as follows:

1. If two polygon segments p1 and p2 are modified parallels,
   and a centre line lies between p1 and p2,
   then accept p1 and p2.
2. If a polygon segment p2 is adjacent to an accepted segment p1,
   and p2 overlaps with a centre line,
   then accept p2.

3. If two polygon segments p1 and p2 are modified parallels,
   and there exists a polygon segment p3 collinear to p1,
   and p2 overlaps with p3,
   and there exists a centre line between p2 and p3,
   then accept p3.
4. If between two accepted polygon segments there exists exactly one segment p3,
   and if p3 is an extended parallel,
   then accept p3.

Rule 1, which is used to start the combination process, is rather conservative for the reasons explained above. Rule 2 extends the results, mainly in intersection areas. With rule 3 small gaps between segments can be bridged provided that a segment on the opposite side and a centre line exist. Rule 4 is the only rule which does not require a centre line and was introduced, because in intersection areas centre lines are sometimes not extracted as outlined before.

The results of the combination can be seen in figure 10a and b. Comparing them to the figures 3 and 9 improvements can be observed. Besides the elimination of a number of false segments still present in figure 9, the missing left road side in the lower centre part of figure 9a has been recaptured. The problems connected with the trees in figure 9b, however, could not be solved, since shadows and occlusions are not addressed in the current road model.

Fig. 10a: Final results, case 1

Fig. 10b: Final results, case 2

4. DISCUSSION AND CONCLUSION

We have presented some preliminary results obtained by our hierarchical road extraction approach. Roads were extracted in two different ground resolutions separately and subsequently the results were combined using a number of rules. Essentially, roads found on both levels were accepted.

We think the results give reason to believe that a hierarchical approach is more successful than a single resolution one. However, many problems still remain unsolved. To name just a few:

- How many and which levels of resolution should be considered?
- How does a more elaborate set of rules for the combination of the various outputs look like?
- How does feature tracking through the pyramid compare to the separate extraction of roads and the subsequent combination of the results?
- Is it beneficial to combine various algorithms on any one resolution level (as demonstrated e.g. by McKeown et al. 1988)?
- How to design a strategy, in which particular parts of the image, which have already been processed are revisited in order to verify hypotheses obtained elsewhere? Should the same algorithm be used as before with a different set of threshold parameters or a different algorithm?
- How to incorporate the concept of self diagnosis into the strategy, such that a measure of reliability of the results can be given (see e.g. Förstner 1991 for a more detailed discussion on this topic)?

In our future research we will address these questions and extend the presented approach to make it more versatile. Further improvements are expected to come from the use of colour and a DTM representing the visual surface, as additional sources of information. An analysis of this DTM can provide cues for 3D objects (see e.g. Al-Tahir, Schenk 1992; Haala 1994) and thus for possible occlusions and shadows. Therefore, it can for instance help to solve the problems encountered with the trees in figure 10b.

Another issue which needs more attention is topology. Observing figure 3 again it should be possible to predict the approximate location of road intersections with some degree of reliability. Topology is also needed in the construction of a road network.

We will also study ’context’ (the interpretation of individual parts of the image based on the interpretation of lager parts thereof). Context is currently a major issue in computer vision (e.g. Strat 1992; 1994). While it is intuitively clear that context plays an important role in image interpretation, its application in detail is less evident, because the use of context implies that other knowledge of the scene already exists. Thus, questions such as ’How much of the surroundings need to be taken into consideration?’, ’Which additional objects are crucial to the extraction of the desired ones?’, ’Where does this additional information come from?’ arise. Most of the answers are, of course, dependent on the application. However, in the task of updating an existing GIS database the old status can constitute at least some of this knowledge.

5. ACKNOWLEDGEMENTS

We would like to acknowledge the vital contributions of Heinrich Ebner, Wolfgang Eckstein, Clemens Glock, Helmut Mayer, Werner Meixner, and Bernd Radig to this research. Funding was provided by the Deutsche Forschungsgemeinschaft under grant no. Eb 74-8/1.

6. REFERENCES

CVGIP:IU	Computer Vision, Graphics and Image Processing: Image Understanding
CVPR	The IEEE Computer Society Conference on Computer Vision and Pattern Recognition
DGK-C	Deutsche Geodätische Kommission, Reihe C
ICPR	International Conference on Pattern Recognition
IJCV	International Journal of Computer Vision
IntArchPhRS	International Archives for Photogrammetry and Remote Sensing
IUW	(D)ARPA Image Understanding Workshop
JPhRS	ISPRS Journal of Photogrammetry and Remote Sensing
PE&RS	Photogrammetric Engineering and Remote Sensing

Al-Tahir R., Schenk T., 1992: On the interpolation problem of automated surface reconstruction, IntArchPhRS (29) B3, 227-232.
Barzohar M., Cooper D.B., 1993: Automatic finding of main roads in aerial images by using geometric - stochastic models and estimation, Proceedings CVPR, 459-464.
Deriche R., 1987: Using Canny's criteria to derive a recursively implemented optimal edge detector, IJCV (1) 5, 167-187.
Edwards J., Gee S.J., Newman A.M., Onishi R., Parks A., Sleeth M., Vilnrotter F., 1992: RADIUS: Research and development for image understanding systems phase 1, Proceedings IUW, 177-184.
Förstner W., 1991: Statistische Verfahren für die automatische Bildanalyse und ihre Bewertung bei der Objekterkennung und -vermessung, DGK-C 370.
Grün A., Li H., 1994: Semi-automatic road extraction by dynamic programming, IntArchPhRS (30) 3/1, 324-332.
Haala N., 1994: Detection of buildings by fusion of range and image data, IntArchPhRS (30) 3/1, 341-346.
Heipke C., 1994: Digitale photogrammetrische Arbeitsstationen, Habilitationsschrift, TU München.
Heipke C., Englisch A., Speer T., Stier S., Kutka R., 1994: Semi-automatic extraction of roads from aerial images, IntArchPhRS (30) 3/1, 353-360.
Krzystek P., Wild D., 1992: Experimental accuracy analysis of automatically measured digital terrain models, in: Förstner W., Ruhwiedel S. (Eds.), Robust Computer Vision, Wichmann, Karlsruhe, 372-390.
Lanser S., Eckstein W., 1992: A modification of Deriche's approach to edge detection, Proceedings ICPR (III), 633-637.
McKeown D.M., 1991: Information fusion in cartographic feature extraction, in: Ebner H., Fritsch D., Heipke C. (Eds.),Digital Photogrammetric Systems, Wichmann, Karlsruhe, 103-110.
McKeown D.M., Denlinger J., 1988: Cooperative methods for road tracking in aerial imagery, Proceedings CVPR, 662-672.
Multhammer R., 1994: Straßenextraktion aus digitalen Luftbildern, Diplomarbeit, Forschungs- und Lehreinheit Informatik IX und Lehrstuhl für Photogrammetrie und Fernerkundung, Technische Universität München (unpublished).
Price K., Huertas A., 1992: Using perceptual grouping to detect objects in aerial scenes, IntArchPhRS (29) B3, 842-855.
Quam L.H., 1978: Road tracking and anomaly detection, Proceedings IUW, 51-55.
Quam L.H., Strat T.M., 1991: SRI image understanding research in cartographic feature extraction, in: Ebner H., Fritsch D., Heipke C. (Eds.), Digital Photogrammetric Systems, Wichmann, Karlsruhe, 111-122.
Ramer U., 1972: An iterative procedure for the polygonal approximation of plane curves, Computer Graphics and Image Processing (1) 244-256.
Ruskoné R., Airault S., Jamet O., 1994: Road network interpretation: a topological hypothesis driven system, IntArchPhRS (30) 3/2, 711-717.
Schenk T., 1994: Concepts and algorithms in digital photogrammetry, JPhRS (49) 6, 2-8.
Schenk T., Li J.-C., Toth C., 1991. Towards an automatic system for orienting digital stereopairs. PE&RS (57) 8, 1057-1064.
Schenk T., Toth. C., 1993: Towards a fully automated aerotriangulation system. American Congress on Surveying and Mapping/American Society of Photogrammetry and Remote Sensing Annual Convention, Technical Papers (3), 340-347.
Strat T.M., 1992: Natural object recognition, Springer, Berlin.
Strat T.M., 1994: Photogrammetry and knowledge representation in computer vision, IntArchPhRS (30) 3/2, 784-792.
Tang L., Heipke C., 1995: Automatic relative orientation of aerial images, accepted for publication in PE&RS.
Tsingas V., 1992: Automatisierung der Punktübertragung in der Aerotriangulation durch mehrfache digitale Bildzuordnung, DGK-C 392.
Zerubia J., Merlet N., 1993: Classical mechanics and road detection in SPOT images, INRIA Rapport de Recherche No. 1889, INRIA, Sophia Antipolis.
Zlotnick A., Carnine P.D., 1993: Finding road seeds in aerial images, CVGIP:IU (57) 2, 243-260.
ZPF, 1995: Zeitschrift für Photogrammetrie und Fernerkundung, special issue on digital photogrammetric workstations (63) 1.