NEAR-REAL-TIME ROAD EXTRACTION FROM SATELLITE IMAGES USING VECTOR REFERENCE DATA

Nikolas Prechtel and Oliver Bringmann
Dresden University of Technology, Institute for Cartography, Germany

Commission II Working Group 5

KEY WORDS: line extraction, data reduction, vector-to-image matching, real-time processing, data integration

ABSTRACT: A method is presented to reduce the huge amount of data generated by an imaging system to the actual objects of interest. This is achieved by a fast algorithm, which is performing a matching of given vector information, localising these objects, with pre-processed image contents. Starting with a rough estimation (GPS) of the relative position of the vector vertices in the image, one tries to set up a transformation matrix to map them to pixel structures of the assumed target class. For the iterative matching process, a new rating criterion is introduced: a potential field derived from a target-specific filtered image. In the context of a small-satellite programme of the Dresden University of Technology, the method has been adapted to road extraction for traffic monitoring.

KURZFASSUNG: Bildgebende Systeme generieren oft eine große Menge an Bilddaten. Es wird eine Methode vorgeschlagen, die nur die eigentlich interessierenden Bildteile extrahiert. Ausgehend von einer groben Positionsschätzung (GPS) werden Vektoren einer Geodatenbasis dem Bild so überlagert, daß die Stützpunkte die betreffenden Bildstrukturen optimal treffen. Dazu wird eine entsprechende Transformationsmatrix ermittelt. Die schnelle Lösung der Optimierungsaufgabe erfordert ein neues Bewertungskriterium: ein Potentialfeld, das aus dem gefilterten Originalbild erzeugt wird. Die Methode wurde im Rahmen eines Satellitenprogrammes der Technischen Universität Dresden angewandt. Hierbei diente sie der Straßenextraktion aus Bilddaten zur nachfolgenden Verkehrsbeobachtung.

1 INTRODUCTION

1.1 Motivation an Problem Specification

Image series which are generated by high-resolution cameras mounted on low-orbiting small satellites can typically not be classified on-board. But low orbits and very high resolution are leading to a huge data flow, which conflicts with limited down-link capacities of a proposed "low-budget" system. A significant reduction of the data flow in real-time has to precede the data transmission and seems possible when, as in our case, only small parts of the real-world objects, mapped by the camera, pertain to the objects of interest. Despite a detailed knowledge of the location of those real-world objects - in the case of the present study the high-ranking parts of the Central-European road network - a direct on-board masking of the image is inhibited by an inadequate determination of the field of view of the sensor during the overflight.

The objective of the study is to present a quick and robust method to improve the near real-time positional information. It makes use of a typical road signal in the image, inside of pre-selected areas of interest, in connection with a corresponding road vector reference. With the help of a precise detection of target locations in an image frame, a closed polygon can be calculated to separate the target from the background areas. Afterwards, the data subset can be handed over to a standard compression, and then fits the performance of the transmission module. Monochrome images (VIS/NIR) of a value range (0...*maxV*) are considered.

1.2 Concept of a Solution

Well-documented solutions for road extraction, on a high semantic level, will not suit the real-time demand (Eckstein, 1996; Barzohar, 1996). Hence, an algorithm had to be

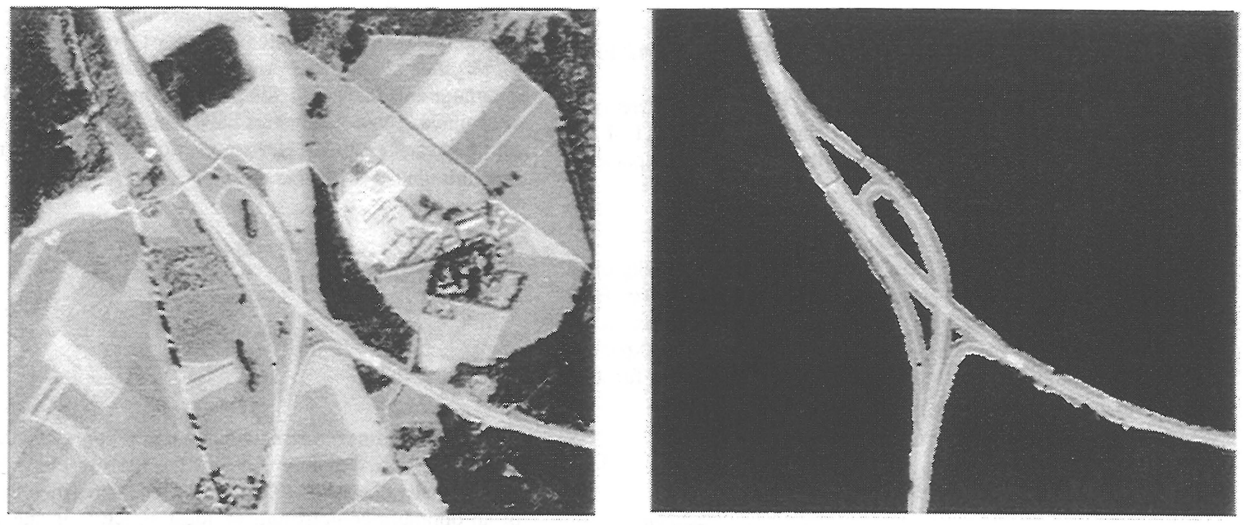

Figure 1: Motorway, input image (left) and output after matching and masking (right).

developed that most directly allows to calculate the transformation parameters, for the matching of the given road vector reference, to the corresponding image structures. The algorithm requires a consistent external data set of road middle-axes augmented by a local estimation of the road width and a global estimation of the mean greyvalue of the road surface. The roads will not be detected in a bottom-up manner, rather the vector reference shall be matched to enhanced image structures in total. By the long focal length of the proposed camera and a near-nadir view of all elements of an image frame, an affine projection model should be suitable to transform the co-ordinates x of the vector reference into the image with a satisfying fit.

A matrix **T** has to be set up in such a way that the homogenous co-ordinates $\mathbf{T}x$ will be adjusted to the road middle-axes in the image. In a first step, potential candidates for the roads in the matrix have to be emphasised.

This is done in several steps. The pre-processing consists of a stretch of the greyvalues around the mean road greyvalue. $b(u)$ indicates the greyvalue of a contrast-stretched image for a pixel with the homogeneous co-ordinate u. Now, the shape criterion is added when the monochrome image is converted into a matrix $s(u)$ of the same size. The new greyvalues are expressing the probability of being lined up along the middle-axis of a road, or what looks like a road in a local context.

This matrix serves as input for the calculation of a potential field $p(u)$. $p(u)$ contains a combined measure of the probability of membership in the road middle-axis class and the geometric distance to the assumed road pixels of the preceding matrix. Visually, $p(u)$ has the appearance of $s(u)$ after having passed a strong smoothing filter.

The second principal task is the numeric solution of a maximisation problem. One looks for an invertable matrix **T** with:

$$\sum_{\forall x} p(\mathbf{T}\,x) = \max! \qquad (1)$$

A direct calculation of **T** is impossible, since p cannot be derived for the parameters of the transformation matrix. The coefficients of **T** are, therefore, determined by a gradient ascent algorithm. The ascent starts with the coefficients of an identical projection.

2 EXTRACTION OF LINES OF A DEFINED WIDTH

To search for road middle-axis pixels in the image, preference is given to a specially adapted local linear filter instead of the application of a vectorisation algorithm. The filter response is written into a matrix $s(u)$.

2.1 Contrast Stretch

A contrast stretch is performed to harmonise the background signal, which, consequently, leads to a more uniform filter response during the calculation of the 'road matrix' s. A greyvalue range *roadVR* with a centre value *roadV* is expanded to the definition range 0 to *maxV*. The following linear function from original (*gv*) to new greyvalues (*gv'*) is used:

$$gv' = \begin{cases} 0 & gv < roadV - \dfrac{roadVR}{2} \\ \left(gv - roadV + \dfrac{roadVR}{2}\right)\left(\dfrac{maxV}{roadV}\right) & otherwise \\ \max V & gv > roadV + \dfrac{roadVR}{2}. \end{cases} \qquad (2)$$

2.2 Linear Filter Design

According to the spectral reflectance of road-surfaces in the visible range, the model assumes bright roads on a darker background. The road curvature shall be low in relation to the road width, an assumption that is in accordance to the objective of an extraction of principal roads outside of towns only. The imaging process is taken into account by a low-pass filtering of an idealised rectangular road-background image model. In a cross-section the proposed filter consists of two symmetrical "wings" with a dike-shaped form (comp. fig. 2). Such a filter pair is applied to four main directions (*North, West, Northwest, Northeast*).

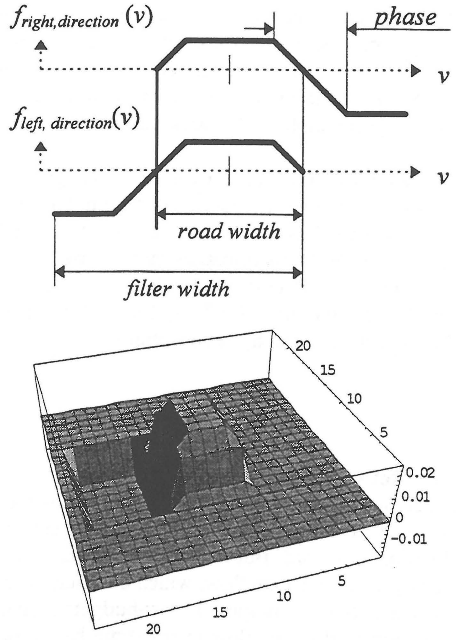

Figure 1: Filter cross-section (above) and perspective view on $f_{left,\ north}$ (below).

In order to account for minor bends of roads and/or small derivations from the maximum response direction, the filter surface is flattened (decreasing weights) to value zero towards both ends in the mentioned main line directions. Such a dike-shaped filter can be generated by a set of fixed parameters: *road width, filter length, filter width,* and *phase*. At this stage, the minimum and maximum filter values (*height of filter*) are not determined. These parameters have been calculated by solving a set of linear equations. Hereto, the following constraints have been introduced (with $side \in \{left, right\}$):

$$\sum_{\forall v} f_{side,direction}(v) = 0\,, \qquad (3)$$

$$\sum_{\forall v:\, f(v)>0} f_{side,direction}(v) = 1\,. \qquad (4)$$

By equation 3 image regions with a constant grey value deliver a response of 0. Equation 4 keeps the filter response inside of the original value range (0...*maxV*). *maxV* is reached in the case of an optimum contrast and an ideal structure in the image (a straight line of constant assumed width, in the same orientation

$$
\begin{aligned}
s_{left,direction}(u) &= \sum_{\forall v} f_{left,direction}(v)\, b(u-v) \\
s_{right,direction}(u) &= \sum_{\forall v} f_{right,direction}(v)\, b(u-v) \\
s_{direction}(u) &= \begin{cases} \sqrt{s_{left,direction}(u)\, s_{right,direction}(u)} & s_{left,direction}(u) > 0 \text{ and } s_{right,direction}(u) > 0 \\ 0 & otherwise \end{cases}
\end{aligned} \tag{5}
$$

as the filter). Since any of the four filter directions leads to two initial values due to the filter pair used, the following step means combining the left and right responses in such a way that line structures are preserved but edges between more spacious, highly contrasting areas are subdued (equations 5).
An open parameter of the equations 5 is the favoured direction of the model road. For eight potential road directions four filters with axes rotated by 0, $\pi/4$, $\pi/2$ and $3\pi/4$ have been generated. The filter shape as described above is flexible enough to cope with a maximum angular deviation of $\pi/8$ of a straight line in the image without showing a strong drop in response. The maximum response of the four directional filter processes is written into matrix s.

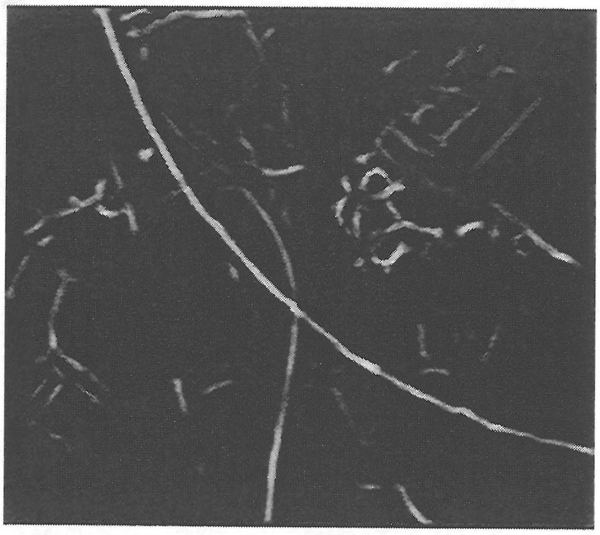

Figure 2: Filter response $s(u)$.

3 POTENTIAL FIELD AROUND DETECTED LINES

The matrix s, with its greyvalues expressing a rating of road membership, is not appropriate to steer the calculation of the transformation matrix **T**. **T'** may be a good approximation of the ideal transformation matrix **T**. Then, a global rating following equation 6 could eventually deliver:

$$\sum_{\forall x} s(\mathbf{T}\, x) < \sum_{\forall x} s(\mathbf{T'}\, x) . \tag{6}$$

For an iterative process like gradient ascent it is, however, necessary that close to a (local) maximum one always gets:

$$\sum_{\forall x} p(\mathbf{T}\, x) \geq \sum_{\forall x} p(\mathbf{T'}\, x) . \tag{7}$$

Therefore, one looks for a matrix p to guarantee this correct behaviour. A fast algorithm with the required performance has been described by Rosenfeld and Pfaltz (Rosenfeld, 1968). It has been introduced into the image processing theory as a distance transformation algorithm under the expression 'recursive morphology'. It has a time complexity of the order $O(n)$ with n indicating the number of pixels. Apart from equation 7 two further conditions for p shall be mentioned:

- Dominant maxima in s shall also be found in p:

  $\forall u$: $\forall u' \in$ *environment*(u): $s(u) > s(u') \Rightarrow p(u) > p(u')$. (8)

- The maximum gradient of p shall be limited:

  $\exists \varepsilon \in N$: $\forall u : \forall u' \in$ *neighbourhood*(u): $| p(u) - p(u') | < \varepsilon$. (9)

The second condition leads to a certain discrete steadiness of p, which s cannot have. This is crucial to the correct behaviour of p as demanded in equation 7. The algorithm of Rosenfeld and Pfaltz has been modified for our purpose.

Figure 3: Potential field $p(u)$.

3.1 Algorithm to generate the potential field

For the calculation of p we need the matrix s and an auxiliary matrix h to be of the same size. After an initialisation step which writes zeros to all relevant border pixels, the process moves through the input image from upper left to lower right. Maximum grey values are propagated through the array and become stored in the auxiliary matrix h (diminished by a value ε). An analogue process now starts from the opposite corner, and moves from the lower right to the upper left.

```
for(int c = 0; c < columns; c++)
        h_0,c = 0
        p_rows-1,c = 0

for(int r = 0; r < rows; r++)
        h_r,0 = 0
        p_r,columns-1 = 0

for(int r = 1; r < rows; r++)
        for(int c = 1; c < columns; c++)
                h_r,c = max(s_r,c, sub(h_r-1,c, ε), sub(h_r,c-1, ε)

for(int r = rows-2; r ≥ 0; r--)
        for(int c = columns-2; c ≥ 0; c--)
                g_r,c = max(h_r,c, sub(g_r+1,c, ε), sub(g_r,c+1, ε)
```

Since the greyvalues have to be clipped at 0, the function sub performing a subtraction has to be modified in the following way:

```
byte sub(byte a, byte b)
        if (a < b) return 0
        else return a - b
```

4 CALCULATION OF THE OPTIMUM TRANSFORMATION PARAMETERS

Unfortunately, the vertices of the vector reference cannot directly be connected to homologue points in the image signal. Therefore, it is impossible to determine the transformation matrix $\mathbf{T}$ with the help of a Direct Linear Transformation Algorithm.
With the initial position parameters x being close to the transformed parameters $\mathbf{T}x$ and the density of road structures (or similar road artefacts) in the enhanced image being low, $\mathbf{T}$ can be calculated using a gradient ascent method (comp. fig. 5). Whenever the problem arises, that a local maximum instead of a global maximum will be found, one has to substitute the gradient ascent method by an algorithm of the type Simulated Annealing (Lengauer, 1990), which is more robust, but on the other hand slower. This implies the determination of a cooling strategy and elementary variations of the transformation parameters, which, obviously, must relate to the accuracy of the original position approximation. In the present paper the simple gradient ascent method will be described.
The transformation of the vertices of the vector reference x shall be given by a function $trans(x, \mathbf{p})$. $trans$ is defined as:

$$trans(x, \mathbf{p}) := \mathbf{T}x. \qquad (10)$$

$\mathbf{p}$ indicates a vector of the transformation parameters, or, in general, the coefficients of the matrix $\mathbf{T}$. The task is the solution of an optimisation problem of the form:

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\arg\max} \sum_{\mathbf{x}} p(trans(\mathbf{x},\mathbf{p})). \qquad (11)$$

The components of $\mathbf{p}$ will be varied by small steps, in total by $\Delta\mathbf{p}$. The corresponding rating of the new configuration is getting calculated in connection to the new parameter vector:

$$\Delta\mathbf{p}_{opt} = \underset{\Delta\mathbf{p}}{\arg\max} \sum_{\mathbf{x}} p(trans(\mathbf{x},\mathbf{p}+\Delta\mathbf{p})). \qquad (12)$$

The best variation is accepted:

$$\mathbf{p} = \mathbf{p} + \Delta\mathbf{p}_{opt}. \qquad (13)$$

Figure 4: Result of optimising procedure.

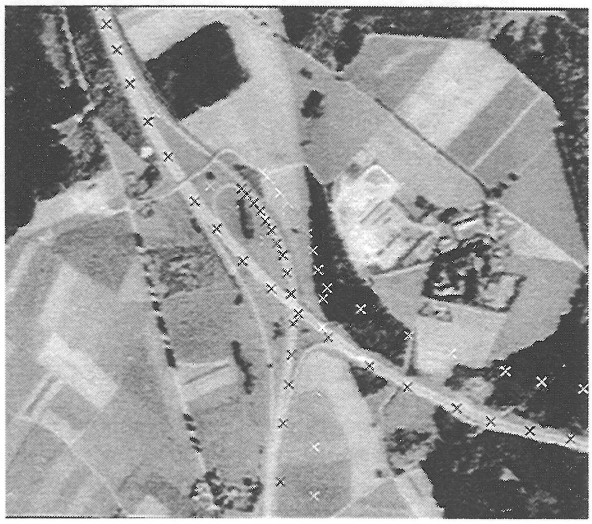

Figure 5: Result mapped on input image.

5 TUNING OF THE METHOD

The adjustment of the vertices x to the image structures does not require to be carried out for all x. On the contrary, it is more effective to select few subsets of the vector reference, and to restrict the processing to them alone. This will naturally allow the calculation of the matrices s and p only for image subsets. In this context, a few criteria for a subset choice in the reference data can be given:

- expectation of a low number of road-like structures in the image which might lead to a confusion with real roads
- favourable mutual position of street segments to guarantee a robust determination of transformation parameters (e.g. short straight sections enclosing right angles)

The combination of the vertices of all selected parts of the road reference with the matrix *p* forms the source for an integral optimisation using methods like gradient ascent or Simulated Annealing.

The incorporation of different street types (reflecting on mean greyvalue and mean width) increases the robustness of the approach. Then, type-specific filter sets will produce various (!) matrices *p* and *s*. A geometric overlap of these type-dependent *s* and *p* is not detrimental at all. On the contrary, a higher robustness of the optimising procedure can be expected, when, for example, two parallel roads of different type will not compete in the calculation of the transformation parameters.

After the calculation of the transformation parameters the masking of the image frame can take place using the complete transformed road reference. A standard compression procedure follows. Then the data are sent to the ground segment. To account for inaccuracies, the transformed road reference will be expanded to tolerance corridors. The width of these corridors will be calculated with an empirical function of the value $p(\mathbf{T}x)$ and the type-dependent local road width.

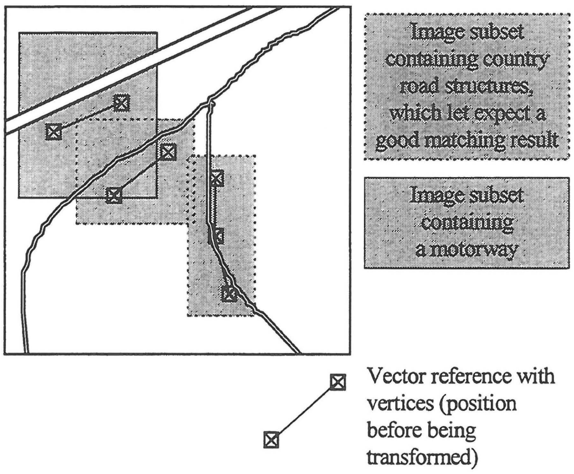

Figure 6: Segments of vector reference labelled for on-board matching.

6 TIME COMPLEXITY

This paragraph is dedicated to a treatment of the number of elementary operations necessary for the objective of an improved position estimation. The following set of operations plays a role:

L	Load/Save
M	Multiplication
A	Addition/Subtraction
C	Comparing
RT	Reading of lookup table

The open parameters of the algorithms are named as follows:

p	number of pixels
fc, fr	filter size by columns and rows
v	number of vertices of the vector reference
o	number of circles until the optimum will be found
para	number of parameters of the transformation.

For the filtering we get the following term:

$$\text{filtering}(\,p, fc, fr\,) = ((\,fc * fr)\,(\,\mathbf{M} + \mathbf{A} + 2\mathbf{L}\,) + 3\mathbf{C} + \mathbf{LT} + 2\mathbf{A})\ p. \tag{14}$$

The generation of the potential relief requires

$$\text{potential}(p) = 2\,(\,2\mathbf{A} + 2\mathbf{V} + 4\mathbf{L}\,) \tag{15}$$

operations. A simple gradient ascent is independent of the number of pixels checked and leads to the following number of elementary operations:

$$\text{gradient\_ascent}(\,s, o, para\,) = v * o * ((\,3para - 1\,) * \text{trafo}(\,para\,) + \mathrm{L}\,). \tag{16}$$

trafo(para) expresses the number of additions and multiplications for the transformation of one pixel vector:

$$\text{trafo}(\,para = 3\,) = 4\mathbf{M} + 6\mathbf{A}, \tag{17}$$
$$\text{trafo}(\,para = 5\,) = 6\mathbf{M} + 6\mathbf{A}. \tag{18}$$

From the preceding formulae one derives the following number of operations accounting for the whole process of position estimation:

$$\text{position} = \text{filtering} + \text{potential} + \text{gradient\_ascent}, \tag{19}$$

$$\text{position} = ((+ fc * fr\,)\mathbf{A} + 2(\,4 + fc * fr\,)\mathbf{L} + (fc * fr)\mathbf{M} + 7\mathbf{C})p + o * v * (\mathrm{L} + (3para - 1) * \text{trafo}(\,para\,)). \tag{20}$$

7 ABBREVIATIONS USED

u	homologue co-ordinates of pixels
$b(u)$	image function with optimised contrast
$s(u)$	rating of pixels according to membership in road middle-axis class
$p(u)$	rating of pixels according to the distances to the next (assumed) road pixel
x	homologue co-ordinates of the vertices of the vector reference
$\mathbf{T}$	transformation matrix performing an affine projection of x to u with an optimal matching of the vector reference to the road structures in the image $b(u)$
$\mathbf{p}$	vector of relevant coefficients of $\mathbf{T}$
v	homologue co-ordinates of a filter matrix
$f(v)$	value of a filter matrix at position v
$trans(x,\mathbf{p})$	transformation of x with parameters $\mathbf{p}$

8 REFERENCES

Barzohar, M. and Cooper, D. B., 1996. Automatic Finding of Main Roads in Arial Images by Using Geometric-Stochastic Models and Estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 7, pp 707-721

Eckstein, W., 1996. Extraction of Roads in Aerial Images using Different Levels of Resolution, Erfassung der Landschaft und ihrer Veränderungen - Einsatz von Luftbildern und Satellitendaten (List, F. K., ed.) - Publikationen der DGPF, Vol. 4, pp 283-292

Lengauer, T., 1990. Combinatorial Algorithms for Integrated Circuit Layout, B.G. Teubner, Stuttgart

Rosenfeld, A. and Pfaltz, J., 1968. Distance Functions in Digital Pictures, Pattern Recognition, Vol. 1, pp 33 - 61