SEGMENTATION AND TEXTURE ANALYSIS

W. Eckstein
Technische Universität München
Forschungsgruppe Bildverstehen (FG BV), Informatik IX
Orleansstr. 34, 81667 München, Germany
eckstein@informatik.tu-muenchen.de

Commission III, Working Group 2

KEY WORDS: Segmentation, Aerial Imagery, Algorithms, Vision, Feature Extraction, Recognition

ABSTRACT

This paper describes the state of the art in segmentation algorithms of aerial images. Different approaches and object classes are described and their advantages and limitations are shown. First the advantage of multiple input data (e.g., color, infrared, DEM) and the information that can be derived from these sources is discussed. Besides sensor data, "synthetic" input images (e.g., using texture filters) are generated to support the segmentation process. After an optional noise cleaning, primitives are extracted in scale space. This offers the possibility of selecting an optimal resolution depending on the size and shape of an object. Using this resolution, the raw segmentation will be stable and conflicts with other object classes will be reduced. Depending on the class of the object the final extraction has to be selected: Compact artificial objects can be segmented using primitives like areas, lines, or points. Linear objects like roads are similar but the borders are curves and the size is not limited. Arbitrary areas like meadows, forests, or fields have an arbitrary border and are mainly defined by their specific texture. Objects like trees or cars have to be treated in a very specific manner. Finally, different base algorithms for segmentation are discussed: Pixel classification is very simple but lacks the use of context. The extraction of primitives (egdes, lines, area, points) can be used as a basis for a wide class of objects. Texture analysis can be used for a rough segmentation of the image. Specialized operations are useful for the extraction of objects like single trees or to support the interpretation process.

1 INTRODUCTION

Before describing the topics of segmentation we have to discuss one important question: *Is there a known algorithm to extract all objects in aerial images?* The answer to this question is *no* and will remain *no* for many years and it is not even clear if there exists any. Segmentation is not just applying one sophisticated procedure and thus extracting all desired objects. On the other hand there exist a lot of more or less specialized algorithms. These have to be selected, depending on the classes of objects to be extracted, the resolution of the image, and the type of sensor.

The reason for this is the complexity of an aerial image. There are completely different classes of objects, like, buildings, roads, rivers, trees, meadows, fields, rocks, ice, hills, cars, poles, bridges, ships, waves, to name but a few (see figure 1). These classes have different extensions (e.g., cars and roads), specific or indifferent shapes (e.g., trucks and forest), uniform or textured surface (e.g., roofs and forest), which also depends on the resolution and can be extracted locally or only globally (e.g., trees versus rivers). In addition, the appearence of objects changes depending on the point of view, the weather, the time of day, and the season.

On the other hand there is a lot of information about the object classes. This knowledge can be used in multiple ways: As the basis for the interpretation, but also to design the segmentation procedure in two directions. Firstly, the selection of sensors and procedures operating on their data defines the static (procedural) knowledge incorporated in the system. Second, shape, topology, and radiometry of object classes can be used to control the segmentation process during runtime (dynamic knowledge). Constructing a system for a "complete" segmentation of an image (i.e., with different object classes) the following points have to be observed:

- Use all input sources that ease the task.
- Select the optimal resolution for every object class.
- Select an optimal strategy for the extraction of every object class.

Neglecting one of these points will limit the system significantly or at least adds a lot of work for the developer.

2 SOURCES

In the case of aerial image analysis a lot of data sources are available. Different sensors which allow a more stable extraction of a special class of objects can be used. Additional information, like the position of the sun (for shadows) or the angle of view (for the interpretation of 3D objects), can be used.

2.1 Color

Most interpretation of aerial images is done based on black and white pictures. The reason is mainly the availability of these pictures, and lower cost for digitizing and the required computer equipment. Many problems can actually be solved using this kind of images. Nevertheless, additional channels, like color or infrared, can ease the task (Ford and McKeown Jr., 1992). Given the task of interpreting suburb regions, for example, green areas are probably lawns, red rectangular areas are candidates for roofs, and small red, yellow or blue rectangular areas on the road are probably cars (see figure 2). Using infrared, the extraction of vegetation is even more stable.

The advantage of color is the simple algorithms for segmentation which are well known from multispectral analysis in the field of remote sensing. In some cases even a simple color transformation like the HSI, HSV, or CIE space, with a successive threshold suffices. But besides the pixel classification region oriented post processing must be used to combine groups of pixels to areas. Morphological operators, like dilation, closing, or binary rank, are very useful in this context. At the left of figure 3 left an example for a pixel classification can be seen. At the right the modified regions after filling of small holes, applying an opening operation

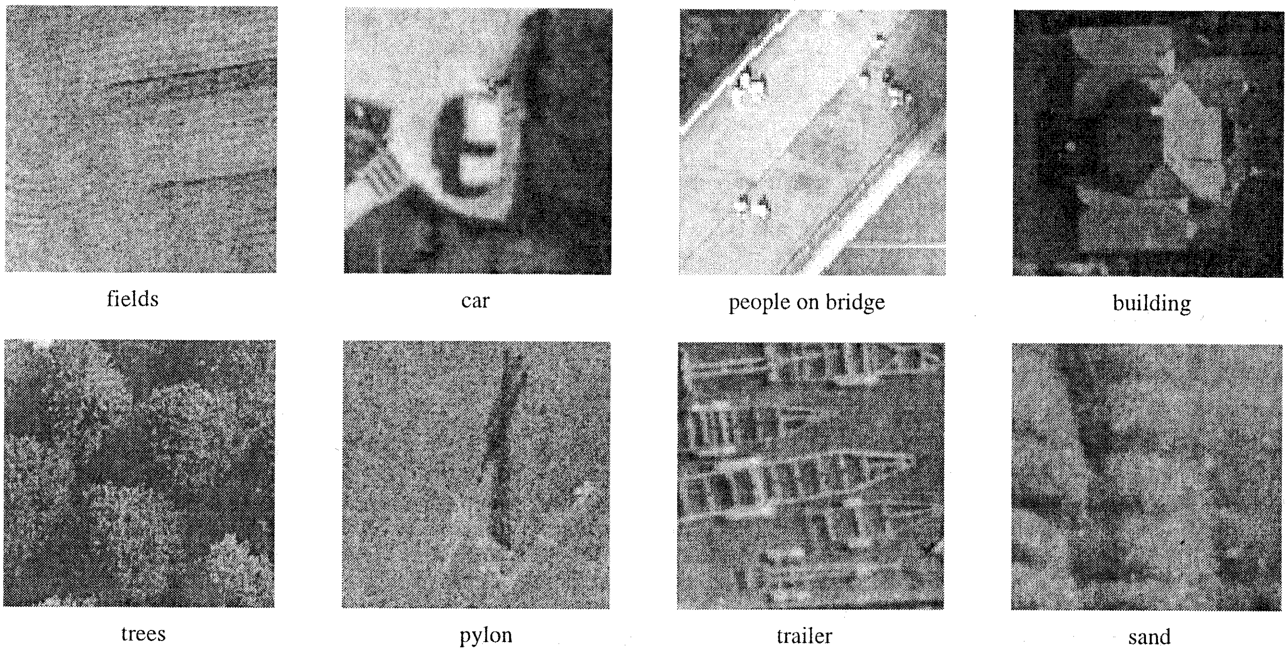


Figure 1: Examples for different object classes in aerial images

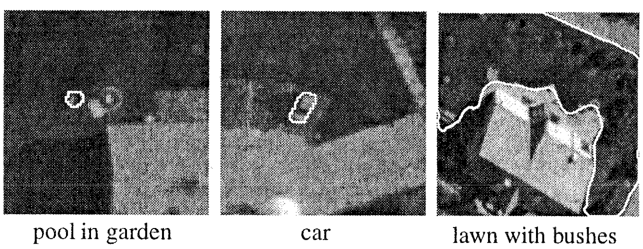


Figure 2: Objects that can be detected easily using color

with a rectangular mask (diameter 5 pixel), and selection of large regions is shown.

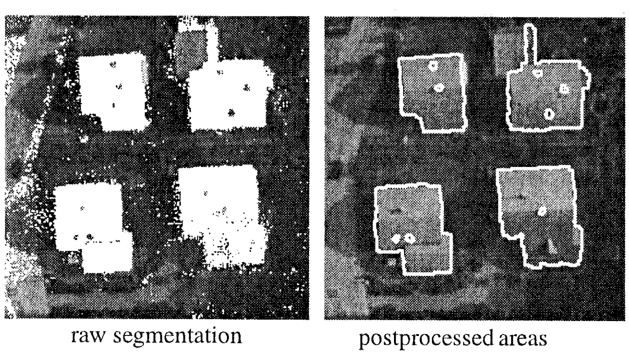


Figure 3: Transformation and selection of regions after pixel classification

One disadvantage using color is the problem of calibration because most images are digitized from pictures. In this case the color features of every object class has to be trained for every film and every scanner.

2.2 Multi View

The extraction of primitives like edges is often incomplete because the objects are partly occluded by other objects or due to unfavorable illumination. This problem can be reduced by using more than one view of the objects (Roux and McKeown Jr., 1994). In the case of aerial images stereo pairs are often available. In figure 4, for example, a building is shown in two different views. The edges in both images are incomplete. But the combination of both segmentations yields a better interpretation with additional information of the 3D structure of the building (Haala, 1994).

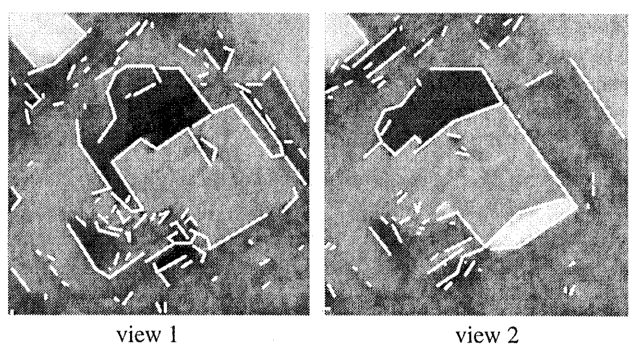
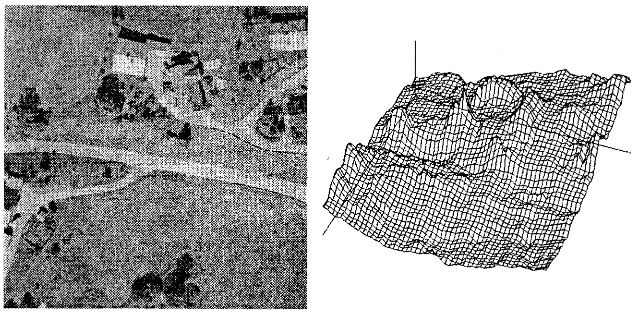


Figure 4: Edges extracted from two different views

2.3 Digital Elevation Model

One completely different type of input data is a digital elevation model (DEM). It can be generated using manual or automatic matching of stereo images or by sensors like a laser scanner. A DEM is useful for the extraction of objects which are higher than their surroundings (e.g., buildings or trees). A popular operator for the extraction of high objects is the gray opening. In case of noisy data the dual rank, which can be seen as an extension of the gray opening, gives better results (Eckstein and Munkelt, 1995). The dual rank consists of two successive rank operators. The first rank operator is applied with the given rank value while the second one uses the "dual" value (i.e., maximum - rank value). Therefore the rank value 1 results in a gray opening, and the value n (maximum) corresponds to a gray closing. In the case of $n/2$ we get two successive median filters. The rank value thus controls the behaviour of the operators. For the extraction of high objects a rank value a few percents above 1 is used. The increase of the rank value should be chosen proportional to the amount of noise pixels.

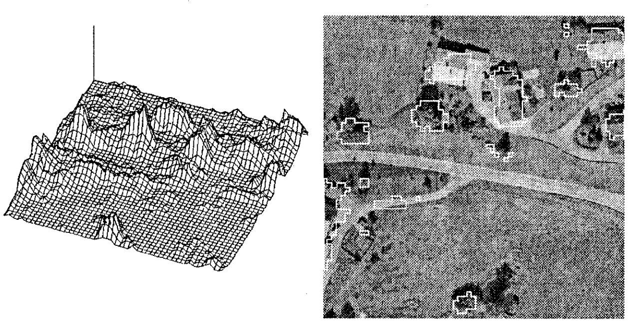
As an example for the segmentation of a DEM figure 5 shows a gray image of a hilly landscape with buildings and trees and the corresponding DEM.



gray image digital elevation model

Figure 5: Hilly landscape with the corresponding DEM

At first the dual rank is applied to the DEM. The result corresponds to the ground without objects (ground DEM). Subtracting this ground DEM from the original DEM yields the result in figure 6, which is similar to the top hat of a gray opening. Only those objects remain which are higher than their surroundings. The extraction of these objects is now simply a threshold operation, where the parameter is chosen according to the desired height. The right image shows the results of the threshold operation. The detection of the road (left below the intersection) is a bit surprising, but the road is actually higher than the surrounding meadows. Another problem in this example are the trees, which are too small to be significant for the resolution of the DEM.



top hat of dual rank extracted objects

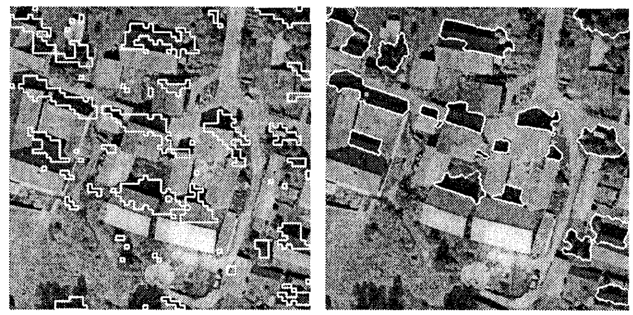
Figure 6: Normalized DEM and the extracted high objects

Shadows are another class of "objects" that can be extracted using the DEM. This is useful, because shadows cause a lot of problems during the interpretation of images since they change the gray values of objects drastically and add egdes or texture-like structures. In the image of figure 7, for example, the road is divided into light and dark areas. The extraction of shadow pixels cannot simply be done by selecting all dark pixels because other dark objects may be present. Instead the illumination of the sun is simulated using the DEM (figure 7 center and right). The segmentation of this image gives the raw shadows.

Due to the low resolution of the DEM the segmentation has to be improved in the gray image using the following steps:

1. Elimination of small areas.
2. All those pixels of the remaining areas are selected which have the intensity of shadows (i.e., their gray values are inside a given range).
3. These pixels are used as seed areas for regiongrowing: Border pixels are added as long as the difference between their gray values and the mean value of the area is below a given threshold. In addition the number of iterations is limited according to the maximal error of the DEM.

The result of this post-processing can be seen at the right of figure 8. These areas can be used to support the interpretation (Lin et al., 1994), for example, to extend small areas of road hypotheses.



raw segmentation shadows

Figure 8: Segmentation of the illuminated DEM and refinement using the gray image

3 NOISE CLEANING

In many cases some kind of preprocessing has be applied to the images before segmentation. One reason is the grain when using maximum resolution of the film, another is the elimination of texture which complicates the segmentation. Different theories and algorithms have been developed to solve these problems. Some important classes are:

Lowpass Filter: The assumption of these filters is that noise has a high frequency. The elimination of noise is therefore done by suppression of high frequencies. Popular representatives of this class are the average and Gauss filter. These filters are very fast but the noise suppression is poor, especially using the average filter, and important image structures like edges are blured.

Rank Operator: These nonlinear operators take linear combinations of the sorted values of all the neighborhood pixels. Conceptually we can visualize the operator as sorting the gray values from the smallest to the largest and taking a linear combination of these sorted values. The most common rank operator for noise cleaning is the median. The median suppresses small lines and points but edges are preserved. Other types of rank operators have been proposed, like variable size and shape of the neighborhood depending on the noise, or the weighted-median filter which adds gray values more than once (depending on the weight) to the sorting list. Further information can be found in (Haralick and Shapiro, 1992).

Wiener Filter: Like inverse and pseudoinverse filters the Wiener filter is used in the field of image restoration. The Wiener filter, which models noise explicitly, makes use of the following model assumptions:

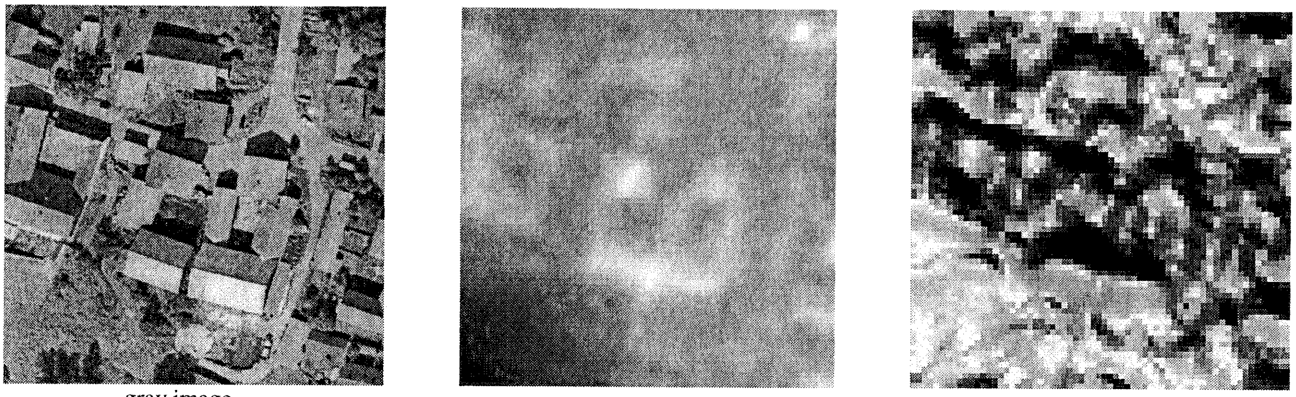


Figure 7: Gray image with corresponding illuminated DEM

- The image distortions can be modelled by a linear system.
- The original image and the observed image can be considered as stationary random.
- There is no correlation between noise and image.
- The noise in the amplitude only (not in the phase) of the Fourier spectrum.

The implementation usually is done via convolution in the Fourier domain. To build the impulse response of the Wiener filter, the spatially invariant impulse response of the image distortion, and the power spectral density of the original image and the noise are needed. Wiener filtering of an observed image produces an estimation of the undistorted original image that is optimal in the sense of a minimal mean square error between the estimated and original image.

Others: Hysteresis smoothing can remove minor fluctuations while preserving the structure of all major transients. The anisotrope diffusion is an iterative, anisotrope smoothing operation on the basis of physical diffusion. Low edges are suppressed while step slopes remain. The sigma filter is an average filter with gray values in the neighborhood that are close in value to the center value. Thus we have a simple local adaption to noise or texture.

In general it is very difficult to decide which of the operators mentioned above has to be selected for a given task because no theory exists which allows a comparison. To have a short impression of the effects one example is given. In figure 9 a roof with noise due the grain of the film and some texture and the elimination using a Gauss filter and the anisotrope diffusion (left to right) can be seen. In this case it is obvious that the result of the anisotrope diffusion is better, because edges are preserved and noise is eliminated.

4 OPTIMAL RESOLUTION

For the extraction of a class of objects one has to select the resolution depending on its shape and radiometric properties. If the resolution is too high the details complicate the segmentation and interpretation. The advantage of a lower resolution is the reduced number of pixels (runtime) and the reduced size of the objects (locality and generalization). If the resolution is too low the objects cannot be extracted at all. In german there is a proverb which explaines the problem: "You can't see the forest due to many trees." If you want to extract a forest there is no need to see the leaves. The selection of the optimal resolution simplifies the problem significantly. Roads, for example, can be detected as lines in a lower resolution (Fischler et al., 1981), (Aviad and Carnine Jr, 1992), (Barzohar and Cooper, 1993), (Barzohar and Cooper, 1995), (Berthod and Serendero, 1988). In figure 10 a house in high resolution (one pixel corresponds to 25 cm) is shown. Applying a Sobel filter yields many edges due to the texture of the tiles. Using a lower resolution (one pixel corresponds to ≈ 0.75 m) results in the edges of the roof.

Looking at the selection of the correct resolution more closely, one finds out that a single resolution does not suffice in many cases: A higher resolution is needed to refine the segmentation.

1. There is an optimal resolution to extract the raw shape of an object called *initial resolution.*
2. In many cases a higher resolution (*refined resolution*) is needed to extract the exact shape of the object and to distinguish it from other similar looking objects.

This leeds to a multi resolution approach to segmentation: Segmentation starts with the initial resolution. The results of this step are verified and improved using the next refined resolution. If necessary, the process is continued with further refined resolutions (Heipke et al., 1995). In figure 11 the extraction of a road in the initial resolution can be seen. Here roads are extracted as lines of a given width. In the refined resolution edges are extracted. These results are used for a precise detection of the road boundaries and the elimination of false candidates. For final results a further refinement is needed to extract road marks.

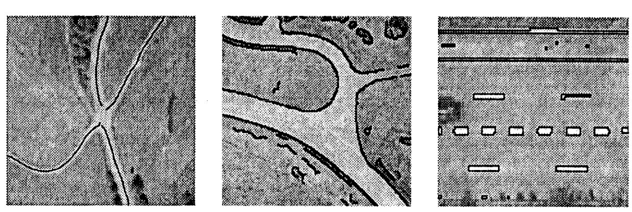
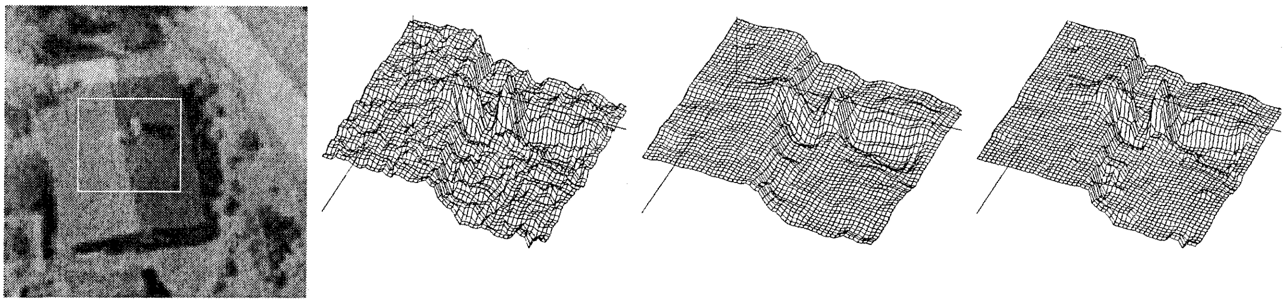


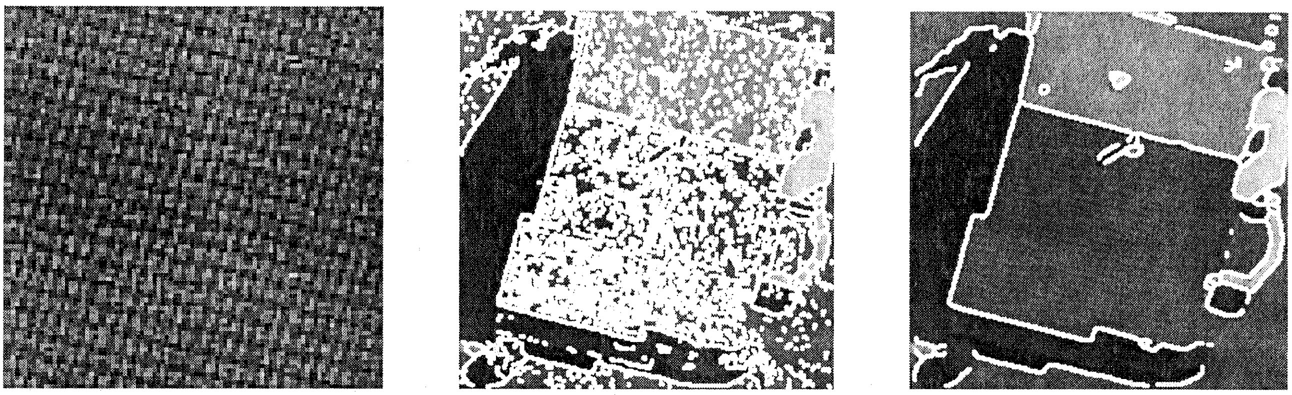
Figure 11: Extraction of primitives of roads in different resolutions (three different scenes)

The resolution hierarchy might be invalid in some case. This can be illustrated by the roads in figure 12. The gray value of the asphalt is simular to its surrounding. Therefore the road cannot be extracted as a line in low resolution. The road is defined only by the marks found in the refined resolution. In this case the



building | original image | Gauss | anisotrope diffusion

Figure 9: Part of an image with noise and some texture and the results of different filters for noise reduction



texture of tiles on the roof | high resolution | low resolution

Figure 10: Edge detection in high and low resolution

interpretation process is very difficult because there is a very low hypotheses for a road because it could not be found in the initial resolution.

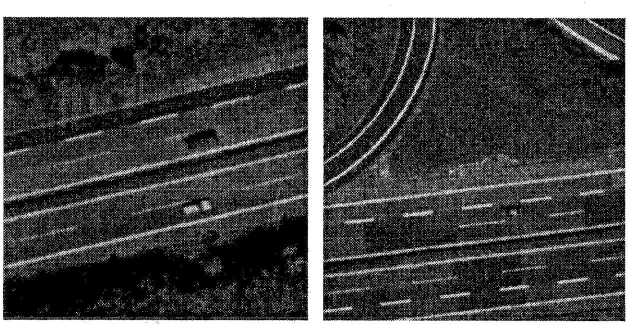


Figure 12: Roads that can be extracted only in high resolution

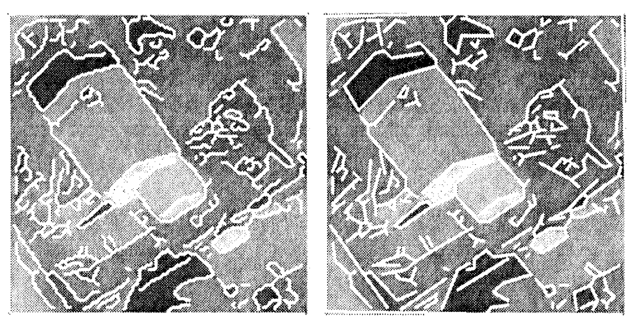
More information on scale space and pyramids can be found in (Gauch and Pizer, 1993), (Lindeberg, 1991), (Lindeberg, 1993).

5 OBJECT CLASSES

As we have seen in figure 1 there are a lot of different object classes in an aerial image. Ideally, one segmentation procedure should be used to extract primitives which are sufficient to recognize objects of all classes. Unfortunately, this is not the case. As we will see in the following subsection there exist specific procedures for a broader class of objects which can be processed in a similar way. But no procedure for all classes exists.

5.1 Compact Artifical Objects

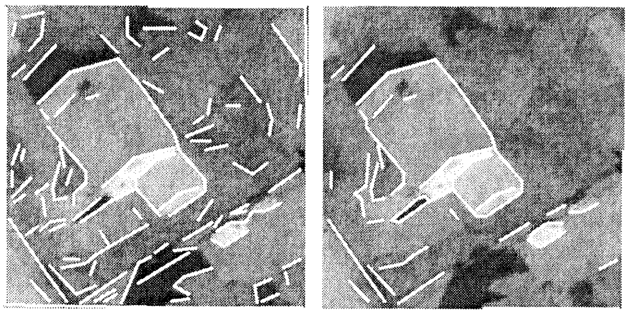
Examples for this class are buildings, cars, trucks, and ships. All these objects are composed of more or less homogeneous areas with polyhydral borders. Therefore, models can be contructed using descriptions of areas, lines, and points together with attributes (e.g., color or size) and relations between the primitives. The interpretation of objects is done by extracting similar primitives (area, edges, junctions) and matching these with the model after an optional grouping (Dolan and Weiss, 1989), (Lin et al., 1994), (Lu and Aggarwal, 1992), (Mohan and Nevatia, 1987), (Mohan and Nevatia, 1992), (Sarkar and Boyer, 1993). In figure 13 a building with extracted edges and an approximation of the contours can be seen. To ease the interpretation the lines are grouped



raw contours | polygons

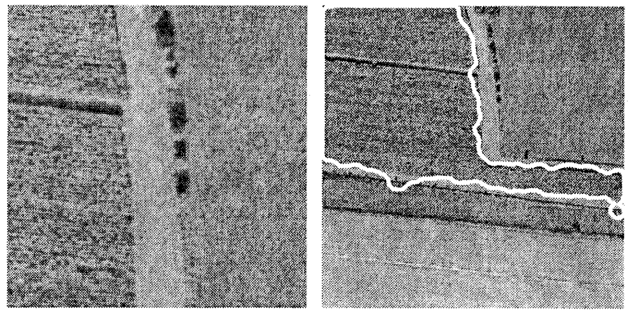
Figure 13: Extraction of edges and polygon approximation

(figure 14): In a first step all parallel lines are selected given a maximal distance and a maximal error for the angle. From these all those pairs are selected which enclose homogenous areas.



parallel lines / selected parallels

Figure 14: Selection of parallel lines enclosing areas with contant gray values



different textures / classification

Figure 16: Separation of areas with different textures

The main structure of this class of objects can be extracted using one resolution. This is choosen in such a manner that areas have homogenous gray values (Förstner, 1995). To distinguish (long) buildings from (short) roads a higher resolution or a DEM is needed.

5.2 Linear objects

Typical representatives are roads, rivers, and railroads (Huertas et al., 1987), (Heipke et al., 1995), (Jedynak and Rozé, 1995), (Heipke et al., 1994), (Ilg, 1990), (Li et al., 1992), (Lipari et al., 1989), (McKeown Jr. and Denlinger, 1988), (Venkateswar and Chellappa, 1992), (Vosselman and de Knecht, 1995), (Zerubia and Merlet, 1993). Roads are similar to the objects above, but the borders are curves and the size (length) is not limited. In addition, lines (in contrast to edges) are needed for the interpretation. As we have seen in section 4 roads are extracted using different levels of resolution. The initial resolution is choosen in such a way that roads have a width of a few pixels. In the highest resolution road marks have a similar width. Because there are different types of roads with typical widths the initial resolution has be to selected appropriately.

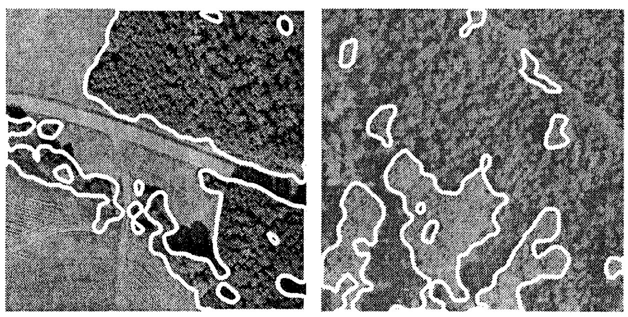
The type of model for roads is different to that for buildings. This is because roads are unbound in principle and do not have a fixed shape. Therefore, the interpretation process is mainly bottom up and the model is used for tasks like grouping and selection of areas which are road candidates and not for matching of primitves with a model of a road. An example of this grouping process can be seen in figure 15. The first picture shows the initial primitives which are grouped (parallel, colinear), selected (homogenity) and combined with the results of the initial resolution.

Linear objects which are more complicated are brooks or railroads: Brooks and rivers often have badly defined banks. Railroads are defined by a combination of lines and a specific texture. For these objects strategies used for roads and arbitrary areas have to be combined.

5.3 Arbitrary areas

Areas like meadows, forests, or fields define another class. They have an arbitrary border and are defined by their specific gray value, color, and texture. In this case it is very difficult to extract any type of primitives. Direct texture analysis is used instead (see section 6.3). A first example can be seen in figure 16. The left picture shows a zoomed part with two fields with different texture. These textures can be distinguished very easily due to the horizontal structure of the left field.

In figure 17 two examples with forest can be seen. The left picture shows the selection of areas with the texture of firs. Looking at the road at the right side the invariance with respect to illumination can be seen. In the right picture deciduous trees are selected.



forest / deciduous trees

Figure 17: Selection of regions with with different type of textures

A model for this class of objects is mainly implicit (e.g., frequency or color distribution). Only some features, like minimal size, width, or typical shapes (fields), can be used.

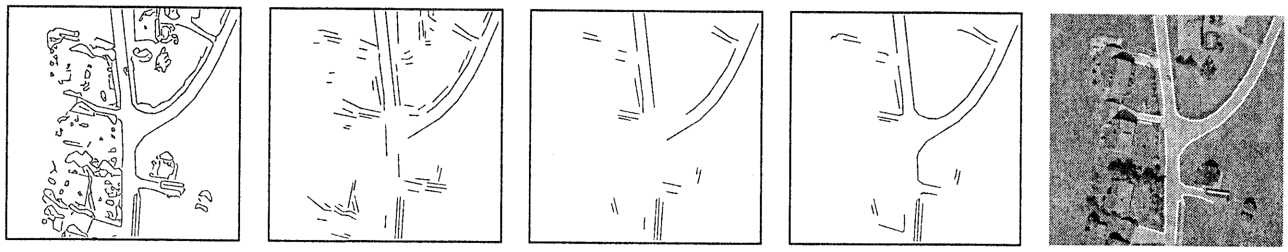
5.4 Special Objects

Objects like trees or persons have to be treated in a very specific manner. A tree, for example, has a complex structure which involves texture as well as shape. In the case of deciduous trees it is very difficult to extract their boundary, especially if they stand close together. As we saw in figure 17 it is possible to detect the texture of deciduous trees. Figure 18 shows two extreme examples: The single trees in the left image can be extracted using the gray values and shape. In addition, 3D information, like shadows or height, is useful. But there is no procedure known to separate the trees in the right image.

A more simple class of trees are firs. In figure 19 one tree and parts of the neighboring trees can be seen. Looking at the 3D plot of the gray values the complex structure of the branches can be seen. One approach to segment these trees is base on the following assumptions (Haenel and Eckstein, 1986):

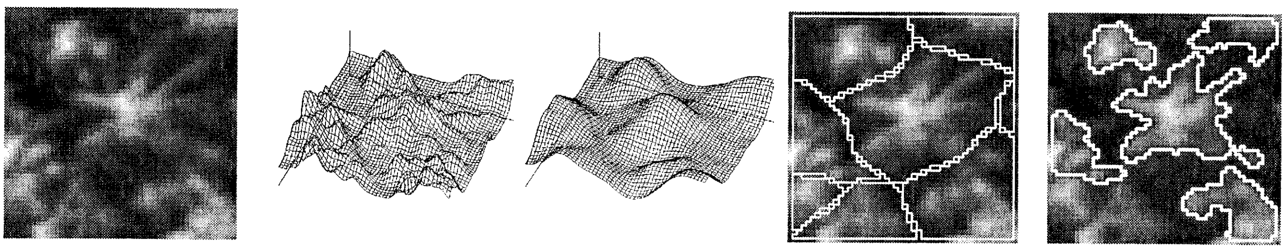
- The top of the tree is brighter than the outer parts.
- All trees a separated by dark areas (shadows).
- The visible (bright) part of the tree has some texture.

Smoothing the image with a Gauss filter the influence of the texture is eliminated and trees can be interpreted as distinct bright blobs. The blobs are segmented using the watershed algorithm (inverse gray values). The boundaries are now along the darkest part of the shadows. To extract the visible part of the trees a local threshold operation is used.



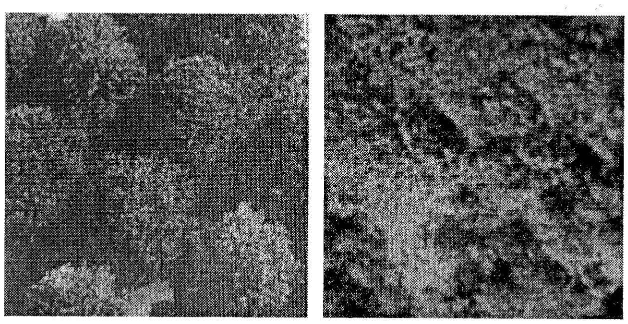
polygons parallel lines homogenous areas extended lines final result

Figure 15: Grouping process for road extraction



bot spruce 3D-plot smoothed image watersheds local threshold

Figure 19: Segmentation of fir trees by calculation of watersheds in a smoothed image



single trees dense forest

Figure 18: Different forms of deciduous trees

6 BASIC SEGMENTATION ALGORITHMS

A variety of algorithms are available as building blocks for the construction of a complete segmentation procedure. Depending on the object class to be extracted, one or more of them is needed.

6.1 Pixel Classification

This is the well known class of point operations, mainly applied to multichannel images in the field of remote sensing. In the case of aerial images the algorithms can be used with color or infrared. As we will see in section 6.3 synthetic channels can the generated with the help of texture filters. The problem of pixel classification is the lack of context. Some kind of context can be added by using resolution pyramids as additional channels or by post processing the classified pixels (closing, dilation, etc.).

6.2 Primitives

The most polular approach for the extracton of objects is to find edges. Different operators have been proposed. Besides simple filters like Sobel, Kirsch, or Prewitt more sophisticated have been developed: (Shen and Castan, 1992), (Canny, 1983), (Canny, 1986), (Lanser and Eckstein, 1992). Edge detection assumes that an object consists of one or more constant areas. In general it is not possible to extract closed contours, so the edges are used as primitives. In many cases they are approximated to polygons. Junctions and points of high curvature are derived as additional primitives. Using an edge detector, blobs, defined as areas of low gradient, can be extacted by a threshold operation with optional postprocessing e.g. opening.

gray image blobs polygons

Figure 20: Two types of primitives: blobs and polygons

A more general approach is given by (Förstner, 1994). Here different types of primitives are extracted simultaneously:

- Homogeneous areas
- Edges and lines
- Points (boundary points of high curvature or junctions)

These features are well defined in the case of artifical objects like buildings (see section 5.1). The approach is based on the average squared gradient defined by

$$\Gamma g = G_\sigma * \begin{pmatrix} g_x^2 & g_x g_y \\ g_y g_x & g_y^2 \end{pmatrix} \qquad (1)$$

where G_σ is symmetric Gaussian function with standard deviation σ. The extraction of primitives is composed of the following steps:

1. Estimation of noise characteristics.
2. Information preserving restoration using a Wiener filter (see section 3).

3. Feature detection (blobs, lines, points) and feature descrimination (junctions, symmetric features, edges, and lines)
4. Feature localization with subpixel accuracy, where features are expected to lie inside the classified pixels.

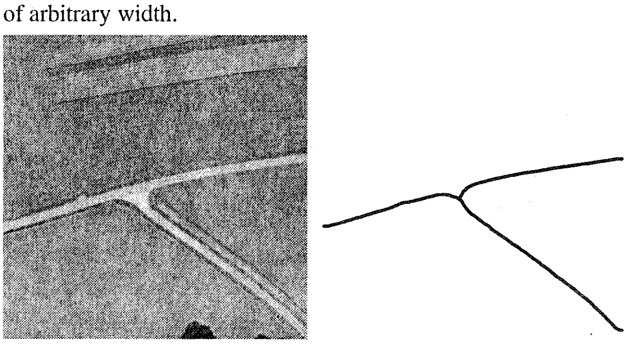
In (Steger, 1996) lines are extracted as primitives. The image is regarded as a function $g(x, y)$ and lines are detected as ridges and ravines in this function by locally approximating the image function by its second order Taylor polynomial (not the facet model like in (Busch, 1994)). The coefficients of the Taylor polynomial are determined by convolving the image with the derivatives of a Gaussian smoothing kernel. In contrast to (Förstner, 1994) the Hessian matrix

$$H = \begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix} \tag{2}$$

is used to extract the local features.

Curvilinear structures in 2D are modeled as curves $s(t)$ that exhibit a characteristic 1D line profile in the direction perpendicular to the line, i.e., perpendicular to $s'(t)$. Let this direction be $n(t)$. This means that the first directional derivative in the direction $n(t)$ should vanish and the second directional derivative should be of large absolute value. To compute the direction of the line locally for each image point the partial derivatives g_x, g_y, g_{xx}, g_{xy}, and g_{yy} of the image are estimated. This is done by convolving the image with the appropriate 2D Gaussian kernels. The direction in which the second directional derivative of $g(x, y)$ takes on its maximum absolute value is used as the direction $n(t)$. This direction is determined by calculating the eigenvalues and eigenvectors of the Hessian matrix.

The use of the Taylor polynomial leads to a single response of the filter to each line. Furthermore, the line position are determined with sub-pixel accuracy and the algorithm scales to lines of arbitrary width.



original image extracted lines

Figure 21: Extraction of lines using the appraoch of Steger

Other articles on the extraction of lines are: (Blaszka and Deriche, 1994a), (Gruen and Agouris, 1994), (Koller et al., 1994), (Koller et al., 1995), (Monga et al., 1995). Further articles on the extraction of image primitives: (Reynolds and Beveridge, 1987), (Blaszka and Deriche, 1994b), (Filbois and Gemmerlé, 1994).

6.3 Texture

A great variety of operators for texture segmentation have been developed. The first class analyses the local frequency distribution based on the idea that every texture has a specific spectrum. The next class extracts local features (texture elements) by which the global texture can be defined. Other approaches use stochastic models for segmentation (Geman and Geman, 1984), (Kato et al., 1991), (Nguyen and Cohen, 1993). Finally, local features like the co occurrence matrix are used to describe textures (Lohmann, 1994).

The texture analysis using spectral decomposition will be explained in more detail. The idea is very simple: a set of n filters has to be defined which extracts the amount of a specific frequency range for the neighborhood of every pixel. Thus we have a vector $t(g)$ of length n desribing the pixel g and its neighborhood with n values. This vector can be used as input for n-dimensional (un)-supervised classification (see section 6.1).

For the implementation of this approach appropriate filters have to be found. A simple set of filters was proposed by (Laws, 1980). He defined 25 filters f_{ij} of size 5×5 constructed from vectors $v \in \{l, e, s, r, w\}$ by convolution: $f_{ij} = v_i^T * v_j$.

$$\begin{array}{lcl} l & = & (\;\;1 \;\;\; 4 \;\;\; 6 \;\;\; 4 \;\;\; 1) \\ e & = & (-1 \;\; -2 \;\;\; 0 \;\;\; 2 \;\;\; 1) \\ s & = & (-1 \;\;\; 0 \;\;\; 2 \;\;\; 0 \;\; -1) \\ r & = & (\;\;1 \;\; -4 \;\;\; 6 \;\; -4 \;\;\; 1) \\ w & = & (-1 \;\;\; 2 \;\;\; 0 \;\; -2 \;\;\; 1) \end{array}$$

Another popular set are the Gabor filters (Shao and Förstner, 1994). They are defined in frequency space and have some nice features: They have orientation selectivity, multiscale property, linear phase and good localization both in spatial and frequency domains.

As a last example for texture filters simple gauss shaped filters can be used. They are invariant with respect to rotation and are defined via center frequency and the deviation. Typical filters of all three classes can be found in figure 22.

An example for the application of two laws filters is given in figure 23. At first the filters f_{ee} and f_{rl} are calculated from the gray image. The so called texture energy is calculated using a lowpass filter (e.g., average) with a large filtermask to generalize the texture. In this case a median filter with circular mask (diameter 50 pixel) was used. These texture energy images can be used as input to pixel classification.

6.4 Specialized Operations

Besides more general segmentation procedures like those of section 6.2 and 6.3 are the specialized filters which emphasize special structures in a gray image like points, lines, or corners.

The corner resonce operator, for example, is defined by (Harris and Stephans, 1988):

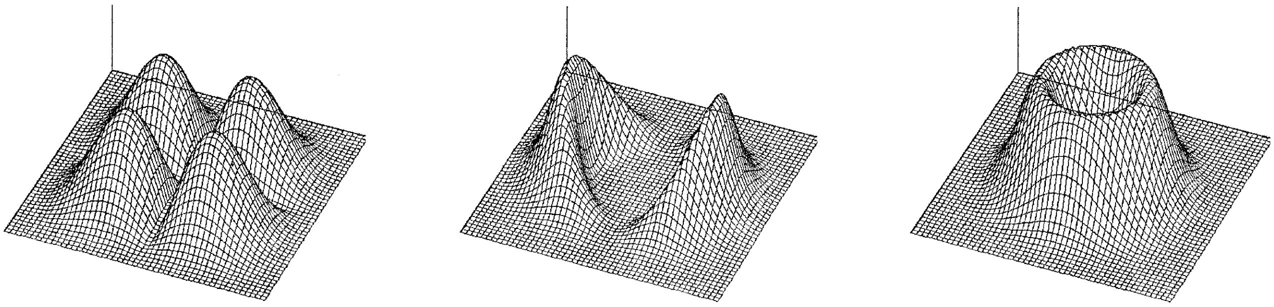
$$\begin{aligned} g^c &= G_\sigma * g_x^2 \cdot G_\sigma * g_y^2 - G_\sigma * (g_x g_y) - \\ &\quad k \left(G_\sigma * g_x^2 + G_\sigma * g_y^2\right)^2 \end{aligned} \tag{3}$$

where g is the gray value and G_σ ist the Gaussian filter with deviation σ. The corner response function is invariant with respect to rotation. A typical value for the factor k is 0.04. In this case corners result in a positive g^c while edges have negative values. An extension of the corner response function is given in formula (4).

$$\begin{aligned} g^c &= G_\sigma * (g_x^a)^2 \cdot G_\sigma * (g_y^a)^2 - G_\sigma * \left(g_x^a g_y^a\right) - \\ &\quad k \left(G_\sigma * (g_x^a)^2 + G_\sigma * (g_y^a)^2\right)^2 \\ g^a &= \sqrt{g_x^2 + g_y^2} \end{aligned} \tag{4}$$

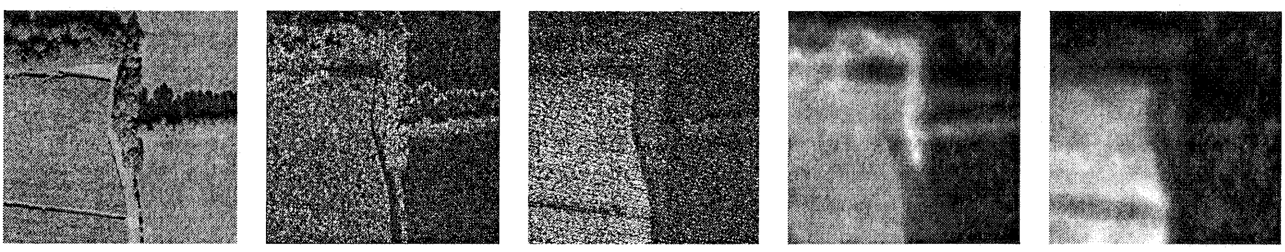
Here the filter is applied to the gradient and not to the original gray values. The response is maximal for highly curved edges. In figure 24 two examples are given. All maximums of the filter above a given threshold are marked with a cross. Most of the dominant points of the buildings are found as well as corners caused by the shadows.

Besides the corner response filter a lot of other filters for corners or "promiment" points have been defined. Some of these can



Laws Gabor Gauss

Figure 22: Frequency response of different texture filter classes



original image filter ee filter rl median of ee median of rl

Figure 23: Result of two different Laws filter

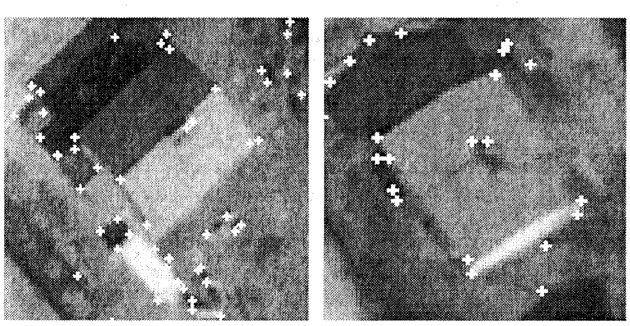


image 1 image 2

Figure 24: Result of extended corner response filter

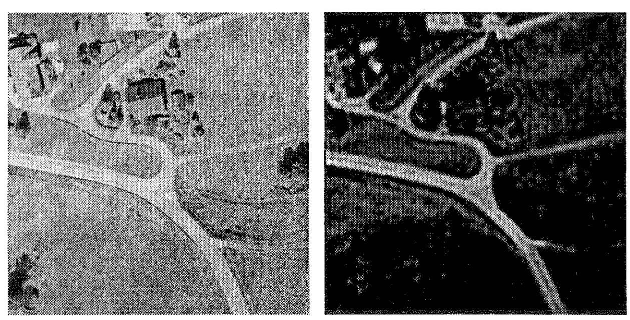
be found in: (Dreschler, 1981), (Deriche and Giraudon, 1993), (Tabbone, 1994), (Förstner and Gülch, 1987). In (Rohr, 1993) a comparison of different operators is given.

As a second example for specialized operators the extraction of lines will be discussed. The aim is to construct a filter, whose emphasizes linear structures of a given width and suppresses structures beeing smaller or wider. We define the frequency f of a line as the frequency of the sinus wave which half period is equal to the the width of the line. Given the minimum and maximum width of the line and thus f_{max} and f_{min} we define a bandpass filter with the following properties:

1. Suppression of frequencies below f_{min}.

2. Suppression of frequencies above f_{max}.

The frequency f_{max} can be choosen higher if sharper edges of the lines are required. In figure 25 an example of such a filter for the extraction of roads can be seen. f_{max} has been choosen higher, so the edges of the road are fairly well defined. The extration of the road is simply a threshold operation. Besides the roads some roofs of similar width are selected. These have to be eliminated using context information.



original image bandpass filter

Figure 25: Emphasizing linear structures of a given with

7 CONCLUSION

Is was shown that the segmentation of aerial images needs task oriented segmentation procedures. Depending on the class of objects, appropriate resolutions and procedures have to be selected. This selection has to be done according to the object model. The segmentation becomes more stable if additional data, like color or a DEM, i, iSS used. One open question is how to merge segmentation results when processing different object classes simultaneously.

All examples of this paper were programmed with the image analysis system $\mathcal{HORUS}$ using the interactive user interface $\mathcal{HORUS}$Develop (Eckstein and Steger, 1996). The image examples have mainly been taken from the ETH-Zürich and the ISPRS testset (Fritsch et al., 1994). All images are available via ftp from: `ftp://ftp.informatik.tu-muenchen.de/pub/rec/images/space/`

REFERENCES

Aviad, Z. and Carnine Jr, P. D., 1992. Road finding for road-network extraction. In: Computer Vision and Pattern Recognition, pp. 814–819.

Barzohar, M. and Cooper, D. B., 1993. Automatic finding of main roads in aerial images by using geometric - stochastic models and estimation. In: Computer Vision and Pattern Recognition, pp. 459–464.

Barzohar, M. and Cooper, D. B., 1995. New geometric stocastic technology for finding and recognizing roads and their features in aerial images. In: A. Gruen, O. Kuebler and P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, pp. 255–264.

Berthod, M. and Serendero, M. A., 1988. Extraction of thin networks on satellite imagery. In: 9th International Conference on Pattern Recognition, pp. 456–458.

Blaszka, T. and Deriche, R., 1994a. A model based method for characterization and location of curved image features. Rapport de Recherche 2451, INRIA, Sophia Antipolis, France.

Blaszka, T. and Deriche, R., 1994b. Recovering and characterizing image features using an efficient model based approach. Rapport de Recherche 2422, INRIA, Sophia Antipolis, France.

Busch, A., 1994. Fast recognition of lines in digital images without user-supplied parameters. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXX, Part 3/1, pp. 91–97.

Canny, J., 1983. Finding edges and lines in images. Technical Report 720, MIT Artificial Intelligence Laboratory, Cambridge, MA, USA.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6), pp. 679–698.

Deriche, R. and Giraudon, G., 1993. A computational approach for corner and vertex detection. In: Computer Vision and Pattern Recognition, pp. 101–124.

Dolan, J. and Weiss, R., 1989. Perceptual grouping of curved lines. In: Image Understanding Workshop, Morgan Kaufmann Publishers, pp. 1135–1145.

Dreschler, L., 1981. Ermittlung markanter Punkte auf Bildern bewegter Objekte und Berechnung einer 3D-Beschreibung auf dieser Grundlage. Dissertation, Fachbereich Informatik, Universiät Hamburg, Hamburg, Germany.

Eckstein, W. and Munkelt, O., 1995. Extracting objects from digital terrain models. In: T. Schenck (ed.), Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes, Proc. SPIE 2572, pp. 43–51.

Eckstein, W. and Steger, C., 1996. Interactive data inspection and program development for computer vision. In: Visual data Exploration and Analysis IV, Proc. SPIE 2656, Vol. 2656.

Filbois, A. and Gemmerlé, D., 1994. From step edge to line edge: Combining geometric and photometric information. In: MVA '94 IAPR Workshop on Machine Vision Applications, pp. 87–90.

Fischler, M. A., Tenenbaum, J. M. and Wolf, H. C., 1981. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. Computer Graphics and Image Processing 15, pp. 201–223.

Ford, S. J. and McKeown Jr., D. M., 1992. Utilization of multispectral imagery for cartographic feature extraction. In: Image Understanding Workshop, Morgan Kaufmann Publishers, pp. 805–820.

Förstner, W., 1994. A framework for low level feature extraction. In: J.-O. Eklundh (ed.), Third European Conference on Computer Vision, Lecture Notes in Computer Science, Vol. 801, Springer-Verlag, pp. 383–394.

Förstner, W., 1995. Mid-level vision processes for automatic building extraction. In: A. Gruen, O. Kuebler and P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, pp. 179–188.

Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise lacation of distinct points, corners, and centres of circular features. In: Proceedings of the Intercomission Conference on Fast Processing of Photogrammetric Data, Informatik aktuell, International Society for Photogrammetry and Remote Sensing, Springer-Verlag, pp. 281–305.

Fritsch, D., Sester, M. and Schenck, T., 1994. Test on image understanding. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXX, Part 3/1, pp. 243–248.

Gauch, J. M. and Pizer, S. M., 1993. Multiresolution analysis of ridges and valleys in grey-scale images. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(6), pp. 635–646.

Geman, S. and Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(6), pp. 721–741.

Gruen, A. and Agouris, P., 1994. Linear feature extraction by least squares template matching constrained by internal shape forces. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXX, Part 3/1, pp. 316–323.

Haala, N., 1994. Detection of building by fusion of range and image data. In: K. E. H. Ebner, C. Heipke (ed.), International Archives of Photogrammetry and Remote Sensing, Vol. XXX, Part 3/1, pp. 341–346.

Haenel, S. and Eckstein, W., 1986. Ein Arbeitplatz zur Halbautomatischen Luftbildauswertung. In: Hartmann (ed.), Proceedings of the 8th DAGM Symposium, IFB, Vol. 125, Deutsche Arbeitsgemeinschaft für Mustererkennung, Springer-Verlag.

Haralick, R. M. and Shapiro, L. G., 1992. Computer and Robot Vision. Vol. I, Addison-Wesley Publishing Company.

Harris, C. and Stephans, M., 1988. A combined corner and edges detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–152.

Heipke, C., Englisch, A., Speer, T., Stier, S. and Kutka, R., 1994. Semi-automatic extraction of roads from aerial images. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXX, Part 3/1, pp. 353–360.

Heipke, C., Steger, C. and Multhammer, R., 1995. A hierarchical approach to automatic road extraction from aerial imagery. In: D. M. McKeown Jr. and I. J. Dowman (eds), Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, Proc. SPIE 2486, pp. 222–231.

Huertas, A., Cole, W. and Nevatia, R., 1987. Detecting runways in aerial images. In: Image Understanding Workshop, Morgan Kaufmann Publishers, pp. 272–297.

Ilg, M., 1990. Knowledge-based understanding of road maps and other line images. In: 10th International Conference on Pattern Recognition, pp. 282–289.

Jedynak, B. and Rozé, J.-P., 1995. Tracking roads in satellite images by playing twenty questions. In: A. Gruen, O. Kuebler and P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, pp. 243–253.

Kato, Z., Zerubia, J. and Berthod, M., 1991. Image classification using markov random fields with two relaxation methods: Deterministic pseudo annealing and modified metropolis dynamics. Rapport de Recherche 1606, INRIA, Sophia Antipolis, France.

Koller, T. M., Gerig, G., Székely, G. and Dettwiler, D., 1994. Multi-scale detection of curvilinear structures in 2-D and 3-D image data. Technical Report BIWI-TR-153, Communication Technology Laboratory, Image Science, ETH Zürich, Zurich, Switzerland.

Koller, T. M., Gerig, G., Székely, G. and Dettwiler, D., 1995. Multiscale detection of curvilinear structures in 2-D and 3-D image data. In: Fifth International Conference on Computer Vision, pp. 864–869.

Lanser, S. and Eckstein, W., 1992. A modification of deriche's approache to edge detection. In: 11th International Conference on Pattern Recognition, pp. 633–637.

Laws, K., 1980. Texture image segmentation. Ph.d. dissertation, Dept. of Engineering, University of Southern California.

Li, S. Z., Kittler, J. and Petrou, M., 1992. Matching and recognition of road networks from aerial images. In: G. Sandini (ed.), Second European Conference on Computer Vision, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 857–861.

Lin, C., Huertas, A. and Neviata, R., 1994. Detection of buildings using perceptual grouping and shadows. In: Computer Vision and Pattern Recognition, pp. 62–69.

Lindeberg, T., 1991. Discrete scale-space theory and the scale-space primal sketch. Ph.d. dissertation, Department of Numerical Analysis and Computing Science, KTH (Royal Institute of Technology), Stockholm, Sweden.

Lindeberg, T., 1993. Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction. Journal of Mathematical Imaging and Vision 3(4), pp. 349–376.

Lipari, C., Trivedi, M. and Harlow, C., 1989. Geometric modeling and recognition of elongated regions in aerial images. IEEE Transactions on Systems, Man, and Cybernetics 19(6), pp. 1600–1612.

Lohmann, G., 1994. Co-occurrence-based analysis and synthesis of textures. In: 12th International Conference on Pattern Recognition, pp. 449–453.

Lu, H. Q. and Aggarwal, J. K., 1992. Applying perceptual organization to the detection of man-made objects in non-urban scenes. Pattern Recognition 25(8), pp. 835–853.

McKeown Jr., D. M. and Denlinger, J. L., 1988. Cooperative methods for road tracking in aerial imagery. In: Computer Vision and Pattern Recognition, pp. 662–672.

Mohan, R. and Nevatia, R., 1987. Perceptual grouping with applications to 3D shape extraction. In: Workshop on Computer Vision, pp. 158–163.

Mohan, R. and Nevatia, R., 1992. Perceptual organization for scene segmentation and description. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(6), pp. 616–635.

Monga, O., Armande, N. and Montesinos, P., 1995. Thin nets and crest lines: Application to satellite data and medical images. Rapport de Recherche 2480, INRIA, Rocquencourt, France.

Nguyen, H. H. and Cohen, P., 1993. Gibbs random fields, fuzzy clustering, and and the unsupervized segmentation of textured images. Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing 55(1), pp. 1–19.

Reynolds, G. and Beveridge, J. R., 1987. Searching for geometric structure in images of natural scenes. In: Image Understanding Workshop, Morgan Kaufmann Publishers, pp. 257–271.

Rohr, K., 1993. Lokalisierungseigenschaften direkter Ansätze zur Ermittlung von Grauwertecken. In: S. J. Poeppl (ed.), Mustererkennung, Informatik aktuell, Deutsche Arbeitsgemeinschaft für Mustererkennung, Springer-Verlag, pp. 647–654.

Roux, M. and McKeown Jr., D. M., 1994. Feature matching for building extraction from multiple views. In: Computer Vision and Pattern Recognition, pp. 46–53.

Sarkar, S. and Boyer, K. L., 1993. Integration, inference, and management of spatial information using bayesian networks: Perceptual organization. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(3), pp. 256–274.

Shao, J. and Förstner, W., 1994. Gabor wavelets for texture edge extraction. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXX, Part 3/2, pp. 745–752.

Shen, J. and Castan, S., 1992. An optimal linear operator for step edge detection. Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing 54(2), pp. 112–133.

Steger, C., 1996. Extracting curvilinear structures: A differential geometric approach. In: Fourth European Conference on Computer Vision, Springer-Verlag.

Tabbone, S., 1994. Detection junctions using properties of the laplacian of gaussion detector. In: 12th International Conference on Pattern Recognition, pp. 52–56.

Venkateswar, V. and Chellappa, R., 1992. Extraction of straight lines in aerial images. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(11), pp. 1111–1114.

Vosselman, G. and de Knecht, J., 1995. Road tracing by profile matching and kalman filtering. In: A. Gruen, O. Kuebler and P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, pp. 265–274.

Zerubia, J. and Merlet, N., 1993. Classical mechanics and roads detection in SPOT images. Rapport de Recherche 1889, INRIA, Sophia Antipolis, France.