DETECTABILITY OF BUILDINGS IN AERIAL IMAGES OVER SCALE SPACE

Martin Drauschke, Hanns-Florian Schuster,Wolfgang Förstner

Institute of Photogrammetry, University of Bonn

KEY WORDS: Building Detection, Scale Space, Feature Extraction, Stable Regions

ABSTRACT:

Automatic scene interpretation of aerial images is a major purpose of photogrammetry. Therefore, we want to improve building detection by exploring the "life-time" of stable and relevant image features in scale space. We use watersheds for feature extraction to gain a topologically consistent map. We will show that characteristic features for building detection can be found in all considered scales, so that no optimal scale can be selected for building recognition. Nevertheless, many of these features "live" in a wide scale interval, so that a combination of a small number of scales can be used for automatic building detection.

1 INTRODUCTION

Building detection from aerial images is a very active research area in photogrammetry, cf. the review in (Mayer, 1999). Early attempts go back into the eighties (Nevatia and Price, 1982, Herman and Kanade, 1987, Huertas and Nevatia, 1988). In most cases roof edges or roof parts have been used to identify complex buildings, as facades usually are more difficult to extract. Though some approaches concentrate on simple building types, such as gabled or hipped roof type buildings, they are not generic enough to deal with the great variety of building structures.

Interestingly, all approaches use image features extracted at a single scale, which however, is either given by the resolution of the images, or in some reasonable way selected by a human interpretor. Of course, in general at resolutions of 5 to 30 cm pixel size at ground level building easily can be detected by humans. They obviously exploit the rich context on top of a building and around it. When building an automatic interpretation system, modeling context is one of the most difficult tasks. We want to reduce the demands for modeling context by automatically selecting the optimal scale for image feature extraction.

Technically, detection is inferring existence from observable image features using some classification procedure; localization is only a side effect, precise boundaries are not of primary concern at this stage, cf. the approach of (Brunn and Weidner, 1998) using an image pyramid. Detection may be based on Bayes' rule $P(B|f(R)) \propto P(f(R)|B)P(B)$, where the posterior probability $P(B|f(R))$ for detecting a building or building part given some features $f(R)$ of a region R requires the likelihood $P(f(R)|B)$ and some a priori information about the occurrence of a building. Training a classifier essentially consists of determining parameters p of an adequate likelihood function $l(p) = P(f(R)|B, p)$. This approach assumes the type of features $f(R)$ to be known.

We want to explore the observability of image features, which may be relevant for building detection. Especially, we want to investigate the suitability of regions extracted over scale space. Small changes of scale often do not affect most of the regions however, may lead to extinction of certain regions by merging with neighbored regions, cf. fig. 1.

Such investigations are important for evaluation the mapping potential in the context of human image interpretation, cf. (Jacobsen, 1997) or for evaluating the observability of objects in images in the context of automatic interpretation, e. g. for road mapping procedures (Mayer et al., 1998, Baumgartner et al., 1999, Pakzad and Heller, 2004).

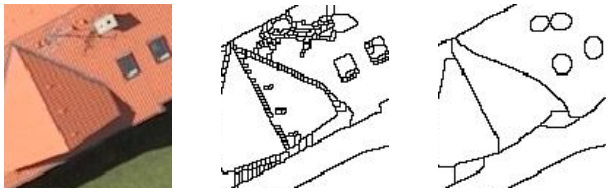

Figure 1: Effect of a scale change onto segmentation. Left: image section of roof part with dormer, chimney, windows and antenna. Middle: segmentation with scale $\sigma = 1$. Right: segmentation with scale $\sigma = 4$. The regions belonging to minor roof parts, e. g. small parts of the dormer's roof or smaller shadows, do not live over a larger range of scales. Most of them merge with other regions with increasing scale.

In our context we need to consider the complexity of roof structures when deciding on the type of image features. Whereas *reconstruction* implicitly aims at a geometric description, and therefore uses features based on edges, cf. e. g. (Nevatia and Price, 1982), or edge-aggregates such as corners, cf. e. g. (Lang and Förstner, 1996), *detection* appears to better rely on features based on regions, especially their form.

Scale space for region extraction has already been investigated, cf. the review (Harvey et al., 1997). In contrast to the blob detection approach of (Lindeberg, 1993), we are interested in a complete partitioning of color images, not restricting to local maxima of intensity or a certain color. Therefore, we propose to use the watersheds of the gradient image, cf. also (Olsen and Nielsen, 1997). Additionally, we adopt the idea of finding maximally stable regions over scale, which are regions whose area does not change over scale, similar to the approach of MSER (Matas et al., 2002) which searches for region which are stable over intensity level sets. At the moment we do not exploit the hierarchical structure of the regions, as e. g. (Bangham et al., 1999) and (Kuiper et al., 2003).

The goal of this paper is to investigate the suitability of such regions for building detection. We will derive a statistics about the scale occurrence of certain roof parts, such as triangular or rectangular roof planes, which is a first attempt to derive the likelihood function for building part detection. Using only a single scale will turn out not to be sufficient for capturing the region information contained in an image, as we will show in our empirical investigations.

The paper is organized as follows: Section 2 describes the segmentation procedure. Sect. 3 presents our approach to measure the stability of regions. Sect. 4 investigates the suitability of the regions over scale space. Sect. 5 discusses the results and gives an outlook on future research.

2 SEGMENTATION

Our segmentation is based on the watershed boundaries derived from a sequence of images in the Gaussian scale space of the gradient magnitude.

The Gaussian scale space is built with logarithmicly ranged scales $2^{i/n}\sigma_0, i = -N_1, ..., 0, ..., N_2$ starting at $\sigma_{-N_1} = 2^{-N_1/n}\sigma_0$ and leading to $\sigma_{N_2} = 2^{N_2/n}\sigma_0$. In our experiments we use $\sigma_0 = 1$, $n = 10$, and $N_2 = 30$, thus scales between 1 and 8 with steps of a ten'th octave. The $N_1 = 17$ scales between 0 and 1 continue the arrangement of the larger scales into the scale between 0.3 and 1. Smaller scales are useless to compute, because the smoothing has nearly no effect.

For each scale $\sigma = \sigma_i$ the three band image $\boldsymbol{f} = [f_c], c = 1, 2, 3$ is convolved with a Gaussian filter $G(x, y, \sigma)$:

$$\boldsymbol{f}(x, y, \sigma) = \boldsymbol{f}(x, y) * G_\sigma(x, y). \tag{1}$$

As input function for the watershed algorithm we use the total gradient of the color images $\boldsymbol{f}(x, y, \sigma)$ as homogeneity measure: For each channel $f_c(x, y, \sigma)$ we compute the squared gradients $\|\nabla f_c(x, y, \sigma)\|^2$. In order to compensate for the different noise characteristics in the three color channels the homogeneity then is the sum of the squared gradients over all channels c weighted with the inverse of the variance σ_{n_c} of the noise

$$g(x, y, \sigma) = \sqrt{\sum_{c=1}^{3} \frac{\|\nabla f_c(x, y, \sigma)\|^2}{\sigma_{n_c}^2}} \tag{2}$$

in each channel. (Brügelmann and Förstner, 1992) have shown that the median of the squared gradients, except for a factor, is a good estimation for the noise variance. Therefore, we apply this approach and get the channel specific noise variance by

$$\sigma_{n_c}^2 = \text{med}_{x,y}(||\nabla f(x, y, \sigma)||^2). \tag{3}$$

In order to eliminate noise effects we use as input function for the watershed algorithm

$$h(x, y, \sigma) = \max(g(x, y, \sigma), m_g) \tag{4}$$

where

$$m_g = \text{med}_{x,y}(g(x, y, \sigma)). \tag{5}$$

The watershed algorithm takes the local minima of the input function as seed points and performs a region growing. This gives us a complete partitioning of the image. The result is a label image

$$l(x, y, \sigma) = \text{WS}\,[h(x, y, \sigma)] \tag{6}$$

that has the same labels at the catchment region of the local minima. It can be thought as flooding the basins, if the input function is seen as height values of a virtual landscape. All border pixels of watershed regions are labeled 0.

3 STABILITY OF REGIONS OVER SCALE SPACE

Regions which show only little variation over a certain scale range can be termed stable. There are various metric and topological criterions for measuring stability of regions, but nevertheless the area is the most important one: The region size changes dramatically when regions merge or split. Other region's properties do not change that much over scale.

For obtaining the stability of the L_σ regions $R(l, \sigma), l = 1, ..., L_\sigma$ at scale σ , we compute the area $|R(l, \sigma)|$ of each region from the histogram of $l(x, y, \sigma)$ in (6). We build a set of images where each region is labeled with its area

$$a(x, y, \sigma) = |R(l(x, y, \sigma), \sigma)| \tag{7}$$

We then analyse the *area function*

$$a(\sigma|x, y) \tag{8}$$

for manually selected points (x, y) over the scales. Taking points in regions with a selected content allows to investigate the stability of these regions, thus their usefulness for detection.

In order to evaluate the stability of the area from the area function $a(\sigma|x, y)$, we have to consider the uncertainty of the area of regions. Areas can be categorized as stable in case their area lies in the error band over a large enough range of scales. We require stability over at least 10 scale space layers, i. e. over at least one octave.

The uncertainty of the area A of a polygon $[\boldsymbol{p}_j]$ with J chord lengths d_j between the two neigboring points $\boldsymbol{p}_{j-1}$ and $\boldsymbol{p}_j$ can be shown to be

$$\sigma_A^2 = \frac{1}{4}\sum_{j}^{J} d_j^2\,\sigma_p^2 \tag{9}$$

if all points have the same standard deviation σ_p and taking the indices j cyclically, cf. (Förstner, 1999). In case of dense points and a smooth boundary the standard deviation of the area

$$\sigma_A = \frac{U}{\sqrt{J}}\sigma_p \tag{10}$$

reveals to be only dependent on the length U of the boundary, the number of border points J and the standard deviation σ_p of the points. We use a 3 σ_A-error bands assuming a positional error of $\sigma_p = 0.5$ [pixel] to estimate the accuracy of the region's boundary.

4 EMPIRICAL INVESTIGATION

The empirical investigation aims at exploring the usefulness of regions over scale space for building detection. A region is useful, if we can expect that features, which are distinctive for separating building parts from non-building parts, can be derived automatically. Therefore we select roof regions in a supervise mode by picking a point in the region, identify the scale range for stable regions from the area function $a(\sigma|x, y)$ and evaluate these regions visually with respect to their usefulness. Though this is subjective, it gives a clear indication whether there is a chance at all, that stable and relevant regions may be found. We also want to find out whether there are characteristic scales for different classes of roof regions.

4.1 Basic categories of roof planes

Since roofs are the most distinctive parts of a building seen in an aerial image, there seems to be no limit for the complexity of urban roof structures. We model our roof prototypes by examining their roof planes.

Although there exist various catalogs of basic roof styles due to roof construction, we define roof prototypes in a different way. This is due to the high complexity of roof structures, which can be only partially categorized by classical roof styles. Therefore our detection scheme does not aim at such a categorization of complete roofs, but only on categories of roof planes.

Fig. 2 shows some basic roof styles of suburban buildings. Each of these roofs can be modeled by triangles and tetragons. Some planes of half-hipped roofs are hexagons. Of course, more complex buildings, e. g. L-shaped buildings, show other shapes, e. g. parallelograms or skew trapezoids. However, taking roof regions as key-features for triggering building detection does not require distinguishing between planes of major roof and those of dormers (see fig. 3).

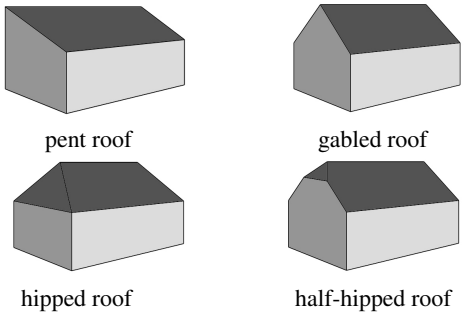

Figure 2: Common types of Middle European roofs

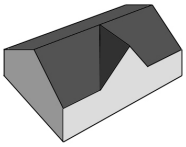

Figure 3: Example for building with a dormer

We therefore represent each roof by the roof planes together with their geometric traits and by their adjacency graph, possibly including attributes of the type of neighborhood.

Besides dormers the roof planes can get disturbed by other objects fixed on the roof. Examples of these objects are chimneys, windows and solar cells (see fig. 1, left). Furthermore, roofs can be occluded by trees or other buildings and their shadows. These disturbances may affect the region detection. Whereas some of these, e. g. antennas and their shadows, will not be visible at lower resolution, thus at other scales, others such as occlusions will change the form of the extractable regions, however, be visible over a larger range of scales.

At this stage of our investigation we are only interested in the observability of stable regions, which show roof-type structures.

4.2 Test Data

Our experiments are based on aerial image data, showing suburban buildings in the cities of Bonn (Germany), Graz (Austria), and Toyonaka (Japan).

Bonn: We consider 13 aerial images taken over Bonn, having a ground resolution of 10 cm, cf. figs. 4 and 5. The first example image shows a scene of a shopping area with bigger flat and gabled roofs. In the other example image, there is a suburban scene with gabled roofs often having additional roof parts as dormers or windows. Due to the time of image acquisition in winter, the vegetation around the buildings does not show a strong contrast. Additionally, as the position of the sun is quite low, the shadows often reach to the neighboring building.

Figure 4: Image section from Bonn, shopping area, 10 cm ground resolution

Figure 5: Image section from Bonn, residential area, 10 cm ground resolution

Graz: Our 14 test images of Graz have a ground resolution of 8 cm (cf. fig. 6). Most of the roofs are covered with red tiles, the buildings are surrounded by fresh vegetation. There are only small shadows in the picture, the roof planes are only disturbed by other objects, such as chimneys or solar cells. The images show many gabled and cross gabled roofs.

Figure 6: Image section Graz, residential area, 8 cm ground resolution, kindly provided by Vexcel Imaging GmbH in Graz

Toyonaka: Our 9 test images of Toyonaka have a ground resolution of 7 cm (cf. fig. 7). The concentration of buildings differs strongly from the other test images. Roofs with colorful tiles are detectable by eye very well. Due to the low position of the sun, shadows make it difficult to distinguish between dark covered roofs. There are no other objects on the roofs, but most of the houses are extended by additional building parts. Our test image does not show any vegetation and is weak in contrast. Most of buildings have hipped or pyramid roofs.

Figure 7: Image section Toyonaka, dense residential area, 7 cm ground resolution, kindly provided by Vexcel Imaging GmbH in Graz

4.3 Experimental Results

We selected roof planes manually to observe the stability of their area in scale space. Tab. 1 shows the number of stable regions we found in all images. We distinguished the regions by their shape. The row of more complex shaped regions refers to those regions, which have melted together with other regions of the roof still forming characteristic roof shapes. Less than 20% of the selected regions were not stable at all, these regions are not taken into account any further.

Shape	Bonn	Graz	Toyonaka	Σ
triangle	20	71	151	242
square	55	95	16	166
rectangle	205	373	152	730
trapezoid	39	60	115	214
more complex	56	60	176	292
Σ	375	659	610	1644

Table 1: Statistics of selected regions which turned out to be stable.

Fig. 9, an extract of fig. 5, demonstrates that at various scales relevant roof areas can be detected. We obtain a thin recangular shaped roof plane, which merges with another one at $\sigma \approx 3$ [pixel]. As long as the balconies form bays at the bottom of the region, it is not considered to be stable. In contrast to the balcony bays, the hole of the region belonging to the window is very stable.

Fig. 10, an extract of fig. 4, demonstrates that even smaller roof parts as dormer roofs can be stable, too. From scale $\sigma = 4$ [pixel] on, the region merges with a vegetation area in front of the building.

Fig. 11, an extract of fig. 5, shows the smoothing deforms roof part with increasing scale. The region is stable over various scale space layers, but the shape of the original region changes from a triangle to a circle within the last 10 layers, resp. starting at $\sigma \approx 4$ [pixel].

Fig. 12, an extract of fig. 7, shows a problem of our manual region selection: The Japanese roofs are often strong textured in opposite to the most European roofs. In this case, one almost always selects a border point of region at at least one scale space layer between $\sigma = 0$ and 1 [pixel]. So, the area function $a(\sigma|x, y)$ is not determined from a region but from all border pixels in the image (which have the same label: 0). From $\sigma = 1$ [pixel] on, the roof planes are well observable.

The results of our investigation are shown in fig. 8. It is organized in a max-min-diagram that shows the maximal versus the minimal level of scale for the rectangular roof parts, measuring the scales in dm at ground level. The other roof parts show similar results, the graphics would present nearly the same range of positions, the density of the drawn dots would only be less.

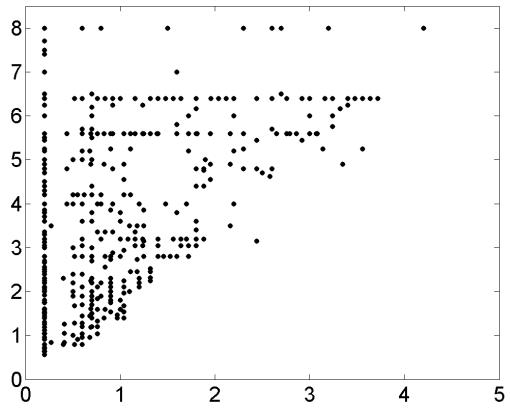

Figure 8: The results of the investigation on rectangular regions are drawn in a max-min-interval diagram. The results of the different shaped regions are similar to those as tab. 2 shows.

Obviously, we have very stable regions, where the minimum scale is small and the maximum scale is large, namely those in the upper left of the diagram. We also find regions which only live in large scales as those few in the upper right of the diagram. We finally find regions which only exist at small scales, i. e. those in the lower left.

There is no certain scale where uniquely formed regions can be found. Also, choosing a certain scale for finding regions, say 3 dm, would only allow to detect those regions which are in the upper left rectangular having its lower right corner at (3,3), thus missing quite some relevant regions, in the lower left and the upper right of the diagram.

The relevance of the extracted regions has only been evaluated in general: over appr. 80% of all regions are stable over at least an octave in the investigated scale range.

Tab. 2 also contains the range of the minimal and maximal scales in [dm]. The different forms appear in all scale ranges. As the minimum and maximum ranges of the scales almost totally overlap, selecting a single scale in this overlap would lead to a loss in region detection, e. g. when choosing $\sigma = 3$ dm and searching for rectangular roof regions.

Type	$\sigma_{\min}$ [dm] min – max	$\overline{\sigma_{\min}}$ [dm]	$\sigma_{\max}$ [dm] min – max	$\overline{\sigma_{\max}}$ [dm]
triangle	0.21 – 3.56	0.58	0.64 – 8.00	3.13
square	0.21 – 4.00	0.61	0.56 – 8.00	3.37
rectangle	0.21 – 4.15	0.74	0.56 – 8.00	3.95
trapezoid	0.21 – 4.15	0.59	0.56 – 8.00	3.48
others	0.21 – 4.25	0.86	0.72 – 8.00	3.64

Table 2: Range of minimal and maximal scales over all stable regions, additionally their means, distinguishing the shapes of regions.

5 CONCLUSION

This paper is a first investigation into the detectability of building roofs via regions which are stable in scale space. The stability of a region can be measured by the scale range where the region's area does not change significantly.

We used the watershed algorithm on the averaged and weighted gradient magnitude image for image partitioning. The weights

are the inverse of the noise variance in the different channels. These regions turned out to be quite stable over scale.

We have shown, that regions that represent roofs and roof parts in aerial images can only be extracted in certain intervals of scale. However, there is no optimal scale for the extraction of roof parts in aerial images. It is necessary to automatically choose the scale for each region.

The usefulness of stable regions was explored. Over appr. 80% of the roof regions, which were selected manually, lead to image regions which were stable and promised to have attributes for reliable detection.

We are currently investigating the automatic extraction of regions which are stable over scale space. We are setting up an annotated image database, which makes it possible to train our building detectors.

The approach should be easily transferred to other types of images, as such region detectors exploiting scale space can be expected to play at least a prominent role as point type detectors.

Acknowledgements

This work was done within the project "Ontological scales for automated detection, efficient processing and fast visualisation of landscape models" which is supported by the German Research Foundation (DFG). The authors would also like to thank Ribana Roscher and Susanne Wenzel for assisting at the evaluation.

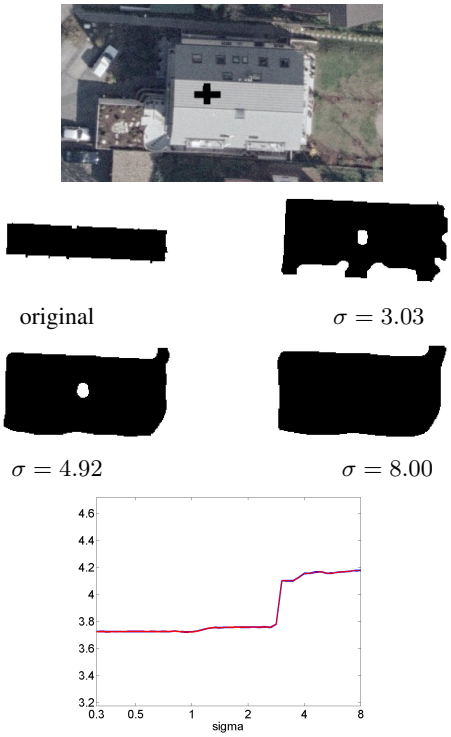

Figure 9: Example 1 of the development of a region: $\log(a(\sigma|x,y))$. Starting with a thin rectangular roof plane, the region merges at higher levels with other roof planes and roof objects. Graph: Relation between smoothing scale and region sizes together with the error bands. Stable Regions allude to concrete scales.

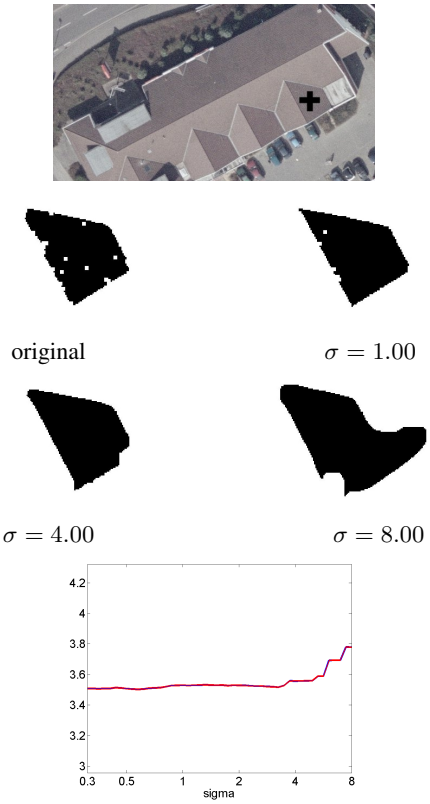

Figure 10: Example 2 of the development of a region: $\log(a(\sigma|x,y))$. Starting with a trapezoid part of a dormer, the region merges at higher levels with a vegetation area in front of the building. Graph: Relation between smoothing scale and region sizes together with the error bands.

REFERENCES

Bangham, J. A., Moravec, K., Harvey, R. and Fisher, M., 1999. Scale-space trees and applications as filters for stereo vision and image retrieval. In: BMVC, pp. 113–143.

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W. and Ebner, H., 1999. Automatic road extraction based on multi-scale, grouping, and context. Photogrammetric Engineering & Remote Sensing 65, pp. 777–786.

Brügelmann, R. and Förstner, W., 1992. Noise estimation for color edge extraction. In: W. Förstner and S. Ruwiedel (eds), Robust Computer Vision, Wichmann, Karlsruhe, pp. 90–107.

Brunn, A. and Weidner, U., 1998. Hierarchical bayesian nets for building extraction using dense digital surface models. Journal for Photogrammetry & Remote Sensing 53(5), pp. 296–307.

Förstner, W., 1999. Areas and their uncertainty. Technical report, Institute of Photogrammetry, University of Bonn.

Harvey, R., Bangham, J. A. and Bosson, A., 1997. Scale-Space Filters and Their Robustness. Vol. Lecture Notes in Computer Science 1252, pp. 341–344.

Herman, M. and Kanade, T., 1987. The 3d mosaic scene understanding system: Incremental recognition of 3d scenes from complex images. In: Fischler/Firschein (ed.), Readings in Computer Vision, Kaufmann, pp. 471–482.

Huertas, A. and Nevatia, R., 1988. Detecting building in aerial images. CVGIP 41, pp. 131–152.

Jacobsen, K., 1997. Comparison of information contents of different space images. In: Joint Workshop "Sensors and Mapping from Space".

Kuiper, A., Florack, L. and Viergever, M., 2003. Scale space hierarchy. JMIV 18, pp. 169–189.

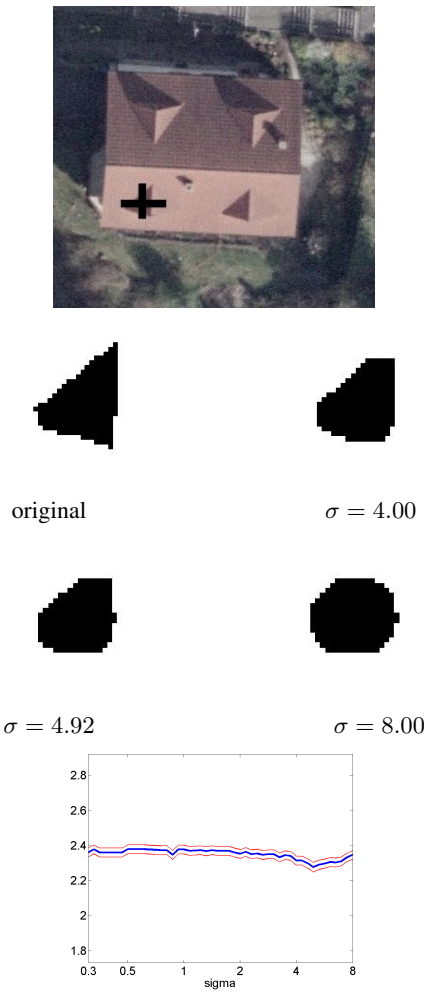

Figure 11: Example 3 of the development of a region: $\log(a(\sigma|x,y))$. Deformation of a region's shape by image smoothing. Graph: Relation between smoothing scale and region sizes together with the error bands.

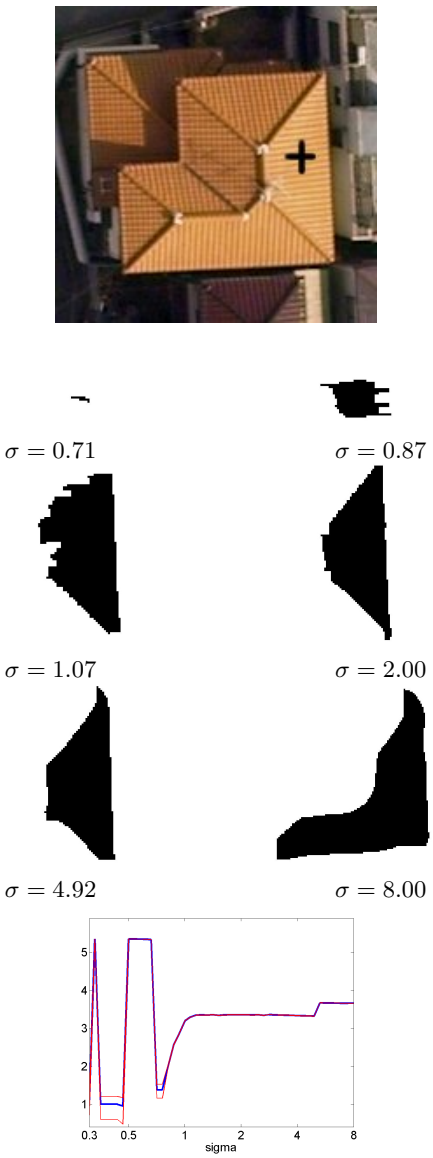

Figure 12: Example 4 of the development of a region: $\log(a(\sigma|x,y))$. The strong oscillation of the area refers to wrongly selected regions. Graph: Relation between smoothing scale and region sizes together with the error bands.

Lang, F. and Förstner, W., 1996. Surface reconstruction of man-made objects using polymorphic mid-level features and generic scene knowledge. In: International Archives of Photogrammetry and Remote Sensing, Part B3, Vol. 31, pp. 415–420.

Lindeberg, T., 1993. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. IJCV 11(3), pp. 283–318.

Matas, J., O.Chum, M.Urban and T.Pajdla, 2002. Robust wide baseline stereo from maximally stable extremal regions. In: BMVC, Vol. 1, pp. 384–393.

Mayer, H., 1999. Automatic object extraction from aerial imagery - a survey focusing on buildings. Computer Vision and Image Understanding 74, pp. 138–149.

Mayer, H., Laptev, I. and Baumgartner, A., 1998. Multi-scale and snakes for automatic road extraction. In: Fifth European Conference on Computer Vision (2), pp. 720–733.

Nevatia, R. and Price, K. E., 1982. Locating Structures in Aerial Images. IEEE Transactions on Pattern Analysis and Machine Intelligence (5), pp. 476–484.

Olsen, O. F. and Nielsen, M., 1997. Multiscale gradient magnitude watershed segmentation. In: ICIAP'97 - 9th International Conference on Image Analysis and Processing, Vol. Lecture Notes in Computer Science 1310, pp. 9–13.

Pakzad, K. and Heller, J., 2004. Automatic scale adaptation of semantic nets. Publikationen der DGPF 13, pp. 67–76.