# Building Detection in Edge-Representation of Aerial Images by Means of Heuristic Histogram Analysis

A. Pogoda, V. Steinhage
Institut für Informatik I, Univerität Bonn
Römerstr. 164, 53117 Bonn

http://arthur.cs.uni-bonn.de/Building
e-mail: pogoda|steinhag@cs.uni-bonn.de

**Abstract:** We present a new approach for the detection of buildings in digital aerial images. It uses heuristic analysis of edge orientation and edge position histograms, that are performed within a sampling of an edge based representation of the input image. The result is a map that defines a belief value for every position in the input image pleading for the occurrence of buildings at the corresponding position.

# 1 Motivation

Three-dimensional building extraction from digital images becomes an issue of increasing importance for a large number of applications in town planning, architecture, environmental investigations etc.

Image data usually reveal irrelevant information on the one hand and loss of relevant information on the other hand, as can be seen in figure 1. Furthermore the complexity of building shapes makes a sufficiently complete modeling of buildings and their appearances in real images necessary.

In [Kort *et. al.* 96] we presented a model based approach for building extraction that uses template matching of the image data and projective appearances[1] of building models. A problem that arose from this work was the determination of worthy initial searching areas where the template matching it to be performed.

Considerable methods for searching such areas of high probability for the occurrence of buildings are (1) a semiautomatic, operator driven preselection of image areas (see [Lang and Schickler 93]), (2) local elevation peaks in the case of a given digital elevation model (see [Weidner and Förstner 95]), (3) a model based scoring of image areas.

Local elevation models are expensive and often unavailable for aerial images. In [Braun and Kolbe 95] we gave a detailed overview about approaches to 3D building extraction from aerial images

In this work we present an approach that uses statistical and heuristic interpretations — guided by a general building model — of an edge based image representation to define such areas. Our aim is to sample the edge representation of the input image, to evaluate the configurations of edges at the sampling positions and to construct a map for this image that defines for each point of the input a belief value for the occurrence of a building.

[1] These appearances are represented as an *aspects hierarchy*, see [Dickinson *et. al.* 92]

# 2 Building Model

Our input is an edge representation of the original gray level image that was obtained by passing the original through a general-purpose feature extraction module, see [Fuchs and Förstner 95]. Note that this initial step employs a very general object model, not specific for building detection, i. e. it assumes homogeneous, smooth faces with continuous edges.

We currently use images with a resolution of about $500 \times 500$ – $1000 \times 1000$ pixels that contain about 10–20 buildings which occupy about 20% of the visible area. The images are taken from nearly overhead views. Figure 1 shows an exemplary gray level image as well as the extracted edge segments.

Figure 1: The input image (left) and the extracted edge segments (right).

Although plenty of different instances of buildings exist, there are a few characteristics that most of them resp. their projections have in common:

**O1**: specific dimensions

**O2**: orthogonal corners between walls or even orthogonal ground plans that produce orthogonal arrangements of edges in the projection[2]

**O3**: parallel wall groups and therefore parallel edge groups from which at least some edges are well detectable

The results of edge detecting processes are in general incomplete and noisy due to phenomena like occlusions, shadows, low contrasts etc., furthermore the edge based representation of such an image typically contains 2000–3000 edges. This shows the complexity of the correspondence problem and causes a naive approach like finding probable matchings between the ungrouped image data and the model to fail.

Moreover we cannot expect (1) to find edges that uniquely correspond to complete building edges, (2) to refer immediately to above observations **O1** – **O3**, (3) to determine if these edges might denote buildings.

# 3 Grouping of Image Structures

Obviously we need an information aggregating technique for deciding whether a given area of the image shows a building or not. We will show how **O1** — **O3** can be met

[2] As we deal with nearly overhead views, even sloped roof edges form nearly orthogonal angles.

by an analysis of histograms in a subregion of the image.

To start with, we define this subregion in the following subsection and show how its choice meets observation **O1**. In subsections 3.2 and 3.3 we define two types of histograms:

- orientation histograms
- distance histograms

These hold the opportunity to meet **O2** resp. **O3**.

## 3.1 Clipping area

We isolate a circular region $A$ of the image with centerpoint $M = (x, y)$. Only edges $e_i$ that fit completely into this circle — denoted by $e_i \in A$ — will be considered in the following steps in which we will compute histograms of the edges within this circle. With different choices for $M$, we can perform the sampling process.

The wise choice of this clipping area's radius $r$ is essential for the detection result as it defines the estimated extent of typical buildings in the image in the further processes and therefore prevents us from considering edges that are too long to denote building elements, thus the radius $r$ is a parameter that corresponds to **O1** in section 2. This radius can be derived by the a priori knowledge of building dimensions on the one hand and the resolution of the given aerial image on the other hand.

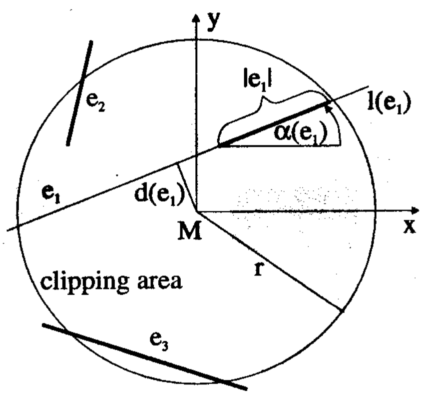



Figure 2: The circular clipping area. The edges $e_2$ and $e_3$ are rejected, as they do not fit completely. The edge $e_1$ is accepted, it's properties are displayed in the figure.

Let $l(e_i)$ be the line that corresponds to an image edge $e_i$, i.e. the extension of $e_i$ to an infinite line. Each considered edge $e_i \in A$ is assigned three values:

| | |
|---|---|
| $0 < \|e_i\| \leq 2r$ | the euklidian length of $e_i$ |
| $0 \leq \alpha(e_i) < \pi$ | the angle between $l$ and the horizontal axis |
| $-r \leq d(e_i) \leq r$ | the minimum distance of $l$ and $M$ |

See also figure 2 that shows these quantities.

The values $\alpha(e_i)$ and $d(e_i)$ define the Hesse form of the line $l(e_i)$ in a coordinate system with the origin placed at $M$. The computation of $|e_i|$ and $\alpha(e_i)$ can be done once in the initial step whereas $d(e_i)$ depends on the clipping area that we currently consider.

## 3.2 Orientation Histograms

For a considered clipping area $A$ with center $M$ we define an orientation histogram $O$:

$$O(\beta, A) = \sum_{\substack{e_i \in A \\ \alpha(e_i)=\beta}} |e_i|^{w_O} \quad \text{with} \quad 0 \le \beta < \pi$$

The parameter $w_O$ defines a weighting of the edges depending on their lengths, e.g. for $w_O = 2$ the edges $e_i$ contribute with the square of their lengths to the histogram. Figure 3 shows two exemplary clipping areas and the corresponding histograms. We denote the average $\int_0^\pi O(\beta, A)\ d\beta/\pi$ of a histogram $O$ with $\bar{O}$. Extracting peaks in

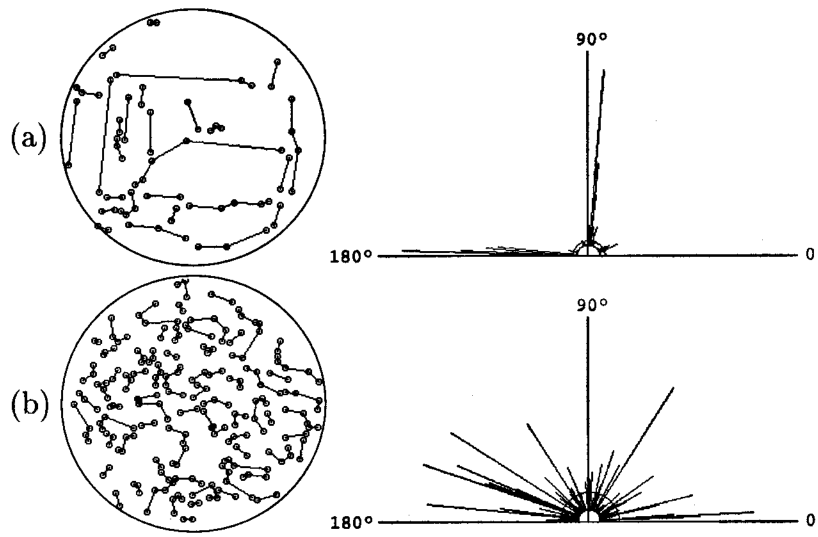



Figure 3: Clipping areas and corresponding orientation histograms. (a) shows the histogram of a clipping area with a building, (b) shows a histogram of a non-building area. The outer semicircle represents the histogram's average $\bar{O}$.

these histograms means considering intervals that define local maxima. These maxima vote for preferred orientations of edges in the image. Section 4 explains how these preferred orientations can be evaluated to meet observation **O2**.

As we deal with discrete angle values and discrete histogram resolutions, we always regard an orientational peak $P_O$ together with an interval

$$P_O = [\rho - \epsilon_l, \ldots, \rho, \ldots, \rho + \epsilon_r] \quad \text{with} \quad O(\rho, A) > O(\tau, A) \quad \forall \tau \neq \rho$$

for which the subintervals $[\rho - \epsilon_l, \ldots, \rho]$ and $[\rho, \ldots, \rho + \epsilon_r]$ with the such defined left and right borders $\rho - \epsilon_l$ and $\rho - \epsilon_r$ are monotonic, to derive a tolerance range for the further steps. By iterating the maximum finding process we obtain a series of peak positions. To ensure termination of the iteration we extract only the first $n_O$ peaks.

Furthermore we assume that a peak's maximum must be larger than a threshold $O_{\min} = c_O \cdot \bar{O}$, where $c_O$ is a configurable, domain dependent parameter that describes the ratio between the histogram's average and this lower bound for peaks. This linear dependency makes the choice of the threshold independent from the histogram's absolute values. Finally we assume that at least one edge that contributed to the histogram at a given peak position must have a minimum length $l_O$. These heuristic values $c_O, l_O$ and $w_O$ serve to distinguish between significant peaks and accidental, small peaks of histograms that do not show building specific configurations of edges.

The parameter for minimum length, $l_O$, again corresponds to observation **O1**, as it defines a minimum length of building edges. As edges might be split by the feature extraction module, its concrete value — about $0.5r$ — is somewhat smaller than the typical length of complete edges specific for buildings.

The value for $c_O$ just as for $w_O$ is not critical as we will show in section 5. Experimental results show that $c_O \approx 3$, $w_O \approx 2$ is a good choice.

The choice of $n_O$ strongly affects the performance of the later steps as it restricts the number of peaks that have to be analysed. A value of about 5 is sufficient for the desired application as building specific configurations typically occur within these largest peaks.

## 3.3 Distance Histogram

Analogously to the construction of the orientational histograms we define for a given orientational peak $P_O = [\rho - \epsilon_l, \ldots, \rho, \ldots, \rho + \epsilon_r]$ and a clipping area $A$ the distance histogram:

$$D(z, A, P_O) = \sum_{\substack{e_i \in A \\ \alpha(e_i) \in P_O \wedge d(e_i) = z}} |e_i|^{w_D} \quad \text{with} \quad -r \leq z \leq r$$

Here the interval is not cyclic, a peak $P_D$ of such a histogram votes for a stripe shaped image area with strong agglomeration of nearly parallel edges. We call these stripe shaped areas *agglomeration axes.* The dependencies between the orientation and the distance histograms can be looked upon as a two level hierarchy as displayed on the lefthand side of figure 4. This figure also displays how the peaks of the distance histograms define agglomeration axes in the image. These agglomeration axes can be utilized to meet observation **O3**.

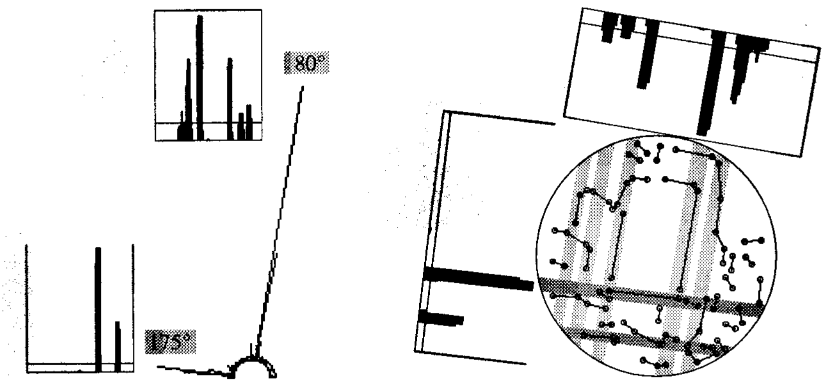



Figure 4: Histogram hierarchy of a clipping area defines axes in the image with agglomeration of nearly parallel edges.

The principles for finding peaks are the same, just as the meaning of the parameters $w_D$, $c_D$, $n_D$ and $l_D$. The concrete values of these parameters differ slightly from that for $w_O$ resp. $c_O$ because we made the observation that only very few edges contribute to a distance histogram whereas many edges contribute to an orientational histogram. The reason is that taking only edges with a specific orientation into consideration for such an histogram means a strong preselection.

Furthermore we already know that the edges that contributed to a peak must have been significant, because they got over the selection caused by the previous computations. Thus we do not have to weight them as strong as before, profitable values are $w_D \approx 1$ and $c_D \approx 1$. Note that $w_D = 1$ does not mean that there is no weighting at all, but each edge $e_i$ contributes with its length to the histogram.

The parameter $l_D$ for minimum length should be also chosen different from $l_O$. In the orientational histograms any edge of a specific orientation was considered, regardless of its distance. Therefore — in the case of ridges or gutters — any ridge or gutter contributed to an orientational peak. In **O3** we stated that some of them might be decayed to a large extent. If we consider single ridges or gutters now, we have to accept somewhat smaller edges, too. For this reason we assign a smaller value to $l_D$, approx. $0.2r$.

Finally, we set $n_D = 5$ just as we did for $n_O$, because a projection of a building typically can show many parallel edges, especially in the case of an inaccurate overhead view.

# 4 Evaluation Strategy

As mentioned before, the first and most important source of external knowledge is the choice of the clipping area's radius, as it defines the approximate extents of buildings visible in the image. Structures like roads will not be detected in most cases, as they produce too long edge segments in the image, that will be rejected by the clipping area. On the other hand small details will produce minor contribution to the histograms because of the weighting exponents $w_O$ resp. $w_D$.

Analogously we can find histogram characteristics that correspond to the other observations of section 2; we meet **O2** by expecting a nearly orthogonal angle between two of the computed orientational peaks. For simplicity reasons let an orientational peak $P_O$ be characterized only by its maximum value $\rho$ rather than by its surrounding interval.

**C1** : $$\exists P_1, P_2 \text{ peaks of } O(\beta, A) \ : |P_1 - P_2| = \pi/2 \pm \theta$$

The parameter $\theta$ stands for a tolerance interval, a value of about $5°$ suffices[3]. If no such orthogonal configuration is found, there is no need to compute the distance histograms at all, what results in a significant speedup of the computation. Otherwise further constraints are necessary to distinguish between buildings and accidental similar configurations caused by crossroads and other orthogonal structures.

A further constraint is the assumption that at least two agglomeration axes are detected in the distance histograms that correspond to the two orthogonal directions $P_1$ and $P_2$:

**C2** : $$\exists P_{i1}, P_{i2} \text{ peaks of } D(z, A, P_i) \quad ; \quad i = 1,2$$

This meets observations **O3**.

Moreover the agglomeration axes in both of the directions should have a minimum distance to each other to ensure that they describe image structures that occupy a sufficiently large area to denote buildings:

**C3** : $$|P_{i1} - P_{i2}| > z_0 \quad ; \quad i = 1,2$$

This is another expectation about the typical building dimensions and therefore also corresponds to **O1**. Note that this constraint implies that $r > |P_{i1} - P_{i2}|$, as we only consider axes within the clipping area when computing the histograms. The value is similar to $l_O$, $z_O \approx 0.5r$, as they both describe minimum dimensions which are perpendicular to each other.

[3] Although we restricted our approach to operate on overhead views, we must tolerate slight violations of this restriction.

The values for all of these parameters can be derived from the domain dependent size of the clipping area what again emphasizes this parameter's outstanding importance, but as there might occur different house sizes within one image it might be more profitable to define this constraint for minimum and that for maximum dimensions — the radius $r$ of the clipping area — independent from each other.

The response $\Phi$ of the such obtained filter, applied on a clipping area $A$ votes for or against the occurrence of a building within $A$:

$$\Phi(A) = \begin{cases} 1 & \text{if} \quad \mathrm{C1} \wedge \mathrm{C2} \wedge \mathrm{C3} \\ 0 & \text{else} \end{cases}$$

By sampling the image in discrete points $M_j = (x_j, y_j)$ we can construct a map $B(x,y)$ of belief values; starting with a map that initially contains the value $B(x,y) = 0$ for every position $(x,y)$, the values within the currently considered clipping area $A$ are incremented, if this area pleads for the occurrence of a building. Thus every point $(x,y)$ in the result map is assigned a belief value $B$ as below:

$$B(x,y) = \sum_{\substack{A \\ (x,y)\in A}} \Phi(A)$$

The map's regions with high scores maximize the probability for finding a building in these areas.

**Algorithm:**

**Input:** Edge represented image $I$
**Output:** Result map $B(x,y)$, same size as $I$.

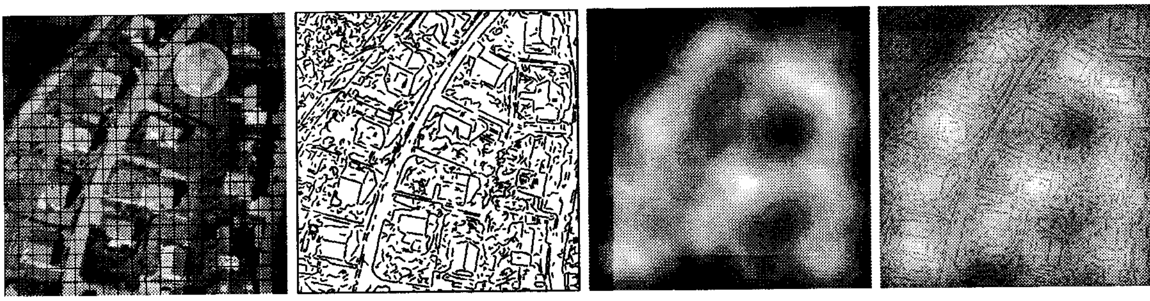
initialize $B$ with 0
**foreach** point $S_j = (x_j, y_j)$ of a sampling scheme **do**
  create clipping area $A(S_j)$
  compute orientation $(O(A,\beta))$ and distance $(D(A,z,P_i))$ histograms of $A(S_j)$
  **if** $\Phi(A(S_j)) = 1$
    **foreach** point $(x_k, y_k)$ of $B$ with $(x_k, y_k) \in A(S_j)$ **do**
      increment belief value: $B(x_k, y_k) \leftarrow B(x_k, y_k) + 1$
    **od**
  **fi**
**od**

# 5 Results

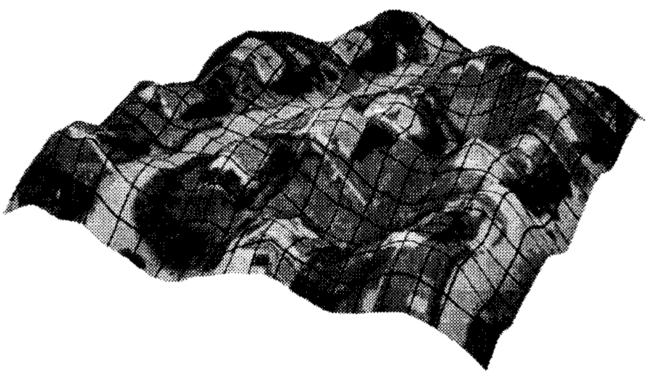
Figure 5 shows the result of a sampling process of the edge representation of the input image. We get a smooth gray level image. The bright maximums stand for a high belief value for the occurrence of buildings at the corresponding positions. The clipping area's radius was $r = 10\%$ of the image-width [4], the brightened circle in figure 5 (a) displays its size. The computation was done using using $200 \times 200$ sample positions.

The input image contains 14 buildings, from which 13 could be detected with a high belief value in the result map. The gambrel roof building in the upper right corner was not detected successfully. The reason is, that it is located at the image's border, therefore only few sample positions contributed to the belief value.

[4] Just as 10'% of the height, as this image is square



(a) (b) (c) (d)



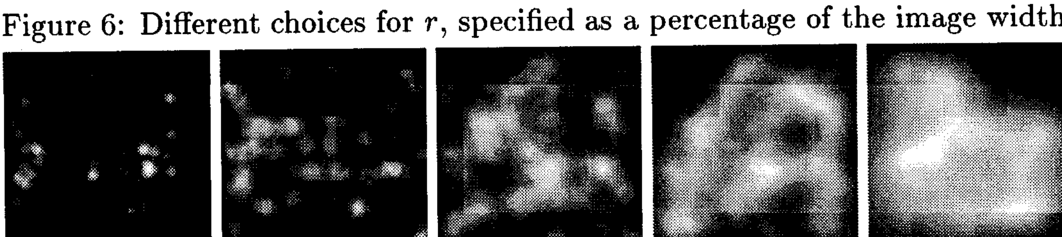
(e)

Figure 5: Input (a) with a clipping area as used in this example, extracted edges (b). The result of histogram analysis (c) encodes the belief value as the intensity. Image (d) is a mixture of (b) and (c). Image (e) encodes the belief value as the height of the displayed surface. Its texture is the input image. The grid helps to visualize the 3D structure of (e), it does *not* represent the sampling positions !

The high importance of the radius $r$ can be seen in figure 6, where we performed the sampling process for the example of figure 5 with different choices for the radius, conserving the remaining parameters. The results are best (at $r = 10\%$), if the radius is chosen such, that the buildings fit the clipping area.

Figure 7 demonstrates the robustness of the approach as its results remain constant within a wide range with regard to the choice of the control parameters — except the radius, of course. The value for $c_O$ can be chosen safely within a range of about 1–5, that for $w_O$ within 1.5–3. The choice of the sample rate is quite stable, too. Figure 5 (c) arose from a $200 \times 200$ sampling, whereas we performed $100 \times 100$ samples for figure 6.

Figure 6: Different choices for $r$, specified as a percentage of the image width.

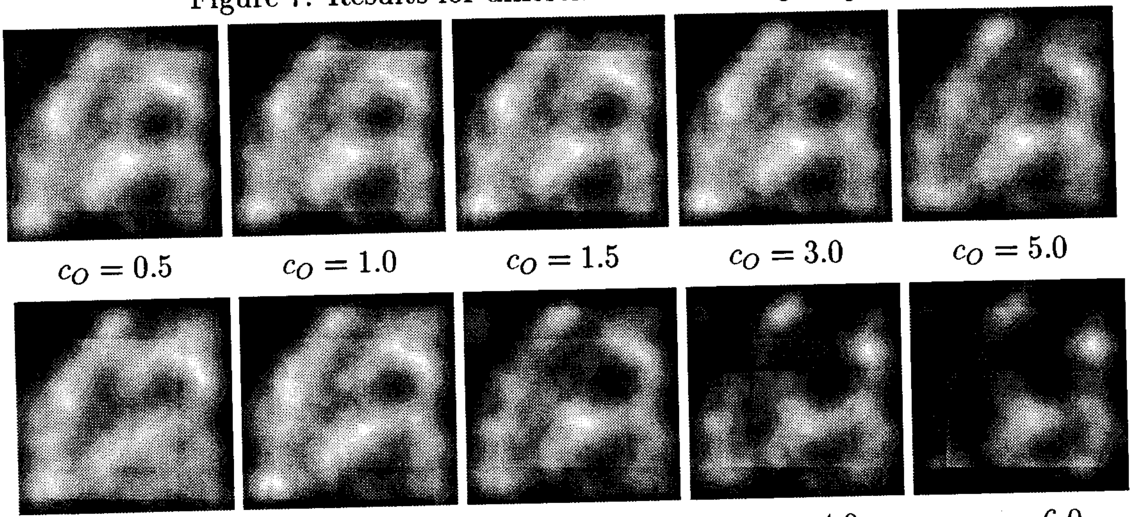


$r = 2.5\%$ $r = 5\%$ $r = 7.5\%$ $r = 10\%$ $r = 17.25\%$

The computation of a typical image as described in section 2 takes between 10–45 min. on an Intel 486-66 system at a sample rate of about $100 \times 100$, depending on the choice of parameters. It should be noted that the sampling process is suitable to a high degree for parallel computing.

Figure 8 shows the results of our approach applied on other input images. We used the same parameter set as for the image in figure 5. The images contain buildings of different dimensions, what makes a wise choice for the radius of the clipping area

Figure 7: Results for different choices for $c_O$ resp. $w_O$.



$c_O = 0.5$ $c_O = 1.0$ $c_O = 1.5$ $c_O = 3.0$ $c_O = 5.0$

$w_O = 1.0$ $w_O = 1.5$ $w_O = 3.0$ $w_O = 4.0$ $w_O = 6.0$

difficult. If the radius is too small, large buildings remain undetected, if it is to large, many disturbing edges contribute to their histograms.

# 6 Conclusion and Future Work

We reduced the problem to determine worthy areas to be traced by the later template matching steps to a maximum finding problem in the result map. By registering also the orientation of the two orthogonal directions found in the orientational histograms, better results can be achieved, because different regions that might melt together sometimes, can be distinguished by means of this extra information.

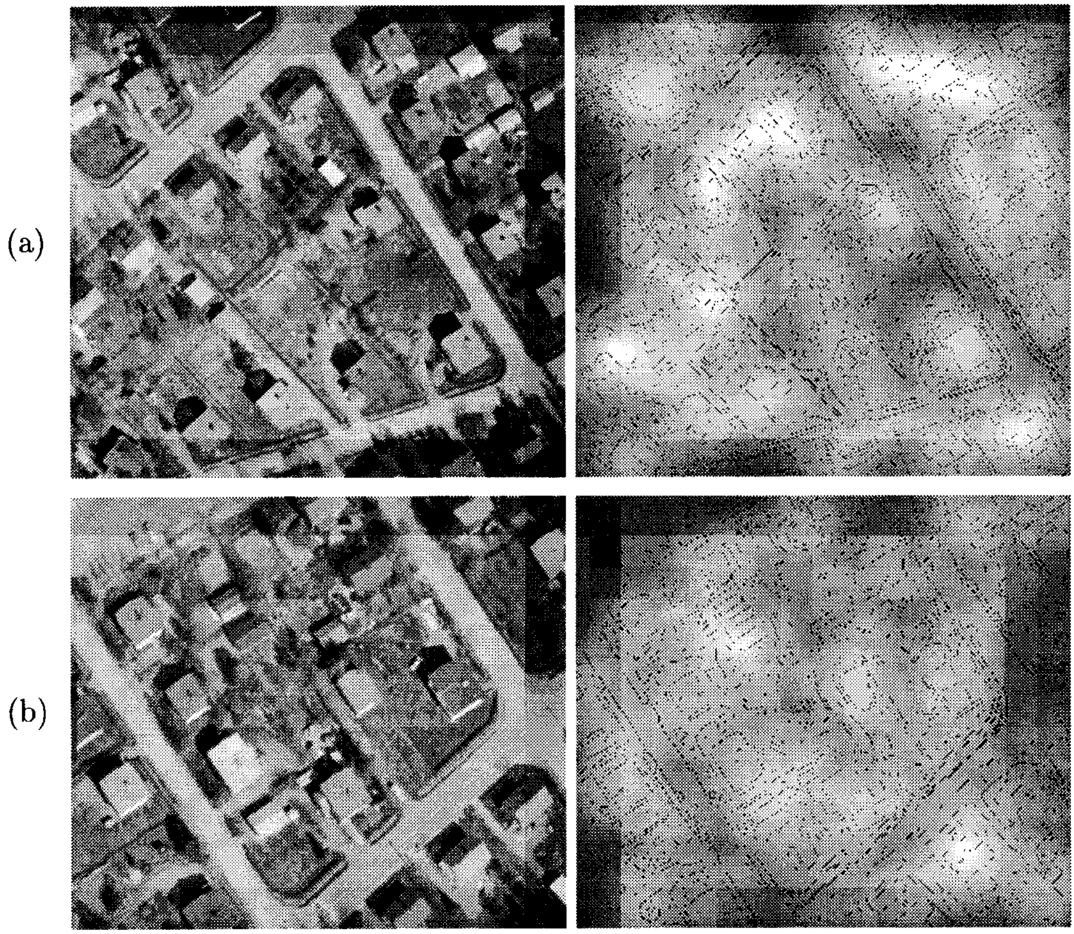
Moreover valuable extra knowledge could be extracted from the histograms for the template matching, e. g. main orientation of the buildings, axes to be traced first, etc.

We propose an employment of the approach we presented as the second level within a hierarchical concept with the following levels of abstraction:

1. general objects (not only suitable for building recognition) with homogeneous, smooth surfaces and continuous edges as already mentioned in section 2
2. general model of buildings (not distinguishing between special building types), visible in aerial images
3. domain specific aspect graphs of a finite set of spatial models of different building types

*Acknowledgements* – This work was done largely within the project "Semantic Modeling and Extraction of Spatial Objects from Images and Maps", especially in the subproject "Building Extraction" which is supported by the Deutsche Forschungsgemeinschaft (DFG). We thank the DFG for supporting our work, moreover the Institute for Photogrammetry, Bonn University, for the provision of aerial images and a feature extraction module. Furthermore we profitted from discussions with our cooperation partners, especially we thank André Fischer for his valuable comments on the early drafts of this paper.

Figure 8: Results of the sampling process for different exemplary input images.



# References

[Braun and Kolbe 95] C. Braun, T.H. Kolbe, F. Lang, W. Schickler, V. Steinhage, A.B. Cremers, W. Förstner and L. Plümer, *Models for Photogrammetric Building Reconstruction.* Computer & Graphics, 19(1): 109–118, 1995.

[Dickinson *et. al.* 92] S. Dickinson, A. Pentland, A. Rosenfeld, *Shape Recovery Using Distributed Aspect Matching.* IEEE T-PAMI, 14(2), 1992

[Fuchs and Förstner 95] C. Fuchs and W. Förstner, *Polymorphic Grouping for Image Segmentation.* 5th ICCV, Boston, 175–182, 1995.

[Lang and Schickler 93] F. Lang and W. Schickler, *Semiautomatische 3D-Gebäudeerfassung aus digitalen Bildern.* Zeitschrift für Photogrammetrie und Fernerkundung, 5: 193–200, 1993

[Kort *et. al.* 96] A. Kort, A. Pogoda, V. Steinhage, *Employing Aspect Hierarchies for Building Detection in Digital Images.* 18th DAGM-Symposium, Heidelberg, 217–224, Springer 1996.

[Weidner and Förstner 95] U. Weidner, W. Förstner, *Towards Automatic Building Extraction from High Resolution Digital Elevation Models.* ISPRS, 50(4): 38–49, 1995.