# 3D Building detection and modeling using a statistical approach

Matthieu Cord[1][2] and David Declercq[1]

[1] ETIS, ENSEA/UCP/CNRS UPRESA 8051, Cergy-Pontoise, France, [cord,declercq]@ensea.fr

[2] ESAT / PSI, KUL, Leuven, Belgium, mcord@esat.kuleuven.ac.be

October 27, 2000

**Abstract**

In this paper, we address the problem of building reconstruction in high resolution stereoscopic aerial imagery. We present a hierarchical strategy to detect and model buildings in urban sites, based on a global focusing process, followed by a local modeling. During the first step, we extract the building regions by exploiting to the full extent the depth information obtained with a new adaptive correlation stereo matching. In the modeling step, we propose a statistical approach, which is competitive to the sequential methods using segmentation and modeling. This parametric method is based on a multi-plane model of the data, interpreted as a mixture model. From a Bayesian point of view, the so-called augmentation of the model with indicator variables allows using stochastic algorithms to achieve both model parameter estimation and plane segmentation. We then report a Monte Carlo study of the performance of the stochastic algorithm on synthetic data, before diplaying results on real data.

## Introduction

Automatic techniques for building reconstruction are important for various application fields. Highly accurate and up-to-date 3D building cartographic information is essential in all investigations concerning telecommunication, urbanism, environmental modeling, *etc.*. In this context, man-made features extraction has been widely studied, especially in urban areas [1] [2].

Many approaches based on feature grouping have been developed to model simple shape buildings in mid-resolution aerial imagery [3] [4].

New data types (high resolution aerial images, such as a few centimeters per pixel; digital color camera images) are now available. That allows extracting more accurate three dimensional building descriptions out of urban site images. Actually, the quality of the result really depends on the type of scenes that we process. In dense urban sites, automatic recognition and reconstruction are very difficult tasks because of the complexity and the diversity of the scene objects. To overcome these difficulties, most techniques use a prior focusing step on regions of interest (ROI). The reconstruction may then be locally carried out. The region selection can be interactively done by a human operator [5] [6], or using GIS data (for instance, by projecting cadastrial maps [7] into the dataset, or by exploiting digital elevation models (DEM) [8] [9] [10]).

The problem is then to extract and reconstruct the buildings, region by region. Many image-based and DEM-based approaches have been carried out during the last years, using building databases, parametric and prismatic models, image segmentation and grouping. However, DEM are usually not dense and accurate enough to be efficiently used during the building reconstruction process.

We present in this paper a sequential building reconstruction method for high resolution monochromatic aerial image pairs. It is based on a global focusing step, followed by a statistical method for roof area modeling. Figure 1 shows our system scheme: we first compute a dense and accurate digital elevation model from the image pair; thanks to this depth information, buildings are detected as height blobs (after detection, height blobs are classified as vegetation or building) (section 1). The regions labeled building are 3D regions corresponding to multi-slope roof structures that we model separatly (section 2). This part is the main contribution of our work. The aim is to show how stochastic modeling approaches may be helpful for building reconstruction and 3-D roof recognition.

Sequential processings are usually made to achieve the multi-plane region research. First, a radiometric or range data segmentation is done, and then, a plane parameter identification is carried out [11]. Even in the case of techniques based on fitting rigid models to the data [12] [7], segmentation is usually used to make the matching problem linear. However, the segmentation step is difficult to adjust and provides some artifacts, as over- or under-segmentations. For example, when there is an under-segmented area, the data can not be correctly fitted by a plane during the second step. Because of the planar segmentation limitation, we have considered a completely different approach which can globally solve the identification of more than one plane in a single region and which can deal with the particular spatial distribution of range data (high level noise, no real definition of neighborhood, etc). The method is based on the interpretation of a multi-plane model for the data, as a mixing model. The problem becomes non linear and can be analyzed with a Bayesian inference, solved with stochastic algorithms.

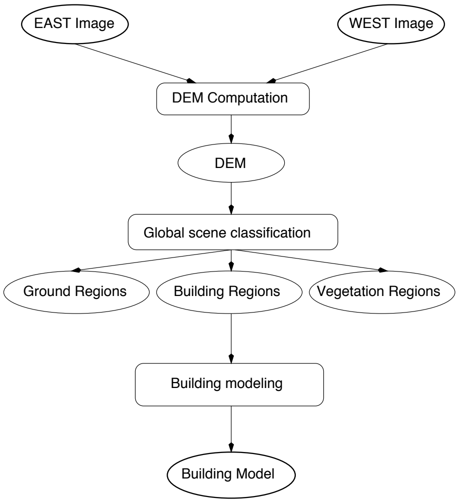


Figure 1: *Hierarchical scheme for building detection and modeling.*

# 1 Building detection

The building detection step aims at extracting areas corresponding to building structures. Because of the complexity of the urban high resolution data, monoscopic approaches (only using perceptual grouping and geometric models) are very limited. The 3D information can be very helpful for consistent and efficient grouping. Furthermore, we think that DEM properties such as density, reliability, accuracy, depth discontinuities localization are a key point for building detection and reconstruction. That is why we will take special care of the DEM calculation.

We propose on figure 2 the main steps of our global image processing. It is based on the DEM computation in order to segment images and select above-ground regions, which are then separated into building regions and vegetation regions.

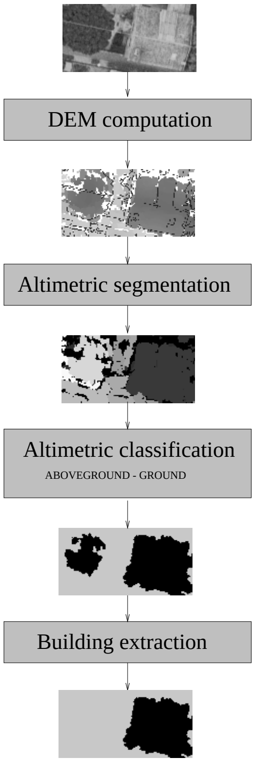


Figure 2: *Global analysis scheme.*

## 1.1 Digital elevation model computation

Area-based matching techniques usually provide dense disparity maps. Unfortunately, the fixed template size matching is not able to track narrow depth discontinuities. To overcome this problem, adaptive size templates are preferred [13] [14].

As presented in [15], we have introduced an adaptive shape window matching using contour image features to define the window shape: only the pixels on the same side of a contour and connected to the center pixel are used for the correlation score. The depth discontinuities are then preserved and precisely located.

However, this method is not efficient when the contour line is broken; in this case, the matching, using all the points of the starting square correlation window, is no more adaptive.

Anyway, for high resolution image matching, a large window size is necessary to take the poorly textured surfaces into account. It is thus interesting to use a template weighting function to reinforce the influence of the central pixels [16]. Usually, Gaussian weight functions are used to calculate the template weights.

We propose a new adaptive correlation scheme based on a cooperation between our adaptive shape technique and Gaussian weighting template correlation methods. The idea is to prevent the diffusion effects due to the contour discontinuities. We change the classical isotropic Gaussian weigthing to a geodesic weighting, propagating on all inter-contour area (figure 3).

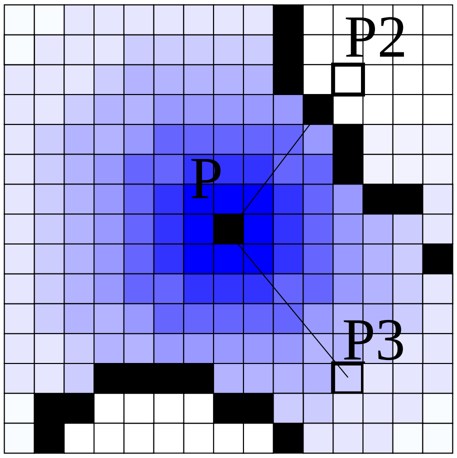


Figure 3: *Geodesic adaptive correlation scheme.* Each pixel of the window has a weight value depending on its distance from the central pixel $P$. The weight value of the pixel $P2$ is much smaller than the value of the pixel $P3$ because the $(P, P2)$ geodesic distance (without crossing contours) is quite larger than the $(P, P3)$ distance.

Our geodesic adaptive template $M_g^a$ is built for each pixel $(i, j)$ of image $I_1$ (slave image) as following:

- $\forall (x, y) \in \left[-\frac{w-1}{2}, \frac{w-1}{2}\right]^2$ :
- IF there is a 4-connected way between $I_1(i + x, j + y)$ and $I_1(i, j)$
- THEN

$$M_g^a(x, y) = exp\left(-\frac{(d_{geod}(I_1(i + x, j + y), I_1(i, j)))^2}{2\sigma^2}\right)$$

  with $d_{geod}(P_1, P_2)$ the shortest way between $P_1$ and $P_2$ without crossing any contours.

- ELSE

$$M_g^a(x, y) = 0$$

$w$ adjusts the starting template size and $\sigma$ the weighting influence.

**Remark**

*Multi-resolution matching strategy is most of the time used together with template-based matching techniques to overcome computational problems [17] [18]. Adaptive window shaping techniques have to be efficiently combined with multi-scale matching processing. Our multi-resolution strategy is based on the geodesic adaptive matching technique applied at each level of the multi-resolution process. It is coupled with a validation process to avoid the matching error propagation. We use a symmetric validation based on the two way filtering technique [19].* ■

## 1.2 Building extraction

Building classification may be obtained from graylevel and texture analysis, colour analysis, or 3D local analysis [5] [20] [21] [7].

Due to the great accuracy and reliability of our DEM, we have chosen to extract building areas using a 3D analysis.

The algorithm described in section 1.1 provides dense, accurate and discontinuity preserving DEM. Thus it is possible to obtain a segmentation of this DEM by a classical region growing algorithm. We define the local pixel aggregation with an altitude criterion: if the altitude difference between two neighboring pixels is less than a threshold $t_{seg}$, they are merged in the same region. Pixels having no altitude (the ones having no corresponding point in both images) are not considered in the computation. The threshold $t_{seg}$ is chosen so as to control the maximal possible slope of homogeneous regions.

We then select the above-ground regions as the ones being sufficiently high upon the ground. We compute the ground altitude as the mean altitude of the lowest region, and all the regions having a mean altitude higher than a threshold are considered as above-ground. The other regions are labeled ground.

**Remark**

*After this first classification, we compute an adjacency graph of altimetric regions, and regions are* *merged relying on two criteria: neighboring above-ground regions are merged if they have the same mean altitude, and some small isolated regions are eliminated from the above-ground description [22].*

■

Finally, we make a building region extraction from the above-ground regions using the repartition of the normal's directions inside the region: for each point of the considered region, we compute from a $5 \times 5$ neighborhood the normal of the local 3D surface (using a mean squares estimation). We then consider the histogram of normals on the region. For a vegetation like region, normals are sparsed and there is no privileged direction. This is caracterised by a flat histogram. On the opposite, for a building region, there are some privileged directions, and corresponding peaks appear in the histogram shape. Thus, a simple thresholding of the normal histograms enables us to decide whether the region is building or vegetation (see [23] for more details).

# 2 Building modeling

At the end of the global scene analysis step, buildings have been detected. We propose in this section to model building roofs. Thanks to the high image resolution, it becomes possible to separate the different parts of a building with a multi-roof aggregated structure. Instead of the classical methods making segmentation by plane, or grouping primitives by plane, we carry out a non linear optimization method, which enables us to avoid the segmentation step [24].

First, we present the structure of the model, and its implications to the roof identification and classification. We then write the posterior densities of the parameters that we want to identify and we end with the description of the chosen Bayesian sampler: a stochastic EM algorithm.

## 2.1 Model for multi-slope roofs

### Notations

*We note $\mathcal{R}$ the building region we are working on. It contains $N$ points, that are supposed to be distributed as $p$ planes in the Euclidean space. $X_{1\to N} = (x_{1\to N}, y_{1\to N}, z_{1\to N})$ are the vectors which contain the spatial coordinates of each data sample $X_t$, and $\mathbf{a} = (a_1, \ldots, a_p)$, $\mathbf{b} = (b_1, \ldots, b_p)$,*

$\mathbf{c} = (c_1, \ldots, c_p)$ *are the vectors of plane parameters.* ■

The tridimensionnal model of $N$ points distributed as $p$ planes is defined as:

$$\sum_{k=1}^{p} (a_k x_t + b_k y_t + z_t - c_k) \, \mathbb{1}_{X_t \in \mathcal{P}_k} = \varepsilon_t \quad \forall \, t \in \{1, \ldots N\} \tag{1}$$

where $\mathbb{1}_S$ is the indicator function of set $S$, and $a_k x + b_k y + z - c_k = 0$ is the equation of the plane $\mathcal{P}_k$. $\varepsilon_t$ is a Gaussian noise with zero mean and variance $\sigma_\varepsilon^2$ and which represents the modeling error, that is the distance between a sample and the model for all $t$.

Such a model can be statistically interpreted as switching model, which is a particular case of the so-called *data augmentation models* [25, 26]. The principle of those models is that there exists a hidden process, generally a hidden Markov chain, which contains some important information for the identification of the model. It is very interesting to complete the data description by adding variables which describe this hidden process. For example, such variables can represent missing data as well as explanatory variables in prediction models. When they are introduced in the densities (likelihood, posterior, etc) of the model, they make them easier to work with and thereby simplify the model estimation.

This type of augmented model is usually considered in a Bayesian framework, because the Bayes inversion formula allows an efficient use of the augmented stochastic variables, as we will see below. We will then try to identify our model (1) with a Bayesian inference. The first step is to express the global posterior density, given by:

$$p(\boldsymbol{\theta} | X_{1 \to N}) = \int_{\tilde{\mathcal{X}}} p(\boldsymbol{\theta} | X_{1 \to N}, \tilde{\mathbf{X}}) \, p(\tilde{\mathbf{X}} | X_{1 \to N}) \, d\tilde{\mathbf{X}} \tag{2}$$

with $\tilde{\mathbf{X}}$ being the augmented variables.

In our case, we introduce one augmented variable $\tilde{\mathbf{X}}_t$ for each observation $X_t$, which describes the belonging of the data to the different planes $\mathcal{P}_k$ of the model. We will refer to $\tilde{\mathbf{X}}_t$ as *indicator variables* or *state variables* because when augmented, the model has a Markovian representation (or a state space representation). This augmented state variable is a vector of size $p$ which takes its values in the discrete set $\{0, 1\}^p$. Each component $k$ of the vector $\tilde{\mathbf{X}}_t$ is therefore defined by the probabilities of the corresponding data to belong to the plane $\mathcal{P}_k$.

The next step is then to deal with the posterior density in (2) to build estimators of the parameters $\boldsymbol{\theta}$. Classical Bayesian estimators are the maximum *a posteriori* (MAP)

$$\hat{\boldsymbol{\theta}} = arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|X_{1\to N})$$

or the expectation *a posteriori*

$$\hat{\boldsymbol{\theta}} = \mathbb{E}\left[\boldsymbol{\theta}|X_{1\to N}\right]$$

As one can see in (2), the problem is that the integration has to be done with respect to a huge number of integrands because there is a vector $\tilde{\mathbf{X}}_t$ for each pixel $X_t$ in the region. In these types of Bayesian inferences, we must turn to stochastic algorithms, which aim to produce - by sampling - data which are asymptotically distributed as $p(\boldsymbol{\theta}|X_{1\to N})$ [27].

A well known possible stochastic algorithm, called the Gibbs sampler [28], relies on iterative sampling to build such a process $(\boldsymbol{\theta}^{(i)})$:

$$\begin{array}{ll} 1. & \tilde{\mathbf{X}}^{(i+1)} \sim p(\tilde{\mathbf{X}}|\boldsymbol{\theta}^{(i)}, X_{1\to N}) \\ 2. & \boldsymbol{\theta}^{(i+1)} \sim p(\boldsymbol{\theta}|\tilde{\mathbf{X}}^{(i+1)}, X_{1\to N}) \end{array} \tag{3}$$

**Remark**

*Tanner and Wong [29] have proposed a similar iterative scheme inspired by the EM algorithm [30], but the convergence has been found much slower [27].* ■

We have adopted this sequential scheme to build MAP (maximum a posteriori) estimators of the parameters $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ of our multi-slope model (1). The estimators are obtained with a stochastic version of the well-known EM algorithm, which maximizes the posterior density $p(\boldsymbol{\theta}|X_{1\to N})$. In the next two sections, we describe each step of the process: *(i)* first we simulate the augmented variables $\tilde{\mathbf{X}}_{1\to N}$ (imputation step), *(ii)* and then we sample data from the posterior densities of the parameters and hyper-parameters (posterior step).

## 2.2 Imputation step

From the definition of the indicator variables, we have

$$Prob\left(\tilde{\mathbf{X}}_t[k] = 1\right) = Prob\left(X_t \in \mathcal{P}_k\right)$$

where $\tilde{\mathbf{X}}_t[k]$ is the $k^{th}$ component in the random vector $\tilde{\mathbf{X}}_t$.

Knowing the previous estimates of the parameters at the previous step $\boldsymbol{\theta}^{(i)}$, defining $p$ estimated planes ($\mathcal{P}_k^{(i)}$, $k = 1..p$), the probability of a sample data to belong to the plane $\mathcal{P}_k^{(i)}$ is Gaussian, due to the Gaussianity of the error:

$$\forall\ k = 1 \rightarrow p, \qquad Prob\left(X_t \in \mathcal{P}_k^{(i)}\right) \propto \exp\left(-\frac{\left(a_k^{(i)} x_t + b_k^{(i)} y_t + z_t - c_k^{(i)}\right)^2}{2\left(\sigma_\varepsilon^2\right)^{(i)}}\right)$$

We sample the new augmented variables $\tilde{\mathbf{X}}_t^{(i+1)}$ from the density $p(\tilde{\mathbf{X}}|\boldsymbol{\theta}^{(i)}, \mathbf{X}_{1\rightarrow N})$. This density is taken as a multinomial distribution $\mathcal{M}$, usually used in mixing distribution problems [31].

$$\tilde{\mathbf{X}}_t^{(i+1)} \sim p\left(\tilde{\mathbf{X}}_t | X_t, \mathbf{a}^{(i)}, \mathbf{b}^{(i)}, \mathbf{c}^{(i)}\right) = \mathcal{M}(1; \alpha_1, \ldots, \alpha_p) \tag{4}$$

with

$$\left| \begin{array}{l} \alpha_k \propto Prob\left(X_t \in \mathcal{P}_k^{(i)}\right) \quad \forall\ k = 1 \ldots p \\ \displaystyle\sum_{k=1}^{p} \tilde{\mathbf{X}}_t^{(i+1)}[k] = 1 \end{array} \right. \tag{5}$$

A random variable sampled from this density is then a $p$ variate vector with only one nonzero component. The key point of the global augmented model is the derivation of the weights $\alpha_k$. For more details on indicator variables in mixture or switching models, we refer to [31].

## 2.3 Posterior step

The second step of our process (given in eq. 3) deals with the sampling of the conditional posterior densities of the parameters. We start from the global posterior density derived from the classical Bayes rule:

$$p\left(\mathbf{a}, \mathbf{b}, \mathbf{c}, \sigma_\varepsilon^2 | X_{1\rightarrow N}, \tilde{\mathbf{X}}_{1\rightarrow N}\right) \propto p\left(X_{1\rightarrow N} | \mathbf{a}, \mathbf{b}, \mathbf{c}, \sigma_\varepsilon^2, \tilde{\mathbf{X}}_{1\rightarrow N}\right)\ \pi\left(\mathbf{a}, \mathbf{b}, \mathbf{c}\right)\ \pi\left(\sigma_\varepsilon^2\right) \tag{6}$$

where $\pi\left(\mathbf{a}, \mathbf{b}, \mathbf{c}\right)$ and $\pi\left(\sigma_\varepsilon^2\right)$[1] are respectively the prior distributions of the plane coefficients and the variance of the error.

Markov field approaches are often used in image processing [32] to express spatial dependence on data. Unfortunately, the range data used for building reconstruction are not everywhere dense and

[1] the complete description of the model includes also the hyper-parameters, which is in our case the variance of the error. This variance has to be estimated (optimized) too, and therefore must appear in the full posterior density.

above all not regularly sampled. Obtained from the correlation matching, homogeneous areas provide very sparse range data whereas textured areas are well matched and provide dense data. We have then prefered not to take spatial local dependences into account, and we have considered our data as spatially independent data. From that independence, the global posterior density (6) can be developed as follows:

$$p\left(\mathbf{a},\mathbf{b},\mathbf{c},\sigma_{\varepsilon}^{2}|X_{1\rightarrow N},\tilde{\mathbf{X}}_{1\rightarrow N}\right)\propto\prod_{t=1}^{N}p\left(X_{t}|\mathbf{a},\mathbf{b},\mathbf{c},\sigma_{\varepsilon}^{2},\tilde{\mathbf{X}}_{\mathbf{t}}\right)\pi\left(\mathbf{a},\mathbf{b},\mathbf{c}\right)\ \pi\left(\sigma_{\varepsilon}^{2}\right)\tag{7}$$

Actually, because the modeling error is considered as Gaussian, the conditional likelihood of a sample data is Gaussian (cf. eq. (1)):

$$p\left(X_{t}|\mathbf{a},\mathbf{b},\mathbf{c},\tilde{\mathbf{X}}_{t}\right)=\frac{1}{\sqrt{2\pi\sigma_{\varepsilon}^{2}}}\exp\left(-\frac{\left(a_{\tilde{\mathbf{X}}_{t}}x_{t}+b_{\tilde{\mathbf{X}}_{t}}y_{t}+z_{t}-c_{\tilde{\mathbf{X}}_{t}}\right)^{2}}{2\sigma_{\varepsilon}^{2}}\right)\tag{8}$$

with the notation $a_{\tilde{\mathbf{X}}_t} = a_k$ if $\tilde{\mathbf{X}}_t[k] = 1$ (remember from the previous section, that only one component of $\tilde{\mathbf{X}}_t$ is nonzero).

In Bayesian framework, it is convenient to make use of conjugate prior, that is prior which doesn't change the density family when multiplied by the likelihood term (8). This is especially convenient when this augmented likelihood in terms of the parameters belongs to the exponential family, which is the present case. The conjugate priors for the parameters are normal:

$$\pi(a_k)=\mathcal{N}(0,\sigma_{prior}^{2})\qquad\forall k\in\{1\ldots p\}\tag{9}$$

$\mathbf{b}$ and $\mathbf{c}$ have exactly the same prior. While taking a large value for $\sigma^2_{prior}$, the prior density is still conjugate and becomes nearly noninformative. A noninformative strategy is very interesting here since we do not want to privilegiate special range values for the plane parameters $(\mathbf{a},\mathbf{b},\mathbf{c})$.

The variance of the error has an inverse Gamma conjugate prior:

$$\pi(\sigma_{\varepsilon}^{2})=\mathcal{IG}(\lambda_{prior},\tau_{prior})\tag{10}$$

where the density function of the law $\mathcal{IG}(\lambda,\tau)$ is written as follows (using the classical Gamma function $\Gamma$):

$$\mathcal{IG}(x|\lambda,\beta)=\frac{\beta^{\lambda}}{\Gamma(\lambda)}\frac{e^{-\frac{\beta}{x}}}{x^{\lambda+1}}\mathbb{1}_{[0,+\infty[}(x)$$

A nearly noninformative behavior corresponds to $\lambda_{prior} >> \tau_{prior}$.

The full conditional posterior densities are then expressed according to eq. (7) and eq. (8):

- Plane coefficients $(\mathbf{a}, \mathbf{b}, \mathbf{c})$

$$\begin{aligned} & p\left(a_k | X_{1\to N}, \tilde{\mathbf{X}}_{1\to N}, b_k, c_k, \sigma_\varepsilon^2\right) \\ & \quad \propto p\left(X_{1\to N} | \tilde{\mathbf{X}}_{1\to N}, a_k, b_k, c_k, \sigma_\varepsilon^2\right) \ \pi\left(a_k\right) \\ & \quad \propto \prod_{t=1}^{N} \exp\left( -\frac{\left(a_{\tilde{\mathbf{X}}_t} x_t + b_{\tilde{\mathbf{X}}_t} y_t + z_t - c_{\tilde{\mathbf{X}}_t}\right)^2}{2\sigma_\varepsilon^2} \right) \pi\left(a_k\right) \\ & \quad \propto \mathcal{N}\left(m_{a_k}, \sigma_{a_k}^2\right) \end{aligned} \tag{11}$$

where the mean and the variance are given by

$$\sigma_{a_k}^2 = \left( \frac{1}{\sigma_{prior}^2} + \frac{\sum_{t=1}^{N} x_t^2 \tilde{\mathbf{X}}_t[k]}{\sigma_\varepsilon^2} \right)^{-1} \tag{12}$$

$$m_{a_k} = -\frac{\sigma_{a_k}^2}{\sigma_\varepsilon^2} \sum_{t=1}^{N} x_t \left(b_k y_t + z_t - c_k\right) \tilde{\mathbf{X}}_t[k] \tag{13}$$

**b** and **c** have the same kind of distribution and are straightforwardly deduced from (11) - (13).

- Noise variance

$$\begin{aligned} & p\left(\sigma_\varepsilon^2 | X_{1\to N}, \tilde{\mathbf{X}}_{1\to N}, a_k, b_k, c_k\right) \\ & \quad \propto p\left(X_{1\to N} | \tilde{\mathbf{X}}_{1\to N}, a_k, b_k, c_k, \sigma_\varepsilon^2\right) \ \pi\left(\sigma_\varepsilon^2\right) \\ & \quad \propto \mathcal{IG}\left(\lambda_\varepsilon, \tau_\varepsilon\right) \end{aligned} \tag{14}$$

with

$$\left| \begin{array}{l} \lambda_\varepsilon = \dfrac{N}{2} - 1 + \lambda_{prior} \\ \tau_\varepsilon = \dfrac{1}{2} \sum_{t=1}^{N} \left(a_{\tilde{\mathbf{X}}_t} x_t + b_{\tilde{\mathbf{X}}_t} y_t + z_t - c_{\tilde{\mathbf{X}}_t}\right)^2 + \tau_{prior} \end{array} \right. \tag{15}$$

## 2.4 Stochastic algorithm for model identification

A Bayesian sampler which will provide parameter estimators starts with the imputation of the indicator variables $\tilde{\mathbf{X}}_t$ according to their multinomial distribution (4), and then makes use of the posterior densities above described to sample the parameters. A Gibbs sampler could be a relevant choice because one can easily sample all the posterior densities. However, these densities belong to the exponential family and their maximization does not require a lot of efforts. We have then chosen a Stochastic EM algorithm [33] which consists in two steps: *(i)* first, the Expectation step is achieved by stochastic imputation, which provides an estimator of the expectation of the posterior log-density, *(ii)* the Maximization step is the same as in the classical EM algorithm, and computes the maximum of the conditional posterior densities.

This algorithm samples a Markov chain of the plane parameters, which converges on its stationary density under weak conditions (see [25] for instance). Another attractive advantage of this algorithm is its low complexity, which is of order $\mathcal{O}(pN)$.

*SEM Algorithm for roof reconstruction: loop for iteration (i) to (i+1)*

```
Start with $\mathbf{a}^{(i)}$, $\mathbf{b}^{(i)}$, $\mathbf{c}^{(i)}$, $\left(\sigma_\varepsilon^2\right)^{(i)}$,

 1. Imputation step:  indicator variables sampling

    • compute $\forall\ k = 1 \to p$,
          $\beta_k = \exp\left(-\frac{\left(a_k^{(i)} x_t + b_k^{(i)} y_t + z_t - c_k^{(i)}\right)^2}{2\left(\sigma_\varepsilon^2\right)^{(i)}}\right)$

    • normalize the weights,
          $\alpha_k = \frac{\beta_k}{\sum_{l=1}^{p} \beta_l} \qquad \forall\ k = 1 \to p$

    • sample $\tilde{\mathbf{X}}_t^{(i+1)} \sim \mathcal{M}\left(1; \alpha_1, \dots, \alpha_p\right)$

 2. Posterior step:  maximization (see Eq.  (13))

    • $\forall\ k = 1 \to p$    $a_k^{(i+1)} = m_{a_k}\left(\tilde{\mathbf{X}}_{1\to N}^{(i+1)}, b_k^{(i)}, c_k^{(i)}, \left(\sigma_\varepsilon^2\right)^{(i)}\right)$
    • $\forall\ k = 1 \to p$    $b_k^{(i+1)} = m_{b_k}\left(\tilde{\mathbf{X}}_{1\to N}^{(i+1)}, a_k^{(i+1)}, c_k^{(i)}, \left(\sigma_\varepsilon^2\right)^{(i)}\right)$
    • $\forall\ k = 1 \to p$    $c_k^{(i+1)} = m_{c_k}\left(\tilde{\mathbf{X}}_{1\to N}^{(i+1)}, a_k^{(i+1)}, b_k^{(i+1)}, \left(\sigma_\varepsilon^2\right)^{(i)}\right)$
    • $\left(\sigma_\varepsilon^2\right)^{(i+1)} = \frac{1}{N + 2\lambda_{Prior} + 2}\left(\sum_{t=1}^{N}\left(a_{\tilde{\boldsymbol{X}}_t}^{(i+1)} x_t + b_{\tilde{\boldsymbol{X}}_t}^{(i+1)} y_t + z_t - c_{\tilde{\boldsymbol{X}}_t}^{(i+1)}\right)^2 + 2\tau_{Prior}\right)$
```

|  |  | $\sigma^2$ (in meters) |  |  |  |
|---|---|---|---|---|---|
|  |  | 0.04 | 0.12 | 0.25 | 0.5 |
| N | 500 | 100% | 100% | 100% | 92% |
|  | 1000 | 100% | 100% | 100% | 96% |

Table 1: Monte Carlo results of the proposed algorithm on synthetic data with 2 planes.

|  |  | $\sigma^2$ (in meters) |  |  |  |
|---|---|---|---|---|---|
|  |  | 0.04 | 0.12 | 0.25 | 0.5 |
| N | 500 | 93% | 93% | 89% | 81% |
|  | 1000 | 95% | 95% | 93% | 89% |
|  | 3000 | 98% | 97% | 97% | 93% |

Table 2: Monte Carlo results of the proposed algorithm on synthetic data with 3 planes.

# 3 Simulations

## 3.1 Results on synthetic data

In order to demonstrate the validity of the proposed stochastic algorithm before applying it to real data, we have made a Monte Carlo study of its performance. For each Monte Carlo experiment, we have randomly generated $N$ samples spatially distributed as a mixing of 2 or 3 planes. The samples have been corrupted by an additive white Gaussian noise with variance $\sigma^2$.

We have reported in tables 1 and 2, the percentage of good plane detection for several sample sizes $N$ and several noise powers $\sigma^2$. We decided that the planes were successfully detected when the mean square error between the true parameters and the estimated ones was less than a threshold, chosen empirically. Note that a noise power of approximately 4 centimeters (the noise with least power in our table) corresponds to a real data case.

The results contained in tables 1 and 2 clearly demonstrate the very good behaviour of our algorithm. Moreover, this study shows that the algorithm can operate at noise powers far greater than observed real noise. As it was expected, the detection percentage grows with the number of observed points and when the noise power decreases.

### 3.2 Results on real data

We turn now to real scene results. We first explain the pre-processing that led us to the 3D samples for the modeling process. The results of our algorithm are then compared to a man-made IGN database.

The test images are stereo pairs of 8 centimeters resolution supplied by the I.G.N. (Institut Géographique National) and cover the french city of Colombes. One of the stereo pairs is provided in figure 4 ($1000 \times 600$ pixels).

We make an edge detection using a Canny-Deriche edge detector [34, 35][2], and, thanks to the contour map, we compute the adaptive geodesic template stereo matching. The starting window size $w$ is fixed to $15 \times 15$ which doesn't cover more than 1.44 square meters. $\sigma$ is chosen in such a way that the weight of the window corner point is two times smaller than the weight of the center point.

After the DEM computation, we make the altimetric segmentation with the threshold $t_{seg} = 20\ cm$ ( § 1.2). The above-ground regions are those which are at least 5 meters (about one stair) above the ground altitude. All of them have privileged normal directions and are classified as building (fig. 5); in this part of the whole scene, there is no above-ground vegetation region[3]. We have tested our matching and 3D building detection scheme on many stereo pairs and we have made an evaluation thanks to a database reference (also supplied by I.G.N.). It results that, regarding the roofs, on the base of about 500.000 pixels treated, 95% of pixels are matched, 96% of the matched pixels are reliable (viz the reconstructed corresponding 3D point is less than 50 $cm$ away from the reference), and the altimetric map is very accurate (only 15 $cm$ for the standard deviation of the error on z-value) [23]. As far as the processing time is concerned, our matching is no more time consuming than a classical cross-correlation scheme, because the adaptive template computation time is weak in comparison with the time to compute the curve of the similarity scores.

For the modeling, we therefore work on 3D data sets corresponding to each building region. A roof example with 2 slopes is depicted in figure 6.a.

We show on figure 7 the Markov chains that were generated with our stochastic algorithm. Each

[2] We have adjusted the derivator filter thanks to limit values introduced in [36] in order to detect close contours without error localization.

[3] In [23] many classifications with vegetation regions are presented.

column represents the three coefficients $(a, b, c)$ of a plane in the scene. We can see that the third plane rapidely converges, and that the first two wait approximately $40 - 50$ iterations to achieve convergence. This is the number of iterations needed for the sampling scheme to catch the *a posteriori* mode of the model distribution. If $K$ iterations are necessary to achieve convergence, the complexity of our algorithm is actually in $\mathcal{O}(KNp)$. The problem of choosing $K$ in practice - that is when we decide that the algorithm has achieved convergence - is a real issue. We have decided to implement an intuitive but not optimal scheme : we stop the algorithm when the variance of the last 20 generated parameters is less than a treshold. The behaviour of the algorithm is also depicted in figure 6 in 3D form.

The order of the model, which is the number of planes, is choosen *a priori*. We made the assumption that the order $P$ was known because the data sets that we work with in practice are often from scenes with a small number of roof slopes (2 or 3). The simplest strategy to estimate $P$ is then to run the algorithm for several values of $P$, and choose the model that exhibits the greatest likelihood (or any other model selection criterion: Akaïke, etc). It is possible to consider the order of the model as a random variable that we need to estimate, but this kind of model would lead to more complicated sampling schemes [37].

We have tested our building modeling scheme on many regions of different stereo images. For each 3-D region, planes are generally well detected and adjusted. Moreover, we have compared our results to the man-made I.G.N. database: on the 30 tested building models, the mean square error on z-value never exceeds 20 $cm$, that confirms the accuracy of our modeling. We display on figure 8 the result of the stochastic algorithm for one-building region. We keep the planimetric coordonnates of the reference and we have computed the $z$ value using our estimated planes. On figure 9 the results for the whole scene of the figure 4 are displayed. There is one single roof, three two slope roofs, and one three slope roof. On these data, there is no problem to build efficient models close to the reference, even in the case of the three slope roofs (the Markov chains obtained from our stochastic algorithm in this region are those displayed on figure 7). The final values of the state variables give us an additional result: an image segmentation may be carried out using the state variables.

When the roof structure becomes very complicated, difficulties may appear in finding the right planes. It might happen if there are more than three slopes in the scene, or if the scene contains

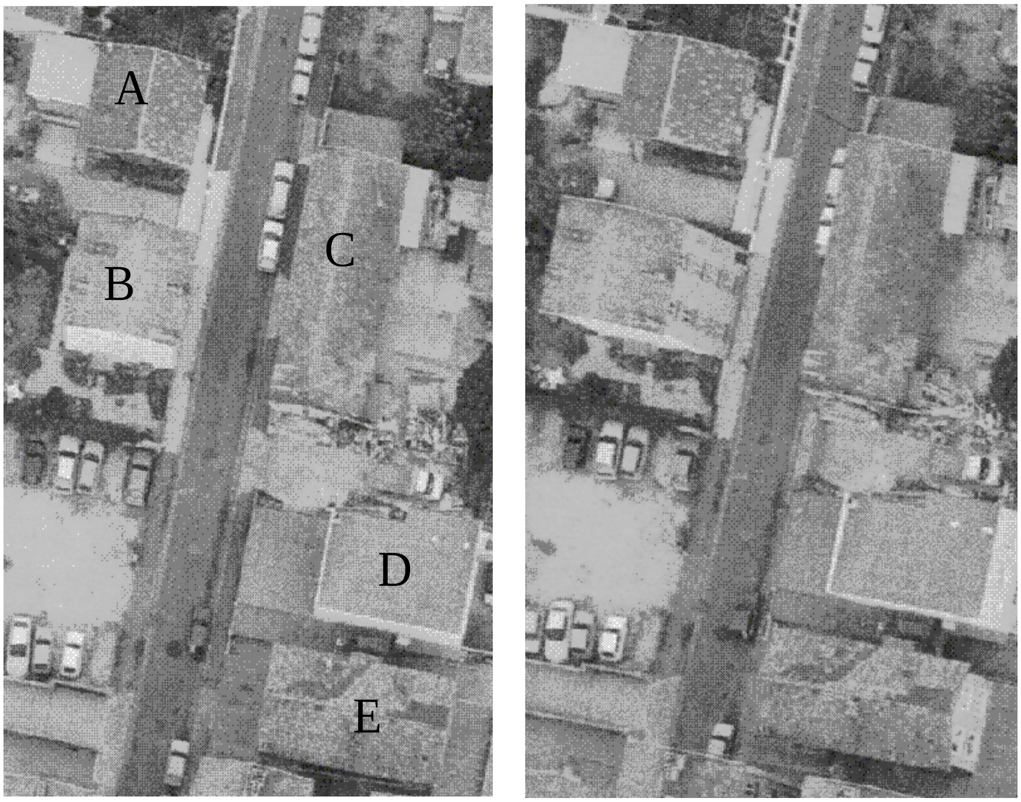


Figure 4: *High resolution digitized aerial stereo-photographs.* Building regions which have been detected are noted from A to E.

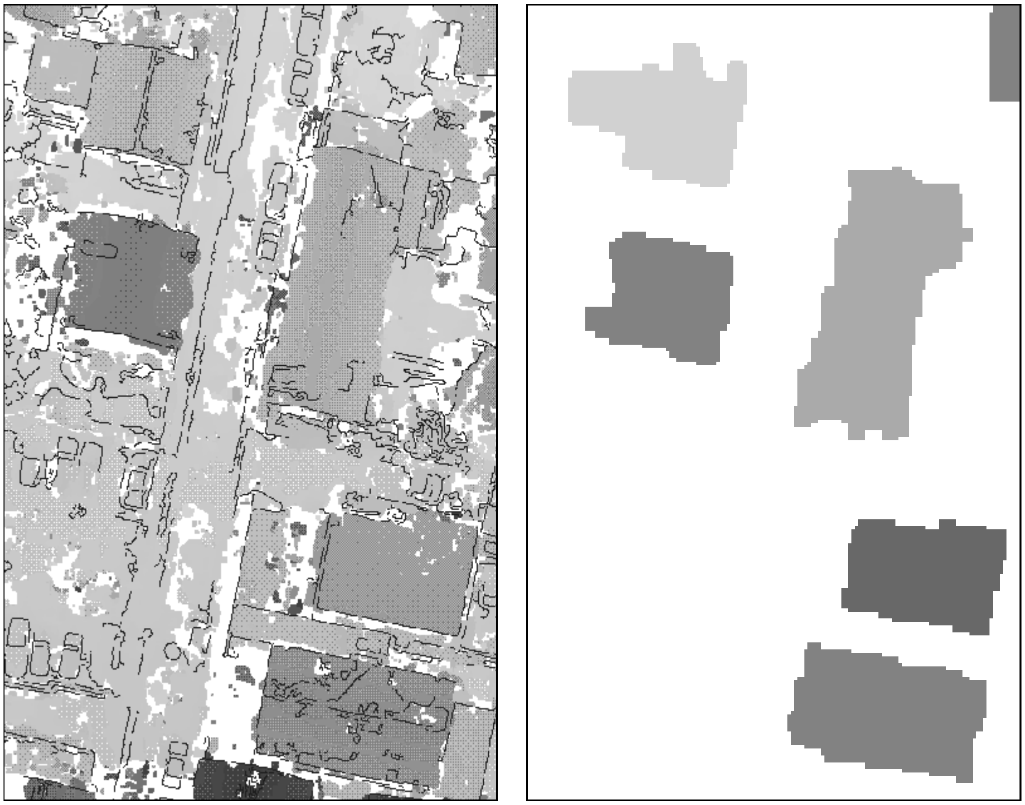
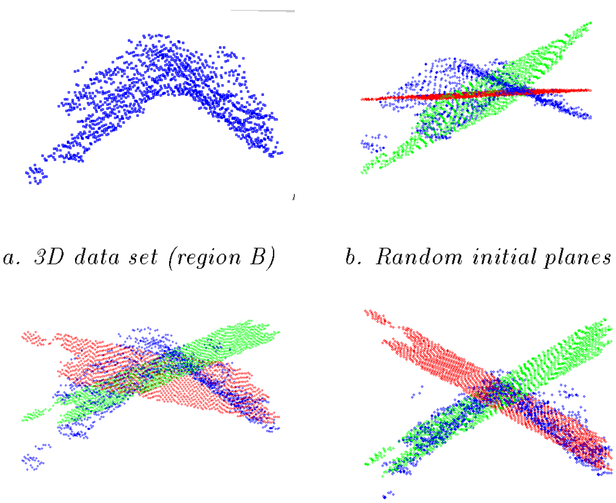


Figure 5: *Building detection.* We display on the left figure the result of the matching scheme and on the right figure the five building regions which have been detected by the global focusing process.



*a. 3D data set (region B)*

*b. Random initial planes*

*c. 10th iteration planes*

*d. 30th iteration planes*

Figure 6: *Algorithm convergence on the building region B of the figure 4.* It is a roof region with two slopes. After only thirty iterations, the convergence is achieved.

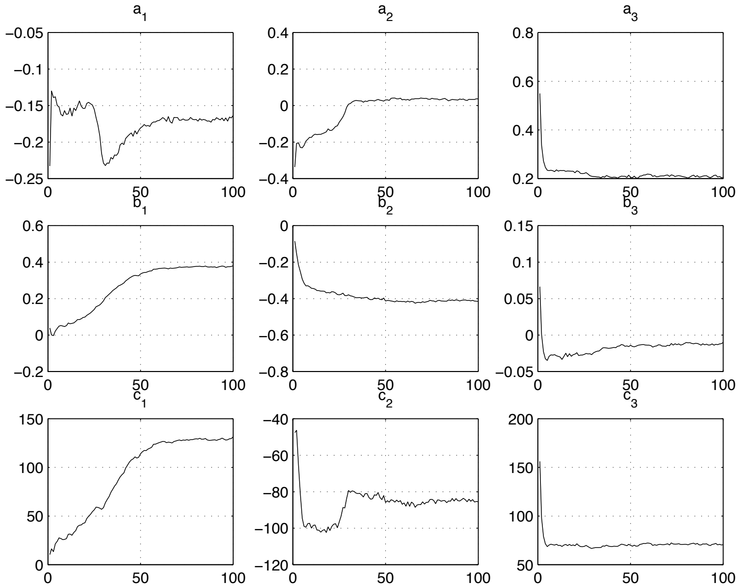


Figure 7: *SEM Markov chains convergence for the plane coefficients corresponding to the region A (fig. 4).*

artefacts such as chimneys. In that case, it could be helpful to introduce a sequential process which enables us to find first the largest plane, to remove the data belonging to it, and to start again the process on the remaining data.

# Conclusion

We have described an automatic multi-slope roof building detection and modeling from high resolution digitized aerial stereo-photographs. Our method is a hierarchical technique based on a global building detection step and a local modeling.

The first part of our process concerns the focalisation step: we do stereo computation and a 3D data analysis to isolate regions of interest, viz as far as we are concerned, the building regions. The process is based on a new efficient digital elevation model computation. That has allowed us to obtain very accurate and dense data, while preserving the depth discontinuities. Due to these 3D map characteristics, we have carried out an altimetric segmentation of the scene, and we have made an efficient building detection.

As for the second part, the building modeling, our method uses a stochastic optimization technique.

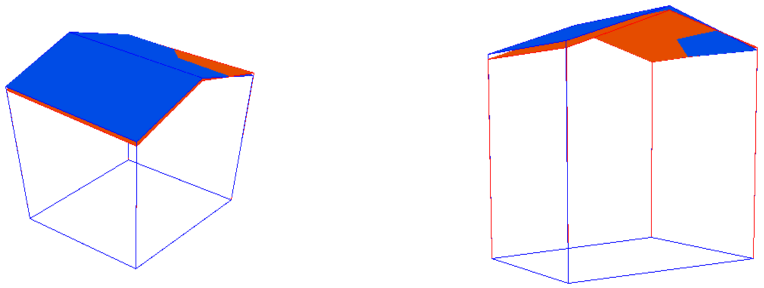


Figure 8: *Perspective view of the 3D model corresponding to the building noted B.* The building reconstruction is carried out using the plane parameters provided by our stochastic algorithm, and the result is compared with the 3D data of the reference (light color).

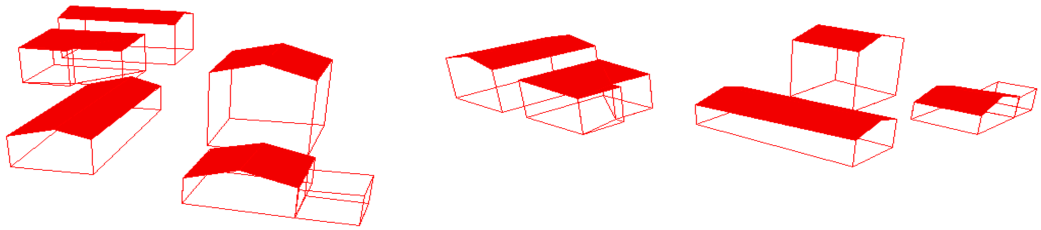


Figure 9: *Two perspective views of the 3D models corresponding to the five buildings of the fig. 4.* The building reconstruction is carried out using the plane parameters provided by our stochastic algorithm.

We have developed a data model to express any 3D multi-slope roof distribution. The starting model has been completed by augmented variables dynamically expressing the belonging to the different slopes. This method works without any prior knowledge on the shape of the roof except that it is composed by planes. When classic matching model methods are *de facto* limited, our modeling can accept any slope roof configuration. There is no restriction on the distribution of the different planes. Statistical approaches are decisive to process very complex non linear signals without segmentation, and we believe that our modeling deriving from stochastic models is an improvement to building recognition and shape reconstruction in urban sites. Furthermore, the principle of dynamical stochastic sampling coupled with the parameter up-dating could be also applied to select a type of regions from a segmentation.

This system is complete and well-suited to process in dense urban areas, which are usually the most difficult areas.

Further developments regard an extention of the modeling method using a rejection class which should detect the outliers provided by artifacts in the detected regions (discontinuities on the roofs, bad region detection, etc). That will enable us to take small building structures such as chimney tops into account.

## References

[1] A. Grün, O. Kübler, and P. Aggouris, editors. *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona (CH), 1995. Birkhäuser Verlag.

[2] A. Grün, O. Kübler, and P. Aggouris, editors. *Automatic Extraction of Man-Made Objects from Aerial and Space Images II*, Ascona (CH), 1997. Birkhäuser Verlag.

[3] R. Mohan and R. Nevatia. Using Perceptual Organization to Extract 3-D Structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(11):1121–1138, 1989.

[4] C.O. Jaynes, F.R. Stolle, H. Schultz, R.T. Collins, A.R. Hanson, and E.M. Riseman. Three-Dimensional Grouping and Information Fusion for Site Modeling from Aerial Images. In *Proc. Arpa Image Understanding Workshop*, pages 479–490, Palm Springs, 1996.

[5] O. Henricsson and E. Baltsavias. 3-D Building Reconstruction with ARUBA: A Qualitative and Quantitative Evaluation. In A. Grün, E. Baltsavias, and O. Henricsson, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images II*. Birkhäuser Verlag, Basel, 1997.

[6] T. Moons, D. Frère, J. Vandekerckhove, and L. Van Gool. Automatic modeling and 3d reconstruction of urban house roofs from high resolution aerial imagery. In *Proc. of ECCV*, June 1998.

[7] N. Haala, C. Brenner, and C. Statter. An integrated system for urban model generation. In *ISPRS Commission II Symposium*, Cambridge, England, July 1998.

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

[9] C. Baillard, O. Dissard, O. Jamet, and H. Maître. Extraction and textural characterization of above ground areas from aerial stereo pairs: a quality assessment. *Photogrammetry and Remote Sensing*, 53(2):130–141, 1998.

[10] C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 559–565, June 1999.

[11] H. Maître and W. Luo. Using models to improve stereo reconstruction. *I.E.E.E. Trans. on Pattern Analysis and Machine Intellignce*, 14(2):269–277, 1992.

[12] T. Läbe and E. Gülch. Robust Techniques for Estimating Parameters of 3D Building Primitives. In *Proc. of ISPRS Comm. II Symposium*, Cambridge, UK, July 1998. ISPRS.

[13] T. Kanade and M. Okutomi. A Stereo Matching Algorithm with an Adaptative Window: Theory and Experiment. *I.E.E.E. Trans. on Pattern Analysis and Machine Intelligence*, 16(9):920–932, sept. 1994.

[14] J.-L. Lotti and G. Giraudon. Adaptive window algorithm for aerial image stereo. In *12th Int. Conf. on Pattern Recognition*, pages 701–703, Jerusalem, 1994.

[15] N. Paparoditis, M. Cord, M. Jordan, and J.-P. Cocquerez. Building Detection and Reconstruction from Mid and High Resolution Aerial Images. *Computer Vision and Image Understanding*, 72(2):122–142, Nov. 1998.

[16] H. Schultz. Terrain reconstruction from widely separated images. In *SPIE Conf. on Integrating Photogrammetric Techniques with Scenes Analysis and Machine Vision II*, volume 2486, pages 113–123, Orlando-Florida, Avril 1995.

[17] M.J. Hannah. A System for Digital Stereo Matching. *Photogrammetric Engeenering and Remote Sensing*, 55(12):1765–1770, 1989.

[18] U. Leloglu, M. Roux, and H. Maître. Dense Urban DEM with Three or More High-Resolution Aerial Images. In *ISPRS Symposium on GIS - Between Visions and Applications*, Stuttgart, Germany, 1998.

[19] P. Fua. Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities. In *Int. Joint Conf. on Artificial Intelligence*, pages 1292–1298, 1991.

[20] W. Eckstein and O. Munkelt. Extracting objects from digital terrain models. In *Remote sensing and reconstruction of 3D objects and scenes*. SPIE, 1995.

[21] C. Hug. Extracting artificial surface objects from airborne laser scanner data. In A. Grün, E. Baltsavias, and O. Henricsson, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images II*, Ascona (CH), 1997. Birkhäuser Verlag.

[22] M. Cord, M. Jordan, J.-P. Cocquerez, and N. Paparoditis. Automatic Extraction and Modelling of Urban Buildings from High Resolution Aerial Images. In *ISPRS, Automatic Extraction of GIS Objects from Digital Imagery*, volume 32, pages 187–192, München, September 1999.

[23] M. Cord, M. Jordan, and J.P. Cocquerez. Accurate building structure recovery from high resolution aerial imagery. *Computer Vision and Image Understanding*, Submitted Sept. 99.

[24] M. Cord and D. Declercq. Bayesian model identification: Application to building reconstruction in aerial imagery. In *ICIP-99*, Kobe, Japan, Oct. 1999.

[25] M.A. Tanner. *Tools for Statistical Inference : Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Spinger-Verlag, New York, 1993.

[26] N. Shephard. Partial non-gaussian state space. *Biometrika*, 81:115–131, 1994.

[27] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 1998.

[28] G. Casella and E. George. Explaining the gibbs sampler. *Ann. Stat.*, 46:167–174, 1992.

[29] M.A. Tanner and W. Wong. The calculation of posterior distributions by data augmentation. *J. Amer. Stat. Assoc.*, 82:528–550, 1987.

[30] A. Dempster N. Laird and D. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *J. Royal Stat. Soc.*, B-39:1–38, 1977.

[31] J. Diebolt and C.P. Robert. Estimation of finite mixture distributions through bayesian sampling. *J. Royal Stat. Soc.*, 56(2):363–375, 1994.

[32] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:721–740, 1984.

[33] G. Celeux and J. Diebold. The sem algorithm: a probabilistic teacher algorithm derived from the em algorithm for the mixture problems. *Comp. Stat. Quart.*, 2:73–82, 1985.

[34] J. Canny. A computational approach to edge detection. *IEEE Trans. on P.A.M.I.*, 8(6):679–697, Nov. 1986.

[35] R. Deriche. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.

[36] M. Cord, F. Huet, and S. Philipp. Optimal Adjusting of Edge Detectors to Extract Close Contours. In *Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, June 1997.

[37] P.J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

# List of Tables

# List of Figures