

# RECONSTRUCTION OF 3D LINEAR PRIMITIVES FROM MULTIPLE VIEWS FOR URBAN AREAS MODELISATION

Franck Taillandier<sup>a</sup>, Rachid Deriche<sup>b</sup>

<sup>a</sup> Institut Géographique National/MATIS, 2-4 avenue Pasteur, 94165 Saint-Mandé - franck.taillandier@ign.fr

<sup>b</sup> INRIA/ROBOTVIS, 2004 Route des Lucioles - BP 93 - 06902 - Sophia-Antipolis Cédex, France - Rachid.Deriche@sophia.inria.fr

**KEY WORDS:** 3D Reconstruction, Feature Matching, Multiple Aerial Images, Accuracy, 3D Segments

## ABSTRACT:

In this paper, a new method for reconstruction of 3D segments from multiple images in urban areas environment is presented. Compared to previous algorithms, this one performs the matching of 2D segments in the Object Space through a sweep plane technique, thus avoiding the combinatorial exploration of all possible correspondences and handling images in a *symmetric* way. Furthermore, a method for reconstruction of 3D line from 2D lines, which takes into account the uncertainty on the parameters that define the 2D lines is also presented. It enables to get normalized residuals, which is used as a geometric criterion usable *whatever the number of images* is, to assess or reject potential correspondences. This criterion along with an unicity criterion is at the heart of the algorithm to prune the set of possible correspondences and to keep only reliable matches. Promising results are presented on simulated and real data. They show the ability of the algorithm to overcome detection errors in images and its robustness to occlusions in some images.

## 1 INTRODUCTION

### 1.1 Context

Reconstruction of buildings in urban areas is a very hard problem with regard to the complexity of the scenes. In this context, the extraction of reliable 3D primitives is a key step and an important preliminary to facilitate the generation of hypotheses of buildings or to detect models of buildings (Baillard et al., 1999; Fuchs and Le-Men, 1999; Fischer et al., 1998; Willuhn and Ade, 1996; Noronha and Nevatia, 2001). 3D segments are essential in order to make a model of urban scenes, because of both their geometric reliability and their ability to caricature the scene. However, matching of segments is a difficult problem mainly because of the segmentation artifacts visible in the images: although the line orientation is generally well known, boundary points are not reliable and the polygonal approximation often behaves differently in the images. However, 3D geometric constraints are very strong and can efficiently discriminate potential matches. One of the key element for an algorithm of 3D segments reconstruction in multi-view is to avoid the tedious exploration of all possible matches and to handle images in a symmetric way without giving any image a special role.

### 1.2 State of the Art

Detection and reconstruction of 3D segments is a well-known problem and has received much attention for years from the scientific community. In the case of stereoscopic views, the geometric constraint is reduced to an overlap constraint, which led the authors to use two kinds of strategies. The first one consists in matching graphs of segments, which introduces very strong constraints (Ayache, 1989; Horaud and Skordas, 1989). These methods give a priori more reliable results but are very sensitive to segmentation errors that alter the appearance of graphs from one image to another. Besides, they have a high complexity, which makes the generalization to the multi-view case very difficult. The other strategy is based on the use of geometric or radiometric attributes such as orientation, length, overlap (Medioni and Nevatia, 1985; Zhang and Faugeras, 1992; Gros et al., 1998) or radiometric neighborhood (Schmid and Zisserman, 1997). The set of these attributes is also very sensitive to segmentation errors and depends on the conditions in which views were taken.

In multi-view, the use of trifocal tensor ensures a strong geometric constraint (Hartley, 1995; Shashua, 1994; Torr and Zisserman, 1997; Papadopoulos and Faugeras, 1998) and promising results

are supplied by (Schmid and Zisserman, 1997) who show that the introduction of this constraint enables to get much less ambiguous matches. All these techniques face three main problems:

**combinatory** : all the triplets of segments must be tested

**fusion** : information from the triplets need to be merged

**dissymmetry** : one *always* uses a reference image.

The algorithm proposed in this article overcomes these drawbacks and shows a new approach for the reconstruction of 3D segments *from calibrated views*. First, it performs the point to point matching in the Object Space, thus avoiding the tedious combinatorial exploration of all possible correspondences and handling images in a true symmetric way. Second, this article shows a new reconstruction method that takes into account the uncertainty on the parameters and enables to get a statistic score usable whatever the number of images is, in order to assess or reject potential matches. Potential matches are then pruned based on this geometric criterion as well as on an unicity criterion. Results, presented on simulations and real data, are promising.

## 2 MATCHING IN OBJECT SPACE

### 2.1 A Sweep Plane Approach for 3D Edge Points

The first step in the reconstruction of 3D primitives, known as the most difficult one, is the matching of 2D primitives. This problem, deeply studied in the stereoscopic case, is hard in multi-view because the exhaustive search of all possible correspondences has a crippling combinatorial. Our goal, here, is to provide a method satisfying the 3 principles of true multi-image as stated by Collins (Collins, 1995):

- the method generalizes to *any* number of images greater than 2,
- the algorithmic complexity is  $\mathcal{O}(n)$  in the number of images
- all the images are treated equally (no "reference" image)

The reader can refer to this article for further bibliography on the multi-view topic and the techniques developed in the literature meeting or not these three conditions.

In his article, Collins shows a method that consists in finding 3D edge points by matching 2D edge points directly in the Object

Space. By successively considering all the voxels in the discretized 3D space, Collins determines whether there should be a 3D edge point or not, according to the number of rays hitting the given voxel. The matching is done through a sweep plane algorithm. The main drawback is that no explicit link is made between matched primitives. Collins uses a statistic criterion to assess whether there is a 3D edge point or not, without linking the 2D edges. No use of topology in images can thus be made. The algorithm described here extends this method and stores potential matchings of 2D segments.

## 2.2 Matching of 2D Segments

**Images of Segments:** An edge extraction followed by a polygonal approximation are first performed in each image. Therefore, one gets a set of images of labels called hereafter *images of segments* (an edge point belongs to a labelled segment). More formally,  $\mathcal{C}_i$  refers as the set of labels in the image  $i$  and  $\mathcal{C}_i^\infty = \mathcal{C}_i \cup \{\Phi\}$  where  $\Phi$  indicates the absence of edge. A pixel  $p$  in an image of segment  $I_i$  verifies then  $p \in \mathcal{C}_i^\infty$ .

**Sweeping plane:** The matching process is based on the move of a “sweeping plane” along a line that is normal to it, as it is presented in (Collins, 1995). Conveniently but without loss of generality, in the case of aerial images, we will choose a horizontal plane  $z = z_k$  partitioned into a grid of cells whose axis are chosen aligned with the X-Y axes of the scene. Thus, each cell  $(x, y)$  on this grid defines a 3D voxel in the space  $(x, y, z_k)$ . By sweeping from  $z_{min}$  to  $z_{max}$ , this plane samples a volume of interest discretized into a set of voxels.

Let us notice that the optimal cell size of the grid as well as the step of the sweeping procedure in  $z$  can be automatically defined, knowing the absolute positions of the focal points of the images (result of an aerotriangulation process in the case of aerial images) and the intrinsic parameters of the camera.

**Ground Grids:** For a given sweeping plane  $z = z_k$ , each image of segments is projected onto the grid leading to a “ground grid”  $G_i$ . Each edge point “votes” for a set of cells surrounding the intersection of its viewing ray with the sweeping plane and roughly subtended by a pixel shaped cone of viewing rays emanating from the edge point. This set is specified by the approximate Jacobian of the transformation that maps image  $i$  onto the grid (Collins, 1995). Thus, we account for the fact that close-up images give finer localization than further images.

Since all the “ground grids”  $G_i$  are in a common referential defined by the grid of the sweeping plane, for each voxel  $(x, y, z_k)$  of this grid, one can thus determine some hypotheses of matches between labels and therefore between segments of the different images (Figure 1). A match hypothesis is formally defined by the  $N$ -uplet:

$$h = (c_1, c_2, \dots, c_N) \text{ where } c_i \in \mathcal{C}_i^\infty \quad (1)$$

and the set of matches hypotheses computed in a given voxel  $\mathbf{V} = (x, y, z)$  is  $\mathcal{H}_{\mathbf{V}}$ .

**Associations and  $\Omega$ :** By sweeping the plane in the Object Space from  $z_{min}$  to  $z_{max}$ , one can easily collect all the *potential* correspondences. these are segments that have, at least, one voxel of intersection when they are projected through the scene. Since the matching is done “voxel by voxel”, this method of matching enables us to keep the number of occurrences  $v(h)$  of a given hypothesis (the number of voxels in which this hypothesis was met), which is proportional to the overlap of the 2D segments projected in the 3D space in the case of a correct match:

$$v(h) = \text{Card}\{\mathbf{V}/h \in \mathcal{H}_{\mathbf{V}}\} \quad (2)$$

This valuation will be used afterwards to discriminate between hypotheses. An *association* is then defined by a couple

$(h_\omega, v_\omega) = (h, v(h))$ . The set of all the associations is  $\Omega = \bigcup \omega$ . An order is assigned to each association related to the number of segments that really match, namely:

$$\text{ord}(\omega) = \text{ord}(h_\omega) = \text{Card}\{c_i, i = 1 \dots N/c_i \neq \Phi\} \quad (3)$$

The set of associations calculated through this sweeping plane technique defines a set of hypotheses, which need to be assessed or rejected. We present in section 3 a method of reconstruction that gives a statistic score that will be used to prune  $\Omega$ . A total order law will also be defined in section 4 so that maximal associations should be kept while ensuring an unicity criteria. Let us mention that this method can also easily integrate point by point correlation method (Paparoditis et al., 2000) along the 3D segment, which could be used to discriminate between hypotheses. However this criterion may be not reliable because most of the time, the segments delineate facades which are seen in very different ways according to the point of view.

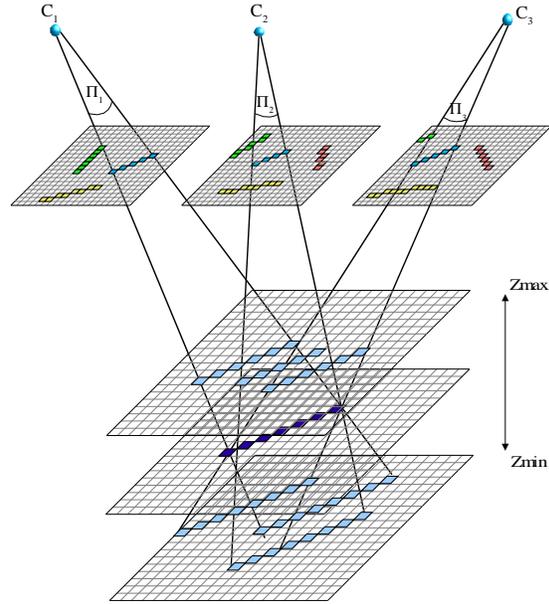


Figure 1: Technique of the sweeping plane. In this figure, we show the projection of images of segments at an given altitude. Since the referential is the same for all the ground grids, the projections are shown by accumulation on the same grid.

## 3 RECONSTRUCTION OF 3D SEGMENTS

### 3.1 Straight Lines in Space

In order to represent a 3D straight line, we decide to use the complete **abpq** representation (Ayache, 1989; Zhang and Faugeras, 1992) described by the three maps:

$$\begin{cases} \text{Map 1 : } x = az + p, & y = bz + q \\ \text{Map 2 : } y = ax + p, & z = bx + q \\ \text{Map 3 : } x = ay + p, & z = by + q \end{cases} \quad (4)$$

In each map, the 3D line is represented as the intersection of two planes. Map 1 can represent straight lines non parallel to  $\mathcal{P}_{XY}$ , Map 2 straight lines non parallel to  $\mathcal{P}_{YZ}$  and Map 3, straight lines non parallel to  $\mathcal{P}_{XZ}$ .

### 3.2 Iterative Reconstruction

The goal is to compute the line  $\mathcal{D}$ , which represents intersection of  $n_s$  planes  $\Pi_i$  where  $\Pi_i$  is defined as the plane going through the 2D segment  $S_i$  in image  $i$  and the corresponding focal point  $C_i$  (Figure 1).

In the following, after a recall of the relationships between the equations of lines in the images and the equations of planes  $\Pi_i$ , we describe a method taking into account the uncertainty on the parameters of the line underlying the segment  $S_i$ :  $(\theta_i, \rho_i)$  and giving a valid criterion on the quality of the reconstruction.

**From 2D Lines to Planes  $\Pi_i$ :** A straight line  $\mathcal{D}_i$  in the plane can be represented by the parameters  $\theta_i$  and  $\rho_i$  and the equation:

$$u \sin \theta_i - v \cos \theta_i + \rho_i = 0 \quad (5)$$

where  $u$  and  $v$  are the coordinates of a pixel.

In the following, the variance-covariance matrices  $\Lambda_{\mathcal{D}_i}$  linked to the uncertainty on these parameters are assumed to be known. These matrices can indeed be computed after the polygonal approximations steps (Deriche et al., 1992).

The camera parameters are assumed to be known as well, by hypothesis. Thus, one knows, for each camera the 3x4 perspective transformation matrix  $\mathcal{T}$ :

$$\mathcal{T} = \begin{bmatrix} \mathcal{T}_{11} & \mathcal{T}_{12} & \mathcal{T}_{13} & \mathcal{T}_{14} \\ \mathcal{T}_{21} & \mathcal{T}_{22} & \mathcal{T}_{23} & \mathcal{T}_{24} \\ \mathcal{T}_{31} & \mathcal{T}_{32} & \mathcal{T}_{33} & \mathcal{T}_{34} \end{bmatrix} \quad (6)$$

Considering the pinhole camera model, a 3D point  $\mathbf{M}(X, Y, Z)$  can be linked to its projection  $\mathbf{m}(u, v)$  in the image by the relation  $\tilde{\mathbf{m}} = s\mathcal{T}\tilde{\mathbf{M}}$  with  $s \neq 0$  in homogeneous coordinates. Substituting  $u$  and  $v$  from the latter equation in equation (5) gives  $\tilde{M}^t \mathbf{v} = 0$  with

$$\mathbf{v} = \begin{bmatrix} \mathcal{T}_{11} \sin \theta - \mathcal{T}_{21} \cos \theta + \mathcal{T}_{31} \rho \\ \mathcal{T}_{12} \sin \theta - \mathcal{T}_{22} \cos \theta + \mathcal{T}_{32} \rho \\ \mathcal{T}_{13} \sin \theta - \mathcal{T}_{23} \cos \theta + \mathcal{T}_{33} \rho \\ \mathcal{T}_{14} \sin \theta - \mathcal{T}_{24} \cos \theta + \mathcal{T}_{34} \rho \end{bmatrix} \quad (7)$$

For each 2D segment  $S_i$ , the corresponding plane  $\Pi_i$  can therefore be represented by a 4-parameter vector  $\mathbf{v}_i = [\alpha_i, \beta_i, \gamma_i, \delta_i]^t$  defined in equation (7). In order to compute the straight line  $\mathcal{D}$ , intersection of the planes  $\mathbf{v}_i$ , one uses the complete **abpq** representation described in 3.1. The reconstruction should be performed in the three maps in order to be able to represent each straight line solution. At the end of the process, we choose the map where  $\|ab\|$  is minimum. That represents the case where both intersecting planes are the closest to the orthogonal configuration for which the intersection is better defined. In the following, we will only describe the computations for Map 1 since they are the same in the other maps. Then each point on the straight line  $\mathcal{D}$  represented in this map satisfies equation (8):

$$[az + p, bz + q, z, 1]^t \mathbf{v}_i = 0 \quad \forall i \in [1..n_s] \quad (8)$$

For the  $n_s$  segments,  $2n_s$  equations are obtained:

$$\begin{aligned} f_{2i} &= a\alpha_i + b\beta_i + \gamma_i = 0 \\ f_{2i+1} &= p\alpha_i + q\beta_i + \delta_i = 0 \end{aligned} \quad \forall i \in [1..n_s] \quad (9)$$

In order to find out the 4 unknown  $a, b, p, q$ , we search for  $\mathbf{X} = [a, b, p, q]^t$  solution of the linear system  $\mathbf{A}\mathbf{X} = \mathbf{b}$  where the  $\mathbf{A}$ -matrix and the  $\mathbf{b}$ -vector, whose sizes are respectively of  $2n_s \times 4$  and  $2n_s$ , are given by:

$$\mathbf{A} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ \alpha_i & \beta_i & 0 & 0 \\ 0 & 0 & \alpha_i & \beta_i \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \vdots \\ -\gamma_i \\ -\delta_i \\ \vdots \end{bmatrix} \quad (10)$$

The least square solution of this system, already supplied by (Zhang and Faugeras, 1992) minimizes the sum of squares of the residuals:

$$\mathcal{S} = \sum_i \dots + f_{2i}^2 + f_{2i+1}^2 + \dots \quad (11)$$

The classical solution is  $\hat{\mathbf{X}} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{b}$ .

**Iterative Method:** The previous method assumes implicitly that the same variance is taken for each residual (the least-square solution is only optimal under this condition). The same importance is thus given to each plane whereas the variances  $\Lambda_{v_i}$  depend a lot on the variance of the parameters of the lines in images. In order to take this uncertainty into account, a weighted least-square solution is used and one searches for the minimization of:

$$\tilde{\mathcal{S}} = \sum_i \dots + \frac{f_{2i}^2}{\sigma_{f_{2i}}^2} + \frac{f_{2i+1}^2}{\sigma_{f_{2i+1}}^2} + \dots \quad (12)$$

This resolution is the same as the resolution of the weighted least-square problem  $\mathbf{K}\mathbf{A}\mathbf{X} = \mathbf{K}\mathbf{b}$  with:

$$\mathbf{K} = \text{diag}\left(\frac{1}{\sigma_{f_0}^2}, \dots, \frac{1}{\sigma_{f_{2*n}+1}^2}\right) \quad (13)$$

when  $K$  is constant, the problem is sorted out using the same methodology implemented in the classical least-square minimization, namely  $\hat{\mathbf{X}} = (\mathbf{A}^t \mathbf{K}^t \mathbf{K} \mathbf{A})^{-1} \mathbf{A}^t \mathbf{K}^t \mathbf{K} \mathbf{b}$ . In our case, however, the variance parameters on the residuals  $\sigma_{f_j}$ , which can be computed from the relations ((9)) and ((7)) and equation ((14)) (Xu and Zhang, 1996), assuming  $\Lambda_{\mathcal{D}_i}$  known, depend on  $\mathbf{X}$ .

$$\begin{aligned} \sigma_{f_{2i}}^2(\mathbf{X}) &= \frac{\partial f_{2i}^t}{\partial \mathbf{v}_i} \frac{\partial \mathbf{v}_i^t}{\partial \mathcal{D}_i} \Lambda_{\mathcal{D}_i} \frac{\partial \mathbf{v}_i}{\partial \mathcal{D}_i} \frac{\partial f_{2i}}{\partial \mathbf{v}_i} \\ \sigma_{f_{2i+1}}^2(\mathbf{X}) &= \frac{\partial f_{2i+1}^t}{\partial \mathbf{v}_i} \frac{\partial \mathbf{v}_i^t}{\partial \mathcal{D}_i} \Lambda_{\mathcal{D}_i} \frac{\partial \mathbf{v}_i}{\partial \mathcal{D}_i} \frac{\partial f_{2i+1}}{\partial \mathbf{v}_i} \end{aligned} \quad (14)$$

The following iterative scheme is thus finally used:

```

r=0
resolution of  $AX^{(0)} = b$ 
do
  computations of  $\sigma_{f_i}^{(r)}(X^{(r)})$  with (14)
  resolution of  $KAX^{(r+1)} = Kb$ 
  with  $K = \text{diag}\left(\frac{1}{\sigma_{f_1}^{(r)}}, \frac{1}{\sigma_{f_2}^{(r)}}, \dots, \frac{1}{\sigma_{f_{2n+1}}^{(r)}}\right)$ 
  if  $((X^{(r+1)} - X^{(r)})^2 < \text{threshold})$  convergence=true
  else convergence=false
  r=r+1
while  $(\neg (\text{convergence}) \vee (r \geq \text{NMAX}))$ 
end

```

### 3.3 Final Reconstruction and Qualification

The extremity points are finally computed by projecting each 2D extremity on the 3D line. We chose an union strategy, which consists in taking the union of all the 2D segments projected on the 3D line.

One of the key points of this approach is that residuals are *normalized*. Assuming that the errors follow a Gaussian law, the sum of the squares  $\tilde{\mathcal{S}}$  of these residuals follows a  $\chi^2$  law with  $2*n_s - 4$  degree of freedom (there are indeed  $2 * n_s$  equations and 4 parameters). This result can give a good qualification criterion and enables to assess or reject a given match. We will use this result in the final algorithm to select correct matches.

## 4 GLOBAL ALGORITHM

### 4.1 Extraction of 2D Segments

First, an edge extraction is performed on each image, using a classical gradient operator (Deriche, 1987) followed by the hysteresis detection of local maxima in the direction of gradients. Edges are then linked and polygonized. One uses an iterative merging process based on the maximum residual of the orthogonal regression: The polylines whose merging gives a minimal maximum residual are first merged. A tolerance on the polygonal approximation enables us to stop the process when the merging has a maximum residual above a threshold given by the user  $s_{pol}$ . Once the polygonal approximation is done, the parameters  $\theta$  and  $\rho$  of the lines underlying the segments as well as the *variance-covariance matrix* of these parameters are estimated by using the results of (Deriche et al., 1992) and assuming that the edge points detected by the Canny-Deriche operator (Deriche, 1987) have a variance given by:

$$\Lambda = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \quad (15)$$

where  $\sigma$  can be determined through the ratio signal/noise in the images.

### 4.2 Pruning of $\Omega$

**Geometric pruning:** By using the matching technique described in section 2, a set of associations is generated. In this algorithm, only the association which match more than three 2D segments are taken into account (the geometric criterion is indeed valid under this condition). An association is assumed to be valid only if the score given by the reconstruction (see section 3) passes the  $\chi^2$  test with a probability  $\mathcal{P}$  defined by the user.  $\Omega$  is pruned by using this geometric constraint.

**Pruning on an unicity criterion:** A total order law  $\prec$  is then defined on  $\Omega$ :

$$\omega \prec \omega' \Leftrightarrow \begin{cases} ord(\omega) < ord(\omega') \\ \text{or else } v_\omega < v_{\omega'} \\ \text{or else } \tilde{S}_{h_\omega} < \tilde{S}_{h_{\omega'}} \end{cases} \quad (16)$$

$\tilde{S}$  represents the sum of the squares of normalized residuals of the reconstruction. Thus, this relation gives the priority to the associations that have a high number of matched segments and then the associations whose number of matched edge points is high. This relation enables us to adopt a “winner takes all” strategy *while keeping the symmetry in the problem and without giving any image a special role*. Thanks to the set  $\Omega$  and to the relationships given in (16), associations can be sorted. Iteratively, each maximal association is chosen and  $\Omega$  is pruned by checking the following unicity constraint: a segment in one image can only belong to one association. At the end of the process, the set of associations validated as correct is obtained and 3D segments can be reconstructed.

## 5 RESULTS

### 5.1 Comparison with the $\chi^2$ Law

The geometric pruning is based on the comparison test of the normalized sum of the squares of residuals with the  $\chi^2$  law and it is therefore important to assess the behavior of the reconstruction with this law. In order to check this point, the following random test has been performed:

- 1 project a 3D segment sampled in 6 images,
- 2 add a Gaussian noise of parameter  $\sigma$  on each 2D point in the images,

- 3 reconstruct the 3D segment by computing the normalized sum of squares of residuals  $\tilde{S}_t$  for the test  $t$ .

By iterating steps 2 and 3, one can compute the proportion  $g(x)$  of the tests for which  $\tilde{S}_t$  is greater than a number  $x$ :

$$g(x) = \frac{Card\{t/\tilde{S}_t > x\}}{Card\{t\}} \quad (17)$$

If  $\tilde{S}_t$  follows the  $\chi^2$  law with  $2 * 6 - 4 = 8$  degrees of freedom,  $g$  must follow the probability law  $\mathcal{P}(\chi_{[8]}^2 > x)$  given by equation (18):

$$\mathcal{P}(\chi_{[8]}^2 > x) = \frac{\int_x^\infty u^{(8/2)-1} * \exp(-u/2) du}{\int_0^\infty u^{(8/2)-1} * \exp(-u/2) du} \quad (18)$$

Results presented in Figure 2 show that the comparison of  $g(x)$  with this law is fairly good. The theoretical curve and the curve obtained through the statistic test follow the same tendency. The observed differences are likely due to the first order approximations made in the computations of the variances at the different steps of the reconstruction (Deriche et al., 1992). The general shapes are however close enough to justify the use of the  $\chi^2$  statistic test to reject incorrect matches.

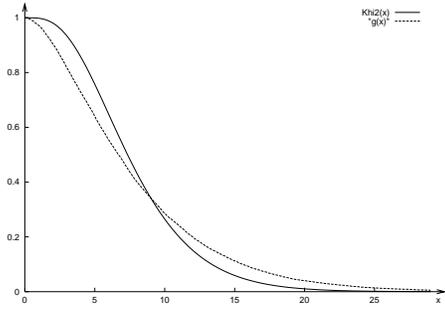


Figure 2: comparison of residuals with the  $\chi^2$  law.  $\sigma = 2$ , number of tests : 10000

### 5.2 Simulations

Some simulations have been performed with a building model from the BD TRAPU©\* (Figure 3). This model is made up of a set of polygonal facets whose edges are extracted. They are then projected in several images (6 in our test). In order to test the matching abilities of our algorithm independently of the segmentation errors present in the images, *we perform no edge detection in this simulation*. The 2D segments are then directly the projections of existing 3D segments. In all the simulations, the value  $\mathcal{P} = 0.9$  and a volume of interest of  $160 \times 170 \times 40 \text{ m}^3$  were used. We first validated our algorithm onto noisy and noise-free simulated data to test its capabilities when *all* the segments are seen in the 6 images. In this case, no mismatch was found, all the segments were reconstructed (except in the noisy case where the choice of  $\mathcal{P} \neq 1$  induced a few rejects). In order to also assess the capabilities of the algorithm to handle cases where the segments are visible in whatever number (greater than 3) of images, another simulation was performed with hidden faces and noise (Figure 4): the segments are projected in the images by taking into account hidden faces and by adding some noise to the points. The reconstruction is still correct as the matches accepted by the algorithm are all correct. Necessarily, some of the true matches were rejected due to the fact that  $\mathcal{P} \neq 1$ . The algorithm has well

\*“TRAcE de Perspectives Urbaines”. IGN Copyright.

handled the case where the segment is only seen in a number of images smaller than  $N = 6$ .

In the case of a perfect line detector, the results have thus shown that the generation of matching hypotheses is correct and that the algorithm extracts the correct association independently on the number of images.

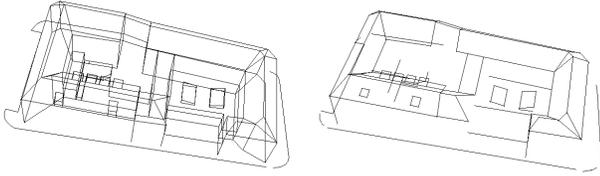


Figure 3: Reference model used in all our simulations

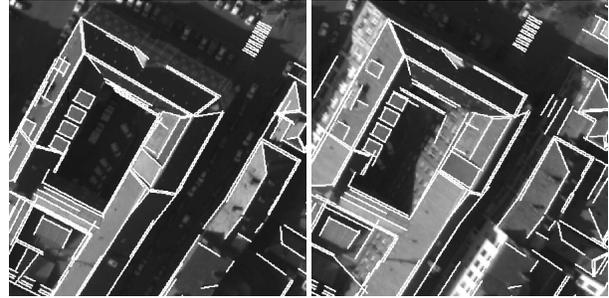
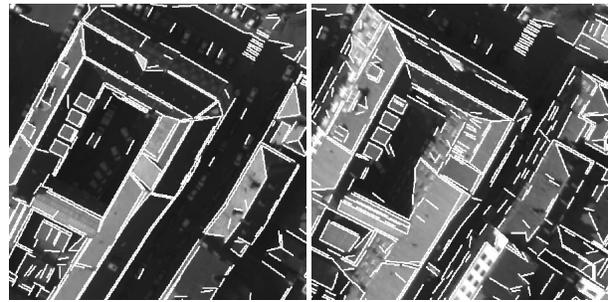


Figure 4: reconstruction with noise and hidden faces

### 5.3 Real Images

With real images, the experiment has been made with 6 images of a building. For this set of images taken by the CCD camera of IGN, a pixel in an image represents roughly 20cm on the ground. The parameters used in the following are:

- size of the volume of interest:  $160 \times 170 \times 40 \text{ m}^3$
- alpha for the Canny-Deriche filter: 1.5
- hysteresis thresholds:  $sB=0 \text{ sH}=5$
- polygonisation threshold: 1 pixel
- minimum size for a 2D segment in an image: 20 pixels
- number of images required for an association: 4
- $\mathcal{P} = 0.9$

The results (Figure 5) show that 413 3D segments have been reconstructed. Given the number of segments detected in each image (between 845 and 1100), this result is fairly satisfactory. The results show a good restitution of details, as for instance parallel and very close lines that are difficult to discriminate. The algorithm overcomes some artifact problems due to the detection in images like broken segments or undersegmentation in some images for instance. Besides, there is no false match between 2D segments in different images.

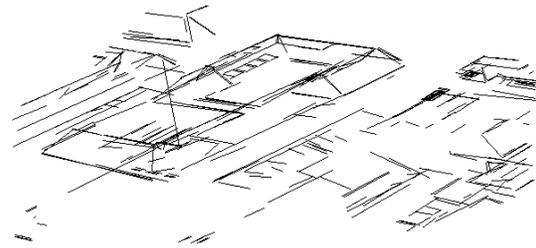


Figure 5: 2 extracts of the 6 images used in this experiment. Above: segments detected. Middle: reconstructed segments projected on the images. Below: lateral view of the reconstruction.

## 6 EXTENSIONS

### 6.1 Restriction of the Search Space

In order to reduce the search space and also to avoid mismatches, a search volume can be derived from the dilatation of a DEM. One can thus restrain the valid voxels and force the reconstructed segments to have their extremities in the search volume.

### 6.2 Extension to 2 Images

In order to overcome most of the undersegmentation problems, the algorithm can easily be extended to integrate hypotheses with 2 matched 2D segments only. In this case, of course, the geometric criterion is not used. Instead of it, an overlap constraint has been set up to 0.5 (ratio of the common part over the union part). Figure 6 shows that a lot of segments are reconstructed using these both extensions. Nevertheless, some mismatches appeared, certainly due to the poor geometric constraint in the last case.

These two extensions enables to treat regions where the number of available images is relatively low or to treat wide areas while keeping a reasonable running time as in Figure 6, where 3500 3D segments were reconstructed from 5 images using a DEM to constrain the search space.

## 7 DISCUSSION

### 7.1 Advantages

The main interest of the method described above, on the first hand is to deal with all images in a symmetric way, without giving any image a special role, and on the other hand to test all the possible associations, without any combinatorial explosion. A new method of reconstruction of 3D segments has also been presented. Although certainly less precise than any bundle adjustment technique that avoids the propagation of first order errors, the method enables to give a valid criterion usable whatever the number of images is to assess or reject a potential match.

### 7.2 Problems

The actual algorithm does not handle the uncertainty in the camera parameters and takes only into account the variance on the line parameters. A modelisation of the influence of these errors on  $\Lambda_{v_i}$  should be thought over. As far as the extensions are concerned, the algorithm would certainly benefit from correlation score in the case of matching segments in only 2 short-range images.

### 7.3 Future Work

We plan to improve the “winner takes all” scheme and trying to merge associations in order to refine the geometric precision.

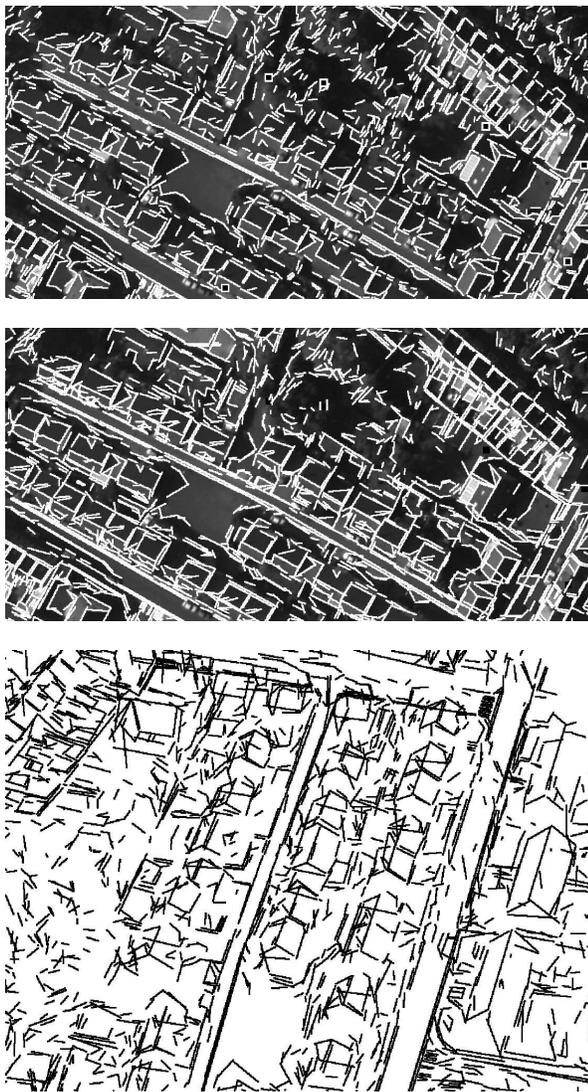


Figure 6: 2 extracts of the 5 images used in this experiment. Above: segments detected in them. Middle: reconstructed segments projected on them. Below: lateral view of reconstruction

These 3D segments will be used as basic primitives to make the generation of hypotheses of buildings easier and thus constrain the search of buildings models that matches the best the reality.

## 8 CONCLUSION

We have presented an algorithm that enables matching and reconstruction of 2D segment in a multiple calibrated images context. The algorithm performs the matching in the Object Space and does not give thus any image a special role. The algorithm tries to keep matches that ensure maximum overlaps. Furthermore, the matching is in  $\mathcal{O}(N)$  and thus avoids the tedious exploration of all possible correspondence while taking into account all the possible associations. The article also presents a method of reconstruction of 3D segments that takes into account the uncertainty on the determination of the parameters of the 2D lines underlying the segments. This method gives normalized residuals, which enables us to qualify the reconstruction. The results of this algorithm have already been used with points obtained through a correlation process in the caricature of an urban scene and provides very promising results. (Paparoditis et al., 2001).

## REFERENCES

- Ayache, N., 1989. *Vision Stoscopique et Perception Multisensorielle: Application la Robotique Mobile*. InterEditions, Paris.
- Baillard, C., Schmid, C., Zisserman, A., and Fitzgibbon, A., 1999. Automatic line matching and 3D reconstruction of buildings from multiple views. In *IAPRS*, volume 32.
- Collins, R., 1995. A space-sweep approach to true multi-image matching. Technical report, Computer Science Department, Univ. of Massachusetts.
- Deriche, R., 1987. Using canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187.
- Deriche, R., Vaillant, R., and Faugeras, O., 1992. In *Theory and Applications of Image Analysis*, chapter From Noisy Edges Points to 3D Reconstruction of a Scene : A Robust Approach and Its Uncertainty Analysis, pages 71–79. World Scientific. Series in Machine Perception and Artificial Intelligence.
- Fischer, A., Kolbe, T., Lang, F., Cremers, A., Förstner, W., Plümer, L., and Steinhage, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *CVIU*, 72(2):163–185.
- Fuchs, F. and Le-Men, H., 1999. Building reconstruction on aerial images through multi-primitive graph matching. In *2nd IAPR Workshop on Graph-based Representations*, Vienna, Austria.
- Gros, P., Bournez, O., and Boyer, E., 1998. Using local planar geometric invariant to match and model images of line segments. *CVIU*, 69(2):135–155.
- Hartley, R., 1995. A linear method for reconstruction from lines and points. *ICCV*.
- Horaud, R. and Skordas, T., 1989. Stereo correspondence through feature groupings and maximal cliques. *PAMI*.
- Medioni, G. and Nevatia, R., 1985. Segment-based stereo matching. *CVGIP*.
- Noronha, S. and Nevatia, R., 2001. Detection and modeling of buildings from multiple aerial images. *PAMI*, 23(5):501–518.
- Papadopoulo, T. and Faugeras, O., 1998. A new characterization of the trifocal tensor. In *ECCV*.
- Paparoditis, N., Maillet, G., Taillandier, F., Jibrini, H., Jung, F., Guigues, L., and Boldo, D., 2001. Multi-image 3D feature and DSM extraction for change detection and building reconstruction. In Verlag, B. B., editor, *Automatic Extraction of Man-Made objects from aerial and space images*, Ascona.
- Paparoditis, N., Thom, C., and Jibrini, H., 2000. Surface reconstruction in urban areas from multiple views of aerial digital frames. In *IAPRS*, volume XXXIII, Amsterdam.
- Schmid, C. and Zisserman, A., 1997. Automatic line matching across views. In *CVPR*, pages 666–671.
- Shashua, A., 1994. Trilinearity in visual recognition by alignment. *ECCV*.
- Torr, P. and Zisserman, A., 1997. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605.
- Willuhn, W. and Ade, F., 1996. A rule-based system for house reconstruction from aerial images. In *ICPR*, pages 885–889.
- Xu, G. and Zhang, Z., 1996. *Epipolar Geometry in stereo, motion and object recognition*. Kluwer Academic Publishers.
- Zhang, Z. and Faugeras, O., 1992. *3D scene analysis: a stereo based approach*. Springer.