3D INFORMATION EXTRACTION BY STRUCTURAL MATCHING OF SAR AND OPTICAL FEATURES

Florence TUPIN and Michel ROUX

GET - Télécom-Paris - UMR 5141 LTCI - Département TSI
46 rue Barrault, 75013 Paris - France
florence.tupin@enst.fr

KEY WORDS: SAR imagery, aerial imagery, optical imagery

ABSTRACT

The subject of this paper is the extraction of 3D information using an optical image and a SAR image. This aim is made difficult by the very different appearances of the landscape and the man-made features in both images. The proposed method is based on the extraction of SAR primitives (bright linear and point features). They are then matched to the optical image using the directional and modulus gradient map computed on the optical image. A set of matches and associated heights is selected for each primitive. In a first step, the digital terrain model is computed in an iterative way around the predominant height of the SAR primitives. In a second step, using the external knowledge of a building map, the mean height of each building is computed by means of the previous DTM and the remaining matching primitives. The results obtained on an industrial test area are evaluated.

1 INTRODUCTION

Three-dimensional reconstruction in urban areas is a very important subject. Indeed, 3D models are very useful for many applications like urban extension monitoring, disaster (flood, earthquake) monitoring, mobile phone cells planning, etc. There are now many ways to obtain digital elevation models (DEM): on one hand, classical stereovision approaches with optical data are already operational; on the other hand, radargrammetric or interferometric approaches with SAR (Synthetic Aperture Radar) data are still in a research stage (Gamba et al., 2000) (Simonetto et al., 2003) (Quartulli and Datcu, 2003) (Tison et al., 2004).

We are interested here in the use of heterogeneous sources, specially the case of optical data and SAR data. Three-dimensional reconstruction using an optical and a SAR images is theoretically possible, although the accurate knowledge of the sensor parameters is necessary. Besides, since the images of both sensors are radiometrically and geometrically different (speckle presence and distance sampling in SAR images, leading to geometrical distortions like overlay and shadow), point matching algorithms with correlation based methods are un-usable. In this paper we propose a new method for the extraction of 3D information dedicated to these two sensors over semi-urban areas. Preliminary results using the external knowledge of a building map are presented.

There have been many works dealing with the fusion of optical and SAR data. Two main approaches can be distinguished. The pixel-based approaches, which often rely on the use of the joint probability density functions (Schistad et al., 1994) (Lombardo et al., 2003) for segmentation or classification purposes, and the primitive-based approaches which match features of the optical and SAR data for registration purposes (Dare and Dowman, 2000) or region of interest delimitation (Hellwich et al., 2000) (Tupin and Roux, 2003).

In the context of 3D information extraction, only figural approaches are possible. The proposed method is based on the extraction of SAR strong backscatterers. Indeed, due to the surface smoothness in urban areas (relatively to the wavelength), backscattering phenomena strongly rely on the surface orientation compared to the incidence direction. Therefore, specific parts of the buildings usually appear as very bright features (linear and/or punctual), whereas the rest of the building is more or less of the same grey level as the background. The radiometrically high features correspond to the corners between the walls of the buildings and the ground, some parts of the roofs or of the walls; more generally, everywhere a dihedral or trihedral configuration appears with a specific orientation towards the sensor. These features are first extracted in the SAR image, providing a set of salient features. This step is presented in section 2.

These primitives are then projected on the vector gradient map of the optical image using a range of possible heights (section 3). A set of best matches is then selected for each primitive. In a first stage, the digital terrain model (DTM) is extracted in an iterative way around the predominant height of the SAR primitives (most of them lie on the ground since they correspond to the wall/ground corner of the buildings). This part is described in section 4. In a second step, using the external knowledge of a building map, the mean height of each building is computed by means of the previous DTM and the matched primitives (section 5). Eventually, the results obtained on a test area are evaluated (section 6). The whole synopsis of the method is shown figure 1.

2 SAR PRIMITIVE EXTRACTION

As said in the introduction, the appearance of the objects in the SAR image are strongly related to their geometrical properties compared to the along track direction and the incidence angle, and to their roughness compared to the wavelength (Hardaway et al., 1982) (Tupin et al., 2002). Due to multiple bounce scatterings, many very bright features corresponding to dihedral or trihedral configurations are present in the data (Franceschetti et al., 2002). They correspond to wall / ground corners, balconies, chimneys, posts, street lamps, etc. Except these features, relatively few information is available on the building surfaces due to the speckle. Indeed, since rooftop surfaces are quite smooth compared to the wavelength (9cm in S band), their mean radiometry is often rather close to the one of the ground.

The extracted primitives are thus the very bright features of the SAR image, either punctual or linear, which correspond to dihe-

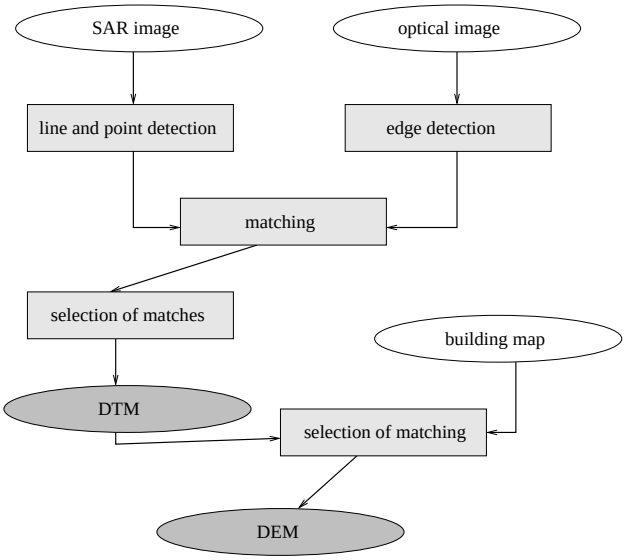

Figure 1: Synopsis of the proposed method. Input data are the original SAR image and optical image (knowing the sensor parameters) and a map of buildings for the DEM reconstruction step.

dral or trihedral configurations and therefore should be visible on the optical data as well.

2.1 Line detector

The line detector has previously been proposed in (Tupin et al., 1998). It is based on the fusion of the results from two line detectors D1 and D2, both taking the statistical properties of speckle into account. Both detectors have a constant false-alarm rate - CFAR detectors- (that is, the rate of false alarms is independent of the average radiometry of the considered region, as defined in (Touzi et al., 1988)). Line detector D1 is based on the ratio edge detector (Touzi et al., 1988), widely used in coherent imagery as stated before. Detector D2 uses the normalized centered correlation between 2 populations of pixels. Both responses from D1 and D2 are merged in order to obtain a unique response as well as an associated direction for each pixel. The detection results are post-processed to provide line segment candidates.

We just recall here the line detector expressions (a detailed study can be found in (Tupin et al., 1998)). The response of the ratio edge detector between 2 regions i and j of empirical means μ_i and μ_j is r_{ij}:

$$r_{ij} = 1 - \min(\frac{\mu_i}{\mu_j}, \frac{\mu_j}{\mu_i})$$

and the response to D1 as $r = \min(r_{12}, r_{13})$, the minimum response of a ratio edge detector on both sides of the linear structure.

The cross-correlation coefficient ρ_{ij} between 2 regions i and j can be shown to be:

$$\rho_{ij}^2 = \frac{1}{1 + (n_i + n_j)\frac{n_i\gamma_i^2\overline{c_{ij}}^2 + n_j\gamma_j^2}{n_i n_j(\overline{c_{ij}} - 1)^2}}$$

where n_i is the pixel number in region i, $\overline{c_{ij}} = \frac{\mu_i}{\mu_j}$ is the empirical contrast between regions i and j, and γ_i the variation coefficient (ratio of standard deviation and mean) which adequately measures homogeneity in radar imagery scenes. This expression depends on the contrast between regions i and j, but also takes into account the homogeneity of each region, thus being more coherent than the ratio detector (which may be influenced by isolated values). In the case of a homogeneous window $\mu_i = \mu_j$, ρ_{ij} equals 0 as expected. As for D1, the line detector D2 is defined by the minimum response ρ of the filter on both sides of the structure: $\rho = \min(\rho_{12}, \rho_{23})$.

Then both responses are merged using an associative symmetrical sum $\sigma(x, y)$, as defined in (Bloch, 1996):

$$\sigma(x, y) = \frac{xy}{1 - x - y + 2xy} \text{ with } x, y \in [0, 1]$$

A theoretical and simulation based study could be used to define the adapted threshold depending on a false alarm and a detection rate. In fact, due to the unknown distribution of the bright pixels along the building / ground corner, such a study is quite difficult in this case and the detection threshold has been empirically chosen.

The set of detected linear features (i.e segments) will be denoted S_l in the following.

2.2 Target detection

Once again a CFAR detector is used. It has first been proposed by Lopes (Lopes et al., 1990) and is based on the ratio of the intensity means of the target and the surrounding background. Therefore, the moving window is subdivided into two areas, a cross shaped area centered on the center of the window and the area resulting of the suppression of the cross in the window.

Again, a theoretical and simulation based study could be used to define the adapted threshold depending on a false alarm and a detection rate. In fact, due to the unknown distribution of the bright pixels of the buildings or man-made objects, an empirical threshold of 2 has been used.

The set of detected punctual features will be denoted S_p in the following.

3 PROJECTION AND MATCHING

3.1 Projection equations

To project points from optical to SAR data and inversely we need some transformation functions. They are based on the computation of the 3D coordinates of the point and on the knowledge of the sensor acquisition system parameters.

The principle of the SAR system is based on the emission of electromagnetic waves which are then backscattered by the elements lying on the ground. For a given time t of acquisition, the imaged points lie in the intersection of a sphere of range $R = ct$ and a cone related to the depointing of the antenna. More precisely, let us denote by S the sensor position, by $\vec{V}$ the speed of the sensor, and by θ_D the Doppler angle which is related to the Doppler frequency f_D and the speed by $\cos(\theta) = \frac{\lambda f_D}{2|\vec{V}|}$, the SAR equations are then given by:

$$\begin{aligned} SM^2 &= R^2 \\ R\sin(\theta_D)V &= S\vec{M}.\vec{V} \end{aligned}$$

Knowing the line i and column j of a pixel and making a height hypothesis h, the 3D coordinates of the corresponding point M

are recovered using the previous equations. R is given by the column number j the resolution step δR and the Nadir range R_o, by $R = j \times \delta R + R_o$. Thus the 3D point M is the intersection of a sphere with radius R, the Doppler cone of angle θ_D and a plane with altitude h. The coordinates are given as solutions of a system with 3 equations and 2 unknowns (since the height must be given).

Inversely, knowing the 3D point M allows to recover the (i, j) pixel image coordinates, by computing the sensor position for the corresponding Doppler angle (which provides the line number) and then deducing the sensor - point distance, which permits to define the column number, since $j = \frac{R - R_o}{\delta R}$.

The geometrical model for optical image acquisition is completely different and is based on the optical center. Each point of the image is obtained by the intersection of the image plan and the line joining the 3D point M and the optical center C. The collinear equations between the image coordinates (x_m, y_m) and the 3D point M (X_M, Y_M, Z_M) are given by:

$$x_m = \frac{a_{11}X_M + a_{12}Y_M + a_{13}Z_M + a_{14}}{a_{31}X_M + a_{32}Y_M + a_{33}Z_M + a_{34}}$$

$$y_m = \frac{a_{21}X_M + a_{22}Y_M + a_{23}Z_M + a_{24}}{a_{31}X_M + a_{32}Y_M + a_{33}Z_M + a_{34}}$$

where the a_{ij} coefficients include internal and external parameters. Once again, a height hypothesis is necessary to obtain M from an image point (x_m, y_m).

3.2 Processing of the optical image

The SAR features are supposed to correspond to radiometric discontinuities in the optical image. Therefore an gradient operator is applied to the optical data. The operator proposed by Deriche (Deriche, 1987), which is a RII filter built using the formalism of Canny (Canny, 1986) has been chosen. The two outputs of the filter (magnitude and direction of the gradient) are used in the following. An example is shown on figure 2.

3.3 Matching

The real position of the SAR feature f in $S_l \cup S_p$ is searched by projecting it on the optical image for a set of height hypotheses. Let us denote by $\mathcal{I}(f, h)$ the set of pixels in the optical image corresponding to the projection of f using the height hypothesis h. For a segment, it is done by projecting both extremities on the optical data and linking them. This is thus an approximation but it is valid for short segments. Figure 3 illustrates the projection of one point of the SAR image in the optical image for a set of heights.

The tested height set $S_h = [H_{min}; H_{max}]$ is chosen to contain the true height of the primitive (typically in urban area, the maximal size of the buildings can be used as upper bound of the interval). Figure 4 shows the variation interval of the SAR primitive.

For each tested height $h \in S_h$, a score $s(f, h)$ is computed as the average of the gradient responses. In the case of the linear features, a constraint on the gradient direction is also introduced. Denoting by $e(i, j)$ and $d(i, j)$ the gradient responses (magnitude and direction) for pixel (i, j), $s(f, h)$ is then given by:

$$s(f, h) = \frac{1}{card(\mathcal{I}(f, h))} \sum_{(i,j) \in \mathcal{I}(f,h)} e(i, j)\Delta(d(i, j), \mathcal{O}(f))$$

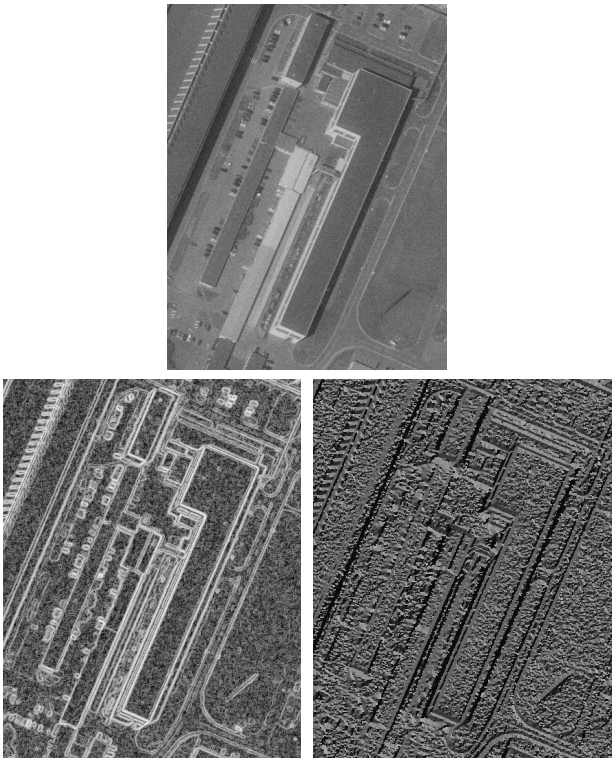

Figure 2: Part of the optical image (on the top), magnitude (left bottom) and direction (right bottom) of the gradient.

with $\mathcal{O}(f)$ the direction of the primtive f, and:

$$\begin{aligned} \Delta(d(i, j), \mathcal{O}(f)) &= 1 \text{ if } | d(i, j) - \mathcal{O}(f) | \leq \frac{\pi}{6} \\ &= 0 \text{ otherwise} \end{aligned}$$

For a point feature, $s(f, h)$ is the gradient magnitude of the projected point.

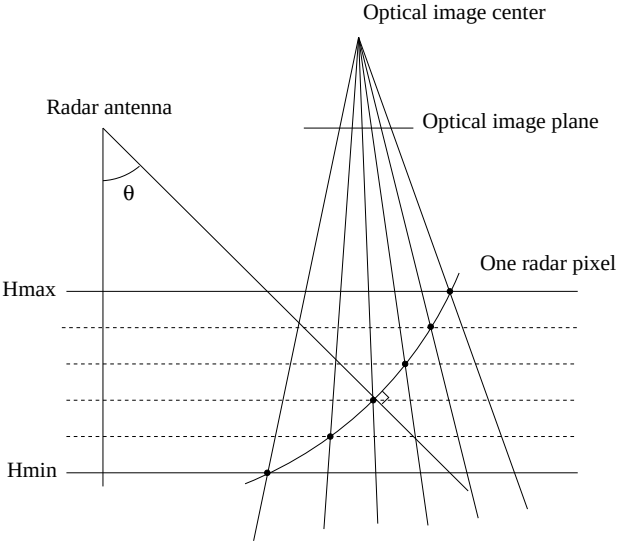

Figure 3: A point of the SAR image is projected in the optical image for different heights.

For each primitive $f \in S_l \cup S_p$, the three best scores $s(f, h)$ and the associated heights h are kept, with the condition to be higher than a fixed threshold th_s. Figure 5 shows the three best matches for a small test area. In point of fact, in the remaining of the article the only best match will be considered.

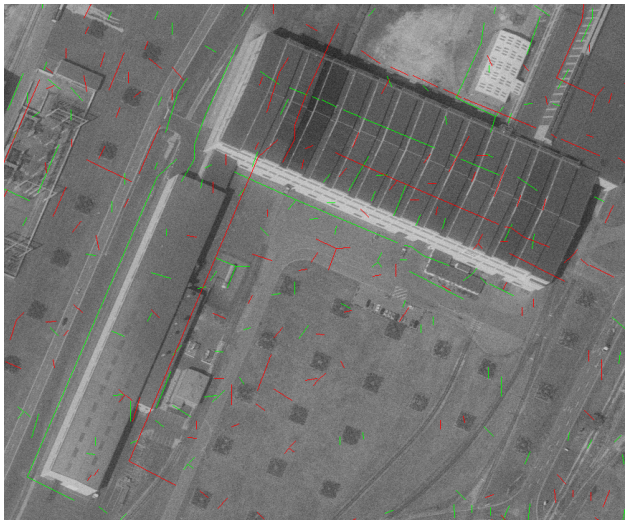

Figure 4: Projection of the linear SAR primitives on the optical image for a height of $H_{min} = 3m$ (in green) and $H_{max} = 40m$ (in red) which are the bounds of S_h for this scene.

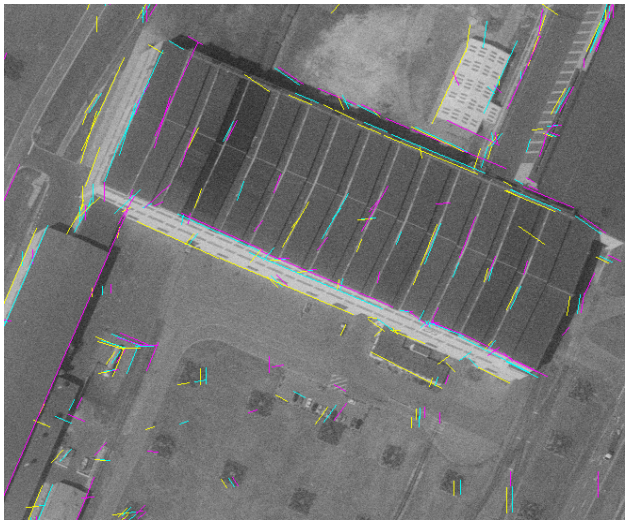

Figure 5: Projection on the optical image of the three best matches of linear SAR primitives.

4 GROUND ELEVATION

Most of the SAR primitives of S_l are due to the response of wall/ground corners to the line detector. It means that their associated heights correspond to the ground height. Therefore a first step is the ground estimation. In a quite restrictive approach, we suppose that the ground is rather flat and that all the wall/ground corners have a relatively close height. Therefore, an height histogram is built, filtered with a Gaussian kernel and its principal mode h_g is computed. The height h_g is supposed to be close to the mean height of the ground and is used to select the primitives of S_l which lie on the ground using a new search step around h_g with a more restrictive height interval ($h \in [h_g - \delta_h; h_g + \delta_h]$) and a lower threshold th'_s ($< th_s$). Figure 6 shows the match obtained with this approach.

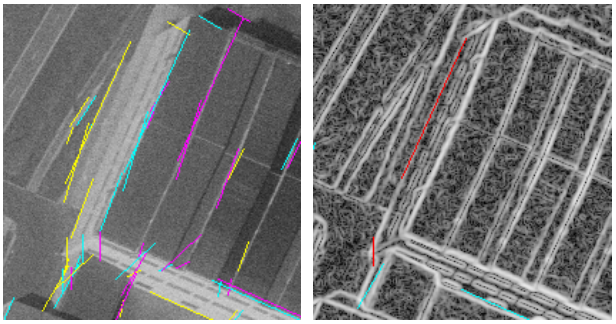

Figure 6: On the left the three best matches using the whole S_h interval; on the right, the match given by the refined search around the ground height.

The segments which have been matched with this refined approach around the supposed ground height are then put in a subset S_l^g and will not be considered in the following. They can be used to build a DTM for instance using a Delaunay triangulation using the heights of the segment extremities.

As for the primitives of $(S_p \cup S_l) \setminus S_l^g$, they should not lie on the ground. They are thus supposed to belong to internal structures of the buildings and should give information on their elevation.

5 BUILDING RECONSTRUCTION

For this preliminary study an external knowledge is used which is a map of the buildings present in the scene. It is used to select the more often detected height inside the building footprint. This is of course a very basic method to detect the building height but the aim here is to analyze the potential of SAR / optical matching for 3D reconstruction.

Let us remark here that no prior knowledge about the building shape is introduced. Specially, the detection of the segments lying on the ground has been done without this map. Particularly, it means that some segments have been falsely suppressed during the ground elevation estimation step if by coincidence they have been matched with an edge in the optical image with an associated height close to h_g: it is especially the case for rooftops with a periodical structure (see Figure 5).

6 RESULT EVALUATION

The results are evaluated using an height map which has been computed by classical stereovision using two optical images. The ground height is well evaluated with 8.5m. The results for the DEM reconstruction highly depend on the kind of building and specially on the structures lying on the roof.

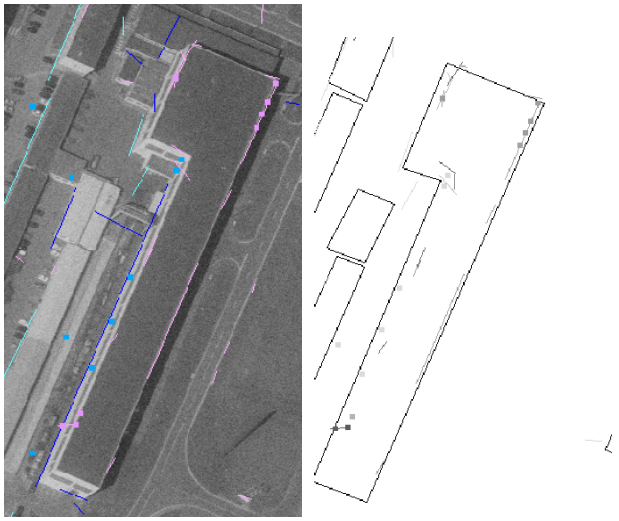

Figure 7: Result on a building: on the left the matched SAR primitives superimposed on the optical image (in blue corresponding to ground matches and in pink corresponding to the building); on the right the building footprint given by the map and the matches.

For instance, for the building shown on figure 7, a mean altitude of 20m is obtained which is the true elevation (i.e an height of 11.5m). This good result is due to the small parapet at the end of the roof which is correctly matched with the corresponding feature in the optical image.

This is quite difficult to give general conclusions about the potential of the method since the SAR image appearance strongly influence the result. But some general remarks can be made, both on the intrinsic limitations of SAR / optical 3D reconstruction, and about the method and the improvements which should be made and are subject of further work.

Concerning the intrinsic potential of SAR / optical 3D information:

- one of the main limitation is of course the assumption that a strong reflector in the SAR image correspond to an edge in the optical data; in fact, it has often been verified on our test set, even for short edge;
- the second main limitation is the nature of the SAR data; indeed for many buildings with rather smooth roof, the backscattered signal is uniform (although noisy); in this case there is no 3D information.

Concerning the improvements to be made:

- some good candidates to match are the parapets ending the roof; a problem arises from the shadow presence in the optical image since a match with higher response can be caused by this shadow (an example is shown figure 8);
- the linear features are supposed to be horizontal (for a segment, both extremities should be at the same height); in practice for part of the roof with a "V" shape, this constraint can provide some problem;
- an important limitation is that only the best match of a SAR primitive is used; furthermore, the method does not introduce any contextual relationship between the primitives (this is done in a coarse way in the estimation of the building height by imposing to all the features to have the same height); in practice the inverse strategy should be tested: an height hypothesis is made for the building and the corresponding score is computed (allowing a small tolerance around the tested height);
- as said before, the suppression of the features corresponding to ground/wall corners is very important (if they are badly matched a wrong height would be given); but on the other hand, the method proposed is too severe and many features are suppressed due to accidental matching at the ground level; both ground estimation and building reconstruction should be performed simultaneously with a complementary constraint: ground features cannot appear inside a building.

Figure 8 presents a building which perfectly illustrate these problems. The predominant height is the good one (26m corresponding to the beginning of the roof) thanks to 6 well matched points and 2 well matched linear features on the roof. But some linear features are matched against the shadow, thus giving a wrong height. Besides, many other linear features of the roof have been suppressed during the ground estimation because of the periodical structure of the roof: they have been matched by accident with the optical image to an height within the ground interval $[h_g - \delta_h; h_g + \delta_h]$.

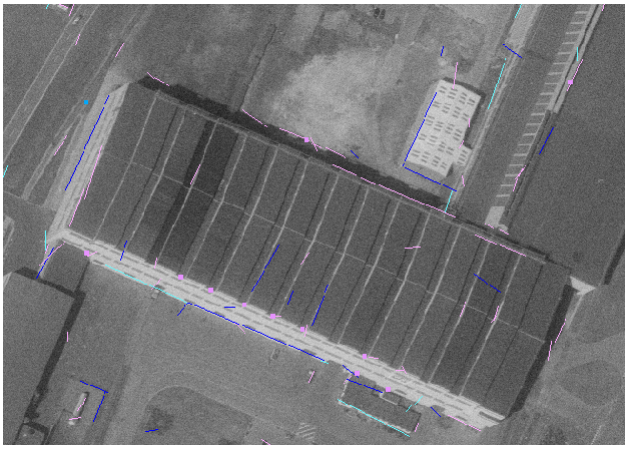

Figure 8: Building with a periodical structure generating some mis-matchings

7 CONCLUSION

This article has presented a preliminary study for height information extraction using a SAR data and an optical data. The results are encouraging if bright features appear in the SAR image since they often correspond to a corner also visible in the optical data. Nevertheless, such method gives very sparse results (only a few features are matched) and the 3D reconstruction step should be made in the same time as the scene interpretation.

Acknowledgment The authors would like to thank Emmanuel Meier and Michel Auger for some algorithmic developments.

REFERENCES

Bloch, I., 1996. Information combination operators for data fusion: A comparative review with classification. IEEE Transactions on Systems, Man and Cybernetics SMC-26(1), pp. 52–67.

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

Dare, P. and Dowman, I., 2000. Automatic registration of SAR and SPOT imagery based on multiple feature extraction and matching. IGARSS'00 pp. 24–28.

Deriche, R., 1987. Using Canny's criteria to derive a recursively implemented optimal edge detector. International Journal of Remote Sensing 2, pp. 167–187.

Franceschetti, G., Iodice, A. and Riccio, D., 2002. A canonical problem in electromagnetic backscattering from buildings. IEEE Transactions on Geoscience and Remote Sensing 40(8), pp. 1787–1801.

Gamba, P., Houshmand, B. and Saccani, M., 2000. Detection and extraction of buildings from interferometric SAR data. IEEE Transactions on Geoscience and Remote Sensing 38(1), pp. 611–618.

Hardaway, G., Gustafson, G. and Lichy, D., 1982. Cardinal effects on Seasat images of urban areas. Photogrammetric Engineering and Remote Sensing 48(3), pp. 399–404.

Hellwich, O., Günzl, M. and Wiedemann, C., 2000. Fusion of optical imagery and SAR/InSAR data for object extraction. International Archives of Photogrammetry and Remote Sensing, Amsterdam XXXIII(B3), pp. 1327–1329.

Lombardo, P., Oliver, C., Pellizeri, T. and Meloni, M., 2003. A new Maximum-Likelihood joint segmentation technique for multitemporal SAR and multiband optical images. IEEE Transactions on Geoscience and Remote Sensing 41(11), pp. 2500–2518.

Lopes, A., Touzi, R. and Nezry, E., 1990. Adaptative speckle filters and scene heterogeneity. IEEE Transactions on Geoscience and Remote Sensing 28(6), pp. 992–1000.

Quartulli, M. and Datcu, M., 2003. Information fusion for scene understanding from interferometric SAR data in urban environments. IEEE Transactions on Geoscience and Remote Sensing 41(9), pp. 1976–1985.

Schistad, A. H., Jain, A. K. and Taxt, T., 1994. Multisource classification of remotely sensed data: Fusion of landsat TM and SAR images. IEEE Transactions on Geoscience and Remote Sensing 32(4), pp. 768–778.

Simonetto, E., Oriot, H. and Garello, R., 2003. Radargrammetric processing for 3-D building extraction from high-resolution airborne SAR data. IGARSS'03, Toulouse (France) pp. 780–782.

Tison, C., Tupin, F. and Maître, H., 2004. Extraction of urban elevation models from high resolution interferometric SAR images. In: EUSAR 2004, Ulm (Allemagne).

Touzi, R., Lopes, A. and Bousquet, P., 1988. A statistical and geometrical edge detector for SAR images. IEEE Transactions on Geoscience and Remote Sensing 26(6), pp. 764 – 773.

Tupin, F. and Roux, M., 2003. Detection of building outlines based on the fusion of SAR and optical features. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 71–82.

Tupin, F., Houshmand, B. and Datcu, M., 2002. Road detection in dense urban areas using SAR imagery and the usefulness of multiple views. IEEE Transactions on Geoscience and Remote Sensing 40(11), pp. 2405–2414.

Tupin, F., Maître, H., Mangin, J.-F., Nicolas, J.-M. and Pechersky, E., 1998. Detection of linear features in SAR images: application to road network extraction. IEEE Transactions on Geoscience and Remote Sensing 36(2), pp. 434–453.