

PRODUCTION OF URBAN DSMS COMBINING 3D VECTOR DATA AND STEREO AERIAL IMAGERY

C. Baillard

SIRADEL, 3 Allée Adolphe Bobière CS 24 343, 35043 Rennes, France – cbaillard@siradel.com

KEY WORDS: DSM, DTM, 3D city models, vector data, aerial imagery, stereo matching

ABSTRACT:

The approach described in this paper combines external 3D vector data and stereo aerial imagery to compute Digital Surface Models (DSMs) and Digital Terrain Models (DTM) in dense urban areas. The method uses a stereo matching algorithm based on dynamic programming. The contribution of this work is the intensive use of vector information at every stage of the process. The vectors define an input elevation map for the matching algorithm. They also locally determine the height range of the scene, in order to reduce research area. They are used to define an adaptive correlation window. The resulting DSM is analysed and filtered according to the input vectors. Finally vector information is used for quality control as reference for assessing the accuracy of the result. The output of the algorithm is a DSM where the complete street network and the building façades are accurately located. A DTM is also computed. The results of the method are presented for 3 test areas with a pixel size of 21 or 28cm. Results show that it is possible to compute a reliable DSM using only two stereo images, as long as appropriate external vector information is available.

1. INTRODUCTION

The automatic production of 3D city models has been a major challenge for years. The availability of such models is valuable for various applications: telecommunication, urban planning, tourism, etc. Thanks to the evolution of new sensors (digital camera, LIDAR data, high resolution satellites) and the development of efficient algorithms, the automatic production of urban DSMS (including buildings and trees) is now possible. Digital cameras provide better quality images than traditional cameras, with an improved contrast in shadows (Maas *et al.*, 2000). Digital images coupled with multiple view algorithms can lead to very good results even over dense urban areas (Cord *et al.*, 1999; Papanoditis *et al.*, 2000; Roux and Maître, 2001). Lidar technology has also evolved enormously over the last few years, and it is now appropriate to urban areas (Maas, 2001; Palmer and Shan, 2001; Zinger *et al.*, 2002). High-resolution satellite imagery is less accurate but still competitive (Fraser *et al.*, 2002).

Although the industrial production of urban DSM has become possible, the automatic computation of 3D vector data is much more difficult. Promising results have been obtained from multiple views (Baillard and Zisserman, 1999 ; Kim *et al.*, 2001), but there always are contours missing in complex dense urban areas (low contrasted building contours, hidden parts, complex-shaped buildings, etc.). In an industrial production context, the building outlines must be accurately represented in all cases, with very few errors. For applications requiring high quality vector data, a manual intervention (or external information) is still necessary, implying high production costs. SIRADEL is a company specialized in the production of urban cartographic data for radio planning applications. For this purpose, a large number of European cities have been manually captured. The produced data sets are characterized by reliable and accurate 3D vectors (buildings, bridges, water, main vegetation). Raster data (DSM, DTM) and image data (orthomosaics) are derived from the vectors. The data accuracy is below 1 meter, and the detail level is fully appropriate to the determinist radio propagation models. Recently, the company has extended its activity to new sectors: real estate, tourism, communication, territorial communities, etc. The need for a new type of DSM has appeared, because the DSM currently

produced is based on a piecewise horizontal surface model, which does not describe oblique roofs and roof superstructures.

The aim of this study is to compute a DSM including oblique roofs and small details, without any additional cost. It should only use available data, i.e. stereo aerial imagery and 3D vector data previously produced for radio network planning. Previous work has focused on the use of external 2D vector information for refining an available DSM (Jibrini *et al.*, 2000; Brenner *et al.*, 2001; Vosselman and Suveg, 2001). The contribution of this work is the intensive use of external 3D vector information during the matching process itself.

Section 2 presents the characteristics of the input data. A method for producing DSM by combining available vector information and stereo images is presented in section 3. Results are discussed in section 4, before conclusion in section 5.

2. DESCRIPTION OF THE INPUT DATA

In this study, the following information has been used:

- Stereo aerial imagery,
- 3D vector data (manually captured by photogrammetric process)

Aerial imagery is available at a scale between 1:15000 and 1:25000 (focal length 157mm), scanned at 14 μ m. The pixel size is between 21 and 35cm. The stereo pairs have an overlap of 60% intra-band and 20% inter-band. The orientation of the images is performed with an analytical stereo plotter and results in an absolute positioning accuracy of 30cm. Each image pair is re-sampled using the epipolar geometry.

The external 3D vectors have been previously produced by manual photogrammetric capture and automatic topology analysis (see example Figure 1). Their absolute accuracy is around 0.60m both in planimetry and altimetry. The vector data are classified into two groups: ground and aboveground.

The ground is defined by a set of 3D lines including roads, railway, hydrography, breaklines (orographic characteristics of the terrain). The aboveground is defined by a set of polygons and points, describing buildings, vegetation, and bridges. The 3D points characterize the highest point inside each polygon. Building blocks are divided into several independent adjacent polygons as soon as the difference in height between two adjacent buildings exceeds 2 meters. The polygons are

topologically consistent, with no overlapping and no intersection. Each aboveground object is thus characterized by :

- A reliable closed 2D contour (polygon),
- Approximate z-values for the contour
- A 3D point characterizing its maximal elevation

In particular, the minimal and the maximal elevations of each polygon are known.



Figure 1: Example of external data (Deauville)

3. PRESENTATION OF THE METHOD

3.1 General Strategy

Due to the low overlap of source images, we are limited to two-view matching techniques. Pixel matching has been preferred to model-based matching techniques because it is more appropriate to complex dense urban areas. External vector data are used as much as possible, in order to compensate for the poverty of the input data. The method is summarized in Figure 2, and is divided into 4 main stages:

1. Initialisation of input data
2. Image matching using input maps derived from vectors (starting at lowest resolution),
3. DSM analysis and filtering using input vectors,
4. Auto-evaluation of the method using reference vectors.

The four stages of the method are described in the following subsections.

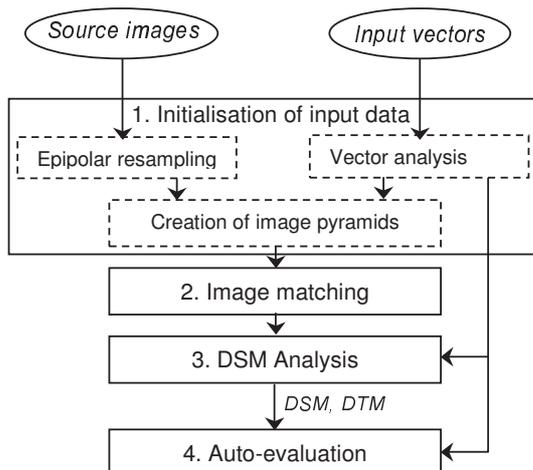


Figure 2: Overview of the method

3.2 Initialisation of input data

The initialisation consists of 3 steps:

- Image re-sampling using epipolar geometry,
- Vector analysis for the creation of “input maps”,
- Creation of image pyramids (multi-resolution scheme).

Vector Analysis

First a set of reference vectors is selected from the input data set. They consist of “reference ground points” and “reference building points”. The reference ground points consist of randomly selected 10% of the input ground vector data. The reference building points consist of all 3D points characterizing the maximal height of each building polygon. These reference points will be used for assessing the accuracy of the results.

Then 3 kinds of maps are computed by projecting the input vectors onto image planes (see examples in Figure 4). These maps will be used as input data by the matching algorithm, along with the stereo pair :

- Input elevation map,
- Building maps,
- Elevation range maps.

The **input elevation map** is obtained by mapping z-values of input vectors (other than the reference points) onto the left image plane. Only the points “visible” in both images (not hidden by an above-ground object) are kept.

The **building maps** are obtained by projecting the 3D building contours onto each image plane (left and right). Each building polygon is associated to a unique label in each map. The hydrography is also mapped onto these images.

The **elevation range maps** consist of 2 maps ZMIN and ZMAX named “minimal elevation map” and “maximal elevation map”. They are computed using three parameters (see schema in Figure 3):

- **max_ground_diff**: maximal difference between the terrain elevation $z(x,y)$ and the elevation $z_0(x,y)$ estimated from the input ground vector data,
- **max_aboveground_diff**: maximal difference between the elevation of an input above-ground vector point, and the elevation of the corresponding object in the real scene.
- **max_height**: maximal height of aboveground objects.

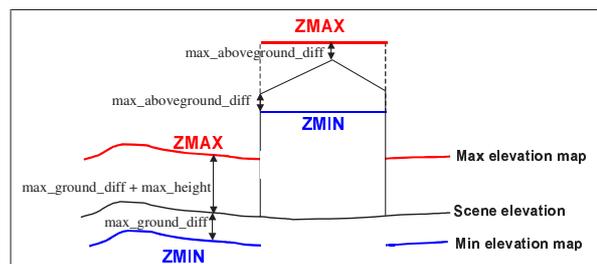


Figure 3: Computation of elevation range maps ZMIN and ZMAX

Creation of image pyramids

The input aerial images and the created maps are finally sub-sampled at various resolution levels in order to create image pyramids (3 levels are used).

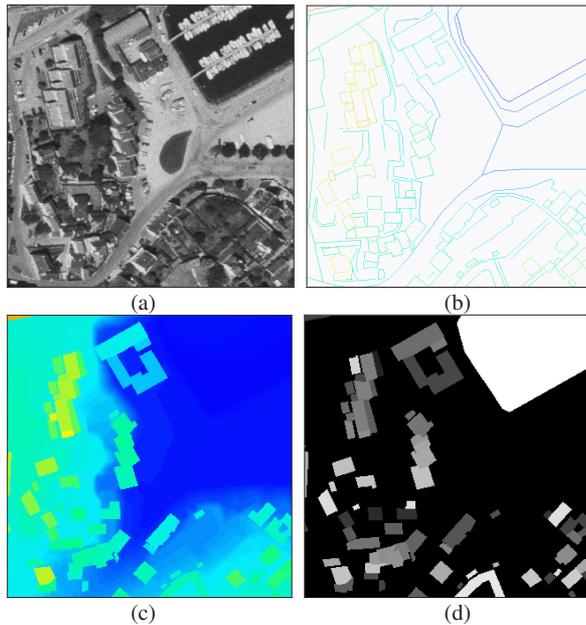


Figure 4: example of maps derived from input vectors.
 (a): left source image (extract from Figure 6a) ;
 (b): input elevation map;
 (c): minimal elevation map;
 (d): building map in left image.

3.3 Image matching

The matching algorithm is based on dynamic programming and derived from the method described in (Baillard and Dissard, 2000). The matching problem is expressed as an optimisation problem for each epipolar line pair. It consists of finding a path with minimal cost in a 2D graph defined by conjugate epipolar lines. The cost function is recursively computed using 2 types of elementary cost functions :

- $match(i,j)$ characterizing the match between pixels i and j ,
- $occ(i)$ characterizing the occlusion of pixel i (when pixel i can not be seen in one image).

This method is particularly appropriate to dense urban scenes, because each height discontinuity along scanline in one image is explicitly associated to an occlusion in the other image. The matching cost $match(i,j)$ is directly related to the correlation score between the 2 images at pair (i,j) ; the occlusion cost $occ(i)$ is constant. Dynamic programming is applied within 2 steps: first with a strong similarity constraint (low occlusion cost), then with a strong geometry constraint (high occlusion cost). The algorithm presented in (Baillard and Dissard, 2000) has been extended in two ways : it has been integrated within a multi-resolution scheme (the correlation result at one resolution level is used to update the elevation range maps at the next level, see Figure 5), and the input maps previously discussed are taken into account as follows:

- The input elevation map constrains the optimal path by defining input matched points,
- The elevation range maps define allowed and forbidden areas for the path,
- The building maps are used to weight the correlation score between 2 pixels: only the pixels of the correlation window with the same label as the central point in the building map are taken into account into the correlation score.

The elevation map computed at the last resolution level is finally projected onto the object space to produce a raw DSM.

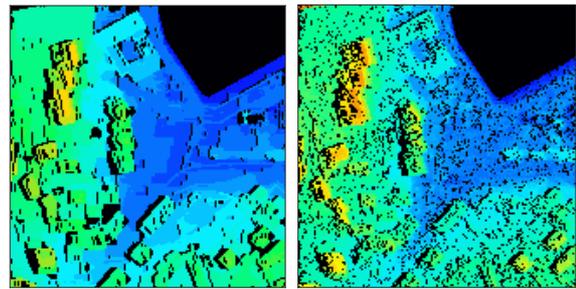


Figure 5: computed elevation maps at resolution level 3 (left) and resolution level 1 (right)

3.4 DSM analysis

This final stage aims at removing noise from the raw DSM and deriving a DTM from it.

Computation of a DTM

First “reliable” ground points are selected from the DSM by combining various criteria and filtering methods:

- Study of relative height (comparison with neighbourhood),
- Study of the slope,
- Minimum region size (regions defined by homogeneous height),
- Minimum correlation score,
- Top hat morphological filtering.

The selected points are then sampled and triangulated in order to derive a dense DTM.

Computation of the final DSM

The aboveground objects corresponding to input polygons are corrected. In particular, the location of the building façades is refined in order to match the input contours. Additionally, uncaptured but “significant” aboveground objects (trees, sheds, etc) are detected by a joint analysis of DTM and raw DSM.

The final DSM is obtained by superimposition of the aboveground objects with the DTM.

3.5 Auto-evaluation

Two kinds of assessments are performed. The accuracy of the ground points (taken from DTM) is assessed using the reference ground points. The accuracy of aboveground points (taken from DSM) is assessed using the reference building points (highest point of each building). The assessment results in 6 characteristic figures: number of reference points, error bias and root mean square, computed for both ground and above-ground.

4. EXPERIMENTAL RESULTS

4.1 Description of test data

Three test areas are presented:

- Deauville1, pixel size=28cm, area 589mx589m = 0.34km²
- Deauville2, pixel size=28cm, area 587mx587m = 0.34km²
- Kerlaz, pixel size=21cm, area 437mx437m= 0.19km²

Input images are shown in Figures 6a, 7a, 8a. Input vectors are shown in Figures 6b, 7b, 8b. The algorithm requires no tuning, and all internal parameters were fixed during the tests.

4.2 Result analysis

The results are illustrated in Figures 6c-f, 7c and 8c. The computed accuracy is summarized in Table 1.

Table 1 : Accuracy of the computed DTM / DSM (in meters)

	Ground			Aboveground		
	Npts	Bias	RMS	Npts	Bias	RMS
Deauville1	54	-0.12	0.95	318	-0.73	1.11
Deauville2	130	0.07	1.29	563	-0.77	1.46
Kerlaz	254	0.14	0.94	698	-0.73	1.08

The accuracy of the computed DTM is around 1m. It is slightly more on Deauville2 because of a strong slope (more than 60m difference in height over a distance of 400m).

The final DSM describes all buildings and the street network. In Deauville2 (see Figure 8c), the street network clearly appears although it is not entirely visible in the images. Thanks to the external vector data, the building façades are accurately located. Most oblique roofs are clearly represented, as well as single trees (see zoom on Kerlaz, Figure 6f). More generally, the algorithm copes very well with big slopes and high buildings. The accuracy of the reference building points varies between 1.08m (Kerlaz) and 1.46m (Deauville2), with a systematic negative bias around 0.75m (standard deviation around 1m). This is due to the definition of the reference points, which are the highest points on the building roofs: like any algorithm using a correlation window in image space, our algorithm tends to smooth thin peaks like roof ridges.

5. CONCLUSION

A method for computing DSM in dense urban areas has been presented. It combines stereo aerial imagery and 3D vector data. The vector information is used in 5 different ways:

- the vector data define a 3D wire frame model initialising the algorithm;
- the elevation range is given by the input vector data;
- building polygons are used as correlation masks for computing scores (adaptive correlation window);
- the computed DSM is filtered using vector information;
- a part of the input vectors is kept for quality control, in order to assess the accuracy of the final DSM.

The accuracy of resulting DSM and DTM varies between 1 et 1.50m depending on the scene complexity. It demonstrates that it is possible to compute a good-quality DSM using only two stereo images, as long as appropriate external vector information is available. Additionally, the accuracy of the results is assessed during the process, which is a major advantage for industrial production.

Although the resulting accuracy is acceptable for some applications, the quality of the results could certainly be improved by using multiple views, when these are available. Further work should focus on DSM refinement using building reconstruction techniques. In particular, vector information can be used for instantiating roof models when possible (Jibrini, 2000; Brenner, 2001; Vosselman, 2001). Finally, a complementary study will determine how the quality of the input vectors does effect the accuracy of the final DSM.

REFERENCES

C. Baillard, O. Dissard, 2000. A Stereo Matching Algorithm for Urban Digital Elevation Models. In *ASPRS Photogrammetric Engineering and Remote Sensing*, 66 (9), pp.1119-1128.

C. Baillard, A. Zisserman, 1999. Automatic reconstruction of piecewise planar models from multiple views. In *Proc. of CVPR'99*, Fort Collins, pp. 559-565.

C. Brenner, N. Haala, D. Fritsch, 2001. Towards fully automated 3D city model generation. In *Automatic Extraction of GIS Objects from Digital Imagery*, *IAPRS Vol. 32*, pp.47-57.

M. Cord, M. Jordan, J.-P. Cocquerez, N. Paparoditis, 1999. Automatic extraction and modelling of urban buildings from high resolution aerial images. In *Automatic Extraction of GIS Objects from Digital Imagery*, *IAPRS Vol. 32*, pp 187-192.

C. S. Fraser, E. Baltsavias, A. Gruen, 2002. Processing of Ikonos imagery for submetre 3D positioning and building extraction. *ISPRS j. of Photogrammetry & Remote Sensing*, 56(3), pp. 177-194.

H. Jibrini, M. Pierrot Deseilligny, N. Paparoditis, H. Maître, 2000. Automatic building reconstruction from very high resolution aerial stereopairs using cadastral ground plans. In *Geoinformation for all, Int. Archives of Photogrammetry and Remote Sensing, Vol.33*, Amsterdam (NL), July 2000.

Z. Kim, A. Huertas, R. Nevatia, 2001. A model-based approach for multi-view complex building description. In *Automatic Extraction of Man-Made Objects from Aerial and Space images (III)*, pp. 181-193.

H. G. Maas, C. Thom, J. P. Souchon, C. K. Toth, 2000. Digital camera experts present their views, "Three-Line Linear" versus "Multi-Head Array". *GIM International*, pp. 68-71, may 2000.

H. G. Maas. The suitability of airborne laser scanner data for automatic 3D object reconstruction, 2001. In *Automatic Extraction of Man-Made Objects from Aerial and Space images (III)*, pp. 291-296.

T. C. Palmer, J. J. Shan, 2001. Urban modelling from Lidar data in an integrated GIS environment. In *Proc. of ASPRS Conference*, Saint Louis, MO, may 2001.

N. Paparoditis, C. Thom, H. Jibrini, 2000. Surface reconstruction of urban areas from highly overlapping digital aerial images. In *Geoinformation for all, Int. Archives of Photogrammetry and Remote Sensing, Vol.33*, July 2000.

M. Roux, H. Maître, 2001. Some more steps towards 3D reconstruction of urban areas from multiple views. In *Automatic Extraction of Man-Made Objects from Aerial and Space images (III)*, pp. 135-147.

G. Vosselman, I. Suveg, 2001. Map based building reconstruction from laser data and images. In *Automatic Extraction of Man-Made Objects from Aerial and Space images (III)*, pp. 231-239.

S. Zinger, M. Nikolova, M. Roux, H. Maître, 2002. 3D resampling for airborne Laser data of urban areas. In *IAPRS Vol.34*, n° 3A, pp. 418-423.

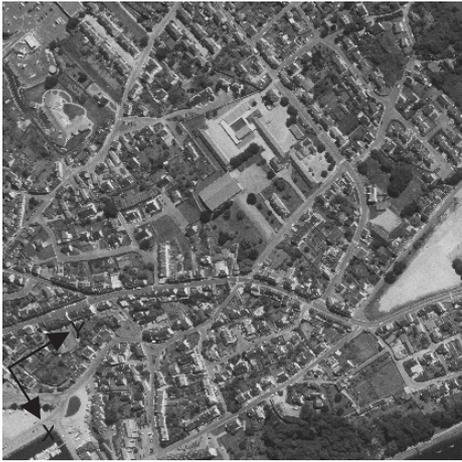


Figure 6a: Kerlaz , left image

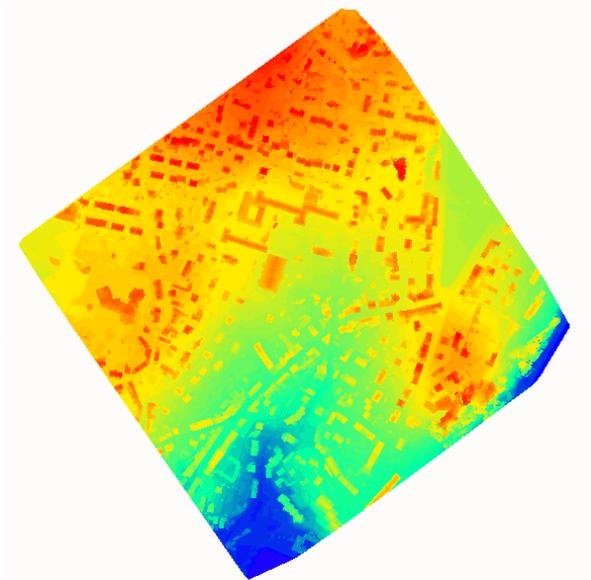


Figure 6d: Kerlaz, final DSM after post-processing



Figure 6b: Kerlaz, input vectors and footprint of the image

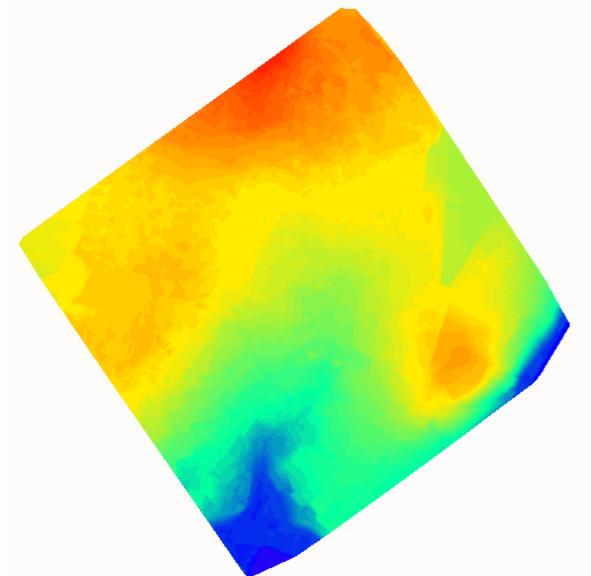


Figure 6e: Kerlaz, final DTM

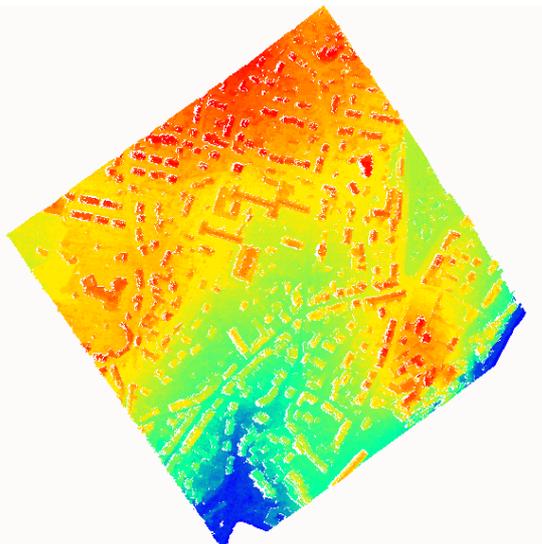


Figure 6c: Kerlaz, raw DSM after image matching
(height range 0-52m)

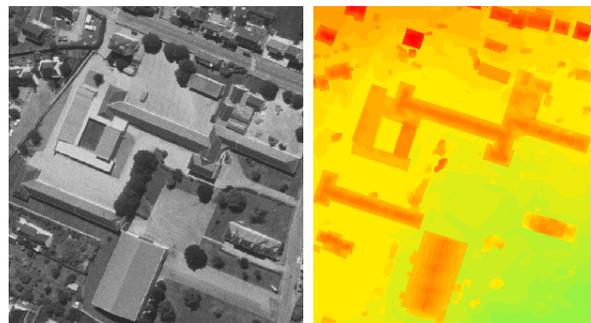


Figure 6f: Kerlaz, zoom on source image (left) and DSM
(right); Oblique roofs and single trees are clearly represented



Figure 7a: Deauville1, left image



Figure 8a: Deauville2, left image



Figure 7b: Deauville1, input vectors and footprint of the image

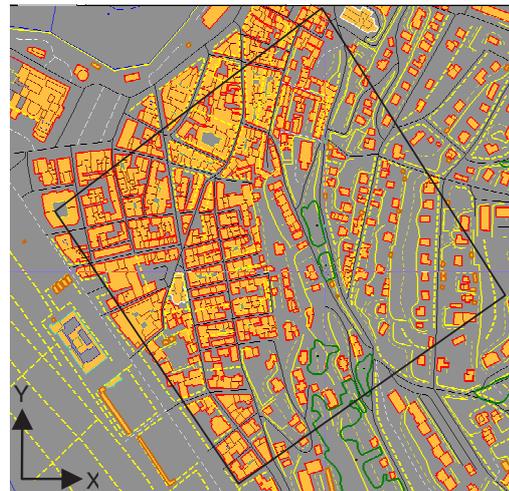


Figure 8b: Deauville2, input vectors and footprint of the image

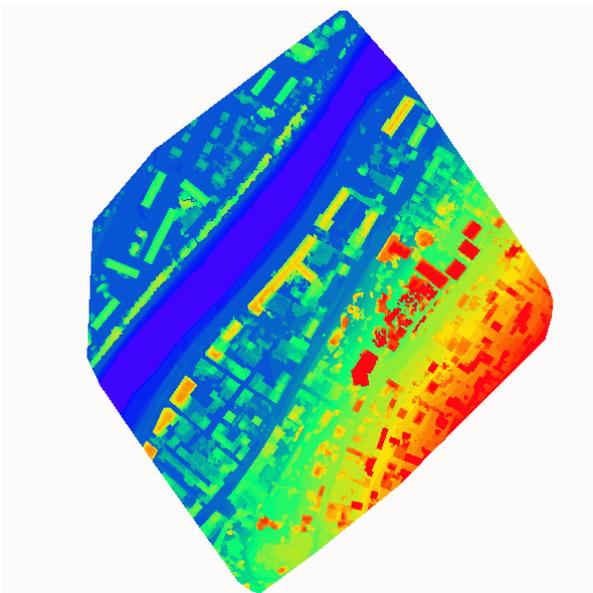


Figure 7c: Deauville1, final DSM
(height range 0-47m)

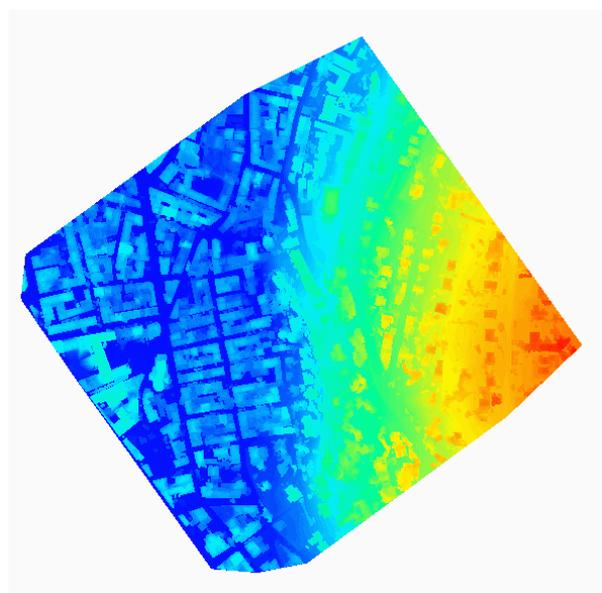


Figure 8c: Deauville2, final DSM
(height range 1-88m)