ORTHO-RECTIFIED FACADE IMAGE BY FUSION OF 3D LASER DATA AND OPTICAL IMAGES

Ayman Zureiki and Michel Roux

Institut TELECOM; Telecom ParisTech; CNRS LTCI
46 rue Barrault, F-75634 Paris Cedex 13, France.
{Ayman.Zureiki, Michel.Roux}@Telecom-ParisTech.fr

KEY WORDS: Facade ortho-image, point cloud, data fusion, multi-sensor, 3D reconstruction

ABSTRACT

Obtaining three-dimensional models of urban scenes from sensor data is a challenging topic. The construction of 3D city models requires fusion of data coming from different sensors, such as laser scanners and cameras. At a first phase, we will process building facades, which will be considered as textured 3D planes. The acquisition system is equipped with two laser scanners and many (twelve) cameras. The laser scanner provides 3D point clouds of a huge size. We detail how to segment the point cloud in order to compress the information to be included in the final model and to extract the principal plane. Afterwards, we compute the homography that maps the camera image to the facade image (image fixed on the facade porter plane). The important point in our approach is that the obtained facade images are totally geo-referenced, i.e. the 3D position of each pixel is known in a global (world) reference frame.

1 INTRODUCTION

Three-dimensional modelling of urban districts is an intersting and challenging topic. It could open the way for many different applications such as urban planning, transport management, navigational aids, civil security, historical preservation, and more odd aspects such as museography, cultural and touristic information services, and virtual reality applications.

Building of coherent 3D models of large outdoor scenes by using real sensors data gains more interest recently. Scanning every square metre of an area enables the production of a geospecific representation of a whole city (even on the scale of Paris, for example). Our research is done according to the project TerraNumerica, which tries to develop a set of necessary technologies to produce large-dimension and high-resolution 3D representations of urban territories automatically and accurately.

An important issue concerning the generated 3D representation is that the final result is often a dense complicated mesh. So, considerable postprocessing (mesh simplification, hole filling) is required to make this representation more usable by graphic applications. (Stamos et al., 2006) present a 3D modelling approach that models large planar surfaces of the scene as planar primitives (extracted via a segmentation pre-process), and non-planar areas as mesh primitives. In that respect, the final model is significantly compressed. Our approach will be similar, but we will use indeed optical cameras to create ortho-images of extracted facade planes. In fact, in 3D city modelling, real texture data (images) are required to make buildings more realistic and authentic, and easier to be recognised by a user. (Kang et al., 2005) present an algorithm employing both the constraint of straight lines bundle and the constraint of known orientation of parallel lines in object space in order to generate the whole facade texture for each building by a strip method. They use the vanishing points detection (Kalantari et al., 2008) to produce the ortho-image.

In this paper, we use data fusion (laser - camera) to obtain the facade image. The major point in our approach is that we suppose that the facade image is fixed on the principal plane of the facade which is extracted by point cloud segmentation. By assuring a fix ratio pixel-meter in the facade image, our facade image will have a metric information, i.e. the 3D position of each pixel in the facade image will be known. Another important aspect, the ratio pixel/meter will be the same for all facades images.

The expected model will be a textured three-dimensional geo-referenced model, i.e. a model where each point and surface is referenced in a global (world) frame, with additional colorimetric information. In this paper, we will consider buildings facades only, and we aim to obtain a textured 3D plane representation of facades as a first and simple model. In this context, we use sensors data acquired by our partners in the project: IGN [1] (*Institut Géographique* National) and Trimble [2]. The IGN provides us with registered 3D point clouds and geo-referenced images. The IGN uses a special vehicle (called Stereopilis) equipped with two laser scanners and with twelve cameras, in addition to a localisation system based on GPS and an inertial measurement unit. Trimble provies other geo-referenced images of high resolution.

[1] http://www.ign.fr/
[2] http://www.trimble.com/

This article is organised as follow: section 2 describes the segmentation of a point cloud and the extraction of planar surfaces. In section 3 we detail how to compute the homography between camera image and facade image, an image that is fixed on the 3D plane extracted by the segmentation step. Experimental results are presented in section 4. We conclude in section 5 and propose some future works.

2 POINT CLOUD SEGMENTATION AND PLANE EXTRACTION

3D sensors (laser range finder, stereo vision, etc.) provide point clouds (sets of 3D points) of different sizes. Compressing such a point cloud into some features (surfaces or regions) without loosing essential information is of great importance. This is known as the 3D segmentation problem: how to divide the point cloud into features, i.e. how to bind each point with a label identifying to which feature it belongs, so that points belonging to the same surface or region are given the same label. Segmenting a point cloud is generally considered a very difficult problem due to very large size of datasets, complexity, and variation of numbers and types of features that could be present in the scene.

Planar segmentation is the extraction of planes from a point cloud. It has been well studied in computer graphics in order to perform real-time rendering of complex models (Heckbert and Garland, 1997). There is a major difference between 3D urban reconstruction and computer graphics. Data in 3D reconstruction are issued from sensors and thus they are erroneous, while models in computer graphics are supposed to be without errors. The decimation algorithms in computer graphics aim to accelerate rendering and not to deal with errors.

(Horn and Schmidt, 1995) extract plane using Hough Transformation, which converts the complex problem of model fitting into finding the peak in the parameters space (Illingworth and Kittler, 1988). (Sequeira et al., 1999) use a hybrid method of region-based and edge-based to perform the segmentation and assure the alignment of consecutive data using an Iterative Closest Point algorithm (ICP). (Liu et al., 2001) use Expectation Maximisation (EM) to create a 3D map of planar segments. (Kohlhepp et al., 2004) extract planes in real time by using a grouping algorithm of scan lines. This algorithm assembles neighbour line segments in an efficient way, but it requires data line segmentation in each scan line.

(Hähnel et al., 2003) propose a simplification algorithm based on mesh simplification. They extract planes by using an approach based on region-growing by starting from an arbitrary point, then trying to enlarge the region in all directions. (Weingarten, 2006) proposes some improvement to this algorithm by starting region seed from the most flat point in the cloud (minimum local error), and by taking advantages of the range image structure to simplify the research of neighbour points. (Harati et al., 2007) proposes a method based on bearing angle, which is the angle of the laser beam with the reflecting surface.

Region-growing algorithms perform better when the point cloud is structured as a depth image, in which each point (pixel) has neighbourhood relations (left, right, up, down) with neighbour pixels. In a non-structured point cloud, these relations are not defined, which make region-growing methods less appropriate.

2.1 Our approach

Our acquisition system provides point clouds of very large size. For example, for a part of Soufflot street (in Paris), the point cloud of one side of the street contains more than 3 million points. The obtained point cloud is not structured. The segmentation approach is composed of three steps:

- dividing the point cloud into blocks representing Buildings.
- extracting the main plane (facade) of each block.
- generating the ortho-facade image as we detail in section 3.

For the building detection, first we distribute the 3D points into cells of an octree. Then we use connected component analysis to separate the point cloud into different connected zones, each representing a building. The resulting set of points of each building has an important thickness due to the presence of balconies and sculptures in the facades. So we must use a robust method for determining the principal plane likes RANdom Sample Consensus (RANSAC) (Fischler and Bolles, 1981). In the later algorithm, we randomly choose three points, find the plane passing through them, then polling all the other points to see whether they are on the plane (with a predefined threshold) or not, we repeat this operation a certain number of times (given by the algorithm), and keep at the end the plane which has the larger number of points with the smaller accumulated global error.

Figure 1 illustrates a point cloud of a part of Soufflot street, figure 2 represents a building block, where the points belonging to the facade main plane are represented in figure 3.

2.2 Facade Local Reference Frame

Let Π be a 3D plane defined by its normal vector $\mathbf{n} = \{\varphi, \psi\}$ and by its distance to the origin O_w in the global reference fame R_w (see figure 4).

We are looking for an orthonormal reference frame for this plane. We choose the barycenter point B of the 3D points belonging to this plane as the origin (O_p) of the local frame R_p, and the axis Y_p to be parallel to the normal vector $\mathbf{n}$. Let $\vec{i}_w, \vec{j}_w, \vec{k}_w$ be the unit vectors along axes O_wX_w, O_wY_w, O_wZ_w respectively, and $\vec{i}_p, \vec{j}_p, \vec{k}_p$ be the unit vectors along axes O_pX_p, O_pY_p, O_pZ_p respectively. Let $\vec{i}_p$ be:

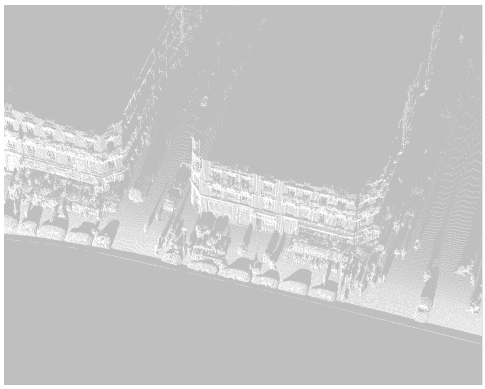

Figure 1: Point cloud

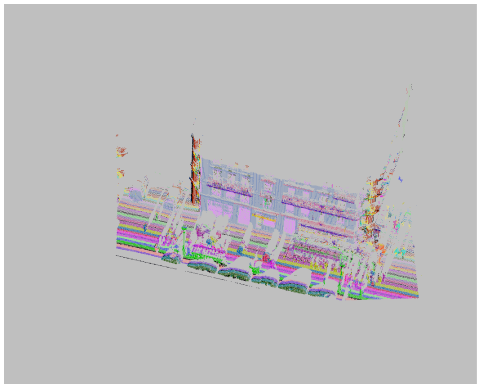

Figure 2: Building block

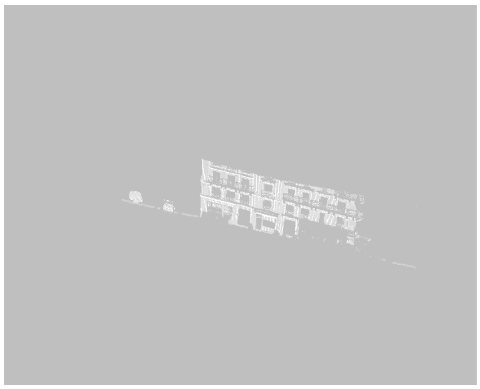

Figure 3: Facade main plane points

$$\vec{i}_p = \begin{bmatrix} \sin\varphi \\ -\cos\varphi \\ 0 \end{bmatrix} \quad (1)$$

This vector can be interpreted as the unit vector of the intersection line between the plane Π and the plane $Z_w = 0$ (if they are not parallel).

with the following notation:

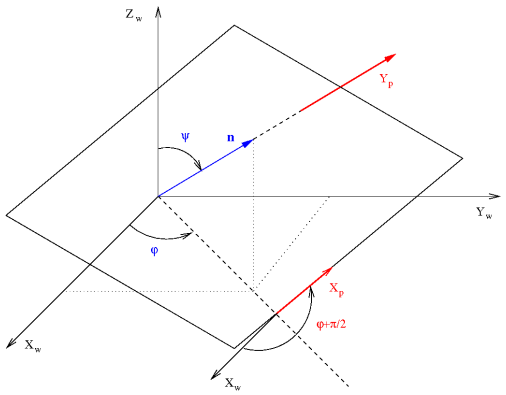

Figure 4: Local 3D plane reference frame

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \mathbf{R}_{w,p} \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} + \mathbf{t}_{w,p} \quad (2)$$

The rotation matrix and the translation vector from global to the local frames are:

$$\mathbf{R}_{w,p} = \begin{bmatrix} \sin\varphi & \cos\varphi\sin\psi & -\cos\varphi\cos\psi \\ -\cos\varphi & \sin\varphi\sin\psi & -\sin\varphi\cos\psi \\ 0 & \cos\psi & \sin\psi \end{bmatrix} \quad (3)$$

$$\mathbf{t}_{w,p} = \begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} \quad (4)$$

3 GENERATION OF FACADE ORTHO-IMAGE

Plane extraction from the point cloud provides the principal plane of a facade, we must generate now its image. Acquisition of facade image is a delicate issue for many reasons. First, camera images are taken from different distances and viewing angles for different buildings, so there is not a unique zoom factor for all them. Second, the acquisition system have 12 cameras, and even though they are of the same type, their intrinsic parameters are not identical. Third, we may use different images from different cameras of from the same camera with different viewing point to generate a facade image. In order to resolve this problem, our approach is to unify the resolution of the resulting facade image by adding metric information to each pixel. In fact, we impose a fix ratio pixel/metre for all facades' images, so that for example, a pixel in the resulting facade image corresponds to a real length of $1cm$ in the facade 3D plane. In addition, we suppose that a facade image is

an image fixed to the 3D main plane of this facade (thus, we neglect errors due to balconies and sculptures).

We will compute the facade image by a homography between the camera image plane and the facade 3D plane.

3.1 Homography Calculation

Definition: The mathematical map relating different views of a planar surface by one or many projective cameras is a homography (Faugeras et al., 2001).

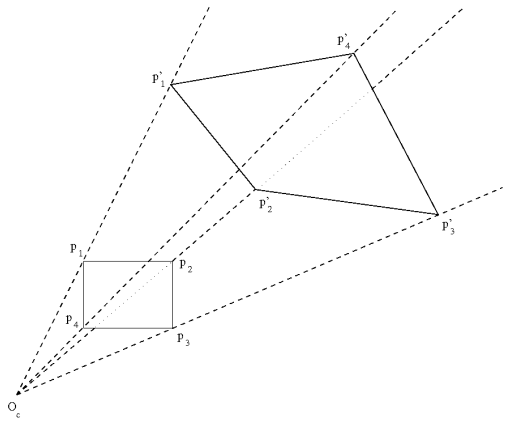

Figure 5: Homography Calculation

We detail how to project a camera image onto a facade image. This is done by a homography relating the two planes. A bi-dimensional homography is a transformation relating the projections $p = (u, v)$ and $p^{'} = (u^{'}, v^{'})$ of each 3D point P belonging to a plane Π, see figure 5. A Homography is defined by a 3x3 matrix $\mathbf{H}$.

$$\lambda \begin{pmatrix} u_i^{'} \\ v_i^{'} \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \quad (5)$$

hence:

$$\begin{cases} u_i^{'} = \frac{h_{11}\, u_i + h_{12}\, v_i + h_{13}}{h_{31}\, u_i + h_{32}\, v_i + 1} \\ v_i^{'} = \frac{h_{21}\, u_i + h_{22}\, v_i + h_{23}}{h_{31}\, u_i + h_{32}\, v_i + 1} \end{cases} \quad (6)$$

Thus, to compute a 2D homography, we need at least 4 corresponding points. To find the matrix $\mathbf{H}$, we can resolve the following linear system $A\,X = B$ with:

$$A = \begin{bmatrix} \vdots & & & & & & & \\ u_i & v_i & 1 & 0 & 0 & 0 & -u_i\, u_i^{'} & -v_i\, u_i^{'} \\ 0 & 0 & 0 & u_i & v_i & 1 & -u_i\, v_i^{'} & -v_i\, v_i^{'} \\ \vdots & & & & & & & \end{bmatrix} \quad (7)$$

$$X = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \qquad B = \begin{bmatrix} \vdots \\ u_i^{'} \\ v_i^{'} \\ \vdots \end{bmatrix} \quad (8)$$

If there are more than 4 corresponding points, the linear system could be solved by pseudo-inversion of the matrix A.

3.2 Resampling of Facade Image

Once the homography matrix $\mathbf{H}$ is determined, we use it to calculate the pixels of the facade image. In fact, for a pixel in the facade image, we transform it with the matrix $\mathbf{H}$, which give us a point in the camera image surrounded by 4 pixels. A bi-linear interpolation is used to deduce the facade pixel value.

We note that when using more than one camera image to construct the facade image, we may have more than one candidate pixel value due to overlapped zones in cameras' images. We may use mosaic strategies to determine the values of facade image pixels (we don't present this approach in the current paper).

In our approach, we keep three-dimensional information of facade image pixels, i.e. we keep information about the transformation between the global and local frames (as we detailed in section 2.2) and pixel/metre ratio information. Hence, we can supply the 3D coordinates (x_w, y_w, z_w) in the global frame for any pixel (i, j) in the facade image, and vice-versa. This has a special interest in the three-dimensional model of urban scenes. In practice, this can be translated as : "What are the 3D coordinates of the second window of the third floor in the building number 13 street Soufflot in Paris in a global frame!".

4 RESULTS AND DISCUSSION

Figure 6 illustrates an image taken by a camera for the building at 13 Soufflot street in Paris, and the figure 7 represents the facade ortho-image resulting from our algorithm. Another example is given by the figure 8 in which the camera is mounted on the acquisition vehicle in vertical manner. The rectified facade image is given in figure 9. We can notice that the facade ortho-image is well rectified, which can point out that the facade holding plane is well estimated, and our acquisition system is well calibrated. There are still other sources of errors like error in camera localisation (GPS uncertainty), errors in point cloud registration, and distorsion of camera images which we have neglected in our work

(Kang et al., 2005) and (Kalantari et al., 2008) propose an algorithm using automatic detection of vanishing points

Figure 6: Example 1: initial camera image

Figure 7: Example 1: facade ortho-image

Figure 8: Example 2: initial camera image

Figure 9: Example 2: facade ortho-image

in urban scene images. They used it to rectify facade images. Even if the obtained images are of good quality and well rectified, they are of unknown scale factor, i.e. there is no information about the ration pixel:metre. This comes from the fact they use only one information source (one camera). Whereas, in our approach, the use of two information sources (3D laser + camera) let us keep a unique scale factor for all the resulting facades images. This illustrates benefits we could gain by using multi-sensor data fusion!

5 CONCLUSION

In this article, the segmentation of point cloud is detailed with emphasis on extraction of planar surfaces. Our segmentation approach to extract facade main plane was tested on point clouds acquired by laser scanning. The results, which appear reliable, were used immediately in the generation of the facade image. Then we have proposed a method to resample the facade ortho-image based on multi-sensor data fusion. Our approach used the 3D plane extracted by the segmentation step as a porter of the facade image, then calculate the homography relating the camera image plane and the facade plane. Our main contribution resides in the addition of metric information to the obtained

facade images. In fact, we keep 3D information for each pixel in the facade image. Hence, we can find the 3D position in the global reference frame for each pixel in the facade image, and vice-versa.

In future work, we plan to perform more experimental evaluation in order to construct a textured 3D model on an entire district (for example). Taking into consideration the distorsion of camera image and other sources of errors will be studied also.

Currently we consider only planar surfaces, so in the next step we envisage detection and modelling of other artifacts (landlamp, cars, etc.) in order to obtain a more realistic 3D city model.

ACKNOWLEDGEMENTS

The work reported in this paper has been performed as part of the Cap Digital Business Cluster TerraNumerica project. The authors would like to thank in particular the Institut Géographique National(`http://www.ign.fr/`) and Trimble (`http://www.trimble.com/`), for providing their georeferenced data.

REFERENCES

Faugeras, O., Luong, Q.-T. and Papadopoulou, T., 2001. The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications. MIT Press.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM pp. 381–395.

Hähnel, D., Burgard, W. and Thrun, S., 2003. Learning compact 3D models of indoor and outdoor environments with a mobile robot. Robotics and Autonomous Systems 44(1), pp. 15–27.

Harati, A., Gachter, S. and Siegwart, R., 2007. Fast range image segmentation for indoor 3D-SLAM. In: The 6th IFAC Symposium on Intelligent Autonomous Vehicles (IAV), Toulouse, France.

Heckbert, P. S. and Garland, M., 1997. Survey of polygonal surface simplification algorithms. Technical report, Carnegie Mellon University.

Horn, J. and Schmidt, G., 1995. Continuous localization for long-range indoor navigation of mobile robots. In: IEEE International Conference on Robotics and Automation, pp. 387–394.

Illingworth, J. and Kittler, J., 1988. A survey of the Hough transform. Computer Vision, Graphics, and Image Processing 44(1), pp. 87–116.

Kalantari, M., Jungd, F., Paparoditisa, N. and Guedon, J.-P., 2008. Robust and automatic vanishing points detection with their uncertainties from a single uncalibrated image, by planes extraction on the unit sphere. In: ISPRS Congress, Proceedings of Commission III, Beijing, China.

Kang, Z., Zhang, Z. and Zhang, J., 2005. A strip method of image mosaic for the vehicle-based imagery. In: Proceedings of the 4th International Conference on Environmental Informatics (ISEIS), Vol. 3, pp. 306–314.

Kohlhepp, P., Pozzo, P., Walther, M. and Dillmann, R., 2004. Sequential 3D-SLAM for mobile action planning. In: Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems (IROS), pp. 722–729.

Liu, Y., Emery, R., Chakrabarti, D., Burgard, W. and Thrun, S., 2001. Using EM to learn 3D models of indoor environments with mobile robots. In: International Conference on Machine Learning (ICML), San Francisco, CA, USA, pp. 329–336.

Sequeira, V., Ng, K., Wolfart, E., Goncalves, J. and Hogg, D., 1999. Automated reconstruction of 3D models from real environments. ISPRS Journal of Photogrammetry and Remote Sensing 54(1), pp. 1–22.

Stamos, I., Yu, G., Wolberg, G. and Zokai, S., 2006. 3D modeling using planar segments and mesh elements. In: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT), pp. 599–606.

Weingarten, J., 2006. Feature-based 3D SLAM. PhD thesis, École Polytechnique Fédérale de Lausanne.