Eurographics Workshop on Urban Data Modelling and Visualisation (2014)
V. Tourre and G. Besuievsky (Editors)

Incremental plane-based reconstruction and exploration of buildings in old postcards

Lara Younes , Barbara Romaniuk and Eric Bittar

University of Reims Champagne-Ardenne, CReSTIC - SIC
Reims, France

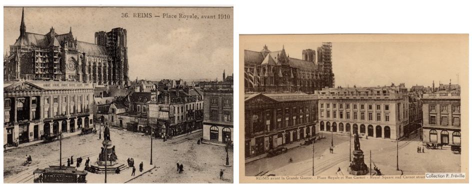
Abstract

We present an approach for 3D reconstruction of a city model over the time from a collection of old postcards of the city of Reims. The planar structure of the buildings façades constraints the dense reconstruction of the city. We use a feature matching technique while proposing the registration of façades in the images and there use for the reasoning about there visibility in the images. This system is semi-automatic, it requires a user control in the complicated case where no matches are found to link an image to the dataset. The image data set is sparse and the urban space evolves over time.

1. Introduction

The present work is a part of a project that consists in building a 3D reconstruction the city of Reims over the time from old postcards representing the most prominent buildings of its history at different times. We aim at offering the citizen the opportunity to virtually navigate in the city through space and time using an interactive Geographic Information System (GIS). The system should be open for the user enhancement with additional data like postcards or pictures, as well as textual information about the city.

In this paper, we present a method to incrementally build a 3D model of buildings from a dataset of old postcards (figure 1a). The model will be visualized over a geographic map (old cadastral map) and the navigation in the system offers the ability for the exploration of the urban environment. Our approach gives automatic and interactive possibilities. The poscards represent a set of uncalibrated and non geolocated images of low resolution where text and stamps can be found. We take advantage of these data over the postcards to tag the photos in time and space. Additional information can be registered thanks to the collectors of the postcards. It may include historical summaries describing the evolution of the monuments over the time. We refer to this set of information, including date and space tags, as *meta-data* of a postcard.



(a) A group of buildings on the right side (left image) was later replaced by a unique one (right image).

(b) Images with no orthogonality constraints.

Figure 1: Examples of images in our dataset.

2. State of the art

A survey of urban reconstruction has been presented in [MWA*13]. A classification of existing reconstruction approaches into data-driven, model-driven or both combined together is proposed. The approaches are categorized based on the output of the systems. We present in the following a review of image-based and model-based techniques for 3D reconstruction. Other reconstruction techniques such as Li-

DOI: 10.2312/udmv.20141077

DAR and Aerial reconstruction are not within the scope of the paper.

Since the tremendous growth of photography collections over the internet, image-based techniques using feature matching and structure from motion approach were widely used for 3D reconstruction especially in urban environment. The relation of a pair of views is studied in the context of an epipolar geometry based on the pinhole camera model and correspondences between the views. This allows the estimation of the camera poses and a point cloud reconstruction of the scene. Multiview stereo completes the reconstructed model while estimating additional camera poses and 3D points in the cloud. This technique is widely used independently or as a preprocessing stage for other techniques.

[AFS*11, SSS06] suggest the use of image collections for the reconstruction of 3D models of touristic places. A sequence of uncalibrated images is used in [PKVVG00] for the reconstruction of 3D architectural models with no restrictions in camera motions. [BBGC07] present a method for the reconstruction of a metric scene and estimation of the poses of a constant focal length camera. They use the calibration information of the images to undertake intrinsic parameters estimation. [Lee09] uses omnidirectional street views images for 3D reconstruction and geo-referencing of urban environment based on structure from motion (SfM) and bundle adjustment together with 3D sky motion estimation.

Combined model- and image-driven approaches make use of the geometric structure of buildings for the 3D reconstruction. Orthogonality and parallelism constraints of geometrical structures are used [DTM96] to recover the camera poses. This requires three dominant orientations in the image. Interactive photo builder [CRB99] is based on such constraints of architectural scenes for 3D reconstruction of building models from uncalibrated cameras taken from arbitrary viewpoints. [ZK03] detect line structures for dominant orientations in the images that are merged later in hypothesised rectangles for initial pose estimation. They extend the reconstruction to multi-view photographies enabling wide baseline pose estimation while matching rectangular structures. Another type of model-based application is single camera calibration by [Bal01]. Rectangular regular textures and real dimensions constraints of the buildings are required. A color segmentation is used to detect seed points for which real world coordinates are assigned in order to estimate the camera pose.

Model based approaches rely on the presence of particular structures of urban buildings. Orthogonality and parallelism constraints as well as regular rectangular structures are required for the pose estimation and model reconstruction. Those are usually present in urban buildings. In our study case, old buildings has a particular architectural structure and do not necessarily have regular structures. In addition, the presence of three dominant directions is not always satisfied in old postcards (figure 1b), which does not allow the use of orthogonality constraint for pose estimation. We therefore work with image-based techniques while trying to adapt the methods to our specific application case. We deal with a sparse dataset. Calibration parameters and geographic localization are not available.

In this paper we present an approach for reconstructing a 3D model of the city that illustrates prominent buildings and there variations over the time. In the second section we present an overview of the method followed by a detailed description of every step in the process. We present and discuss a set of obtained results. We then conclude and present future works.

3. Incremental 3D reconstruction based on an SfM initialization

Our approach concerns the 3D reconstruction of a city model over time. We are essentially interested in modelling the buildings in the images. The surrounding environment can be witnessed through the images context in the ultimate visualisation of the navigation system.

We overcome the sparsity of the data set by taking advantage of the building geometry. Communally a building is composed of façades. In a low level configuration, a façade is considered as a planar surface without taking into account the details of higher levels that can occur in the architectural structure (windows, balconies, doors...). This low level configuration is exploited to constraint the reconstruction of a multi-plane model.

Our system is feature-based and explores matches between images. The matches are used at different stages of the process to fulfil many requirements. Features based techniques are used to extract and identify façades in the images as well as for the reasoning about the visibility of the façades in the images. As well, the reconstruction of a 3D model depends on the identified feature matches. The result is a set a multi-plane reconstruction and the associated calibrated images. The calibration parameters are the intrinsic and extrinsic parameters. The first defines the projection of the reconstructed model, the second the pose (location and orientation) of the camera that was used to capture the image.

The proposed incremental approach is aimed to allow submitting new data for the system enhancement. Regarding the complexity of the images properties, the system is not fully automated and requires an interaction of an end user to link an input image to the registered dataset when no automatic matches are detected. In the following we expand the different steps of the method.

3.1. Image matching

We suggest the use of feature based technique for image matching. When no automatic matches to a query image are detected, a user intervention is required for the selection of the polygons associated with a building façade in the image.

Feature based image matching has been widely studied in the literature. Many methods for the detection and the description of salient points in the images were proposed and evaluation studies compares the performances and the robustness of such methods in different fileds of application [MS05, DAP11]. In previous works, [YRB13] we presented the motivation behind the choice of features extraction methods that perform well in our study case. We recommended SIFT [Low04] algorithm to extract features for matching our images. We proposed the use of FAST [RPD10] detector combined with DAISY [TLF10] descriptor for the completion of the results in the specific cases where the extracted features matches are not enough for the reconstruction.

In order to reduce the search for correspondences [SSS06] uses EXIF tags of the photographies. We use meta-data tagging of the postcards presented at the end of section 1. The images taged to the same city district are matched and grouped in bundles of matches. Every query image is included in one or many bundles after its automatic match to the dataset. A bundle is a set of images that has strong matching constraint (more than 10 matches per image of same district). Bundles can be weakly connected (less than 10 matches), or may not be connected with automatic matches but only by meta-data assumption of belongingness to the same building or district.

3.2. Façades identification

The features matches are exploited for visible façades identification in the images. The starting point of the process is an image registered by the user. That is, the user selects the set of polygons in the image that are associated to the different visible planar façades of the building. A registered image is an image where corners delimiting planar façades are identified.

To extract and identify façades in a query image that matches with one or multiple images we use the registered image with which it has the greatest number of matches per visible façades. Multiple homographies are computed for the multiple façades to extract there associated polygons in the query image.

We can take advantage of multiple homographies estimation for outliers rejection with the RANSAC [FB81] method. This allows handling complex cases with large number of outliers amongst automatic matches.

The set of registered façades is not constant for all the images and depends on the buildings parts that are visible in the images as well as the result of the matching step. Also, our method is designed to deal with partially visible façades.

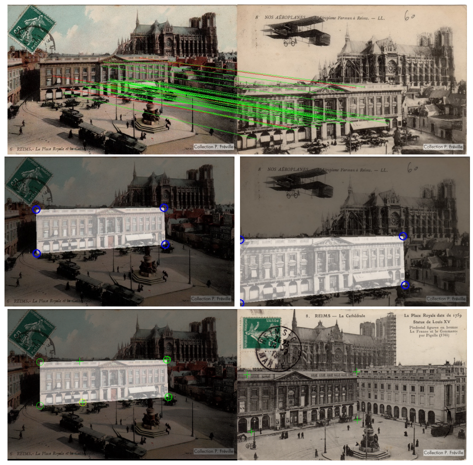


Figure 2: Façades identification (middle row) based on automatically matched features in the images (top row) and previously registered façade. Interactive selection of façade corners (bottom row). The example represent a partially visible façade case.

3.3. Multi-View Stereo SfM for model basis initialization

Common image based 3D reconstruction technique is multi-view stereo structure from motion (MVS-SfM). It relies basically on feature matches between images and camera *pinhole* model. [HZ03] present a detailed geometry analysis of a rigid scene and the mathematical model for solving 3D coordinates when matches between a pair of calibrated images are known.

3.3.1. Camera calibration

In our case no calibration information are available for the old postcards. We consider a null skew for the cameras with square pixels. We do not study the case of potential distortion in the images. The computation of the cameras intrinsic parameters is then limited to the estimation of the focal length. [GF10] suggested a camera self calibration method based on an enumerated integral space of focal length. This is prompted by the finiteness of the acquisition device. [ZJM12] uses this proposed method in a plane tracking context for camera calibration. In our context, we adapt the method. The feature based estimated motion is scored over the enumerated integral of focal length f_i sampled over a logarithmic space $[0.3 f_0, .., 3.0 f_0]$ where f_0 is the sum of half width and height of the image. The result for camera calibration is the focal length f such that

$$f = \operatorname*{Arg\,min}_{f_i} S(f_i).$$

$S(f_i)$ is the reprojection error of the projective reconstruction for the focal length f_i. The reprojection error is the mean distance between the predicted features used for pose estimation and the reprojection of its reconstructed 3D position from the motion parameters.

In addition to camera calibration, given the assumption that the cameras has captured images of the same buildings, we constraint the selection of the focal length so that the cameras optical centers lie above the ground level of the reconstructed model.

3.3.2. Multiview stereo (MVS) reconstruction

The SfM technique is used as a preprocessing stage in this approach. The calibration method presented in the previous section is performed. We construct a basis MVS model for our system as suggested in [SSS06]. The set of images exploited in the reconstruction process is limited to a sparse set of images in our dataset. So far few bundles of matched images are included in the reconstruction process. The reconstructed points cloud is not representative of the buildings models over time and the reconstructed points cloud is not dense due to the sparsity of the dataset.

We exploit the façades planarity constraint to transform the points cloud to a multi-plane model. This allows overriding the sparsity of the reconstructed cloud. We take advantage of this to make the system incremental and open for enhancement by additional data. A least square linear system resolution is used for plane fitting of the points in the cloud associated with each façade. The registered façades polygons are used to compute the 3D coordinates of the façades given the estimated intrinsic and extrinsic parameters associated with each image. Once a façade is reconstructed in 3D it is held fixed in the process. Up to this stage, all the features in the registered images that belong to the reconstructed façades can be computed in 3D. We use the computed camera intrinsic and extrinsic parameters together with the façades metric equations for the computation of these coordinates as a ray-to-plane intersection. We refer to those points as *tracks* of the building model.

3.4. Incremental calibration and pose estimation

The reconstructed model from previous section is used as a basis for further enhancement in an incremental fashion. We discuss here the process of connecting images bundles of different periods of time, as well as the submission of a query image to the system. We distinguish two study cases. First an image may not match automatically to any of the images in the registered dataset despite the fact that its *metadata* indicates its inherence of specific reconstructed area. The submitted image may otherwise match to one or several images in the registered dataset that is, it belongs to a bundle of matched images that has already been included in the process.We expand every case in the following.

3.4.1. Non matches with the database

Due to photometric properties and scene geometry evolution in the images, it can occur that a query image does not match any image in the dataset associated with the same district. The reconstructed façades in the associated district are displayed and the user is invited to select the visible façades in the query image. The meta-data of the image are used to relate it to the bundle of images corresponding to the image district.

To calibrate the query image to a reconstructed model, we suggest a process of *propagation*. The previously computed multiple homographies per extracted façades polygons between images allow the propagation of the *tracks* to the query image. In other words, the position of the pixels corresponding to the *tracks* of the façades are determined to define a set of 2D-to-3D correspondences. The same principle for camera calibration in section 3.3 is used for intrinsic parameters estimation. A Direct Linear Transformation (DLT) method followed by a Levenberg-Marquardt (LM) optimization solves for extrinsic parameters. The pose is estimated iteratively starting with a initial estimation of the pose relative to the dominant façade that occupy the largest space of the image. The result is then optimized with LM method iteratively to the other visible façades.

3.4.2. Existing matches with the database

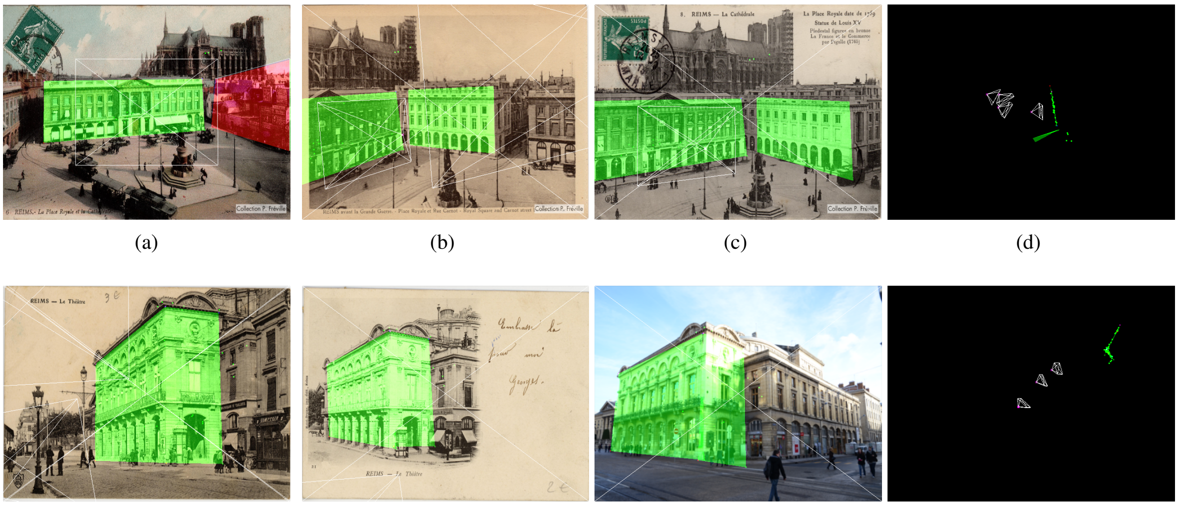
While incrementally adding images to the system, the ideal case is that the added query image matches to the dataset, *ie*. to one or many registered images in the dataset. The matches are used to identify the corresponding *tracks*. The set of 2D-to-3D correspondences identified is used to calibrate and estimate the pose of the camera iteratively relatively to the visible façades in the query image similarly to section 3.4.1. The image is registered and added to the system.

When matches does not lie on any of the registered façades, they are not used to calibrate and estimate the pose of the query image. In this a case, this set of matches is triangulated and a 3D points cloud is computed. New façades, not yet been registered, can then be added to the model. This extends our model to a dynamic city model that evolves over the time.

4. Results

In this section we present results of our method over a set of images in two districts of the city. The *Royal Place* (upper row of figure 3) has witnessed a remarkable evolution over the time and some buildings has been replaced after demolition. The theatre (lower row of figure 3) has not been destroyed over the time, nevertheless the challenges of the reconstruction task remain in the images properties. In figure 3 we show the results for both districts. The model is reconstructed based on feature matches followed by plane fitting to multiple façades.

The estimated calibration (intrinsic and extrinsic) parameters determine the projective transformation that projects the reconstructed façades into their corresponding locations in



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3: A result of the reconstruction of a set of façades. An overlay of the façade in the image shows its projective preview under the estimated parametrization. The color of the façade gives the information about the visibility of the façade in the image. The global scene in 3D is witnessed in the images at the right.

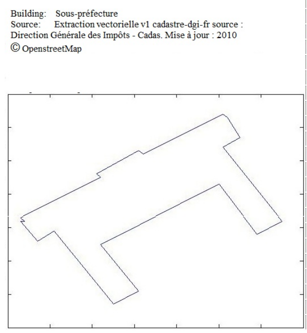
each image. In the figures 3a, 3b, 3c, 3e, 3f and 3g, the aforementioned transformation is used to show the reconstructed façades augmented in each of the images. The global scene can be witnessed in the images at the right of each row (figures 3d and 3h).

The system is based on a preprocessing step of features extraction and matching. The façades are identified by reasoning over there visibility in the images through feature correspondences between the images. For comprehensive perception of the visibility of the façades in the images we use color indexes for the façades augmentation. The façade is visualized in a green color in projective visualisation of an image calibration in which the façade is visible. Otherwise, the fa cade is projected in a red color.

5. Conclusion and future works

In this paper, we have proposed a hybrid approach based plane and points cloud reconstruction of a city model from a dataset of old postcards. We take advantage of the building low level representation as a set of polygons to override the sparseness of the reconstructed points cloud with classical structure from motion techniques. We reconstruct a dense model of multi-planes buildings. Our approach uses matches to identify and extract façades and reason about there visibility to deal with the temporal evolution of the reconstructed city model. It is conceived to allow users submitting additional data for the enhancement of the system.

In future works, we aim to geo-register the reconstructed model over a geographic map. We use *openstreetmap* [OSM] data for the extraction of buildings *landmarks* (figure 4) from a geographic information system. We suggest the alignment of the geographic map to the reconstructed model while adjusting the scale factor of the latter. The aim is to conceive a 4D Geographic Information System representing the evolution of the city model over the time. It allows the user to virtually navigate in space and time while exploring the collection of available postcards testifying the history of the city. Additional data submission will be open for the user. This aims at the enhancement of the system.



Lattitude	Longitude
49.2551224	4.0348706
49.2553210	4.0346808
49.2553476	4.0347429
49.2553975	4.0346964
49.2553963	4.0346917
49.2554076	4.0346826
49.2553022	4.0344239
49.2553122	4.0344129
49.2552533	4.0342671
49.2552433	4.0342781
49.2551366	4.0340157
49.2551260	4.0340267
49.2551227	4.0340238
49.2550720	4.0340712
49.2550979	4.0341315
49.2549007	4.0343167
49.2549337	4.0343985
49.2550596	4.0342808
49.2552155	4.0346627
49.2550875	4.0347832
49.2551224	4.0348706

Figure 4: Example of *landmarks* of the reconstructed building shown in previous section (figure 3a). The geographic coordinates (longitude and latitude) are extracted from *openstreetmap*. They are converted to metric coordinates and represented in the euclidean space.

Toward a better comprehension and perception of the city model context and its state evolution over the time, we propose to involve a transparency display as a function of the visibility of the façades in the images. The meta-data will be used for the classification of the images by date periods up to an uncertainty of the information collected about the old postcards.

6. ACKNOWLEDGMENTS

The authors are grateful to Olivier Rigaud and the association amieCarte51 for providing old Rheims postcards. This project is funded by the Champagne-Ardenne region in France.

References

[AFS*11] AGARWAL S., FURUKAWA Y., SNAVELY N., SIMON I., CURLESS B., SEITZ S., SZELISKI R.: Building rome in a day. *Communications of the ACM 54*, 10 (2011), 105–112. 2

[Bal01] BALTES J.: Camera calibration using rectangular textures. In *Robot Vision*, Klette R., Peleg S., Sommer G., (Eds.), vol. 1998 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001, pp. 245–251. 2

[BBGC07] BOCQUILLON B., BARTOLI A., GURDJOS P., CROUZIL A.: On constant focal length self-calibration from multiple views. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR 07* (2007), pp. 1–8. 2

[CRB99] CIPOLLA R., ROBERTSON D., BOYER E.: Photobuilder-3d models of architectural scenes from uncalibrated images. In *IEEE International Conference on Multimedia Computing and Systems, 1999* (1999), vol. 1, pp. 25–31 vol.1. 2

[DAP11] DAHL A., AANÃĘS H., PEDERSEN K.: Finding the best feature detector-descriptor combination. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)* (2011), pp. 318–325. 3

[DTM96] DEBEVEC P. E., TAYLOR C. J., MALIK J.: Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics* (1996), pp. 11–20. 2

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (1981), 381–395. 3

[GF10] GHERARDI R., FUSIELLO A.: Practical autocalibration. In *Computer Vision - ECCV 2010*, Daniilidis K., Maragos P., Paragios N., (Eds.), vol. 6311 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, pp. 790–801. 3

[HZ03] HARTLEY R., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, 2 ed. Cambridge University Press, New York, NY, USA, 2003. 3

[Lee09] LEE T.: Robust 3D street-view reconstruction using sky motion estimation. In *2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)* (2009), pp. 1840–1847. 2

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. 3

[MS05] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence 27*, 10 (2005), 1615–1630. 3

[MWA*13] MUSIALSKI P., WONKA P., ALIAGA D. G., WIMMER M., VAN GOOL L., PURGATHOFER W.: A survey of urban reconstruction. *Computer Graphics Forum 32*, 6 (2013), 146–177. 1

[OSM] http://openstreetmap.fr/. 5

[PKVVG00] POLLEFEYS M., KOCH R., VERGAUWEN M., VAN GOOL L.: Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal of Photogrammetry and Remote Sensing 55*, 4 (Nov. 2000), 251–267. 2

[RPD10] ROSTEN E., PORTER R., DRUMMOND T.: Faster and better: a machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 1 (2010), 105–119. 3

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: Exploring photo collections in 3d. *ACM Trans. Graph. 25*, 3 (July 2006), 835–846. 2, 3, 4

[TLF10] TOLA E., LEPETIT V., FUA P.: DAISY: an efficient dense descriptor applied to wide baseline stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence 32*, 5 (2010), 815–830. 3

[YRB13] YOUNES L., ROMANIUK B., BITTAR E.: Can feature points be used with low resolution disparate images? (application to postcard data set for 4D city modeling). In *Eighth International Conference on Computer Vision Theory and Applications (VISAPP)* (2013), vol. 1, pp. 481–486. 3

[ZJM12] ZHOU Z., JIN H., MA Y.: Robust plane-based structure from motion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2012), pp. 1482–1489. 3

[ZK03] ZHANG W., KOSECKA J.: Extraction, matching and pose recovery based on dominant rectangular structures. In *Higher-Level Knowledge in 3D Modeling and Motion Analysis, 2003. HLK 2003. First IEEE International Workshop on* (Oct 2003), pp. 83–91. 2