Eurographics Italian Chapter Conference (2006)
G. Gallo and S. Battiato and F. Stanco (Editors)

Solutions to 3D Building Reconstruction from Photographs

G. M. Farinella, G. Mattiolo

Dipartimento di Matematica e Informatica, University of Catania, Italy
Image Processing Laboratory
http://www.dmi.unict.it/~iplab

Abstract

The 3D model reconstruction of buildings from uncalibrated photographs allows new useful Computer Graphics and Computer Vision applications. In this paper we survey some solutions proposed in literature and in commerce. The common and different methodology are reported and described. Finally we try to address the 3D reconstruction problem of architectural scenes presenting our solution: BolBol. We demonstrate our reconstruction process and some results obtained applying BolBol to the modeling of real buildings.

Categories and Subject Descriptors (according to ACM CCS): A.1 [**General Literature**]: Introductory and Survey; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism - *General*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis - *Stereo*; J.6 [**Computer Applications**]: Computer-Aided Engineering - *Computer-aided design (CAD).*

1. Introduction

Some new and emerging applications for 3D information system of a city require more and different type of information than those required by 2D traditional plans. These new applications include for example trip planning by Internet or advanced guide for pedestrians through digital personal assistant. Rather then providing a flat and schematic plan, one wishes a system that combines navigation and communication, presenting touristic, cultural and commercial information all in 3D. Another applicaion may be car advanced navigation interfaces that exploit 3D environment visualization to guide the driver to the desired destination. More traditional application fields include architectural and town planning, research about urban climate and environment, studies of pollution diffusion dynamics, and sound and electromagnetic waves propagation. These studies can be particularly valuable in planning and in the choice of installation point for telecommunication antennas.

This paper deals with the problem of 3D model reconstruction of buildings and architectural scene from their sparse uncalibrated photographs. We survey some solutions proposed in literature and in commerce and finally try to address the 3D architectural scene reconstruction behind our solution.

The remainder of this paper is organized as follows. In Section 2 we survey some examples in literature and in commerce, describing their different approaches and philosophies. In particular we focus our attention on the research done in this field by Roberto Cipolla and Paul E. Debevec. Their systems named respectively PhotoBuilder and Façade have inspired us in realizing the solution proposed in this paper: BolBol. In section 3 BolBol is discussed, while in Section 4 we demonstrate our reconstruction process and some results obtained applying BolBol to the modelling of some real buildings. Finally, conclusion are summarized in Section 5.

2. Previous Works

Many researchers proposed to exploit the rigorous constraints of parallelism and orthogonality clearly present in most of the buildings to create 3D models. In this research field many contributions are due to Roberto Cipolla; he and his team at Cambridge University have realized an advanced system, called PhotoBuilder [CRB99], that have improved the state of the art of 3D reconstruction. PhotoBuilder is able to exploit buildings common geometrical features to initially identify camera unknown parameters without any a priori information [CB98]. The 3D building reconstruction algorithm used by PhotoBuilder consist of four stages:

1. in the first phase the user define a set of interesting primitives, such as segments and cuboids, from the available input images. PhotoBuilder helps the user to precisely localize those primitives thanks to corner detection algorithms. The information extracted in this phase are required to define the image plane projections of primitives such as 3D points, segments and plane polygons. Information from maps and plans can be also taken into account.
2. The orthogonality and parallel constraints in geometric primitives derived by previous step are used for camera calibration. For intrinsic camera parameters estimation PhotoBuilder uses the vanishing points obtained from lines of primitives detected in the input 2D images
3. The next step requires to compute projection matrices for each viewpoint. Using these matrices, camera orientation and translation relative to a fixed reference viewpoint can be calculated. Exploiting the information reached in the previous step epipolar constraints can be defined and the projection matrices are refined. Additional information related to motion between the viewpoints are also considered in projection matrices refinement.
4. Eventually, projection matrices relative to every photo are used to find correspondences for every couple of photos allowing to compute 3D triangles that compose the final 3D model of the building. Each triangles can be painted with a texture obtained from input photos.

Using PhotoBuilder, Roberto Cipolla realized the 3D model reconstruction of some Colleges buildings in Cambridge University Campus; some photorealistic VRML example of 3D building reconstruction by Photobuilder are available in the related project page [CR99].

Another well known and successful system for building reconstruction is Façade, designed and implemented by P. Devebec and his team. In his works [DTM96] [Deb96] Debevec clearly distinguishes his modus operandi from that proposed by others; in particular he identifies two classes of methodologies:

- *Geometry-Based*:, the user watching photographs and exploiting other type of information about the building to reconstruct, creates the model using a modelling tool;
- *Image-Based*: adopting Computer Vision algorithms based on stereo vision and epipolar geometry it is possible to find the depth map of the building in a semi-automatic way.

Debevec mediates between these two philosophies to propose a new technique that he defines *Hybrid*. He wishes to combine both approaches taking the best of them and minimizing their drawbacks. The first phase of the *Hybrid process* is photogrammetric: the user has to interactively build a raw 3D model, combining existing geometric primitives or eventually creating new ones. He also has to specify the associations between model entities and lines on the photo. Most of the critics moved to Debevec are about this phase, considered too laborious and time-consuming for the user. The raw model and the correspondences defined by the user are presented as input to the next phase, that consists in an optimization process that tries to find unknown real model and camera parameters (3D position and orientation). At the end of this process one has a 3D model, respecting dimensions and proportions of the original building, and he knows all information about camera and the point of view from which every photo was taken.

Once the model has been obtained, the successive step is aimed to project a texture onto the model, so that you can obtain a reconstruction visually very similar to reality. A difficult issue is that the whole effect appears natural and realistic only observing the textured building model from the same points of view of the photos from which textures were extracted: moving away the observer would expect variations in material reflection and in the projected shadows, that can't happen. This problem is known as the painted shoebox effect.

To avoid this drawback Debevec proposes a technique known as view-dependent texture mapping: the idea behind that is to project all the available textures from different photographs on the object, but assigning a value to every point of the surface that can be thought as a weighted sum of the contribute of every projected texture: weights depend on the point of view from which you observe the surface. The contributes of the texture projected from points of view next to that you want to render are more important; naturally the effect is as good as the artificial view you want to render is close to the available real ones. At this point the Image-Based nature of the process developed by Debevec is revealed; the idea consists in exploiting stereo vision to find a depth map of the building in order to recover those details that were not modelled in previous steps. This allows the user to create completely artificial views of the building from point of views as arbitrary as possible. A known problem with stereo vision algorithms is that they suffer when the images to pair are taken from very distant points of view. The situation here is different because we have a model to exploit, that was reconstructed in the previous steps: different images of the same scene become more similar when projected on the model of the scene, making the pixel neighbourhoods comparison easier and safer; for this reason Debevec calls his approach *Model-Based Stereo*. Façade was used to reconstruct a lot of buildings, among which in particular the Campanile, the Berkeley clock tower: the result is an incredible movie available in Internet and that received international acclaim [Deb05].

In 1997, G. Borshukov, now Computer Graphics Supervisor at Electronic Arts, introduced some extensions in Façade to allow the modelling of arches and revolution surface [Bor97], not supported in the original project. These geometric features are important in 3D building reconstruction because they are usually present in the cultural heritage buildings.

The strategy developed by Roberto Scopigno et al. [TCRS00] aims to make the role played by the user in the reconstruction process intuitive, fast and precise at the same time. Their choice is to model each face of the building with rectangles sharing edges or vertices: this reduces the application field to relatively simple buildings. The first target is to find the planes upon which the rectangular primitives lay. The user has to provide three sets of segments in three orthogonal directions; hence three vanishing points are detected and each pair is used to identify a different orthogonal plane. Finally the three planes are used to calibrate camera for each image. Thanks to this information, successive plane identifications require only two sets of lines. The next step consists in identifing the rectangles composing the building: in this phase the user is helped by the system interface while adding new rectangles, modifing previously created shapes and specifing adjacency between them with good precision. A remarkable feature of this system is the general low computational load required to achieve a model, allowing the user to visualize the ongoing building reconstruction progress.

Among the more known commercial tools there is RealViz ImageModeler [Rea05]: even in this case the input is represented by images collected by the user. The user has to identify in every image points (markers) correspondent to the same vertex (locator) of the object to reconstruct: we point out the fact that ImageModeler works with vertices, while in Façade the main entities are lines. It's considered the possibility to import CAD data relative to the object to model. From such information ImageModeler is able to find camera 3D position, focal distance and distortion for every photo provided to the application. To improve the result of calibration process it's possible to insert further information such as right angles or planar constraints among different locators. Once the camera has been calibrated and the desired reference framework and the suitable scale factor have been fixed, it's possible to go to the real phase of modelling: the user chooses some geometric primitives and creates relationships between their vertices and the locators defined in the photographs. To model more complex shape objects ImageModeler interface is able to assist the user during manipulation of primitive components, such as edges and vertices. At last ImageModeler allows the user to extract and edit the texture relatives to objects of interests and project them into the model. The achieved result can be easily exported in very common formats: VRML, 3ds Max, Maya and Lightwave 3D.

The last system we cite here is Voodoo Camera Tracker [DLfIotUH05], a non-commercial software tool developed for research purpose at the Laboratorium für Informationstechnologie of University of Hannover. It is freely downloadable at http://www.digilab.uni-hannover.de/docs/manual.html. A difference with the previously seen applications is that the Voodoo Camera Tracker tries to estimate camera parameters and reconstructs a 3D scene from video sequences. Relevant potentialities are automatic detection of feature points with sub-pixel accuracy, thanks to a corner detector that implements various strategies (Harris [HS88], Susan [SB95]). Automatic correspondence analysis is also possible: matching points are identified in every frame of the sequence by using intensity cross-correlation algorithms. According to the authors, possible applications are film production, 3D reconstruction or video coding. The estimated parameters can be exported to Softimage 3D, 3D Studio Max or Blender.

3. BolBol

BolBol is the application that we have developed following the *Model-Based Stereo* approach whose objective is to reconstruct the model of a building from information provided by user and by photographs. The application was developed partially in Java, exploiting in particular the Java 3D API to build the interface and to visualize 3D models. It partially uses numerical routines in Octave/Matlab to implement the parameters estimation routines. In this Section we'll describe the process of reconstruction implemented in BolBol[†].

First of all the user is required to create a raw model of the building, using the available set of geometric primitives provided by BolBol user interfaces (Figure 5): until now they include boxes, pyramides and prisms. The creation of the model consists in adding and combining fundamental entities called blocks. You start adding a root block, that is the only one without a parent block: every other block has one parent, specified at the moment of its creation. Every block is characterized by its dimensions and by the translation relative to its parent: until now it's not supported the possibility that a block expresses a rotation relative to its parent block. Equality relationships can be defined among the dimensions of different blocks and even among those of the same block, through the BolBol User Interface: this fact implies a reduction in the number of parameters to estimate and then in a smaller load for the optimization algorithm.

Supposing no rotation, spatial relationships between parent and child blocks help to reduce the model degrees of freedom of the model. For example consider a model composed by a pyramide and a cube; the first lies upon the other along world Z axis. This implies that the Z component of the translation that links these primitives can be thought as a composition of the maximum and minimum Z bounds of their container blocks. In this case one parameter has been eliminated from the model.

To achieve the coordinates of a point in a block in the main

[†] The name is a personal reference of the authors.

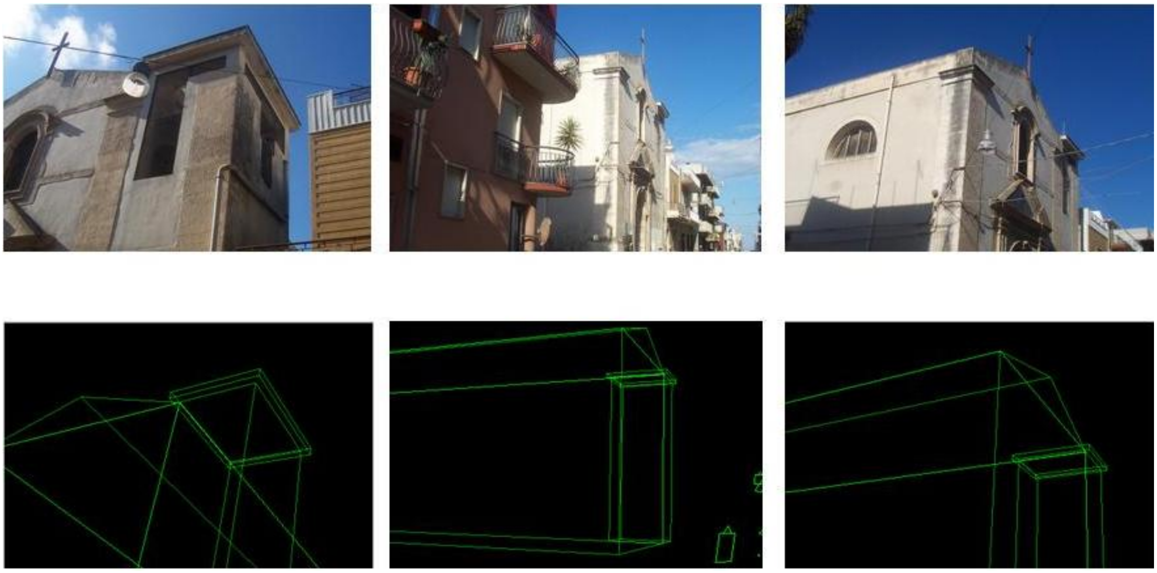


Figure 1: *Some views of San Sebastiano's church in Carlentini; real (top) and reconstructed (bottom).*

reference framework, first of all you need to compute them in the local reference framework of the owner block. Then you need to evaluate the absolute translation of the block in the main reference framework. These requires to initially consider the translation relative to the parent block; this has a translation relative to his parent and so on as far as the root block whose absolute translation is perfectly known. The desired translation is equal to the composition of all translational contributes of the blocks that precede the considered one. Carefully looking at it, the structure of the model in BolBol is tree-like, whose root is represented exactly by the root block.

Besides the definition of the raw model, an other important task for the user, is to create some associations which link model entities with elements of the photographs taken of the real building. In this phase the user typically defines associations among model edges and image lines. Perhaps this is the first aspect suitable to improvement in BolBol: complex buildings require a sensible contribute in terms of time and attention from the user and future versions will include solutions to this problem, which affects Façade too.

The kernel of the 3D reconstruction process in BolBol resides in the optimization phase. The input of the optimization phase is composed by:

- the raw model created by the user;
- the associations defined by the user between the model and the images;
- the intrinsic parameters of the camera used to take them.

The optimization phase provides the values of the variables that have to be estimated:

- the minimum set of model dimensional parameters, considering the constraints;
- the minimum set of model translational parameters, considering the constraints;
- for every photo provided as input to the system, the extrinsic parameters of the camera.

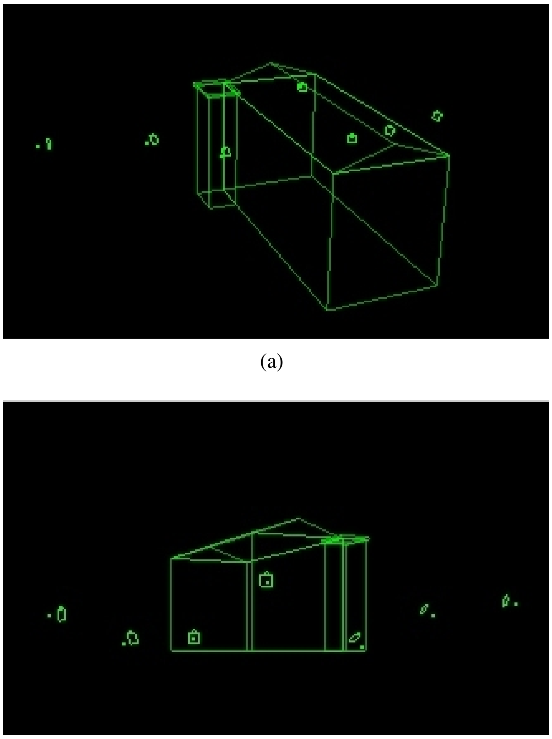
Camera intrinsic parameters must be provided to the system, hence they're not a part of the reconstruction process: however they can be derived from a previous camera self calibration process as described in literature [Hem03].

The optimization process can be divided into three distinct phases [Deb96]. This partition derives from the possibility of achieving in most of cases initial estimates of the optimization variables. These are provided as input to the final optimization process, and extended to the entire solution space. This multi-step nature often allows to avoid well-known local minima, undoubtedly present in the target function of the final phase, and to reduce the time of the optimum search. The first estimate is relative to rotation matrices for every photo provided as input to the system: the constraint to respect here is that the matrix is a special orthogonal one. Thanks to it, you can obtain a second estimate relative to the position of the camera per photo and to model parameters, that must be all positive. Target of the last step is to refine what was found in previous phases, again based on photos and associations. The desired result is that the reconstructed 3D model is as near as possible to the building depicted in photographs; it means that when framed from the same points of view of photographs, the model should produce the same projections visible in input images. It's then possible to define the objective function relative to the final phase of the optimization process so that it provides a measure of the error between expected and actual results [TK95].

The main problem in the approach previously described is that all buildings composed by blocks without edges, such as cupolas, columns and arches are excluded from the reconstruction process, although they very frequently appear in architecture. It's possible to exploit previous reconstruction work to support more complex solid primitives. We point out that in the approach we adopted following Borshukov such primitives, before excluded by Façade, are not reconstructed through an other optimization process, but using the information about extrinsic camera parameters and about the blocks of the same building collected in previous steps.

Finally the possibility of projecting textures extracted from photographs into the model and of exporting process output in VRML have also been included in BolBol.

4. Application Examples



(a)

(b)

Figure 2: *Artificial views of San Sebastiano's church with position and orientation of camera for every used photographs.*

We tested BolBol by reconstructing some buildings, chosen for their relative architectural simplicity and for the possibility to take interesting views in order to start the process.

The first test building is the San Sebastiano's church in Carlentini, Syracuse, Italy (Figure 1). For our purpose we used seven photos of the building (Figure 2).

To verify the capacity of BolBol to reconstruct buildings with complex architectural components such as arches, we realized the reconstruction of the façade of Costantino's Triumph Arch in Rome, using only a couple of photographs (Figure 3). Notice that a part of the background has been projected into the inferior surface of the arch: this is due to the fact that the depth of the building was provided by the user. Unfortunately, the depth of the Triumph Arch couldn't be inferred only using the available images since they have taken only from point of view facing the monument.

Another example of 3D building reconstruction using BolBol is reported in Figure 4: the King's College Gibbs Building in Cambridge. As in the previous example, the depth of the Gibbs building has been fixed a priori through BolBol interface (Figure 5), because the photographs used in this experiment don't allow to capture it.

5. Conclusions

In this paper we reviewed the state of the art for 3D building semi-automatic reconstruction from uncalibrated photos and introduced BolBol. This application has been designed and implemented to be able to achieve 3D model of building starting from real building photographs and information provided by the user. BolBol has been tested on some real buildings, showing its ability to catch building geometry and to infer extrinsic camera parameters in a satisfactory way. Future works can be aimed to include camera autocalibration process in BolBol and to find some useful heuristics that can be help the user in construction of the raw model in a more simple way. Indeed actually the reconstruction even of simple buildings may require a heavy work from the user and much work has to be done to address usability and simplicity issues.

Acknowledgements

The authors wish to thank Prof. Giovanni Gallo and Prof. Roberto Cipolla for their valuable and constructive suggestions.

References

[Bor97] Borshukov G. D.: *New Algorithms for Modeling and Rendering Architecture from Photographs*. Master's thesis, University of California at Berkeley, 1997.

[CB98] Cipolla R., Boyer E.: 3d Model Acquisition from Uncalibrated Images. In *IAPR Workshop on Machine Vision Applications, Chiba, Japan* (November 1998), pp. 559–568.

[CR99] Cipolla R., Robertson D.: Photobuilder:

3D Model Acquisition from Photographs. http://svr-www.eng.cam.ac.uk/research/vision/photobuilder/aim.html, 1999.

[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,Firenze* (June 1999), vol. I, pp. 25–31.

[Deb96] Debevec P. E.: *Modeling and Rendering Architecture from Photographs*. PhD thesis, University of California at Berkeley, 1996.

[Deb05] Debevec P. E.: The Campanile Movie. http://www.debevec.org/Campanile, 2005.

[DLfIotUH05] Digilab Laboratorium für Informationstechnologie of the University Hannover: Voodoo camera tracker: A tool for the integration of virtual and real scenes. http://www.digilab.uni-hannover.de/docs/manual.html, 2005.

[DTM96] Debevec P. E., Taylor C. J., Malik J.: Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach. *Computer Graphics 30*, Annual Conference Series (1996), 11–20.

[Hem03] Hemayed E.: A Survey of Camera Self-Calibration. In *AVSBS03* (2003), pp. 351–357.

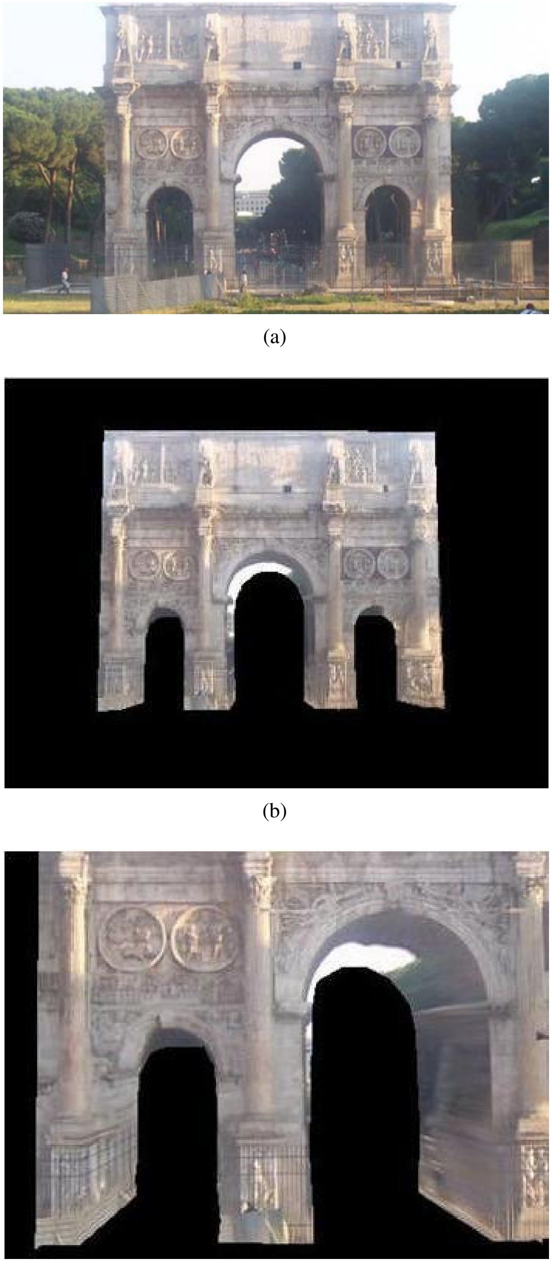
[HS88] Harris C., Stephens M.: A Combined Corner and Edge Detector. In *Fourth Alvey Vision Conference* (1988), pp. 147–151.

[Rea05] RealViz: RealViz imageModeler. http://www.realviz.com/products/im/index.php, 2005.

[SB95] Smith S. M., Brady J. M.: *SUSAN – A new approach to low level image processing*. Tech. Rep. TR95SMS1c, Chertsey, Surrey, UK, 1995.

[TCRS00] Tarini M., Cignoni P., Rocchini C., Scopigno R.: Computer assisted reconstruction of buildings from photographic data. In *5th IEEE Workshop on Vision, Modeling and Visualization* (November 2000), pp. 213–220.

[TK95] Taylor C. J., Kriegman D. J.: Structure and Motion from Line Segments in Multiple Images. *IEEE Trans. Pattern Anal. Mach. Intell. 17*, 11 (1995), 1021–1032.



(a)

(b)

(c)

Figure 3: *In (a) a photo of Constantino's Arch in Rome. In (b) and (c) reconstructed models of Costantino's Arch, recovered by a texture extracted from photographs.*

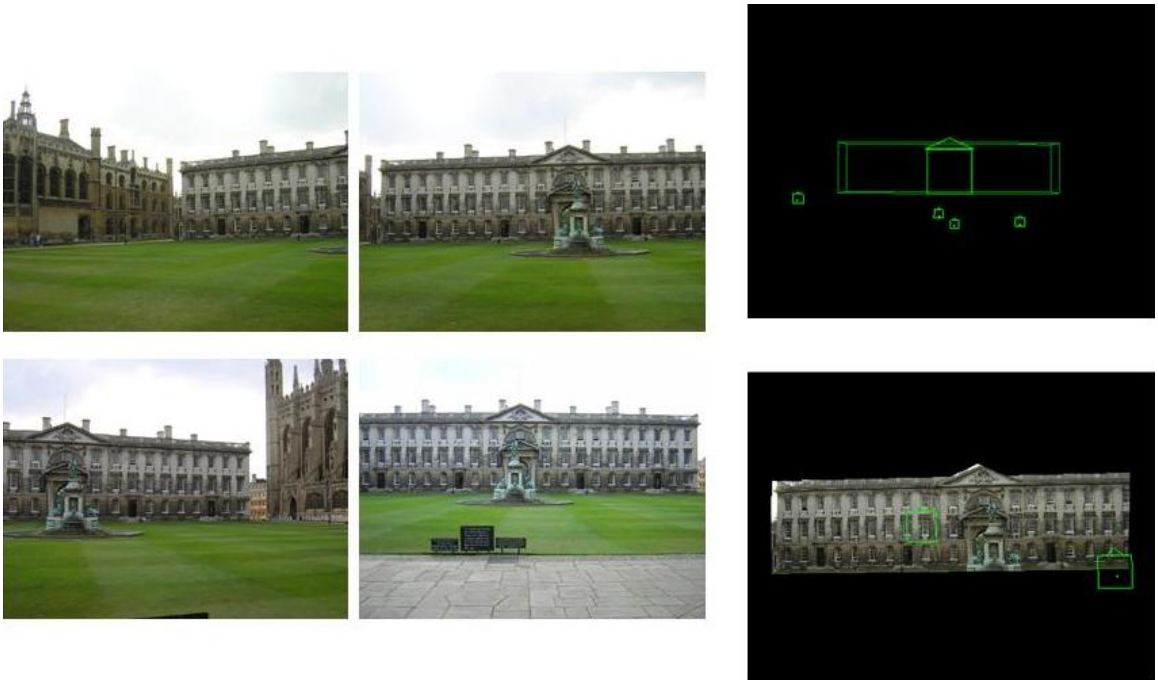
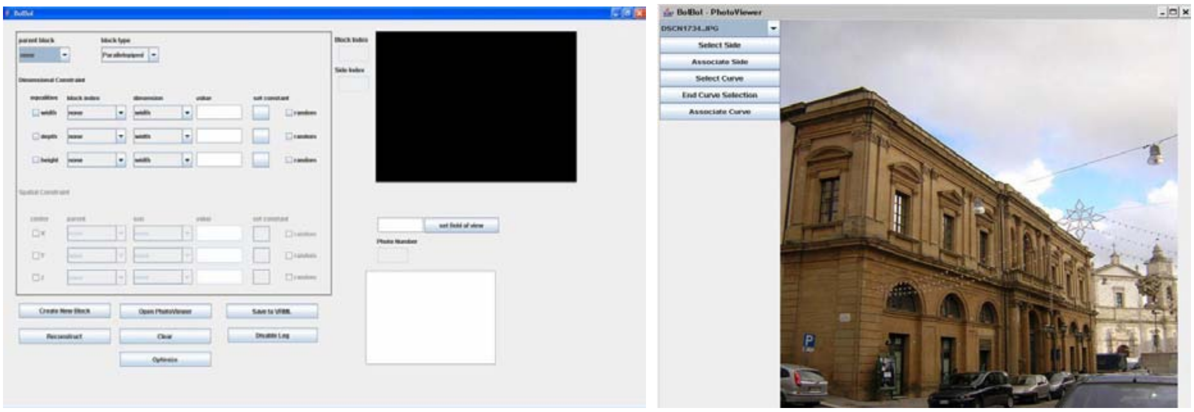


Figure 4: *Views of Gibbs Building del King's College in Cambridge. Original on the left, reconstruction (with and without texture) on the right.*



(a)

(b)

Figure 5: *BolBol user interfaces. The Main (a) window allow the user user to create a raw model of the building, using the available set of geometric primitives. The PhotoViewer window in (b) allow to create some associations which link model entities with elements of the photographs taken of the real building.*