OCCLUSION-FREE 3D REALISTIC MODELLING OF BUILDINGS IN URBAN AREAS

M. Varshosaz

Department of Photogrammetry and Remote Sensing,
Faculty of Suryveying, KNT University of Technology
Vali-Asr Street, Mirdamad Cross, Tehran, Iran
varshosazm@kntu.ac.ir

KEY WORDS: Photogrammetry, Textures, Buildings, Modelling, Virtual Reality

ABSTRACT:

Currently the creation of textures for construction of Visually Realistic Models (VRMs) of buildings is a time consuming manual process involving the use of a number of separate software applications and substantial skill by the user. This paper presents the development of a novel procedure for the semi-automatic construction of textures, which makes use of a Charge Coupled Device (CCD) camera mounted on a motor-driven theodolite to acquire a digital texture database. The images are indexed using the angular horizontal and vertical readings from the theodolite. Given the coordinates of a building façade, an Automatic Texture Processing Tool (ATPT) finds the corresponding images covering the face and forms its texture. The system suggested here is substantially faster and easier to use compared the current methods available in the market. In addition, it can be used to easily merge textures of the same face in order to remove undesirable occlusions from textures.

1. INTRODUCTION

Developing Visually Realistic Model (VRM) of a building requires a 3D geometric model rendered with textures from building sides. This encompasses two distinct tasks namely the acquisition of building geometry and the processing of textures. Recent years have seen a number of developments in geometric modelling of buildings in 3D. The semi-automatic system developed at Bonn University (Gülch *et al.*, 1998) is an example which aims on fast production of simple and complicated models of buildings in urban areas.

The second, and difficult, task is the processing of textures. Textures are real images, taken from building façades, used to improve the visual impact of the models. They can also be used as a substitute to see/access the details that are missing in the geometric model. Processing textures for the photo-realistic documentation of buildings is carried out in two main steps: image acquisition and image processing.

The aim of image acquisition is to provide digital input for the production of textures. Images can be obtained from either terrestrial or aerial photographs. Aerial photographs are used to cover details of roofs, while terrestrial images are used for the modelling of building façades. In rare situations, where the façades of buildings are not occluded, aerial photographs may also be used to extract textures of facades. However, due to the vertical nature of aerial photographs building facades are either invisible or subject to large perspective distortions. Therefore, the photo-realism of building facades can only be achieved if terrestrial images are acquired and mapped to their corresponding building faces (Brenner and Haala, 1998).

Once the images are captured they have to be processed in order to form textures. This includes setting up the correspondence between the geometry and texture, rectifying and mosaicing the images, and enhancing the resulting textures by adapting the colour, intensity, saturation, and clearing occlusions shadows as required. To create textures, the first step is to set up the correspondence between the image and object coordinate systems. As the interior and exterior orientation of images is generally unknown, usually this is done manually by measuring the image coordinate of few control points whose positions on the ground are known. Depending on the type of the camera used, the images may show a panoramic or perspective view of objects, which need to be compensated for through a rectification process. The images covering a particular building face are then mosaiced to form the complete texture of the face.

Occlusion by static or moving objects is another problem in texture processing. The effect of occlusions can be reduced by having a good plan prior to data acquisition to gain an obstruction-free view of the buildings. A partial solution to remove occlusions is to apply computer graphic methods filling in the missing parts with artificial or image-based templates. Techniques like this, however, may result in unrealistic impression of objects in the model. Besides, even with imagery of similar information to that missing, different intensity, hue, etc. may result in discontinuities over the created texture.

Looking at the literature, it can be realised that there exist no fully automatic texture processing technique. Attempts so far have mainly concentrated only on the automatic acquisition of images. The modelling process is highly dependent on human interaction (Debevec, 1999) and, thus, improvement in this area is extremely desirable (Maresch, 1997).

The system presented here is an attempt to simplify and speed up the texture-processing task. It uses a CCD camera to provide digital input for textures, which are then processed automatically. The details of how the system works can be found in Varshosaz (1997). The aim here is to briefly review the main components of the system and show how it can be used to remove occlusions from textures.

2. TEXTURE CREATION: SYSTEM OVERVIEW

Texture processing aims at the provision of rectified images for all the visible faces of buildings. In the system proposed here, terrestrial imagery is used as the texture source. The image acquisition unit was developed at University College London (Chapman, *et al.*, 1994), the main characteristics of which are:

- Capturing images using an off-the-shelf CCD camera mounted on a servo-driven theodolite;
- Taking images automatically in a step-wise manner at preset locations to cover a complete field of view of 360° by 90° in horizontal and vertical directions respectively;
- Indexing the images using the theodolite angles recorded for each image.

Indexing is an important feature of the system. In other words, in order to localise the coordinate system of each images, usually a number of control points are pre-marked in the image capture area. However, here, as the images are related to each other through their recorded vertical and horizontal angles, localisation is done only once for each station, not each image. The result is a grate reduction in the localisation time and the number of control points needed. As a result of indexing, the relative angular location of images with respect to each other is known. Therefore, by knowing only the exterior orientation of image capture stations, the relative orientation of images at all camera positions can be determined.

To process textures, the following components are required:

- A Numerical Frame of Reference (NFR);
- A Texture Image Database (TID);
- An Automatic Texture Processing Tool (ATPT).

The NFR describes the geometric structure of buildings, while the TID is the database of terrestrial CCD images that contains all the data required to relate the images to each other in space. The heart of TID is a reference file which is a look-up table defining a number of entities for each image:

- Image ID;
- Name of the file containing camera information;
- Station ID at which the image was taken;
- Image dimensions;
- Angular theodolite readings;
- Name and location of image on the computer hard disc.

As shown in Figure 1, to from the texture of a face, the 3D coordinates of its corners, extracted from the NFR, are passed to ATPT that uses the camera calibration, localisation, and data provided by the reference file to find the images covering the face of interest. Based on a given spacing interval a 3D grid, the points of which define the texture points, is fitted to the face.
The basic goal of ATPT is to define the pixel values associated to the grid points, which are processed one at a time. For each point the reference file is searched and an image with most orthogonal view over the point is selected. The coordinates of the point on this image are then computed using the collinearity equations. An image interpolation is finally carried out to estimate the pixel value of the point. The estimated value is then written to a new image, which is in fact the texture of the face. Figure 2 shows an example.
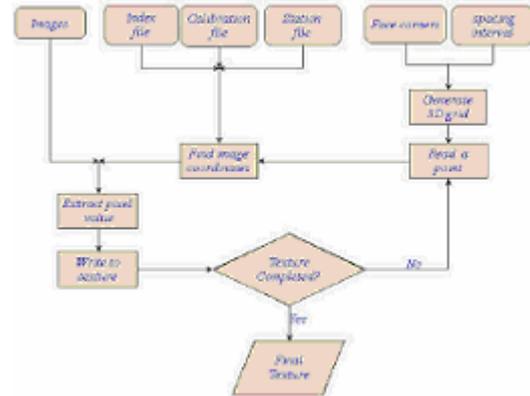

Figure 1: Flowchart of texture creation

(a)   (b)
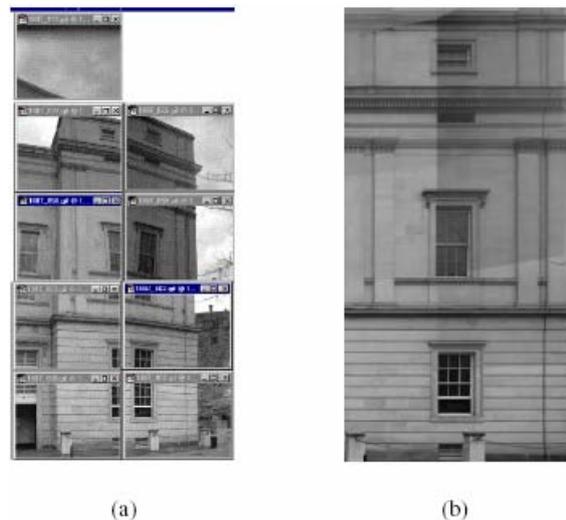
Figure 2. Individual images and the corresponding texture

3. MERGING TEXTURES

Merging refers to fusing images from different stations in order to get a complete or higher quality texture. There are several cases where parts of textures are missing or need improvement. They can be categorised into three groups: occlusions, missing images and perspective distortions.

It was mentioned that to create a texture all features between a camera and a building facade are projected onto the texture plane. Consequently, some parts may be occluded by details like cars and trees near to the camera. The existence of occluded objects in textures does not necessarily mean they have to be removed. On the contrary, features like a tree or a passing person, can even improve the reality of a visual model (Figure 3). However, there are situations where occlusions may cause problems, for instance, where an important point is behind an object or part of a building appears on the texture of another building facade (Figure 4).

The second case where merging may become necessary is when an image covering a certain part is missing. This can happen due to data capture internal errors, over exposure resulting from bright sunlight, or buildings being too high. For example, if buildings are too high it may not be possible for the camera to cover the top parts of the buildings. However, as the instrument is moved along the buildings, such parts are usually covered

Figure 3. Occlusions improving the look and feel of a VRM

Figure 4. An occluded building.

Figure 6. High perspective effect.

In most of the above situations, it is desired to improve the quality of the texture by filling in the gaps or integrating better quality images into the affected areas. Unfortunately, there is no system capable of dealing with such problems easily and efficiently. Some techniques simply fill in the affected part with an artificial pattern or an image-based template. Applying such methods, however, does not lead to a truly realistic solution to the problem, as the improved section may still be different from the way it looks in reality.

One possible solution is to take original images in a way that occlusion does not occur in the resulting textures. For this, each image is examined on site and is replaced with a new one if necessary. Most current digital cameras allow for reviewing images immediately after they are captured. Thus if unwanted details are included or if the desired features are not present, the relevant images are simply deleted and new ones taken. Taking images in such a way, however, requires further on-site supervision and interaction by the user. In addition, this method may not be practical in busy parts of cities.

Fortunately, here, as the terrestrial images are all registered to the same coordinate system and the texture processing is carried out based on object points, the textures of a specific face can easily be merged. This means, if parts of a texture are affected, they can be replaced by textures processed using images from the neighbouring stations.

As a first step, the face containing the affected area and the stations that cover this area are defined by the user. Using ATPT, a low resolution texture for the affected area is created from each individual station. The textures are then examined visually and the one with the best view over the area of interest is selected. Then, a high resolution texture, having the same spacing interval as that of the final texture, is created using the images from the selected station.

The unaffected parts, taken from the original texture, and the affected part taken from the texture just created are fused to form a new texture. Indeed, this texture is the result of merging the two texture parts. As both textures have been created at the same resolution and belong to the same object (i.e. the corresponding face), their dimensions and orientation are the

from the neighbouring stations, which have a smaller vertical angle to the missing part. A texture example is shown in Figure 5 where the instrument could not capture the top part of the building face; thus its texture is incomplete and is shown using white colour. The third situation is when a building face is too long or too high; thus using a single station may cause features in the resulting texture to appear stretched, as the images are captured with high perspectives. Figure 6 shows an example of such, where features are stretched especially at the top.

Figure 5. Texture with a missing part.

same. Thus, no extra work is required to scale, orient, or wrap the new texture to fit the original one, when the merging process takes place. Small mismatches may, however, arise if the perspective views of stations over the area of interest are too different. Therefore, when examining the low resolution textures a station should be selected that not fills the affected area appropriately, but also does not result in large discontinuities in the final texture. The process of merging textures can be done in two ways described below.

3.1 Pick and paint

Some image processing packages like Photoshop ™ allow painting a copy of an image onto another image interactively. For this, samples from a source texture are Picked and Painted onto the same area on the other texture. For convenience, the original texture is defined as the destination, while the one created from an individual neighbouring station is taken as the source.

The Pick and Paint process involves positioning the cursor over the area of interest in the source texture and picking a sample point. The sample point is the location from which the image will begin to be duplicated as Painting operation takes place in the destination texture. The cursor is then pressed at the same point in the destination texture and is dragged to Paint the sampled texture. The process is continued until the whole unwanted area of texture is filled in with the source samples. The resulting texture is saved and a new texture taking parts from each texture is created.

3.2 Forcing ATPT

This approach is more automated compared to the Pick and Paint method. Having defined the station at which the affected area has to be produced, the image coordinates of the affected area along with the station identification are introduced to ATPT. While processing the texture points, the ATPT uses the images of the given station to define the pixel value of the point. The rest of the texture creation process remains the same as that described before.

Compared to the Pick and paint method, Forcing ATPT is more convenient and requires less interaction while the former is best used to demonstrate the merging process. However, the Pick and paint method is more flexible and allows the affected area to comprise more irregular shapes.

3.3 Examples: a missing image

In an experience, textures obtained from two different stations were visually investigated. Despite having a good view over the facade area, the first station failed to fill the upper right part of the texture (Figure 7-a). This part was missing due to an internal error in the data capture unit. The error was, however, only realised later when the data capture phase had finished. Thus, it was decided to use images from another station to fill in the gap instead of recapturing the data.

Fortunately, the second station, besides covering the missing area, had a good orthogonal view of it (Figure 7-b). Figure 7-c shows the texture from the first station, the missing part of which is filled in using the texture from the other station.


(a)   (b)

(c)
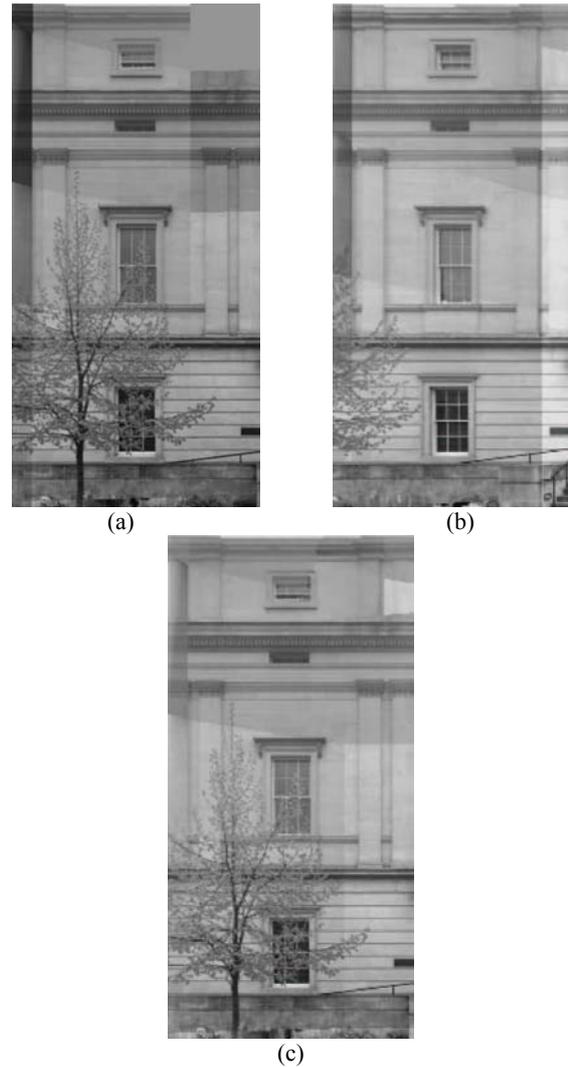
Figure 7. A texture merging example to fill missing part

3.4 Examples: an occluding tree

Here an example is given to show how an occluding tree can be removed from a texture. Figures 8-a to 8-c show textures that all see a building wall from different stations, of which (a) has the most orthogonal view over the area of interest. This texture, however, includes a big tree that occludes parts of the building facade. Comparing the other two textures, it can be seen that (c) does not include the image of the tree and has a good orthogonal view over the tree area. This texture was therefore selected as the source, from which samples are taken and painted over the tree area on (a). For a better demonstration, the merging process was done in two steps. Figures 9-a and 9-b show the original texture with the tree partly (a) and completely (b) removed using the Pick and paint method. As the textures are all created for the same face and at the same resolution, they fit well leading to almost a homogeneous texture.

A further enhancement was to improve the geometric quality of windows on the texture. As can be seen, the windows on the left side of the final texture are slightly stretched out due to perspective distortions. The Pick and paint method can thus be used to improve the quality of the texture reducing the perspective distortions. Figure 10-a shows the final texture, on which some of the windows on the left side are stretched.

(a)

(b)

(c)

Figure 8. Textures of a face from different stations.

(a)

(b)

Figure 9: Improved textures.

(a)

(b)

Figure 10. Improving geometric quality of textures

depends on how similar the sample and the source features are. In other words, this method can only be used if features that are copied have the same size, shape, and orientation.

The above examples show how easily the textures can be merged in order to create more complete and realistic results. The process is convenient and is advantageous over the techniques which simply replace the affected area with a colour or a pattern. It should be noted that, despite being simple and easy to use, the merging technique does not guarantee the removal of all unwanted features or the filling in of all the gaps present in a texture. The success of this method depends mainly on the availability of suitable images on the neighbouring stations.

Since images are captured at different times and position, however, it is quite likely that they form ideal textures when used in conjunction with each other. More importantly, as the textures of a given face are all of the same size, shape, and orientation the merging process is carried out as simply as possible. In other words, the convenience of the procedure is basically due to the appropriate structure of image acquisition and texture processing strategies employed in this project.

4. CONCLUSIONS

This paper reviewed and examined the components of a system developed for fast production of geometrically correct textures for building visually realistic models of buildings. The key concept incorporated in the system is the establishment of an indexed database of oriented images to allow automatic correspondence between images and the geometric model of buildings. This allows finding, rectifying, and mosaicing image portions covering a given face of a building. In addition as textures are all referenced to the actual object space, they can easily be compared and merged if necessary. This in turn, can lead to textures free of occlusions.

As mentioned before, the ATPT uses a geometric model (NFR) to provide a starting point for creating textures. The source for this model can vary, provided that it leads to satisfactory textures. This means not only aerial photographs, which are

To reduce the perspective distortion of the windows, a window in the middle of the texture was copied as the source and pasted over the stretched windows as shown in 10-(b). It should be noted that the safety of this method in terms of "reality"

frequently utilised for building modelling, but also current maps or laser ranging coordinates can easily be used to create textures. This means greater flexibility in the system which suggests it for future applications.

REFERENCES

Brenner, C., Haala, N., 1998. Fast production of virtual reality city models. *International Archives of Photogrammetry and Remote Sensing, 32(4): 77-84.*

Chapman, D.P., Deacon, A.T.D., Hamid, A., 1994. HAZMAP: a remote digital measurement system for work in hazardous environments. *Photogrammetric Record, (14) 83: 747-758*

Debevec, P.E., 1999. Modeling and rendering architecture from photographs. *SIGGRAPH 1999, notes for Course 28.*

Gülch, E., Müller, H., Läbe, T., Ragia, L., 1998. On the performance of semi-automatic building extraction. *Proceedings of ISPRS Commission III Symposium, Columbus, Ohio, July 6-10.*

Maresch, M., 1997. Linear CCD array based recording of buildings for digital models. *PhD thesis, Technischen Universität Graz. April: 132 pages.*

Varshosaz, M., 1999. Visually Realistic Construction of Building Models. PHd. Thesis, University College London, 198 pages.