

TOWARDS AUTOMATIC RECONSTRUCTION OF VISUALLY REALISTIC MODELS OF BUILDINGS

Masood Varshosaz, Ian Dowman, Dave Chapman

Department of Geomatic Engineering, University College London
Gower Street, London, WC1E 6BT, United Kingdom
varshosaz@hotmail.com, idowman@ge.ucl.ac.uk, dchapman@ge.ucl.ac.uk

KEY WORDS: Visually Realistic Models, 3D GIS, Texture mapping, Automatic, Buildings, Modelling.

ABSTRACT

This paper describes an innovation in the field of texture creation for the construction of Visually Realistic Models (VRMs) of buildings in urban areas. The development of an Automatic Texture Processing Tool (ATPT) is described which uses registered CCD images, captured in a special manner, and a Numerical Frame of Reference (NFR) to create textures of individual faces of buildings. The images are acquired on a virtual sphere using a CCD camera which is mounted on a servo-driven theodolite and are indexed using the angular readings from the theodolite. The registration of the images includes the calibration of the data capture unit and the localisation of the imaging stations. In order to process the textures, the coordinates of building faces, extracted from the NFR, are passed to the ATPT. The ATPT finds the corresponding images covering each face and forms the textures by extracting, rectifying, and mosaicing the image portions automatically.

The system not only enables the automatic creation of textures, but also allows one to make highly accurate measurements on the resulting VRM, a task which is not possible using conventional texture processing systems. Textures of a specific face can easily be merged in order to obtain an occlusion-free or higher quality texture. Moreover, independently formed NFRs can be used to initialise the texture creation process.

1 INTRODUCTION

A Visually Realistic Model (VRM) is a digital 3D representation (eg Figure 1) which contains all the information necessary to view and recognise an object from all directions. It not only shows the geometric relationships but also permits realistic renderings of objects. Such three dimensional realistic representation provides designers, planners, the public, and the others with imagery closely coupled with the human experience and, hence, is of great interest (eg Varshosaz, 1997; Brenner and Haala, 1998; El-Hakim *et al.*, 1998).



Fig. 1: Visually Realistic Model of University College London.

The need for VRMs of cities is evident from numerous applications such as city planning, architecture, entertainment, education and training, tourism, and disaster preparedness. In the recent years, there have been a number of advances in the techniques developed for the production of VRMs of buildings. The major improvements have been in the use of laser scanners and the direct use of photographs for modelling and rendering of scenes.

- **Laser range scanning:**
Laser rangefinders are used to reconstruct the scene structure by direct measurement of depth to various points in the scene. Once the geometry of the scene structure is built, the images can be mapped to their real positions to form 3D realistic models. While existing versions of these sensors are generally slow, cumbersome and expensive, active development of this technology is making it of practical use for more and more applications (Debevec, 1999).
- **Image-based rendering techniques:**
In a simple case images are projected to the body of a cylinder having a fixed radius from the camera projection centre. A Quick Time Virtual Reality model (QTVR) is an example of such which is usually built by mosaicing photographs taken on a virtual sphere around the camera. In more sophisticated situations, the model is reconstructed using a set of images and their corresponding depth maps. In other words, using the images the structure of the scene is determined and the images are projected to their determined 3D positions. Some of these techniques employ the computer vision technique of computational stereopsis to automatically construct the structure of the scene from the multiple photographs available. As a consequence, these systems are only as strong as the underlying stereo-correspondence algorithms. This has caused problems because the stereo algorithms have a number of significant weaknesses; in particular the photographs need to have similar viewpoints in order for reliable results to be obtained. These techniques are still subject to research and development and are mainly suitable for small projects.

In general both of the above techniques have the advantage of being able to produce truly 3D and realistic models of buildings. However, they form the model using a huge number of unstructured points. As a result the topology of the scene is either completely lost or is local to the location of the 3D points. In contrast, the system presented in this article, is geometry-based and uses a geometric model to define the position of textures in space and to process the textures of building facades. A Numerical Frame of Reference (NFR), a Texture Image Database (TID), as the source for textures, and an Automatic Texture Processing Tool (ATPT) are used to process the textures. These components are described and examined briefly in the following sections. Further discussions and analysis can be found in Varshosaz (1999) and <http://www.ge.ucl.ac.uk/staff/ucfsmva.html>.

2 THE NUMERICAL FRAME OF REFERENCE

The NFR is a model that shows the geometric structure of buildings which can be rendered using photo-textures to produce realistic models. Most of the work carried out in the reconstruction of 3D models of buildings, has concentrated on the development of techniques for production of NFR models. A NFR is primarily used for making crude 3D measurements and calculating volumes. It carries enough realism to satisfy needs of applications such as small scale town planning or environmental monitoring which require only a simple description of buildings. However, as far as the texture processing is concerned, it is used to define the location, dimensions, and direction of the textures in space. The coordinates of building corners forming the NFR can be derived from various sources and methods. However, as long as this information is available and is of sufficient accuracy the way it is provided is not important.

3 THE TEXTURE IMAGE DATABASE

The TID is an indexed archive of spatially registered CCD images, acquired as described in the next section, and includes the following data:

- Images as captured by the method outlined later;
- Reference file: this file determines a number of elements for each image and enables automatic search for images that cover a particular section of a building face;
- Camera file(s): a camera file includes all details about the internal and external geometry of the CCD camera used for image acquisition;
- Station file: this file includes the exterior orientation parameters of the stations at which the images are taken.

To form a TID the images are captured and the reference file is created. To obtain the station and camera files, the images need to be registered.

3.1 Image capture

The imaging unit (Figure 2) was developed as part of a project called HAZMAP, which was carried out at the Department of Geomatic Engineering of University College London. A main goal of HAZMAP was to minimise human intervention during the image acquisition in hazardous environments (Chapman *et al.*, 1994). The images are captured in digital form and are displayed online on the screen of a computer. The image acquisition is performed using an off-the-shelf Pulnix TM526 CCD camera which is mounted on the motorised Geotronics 510 SN servo-driven Geodimeter.



Fig. 2: Image acquisition unit.

The concept of camera-on-theodolite was originally introduced by Huang and Harley (1990) but is used here for the capture of terrestrial images from building facades in order to speed up the process of texture creation. The system is quite simple and requires no extensive training for its setup and the corresponding operations. At each imaging location (station) a set of image tiles that cover a field of view of 360° (horizontally) by 90° (vertically) is taken. Once captured, the images along with the theodolite angular readings are stored to the hard disc of a PC computer. The process of recording theodolite angles for each image is called indexing. Indexing the images considerably simplifies the task of image registration which is required for the processing of textures. This is because, at any station the angular location of images with respect to each other and the theodolite centre is known. Therefore, by knowing only the exterior orientation of stations, the relative orientation of all images in the database with respect to each other is determined. Using the index data the reference file is created which mainly defines the theodolite angles, dimensions, and the location of each image on the hard disk.

3.2 Registration of CCD images

The registration of images includes the calibration of the data capture unit and the localisation of the imaging stations. Within the camera calibration stage, not only the characteristics of the CCD camera lens but also the parameters relating the geometry of the CCD chip to the rotation centre of the Geodimeter are determined. The images used to calibrate the instrument are usually taken before the image acquisition is started. Having done the calibration, the images are acquired and the stations are localised. Localisation refers to the process in which the exterior orientation elements of the theodolite centre at each imaging station are identified. Central to the calibration and localisation is a mathematical model relating the CCD and object coordinates of a point using an extended form of the Collinearity Equations. Details of this model can be found in Varshosaz (1999). The outcome from the calibration and the localisation stages are the camera and the station files respectively.

Once the images are captured and the reference, camera, and station files are created, the TID is formed and can be used to process the texture of a given face using the ATPT.

4 THE AUTOMATIC TEXTURE PROCESSING TOOL

Given the coordinates of the corners of a building face, the ATPT searches the reference file and automatically extracts, rectifies, and mosaics the image portions covering the face. The output is the texture of the given face. Within the current implementation of the ATPT the faces are assumed to have a rectangular shape. This assumption simplifies the development of the tool and is acceptable for most building facades in urban areas. In addition, the creation of textures for building parts having a different shape follows mainly the same principles as the rectangular faces. The texture processing is carried out for all building faces individually, for each of which a texture is created.

Figure 3 shows how the texture processing is carried out. To from the texture of a face, the 3D coordinates of its corners, extracted from the NFR, are passed to the ATPT that uses the camera calibration, localisation, and reference data to find the images which cover the face of interest. Based on a given spacing interval a 3D grid, the points of which define the elements of the texture, is fitted to the face.

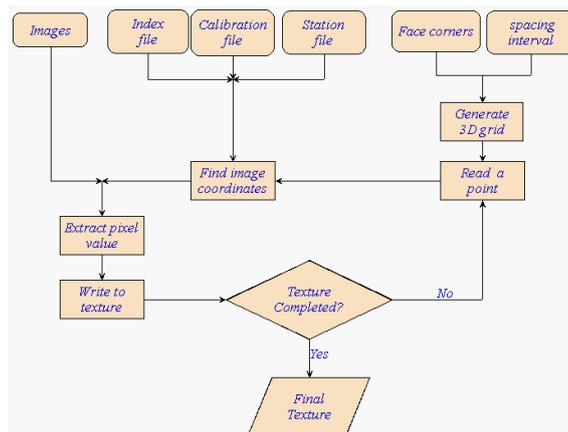


Fig. 3: Flowchart of texture processing

The basic goal of the 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 the 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.

4.1 Finding the best image

For each 3D point, an image has to be found in order to extract the pixel value of the point. Finding the image (Figure 4) is accomplished using the following information:

- H and V: the horizontal and vertical angles, recorded by the theodolite, for the centre of the images;
- Camera calibration parameters;
- Station locations and orientations.

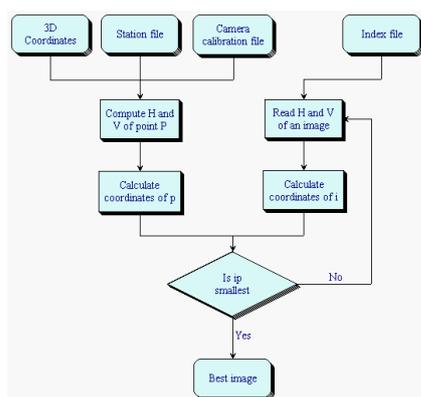


Fig. 4: Find image process

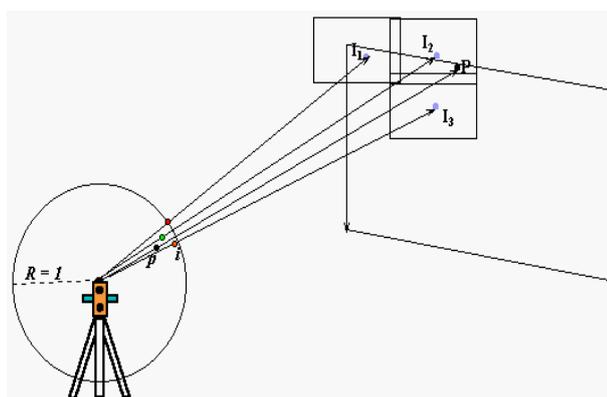


Fig. 5: Distance between the image and point vectors.

This data is used to relate the image and object points via the collinearity equations. The process relies on the computation of a parameter which is used to rank the images; the image with the most appropriate parameter is then selected. This parameter is the distance between two vectors with a size equal to 1. Both vectors start at the CCD camera centre, but one of them points to the centre of the image while the other refers to the point. As shown in Figure 5, the distance ip is used to rank the images. Points p and i are on the vectors pointing to the 3D grid point and the image centre respectively. Once a 3D point is passed to ATPT, the coordinates of p are computed. Then, the angles of the images at the station, determined in the previous stage, are read one set at a time. For each image, the coordinates of i , and consequently the distance ip is calculated. Every time a new image is introduced to ATPT, its ip is computed and compared against the one obtained for the previous image. If the new image has a smaller ip , it is chosen as the best image of the 3D grid point. Otherwise, the old one is kept as the image of the given 3D grid point. This process is repeated for all images at the station. In the end, the image with the minimum ip is defined as the image for the grid point.

4.2 Extracting the pixel value and writing the texture

The pixel value of a 3D grid point refers to the grey level of its projection on the image defined in the previous Section. In order to define this value, the coordinates of the point on the image are computed. Then, using a simple interpolation the grey level associated to the point is estimated. Once the image coordinates of the 3D point are computed, the texture value of the point is to be defined. For this, the image is loaded into the memory by writing all its elements to an array with the same dimensions as those of the image. Each element of this array is identified by a set of integers (i, j) which correspond to the computed y and x respectively. The grey level of the point on the image, is the value of the array at $i = y$ and $j = x$.

Having obtained the pixel value of a 3D grid point, it is placed in the correct position which is defined by its indices (ie row and column) on the grid. The process is repeated for all points of the 3D grid until the complete texture is produced. The task of extracting, rectifying, and mosaicing image portions covering the face of interest, are carried out simultaneously in a pixel by pixel manner. The result is a unique homogeneous rectified image of the building facade which is used as its texture. To actually make use of the texture, it is mapped to the corresponding building face when the VRM is loaded. In other words, the building surface which is initially a blank shaded face, is rendered with the texture used as the render material (Figure 6). The result is a geo-visual model, the components of which are the geometric model and the textures obtained using the ATPT.

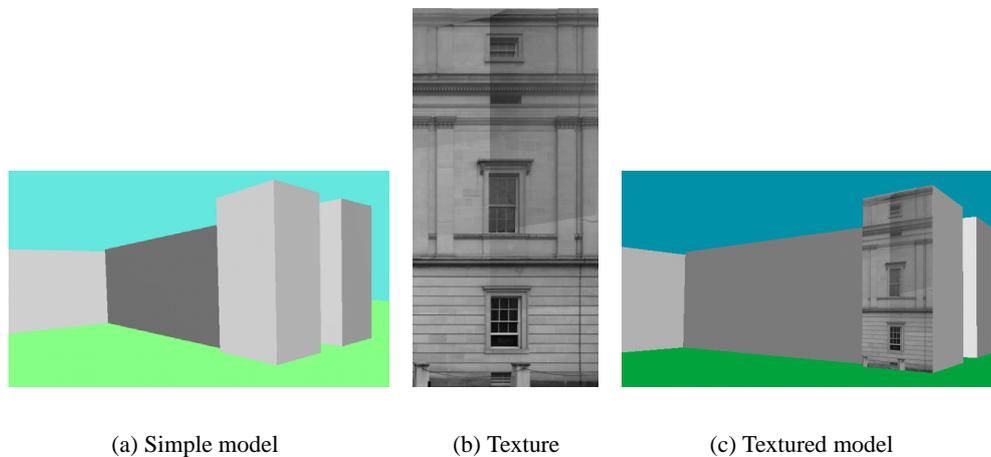


Fig. 6: Building model before and after texture mapping.

5 A CASE STUDY

The area of this case study is one of the buildings of University College London (Figure 7) which is located at the west side of the UCL quadrangle. As already mentioned, to produce the VRM of the building, a NFR is required which is enhanced with textures extracted from the TID. The NFR defines the geometric structure of building while the TID contains the necessary data to extract the textures.



Fig. 7: A view of Chadwick Building.

As shown by Chapman *et al.* (1994), using the registered terrestrial CCD images included in a TID the geometry of buildings can be extracted accurately. Therefore, in this case the TID was used as the source to measure the coordinates of the building corners and to form the textures. The result was the geometric model of the building which was rendered with the textures of its facades. The textures were processed using the ATPT; for each face of the building, a texture was created, the data was structured using VRML for visualisation. Figure 8 shows snapshots from the completed VRM from two different views.



Fig. 8: Views of the completed model.

As can be seen, the texture mapped building is almost identical to that shown in Figure 7 from a visual point of view. It can also be used for making crude 3D measurements and volume-oriented computations, provided that its NFR is structured properly.

The following section examines the ATPT based on the geometric quality of the textures.

5.1 Geometric quality

To evaluate the geometric quality of the textures, two parameters are defined: white pixels and discontinuities. As already mentioned, the textures are created by passing the 3D grid points to the ATPT and extracting the corresponding pixel values. Consequently, the system works properly only if the image coordinates of the grid points are computed correctly. For this, the image selection and the coordinate computation parts of ATPT need to function correctly and accurately. Otherwise, the grid points are projected to wrong places. If the projection of a grid point falls outside the frame of a CCD image, its pixel value is set to 255 which corresponds to white colour. Alternatively, if the projection of a grid point is computed to be within the frame of a CCD image but not on the right place, a wrong value for the pixel of the point is defined. The result of such malfunctioning, is the appearance of discontinuities on the textures.

Figure 9 (a) shows the individual images used to process the texture of one of the building facades. The dimensions of this facade of the building are $17.5 \times 7.5m$, and the dimensions of the corresponding texture are 350×150 pixels. The texture of this face was processed with the spacing interval of $5cm$. The texture processing was carried out with no supervision or manual interactions required. The only interaction was to define the location of a few files and the value of some parameters, like the spacing interval, to initialise the program.

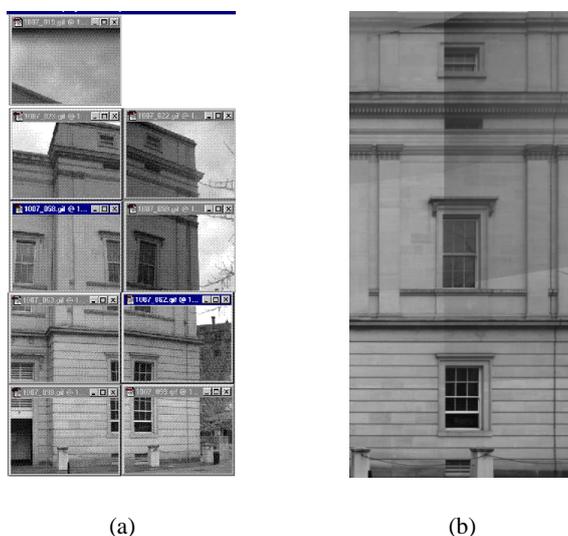


Fig. 9: Individual images and the corresponding texture.

Figure 9 (b) shows the rectified mosaic of the CCD image portions forming the texture of the face. As can be seen, neither white pixels nor discontinuities are present on the texture. This indicates that from a geometric point of view ATPT has functioned properly. However, there are some contrast discontinuities over the right area of the texture, which are the result of the Automatic Gain Control (AGC). AGC is a property of most CCD cameras used to average the image brightness to a consistent level. As a result, darker areas become brighter while lighter ones become darker. It has the desirable effect of compensating for the ambient lighting, but the undesirable effect of changing the signal level of an area of the image, if the lighting conditions change elsewhere in the image. A method to reduce this effect has been developed by Varshosaz (1999).

Figure 10 shows another example. The dimensions of the corresponding facade are $22.7 \times 13.6m$. The texture of this face has 227×136 pixels and was created with a spacing interval of $10cm$. Similar to the previous example, no white pixels or discontinuities are present on the texture. The textures of the rest of the building faces were processed using images from single stations and in all of them no geometric discontinuity or white pixels were observed. These textures were used to form the final VRM presented in Figure 8.



Fig. 10: Another example for a texture created using CCD images.

The results from these tests represent the usability of ATPT in practice as most of the manual work in the processing textures is replaced with unsupervised machine computations. The importance of this can be realised as many textures can be processed as batch files in the computer background, while the user is absent or is concentrated on a different job.

6 CONCLUSIONS AND FUTURE ENHANCEMENTS

This paper showed that complete textures with high geometric quality can be produced automatically using the CCD images captured and structured in the way presented in this project. No interactions are required to relate the geometric model to the images; a job which is usually done by manual observation of several control points on the images.

The components of the system including the image acquisition, the TID formation, and the texture creation are designed such that the formation of VRMs can be performed fully automatically. There is a large research effort at present to produce the geometric models of buildings automatically. Given such a geometric model and the location of the terrestrial stations, the model can be split into individual faces, for each of which a texture is created automatically using the ATPT. The only interactive part is the registration of the images, the status of which will hopefully be improved through the forthcoming research.

As far as the method developed in this project is concerned, to facilitate the process in this direction, further developments are required in the following areas:

- **Data preparation:**
The time to build up a NFR and to register terrestrial images needs to be significantly reduced. Using aerial photographs can greatly improve the speed of NFR creation. However, still there is a large amount of time required to register the terrestrial images. Therefore, techniques capable of automatic bundle adjustment of terrestrial images need to be developed. Only in this case will the production of VRMs become fully practicable and thus acceptable to the community.
- **Image acquisition unit:**
Although the unit is portable and automatic, there is still room for further developments to make it lighter and faster. A possible configuration is to use several CCD cameras, linked to a portable computer, to capture several images at once such that they cover a complete panoramic view. The cameras need to have a rigid position to make sure they will have sufficient geometric quality.
- **Implementation procedures:**
The current version of the ATPT is command-based and thus needs a visual front-end to be easier to use. In addition, currently only textures for rectangular faces can be processed. Thus, improvements are required to cope with non-rectangular shapes. This is usually done using a triangulation technique, ie to triangulate the building face and to produce a texture for each triangle. It should be noted that, the triangulation should be recursively continued until all vertices of a triangle fall on a single image so that its texture can be produced.

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. Modelling and rendering architecture from photographs. *SIGGRAPH 1999*, notes for Course 28.
- El-Hakim, S., Brenner, C., Roth, G., 1998. An approach to creating virtual environments using range and texture. *International Archives of Photogrammetry and Remote Sensing*, Hakodate, 32(5): 331-338
- Huang, Y.D., and Harley, I., 1990. CCD-camera calibration without a control field. *International Archives of Photogrammetry and Remote Sensing*, 28(5): 1028-1034.
- Varshosaz, M., 1997. Building modelling for 3D GIS. MPhil Report, University College London, 126 pages.
- Varshosaz, M., 1999. Visually Realistic Construction of Building Models. PhD. Thesis, University College London, 198 pages.