DOCUMENTATION OF ARCHITECTURAL SCENES USING A HIERARCHICAL METHOD

Nitin Jain S [a,*], Rohini SS [a], Sabarish S M [a], Srinath Sridhar [a], Vidhya R [a]

[a] Institute of Remote Sensing, Dept. of Civil Engineering, Anna University Chennai, Sardar Patel Road
Chennai 600 025, India - (nitin.neo7, jollylyphe, sabarish.sm, srinath1905)@gmail.com, vidhya@annauniv.edu

KEY WORDS: Cultural Heritage, Vision, Modelling, Visualization, Reconstruction, Hierarchical, Three-dimensional

ABSTRACT:

This paper presents a hierarchical two-level method for reconstruction and detailed documentation of architectural scenes from photographs taken using digital cameras. The first level in the method involves creating a *coarse model* of the scene. Multiple images of an architectural scene are automatically matched for corresponding points and *autocalibration* is performed. Subsequently, a projective geometry based approach is adopted and bundle adjustment techniques are applied to create a reconstruction of the scene. The *coarse model* is thus generated with minimal user input, the results of which are presented in this paper.

The second level takes the *coarse model* as the input and creates a detailed documentation of the architectural scene. Documentation consists of a *3D model* and, associated with it, ancillary information about the architectural scene. The 3D model generated may be of a much higher geometric quality than the coarse model. Ancillary information is in the form of an information system containing attributes such as spatial reference, textual information about the structure, historical notes, conservation status and any other information of importance. The second level has two advantages – it avoids field reconnaissance and provides a means for automatic planning of camera positions. The advantages are offered by the concepts of *virtual reconnaissance* and *Automatic Image Acquisition Planning (AIAP)* respectively.

As an extension, an integrated *Architectural Heritage Information System (AHIS)* is proposed. AHIS facilitates efficient management of both spatial and non-spatial ancillary information. The advantages of having a spatial database for architectural heritage will help conservation architects, government agencies and other interested parties to manage resources needed for preservation of built heritage effectively.

1. INTRODUCTION

The awareness of preserving architectural heritage of the world through documentation has been increasing over the past several decades. The applicability of various tools of science for this purpose, apart from archaeology, has also been realised. Photogrammetry has played a central role in this regard. The process of accurate documentation has been hugely aided by automated algorithms, information management systems, visualization tools and general improvements in computers.

Meanwhile, the internet has been growing at an astounding rate. It is so huge a network today that it can almost be called a parallel world, although a virtual one. Just as a person in the real world who wants to visit monuments, the person in the virtual world also wants to visit monuments that are virtual reconstructions of the original.

Reconstruction and documentation are two distinct tasks that the proposed method aims to achieve. By reconstruction, the creation of a 3D model from images, that may not be necessarily an accurate representation, is meant. Documentation means that the scene is captured along with its ancillary information. In this paper, a hierarchical method with reconstruction first and documentation next is described. The reconstruction process uses a multiple view projective geometry approach to modelling. In the second level of the method, the model generated for the first level is used for planning a detailed survey of the scene.

The method described has several advantages including those offered by *virtual reconnaissance* and *Automatic Image Acquisition Planning (AIAP)*. Virtual reconnaissance reduces the number of field visits required. Additionally, the *Architectural Heritage Information System (AHIS)* shows great promise for the future of heritage conservation. Preliminary results of many subunits of the method are shown.

2. RELATED WORK

The result of decades of work by researchers in surveying,

---

* Corresponding author

photogrammetry and computer vision is that different methods for reconstruction and documentation of architectural scenes are available. Photogrammetry and surveying based techniques tend to focus on accurate documentation whereas computer vision techniques focus on automatic modelling and photo-realistic rendering. A brief description of some of the popular techniques is given below.

2.1 Ground Surveying & Modelling

Ground surveying techniques are the oldest techniques where surveyors conduct field measurements of the monument using instruments like the total station and subsequently a virtual model is created using modelling software like CAD. Depending on the scale and details to be recorded, accuracies of the order of millimetres can be obtained (Georgopoulos and Ioannidis, 2004). However, the effort required in transforming the field measurements to accurate and visually appealing 3D models is huge.

2.2 Photogrammetric Techniques

Photogrammetric techniques typically start with a reconnaissance of the scene to be recorded. Metric cameras of very high quality are used. The camera is calibrated to find the interior orientation elements and also the distortion parameters (Wolf and Dewitt, 1983). The photographs taken are processed in advanced digital photogrammetry workstations and this method calls for specially trained users to specify control points on the objects. These workstations are often limited to the generation of contour maps and digital elevation models (DEM). Software like Photomodeler have the capacity to produce 3D models that are both realistic and accurate. However, extensive human interaction is needed.

2.3 Image-Based Techniques

Image-based techniques are, in a way, extension of the photogrammetric technique. However the primary aim is to create 3D models and not the level of accuracy. Work done by Debevec (Debevec, 1996), although not strictly an image-based, uses a composite approach combining a photogrammetric modelling phase with an image-based rendering phase. But the method is not fully automatic. Fully automatic reconstruction of architectural scenes has been explored and demonstrated (Werner and Zisserman, 2002). The work done by Snavely et al. (Snavely et al., 2006) has been developed into a web application called Photosynth. Photosynth generates an interactive 3D environment by combining a set of photographs with main emphasis on visualisation.

3. METHODOLOGY

A two step hierarchical approach is adopted for reconstruction and detailed documentation of architectural scenes. The first step is to create a *coarse model* and the second step is to create a detailed documentation. This hierarchical approach is somewhat similar to the act of *reconnaissance* followed by the actual field work in

surveying. The two step method offers numerous advantages including less number of field visits and automatic image acquisition planning. Each level of the method contains numerous sub-activities that need to be performed in a specific order which is described below.

3.1 Reconstruction

One of the objectives of the hierarchical method is to generate a visually appealing and realistic reconstruction of an architectural scene. The intended audience for this reconstruction are people who want a quick way of visualizing the world in three dimensions. In order to achieve this objective, the first part of the proposed system tries to create a *coarse model* from multiple uncalibrated images of the scene. Additional calibration information about the camera is not needed in this method. Methods for uncalibrated metric reconstruction have been studied and demonstrated (Pollefeys *et al.*, 1999; Hartley, 1994) and these techniques are adopted.

The first step in the coarse model reconstruction process is the pre-processing of the images. This is followed by image sequencing and automatic image correspondence matching which uses invariant descriptors for robust matching of corresponding points. These matches are then used to find camera orientations and relative positions. Autocalibration and bundle adjustment are used to find the 3D positions of the matched points to generate a point cloud. This cloud of points forms the basis of the 3D visualization of the scene.

3.1.1 Input Pre-processing: The pre-processing of input in the case of video sequences is an important step before the image matching process. The method suggested by Kien (Kien, 2005) is adopted which has two steps. The first one is *frame selection* and the second is *sequence segmentation*. Frame selection is a data reduction process in which thousands of frames in a video sequence are reduced to a few hundred for reconstruction. It is essential for proper reconstruction that the chosen frames have good geometric and photometric qualities. The sequence segmentation activity is to ensure that the frames chosen are in the proper sequence, in this case chronologically. This is an essential aspect since the structure from motion algorithm assumes continuous frames.

In the case of still images, the frame selection step is not needed. However, it must be ensured that the images acquired are still in the proper sequence. There are numerous methods to ensure this and one of them is used (Zisserman, 1998). Additionally, for both video and images, certain basic pre-processing tasks like noise removal or normalising illumination are done to obtain better results in matching.

3.1.2 Image Correspondence Matching: Automatic correspondence matching between multiple images is usually performed in two stages. The first is a feature extraction stage and the second is a feature matching stage. In the feature extraction process *keypoints* are identified in an image using edge detectors and corner detectors. Other detectors like SIFT (Lowe, 1999) can also be used as they
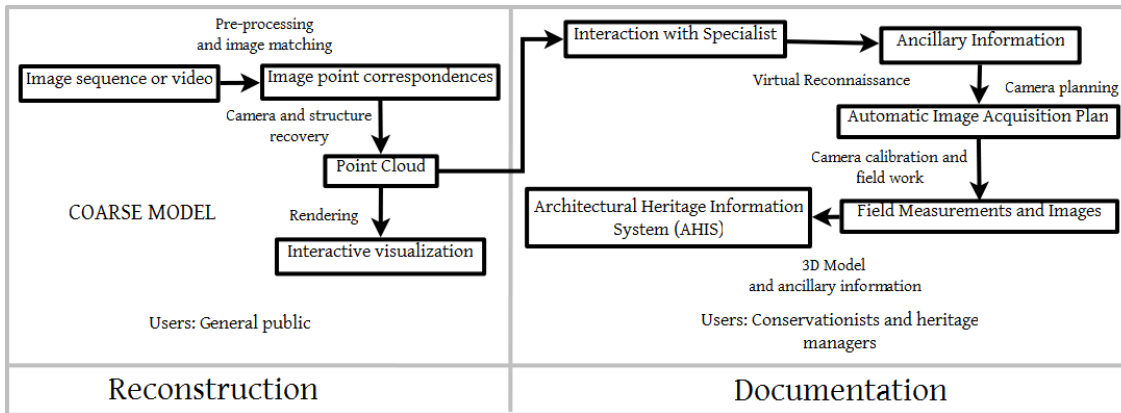
Figure 1. Flow diagram of the two level method

are invariant to image transformations. In the current approach keypoints are identified using the *Harris Corner Detector* (Harris and Stephens, 1998). Subsequently, initial correlation matching is performed. The approach of integrating the correspondence matching process to the underlying geometry of the views is adopted (Zhang *et al.*, 1995). First, the feature matches are used to compute the fundamental matrix. This can be done using a robust estimator like RANSAC. The RANSAC method requires a minimum of 7 point correspondences. The 8-point algorithm may also be used in which case non-linear refinement needs to be done. False correspondences are treated as outliers and are eliminated from the structure recovery process.

3.1.3   Camera Recovery: For two views, once the fundamental matrix has been computed, the cameras can be recovered. The recovered cameras are in the form of the *projectivities* (the camera matrices). These matrices are 3x4 matrices that denote the projective relationship between the point in the world and the imaged point. Thus, for a given pair of images it is possible to define two projectivities $\mathbf{P}$ and $\mathbf{P'}$. These matrices can be retrieved from the fundamental matrix, $\mathbf{F}$ (Hartley and Zisserman, 2003).

For multiple views, the projectivities will have to be found by taking the between-image homographies for consistency. There are several methods to accomplish this and a combination of some of them is adopted in the current implementation (Luong and Vieville, 1996; Pollefeys *et al.*, 1999). The intricacies involved in numerical computation of the camera matrices are briefly discussed by Kien (Kien, 2005).

3.1.4   Structure from Motion: By recovering the cameras their motion in the scene has been fixed. Now, a projective reconstruction of the scene can be obtained from the recovered camera positions. This can subsequently be upgraded to a metric reconstruction (Zisserman et al, 2002).

It is important that the estimation of the 3D points be done so that the *reprojection error* is minimised. The process that does this is called ***bundle adjustment*** (Hartley and Zisserman, 2003). A successful implementation of bundle adjustment based reconstruction is given by Snavely *et al.* (Snavely *et al.*, 2006). A method similar to this is adopted in the current implementation. Snavely *et al.*'s implementation, also aimed at reconstruction of architectural tourist spots, involves an incremental approach. Initially, a single pair of images is taken and then a two frame bundle adjustment is done. Subsequently, another camera is added to the optimisation. However, one important difference between the current implementation and that of Snavely *et al.* is that external information from the image such as EXIF tags are not considered.

The discrete points generated as a result of the bundle adjustment are in projective space. Work on how to upgrade this projective reconstruction to affine and metric has been done by many researchers (Triggs, 1997). The autocalibration technique allows the projective reconstruction to be upgraded to metric. This is the method adopted in the current implementation.

3.2   Detailed Documentation

The second step in the hierarchical method is to provide a means for detailed documentation of architectural scenes. Documentation includes textual information of historic and cultural importance. The aim, therefore, is to make the second step an integrated tool for specialists including archaeologists, conservation architects and others interested in heritage conservation.

One of the unique features of this part of the method is that the number of visits that will have to be made to the structure under study is reduced drastically. This is accomplished by *virtual reconnaissance*. Reconnaissance is often the toughest part of a survey operation and the part that is most crucial. From the *coarse model* generated by the techniques described earlier a physical representation of the scene ready for planning a more detailed survey for documentation can be had.

Due to the greater emphasis on accuracy and recording of details two important aspects come into play. These are camera calibration and control point acquisition. Camera calibration makes the internal camera parameters known

prior to the survey. Control points on the scene are necessary for upgrading the projective or affine model recovered to a Euclidean model. They are also necessary for quality assessment. In the following sections, the individual steps of the documentation process are explained.

3.2.1 Virtual Reconnaissance: Virtual reconnaissance allows the model generated from the methods described in Section 3.1 to be used as virtual reality environment for planning and executing a detailed documentation of the scene.

Virtual reconnaissance is enabled by an interface for navigation and interaction that displays the coarse model. This interface uses the motion planning and 3D navigation system suggested by Li and Ting (Li and Ting, 2000). A path planning algorithm is used with probabilistic roadmap to help users avoid unnecessary manoeuvres due to collisions with the environment. This system is especially suited for architectural scenes.

3.2.2 Interaction with Specialist: The second most important aspect of documentation is the ability of the users to contribute their expertise before the actual data acquisition starts. In this interaction, the user provides information about specific aspects of the scene that must be considered during image acquisition and data gathering. These specifics might include 3D model parameters, impossible camera positions, historical details, local knowledge, etc.

3D model parameters are the parameters that describe the aspects of the model such as accuracy, scale, level of detail, etc. This is required because, in general, the disposition of the cameras is dependent on these parameters. The user will also be able to specify regions in the space of the scene where it is impossible to have the camera - occlusions. This may be due to objects like trees or other buildings.

3.2.3 Camera Calibration: Unlike in the coarse model generation, the 3D model in documentation requires camera calibration. Any camera calibration procedure is interested in determining the focal length and the principal point co-ordinates (Wolf and Dewitt, 1983). Camera calibration is also concerned with finding the distortion parameters of the camera.

For the camera calibration, a projective technique has been used. This method computes the homographies between the world co-ordinates and the image. The calibration process may be done by the so called eight point method that requires eight known points (Kushal *et al.*, 2003).

3.2.4 Automatic Image Acquisition Planning (AIAP): In this stage, the system has all the information that it needs to plan the positions of the camera for image acquisition. The knowledge that the system has gained from the user on accuracy, level of detail, in accessibilities, etc. is used to plan the image acquisition. The geometry of multiple images and baselines of pairs of images are also considered.

An overview of camera planning techniques is given by Christie *et al.* (Christie *et al.*, 2005). Although primarily concerned with graphics and cinematography, the methods described are equally applicable to finding camera positions in the real world. A camera planning agent called **CAMPLAN** has been found to work well and this has been adopted in the current implementation (Halper and Olivier, 2000). Evolutionary algorithms are used in this method. The advantage with the method is that constraints on the camera positions can be explicitly stated which is exactly what is needed for the automatic image acquisition planning. An example of an image acquisition plan is shown in Figure 2.
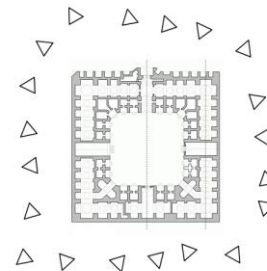

Figure 2. Sample output from the AIAP software module

3.2.5 Field Work and Reconstruction: Field work is essential for creating a 3D model in documentation. However, a considerable amount of the planning activity is automated by the image acquisition planning system. Thus, when compared to conventional survey techniques (Georgopolous et al, 2004) the field work involved here is very less. It is sufficient for the user to visit the field once. During this both control points and images are acquired simultaneously.

The survey of control points can be done using conventional surveying techniques. Total stations and GPS receivers are used for this purpose. These control points are used in the 3D reconstruction process to upgrade the projective model to Euclidean. Also, markers were pasted on the scene to enable easy identification by an automatic marker identification algorithm.

The second step in field work is image acquisition. The camera that has been calibrated is used with the same internal parameters. The acquisition is simplified considerably as the positions are already suggested by the automatic system. An example of a dataset of images acquired is shown in Figure 3.

The reconstruction process used is similar to the one described in Section 3.1 except that the reconstruction here is Euclidean.

3.2.6 Architectural Heritage Information System (AHIS): Apart from maps and 3D models conservationists also need to record textual, historic and cultural information about a particular monument. The management of all this information can often be a very cumbersome task. It seems logical that a spatial information system be

used for this purpose. Such a system is capable of managing both spatial i.e. the 3D model and non-spatial contextual information. This type of system would be modelled based on a Geographical Information System (GIS).

As far as previous work on such an information system is concerned, an Architectural Photogrammetry Information System (APIS) has been proposed (Herbig and Waldhäusl, 1997). The APIS is an internet based system that shall be able to deliver end users with information. This and other systems concentrate more on information delivery than about architectural heritage and less on information management. We propose to use concepts of data management, GIS and database to create an all encompassing information system for documenting monuments.

4. RESULTS AND DISCUSSION

In order to test the applicability of the hierarchical method a test architectural scene was needed. The 90-year old College of Engineering building of Anna University Chennai was chosen. This building is a perfect example of colonial *Indo-Saracenic* style of architecture in India. The building does not come under the purview of the *Archaeological Survey of India*, which takes care of the preservation of buildings more than 100 years old.

The two levels of reconstruction and documentation were carried out for this scene. The reconstruction was accomplished by taking approximately 100 photographs on an 8 megapixel point and shoot camera (Figure 3). These images were then passed to the image sequencing and image matching module of the software.

Figure 3. A subset of images of the CEG Building

The matched images are then considered as a whole by the structure-from-motion algorithm of the software and a projective reconstruction is obtained. In a projective reconstruction, lengths and angles are not preserved. This reconstruction is upgraded to metric by autocalibration. The output point cloud generated for the CEG by our software is shown in Figure 4. This point cloud is comparable to those generated by many other software applications such

as Insight3D.

The point cloud is rendered using a graphics library and key bindings allow interactive visualization of the point cloud. This completes the reconstruction level.

To complete the first few steps of the documentation level the interactive visualization part of the software module was fitted with extra features to allow all external information to be input for *virtual reconnaissance* and AIAP. A sample output from the AIAP module is shown in Figure 2. It shows the plan of a building surrounded by the possible camera positions needed to create a detailed 3D model of the scene. For the documentation process a DSLR camera was used. An independent software module was built for this and the camera was calibrated.

The final step in detailed documentation, AHIS, requires lots of ancillary information about the scene. This was obtained by contacting the Estate Officer of the CEG Building and collecting suitable records for the building. All this information was fed into a GIS database for easy access later. Currently, integrating the 3D model into the AHIS in the form of a 3D GIS is being done. When complete, all the information obtained in both levels of the method described can be stored in a GIS database and retrieved at will. The aim of the work in progress is to create a browser interface for accessing this database via the WWW.

5. CONCLUSION

A two step hierarchical method for creating 3D reconstructions from multiple uncalibrated images and a documentation system that uses the coarse model for detailed documentation has been described. The advantages of the system like virtual reconnaissance and AIAP has also been described along with the recording of ancillary information of the scene and its integration into the AHIS. Results on implementations of many parts of the system have also been discussed.

The applicability of this system is limited only by the ability to implement it. Reconstruction from uncalibrated images is a much explored field and so is rendering of computer models. There is an increasing need for 3D modelling in the present world, with more demand for better visualization capabilities and easy rendering. Less user intervention is one great advantage of this method in contrast to the conventional surveying techniques. Ultimately, one envisions the system being used by conservationists and general public alike for documentation and visualization.
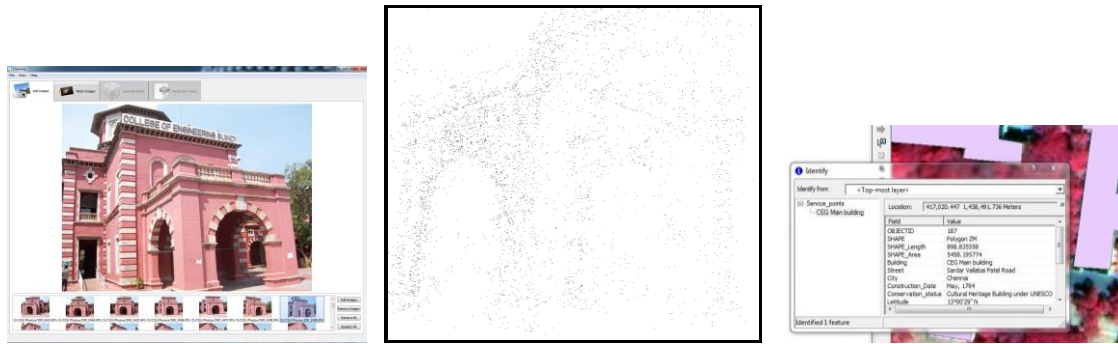
Figure 4. (a) Images loaded in the software for matching (b) the point cloud generated (c) AHIS database

REFERENCES

Christie, M., Machap, R., Normand, J., Olivier, P., and Pickering, J., 2005. Virtual camera planning: A survey. Lecture notes in computer science.

Debevec, P.E., 1996. Modeling and Rendering Architecture from Photographs. PhD thesis, University of California at Berkeley

El-Hakim, S., Beraldin, J.A., and Picard, M., 2002. Detailed 3d reconstruction of monuments using multiple techniques, pp. 13-18.

Georgopoulous, A., and Ionnidis,C., 2004. Photogrammetric and Surverying methods for the geometric recording of archaeological monuments.

Halper, N., and Olivier, P., 2000. Camplan: A camera planning agent, pp. 92-100.

Harris, C., and Stephens, M., 1998. A combined corner and edge detector. Plessey Research Rolce Manor, U.K.

Hartley, R., and Zisserman, A., 2003. *M*ultiple view geometry in computer vision. Cambridge University Press.

Hartley, R.I., 1994. Euclidean reconstruction from uncalibrated views. pp. 237.

Herbig, R.I., and Waldhausl, P.,1997. *APIS- architectural photogrammetry information system, International Archives of Photogrammetry and Remote Sensing.* pp. 23-27.

Kien, D.T., 2005. A review of 3d reconstruction from video sequences. Technical report.

Kushal, A.M., Chanda, G., Shrivastana, M., Banerjee,S., 2003. Multilevel modelling and rendering of architectural scenes.

Li, T.Y., and Ting, H.K., 2000. An intelligent user interface with motion planning for 3D navigation. pp. 177.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. pp. 1150-1157.

Pollefeys, M., Koch, R., and Gool, L.V., 1999. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision.* pp. 7-25.

Pollefeys, M.,VanGool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., and Koch, R., 2004. Visual modeling with a hand-held camera. *International Journal of Computer Vision.* pp. 207-232.

Pritchett, P., and Zisserman, A., 1998. Matching and reconstruction from widely separated views. Lecture Notes in Computer Science. pp. 78-92.

Snavely, N., Seitz, S.M., and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. pp. 835-846.

Summers, F., Atalan, N., Aydin, N., Basagac, O., and Ucar, G., 2002. Documentation of archaeological ruins and standing monuments using photo-rectification and 3d modelling.

Triggs, B., 1996. Factorization methods for projective structure and motion. pp. 845-851.

Vilbrandt, C., Pasko, A., Pasko, G., Goodwin, J.R., and Goodwin, J.M., 2000. Digital preservation of cultural heritage through constructive modeling. pp. 183-200.

Werner, T., and Zisserman, A., 2002. New techniques for automated architectural reconstruction from photographs. Lecture Notes in Computer Science. pp. 541-555.

Wolf, P.R., and Dewitt, B.A., 1983. *Elements of photogrammetry.*

Zhang, Z., Deriche, R., Faugeras, O., and Luong. Q.T., 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artificial Intelligence. pp. 87-119

Triggs, B., 1997. Autocalibration and the absolute quadric. CVPR. pp. 609