Towards Automatic Modeling of Monuments and Towers

El-Hakim, Sabry; Beraldin, Jean-Angelo; Lapointe, Jean-François

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

Towards Automatic Modeling of Monuments and Towers *

El-Hakim, S., Beraldin, J.-A., and Lapointe, J.-F.

June 2002

* published in: IEEE Proceedings of the International Symposium on 3D Data Processing Visualization and Transmission (3DPVT 2002), Padova, Italy. June 19-21, 2002. NRC 44910.

Towards Automatic Modeling of Monuments and Towers

Sabry El-Hakim, J.-Angelo Beraldin, Jean-François Lapointe
Visual Information Technology (VIT) Group
Institute for Information Technology, National Research Council,
Ottawa, Ontario, Canada K1A 0R6
E-mail: {Sabry.El-Hakim, Angelo.Beraldin, Jean-Francois.Lapoints}@nrc.ca

Abstract

Three-dimensional modeling from images, when carried out entirely by a human, can be time consuming and impractical for large-scale projects. On the other hand, full automation may be unachievable or not accurate enough for many applications such as culture heritage documentation. In addition, three-dimensional modeling from images, particularly fully automated methods, requires the extraction of features, such as corners, and needs them to appear in multiple images. However, in practical situations those features are not always available, sometimes not even in a single image, due to occlusions or lack of texture on the surface. Taking closely separated images or optimally designing view locations can preclude some occlusions. However, taking such images is often not practical and we are left with small number of images that do not properly cover every surface or corner. The approach presented in this paper uses both interactive and automatic techniques, each where it is best suited, to accurately and completely model monuments and towers. It particularly focuses on automating the construction of unmarked surfaces such as columns, arches, and blocks from minimum available clues. It also extracts the occluded or invisible corners from existing ones. Many examples, such as Arc de Triomphe in Paris, Florence's St. John baptistery at Santa Maria del Fiori Cathedral, and other monuments and towers from around the world are completely modeled from a small number of images taken by tourists.

1. Introduction

This paper addresses several interconnected issues: full automation versus partial automation, how to handle the inevitable occlusions and lack of features or texture, and the importance of high accuracy to constructing and documenting monuments and towers. We will address only image-based approaches. However, it is important to note that to achieve complete geometric details, range sensors will also be required for sculpted surfaces that are usually found on many monuments [1]. This requires the integration of the two types of data [2].

1.1. Full Automation versus Human Interaction

Three-dimensional modeling from images, when carried out entirely by a human, can be very time consuming and impractical for large-scale projects. Efforts to increase the level of automation are essential in order to broaden the use of this technology. So far, however, the efforts to completely automate the processing, from image capture to the output of a 3D model, are not always successful or applicable [3, 4]. Full automation has been achieved under certain conditions and up to finding point correspondence and camera positions and orientation [5]. Self-calibration and 3D construction still requires a human in the loop either to specify constraints or to perform post processing [6, 7]. Also some sacrifice to the accuracy and fidelity of the created model may result when using full automation (see section 1.3). Automated methods also rely on features that can be extracted automatically from the scene, thus occlusions and un-textured surfaces are problematic. We often end up with areas with too many features that are not all needed for modeling, and areas with no or too few features to produce a complete model. This means that post processing is often required which means that user interaction is still needed. Most impressive results were achieved with highly interactive approaches e.g. [8]. Some interactive approaches with automated features that take advantage of environment constraints proved effective [3, 9]. Other more automated techniques that target specific objects such as architecture [10, 11, 12] have also been developed. If the goal is creating accurate and complete 3D models of medium and large scale objects under practical situations using only information contained in images, then full automation is still in the future.

Full automation is a priority for certain applications such as navigation, telepresence, augmented reality, and where a model is needed fast for decision-making. In those applications, complete details and high accuracy are secondary. For other applications such as documentation and even virtual museums full automation cannot excuse the missing details or lack of accuracy.

1.2. Occlusions and Lack of Texture

Three-dimensional measurement and modeling from images obviously requires that relevant points be visible in the image. This is often not possible either because the points or region of interest are hidden or occluded behind an object or a surface, or because there is no mark, edge, or visual feature to extract. In fact even without multiple objects in the scene and when we can take images from well planned positions, there are not many objects that can be imaged without having portions of its surfaces either invisible or without texture to extract. In objects such as architectures and monuments in their normal settings we are also faced with restrictions limiting the positions from which the images can be taken. Also illumination variations and shadows hamper feature extraction. Not only those factors preclude the modeling of occluded parts but also have negative effect on the modeling of visible parts, for example when applying automatic matching.

1.3. Accuracy of 3D Modeling

Historic monuments and towers are particularly important and thus need to be constructed with high accuracy both for documentation and visualization purposes. To achieve the needed accuracy, one must use the most rigorous approach for 3D modeling from images rather than the simplest or easiest to implement. Tests showed that methods based on projective geometry, although an elegant and efficient approach, result in geometric errors in the range of 4 to 5% [6, 13]. This means that 20-meter tower could have a significant 1-meter error. Photogrammetric methods such as bundle adjustment and proper camera calibration [14, 15, 16], although interactive and not as easy to use as projective methods, give several orders of magnitude smaller error, in the range of 0.01-0.001% on well defined features, depending on camera resolution and lens quality.

2. Outline Of The Approach

Our approach is photogrammetry-based. In order to increase the level of automation, the process takes advantage of properties found in monuments and towers. For example those structures usually have:

- Well defined surface shapes
- Well defined openings such as archways
- Regular blocks attached to flat surfaces
- Many symmetric sections
- Columns with known shape

The approach does not aim to fully automate the procedure nor completely rely on human operator for reasons discussed in section 1.1 above. It provides enough level of automation to assist the operator without sacrificing accuracy or level of details. Figure 1 summarizes the procedure and indicates which step is interactive and which is automatic (interactive operations are grayed). Images are taken, all with the same camera set up, from positions where the object is suitably showing. Parts of the object should appear in two or more images when possible, and there should be a reasonable distance, or baseline, between the images. Several features appearing in multiple images are interactively extracted from the images, usually 12-15 per image. The user points to a corner and label it with a unique number and the system will accurately extract the corner point. Harris operator is used [17] for its simplicity and efficiency. Image registration and 3D coordinate computation are based on the photogrammetric bundle adjustment approach for its accuracy, flexibility, and effectiveness [18] compared to other structure from motion techniques. Advances in bundle adjustment eliminated the need for control points or initial approximate coordinates. Many other aspects required for high accuracy such as camera calibration with full distortion correction have long been solved problems in Photogrammetry [16] and will not be discussed in the remainder of the paper.

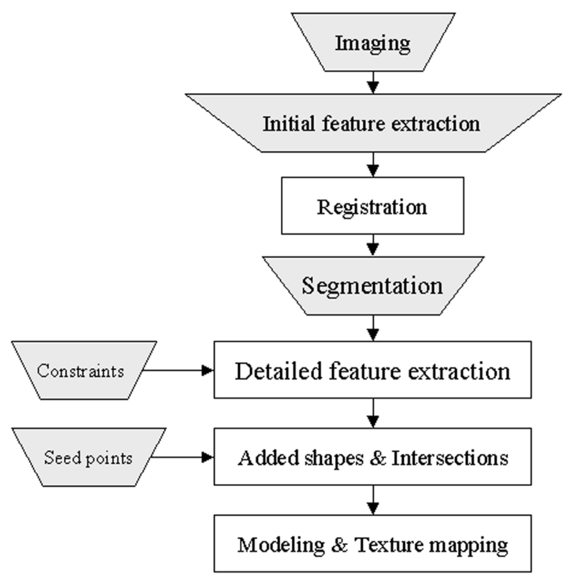

Figure 1. Simplified diagram of the procedure. Interactive operations are grayed.

We now have all camera coordinates and orientations and the 3D coordinates of a set of initial points, all in the same global coordinates system. The next interactive operation is to divide the scene into connected segments to define the surface topology. An automatic corner extractor, again the Harris operator, is used and a matching procedure is applied across the images to add more points into each of the segmented regions. The matching is constrained by the epipolar condition and disparity range computed from the 3D coordinates of the initial points. The bundle adjustment is repeated with the newly added points to improve on previous results and re-compute 3D coordinate of all points.

We now need to add more points in order to reconstruct un-textured surfaces and those that are occluded. Subdivision techniques [19] are used to add points on free-form shapes where some seed points are available. The points are then projected onto the images in order to determine texture coordinates. This results in a smooth appearance of sculptured surfaces.

Since many parts of the scene will show only in one image, an approach to extract 3D information from a single image is necessary [20]. Our approach applies the equation of the surface as a constraint, along with the camera parameters, to the single-image coordinates to compute the corresponding 3D coordinates. For example in many monuments and towers, the walls are planes that are either parallel or perpendicular to each other. The equations of some of the planes can be determined from seed points previously measured. The remaining plane equations are determined using the knowledge that they are either perpendicular or parallel to one of the planes already determined. With little effort, the equations of all the planes on the structure can be computed. From these equations and the known camera parameters for each image, we can determine 3D coordinates of any point or pixel from a single image. This can also be applied to surfaces like quadrics or cylinders whose equations can be computed from existing points. Other constraints, such as symmetry and points with the same depth or same height are also used.

The general rule for adding points on cylinders or columns, arches, and blocks and for generating points in occluded or symmetrical parts is to do the work in the 3D space, like in a CAD system, to find the new points then project them on the images using the internal and external camera parameters. The texture images are edited afterwards to remove the occluding objects and replace them with texture from current or other images. We specifically designed features in the approach to automatically add columns, arches, and blocks. The cylinder is constructed after its direction and approximate radius and position have been automatically determined from four seed points (figure 2-a) using quadric formulation [21]. The ratio between the upper and the lower circle can be set in advance. It is set to less than 1.0 (about 0.85) to create a tapered column. From this information, points on the top and bottom circle of the column (figure 2-b) can be automatically generated in 3D resulting in a complete solid model.

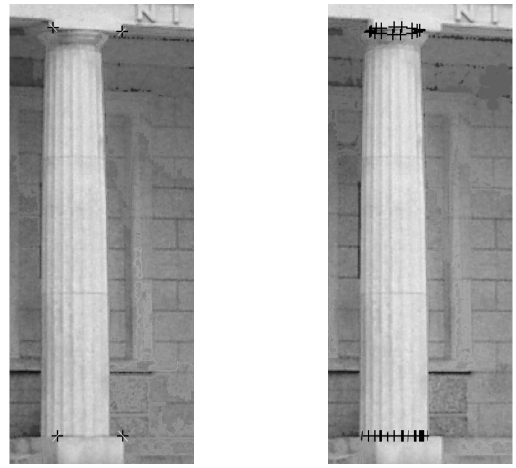

Figure 2. Left (a) 4 seed points are extracted on the base and crown of the column, right (b) column points are added automatically.

Arches are constructed by first fitting a plane to seed points on the wall (figure 3-a). An edge detector is applied to the region (figure 3-b) and points at constant interval along the arch are sampled. For edge detection, a specially designed morphological operator was developed (a variation on [22]). Using the image coordinate of these points (in one image only), the known image parameters (from the bundle adjustment), and the equation of the plane, the 3D coordinates are computed (figure 4).

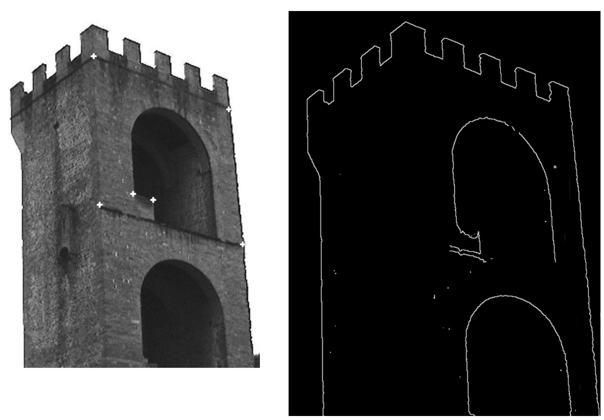

Figure 3. Left (a) shows seed points extracted to fit a plane, right (b) shows edge detector.

It often happens that only part of a monument section, we will call it a block, is visible. For example in figure 5 the bottom part of the block where it meets another block surface is not visible and need to be measured in order to reconstruct the whole block. To solve this problem, we first extract the visible corners from several images and compute their 3D coordinates. We then fit a plan to the top of the base block, using the gray points in figure 5, then project normal to the plane from each of the corners of the block attached to it (the white points). The intersections of each normal will produce a new point (a black point in figure 5) automatically. We now have sufficient points to fully construct the block. More details of the procedure are given in the following examples.

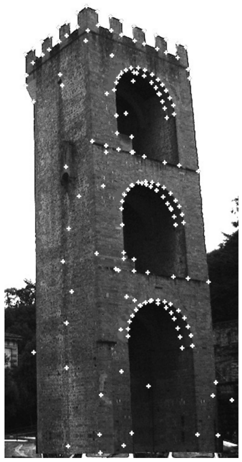

Figure 4. Results of automatic point extraction on corners and selected edges (arches).

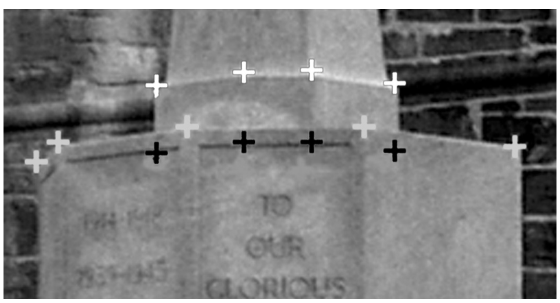

Figure 5. Constructing blocks.

3. Examples

Over the past year, members of our group visited different cities around the world. Whenever possible, they took images covering various interesting monuments. The images were taken during routine tours without any advanced planning of where to take the images. We took the images just like any typical tourist, by walking around the monument and getting the best view under real conditions such as presence of other tourists, vehicles, and other buildings and objects. Several types of digital cameras and regular film cameras (where the film was digitized later) were used. The results were very encouraging and compelling. Over 100 models were created using this approach, each one usually in 1-2 days of work by one person. The number of points and level of interaction and automation obviously varied significantly from one model to another. Usually between 500 – 3000 points were needed, at least 80% of which were generated automatically. Eight examples are presented here (they and several more are on the web [23]), each to illustrate specific feature. They are presented in wire-frame, solid model without texture, and solid model with texture, in figures 6 to 13. In some of the monuments, we found dimensional information available in travel or history books. This information was not used or needed in the model construction, but was valuable in evaluating the accuracy.

Figure 6 shows the Arc de Triomphe in Paris. The Olympus C3030 digital camera (3.1 Mega-pixels) was used (14 images). The arc measures 45 m x 22 m, as indicated in some tourist guides (height varied from one source to another, thus it was not used for evaluation). We used one distance (the 22 m width of one side) to scale our model. From the model, the dimensions on the four sides were: 22 m (fixed for scale), 22.06 m, 44.85 m, and 44.89 m. This gives an error of 0.28%. One should point out that the given dimensions are probably rounded off.

Figure 6. Arc de Triomphe, Paris (14 images). Illustrates automatic arches.

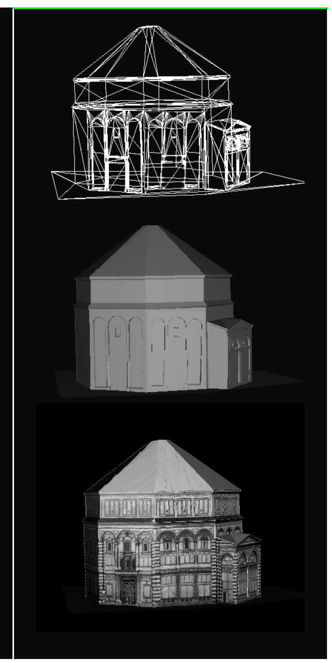

Figure 7. St. John Baptistery, Florence (8 images). Illustrates automatic blocks

The next example is the St. John baptistery in Florence (figure 7). The Olympus E-10 (4 Mega pixels) camera was used to take eight images. The baptistery has eight sides. The actual dimensions were obtained from a plan in a book. The sides average about 13 m in length. Again we will assign 13 m to one side and use it to scale the whole model. The average difference between the model sides and the actual sides is less than 1 cm, or 0.07%. This is significantly better than the accuracy of the Arc de

Triomphe (figure 6). This is due to the better camera used (higher resolution, larger pixel size, and better quality lens) and smaller size object with good feature definition.

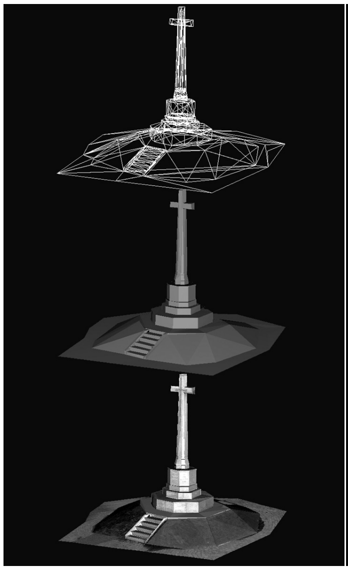

Figure 8. The WWII monument, Quebec City (6 images). Illustrates automatic blocks.

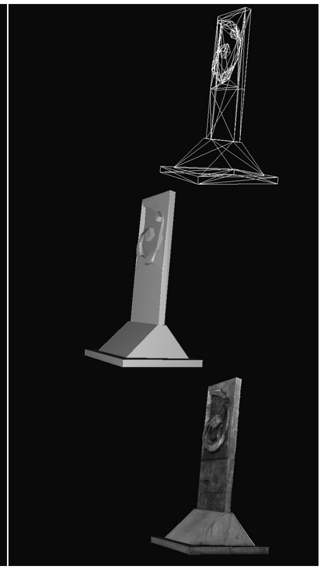

Figure 9. Monument to Galileo, Padova (5 images). Illustrates automatic irregular blocks.

The monument shown in figure 8 consists mostly of blocks, including the steps. After extracting the visible corners, all remaining points needed for complete reconstruction of the monuments were easily added using the block approach described in section 2. Figure 9 shows a relatively uncomplicated monument. Corners of the main structures are first extracted and plane equations of each surface are computed. Sculptured details that are attached to the surfaces are added by automatically extracting the top most points on the sculptures, applying our constrained matching technique to compute their 3D coordinates, then projecting normal from each to the plane to which they attached. The tower shown in figure 10 includes three arch-shaped openings. Points on these arches are automatically measured using the procedure illustrated in figures 3 and 4. The inside points of the arches, even though they do not appear in any image, were measured by intersecting the outside points with the back plane along the normal to that plane. Figure 11 shows a modern monument in Dublin. Only 5 points were measured interactively on the sphere, then a sphere equation is fitted to these points and 1000 more points were added automatically. The examples shown in figures 12 and 13 illustrate automatic modeling of columns, cylinders, steps and blocks.

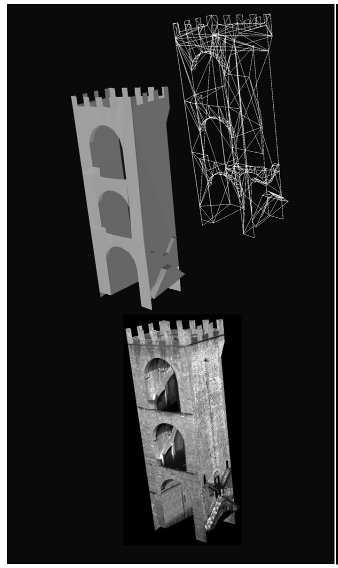

Figure 10. G. Poggi Tower, Florence (8 images). Illustrates automatic arches.

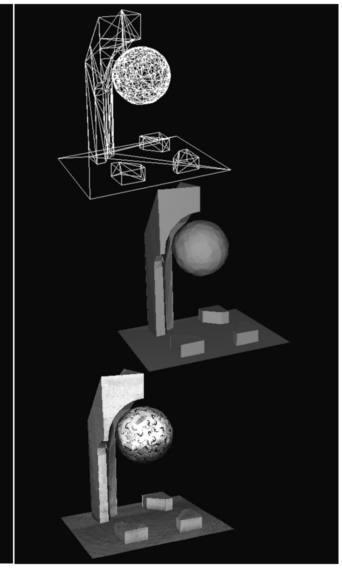

Figure 11. Modern Monument in Dublin (5 images). Illustrates automatic spheres.

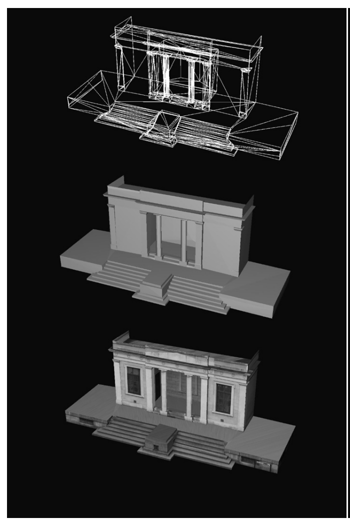

Figure 12. Trinity College building, Dublin (2 images). Illustrates automatic columns and steps.

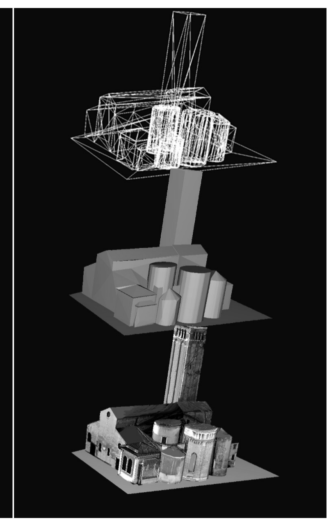

Figure 13. San Giacomo dell'Orio, Venice (6 images). Illustrates automatic cylinders.

4. Conclusion

A semi-automatic approach for constructing medium and large-scale objects, mainly monuments and towers, was presented. Several representative examples from images taken by tourists were given. Parts of the process that can straightforwardly be performed by humans, such as registration, extracting seed points, and topological segmentation, remain interactive. Numerous details plus the occluded and the un-textured parts are added automatically by taking advantage of some of the object characteristics and making some realistic assumptions. Efforts to automate the whole procedure are continuing and will undoubtedly intensify in the future. In the mean time in order to achieve immediate and useful results, parts of the process necessitate human interaction.

5. Acknowledgements

Our colleagues François Blais and Eric Paquet took many of the images. Emily Whiting constructed some models.

6. References

[1] Beraldin, J.-A., F. Blais, L. Cornouyer., M. Rioux, S.F. El-Hakim, R. Rodella, F. Bernier, N. Harrison, "3D imaging system for rapid response on remote sites", IEEE proc. of 2nd. Int. Conf. On 3D Digital Imaging and Modeling (3DIM'99), pp 34- 43, 1999

[2] El-Hakim, S.F., "3D modeling of complex environments", In Proc. Videometrics and Optical Methods for 3D Shape Measurement, San Jose, California, SPIE Vol. 4309, pp 162-173, Jan. 2001.

[3] Shum, H.Y., R. Szeliski, S. Baker, M. Han, P. Anandan, "Interactive 3D modeling from multiple images using scene regularities", European Workshop 3D Structure from Multiple Images of Large-scale Environments – SMILE (ECCV'98), pp 236-252, 1998.

[4] Oliensis, J., "A critique of structure-from-motion algorithms", Computer Vision and Image Understanding, 80(2), pp 172-214, 2000.

[5] Pollefeys, M., R. Koch, M. Vergauwen., L. Van Gool, "Hand-held acquisition of 3D models with a video camera", IEEE proceedings of 2nd. Int. Conf. On 3D Digital Imaging and Modeling (3DIM'99), pp. 14- 23, 1999.

[6] Pollefeys, M., R. Koch, L. Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters", International J. of Computer Vision, 32(1), pp. 7-25, 1999.

[7] Faugeras, O., L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, I. Zoghlami, "3-D reconstruction of urban scenes from image sequences", Computer Vision and Image Understanding, 69(3), pp 292-309, 1998.

[8] Debevec, P., C.J. Taylor, J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach", SIGGRAPH'96, pp. 11–20, 1996.

[9] Gruen, A., "Semi-automatic approaches to site recording and modeling", International Archives of Photogrammetry and Remote Sensing, Volume 33, Part B5A, Commission V, pp 309-318, Amsterdam, July 16-23, 2000.

[10] Liebowitz, D., A. Criminisi, A. Zisserman, A., "Creating Architectural Models from Images", EUROGRAPHICS '99, 18(3), 1999.

[11] Tarini, A., P.Cignoni, C.Rocchini, R.Scopigno, "Computer assisted reconstruction of buildings from photographic data", Vision, Modeling and Visualization 2000 Conference Proc., pp. 213-220, Saarbrucken, DE, November 2000.

[12] Dick, A.R., P.H.Torr, S.J. Ruffle, R.Cipolla, "Combining single view recognition and multiple view stereo for architectural scenes", Proc. 8th IEEE International Conference on Computer Vision (ICCV'01), pp. 268-274, July 2001.

[13] Georgis, N., M. Petrou, J. Kittler, "Error guided design of a 3D vision system", IEEE Trans. PAMI, 20(4), pp. 366-379, 1998.

[14] Brown, D.C., "The bundle adjustment - Progress and prospective." International Archives of Photogrammetry, 21(3): 33 pages, ISP Congress, Helsinki, Finland, 1976.

[15] Fraser, C.S., "Network design considerations for non-topographic photogrammetry", Photogrammetric Engineering & Remote Sensing, 50(8), pp. 1115-1126, 1994.

[16] Fraser, C.S., "Digital camera self-calibration", ISPRS Journal for Photogrammetry and Remote Sensing, 52(4), pp. 149-159, 1997.

[17] Harris C., M. Stephens, "A combined corner and edge detector", Proc. 4th Alvey Vision Conf., pp. 147-151, 1998.

[18] Triggs, W., P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle Adjustment for Structure from Motion, in Vision Algorithms: Theory and Practice, Springer-Verlag, 2000.

[19] Zorin, D.N., Subdivision and Multiresolution Surface Representation. Ph.D. Thesis, Caltech, California, 1997.

[20] van den Heuvel, F.A., "3D reconstruction from a single image using geometric constraints", ISPRS Journal Photogrammetry & Remote Sensing, 53(6), pp. 354-368, 1998.

[21] Zwillinger, D. (ed.), Standard Mathematical Tables, 30th Edition, CRC Press, Inc., West Palm Beach, Florida, pp 311-316, 1996.

[22] Lee, J., R. Haralick, L. Shapiro, "Morphologic edge detection", IEEE Journal of Robotics and Automation, 3(2), pp 142-156, 1987.

[23] http://www.vit.iit.nrc.ca/elhakim/home.html.