VIDEO BASED 3D MODELING

Jussi Heikkinen
Institute of Photogrammetry and Remote Sensing
Helsinki University of Technology
Finland

Commission V, IC Working Group V/III

KEY WORDS: Video recording, Free network, Modeling

ABSTRACT

Recently more and more rapid 3D documentation and less expensive measurement methods have been required in building construction field as well as in archeological documentation projects. One considerable alternative will be the methods utilizing video recording implementation. In this paper we will show a method based on continuous video recording sequences and free network type estimation. The equipment used is a normal analogue amateur video camcoder. The lack of accuracy of image formation is substituted by special recording conditions and over-determination of the problem in question. The video recording is based on formation of circular closed image blocks, where image relations are determined with homogeneous point/feature pairs. No static reference system is required in first stage. The improvement of precision will be achieved by involving numerous video frames in modeling process. The experiments accomplished involve room reconstruction as well as archeological site documentation examples.

INTRODUCTION

In traditional imaging or camera constellation it is essential to plan the geometry of the imaging system. The achievable accuracy of a photogrammetric measuring system highly depends on the accuracy and distribution of control points or control features to be used. Also the geometry of camera constellation plays an important role in a precise measurement system. The imaging device and its accuracy and geometrical stability have to be known before planning the measurement system. These all affect on the final measurement precision and changing one component might have great affect on some other component and affect that way accuracy of the whole measurement system. The standard procedure is to adjust the measurement conditions iteratively to guarantee the accuracy required. The design of the measurement system can be divided into ZOD FOD SOD TOD levels of planning according to Fraser /1/.

Planning is a time consuming process. It is not unusual that the planning of measurements takes major part of the total time of a measuring process including imaging, measurements and computation. Also, the planning itself is expensive. However, in some applications where the precision requested is not so high and the imaging conditions are complicated, some other kind of approach for imaging design can be chosen.

An alternative way to create a coordinate system for measurements is to use circular image blocks. This is an ideal way of imaging for video type of imaging devices. The idea is to form an image block by rotating the camera mounted on the end of a rotating rod around the camera pole. By applying free network type estimation and fixing few parameters a coordinate system for measurements can be constructed.

This type of imaging design does not guarantee the ideal measuring geometry neither it avoids all the possible occlusions. Usually, you have to take multiple image blocks of this kind to satisfy the condition of reconstructing the object. The method does not provide measurements of highest quality, but what it does provide are measurements of moderate quality that are sufficient for some applications or are a good base for a model to be used in a design of more precise measurements in complex modeling conditions.

The benefit for using circular image blocks is to get the data faster and also in case of video recording the amount of data can be increased. The increased amount of data can be used for substituting the low quality of images at some level. By using low cost tape recording devices, also the sufficient storage space is not a problem.

IMAGE BLOCK FORMATION

The idea in a circular image block is the same as in a normal image block formation: you have to have at least three common image points between consecutive image frames. And those points should not lie on the same line. With a circular image block the camera is rotated around one rotation axle at the top of a supporting rod. The idea is to get a closed sequence where the features detected on the first frame can be tied up with the same 3D features seen on the last frames. The turning rod is supposed to lie on a plane, so all the projection centers of camera connected to each frame will lie on the same cylinder surface.

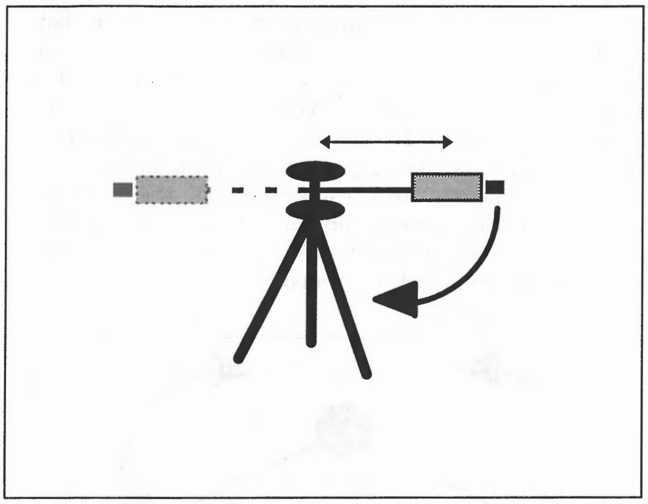

Figure 1 Circular image block formation.

Circular image block estimation

The geometric model we are using here is based on a general pinhole camera model and the interior orientation of the camera used is supposed to be a prior knowledge.

It is essential for such a circular image block estimation to realize that no control points have to be involved. So we can see the estimation as a **n+1** number of relative orientations of images or **n** number of exterior orientations without any control points. How that can be done? One solution is to use free network type estimation and restrict the solution vector with fixed values and by using constraints between orientation parameters.

The projection centers lying on the same cylinder surface, more precisely on same path of a 3D circle, can be given as:

$$|P_i| = r \tag{1}$$

or

$$\sqrt{X_i^2 + Y_i^2 + Z_i^2} = \sqrt{r^2} \tag{2}$$

This equation expresses that the projection centers are lying at the same distance **r** from the center of rotation, which is chosen as an origin for the estimation.

Along rotating the camera around the origin the rotation angles of camera are changing respect to the co-ordinate system but the angle between the image plane and the vector between projection center and the origin stays as constant. This can used as a constraint to stabilize the estimation. This yields to:

$$R_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \bullet P_i = \text{constant} \tag{3}$$

The equation (2) actually does not express that points would be lying in path of a same 3D circle, rather that they only lie on the surface of the same sphere. This can be fixed by allowing the position vector of the projection center of the first frame construct the x-axis of the co-ordinate system in a case of horizontal sequence and the rest of the projection centers are lying on the xy-plane. So all z-values of the projection centers are fixed to zero. The co-ordinates of the first camera position will be then P_1 (r,0,0) where **r** is the radius of the circle.

Estimation of image block

The model used here is based on image bundle blocks and collinearity condition. Another alternative would have been to use independent stereo models as primary computation units. It is true that block adjustment based on stereo models and coplanarity condition does not include unknown 3D object points in estimation which is the case with image bundles. But geometrically thinking those unknown points are there, and you can always eliminate the unknown 3D points out of a LSQ type estimation like /2/ :

$$\begin{aligned} & A\,l = b \\ & A = [A_1 A_2] \\ & N = A^T A = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \\ & \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ & (N_{11} + N_{21} N_{22}^{-1} N_{12}) x_1 = b_1 - N_{21} N_{22}^{-1} b_2 \end{aligned} \tag{4}$$

A normal matrix contains 5x5 elements respect to every camera orientation on the diagonal, except the first camera orientation, which results a 4x4 element on the diagonal of the normal matrix. The constrain equations follow each diagonal element as depicted in Figure 2.

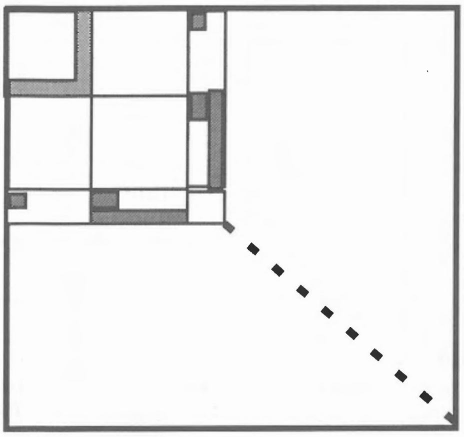

Figure 2 Normal matrix of the image block adjustment.

The system can be solved also by not using any constraints. This can be done by substituting the projection center coordinates with an angle α from x-axis.

$$X_i = r \cdot \cos\alpha_i \qquad (5)$$
$$Y_i = r \cdot \sin\alpha_i$$

Remember Z_i was previously set to zero. Also only the κ-value changes during the rotation and it is the same angle as α which was presented above. So what we get are four unknowns from the first camera position (**r**,ω,φ,κ) and one unknown κ_i for each following camera position. All together it makes **n+3** unknowns.

Modeling condition

When rotating the camera at the top of the turning rod, the geometry to make space intersection from subsequent frames will be poor, since cameras are looking divergently respect to each other. Also in a case when **r** has a small value the viewing angle respect to object point has changed only a little and definitely too little for the range to be defined accurately. The solution will be to use multiple circular sequences and make a co-ordinate transformation based on common points seen on both sequences. From these different sequences you can find the best frames for object reconstruction which also fulfill the geometry requirements.

An interesting feature is that you actually can construct quite good geometry conditions and also get the scale for your measurements for object reconstruction without any control points. This can be done from a single camera stand. Instead of having camera looking along the supporting rod outwards from the origin, we can turn the camera looking to the direction of tangent. With a such constellation a circular block can be formed and unknowns **r**,ω,φ,κ and κ_i can be estimated. After this, camera will be turned to the opposite direction and a new circular block will be formed without changing any other condition. Now we have two circular blocks that have the same origin, the same xy-plane and possibly the same radius. What we still have to resolve is the angle β between x-axes of these two co-ordinate systems. This can be done based on common points seen on both sequences. Now we have frames whose exterior orientation is resolved in the same co-ordinate system and whose optical axes are converging. In such a system the maximum base is **2r**. The scale of the measurements in such a case is strongly dependent on goodness of interior orientation and especially on the accurate estimation of the camera constant. In a real situation, this might not be sufficient and some exterior distance for scale measure might be needed.

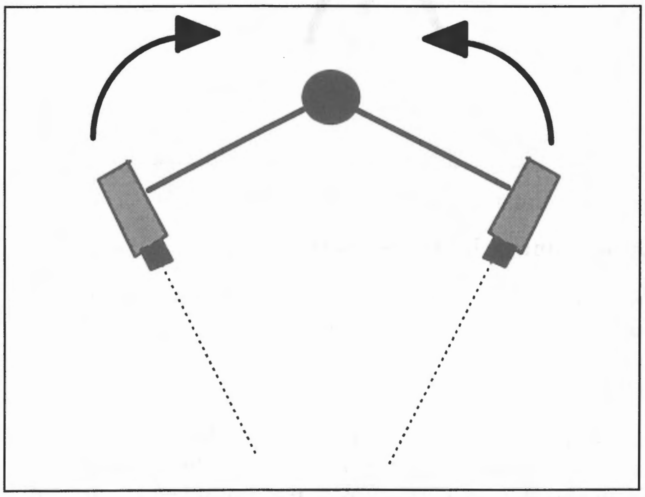

Figure 3 Camera positions from two separate sequencess with a common origin.

IMAGE MEASUREMENTS

While imaging highly three-dimensional objects, which is often the case with close-range applications, automation of image measurements is a demanding task. For template matching highly three dimensional objects are not ideal targets because the method makes assumption of flatness of the surface. As mentioned in previous chapter; while rotating the camera around the origin and when **r** is considerably small, the viewing angle respect to object point does not change much between subsequent frames. This yields that also area based template matching might work fine in such conditions. Especially with video sequences where you can digitize nearly as many frames as you want from a continuos sequence.

Our implementation was to extract the template from the first image. This was guided by the operator. After this, the cross-correlation was computed between template and a next target frame. The best position of correlation with template was detected. This gave good initial values for final LSQ-type template matching resulting in the final location with subpixel accuracy. The next step was to extract a new template from the current target frame at the location of best fit. This was then used for finding the best fit on the following frame. The procedure iterated until the point to be located was out of sight, because of occlusion or the point was out of frame border. The "traveling point"-type effect appeared if extraction was not done with subpixel accuracy.

In our first implementation we only have measured single points, but other type of features like lines and curves can also be included in estimation /3/ /4/.

VIDEO IMAGING

We have used a simple amateur type video camcoder in our first tests. The benefit of using video is that we can have a continuos sequence of frames and we can include nearly as many frames as we want in modeling to substitute the low quality of the images. Perhaps the best benefit is still nearly endless storage space of tape devices used with video recorders.

The amount of noise on video frame is considerable compared e.g. to digital camera images. The noise can be reduced by averaging the frames. Even by averaging three subsequent frames the noise will be reduced considerably. But the camera cannot be moved during the time between frames.

The problem though is the low resolution of video frames. Often the most common commercial frame grabbers are designed for grabbing TV resolution quality images, meaning better resolution SVHS data is converted to lower resolution TV standard format (PAL, NICAM etc.). Also quite many grabbers like to convert grabbed images to JPEG or MPEG format which degrades the geometric properties of images in the image measurements' standpoint of view.

APPLICATIONS

Applications of using such a method are various. The method suites best for applications where no exterior co-ordinate system is available or imaging conditions are otherwise difficult. This method has found its way in archeology, forestry and in modeling the inner space of a building. All the following examples are still on research level and the real estimates of suitability of the method for these applications can be found out in next few years.

Archeology

In archeological excavations you very seldom have any relevant co-ordinate system available. Usually you have to create your own coordinate system out of scraps. This was also the case in excavations which started in Autumn 1997 in Petra, Jordan. There is a Finnish group excavating the site assumed to have been an ancient bysantic monastery on the mountain of Jabal Harun.

The depicted method was applied for constructing the co-ordinate system for the measurements on the excavations. Altogether we took 17 circular sequences. Twelve of them were outside the site and five on the top of the formation. The surface model of the site is depicted on Figure 4. It was constructed based on other material and it is a very course presentation of the site.

What most interested the archeologists was the possibility to create multitemporal models without having to make heavy work-loaded mapping with tachymeter during the excavations.

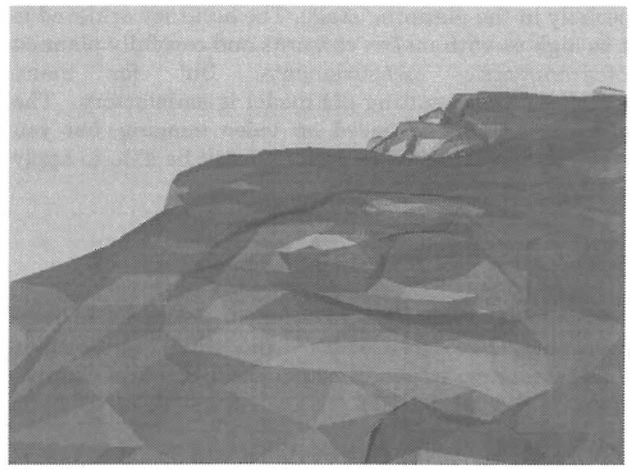

Figure 4 A course surface model of the archealogical site

Modeling a building inner space

This method was first developed for solving the problem of imaging in complex conditions. The common situation in close-range measurements is that you can pose the cameras around the object and achieve good geometry for measurements. When modeling the inner space of the building that is not possible. You have to do the imaging inside the room and try to find out the extent of the room space from there.

The practical problems with room space modeling occurred in situations where in some part of the sequence no texture or distinctive points were present. This was crucial since the method requires that subsequent frames should have enough common points to be tied up to an image block. The problem was then solved by adding artificially some texture on the scene.

While recording the image sequence the exposure and focus have to be untouched. Change of these components affects camera calibration values. It is preferable to calibrate the camera with the same settings. Some amateur camcoders have quite a fancy set of automatic features adjusting exposure, focus as well as compen-sating the random vibration during imaging. One has to be able to turn features off while doing imaging for measuring purpose. This was the case with our camera. The only difficulties occurred when imaging a room space the light coming through the windows made the lightning conditions so uneven that Venetian blinds and artificial light had to be used during imaging.

DISCUSSION

The method depicted here shows an alternative way of solving the 3D modeling problem. This method is based n closed video image sequences without any knowledge of exterior co-ordinate system. Presented approach is capable of producing the 3D model faster than with traditional photogrammetric methods. The time is saved especially in the planning stage. The accuracy achieved is not as high as with metric cameras and carefully planned photogrammetric measurements. But for many applications the resulting 3D model is satisfactory. The preliminary tests are based on video imaging, but you can use digital cameras as well and still be able to apply the same procedures.

REFERENCES

/1/ Fraser, C., 1989. Optimization of networks in non-topographic photogrammetry. In: Non-Tophographic Photogrammetry, edited Karara, ASPRS, Falls Church, Virginia U.S.A. pp. 95-106.

/2/ Mikhail, E., 1976. Observations and Least Squares. IEP-A Dun-Donnelley Publisher, New York, U.S.A., pp. 303-311.

/3/ Heikkinen, J., Object Reconstruction from Images of a Moving Camera. In: International Archives of Photogrammetry and Remote Sensing. Vol. XXXI, Part B5, XVIII ISPRS Congress 9.-19.7.1996, Vienna, p. 220-224.

/4/ Heikkinen, J., Three-Dimensional Modelling from Moving Images with Help of Linear Features.
In: SPIE Proceedings Vol. 2598 Videometrics IV, Editor(s): Sabry F. El-Hakim, p. 255 - 263.