CALIBRATION OF A PTZ SURVEILLANCE CAMERA USING 3D INDOOR MODEL

R.A. Persad*, C. Armenakis, G. Sohn

Geomatics Engineering, GeoICT Lab
Earth and Space Science and Engineering
York University
4700 Keele St., Toronto, Ontario, M3J 1P3 Canada
(ravi071, armenc, gsohn)@yorku.ca

Commission I, WG I/3

KEY WORDS: Calibration, Pose Estimation, Surveillance, PTZ Video Camera, 3D Indoor Models

ABSTRACT:

Security monitoring, event detection and tracking of people are a few of the well-known applications for indoor video surveillance systems, but mainly they are accomplished by visual inspection of the images. There is need to integrate visual surveillance systems with 3D geospatial data to semantically improve interpretation of scene dynamics which is challenge in 2D image analysis. In this study a single view video camera is calibrated with respect to a 3D wireframe model to lay the foundation for further research into people tracking within this 3D space. For the estimation of camera parameters, the proposed method caters for the motions of a typical Pan-Tilt-Zoom (PTZ) surveillance camera. The method takes into consideration the degrees of freedom of the PTZ camera where the perspective centre is considered fixed –no translation change between the image frames- and that the camera rotates about its centre. For the initial estimation of camera parameters we use point correspondence between image to model space followed by line correspondences within image space for updating the parameters after camera motion. Point correspondences are utilized in a self-calibration adjustment to project the 3D model an initial image frame. For the determination of the camera parameters of another overlapping image frame 2, the 3D model is projected into it using the camera parameters of the initial frame. Extracted lines are labelled on image 2 and a distance function is used to describe the relation of camera parameters between the two image frames based on the difference of the orthogonal distances between extracted lines and the projected model lines. Using image and model line correspondence, the distance function is minimized and the initial camera parameters are updated to retrieve those of the second image. The distance minimization approach for estimating camera parameters is tested on simulated and real datasets providing promising preliminary results.

1. INTRODUCTION

Surveillance monitoring video systems predominantly exist in the interiors of commercial, industrial and public buildings. A human operator of such a system is typically faced with the arduous and almost impossible task of looking at multiple monitors in an attempt to fully interpret the 3-dimensional (3D), real-world scene from 2-dimensional (2D) monocular video streams. To overcome this handicap, fusion of 3D knowledge content of the real world with the video media is needed.

3D geospatial data such as building information models (BIMs) and computer-aided design (CAD) models are the next best representation of real world building scenes. Traditionally, 3D models have been well-utilized in many applications such as city design planning and development. However, their potential for '3-dimensionalizing' a surveillance system are yet to be fully realized.

Existing BIM and CAD models can be acquired from repositories of the engineering or architectural firms involved in the building design. Alternatively, with the continuing improvements in architectural modelling software and data collection methods such as laser scanning, 3D model generation can be achieved with relative ease and at increasingly lower costs. Therefore, the high availability of 3D models should not present any hindrance in mass productions of 3D model-based surveillance systems.

The initial research challenge faced in design of a 3D model-based surveillance system is the integration of 2D image with 3D model data sources. This research work aims to address this issue by the calibrating the image frames from the video camera with respect to its 3D indoor representation, thereby integrating these two data types. To relate the 3D model to each image frame of the video stream, camera calibration is necessary for the retrieval of the parameters that can project 3D model space to 2D image space. The advantages of the 3D indoor model are its use as a virtual calibration object for all image frames of the video data and as a framework for accurately monitoring and tracking of objects from a 3D perspective.

The majority of the state of the art surveillance networks are currently employing pan-tilt-zoom (PTZ) cameras for people tracking and event detection. Hence, the experiments in this paper cater for these cameras by simulating PTZ motions. In our work, camera calibration is defined by a partial interior and full exterior camera parameter determination. The full external calibration is defined as the parameter determination of the three rotational pose angles (i.e. pan, tilt, roll) and translations, whereas, the partial internal calibration recovers the focal length. The latter is a function of the camera's zooming capabilities. Since PTZ units are usually mounted to a fixed position and given that our experimentation environments are indoors where camera de-stabilization factors such as wind cease to exist, we assume the translation to be rigid and known within the local coordinate frame provided by 3D model.

2. RELATED WORKS

This paper incorporates issues in research fields such as model-image integration, surveillance using 3D models and PTZ camera calibration. In this section, a summary of some previous related works will be highlighted.

Indoor structures are usually populated with linear features. It is also known a 3D model can be represented as a wireframe model. This category of 3D models can be defined as a topological organization of straight line segments that describes the scene it represents. Significant research into image and wireframe integration has been used in various applications.

3D wireframe models have been used for applications such as robotic-vehicle navigation (Kosaka and Kak, 1992). 3D wireframe model to image fitting has also found significance in vehicle detection and traffic surveillance (Wijnhoven and de With, 2006). A-priori constructive-solid geometry (CSG) model primitives have also been used for semi-automated building extraction (Tseng and Wang, 2003).

Minor strides have been also made in using 3D models for video surveillance, particularly, by augmented reality researchers. Outdoor video surveillance using global positioning system (GPS) technology for dynamic camera calibration and pose tracking via the fusion of dynamic image frames with the use of textured 3D models have been developed (Sebe et al., 2003). Unlike this, our proposed method is constrained to GPS-denied, indoor environments. We instead employ a purely vision-based photogrammetric approach for the calibration of the camera. Other works have also designed systems for tracking and visualization in 3D based on a 'smart camera' network architecture (Fleck et al. 2006). Instead, our long term goals are to effectively utilize already existing PTZ units for this exact purpose.

PTZ camera calibration has received a significant amount of attention in the computer vision community with numerous methods being developed. Fung and David, 2009 used inter-image homographies to develop a calibration approach for a rooftop PTZ camera. Huang et al., 2007 proposed a stereo method to address calibration issues associated with a long-focal-length, PTZ camera. Generally, these methods ignore the incorporation of external control from a 3D source for solving all the camera parameters, particularly for determination of interior parameters. However, the fundamental definition of camera calibration describes the all parameters that link the 2D image to the 3D world. For this reason, we directly use a 3D model, which is the next best representation of the real world, for all stages of the camera calibration, i.e. for interior and exterior parameter determination. In addition to this, the model to image mapping across multiple image frames of a video sequence is a crucial component in the future development of our 3D tracking methods.

3. METHODOLOGY

The following section describes the proposed approach of the research. Essentially, the method has two main components (Figure 1). Two data sets are acquired from an indoor scene: a video sequence that is divided into multiple image frames and a 3D wireframe model. Initially, point correspondences in first image frame and model space are interactively established .Using the well-known collinearity condition, a self-calibration adjustment is used to project the 3D wireframe model into image space via the initially computed camera parameters. Afterwards, the camera pans, tilts or zooms. Thus, the parameters with respect to image 'k' must be determined, where k = 2,....,n with n representing the image frame number. To accomplish this, the initial image frame parameters are used to project the model into image 2, where it is skewed. Using the line correspondence between manually extracted image 2 lines and the projected model lines, an objective function is used to minimize the distance between these lines, thus matching them and updating the initial camera parameters from the previous image frame to that of the current image frame. Figure 1 outlines the methodological framework.

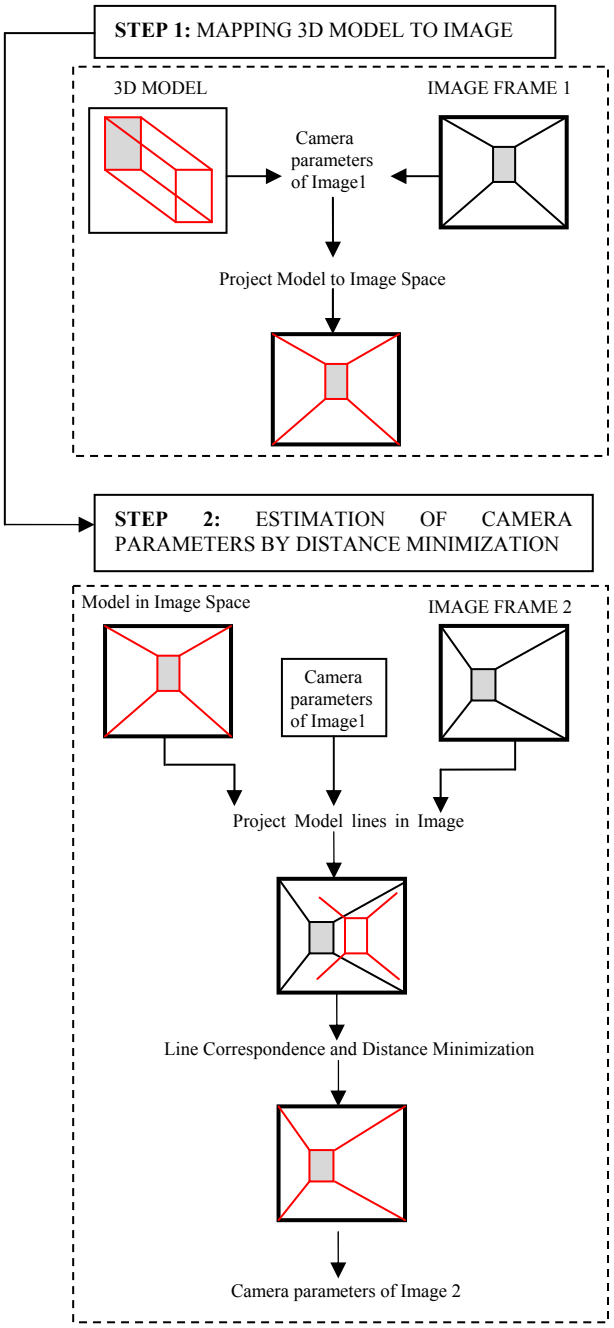


Figure 1. Conceptual outline of the approach.

3.1 Mapping 3D Wireframe Model to Image

The method of camera self-calibration or bundle adjustment with additional parameters (AP's) is used to mathematically represent the collinearity condition that exists between the image and model spaces (equation 1). In this case, the AP's are used to model errors that occur in the imaging process (Fraser, 2001). A least squares solution is applied by adopting the well-known Gauss-Markov (GM) model.

$$u = u_0 - f\frac{m_{11}(X-X^0)+m_{21}(Y-Y^0)+m_{31}(Z-Z^0)}{m_{13}(X-X^0)+m_{23}(Y-Y^0)+m_{33}(Z-Z^0)} - \Delta u$$

$$v = v_0 - f\frac{m_{12}(X-X^0)+m_{22}(Y-Y^0)+m_{32}(Z-Z^0)}{m_{13}(X-X^0)+m_{23}(Y-Y^0)+m_{33}(Z-Z^0)} - \Delta v \quad (1)$$

with

$$\Delta u = \Delta u_0 - \frac{\bar{u}}{f}\Delta f + \bar{u}r^2K_1 + \bar{u}r^4K_2 + \bar{u}r^6K_3 + P_1(2\bar{u}^2 + r^2) + 2P_2\bar{u}\bar{v} - c_1\bar{u} + c_2\bar{v}$$

$$\Delta v = \Delta v_0 - \frac{\bar{v}}{f}\Delta f + \bar{v}r^2K_1 + \bar{v}r^4K_2 + \bar{v}r^6K_3 + P_2(2\bar{v}^2 + r^2) + 2P_1\bar{u}\bar{v} + c_2\bar{u} \quad (2)$$

and

$$r = \sqrt{\bar{u}^2 + \bar{v}^2} = \sqrt{(u-u_0)^2 + (v-v_0)^2} \quad (3)$$

where u,v = image coordinates
u_0,v_0 = principal point
f = focal length
X,Y,Z = model coordinates
X^0,Y^0,Z^0 = ground coordinates of perspective center
m_{ij} = Orthogonal rotation matrix elements that contains the pan and tilt angle, as well as, the roll angle.
Δu and Δv represent the combined physical model according to Beyer, 1992
K_i = radial-symmetric lens distortion
P_i = radial asymmetric and tangential distortion
c_i = Affinity and Shear factors

Initially, the Direct Linear Transformation (DLT) equations in equation (4) (Abdel Aziz and Karara, 1971) are used to provide initial camera parameter approximations for the iterative, non-linear self-calibration process. The DLT is an algebraic solution whose 11 coefficients are afterwards decomposed into the physical camera parameters.

$$u = \frac{P_{11}X^0 + P_{12}Y^0 + P_{13}Z^0 + P_{14}}{P_{31}X^0 + P_{32}Y^0 + P_{33}Z^0 + 1}$$

$$v = \frac{P_{21}X^0 + P_{22}Y^0 + P_{23}Z^0 + P_{24}}{P_{31}X^0 + P_{32}Y^0 + P_{33}Z^0 + 1} \quad (4)$$

During the self-calibration adjustment, the iterations are terminated when the absolute relative approximate error percentage $|\varepsilon_A|$ is less than a pre-specified relative error tolerance $|\varepsilon_S|$, i.e. $|\varepsilon_A|<|\varepsilon_S|$. Afterwards, all the computed parameters are used to project the 3D wireframe model into image space.

3.2 Line Correspondence

Prior to the distance minimization and camera parameter estimation for the second image frame, a preliminary line correspondence procedure is undertaken to associate the projected model lines to their respective image 2 lines. An angle check method (equation 5) to compute line orientation was adopted for this purpose (Jaw and Perny, 2008). Wrong correspondences were interactively checked for and removed from the datasets.

$$\theta = \cos^{-1}\frac{(u_1-u_2)(u_3-u_4)+(v_1-v_2)(v_3-v_4)}{\sqrt{(u_1-u_2)^2+(v_1-v_2)^2.(u_3-u_4)^2+(v_3-v_4)^2}} \quad (5)$$

where θ = angle between 2 lines
(u_1, v_1), (u_2, v_2), = coordinates of projected model line
(u_3, v_3) and (u_4, v_4) = coordinates of image 'k' line

3.3 Minimization of Model to Image Distance

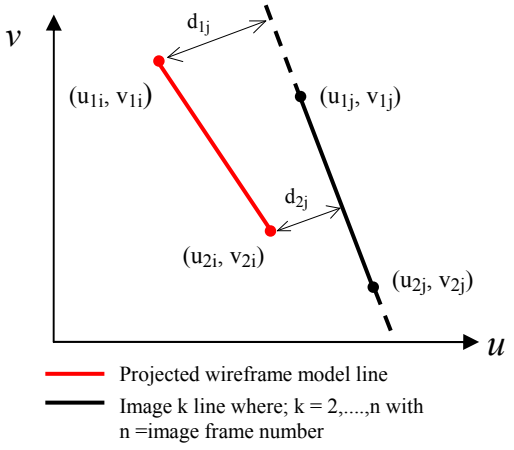


Figure 2. Residual distance d_{ij} between corresponding lines.

The camera parameters of image frame 2 are estimated by adjusting the camera parameters of image frame 1 via distance minimization between each pair of corresponding projected model and image 2 lines. The coordinates of the projected model line (u_{1i},v_{1i}) and (u_{2i},v_{2i}) are functions of image 1 camera parameters. The coordinates (u_{1j},v_{1j}) and (u_{2j},v_{2j}) are functions

of the image 2 camera parameters . Therefore, distance d_{1j} or d_{2j} is represented by a function 'G' of its respective projected model line end point coordinate and both image 2 line coordinates. Figure 2 illustrates this concept and shows distance d_{1j}, where:

$$d_{1j} = G(\{u_{1i}, v_{1i}\}, \{u_{1j}, v_{1j}, u_{2j}, v_{2j}\}) \quad (6)$$

The projected model line endpoint coordinate in equation (6) can be further expressed as a function 'H' of the camera parameters from image 1 as shown in equation (7).

$$d_{1j} = G(\{H(f_1, \omega_1, \phi_1, \kappa_1, X_1^0, Y_1^0, Z_1^0)\}, \{u_{1j}, v_{1j}, u_{2j}, v_{2j}\}) \quad (7)$$

The perpendicular distance of an endpoint from the projected model line (in this case (u_{1i}, v_{1i})) to the line on image 2 is represented as the distance function in equation (8).

$$d_{1j} = \frac{au_{1i} + bv_{1i} + c}{\sqrt{(a^2 + b^2}} \quad (8)$$

Replacing u_{1i} and v_{1i} in equation (8) with the standard collinearity equations and minimizing d_{1j} leads to a non-linear system of the form:

$$A\hat{x} = 0 + d_{1j} \quad (9)$$

where d_{1j} is treated as the residual to be minimized since it is expected that d_{1j} should be equal to zero

A = Jacobian matrix containing partial derivatives of the distance function with respect to unknown camera parameters

$\hat{x}$ = Update vector to initial parameter estimations

The linearized form of equation (9) given in equation (10) is solved rigorously by a least squares adjustment using the GM model.

$$0 = d_{1j}^0 + \frac{\partial d_{1j}}{\partial_{PARAMETER}}\Bigg|_0 \cdot \hat{x}_{PARAMETER} \quad (10)$$

where parameter = $(f_2, \omega_2, \phi_2, \kappa_2, X_2^0, Y_2^0, Z_2^0)$ of image frame 2

By minimizing d_{1j}, and also repeating this for d_{2j}, the parameters of image 1 will be adjusted accordingly to coincide with the required parameters of image 2 as seen in equation (11).

$$(f_2, \omega_2, \phi_2, \kappa_2, X_2^0, Y_2^0, Z_2^0) = (f_1, \omega_1, \phi_1, \kappa_1, X_1^0, Y_1^0, Z_1^0) + (\Delta_{f, \omega, \phi, \kappa, X^0, Y^0, Z^0}) \quad (11)$$

where $(\Delta_{f, \omega, \phi, \kappa, X^0, Y^0, Z^0})$ = change in camera parameters after least squares adjustment

4. EXPERIMENTS AND RESULTS

The method was tested both on numerically synthetic and real datasets that simulate a PTZ camera. The use of simulated data has the significant advantage of ground truth knowledge. This enables us to quantify the general robustness and validity of the method. The distance minimization method applied to both the synthetic and real datasets follows a rigorous least squares adjustment process where the focal length, rotational and translational parameters are all solved for simultaneously.

4.1 Tests with Simulated Data

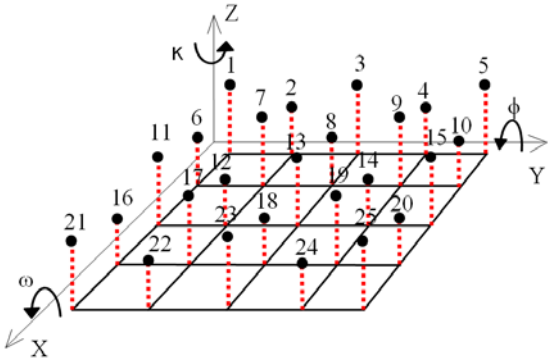


Figure 3. Geometric configuration of simulated model space.

In this section pan, tilt and focal length estimations are analyzed with respect to their determination via the self –calibration adjustment and the distance function minimization. From the simulated model space coordinates in Figure 3, synthetic image sequences containing image lines at different pan, tilt and focal length values were generated. Two frames were produced with slight changes to the focal length and rotational parameters. This was done in order to simulate these motion changes as they alter from one image frame to the next. In the first image frame parameter values were derived via the self-calibration adjustment. Using these self-calibration results as initial estimations, the parameters for the second image frame were derived using the distance function. Ten line correspondences were used in the distance minimization adjustment (Figure 4).

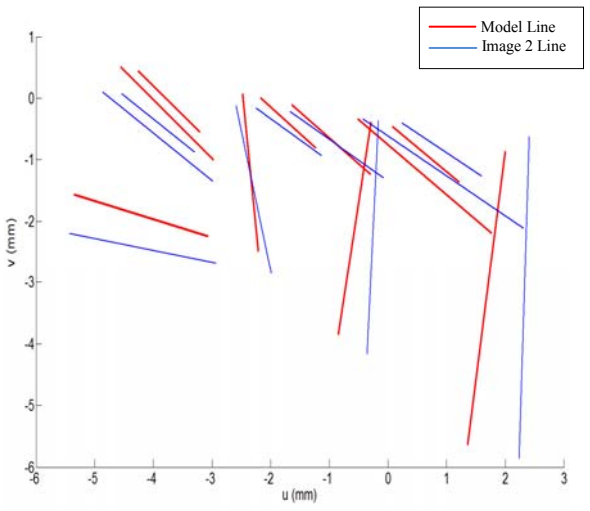


Figure 4. Simulated image and projected model lines.

Using the residuals of the distance minimization adjustment to compute the a posteriori variance factor $(\hat{\sigma}_0)$,the standard deviations (σ) of each unknown were also computed.

$$\hat{\sigma}_0 = \sqrt{\frac{V^T V}{n-m}} \qquad (12)$$

$$\sigma = \hat{\sigma}_0 \cdot \sqrt{Q_{xx}}$$

where V = observation residuals from adjustment
n = number of observation equations
m= number of parameters
Q_{xx} = a diagonal element of the covariance matrix

Table 1 show the results of the camera parameters derived by the distance minimization compared with the true, reference values that were used to produce the simulated image 2 lines.

Parameter	$\text{Value}_{\text{Reference}}$	$\text{Value}_{\text{Adjusted}} \pm (\sigma)$
f (mm)	5.5	5.49993±(1.44e-005mm)
TILT/ω (°)	4.0	4.00029±(0.0497")
PAN/ϕ (°)	10.0	9.99993±(0.0195")
ROLL/κ(°)	5.0	5.00031±(0.1297")
X^o *(m)*	105	105.00015±(0.00695cm)
Y^0*(m)*	100	100.00005±(0.01298cm)
Z^0 *(m)*	100	100.00026±(0.01108cm)

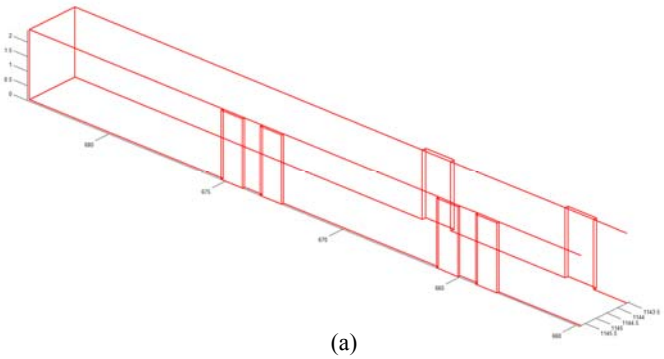
Table 1. Comparison of True vs. Distance Minimization-based Parameter values using simulated data.

4.2 Preliminary Tests with Actual Data

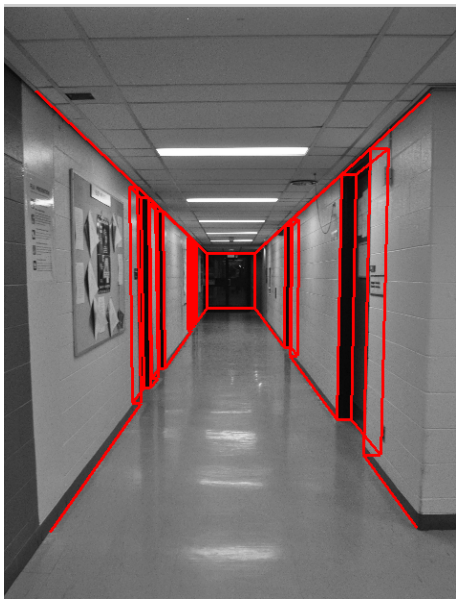
In this section , we carry out testing of the method on real image data and an actual 3D model. However, our full method has not yet been extensively tested on this real data. In particular, the automated line correspondence method was not used in this dataset. Manual correspondence was made between image and model lines. Ten corresponding lines were again used.

The test area chosen was a corridor in an office environment consisting primarily of straight line features from structures such as doors. A NIKON D90 digital camera with a 20mm Nikkor lens mounted onto a camera tripod was used for precise image acquisition, i.e. for minimal or zero shifting of the camera's centre. The camera was attached to a rotational panoramic head to simulate the rotations of an actual PTZ camera. A two-frame image sequence rotating in a pan direction, with auto-zoom (slight focal length change) was acquired. A 3D wireframe model of the indoor scene was created from an engineering plan. Figure 5 displays the data used in this section.

To verify our distance minimization approach for determining the camera parameters, we used the self-calibration adjustment approach as a means of obtaining a true source (i.e. reference value) for the second image frame parameters. Table 2 shows the results of the experimentation with real data.



(a)



(b)

(c)

Figure 5. Real Dataset. (a) 3D indoor wireframe model (b) 3D model projected to image frame 1 (c) Ten digitized lines on image frame 2.

Parameter	$Value_{Reference}$	$Value_{Adjusted} \pm (\sigma)$
f (mm)	22.18	22.10±(0.004mm)
TILT/ω (°)	-6.162	-6.134±(8.35")
PAN/ϕ (°)	-8.517	-8.668±(7.78")
ROLL/κ(°)	-0.714	-0.631±(22')
X^o *(m)*	101.155	101.116±(1.57cm)
Y^0*(m)*	101.419	101.473±(5.58cm)
Z^0 *(m)*	126.709	126.906±(11.19cm)

Table 2. Comparison of True vs. Distance Minimization-based Parameter values using actual data.

5. ANALYSIS OF RESULTS

To analyse the quality of results obtained for both the simulated and real datasets, the relative parameter error is computed using equation (13). Table 3 highlights these error values. The residual values from the distance minimization adjustment were also examined. These accuracy measures are shown in table 4.

$$\varepsilon = \frac{\left| Parameter_{\ REFERENCE} - Parameter_{\ COMPUTED} \right|}{\left| Parameter_{\ REFERENCE} \right|} \qquad (13)$$

Parameter	ε % *(Simulated)*	ε % *(Actual)*
f	0.0010	0.360
TILT/ω	0.0073	0.450
PAN/ϕ	0.0007	1.770
ROLL/κ	0.0062	11.62
X^o	0.0001	0.039
Y^0	0.0001	0.053
Z^0	0.0003	0.155

Table 3. Relative parameter error percentages.

Observations	RMSE (mm)	Mean (mm)	Std. Dev. (mm)	Max residual (mm)	Min residual (mm)
Simulated	2.0e-005	4.9e-007	2.1e-005	3.8e-005	-2.9e-005
Actual	0.028	-0.0013	0.028	0.069	-0.058

Table 4. Distance Minimization Adjustment residuals.

In table 3, the relative errors for the simulated data were below 0.01% and very small as expected. For the real data in table 4, the mean value is almost zero indicating a valid model and minimal systematic errors in the observations. Standard deviation of the residuals is in the range of 4 pixels reflecting the error of the approach. The roll angle had the largest relative error. This is presumably due to zero lines being extracted on image 2 in the direction of the X-axis as seen in figure 5(c) .Thus, it can implied that the accuracy of the computed rotations from our method relies upon the number and directional configuration of the extracted image 2 lines.

6. CONCLUSIONS

A model-based method of calibrating a PTZ camera has been presented. A distance minimization approach was used to determine camera calibration parameters of latter image frames given the parameters of an initial frame with sufficient overlap. Using simulated and real datasets, the proposed method provides encouraging preliminary results. The method is expected to contribute to the monitoring and tracking of objects within a given 3D model. Further work includes implementation of the method with a higher degree of automation and an assessment of the method with an actual PTZ camera.

ACKNOWLEDGEMENTS

The financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Geomatics for Informed Decisions (GEOIDE) Network of Centres of Excellence is greatly appreciated by the authors.

REFERENCES

Abdel-Aziz, Y. I., and Karara, H. M., 1971. Direct linear transformation into object space coordinates in close-range photogrammetry. *Proc. Symposium on Close-Range Photogrammetry*, Urbana, Illinois, pp. 1-18.

Beyer, H., 1992. Geometric and radiometric analysis of a CCD-camera based photogrammetric close-range system. Mitteilungen, *Nr. 51, Institut fur Geodasie and Photogrammetrie*, ETH Zurich.

Fleck, S., Busch, F. and Biber, P., 2006. 3D Surveillance a Distributed Network of Smart Cameras for Real-Time Tracking and its Visualization in 3D, *IEEE Computer Vision and Pattern Recognition, Workshop on Embedded Computer Vision*, New York, pp: 118.

Fraser, C.S., 2001. Photogrammetric Camera Component Calibration: A Review of Analytical Techniques", in *Calibration and Orientation of Cameras in Computer Vision, Springer Series in Information Sciences* 34, Eds.: Grün, Huang, pp: 95-121.

Fung, N. and David, P., 2009. Implementation of Efficient Pan-Tilt-Zoom Camera Calibration. *Technical Report: ARL-TR-4799*, U.S. Army Research Laboratory.

Huang, X., Gao, J. and Yang, R. , 2007 Calibrating pan-tilt cameras with telephoto lenses. *In ACCV*, pp: 127–137.

Jaw, J.J., and Perny, N.H., 2008. Line Feature Correspondence between Object Space and Image Space, *Photogrammetric Engineering & Remote Sensing*, 74(12), pp: 1521–1528.

Kosaka, A. and Kak, A.C., 1992. Fast Vision-Guided Mobile Robot Navigation Using Model-Based Reasoning and Prediction of Uncertainties, *Computer Vision, Graphics, and Image Processing— Image Understanding*, 56(3), pp: 271-329.

Sebe, I.O., Hu, J. ,You, S. and Neumann, U, 2003. 3D video surveillance with Augmented Virtual Environments, *First ACM SIGMM international workshop on Video surveillance*, November 02-08, 2003, Berkeley, California.

Tseng, Y.H. and Wang, S., 2003. Semiautomated Building Extraction Based on CSG Model-Image Fitting. *Photogrammetric Engineering & Remote Sensing*, 69(2), pp: 171-180.

Wijnhoven, R., de With, P, 2006. 3D Wire-frame Object-Modeling Experiments for Video Surveillance. *In Proc. of 27th Symposium on Information Theory*, Benelux, pp: 101–108.