RELATIVE ORIENTATION ESTIMATION OF VIDEO STREAMS FROM A SINGLE PAN-TILT-ZOOM CAMERA

Taeyoon Lee [a,] *, Taejung Kim [a], Gunho Sohn [b], James Elder [a]

[a] Department of Geoinformatic Engineering, Inha University, 253 Yonghyun-dong, Nam-gu Incheon, Korea - etaeyoon@inha.edu, tezid@inha.ac.kr
[b] GeoICT Lab, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada – gsohn@yorku.ca
[b] Centre for Vision Research, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada – jelder@yorku.ca

Commission I, WG I/5

KEY WORDS: Single Pan-Tilt-Zoom Camera Calibration, Collinearity Condition, Orientation, Self-calibration

ABSTRACT:

As a surveillance camera for the security of urban area, a single pan-tilt-zoom (PTZ) camera has been used. For effective application of the video sequence of the single PTZ camera, accurate orientation or calibration of the video sequence is needed. This paper proposes a mathematical model based on the collinearity condition equation for relative orientation estimation between frames from video sequence of a single PTZ camera. The basic concept of this mathematical model is similar to the homography concept, which explains the relationship between two images. However the proposed model is based on the collinearity equation and the model can estimate the varying focal length and relative rotation angles directly from corresponding points between image frames. For the relative orientation estimation by the model, the ratio between focal length and CCD size of the first frame is required as initial input. In this paper, we estimated the ratio by two vanishing points. The corresponding points were extracted by Scale Invariant Feature Transform (SIFT) and RANdom SAmple Consensus (RANSAC) automatically. Photos taken by digital camera and a video sequence of a single PTZ camera were used for experiments. The experiment results show the proposed method can estimate the relative orientation from the video sequence automatically.

1. INTRODUCTION

In modern society, cameras have been used for the security in many places. Usually they are fixed in location with pan, tile and zoom capability. For the effective application of a video stream data of surveillance cameras and for extraction of the accurate position information from the video sequence, accurate calibration or orientation of the camera is an essential process.

Some calibration methods for the single PTZ camera have been proposed. Many of the proposed methods are based on the concept of the image of the absolute conic (IAC) and the concept of homography between two views (Agapito et al., 2000; Kim and Hong, 2000; Fung and David, 2009). These methods (Agapito et al., 2000; Fung and David, 2009) extracted internal parameters from IAC matrix estimated using some constraints such as the zero-skew, the square-pixel and distortion. They estimated external parameters by internal parameters extracted. Kim and Hong (2000) proposed the method which calculated rotation angles and then varying focal length sequentially. We think relative internal parameters extracted from previous methods may include errors as these methods minimize errors for IAC parameters, not for internal parameters of the camera.

In this paper, we propose a mathematical model and automatic method for relative orientation estimation between successive frames from a video sequence of a single PTZ camera. In this paper, relative orientation means rotation angles (pan and tile angles) and varying focal lengths between successive frames. The proposed model uses the concept which is similar to the homography between two views, but the model is in form of the collinearity condition equation, so it can directly estimate relative orientation parameters from the video sequence of a single PTZ camera.

2. PROPOSED METHOD

2.1 Model based on the Collinearity Equation

Proposed mathematical model explains the relationship between two images. Collinearity condition can be express as E.q. (1), The

$$\begin{pmatrix} x - x_0 \\ y - y_0 \\ -f \end{pmatrix} = \lambda R \begin{pmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{pmatrix} \tag{1}$$

where f is the focal length, (x, y) is the image coordinates, (x_0, y_0) is the coordinates of the principle point, (X_0, Y_0, Z_0) is the coordinates of projection centre in the object coordinate system, (X, Y, Z) is the object coordinates in the object coordinate system, R is rotation matrix and λ is a scale factor. E.q. (1) can simply express as $\vec{x} = \lambda R(\vec{X} - \vec{X}_0)$.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

In the case of a stationary camera, we can assume that the projection centre of the camera is at the origin of the object coordinate system, so $\vec{X}_0$ is $(0,0,0)^T$.

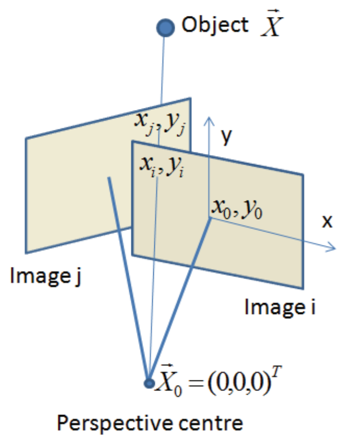

Figure 1. The geometric relationship between a single PTZ camera

Hence, if there are two images (i and j) and an object point $\vec{X}$ is projected on the two images (Figure 1), the image i and j can be express as

$$\begin{aligned} \vec{x}_i &= \lambda_i R_i \vec{X} \\ \vec{x}_j &= \lambda_j R_j \vec{X} \end{aligned} \quad (2)$$

Eliminating $\vec{X}$ from E.q. (2) we can obtain the relationship between the image i and j:

$$\vec{x}_i = \frac{\lambda_i}{\lambda_j} R_i R_j^T \vec{x}_j \quad (3)$$

In relative relationship between the image i and j, if the image i is the first image and j is next image, the all rotation angles of the image i are 0°. Hence E.q (3) is $\vec{x}_i = \mu R_j^T \vec{x}_j$, where μ is λ_i / λ_j. This relationship can be derived into the form of the collinearity equations by cancelling out the scale factor μ:

$$\begin{aligned} x_j - x_{0j} &= -f_j \frac{r_{11}(x_i - x_{0i}) + r_{21}(y_i - y_{0i}) + r_{31}(-f_i)}{r_{13}(x_i - x_{0i}) + r_{23}(y_i - y_{0i}) + r_{33}(-f_i)} \\ y_j - y_{0j} &= -f_j \frac{r_{12}(x_i - x_{0i}) + r_{22}(y_i - y_{0i}) + r_{32}(-f_i)}{r_{13}(x_i - x_{0i}) + r_{23}(y_i - y_{0i}) + r_{33}(-f_i)} \end{aligned} \quad (4)$$

where r_{ij} are elements of R_j. The relative orientation of next image j can be estimated by E.q. (4) and non-linear least square estimation technique.

2.2 Initial Focal Length and CCD Cell Size

For estimation of the relative orientation by E.q. (4), a ratio (f/CCS) between an initial focal length (focal length of the first image) and CCD Cell Size (CCS) is needed. For the ratio estimation, the method using two vanishing points of a single image was used (Guillou et al., 2000). In this method, the vanishing points are estimated by least square estimation and coordinates of lines on the image. For estimation of the vanishing points, we assumed that the slops (m_i) in N line equations (E.q. (5)), estimated by coordinates of lines, had error and the vanishing point (x_v, y_v) were estimated by E.q. (5) and non-linear least square estimation.

$$\begin{aligned} y_v &= x_v m_i + n_i \\ m_i &= \frac{(y_v - n_i)}{x_v} \quad \text{i = 1, …, N (x)} \end{aligned} \quad (5)$$

The coordinates of the lines were extracted by both hough transform and selecting manually.

2.3 Automatic Estimation of the Relative Orientation

After estimation of the ratio (f/CCS) between focal length and PSC in the first image, the proposed method begins to estimate the relative orientation between continuous two frames of a video sequence automatically. In this paper, the successive two frame images are extracted by OpenCV 2.0 library (OpenCV2.0). From the two frame images, accurate corresponding points are extracted by SIFT (Lowe, 2004) and RANSAC technique (Fischler and Bolles, 1981). Then the relative orientation is estimated by E.q. (4) and the corresponding points. In this paper, for SIFT processing, we used an open source by Ross Hess (Hess). Figure 2 shows corresponding points which is extracted by SIFT and RANSAC.

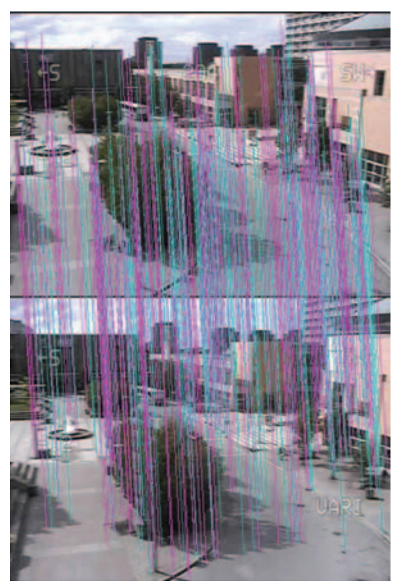

Figure 2. The corresponding points by SIFT and RANSAC

3. EXPERIMENTS

3.1 Dataset used

For experiments, photos taken by digital camera and a video sequence from a security camera were used. Table 1 shows the spec of the digital camera (Photonotes.org; Canon).

Table 1. The spec of digital camera used

North American name	EOS Digital Rebel XTi
Maximum imaging output dimensions	3888 x 2592
Physical sensor size	22.2 x 14.8 mm
Pixel size on sensor chip (PSC)	5.7 μm square
Lens	EF-S 18-55 mm

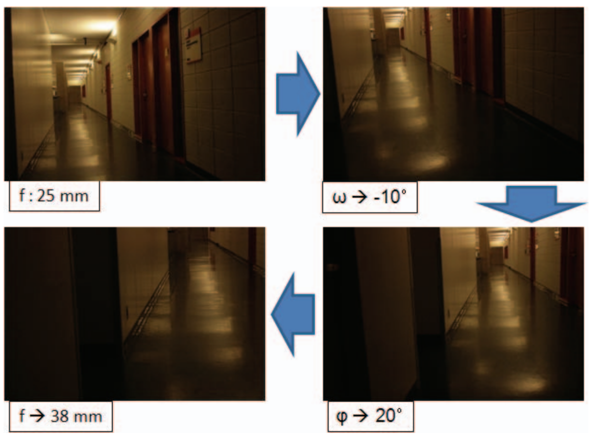

Figure 3. The photos taken by digital camera and the parameters of the relative orientation measured manually

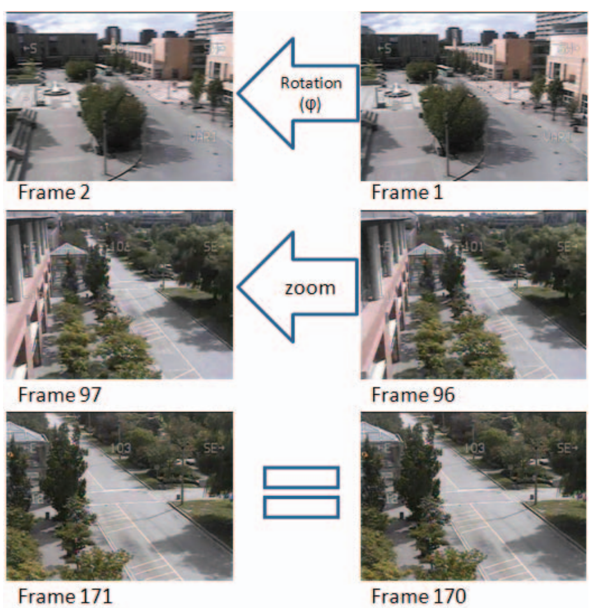

Figure 4. Frame images extracted from the video sequence

Figure 3 shows the photos and reference values. These photos were taken in each case of pan, tile and zooming. We measured the pan and tile angles manually, so the angles may be included measurement errors, but we think the angles can be used as reference angles. This camera records the focal length to Exchangeable Image File Format (EXIF) when a photo is taken. In this experiment, we used the focal length recorded to EXIF as reference. The size of the photos (Figure 3) is 3888 x 2592 pixels.

Figure 4 shows some frame images which were extracted from the security camera. The video sequence was obtained from York University. We could not know internal parameters of the security camera. Therefore, we had to estimate f/CCS by vanishing point analysis.

3.2 Results

Table 2 shows the f/CCSs which were estimated by two vanishing points and the f/CCSs which were calculated by reference values. These results show the method by two vanishing points from a single image can estimate the f/CCS, but the f/CCS may have some errors. This paper focused on the relative orientation estimation, so we ignored small errors and used the f/CCS estimated by two vanishing points as initial value.

Table 2. f/CCS from reference and f/CCS estimated by two vanishing points

Num	f/CCS calculated from reference	f/CCS estimated by two vanishing points
1	4386.0 (f: 25 mm)	4380.93
2	3157.9 (f: 18 mm)	3560.2

Table 3 shows the parameters of the relative orientation which are estimated by the proposed method. In Table 3, "ω → -10°" in "Parameter changed" means that the image is rotated from the first image through 10 degrees clockwise on x-axis in Figure 1. "φ → 20°" means the image is rotated through 20 degrees on y-axis. Case 1 of Table 3 shows the result which is estimated from photos of Figure 3.

Table 3. The results of the relative orientation estimation by the proposed method and the photos

Parameter changed	ω → -10°	φ → 20°	Zooming (Reference)	ω-φ-f changed
Case 1: f/CCS estimated from the first image : 4380.93				
f/CCS	4349.5	4458.3	6518.5 (6666.7)	6587.7 (6666.7)
ω (°)	-11.1	-0.0	-0.0	-11.2
φ (°)	0.1	19.7	-0.0	19.8
Case 2: f/CCS estimated from the first image : 3560.2				
f	3586.7	3515.3	5286.3 (4736.8)	5257.3 (4736.8)
ω (°)	-9.9	-0.0	0.0	-10.2
φ (°)	-0.0	17.8	0.0	17.6

As shown in Table 3, the f/CCSs estimated were different with reference, but the ratio of f/CCS estimated by the proposed method between the images, when camera is zoomed, was similar to the ratio calculated using reference. In "Case 2", there was 3° difference between the φ (17°) estimated and reference φ (20°). The difference was caused by inaccurate initial f/CCS (Table 2).

In the experiments, the radial distortion was not considered. The radial distortion is well known as one error source (Agapito et al., 2000). Even though the radial distortion was not removed from the images, relative f/CCS and rotation angles estimated by the proposed method were generally accurate.

Table 4 shows the parameters of the relative orientation estimated by the proposed method from the video sequence of the security camera. In the experiment, reference values for the video sequence were not known. This paper does not show accuracy of relative orientation of the video sequence by the proposed method. We guessed relative orientation estimated may be accurate relatively when we consider the results using the images of digital camera (Table 3).

Table 4. The results of the relative orientation estimation by the proposed method and the video sequence

	Frame 1-2	Frame 96-97	Frame 170-171
Initial f	700.4	977.0	1871.7
f/CCS	701.1	1014.2	1871.7
ω (°)	0.0	0.2	-0.0
φ (°)	2.2	-0.2	-0.0
RMSE (pix.)	1.6	0.5	0.1

In Table 4, “Frame 1-2, Frame 96-97, Frame 170-171” mean the two frame number (Figure 4). “Initial f“ of “Frame 1-2” is f/CCS estimated by two vanishing points. The initial f of rest frames were estimated from previous frame images by the proposed method. When we observed the changes of zoom and rotation angles between successive frames by eyes, we could check which the changes were similar to the relative orientation estimated by the proposed method (Table 4).

In this paper, the accuracy of the model established is expressed by Root Mean Square Error (RMSE). RMSEs were calculated using corresponding points, which were not used for the model establishment, as check points. Through the experiment results, we checked that there was some correlation between the rotation angle (φ) estimated and an accuracy of the proposed model (Figure 5). As shown in Figure 5, the more rotation angle (φ) is increased, the more RMSE is increased.

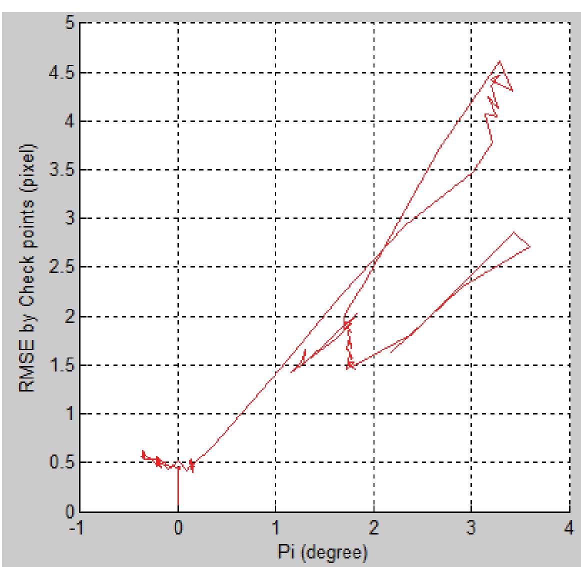

Figure 5. The relationship between the model accuracy and rotation angle (Pi)

Figure 6 shows the change of the focal length estimated and of the RMSE by time (stream of the frames). The change of the focal length was similar to zooming of the video sequence. In Figure 6, when the focal length estimated changed around 800, RMSE was increased. The reason for this is the rotation angle (φ). The angle (φ) is increased between frames of the graph which the focal length changed around 800.

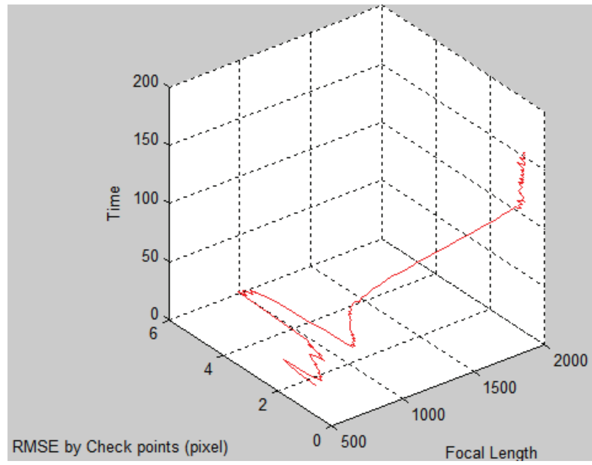

Figure 6. The change of focal length and RMSE by time, the frame order of the video sequence

4. CONCLUSION

This paper proposed a mathematical model based on the collinearity condition equation for relative orientation estimation between frames from the video sequence of a single PTZ camera. This paper also proposed automatic relative orientation of the video sequence based on the proposed model. By the proposed model, we tried to estimate accurate f/CCS and rotation angles directly from the relationship between two frame images. The proposed method can estimate the relative orientation using only the video sequence. For this, initial f/CCS is estimated by two vanishing by non-linear estimation. The experiment results show that the accuracy of the initial f/CCS influences the accuracy of the relative orientation by the proposed method. For automatic processing, OpenCV library, SIFT and RANSAC techniques are used. The results, which are estimated from the photos taken by digital camera, show that the proposed method can estimate accurate parameters without correction of the radial distortion. The results, which are estimated from the video sequence of the security camera, show that the correlation between the rotation angle and RMSE exits. Though these experiment results, we checked the proposed method could estimate the relative orientation from the video sequence of a single PTZ camera.

For obtaining the more accurate information, the proposed method needs to include estimation of the accurate initial f/CCS, correction of the radial distortion, consideration of the correlation between the rotation angles and the modelling error or focal length changed. This paper presented automatic processing for the relative orientation by well known techniques, but this method was slow. For efficient application of security camera, real time processing may be needed.

ACKOWLEDGEMENTS

The work in this paper was supported by a grant (07KLSGC03) from Cutting-edge Urban Development-Korean Land Spatialization Research Project funded by Ministry of Land, Transport and Maritime Affairs of Korea, by the supporting project to educate GIS experts funded by Ministry of Land, Transport and Maritime Affairs of Korea, and by the project 'Three-dimensionalizing surveillance network" funded by GEOIDE of Canada.

REFERENCES

Agapito, L., Hayman, E., and Reid, I., 2001. Self-calibration of rotating and zooming cameras. *International Journal of Computer Vision*, 45(2), pp. 107-127.

Fischler, M. A., and Bolles, R. C., 1981. Random sample consensus – A paradigm for model-fitting with applications to image-analysis and automated cartography. *Communications of the ACM,* 24(6), pp. 381-395.

Guillou, E., Meneveaux, D., Maisel, E., and Bouatouch, K., 2000. Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. *The Visual Computer*, 16, pp. 396-410.

Kim, H., and Hong, K. S., 2000. A practical self-calibration method of rotating and zooming cameras. *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 1, pp. 354-357.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), pp. 91-110.

Canon, http://www.usa.canon.com (accessed 2010)

Fung, N., and David, P., 2009. Implementation of efficient pan-tilt-zoom camera calibration. Army Research Laboratory. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA497476&Location=U2&doc=GetTRDoc.pdf (accessed 2010)

Hess, R., http://web.engr.oregonstate.edu/~hess/ (accessed 2010)

OpenCV2.0, http://sourceforge.net/projects/opencvlibrary/ (accessed 2010)

Photonotes.org, http://photonotes.org (accessed 2010)