Robust Camera Calibration Tool for Video Surveillance Camera in Urban Environment

Sung Chun Lee and Ram Nevatia
Institute for Robotics and Intelligent Systems,
University of Southern California
Los Angeles, California 90089, USA
SungChun.Lee|nevatia@usc.edu

Abstract

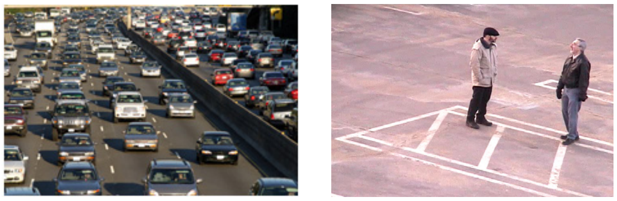
Video surveillance applications such as smart room and security system are prevailing nowadays. Camera calibration information (e.g. camera position, orientation, and focal length) is very useful for various surveillance systems because it can provide scene knowledge and limit search space for object detection or tracking. In this paper, we describe a camera calibration tool that does not require any calibration object or specific geometric objects by using vanishing points. In urban environment, vanishing points are easily obtainable since there exist many parallel lines such as street lines, light poles, buildings, etc in either outdoor or indoor scene images. Experimental results from various surveillance cameras are presented.

1. Introduction

Video surveillance camera have been used for various applications such as traffic monitoring, security system, post incident analysis, etc. Originally these video surveillance systems were designed for human operator to watch concurrently or to record video data as archive for later analysis. As the amount of cameras is significantly increasing and the quantity of the archived video data becomes unmanageable by human operator, intelligent video surveillance systems have been introduced. Recently, computer vision researches have been heavily involved in intelligent video surveillance applications.

Many intelligent video surveillance researches mainly focus on pedestrians or vehicles detection. Camera calibration information (focal length, its position and orientation) is very useful to reduce search space as well as false alarms for object detection process. However, due to its initial purpose of the video surveillance system, many already-installed cameras and the archived video data are not calibrated.

Camera parameters can be obtained by using standard methods if a calibration object or measurements of enough 3D points in the scene are available [13], [6]. Such measurements, however, are not always available and it is not applicable for the already-recorded video data. Recent approaches use scene geometry information (e.g. two non-planar surfaces in [7] or floor plan information in [11], [2]) or human detection and tracking information [9], [8] to estimate the camera calibration. However, a major limitation of these methods is that these special geometric objects or humans may not be seen in some video surveillance camera environments. For instance, there is no human walking around in the traffic monitoring video cameras as shown in Figure 1 (a) and due to small field of views because the camera usually looks down, 3D structural scene may not be visible as shown in Figure 1 (b). In this paper, we present a robust camera calibration tool to apply it to static video surveillance cameras or even the already archived video data.



(a) Traffic monitoring example (b) Small field of view example

Figure 1. Examples of video surveillance data.

Vanishing points of parallel lines have proven to be useful features for this task [1], [3], [4]. In urban environment, vanishing points are easily obtainable since there exist many parallel lines such as street lines, light poles, buildings, etc in either outdoor or indoor scene images. Caprile and Torre described a method to use vanishing points to recover intrinsic parameters [1]. Criminisi et al. have developed a method to estimate intrinsic parameters using Cholesky decomposition [4] and to reconstruct scene geometry by using projective geometry methods; they do not explicitly compute the extrinsic parameters. Cipolla et al present a method [3] to compute both intrinsic and extrinsic parameters by using three orthogonal vanishing points and one reference point.

We describe an approach similar to [3] and [4] for estimating both intrinsic and extrinsic parameters from three orthogonal vanishing points and an object of known height. However, it is difficult to directly use these methods for our application as in an urban site, due to a limited field of view, often only two vanishing points may be visible. In the case that two orthogonal vanishing points are found, a third orthogonal vanishing point can then be recovered by an approximate solution with some assumptions.

Section 2 explains how to extract vanishing points followed by camera calibration method with the extracted vanishing points in the section 3. Some results and evaluations are discussed in section 4.

2. Vanishing Point Extraction

Vanishing points can be extracted from a group of parallel lines in an image, which can be obtained by an automatic approach or given by user. In many cases of video surveillance cameras, it is not easy to extract vanishing points automatically. In addition, most video surveillance cameras are installed in one location permanently which requires the camera calibration process only once. It suggests that a user interactive vanishing point extraction method should be required in our system to be used robustly.

2.1. User interactive vanishing points extraction

In many cases, it is hard to extract vanishing points automatically as shown in Figure 2. It is harder to find a set of orthogonal vanishing points to be used for the camera calibration without user's interaction. Our system provides an interactive tool for the user to draw lines to get vanishing points in any cases. The user draws a set of parallel lines and the system computes the intersection (i.e. vanishing point) of the drawn lines by using a least square error method (SVD). Some examples of user interaction of extracting vanishing points are shown in Figure 2. Because the vanishing points are commonly located at the very far away from the image center, we display the lines (red lines) formed by the user drawn lines (yellow lines) and the extracted vanishing point.

2.2. Missing vanishing point inference

Three orthogonal vanishing points are required to compute the rotation of the camera. However, in some cases, there is the case of only two orthogonal vanishing points to be extracted even interactively.

A third orthogonal vanishing point can then be recovered if we are given the principal point (or assume it to be in the center of the image) by using the geometry shown in Figure 3. We assume the principal point, V_0 to be the center of image, which is the orthocenter of the triangle formed from three orthogonal vanishing points, then new lines l_{n1} and l_{n2} can be drawn as depicted in Figure 3. The intersection point of l_{n1} and l_{n2} is the recovered orthogonal vanishing point V_n.

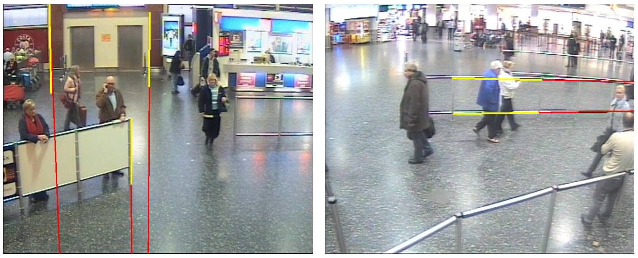


Figure 2. Interactive vanishing points extraction (yellow lines: user's drawings, red lines: new lines formed by the computed vanishing point.

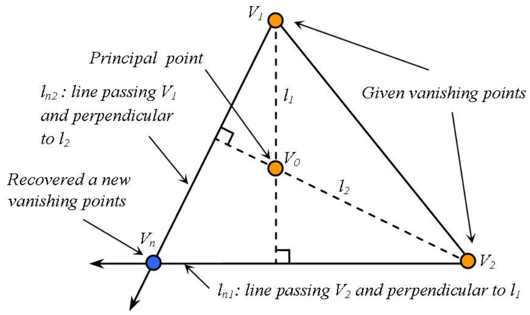

V_1
Principal point
l_{n2} : line passing V_1 and perpendicular to l_2
l_1
Given vanishing points
V_0
l_2
Recovered a new vanishing points
V_n
V_2
l_{n1}: line passing V_2 and perpendicular to l_1

Figure 3. Inferring a new orthogonal vanishing point using two known orthogonal vanishing points.

3. Camera Calibration

In this section, we describe how to estimate the position and orientation of the camera from vanishing points. We divide the camera calibration problem into two components: rotation estimation and position estimation. In our approach, we assume the aspect ratio of the camera to be known as 1.0 and the skewness to be zero.

3.1. Estimation of camera rotation

Vanishing point is the projected image point of the intersection of 3D parallel lines (i.e point at infinity) When two 3D line groups are orthogonal, two vanishing points are defined to be orthogonal vanishing points. When three mutually orthogonal vanishing points are given, the camera orientation and some camera internal parameters such as focal length and principal points can be recovered [1], [3]. Since the three vanishing points are the image projection of the points at infinity of the 3D world coordinate vectors, the following equation can be derived:

$$(V_x \ V_y \ V_z) = P(I_x \ I_y \ I_z) \tag{1}$$

where $I_x = (1, \ 0, \ 0, \ 0)^T, I_y = (0, \ 1, \ 0, \ 0)^T, I_z = (0, \ 0, \ 1, \ 0)^T$ are the world coordinate vectors, and V_x, V_y, V_z are 3×1 scaled vectors of vanishing points, and P is a 3D to 2D projective transformation matrix (3×4).

Cipolla et al. [3] derive the following equation to get the external rotation matrix, R_{wc} using three orthogonal vanishing points:

$$R_{wc} = \begin{pmatrix} \frac{\lambda_1(u_1-u_0)}{f} & \frac{\lambda_2(u_2-u_0)}{f} & \frac{\lambda_3(u_3-u_0)}{f} \\ \frac{\lambda_1(v_1-v_0)}{f} & \frac{\lambda_2(v_2-v_0)}{f} & \frac{\lambda_3(v_3-v_0)}{f} \\ \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix} \tag{2}$$

where $\lambda_1, \lambda_2, \lambda_3$ are scaling factors, $(u_0, \quad v_0)$ is the principal point, f is the focal length, and $(u_1, \ v_1), (u_2, \ v_2), (u_3, \ v_3)$ are x, y, and z directional vanishing points respectively under no skew and the known aspect ratio (= 1.0) assumptions.

The unknown principal point $(u_0, \ v_0)$ can be obtained as the ortho-center of the triangle formed by three orthogonal vanishing points [1], [3]. We had to compute the other unknowns, $\lambda_1, \lambda_2, \lambda_3$ and the focal length f to get the external rotation R_{wc}.

Since each column of the rotation matrix R_{wc} is orthogonal to the other columns, the following equations can be generated to compute the focal length f by inner products of three pairs of two columns:

$$\begin{aligned} \lambda_1\lambda_2(\tfrac{(\vec{x}_1-\vec{x}_0)\cdot(\vec{x}_2-\vec{x}_0)}{f^2} + 1) = 0 \\ \lambda_1\lambda_3(\tfrac{(\vec{x}_1-\vec{x}_0)\cdot(\vec{x}_3-\vec{x}_0)}{f^2} + 1) = 0 \\ \lambda_2\lambda_3(\tfrac{(\vec{x}_2-\vec{x}_0)\cdot(\vec{x}_3-\vec{x}_0)}{f^2} + 1) = 0 \end{aligned} \tag{3}$$

where $\vec{x}_i = (u_i, \ v_i)^T$.

Then, we can derive a new equation to obtain ${\lambda_1}^2$ as following [3]:

$${\lambda_1}^2 \doteq \frac{|(\vec{x}_2 - \vec{x}_3)||(\vec{x}_0 - \vec{x}_3)| \sin b}{|(\vec{x}_2 - \vec{x}_3)||(\vec{x}_1 - \vec{x}_3)| \sin a} \doteq \frac{\text{area of} \triangle \ \vec{x}_0\vec{x}_2\vec{x}_3}{\text{area of} \triangle \ \vec{x}_1\vec{x}_2\vec{x}_3} \tag{4}$$

where $\vec{x}_i = (u_i, \ v_i)^T$, a is an angle between two vectors $(\vec{x}_2 - \vec{x}_3)$ and $(\vec{x}_1 - \vec{x}_3)$, b is an angle between two vectors $(\vec{x}_2 - \vec{x}_3)$ and $(\vec{x}_0 - \vec{x}_3)$, and $\times$ is a cross product of vectors.

When we interpret Equation (4), both numerator and denominator imply the areas of the triangles formed by $\vec{x}_0, \vec{x}_2, \vec{x}_3$ and $\vec{x}_1, \vec{x}_2, \vec{x}_3$ respectively. We can obtain ${\lambda_2}^2$ and ${\lambda_3}^2$ by computing areas of the other triangles.

3.2. Estimation of the camera position

Given the external rotation matrix and a 3D to 2D point correspondence, we create a vector passing through the center of projection, which is the location of the camera. With the world to camera rotational matrix R_{wc}, the known principal points $(u_0, \ v_0)$, and focal length f from Equation (2), we compute a 3D directional vector $C_{di}(x_{di}, \ \ y_{di}, \ \ z_{di})$ from the 2D image point p_i (= $(x_i, \ \ y_i)$):

$$C_{di} = \begin{pmatrix} x_{di} \\ y_{di} \\ z_{di} \end{pmatrix} = R_{cw} \begin{pmatrix} x_i - u_0 \\ y_i - v_0 \\ f \end{pmatrix} \tag{5}$$

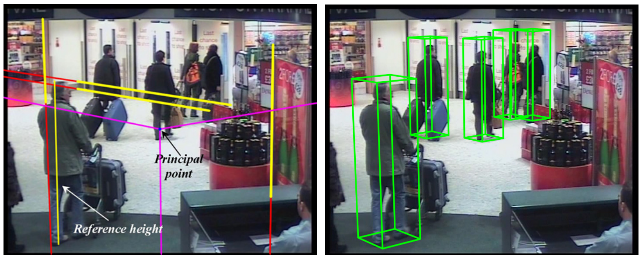
where R_{cw} is a camera to world rotational matrix (= ${R_{cw}}^T$).

We form a 3D line vector with the 3D directional vector C_{di} in Equation (5) and the 3D point correspondence provided, which passes through the center of projection. The 3D position of the camera must be on the intersection of more than one of these vectors. With two or more than two 3D to 2D point correspondences, the position of the camera can be obtained by intersecting the 3D line vectors.

In our system, the user selects any size-known object such as human, street line, or light pole from the scene by drawing a image line segment and provides the metric size of the chosen object.

4. Results and Discussion

We have tested our calibration method with some archived surveillance videos (TrecVID 2008) [12]. Figure 4 shows an example of how to extract vanishing points and to estimate camera calibration parameters. Because all three orthogonal vanishing lines are not locatable in this scene, the user draws only two sets of parallel lines (yellow lines - one in vertical and the other in horizontal directions) as shown in Figure 4. We estimate all three orthogonal vanishing points using the method described in Section 2.2. Figure 4 (a) shows the vanishing lines connecting the user drawn lines and the extracted vanishing points (red lines). In addition, the user indicates two point correspondences between the image line (e.g. yellow line of a man at bottom) and its world metric (e.g. man's approximate height) to compute the camera location.



(a) (b)

Figure 4. Results of camera calibration.

In Figure 4 (b) illustrates the result of the estimated camera calibration. We create a 3D box model that is approximate human size (50cm x 50cm x 180cm). When the user clicks a point at the bottom of the human at image, we project the 3D human model to the image as shown in Figure 4 (b). We can qualitatively validate the camera calibration result using the projected 3D human models. It may not be accurate as the other calibration method, but it is efficiently enough to aid the performance of object detection and tracking algorithms.

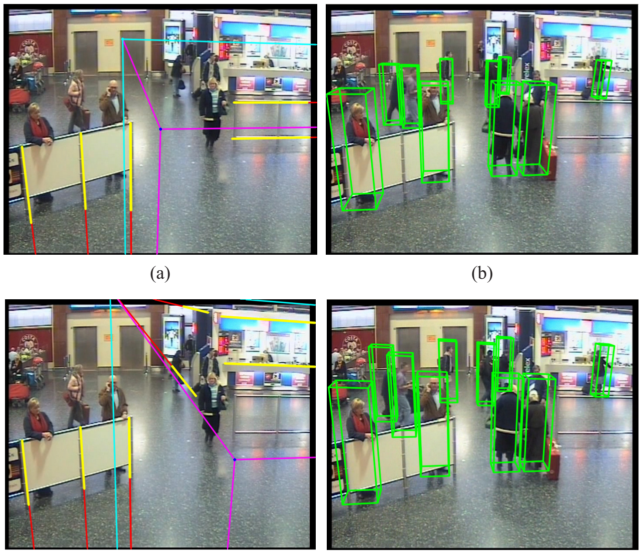
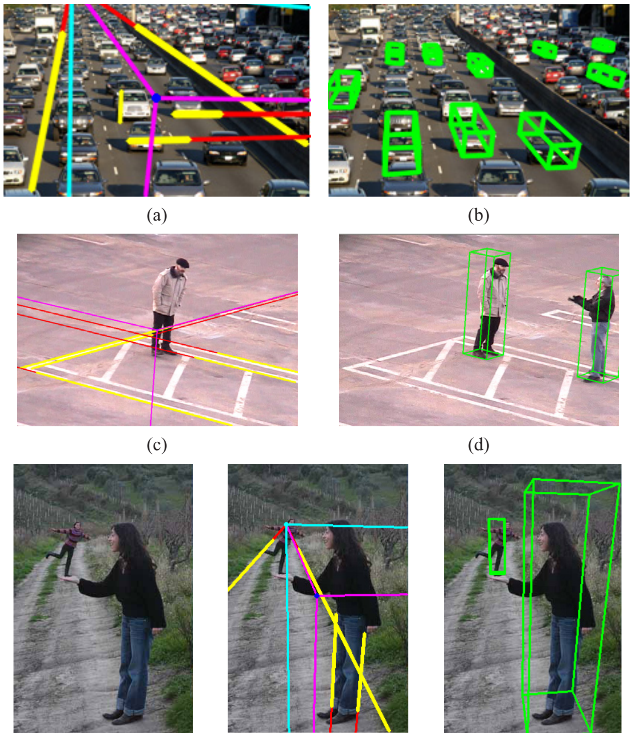

(a) (b) (c) (d)

Figure 5. Comparison between the proposed camera calibration method using two and three orthogonal vanishing points extractions: (a),(c) - Two and three vanishing points extractions, (b),(d) - Projected human models.

Figure 5 shows the comparison between two cases with the same scene: two and three orthogonal vanishing points. Human projections near the camera in both cases are similar in terms of size and orientation as depicted in Figure 5 (b) and (d). Due to radial distortion, however there are errors in human projections near the horizontal line of Figure 5 (b). However, the error range is not significant

Figure 6 (a) and (b) show some examples of traffic monitoring cameras. This time, we use a 3D box model for vehicle with the dimension of (480cm x 180cm x 160cm) to project and display. Figure 6 (c) - (e) shows some difficult cases of camera calibration. Figure 6 (c) is the case of no visible vertical vanishing lines and we can calibrate this video camera using two available vanishing lines as shown in Figure 6 (d). Figure 6 (e) is more challenging case, which is not from video surveillance camera. We found this photo from the Internet and test our calibration method to this picture. As shown in Figure 6 (e), the size variation is too big between two people in this picture. Due to camera calibration information, we can verify that the person in far distance is still valid size of human. It suggests that our camera calibration method should help to reduce 'false alarm' and 'missed detection' as well as computation time in object detection and tracking.


(a) (b) (c) (d)

(e) Image illusion picture case (non-surveillance camera)

Figure 6. Result of camera calibration in difficult cases.

4.1. Quantitative evaluation

Regarding accuracy evaluation, we tested our calibration method with "PETS (Performance Evaluation of Tracking and Surveillance) 2009" datasets [10]. They provide calibration data for 8 video cameras whose parameters were obtained using the Tsai method [13]. The world ground plane is assumed to be the Z=0 plane in both methods. Because the world coordinate systems between both methods are different, i.e. the world origin position and the definition of three (x, y, z) basis axises are different, it is not easy to compare the camera parameters directly. Instead, we measure the reprojection errors in image plane to evaluate the accuracy. In order to discard the radial distortion effect in the evaluation, we randomly choose the evaluation points near the center of the image.

First we define the center point of image as (c_{xi}, c_{yi}) and its back-projected world point on the world ground as $(C_{xw}, C_{yw}, 0)$ using the ground truth camera calibration information (Tsai method) and the camera software provided by the courtesy of project ETISEO [5]. Then, we randomly generate 100 world points, (E_{xw}, E_{yw}, E_{zw}) as following:

$$E_{xw} = C_{xw} + d_{xw}; E_{yw} = C_{yw} + d_{yw}; E_{zw} = d_{zw} \quad (6)$$

where d_{xw} and d_{yw} are random range between -2.5 meter and 2.5 meter and d_{zw} is random range between 0 meter and 2 meter.

We reproject these 100 points into the image plane as (e_{xi}, e_{yi}) to be the ground truth points. Now, we generate the test image points produced by our calibration data. Since we use the different world coordinate system, we have to conduct some transformations to get the same reprojecting image points. The detail description of how to generate the test image points is followed.

We define the ground projecting points (z=0) of the selected world points in Equation 6 as $(E_{xw}, E_{yw}, 0)$ and reproject them to the image plane using our camera calibration data and get (e'_{xi}, e'_{yi}). Then, we generate a temporary set of the world points in our world coordinate system by back-projecting (e'_{xi}, e'_{yi}) to the world ground plane as $(E'_{xw}, E'_{yw}, 0)$. A new set of world points in our world coordinate system can be defined as:

$$S_{xw} = E'_{xw}; S_{yw} = E'_{yw}; S_{zw} = d_{zw} \tag{7}$$

where d_{zw} is the same range value used in Equation 6.

Note that the world points, (S_{xw}, S_{yw}, S_{zw}) and (E_{xw}, E_{yw}, E_{zw}) are physically the same points even though the coordinate values are different due to the different world coordinate systems. Their reprojection points are supposed to be the same in the image coordinate. We produce the test image points, (s_{xi}, s_{yi}) by reprojecting (S_{xw}, S_{yw}, S_{zw}) to the image plane using our calibration data. We estimate the reprojection errors by measuring the image distance between (e_{xi}, e_{yi}) and (s_{xi}, s_{yi}) as following:

$$\frac{1}{N}\sum_{i=1}^{N}\sqrt{(e_{xi} - s_{xi})^2 + (e_{yi} - s_{yi})^2} \tag{8}$$

where N is the number of evaluating points (= 100 in this evaluation).

Table 1. Reprojection errors of 8 Cameras.

Camera No	Average distance (Pixels)
Camera 1	3.36
Camera 2	1.64
Camera 3	2.48
Camera 4	1.06
Camera 5	3.52
Camera 6	4.37
Camera 7	8.33
Camera 8	4.64

The average reprojection error per each camera is shown in Table 1. Because Cameras 1 through 4 are mid-range distance, the most of the evaluating world points that are located within the radius of 2.5 meter from the center point, $(C_{xw}, C_{yw}, 0)$ are projected near the center of the image as shown in Figure 7. Their distance error range is approximately 1 to 4 pixels as indicated in Table 1. In case of Camera 5 through 8, they are near-range distance and their distance errors become bigger ranging from 4 to 10 pixels. In case of camera 7, there are some inaccurate ground truth points in left part of the image that causes large distance errors as shown in Figure 7 (g).

Figure 7 displays the reprojected points (red circle), the ground truth points (yellow circle), and the distance (blue line) between these points per each image. Please note that there are the common points (e'_{xi}, e'_{yi}) (magenta circle) reprojected from the ground plane by using both calibration methods. White lines are generated by using our calibration data while cyan lines are drawn by Tsai method.

5. Conclusion

We presented a camera calibration tool that does not require any calibration object or specific geometric objects. Instead, we use any scene objects that have parallel lines such as street lines, light poles, buildings which are commonly found in urban video surveillance cameras. We implemented a simple and efficient interactive vanishing points extracting tool because in many cases of video surveillance cameras, it is hard to extract vanishing points automatically. We derived a method to infer the third orthogonal vanishing points using two vanishing ones to estimate camera calibration. We have showed the efficiency of our calibration methods through many experiments. Future work will be focus on handling the case that there is no vanishing points extracted, e.g. affine camera. However, this is not common in video surveillance cameras because they are installed at near-range distance with 30 to 60 degree tilt angle that causes the perspective projection not affine transform.

References

[1] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–139, 1990. 64, 65, 66

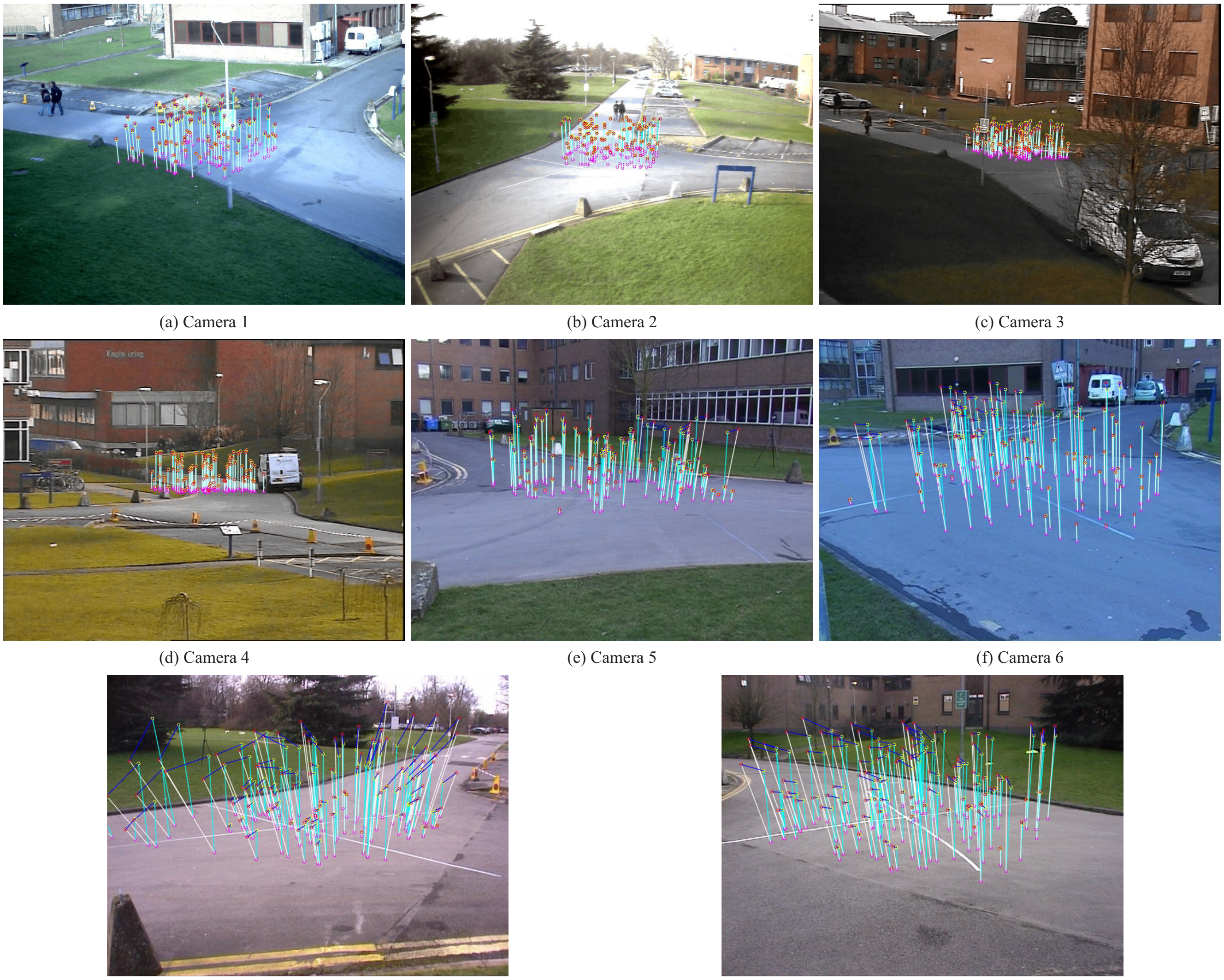
[2] T.-J. Cham, A. Ciptadi, W.-C. Tan, M.-T. Pham, and L.-T. Chia. Estimating camera pose from a single urban ground-view omnidirectional image and a 2d building outline map. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 64

[3] R. Cipolla, T. Drummond, and D. Robertson. Camera calibration from vanishing points in images of architectural scenes. *In Proceedings of British Machine Vision Conference*, 2:382–391, 1999. 64, 65, 66

[4] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000. 64, 65

[5] ETISEO. Etiseo. *http://www-sop.inria.fr/orion/ETISEO/*, 2006. 67

[6] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. *Cambridge University Press*, 2000. 64


(a) Camera 1

(b) Camera 2

(c) Camera 3

(d) Camera 4

(e) Camera 5

(f) Camera 6

(g) Camera 7

(h) Camera 8

Figure 7. Camera calibration comparison results: reprojection errors estimated by the distance (blue line) between the ground truth points (yellow circles) and the test image points (red circles).

[7] J.-H. Kim. Linear stratified approach for 3d modelling and calibration using full geometric constraints. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2144–2151, 2009. 64

[8] F. Lv, T. Zhao, and R. Nevatia. Camera calibration from video of a walking human. *Pattern Analysis and Machine Intelligence*, 28(9):1513–1518, 2006. 64

[9] B. Micusik and T. Pajdla. Simultaneous surveillance camera calibration and foot-head homology estimation from human detections. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 64

[10] PETS. Performance evaluation of tracking and surveillance (pets). *http://www.cvg.rdg.ac.uk/PETS2009/*, 2009. 67

[11] I. Rishabh and R. Jain. Interactive semantic camera coverage determination using 3d floorplans *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 64

[12] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, 2006. 66

[13] R. Tsai. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987. 64, 67