

IMAGE-BASED PAN-TILT CAMERA CONTROL IN A MULTI-CAMERA SURVEILLANCE ENVIRONMENT

Ser-Nam Lim, Ahmed Elgammal and Larry S. Davis

University of Maryland, College Park
Computer Vision Laboratory, UMIACS
{sernam,Elgammal,lsd}@umiacs.umd.edu

ABSTRACT

In automated surveillance systems with multiple cameras, the system must be able to position the cameras accurately. Each camera must be able to pan-tilt such that an object detected in the scene is in an vantage position in the camera's image plane and subsequently capture images of that object. Typically, camera calibration is required. We propose an approach that uses only image-based information. Each camera is assigned a pan-tilt zero-position. Position of an object detected in one camera is related to the other cameras by homographies between the zero-positions while different pan-tilt positions of the same camera are related in the form of projective rotations. We then derive that the trajectories in the image plane corresponding to these projective rotations are approximately circular for pan and linear for tilt. The camera control technique is subsequently tested in a working prototype.

1. INTRODUCTION

In a surveillance environment with multiple cameras monitoring a scene, the first task is to position the cameras at the location of a detected object. [1, 2] provides details about projective rotations, which is defined to be the homography that corresponds to pure Euclidean rotation. In [3], it was proposed that a ground-plane coordinate system could be setup by merely watching the objects entering and leaving the scene and then recovering the image-plane to local-ground-plane transformation of each camera. [4] described using the color model of objects to integrate information from multiple cameras. [5] uses a Markov Chain Monte Carlo approach for object identification between multiple cameras. [6] uses the relationship between the FOVs of different cameras to uniquely label detected objects. Others [7, 8] uses the more traditional approach of calibrating the cameras by providing different approaches to determine the intrinsic and extrinsic parameters.

In this paper, a camera with wide-angle lenses viewing the monitored scene is designated as the master camera [9]. The master camera performs all detection and tracking while a set of available pan-tilt-zoom (PTZ) cameras will be assigned to zoom in and get clear views (with good resolution) of objects detected by the master camera. Each camera is assigned a pan-tilt zero-position, ϕ_0 and ψ_0 [2]. Homographies between these zero-positions are pre-computed. This provides only the calibration between different cameras in their zero-positions. Different pan-tilt positions of the same camera are related to each other by projective rotations. We show that the projective rotations conjugated to pan and tilt are approximately circular and linear trajectories in the image plane respectively. As a result, cameras can be positioned accu-

rately by first mapping the image coordinates of a detected object in the master camera's zero-position to the assigned camera's zero-position using the pre-computed homographies. The mapped image coordinates is in turn mapped to the current position of the assigned camera by using the projective rotations relating different pan-tilt positions of the same camera. This approach is much easier than previous work since it utilizes only information available in the images and experimental results show it to be highly accurate and reliable.

2. CAMERA CONTROL TECHNIQUE

In this paper, we propose to model the trajectories of image points caused by projective rotations because that requires only image-based information. In particular, we will now describe a camera control technique that positions the cameras by panning first follow by tilting the cameras, based on projective rotations.

2.1. Relation Between Image Trajectory and Projective Rotations

Fig. 1 shows a top view of the camera geometry. Let the change in pan angle be ψ_c starting from any initial pan position and the tilt angle be ϕ , where the tilt angle is defined as the angle between the rotation axis and the ray joining the pan-tilt joint and the camera's center of projection. Given a point p with world coordinate

$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$, we start with the camera at $\psi_c = 0$ and some tilt ϕ . We

have the world coordinate system axes orientated as in Fig. 1 and r and f as the radius of rotation and focal length respectively. At the current tilt ϕ , the center of projection is

$$C = \begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix} = \begin{bmatrix} 0 \\ -r \cos \phi \\ r \sin \phi \end{bmatrix} \quad (1)$$

If we change the pan angle by panning with respect to the world coordinate system's y-axis, we get a rotation matrix R , for a right-handed system that gives C_{new}

$$R = \begin{bmatrix} \cos \psi_c & 0 & -\sin \psi_c & 0 \\ 0 & 1 & 0 & 0 \\ \sin \psi_c & 0 & \cos \psi_c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$C_{new} = \begin{bmatrix} -r \sin \phi \sin \psi_c \\ -r \cos \phi \\ r \sin \phi \cos \psi_c \end{bmatrix} \quad (3)$$

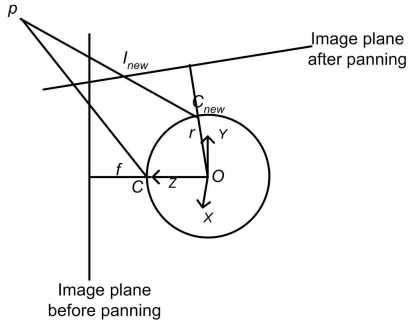


Fig. 1. Top view - camera geometry

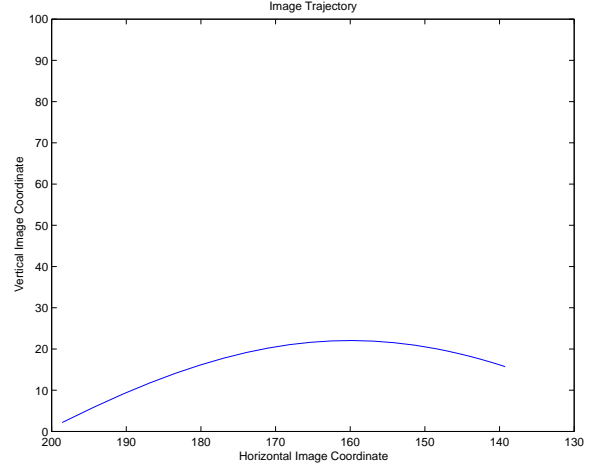


Fig. 2. Image trajectory

The coordinate of p in the camera coordinate system thus changed by first applying translation opposite to the initial world coordinate of C and then R in the reverse direction

$$\begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \\ 1 \end{bmatrix} = \begin{bmatrix} X \cos \psi_c + Z \sin \psi_c \\ Y + r \cos \phi \\ -X \sin \psi_c + Z \cos \psi_c - r \sin \phi \\ 1 \end{bmatrix} \quad (4)$$

Let the camera-to-image mapping M_{CI} be

$$M_{CI} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

where (u_0, v_0) is the principle point, $\alpha_u = fk_u$, $\alpha_v = -fk_v$, $k_u =$ horizontal inter-pixel distance and $k_v =$ vertical inter-pixel distance. We map Eqn. 4 from 3D to 2D

$$\begin{bmatrix} x_{cam} \\ y_{cam} \\ f \end{bmatrix} = \begin{bmatrix} f \frac{X \cos \psi_c + Z \sin \psi_c}{-X \sin \psi_c + Z \cos \psi_c - r \sin \phi} \\ f \frac{Y + r \cos \phi}{-X \sin \psi_c + Z \cos \psi_c - r \sin \phi} \\ f \end{bmatrix} \quad (6)$$

followed by applying Eqn. 5 to get the image coordinate of p

$$I_{new} = \begin{bmatrix} I_x \\ I_y \end{bmatrix} = \begin{bmatrix} \alpha_u \frac{X \cos \psi_c + Z \sin \psi_c}{-X \sin \psi_c + Z \cos \psi_c - r \sin \phi} + u_0 \\ \alpha_v \frac{Y + r \cos \phi}{-X \sin \psi_c + Z \cos \psi_c - r \sin \phi} + v_0 \end{bmatrix} \quad (7)$$

2.2. Projective Rotations Conjugated to Pure Camera Pan

Eqn. 7 can be used to derive the relation between I_x and I_y

$$I_y = \frac{\alpha_v(Y + r \cos \phi)}{\alpha_u \sqrt{X^2 + Z^2} \cos(\psi_c - \tan^{-1} \frac{Z}{X})} (I_x - u_0) + v_0 \quad (8)$$

Since we are considering pure camera pan, ϕ is a constant. Hence as the pan angle changes, I_y and I_x change with respect to each other in the form of a $\frac{1}{\cos}$ curve. Note that if the world point is behind the image plane, I_x and I_y becomes undefined, so we are only concerned with the part of the curve that corresponds to trajectory while the world point is in front of the image plane. To model this characteristic of the projective rotation for pure camera

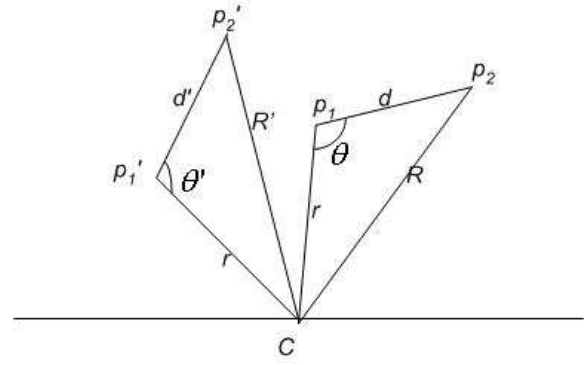


Fig. 3. Trajectories for panning have approximately same circle center

pan, we propose using a circular trajectory that best fits the part of the curve when the world point is in front of the image plane. This is because the term v_0 in Eqn. 8 would flatten out the curve largely since we would expect $|v_0|$ to be comparatively large, making it close to a circular curve. Fig. 2 shows the partial plot of the trajectory of an image point while it is in front of the image plane. Note that the image coordinate system used here has x-value increasing leftward due to the projective matrix.

Different image points have different centers of trajectory. However, we know that r is very small compared to the coordinates of p in Fig. 1. In Fig. 3, this means that $d \approx d'$ and $\theta \approx \theta'$ even as the camera is panned, where p_1 and p_2 are the image points of two different points in the scene, p_1' and p_2' are the corresponding image points after the camera panned, r is now the radius of trajectory for p_1 and p_1' and C is the center of trajectory for p_1 and p_1' . Therefore, $R \approx R'$ i.e. different trajectories for panning have the same circle center.

2.3. Projective Rotations Conjugated to Pure Camera Tilt

If the pan angle is fixed while we change the tilt angle, it is clear that the horizontal image coordinate remains constant if r is small and therefore $r \sin \phi \ll -X \sin \psi_c + Z \cos \psi_c$. Using Eqn. 7, we have

$$I_x \approx \alpha_u \frac{X \cos \psi_c + Z \sin \psi_c}{-X \sin \psi_c + Z \cos \psi_c} + u_0 \quad (9)$$

Similarly, for the vertical image coordinate, the term $-r \sin \phi$ in the denominator can be eliminated. However, we cannot do so in the numerator since $\alpha_v r \cos \phi$ cannot be ignored because we expect α_v to be large

$$I_y \approx \alpha_v \frac{Y + r \cos \phi}{-X \sin \psi_c + Z \cos \psi_c} + v_0 \quad (10)$$

Eqn. 10 thus tells us that given a constant-sized tilt step δ , the change in I_y remains constant. This characteristic is useful for controlling the tilt position of a camera.

3. IMPLEMENTATIONS

An adaptive background subtraction algorithm, details being given in [10] is used for detection. The background subtraction results in foreground regions (blobs). These blobs are tracked by establishing correspondences between blobs in each new frame in a way similar to [11]. A recursive Kalman filter is used to track and predict the image location of the blobs in the master camera image coordinate system, which is passed to the assigned camera for positioning. For camera assignment, each camera besides the master camera uses a binary semaphore to indicate whether it is in a busy state. The first available camera will be assigned. To find the center of circular trajectory for panning, we collect n image coordinates of the same point as the camera is panned, where $n \geq 3$ and then run the algorithm in [12] to get the center of the circular trajectory that these points lie on. Each camera is also fitted with a zoom function that gives the magnification factor. This zoom function is used to compute the amount of zooming in required in order to acquire images of certain resolution.

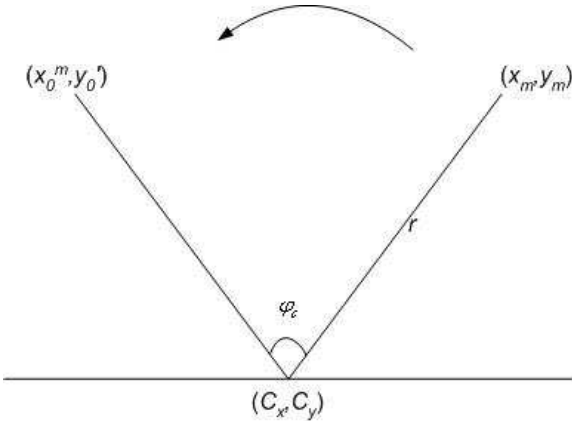


Fig. 4. Moving camera between different pan-tilt positions

The camera control technique starts by computing homographies between zero-positions of the cameras. We simply use four

pairs of corresponding points between two cameras' zero-positions and compute the homographies. Despite this, the accuracy of the prototype is good as given in Sec. 4 because most of the tracked objects have heights close to the ground plane so that the homographies work reasonably. We will first compute the center of image trajectories for the master camera, (C_x, C_y) . With all cameras starting in their respective zero-positions, when the master camera detects an object and predicts its future image coordinate to be (x_m, y_m) , given its pan as ψ_m and tilt as ϕ_m , it will first have to convert (x_m, y_m) to (x_0^m, y_0^m) , the corresponding image coordinate in its zero-position. This can be achieved by first panning with respect to (C_x, C_y) . Since we know the pan angle for the zero-position, ψ_c is known. In addition, we know the radius r of trajectory is the Euclidean distance between (C_x, C_y) and (x_m, y_m) . Referring to Fig. 4, we have

$$r = \sqrt{(C_x - x_m)^2 + (C_y - y_m)^2} \quad (11)$$

$$x_0^m = r \cos(180 - \psi_c - \cos^{-1} \frac{|C_x - x_m|}{r}) + C_x \quad (12)$$

$$y_0^m = C_y - r \sin(180 - \psi_c - \cos^{-1} \frac{|C_x - x_m|}{r}) \quad (13)$$

where y_0^m is not yet the vertical image coordinate for the zero-position, but since we know the change in I_y is constant with a constant-sized tilt step from Sec. 2.3, the change in tilt angle to get to the zero-position can be used to compute y_0^m . If the assigned camera is i , then the image coordinate (x_0^i, y_0^i) in the zero-position of camera i is

$$I_i = H \begin{bmatrix} x_0^m \\ y_0^m \\ 1 \end{bmatrix} \quad (14)$$

where $I_i = \begin{bmatrix} x_0^i \\ y_0^i \\ 1 \end{bmatrix}$ and H is the homography between zero-positions of the master camera and camera i . Using Eqn. 11 to Eqn. 13 should give us the image coordinate in the current position of camera i . To move camera i so that the detected object is in a vantage position in camera i 's image plane, such as the center of the image, we can let x_0^m be the image center. We can then derive y_0^m in Eqn. 13 and compute the change in tilt so that the vertical image coordinate is in the center of the image.

4. RESULTS

We randomly sample a number of different image points in the master camera and move the assigned camera so that the image points would be in some desired position. Euclidean differences between the resulting and desired positions are then recorded. Our prototype has three PTZ cameras, one of which is designated as the master camera. They are mounted at high positions and have common field of view. Fig. 5 shows that the error of our camera control method is very small, where the maximum error is less than 6 pixels. Fig. 6 shows images of the prototype in action, demonstrating a high accuracy in detection, camera control and camera zoom. Video clips can be downloaded from <http://www.cs.umd.edu/~sernam/icme03/video/>. For the first four sets, the first video clip in each set shows the master camera performing detection and tracking while the second and third video clips show camera 1 and camera 2 respectively moving to position themselves in vantage positions, in respond to the objects tracked by the

master camera in the first video clip. The first four sets of video clips only move the assigned cameras to vantage positions without zooming in, to demonstrate the accuracy of the camera control method. The next set of video clips shows the assigned cameras moving to vantage positions and zooming in. The last three sets of video clips use a different camera as the master camera.

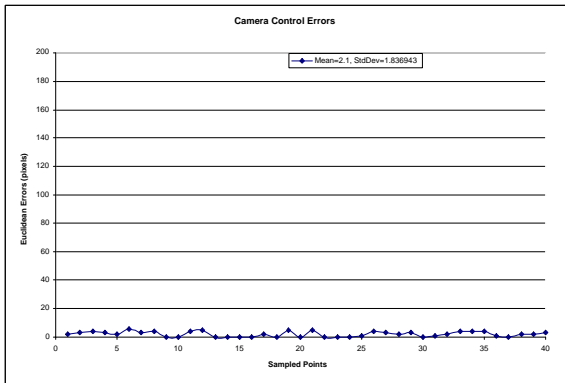


Fig. 5. Error in camera control method is small

5. CONCLUSIONS

In this paper, we have presented an image-based pan-tilt camera control technique by deriving the underlying characteristics of the projective rotations corresponding to camera pan-tilt. This image-based model greatly simplifies the task of camera control, since no intrinsic and extrinsic parameters of the cameras are required; instead we only need easily attainable information about the cameras. This contribution would facilitate future development of a multi-camera surveillance system. Effectiveness and reliability of the technique have also been tested rigorously in a real-time prototype system.

6. REFERENCES

- [1] Paul A. Beardsley Andrew Zisserman and Ian D. Reid, "Metric calibration of a stereo rig," in *IEEE Workshop on Representation of Visual Scenes, Cambridge, Massachusetts*, Jun 1995.
- [2] Andreas Ruf and Radu Horaud, "Projective rotations applied to a pan-tilt stereo head," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*, 1999, pp. 144–150.
- [3] Graeme A. Jones, John-Paul Renno, and P. Remagnino, "Auto-calibration in multiple-camera surveillance environments," in *Proc. of IEEE International Workshop on Performance Evaluation and Surveillance*, 2002.
- [4] P. Remagnino J. Orwell and G.A. Jones, "Multi-camera color tracking," in *Second IEEE Workshop on Visual Surveillance, Fort Collins, Colorado*, Jun 1999.
- [5] Michael Ostland Hanna Pasula, Stuart Russell and Yašacov Ritov, "Tracking many objects with many sensors," in *Proceedings of IJCAI-99, Stockholm*, 1999.



Fig. 6. Two sets of images of camera action (a) Master camera detects (b) Assigned camera positions (c) Assigned camera zooms

- [6] Zeeshan Rasheed Omar Javed, Sohaib Khan and Mubarak Shah, "Camera handoff: Tracking in multiple uncalibrated stationary cameras," in *IEEE Workshop on Human Motion, Austin, Texas*, Dec 2000.
- [7] P.M. Ngan abd R.J. Valkenburg, "Calibrating a pan-tilt camera head," in *Image and Vision Computing Workshop, New Zealand*, 1995.
- [8] Robert T. Collins and Yanghai Tsing, "Calibration of an outdoor active camera system," in *IEEE Computer Vision and Pattern Recognition, Fort Collins, Colorado*, Jun 1999.
- [9] Kenneth Dawson-Howe, "Active surveillance using dynamic background subtraction," 1996.
- [10] Ahmed Elgammal, David Harwood, and Larry S. Davis, "Nonparametric background model for background subtraction," in *Proc. of 6th European Conference of Computer Vision*, 2000.
- [11] Ismail Haritaoglu, David Harwood, and Larry S. Davis, "W4:who? when? where? what? a real time system for detecting and tracking people," in *International Conference on Face and Gesture Recognition*, 1998.
- [12] G. Taubin, "Estimation of planar curves, surfaces and non-planar space curves defined by implicit equations with applications to edge and range image segmentation," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1115–1138, 1991.