Calibrating pan-tilt cameras in wide-area surveillance networks

James Davis†
Honda Research Institute
Mountain View CA USA
jedavis@ieee.org

Xing Chen†
NVIDIA Corporation
Santa Clara CA USA
xcchen@ieee.org

Abstract

Pan-tilt cameras are often used as components of wide-area surveillance systems. It is necessary to calibrate these cameras in relation to one another in order to obtain a consistent representation of the entire space. Existing methods for calibrating pan-tilt cameras have assumed an idealized model of camera mechanics. In addition, most methods have been calibrated using only a small range of camera motion.

This paper presents a method for calibrating pan-tilt cameras that introduces a more complete model of camera motion. Pan and tilt rotations are modeled as occurring around arbitrary axes in space. In addition, the wide area surveillance system itself is used to build a large virtual calibration object, resulting in better calibration than would be possible with a single small calibration target. Finally, the proposed enhancements are validated experimentally, with comparisons showing the improvement provided over more traditional methods.

1. Introduction

Observing and capturing the motion of people in wide areas is useful in a number of applications including surveillance. The network of many cameras needed to observe a wide area is often augmented by using a few pan-tilt cameras that can be actuated to follow targets over time. These pan-tilt cameras can be narrowly focused so that they provide high-resolution imagery of the region that they observe. This paper describes a method for calibrating the pan-tilt cameras in such a network.

Existing methods of calibrating pan-tilt cameras have either assumed expensive precision equipment that simplifies the camera model, or relatively small working volumes that simplify the calibration procedure.

By using a tracked object to build a large virtual calibration target, this work extends calibration to wide area environments. Further a more complete model of pan-tilt cameras is employed, making the calibration suitable for use with low cost pan-tilt mechanisms. Finally, the method is validated and compared against previous calibration models.

2. Related Work

Most existing methods for calibrating a pan-tilt camera assume a rather simplistic geometric model of motion, in which axes of rotation are orthogonal and aligned with the camera imaging optics [2, 3, 5, 6, 13]. This simplistic model is insufficient to account for the motion of inexpensive pan-tilt mechanisms. Of course more complex models have been proposed, e.g. Shih gives the details of calibrating a stereo head with multiple actuated degrees of freedom, however orthogonally aligned rotational axes are still assumed [10]. This work introduces a more general model for pan-tilt camera motion.

Existing camera calibration systems typically make use of a relatively small calibration target. While a small target is appropriate for stationary cameras [7, 12, 14] and small working volumes, it is not sufficient to calibrate cameras that are actuated to cover a wide area. In order to obtain high quality calibration, the target should fill the working volume that is observable by the camera. When a single small calibration target is used, it can only be observed from a narrow range of pan-tilt orientations. Some researchers have attempted to move a small target under precise motorized control [10]. A more desirable target would fill the space, and be visible from all possible pan-tilt orientations.

When calibrating stationary cameras, researchers have often addressed the need for wide area calibration targets by constructing large physical targets [9]. This method becomes increasingly difficult as the size of the working volume increases. An alternative is to construct a virtual calibration target by moving a single identifiable marker throughout the working volume and using the markers position over time as a basis for calibration [1, 4]. We believe the latter method scales better to large environments and apply the method to the domain of pan-tilt calibration.

3. Method

Our method improves upon existing work in two specific ways: we use an improved model of camera pan-tilt motion, and we track the motion of a point feature to provide a large virtual calibration target.

3.1. Simple pan-tilt camera model

Most previous researchers using pan-tilt cameras have used a fairly simple camera model. Since the camera

† Affiliated with Stanford University at time of research

J. Davis and X. Chen, "Calibrating pan-tilt cameras in wide-area surveillance networks," IEEE International Conference on Computer Vision, 2003.

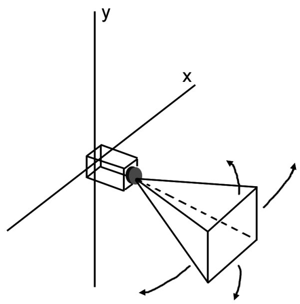

Figure 1 Simple pan-tilt camera motion model. Pan and tilt are modeled as axis aligned rotations around the camera's center of projection.

pans and tilts, the camera motion is modeled as two idealized rotations around the origin, followed by a perspective camera transformation, as shown in Figure 1. This transformation can be written as a sequence of matrix operations

$$\begin{bmatrix} I_x \\ I_y \\ I_w \end{bmatrix} = \mathbf{C}\,\mathbf{R_y}\,\mathbf{R_x} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1)$$

where $[x\ y\ z]^T$ is a point in world coordinates. A rotation around the x-axis, $\mathbf{R_x}$, and then around the y-axis, $\mathbf{R_y}$, correspond to tilt and pan respectively. Finally a perspective camera transform, **C**, results in the image plane coordinates at which the point is observed, $(I_x/I_w\,,\ I_y/I_w)$.

For an ideal pan-tilt camera this model is sufficient. However, the assumption that the pan and tilt axes of rotation are orthogonal, aligned with the image plane, and through the camera's nodal point are frequently violated. The usual solution to this problem is careful engineering, so that the assumptions are as nearly true as possible [11]. Since our system uses commercially integrated pan-tilt cameras that significantly violate the assumptions and can not be easily modified, a better model is needed.

3.2. Improved pan-tilt camera model

When a pan-tilt camera is assembled, it is difficult to ensure that the axes of rotation intersect the optical center of the camera imaging system. In addition, the axes are unlikely to be exactly aligned. In order to model the actual motion of the camera, new parameters can be introduced into the camera model. As shown in Figure 2, rather than being coordinate frame aligned, the pan and tilt degrees of freedom can be modeled as arbitrary axes in space. The imaging plane and optics are modeled as a rigid element that rotates around each of these axes. This model more closely approximates the actual camera geometry, and can be written as

$$\begin{bmatrix} I_x \\ I_y \\ I_w \end{bmatrix} = \mathbf{C}\ \mathbf{T_{pan}}\ \mathbf{R_{pan}}\ \mathbf{T_{pan}^{-1}}\ \mathbf{T_{tilt}}\ \mathbf{R_{tilt}}\ \mathbf{T_{tilt}^{-1}} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2)$$

where $\mathbf{R_{tilt}}$ is a 3x3 rotation matrix which rotates *tilt* angles around a vector in the direction of the tilt-axis. $\mathbf{R_{pan}}$ is analogous. These axes do not necessarily pass through the origin, and $\mathbf{T_{pan}}$ and $\mathbf{T_{tilt}}$ represent translation vectors from the origin to each axis. Thus the projection of a point in space onto the image plane can be found as a function of current camera *pan* and *tilt* parameters.

The above models assume that the angular rotation around each axis is known. Of course the motor control of each axis occurs in terms of some unknown units, and a calibrated mapping to angular units is required. In the case of the Sony EVI-D30 cameras used in our system, rotations are specified to the camera in degrees, and empirical tests suggest that angular rotation was in fact well calibrated by the manufacturer.

3.3. Calibrating the model

The above model specifies the generic relations between geometric components. Each camera will vary within this general model, and determining the specific parameters of the model is typically referred to as calibration. In the case of this model, the parameters to be determined are the two axes of rotation in addition to the standard set of camera intrinsic and extrinsic parameters.

One reason that the simple camera model described above is often used despite its inadequacies is the ease of calibration. Since the axes of rotation are given relative to the camera image plane, it suffices to calibrate the camera in a single configuration and then apply the model when a new set of pan-tilt rotations is required. In the more gen-

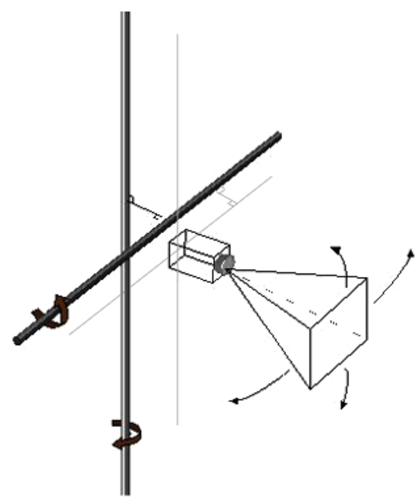

Figure 2 Improved model of camera motion. Pan and tilt motions are modeled as rotations around arbitrary axes in space.

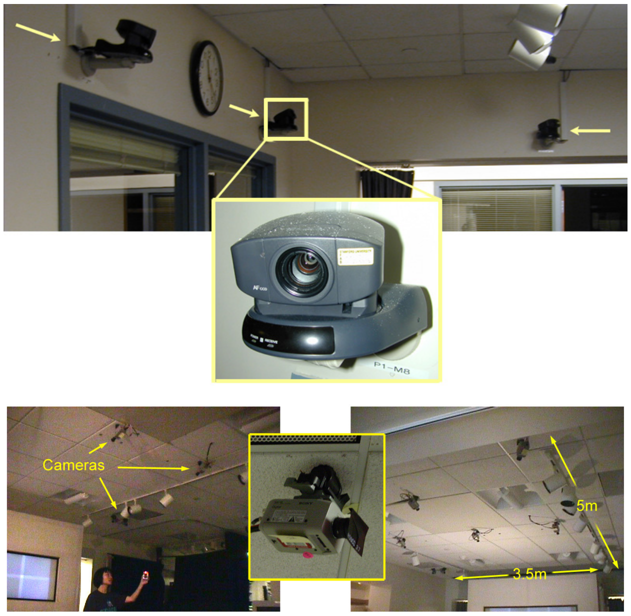

Figure 3 Both static and pan-tilt cameras mounted on the ceiling of a laboratory as part of a wide-area tracking system.

eral case of arbitrary axes, a great deal of additional data needs to be collected in order to robustly determine all parameters.

Traditional camera calibration proceeds by arranging a set of known 3D features in the camera's view frustum. These are typically attached to some calibration target. The projection of each 3D feature is then observed on the camera image plane. These observations are 2D data points. Camera parameters can be determined by solving

$$\arg\min_{\Phi} \{ \mathbb{C}(\Phi, \mathbf{X_{3D}}) - \mathbf{X_{2D}} \} \qquad (3)$$

where $\mathbb{C}$ is the camera model that defines the projection of points onto the image plane, $\mathbf{\Phi}$ is the set of camera parameters to be determined, $\mathbf{X_{3D}}$ is the vector of 3D feature locations, and $\mathbf{X_{2D}}$ is the vector of corresponding image plane observations. Since $\mathbf{\Phi}$ often includes radial lens distortion terms in addition to extrinsic geometric pose, minimizing this equation is usually framed as a non-linear search problem.

In order to obtain a good estimate of the parameters, Φ, it is necessary that the spatial features, $\mathbf{X_{3D}}$, cover the working volume, and that the observations, $\mathbf{X_{2D}}$, cover the image plane. In the event that coverage is not comprehensive, parameters that fit the observed space will be obtained. If a target later moves outside of the observed calibration range, the model will extrapolate into the new area. Since data extrapolation is known to be unreliable relative to interpolation, it is beneficial to obtain maximum coverage [8].

In the case of the pan-tilt camera model considered here, two axes of rotation are added to the set of parameters, Φ. Since the current pan and tilt parameter settings must also be included, the equation now becomes

$$\arg\min_{\Phi} \{ \mathbb{C}(\Phi, \boldsymbol{pan}, \boldsymbol{tilt}, \mathbf{X_{3D}}) - \mathbf{X_{2D}} \} \qquad (4)$$

where ***pan*** and ***tilt*** are vectors specifying the setting corresponding to each feature-observation pair. As before, good calibration requires adequate coverage of the space. In this case, coverage of the range of pan and tilt parameters is implied, as well as coverage of space.

Thus the procedure for calibrating this camera model is an iteration of the procedure used for a static camera. The pan-tilt parameters are set to some value, .e.g. (0º, 0º). A set of feature-observation pairs is obtained. Then the pan-tilt parameter values are changed to some new value, .e.g. (100º, 50º) and an additional set of feature-observation pairs is obtained. The procedure is repeated until the range of camera motion has been covered. All

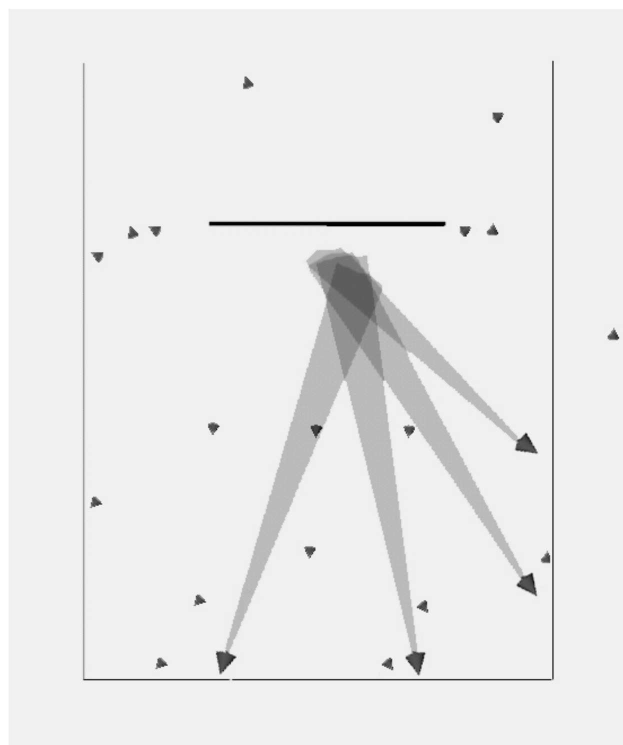

Figure 4 Top down view of pan-tilt camera placement. Cameras can be directed such that their view frustums intersect at the desired target feature location. Small triangles indicate the position of static cameras in the tracking network.

vectors are concatenated and the equation is minimized.

3.4. Acquiring calibration data

Features and observations must be placed to cover the volume observed by the pan-tilt camera. Since the camera can turn, this is potentially a very wide area. It is cumbersome to build a sufficiently large calibration object. However, if a small calibration target is used, it will only be visible from a narrow range of pan-tilt settings, resulting in poor calibration. A small target could be moved to several locations, but this motion would have to be precisely measured and accounted for in some way.

We intend to use the pan-tilt cameras as part of a wide-area tracking system. Figure 3 shows some of these cameras mounted on the ceiling of or laboratory. Figure 4 shows a top down diagram of the placement of both static and pan-tilt cameras. The working volume for this arrangement is approximately 5.0m × 3.5m × 2.0m.

The tracking system itself can be used to build a virtual calibration object. We use the method of [4] to calibrate the set of stationary cameras in our tracking network. Using the stationary cameras only, the motion of a single feature can be tracked over time. Since the 3D location of this path is known, and the point can also be observed by the pan-tilt cameras, it can be used as a virtual calibration object.

We use an LED as shown in Figure 5 as our point feature. This feature can be easily identified by both our tracking system and in the images returned from our pan-tilt cameras. Calibration data is captured for a number of pan-tilt parameter settings. For each setting of (pan,tilt) the LED is moved so that it covers the currently observable working volume. Since the observable area will change with different settings of the pan-tilt parameters, the path of the target must be adjusted according to the current settings. Figure 6 shows a top down view of the path recorded to calibrate one pan-tilt camera. The darker path represents the motion recorded during a single setting of pan-tilt parameters, while the lighter path shows the motion from all settings superimposed.

Figure 5 A bright LED used as a feature easily identifiable in both a tracking system and by the pan-tilt cameras.

Note that although the paths defining 3D target position are gathered individually, they are all relative to the same global coordinate system. By using the tracking system to record all traces, we insure that a single large virtual calibration object is constructed in a single reference frame, even though the motion is in fact recorded in separate sessions.

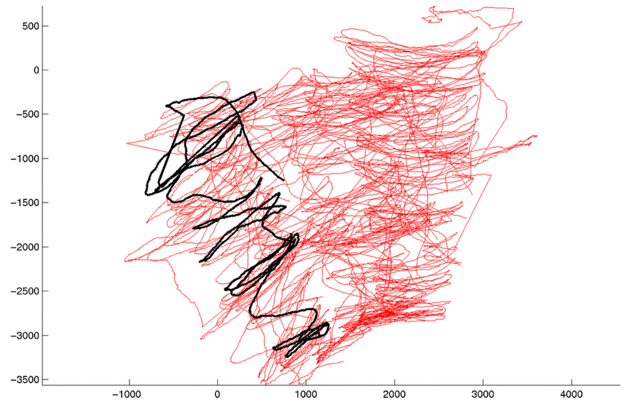

Figure 6 Top down view of the 3D path taken by a LED as it is moved through the working volume. The darker trace is the path for a single setting of pan-tilt parameters, while the lighter trace is a concatenation of all traces recorded for the given camera. Axes units are in millimeters.

4. Evaluation

In order to evaluate the effectiveness of our calibration, we acquired both calibration and test data. The test data was captured in an identical manner as the calibration data, but the two sets were kept separately. Table 1 gives the pan-tilt settings for which calibration and test data were captured.

The quality of camera calibration can be measured in terms of image plane projection error. This is essentially a measure of the camera model's ability to explain data. If the model is of high quality, then the projection of target locations onto the image plane will fall close to actual observations. In all cases we calibrate using the calibration data, and evaluate the quality of the calibration using the projection error of the test data.

Since a virtual calibration trace is captured at 30 Hz, a large number of calibration points are observed. Many of these points are very near one another and do not materially contribute to a good calibration. We seek to minimize the number of calibration features, while ensuring a good sampling of the working volume. To achieve this, the traces are uniformly sub-sampled in time, so that only 50 samples remain. Figure 7 shows a single trace projected onto the image plane of a pan-tilt camera. The remaining calibration samples are shown as points over the original trace.

4.1. Necessity of a full range of motion

It is important that calibration data spans the full range of pan-tilt motion. Earlier techniques that used only a small stationary calibration target were limited to a small range of motion when obtaining calibration data. In order to demonstrate the importance of using a large calibration target, we calibrated a pan-tilt camera using different subsets of the available calibration data. It is expected that using calibration data from only a small range of pan-tilt values will obtain a poor calibration, while using calibration data covering the entire range will perform well. Figure 8 shows a plot of angular pan-tilt coverage vs. average projection error. For a given angular coverage, α, all calibration sets for which the following relation holds are used to calibrate the camera.

$$\sqrt{pan^2 + tilt^2} < \alpha$$

As expected, using a greater range of pan-tilt motion to calibrate the camera produced a better fit to the actual motion of the camera.

4.2. Necessity of improved camera model

The collected data was used for camera calibration under three conditions. In all cases the intrinsic camera parameters including focal length and lens distortion were calculated as a preprocess using the method of Heikkila [7]. In the first condition, only the data from the $(pan = 0, tilt = 0)$ parameter setting was used. Under this condition the external pose was calculated using standard camera calibration, and the rotational axes were assumed to be orthogonal to and aligned with the optical imaging plane. Since data from only a limited portion of the configuration space was used, it is expected that errors in the extrapolated regions of the space will be relatively high. In the second condition, data from all pan-tilt parameter settings was aggregated and used to find the best fit external pose of the camera, still under the assumption that rotations are around orthogonal aligned axes. This condition is similar to the simple model of pan-tilt motion used in nearly all other research. Finally, the model proposed in this work was investigated. All data was used to calibrate both the external pose of the camera, and a pair of arbitrary rotational axes.

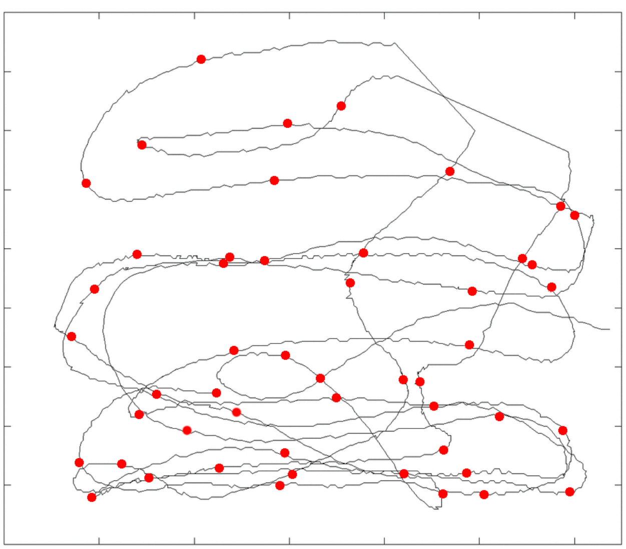

Figure 7 Image plane observation of the path of a LED feature as it is moved through the working volume. Since the trace has many similar points, it is subsampled to only 50 points, shown as dots in this image.

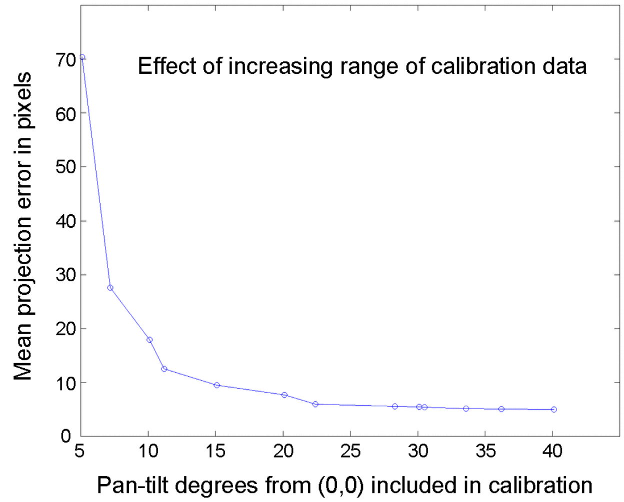

Figure 8 Plot of angular pan-tilt coverage vs. calibration quality. Including calibration data from a larger range of motion improves the quality of the calibration.

Figure 9 shows histograms of the projection error using each camera model. As expected, using calibration data recorded at only the $(pan = 0, tilt = 0)$ parameter setting performs poorly. Using all available calibration data, the ability to predict image point projection is noticeably improved. However the model is still not capable of fully accounting for the camera's motion. Using the more complete pan-tilt model proposed in this paper leads to even further improvement in calibration.

5. Conclusions and future work

This work has presented a method for calibrating pan-tilt cameras that are part of a wide area surveillance or tracking system.

A camera model that accounts for the inadequacies of inexpensive pan-tilt mechanisms was introduced and validated. Empirically, the new model was better able to predict the location of scene features on the camera image plane than the traditional calibration model.

In addition, the methodology for obtaining calibration data was extended to make use of an existing tracking framework. Rather than rely on a small target which can cover only a small portion of the parameter space, a large virtual calibration object is constructed which covers the entire working volume. The importance of covering the pan-tilt parameter space was shown experimentally.

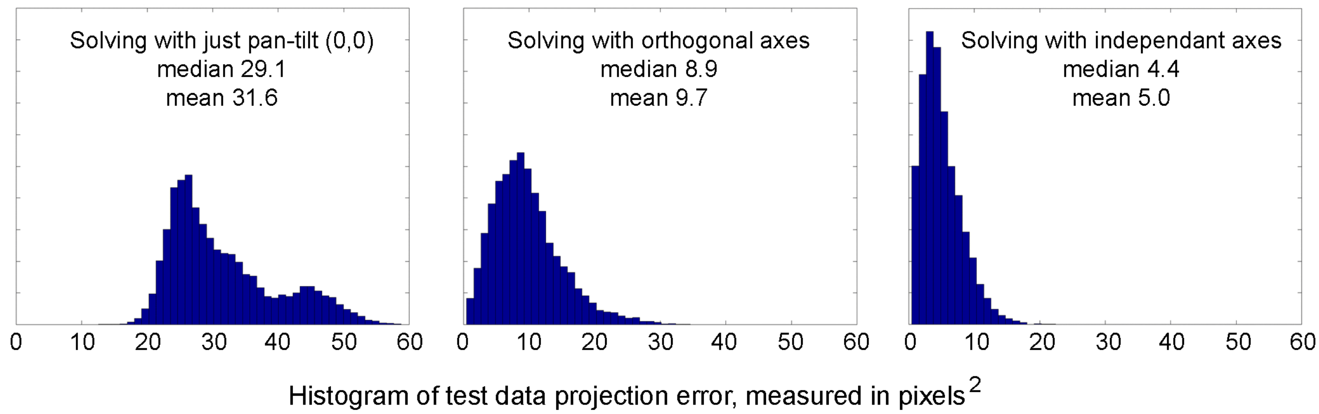

Figure 9 Histograms of projection error under three calibration conditions. (Left) Only data from the (0,0) pan-tilt setting is used, and axes are assumed to be orthogonal and aligned with the image plane. (Middle) All calibration data is used, but axes are still assumed to be orthogonal and aligned to the image plane. (Right) All calibration data is used and axes are arbitrarily positioned in space. The most general model can be seen to fit the test data significantly better.

Despite the success of this method for our application, improvements are possible. This work has focused on pan-tilt cameras that exist as part of surveillance network containing static cameras that have been previously calibrated. It should be possible to calibrate a network containing only pan-tilt cameras by temporarily fixing half of the cameras to be stationary, calibrating as described, and then repeating for the remaining half.

Table 1

Calibration Data Sets	
0	0
10	0
-10	0
-20	0
-30	0
-40	0
-30	-5
-30	-15
-30	-20
-20	-20
-10	-20
0	-20
0	-15
0	-5
-5	0
-5	-5
-10	-5
Test Data Sets	
-30	-10
0	-10
-5	-10
-10	-10
-20	-10
Pan	**Tilt**

6. Acknowledgements

Pat Hanrahan provided useful insights on much of this work. Intel, Interval, and Sony provided partial funding.

7. References

[1] A. Azarbayejani and A. Pentland, "Real-time Self-calibrating Stereo Person Tracking Using 3-D Shape Estimation from Blob Features," Proceedings of the 13th International Conference on Pattern Recognition, vol. 3, pp. 627-32, 1996.

[2] J. P. Barreto, P. Peixoto, J. Batista, and H. Araujo, "Tracking multiple objects in 3D," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'99), 1999.

[3] A. Basu and K. Ravi, "Active camera calibration using pan, tilt and roll," IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), vol. 27, pp. 559-66, 1997.

[4] X. Chen and J. Davis, "Wide Area Camera Calibration Using Virtual Calibration Objects," IEEE Computer Vision and Pattern Recognition, 2000.

[5] R. T. Collins and Y. Tsin, "Calibration of an outdoor active camera system," presented at IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999.

[6] S. Fry, M. Bichsel, P. Muller, and D. Robert, "Tracking of flying insects using pan-tilt cameras," Journal of Neuroscience Methods, vol. 101, pp. 59-67, 2000.

[7] J. Heikkila and O. Silven, "A Four-Step Camera Calibration Procedure with Implicit Image Correction," IEEE Computer Vision and Pattern Recognition, pp. 1106-1112, 1997.

[8] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, Numerical Recipes in C : The Art of Scientific Computing, 2nd Edition: Cambridge University Press, 1992.

[9] P. Rander, "A Multi-Camera Method for 3D Digitization of Dynamic, Real-World Events," Robotics Institute, Carnegie Mellon University, 1998.

[10] S.-W. Shih, Y.-P. Hung, and W.-S. Lin, "Calibration of an active binocular head," IEEE Transactions on Systems, Man & Cybernetics, Part A (Systems & Humans), vol. 28, pp. 426-42, 1998.

[11] C. Tomasi and J. Zhang, "How to rotate a camera," presented at ICIAP '99 - 10th International Conference on Image Analysis and Processing, 1999.

[12] R. Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," IEEE Trans. Robotics and Automation, vol. 3, pp. 323-344, 1987.

[13] D. C. Woo and D. W. Capson, "3D visual tracking using a network of low-cost pan/tilt cameras," presented at Canadian Conference on Electrical and Computer Engineering Conference Proceedings, 2000.

[14] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 1330-1334, 2000.

J. Davis and X. Chen, “Calibrating pan-tilt cameras in wide-area surveillance networks,” IEEE International Conference on Computer Vision, 2003.