A Fast Linear Registration Framework for Multi-Camera GIS Coordination*

Karthik Sankaranarayanan James W. Davis
Dept. of Computer Science and Engineering
Ohio State University
Columbus, OH 43210 USA
{sankaran,jwdavis}@cse.ohio-state.edu

Abstract

We propose a novel registration framework to map the field-of-coverage of pan-tilt cameras to a GIS (Geographic Information System) planar coordinate system. The camera's field-of-coverage is obtained by building a spherical panorama using an efficient active camera model. The pan-tilt orientation rays from the panoramic image are projected onto a GIS orthophoto ground plane and registered using a transformation matrix. The parameters of the transformation are learned in linear time using least squares. The proposed model is experimentally evaluated by registering panoramas from multiple cameras with an orthophoto and then overlaying them with GIS metadata ground truth to validate the accuracy of registration. We also demonstrate the applicability of such a GIS-based framework to a multi-camera, master-slave active tracking system.

1. Introduction

A typical outdoor urban surveillance system consists of multiple pan-tilt-zoom (PTZ) cameras overlooking different areas. To utilize these distributed cameras in a coherent way, it is important to establish a common reference frame to which each of these cameras can be mapped. This enables computer vision tasks on a standard coordinate frame of reference, such as multi-camera coordination and tracking. A Geographic Information System (GIS) provides a natural schema for establishing such a common frame of reference because it not only provides solid ground truth but more importantly provides semantic information (e.g., locations of roads, buildings, sensitive areas, etc.) for use in applications such as tracking and activity analysis. Such labelings can help provide semantically rich scene interpretation for urban surveillance.

In this paper, we propose a model and a method to register each camera's complete field-of-coverage (all possible views of a PTZ camera) to a GIS coordinate system. The registration is performed by means of a mapping from the camera's pan-tilt space (view coverage) to the GIS ground plane. Each camera's complete field-of-view is available by creating a panoramic image from smaller overlapping views. This yields a single 360×90 (pan$\times$tilt) image view from the camera (see Fig. 2(c)). Since this panoramic image simulates a linear fisheye lens view, we then perform a "de-fishing" operation to warp the panorama onto an overhead ground plane view. We employ a GIS orthophoto image of the area for the mapping (see Fig. 1). A GIS orthophoto is a high-resolution raster image produced from an aerial photograph and contains metadata used to uniquely identify and position the information on a world coordinate system. The registration between the panorama and the orthophoto is performed by projecting rays of the pan-tilt orientations onto the orthophoto ground plane and then mapping corresponding feature points using an affine transformation.

By registering each of the cameras to a common reference (GIS orthophoto), we now know the pan-tilt orientations of each camera required to view the same ground location (for all ground locations). Therefore, one could select a specific location in the scene and have each of the cameras automatically orient to that location. This is very important in enabling coordinated computer vision tasks such as multi-camera tracking, handoff, and 3D reconstruction. Thus, a mapping from all cameras to a common GIS reference can provide a unified, comprehensive view of the world. We examine the proposed model by mapping panoramic images from multiple cameras onto a GIS orthophoto and then verify the accuracy of registration. We also demonstrate an application of the proposed model with a real-time, multi-camera coordinated tracking system.

In Sect. 2 we discuss previous approaches to this registration and mapping problem. Section 3 describes in detail the proposed framework. Section 4 covers the set of experiments performed to validate the model, and we conclude with a summary in Sect. 5.

*Appears in *IEEE International Conference On Advanced Video and Signal Based Surveillance*, September 2008.

Figure 1. GIS orthophoto of the target area.

2. Related Work

Previous work related to establishing relationships between cameras in a distributed system have typically dealt with the problem by proposing different methods for finding homographies between each pair of cameras. In [5], tracking data between pairs of cameras are employed and centroids of moving objects are used to solve for correspondences. Another way to obtain these correspondences between objects detected in cameras is by following feature matching techniques [4] or by using geometric methods [3, 1]. However, these techniques establish relationships between pairs of cameras and do not share a common coordinate system. These techniques also rely upon recovering the camera locations whereas our proposed model provides additional degrees of freedom by mapping directly from the pan-tilt camera panorama to the GIS ground plane, thus eliminating the need to know the location of the camera. Also, learning homographies is based on pairs of views and not their entire field-of-coverage. In [2], a catadioptric omnicamera is used to establish a mapping from a limited resolution omnidirectional view to the ground plane. However this mapping is employed by means of a lookup table and the inverse mapping from the ground plane back the camera view cannot be expressed analytically, thus hindering analysis sourced from ground plane information.

3. Framework

The proposed model utilizes spherical panoramas [8] generated for each camera, to perform the mapping and registration with the GIS orthophoto ground plane. The panorama is "defished" and then registered using an affine transformation to a GIS-based overhead view by projecting the rays of view from the panorama onto the ground plane. The associated parameters involved in this transformation are learned by matching corresponding feature points from the source panorama and the target GIS orthophoto.

3.1. GIS orthophoto

A GIS orthophoto is a high-resolution raster image produced from an aerial photograph/image (or mosaic of images) that has been corrected for particular distortions. Figure 1 shows the GIS orthophoto (approx. 1000×1000 pixels) for our region of interest. Such imagery is publicly available and we attained this image from the state geographic information office. Our GIS dataset also contains vector layers (in the form of shape files) that identify locations and boundaries of roads, buildings, parking lots, etc. Embedded within the raster and vector layers is metadata which is used to uniquely identify and position the information on a world coordinate referencing system. This dataset provides the ground reference system for registration, and the associated vector layers features provide a form of ground truth for evaluation of the registration process.

3.2. Active Camera Model and Panoramas

In our previous work [8], we proposed an efficient active camera model for pan-tilt cameras. This model is used to map image coordinates directly to the camera's pan-tilt orientations in constant time. The model is based on the elliptical locus of the projections of a fixed point on the original image plane of a moving camera. The parametric location of this point along the ellipse defines the change in camera orientation. The lengths of the major and minor axes of this ellipse are used to calculate the required changes in pan and tilt. This model does not require any knowledge of camera parameters other than the focal length and hence is simple to calibrate. Since this model provides a mapping from the image coordinates (x, y) to the world camera orientations (pan, tilt), we can use this model to effectively build spherical panoramas. Such panoramas provide a natural and intuitive way to represent the complete pan-tilt field-of-coverage of cameras in wide-area surveillance systems.

To construct a spherical panorama, we orient the camera to a set of pre-configured pan-tilt locations (computed automatically), and capture a set of images in such a way that they cover the complete pan-tilt space. Figure 2(a) and (b) show an outdoor PTZ dome camera and one such component image from the camera. Using the camera model [8], every pixel from each component image is mapped to its corresponding pan-tilt location in the 360×90 space. This pan-tilt pixel data plotted on a polar coordinate system produces a spherical panorama, as shown in Fig. 2(c). Further details of this camera model and panorama construction are provided in [8].

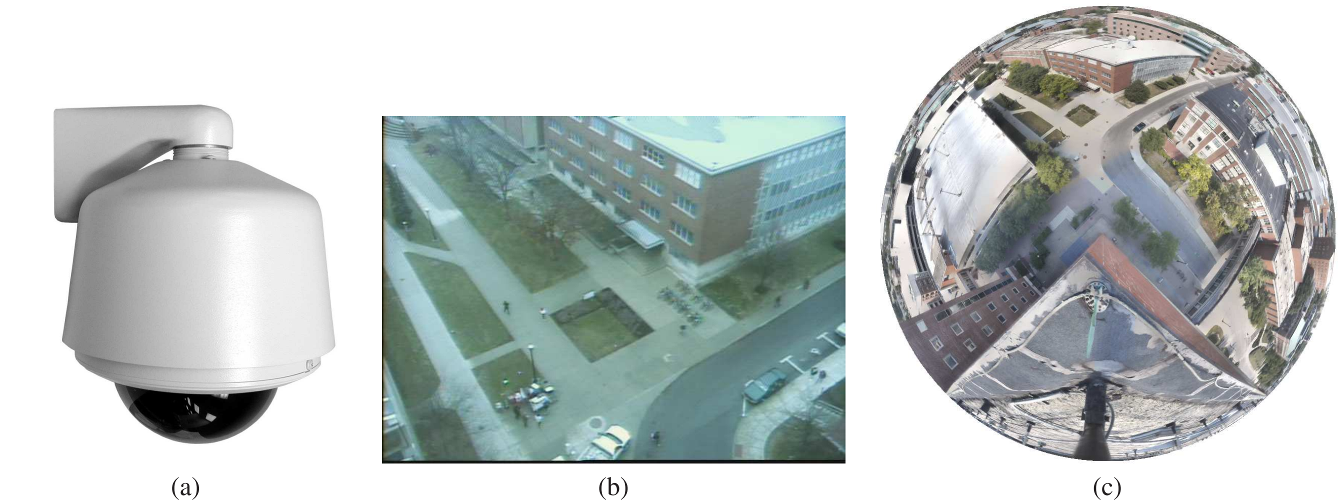


Figure 2. Camera panorama. (a) PTZ dome camera. (b) Sample component image from the camera. (c) Resulting spherical panorama.

In our method, spherical panoramas are modeled as a linear fisheye [6], based on the principle that the distance between an image point and the principle point is linearly dependent on the angle of incidence of the ray. In the camera's pan-tilt space, this means that each pan-tilt location is represented by a point in the panoramic image whose angle subtended with the X-axis indicates the pan of that location and the distance of the point from the center represents its tilt. Thus, the radius varies linearly with tilt angle. We point the reader to [6] for further details on fisheye lens model.

3.3. Registration Model

In this section, we describe how to perform the panorama "defishing" operation and affine transformation between the spherical panorama and the GIS aerial orthophoto. Figure 3 presents a geometric description of the model. A pan-tilt camera is located at point O and is currently oriented at a pan angle θ and tilt angle ϕ. The hemispherical dome D represents the projection of the world onto this pan-tilt camera. The plane X_gY_g represents the GIS ground plane (i.e., orthophoto image) with its origin being the orthophoto image origin. The X and Y axes, on the other hand, represent the coordinate axes aligned with the panoramic image such that the pan angle θ is the angle subtended with X. Point P is an arbitrary point on the ground plane and point Q is its projection on the pan-tilt space of the camera. Let x and y be the coordinates of the point P on the XY frame of reference and let h (unknown) be the height of the camera (in pixels). Then, from the geometry of Fig. 3 we see that

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \tag{1}$$

$$\phi = \tan^{-1}\left(\frac{\sqrt{x^2+y^2}}{h}\right) \tag{2}$$

Rewriting these equations for x and y,

$$x = \frac{h \cdot \tan\phi}{\sqrt{1+\tan^2\theta}} = \frac{h \cdot \tan\phi}{\sqrt{\sec^2\theta}} \tag{3}$$

$$y = x \cdot \tan\theta \tag{4}$$

Further, simplifying the above equations, we get

$$x = h \cdot \tan\phi \cdot \cos\theta \tag{5}$$

$$y = h \cdot \tan\phi \cdot \sin\theta \tag{6}$$

The above mapping gives us the projection of the spherical panorama onto the ground plane. This mapping model assumes that the height of the camera h is known, and that the location of the camera center (point M) on the ground plane is known. Note that this projection is not yet aligned with the orthophoto axes X_gY_g. We will show how we eliminate the need to learn these parameters and align with X_gY_g by means of a linear transformation. Let (x_g, y_g) be a ground truth location point in the coordinate space X_gY_g and let (x, y) be the corresponding location in the coordinate space XY (dictated by the pan origin of the camera). Then the linear transformation matrix which gives the transformation between these two points can be written as:

$$\begin{bmatrix} x_g \\ y_g \\ 1 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & t_x \\ k_3 & k_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

where t_x and t_y are the translation parameters and k_1, k_2, k_3, and k_4 are the affine parameters. From Eqns. (5) and (6), by substituting $c_1 = \tan\phi \cdot \cos\theta$ and $c_2 = \tan\phi \cdot \sin\theta$, we get

$$x = c_1 \cdot h \tag{8}$$

$$y = c_2 \cdot h \tag{9}$$

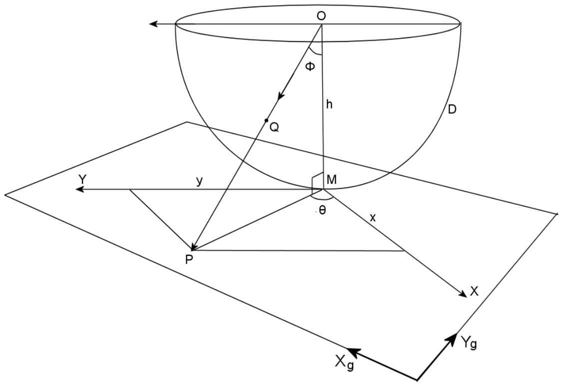


Figure 3. Geometry of the registration model.

Substituting these values in Eqn. (7) and moving the parameter h into the linear transformation matrix, we get

$$\begin{bmatrix} x_g \\ y_g \\ 1 \end{bmatrix} = \begin{bmatrix} k_1 h & k_2 h & t_x \\ k_3 h & k_4 h & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ 1 \end{bmatrix} \quad (10)$$

$$\begin{bmatrix} x_g \\ y_g \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ 1 \end{bmatrix} \quad (11)$$

Therefore the projection and the affine transformation have been combined into one matrix whose six parameters govern all the necessary degrees of freedom. These parameters can be learned by rewriting the Eqn. (11) as

$$\begin{bmatrix} c_1 & c_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & c_1 & c_2 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} x_g \\ y_g \end{bmatrix} \quad (12)$$

and can be solved using multiple corresponding points. By matching points on the panorama (x, y) to the orthophoto ground plane points (x_g, y_g), we obtain a set of correspondences. For each point on the spherical panorama, we know its pan-tilt (θ, ϕ) (using the linear fisheye model), hence we can get c_1 and c_2 for each point. These values can then be employed in a least squares formulation for Eqn. (12) to learn the six parameters.

The result provides a direct mapping from the camera's pan-tilt space to the orthophoto ground plane. Similarly, the inverse of this system gives us the reverse mapping from the orthophoto back to the pan-tilt space.

4. Experiments

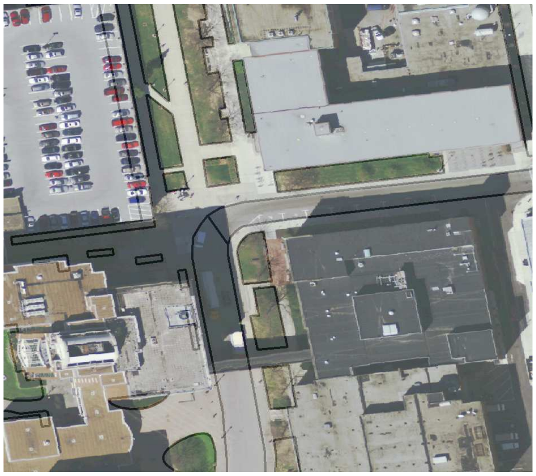
The proposed framework provides a model for mapping the pan-tilt orientation from a camera to a GIS orthophoto ground plane. To examine and test this model, we performed two main classes of experiments. The first involves registering panorama image data from multiple cameras onto an orthophoto and then verifying the accuracy of registration. The next set of experiments examines an application of the proposed framework to a real-time, multi-camera tracking system.

4.1. Registration

Figure 2(c) shows an example spherical panorama which displays the complete field-of-coverage for a pan-tilt camera and Fig. 1 shows the GIS orthophoto raster for the same region (though with different extents of coverage). Our task was to transform panorama image data from multiple cameras onto this orthophoto space. We first identified a set of feature points on a panorama that also appear in the orthophoto. In our experiments, we used 10-12 such feature points (minimum of 6 points required). We then computed the pan-tilt orientation of each of those points in that camera's pan-tilt space using the angle subtended by each point with the X-axis (which indicates the pan of that location) and its distance from the center of the panorama (which represents its tilt angle). From this we calculated the c_1 and c_2 parameters for each of these feature points. Next, we marked the points corresponding to each of these locations on the GIS orthophoto. This gave their x_g, y_g locations in the $X_g Y_g$ GIS planar coordinate frame. Using these values (c_1, c_2, x_g, y_g) for multiple points, we calculated the affine registration employing a least squares formulation of Eqn. 12. We used the resulting transformation matrix to register the panorama with the GIS orthophoto. We did this procedure for three different cameras (each on a different building).

To verify the accuracy of the panorama-to-orthophoto registrations, we employed the additional GIS vector layers that contain precise outlines of the roads and grass patches. The outline features were manually traced from the orthophoto and are shown in Fig. 4(a). To evaluate the registration results, we overlaid the the GIS outlines on top of the transformed panoramas (now aligned to the orthophoto) and validated the accuracy of the registration. Figures 4(b),(c), and (d) show the transformed panoramas overlaid with the GIS outline features. It can be seen that the GIS outlines register very well with the different ground plane features in the transformed panoramas. The parallax of buildings, trees, and other vertical structures will occur (as expected), as we were mapping only to ground-level features. These results demonstrate the accuracy of the registration method and thus validate the correctness of the mapping framework.

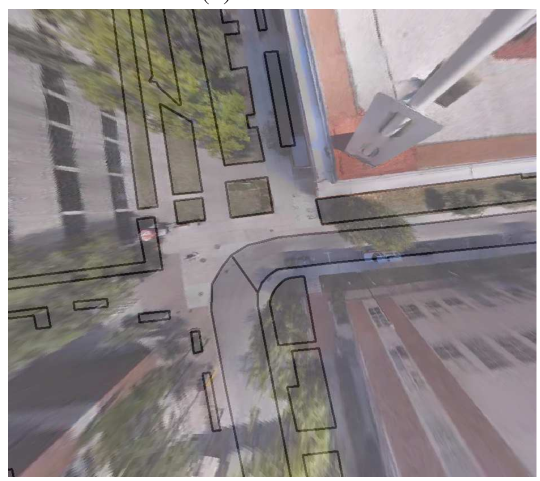
A point to be noted is that the distances of the transformed points from the orthophoto origin grows as a tangent function of tilt. Hence, points at the horizon in the panorama (tilt 90) map to the ground plane at ∞. There-



(a) Orthophoto

(b) Camera-1

(c) Camera-2



(d) Camera-3

Figure 4. Registration validation. (a) GIS orthophoto overlaid with feature outlines. (b)-(d) Transformed panoramas with feature outlines.

fore, instead of transforming the complete panorama, we place a limit on the tilt angle to localize the projection.

4.2. Multi-camera coordinated tracking

Using the proposed model, we next developed and examined an active tracking application using multiple cameras in a coordinated master-slave system. Such a system comprises of a master camera which actively tracks the target across the scene and multiple slave cameras who update their orientations to follow the same target (using the tracking information resulting from the master camera).

The tracking algorithm used in this system is the covariance tracker [7]. The algorithm uses a covariance of color and spatial features to model the target and uses template-based matching to track the target in successive frames. The tracker in our system is initialized manually and as the target moves, the master camera calculates its required change in pan and tilt using the active camera model described in Sect. 3.2. This enables it to keep the target centered in its view. The current tracking position from the master camera is registered into the GIS ground plane, from which the reverse mapping of that position into the slave cameras pan-tilt orientations is used to re-orient the slave cameras. Thus, while the master camera actively tracks the target, the slave cameras simultaneously provide different views (from different distances) of the same target.

Figure 5 shows the results from this multi-camera tracking system using 3 Pelco Spectra III SE wide-area PTZ surveillance cameras mounted on top of buildings of different heights. Figure 5(a) shows the raw (not smoothed) tracks of the target (tracked by the master camera) overlaid on the orthophoto image. The image chips in Fig. 5(b) show close-ups of the target being tracked at different time instances (marked in Fig. 5(a)). The field-of-coverage of the 3 different cameras are shown by the ellipses. As the target moves across the scene, it crosses regions of overlapping coverage (where the target is visible from multiple cameras) as well non-overlapping coverage (where the target is not visible from one or more slave cameras). In Fig. 5(a), the red track starts off with the target visible from both the master and one slave camera, then crosses over into a region of visibility from all 3 cameras and subsequently moves onto an area of only the master and slave 2's visibility. The blue track, on the other hand, starts with the master and slave 1 following the target, then crossing over onto an area of complete overlap followed by a region where the target is visible only from the master and slave 2.

The tracking results obtained with these three cameras demonstrate the applicability of the proposed model to an accurate multi-camera coordinated active tracking system in real-time.

5. Summary

We proposed a novel framework to register the field-of-coverage of multiple pan-tilt cameras to a common GIS planar coordinate system. This mapping is obtained by utilizing linear fisheye panoramas thus giving the pan-tilt locations of any point in the field of a camera. These panoramas are then projected onto a GIS ground plane and registered using a transformation matrix, the parameters of which are learned by establishing point correspondences between the panoramic view and the GIS plane. We experimentally evaluated the proposed framework by registering panoramas from multiple cameras onto a GIS orthophoto and validated the accuracy of registration using GIS ground features. We also demonstrated the applicability of such a GIS mapping framework by means of a multi-camera master-slave active camera tracking system. The proposed method currently employs a fixed zoom level (multiple zoom levels are to be addressed in future work) and requires an absolute pan-tilt positioning mechanism for the camera (common on many quality PTZ cameras).

Since we now have the information from each camera mapped to a common GIS framework, in the future, we plan to utilize the abundant semantic information offered to us by the GIS system to perform intelligent scene interpretations and event analysis. This framework can also be utilized in coordinated computer vision tasks such as multi-camera tracking, camera-handoff, and 3D reconstruction.

6. Acknowledgements

This research was supported in part by the National Science Foundation under grant No. 0236653. We additionally thank Jon Downey of Woolpert for assistance with the orthophoto used in this research.

References

[1] J. Black and T. Ellis. Multi camera image tracking. *Image and Vision Comp.*, 24(11):1256–1267, November 2006.

[2] J. Gaspar and J. Santos-Victor. Visual path following with a catadioptric panoramic camera. In *International Symposium on Intelligent Robotic Systems - SIRS*, pages 139–147, 1999.

[3] S. Khan and M. Shah. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 25(10):1355–1360, October 2003.

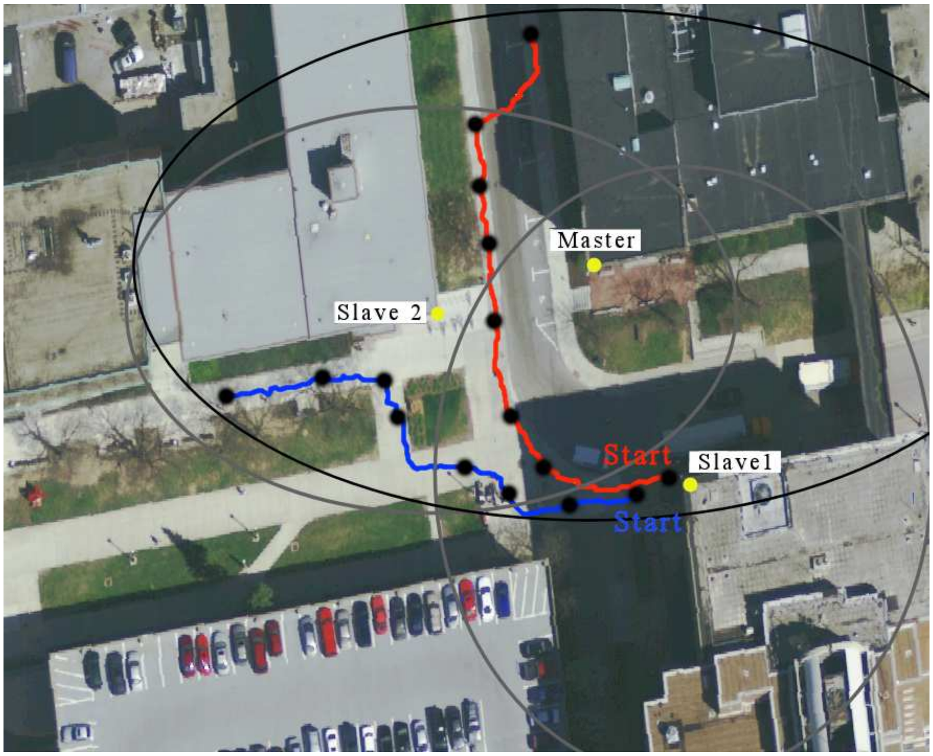
[4] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for easyliving. In *IEEE International Workshop on Visual Surveillance*, pages 3–10, 2000.

[5] L. Lee, R. Romano, and G. Stein. Monitoring activities from multiple video streams: Establishing a common coordinate frame. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 22(8):758–767, August 2000.

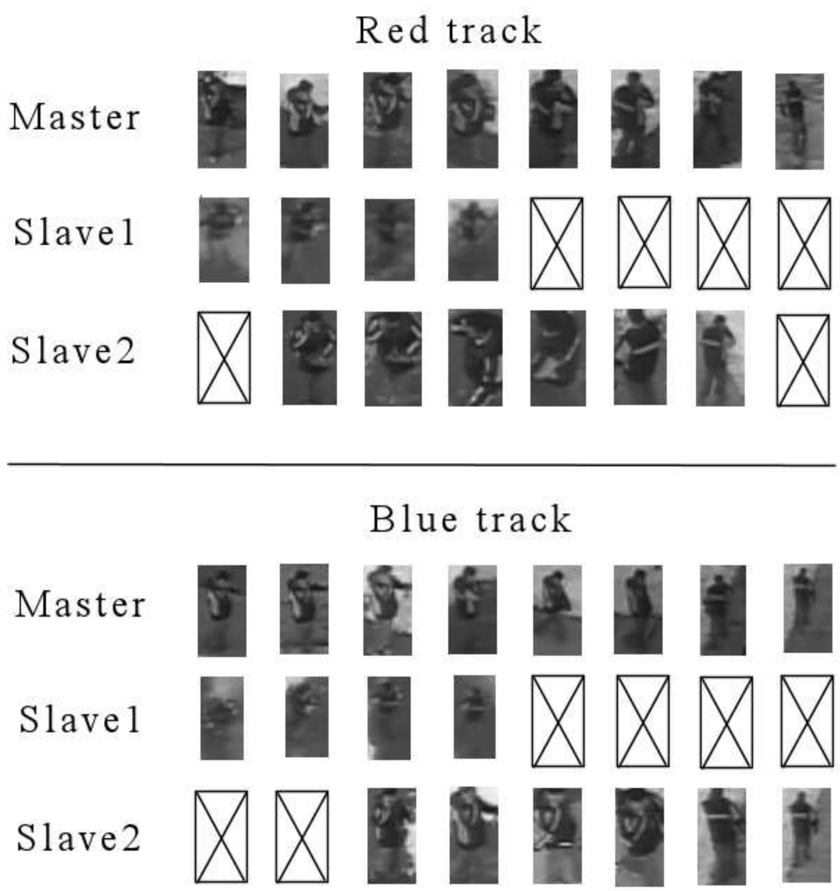
[6] K. Miyamoto. Fisheye lens. *Journal of Optical Society of America*, pages 1060–1061, 1964.

[7] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on means on Riemannian manifolds. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.

[8] K. Sankaranarayanan and J. Davis. An efficient active camera model for video surveillance. In *Proc. IEEE Workshop on Applications of Computer Vision*, 2008.



(a)



(b)

Figure 5. Tracking results. (a) Tracks overlaid on the GIS orthophoto. Ellipses show regions of coverage for the 3 cameras. (b) Image chips for the locations of the black dots in (a). Crosses indicate the target is out of the visible area.