Coordinated Steering for an Uncalibrated Pan-Tilt-Zoom Camera Array

Kit Axelrod and Surya Singh
Australian Centre for Field Robotics
University of Sydney
Sydney, NSW, 2006 Australia
kaxe6043@uni.sydney.edu, spns@acfr.usyd.edu.au

Abstract

Pan-tilt-zoom (PTZ) cameras complement many robotic applications. A coordinated camera steering method is presented to orient a set of commodity cameras to gaze at the same object or area of interest in the workspace. This method solves for a relevant camera model (intrinsics and extrinsics) and uses this to compute the geometry required for motion relative to a tracking signal.

This paper modifies auto-calibration of intrinsics and self-surveying methods of extrinsics for on-line gaze control operations over off-line reconstruction. To this feature tracking is added for feedback to compensate for servo imprecision and asynchrony. The performance of the approach is validated through a cooperative steering task on an array of PTZ cameras. Results show successful automatic steering and mean calibrations within 5% of estimates generated using reference calibration techniques.

1 Introduction

Pan-tilt-zoom (PTZ) cameras, and active camera systems in general, are increasingly common in robotic systems with applications from outdoor surveillance [Collins and Tsin, 1999] to tracking [Everts et al., 2007] to mosaic generation [Sinha and Pollefeys, 2006] and robot navigation [Civera et al., 2009]. These systems can provide both coverage (through pan and tilt motions) and resolution (through zooming). In environments where activity is dynamic, such systems can follow the principal activity. The additional degree(s) of freedom, while providing flexibility, require control. Hence, camera steering is central to these and other applications.

This paper considers the case of steering an array of multiple uninitialized asynchronous PTZ cameras in an unstructured (field) environment. Extensions are introduced to pairwise homography based auto-calibration and structure for motion based self-surveying procedures to estimate the intrinsic and extrinsic parameters. These can then be used with a perspective camera geometry to coordinate camera motions towards an overlapping area of interest (as in this work) or towards different regions (for coverage). To accommodate for servo latency and error, which is particularly noticeable in low-cost commodity hardware, further extensions are provided in the control scheme to feed-forward for the gaze error and frame capture timing variations, as these errors are particularly evident in (imprecise) commodity pan-tilt servo camera mechanisms and as these effects tend to amplify over the delays introduced by slower, asynchronous communications links.

A motivating application of a distributed array of PTZ cameras (such as that in Fig. 1) is telerobotics. In particular, to obtain multiple views of object(s) (in a given area of interest) as seen by a selected master camera. In this way the operator has to only direct one camera (through manual steering or through feature selection from an object library). The rest of the cameras automatically steer in pursuit.

The first part of the paper analyses and consolidates a number of methods proposed for pan-tilt-zoom camera self-calibration to produce a full intrinsic calibration and extrinsic calibration up to scale. The later sections register this calibration data into a map of the camera array and solve for elevation and azimuth gazing angles. The paper concludes with experimental results and tracking performance for the system.

2 Approach and Related Work

Two main approaches may be considered for camera steering. The approach adopted in this paper is to track specific locations or regions in the operating workspace. This is in contrast to visual servoing [Badri et al., 2007] in which scene features (either dense or sparse) are tracked in the image space. A visual servoing approach at best provides weak coordination as it can be executed centrally with one camera sending feature vectors

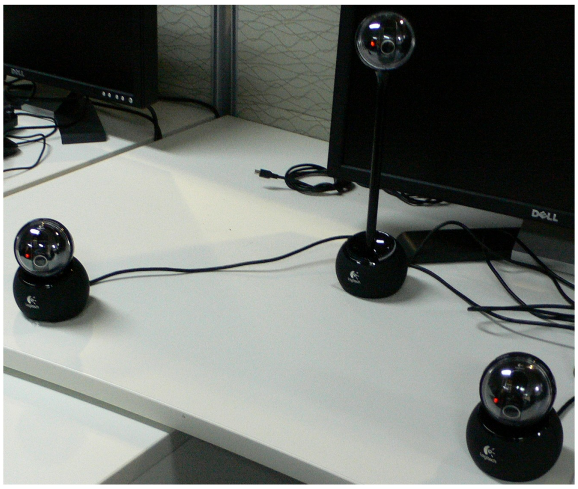


Figure 1: An example pan-tilt-zoom camera array. The cameras are steered to maintain gaze towards a common point in the workspace.

to neighbouring cameras or independently on each camera (with the cameras coordinated through the scene). However, the explicit approach provides direct coordinated control of camera steering with all cameras pointing at the same point (or area). It can handle multiple objects in the field of view and potential occlusions. It does not require an extensive feature database with all the potential observation poses for objects as would be needed in a visual servoing approach. However, it requires a camera model and, in particular, an estimate of camera intrinsics and extrinsics (see architecture in Fig. 2).

Typically the camera model and calibration is assumed. For example, the EyeVision system in sports requires good synchronization and calibration [Tsuji et al., 2003]. However, the PTZ camera geometry can be exploited to assist with this part of the process (see also section 3). An approach based on pairwise homographies (rotation-only auto-calibration methods) is introduced in [Faugeras et al., 1992] and [Hartley and Zisserman, 2003]. This approach is used by [Sinha and Pollefeys, 2006] to aid with mosaic generation. A similarly derived approach [Everts et al., 2007], considers tracking with a pair of PTZ cameras, but uses a circular calibration target. Even though the cameras could be calibrated directly using a pattern [Zhang, 2000], an auto-calibration approach is useful as the calibration can vary (e.g., as the camera is zoomed) or drift (e.g., due to the backlash and servoing errors in the pan-tilt mechanism). Furthermore, as noted in [de Agapito et al., 1999], homography only solutions can be unstable unless there is camera motion about the depth (or principle) axis as this can cause problems with lack of convergence in the the final solution. This approach addresses this by integrating observability analysis into the motion control scheme (e.g., by adding additional tilt motions to a pan-only sequence).

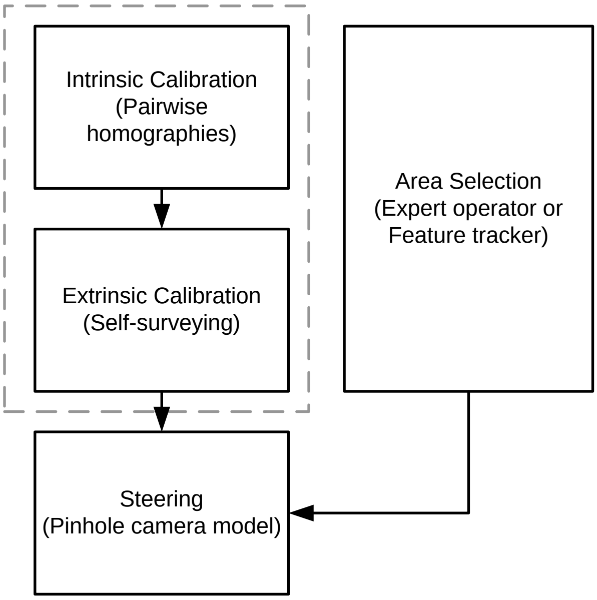


Figure 2: Coordinated camera steering using a direct approach in which points are explicitly tracked in the workspace.

The extrinsics can be recovered using a self-surveying camera array concept where a group of networked cameras calibrated in a system-wide reference frame are used to track a moving target [Matsuoka et al., 2007]. This achieves high accuracy; however, target motion and camera position can not be computed until post-hoc analysis using a structure-from-motion approach. It is a fundamental assumption that the positions of cameras in the array do not change after calibration, which is justified for PTZ (rotating) cameras.

With this the cameras can be steered using perspective camera geometry. Given point in the workspace and a known camera location, the pan and tilt orientations for the projection ray from the principal point can be solved using trigonometric relations.

3 Camera Calibration

There are various mechanisms that exploit scene structure for camera calibration. This process is particularly important for commodity cameras having imprecise mechanisms. The process is four fold: first calculate distortion, then pan-tilt hysteresis variation, then image center, and then extrinsics.

3.1 Intrinsics

Distortion

Before feature matches may be used to calculate Homographies and Fundamental Matrices from a captured dataset, it is necessary to account for radial distortion present in the images. Similar to [Lanz, 2004], the inverse distortion model is shown as follows. An actual projection (p_x, p_y) is related to its undistorted point $(\hat{p}_x, \hat{p}_y)$ by a radial displacement from the center of distortion (which has been confirmed frequently to be near the principle point of the camera for all practical purposes) by the relation below:

$$\begin{pmatrix} \hat{p}_x \\ \hat{p}_y \end{pmatrix} = \begin{pmatrix} c_x \\ x_y \end{pmatrix} + L\left(r\right) \begin{pmatrix} p_x - c_x \\ p_y - c_y \end{pmatrix} \tag{1}$$

$$r = \sqrt{\left(p_x - c_x\right)^2 + \left(p_y - c_y\right)^2} \tag{2}$$

The principle point is estimated by constructing feature matches between increasing levels of zoom. As optical flow is radial about the principle point during zoom, a calibration pattern may be generated by matching image features during zoom. In the case of digital zoom flow is central about the digital image centre, it is possible to introduce motion about the principle point by adding a camera rotation between images, while keeping the high zoom image within the field of view of the low zoom image. Performing this process reciprocally between two views cancels the translation component of the optical flow, leaving an estimate for the image centre.

Hysteresis

In a correctly operating and calibrated camera, feature translations are parallel and in the opposite direction of the pan or tilt motion. Tracking image features across equal pan steps produces a vertical distortion pattern as shown in Fig. 3. Contrary to what is suggested by [Lanz, 2004], the distortion pattern appears as a pincushion with distortion increasing with radial distance from the image centre. That is to imply, that after a cycle of motions, if the steering mechanism is precise, that the camera should servo to the same position. It has been observed that for the open-loop pan-tilt steering mechanisms this is not the case. Furthermore, it is possible find the displacement which produces the least curvature which must the vertical location of the principle point c_y . A similar method may be used to determine the horizontal location c_y by tracking points through tilt-steps.

Using the principle point determined by this method it is then possible to linearize the point matches and obtain the coefficients of the Taylor expansion $L\left(r\right) = l_0 + l_1 r + l_2 r$ [Lanz, 2004].

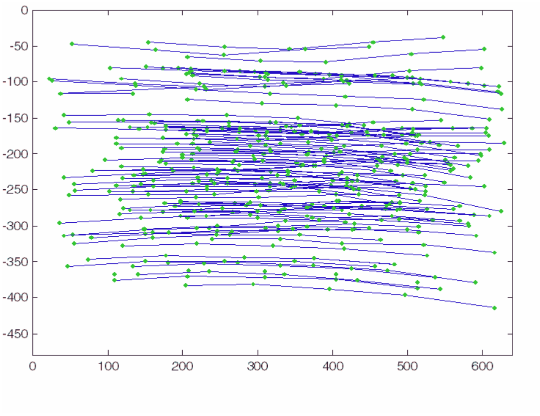


Figure 3: Vertical distortion pattern (blue lines) from feature matches (green dots) for a pan motion (units are in pixels).

Focal Length and Aspect Ratio

The focal length and pixel aspect ratio for a pan-tilt-zoom camera may be calculated at each zoom level by using the Homography relation between images of a rotating camera. The standard camera model is used, where x is 2D projective mapping of the 3D point **X**, and the camera projection matrix may be decomposed into the intrinsic matrix K and the extrinsic matrices **R**, rotation, and **t**, translation, as shown.

$$x = PX \tag{3}$$

$$P = K\left[R - Rt\right] \tag{4}$$

$$K = \begin{pmatrix} \alpha f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

Where α is the pixel aspect ratio, s is the pixel skew, f is the focal length and (p_x, p_y) is the principle point. To solve for **K** we exploit the homography relation which is calculated for a rotation as follows:

$$x' = Hx \tag{6}$$

From the camera model, with zero translation:

$$x = K\left[R\right]X, x' = K'\left[R'\right]X \tag{7}$$

The relation between x and x' is therefore written as:

$$x' = K'R'R^{-1}K^{-1}x \tag{8}$$

Defining the relative rotation between the two views as $R_{rel} = R'R^{(-1)}$ the homography H must be equal to the product below:

$$H = K' R_{rel} K^{-1} \tag{9}$$

Since both images are produced using the same camera, the intrinsic matrices K and K' are identical, the final relation used is:

$$H = K R_{rel} K^{-1} \tag{10}$$

The method used to calculate K is therefore to first capture a dataset grid of images at known pan and tilt steps. From the gazing angle differences the relative rotation between images i and j, $R_{rel_{i,j}}$ is calculated, while the homography, $H_{i,j}$ is obtained using a feature-matching algorithm based on RANSAC [Hartley and Zisserman, 2003]. To ensure that a valid homography may be found between images i and j only adjacent images in grid are processed.

Homographies obtained from the dataset frequently carry some noise due to low precision of point matches available. This noise is especially pronounced when the images captured are small in size or features are improperly matched. Applying basic outlier rejection produces a set of homographies which are more representative of the actual camera rotation between images.

Due to the absence of image registration at this point, the precision of rotations is not known and must be assumed to be adequately approximated by the steering angle. Differences in the homographies resulting from slight variations in the actual camera rotation steps are minimized by calculating an estimated mean homography ($\mathbf{H_{est}}$) from each type of pair-wise comparison considered. This may be achieved by re-projecting a set of points over the set of noisy homographies to determine maximum likelihood re-projected positions. Matches from the new point positions to the originals are now used to calculate a mean estimate homography by the same method as above.

There is another opportunity to obtain an estimate for the principle point here. Defining an image direction $\mathbf{d_{perp}}$ which is perpendicular to the flow from the expected rotation R_{rel}, a line of points in the direction $\mathbf{d_{perp}}$ is projected over $\mathbf{H_{est}}$. The distance between points and their projections has some component in the direction of rotation and some component in the direction $\mathbf{d_{perp}}$, which is assumed to be minimal for the nearest estimate of the principle point coordinate in the direction of $\mathbf{d_{perp}}$. Similarly performing a perpendicular rotation will yield the other coordinate.

Solving the homography relation for K is a process of minimizing reprojection error. This error is measured as the distance between a point projected over H_{est} and its approximation from $KR_{rel}K^{-1}$.This is posed as:

$$\varepsilon_i = [H_{est}]x_i - [KR_{rel}K^{-1}]x_i \tag{11}$$

With individual errors summed for all points compared. The sum of squared errors is minimized using a non linear least squares optimization to produce the parameters of $\mathbf{K}$.

3.2 Extrinsics

Extrinsic parameters, or the position and reference direction of the cameras, are estimated using a two prong approach. The first is to get an initial estimate (up to scale) from using the fundamental matrix $\mathbf{F}$, which can be calculated from point matches between two images (detailed in Ch. 11 of [Hartley and Zisserman, 2003]). The extrinsics are then determined by relating views from the two cameras and intrinsic matrices for both cameras ($\mathbf{K}$,$\mathbf{K'}$).

These estimates are then used in conjunction with a self-surveying method [Matsuoka et al., 2007] as this takes advantage of the motion in the scene. This extrinsic information is found using a variant of the structure from motion (SFM) algorithm. SFM algorithms typically use camera motion to recover static scene structure; however, reversing this approach allows for the computation of the static camera geometry from scene motion. The initial estimate from above produce an initial guess that SFM and bundle adjustment iteratively refine.

4 Camera Steering

The tracking camera may be steered manually by specifying gazing elevation and azimuth angles. Alternatively, it is possible to obtain a steering angle automatically from a tracking application which maintains the target at the centre of the field of view. The camera geometry is shown in Fig. 5 and Fig. 4.

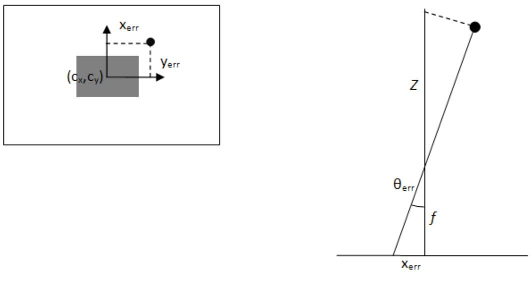


Figure 4: Camera steering control geometry

The error θ_{err} may be calculated individually for azimuth and elevation. The pinhole camera model demonstrates that the error angle for both axes is a function of the target displacement, input as the visible displacement on the image x_{err}, and the focal length f. The error angle is calculated about the focal point of the camera model for an idealized rotation-only camera. It is possible to calculate the error angle about the actual centre of rotation of a pan-tilt-zoom camera if it

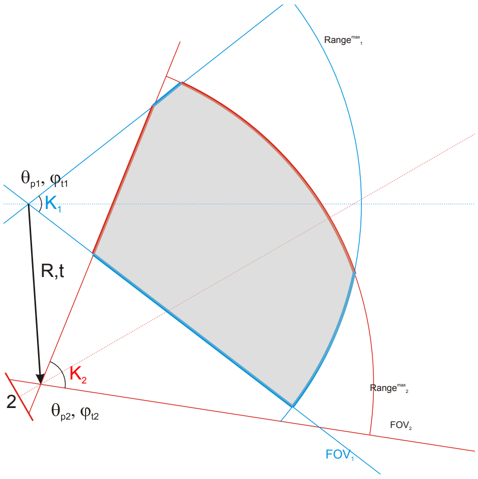


Figure 5: Camera view geometry (top view). This shows that the objective of the steering control is to have an overlap of the field of view.

is known using a more complex model such as that suggested by [Kanatani, 1988]. However, in practise this is not necessary because the distance to the target Z outstrips the distance from the focal point to the centre of rotation. We calculate the error angle as:

$$\theta_{err} = \tan^{-1}\left(\frac{x_{err}}{f}\right) \tag{12}$$

Camera gaze or steering regulation has to be performed with care as commodity security and/or personal systems have unregulated servoing (i.e., their positioning mechanisms operate in an open-loop manner). These systems often have delay from the driver (i.e., USB is an asynchronous bus) and the camera motion mechanism. While it is possible to monitor keypoints are perform visual servoing, this requires correspondences, which are not always available (especially with wide baselines and lighting variations). Thus, an adaptive (proportional) feed-back control approach is taken in which the proportional gain is varied based on the values obtained from the hysteresis estimation (see also Sec. 3.1).

A metric is constructed to quantize the necessity to turn to maintain tracking. Due to limitations in the hardware, turning the camera is costly in terms of tracking images lost due to blurring and loss of frame-rate during camera rotation. On better hardware this step may be omitted. The size of error in either principle direction is taken as an indicator of the need to turn, with a null zone specified in the central region where it is assumed there is no turning required to maintain visibility. This steering method is used regardless of the target identification method, specifically both motion detection and feature-matched tracking approaches may return an error metric in the correct form.

5 Experimental Validation

The PTZ steering was validated using two off-the-shelf pan-tilt USB cameras. The main result is illustrated in Fig. 6 in which a second "steered" camera automatically follows a master "tracking" camera. Additionally, the tracking camera has a SURF detector to aid with target labelling.

5.1 Calibration

Several experiments were conducted to validate the system detailed. The calibration method was tested by obtaining several sets of image panoramas for a number of cameras and performing the calibration methods described. The parameters obtained were validated by comparing to values obtained from the Matlab camera calibration toolbox [Bouguet, 2004] and applying to an actual steering scenario.

5.2 Estimation of the Image Center

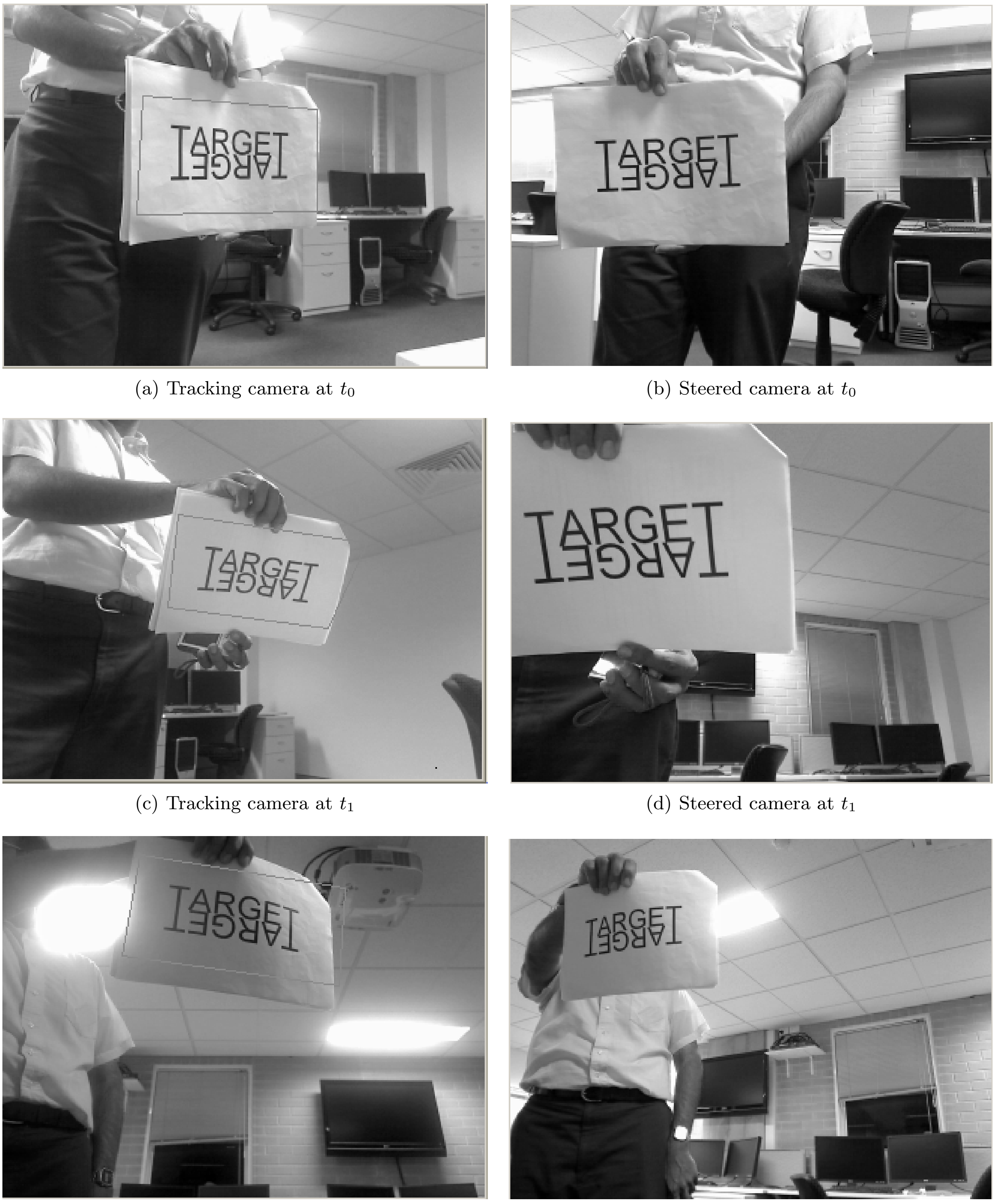
The image centre was estimated for two cameras using the homography reprojection technique described in the Intrinsics section. Horizontal and Vertical mean homographies were calculated from the pair-wise homographies and a perpendicular line was re-projected. Locations of minimal displacement in the perpendicular direction were selected as the image centre estimates.

The intrinsic camera parameters were calculated for two cameras using the method described. The homographies from both datasets were considered in each case as horizontal pairs, vertical pairs and diagonal pairs consisting of respectively a single step right, down and both. Example results from one of the cameras are shown in Table 1 (results are similar for the other cameras).

Estimated from:	f	p_x	p_y
Horizontal homographies	250.51	160.52	118.93
Vertical homographies	242.99	176.93	112.14
Diagonal homographies	284.98	91.30	155.74
All (H/V/D) homographies	254.06	153.91	123.22
Calibration Toolbox	270.09	161.30	122.89

Table 1: Auto-calibration comparison for intrinsic values

An auto-steering task was used to validate PTZ control and steering. To improve computation speed, the feature filter selected was SURF [Bay et al., 2008]. Feature matching was done by comparing SURF descriptors using a nearest neighbour algorithm. It was found that given the error signal obtained as the image location of the centre of the target, the camera can easily be steered correctly to maintain the moving target within the field of view of the camera.



(a) Tracking camera at t_0

(b) Steered camera at t_0

(c) Tracking camera at t_1

(d) Steered camera at t_1

(e) Tracking camera at t_2

(f) Steered camera at t_2

Figure 6: Results of target object tracking with the tracking camera (or master camera) tracking the object (using SURF features) and the steered camera (or slave camera) following such that it is focused on the same region in the workspace automatically. The steered camera was not calibrated *a priori*. The bounding box in the master camera tracking frame is added by SURF feature detector. Notice that the target remains in the scene even though the steering commands for the slave camera are different that that for the master camera

Two cameras were set up to confirm that feature based tracking through one camera may be used to steer the other toward the target. Once again from an error signal on the tracking camera, the pan and tilt steering required to track the target were applied to the camera and similar commands calculated for the driven camera. Although there was significant lag between steering commands applied to the tracking camera and the slave, this is most likely an implementation error arising from poor hardware handling.

5.3 Kinematics

The pan and tilt mechanism kinematics for the PTZ camera (Logitech Orbit) are given by:

$$ {}^{\text{tilt}}_{\text{pan}}\mathbf{R} = \begin{pmatrix} \cos(\theta_p) & \sin(\theta_p)\sin(\phi_t) & \cos(\phi_t)\sin(\theta_p) \\ 0 & \cos(\phi_t) & -\sin(\phi_t) \\ -\sin(\theta_p) & \cos(\theta_p)\sin(\phi_t) & \cos(\theta_p)\cos(\phi_t) \end{pmatrix} \tag{13} $$

However subsequent rotations have to be handled with care as pan is relative to the base and tilt is relative to a moving frame.

Due to motion imprecision the cameras frequently over or undershoot when moving. In this implementation, this is compensated through the use a threshold (or integral error) function. While the target remains within a particular margin of the desired location, the imperative to move is zero. Once it set-point crosses this, the imperative to move is the integral of the cost function over time. While it is possible to dynamically adjust this parameters (e.g., to handle high-frequency fluctuations), a consequence of this implementation is that tracking precision is lost as movement commands are issued only when there is large target movement.

6 Conclusions

This paper expands on self-surveying concepts developed [Matsuoka et al., 2007] by covering some of the deficiencies in the setup of the self-surveying camera array through techniques which exploit the functionality of a pan-tilt-zoom camera. The camera steering approach presented is able to effectively control an active array of pan-tilt-zoom (PTZ) cameras in a coordinated manner.

It does so by adapting homographic auto-calibration ideas to exploit PTZ motion. Further, the results are optimized for on-line control (with field of view overlap) more so than off-line metric reconstruction. Extrinsic parameters are estimated by extending self-surveying ideas that themselves are based on SFM and bundle adjustment. This is then combined with perspective geometry to generate control set-points that then are used to drive the camera motion.

7 Acknowledgement

This work is supported by the ARC Centre of Excellence program funded by the Australian Research Council (ARC) and the New South Wales State Government and the Rio Tinto Centre for Mine Automation. Part of this research effort was conducted as part of the AMME4111 Thesis course at the University of Sydney.

References

[Badri et al., 2007] Badri, J., Tilmant, C., Lavest, J., Pham, Q., and Sayd, P. (2007). Camera-to-Camera Mapping for Hybrid Pan-Tilt-Zoom Sensors Calibration. *Lecture Notes in Computer Science*, 4522:132.

[Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359. Camera calibration;Feature description;Interest points;Local features;Speeded-Up Robust Features (SURF);.

[Bouguet, 2004] Bouguet, J. (2004). Camera calibration toolbox for matlab.

[Civera et al., 2009] Civera, J., Bueno, D. R., Davison, A. J., and Montiel, J. M. M. (2009). Camera self-calibration for sequential bayesian structure from motion. In *2009 IEEE International Conference on Robotics and Automation (ICRA2009)*, Kobe International Conference Centre Kobe, Japan.

[Collins and Tsin, 1999] Collins, R. and Tsin, Y. (1999). Calibration of an outdoor active camera system. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 528–534. IEEE, LOS ALAMITOS, CA,(USA).

[de Agapito et al., 1999] de Agapito, L., Hartley, R., and Hayman, E. (1999). Linear self-calibration of a rotating and zooming camera. In *Proc. of the Conference on Computer Vision and Pattern Recognition*, volume 1, pages 15–21.

[Everts et al., 2007] Everts, I., Sebe, N., and Jones, G. (2007). Cooperative object tracking with multiple ptz cameras. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 323–330.

[Faugeras et al., 1992] Faugeras, O., Luong, Q., and Maybank, S. (1992). Camera self-calibration: Theory and experiments. In *Computer Vision–ECCV'92: Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19-22, 1992, Proceedings*, page 321. Springer.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision.* Cambridge Univ. Press.

[Kanatani, 1988] Kanatani, K. (1988). Transformation of optical flow by camera rotation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(2):131–143.

[Lanz, 2004] Lanz, O. (2004). Automatic lens distortion estimation for an active camera. In *International Conference on Computer Vision and Graphics*.

[Matsuoka et al., 2007] Matsuoka, M., Chen, A., Singh, S. P. N., Coates, A., Ng, A. Y., and Thrun, S. (2007). Autonomous helicopter tracking and localization using a self-surveying camera array. *The International Journal of Robotics Research*, 26(2):205–215.

[Sinha and Pollefeys, 2006] Sinha, S. N. and Pollefeys, M. (2006). Pan-tilt-zoom camera calibration and high-resolution mosaic generation. *Computer Vision and Image Understanding*, 103(3):170 – 183. Active camera networks;Image mosaicing;Pan-tilt-zoom camera calibration;Radial distortion;Radiometric alignment;Zoom calibration;.

[Tsuji et al., 2003] Tsuji, H., Shimokawa, K., Fujita, J., and Kawachi, N. (2003). MHI Robot Technology for Eye Vision Pan-Tilt. *Mitsubishi Juko Giho*, 40(5):274–277.

[Zhang, 2000] Zhang, Z. (2000). A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334.