Pan-Tilt-Zoom Camera Calibration and High-Resolution Mosaic Generation

Sudipta N. Sinha * and Marc Pollefeys

Department of Computer Science,
University of North Carolina at Chapel Hill,
CB # 3175, Sitterson Hall, Chapel Hill, NC 27599, USA.

Abstract

In this paper we discuss the problem of estimating parameters of a calibration model for active pan-tilt-zoom cameras. The variation of the intrinsic parameters of each camera over its full range of zoom settings is estimated through a two-step procedure. We first determine the intrinsic parameters at the camera's lowest zoom setting very accurately by capturing an extended panorama. The camera intrinsics and radial distortion parameters are then determined at discrete steps in a monotonically increasing zoom sequence that spans the full zoom range of the camera. Our model incorporates the variation of radial distortion with camera zoom. Both calibration phases are fully automatic and do not assume any knowledge of the scene structure. High-resolution calibrated panoramic mosaics are also computed during this process. These fully calibrated panoramas are represented as multi-resolution pyramids of cube-maps. We describe a hierarchical approach for building multiple levels of detail in panoramas, by aligning hundreds of images captured within a 1-12X zoom range. Results are shown from datasets captured from two types of pan-tilt-zoom cameras placed in an uncontrolled outdoor environment. The estimated camera intrinsics model along with the cube-maps provides a calibration reference for images captured on the fly by the active pan-tilt-zoom camera under operation making our approach promising for active camera network calibration.

Key words: Pan-Tilt-Zoom Camera Calibration, Active Camera Networks, Image Mosaicing, Zoom-Calibration, Radial Distortion, Radiometric Alignment.

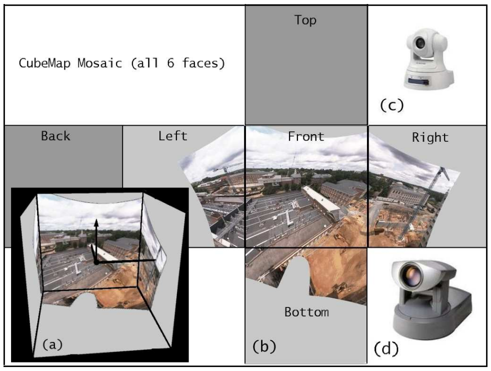
* Corresponding Author
Email addresses: ssinha@cs.unc.edu (Sudipta N. Sinha), marc@cs.unc.edu (Marc Pollefeys).

1 Introduction

The use of active pan-tilt-zoom (PTZ) cameras in wide-area surveillance systems, reduces the actual number of cameras required for monitoring a certain environment. During operation, each PTZ camera can act like a high-resolution omnidirectional sensor, which can potentially track activities over a large area and capture high-resolution imagery around the tracked objects. While omnidirectional cameras simultaneously observe a scene with a large field of view (FOV) often from a single viewpoint, they typically capture low-resolution images and have a limited range of scale. High-resolution panoramic cameras need specialized hardware and can be extremely expensive. However environments that are static or where events happen around a small region, do not require simultaneous imaging. For static scenes, multiple images captured over time can be aligned and composited into a complete panorama using image mosaicing algorithms [1,12,13]. PTZ cameras by virtue of their large zoom range can view a scene at a greater range of scale compared to an omnidirectional camera. At its finest scale, it can capture high-resolution imagery whereas a large range of pan and tilt gives it a large virtual FOV. Hence the PTZ camera combines the best of both worlds at an affordable cost.

A network of such active cameras could be used for 3D modeling of large scenes and reconstruction of events and activities within a large area, provided pixels captured from an active camera under operation could be accurately mapped to visual rays in 3D space. This paper describes a fully automatic method for calibrating such a model for pan-tilt-zoom cameras that does not require physical access to the cameras or the observed space. A model for the camera's intrinsic parameters is estimated from images captured within its full range of pan, tilt and zoom configurations. Our method is inherently feature-based, but does not require a calibration object or specific structures in the scene.

Past work on active camera calibration has mostly been done in a laboratory setup using calibration targets and LEDs or at least in a controlled environment. Some of these include active zoom lens calibration by Willson et. al. [14,9,15], self-calibration from purely rotating cameras by deAgapito [4], and more recently pan-tilt camera calibration by Davis et. al. [6]. Our approach towards zoom calibration is simpler than that of Wilson [15] who computed both focal length and radial distortion at many different zoom settings [15] and is similar to that of Collins et. al. [5], who calibrated a pan-tilt-zoom active camera system in an outdoor environment. However we extend the lens distortion model proposed by Collins [5] who assumed constant radial distortion, estimated it only at a particular zoom level and modeled its variation using a magnification factor. We actually estimate the radial distortion caused by optical zoom of the camera, the effect of which varies with camera zoom. Our method computes the intrinsic parameters of a PTZ camera from images taken by the rotating and zooming camera in an unknown scene. The intrinsic parameters at the lowest zoom are first computed by estimating homographies between



Feature	VB-C10	SNC-RZ30
Pan Angle Range	200.0^o	340.0^o
Tilt Angle Range	120.0^o	115.0^o
Zoom Range / Steps	$16X$ / 4450	$25X$ / 16000
fov at z_0	47.5^o	45.0^o
fov at z_{max}	3.0^o	2.0^o
Potential # pixels	1 Gpixels	3 Gpixels

Fig. 1. Cube-map Mosaic of 84 images computed from a PTZ Camera mounted on a building-top. (a) mapped on a cube.(b) 6 faces unfolded on a plane. PTZ cameras used in our experiements - (c) Sony SNC-RZ30 (d) Canon VB-C10. See table for specifications.

multiple images acquired by a rotating camera. Using bundle adjustment [11], the homography model is extended to take radial distortion into account and obtain a complete calibrated panorama of the scene with sub-pixel alignment error (see Figure 1). We next use an image sequence from the full zoom range of the camera to estimate the variation of its intrinsics with zoom. In our calibration model, an active PTZ camera is a virtual, static omnidirectional sensor. Next multiple images captured at increasing zoom levels are aligned to the calibrated panorama to generate a multi-resolution cube-map panorama.

We presented preliminary work on PTZ camera calibration and multi-resolution cube-map generation in [2] and [3] respectively; this paper contains a comprehensive description of the work and shows its importance for active PTZ camera calibration. We do not address the extrinsic calibration of a PTZ camera network; the methods for conventional camera network calibration as proposed in [16,17], could be extended to PTZ cameras. The paper is organised as follows. Section 2 introduces the camera model while Section 3 explains the calibration procedure in detail followed by experimental results. Section 4 addresses the construction of multi-resolution panoramas using a method that overlaps with the calibration algorithm. We conclude with discussions and scope for future work in Section 5.

2 Theory and Background

2.1 Camera Model

We chose to use a simple pan-tilt-zoom (PTZ) camera model and make a tradeoff for simplicity over exactness in our choice, similar to [4,5]. Our model assumes that the center of rotation of the camera is fixed and coincides with the camera's center of projection. More general camera models have been proposed [6,14] for PTZ cameras that violate this assumption. Davis and Chen [6] model the pan and tilt

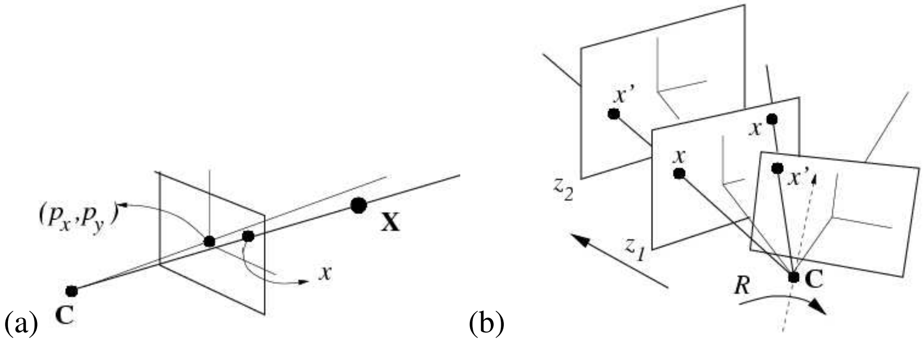


Fig. 2. (a) The pin-hole camera model. (b) Camera rotation and zoom.

rotations to be occuring about arbitrary axes in space; they estimate them in their calibration method. However when cameras are used outdoors or in large environments, the deviation of the center is negligible compared to the average distance of the observed features, which are typically distant. Our experiments with the Canon VB-C10 and Sony SNC-RZ30 surveillance cameras (refer to the table in Figure 1 for relevant specifications) have shown this assumption to be reasonably accurate.

In the pin-hole camera model (see Figure 2(a)) for the perspective camera, a point $\mathbf{X}$ in $3D$ projective space $\mathbf{P}^3$ projects to a point $\mathbf{x}$ on the $2D$ projective plane $\mathbf{P}^2$ (the image plane). This can be represented by a mapping $f : \mathbf{P}^3 \rightarrow \mathbf{P}^2$ such that $\mathbf{x} = \mathbf{PX}$, $\mathbf{P}$ being the 3×4 rank-3 camera projection matrix. This matrix $\mathbf{P}$ can be decomposed as shown in Eq. 1.

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \;\; -\mathbf{Rt}] \qquad \mathbf{K} = \begin{pmatrix} \alpha f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

where $\mathbf{K}$ represents the camera intrinsics while $\mathbf{R}$ and $\mathbf{t}$ represents the camera position and orientation with respect to the world coordinate system. The matrix $\mathbf{K}$ can be expressed in terms of α, s, f, p_x and p_y (see Eq. 1), where α and s are the camera's x:y pixel aspect ratio and skew respectively; f its focal length measured in pixel in the vertical direction; (p_x,p_y) its principal point in the image. Since we model the camera's pan and tilt movements by pure rotations about its projection center $\mathbf{C}$, we choose it as the world origin and set $\mathbf{t} = \mathbf{0}$. Our goal is to estimate the unknown parameters of a model for $\mathbf{K}^{p,t,z}$ that provides the intrinsics for any (pan= p; tilt= t; zoom= z) configuration within the admissible PTZ ranges of the camera. The principal point (p_x,p_y) and focal length f depend only on the camera zoom z. Thus they are denoted by f^z and (p_x^z,p_y^z) respectively. α and s, (we assume $s = 0$) are constants for a particular camera. Hence the unknown intrinsics we wish to estimate are of the following form.

$$\mathbf{K}^{p,t,z} = \mathbf{K}^{z} = \begin{pmatrix} \alpha f^{z} & s & p_x^z \\ 0 & f^{z} & p_y^z \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

Most cameras deviate from a real pin-hole model due to radial distortion; this effect decreases with increasing focal length. The $3D$ point $\mathbf{X}$ which in the pin-hole model projects to $\mathbf{x}$ = $[\tilde{x}\ \tilde{y}\ 1]^T$ actually gets imaged at (x_d,y_d) due to the radial distortion function $\boldsymbol{\mathcal{R}} : \boldsymbol{R^2} \rightarrow \boldsymbol{R^2}$ (see Eq. 3). $\tilde{r}$ is the radial distance of $\mathbf{x}$ from the center of distortion (x_c,y_c) and $\mathbf{L}(\tilde{r})$ is a radially symmetric distortion factor. κ_1 and κ_2 are the coefficients of radial distortion. The radial distortion model at zoom z, $\boldsymbol{\mathcal{R}}^z$ is parameterized by $(\kappa_1^z, \kappa_2^z, x_c^z, y_c^z)$, the respective parameters at zoom z. Based on properties observed by Wilson [14], we constrain the principal point (p_x^z, p_y^z) to be the same as the distortion center (x_c^z, y_c^z) in our camera model.

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = \mathbf{L}(\tilde{r}) \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}, \quad \tilde{r} = \sqrt{\tilde{x}^2 + \tilde{y}^2}, \;\; \mathbf{L}(r) = 1 + \kappa_1 r^2 + \kappa_2 r^4 \tag{3}$$

We determine calibration over the full zoom range by estimating $\mathbf{K}^z$ and $\boldsymbol{\mathcal{R}}^z$ at equal steps of zoom on a logarithmic scale, between z_0 and z_{max}, the minimum and maximum optical zoom levels respectively. Once all the parameters have been estimated at these discrete zoom levels, the complete intrinsics at any zoom can be obtained by piecewise linear interpolation.

2.2 Rotating and Zooming Cameras

Here we consider the case of a rotating and zooming camera. Let $\mathbf{x}$ and $\mathbf{x}'$ be the images of $\mathbf{X}$ taken at two different instants by a camera that is either zooming or rotating (see Figure 2(b)). These points, $\mathbf{x}$ and $\mathbf{x}'$ are related to $\mathbf{X}$ as $\mathbf{x} = \mathbf{K}[\mathbf{R}\ \mathbf{t}]\mathbf{X}$ and $\mathbf{x}' = \mathbf{K}'[\mathbf{R}'\ \mathbf{t}]\mathbf{X}$ where $\mathbf{t} = \mathbf{0}$. Hence $\mathbf{x}' = \mathbf{K}'\mathbf{R}'\mathbf{R}^{-1}\mathbf{K}^{-1}\mathbf{x}$. In our model, the intrinsics remain the same for pure rotation at constant zoom; hence this equation reduces to $\mathbf{x}' = \mathbf{K}\mathbf{R_{rel}}\mathbf{K}^{-1}\mathbf{x}$ where $\mathbf{R_{rel}} = \mathbf{R}'\mathbf{R}^{-1}$ represents the relative camera rotation about its projection center between the two views and $\mathbf{K}$ is the camera intrinsic matrix for that particular zoom level. Similarly for a camera zooming in a fixed direction with a fixed projection center, $\mathbf{x}' = \mathbf{K}'\mathbf{K}^{-1}\mathbf{x}$. These homographies are represented by $\mathbf{H_{rot}}$ and $\mathbf{H_{zoom}}$ (see Eq. 4).

$$\mathbf{H_{rot}} = \mathbf{K}\mathbf{R_{rel}}\mathbf{K}^{-1} \qquad \mathbf{H_{zoom}} = \mathbf{K}'\mathbf{K}^{-1} \tag{4}$$

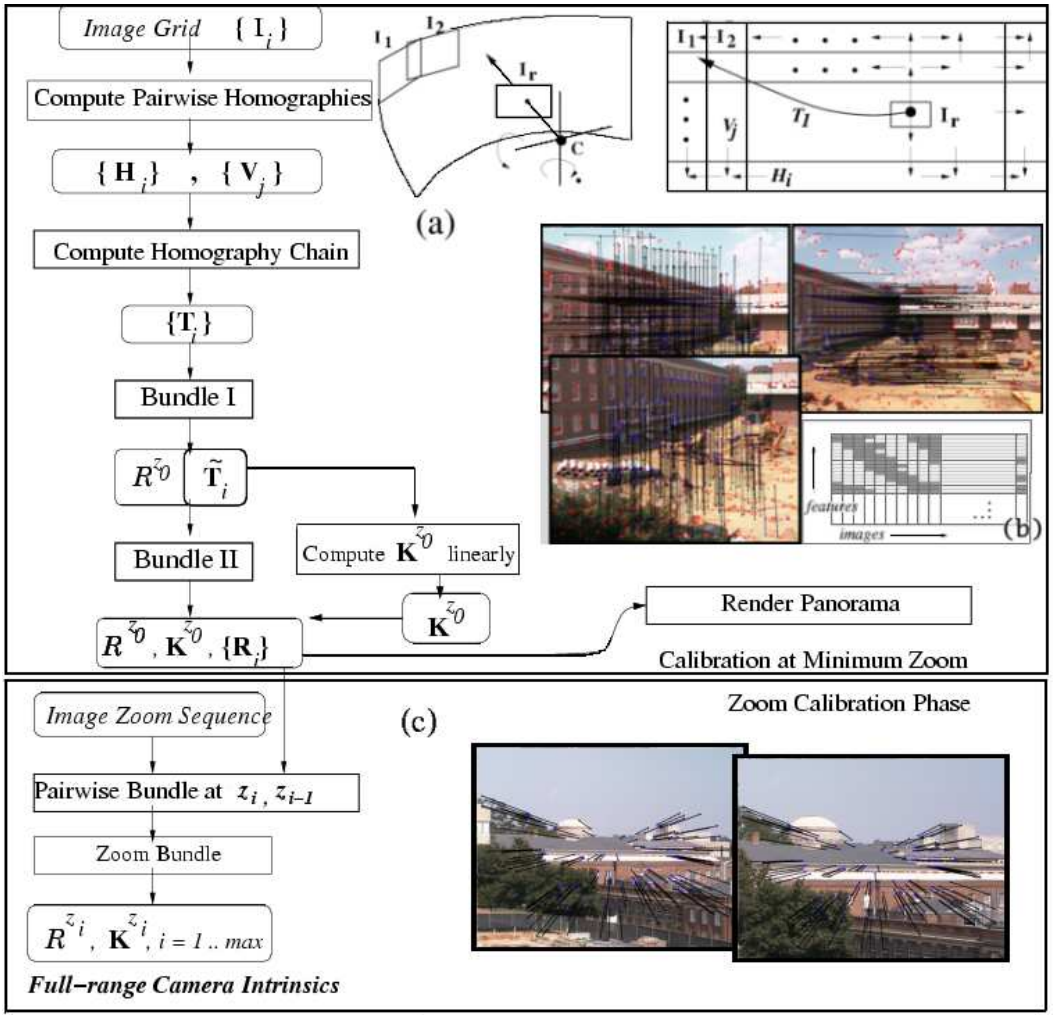


Fig. 3. Overview of calibration: (1) Intrinsics Calibration at minimum zoom. (2) Zoom Calibration. (a) Images $\{\mathbf{I}_i\}$ captured during rotation at fixed zoom and the mosaic computed with respect to $\mathbf{I}_r$. (b) Horizontally and vertically adjacent images in the grid shown with the corresponding matches. Feature lists built from features visible in multiple images are illustrated. (c) Successive images from a zoom sequence shown with the feature matches.

3 Camera Calibration

Our calibration algorithm works in two phases. The camera intrinsics are first estimated at the camera's minimum zoom and then computed for an increasing zoom sequence. Figure 3 gives a overview of the whole procedure. The notation used here is as follows. $\{\mathbf{I}_i\}$ are images acquired by the rotating camera. $\{\mathbf{H}_i\}$ and $\{\mathbf{V}_j\}$ represent the homographies between horizontal and vertical adjacent pairs in the image grid. $\{\mathbf{T}_i\}$ represents the homographies with respect to a reference image $\mathbf{I}_r$. The optimal homographies $\{\tilde{\mathbf{T}}_i\}$, computed through *Bundle I*, are used to obtain an approximate intrinsics $\mathbf{K}^{z_0}$ at the lowest zoom. $\{\tilde{\mathbf{R}}\}$ represents camera rotation matrices and $\boldsymbol{\mathcal{R}}^z$ stands for the radial distortion model at zoom level z.

3.1 Computing Intrinsics at Minimum Zoom

The first step towards calibration is determining the intrinsics at minimum zoom. This involves computing homographies $\mathbf{H}_{\mathbf{rot}}$ between images taken from a rotating camera (see Eq. 4). During a capture phase, images are acquired in a spherical grid (see Figure 3(a)) for certain discrete pan and tilt steps. Next the homographies between every adjacent horizontal pair of images, $\mathbf{H}_i$ and between every adjacent vertical pair, $\mathbf{V}_j$ in the grid are robustly computed. Harris corners extracted from these images are matched using normalized cross-correlation (NCC). This is followed by a RANSAC-based homography estimation and non-linear minimization. The details are described in [7] (Chap.3, page 108). Figure 3(b) shows a horizontal image pair and a vertical pair with the respective matched features. One of the images in the grid, $\mathbf{I}_r$ is chosen as the reference image and homography $\mathbf{T}_i$ is computed for every image $\mathbf{I}_i$, by composing a sequence of transformations, $(\cdots \mathbf{H}_a, \mathbf{H}_b \cdots \mathbf{V}_c, \mathbf{V}_d..)$ along a connected path between $\mathbf{I}_i$ to $\mathbf{I}_r$ in the image grid as illustrated in Fig. 3(a). An accurate estimate of $\mathbf{T}_i$'s for all the images would allow multi-image alignment in the image plane of $\mathbf{I}_r$. Since residual errors accumulate over the composed homographies, the final mosaic obtained by aligning all the images at this stage contains significant registration errors.

Global image alignment and sub-pixel registration is achieved using an efficient sparse implementation of bundle adjustment [7,11]. It is initialized from the homographies $\{\mathbf{T}_i\}$ and a global feature list (see Figure 3(b)), obtained from the pairwise matches. Bundle Adjustment performs global minimization which produces the maximum likelihood estimates of the model parameters when the reprojection error is assumed to be zero-mean Gaussian noise. The reprojection error is minimized; the homographies $\{\mathbf{T}_i\}$, the panorama features $\{\mathbf{X}^j\}$ (see Figure 4) and the radial distortion parameters $\boldsymbol{\mathcal{R}}^{z_0}$ are estimated by evaluating the expression:

$$arg\ min_{\{\mathbf{T}_i\},\mathcal{R}^{z_0},\{\mathbf{X}^j\}} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{D}(\mathbf{x}_i^j, \mathcal{R}^{z_0}(\mathbf{T}_i \mathbf{X}^j))^2 \tag{5}$$

$\mathbf{D}$ is the distance between image points; $\{\mathbf{x}_i^j\}$ are the observed features; m and n are the feature-count and image-count respectively. This is called *Bundle I* in Figures 3 and 4. The accurate homographies $\{\tilde{\mathbf{T}}_i\}$ computed here are used to estimate $\mathbf{K}^{z_0}$, the intrinsics at zoom z_0 using Hartley's method [8]. $\mathbf{K}^{z_0}$ is used to initialize *Bundle II*, which estimates $\mathbf{K}^{z_0}$, $\{\mathbf{R}_i\}$ and $\boldsymbol{\mathcal{R}}^{z_0}$ by evaluating the following expression:

$$arg\ min_{\mathbf{K}^{z_0},\{\mathbf{R}_i\},\mathcal{R}^{z_0},\{\mathbf{X}^j\}} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{D}(\mathbf{x}_i^j, \mathbf{K}^{z_0}.(\mathcal{R}^{z_0}(\mathbf{R}_i \mathbf{X}^j)))^2 \tag{6}$$

Every $\mathbf{X}^j$ (in $2D$ homogeneous coordinates) is parameterized as (a_j,b_j) where the third coordinate is +/-1 depending on which face of the unit-cube $\mathbf{X}^j$ projects to.

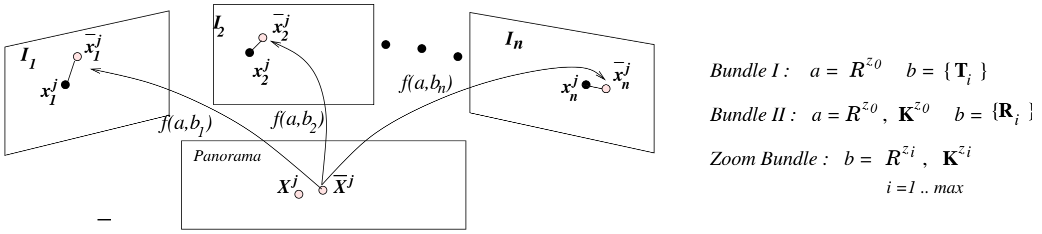

I_1
$\bar{x}_1^j$
x_1^j
I_2
$\bar{x}_2^j$
x_2^j
I_n
$\bar{x}_n^j$
x_n^j
$f(a,b_1)$
$f(a,b_2)$
$f(a,b_n)$
Panorama
X^j
$\bar{X}^j$
Bundle I : $a = R^{z0}$ $b = \{\mathbf{T}_i\}$
Bundle II : $a = R^{z0}$, $\mathbf{K}^{z0}$ $b = \{\mathbf{R}_i\}$
Zoom Bundle : $b = R^{zi}$, $\mathbf{K}^{zi}$
$i = 1 .. max$

Fig. 4. Bundle Adjustment: $f(a, b)$: parameters → measurements. a: common parameters, b: view-dependent parameters. The parameters for our different bundles (Sections 3.1, 3.2) are shown. Initial and optimal estimates of the panorama point are $\mathbf{X}^j$ and $\bar{\mathbf{X}}^j$ respectively.

3.2 Zoom Sequence Calibration

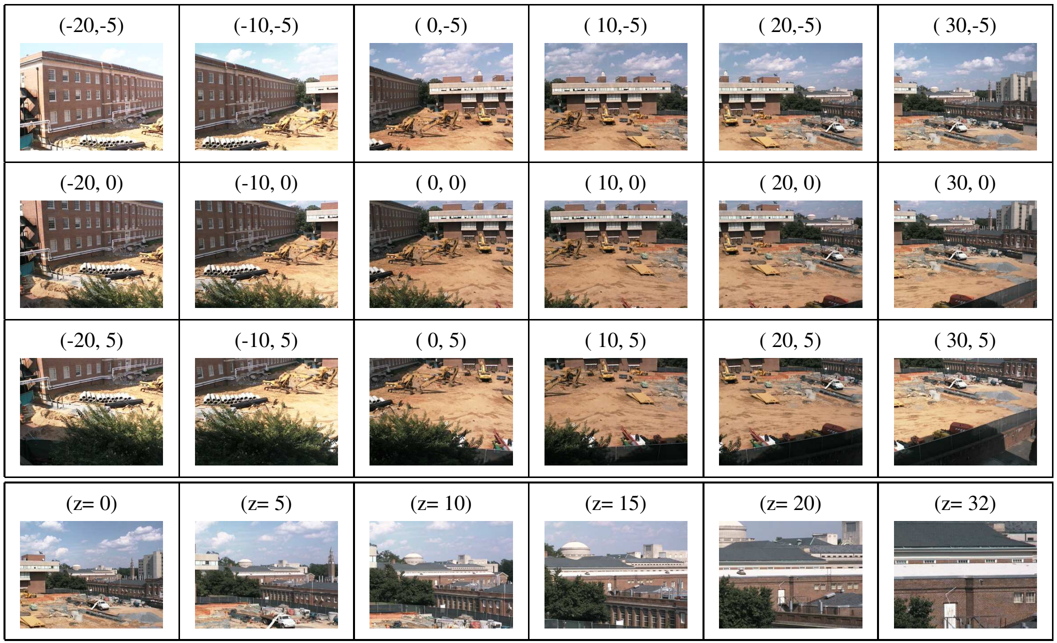
Full range zoom calibration can be done by building a mosaic and repeating the process described in Sec. 3.1 at multiple camera zoom levels. This would be slow since more images are needed as the camera zooms in and its effective field-of-view decreases [8]. Instead we calibrate an image sequence captured in a fixed direction with the camera progressively zooming in. First homographies $\{\mathbf{H_{zoom}}\}$,(see Eq. 4) are estimated between every image pair in this sequence. Next parameters $\mathbf{K}^{z_i}$ and $\boldsymbol{\mathcal{R}}^{z_i}$ at zoom z_i for every i is iteratively computed by a bundle adjustment on images at zoom steps z_i and z_{i-1}. This is possible since $\mathbf{K}^{z_{i-1}}$ and $\boldsymbol{\mathcal{R}}^{z_{i-1}}$ were computed in the previous iteration and $\mathbf{K}^{z_0}$ and $\boldsymbol{\mathcal{R}}^{z_0}$ were obtained from *Bundle II* (see Section 3.1). The following expression is evaluated at each zoom step z_i.

$$arg\ min_{\ \mathbf{K}^{z_i},\mathcal{R}^{z_i},\{\mathbf{X}^j\}} \sum_{j=1}^{m} \sum_{k=i-1}^{i} (\mathbf{D}(\mathbf{x}^j_{z_k}, \mathbf{K}^{z_k}\mathcal{R}^{z_k}(\mathbf{X}^j))^2) \tag{7}$$

The uncertainties associated with the estimates of κ_1 and κ_2 are propagated from zoom level z_i to the next level in this bundle. These uncertainty estimates are used to determine the zoom level at which the effect of each coefficient becomes negligible. Each coefficient is clamped to zero at that particular level and subsequent levels. A full bundle adjustment, *Zoom Bundle* (see Figure 4) then refines the calibration by minimizing the reprojection error over the whole zoom sequence by evaluating:

$$arg\ min_{\ \{\mathbf{K}^{z_i}\},\{\mathcal{R}^{z_i}\}\ \forall i=1\ ...\ max,\ \{\mathbf{X}^j\}} \sum_{j=1}^{m} \sum_{i=0}^{max} \mathbf{D}(\mathbf{x}^j_{z_i}, \mathbf{K}^{z_i}\mathcal{R}^{z_i}(\mathbf{X}^j))^2 \tag{8}$$

where m and max are the feature-count and image-count respectively. Estimating radial distortion only from a zoom sequence has inherent ambiguities since a distortion at a particular zoom can be compensated by a radial function at another zoom. We avoid this ambiguity by keeping the intrinsics computed at z^0 fixed.



(a)

Fig. 5. (a) Image Dataset used in Calibration : Top 3 rows: 18 images captured at a fixed zoom (z=0), pan and tilt angles (p,t) shown above the images are in degrees. The bottom row shows 6 frames from a zoom sequence of 36 images for a fixed pan and tilt angle.

3.3 Experimental Results

Here we present results from fully calibrating two Canon VB-C10 and two Sony SNC-RZ30 pan-tilt-zoom cameras in an outdoor environment. The cameras are placed near two adjacent windows about 3-4 meters apart looking out at a construction site roughly 100×120 meters in area. This setup reduced each camera's available field of view for pan to only 150^o. Hence only the front face of the cube-maps we build are interesting and hence shown. Figure 5 shows a few images from a pan-tilt and a zoom sequence respectively which are used in the calibration.

The recovered intrinsics for the four cameras as a function of zoom are shown in Figures 6 and 7. The principal point was found to move in a straight line for difference zoom sequences. The motion was most noticeable at high zooms. The VB-C10 had a linear mapping of focal length to zoom whereas the SNC-RZ30's focal length was non-linear. The pixel aspect ratio of the VB-C10's and SNC-RZ30's were found to be 1.09 and 0.973 respectively while the skew was assumed to be zero. Repeated zoom sequence calibration for the same camera from different datasets (Figure 6) showed the focal length estimation to be quite repeatable. The coefficients of radial distortion in our model, κ_1 and κ_2 were estimated along with their respective uncertainties (Figure 7). These uncertainties were used to clamp κ_1 and κ_2 to zero at particular zoom steps during *Zoom Bundle* (described in Sec 3.2). The mean reprojection error from the final *Zoom Bundle* for 35-40 images, with roughly 200-300 feature matches for every successive pair was within 0.43 pixels.

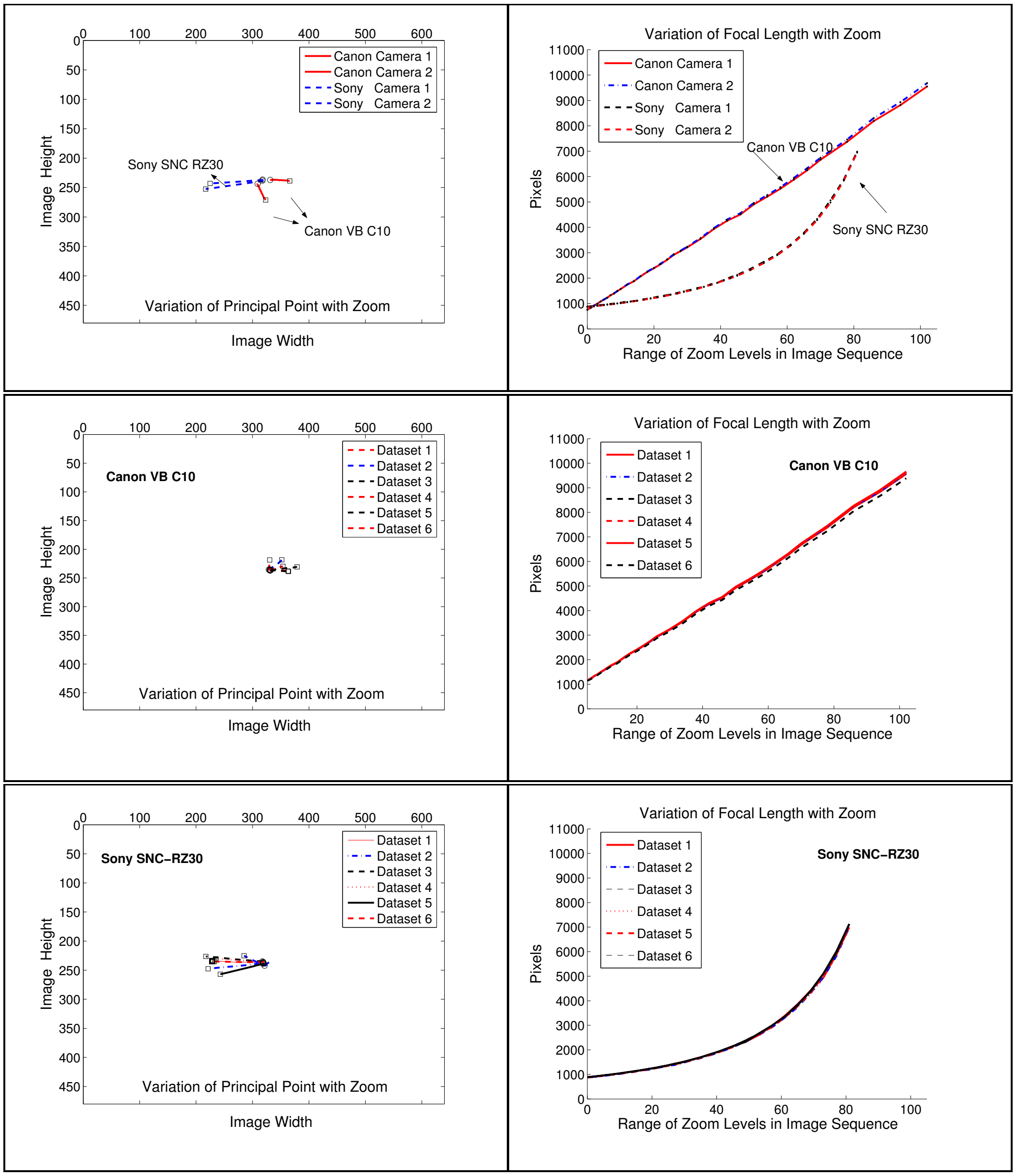


Fig. 6. (Top Row) The variation of the principal point and focal length with zoom is shown for each of the four cameras in our experiments. (Middle Row) Calibration results of a particular Canon VB-C10 camera from six different image sequences. (Bottom Row) Calibration results of a particular Sony SNC-RZ30 camera from six different images sequences.

4 High-resolution calibrated panoramic mosaics

Our approach described in Section 3, similar to that of [13] allows multi-image alignment with sub-pixel accuracy and creates high-resolution mosaics from images acquired by a rotating camera at fixed zoom (see Figures 1, 8). Since the unknown focal length f is computed during calibration, the cube-map face is chosen to be of size $2f \times 2f$, since this preserves the pixel resolution of the original images. Figure 8 shows a panorama with a single cube-map face at resolutions of 6k × 6k

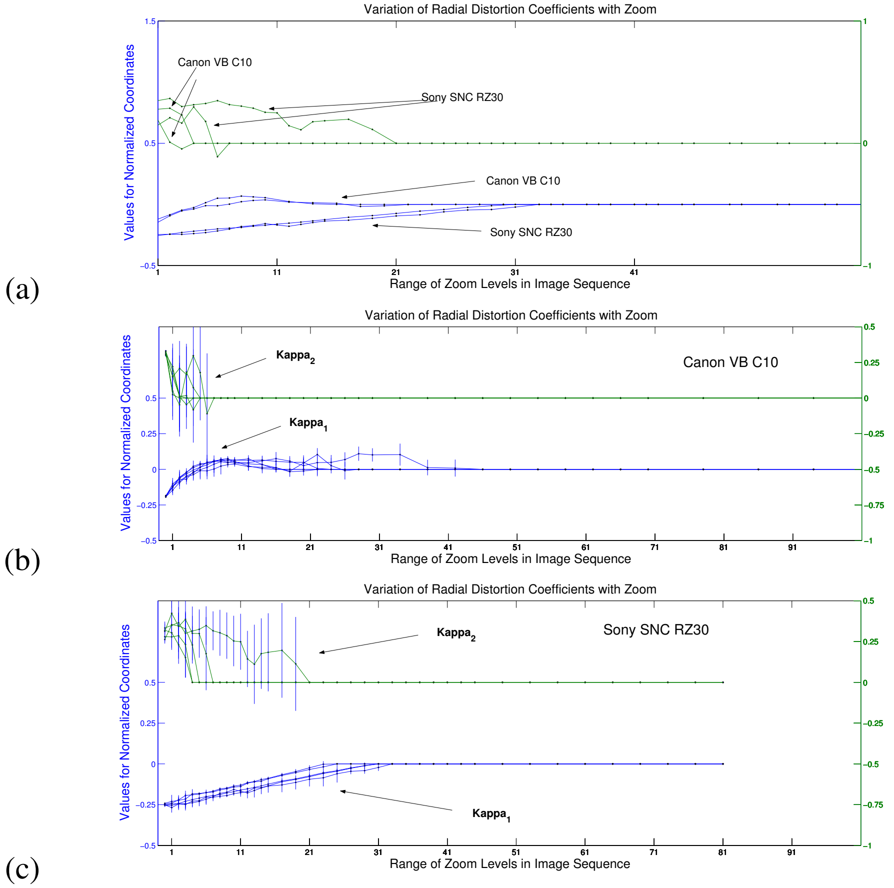


Fig. 7. (a) The variation of the radial distortion coefficients, κ_1 and κ_2, with zoom is shown for our 4 cameras. The uncertainties associated with κ_1 and κ_2 are shown using error bars. κ_2 has larger uncertainty compared to κ_1 which is estimated for a longer zoom range. (b) and (c) shows the variation of κ_1 and κ_2 for repeated calibration of a Canon and Sony camera, each using 6 different image sequences.

pixels, rendered from 119 images (f=3120 pixels approx. 5X zoom) in about 20-25 mins. We now describe a hierarchical approach aimed at capturing even greater detail, that utilizes the large zoom range and perform improved blending.

4.1 The multi-resolution approach

Conventional mosaic algorithm [12,13] would be infeasible for stitching hundreds of images all captured at high-zoom to build extremely high resolution full-view mosaics. For instance, assuming 50% overlap between adjacent images, the SNC-RZ30 must capture 21,600 images at 25X zoom (full FOV of 3.16π steradians) while the VB-C10 needs 7800 images at 16X zoom (full FOV of 2.55π steradians). By adopting a coarse to fine multi-resolution scheme, where images captured at a particular zoom are aligned with a mosaic built from images at half the present zoom, approximately half of the above image count would be needed at full zoom.

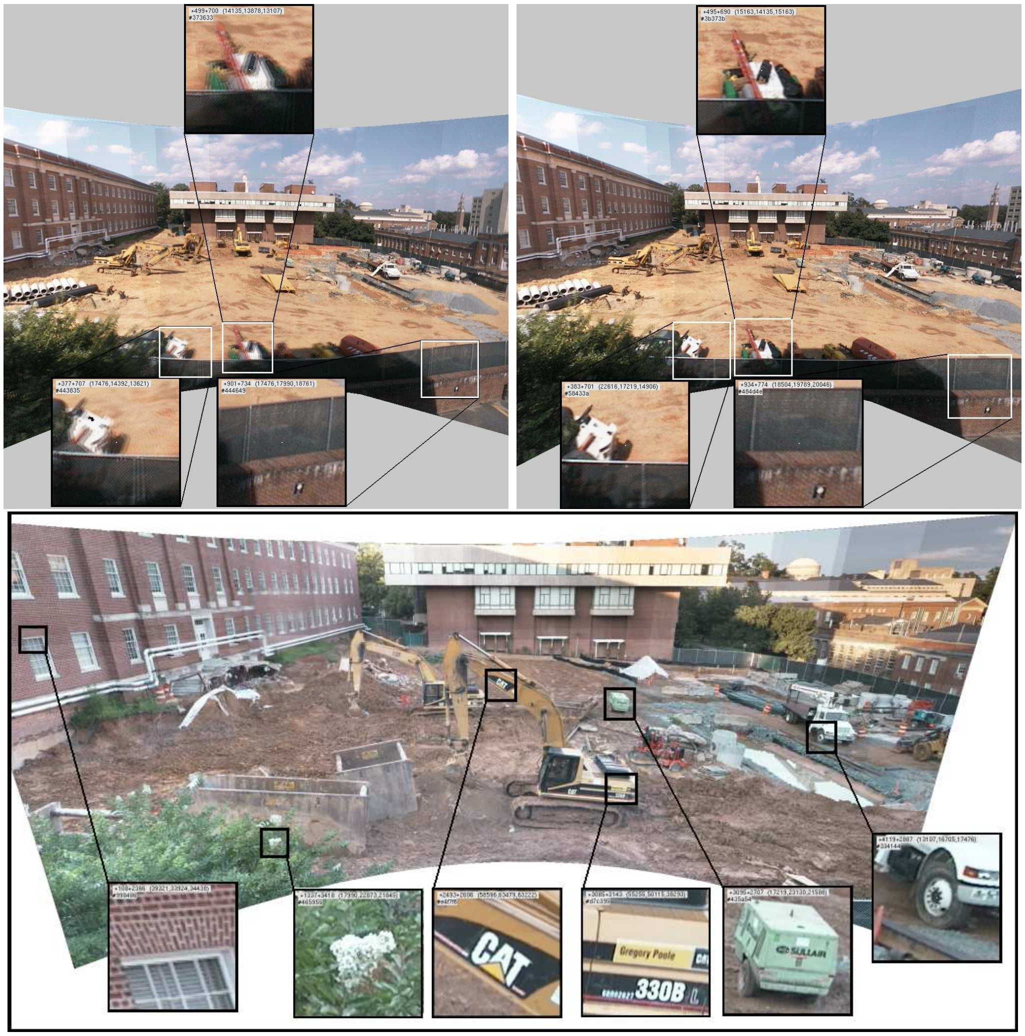


Fig. 8. The front face of the computed cube-maps: (a) Radial distortion was ignored in the camera model (Note that straight lines in the world are not imaged as straight lines). (b) Accurate panorama created after incorporating radial distortion. (c) Part of a high-resolution panorama (6000 × 6000 pixels) built from 119 images at 5X zoom. Note the zoomed-in regions of the panorama, displayed in the original scale.

The multi-resolution framework itself does require additonal images to be captured at intermediate zooms. However by using a top-down pruning scheme, we expect to reduce the number of images captured by avoiding high zoom in areas where detail is absent (see Section 4.4).

Figure 9(a) gives an overview of our approach. Phase I, deals with building the base cube-map $\mathbf{C}_0$ for the lowest zoom (this overlaps with the calibration procedure; see Section 3). Section 4.2 discusses photometric calibration. This allows a consistent blending of the base cube-map. Phase II outlined in Figure 9(b) involves building a cube-map, $\mathbf{C}_z$ of size $2N$ x $2N$ pixels from images captured at zoom level z using the cube-map $\mathbf{C}_{z-1}$ of size N x N computed previously from images at roughly half the zoom. Figure 9(c) summarizes the geometric and photometric alignment

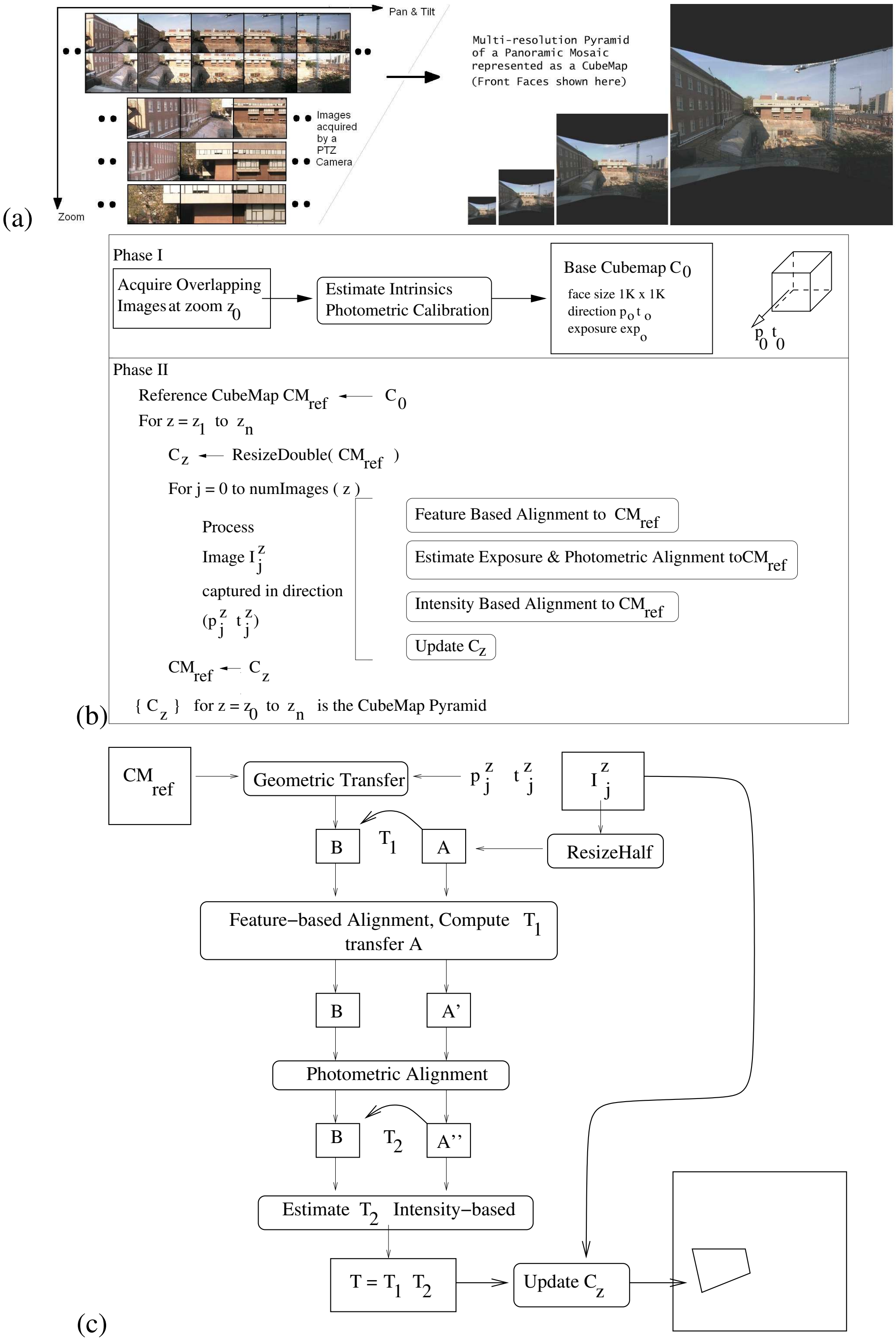

(a)
Pan & Tilt
Zoom
Images acquired by a PTZ Camera
Multi-resolution Pyramid of a Panoramic Mosaic represented as a CubeMap (Front Faces shown here)
(b)
Phase I
Acquire Overlapping Images at zoom z_0
Estimate Intrinsics Photometric Calibration
Base Cubemap C_0
face size 1K x 1K
direction $p_o t_o$
exposure exp_o
$p_0 t_0$
Phase II
Reference CubeMap $CM_{ref} \leftarrow C_0$
For $z = z_1$ to z_n
$C_z \leftarrow$ ResizeDouble(CM_{ref})
For j = 0 to numImages (z)
Process Image I_j^z captured in direction $(p_j^z \; t_j^z)$
Feature Based Alignment to CM_{ref}
Estimate Exposure & Photometric Alignment toCM_{ref}
Intensity Based Alignment to CM_{ref}
Update C_z
$CM_{ref} \leftarrow C_z$
$\{ C_z \}$ for $z = z_0$ to z_n is the CubeMap Pyramid
(c)
CM_{ref}
Geometric Transfer
$p_j^z \; t_j^z$
I_j^z
B
T_1
A
ResizeHalf
Feature−based Alignment, Compute T_1 transfer A
B
A'
Photometric Alignment
B
T_2
A''
Estimate T_2 Intensity−based
$T = T_1 \; T_2$
Update C_z

Fig. 9. (a) Overview of our multi-resolution method. (b) The two Phase approach in Coarse to Fine Cube-map pyramid construction. (c) The Image Alignment step during mosaicing.

step performed on each image. The recorded pan p_j^z, tilt t_j^z associated with every captured image $\mathbf{I}_j^z$ is used to generate an image from the calibrated cube-map $\mathbf{C}_{z-1}$ at half the resolution. For a perfectly repeatable camera, these two images denoted by $\mathbf{A}$ and $\mathbf{B}$ in Fig. 9(b) should be perfectly aligned. The SNC-RZ30 and VB-C10 however require additional alignment because of the inherent non-repeatability of the PTZ controls. The feature based method [7] (Chap.3, page 108) used during calibration (see Section 3), which is invariant to intensity changes in the two images, is used here too. Once the images are aligned, the exposure of this new image, $\mathbf{A}'$ is estimated (see Section 4.2). Once the cube-map at zoom level z is built, its becomes the base cube-map for the next level. Every level of the cube-map pyramid is initialized from the previous level by bilinear interpolation.

4.2 Robust Radiometric Calibration

The intrinsic calibration of a PTZ camera is extended here to include photometric calibration. The camera senses a high range of radiance (brightness intensity) in the scene while acquiring images in auto-exposure mode. Hence the captured images have different exposures and must be transformed to a common exposure before they can be blended into a single mosaic. The camera's response function is robustly estimated from the overlapping images captured at its lowest zoom in Phase I and the exposures of all the images are computed using the method described in [10]. This method works by estimating the brightness transfer function (BTF) between a pair of overlapping images by fitting a curve in the joint histogram space of the two images using dynamic programming. The camera response function is obtained from the BTF's by solving a least-square problem. The pixel correspondences required by this method are obtained from the accurate sub-pixel image alignment step described in Section 3. Once the camera's response function is known, the exposure of every subsequent zoomed-in image captured in Phase II can be estimated using the same method after registering it to the base cube-map (of known exposure). The results of blending the stitched images after photometric alignment is shown in Fig. 10(a,b).

4.3 Calibrated Panorama for Closed Loop Control of Active PTZ Camera

An open-loop calibration system for the PTZ camera which relies only on precise PTZ controls will be inaccurate in general due to errors from various sources; the camera's inherent non-repeatable controls, small changes in camera pose during operation or lack of stability of lens system at high zoom. To deal with such errors, a closed loop system using a pre-calibrated cube-map should be used. Figure 11 shows two examples of such a repeatable PTZ camera system. The current image from the PTZ Camera is aligned to an image generated from the cube-map. Reg-

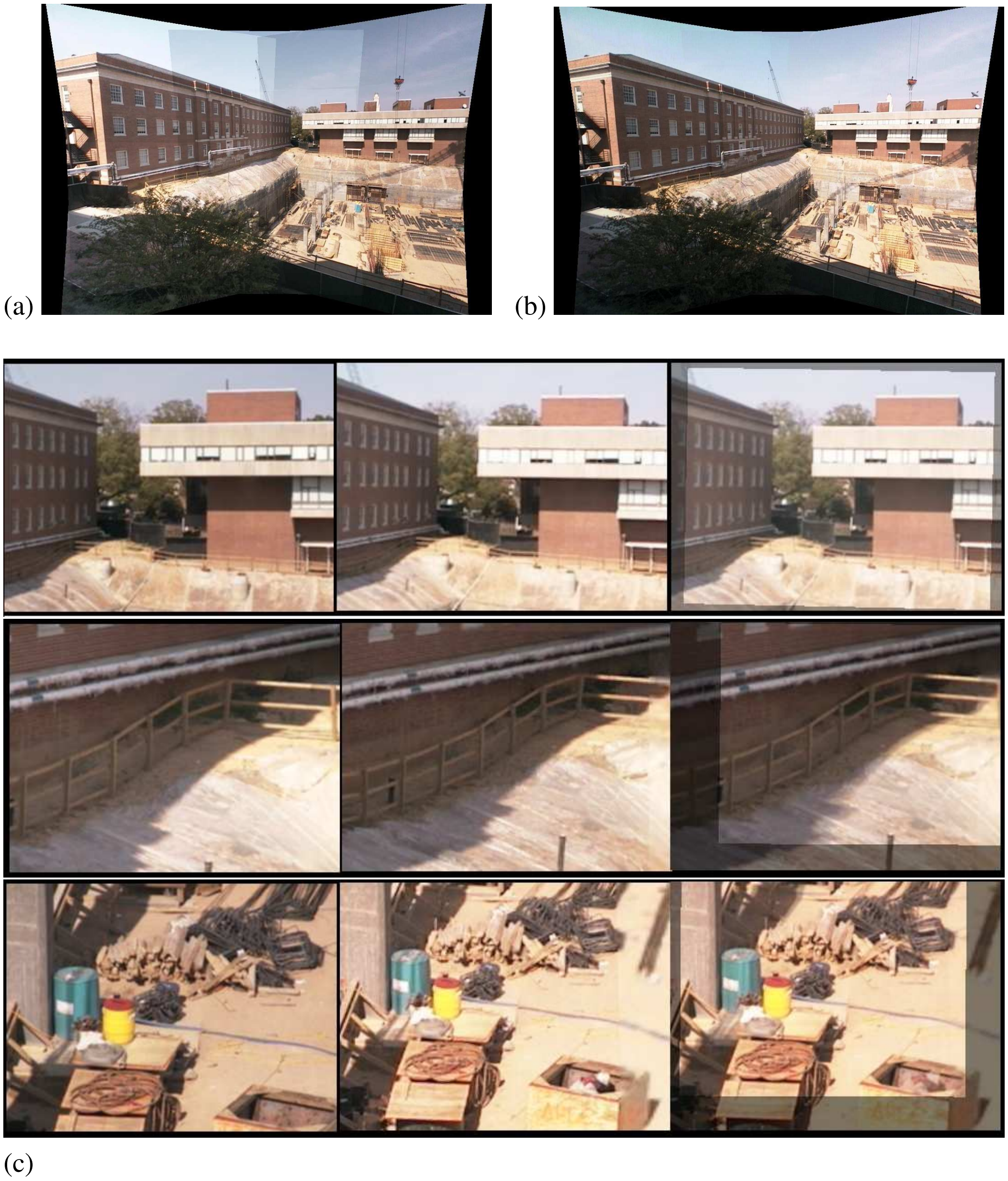


Fig. 10. Front face of a base cube-map rendered (a) without photometric alignment and (b) with photometric alignment. (c) Image Alignment: The 3 columns show the captured frame, the corresponding image generated from the cube-map and the aligned image pair (the first one overlaid on the second) respectively. In the middle row, despite moving shadows (images were taken far apart in time), the static parts of the scene are accurately aligned.

istration with the cube-map provides the ability of repeatably addressing the same pixel in the camera cube's reference from an active camera. In our current system, we use a robust feature-based image alignment algorithm which deals with illumination change as well as the presence of foreground objects missing in the precomputed mosaic. Since this misalignment error should be quite small for most zoom values, we could compute this local alignment in real-time as a simple 2-parameter RANSAC should be sufficient to compute this transformation (see Figure 11).

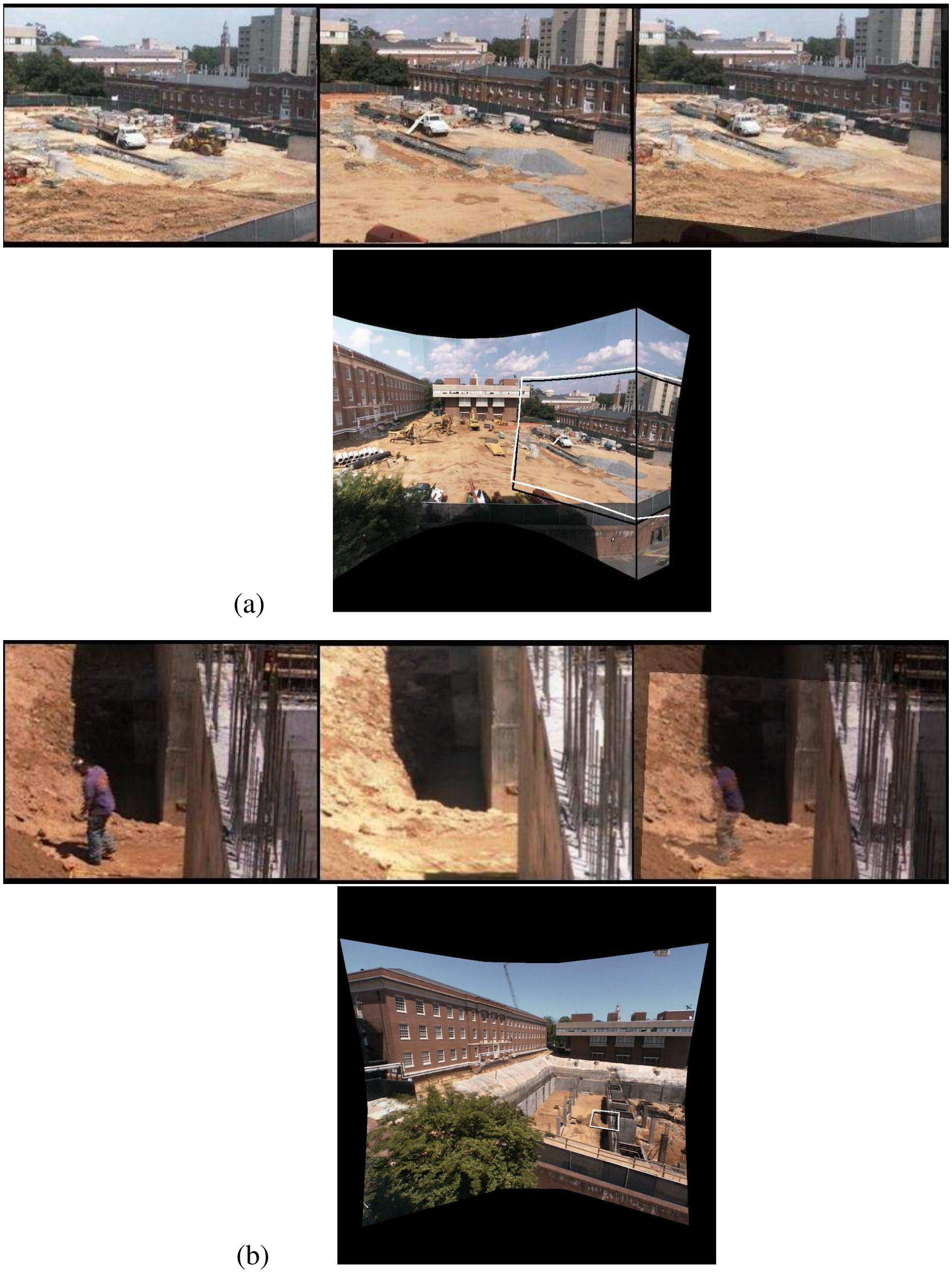


Fig. 11. Closed-loop Control for an Active PTZ Camera in operation using calibrated cube-maps : *(for both examples below)* (Left) Image seen by camera in configuration *(p,t,z)*, (Middle) Image generated from cube-map using the current configuration, (Right) The camera's frame aligned with the generated image. (Bottom) The camera's frame is super-imposed on the cube-map in white while the generated image is super-imposed in black. (a) Example 1 : Camera operating in a construction scene, 6 days after the cube-map was calibrated. Note the changes that have occured in the scene. (b) Example 2 : Camera zoomed in on a moving person under different illumination conditions.

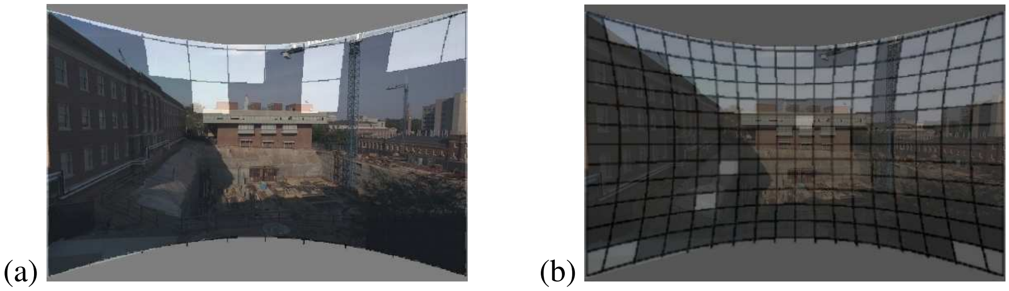


Fig. 12. Grey-shaded regions on the base mosaic (1X zoom), indicate where images were captured at zoom levels (a) 4X and (b) 8X respectively. Most of the sky was skipped.

4.4 Image Acquisition

The computational infeasibility of directly constructing a high resolution mosaic was described in Section 4.1. Building the mosaic pyramid in a coarse to fine fashion requires multiple acquistion passes, which captures the scene at a range of scales. This requires us to inspect images at a coarser scale (low zoom) to decide which parts of the scene contain detail. Often large portions of the scene contain textureless regions, for eg. the sky, walls, roads. We avoid zooming into such homogeneous regions and reduce the number of image acquired considerably.

To quickly acquire images in a scene, we do not wait to first build the calibrated base cube-map before subsequent passes at higher zoom. Instead an approximate calibration is used to backproject pixels into rays and effectively decide on the basis of texture analysis, whether the image at a specific PTZ value should be captured or skipped. An image block, where the eigen values of its second moment matrix are large, is mapped to a ray using the corresponding pan and tilt values, which is inserted into a kd-tree [1]. While scanning the scene in the next pass, a range query within this kd-tree returns a ray-count within the camera's FOV. Viewing directions corresponding to a low count contain mostly textureless regions in the scene. These images are skipped at the current and subsequent zoom levels. Our approach will miss texture present at finer scales which are filtered at coarser scales. However this allows us to directly acquire a compressed version of a very high resolution image instead of acquiring a raw image and then compressing it using lossy techniques. The result of pruning at two higher zoom levels is shown in Fig. 12.

4.5 Experimental Results

We built two cube-map pyramids, one each from images captured by a Sony SNC-RZ30 and a Canon VB-C10 camera placed outdoors looking at a construction site (see Fig. 13). The 1024 x 1024 pixel (face size) base cube-maps were built by stitching 15 and 9 overlapping images respectively. In each case the multi-resolution pyramids had five levels upto a resolution of 16K x 16K pixels. The camera cap-

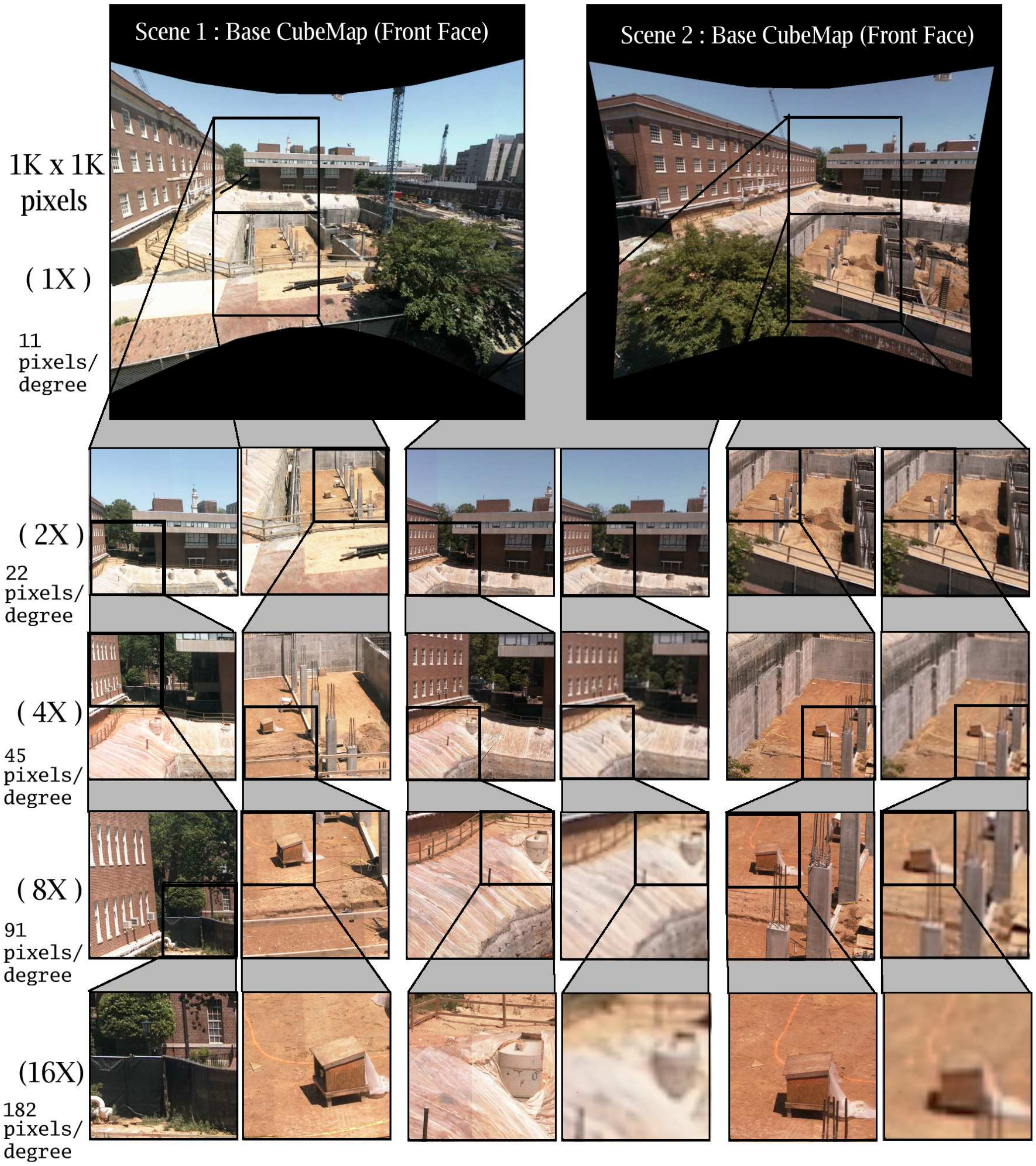


Fig. 13. Results: Two Cube-maps with 5 levels were built; each face had **1K**, **2K**, **4K**, **8K** and **16K** square pixels. Certain zoomed-in sections (512 x 512 actual image pixels) are shown above. Column 1 and 2 show the Levels of Detail for two parts of Scene 1. Column 3 and 5 show two parts of Scene 2 at different levels of detail. Compare the resolution with Column 4 and 6 showing the same view enlarged from the $1K$ x $1K$ base cube-map.

tured 15-20 images ie. 5-6.5 Mpixels at 3X zoom. About 70-95 images were captured at 6X zoom, which produced 21.5-29 Mpixels. Finally 300-350 images were captured at 12X zoom, out of which 200-250 were successfully aligned and hence contributed 62-77 Mpixels. These unique pixels in addition to the pixels interpolated from lower levels in the pyramid made up the faces of all the cube-maps. Scene1 and Scene2 (Fig. 13) were processed in 1-1.5 hrs on a 1.5 GHz Notebook Computer with 512 MB RAM. The original images were 640 x 480 pixels (1:10 compressed jpg) and were captured over the local ethernet. In our multi-scale cube-map pyramid construction, we used a tile-based representation for the large cube-map faces and processed them out-of-core using a image tile cache with FIFO block replacement strategy. This implementation is scalable and can potentially create gigapixel images for full-view panoramas by processing upto a few thousand images to build six full cube-map faces at $16K$ x $16K$ pixel resolution.

5 Conclusions

We have presented an automatic method for calibrating an active PTZ camera (typically used for surveillance) observing an unknown scene. The camera intrinsics are estimated over its full range of pan, tilt and zoom by robustly computing homographies between images acquired by a rotating and zooming camera. Our calibration algorithm also computes accurate calibrated panoramas at multiple levels of detail. We are currently working towards recovering the calibration of active cameras in operation. A lack of repeatability is being addressed by building an efficient closed-loop system, that re-estimates the calibration everytime the camera moves by robustly aligning a new image with the calibrated panorama (the calibration reference). This step works in the presence of new foreground objects and could be performed in real-time. This will allow the practical calibration of active PTZ camera networks for 3D modeling and other surveillance applications.

Acknowledgements

This work was partially supported by the NSF Career award IIS 0237533 and the DARPA/DOI award NBCH 1030015.

References

[1] M. Brown and D.G. Lowe. *Recognizing Panoramas.* In Proc. of ICCV 2003, Vol 1, pages 1218-1225, October 2003.

[2] S. N. Sinha and M. Pollefeys, *Towards Calibrating a pan-tilt-zoom camera network.* In OMNIVIS 2004, 5th Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras (in conjunction with ECCV 2004), Prague, Czech Republic.

[3] S. N. Sinha, M. Pollefeys and S. J. Kim, *High-Resolution Multiscale Panoramic Mosaics from Pan-Tilt-Zoom Cameras.* In Proc. of the 4th Indian Conference on Computer Vision, Graphics and Image Processing, pp. 28-33, ICVGIP Kolkata, India, Dec 16-18, 2004.

[4] L. de Agapito, R. Hartley, and E. Hayman. *Linear selfcalibration of a rotating and zooming camera.* In Proc. IEEE Int. Conf. CVPR99, pages 15–21, 1999.

[5] R.T. Collins and Y. Tsin. *Calibration of an outdoor active camera system.* In Proceedings of CVPR99, pages 528–534. IEEE Computer Society, June 1999.

[6] J. Davis and X. Chen, *Calibrating pan-tilt cameras in wide-area surveillance networks.* In Proc. of ICCV 2003, Vol 1, page 144-150, October 2003.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry In Computer Vision,* Cambridge University Press, 2000.

[8] R.I. Hartley. *Self-calibration of stationary cameras.* International Journal of Computer Vision, 22(1):5–23, 1997.

[9] B. J. Tordoff and D. W. Murray. *Violating rotating camera geometry: The effect of radial distortion on self-calibration.* In Proc. of ICPR, 2000.

[10] S.J. Kim and M. Pollefeys. *Radiometric Alignment of Image Sequences.* In Proceedings of CVPR04, pages 645-651. IEEE Computer Society, June 2004.

[11] B. Triggs, P. McLauchlan, R. Hartley, A. Fiztgibbon, *Bundle Adjustment: A Modern Synthesis*, In B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice, LNCS Vol.1883, pp.298-372, Springer-Verlag, 2000.

[12] H.Y. Shum and R. Szeliski, *Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment* in Proc of CVPR 2000, pages 101-130. IEEE Computer Society, 2000.

[13] H.S. Sawhney and R. Kumar, *True Multi-Image Alignment and Its Applications to Mosaicing and Lens Distortion Correction*, In Proc of CVPR97, pp. 450-456, 1997.

[14] R. Willson and S. Shafer, *What is the Center of the Image?* in Proc. IEEE Conf. of CVPR, pp. 670-671, 1993.

[15] R.G. Willson.*Modeling and Calibration of Automated Zoom Lenses.* PhD thesis, Carnegie Mellon University, 1994.

[16] S.N. Sinha and M. Pollefeys. *Camera Network Calibration from Dynamic Silhouettes,* In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition, pages 195-202. IEEE Computer Society, June 2004.

[17] M. Antone and S. Teller. *Scalable, Extrinsic Calibration of Omni-Directional Image Networks,* In the International Journal of Computer Vision, 49(2/3), pp. 143-174, Sep/Oct 2002.