UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Geometric and Radiometric Calibration for Pan-Tilt Surveillance Cameras

Ricardo Manuel Florindo Galego
(Licenciado)

Dissertação para a obtenção do Grau de Mestre em
Engenharia Electrotécnica e de Computadores

Júri:

Presidente:	Doutor Carlos Jorge Ferreira Silvestre
Orientador:	Doutor José António da Cruz Pinto Gaspar
Co-orientador:	Doutor Alexandre José Malheiro Bernardino
Vogal:	Doutor José Manuel Bioucas Dias

Junho de 2011

Contents

Resumo

A auto-calibração geométrica de câmaras pan-tilt é um problema bem conhecido para o qual já existem soluções robustas propostas na literatura. Por oposição, a fusão de níveis de brilho (cores) de pixeis correspondentes em imagens distintas com sobreposições, tendo em vista a construção de cenários panorâmicos (mosaicos), apesar de também ser um problema muito abordado, continua a merecer investigação tanto ao nível da modelização do processo de radiometria como da procura de soluções práticas e robustas.

Neste trabalho, é proposta uma metodologia para a calibração de efeitos radiométricos inerente à operação de câmaras pan-tilt, para aplicações de vídeo vigilância baseadas em representações cúbicas (mosaicos) dos cenários. A calibração radiométrica é baseada na estimação da distorção na imagem, causada por vignetting, usando os graus de liberdade proporcionados pelo pan-tilt da câmara em vez de padrões de calibração. Experiências com imagens reais mostram que a calibração geométrica reduz a variância nas representações de background, permitindo uma deteção de eventos mais eficaz em algoritmos baseados em background-subtraction.

Palavras chave: Câmara pan tilt zoom, formação da imagem, correção de vignetting, modelo radiométrico, calibração geométrica, deteção de eventos, sistemas de segurança

Abstract

It is a well know result that the geometry of pan and tilt (perspective) cameras auto-calibrate using just the image information. However, applications based on panoramic background representations must also compensate for radiometric effects due to camera motion.

In this dissertation we propose a methodology for calibrating the radiometric effects inherent in the operation of pan-tilt cameras, with applications to visual surveillance in a cube (mosaicked) visual field representation. The radiometric calibration is based on the estimation of vignetting image distortion using the pan and tilt degrees of freedom instead of color calibrating patterns. Experiments with real images show that radiometric calibration reduce the variance in the background representation allowing for more effective event detection in background-subtraction-based algorithms.

Keywords: Pan and Tilt Zoom Camera, Image formation, Vignetting correction, Radiometry model, Geometric calibration, Visual event detection, Surveillance.

List of Figures

List of Tables

Chapter 1

Introduction

Modern video surveillance systems are mostly composed by fixed cameras. Although patent the great potential of using the pan-tilt cameras, for instance the flexibility of selecting on line the fields-of-view, one finds specific issues and challenges inhibiting the widespread usage of these cameras. One of the issues is the slightly increased prices of the pan-tilt cameras. But this is becoming of little relevance as in fact technology evolution is making the prices very competitive. It is actually on the level of the applications that one finds the important issues and challenges.

A basic task of fixed surveillance cameras is the detection of changes on the background which may indicate intrusions. It is already a hard task considering for example the daylight change. Daylight change is still motivating much research, but one finds by now many commercial systems performing on line adaptation of the background model and therefore finding in essence just the relevant background changes.

On pan-tilt cameras, the self-motion is in fact an aspect of much more concern, and challenge, in real applications. Self-motion is a source of global image change that makes harder the task of modeling the background to do event detection based on background subtraction. Considering however that usually one knows the pan and tilt rotation angles, and that one could save the background at all possible pan and tilt orientations, then one has once more a background image to compare with a currently (run time) acquired image. Mosaic based representations come to help here, as a way to avoid saving all images, which would mean a too much large memory requirement.

A mosaic based representation allows saving only once the brightness (color) of a single scene point while considering all the possible images created by all the possible pan-tilt combinations. Many pan-tilt cameras can however exceed a $180^o \times 180^o$ field of view, some actually reach close to $360^o \times 360^o$, making therefore planar mosaics not possible. Choosing a geometric support (geometric representation) for the mosaic is therefore a relevant question.

There are various ways to represent geometrically the background [5, 28]. In this dissertation we use the cube based representation as it allows a complete $360^o \times 360^o$ field-of-view with simple homography transformations. Defined the representation, background differencing can then be used to find intrusions (events), provided one has a good characterization of the uncertainty of the background model. There are two main sources of uncertainty: inaccurate knowledge of the geometry of the camera and poor characterization of the conversion of scene radiances into image brightnesses. The geometric uncertainty, caused by uncertainty in the intrinsic parameters of the camera and the pan and tilt angles, is tackled by calibration. Radiometric uncertainty is mainly due to the nonlinearity of the radiometric response function and to vignetting, a decreasing gain for increasing radial distances in an image

[20, 30]. In this thesis we tackle both the uncertainty sources.

1.1 Related Work

Geometric and radiometric calibration, are two aspects largely studied. Considering the geometric calibration, there are various methods to calculate the intrinsics parameters of a camera. The work of Hartley[15] is one of the most well known for the geometric calibration of a pin-hole camera that rotates about a fixed position. Hartley introduced the infinite homography concept to help estimating the intrinsics of a fixed but rotating camera, without knowing its rotation (odometry), using three overlapping images. Following Hartley, Agapito et al. proved that the intrinsics can also be estimated for a rotating camera with varying intrinsic parameters (zoom) [2, 1] exploring the characteristics of the dual of the image of the absolute conic (DIAC). Zhang estimates the intrinsics of a camera as well as its radial distortion [31], imaging a planar chess pattern placed at multiple unknown orientations. Brown and Lowe [6] build a panorama, without knowing the intrinsic parameters of the camera or its direction, using SIFT features to calculate the homographies between images. Sinha and Pollefeys apply similar concepts to estimate how to stitch images acquired by a system of multiple pan-tilt cameras and therefore build a panorama in a collaborative manner [29]. While the geometric relationship of images is a subject already well understood, the radiometric stitching, or blending, is still a subject of much research.

Given the correction of the geometry of the camera it is possible to build a panorama but one usually perceives large discontinuities at the borders of the overlapped images. A possible way to stitch images together is *feathering*, a technique proposed quite early that smooths images at the stitching borders using e.g. image pyramids [3]. An improved manner, proposed in [21], was based on minimizing a cost function designed in the gradient domain. Eden et al. developed a way to stitch images together with a large variety of exposures, scene motion and misregistration [8]. More recently there was also developed a way to stitch images with structure distortion [16]. Although these methods render visually appealing mosaics, they are not well suited for surveillance applications, such as the ones based on background subtraction where one needs to have a good similarity between background and run-time images so differences effectively correspond to the intrusions that one wants to detect.

Hence, many researchers invested in understanding the physical reasons for the differences found at the image stitching seams. Stitching methodologies started to encompass radiometric calibration to estimate the radiometric response and vignetting functions of a camera. A camera response model, introduced by Grossberg and Nayar, was based on a large database of response functions obtained from well controlled illumination and color pattern setups [13, 14].

A. Litvinov and Y. Schechner simultaneously estimate the radiometric response and vignetting, based on a frame sequence acquired during camera motion, they also show that is not possible to uniquely solve for a vignetting function without applying additional constraints, when the radiometric response curve is unknown [24]. Goldman and Chen [11] and Goldman [12] show that only exist two types of fundamental photometric ambiguities, the scale and gamma. They also show that vignetting and exposure variations can be removed from the images without resolving these ambiguities.

Kim and Pollefeys estimate the radiometric response function, the exposure and white balance between images from a sequence of images [19]. They proposed to solve radiometric and vignetting response functions for high dynamic range mosaics, from a set of images with different and unknown exposures values [20]. In both cases it is needed to change exposure between images to calculate the response function. Pollefeys et al. [18] also propose

solving radiometric response function and at the same time joint feature tracking.

Lin et al. propose to solve the radiometric response function from images without changes in the exposure, using histograms of the edges regions [22, 23]. Another method proposed by Lin and Matsushita [26] uses the noise distribution to estimate the radiometric response function. However, vignetting is not considered in the estimation of the radiometry. It is also proposed by Lin et al. the correction of vignetting from a single image [33, 32]. However, their method requires segmenting large piecewise flat regions in the image, which is highly dependent on the scene contents.

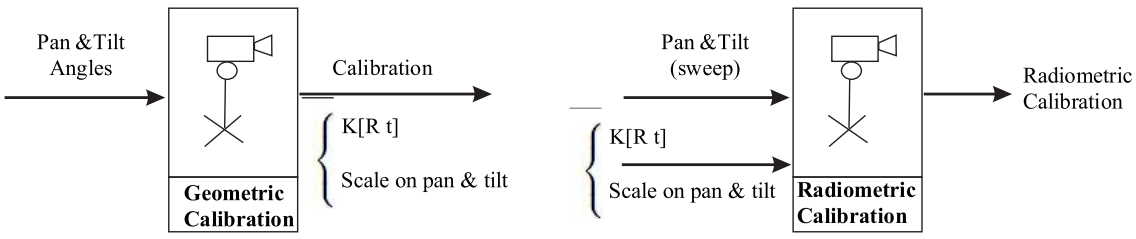
Alternatively, Wonpil Yu proposed to correct vignetting based on a white pattern [30]. The white image, decreasing in brightness towards the borders due to a vignetting distortion function, was approximated with a $2D$ hypercosine function. This is however cumbersome due to the requirement of using very large patterns when the cameras to calibrate are far away, e.g. outdoors at a second level floor, as it is usual with surveillance cameras.

Our approach is briefly described in the next section.

1.2 Problem Formulation and Approach

The work described in this thesis focus on applying pan-tilt cameras to automatic surveillance systems. Surveillance with pan-tilt cameras involves various research and development subjects such as controlling the pan and tilt angles, processing the video stream, modeling the background, etc. Camera geometry and radiometry are key aspects to all the referred subjects, and thus, their modeling and calibration form the central part of the work of this thesis. Concisely, our work comprises three main aspects: modeling and calibrating the pan-tilt camera geometry, modeling and calibrating the camera radiometry, and generating background models.

Geometric and radiometric calibration are briefly described in the Fig 1.1. The input of the geometric calibration consists of images acquired at various camera orientations. The pan and tilt, commanding and odometry readings, may also be used in the calibration. The output of the geometric calibration are the intrinsic parameters of the camera, K, and a scaling on the pan and tilt angle values in the case that the pan and tilt angles were not input in metric units. The radiometric calibration starts by sweeping the range of valid pan and tilt angles, and capturing images of a static scene. Given the pan and tilt odometry and the camera intrinsic parameters, one observes that the same 3D scene point is observed with various brightness values, depending on the camera orientation. The output of the radiometric calibration consists of parameters that tune a radiometric correction function.



(a) Geometric Calibration (b) Radiometric Calibration

Figure 1.1: Calibration setup.

The third main aspect is building the background representation. Various manners can be used to represent geometrically the background e.g. planar mosaic, a cylinder, a sphere or a cube [5, 29]. Figure 1.2 shows the

process of building a background representation. To build the background it is required to know the pose of the camera, the camera intrinsics parameters and to correct radiometrically the captured images.

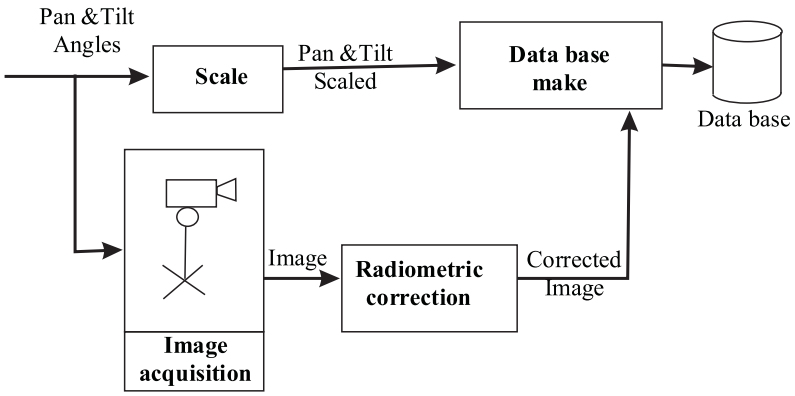


Figure 1.2: Building a background.

Having the background built one can compare the information of the background representation with the one captured by the camera. This comparison can be done trough background subtraction. Surveillance can be done as it is shown on figure 1.3.

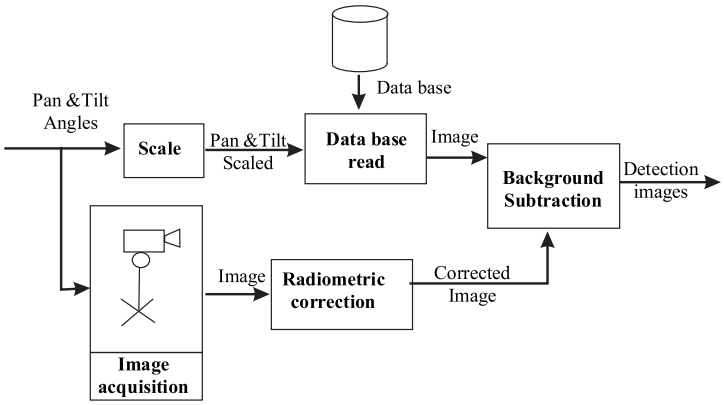


Figure 1.3: Background subtraction between images acquired by a camera and background representation.

1.3 Thesis Structure

Chapter 1 introduces the problem to approach in the thesis, and presents a short discussion on the state of the art. Chapter 2 presents the panoramic scene representation, more specifically the Pan-tilt Camera model, the cube base representation and several algorithms of geometric calibration. Chapter 3 describes the radiometric model of a camera as well as the event detection algorithm used. Chapter 4 provides an overview of the different experiments performed to test the proposed methodologies. Chapter 5 summarizes the work performed and highlights the main achievements in this work. Moreover, this chapter proposes further work to extend the activities described in this document.

1.4 Publications

The work described hereafter was partially published in Recpad [9] and in VISAPP [10]. The work published in Recpad focus mainly on modeling the vignetting and its correction. The VISAPP paper complements the previous work with a probabilistic view of the benefits of vignetting correction.

Chapter 2

Panoramic Scene Representation

In this chapter we describe the geometric model of a pan-tilt camera and propose using a cube based representation for the background. After introducing the geometric model, we describe various calibration methods based on calibration chess patterns, tracked (matched) image features and/or pan-tilt odometry.

2.1 Pan-tilt Camera

A pan-tilt camera combines a perspective camera with pan and tilt degrees of freedom. Geometrically, it is characterized by a projection model having a static projection center position and a varying orientation that allows surveying the scene. The projection model is therefore represented by

$$m \sim PM = K[R \mid t]M, \tag{2.1}$$

where $M = [x\ y\ z\ 1]^T$ and $m = [u\ v\ 1]^T$ are 3D world and 2D image points, $\sim$ denotes equality up to a scale factor, and $P = K[R\,t]$ is a 3x4 matrix projection matrix, composed by the intrinsic parameters matrix,

$$K = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.2}$$

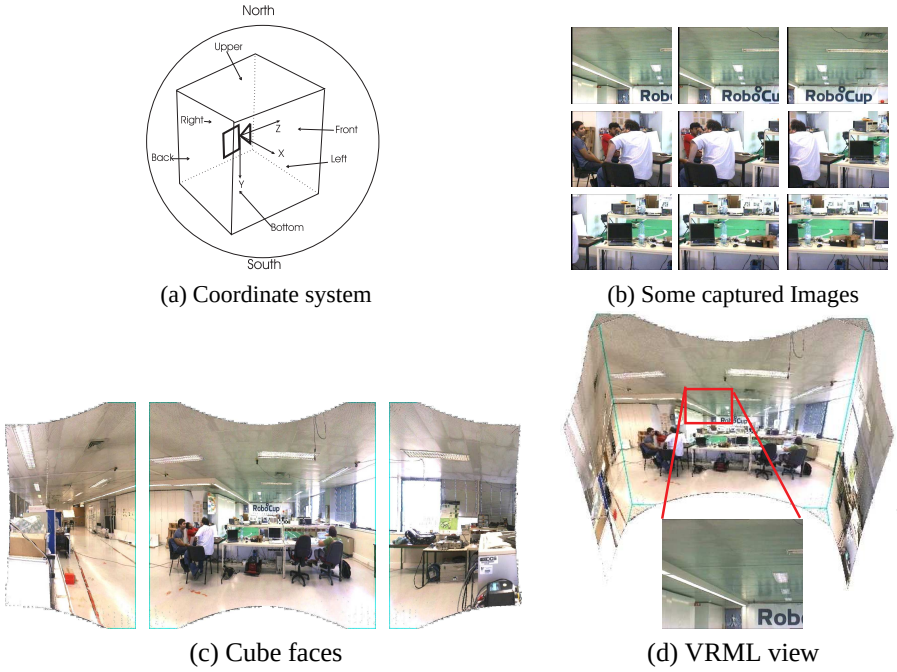
and the extrinsic parameters, R and t, representing the orientation and position of the camera with respect to the world coordinate system. The intrinsics parameters matrix is defined by five parameters. The parameters (u_0, v_0) denote the coordinates of the principal point in pixel units, f_u and f_v are horizontal and vertical focal lengths i.e. scale factors relating pixels to distances, and s represents the skew coefficient between the u and the v axis, in our case we consider it zero.

Without loss of generality we assume that the projection center is at the origin, $t = 0_3 = [0\ 0\ 0]^T$. The rotation

matrix, R depends on the pan and tilt angles, (α, β), i.e. $R = f(\alpha, \beta)$:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\beta) & -sin(\beta) \\ 0 & sin(\beta) & cos(\beta) \end{bmatrix} \times \begin{bmatrix} cos(\alpha) & 0 & sin(\alpha) \\ 0 & 1 & 0 \\ -sin(\alpha) & 0 & cos(\alpha) \end{bmatrix} = \begin{bmatrix} cos(\alpha) & 0 & sin(\alpha) \\ sin(\beta)sin(\alpha) & cos(\beta) & -sin(\beta)cos(\alpha) \\ -sin(\alpha)cos(\beta) & sin(\beta) & cos(\beta)cos(\alpha) \end{bmatrix}. \tag{2.3}$$

Figure 2.1 (a) shows the coordinates system of the pan-tilt camera on the world and on a cube based representation.



(a) Coordinate system (b) Some captured Images

(c) Cube faces (d) VRML view

Figure 2.1: Cube based background representation. (a) the coordinate system of the cube, (b) a number of captured images, (c) cube mosaicked-faces and (d) a VRML view of the cube model.

Having characterized the pan-tilt projection model, we can know describe the construction and query of a panoramic scene representation.

2.2 Cube Based Representation

A small FOV pan-tilt camera scans a wide scenario by densely sampling the full range of pan and tilt angles. Combining all the images into a reduced representation, such as a single texturing surface, is convenient to save memory and to help synthetize views not actually acquired by the camera but still necessary for example to do background subtraction.

The background scene of a pan-tilt camera can be represented in various ways, such as a plane, a cylinder, a sphere or a cube. In particular we select the cube based representation as it can handle a complete spherical field-of-view (FOV), $360^o \times 360^o$, which is not possible in the planar or cylindric mosaics, and maps perspective images to/from the background using just homographies (as compared to using spherical mappings). See Fig. 2.1.

For the cube coordinates we adopt the same coordinates of the camera, i.e. z points forward, x right and y down, as it show on Figure 2.1 (a).

Given a sequence of images, acquires at various pan-tilt pose and covering the complete FOV of the pan-tilt camera, we combine these images to texture the six faces of a cube based background representation. Building the cube base representation is a two steps process: (i) obtaining a back-projection for each image point and finding the right face of the cube(ii) projecting the back-projection into the right face. These two steps are detailed in the following subsection.

2.2.1 Back-projection and Identification of the (Re)projection-Face

If the intrinsics and the orientation of the camera are known, then each image point can be back-projected to a 3D world point

$$[x\ y\ z]^T = (KR)^{-1}m. \tag{2.4}$$

Identification of (re)projection face using the critical latitude If the world point obtained by Eq.2.4 has non-null x or z coordinates, then it can be scaled to be on the lateral faces [1] of a cube having edge lengths $2L$ with

$$[x_c\ y_c\ z_c]^T = [x\ y\ z]^T * L/max(|x|, |z|). \tag{2.5}$$

For each point $[x_c\ y_c\ z_c]^T$ we define the *critical latitude*, $\varphi_c(\theta)$ as the angle marking the transition between the lateral faces of the cube and the top or bottom faces. It is an angle dependent on the longitude, but can be computed directly using the point coordinates, $\varphi_c(\theta) = atan(L/\sqrt{x_c^2 + z_c^2})$ (see Fig. 2.2).

Noting that $|x_c| \leq L \wedge |z_c| \leq L$, then one has $\varphi_c(\theta) \leq 45^o$ and $\varphi_c(\theta) \geq atan(1/\sqrt{2}) \cong 35^o$.

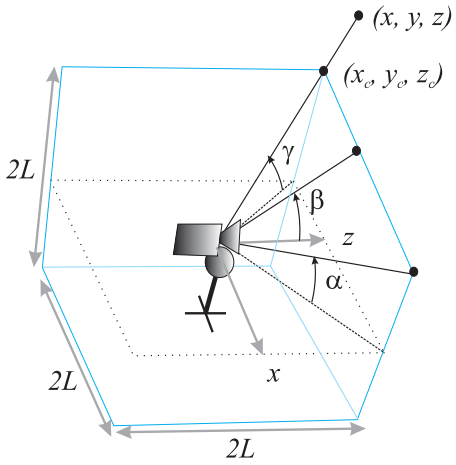


Figure 2.2: Projection of a world point (x, y, z) to the cube surface (x_c, y_c, z_c), and illustration of the *critical latitude*, φ_c. The angles α, β and γ are three examples of φ_c angles considering -45^o, 0^o and $+45^o$ latitudes respectively.

The *critical latitude*, introduced based on a cube with edge side $2L$, can in fact be computed independently of the cube size. Noting that $|x_c| \leq L \wedge |z_c| \leq L$, then one has $\varphi_c(\theta) \leq 45^o$ and $\varphi_c(\theta) \geq atan(1/sqrt2) \cong 35^o$.

[1] Lateral faces enlarged to infinity in the y direction.

Using directly the back-projected point results in a cube size independent expression:

$$\varphi_c(\theta) = atan\left(max(|x|,|z|)/\sqrt{x^2+z^2}\right). \tag{2.6}$$

Defined the *critical latitude*, and converting world coordinates to spherical coordinates longitude, $\theta = \arctan(x/z)$ and latitude, $\varphi = \arctan\left(-y/\sqrt{x^2+z^2}\right)$ one can match the image points with the correct faces of the cube using a simple set of rules. For example a point characterized by the direction (φ, θ) is mapped to the top or bottom faces of the cube if $\varphi \geq \varphi_c(\theta)$ or $\varphi \leq -\varphi_c(\theta)$. If the point is not mapped to the top/bottom faces, than it can be mapped to the front, back, left or right faces if $|\theta| \leq 45^o$, $|\theta| \geq 135^o$, $45^o < \theta < 135^o$ or $-135^o < \theta < -45^o$, respectively. See table 2.1.

Condition	Cube face
$\varphi \geq \varphi_c(\theta)$	Top
$\varphi \leq -\varphi_c(\theta)$	Bottom
$\lvert\varphi\rvert < \varphi_c(\theta) \wedge \lvert\theta\rvert \leq 45^o$	Front
$\lvert\varphi\rvert < \varphi_c(\theta) \wedge \lvert\theta\rvert \geq 135^o$	Rear
$\lvert\varphi\rvert < \varphi_c(\theta) \wedge 45^o < \theta < 135^o$	Right
$\lvert\varphi\rvert < \varphi_c(\theta) \wedge -135^o < \theta < -45^o$	Left

Table 2.1: Matching 3D directions (φ, θ) with cube faces.

Identification of (re)projecting face using the back-projection components Another way to select the correct cube face where to project a point can be determined directly by inspecting the point coordinates.Defining $v = max(|x|,|y|,|z|)$, one has that $[x\ y\ z]^T$ is imaged in the right, left, bottom, top, front or back face of the cube if $v \equiv x, -x, y, -y, z$ or $-z$, respectively.

Condition	Cube face
$v \equiv -y$	Top
$v \equiv y$	Bottom
$v \equiv z$	Front
$v \equiv -z$	Rear
$v \equiv x$	Right
$v \equiv -x$	Left

Table 2.2: Matching 3D directions v with cube faces.

2.2.2 (Re)projection of the back-projection

A cube representation can be defined as a set of six fixed cameras, where each face of the cube represents a different camera. All the six cameras share the same optic center, $t = [0\ 0\ 0]^T$, but have different orientations. Each camera must has a FOV of $90^o \times 90^o$, and the same number of pixels in length and width.

Having identified the cube faces for mapping the image points, the mapping process consists simply in projecting the back-projections of the image points using a projection matrix $P_{WF} = K_F[R_{WF}\ 0_{3\times1}]$, where K_F is an intrinsic parameters matrix characterizing the resolution (size) of the cube faces, and R_{WF} are rotation matrices defining optical axis orthogonal to the cube faces.

More precisely, if one considers that each cube face has $N \times N$ pixels [2] then

$$K_F = \begin{bmatrix} (N+1)/2 & 0 & (N-1)/2 \\ 0 & (N+1)/2 & (N-1)/2 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.7}$$

which represents a perspective camera with a $90^o \times 90^o$ field of view and an image coordinate system such that the top-left pixel is $(1, 1)$. The rotation matrices R_{WF} in essence rotate the 3D points closest to each of the faces of the cube towards the front face.

In more detail, R_{WF} is $I_{3\times3}$, $Rot_Y(180^o)$, $Rot_Y(-90^o)$, $Rot_Y(+90^o)$, $Rot_X(-90^o)$ or $Rot_X(+90^o)$ for the front, back, left, right, top or bottom cube faces, respectively. Where $Rot_Y Rot_X$ are the following matrices:

$$Rot_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\beta) & -sin(\beta) \\ 0 & sin(\beta) & cos(\beta) \end{bmatrix} \tag{2.8}$$

$$Rot_Y = \begin{bmatrix} cos(\alpha) & 0 & sin(\alpha) \\ 0 & 1 & 0 \\ -sin(\alpha) & 0 & cos(\alpha) \end{bmatrix} \tag{2.9}$$

In summary, an image point m_i is mapped to a point on a cube face m_{Fi} as:

$$m_{Fi} \sim K_F R_{WF} R^{-1} K^{-1} m_i \tag{2.10}$$

where $\sim$ denotes equality up to a scale factor.

Final note, in order to map an image to the cube, one has to know precisely the camera orientation, R and the intrinsic parameters, K. In this work we assume that R is given by the camera control system, while K is calibrated using corresponding points found in images taken at various pan-tilt poses.

Given the correspondences between the image points and cube faces, it is possible to project the back-projections of the image points to the correct cube face using Eq. 2.10.

2.3 Geometric calibration

The geometric calibration of a pan-tilt camera involves in essence estimating the intrinsics parameters. In this section we review and propose a number of calibration and auto-calibration methodologies for pan-tilt cameras.

The method presented in Sec. 2.3.2, uses the odometry of the camera to captures various poses of it and correlate the features between the images captured, this method was also used in [9]. Another method, is a well known method that, consist in changing the poses of a chess pattern in front of the camera, Sec. 2.3.1. In Sec. 2.3.3 we present a method that calibrates the camera using various posses of it without knowing its rotation. The last method, Sec. 2.3.4 is uses the same algorithm in method 3 but it assumes that the odometry of the camera is reliable but not calibrated. In all the methods, calibrating the camera consists in finding the parameters of Eq. 2.2, that is finding the values of focal length, skew and central point, $[f_u\ f_v\ s\ u_0\ v_0]$. In the method of Sec. 2.3.4 we do also

[2] In order to approximately match the resolutions of the camera and the cube faces we usually select $N = W \times 90^o/\alpha$, where W is the width of the images acquired and α denotes the horizontal field of view of the pan-tilt camera (hold still).

estimate a scaling between the pan-tilt commands (interpreted by the camera as logic, non-metric, units) and the pan-tilt metric angles.

2.3.1 Method 1: Camera calibration using a planar chess pattern

Jean Ives Bouguet's method [4] estimates the intrinsics parameters of a camera, including skew and radial distortion, by simply showing to the camera a planar chess pattern at various orientations (see Fig. 2.3). This method is based on Zhang's work [31], which explores the characteristics of the image of the absolute conic (IAC).

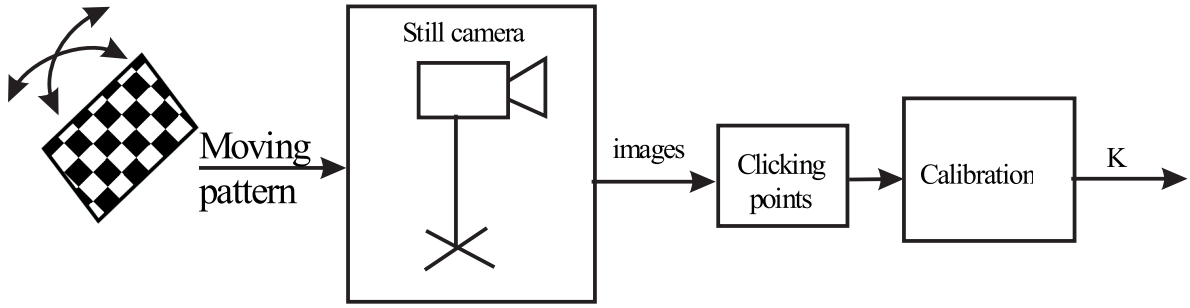


Figure 2.3: Camera calibration using a planar chess pattern.

This method is based on the assumption that the pattern is planar, and thus its surface can be characterized by $z = 0$ given an appropriate choice of the world coordinate system. The choice of $z = 0$ shows that there is a one-to-one relationship between the chess pattern points, (x, y) and the image points, m:

$$m \sim K[R \mid t]M \tag{2.11}$$
$$\sim K[r_1\ r_2\ r_3\ t][x\ y\ 0\ 1]^T \tag{2.12}$$
$$\sim K[r_1\ r_2\ t][x\ y\ 1]^T \tag{2.13}$$

where r_i denotes the column i^{th} of the rotation matrix, R, and t is the origin of the chess pattern in the camera frame. In other words, one has that the 2D point m is related with the 3D point M by a projective transform, H. From Eq. 2.13, we may write:

$$H = [h_1\ h_2\ h_3] \sim K[r_1\ r_2\ t] \tag{2.14}$$

Knowing that r_1 and r_2 are orthonormal, we have two constraints on the columns of H:

$$h_1^T K^{-T} K^{-1} h_2 = 0 \tag{2.15}$$
$$h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \tag{2.16}$$

From one homography H we can use Eq. 2.15 and Eq. 2.16 to write two constraints on the intrinsics parameters. Since the intrinsics parameters are composed by 6 parameters, this algorithm needs at least 3 homographies. Note that a camera or the pattern can change their positions between image acquisitions, as one can reset $z = 0$ to match the surface of the plane, without loss of generality, since t is not included on the constraints.

However this method is not automatic as it needs a person to go in front of the camera with a chess pattern. Additionally, in order to obtain precise results, the pattern has to occupy a large area of the image. This is not prac-

tical, for example when a surveillance camera is mounted high above the ground. In this case a person calibrating the camera cannot come close to the camera and is too cumbersome making and moving very large calibration pattern. Hence, in this work we propose using alternative calibration methodologies based on the texture existing on the scene and the motion of the pan-tilt camera. In other words, we propose using self-calibration methodologies.

2.3.2 Method 2: Camera calibration using pan-tilt odometry and features

The calibration method proposed in this section relies on the motion and odometry of the pan-tilt camera, making it impossible to use in a fixed camera. This method is automatic as it uses the texture of the scene and the odometry of the camera making unnecessary a person showing a calibration-chess-pattern. The method is based on matching scene points observed by the pan-tilt camera oriented to different directions. The scene points are characterized as directions, which depend on the camera orientation and its intrinsic parameters, K. Finding the correct K allows obtaining zero errors between the directions of matching scene points.

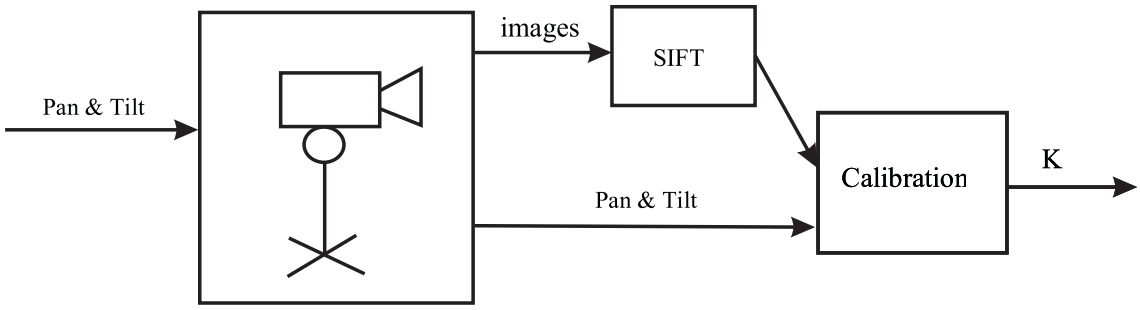


Figure 2.4: Calibration system based on a precise and accurate odometry, correspondences are calculate by SIFTs.

Then, we optimize K, a function of $\vartheta = [u_o \; v_o \; f_u \; f_v]^T$, i.e. the principal point and the focal lengths, by minimizing the back-projections of corresponding image points in two images, m_{1i} and m_{2i} obtained by matching SIFT features [25]. The optimization problem is therefore:

$$\vartheta^* = arg_\vartheta min \sum_i \|R_1^{-1} h({K_{(\vartheta)}}^{-1} m_{1i}) - R_2^{-1} h({K_{(\vartheta)}}^{-1} m_{2i})\|^2 \tag{2.17}$$

where R_1 e R_2 denote the (known) rotation matrices representing the poses of the camera for acquiring the images and $h(.)$ denotes normalization to unit norm.

2.3.3 Method 3: Image based camera calibration

In this section we consider that some pan-tilt cameras may not provide information of their pan and tilt angles, and thus one has to do the calibration based only on image information (texture). In this vein, Hartley and Agapito [1] introduced a self-calibration method that estimate the intrinsics parameters of a rotating camera.

Given a 3D world point M, it has different projections on different images, $m_i = P_i M$. Since M is the same on all the images and P_i is a rank-3 matrix which may be decomposed as $P_i = K_i R_i$, we can calculate the relation between the images with $m_i = P_i P_j^{-1} m_j$.

The relation between any two images with the same optical center is therefore a projective transformation, the so called *infinite homography*, H_{ij}. This homography can be calculated by finding matching points, in our case

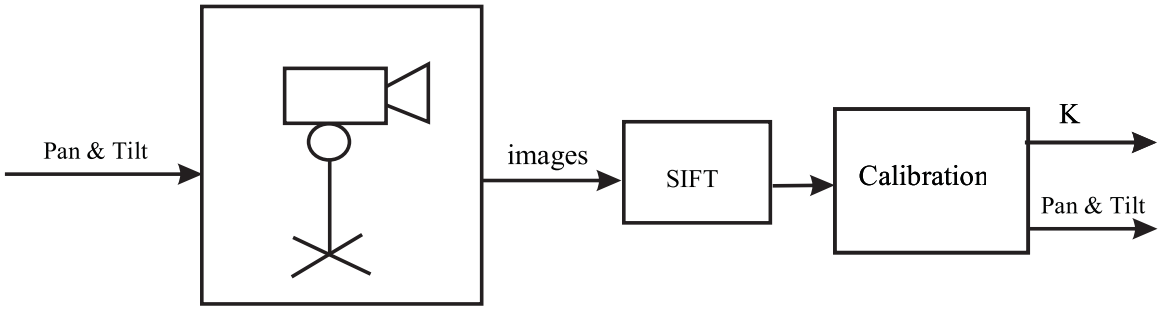


Figure 2.5: Calibration system introduced by Agapito, no pan-tilt information is needed but it can be estimated.

SIFT features, between pairs of images:

$$H_{ij} = P_i {P_j}^{-1} = K_i R_i {R_j}^{-1} {K_j}^{-1} = K_i R_{ij} {K_j}^{-1}. \tag{2.18}$$

The homography can be computed in a least squares sense having at least four matched points. As R_{ij} is a rotation matrix, it satisfies $R_{ij} = {R_{ij}}^{-T}$ and thus one obtains

$$H_{ij} K_j K_j^T H_{ij}^T = K_i K_i^T \tag{2.19}$$

a constraint mixing the intrinsics parameters matrices with the homography between two images. In the cases that $K_i = K_j = K$, i.e. images with constant zoom, Eq.(2.19) can be solved in a least squares sense as a set of linear equations, $Ax = 0$, in the entries of IAC, here denoted as ω,

$$H_{ij}^{-T} \omega H_{ij}^{-1} = \omega. \tag{2.20}$$

A number of constraints can then be imposed on ω in order to facilitate its estimation:

1. Zero-skew : If $s = K_{12} = 0$, then $\omega_{12} = 0$.
2. Square-pixels : If $s = 0$ and $\alpha_x = \alpha_y$, then $\omega_{11} - \omega_{22} = 0$.
3. Known principal point : If $s = 0$ and $x_0 = 0$, then $\omega_{13} = 0$. Similarly if $y_0 = 0$ then $\omega_{23} = 0$.

Considering for example the first constraint, zero-skew, one has:

$$\begin{aligned} \omega &= K^{-T} K^{-1} \\ &= \begin{bmatrix} 1/\alpha_x^2 & 0 & -x_0/\alpha_x^2 \\ 0 & 1/\alpha_y^2 & -y_0/\alpha_y^2 \\ -x_0/\alpha_x^2 & -y_0/\alpha_y^2 & 1 + x_0^2/\alpha_x^2 + y_0^2/\alpha_y^2 \end{bmatrix} \end{aligned}$$

A closed form solution can then be obtained. Since ω is symmetric, its six upper triangular entries can be collected in a vector, $w = [w_1\ w_2\ w_3\ w_4\ w_5\ w_6]^T$, which suffices to parametrize ω. From Eq. 2.20 each homography H_{0i} gives six equations, A, that can be written as $Aw = 0$. With six equations and six unknowns, w, one can find a least-squares solution using singular value decomposition, that is finding w that minimizes $\|Aw\|$ with $\|w\| = 1$. This linear algorithm performs fast, as compared to iterative algorithms, and is simple to implement.

Note that a cube based background representation can be constructed without knowing the pan and tilt angles of the camera, since all the rotations can be calculated from:

$$R_{ij} = K^{-1} H_{ij} K \tag{2.21}$$

This method, as well as Bouguet's method described in Sec. 2.3.1, explores the characteristics of the IAC. The main difference between the methods is that Bouguet's method accepts a fixed camera but requires observing a known, mobile, planar chess pattern, while the method described in this section just requires a pan-tilt camera and allows any (static) textured scene.

2.3.4 Method 4: Image based camera calibration and pan-tilt odometry scaling

This method is designed for cameras having a precise odometry but whose reference values are not metric. More precisely, the reference values are specified in logic units characterizing an (unknown) pan and tilt operating range. As compared with the methodology of the previous section, the objectives and the main steps are similar, i.e. involve estimating the intrinsic parameters K (see fig. 2.6), but in the end one wants, additionally, to take advantage of the precise odometry provided by the camera.

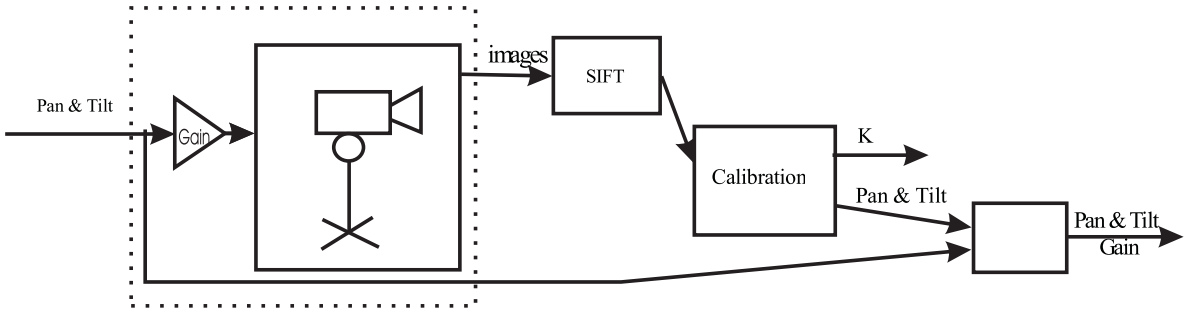


Figure 2.6: System for calibrating the camera intrinsic as well as the odometry.

Assuming that logic and metric pan-tilt angles differ just by scaling factors, i.e. $\alpha_m = \rho_1 \alpha$ and $\beta_m = \rho_2 \beta$ where (α, β) and (α_m, β_m) denote pan, tilt angles, in logic and metric units resp., in essence one wants to find (ρ_1, ρ_2) such that is possible to command the pan-tilt camera precisely and in metric units.

Using the methodology introduced in the previous section one obtains the intrinsic parameters K, and thus one can estimate the rotations R_{ij} from Eq. 2.21. The rotation matrices estimated from the homographies can therefore be used to find the scale factors. Using the Frobenius norm to compare rotation matrices, allows creating a cost function that globally penalizes errors in the scaling factors $\rho = [\rho_1 \ \rho_2]$,

$$\rho^* = arg_\rho min \sum_{ij} \|R(\rho_1 \alpha_{ij}, \rho_2 \beta_{ij}) - R_{ij}\|_F^2 \tag{2.22}$$

where $R(\rho_1 \alpha_{ij}, \rho_2 \beta_{ij})$ denotes a rotation matrix built from the pan and tilt angles as detailed in Eq. 2.3. The pan and tilt angles, α_{ij} and β_{ij}, are read from the odometry and scaled by ρ in order to build $R(.)$. We use the Levenberg-Marquardt algorithm to optimize Eq.2.22. Having estimated ρ effectively allows one to turn a camera with precise odometry, but defined up to an unknown scale factor, into a camera commanded with precise metric units.

2.3.5 Comparing the calibration methods

The advantage of using a pattern calibration, Sec. 2.3.1, is that it can be used in all types of cameras. However it needs a person to move the pattern and it needs that the pattern occupies most of the image, which in some cases it is not possible or practical. The calibration method that uses pan-tilt odometry and features, in Sec. 2.3.2, can estimate the intrinsics without requiring a person showing a calibration pattern, however it needs a camera with an accurate and precise odometry. The method proposed by Agapito et al. [2, 1] is also autonomous and does not need the pan-tilt information from the odometry. The last method presented can be used to turn a camera with precise odometry, but defined up to an unknown scale factor, into a camera with precise and accurate odometry.

On this Chapter we approached the problems of the geometrically modeling a pan-tilt camera, representing the background, selecting and defining a mosaic model and present various methods of geometric calibration of a camera. On the next chapter we will address the Radiometric model of a camera and present a Vignetting correction method to improve event detection.

Chapter 3

Radiometry and Event Detection

In many computer vision systems the image intensity its considered to be proportional to the scene radiance. This is not true in general, due to various factors as lesser optical gain while moving away of the optical axis (vignetting) [19], the electronic or chemical photo-detector conversion of the image brightness into image irradiance [13, 14] or the radiometric response function of the camera. Without calibrating the color information, which is distorted by camera system components, the results of brightness intensity image analysis may conceal or inaccurately represent important intensities characteristics of objects in images. In other words, applications requiring the true colors of the objects, such as comparing images of objects acquired with different settings, imply correcting the non-linear radiometric mapping into a linear one by calibrating the radiometric response of the camera system.

In this chapter we study and propose correction methodologies for the referred radiometric problems. We start by describing the radiometric model of image formation process, having as principal components the radiometric response function and vignetting, and propose a patternless-methodology to estimate and correct the vignetting in pan-tilt cameras, and show that it lowers the variance of the background representation of a scene, the attenuation of the variance helps then improving the event detection.

3.1 Radiometric Image Formation Model

Most cameras compress the dynamic range of the scene, causing the formation of a nonlinear relationship between the image intensities output and the scene radiance. This nonlinear relationship is purposively introduced by manufacturers, so that the camera can compress a wide range of irradiance values within a fixed range of measurable image brightness values. There are also some other effects in the image formation that are not intended by the manufacturers. Some of the effects are naturally inherent to the camera, such as tilt-effect and natural or optical vignetting [17].

Figure 3.1 shows a flow diagram with the basic radiometric transformations present in the image formation model of the most common cameras: tilt-effect and vignetting, V, and the radiometric response function (RRF), f. The vignetting can be decomposed in natural and optical, and models a space variant gain on the optics of the imaging system. The radiometric response function, is usually a non-linear function modeling the response of the CCD (charge-coupled device) sensors or film of a camera.

In this section we describe the various processes that radiance suffers until it is transformed in to pixel intensi-

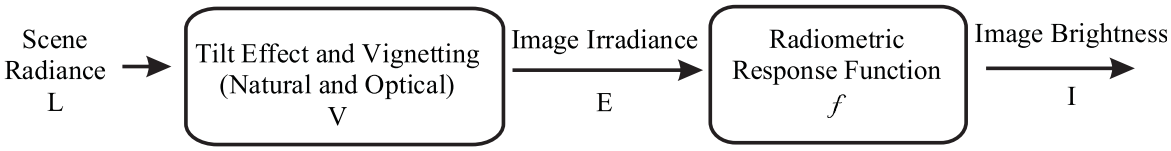


Figure 3.1: Flow diagram showing the two basic transformations, V and f, that map scene radiance L to image brightness I.

ties. Firsts we introduce the Kang-Weiss model [17] which describes the process between the moment light reflects in an object until it gets to the CCD sensor or film. The next process models the radiometric response function of a camera as introduced by Grossberg and Nayar [13, 14]. In the end of the section we show a complete model integrating the Kan-Weiss and Grossberg-Nayar models.

3.1.1 Kang-Weiss Radiometric Model and Vignetting

The Kang and Weiss radiometric model [17] models the light since it reflects in an object until it gets to the charge-coupled device (CCD), or the camera film. The model encompasses three main components: *Natural Vignetting*, *Optical Vignetting* and *Tilt Effect*. These components are detailed in the following.

Natural Vignetting If the object is a plane of uniform radiance exactly perpendicular to the optical axis, then the illuminance of its image can be observed to gradually decrease with distance away from the principal point of the camera. It can be shown that irradiance, E is proportional to radiance, L_0 but varies across the field of view in proportion with the fourth power of the cosine of the field angle.

$$E'(\theta) = L'_0 \cos^4 \theta \tag{3.1}$$

$$= L'_0 \frac{1}{(1 + (r/f)^2)^2} = L'_0 \beta \tag{3.2}$$

Where f is the focal length, r is the distance from the principal point, and considering the aspect ratio 1 and the principal point $[u\ v]^T = [0\ 0]^T$.

Optical Vignetting The optical vignetting is caused by the partial obstruction of light from the object space to the image space. Field-step, pupils or the lens rim are the most common causes of the partial cut off on the cone of lights rays emitted from an object point towards the lens. This obstruction produces a fall off in light intensity towards the image periphery. The intensity of a pixel lost due to optical vignetting can be expressed as the approximation:

$$E'_{vig}(\theta) = (1 - \alpha r) E'(\theta) \tag{3.3}$$

where α is a constant that models the loss of light due to vignetting. This linear relationship can be extended to a more general polynomial form [32]:

$$E'_{vig}(\theta) = (1 - \alpha r_i - \ldots - \alpha_p r_i^p) E'(\theta) \tag{3.4}$$

where p represents a polynomial order that can be arbitrarily set according to a desired precision.

Tilt Effect Closer points in the world appear brighter due to the inverse square law of illumination. Difference in the depth of the scene imply therefore intensity variations. The tilt effect on a planar (constant radiance) object results in an elliptical isointensity distribution. The tilt, γ applied effect to Eq. 3.2 yields the following formulation [17]:

$$E'(\theta) = L'_0 \cos\tau \left(1 + \frac{\tan\tau}{f}(u\sin\chi - v\cos\chi)\right)^3 \cos^4\theta \tag{3.5}$$

$$E'(\theta) = L'_0 \gamma \cos^4\theta = L'_0 \gamma\beta \tag{3.6}$$

where the tilt parameter τ describes the rotation angle of a planar scene surface around an axis parallel to the optical axis and χ is the rotation angle around the x-axis of this rotated plane [32].

Combining Eq. 3.3 and Eq. 3.6 one obtains:

$$E'_{all}(\theta) = (1 - \alpha r)\gamma\beta L'_0 \tag{3.7}$$

This model is useful to calibrate the vignetting effect of a flat textureless surface under uniform illumination [17]. Another application is single-image vignetting correction [33, 32]. This model takes into account the various types of vignetting, natural and optical.

Kang and Weiss work [17] shows that it is possible to calibrate the vignetting of a camera using just a flat, textureless Lambertian surface and constant illumination. This can be done considering the effect of the tilt of the camera and using the effects of off-axis illumination and vignetting, i.e. natural and optical vignetting, which result in reduction of light into the camera at off-axis angles.

Figure3.2 shows a sequence of images with synthetic vignetting. The vignetting effect was applied to all images, increasing from (a) to (d). As the figure (a) shows the person is visible since the image has almost no vignetting at all, but on figure (d) where was applied a severe vignetting, the person is almost seamless.

3.1.2 Radiometric Response Function

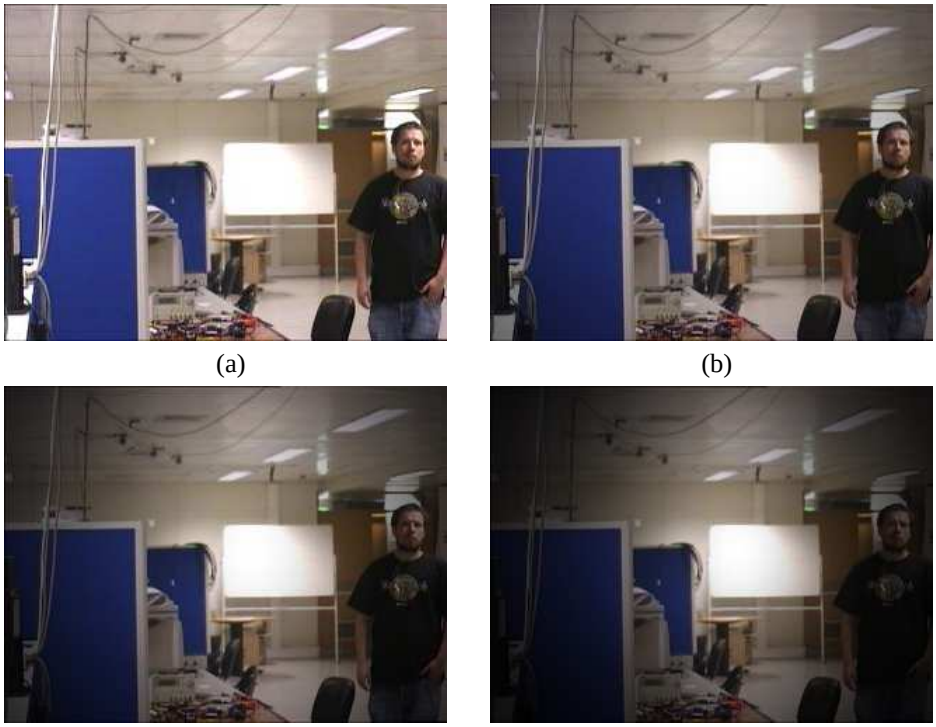
Most cameras compress the dynamic range of the scene, introducing a nonlinearity between the image irradiance and the captured image irradiance. The nonlinear relationship, usually named the Radiometric Response function, $f(\mathbf{.})$ can be expressed as follows :

$$I(m) = f(kE(m)) \tag{3.8}$$

where $I(m)$ is the image intensity at a pixel $m = [u\ v]^T$, k is the shutter time and $E(m)$ the value of the irradiance that its observed by the CCD sensor.

Nayar and Grossberg [13] collected a diverse database of 201 real-world camera response functions, the so called Database of Radiometric Functions (DoRF). They gathered response curves for a wide variety of photographic films, CCD sensors, and digital cameras. Each of the three color channel were treated as individual response curves. The curves gathered include color as well as black and white films, consumer and professional films, still and motion picture films. Using this database they have shown that real-world responses occupy a small part of the theoretical space of all possible responses.

The DoRF allows on finding a low dimensional approximation of theoretical space of response functions, that is close to the empirical data. This can be achieved by applying Principal Component Analysis, PCA, to do the



(a) (b)

(c) (d)

Figure 3.2: Images with synthetic vignetting. Vignetting is increasing from (a) to (d), in (a) the vignetting seamless and (d) as severe vignetting.

training curves to find a basis $[f_0\ h_1\ h2\ ...\ h_M]$ of the function:

$$f_0(E) + \sum_{n=1}^{M} c_n h_n(E) \tag{3.9}$$

To create the PCA, Grossberg and Nayar used a training set of 175 response curves and a testing set with 26 curves, with $M = 25$. The cumulative energies associated with the eigenvalues increase rapidly. In fact 3 eigenvalues explain more than 99.5% of the energy. This suggests that even a 3-parameter model should work reasonably well for most response functions. Fig 3.3 shows the base curve of the EMoR[1] model.

3.1.3 Complete Model

Having the defined the Kang-Weiss radiometric model $I_{iR}(m) = k\varphi L(m)$, vignetting, gain $V(m)$, and the radiometric response function, $f(\cdot)$, the complete model of the image formation process can be described as [13, 30, 20, 32] :

$$I(m) = f(k\varphi L(m)) \simeq f\left(kV(m)L(m)\right) \tag{3.10}$$

where $I(m)$ is the image intensity at the image point m, $f(.)$ is the radiometric response function, k is the exposure time, $L(m)$ the radiance of a scene point imaged at m, and $V(m)$ is the vignetting gain at m. Note that both $f(.)$

[1]Data acquired from `http://www.cs.columbia.edu/CAVE/software/softlib/dorf.php`

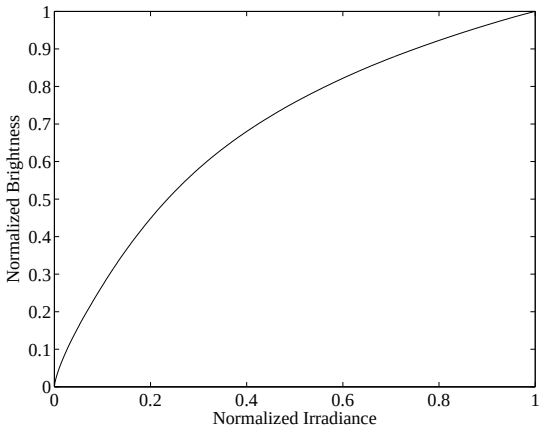


Figure 3.3: The mean of 175 camera responses in DoRF used as the base curve, f_0, in the EMoR model.

and $V(m)$ have nonlinear natures, $f(.)$ depends on the pixel brightness / color, while $V(m)$ depends on the pixel location, such that central pixels tend to be unmodified, i.e. $V(m) = 1$ and pixels in the border of the image have lesser brightness ($V(m) < 1$).

In the following we introduced the patternless methodology for estimating vignetting that is based on a mosaicked image, the Constant Radiance Image, which is composed from a number of images taken at various pan-tilt poses.

3.2 Constant Radiance Images and Radiometric Calibration

A static object illuminated by a constant light source emits a constant radiation. Contrarily to the radiance, the observed irradiance at the image plane of a moving pan-tilt camera is not constant. It varies with the pan-tilt pose of the camera e.g. due to vignetting.

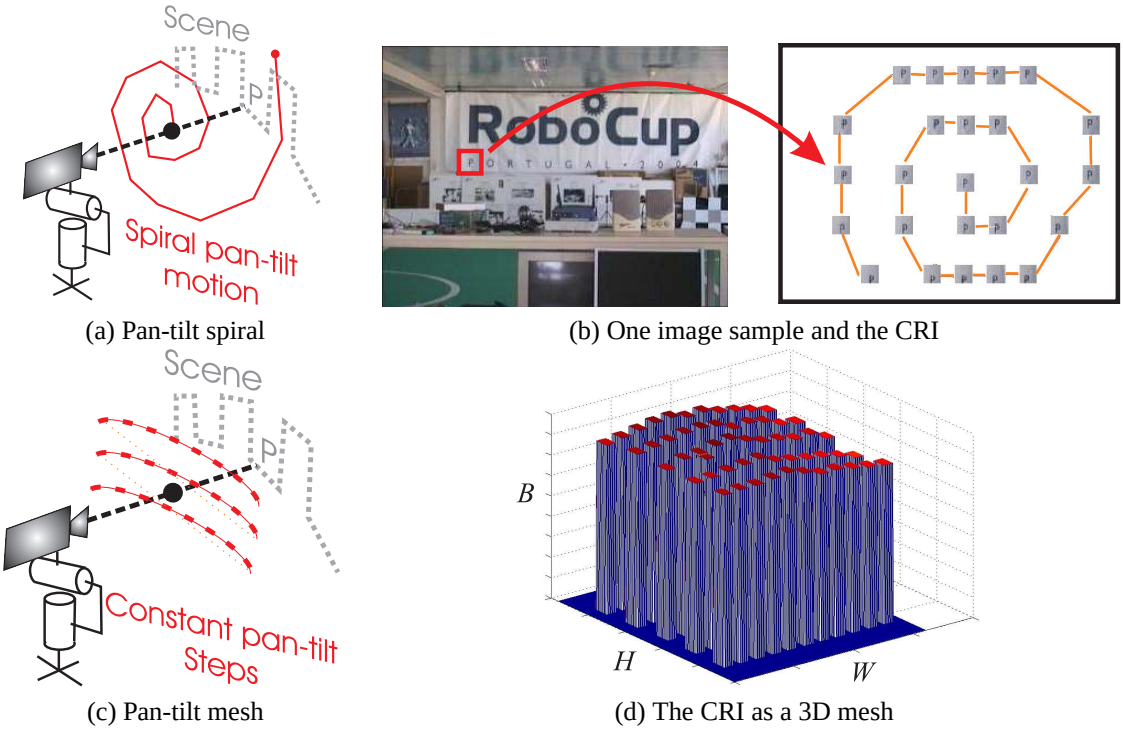
$$\begin{cases} m_1 = \mathcal{P}_1(M_1) \\ m_2 = \mathcal{P}_2(M_2) \\ M_1 = M_2 \end{cases} \quad \not\Rightarrow \quad I(m_1) = I(m_2) \tag{3.11}$$

In other words, in general $f(kV(m_1)L(m_1)) \neq f(kV(m_2)L(m_2))$ even if $L(m_1) = L(m_2)$.

In order to describe the varying irradiance of a single world point captured by moving pan-tilt cameras, it is convenient to construct what we define as Constant Radiance Images. These images represent the irradiance of a single world point when it is observed at different image coordinates.

The construction of a Constant Radiance Image, C_{mo} with a pan and tilt camera starts simply by choosing one image point, $m_o = [u_o\ v_o\ 1]^T$, computing its back-projection to a 3D point, and then moving (rotating) the camera, R_i, and re-projecting the 3D point to obtain the new image point m_i:

$$m_i \sim K R_i R_o^{-1} K^{-1} m_o. \tag{3.12}$$



(a) Pan-tilt spiral

(b) One image sample and the CRI

(c) Pan-tilt mesh

(d) The CRI as a 3D mesh

Figure 3.4: Building a Constant Radiance Image (CRI). Camera motion (a) and sample (patch) collection (b). A regular grid of patches (d) is obtained with approx. constant pan-tilt steps (c).

This allows us to make a color image C_{mo}, where we can observe the vignetting effect of the camera:

$$C_{mo}(m_i) = I_i(m_i). \tag{3.13}$$

Figure 3.4 shows the construction of one C_{mo}, and illustrates the typical aspect of a vignetting effect.

3.2.1 Vignetting Correction

Assuming that (i) one estimates the radiometric response function $f(.)$, using e.g. the method in [22], then one can remove the effect of $f(.)$ by redefining $I(m) \rightarrow f^{-1}(I(m))$, (ii) the exposure time k is the same for background construction and event detection, then it is no longer a distinguishing factor and we can use without loss of generality that $k = 1$, and (iii) the maximum of the vignetting gain is unitary, i.e. keeps unchanged a number of central pixels of the original image, then one finds that one Constant Radiation Image characterizes the vignetting. More precisely, $C_{m0}(m) = f^{-1}\left(f\left(kV(m)L(m)\right)\right) = V(m)L(m)$, and the vignetting gain can be estimated as $V(m) = C_{mo}(m)/max\left(C_{mo}(m)\right)$.

Given the estimated vignetting function, $V(m)$, one desires to apply a correction function, $V_c(m)$ that approximates the captured image equal to the radiance image:

$$I_c(m) = V_c(m)I(m) = V_c(m)V(m)L(m) \tag{3.14}$$

i.e. one wants $V_c(m) = V^{-1}(m)$, which means $V_c(m) = max(C_{mo}(m))/C_{mo}(m)$. This approach has however

two problems, it requires a dense pan-tilt sweeping to fill all the pixels of a Constant Radiance Image, and it is affected by image noise. We propose therefore an optimization methodology having a smooth interpolating (parametric) vignetting correction function.

The parametric vignetting correction function is in general a function that keeps the center pixels unchanged, and gradually enhances (augments) the brightness of the pixels closer to the border. In the literature one finds for example sums of even powers of radial distances (see e.g. [20]). In this work we follow the suggestion of [30]:

$$V_c(m;a) = cosh(a_1(u-u_p))cosh(a_2(v-v_p)) + a_3 \tag{3.15}$$

where $m = [u\ v\ 1]^T$ is an image point, $m_p = [u_p\ v_p\ 1]^T$ is the principal point, and the vector $a = [a_1\ a_2\ a_3]^T$ contains the parameters characterizing the correction.

Having defined the fitting function, we can know describe the optimization procedure to find the vignetting correction as:

$$a^* = arg_a min \sum_m \left(\frac{max(C_{mo}(m))}{C_{mo}(m)} - V_c(m;a) \right)^2 \tag{3.16}$$

which can be solved iteratively with the Levenberg-Marquardt algorithm.

Concluding, the ability of the pan-tilt camera to image the same small Lambertian object (constant grayscale patch) at various locations of the image effectively constructs an image, the Constant Radiance Image, which allows observing the vignetting gain.

Note that differently to [30], this vignetting correction methodology does not require using a calibration pattern.

3.2.2 Background Uncertainty

A panoramic background representation comprises the superposition of various images, acquired at different pan-tilt poses. Thus, the same 3D object seen at various pan-tilt poses, despite having a constant radiance, has a varying (captured) irradiance imposed by the vignetting. A background model is usually represented by the mean value and variance of the irradiance at each background location M, respectively $\mu_{B(M)}$ and $\sigma^2_{B(M)}$. Without vignetting correction the "gray level" value of a background location will change as the camera rotation changes. The values of the background thus depend not only on image noise but also on the changes due to vignetting in the imaged pixel $V(m)$, which can now be considered a random variable with mean, $\mu_{V(m)}$, and a variance, $\sigma^2_{V(m)}$:

$$B(M) = L(M)V(m) + \eta \tag{3.17}$$

where η is a zero mean noise process, and $L(M)$ denotes the radiance that is expected to be observed at the background pixel M. Taking expected values we get:

$$\begin{cases} \mu_{B(M)} = L(M)\ \mu_{V(m)} \\ \sigma^2_{B(M)} = L^2(m)\ \sigma^2_{V(m)} + \sigma^2_\eta \end{cases} \tag{3.18}$$

where σ^2_η is the noise variance. The vignetting correction allows to decrease the variance at the superposition as shown next.

Considering that the processes of vignetting and vignetting-correction can be characterized by a mean gain, $\mu_{V_c(m)V(m)}$, and a variance of gains, $\sigma^2_{V_c(m)V(m)}$, then we have that the background mean and variance are

$\mu_{B(M)} = L(M)\ \mu_{V_c(m)V(m)}$ and $\sigma^2_{B(M)} = L^2(m)\ \sigma^2_{V_c(m)V(m)} + \sigma^2_\eta$. In the case of not having vignetting correction, $V_c(m) = 1$, we have that the vignetting directly effects on the image, $\mu_{V_c(m)V(m)} = \mu_{V(m)}$ and $\sigma^2_{V_c(m)V(m)} = \sigma^2_{V(m)}$. On the other hand, if one has a perfect correction, $V_c(m) = V(m)^{-1}$, then we have a perfect observation of the scene radiance $\mu_{B(M)} = L(M)$ and a zero variance on the background representation, $\sigma^2_{B(M)} = L^2(M) \times 0 = 0$.

Simulated white scenario To prove the benefits of correcting the Vignetting we show next a simulated scenario experiment. In this experiment the scene luminance, $L(M)$ is constant. The reference vignetting correction gain, $V_c(m; a_r)$ with $a_r = [a_{r1} a_{r2} a_{r3}]^T$, is the one obtained from a real camera (see Sec. 3.2.1). Vignetting distortion is defined as $V(m; a_r) \doteq 1/V_c(m; a_r)$.

In order to compare various combinations of vignetting distortion and correction gains, we vary both gains in a parametric manner, by scaling the parameters. More precisely, we use $V(m; \alpha a_r)$ with $\alpha \in \{0, .3, .6, 1, 1.3, 1.6\}$, and $V_c(m; \beta\alpha a_r)$, with $\beta \in \{0, .5, 1, 1.1, 1.2, 1.5\}$. Note that $\alpha = 1$ corresponds to introducing the reference vignetting, while $\alpha = 0$ corresponds to not introducing vignetting. Similarly, $\beta = 0$ and $\beta = 1$ correspond to no vignetting correction and to perfect correction, respectively.

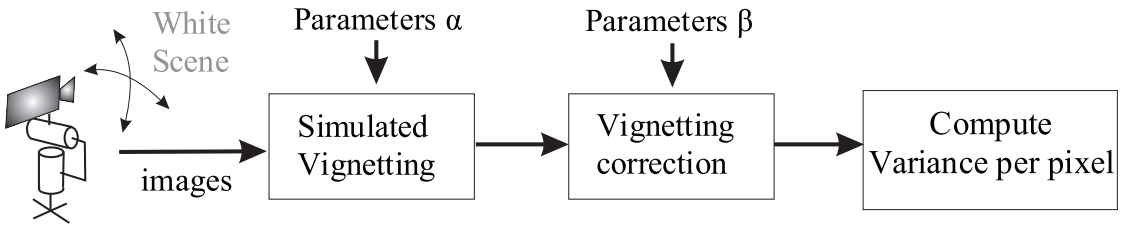


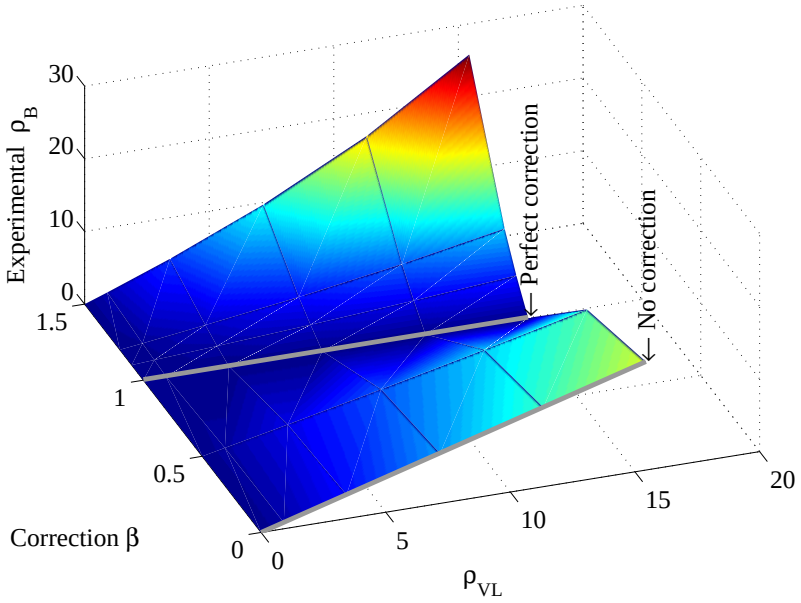
Figure 3.5: Setup of the simulated white scenario.

Figure 3.6 (a) shows the standard deviation of a background pixel, $\rho_{B(m)} = \sqrt{\sigma^2_{B(m)}}$, estimated empirically from a set of overlapping images acquired at various pan-tilt poses, for various combinations of $V(m; \alpha a_r)$ and $V_c(m; \beta a_r)$. In the horizontal axes, Fig. 3.6 (a), has the theoretical estimate of the background standard deviation considering there is no vignetting correction, i.e. $\rho_{VL} = \sqrt{L^2\sigma^2_{V(m)}}$, and the parameter β regulating the amount of vignetting correction. From Eq.3.18 with $\sigma^2_\eta = 0$ one has that $\rho_{B(m)} \equiv \rho_{VL}$ only when $\beta = 0$. Otherwise, one may have $\rho_{B(m)} = 0$ when the vignetting correction removes perfectly the vignetting distortion, $\beta = 1$, which is confirmed by the plot.

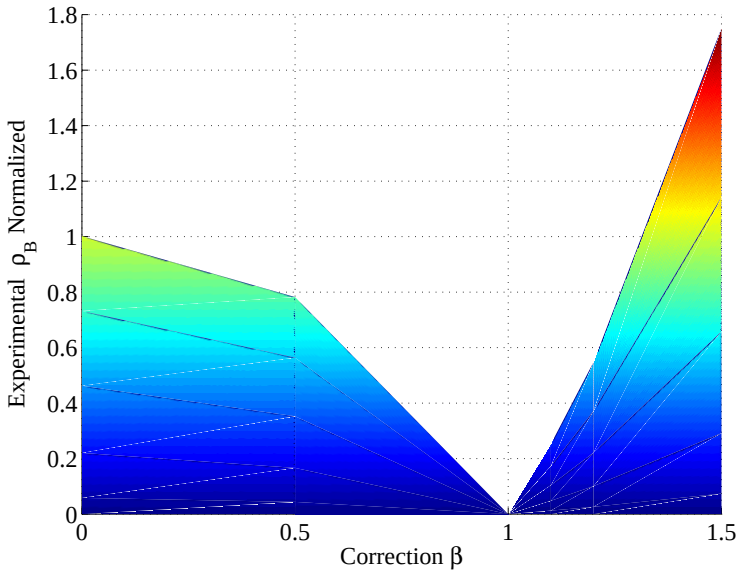
3.3 Event Detection

Identifying objects appearing in the field of view (events) is a fundamental and critical task insurveillance systems. A common approach is to perform background subtraction, which is a class of techniques that is used for segmenting out objects of interest from a data base model, this case is a mosaic.

Some robust methods were analyzed, such as mixture of Gaussians, Kernel density estimation, however, as we can see in Piccardi's work [27] this methods require a vast amount of memory to perform well. Due to memory restriction it was choose the single Gaussians model, that only need to storage the amount of information corresponding to three backgrounds, one that accumulates the sum of all the pixels values, $\sum_{i=1}^{N_{uv}} I_{iuv}$, one that stores



(a)



(b)

Figure 3.6: Background standard deviation vs vignetting effects and corrections (a). The correction β is adimensional, while ρ_{VL} and ρ_B have their dimensions defined in a 8 bits gray-scale space. (b) is a cut on the graphic with constant ρ_{VL}, the scale was changed to be one when the correction is not applied.

the sum of the square of the pixels values,$\sum_{i=1}^{N_{uv}} I_{iuv}^2$, and the other that saves the number off times a pixel is seen, N_{uv}. The mean and variance are them computed as:

$$B_{uv} = \frac{\sum_{i=1}^{N_{uv}} I_{iuv}}{N_{uv}} \tag{3.19}$$

$$\Sigma_{uv}^2 = \frac{\sum_{i=1}^{N_{uv}} I_{iuv}^2 - (\sum_{i=1}^{N_{uv}} I_{iuv})^2 / N_{uv}}{N_{uv} - 1} \tag{3.20}$$

The first method used is done by comparing the currently captured image, vignetting-corrected, I_{uv} with the corresponding image retrieved from the background database, B_{uv}, using the log likelihood function, L_{uv}:

$$L_{uv} = \frac{-0.5(I_{uv} - B_{uv})^2}{\Sigma_{uv}^2} - 0.5\ln(\Sigma_{uv}^2) - 0.5\ln(2\pi) \tag{3.21}$$

where Σ_{uv}^2 denotes the background variance. A pixel (u, v) is considered active, foreground, if the likelihood, L_{uv}, is larger than a threshold in at least two of the three (RGB) components of I_{uv}.

The second method uses the knowledge of standard deviation, considering that the noise in each pixel is Gaussian, one knows that, for instance, two standard deviations, σ, from the mean account for 95.45 percent of the set. The normalization of the standard deviation is calculated with the next equation:

$$\left(\frac{(I_{uv} - B_{uv})^2}{\Sigma_{uv}^2}\right)^{1/2} \leq k \tag{3.22}$$

where k is the desired threshold. A pixel is considerer background if Eq. 3.22 is true, otherwise it is active.

Chapter 4

Experiments

This chapter describes the experiments performed to validate the methodologies presented and introduced in previous chapters. The experiments address two main aspects: testing the geometric calibration methods and testing the benefit of vignetting correction in event detection based on back-ground subtraction with pan-tilt cameras.

4.1 Experimental Setup

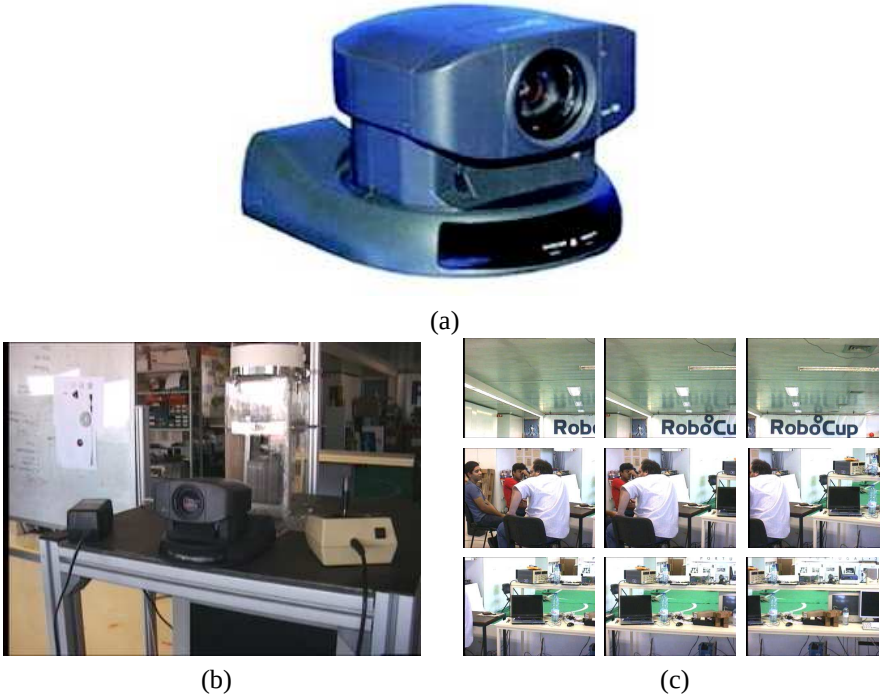
All the experiments were made indoors, namely at the robotics laboratory of the Institute for Systems and Robotics, Lisbon, Portugal, a research institute at IST. We use a camera Sony EVI D30 [7], see Figure 4.1 (a) an analog PAL camera and controlled in discrete positioning units [1]. The camera has a maximum Pan-tilt range of $\pm 100^o$ and $\pm 25^o$ respectively, with a minimum steps of $\pm 0.85^o$ and $\pm 0.64^o$ (estimation based on Sec. 2.3.4). The maximum horizontal and vertical FOV of the camera is 48.8^o and 28^o degrees, respectively, while the minimum FOV is 4.3^o and 3.2^o corresponding to the maximum zoom settings. More information can be obtained in the camera specifications brochure[2].

In all our experiments we used a resolution of 320 pixels horizontally by 240 pixels vertically. The zoom used was set constant, implying a FOV of $17^o \times 13^o$, estimated using the calibration methodology detailed in Sec. 2.3.4. In all the experiments we set the auto gain OFF, the auto iris OFF and made the shutter time constant. By doing these we imply that all the images have the same dynamic range. The images were acquired with a frame grabber based on the BT878 chip.

4.2 Vignetting Correction

This experiment aims at evaluating the vignetting artifacts in a mosaic. Hence emphasizing the advantages of vignetting correction. First we show the vignetting effect using the already defined CRI introduced in Section 3.2. We collected seven patches of the central image, $(pan, tilt) = (0^o, 0^o)$, and analyzed the vignetting effect in the u and v direction of the image. Figure 4.2 (a) shows the patches that were analyzed. Making a CRI for one of the patches as seen in Fig 4.2(c),allows estimating the vignetting function using Eq. 3.16.

[1] `http://bssc.sel.sony.com/Professional/docs/manuals/evid30commandlist1-21.pdf`
[2] `http://www.ave-rassmann.de/pdf/evi_d31_p.pdf`



(a)

(b) (c)

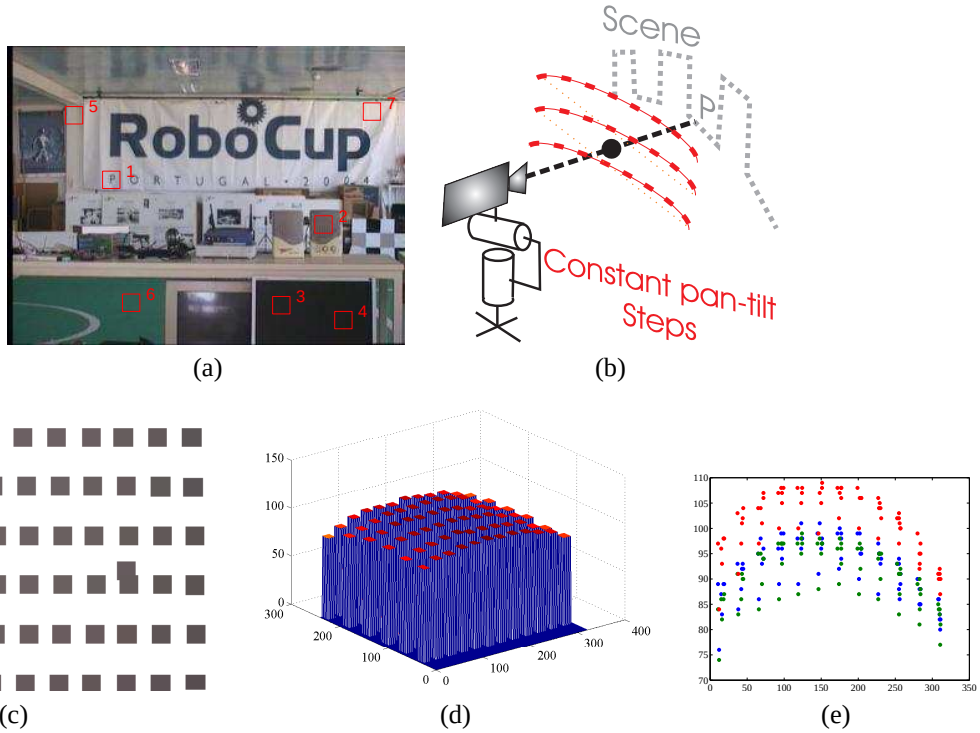
Figure 4.1: Pan-tilt setup. Camera Sony EVI D31 (a). Camera location in the laboratory, in top of Labmate (b). A number of the images captured by the camera (c).

To prove the accuracy of the parameters estimated, we printed a sheet of paper containing sixteen patches with gray levels, Fig 4.3 (a). A CRI was created for each of the patches on the sheet. Afterwards the various CRI were compared with the vignetting function estimated. Figure 4.3 (b) shows the central line of the various CRI and the correspondent vignetting function applied to the various gray levels. As it can be seen the difference between the real data and the synthesized CRI based on the vignetting function is small. One can conclude therefore that the algorithm performs a good approximation. The small errors observed can be explained by the noise created in the camera and by the assumption of the radiometric response function being linear.

Figure 4.4 shows the blending of two image patches having constant color. In (a) one sees clearly two different areas, separated by an image seam caused by the vignetting. In (b) vignetting correction makes barely visible the difference between the two patches. Figure 4.4 (c) shows more clearly the difference between the images. The profile on the top corresponds to the image patches without vignetting correction, while the profile on the bottom shows the effect of vignetting correction. Ones sees that the profile of the images without vignetting correction has a clearly lighter region to the left of the seam.

4.3 Single Image Event Detection

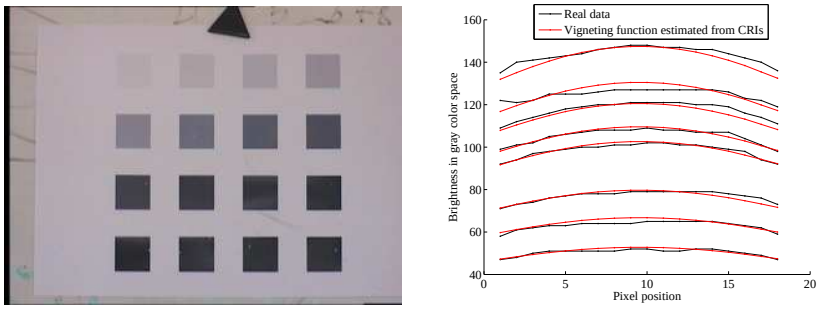
In this experiment we used the pan-tilt camera to scan a room and create two background representations: one lacking and the other having vignetting-correction. The images with events to be detected, were created afterwards using a video projector superimposing text (digits) towards the ceiling of the room, Fig. 4.5 (c). Figure 4.5 (d) shows correct detections of digits, in both cases, despite the variance of the background motivated by the imaging



(a) (b)

(c) (d) (e)

Figure 4.2: Constant resolution images (CRI). Image taken at $pan = 0^o$ and $tilt = 0^o$ and used to initialize the creation of various CRI, one for each of the patches numbered 1 to 7 and shown in red (a). CRI associated to patch number two (c), perspective view of the red component of the CRI shown as a mesh plot (d) and RGB values of the CRI versus the horizontal image coordinate (e).



(a) (b)

Figure 4.3: Validation of the vignetting estimation process. (a) Sheet of paper containing 16 patches with constant gray levels. (b) The central lines of the 8 CRI obtained from the 8 gray level patches compared with the synthesized CRI based on the estimated vignetting function.

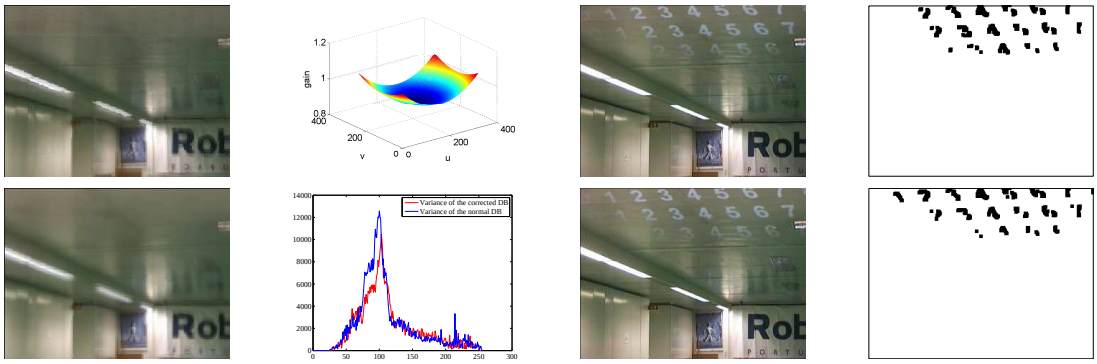
(a) Mosaic seam with uncorrected vignetting

(b) Mosaic seam with corrected vignetting

(c) Scan lines

Figure 4.4: Image blending. (a) images blending without vignetting correction. (b) Image blending with vignetting correction. (c) Central line profiles of (a) and (b), shown on top and bottom, respectively.

distortions (e.g. radiometric response function and vignetting), however it is clear that the background subtraction on the images with vignetting correction is more effective, e.g. detects more events (digits). In addition, the results show that vignetting correction implies a lower variance in the background representation (Fig. 4.5, column b), which allows improving the detection without changing the thresholds of the background subtraction algorithm. The vignetting correction allows on average gains of $-9dB$ of pixel standard deviations between the two mosaics built.



(a) Database

(b) Gain and Variance

(c) Run time

(d) Detections

Figure 4.5: Event detection experiment given a single image and vignetting correction information. Vignetting corrected images are shown in the bottom row, columns a and c. Notice the digits superimposed by a video-projector on the ceiling (column c). Vignetting correction allowed to increase the number of detections without raising the false alarms (column d, bottom row).

4.4 Mosaic Construction

The test described in this section aims at understanding how the geometric calibration affects the mosaic construction.The pan-tilt camera sweeps the background capturing images. In particular, 3501 frames are used for each calibration algorithm, namely camera calibration using pan-tilt odometry and features (Sec. 2.3.2) and image based

camera calibration and pan-tilt odometry scaling (Sec. 2.3.4). The results of this experiment are two cube background representations where the FOV of the cube is the pan-tilt range of the camera. Since the range of the camera is much less then 360×360 there are just some relevant faces of the cube namely the front, right and left face, as it can be seen in Fig 4.6. The gray areas are the ones that can not be seen by the camera at this zoom. Inspecting Fig 4.6 one can conclude that both calibrations allow obtaining visually appealing panoramas. However when it comes to build mosaics for surveillance the calibration from Sec. 2.3.2, using pan-tilt odometry and features, may create some problems because the metrics are distorted. An evident problem created by distorted metrics is when the camera is observing scenes close to ± 180 degrees of vertically or horizontally. In other words the Sec. 2.3.2 will not give accurate result if the camera observes a point at x degrees and than observes the same point but this time with a rotation of $x + 2\pi$.

(a) Calibration using odometry and features

(b) Image based camera calibration, pan-tilt odometry scaling

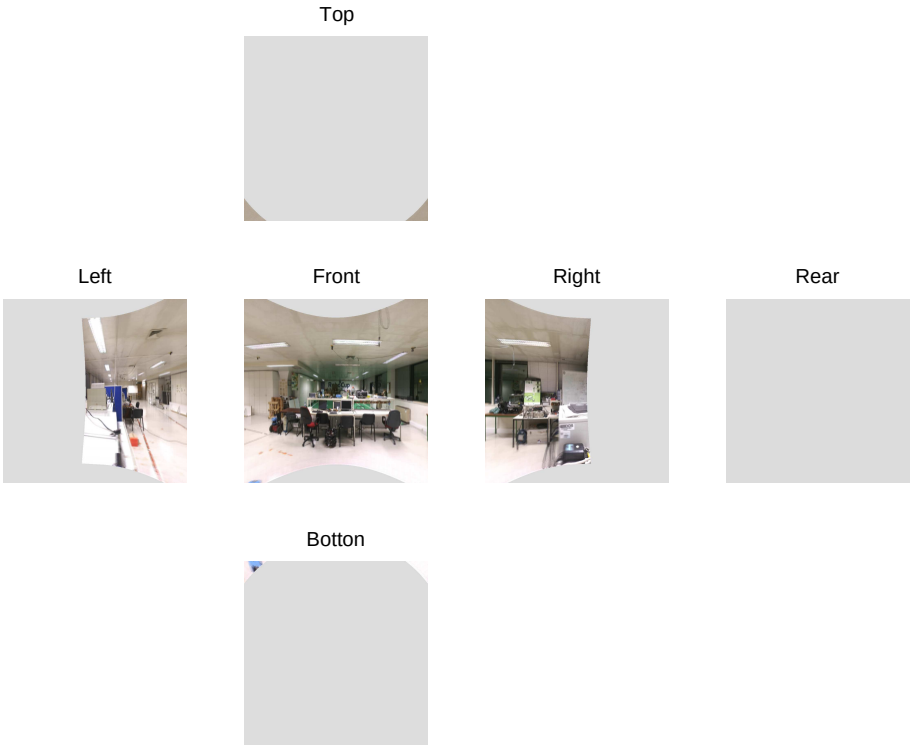
Figure 4.6: Mosaics with different geometric calibration.

4.5 Event Detection based on a Pan-tilt Background Mosaic

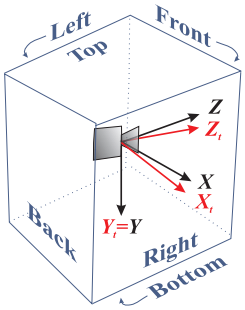
In this experiment the run times images do not have to be acquired at the same poses where the background images were acquired. Instead, one builds first a mosaic representing the background and them acquires the region of interest of the background matching current (run-time) images. This methodology implies that each background image query is in fact a composition of overlapping, previously acquired, images.

Fig. 4.8(b) and (c) show background mosaics built without and with vignetting correction, respectively. These representations result from 347 images, acquired with steps of 2^o in pan and 3.5^o in tilt.

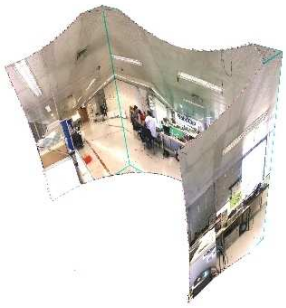
As in Sec. 4.3 the images with events to be detected during the run time were created afterwards using a video projector superimposing text (digits) towards the ceiling of the room. The digits are progressively less



(a) Faces of the cube



(b) Coordinate systems



(c) Left-side view

(d) Frontal view

Figure 4.7: Cube based representation of the scenario. In this case only three faces, left, front and right, are required since the camera field of view is limited to about 180^o and 90^o, in the horizontal and vertical directions, respectively (a). Coordinate systems of the cube, $\{X, Y, Z\}$ and the pan-tilt camera, $\{X_t, Y_t, Z_t\}$ with zero tilt and non-zero pan (b). Left-side and frontal VRML views of the background model are shown in (c) and (d).

(a) Database (b) Database mosaic (c) With vignetting correction

Figure 4.8: Data base construction. Two database images of a set of 347 (a), the database mosaic (b) and the database mosaic obtained after vignetting correction (c)

visible toward the borders in order to test the limits of the proposed event detection methodology. Two run-time images are shown in Fig. 4.9(a). For comparison purposes, a mosaic built from 48 run-time images is shown in Figs. 4.9(b,c), without and with vignetting correction, respectively. The pan and tilt steps are about 5^o and 10.5^o, being therefore much larger than in the database images. Note that the global field of view does not match exactly the one of the background mosaic (database).

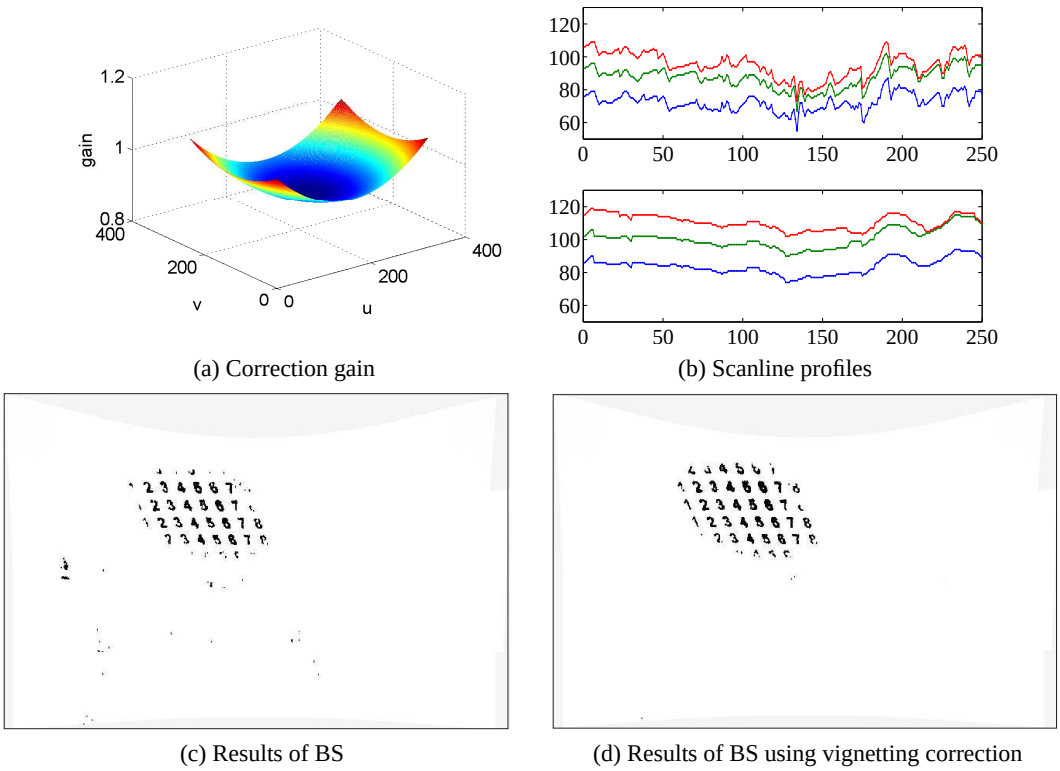
(a) Run time (b) Run time mosaic (c) With vignetting correction

Figure 4.9: Run time view. Two run time images (a), a mosaic built from 47 run time images (b), and the same mosaic with vignetting correction (c).

Event detection methodology is based on comparing the run time images (not the run time mosaics, which in general are not available) with matching database images extracted from the background mosaics, Fig. 4.8(b). In the case of using vignetting correction it is applied to the run time images before comparing them with the background, Fig. 4.8(c).

Figures 4.10(a) and (b) show the estimated vignetting correction function and the change motivated by vignetting correction on a scan-line of a mosaic (we are displaying just $1/3$ of the scan-line). As expected, the vignetting correction gradually enhances (augments) the brightness values when walking towards the image periphery, and the mosaic scan-lines are much smoother after vignetting correction.

Figures 4.10(c) and (d) show correct detections of digits, using or not vignetting correction, however one has more true positives and less false positives when there is used the vignetting correction. The brightness differences motivated by the vignetting when compared with the background, built just keeping at each location the most recent image pixel or averaging all superimposed-image-pixels hitting that location, are significantly more relevant



(a) Correction gain

(b) Scanline profiles

(c) Results of BS

(d) Results of BS using vignetting correction

Figure 4.10: Event detection global results.(a) Vignetting correction gain,(b) scan-lines of figure4.8 (b) and (c), event detections without using vignetting correction (c) and event detections using vignetting correction (d).

than when using the vignetting correction.

Chapter 5

Conclusion and Future Work

The work described within this thesis is focused on improving the construction of visual mosaics with pan-tilt and zoom cameras. An application of the mosaics is video surveillance of mostly static scenes, e.g. museums where any change are to be found as soon as possible. Two main technical questions were originally identified in this work, (i) how to geometrically calibrate the cameras and (ii) how to mitigate the radiometric non-linearities, such as vignetting, on the background subtraction algorithms?

While the geometric calibration has already been shown to be possible to do automatically with a pan-tilt camera, the combination of vision based calibration with the pan-tilt control is still an issue requiring attention. The radiometric non-linearities have been estimated (and then compensated) mainly by varying the exposure time [18], for the case of fixed or translating cameras, i.e. not specifically the pan-tilt cameras.

On the geometric calibration aspect we studied and tested various algorithms. The methods approached were four: (i) camera calibration using a planar chess pattern, (ii) camera calibration using pan-tilt odometry and features, (iii) image based camera calibration and (iv) image based camera calibration and pan-tilt odometry scaling. The method of camera calibration using a planar chess pattern, method (i) is not practical for cameras mounted high above the floor. The methods (ii) and (iv), and consequently (iii) as part of (iv), were the ones mostly used in our experiments. The camera calibration using pan-tilt odometry and features, method (ii), shows good results however it requires cameras equipped with commanding, and feedback, of pan and tilt angles in metric units. Given that most common current pan-tilt surveillance systems encompass precise commanding and feedback of pan and tilt angles, in logic (non metric) units, the image based camera calibration and pan-tilt odometry scaling algorithm, method (iv) is the recommended calibration method. This method provides the scaling factors that allow using pan and tilt angles sent to / read from the camera, instead of having to estimate them constantly with method (iii), and therefore allowing a much faster operation.

In order to estimate and compensate the radiometric non-linearities of a pan-tilt camera we started by constructing a general model of the radiometric process common to all cameras. Then, we paid special attention to vignetting, an effect subject to change with just some tuning of the lens. A model of vignetting was proposed, using a combination of two hyperbolic cosines. To fit the vignetting model to the pan-tilt camera we introduced the concept of a constant radiance image. The concept of constant radiance image proved to be useful not only as for model fitting, but also as a vignetting visualization tool for pan-tilt cameras.

Experiments with real date show that the vignetting correction lowers the variance of the background in $9dB$ (average on the field of view of the camera). Experiments show also that mosaics with vignetting correction

allows the background subtraction algorithm to have an higher percentage of true positives and a lower percentage of false positives, while comparing with uncorrected mosaics, using the same thresholds. Simulated scenarios with vignetting show that vignetting correction improves, lowers, the background variance even if the estimated parameters of the model are not perfectly corrected. These proves that vignetting correction is necessary, to get higher performance in every automatic pan-tilt surveillance systems based on background subtraction algorithms.

In future work we will focus on maintaining minimized variance representations accompanying the daylight change. This leads to the problem of how to update the background while running the surveillance algorithm. Another important issue that was not studied in this work is estimation of the radiometric response function. Its estimation will also improve the vignetting correction, since both problems are correlated. We also plan to extend the framework for multiple pan-tilt-camera scenarios, where questions such as how to define common (comparable) regions of interest for a multiple pan-tilt-camera setup, or how to share comparable information between the cameras, would arise.

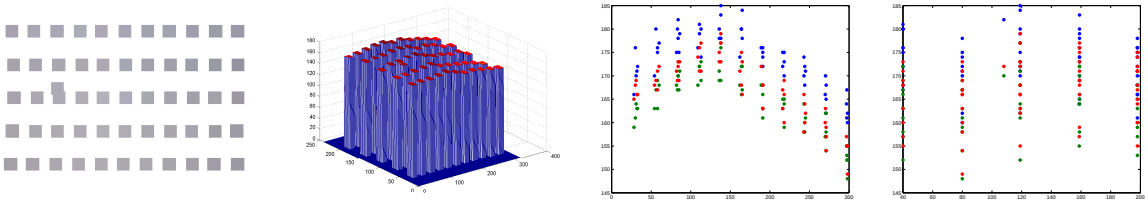
Appendix A

Examples of Constant Radiance Images

In here is shown all the information collected from the experiment on Sec. 4.2. The information is acquired from the patches signed on the figure A.1.

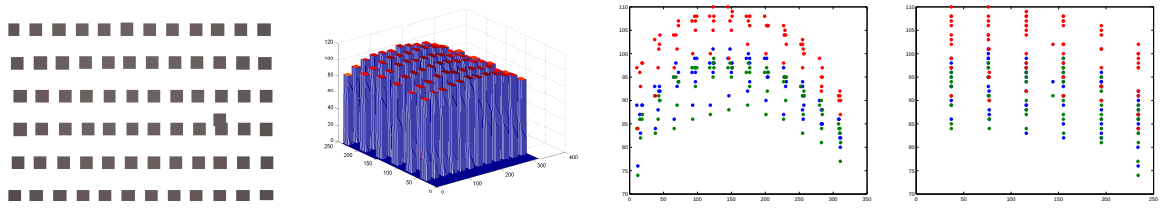
Figure A.1: Constant resolution images (CRI). Image taken at $pan = 0^o, tilt = 0^o$ used to initialize the creation of various CRI, one for each of the patches numbered 1 to 7 and shown in red.

All the images below, Figs[A.2 A.3 A.4 A.5 A.6 A.7 A.8] represent a data collect from a single patch. Each figure as a perspective of the CRI on an image its patch(a). A perspective view of the red component as a mesh plot (b), a RGB values of the CRI versus the horizontal and vertical image coordinates, (c) and (d) respectively.



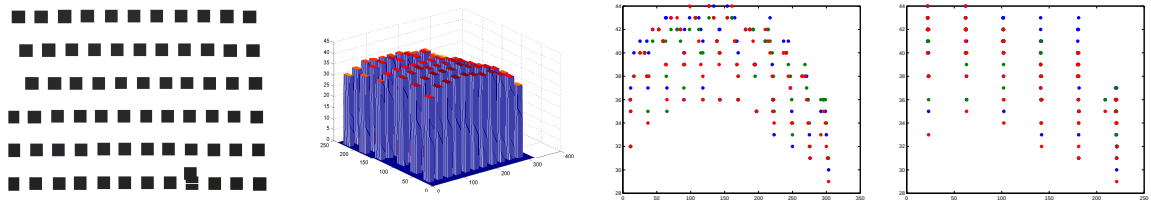
(a) (b) (c) (d)

Figure A.2: Constant radiance image of the patch number 1. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.



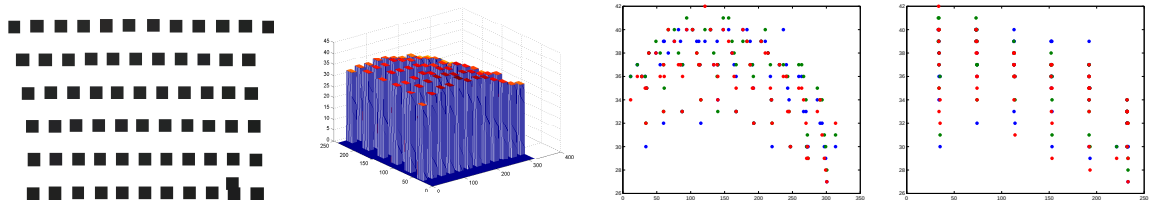
(a) (b) (c) (d)

Figure A.3: Constant radiance image of the patch number 2. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.



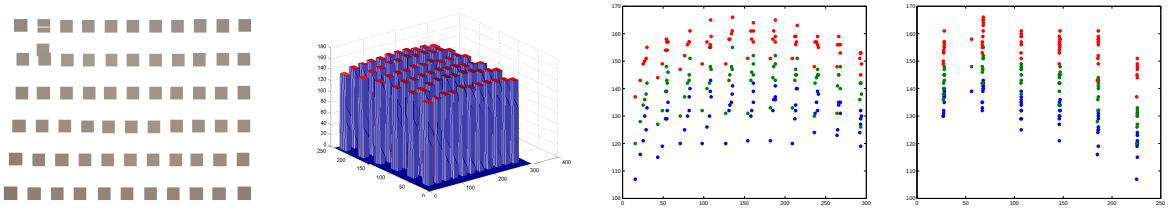
(a) (b) (c) (d)

Figure A.4: Constant radiance image of the patch number 3. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.



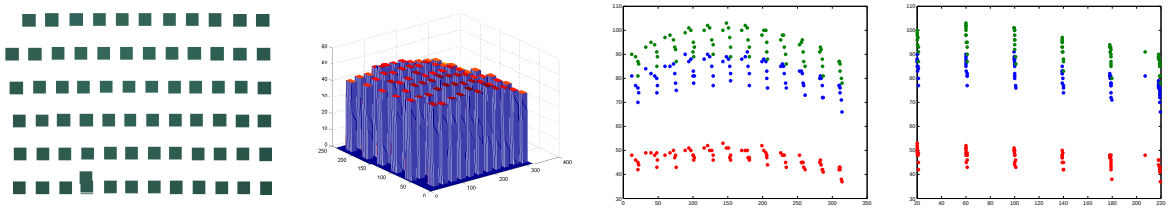
(a) (b) (c) (d)

Figure A.5: Constant radiance image of the patch number 4. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.



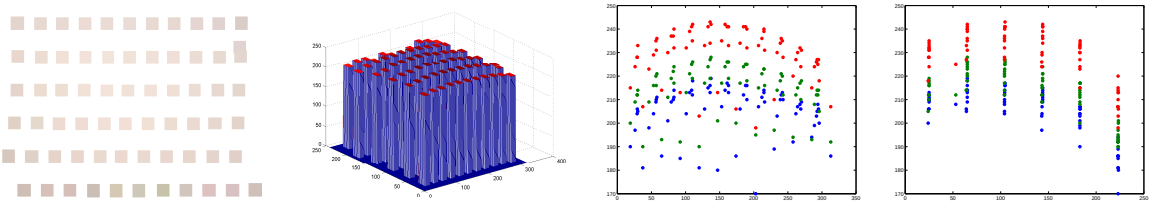
(a) (b) (c) (d)

Figure A.6: Constant radiance image of the patch number 5. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.



(a) (b) (c) (d)

Figure A.7: Constant radiance image of the patch number 6. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.



(a) (b) (c) (d)

Figure A.8: Constant radiance image of the patch number 7. (a) Image of a CRI; (b) Mesh of CRI, red component; (c) Plot of u direction on CRI; (d) Plot of v direction on CRI.

Bibliography

[1] L. Agapito, R. Hartley, and E. Hayman. Linear calibration of a rotating and zooming camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 15–21. IEEE Computer Society, Fort Collins, Colorado, USA, June 1999.

[2] L. De Agapito, E. Hayman, and I. Reid. Self-calibration of a rotating camera with varying intrinsic parameters. In *In Proc 9th British Machine Vision Conf, Southampton*, pages 105–114, 1998.

[3] C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6):33–41, 1984.

[4] Jean-Yves Bouguet. Camera calibration toolbox for matlab. `http://www.vision.caltech.edu/bouguetj`.

[5] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *Int. Journal of Comp. Vision*, 74(1):59–73, 2007.

[6] M. Brown and D.G. Lowe. Recognising panoramas. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1218 –1225 vol.2, oct. 2003.

[7] Sony Corporation. Command list, evi d30/d31. Website. `http://bssc.sel.sony.com/Professional/docs/manuals/evid30commandlist1-21.pdf`.

[8] Ashley Eden, Matthew Uyttendaele, and Richard Szeliski. Seamless image stitching of scenes with large motions and exposure differences. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, pages 2498–2505. IEEE Computer Society, 2006.

[9] Ricardo Galego, Alexandre Bernardino, and José Gaspar. Surveillance with pan-tilt cameras: Background modeling. In *RECPAD 2010. 16^{th} Portuguese Conference on Pattern Recognition*, Oct. 2010.

[10] Ricardo Galego, Alexandre Bernardino, and José Gaspar. Vignetting correction for pan-tilt surveillance cameras. In *VISAPP 2011. International Conference on Computer Vision Theory and Applications*, Mar. 2011.

[11] Dan B. Goldman and Jiun-Hung Chen. Vignette and exposure calibration and compensation. *Computer Vision, IEEE International Conference on*, 1:899–906, 2005.

[12] D.B. Goldman. Vignette and exposure calibration and compensation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12):2276 –2288, dec. 2010.

[13] M.D. Grossberg and S.K. Nayar. What is the space of camera response functions? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume II, pages 602–609, Jun 2003.

[14] M.D. Grossberg and S.K. Nayar. Modeling the space of camera response functions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(10):1272 –1282, oct. 2004.

[15] Richard I. Hartley. Self-calibration from multiple views with a rotating camera. In *In Proceedings of the European Conference on Computer Vision*, pages 471–478. Springer-Verlag, 1994.

[16] Jiaya Jia and Chi-Keung Tang. Image stitching using structure deformation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(4):617 –631, apr. 2008.

[17] Sing Bing Kang and Richard S. Weiss. Can we calibrate a camera using an image of a flat, textureless lambertian surface? In *ECCV (2)*, pages 640–653, 2000.

[18] Seon Joo Kim, J.-M. Frahm, and M. Pollefeys. Joint feature tracking and radiometric calibration from auto-exposure video. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1 –8, Oct. 2007.

[19] Seon Joo Kim and M. Pollefeys. Radiometric alignment of image sequences. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:645–651, 2004.

[20] Seon Joo Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE T-PAMI*, 30(4):562 –576, 2008.

[21] Anat Levin, Assaf Zomet, Shmuel Peleg, and Yair Weiss. Seamless image stitching in the gradient domain. In *In Proceedings of the European Conference on Computer Vision*, 2006.

[22] Stephen Lin, Jinwei Gu, Shuntaro Yamazaki, and Heung-Yeung Shum. Radiometric calibration from a single image. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:938–945, 2004.

[23] Stephen Lin and Lei Zhang. Determining the radiometric response function from a single grayscale image. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:66–73, 2005.

[24] A. Litvinov and Y.Y. Schechner. Radiometric framework for image mosaicking. *J Opt Soc Am A Opt Image Sci Vis*, 22(5):839–48, 2005.

[25] D.G. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150 –1157 vol.2, Sep. 1999.

[26] Y. Matsushita and S. Lin. Radiometric calibration from noise distributions. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –8, jun. 2007.

[27] M. Piccardi. Background subtraction techniques: a review. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 4, pages 3099 – 3104 vol.4, Oct. 2004.

[28] S. N. Sinha and M. Pollefeys. Towards calibrating a pan-tilt-zoom camera network. In *Department of Computer Science, University of North Carolina at Chapel Hill*, pages 91–110, 2006.

[29] Sudipta N. Sinha and Marc Pollefeys. Pan-tilt-zoom camera calibration and high-resolution mosaic generation. *Comput. Vis. Image Underst.*, 103(3):170–183, 2006.

[30] Wonpil Yu. Practical anti-vignetting methods for digital cameras. *Consumer Electronics, IEEE Transactions on*, 50(4):975 – 983, nov. 2004.

[31] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *in ICCV*, pages 666–673, 1999.

[32] Yuanjie Zheng, S. Lin, C. Kambhamettu, Jingyi Yu, and Sing Bing Kang. Single-image vignetting correction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2243 –2256, dec. 2009.

[33] Yuanjie Zheng, Stephen Lin, and Sing Bing Kang. Single-image vignetting correction. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:461–468, 2006.