Geometric and Radiometric Calibration for Pan-Tilt Surveillance Cameras

Ricardo Galego Alexandre Bernardino José Gaspar
Institute for Systems and Robotics, Instituto Superior Técnico / UTL
ricardo.galego@gmail.com, {alex,jag}@isr.ist.utl.pt

***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 paper 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.**

I. INTRODUCTION

Surveillance with pan-tilt cameras is often based on (static) background representations. Whereas in fixed camera settings the background can be modeled with a single image, with pan-tilt cameras we must adopt representations suited to enlarged fields of view. There are various ways to represent geometrically the background [5], [20]. In this paper we use the cube based representation as it allows a complete $360^o \times 360^o$ field-of-view with simple homography transformations. Such a representation can be built by sweeping the camera along the available range of pan-tilt degrees-of-freedom and creating a mosaic of the acquired images projected on the cube. Once the mosaic is built, background differencing can then be used to find intrusions (events), provided one has a good characterization of the uncertainty of the model.

There are two main sources of uncertainty in the process of building a panoramic background representation: inaccurate knowledge of the geometry of the camera and poor characterization of the conversion of scene radiances into image brightnesses. The geometry, defined by 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 [14], [22]. In the following, we tackle both the uncertainty sources.

Geometric and radiometric calibration, are two aspects largely studied and documented in the literature. The work of Hartley [11] 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 homographies concept that link overlapping images acquired by a rotating camera, and allow estimating the intrinsic parameters of the camera and the performed rotations. Following Hartley, Agapito et al. proved that the geometric calibration can also be done for a rotating camera with varying intrinsic parameters, namely zoom [2], [1]. 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 [21]. 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]. In order to avoid over-smoothing image details (edges) Brown and Lowe [6] proposed a multi-band blending of the images, which blends low frequencies over a large range and blends high frequencies just over a short range. An improved manner, proposed in [15], 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 variation of exposures, scene motion and misregistration [7]. More recently there was also developed a way to stitch images with structure distortion [12]. Although these methods render visually appealing mosaics, they do not take into account the utilization of such representations for surveillance applications, such as the ones based on background subtraction where one needs to have similar background and run-time images.

Hence, many researchers invested in understanding the physical reasons for the differences found at the image stitching seams, and stitching methodologies started to encompass radiometric calibration to estimate the radiometric response and vignetting functions of a camera. Grossberg and Nayar introduced a camera response model based on a large database of response functions obtained from well controlled illumination and color pattern setups [10]. 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

(a) Coordinate systems (b) Sample images (c) Cube faces (d) VRML view

Fig. 1: Cube based background representation. (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) A number of the images captured to build a mosaic. (c) Left, front and right cube mosaicked-faces. (d) VRML view of the cube model showing one of the acquired images.

[18]. Goldman and Chen [8] and Goldman [9] 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, having in mind applications that need to compare images acquired with different settings [13]. Later, they proposed finding the radiometric and vignetting response functions, also from a set of images with different and unknown exposures values, but specifically targeting the construction of high dynamic range mosaics [14]. In both works the exposure time is required to change in order to allow estimating the response function. Lin et al. proposed to estimate the radiometric response function from images without changes in the exposure, using histograms of the edges regions [16], [17], [19]. [17]. However, vignetting is not considered in the estimation of the radiometry. Zheng et al. also proposed the correction of vignetting from a single image, but requiring large piecewise flat regions in the image, which is highly dependent on the scene contents [25], [24]. Alternatively, Wonpil Yu proposed to correct vignetting based on a white pattern [22]. The white image, decreasing in brightness towards the borders due to a vignetting distortion function, was approximated with a $2D$ hypercosine function. This calibration methodology is however cumbersome due to the requirement of having to use 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.

In our work we propose using the geometric calibration procedures appropriate for pan-tilt cameras, and propose exploring the pan and tilt degrees of freedom for the radiometric calibration, instead of requiring large constant color areas in the scenarios, or color calibrating patterns.

This paper is organized as follows: Section 1 introduces the problem, and presents a short discussion on the related work Section 2 describes the geometrical model and the background representation for pan-tilt cameras, Section 5 discusses and proposes methodologies to calibrate the camera geometrically, Section 4 discusses and proposes methodologies to correct the effect of vignetting on the background variance and event detection, Section 5 shows experiments testing the proposed methodologies, and finally Section 6 summarizes the work performed and highlights the main achievements in this work.

II. PANORAMIC SCENE REPRESENTATION

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 that are operations much simpler than, for example, the spherical mappings (see Fig. 1).

Building the cube based representation is a two steps process: (i) obtaining a back-projection for each image point and (ii) projecting the back-projection to the right face of the cube. If one knows the intrinsic parameters matrix, K and the orientation R of the camera, then each image point, m can be easily back-projected to a 3D world point:

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

Projecting the world point to the right face of the cube involves determining the face, namely front, back, left, right, top or bottom (see Fig. 1a), and then computing the 2D coordinates within that face. The cube face where to project a world point is 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 $v \equiv x$, $v \equiv -x$, $v \equiv y$, $v \equiv -y$, $v \equiv z$ or $v \equiv -z$, respectively.

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 [1] then

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

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

[1] 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).

3D points closest to each of the faces of the cube towards the front face. In more detail,

$$R_{WF} = \begin{cases} I_{3\times3} & \text{if front face;} \\ Rot_Y(180^o) & \text{if back face;} \\ Rot_Y(-90^o) & \text{if left face;} \\ Rot_Y(+90^o) & \text{if right face;} \\ Rot_X(-90^o) & \text{if top face;} \\ Rot_X(+90^o) & \text{if bottom face.} \end{cases}$$

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

$$m_{Fi} \sim K_F R_{WF} R^{-1} K^{-1} m_i \tag{3}$$

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

Final note, in order to map an image into the background (cube), 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.

III. GEOMETRIC CALIBRATION

The geometric calibration of a pan-tilt camera involves, in essence, estimating the intrinsics parameters matrix,

$$K = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where (u_0, v_0) denotes 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 this section we review and propose a number of calibration and auto-calibration methodologies.

A. 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. This method is based on Zhang's work [23], which explores the characteristics of the image of the absolute conic (IAC). This method is not automatic as it needs a person to show the chess pattern. Additionally, in order to obtain precise results, the pattern has to occupy a large area of the image. This is an issue in practice for example when a surveillance camera is mounted high above the ground, as in that case the person holding the chess pattern cannot come close to the camera and is too cumbersome making and moving very large calibration patterns.

Hence, in this work we propose using alternative calibration methodologies, which are based on the texture existing on the scene and use the mobility of the pan-tilt camera. In other words, we propose using self-calibration methodologies.

B. Camera calibration using pan-tilt odometry and features

This method is automatic as it uses the texture of the scene and the odometry of the camera making unnecessary the person showing the chess pattern. The method is based on matching scene points observed by the pan-tilt camera placed at different orientations. 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.

We optimize K as 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:

$$\vartheta^* = arg_\vartheta min \sum_i \|R_1^{-1} h(K_{(\vartheta)}{}^{-1} m_{1i}) - R_2^{-1} h(K_{(\vartheta)}{}^{-1} m_{2i})\|^2 \tag{5}$$

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. The optimization is done using the Levenberg-Marquardt optimization methodology.

C. 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 estimates the intrinsic 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 $P_i = K_i R_i$, we can calculate the relation between the images $m_i = P_i P_j^{-1} m_j$. The relation between images is therefore a projective transformation, the so called *infinite homography*,

$$H_{ij} = P_i P_j{}^{-1} = K_i R_i R_j{}^{-1} K_j{}^{-1} = K_i R_{ij} K_j{}^{-1}. \tag{6}$$

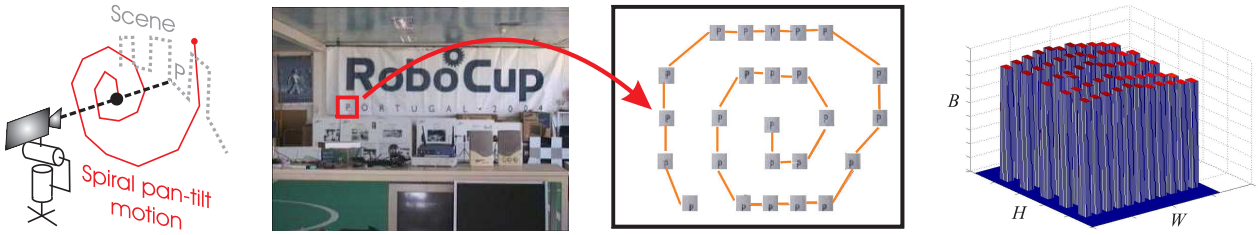
As R_{ij} is a rotation matrix, it satisfies $R_{ij} = R_{ij}{}^{-T}$ and thus one obtains

$$H_{ij} K_j K_j^T H_{ij}^T = K_i K_i^T \tag{7}$$

a constraint mixing the intrinsic parameters matrices with the homography between two images. In the cases that $K_i = K_j = K$, i.e. images with constant zoom, Eq.(7) can be solved in a least squares sense as a set of linear equations, $Ax = 0$, in the entries of $K^{-T}K^{-1}$ (details in [1]).

D. 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, but in the end one wants, additionally, to take advantage of the precise odometry provided by the camera.



(a) Pan-tilt spiral (b) One image sample and the CRI (c) The CRI as a 3D mesh

Fig. 2: Building a Constant Radiance Image (CRI). Camera motion (a) and sample (patch) collection (b). (c) Regular grid of patches, with brightness B, in a $W \times H$ CRI, obtained with approx. constant pan-tilt steps (d).

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

$$R_{ij} = K^{-1} H_{ij} K. \tag{8}$$

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 $\rho = [\rho_1 \ \rho_2]$,

$$\rho^* = arg_\rho min \sum_{ij} \| R(\rho_1 \alpha_{ij}, \rho_2 \beta_{ij}) - R_{ij} \|_F^2 \tag{9}$$

where $R(\rho_1 \alpha_{ij}, \rho_2 \beta_{ij})$ denotes a rotation matrix built from the pan and tilt angles, α_{ij} and β_{ij} read from the odometry and scaled by ρ. We use the Levenberg-Marquardt algorithm to optimize Eq.9. 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.

IV. UNCERTAINTY ANALYSIS AND EVENT DETECTION

In this section we describe the radiometric model of the 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.

The effect of the radiometric response function and vignetting in the image formation process can be described as [10], [22], [14], [24] :

$$I(m) = f(kV(m)L(m)) \tag{10}$$

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 $f(.)$ 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 ($V(m) < 1$).

The patternless methodology for estimating vignetting is based on a mosaicked image, the Constant Radiance Image, which is composed from a number of images taken at various pan-tilt poses. The construction of this composed image is described in the following.

A. Constant Radiance Images

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. 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:

$$\begin{aligned} C_{mo}(m_i) &= I_i(m_i) \\ &= I_i(K R_i R_o^{-1} K^{-1} m_o). \end{aligned} \tag{11}$$

Figure 2 shows the construction of one C_{mo}, and illustrates the typical aspect of a vignetting effect.

Assuming that (i) one estimates the radiometric response function $f(.)$, using e.g. the method in [16], 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, $C_{m0}(m) = f^{-1}(f(kV(m)L(m))) = V(m)L(m)$, and the vignetting gain can be estimated as $V(m) = C_{mo}(m)/max(C_{mo}(m))$.

Note that this is a convenient method to estimate the vignetting as one does not have to image a calibration pattern as in [22]. 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.

B. Vignetting Correction

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:

$$I_c(m) = V_c(m)I(m) = V_c(m)V(m)L(m) \tag{12}$$

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. [14]). In this work we follow the suggestion of [22]:

$$V_c(m;a) = cosh(a_1(u-u_p))cosh(a_2(v-v_p)) + a_3 \tag{13}$$

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:

$$a^* = arg_a min \sum_m \left(\frac{max(C_{mo}(m))}{C_{mo}(m)} - V_c(m;a) \right)^2 \tag{14}$$

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

C. Radiometric Background Modeling

Given $V_c(m)$, we can now correct all acquired images, $I(m) := V_c(m;a)I(m)$, which is beneficial for the construction of panoramic background representations.

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 eack background location M, respectivelly $\mu_{B(M)}$ 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, $\sigma^2_{V(m)}$:

$$B(M) = L(M)V(m) + \eta \tag{15}$$

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:

$$\begin{cases} \mu_{B(M)} = L(M)\ \mu_{V(m)} \\ \sigma^2_{B(M)} = L^2(m)\ \sigma^2_{V(m)} + \sigma^2_\eta \end{cases} \tag{16}$$

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 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, $\sigma^2_{B(M)} = L^2(M) \times 0 = 0$.

D. Event Detection

Event detection is done by comparing a currently captured image, vignetting-corrected, $I(m)$ with the corresponding image retrieved from the background database, $B(m)$, using the background variance, $\sigma^2_{B(m)}$ as a normalizing factor:

$$D(m) = \left((I(m) - B(m))^2 / \sigma^2_{B(m)} \right)^{1/2} \tag{17}$$

A pixel m is considered active, i.e. foreground, if $D(m) \geq 3$.

V. EXPERIMENTS

This section describes two experiments one testing the relationship between the variance of the vignetting gains within a simulated white scenario such that the images exhibit directly the vignetting effect, and the second experiment consisting of event detection on a real setup.

A. Simulated White Scenario

In this simulated experiment we test the relationship between scene luminance $L(M)$, vignetting $V(m)$, vignetting correction $V_c(m)$, and the background variance $\sigma^2_{B(m)}$ as presented in Eq.16. The scene radiance (luminance), $L(M)$ is constant. In order to have a reference vignetting distortion, $V(m)$ similar to a real one, we use a set of parameters $a_r = [a_{r1} a_{r2} a_{r3}]^T$ estimated for the vignetting correction gain on a real camera, $V_c(m)$, and then define $V(m;a_r) \doteq 1/V_c(m;a_r)$.

In our experiment we vary both the vignetting and the vignetting correction gains in a parametric manner, by scaling the parameters, i.e. using $V(m;\alpha a_r)$ 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.

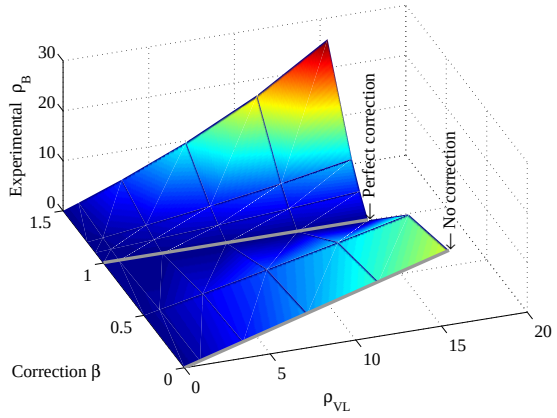


Fig. 3: Background standard deviation vs vignetting effects and corrections. The correction β is adimensional, while ρ_{VL} and ρ_B have their dimensions defined in a 8 bits gray-scale space.

Figure 3 shows in the vertical axis an experimental estimate of the standard deviation of a background pixel, $\rho_{B(m)} = \sqrt{\sigma^2_{B(m)}}$, obtained empirically from the cube based representation constructed 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 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.16 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.

B. Event Detection in a Real Scenario

In our experiments we use a Sony EVI D30 to scan a room and create two background representations: one lacking and the other one having vignetting-correction (Fig. 4(b) and (c), resp.). The background representations result from 347 images, acquired with steps of 2^o in pan and 3.5^o in tilt.

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 visible toward the borders in order to test the limits of the proposed event detection methodology. Two run-time imagesare shown in Fig. 4(d). For comparison purposes, a mosaic built from 48 run-time images is shown in Figs. 4(e,f), 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. The fields of view do not match exactly the ones of the images of the database.

Event detection involves comparing the run time images with the matching database images extracted from the background mosaics (Fig. 4(c)). In the case of using vignetting correction it is applied to the run time images before comparing them with the background (Fig. 4(c)).

Figures 4(g,h) 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(i) and (j) 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 than when using the vignetting correction. This is justified by observing that an average reduction of about $9dB$ in the mosaic and run time images associated to vignetting correction.

VI. CONCLUSION

In this paper we proposed a vignetting correction method for pan-tilt cameras. Experiments have shown that the correction allows building (mosaicked) scene representations with less variance and therefore more effective for event detection. Future work will focus on maintaining minimized variance representations accompanying the daylight change.

REFERENCES

[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] 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.

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

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

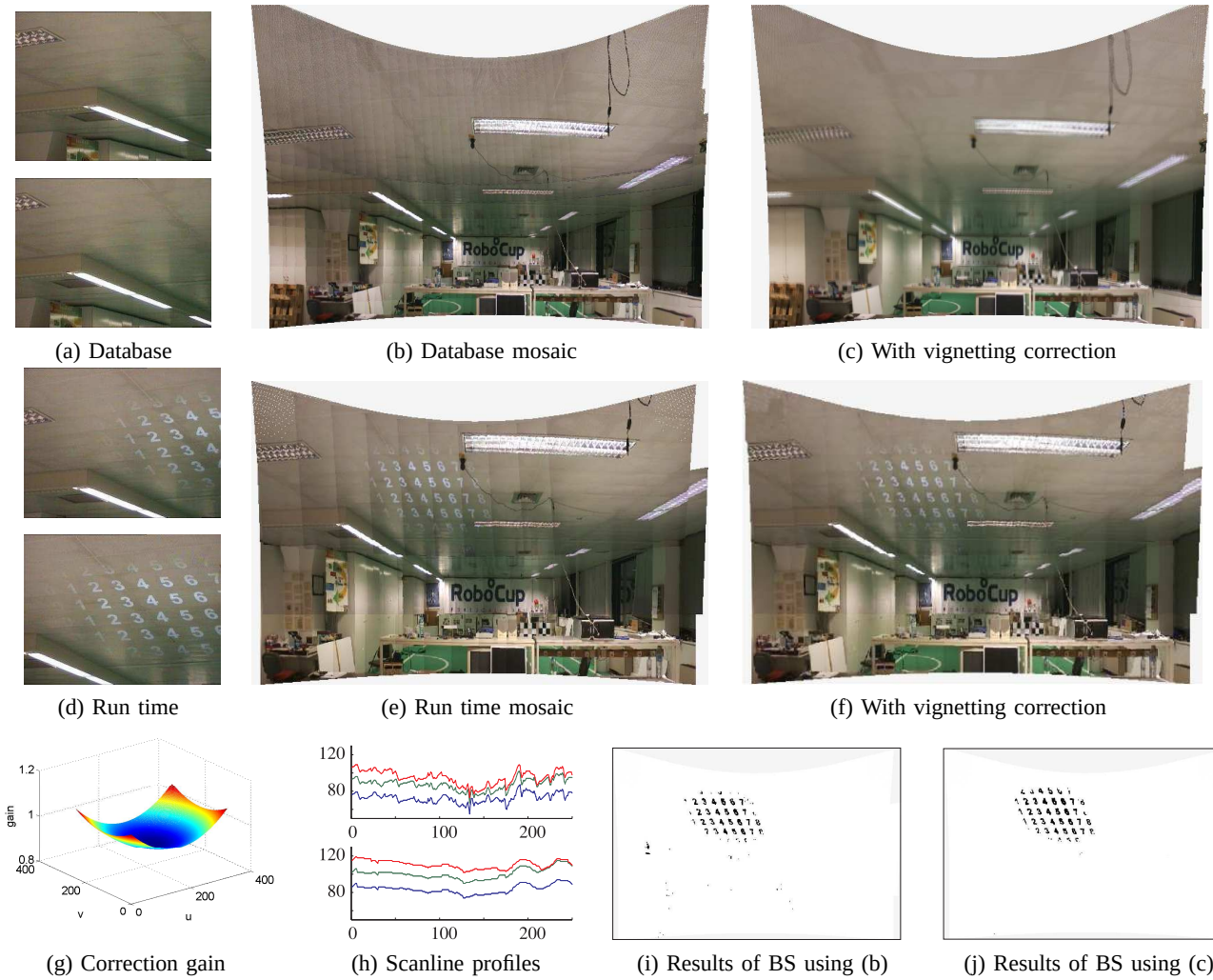
[10] M.D. Grossberg and S.K. Nayar. What is the space of camera response functions? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume II, pages 602–609, Jun 2003.

[11] 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.

[12] 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.

[13] 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.

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



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

(d) Run time (e) Run time mosaic (f) With vignetting correction

(g) Correction gain (h) Scanline profiles (i) Results of BS using (b) (j) Results of BS using (c)

Fig. 4: Event detection experiment. (a,b,c) Two database images of a set of 347, the database mosaic before and after vignetting correction. (d,e,f) Two run time images, a mosaic built from 47 run time images, and the same mosaic with vignetting correction. (g) Vignetting correction gain. (h) Top and bottom plots are scanlines of (b) and (c), respectively. (i,j) Display in mosaics of events found in the run time images (some examples in (d)) using the database mosaics (b) and (c).

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

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

[24] 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.

[25] 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.