VIGNETTING CORRECTION FOR PAN-TILT SURVEILLANCE CAMERAS

Ricardo Galego, Alexandre Bernardino and José Gaspar

Institute for Systems and Robotics, Instituto Superior Técnico / UTL, Lisboa, Portugal

Keywords: Image formation, Vignetting correction, Pan-Tilt cameras, Visual event detection, Surveillance.

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.

1 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. 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 nonlinear-transformation of the radiometric readings. 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 (Kim and Pollefeys, 2008; Yu, 2004). In the following, we tackle both the uncertainty sources.

Geometric and radiometric calibration, are two as-

pects largely studied and documented in the literature. Hartley (Hartley, 1994) introduced the infinite homographies that link overlapping images acquired by a rotating camera, and allow estimating the intrinsic parameters of the camera and the performed rotations. Agapito et al. proved that the geometric calibration can also be done for a rotating camera with varying intrinsic parameters (zoom) (Agapito et al., 1999). Sinha and Pollefeys apply the same concept to estimate how to stitch images acquired by a system of multiple pan-tilt cameras and therefore build a panorama in a collaborative manner (Sinha and Pollefeys, 2006).

Methodologies for image blending, such as *feathering*, have been proposed quite early (Anderson et al., 1984). Brown and Lowe (Brown and Lowe, 2003) 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 (Levin et al., 2006), was based on minimizing a cost function designed in the gradient domain. 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.

More recent research focused in understanding the physical reasons for the differences found at the image stitching seams. Stitching methodologies started
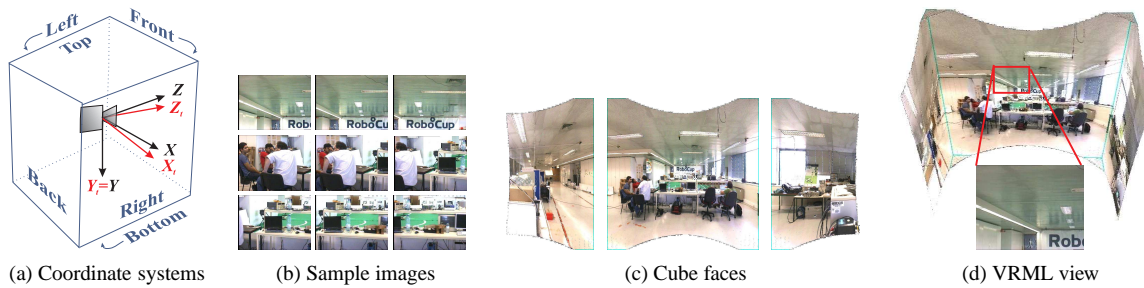
| (a) Coordinate systems | (b) Sample images | (c) Cube faces | (d) VRML view |

Figure 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.

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 (Grossberg and Nayar, 2003). Kim and Pollefeys proposed estimating the vignetting and radiometric response functions for high dynamic range mosaics, from a set of images with different exposures values (Kim and Pollefeys, 2008). Lin et al. proposed to estimate the radiometric response function from images without changes in the exposure, using histograms of the edges regions (Lin and Zhang, 2005). 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 (Zheng et al., 2009).

Alternatively, Wonpil Yu proposed to correct vignetting based on a white pattern (Yu, 2004). 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 this work we propose therefore using the geometric calibration procedures adapted to pan-tilt cameras, and propose exploring the pan and tilt degrees of freedom instead of requiring large constant color areas in the scenarios, or color calibrating patterns.

This paper is organized as follows: Section 2 describes the geometrical model and the background representation for pan-tilt cameras, Section 3 discusses and proposes methodologies to correct the effect of vignetting on the background variance and event detection, Section 4 shows experiments testing the proposed methodologies, and finally Section 5 summarizes the work and draws some conclusions.

2 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 (as compared to using 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 $[x\ y\ z]^T = (KR)^{-1}m$. 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 $v \equiv x, -x, y, -y, z$ or $-z$, 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. The rotation matrices R_{WF}

(a) Pan-tilt spiral   (b) One image sample and the CRI   (c) The CRI as a 3D mesh
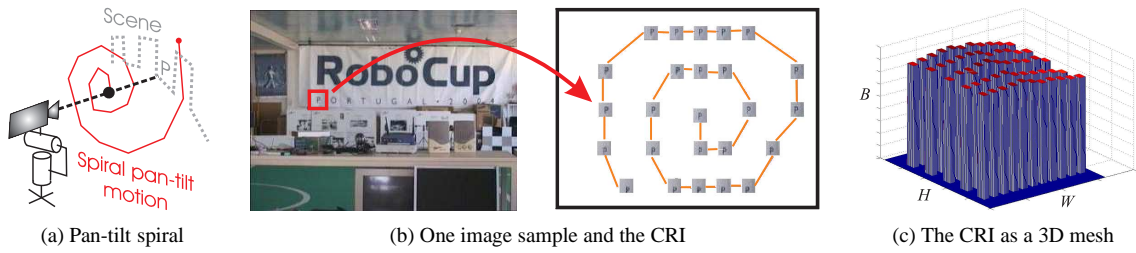
Figure 2: Building a Constant Radiance Image (CRI). Camera motion (a) and sample (patch) collection (b). (c) Regular grid of patches, with brightness B, in a $W \times H$ CRI, obtained with approx. constant pan-tilt steps (d).

in essence rotate the 3D points closest to each of the faces of the cube towards the front face. For example R_{WF} is $Rot_Y(180^o)$ or $Rot_X(-90^o)$, for the back or top cube faces, respectively. In summary, an image point m_i is mapped to a point on a cube face m_{Fi} as:

$$m_{Fi} \sim K_F R_{WF} R^{-1} K^{-1} m_i \qquad (1)$$

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.

3 UNCERTAINTY ANALYSIS AND EVENT DETECTION

In this section we describe 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.

The effect of the radiometric response function and vignetting in the image formation process can be described as (Grossberg and Nayar, 2003; Yu, 2004; Kim and Pollefeys, 2008; Zheng et al., 2009) :

$$I(m) = f(kV(m)L(m)) \qquad (2)$$

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.

3.1 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:

$$\begin{aligned} C_{mo}(m_i) &= I_i(m_i) \\ &= I_i(KR_iR_o^{-1}K^{-1}m_o). \end{aligned} \qquad (3)$$

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 (Lin et al., 2004), then one can remove the effect of $f(.)$ by redefining $I(m) \to f^{-1}(I(m))$, (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 $V(m) = C_{mo}(m)/max(C_{mo}(m))$.

3.2 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:

$$I_c(m) = V_c(m)I(m) = V_c(m)V(m)L(m) \qquad (4)$$

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. (Kim and Pollefeys, 2008)). In this work we follow the suggestion of (Yu, 2004):

$$V_c(m;a) = cosh(a_1(u - u_p))cosh(a_2(v - v_p)) + a_3 \qquad (5)$$

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:

$$a^* = arg_a min \sum_m \left(\frac{max(C_{mo}(m))}{C_{mo}(m)} - V_c(m;a) \right)^2 \qquad (6)$$

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

3.3 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, $\sigma^2_{V(m)}$:

$$B(M) = L(M)V(m) + \eta \qquad (7)$$

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

$$\begin{cases} \mu_{B(M)} = L(M)\,\mu_{V(m)} \\ \sigma^2_{B(M)} = L^2(m)\,\sigma^2_{V(m)} + \sigma^2_\eta \end{cases} \qquad (8)$$

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

3.4 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:

$$D(m) = \left((I(m) - B(m))^2 / \sigma^2_{B(m)} \right)^{1/2} \qquad (9)$$

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

4 EXPERIMENTS

This section describes two experiments: (i) 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 (ii) event detection on a real setup.

4.1 Simulated White Scenario

In this experiment the scene luminance, $L(M)$ is constant. The 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). 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, we vary both both of them in a parametric manner, by scaling the parameters. More precisely, we use $V(m; \alpha a_r)$ and $V_c(m; \beta a_r)$, with $\alpha \in \{0, .3, .6, 1, 1.3, 1.6\}$ and $\beta \in \{0, .5, 1, 1.1, 1.2, 1.5\}$. 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 shows in the vertical axis an experimental estimate of the standard deviation of a background pixel, $\rho_{B(m)}$, which is obtained as the square root of the variance, $\sigma^2_{B(m)}$ estimated empirically from the cube based representation constructed from a set of images acquired at various pan-tilt poses. 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.8 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.
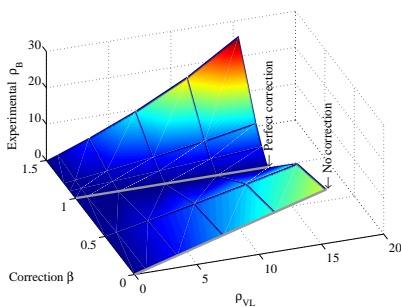

Figure 3: Background standard deviation vs vignetting effects and corrections. The correction β is adimensional, while ρ_{VL} and ρ_B have their dimensions defined in a 8 bits gray-scale space.

4.2 Event Detection in a Real Scenario

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.). These representations result from 347 images, acquired with approximately 2° in pan and 3.5^o in tilt steps.

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 images are 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° 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.

Our 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 (Figs. 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,j) show correct detections of digits, using or not vignetting correction, however one sees more detections (true positives) and less false detections (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.

5 CONCLUSIONS

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.

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

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

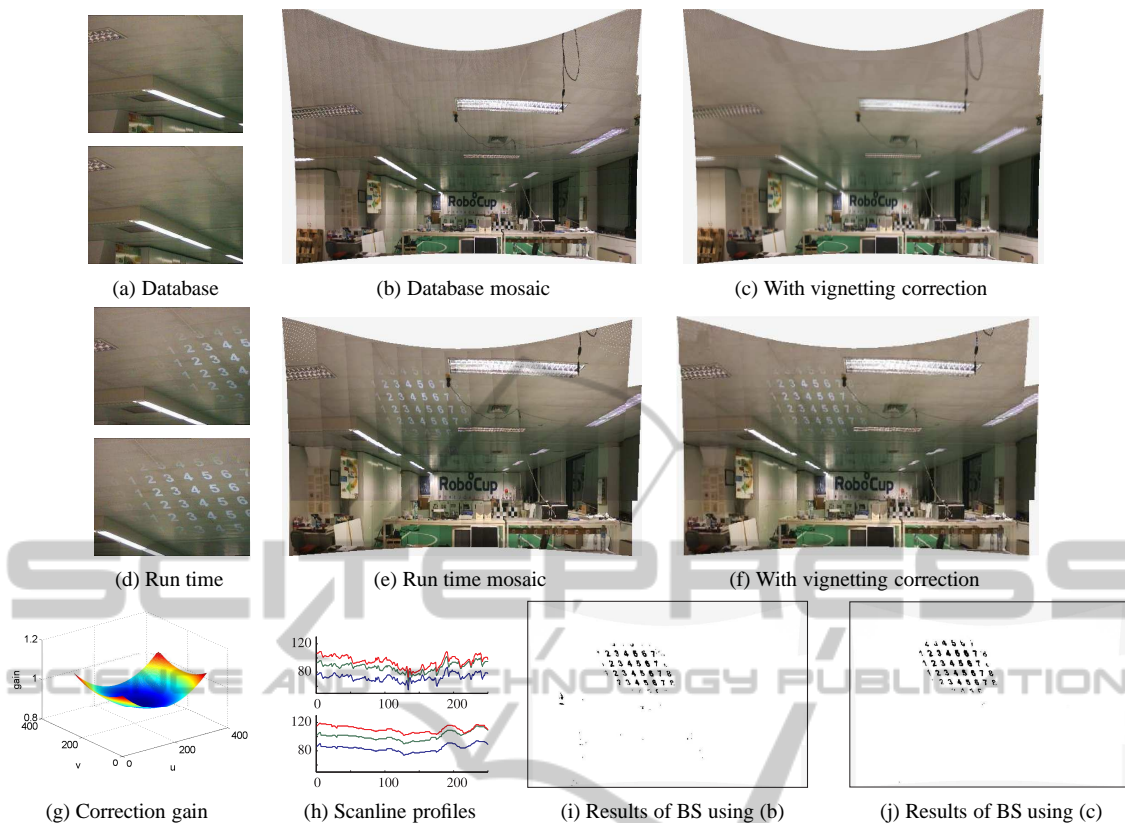(g) Correction gain   (h) Scanline profiles   (i) Results of BS using (b)   (j) Results of BS using (c)

Figure 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).

ACKNOWLEDGEMENTS

REFERENCES

Agapito, L., Hartley, R., and Hayman, E. (1999). 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.

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

Brown, M. and Lowe, D. (2003). Recognising panoramas. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1218 –1225 vol.2.

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

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

Kim, S. J. and Pollefeys, M. (2008). Robust radiometric calibration and vignetting correction. *IEEE T-PAMI*, 30(4):562 –576.

Levin, A., Zomet, A., Peleg, S., and Weiss, Y. (2006). Seamless image stitching in the gradient domain. In *In Proceedings of the European Conference on Computer Vision*.

Lin, S., Gu, J., Yamazaki, S., and Shum, H.-Y. (2004). Radiometric calibration from a single image. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:938–945.

Lin, S. and Zhang, L. (2005). Determining the radiometric response function from a single grayscale image. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:66–73.

643

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

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

Zheng, Y., Lin, S., Kambhamettu, C., Yu, J., and Kang, S. B. (2009). Single-image vignetting correction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2243 –2256.