ENHANCED MOSAIC BLENDING USING INTRINSIC CAMERA PARAMETERS FROM A ROTATING AND ZOOMING CAMERA

Dae-Woong Kim and Ki-Sang Hong

Dept. of E.E. in POSTECH, Pohang, Korea
{dwkim, hongks}@postech.ac.kr

Abstract

This paper presents a global approach to mosaic blending, reconstructing original scene radiances up to scale, and simultaneously estimating intrinsic camera parameters. By incorporating a geometric constraint of rotating and zooming cameras and a constraint of non-linear mapping functions into robust optimization problems, we ascertained optimal blending parameters for multiple images.

1. INTRODUCTION

Most work in image mosaics focuses on recovering a projective transformation between a pair of images captured by a rotating and zooming camera [1][2]. Although the estimation of the projective transformations is accurate, the creation of a high quality mosaic requires color and luminance correction to avoid seams due to photometric variations across image boundaries. Traditionally, the objective of blending is to eliminate such boundary seams using intensity information in the overlapping region of a mosaic [2][3][4]. However, the original factors causing the boundary seams were not considered. Boundary seams in the mosaic are caused by some non-linear mapping properties of cameras. Therefore, once parameters of the mapping function are estimated, the blending would be a simple average problem.

Toward this goal, in this paper, we present a new blending method by self-calibrating intrinsic camera parameters, and simultaneously reconstructing original scene radiances up to scale from the motion information. There have been some efforts to reconstruct scene radiances with a still camera [5][6][7]. Closer to our paper is the work of Hasler and Süsstrunk [8] which proposed methods that can be used to correct color mismatches between images. However, they used only pairwise local relations while our method is a global approach in that it uses multiple images in a single framework by estimating camera parameters and scene radiances simultaneously with a more accurate model. Also, constraints related to rotating and zooming cameras and non-linear mapping functions are introduced. To estimate parameters robust to outliers such as misaligned pixels or moving objects in the overlapping region, we adapted a skipped mean estimator, and incorporated it with Levenberg Marquardt (LM) optimization along with the bundle adjustment technique. We call our method *enhanced blending* because the final blended mosaic is seamless and clear.

Linear and non-linear effects causing discontinuities in blending are briefly described in section 2 and the estimation algorithm is explained in detail in section 3. With experiments in section 4, we conclude our work with some remarks in section 5.

2. RADIOMETRIC RESPONSE

Most vision algorithms assume that the scene radiance (L) at a point on an object is linearly related to the irradiance (E) at the corresponding point in the image sensor array, given by

$$E = \left[\frac{\pi}{4}\left(\frac{d}{f}\right)^2 \cos^4\theta\right] L, \tag{1}$$

where d is the diameter of lens, f is the focal length of the camera, and θ is the angle that the ray from the image point to the center of the lens makes with the optical axis. The parameters are depicted in Fig 1 (b), where $\cos\theta$ is represented by the focal length and distance (r) from the image center (or the principal point): $\cos\theta = \sqrt{f^2/(f^2+r^2)}$.

In addition, most photographic lenses exhibit optical vignetting to some degree which is caused by a partial obstruction of light from the object space to image space. The geometry of vignetting for two-lens system is shown in Fig 1 (a) where the shaded part of the beam never reaches the second lens, causing the irradiance to drop in the image periphery. Assuming that entrance and exit pupils are circular, and the off-axis angle θ is small, the loss of light due to vignetting can be expressed as the approximation [7] given by

$$E = (1 - \alpha r)L, \tag{2}$$

where $\alpha(> 0)$ is the darkening coefficient.

This work was partly supported by the Korea Research Foundation and the Brain Korea 21 Project.

The relationships in Equations (1) and (2) show that irradiance is proportional to scene radiance. Also, irradiance falls off as the light ray deviates from the optical axis due to $\cos^4\theta$ and the vignetting term.

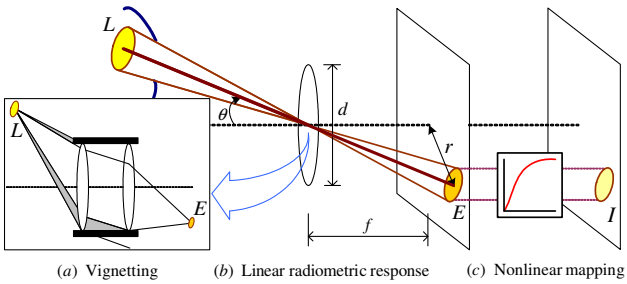


Fig. 1. Relations between intensity (I) and radiance (L).

Finally, the irradiance reaching the image sensor is non-linearly mapped to produce high quality image intensity I depending on the image acquisition devices [5][6][8], as shown in Fig 1 (c).

Let i be an image index, t_i the exposure time of I_i, $\mathbf{x}_k$ a scene point, $\mathbf{x}_{ik}$ the corresponding pixel in I_i, and g the non-linear mapping function of the camera. Then, the final relation between intensity $I_i(\mathbf{x}_{ik})$ and scene radiance $L(\mathbf{x}_k)$ can be expressed as

$$E_r = I_i(\mathbf{x}_{ik}) - g\big(\mathcal{G}_i(\mathbf{x}_{ik})L(\mathbf{x}_k)\big) = 0, \tag{3}$$

where

$$\mathcal{G}_i(\mathbf{x}_{ik}) = \frac{\pi}{4} t_i (1 - \alpha_i r_i(\mathbf{x}_{ik})) \Big(\frac{d_i f_i}{f_i^2 + r_i(\mathbf{x}_{ik})^2} \Big)^2. \tag{4}$$

3. SELF-CALIBRATION FROM A ROTATING AND ZOOMING CAMERA

In this section, we describe the overall procedure of our algorithm. For simplicity, consider two images I_i and I_j obtained from a rotating and zooming camera. Then, two matching points $\mathbf{x}_{ik}$ and $\mathbf{x}_{jk}$ are on the same ray in 3D space passing through the camera center, as shown in Fig 2. The point relationship can be represented by a 3×3 homography, H_{ij}, given by

$$\mathbf{x}_{jk} = H_{ij}\mathbf{x}_{ik} = K_j R_{ij} K_i^{-1} \mathbf{x}_{ik}, \tag{5}$$

where R_{ij} is the rotation matrix between two images and K_i is the camera calibration matrix in the form of

$$K_i = \begin{bmatrix} f_i & s f_i & p_x \\ 0 & \gamma f_i & p_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

The parameters in K_i represent the properties of the image formation system: f_i is the focal length of I_i, γ is the aspect ratio, s is the skew, and (p_x, p_y) is the principal point of the camera. It is noted that the intrinsic parameters in K_i are closely related to the distance r_i, given by

$$r_i(\mathbf{x}_i) = \sqrt{\Big[x_i - p_x - \frac{s}{\gamma}(y_i - p_y)\Big]^2 + \Big[\frac{1}{\gamma}(y_i - p_y)\Big]^2}, \tag{7}$$

where ($\mathbf{x}_i = [x_i, y_i]^T$) is the image coordinate in I_i.

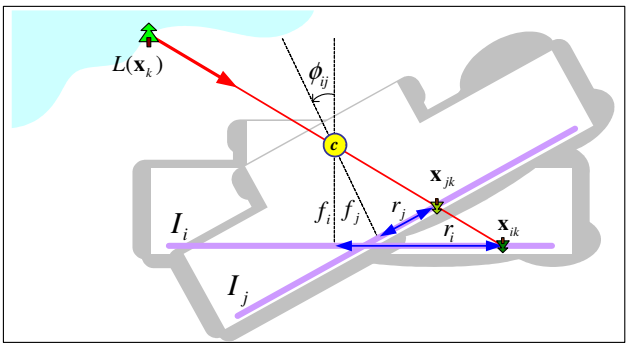


Fig. 2. Geometry of a rotating (ϕ_{ij}) and zooming camera.

In Fig 2, we observe that intensities of two matching points $\mathbf{x}_{ik}$ and $\mathbf{x}_{jk}$ are from the same radiance $L(\mathbf{x}_k)$ because the camera center is the same for both images. Hence, if exact homographies of all images are known, and assuming that all images are corrupted by zero mean Gaussian noise and distorted by non-linear mapping in Equation (4), all related parameters can be found by minimizing the sum of squared errors in Equation (3) for all pixels in the overlapping region.

Notice that given a constant c, replacing $\mathcal{G}_i(\mathbf{x}_{ik})L(\mathbf{x}_k)$ with $\mathcal{G}_i(\mathbf{x}_{ik})cc^{-1}L(\mathbf{x}_k)$ may result in the same energy. Thus, $c^{-1}L(\mathbf{x}_k)$ can be another solution for scene radiances. This ambiguity can be solved by fixing the radiance of the principal point in the reference frame I_1, because the intensity value is not affected by the darkening effects. By replacing Equation (4) with

$$\tilde{\mathcal{G}}_i(\mathbf{x}_{ik}) = D_i (1 - \alpha_i r_i(\mathbf{x}_{ik})) \Big(\frac{f_1 f_i}{f_i^2 + r_i(\mathbf{x}_{ik})^2} \Big)^2, \tag{8}$$

we can make $L(p_x, p_y)$ equal to $g^{-1}(I_1(p_x, p_y))$ with $D_1 = 1$. Then, D_i should be a constant representing the ratios depending on the lens diameters and exposure times between I_1 and I_i. For some cases, it also contains the ratio of white balances and illumination change. In this sense, scene radiances are reconstructed only up to scale where the scale is related to the reference frame.

3.1. Robust Statistics

Among a bulk of noisy data in the overlapping region, there may be outliers which contribute too much to the overall

solution in a negative direction. In our problem, although there is a small error in the estimated homographies causing small errors in all point matchings, there may be large intensity errors in the edge region. Also, if there are moving objects in the overlapping region, which is a general case in the mosaic problem, the intensity data of the region may be outliers.

A way of overcoming the outlier problem is by robust statistics [9]. In our case, intrinsic parameters $\Theta = [f_1, ..., f_N, D_2, ..., D_N, g, \alpha, p_x, p_y, \gamma, s]^T$ and scene radiance L of all pixels are found with Levenberg Marquardt (LM) optimization minimizing

$$E_\rho = \frac{1}{|N\Omega|} \sum_{i=1}^{N} \sum_{\mathbf{x}_k \in \Omega} \rho\big(I_i(\mathbf{x}_{ik}) - g(\tilde{\mathcal{G}}_i(\mathbf{x}_{ik})L(\mathbf{x}_k))\big), \quad (9)$$

where $\rho(x)$ is a robust error norm of residual error x, Ω is the overlapping region of the final mosaic image, and N is the number of images.

The choice of the ρ-functions results in different robust estimators and the robustness of a particular estimator refers to its insensitivity to outliers. In this paper, we used a Huber-type skipped means estimator [9]: This rejects everything which is more than 5.2 median deviations away from the median and takes the mean of the remainder; and the second derivative is always positive which is a very important property for incorporating the robust estimator into LM optimization. Because our algorithm finds too many parameters in L, direct implementation of LM optimization is not efficient. However, noting that each radiance depends only on each pixel position, our formulation can be applied to the bundle adjustment technique[1] utilizing the diagonal structure of the normal equation in the LM optimization.

3.2. Using Priors

There may be some known properties for a given problem, and using prior knowledge as constraints may prevent degenerate solutions. In this subsection, we introduce two priors.

Considering a rotating and zooming camera, we can use a constraint on H_{1i} for $i = 1 \sim N$. Since $R_{1i} = K_i^{-1} H_{1i} K_1$ is a rotation matrix in Equation (5), it satisfies the property that $R_{1i} = R_{1i}^{-T}$. This can be equivalently represented by

$$\omega_i^* = K_i K_i^T = H_{1i} K_1 K_1^T H_{1i}^T = H_{1i} \omega_1^* H_{1i}^T, \quad (10)$$

where ω^* is called the dual image of the absolute conic. This equation is known as the infinite homography constraint (IHC). It relates camera calibration matrices to infinite homographies and has been used as a measure for self-calibration [10]. If H_{1i} is obtained from a rotating and zooming camera, IHC should be a strong constraint with

$$E_h = \|\omega_i^* - H_{1i} \omega_1^* H_{1i}^T\|_F^2, \quad (11)$$

where F denotes Frobenius norm.

As for the non-linear mapping function g, we adapt the polynomial model in [6] to give a broad flexibility to the shape, given by

$$g(x) = a_1 x + a_2 x^2 + ... + a_{n-1} x^{n-1} + (1 - \sum_{i=1}^{n-1} a_i) x^n, \quad (12)$$

and this should always be a monotonically increasing shape. Because the sigmoid function ($Sig(x) = (1 + e^{-ax})^{-1}$) is very close to the unit step function with a large value of $a (= 10^{-6})$, the negative gradient is not permitted by using the constraint given by

$$E_s = \int_0^1 \Big[1 - Sig(\frac{\partial g(x)}{\partial x})\Big]^2 dx. \quad (13)$$

Then, the final solutions are estimated with

$$\hat{\Theta}, \hat{L} = \arg\min_{\Theta, L} \Big[E_\rho + \lambda_1 E_h + \lambda_2 E_s\Big], \quad (14)$$

where λ_1 and λ_2 are Lagrange multipliers, and the two constraints can be easily included in the bundle adjustment because they depend only on Θ.

It is noted that simultaneous estimation of Θ, L makes our algorithm use multiple images in a single framework. It is the reason we call our method a global approach. With Θ, enhanced images $I_i^{en} (= \tilde{\mathcal{G}}_i^{-1} g^{-1}(I_i))$ can be obtained for all pixels, and the final mosaic is constructed by an average blending. We call this enhanced blending in this paper.

4. EXPERIMENTS

In this section, we show results of our algorithm using real digital images. The homographies are calculated using the method in [1] without accumulation errors. The initial values of the parameters in K_i are obtained using the method in [10] with the square pixel assumption ($\gamma = 1, s = 0$), and we set $n = 5$, $\lambda_1 = 10^{-2}$, and $\lambda_2 = 1$.

In Fig 3, we show blending results from twelve digital images taken with a Sony digital camera (DSC-P72) in auto mode, and the estimated intrinsic camera parameters are depicted in Fig 4. As expected, the average blending result shows strong boundary seams in the sky region, while our method shows no boundary seams. The reconstructed non-linear mapping function shows smoothly increasing shapes.

In Fig 5, we consider the case where misalignments and moving objects (cars) exist in the overlapping region. By using robust statistics, a seamless blending result is obtained. Also, the average shape of the non-linear mapping function is close to the shape in Fig 4.

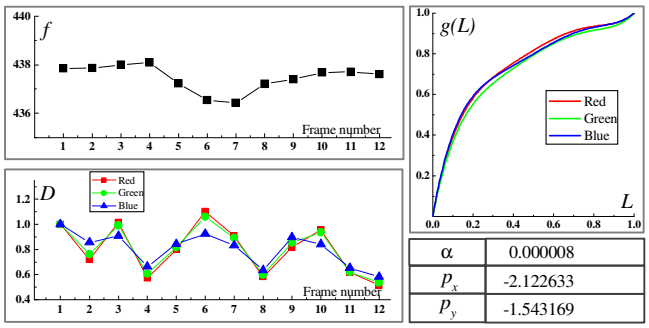
[1] http://www.esat.kuleuven.ac.be/~pollefey/tutorial/node76.html

5. CONCLUDING REMARKS

By introducing new energy terms, we robustly estimated blending parameters and scene radiances. With the reconstructed radiance images from a rotating and zooming camera, we produced enhanced mosaic results without boundary seams. Because our method deals with multiple images in a single framework, we expect that it can be easily modified to the problems of super resolution and global mosaic improving matching quality.

Regarding traditional self-calibration of a rotating and zooming camera, the IHC in Equation (11) was used as a strong measurement. In practice, the assumption of pure rotation is often violated because the camera center and rotation center do not coincide perfectly. This causes translations and large variations between feature matchings, also resulting in large errors in estimated intrinsic camera parameters. Thus, if there is considerable translation like a handhold camera, IHC may not be a strong constraint. This is the reason we set $\lambda_1 = 10^{-2}$ in experiments. In future work, we need to include some analysis about the effect of translation in our formulation.

Fig. 3. Blending with twelve digital images (320x240).



α	0.000008
p_x	-2.122633
p_y	-1.543169

Fig. 4. Estimated intrinsic parameters of Fig 3.

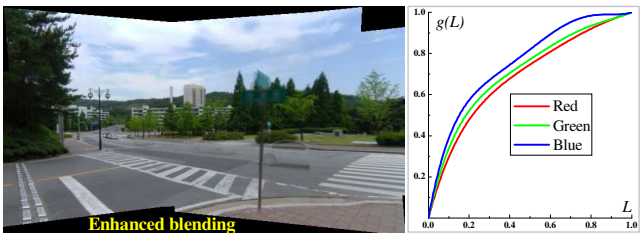


Fig. 5. Results from ten images with moving cars.

6. REFERENCES

[1] D. W. Kim and K. S. Hong, "Fast global registration for image mosaicing," ICIP, pp. 14-17, 2003.

[2] J. Davis, "Mosaics of scenes with moving objects," CVPR, 1998.

[3] P. J. Burt and E. H. Adelson, " A multiresolution spline with applications to image mosaics," ACM transactions on Graphics 2(4), pp. 217-236, 1983.

[4] H. Nicolas, "Optimal criterion for dynamic mosaicking," ICIP, pp.138-142, 1999.

[5] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," SIGGRAPH 97, pp. 369-378, 1997.

[6] T. Mitsunaga and S. K. Nayar, "Radiometric self-calibration," CVPR, pp. 374-380, 1999.

[7] S. B. Kang and R. Weiss, "Can we calibrate a camera using an image of a flat, textureless Lambertian surface?," ECCV, pp. 640-653, 2000.

[8] D. Hasler and S. Süsstrunk, "Mapping colour in image stitching applications," Journal of Visual Communication and Image Representation, Vol. 15, pp. 65-90, 2004.

[9] F. R. Hampel. E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, "Robust statistics: The approach based on influence function," New York: Wiley, 1986.

[10] L. de Agapito, E. Hayman, and I. Reid, "Self-calibration of rotating and zooming cameras," International Journal of Computer Vision 45(2), pp. 107-127, 2001.