Advanced Materials Research Vols. 452-453 (2012) pp 1025-1029
© (2012) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMR.452-453.1025

Sequential Homography-Based Alignment for HDR image generation

Luigi Barazzetti[1,a], Fabio Remondino[2,b] and Marco Scaioni[1,c]

[1]Politecnico di Milano, p.za Leonardo da Vinci 32, Milan, Italy

[2]FBK Trento, via Sommarive 18, Trento, Italy

[a]luigi.barazzetti@polimi.it, [b]remondino@fbk.eu, [c]marco.scaioni@polimi.it

Keywords: Alignment, Feature, HDR, Homography, Image, Matching

Abstract. Tripoding the camera is a standard solution to acquire aligned images useful for High Dynamic Range photography. On the other hand, the chance to use a hand-held digital camera is surely more practical and attractive for photographers. In this paper we propose a registration algorithm that recovers the alignment of several bracketed images using a progressive combination of homographies estimated from a set of image correspondences.

Introduction

High Dynamic Range (HDR) technology was developed to capture the full dynamic range of light in real scenes. Nowadays, modern charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensors have spatial resolutions superior to 20 Mpixels. In general, it is normal for customers and non-expert people to judge the quality of a camera from the number of pixels, although there are more important parameters, like e.g. the pixel size.

On the other hand, the discretization carried out by digital cameras not only involves the spatial domain, but also the radiometric resolution connected to the image depth (# of bits per pixel). A color image can be therefore intended as a discrete function $c = f(x, y)$ where light intensity information is associated to each pixel location (x, y). For a standard digital camera, c includes three channels (r, g, b) corresponding to red, green and blue components. To obtain realistic results, photographic cameras should capture a sufficient dynamic range (the ratio between the maximum and minimum pixel values) in order to simulate human vision.

The human visual system is sensitive to wavelengths in the visible spectrum and can deal with changes in brightness in order to preserve the original color of the scene. Moreover, the eye has optimal adaptation mechanisms that allow one to see real scenes with dynamic ranges exceeding 4-5 orders of magnitude [1]. A common digital camera acquires only Low Dynamic Range (LDR) images and thus the full dynamic range of light cannot be correctly represented.

To overcome this sensor's limitation, a set of multiple LDR images taken at different exposures can be used. It is recommended to employ the camera RAW mode and adjust the shutter speed (some cameras can bracket automatically), preferring low ISO setting to reduce noise. Several packages for HDR composition are today available on the market (e.g. Hydra, Photomatix, Photo Merge, Photoshop). Their use is simple and almost fully automated, making HDR imagery a very attractive field for beginners and not only for expert photographers or people involved in 3D modeling and virtual reality (e.g. [2-3]).

Before generating a HDR image, the sensor's spectral response function must be determined. Then, images acquired at different exposures can be fused into a single high dynamic radiance map.

There exist several methods to solve for the problem, as best summed up in [4]. However, one of the most significant contribution to the diffusion of HDR photography was given by [5] where an innovative solution based on the hypothesis that the response function is invertible and smooth was given.

HDR algorithms rely on pixel correspondences and therefore the camera should be mounted on a tripod to ensure that all images will align correctly. In addition, the scene should be as static as possible in order to avoid artifacts. In the case of hand-held shots, a preliminary geometric transformation is needed to register bracketed images and remove any misalignments.

Sequential registration of hand-held shots

Although many expert photographers use a tripod for the acquisition of bracketed images, the possibility of acquiring hand-held shots remains of primary importance for most practical situations.

In this latter case, possible misalignments between the bracketed images can be present thus requiring a preliminary alignment procedure to compensate for such shifts. The key idea is an automated, quick, precise and robust method capable of estimating a set of geometric transformations that resample all images with respect to a reference one, so that the new data will be consistent among themselves.

As HDR photography is becoming very popular, the method should be as automatic as possible in order to be used by non-expert operators.

The main problem is the choice of an appropriate mapping during the resampling phase. The proposed solution is based on a homography (i.e. an 8- parameter perspective transformation), as it is a rigorous mapping between images acquired after rotating the camera around its perspective center. Under this assumption, we assume that the image perspective center $\boldsymbol{X}_0$ does not move, while rotations could occur. The pinhole camera matrix model [6] expresses the alignment between the perspective center $\boldsymbol{X}_0$ (normally set to $\boldsymbol{0}$), the image point $\boldsymbol{x}$ and object points $\boldsymbol{X}$, through a 3×4 projection matrix **P**:

$$\boldsymbol{x} = \mathbf{P}\boldsymbol{X} = [\mathbf{R}\ \boldsymbol{0}]\boldsymbol{X}\,. \tag{1}$$

The matrix **P** can be split into a matrix product involving a calibration matrix **K**, a rotation matrix **R** and the perspective center vector $\boldsymbol{0}$.

A pair of images taken with a rotating camera will differ by the numerical values of rotation matrices. Without loss of generality, we can write the following relationships:

$$\boldsymbol{x} = \mathbf{K}[\mathbf{I}\ \boldsymbol{0}]\boldsymbol{X} \qquad \boldsymbol{x}' = \mathbf{K}[\mathbf{R}\ \boldsymbol{0}]\boldsymbol{X} \tag{2}$$

and, with some simple substitutions:

$$\boldsymbol{x} = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\boldsymbol{X} = \mathbf{H}\boldsymbol{x}\,. \tag{3}$$

Therefore, the problem is the estimation of a 3×3 matrix **H** encapsulating a homographic transformation between the image pair. **H** can be derived using a set of image-to-image correspondences automatically extracted with feature-based operators, like SURF [7]. In our implementation, each interest point is coupled with a 128-element vector and all the vectors are compared using the norm of the differences, along with the ratio test between the first two candidates. A kd-tree search is run to reduce CPU time during the vector comparisons [8]. The method proved to be efficient even with images taken at different exposures, especially if the exposure time does not vary significantly.

For this reason, all images are preliminary sorted by considering the exposure value stored in the metadata associated to each single image. Matching is then performed between pairs of consecutive images (1-2, 2-3, …) in order to reduce CPU time and potential mismatches. Finally, a robust estimator (RANSAC [9]) is used to remove wrong correspondences and yields a set of corresponding points useful for estimating **H** via a Least Squares (LS) method.

The final result is a sequence of homographies ($\mathbf{H}_{2,1}$, $\mathbf{H}_{3,2}$,…, $\mathbf{H}_{n,n-1}$) that are sequentially used to resample all images. The central image i of the sequence is assumed as reference and, assuming a relationship $\boldsymbol{x}_k=\mathbf{H}_{k,k}^{-1}\boldsymbol{x}_{k-1}$ between consecutive images, any other is processed using the following criteria:

$$\begin{cases} \boldsymbol{x}^i = \mathbf{H}_{i+1,i}^{-1}\mathbf{H}_{i+2,i+1}^{-1}\cdots\mathbf{H}_{j,j-1}^{-1}\,\boldsymbol{x}^j & \text{if } i < j \\ \boldsymbol{x}^i = \mathbf{H}_{i,i-1}\mathbf{H}_{i-1,i-2}\cdots\mathbf{H}_{j+1,j}\,\boldsymbol{x}^j & \text{if } i > j \end{cases} \tag{4}$$

Fig. 1 shows an example of data processing, where a set of 9 bracketed photos was geometrically aligned to generate a final HDR mosaic. The camera is a Nikon D700 with a 50 mm lens. The interpolation technique chosen is the bilinear method.

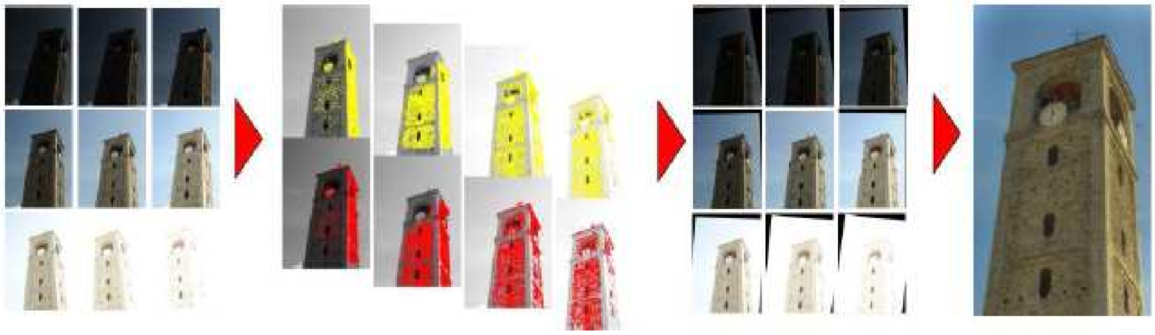

Fig. 1. The pipeline for image registration: matching, image resampling, and HDR mosaicing.

In this example a strong rotation of the camera (check the tower slope) was generated to obtain unfavorable conditions. The method is very robust against this effect, providing sequences of homographic transformations estimated with sub-pixel precision (sigma-naught of LS solution was ±0.7 pix between consecutive images used at their raw resolution). The method is also quite robust against camera translations, except when parallax errors occur.

Accuracy analysis

A visual inspection of results can be carried out by superimposing the resampled bracketed images. Experiments with different datasets provided nice visual results, although a numerical evaluation is pending. The main problem is the direct comparison of the gray values as images have different exposures. For this reason, we applied the alignment algorithm to the same image placed in different positions of the sequence.

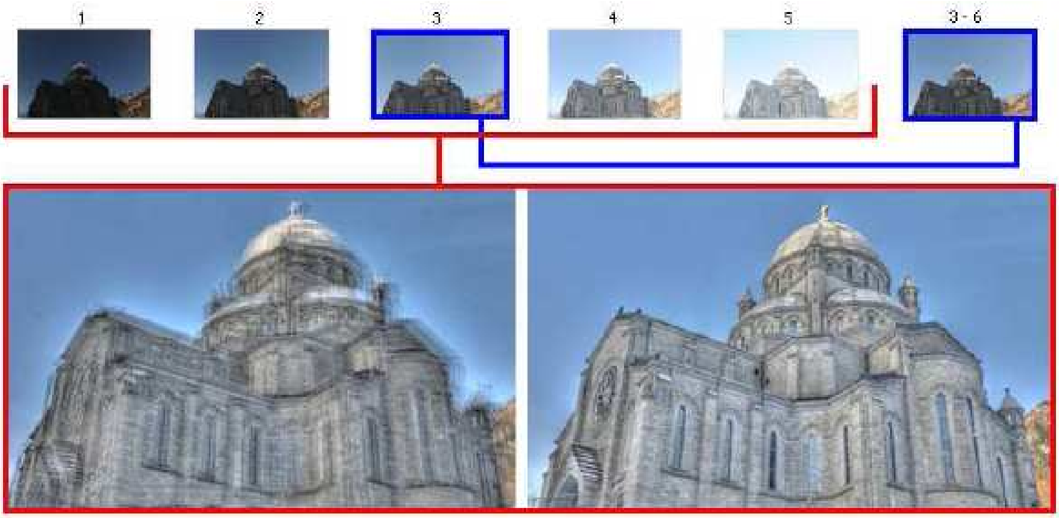

Fig. 2. Bracketed images (1-5) used to generate the HDR mosaic: results without and within the registration algorithm. The image 6 (equal to 3) was employed for the experiment shown in Fig. 3 (the object is the Sanctuary of Re, Italy).

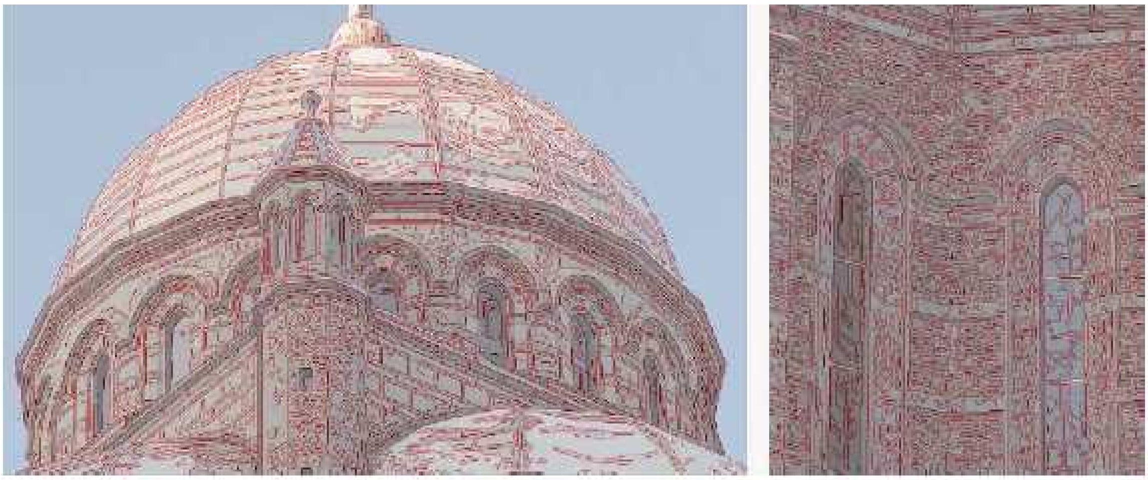

Fig. 3. The central image of the sequence with the edges extracted from image 6 after the resampling process.

A numerical comparison between image 3 and the modified image 6, after a 8 bit grayscale conversion, was carried out by using pixel differences. The mean of all differences was 0.08 pix, and the standard deviation ±0.63 pix. These results confirm the quality of the registration method.

Lastly, a visual comparison between the third image and the edges extracted from the resampled image 6 is shown in Fig. 3. Here, edges were extracted with the Canny operator on image 6 and were superimposed to the central image of the sequence. As can be clearly seen, results are consistent and confirm the statistical evaluation.

Finally, shown in Fig. 4 are some other examples collected in Sondrio (Italy) with a Nikon D700 equipped with different lenses.

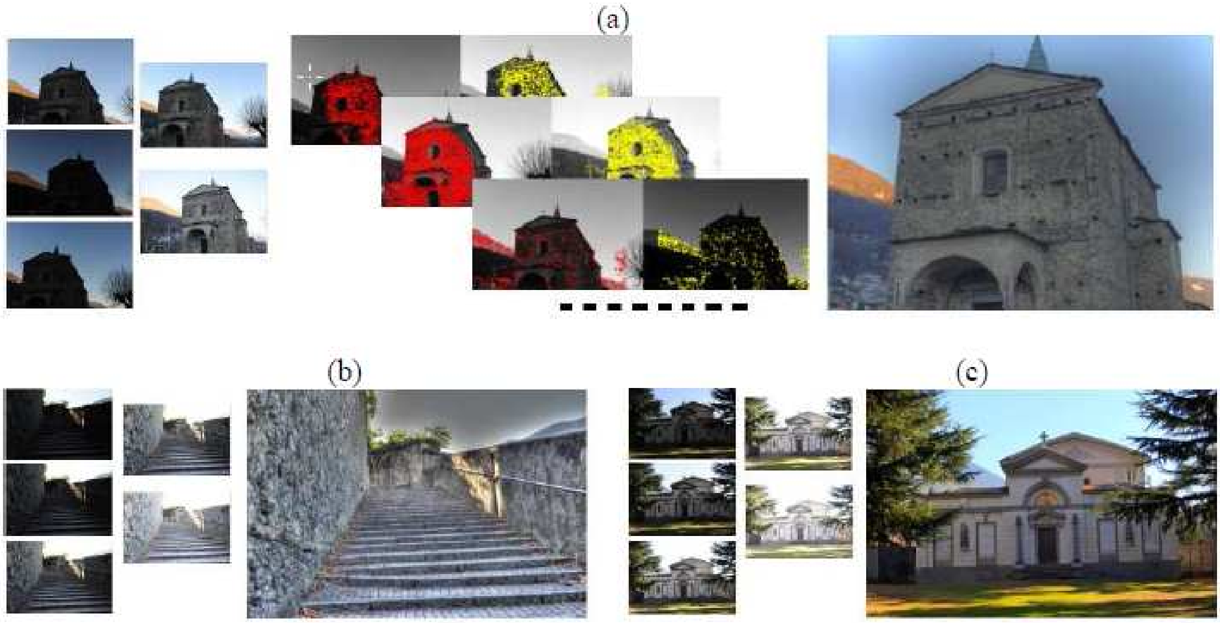

Fig. 4. Some HDR mosaics created with the proposed procedure. All objects are located in Sondrio (Italy): (a) church of Sant'Antonio, (b) an ancient staircase, and (c) Tempio Ossario in Parco della Rimembranza.

Conclusions

The paper presented a methodology to register a sequence of bracketed photos taken with a hand-held camera, where possible misalignments could be present due to human movements during image acquisition. These misalignments must be removed, otherwise the HDR image generation procedure will create blurred results. The developed procedure is robust and can compensate for most misalignments, especially if parallax errors are avoided. The use of a cascaded matching, where one uses only consecutive image pairs, allows a reduction of CPU time. As the registration transformations (mainly based on a homography) are estimated via Least Squares, the method can also provide a statistical evaluation based on the adjustment statistics.

References

[1] G. Qiu, J. Guan, J. Duan and M. Chen: *Tone Mapping for HDR Image using Optimization – A New Closed Form Solution*. ICPR 2006, 4 pages.

[2] F. Remondino and A. Rizzi: *Reality-based 3D Documentation of World Heritage Sites*: *Methodologies, Problems and Examples*. 22nd CIPA Symposium, October 11-15, 2009, Kyoto, Japan, 8 pages.

[3] S. Campana, M. Sordini and A. Rizzi: *3D Modeling of a Romanesque Church in Tuscany: Archaeological Aims and Geomatics Techniques*. Proceedings of 3rd International Workshop 3D ARCH 2009, 7 pages.

[4] A. Sa, P.C. Carvalho and L. Velho: *High Dynamic Range Image Reconstruction*. Morgan & Claypool Publishers, 53 pages.

[5] P.E. Debevec and J. Malik: *Recovering High Dynamic Range Radiance Maps from Photographs*. Proc. of Siggraph, 1997, 10 pages.

[6] R. Hartley and A. Zisserman: *Multiple View Geometry in Computer Vision*. Second edition. Cambridge University Press, Cambridge, 2004, 672 pages.

[7] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool: *Speeded-up robust features (SURF)*. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[8] L. Barazzetti, F. Remondino and M. Scaioni: *Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation.* The Photogrammetric Record, 25(132), 2010, pp. 356–381.

[9] M.A. Fischler and R.C. Bolles: *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Communications of the ACM, 1981, 24(6), pp. 381-395.