International Journal of Advanced Smart Convergence Vol.4 No.2 6-13 (2015)
http://dx.doi.org/10.7236/IJASC.2015.4.2.6

IJASC 15-2-2

Multi-camera based Images through Feature Points Algorithm for HDR Panorama

Jung- Ho Yeong

Department of Accounting, Seowon University, 377-3 Musimseoro Seowon-gu Cheongju, Chungbuk South Korea
pyuksuk@seowon.ac.kr

Abstract

With the spread of various kinds of cameras such as digital cameras and DSLR and a growing interest in high-definition and high-resolution images, a method that synthesizes multiple images is being studied among various methods. High Dynamic Range (HDR) images store light exposure with even wider range of number than normal digital images. Therefore, it can store the intensity of light inherent in specific scenes expressed by light sources in real life quite accurately. This study suggests feature points synthesis algorithm to improve the performance of HDR panorama recognition method (algorithm) at recognition and coordination level through classifying the feature points for image recognition using more than one multi frames.

Keywords: *image-based lighting, aerial image, SIFT, RANSAC.*

1. Introduction

Recently, many programs for identifying acquired images utilizing the mobile phone cameras are being developed. Researches related to the building recognition are actively done to provide the shop information in the buildings for the users [1-2]. The types of input images for building recognition are divided into aerial image and city building image. In most of the city building images, buildings coexist with surrounding objects such as the pedestrians, tress, and the sky. Developments of display techniques to print out high-resolution images are underway, and Full-HD, 4K-UHD, 8K-UHD are now appearing. Recently, methods for UHD broadcasting are constantly explored. "YouTube", the world's biggest video site, is also supporting the resolution up to 4K. Since it is a free platform where ordinary users can upload, equipment's for high-resolution image acquisition need to be popularized.

When photographing the real images, it is more important than anything else to store and utilize all the lighting information accurately. Using image-based lighting (IBL) method [3-4], we can express lighting information similar to the real world. In order to do this, it is necessary to store all the existing lighting information that affects the color and brightness in panorama images. Panorama images are used to provide lighting information that affects the environment map in the course of rendering, and we can create natural and realistic scenes using IBL arithmetic function based on environment mapping that are basically included in commercial graphic software.

Panorama images used for environment mapping work as the most important factor that decides how

Manuscript Received: Sept. 5, 2015 / Revised: Sept. 30, 2015 / Accepted: Oct. 15, 2015
Corresponding Author: pyuksuk@seowon.ac.kr
Tel: +82-43-299-8565, Fax: +82-43-299-8560
Department of Accounting, Seowon University

naturally and realistically the real images and graphic model can be synthesized. Depending on how accurately and sufficiently this image store the lighting information of the real world, the quality of the final images is decided. The generally used normal digital cameras record each pixel in quantized whole number with Low Dynamic Range (LDR) 8 to 10 bit. However, since human eyes can recognize light with even wider range than this, using LDR images for the IBL method is not appropriate. Actually, for realistic image synthesization, high dynamic range (HDR) images are used, which save each pixel in real number in the form of floating points. This makes it possible to more accurately express the lighting information in the real world that will be used in the process of rendering [5].

Before the development of HDR cameras, there were researches on making one HDR image through multiple LDR images photographed with altering shutter speed [6]. However, the process of HDR environment map using widely used LDR cameras requires time and technology. Photomerge in professional photoshop is fast but limited. First of all, if the photos tilt more than 3° ~ 10°, it can cause error. The images should be overlapped 40% and it is hard to arrange them when the exposure values of each photo are different because it makes finding the matching point difficult. The smart phones for ordinary people take photographs turning the camera while people are staying still, in only one direction, and are greatly affected by the vibration of the user's hands. Brightness is also fixed on the spot where the first photograph was taken [7-8].

2. System implementation

Figure 1 is the flow of the image acquisition of 9 images using network time protocol (NPT), image alignment, standardization of brightness, image synthesis, and high-definition images. The ways to classify feature points are composed of three steps. In the first step, feature points are extracted adopting scale invariant feature transform (SIFT) [3] algorithm. In the second step, the extracted feature points are matched using multiple frames. In the last step, the acquired multiple pairs of the feature points obtain the alignment of homography using random sample consensus (RANSAC) [9] algorithm, being classified as one for the buildings [10-11].

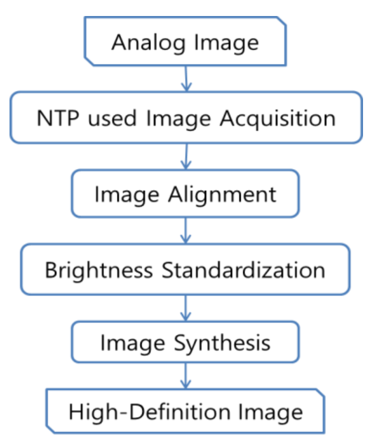

Figure 1. Algorithm development for image acquisition and match

Image acquisitions go through four steps, image alignment, brightness standardization, and image synthesis to achieve high-definition images. Since the purpose is to take photographs of 9 images of the same object at the same time, synchronization through NPT is needed. Image alignment plays a role of adjusting the twisted angles of the acquired images to the images in the middle to enhance the accuracy of image synthesis. Standardization of brightness happens when a camera adjusts the brightness that is changing according to the angles of the shooting. Using methods like white balancing, we can get homogeneous brightness, improve the image quality and enhance the accuracy for image synthesis. Image synthesis extracts the feature points of images using SURF algorithm and calculates their local invariant features. With these values, one image is synthesized through the process of wrapping images [12-13].

2.1 Synchronization of 9 image acquisition through NPT

Since multiple cameras are used, all the photographs (images) need to be taken from the cameras at the same time. Considering the moving objects, if the synchronization between cameras do not match, we will get images for different scenes. Therefore, for more accurate synchronization, a method shown in Figure 2 is used.

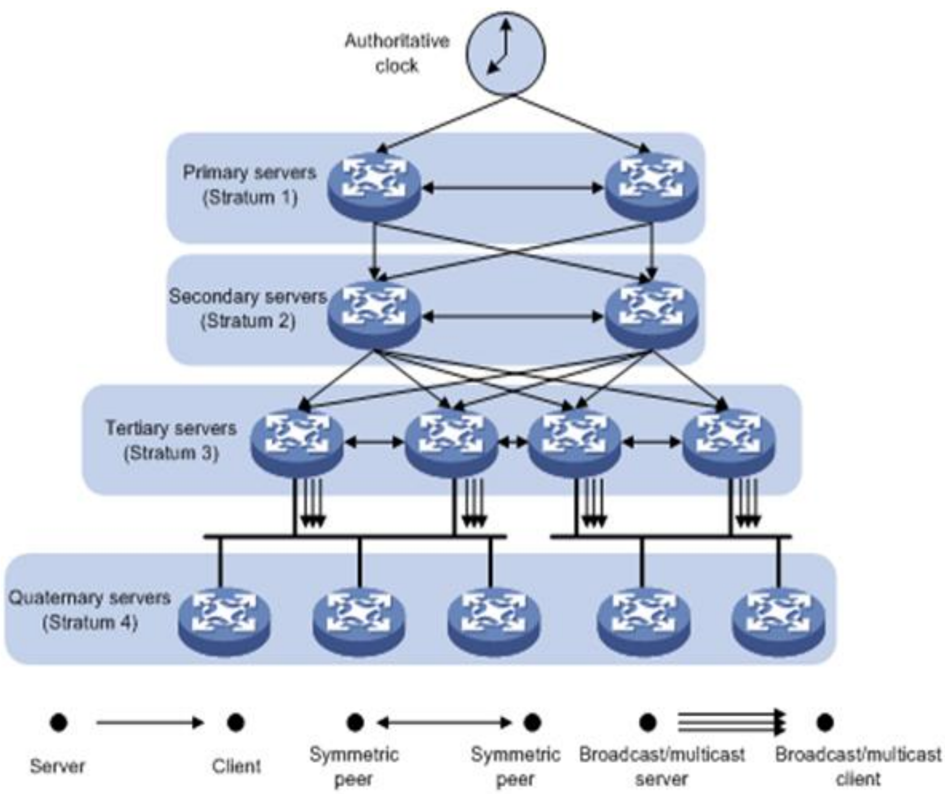

Figure 2. NTP hierarchy

NPT has a tree-shaped hierarchical structure. Stratum 0 refers to equipments such as GPS or cesium molecule clocks that tell us time, and stratum 1 refers to the server directly connected to the clocks. Stratum 1 is synchronized with stratum 0. From Stratum 2, it has a shape of a tree, synchronizing with the upper level servers. This study used nine raspberry pies, each of which was equipped with one camera module. We used 9 camera modules for shooting, and measured the shooting time using NPT at each connected raspberry pie. If the measured time goes over the certain threshold, we consider that synchronization has not been done and try to obtain new images.

2.2 Alignment of images through camera calibration

Since we twist the angles to raise the angle of view when taking photographs, we need to align them for more accurate synthesis [14]. As shown in Figure 3, we calculate the external variables, obtain the orthogonal projection of the image, and convert them.

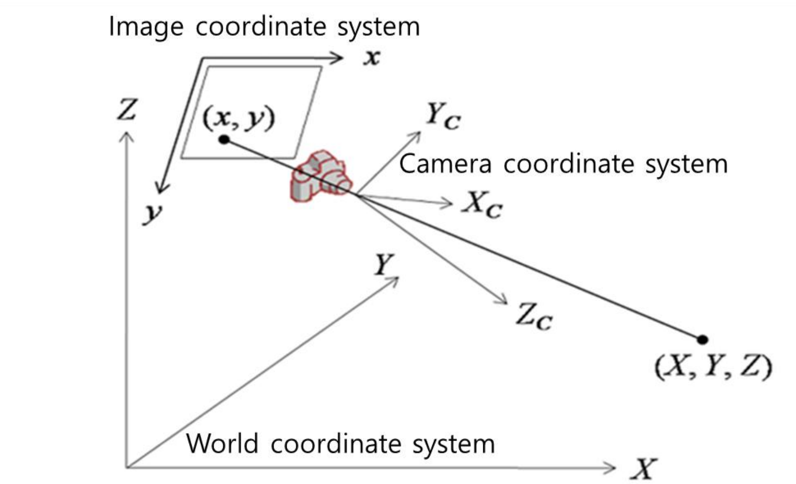

Figure 3. Camera Calibration

Camera Calibration is a process that tries to get the internal factor, which needs to be removed for accurate conversion between two-dimensional flat images and three-dimensional space coordinates. After getting the internal variable, the external variable is calculate as shown in (1) by getting the transfer matrix through matching pairs of 3D world coordinates and 2d image coordinates that are already known or chosen for a sample.   This was done using Solve PNP function in open CV.

$$s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & skex\_cf_x & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{21} & r_{31} & t_1 \\ r_{21} & r_{22} & r_{32} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = A[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

2.3 Standardization of brightness

As the amount of light changes according to the angle of the camera, it can generate images with totally different brightness.   Therefore, to improve the quality of image and the accuracy of image synthesis, we take the step of brightness standardization.

We calculate the probability distribution of brightness using histogram. After getting the value, standardization happens with the value after setting different references to improve the quality of image to a maximum degree. We can obtain images with better quality if we use different methods depending on the kinds of images obtained. In this study, among 9 cameras, we made the camera in the middle as reference for the brightness value, average/median value of all the images, and brightness value through white balancing. Using the final value, we conduct standardization of the whole images based on the standard value.

3. Image synthesis using panorama technology

Panorama technology can divided into 4 steps. After getting the features between the images, we calculate the amount of feature. Using the acquired value, we change the image and get the final synthesized image after synthesizing the changed images [15]. We compare the existing SIFT[16] algorithm with the recall value to examine the performance of the coordination result of the algorithm suggested by this study. The recall value refers to the value that shows the proportion of the exact coordination among the whole matching pairs, as defined in formula (2).

$$\text{recall} = \frac{\text{N correct matches}}{\text{N correspondences}} \times 100\% \tag{2}$$

Here, N correspondences is the total number of finally classified synthesis pairs, and N correct matches is the number of the accurately matched feature points with the naked eye.

There are various kinds of algorithms such as IFT, Harris corner, FAST [17-19], BRISK [20], FREAK, ORB [21] that extract feature points. Among these, SURF [22] is fast and robust against distortion of the images with brightness difference. To extract feature points, SUFR uses Hessian matrix and Hessian shows the second derivative of function. That is, Hessian is the matrix that shows the curvature characteristics of the function. We extract the feature points from the highest point of the matrix formula in Hessian matrix. If the matrix formula is positive number and eigenvalue is the same code, we think of it as interest point. Centered on the interest point found earlier, neighboring Haar wavelet response is calculated. As shown in Figure 4, the blue dot in the original on the right side refers to the distribution of the response values, while the red arrow expresses the sum of the response values as vector. Setting the 60 degree of window and calculating from different directions, the longest vector among them is assigned as orientation.

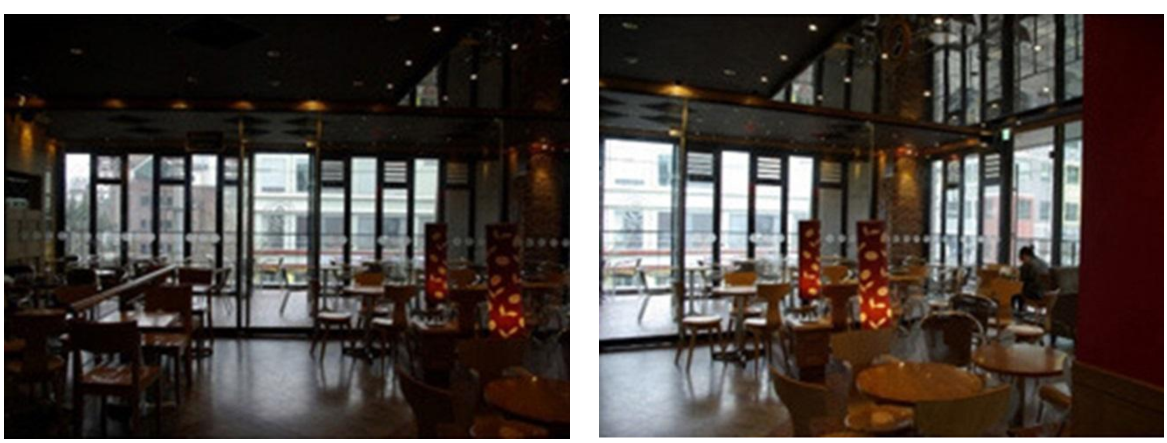

Figure 4. Haar wavelet response

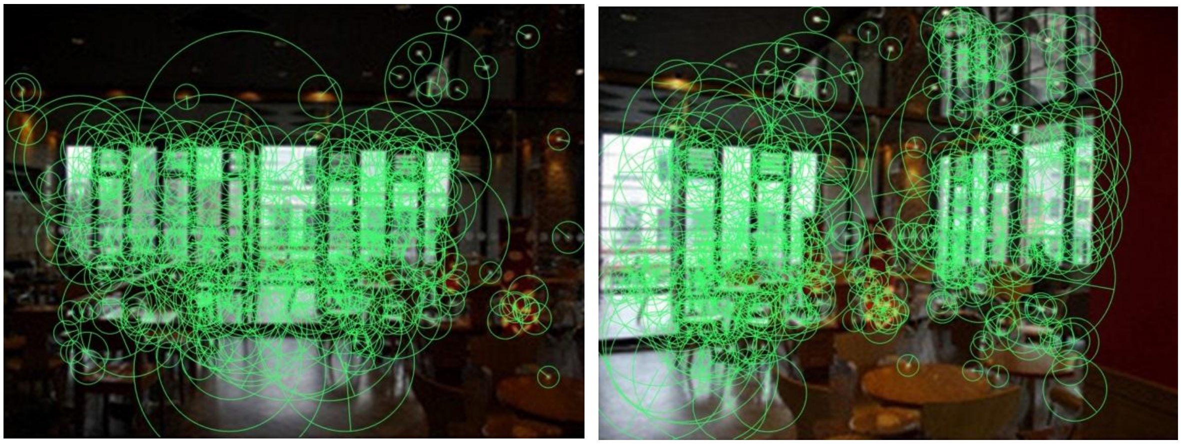

Figure 5. Expresses robust local invariant features

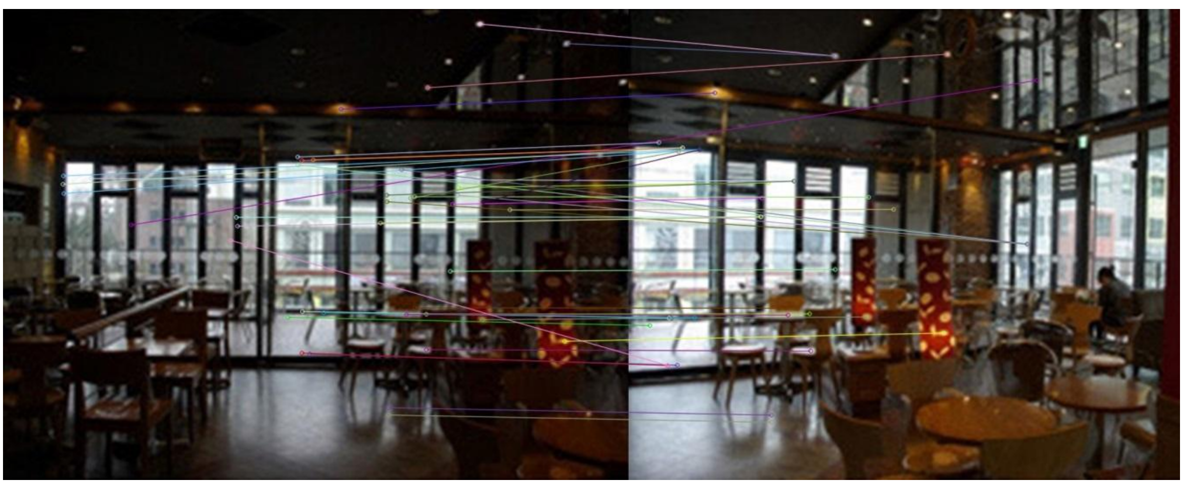

Figure 6. Connects the same feature points

As shown in Figure 5, we perform geometrical manipulation on the images using the correspondent point between images that were found. For synthesis, we need to extend the images unequally, which is called warping. For natural connection of the images, we change the images and combine them as shown in Figure 6. After taking 9 images at the same time, we can synthesize these into one image to acquire a high-definition and high-resolution image. Since we go through the whole process of image alignment and brightness standardization, more accurate image synthesis is performed unlike the other panorama techniques. Also, it improves the quality of images as it adjusts the brightness.

4. Conclusion

This study suggested the methods that can improve the performance of coordination and classify the feature points using multi-frame in the course of image recognition. Extracting feature points of the buildings based on the existing SIFT often lacked the accuracy of coordination since it extracts feature points from the surrounding environment other than the building. This study obtained the homography matrix using images from different viewpoints at the same intervals and classified the feature points adopting the RANSAC algorithm. One homography matrix classifies feature points of one flat surface of the image. The classified remaining feature points are accurately matched all feature points by repeating the process of getting the homography matrix. The recall value of the proposed algorithm has gone up from that of SIFT.

References

[1] D. Miorandia, S. Sicarib, F. D. Pellegrinia, and I. Chlamtaca, "Internet of things: Vision, applications and research challenges," Ad Hoc Networks, Vol. 10, No. 7, pp. 1497-1516, 2012.

[2] H. Ali, G. Paar, L. Paletta, "Semantic indexing for visual recognition of buildings," in: Proc. Int'l. Symposium on Mobile Mapping Technology, pp. 28–31, 2007.

[3] J. Li, W. Huang, L. Shao, and N. Allinson, "Building recognition in urban environments: A survey of state-of-the-art and future challenges," *Information Sciences*, Vol. 277, No. 1, pp. 406-420, 2014. DOI: 10.1016/j.ins.2014.02.112

[4] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004. DOI: 10.1023/B:VISI.0000029664.99615.94

[5] Y. Li, and L. G. Shapiro, "Consistent line clusters for building recognition in CBIR," Pattern Recognition, 2002. Proceedings. 16th International Conference, Vol. 3, pp. 952-956, 2002.

[6] I. Jolliffe, *Principal component analysis*, Wiley StatsRef: Statistics Reference Online, 2002.

[7] D. Cai, X. He, J. Han, *Using Graph Model for Face Analysis*, Department of Computer Science, University of Illinois at Urbana Champaign, 2005.

[8] G. J. Malachlan, *Discriminant analysis and statistical pattern recognition*, Wiley-interscience, New York, 1992.

[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, Vol. 24, No. 6, pp. 381-395, June 1981.

[10] S. Cho, B. Shrestha, B. Hong, and H. Y. Jeong, "Study of Generating Animated Character Using the Face Pattern Recognition," *Lecture Notes in Electrical Engineering*, Vol. 107, pp. 127-133, 2011. DOI: 10.1007/978-94-007-2598-0_13

[11] S. Cho, K. C. Son, J. Lee, and S. H. Lee, "FOLI Technique Algorithm for Real-Time Efficient Image Processing," *Lecture Notes in Electrical Engineering*, Vol. 279, pp1337-1342, 2014. DOI: 10.1007/978-3-642-41674-3_185

[12] E. Rublee, V. Rabaud, and K. Konolige, "ORB: An efficient alternative to SIFT or SURF," 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2564 – 2571, 2011.

[13] Y. Ke and R. Sukthankar. "Pca-sift: A more distinctive representation for local image descriptors," *In Computer Vision and Pattern Recognition*, pp. 506-513, 2004. DOI: 10.1109/CVPR.2004.1315206

[14] W. S. Lee, J. Gu, and M. T. Nadia, "Generating Animatable 3D Virtual Humans from Photographs," Computer Graphics Forum, Vol. 19, No. 3, pp. 1-10, 2000.

[15] Z. Wang, B. Fan, and F. Wu, *FRIF: Fast Robust Invariant Feature*, WANG et al.: Fast robust invariant feature, 1-12, 2013.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, pp.91–110, 2004.

[17] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, “Adaptive and generic corner detection based on the accelerated segment test,” In European conference on Computer vision, pp. 183–196, 2010.

[18] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” In European conference on Computer Vision, pp. 430–443, 2006.
DOI: 10.1007/11744023_34

[19] E. Rosten, R. Porter, and T. Drummond, “Faster and better: A machine learning approach to corner detection,” Pattern Analysis and Machine Intelligence, Vol. 32, No. 1, pp.105–119, 2010.

[20] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “Brief: binary robust independent elementary features,” In European conference on Computer vision, pp. 778–792, 2010.

[21] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: An efficient alternative to sift or surf.” In International Conference on Computer Vision, pp. 2564–2571, 2011.

[22] H. Bay, T. Tuytelaars, and L. V. Gool, “Surf: speeded up robust features,” In European Conference on Computer Vision, pp. 404–417, 2006.