J. Inf. Technol. Appl. Manag. 25(2): 41~52, June 2018
https://doi.org/10.21219/jitam.2018.25.2.041

ISSN 1598-6284 (Print)
ISSN 2508-1209 (Online)

Accurate Camera Self-Calibration based on Image Quality Assessment

Rabia Fayyaz* · Eun Joo Rhee**

Abstract

This paper presents a method for accurate camera self-calibration based on SIFT Feature Detection and image quality assessment. We performed image quality assessment to select high quality images for the camera self-calibration process. We defined high quality images as those that contain little or no blur, and have maximum contrast among images captured within a short period. The image quality assessment includes blur detection and contrast assessment. Blur detection is based on the statistical analysis of energy and standard deviation of high frequency components of the images using Discrete Cosine Transform. Contrast assessment is based on contrast measurement and selection of the high contrast images among some images captured in a short period. Experimental results show little or no distortion in the perspective view of the images. Thus, the suggested method achieves camera self-calibration accuracy of approximately 93%.

Keywords : Image Quality Assessment, Blur detection0, Contrast Assessment, Discrete Cosine Transform, Camera Self-calibration

Received : 2018. 05. 14. Revised : 2018. 06. 13. Final Acceptance : 2018. 06. 15.

* Department of Computer Engineering, Hanbat National University, Daejeon, 34158, Korea, e-mail : rabiafyz@hotmail.com
** Corresponding Author, Department of Computer Engineering, Hanbat National University, Daejeon, 34158, Korea, Tel : +82-42-821-1205, e-mail : ejrhee@hanbat.ac.kr

1. Introduction

Computer vision is an active research that deals with the application of computer-based analysis and interpretation of digital images. With the development of computer vision applications, intensive research has been carried out in robotics, medical imaging, 3D construction and augmented reality, etc. In general, the raw input data are the images captured by digital cameras with CCD or CMOS sensors, which map a three-dimensional real world into two-dimensional image sequences. Since the recorded images are a two-dimensional data, there is a loss of information in this mapping. This mapping is called perspective transformation.

A camera can be modeled by the parameters of perspective transformation which are the intrinsic and extrinsic parameters. These are designed to extract the 3D information of a two-dimensional nature which requires the calibration of the camera. The intrinsic parameters are the internal camera parameters which include focal length, aspect ratio, etc. The extrinsic parameters include information about the position of the camera and its rotation in 3D space. The process of finding these parameters is called camera calibration. Camera calibration is applied in many areas of Computer Vision applications; for instance Stereo Vision for interpreting Depth information [Fayyaz and Rhee, 2014], 3D construction [Wilczkowiak et al., 2001; Bätz et al., 2014], medical imaging [Liu et al., 2015], etc. For effective computer vision applications, it is important for the user to make some assumptions during the camera calibration process.

Many methods have been proposed to calibrate cameras for computer vision applications. Traditionally, the calibration of a camera is done by capturing a set of images of a known artificial grid pattern, such as a chessboard, placed in the scene [Zhang, 2014]. The 3D shape of the calibration provides 3D coordinates associated with the set of reference points on the 2D plane. The 3D coordinates system related to the image projection can be obtained through an optimization process. However, the major drawback of camera calibration by chessboard pattern is that they are applicable only for offline calibration applications. This approach is not applicable in cases where the intrinsic parameters are continuously changing during the image acquisition process due to the focusing or zooming functionality of the camera. This is true for active vision applications, where the camera optical parameters are contentiously controlled in order to simplify vision tasks [Huang et al., 2010].

Thus, for calibrating a camera, another approach is used to automatically formulate scene constraint with little or no prior knowledge about the structure of the images, this approach is called camera self or auto calibration [Huang et al., 2010]. Camera self-calibration can estimate both intrinsic and extrinsic parameters for a single moving camera [Hemayed, 2003], and also for a set of stationary cameras in the different position in 3D space. Many methods of camera self-calibration focused on optimization of camera self-calibration with constant intrinsic parameters, varying intrinsic parameters and with fewer unknowns based on two or three images [Akkad et al., 2013; Baataoui et al., 2012].

In the paper [Akkad et al., 2013], Akkad et al. [2013] used Susan Corner Detector and Normalized Cross Correlation for feature matching. Furthermore, their study performed the practical and theoretical study of a new method of camera self-calibration. In their research, two images are taken from different cameras and characterized by varying intrinsic parameters. This study showed that two images of a planar scene are sufficient to estimate the parameters of the two cameras. Therefore, the constraints are minimized on camera self-calibration. The method is based on demonstration of the relationship between three matches and the relationship between absolute conic for each pair of the images. For the relationship formulation, the author used non-linear cost function.

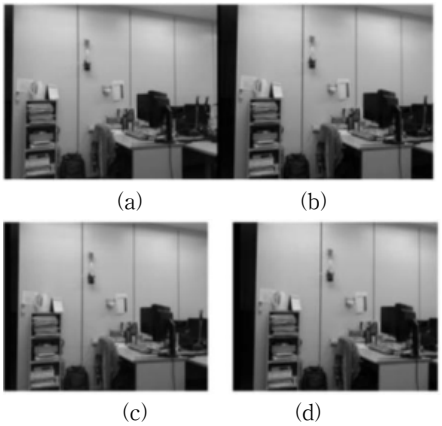
Another technique of self-calibration is presented using CCD camera with constant focal length with a planar scene [Baataoui et al., 2012]. This approach used an equilateral triangle whose two vertices are defined from the matches detected in the image taken from the camera. Based on these vertices, all projection matrices are estimated using homography between the images. The homography is also used to estimate the intrinsic parameters of the camera. Although the method is simple, reliable and robust for the 2D planar scene, the projection matrices can be affected by the effect of environmental changes on the image taken from the camera.

All of the above methods of camera self-calibration do not consider changes in environmental factors such as blurriness, noise, and contrast in images, which affect the accuracy of the perspective projection obtained by the self-calibration method. Camera calibration by a known chessboard pattern used a lot of chessboard images captured in different directions, positions and it required a lot of processing time [Fayyaz and Rhee, 2014]. On the other hand, camera self-calibration automatically calibrates the camera without any known structure in the scene. However, the accuracy of camera self-calibration is affected by environmental factors such as blurriness, noise, and contrast in an image. Blurriness is created by moving objects or camera out of focus, due to the internal vibration, depth of field, etc. The noise is from lighting or environment, and contrast is from very less lighting or dark environment.

In this study, we suggest the use of good quality images taken from two cameras by performing image quality assessment to enhance the accuracy of camera self-calibration. First, we assume the noise in the environment is constant, and we focus on the other two variables blurriness and contrast in images. By assessing the quality of the images, we select those images that have high contrast and no or less blurriness. After selecting the good quality frames, we perform camera self-calibration using SIFT (Scale Invariant Feature Transform) features. By using sharp and high contrast images, the perspective projection will have no distortion as shown in <Figure 1>, and accuracy of camera self-calibration by the suggested method of pre-processing is increased as compared to the conventional camera self-calibration.

The organization of this paper is as follows. The proposed algorithm is described in section 2. The usefulness of the proposed method is dem-

onstrated through simulation in section 3. The conclusion is in section 4.



〈Figure 1〉 Perspective view of input images. (a) Captured image A. (b) Captured image B. (c) Perspective view of image A. (d) Perspective view of image B

2. Image Quality Assessment and Camera Self-Calibration

Camera self-calibration for good perspective projection is a challenging task in environmental variable factors such as blur, noise and low contrast in images. The environmental variable factors are illustrated as follows;

- Blurriness in the image-this is caused by imperfect image acquisition process, applying low pass filter to the image, motion blur, and camera out of focus by movement of objects such as tanks or planes [Gonzales and Woods, 2008; Farias et al., 2003; Oliveira et al., 2014].
- Noise in the image-this factor also comes from environment or lighting condition.
- Low contrast image-this comes from very low lighting or dark environment.

In order to obtain accurate camera self-calibration by considering environmental variable factors, first, we perform pre-processing techniques on images. We perform image quality assessment to select sharp frames with good contrast and perform camera self-calibration using SIFT feature detectors. In other words, blurred images due to the movement of objects, and imperfect image are eliminated by our method. In the next section, image quality assessment is discussed in detail.

2.1 Image Quality Assessment

Image Quality Assessment is the initial step of the suggested method, in which the image is assessed in terms of blurriness and contrast with an assumption of constant environmental noise factors.

2.1.1 Blur Detection by DCT

In the first step, blur detection is performed based on discrete cosine transform (DCT) [Erik et al., 2009] and the sharp or normal images are selected as the good quality images.

In DCT, high frequency shows the edge information in an image. Blurriness cause a reduction in the high-frequency components of an image in DCT. The blurriness is caused by the movement of objects or the camera being out of focus, or an imperfect image acquisition process [Farias et al., 2003]. Thus, in order to capture a good quality image with no blurriness, a method is defined for selecting a good quality image by discarding the blur images using the method shown in definition 1.

Definition 1: Selecting good quality image.

Given I (x, y) as an image, define image as sharp or blur by the following method;

$$\begin{cases} \text{discard blur image} & \text{if } V > 0.8 \text{ and } E < 10 \\ \text{select sharp image} & \text{else } V < 0.9 \text{ and } E \geq 10 \end{cases}$$

Here E and V are defined as

$$E = \sum_{i=0}^{N} |A^2|$$

$$V = \mathrm{Std}(\mathrm{DCT}([X]))$$

Where

X = image

E = energy

V = variation

A = amplitude of frequency i

DCT = discrete cosine transform

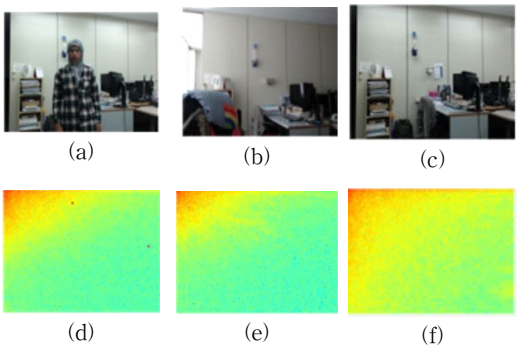
Std = standard deviation

Based on definition 1, an image is classified as a good quality image when it has high energy of more than 10, and has a standard deviation above 0.9 in the high frequency components of DCT. <Figure 2> shows some examples of blurred and sharp images. <Figure 2> (a) is a blur image created from moving object or the camera being out of focus, (b) is a blurred image caused by the camera's internal vibration, and (c) shows a sharp image without any blurriness. <Figure 2> (d), (e) and (f) are their DCT representations, showing the strength of their high and low frequency components. Using definition 1, the blur and sharp images are classified based on the statistical analysis of their high-frequency component.

The statistical analysis of the high-frequency components of the images is given in <Table 1>. It shows that, the energy and standard deviation of the high-frequency components of blurred images are fewer, as compared to the sharp image. Thus, <Figure 2> (a) and (b) are classified as blurred images and (c) as the sharp images.

〈Table 1〉 Statistical Analysis of High Frequency Component of DCT

Parameters	Blurred Image		Sharp image
	From motion	From camera internal vibration	
Energy	8.023	7.770	10.590
Standard deviation	0.325	0.291	1.061
Classification	Blurred	Blurred	Sharp



〈Figure 2〉 Sample of blurred and sharp images. (a) Blur image from motion or camera out of focus. (b) Blur from camera internal vibration. (c) Sharp image. (d)-(f) are DCT representations of (a)-(c)

2.1.2 Contrast Assessment

Next contrast assessment is performed in set of images within a short period of time to get the precise result of good self-calibration. The image with high contrast is selected among N images by the following definition 2;

Definition 2. Selection of high contrast image by contrast measurement.

$$\forall I_N(x,\ y) where I_N(x,\ y) \in I_1(x,\ y),\ I_2(x,\ y),\ I_3(x,\ y) \ldots I_n(x,\ y)$$
$$C(x,\ y) = \arg\max(C(I_N))$$

Where

$$C(I_i) = \frac{I_{\max} - I_{avg}}{I_{\max} + I_{avg}}$$

By the definition 2 for the selection of good contrast image, the image with maximum contrast is selected among the set of images in the short period of time. To get more precise result of selecting good contrast image, the limit of N images can be increased depending on precise application of self-calibration.

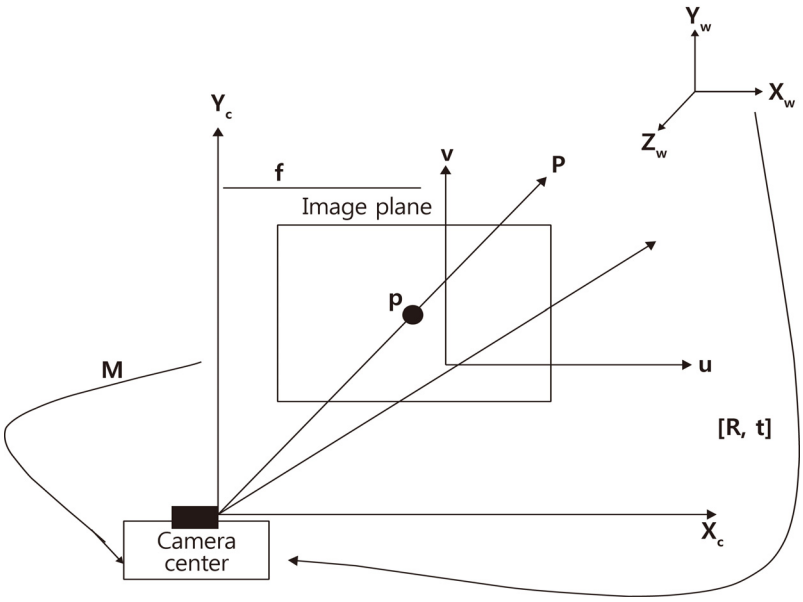
2.2 Camera Self-Calibration

After performing image quality assessment, camera self-calibration is performed automatically to formulate the scene constraints with less or no prior knowledge about the structure of scene compared to the camera calibration method. In this section, the purpose of camera self-calibration and its stages are discussed.

The purpose of camera self-calibration is to get the geometry of a pin-hole camera that is the intrinsic (M) and extrinsic parameters (R, t) in the form of Homography matrix, which maps the real world scene point P to 2D image point p on image plane as shown in <Figure 3>. <Figure 3> Pin-hole camera model.

2.2.1 SIFT Feature Detection

After the quality assessment of an image, feature detection is performed. Feature detection is the low-level image processing technique used in computer vision tasks such as tracking, image-matching, object description, recognition, and



〈Figure 3〉 Pin-hole Camera Model

camera self-calibration. In camera self-calibration. Features are the corner points in the image that are used to measure the structure of the scene.

In this paper, SIFT feature detector [Wu et al., 2013] is used which is an advanced version of Harris corner detector. SIFT features are invariant to scale, translation, rotation, and partially invariant to illumination changes. Previous approaches of features detectors lacked in-variance to scale, and the result is sensitive to projective distortion and illumination changes because the same window is used to detect the key-points with the different scale. The result of SIFT feature detection of input images in <Figure 4> is given in <Figure 5>.

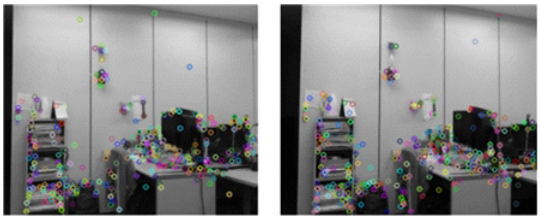
2.2.2 Feature matching

After the feature points detection, feature matching is performed by Normalized cross correlation [Zhao et al., 2006], which is the similarity measurement between two images based on window size as shown in equation (1).

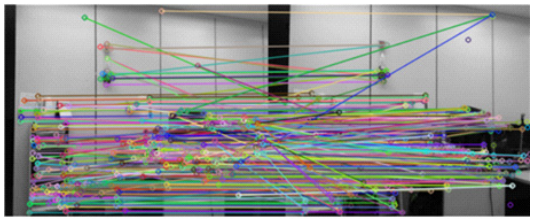
$$NCC(i, j) = \arg\max\left(\frac{\Sigma_{w(X)} - (I_1(\tilde{X}) - \overline{I_1})(I_2\tilde{X} - \overline{I_2})}{\sqrt{\Sigma_{w(X)}(I_1(\tilde{X}) - \overline{I_1})^2 (I_2(\tilde{X}) - \overline{I_2})^{2^2}}}\right) \quad (1)$$

Where $\overline{X} \in w(X_i)$, $\overline{X} \in w(X_j)$, $\overline{I_1}$, and $\overline{I_2}$ are the mean intensities of all the pixels within the window and N is the total number of these pixels and the matching points in <Figure 7> are determined by the above equation. <Figure 6> shows the image matching of one point with many points in the other image that does not produce good homography. To get good features with the minimum distance that generate good homography we use the RANSAC algorithm explained in next section.

〈Figure 4〉 Input Images



〈Figure 5〉 SIFT features Detected Image



〈Figure 6〉 SIFT Features Matching.

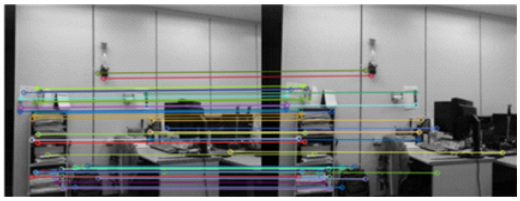
2.2.3 RANSAC Algorithm to Estimate Homography Matrix

After getting correspondence matching points, the RANSAC algorithm [Derpanis, 2010] is used to compute 2D homography by Direct Linear Transform (DLT) algorithm [Abdel-Aziz et al., 2015]. The steps of RANSAC algorithm are as follows;

1. Initialize the number of estimation N, threshold T for distance.

2. For the Ith (1 to N) estimation.
   a) Randomly choose 4 correspondence.
   b) Check whether these points are collinear, if not then redo the step 2.
   c) Compute homography H by normalized DLT from the 4 points pairs.
   d) For each correspondence pair, calculate distance d by above H.
   e) Compute the number of inliers K, which has minimum distance d < t.
   f) If K is big enough, update H as best model with all inliers.
3. Re-estimation H from all the inliers using DLT algorithm.

By RANSAC algorithm, good features shown in <Figure 7> are used to compute the homography matrix, which is used to transform the input image to perspective view that corresponds to the real world image as in <Figure 8>.



〈Figure 7〉 Good Features Matching

〈Figure 8〉 Perspective Projection of Input Image

3. Experiments and Discussion

Accurate camera self-calibration experiments based on image quality assessment has been performed to show the usefulness of the suggested method. The algorithm is implemented on Intel PC i7 in C++ and Open-CV libraries with two Logitech HD C920 web-cams. The experimental data consist of 1926 images of static and moving objects in an office environment. We initially categorized by visual inspection of 1926 images; 241 blurred images and 1685 sharp images. The images were captured using two cameras, and <Figure 9> shows samples of data set for experiment.

〈Figure 9〉 Example of Experiment Data Set

From these images, we determine high-quality images by blur detection and contrast measurement and we discard the low quality images. In the blur detection experiment of the first set of 241 blurred images, 81% (i.e., 195 images) are correctly classified, whereas the remaining 31 images were falsely classified as sharp images, see <Table 2>. Similarly 236 out of the second set of 1685 sharp images were falsely classified as blurred, achieving a discrimination accuracy of 86%, as shown in <Table 3>.

〈Table 2〉 Discrimination of Positive and False Positive Blur from Blurred Images

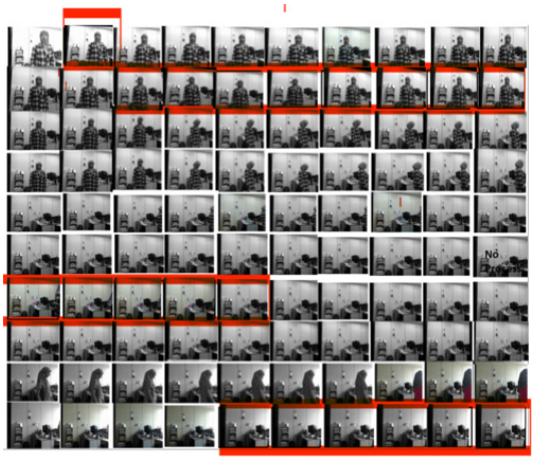
Total images	Blur images		Discrimination rate
	Positive	Falsepo-sitive	
241	195	31	81%

〈Table 3〉 Discrimination of Positive and False Positive Sharp from Sharped Images

Totalimages	Sharp images		Discriminationrate
	Positive	Falsepo-sitive	
1685	1449	236	86%

After getting the high-quality images, we performed camera self-calibration by SIFT features detectors, and generated homography matrix using RANSAC algorithm, and then we transformed the input image to the perspective projection that corresponds to the real world. <Figure 10> shows that there is little or no distortion in the image perspective views compared to conventional camera self-calibration based on SIFT features shown in <Figure 11>.

〈Figure 10〉 Example of Perspective Projection by Suggested Method



〈Figure 11〉 Example of Perspective Projection by SIFT Features

The result of experiments as shown in <Table 4>, shows that out of 1926 input images, 477 and 1301images are discarded by blur detection and contrast assessment respectively. It also shows that 93.2% (i.e., 138 images) of the remaining 148 good quality images result in good perspective views during camera self-calibration. The remaining 10 good quality images result in bad perspective view. This is due to the inaccuracies or false positives that occur at the blur detection stage; some blurred images are falsely classified as sharp images and vice versa. On the other hand, the conventional method of camera self-calibration based on SIFT feature detectors processed all 1926 input images, out of which only 148 (i.e., 7.7%) resulted in good perspective views.

The suggested method produced good perspective projection of images due to the selection of sharp images with high contrast. Another reason for getting good perspective projection is the use of SIFT features, which are invariant to rotation, translation, scaling, and illumination.

〈Table 4〉 Results of Experiments

	Suggested method	Self-calibration
Input images	1,926	1,926
Discarded images by blur	477	0
Discarded images by low contrast	1,301	0
Good perspective view	138	148
Bad perspective view	10	1,778
Self-calibration (w.r.t perspective view)	93.2%	7.7%

4. Conclusion

This paper describes a method for accurate camera self-calibration based on image quality assessment. The camera self-calibration process is performed using the Sift Feature Detection method. First, image quality assessment is perform in order to select high quality images for the camera self-calibration. The image quality assessment stage consists of two processes namely, Blur Detection and Contrast Assessment. The high quality images-sharp images with high image contrast-are selected for the camera self-calibration, discarding the blur and low contrast images. The experimental results showed very little or negligible distortion in the image perspective views during the camera self-calibration process. This confirms that the contrast and blurriness in images cause distortions in image perspective views, and that by discarding such images, it is possible to achieve a more accurate camera self-calibration. The suggested method achieved accuracy of approximately 93% in the camera self-calibration. In instances where the blur detection falsely classifies blurred images as sharp and falsely classifies sharp images as blurred, the blurred images in the camera self-calibration process resulted in bad perspective views and some sharp images were discarded.

In the future, we plan to determine the level of noise that would have negligible effect on camera self-calibration. Additionally, we plan to improve the blur detection method to partial blur detection method to detect small motions in images and improve the accuracy of camera self-calibration.

References

[1] Abdel-Aziz, Y. I., Karara, H. M., and Hauck, M., “Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry”, *Photogrammetric Engineering & Remote Sensing*, Vol. 81, No. 2, 2015, pp. 103-107.

[2] Liu, X., Su, H., Kang, S., Kane, T. D., and Shekhar, R., “Application of Single-Image Camera Calibration for Ultrasound Augmented Laparoscopic Visualization”, *Proceedings SPIE 9415, Medical Imaging 2015 : Image-Guided Procedures, Robotic Interventions, and Modeling*, 94151T, Orlando, USA, 2015 : DOI: 10.1117/12.2082194.

[3] Akkad, N. E., Merras, M., Saaidi, A., and Satori, K., “Robust method for self-calibration of cameras having the varying intrinsic parameters”, *Journal of Theoretical and Applied Information Technology*, Vol. 50, No. 1, 2013, pp. 57-67.

[4] Baataoui, A., Batteoui, I. E., Saaidi, A., and Satori, K., “Camera Self-calibration by an Equilateral Triangle”, *International Journal*

of Computer Application, Special Issue on Software Engineering, Databases and Expert Systems-SEDEX, 2012, pp. 29-34

[5] Bätz, M., Richter, T., Garbas, J. U., Papst, A., Seiler, J., and Kaup, A., "High dynamic range video reconstruction from a stereo camera setup", *Signal Processing : Image Communication*, Vol. 29, No. 2 2014, pp. 191-202.

[6] Derpanis, K. G., *Overview of the RANSAC Algorithm, Technical report Dept. of Computer Science*, York University, 2010, pp. 1-2.

[7] Erik, K., Usman, K., and Gunawan, I. P., "DCT-based local motion blur detection", *Proceedings of International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering IEEE*, Bandung, Indonesia, 2009, pp. 1-6.

[8] Farias, M. C., Mitra, S. K., and Foley, J. M., "Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance", *Proceedings of the 2003 International Conference on Multimedia and Expo*, Baltimore, USA, 2003, pp. I-529-I-532.

[9] Fayyaz, R. and Rhee, E. J., "Hand Segmentation using Depth Information and Adaptive Threshold by Histogram Analysis with Color Clustering", *Journal of Korea Multimedia Society*, Vol. 17, No. 5, 2014, pp. 547-555.

[10] Gonzales, R. C. and Woods, R. E., *Digital Image Processing* (3rd ed.), International edition, Pearson, 2008.

[11] Hemayed, E. E., "A survey of camera self-calibration", *Proceedings of the IEEE Conference in Advanced Video and Signal Based Surveillance 2003*, Miami, USA, 2003, pp. 351-357.

[12] Huang, G., Tian, Y., Wang, Y., Yang, Y., and Tai, X., "Self-Recalibration of PTZ Cameras", *Proceedings of the 2010 International Conference on Machine Vision and Human-machine Interface*, Kaifeng, China, 2010, pp. 292-295.

[13] Oliveira, J. P., Figueiredo, M. A., and Bioucas-Dias, J. M., "Parametric blur estimation for blind restoration of natural images : linear motion and out-of-focus", *IEEE Transactions on Image Processing*, Vol. 23, No. 1, 2014, pp. 466-477.

[14] Wilczkowiak, M., Boyer, E., and Sturm, P., "Camera calibration and 3D reconstruction from single images using parallelepipeds", *Proceedings of the 8th IEEE International Conference on Computer Vision ICCV 2001*, Vancouver, Canada, 2001, pp. 142-148.

[15] Wu, J., Cui, Z., Sheng, V. S., Zhao, P., Su, D., and Gong, S., "A Comparative Study of SIFT and its Variants", *Measurement Science Review*, Vol. 13, No. 3, 2013, pp. 122-131.

[16] Zhang, Z., *Camera Calibration in Computer Vision*, Boston : Springer, 2014.

[17] Zhao, F., Huang, Q., and Gao, W., "Image Matching by Normalized Cross-Correlation", *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing ICASSP 2006*, Vol. 2, Toulouse, France, 2006, pp. 729-732.

Author Profile

Rabia Fayyaz

Dr. Rabia Fayyaz received her doctoral degree in Computer Engineering from Hanbat National University, Daejeon, Korea. She is conducting research on Object Detection, tracking and recognition using stereo vision. Her research interests are in image processing, computer vision, pattern recognition and artificial intelligence.

Eun Joo Rhee

Dr. Eun Joo Rhee is a Professor of Department of Computer Engineering at College of Information Technology, Hanbat National University, Daejeon, Korea. He received his doctoral degree in Electronics Engineering from Chungnam National University. His research interests include in image processing, pattern recognition, computer vision and artificial intelligence. His papers have appeared in IEICE Trans. on Information and Systems, International Journal of Applied Engineering Research, Journal of KIISE, Journal of the Institute of Electronics and Information Engineers, Journal of Information Technology Applications and Management, Journal of Information Technology Application, Journal of the Modern Linguistic Society of Korea, Journal of Korea Multimedia Society, Journal of the Korea Academia-Industrial cooperation Society.