# FAST AFFINE-INVARIANT IMAGE MATCHING BASED ON GLOBAL BHATTACHARYYA MEASURE WITH ADAPTIVE TREE

*Jongin Son, Seungryong Kim, and Kwanghoon Sohn*

Digital Image Media Laboratory (DIML),
School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
Email: khsohn@yonsei.ac.kr

## ABSTRACT

Establishing visual correspondence is one of the most fundamental tasks in many applications of computer vision fields. In this paper we propose a robust image matching to address the affine variation problems between two images taken under different viewpoints. Unlike the conventional approach finding the correspondence with local feature matching on fully affine transformed-images, which provides many outliers with a time consuming scheme, our approach is to find only one global correspondence and then utilizes the local feature matching to estimate the most reliable inliers between two images. In order to estimate a global image correspondence very fast as varying affine transformation in affine space of reference and query images, we employ a Bhattacharyya similarity measure between two images. Furthermore, an adaptive tree with affine transformation model is employed to dramatically reduce the computational complexity. Our approach represents the satisfactory results for severe affine transformed-images while providing a very low computational time. Experimental results show that the proposed affine-invariant image matching is twice faster than the state-of-the-art methods at least, and provides better correspondence performance under viewpoint change conditions.

***Index Terms*—** Image matching, viewpoint robust, feature, ASIFT, fully affine space, Bhattacharyya distance, tree

## 1. INTRODUCTION

Establishing visual correspondence is a fundamental task in computer vision applications such as image stitching [1], localization systems [2], 3-D reconstruction [3], and so on. To estimate reliable correspondence between two images, many literatures have been tried to develop a local feature matching scheme. Generally, local feature-based image matching methods consist of three steps; the feature extraction step, the feature description step, and the feature matching step [4].

First of all, to extract a reliable key-point in an image, many local feature detectors have been proposed such as Harris, Harris-Laplace, Harris-Affine, Difference of Gaussians (DoG), Maximally Stable Extremal Region (MSER), Features from Accelerated Segment Test (FAST) and so on [5][6][7]. Secondly, to describe each key-point properly, intensity-based descriptors such as Binary Robust Independent Elementary Features (BRIEF) [8] and Binary Robust Invariant Scalable Keypoints (BRISK) [9] and order-based descriptors such as center-symmetric local binary pattern (CS-LBP) [10] and Haar-like compact local binary pattern (HC-LBP) [11] have been proposed. As a pioneer work, the Scale Invariant Feature Transforms (SIFT) proposed by Lowe [12] has been one of the most popular approaches due to its high robustness under various environments. In order to reduce computational complexity, Bay et al. proposed the Speeded-Up Robust Features (SURF) [6] which approximates to SIFT and it outperforms other existing methods. Finally, to find the nearest correspondence, a handful of Euclidean distances can be used in practice, such as normalized cross correlation (NCC) [6]. Although these conventional image matching approaches show the satisfactory performance under various environments, they have still limitations to deal with the severe distortion induced by a viewpoint variation or affine variation problem.

To alleviate these limitations, many literatures tried to develop affine-invariant feature detector or descriptor such as Hessian-affine [12], Harris-affine [13], MSER [14], and affine subspace representation (ASR) [15]. However, these approaches have also shown the limited performance on affine variations in real outdoor environments. As one of the most promising works, Morel and Yu have proposed a fully affine invariant framework, i.e., Affine-SIFT (ASIFT), for different viewpoint images matching [16]. The ASIFT simulates the reference and query image to cover the fully affine space. The local SIFT matching is then performed to extract and compare features from these simulations. Through the iterative algorithm, the geometric transformation between the image pair is estimated. Although the ASIFT has shown in reliable matching for various affine variations, it also provides dramatically many outliers and requires a high computational complexity. To overcome the problems of ASIFT, [17] proposed the iterative solver to find homography matrix of two images, which the reference image is then matched with the simulated image. In [5], local stable regions are extracted from the reference image and the query image, and transformed to circular areas according to the second-order moment. Although these methods trying to solve the affine variation problems show the performance under small viewpoint change, they also still have limitations under challenging viewpoint variations and also require high computational complexities.

In this paper, we propose an affine invariant image matching scheme to solve the viewpoint variation problems between reference and query image while providing a low computational complexity. Unlike the conventional ASIFT method which require the extensive local feature matching on each fully affine transformed-images, our approach first find the optimal global affine transformation matching and then estimate the reliable inliers on optimal transformed-images. The optimal affine transformation matrix of two images is found very fast using the Bhattacharyya similarity measure without the loss of information of the sub-sampling. In order to reduce the complexity during the process of estimating the affine transformation matrix, the global matching based on Bhattacharyya distance is conducted with the adaptive tree scheme. Our approach enables one to estimate the reliable affine transformation model with very low

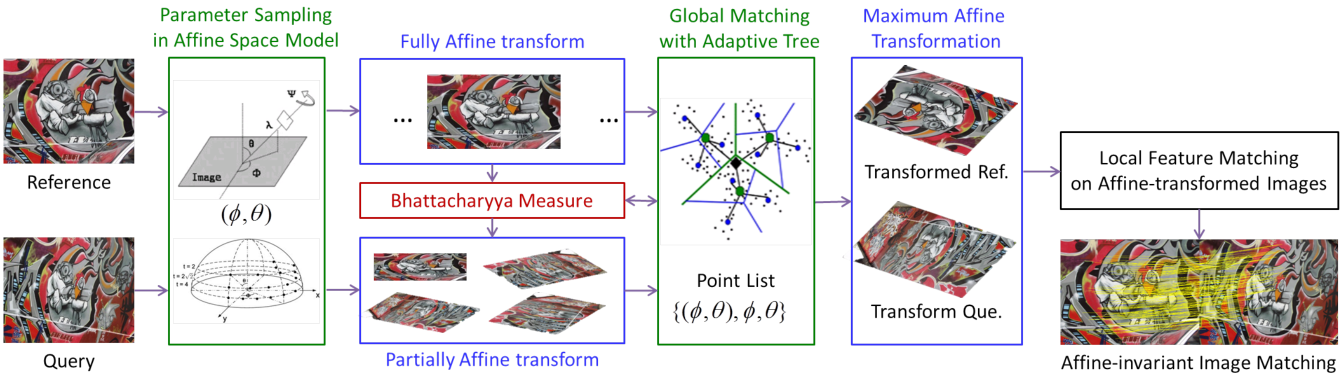



**Fig. 1**. The block diagram of the proposed scheme for affine-invariant image matching.

computational time. In order to evaluate our proposed approach, we tested on four databases taken under varying affine transformation. Experimental results show that our approach provides the reliable performance even under challenging affine variations.

The remainder of this paper is organized as follows. In Sec. 2, we briefly introduce the affine space model. In Sec. 3, we propose a fast affine-invariant image matching to solve viewpoint variation problems. Sec. 4 describes the experimental results for viewpoint variation conditions, followed by the conclusion in Sec. 5.

## 2. REVIEW OF AFFINE TRANSFORMATION

Let an image be $f : \mathcal{I} \to \mathbb{R}$ or $\mathbb{R}^3$, where $\mathcal{I} = \{i = (x_i, y_i)\} \subset \mathbb{N}^2$ is a discrete image domain. Assume that the image $f$ is transformed by any affine transformation matrix $A$, which is the representative of a viewpoint change and can be characterized as follows:

$$\begin{aligned} A &= H_\lambda R_1(\psi) T_t R_2(\phi) \\ &= \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, \end{aligned} \tag{1}$$

where $\phi$ and $t$ are the latitude angles of the camera optical axis and transition tilt, respectively. The $\psi$ angle is the camera spin, and $\lambda$ represents the zoom parameter. Since the tilt can be represented as $t = 1/\cos\theta$ for $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, the affine transformation model is controlled by two variations including $\theta$ and $\phi$ [16]. More detail definition was reviewed in [16]. This affine transformation matrix induces the coordinate of original image as

$$\mathcal{I}' = A \cdot \mathcal{I}. \tag{2}$$

With this relationship, an affine transformed image $f'$ from image $f$ is computed by affine matrix $A$ bas $f' : \mathcal{I}' \to \mathbb{R}$ or $\mathbb{R}^3$.

## 3. FAST AFFINE-INVARIANT IMAGE MATCHING

In this section, we propose robust and fast affine-invariant image matching model based on the global correspondence with Bhattacharyya similarity measure and adaptive tree.

### 3.1. Problem Statements and Overview

As a pioneer work for affine-invariant image matching, the ASIFT approach defines all fully affine spaces of image in the latitude $\theta$ and longitude $\phi$. It applies all affine matrixes through the sampling scheme of latitude and longitude to the reference and query images. By matching all applied images with local feature correspondence, e.g., SIFT, it provide robust image matching for affine variation. However, a sub-sampling scheme used to reduce the computational complexity induces the loss of information. Furthermore, it requires too many unnecessary transformation and local matching, which induces the number of outliers rapidly increased. To tackle these problems, our approach aims to find best viewpoint parameter between two images (reference image $f_r$ and query image $f_q$) through the global matching in the re-sampling between reference and query affine space. It enables us to provide affine invariant image matching between reference and query images.

Fig. 1. shows the overall block diagram of the proposed affine-invariant image matching scheme. First of all, our approach transforms reference image into fully affine space and a query image into partial affine space. The Bhattacharyya coefficient is extracted from each image for the comparison of two images transformed in the virtual affine space. Furthermore, the adaptive tree is organized following the Bhattacharyya distance between images gradationally transformed in the affine space. According to the organized tree, two images find the best affine matrix which has the closest distance. Finally, we estimatiate the local feature correspondence on affine-transformed images.

### 3.2. Global Image Matching with Adaptive Tree

In this section, we introduce fast global matching scheme between virtually generated images and adaptive tree for computational efficiency. As in Sec. 2, affine transformation for reference and query images can be expressed as follows:

$$A_i = H_\lambda R(\psi_i) T_{ti} R(\phi_i), \tag{3}$$

where $i \in \{r, q\}$, and $r$ and $q$ denote reference and query image, respectively.

First of all, the latitude $\theta$ is sampled in geometric progression form (e.g., $1, a, a^2, ., a^n, a > 1$), which induces corresponding tilt as $t = 1/\cos\theta$. The interval of sampling is established as $a = \sqrt{2}, kb/t < 180$, $t_{\max} \approx \sqrt[4]{2}$, $\Delta\phi = 72^\circ$ in light of the exactness and efficiency of creating a viewpoint in three dimensional coordinates. In ASIFT, the longitude $\phi$ is sampled in arithmetic progression form (e.g., $0, b/t, ., kb/t, b \simeq 72$). However, unlike conventional approach such as the ASIFT [16], our approach does not perform the fully affine transformation to reduce the computational complexity and outliers. In our approach, the longitude of reference and query images are differently defined as

$$\begin{cases} \phi_i \in \{0, b/t, ., kb/t, b \simeq 72\}, \\ \phi_j \in \{b/t,\ b = C\}, \end{cases} \tag{4}$$

where $\phi_i$ and $\phi_j$ are reference and query image of longitude. Although the reference image performs the fully affine transformation as $f_r'(t_i, \phi_i)$, the query image is transformed by stages to decide whether to transform or match depending on the matched global matching as in as $f_q'(t_j, \phi_j)$.

In case of query image, the tree is organized as performing the transform and matching from latitude $\theta$ and longitude $\phi$ defined in the affine space model. When organizing the tree, it is divided into two seeds within the range stage. Tree cost is decided by the Bhattacharyya distance of between reference and query images which will be described in Sec. 3.3. Our adaptive tree has following criteria

$$\mathcal{D}_{BH}(f_r'(t_i, \phi_i), f_q'(t_j, \phi_j)) > \mathcal{D}_{BH}(f_r'(t_i, \phi_i), f_q'(t_{j+1}, \phi_{j+1})). \quad (5)$$

That is, as Eq. (5), the parent node in an adaptive tree determines the child node. For example, it has the same the latitude $\theta$ of the tilt stage ($t = \sqrt{2}$, $\phi \in \{0, b/t, 2b/t, 3b/t\}$). At this time, global matching is performed only fixed ranges of longitude $\phi$ when being able to divide the left node and right node and two seeds are provided in the same way as the next tilt stage ($t = 2$, $\phi \in \{0, 2b/t\}\, or\, \{3b/t, 4b/t\}$).

By using this adaptive tree, we efficiently estimate optimal affine parameter of reference and query image as

$$\begin{aligned}(t_r^*, \phi_r^*, t_q^*, \phi_q^*) \\ = argmin_{(t_i,\phi_i,t_j,\phi_j)} \mathcal{D}_{BH}(f_r'(t_i, \phi_i), f_q'(t_j, \phi_j)).\end{aligned} \quad (6)$$

where $(t_r^*, \phi_r^*)$ and $(t_q^*, \phi_q^*)$ is the optimal affine parameters of reference and query image, respectively.

Finally, after transforming reference and query images by using the optimal affine parameters $(t*_r, \phi*_r, t*_q, \phi*_q)$, the local feature correspondence scheme is employed to estimate the affine-invariant correspondence between reference and query images.

### 3.3. Bhattacharyya Similarity Measure

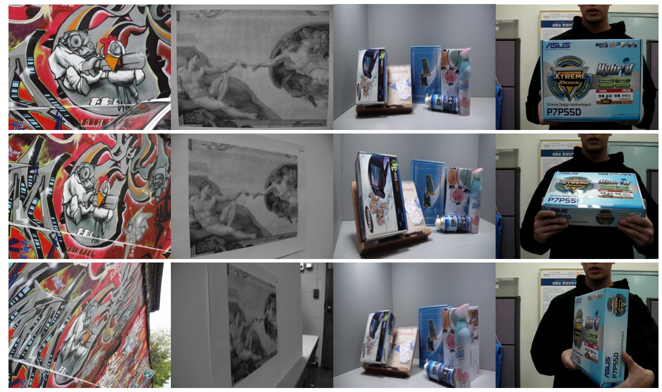
Our approach requires fast global image matching between affine transformed images. In order to estimate a global image distance very fast, we employ the Bhattacharyya distances. Its coefficient is an approximate measurement of the amount of overlap between two statistical samples [18], [19]. The coefficient can be used to determine the relative closeness of the two samples being considered. In order to measure distance simply, we used a 5D histogram of the $Labxy$ values in the image $f_i$ for $i \in \{r, q\}$ as $\mathcal{H}(f_i) = hist(f_i)$, where the space was quantized into $[10, 10, 10, 2, 2]$ bins. For two 5D histograms $\mathcal{H}(f_r)$ and $\mathcal{H}(f_q)$, are calculated as follows:

$$\mathcal{D}_{BH}(f_r, f_q) = \left(1 - \left(\mathcal{H}(f_r)^T \mathcal{H}(f_q)\right)^{1/2}\right)^{1/2} \quad (7)$$

where $\mathcal{D}_{bh}$ means Bhattacharya distance which measures the similarity of between two images. Compared to conventional histogram intersection or Euclidean distance, it provides fast and reliable global image distance [18].

### 3.4. Computational complexity analysis

To analyze the performance in terms of computational complexity, we define the computational time of one SIFT matching as arithmetic 1 for the sake of simplicity. When using the fully affine sampling method designated in ASIFT, complexity of fully transformations and low-resolution SIFT matching are performed for the reference and query image as $\left((1 + (|T_n| - 1)\frac{180}{72})/(K \times K)\right)^2 = 1.5$, where $|T_n|$ means simulated tilt as $\left|\{1, \sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}\}\right|$. $K \times K = 3 \times 3$ is the subsampling factor. The complexity of ASIFT algorithm is therefore $(1.5)^2 = 2.25$ times as much as that of SIFT. However, the ASIFT increases number of feature points and matching dramatically. Therefore, the actually processing time much larger 5 is consumed compared to SIFT. Furthermore, the ISIFT consist of initial local matching for homography estimation, similarity checking, and local matching process. Accordingly, the processing time of ISIFT more than 2 or 3 times is consumed compared to SIFT at least. In case of our method, it does not perform the fully affine transformation for both reference and query images due to using adaptive tree. Also, local matching which has high complexity is replaced by global matching effectively. Proposed algorithm effectively removed and replaced by a part of the high complexity and the computational complexity minimized. Therefore, the processing time of our method consumed 1.2 times compared to SIFT, approximately.




(a) Graffiti (b) Adam (c) Booth (d) Box

**Fig. 2**. Some samples of the test data sets taken under affine deformation. The first rows are samples of reference images, and the second and third rows are samples of query images.

**Table 1**. Descriptions of the test data sets in experiments.

| Datasets | | Resolution | Total frames |
|---|---|---|---|
| Morel [20] | Adam | 600 x 450 | 9 |
| Mikolajczyk [21] | Graffiti | 800 x 600 | 6 |
| DIML [22] | Booth | 320 x 240 | 330 |
| | Box | 640 x 480 | 335 |

## 4. EXPERIMENTS

### 4.1. Experimental Environments

To evaluate the proposed method compared to state-of-the-art methods on affine variations, we build our database taken under varying viewpoint conditions [22] and use public affine variation database [20], [21]. Our viewpoint variation image database [22] is made by rotating objects leftward, center-ward, and upward as shown in Fig. 2. Table 1 shows the description of the affine transformation databases used in our experiments in detail. In order to evaluate the performance of the proposed affine image matching, the following criteria were used to quantify the evaluation results; i) Number of total correct match pairs (NCM), ii) Matching precision (MP) is the radio between the NCM and the number of matches (NCM/Matches) [17]. The higher the MP and NCM value, the more accurate image

**Table 2**. Average matching rates.

| | Adam | | Graffiti | |
|---|---|---|---|---|
| Methods | NCM | MP (%) | NCM | MP (%) |
| SIFT | 11 | 25 | 6 | 25 |
| ASIFT | 332 | 52 | 451 | 42 |
| Harris-affine+ISIFT | 38 | 76 | 18 | 56 |
| MSER+ISIFT | 31 | 54 | 17 | 50 |
| **Proposed method** | 32 | **79** | 27 | **70** |

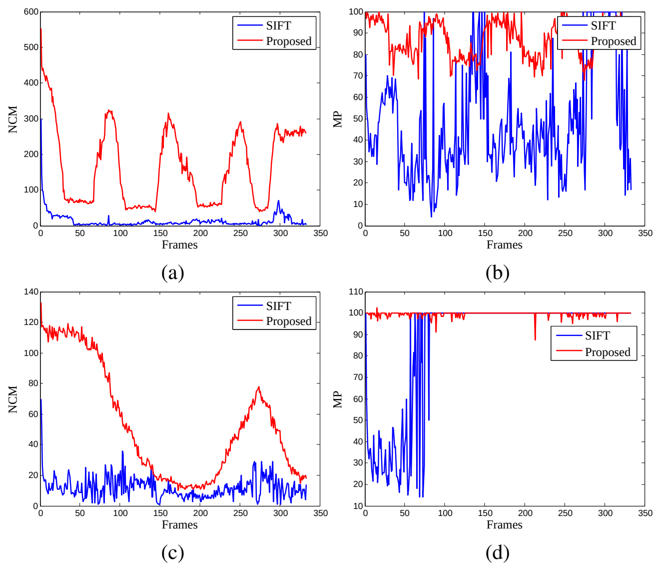


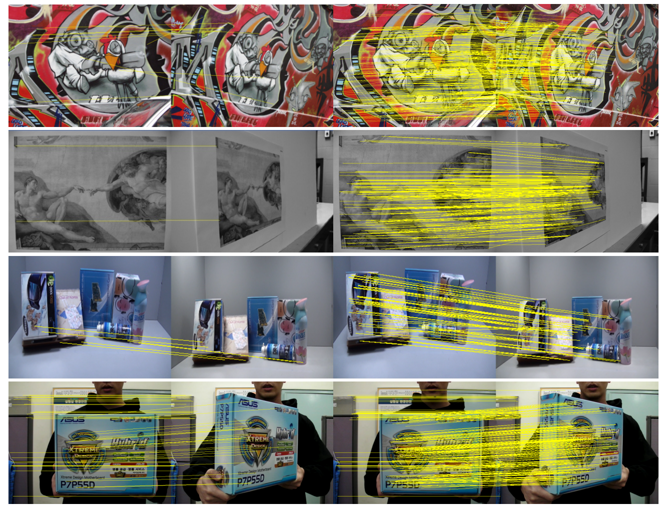

**Fig. 3**. Matching results in DIML video frames. (a) and (b) are NCM and MP of SIFT and proposed algorithm in booth set, and (c) and (d) are NCM and MP of SIFT and proposed algorithm in box set.

matching performance. However, if NCM value is high and MP is low, it means poor reliability. We implemented our approach in C++ on Intel Core i7-4790k CPU at 4.00 GHz, and measured the runtime on a single CPU core without further code optimizations and parallel implementation using multi-core CPUs/GPU.

### 4.2. Experimental Results

In experiments, SIFT [12], Affine-SIFT [16], ISIFT [17], and the proposed approach were evaluated. Furthermore, the Harris-affine [13] and MSER [14] were used for finding the initial homography matrix used in ISIFT [17]. The parameter values of the algorithm used were set as the same ideal value. Table 2 shows that the proposed method has the best matching rate for Adam and Graffiti database, which is better than Harris-Affine + ISIFT. As described in the previous section, the ASIFT provides too many outliers to decrease the reliable matching ratio as in MP. However, our proposed method has shown the satisfactory NCM and MP performance, which means it provides the most reliable matching.

Fig. 3. shows the evaluation on MP and NCM of our approaches and SIFT for box and booth database. The SIFT shows that it has small NCM and uncertain MP under viewpoint change conditions. However, similar to previous experiments, the proposed algorithm shows satisfactory NCM and MP under challenging viewpoint change conditions. Fig. 4 shows qualitative evaluation on image matching results with the proposed algorithm for several



(a) SIFT (b) Proposed

**Fig. 4**. Affine-invariant image matching results for SIFT and proposed method. (from top to bottom) Adam, Graffiti, Booth, and Box.

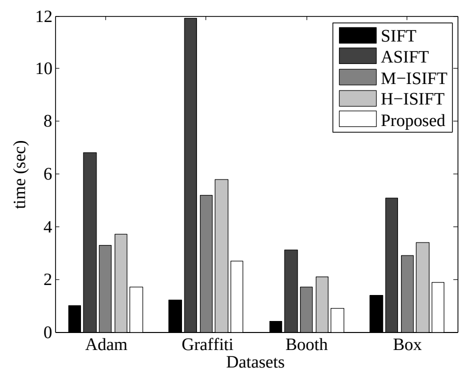



**Fig. 5**. Comparison of complexity using SIFT, ASIFT, H-ISIFT, M-ISIFT and proposed algorithm.

databases. As shown in Fig. 5, the proposed approach reduces the computational complexity dramatically compared to the conventional image matching except for SIFT. The video results and data sets for affine-invariant image matching can be found in [22].

## 5. CONCLUSION

Conventional image correspondence algorithms cannot be applicable to the real outdoor environment where several viewpoint variations occur frequently. To alleviate these problems, we have proposed the affine-invariant image matching to solve the viewpoint problems while providing a low computational complexity. Our approach has tried to find optimal one global correspondence and utilize the local feature matching to estimate the most reliable inliers between two images. We further optimize our approach with the Bhattacharyya similarity measure and adaptive tree in order to improve the complexity performance of the image matching. Experimental results show that our proposed algorithm improves the matching performance for the viewpoint changing conditions with four times faster than the conventional algorithms. In further works, our approach will be reformulated to solve more challenging affine variations.

## 6. REFERENCES


[1] J. Hays, A. Efros, “im2gps: estimating geographic information from a single image,” IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[2] J. Wang, H. Zha, and R. Copolla, “Coarse-to-Fine vision-based localization by indexing scale-invariant features,” IEEE Trans. Syst Man Cybern B Cybern, Vol. 36, no. 2, pp. 413-422, 2006.

[3] M. Ahmed, M. Dailey, J. Landabaso and N. Herrero, “Robust key frame extraction for 3D reconstruction from video streams,” International Conference on Computer Vision Theory and Applications, pp. 231-236 2010.

[4] B. Kim, B. Ham, B. Kim and K. Sohn, “Mahalanobis Distance Cross-Correlation for Illumination Invariant Stereo Matching,” IEEE Trans. Circuits and Systems for Video Technology, vol. 24, no. 11, pp. 1844-1859, 2014.

[5] M. Chen, Z. Shao,D. Li and J. Liu, “Invariant matching method for different viewpoint angle images,” Applied optics, vol. 52, no. 1, pp. 96-104, 2013.

[6] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, “Speeded-Up Robust Features (SURF),” Computer Vision and Image Understanding, vol. 110, no.3, pp.346-359, 2008.

[7] E. Rosten, R. Porter and T. Drummond, “Faster and better : a machine learning approach to corner detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pp. 105-119, 2010.

[8] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha and P. Calonder, “BRIEF : Computing a local binary descriptor very fast,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34, no. 7, pp. 1281-1298, 2011.

[9] S. Leutenegger, M. Chli and R. Y.Siegwart, “BRISK : Binary Robust Invariant Scalable Keypoints,” in Proc. IEEE Conf. International Conference on Computer Vision, 2011.

[10] Guo, Zhenhua, and David Zhang. “A completed modeling of local binary pattern operator for texture classification.” Image Processing, IEEE Transactions on 19.6 (2010): 1657-1663.

[11] Bongjoe Kim, Jihoon Choi, Sanghyun Joo, Kwanghoon Sohn, “Haar-like compact local binary pattern for illumination robust feature matching,” Journal of Electronic Imaging 21 (4) 2012.

[12] K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors”. International Journal of Computer Vision, Vol. 60, No. 1, pp.63-86, 2004.

[13] Mikolajczyk, Krystian, and Cordelia Schmid. “A performance evaluation of local descriptors.” Pattern Analysis and Machine Intelligence, IEEE Transactions on 27.10 pp. 1615-1630, 2005.

[14] Matas, J., Chum, O., Urban, M., Stereo, T: “Robust wide baseline stereo from maximally stable extremal regions.” Proc. of BMVC pp. 414-432. 2002.

[15] Wang, Zhenhua, Bin Fan, and Fuchao Wu. “Affine Subspace Representation for Feature Description.” Computer Vision-ECCV 2014. Springer International Publishing, 94-108, 2014.

[16] Y. Guoshen and J. M. Morel, “ASIFT: A New Framework for Fully Affine Invariant Image Comparison,” SIAM Journal on Imaging Sciences, Vol. 2, No. 2, pp.438-469. 2009.

[17] Y. Yu, K. Huang, W. Chen, and T. Tan, “A novel algorithm for view and illumination invariant image matching,” IEEE Trans. Image Processing 21, 229-240 2012.

[18] Khalid, M. Sohail, et al. “Bhattacharyya coefficient in correlation of gray-scale objects.” Journal of Multimedia 1.1 pp. 56-61, 2006.

[19] D. Comaniciu, V. Ramesh, and P.Meer. “Real-time tracking of non-rigid objects using mean shift,” IEEE international conference on Computer vision and pattern recognition, 2000.

[20] http://www.robots.ox.ac.uk/ vgg/research/affine

[21] http://www.cmap.polytechnique.fr/ yu/research/ASIFT/demo.html

[22] http://diml.yonsei.ac.kr/jison/image matching/